

Wesley Pereira da Silva

Exames Inteligentes: Evidenciação Estatística de Perfis de Aprendizagem, Composição de Banco de Itens Multidimensionais e Personalização de Avaliação

Vitória, ES

2023

Wesley Pereira da Silva

Exames Inteligentes: Evidenciação Estatística de Perfis de Aprendizagem, Composição de Banco de Itens Multidimensionais e Personalização de Avaliação

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Doutor em Informática.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Programa de Pós-Graduação em Informática

Orientador: Dr Elias Oliveira

Vitória, ES

2023

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

S586e Silva, Wesley, 1982-
Exames Inteligentes: : Evidenciação Estatística de Perfis de Aprendizagem, Composição de Banco de Itens Multidimensionais e Personalização de Avaliação / Wesley Silva. - 2023.
200 f. : il.

Orientador: Elias Oliveira.
Tese (Doutorado em Ciência da Computação) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Análise de Aprendizagem. 2. Teoria de Resposta ao Item. 3. Testes Adaptativos Computadorizados. 4. Aprendizagem de Máquina. 5. Processamento de Linguagem Natural. I. Oliveira, Elias. II. Universidade Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 004



Exames Inteligentes: Evidenciação Estatística de Perfis de Aprendizagem, Composição de Banco de Itens Multidimensionais e Personalização de Avaliação

Wesley Pereira da Silva

Tese de Doutorado submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

Aprovada em 19 de Maio de 2023.

Prof. Dr. Elias Silva de Oliveira
Orientador, participação remota

Prof. Dra. Claudine Santos Badue
Membro Interno, participação remota

Prof. Dr. Thiago Oliveira dos Santos
Membro Interno, participação remota

Prof. Dr. Caio Lucidius Naberezny Azevedo
Membro Externo, participação remota

Prof. Dr. Jorge Luis Bazán Guzmán
Membro Externo, participação remota

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória/ES, 19 de Maio de 2023



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
ELIAS SILVA DE OLIVEIRA - SIAPE 6270636
Departamento de Arquivologia - DAr/CCJE
Em 31/05/2023 às 08:50

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/721439?tipoArquivo=O>



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
THIAGO OLIVEIRA DOS SANTOS - SIAPE 2023810
Departamento de Informática - DI/CT
Em 31/05/2023 às 10:30

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/721543?tipoArquivo=O>



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROTOCOLO DE ASSINATURA



O documento acima foi assinado digitalmente com senha eletrônica através do Protocolo Web, conforme Portaria UFES nº 1.269 de 30/08/2018, por
CLAUDINE SANTOS BADUE - SIAPE 1729561
Departamento de Informática - DI/CT
Em 01/06/2023 às 16:59

Para verificar as assinaturas e visualizar o documento original acesse o link:
<https://api.lepisma.ufes.br/arquivos-assinados/722812?tipoArquivo=O>

Dedico esta Tese à minha família, amigos e colegas de trabalho, pelo apoio, compreensão e incentivo, sem os quais não consigo imaginar como seria a superação desta etapa.

Agradecimentos

Expresso aqui meus sinceros agradecimentos ao Instituto Federal do Espírito Santo, em especial, aos servidores da Tecnologia da Informação e da Reitoria que apoiam e estimulam o desenvolvimento profissional.

Igualmente, ao Programa de Pós Graduação em Informática da UFES, e seus professores, em especial, meus orientadores Dra Cluadine Badue e Dr Elias Silva de Oliveira, que sempre foram mais que colaborativos.

Bem como, à Dra Mariana Curi e Dr Jean-Rémi Bourguet, pelo companheirismo e aprendizado que me proporcionaram.

Por fim, aos amigos da UFES e demais locais de atuação, que eu tive o privilégio de conhecer ao longo da jornada e que cada um à sua maneira, me apoiaram na conclusão dessa etapa tão importante.

*“Urge preveni-los do muito que se poderia fazer, com apoio no saber científico, e do
descalabro e pequenez do que se está fazendo.”*
(Darcy Ribeiro)

Resumo

Um desafio comum às áreas de conhecimento é a construção de estratégias de ensino suficientemente generalistas a ponto de se adequar a públicos com perfis de aprendizagem variados. Habitualmente, o docente baliza seu plano de ensino em função de objetivos de aprendizagem, os quais são avaliados por meio da expressão de traços latentes que denotem proficiência por parte do sujeito examinado. Assim, são oportunas técnicas inteligentes e ferramentas tecnológicas que contribuam para o incremento da qualidade do ensino e redução do esforço docente na execução de atividades complexas como, por exemplo: formulação de itens de avaliação, aplicações de provas e provimento de *feedback* aos alunos. O rigor formal na criação de instrumentos de avaliação, tabulação e cálculo das notas, é fator chave para evitar viés na condução da avaliação da aprendizagem e estimação da capacidade do educandos. A performance do aluno é a primeira dimensão a ser considerada na avaliação. O agrupamento de performances similares permite caracterizar grupos que representem perfis de aprendizagem. Autoavaliação e avaliação em pares são técnicas para estimular a autocritica do aluno em relação a si mesmo a aos colegas de turma, buscando desestimular tendenciosidades avaliativas por ensejar coerência do examinando ao exercer o papel de avaliador. Os modelos oriundos da Psicometria possibilitam a caracterização quantitativa dos itens de avaliação, permitindo a mensuração de aspectos qualitativos, como por exemplo: dificuldade, discriminação e propensão a palpite. Com eles, pode-se prever a probabilidade de sucesso do sujeito ao ser avaliado com determinado item. Por fim, o uso de Processamento de Linguagem Natural proporciona a seleção de itens por similaridade do conteúdo com uma expressão de busca, que represente um assunto a ser recuperado num conjunto do banco de itens do teste. Desse modo, buscamos propor um método de criação de trilha de avaliação individualizado, composto por uma sequência de atividades numa determinada ordem adequada à capacidade do examinando. Assim, apresentamos abordagem de teste adaptativo computadorizado inteligente, cuja configuração de execução é ajustável à estratégias de ensino qualitativas, quantitativas e/ou por conteúdo relacionados a termos pré-definidos. A contribuição vislumbrada com tal proposta é de extrapolar espaço de parâmetros bidimensional das avaliações, compostas por performances dos examinandos e pontuações alcançadas por item; para um espaço multidimensional que considere as características dos itens em termos psicométricos e semânticos, bem como, as características dos examinandos e dados históricos de sujeitos com trajetórias similares.

Palavras-chaves: Teoria de Resposta ao Item. Análise de Aprendizagem. Testes Adaptativos Computadorizados. Predição de Performance.

Abstract

A common challenge to the areas of knowledge is the construction of teaching strategies that are sufficiently general to suit audiences with varied learning profiles. Usually, the teacher defines his teaching plan according to learning objectives, which are evaluated through the expression of latent traits that denote proficiency on the part of the subject being examined. Thus, intelligent techniques and technological tools are opportune to contribute to increasing the quality of teaching and reducing the teaching effort in the execution of complex activities such as, for example: formulation of assessment items, application of tests and provision of *feedback* to students. The formal rigor in the creation of instruments for assessment, tabulation and calculation of grades is a key factor to avoid bias in conducting the assessment of learning and estimating the ability of students. Student performance is the first dimension to be considered in the assessment. The grouping of similar performances allows characterizing groups that represent learning profiles. Self-assessment and peer assessment are techniques to stimulate the student's self-criticism in relation to himself and his classmates, seeking to discourage evaluative biases by encouraging the examinee's coherence when exercising the role of evaluator. The logistic models derived from Psychometrics allow the quantitative characterization of the evaluation items, allowing the measurement of qualitative aspects, such as: difficulty, discrimination and propensity to kick. With psychometric models, the probability of success of the subject can be predicted when being evaluated with a certain item. Finally, the use of Natural Language Processing provides the selection of items by content similarity with a search expression, which represents a subject to be retrieved in a set of the test items bank. In this way, we seek to propose a method of creating individualized assessment trail, composed of a sequence of activities in a certain order appropriate to the ability of the examinee. Thus, we present an intelligent computerized adaptive test approach, whose execution configuration is adjustable to qualitative, quantitative and/or content teaching strategies related to pre-defined terms. The contribution envisaged with such a proposal is to extrapolate a two-dimensional parameter space of the evaluations, composed of the examinee's performances and scores achieved by item; for a multidimensional space that considers the characteristics of the items in psychometric and semantic terms, as well as the characteristics of the examinees and historical data of subjects with similar trajectories.

Keywords: Item Response Theory. Learning Analytics. Computer Adaptive Test. Performance Prediction.

Lista de ilustrações

Figura 1 – Sequência de Objetivos de Aprendizagem. Fonte: Elaborado pelo Autor.	28
Figura 2 – Estrutura da Pesquisa. Fonte: Elaborado pelo Autor.	30
Figura 3 – Características dos Grupos.	38
Figura 4 – Evolução Educacional por Ano.	39
Figura 5 – Exame Nacional do Ensino Médio.	41
Figura 6 – Distribuição das Notas Objetivas — CN e CH	42
Figura 7 – Distribuição das Notas Objetivas - LC e MT	43
Figura 8 – Estrutura Microdados x Tendência Objetivas	44
Figura 9 – Médias Nacionais das Notas de Redação por Estado	48
Figura 10 – Domínio Cognitivo. Fonte: Adaptado pelo Autor de (AMORIM, 2018).	62
Figura 11 – Domínio Psico-afetivo. Fonte: Adaptado pelo Autor de (AMORIM, 2018).	64
Figura 12 – Domínio Psicomotor. Fonte: Adaptado pelo Autor de (AMORIM, 2018).	65
Figura 13 – Fluxo de Informação do Experimento. Fonte: (SILVA et al., 2020a)	70
Figura 14 –	72
Figura 15 – Caracterização pelas Correlações e Distâncias	74
Figura 16 – Caracterização Pelas Distâncias	76
Figura 17 – Métricas Mescladas. Fonte: elaborado pelo autor.	77
Figura 18 – Regressões do Primeiro e Segundo Colocados	78
Figura 19 – Regressões do Terceiro e Quarto Colocados	79
Figura 20 – Esquemas Ilustrativos	82
Figura 21 – Conjunto de Dados Artificiais. Fonte: elaborado pelo autor.	86
Figura 22 – Frequência da Média de Acertos. Fonte: elaborado pelo autor.	87
Figura 23 – Curvas de Características	90
Figura 24 – Fluxo do Sistema de Recomendação. Fonte: elaborado pelo autor.	92
Figura 25 – Resíduos Habilidades Estimada Modelo Rasch e Média de Acerto Escalada. Fonte: elaborado pelo autor.	93
Figura 26 – Resíduos Habilidades Estimada Modelo 3PL e Média de Acerto Escalada. Fonte: elaborado pelo autor.	94
Figura 27 – Análise da Heurística do Melhor Número de Grupos. Fonte: elaborado pelo autor.	95
Figura 28 – Desempenho por Cluster. Fonte: Fonte: elaborado pelo autor.	96
Figura 29 – Desempenho por Cluster. Fonte: Fonte: elaborado pelo autor.	97
Figura 30 – Funções de Características e Funções de Informação	98
Figura 31 – Funções de Características e Funções de Informação	99
Figura 32 – Análise de Agrupamento dos Coeficientes da TRI	100
Figura 33 – Coeficientes TRI nos Grupos 1 e 2	101

Figura 34 – Coeficientes TRI nos Grupos 3 e 4	102
Figura 35 – Árvore de Decisão. Fonte: elaborado pelo autor.	105
Figura 36 – Benchmark com Aprendizagem de Máquina. Fonte: elaborado pelo autor.	106
Figura 37 – Curvas de Características do Modelo de Resposta Gradual	112
Figura 38 – Fluxo de Informação do Experimento. Fonte: (SILVA et al., 2021b)	113
Figura 39 – Distribuição de Densidades por Cluster: Métodos Inter-Quartil e K-Means. Fonte: (SILVA et al., 2021b)	115
Figura 40 – Análise de Agrupamento e Correlações das Notas. Fonte: (SILVA et al., 2021b)	116
Figura 41 – Modelo Samejima n=500. Fonte: (SILVA et al., 2021b).	118
Figura 42 – Modelo Samejima n=100k. Fonte: (SILVA et al., 2021b).	119
Figura 43 – Distribuição de Densidades. Fonte: (SILVA et al., 2021b).	120
Figura 44 – Fluxograma de um CAT. Fonte: Elaborado pelo Autor.	124
Figura 45 – Esquema de Funcionamento do Teste Adaptativo	129
Figura 46 – Capas das Provas de 2019	130
Figura 47 – Dados Faltantes — Participante & Instituição	131
Figura 48 – Dados Faltantes — Objetivas & Redação	132
Figura 49 – Frequência de Eixos Cognitivos — LC e MT	133
Figura 50 – Frequência de Eixos Cognitivos — CN e CH	134
Figura 51 – Parâmetros TRI das Provas de LC e MT de 2019	136
Figura 52 – Parâmetros TRI das Provas de CN e CH de 2019	137
Figura 53 – Probabilidade de Sucesso e Informação dos Itens — LC	138
Figura 54 – Classificação Textual das Competências dos Itens de Linguagem e Comunicação	139
Figura 55 – Itens da Prova de LC	140
Figura 56 – Probabilidade de Sucesso e Informação dos Itens - MT	141
Figura 57 – Classificação Textual das Referências dos Itens de MT	142
Figura 58 – Itens da Prova de MT	143
Figura 59 – Probabilidade de Sucesso e Informação dos Itens — CN	144
Figura 60 – Classificação Textual das Referências dos Itens de CN	145
Figura 61 – Itens da Prova de CN	146
Figura 62 – Probabilidade de Sucesso e Informação dos Itens — CH	147
Figura 63 – Classificação Textual das Competências dos Itens de CH	148
Figura 64 – Itens da Prova de CH	149
Figura 65 – Distribuições das Habilidades de LC e MT	150
Figura 66 – Distribuições das Habilidades — CN e CH	151
Figura 67 – Distribuições Conjuntas — LC e MT	152
Figura 68 – Distribuições Conjuntas — CN e CH	153
Figura 69 – Gráfico de Resíduos das Habilidades — LC e MT	154

Figura 70 – Gráfico de Resíduos das Habilidades — CN e CH	155
Figura 71 – Pares Ordenados e Densidades dos Parâmetros de MT	156
Figura 72 – Simulação do CAT	157
Figura 73 – Geral e Por Agrupamento	158
Figura 74 – Por Tamanho do Teste e Variação da Habilidade	159
Figura 75 – Mapas de Contraste da Redação	177
Figura 76 – Mapas de Contraste das Objetivas Interdisciplinares	178
Figura 77 – Mapas de Contraste das Objetivas Interdisciplinares	179
Figura 78 – Mapas de Contraste das Objetivas Interdisciplinares	180
Figura 79 – Mapas de Contraste das Objetivas Interdisciplinares	181
Figura 80 – Mapas de Contraste das Objetivas Interdisciplinares	182
Figura 81 – Região Sudeste - Redação e Provas Interdisciplinares	183
Figura 82 – Região Sudeste - Linguagem-Comunicação e Matemática	184
Figura 83 – Região Sudeste - Ciências Naturais e Ciências Humanas	185
Figura 84 – Região Sul - Redação e Interdisciplinares	186
Figura 85 – Região Sul - Linguagem-Comunicação e Matemática	187
Figura 86 – Região Sul - Ciências Naturais e Ciências Humanas	188
Figura 87 – Região Nordeste - Redação e Interdisciplinares	189
Figura 88 – Região Nordeste - Linguagem-Comunicação e Matemática	190
Figura 89 – Região Nordeste - Ciências Naturais e Ciências Humanas	191
Figura 90 – Região Cento-Oeste - Redação e Interdisciplinares	192
Figura 91 – Região Cento-Oeste - Linguagem-Comunicação e Matemática	193
Figura 92 – Região Cento-Oeste - Ciências Naturais e Ciências Humanas	194
Figura 93 – Região Norte - Redação e Interdisciplinares	195
Figura 94 – Região Norte - Linguagem-Comunicação e Matemática	196
Figura 95 – Região Norte - Ciências Naturais e Ciências Humanas	197
Figura 96 – Região Norte - Ciências Naturais e Ciências Humanas	198

Lista de tabelas

Tabela 1 – Parâmetros de Regressão. Fonte: elaborado pelo autor.	80
Tabela 2 – Correlação das Habilidades	93
Tabela 3 – Top 5 Piores Valores de Correlação de Pearson. Fonte: (SILVA et al., 2021b)	120
Tabela 4 – Top 5 Piores Valores de MSE. Fonte: (SILVA et al., 2021b)	121
Tabela 5 – Top 5 Melhores P-values. Fonte: (SILVA et al., 2021b)	121
Tabela 6 – Top 10 Melhores Valores de MSE. Fonte: (SILVA et al., 2021b)	122

Lista de abreviaturas e siglas

AM	Aprendizagem de Máquina
AVA	Ambiente Virtual de Aprendizagem
CAT	<i>Computerized Adaptive Testing</i>
ENEM	Exame Nacional do Ensino Médio
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IRT	<i>Item Response Theory</i>
ML	Machine Learning
TRI	Teoria de Resposta ao Item
VLE	Virtual Learning Environment
1PL	Modelo de um parâmetro (dificuldade) da TRI
2PL	Modelo de dois parâmetro (dificuldade e discriminação) da TRI
3PL	Modelo de três parâmetro (dificuldade, discriminação e propensão palpite) da TRI

Sumário

	Sumário	23
1	INTRODUÇÃO	27
1.1	Estrutura da Pesquisa	29
1.2	Problema da Pesquisa	31
1.3	Objetivos	32
1.3.1	Objetivo Geral:	32
1.3.2	Objetivos Específico:	32
1.4	Contribuições e Publicações	32
1.5	Organização do Trabalho	33
2	JUSTIFICATIVA	35
2.1	Análise da Avaliação e Predição da Aprendizagem	36
2.2	Breve Panorama Educacional Brasileiro e Exames Nacionais	37
2.3	Microdados do Enem	40
2.3.1	Prova de Redação	45
2.3.2	Provas Objetivas	47
2.3.2.1	Conjunto do Estudo	47
2.3.3	Especificações Pedagógicas	49
2.3.3.1	Competências e Referências de Linguagem e Comunicação	50
2.3.3.2	Competências e Referências de Matemática	52
2.3.3.3	Competências e Referências de Ciências Naturais	53
2.3.3.4	Competências e Referências de Ciências Humanas	59
2.4	Caracterização da Aprendizagem	62
3	AVALIAÇÃO INTELIGENTE DE APRENDIZAGEM APOIADA POR COMPUTAÇÃO	67
3.1	Trabalhos Relacionados	68
3.2	Arquitetura do Experimento	69
3.2.1	Método Proposto	69
3.2.2	Processamento dos Dados	71
3.2.2.1	Cálculo das Correlações	71
3.2.2.2	Calculo das Distâncias	73
3.2.2.3	Classificação via Métricas Agregadas	75
3.3	Resultados e Discussão	77
3.4	Conclusão	80

4	EVIDENCIAÇÃO DE PERFIS DE APRENDIZAGEM E CARACTERIZAÇÃO DE ATIVIDADES	81
4.1	Trabalhos Relacionados	83
4.2	Arquitetura do Experimento	85
4.2.1	Característica dos Dados	85
4.2.2	Identificação Inteligente de Perfis	87
4.2.2.1	A Teoria de Resposta ao Item	87
4.2.2.2	Modelo de 1 Parâmetro	88
4.2.2.3	Modelo de 2 Parâmetros	88
4.2.2.4	Modelo de 3 Parâmetros	88
4.2.2.5	Parâmetros dos Modelos Preditores	89
4.2.2.6	Função de Informação	91
4.2.3	Descrição do Sistema	91
4.3	Resultados e Discussão	92
4.3.1	Perfis de Desempenho	93
4.3.2	Grupos de Habilidades	94
4.3.3	Caracterização dos Itens de Avaliação	95
4.3.4	Comparação com Métodos de Aprendizagem de Máquina	103
4.4	Conclusão	103
5	AVALIAÇÃO DE PROFICIÊNCIA EM REDAÇÃO POR ANÁLISE DE REGRESSÃO	107
5.1	Trabalhos Relacionados	108
5.2	Arquitetura do Experimento	109
5.2.1	Ferramentas e Métodos	109
5.2.1.1	Teoria de Resposta ao Item: Modelo de Resposta Gradual	110
5.2.2	Arquitetura Proposta	113
5.2.3	Caracterização da Aprendizagem	114
5.3	Resultados e Discussão	117
5.3.1	Caracterização dos Itens	117
5.3.2	Avaliação do Tamanho da Porção de Instâncias para Treino e Quantidade de Características	117
5.4	Conclusão	122
6	TESTES ADAPTATIVOS COMPUTADORIZADOS COM AGENTES INTELIGENTES	123
6.1	Trabalhos Relacionados	125
6.2	Arquitetura do Experimento	126
6.2.1	Conjunto de Dados do Experimento	127
6.2.2	CAT - Abordagem Proposta	127

6.3	Resultados e Discussão	128
6.3.1	CAT com Prova de MT	145
6.4	Conclusão	149
7	CONSIDERAÇÕES FINAIS	161
7.1	Contribuições da Pesquisa	161
7.2	Trabalhos Futuros	162
7.3	Conclusão	162
	REFERÊNCIAS	163
	APÊNDICES	175

1 Introdução

A reflexão crítica sobre a qualidade dos processos de ensino-aprendizagem leva à análise da eficiência e eficácia dos subprocessos associados ao aprendizado. Do ponto de vista didático pedagógico, existem teorias que descrevem a estratificação dos níveis de aprendizagem como a corrente teórica discutida no trabalho de [Costa, Júnior e Fernandes \(2019\)](#). Nela, o processo de ensino descrito ocorre enquanto o sujeito é apresentado a novos conhecimentos, por meio de conceitos embarcados em objetivos de aprendizagem.

De modo geral, o processo de ensino-aprendizagem é realizado pela modulação dos objetivos de aprendizagem ([DAVID, 2019](#)), explicitados em planos de ensino que contém: os tópicos de conteúdo a serem abordados por aula; os métodos de exposição e discussões dos assuntos; a lista de referências bibliográficas; e as formas de avaliação. As avaliações, neste contexto, denotam instrumentos de coleta de dados como questionários de perguntas que visam identificar no sujeito avaliado estágios de conhecimento a respeito dos objetos de aprendizagem propostos. Pela abstração da parte documental/legal de projetos de curso, tratados em detalhes na produção acadêmica sobre currículos ([LIMA et al., 2021](#)), podemos considerar para um curso qualquer que os conhecimentos são apresentados em ordem incremental de complexidade, observando os conhecimentos pretéritos necessários para sedimentar novos saberes ([ALVES; MEDEIROS; MELO, 2020](#)).

A título de exemplificação, a Figura 1 propõem, num visual associado a elementos da cultura contemporânea, tópicos relacionados ao aprendizado da disciplina *Sistemas Operacionais* de um curso de graduação em informática de uma instituição superior de ensino do Espírito Santo. A ideia proposta no infográfico é sugerir que níveis de habilidade ordenados gradativamente em termos de capacidade e complexidade estão associados a conjuntos de aprendizagem. A fim de conferir um rigor metodológico na discussão de aprendizagem, consideramos o emprego de elementos da meta-aprendizagem para apresentar, com algum nível de detalhe, os processos associados à forma como se aprende e qual a ordem de conveniência para uma abordagem lógica, sob o aspecto da adequação educacional ([ASCARI; GOTTARDO; PIMENTEL, 2020](#)). Em termos práticos e sucintos, podemos considerar como exemplo as etapas que sujeitos nascidos no Brasil cursam para aprender a ler e escrever, sendo elas: inicialmente a etapa de reconhecer e reproduzir os símbolos básicos como figuras geométricas triviais, letras e números. A seguir, na alfabetização, passa-se a reconhecer as propriedades semânticas dos símbolos e as resultantes de suas inúmeras combinações, conseguindo separar letras e números, bem como agrupá-las para representar quantidades e/ou palavras. Finalmente, espera-se que o sujeito instruído em tais habilidades possa utilizar suas capacidades combinadas para interpretar a realidade e também conseguir descrevê-la, qualitativa e quantitativamente. Portanto,

Figura 1 – Sequência de Objetivos de Aprendizagem. Fonte: Elaborado pelo Autor.



consideramos que conjuntos de habilidades podem caracterizar níveis de aprendizagem. Por meio das características identificadas na composição dos perfis de aprendizagem, modelos preditivos podem ser elaborados e auxiliar na mensuração da eficiência dos processos de ensino. Ora, usar perfis e conhecer sua capacidade para projetar avaliações cujo resultado médio esperado esteja em um intervalo esperado.

Estabelecidos os modelos de aprendizagem dos alunos, têm-se as informações necessárias para prever o desempenho de determinados grupos em certas condições (ROMERO et al., 2007). Definir um processo de avaliação que considere as nuances de cada etapa de avaliação é prover segurança ao resultado da aprendizagem (PLAJNER, 2017). Nesse sentido, a definição de objetivos de aprendizagem, aliados a métodos de avaliação compatíveis com cada estágio do aprendizado, ocasiona uma geração de informações que

evidenciam o processo educacional, oportunizando um maior controle sobre os resultados. Os resultados apresentam como caracterizar, mensurar e avaliar os alunos, itens de avaliação e objetos de aprendizagem, de modo a usar testes adaptativos computacionais para tornar a aprendizagem efetiva.

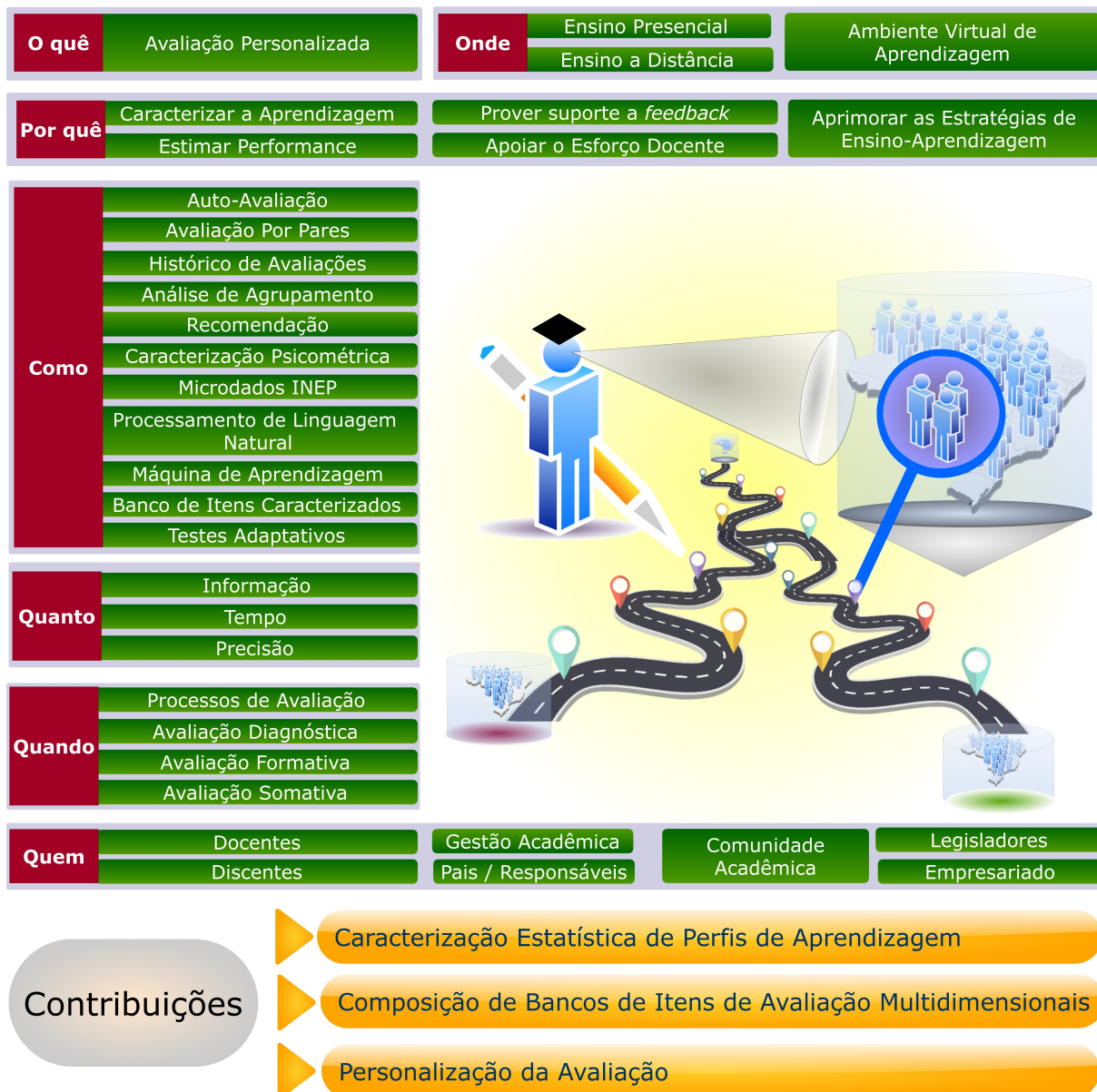
1.1 Estrutura da Pesquisa

A pesquisa proposta pode ser classificada como multidisciplinar, haja vista recorrer a conhecimentos da computação, associados à modelos estatísticos e psicométricos sobre dados educacionais, com vistas a identificar grupos de aprendizagem e estabelecer modelos preditivos. A Figura 2 ilustra os componentes de pesquisa e como eles foram utilizados na compreensão do problema estudado.

- **Avaliação Personalizada:** contém a ideia de selecionar itens de avaliação que sejam compatíveis com a capacidade do aluno avaliado. Para o escopo desta pesquisa, consideremos avaliação personalizada amparada pelo histórico de itens avaliados, o traço de habilidade latente de cada aluno examinado e o banco de itens disponível.
- **Caracterização da Aprendizagem:** caracterizar a aprendizagem coletiva em níveis, representados por distribuições estatísticas, e de parâmetros estimados. Caracterizar a aprendizagem individual, via autoavaliação combinada com avaliação por pares. Predizer desempenho por meio dos modelos da Teoria de Resposta ao Item. Prover *feedback* automático ao examinado a respeito de sua desempenho. Tais informações podem apoiar na redução do esforço docente com a construção de estratégias de ensino personalizadas.
- **Aplicação:** O método proposto funciona independentemente, mas pode ser integrado a ambientes virtuais de aprendizagem, por isso, apoiar o ensino presencial e à distância.
- **Conjunto de Dados de Teste:** Para comprovar a razoabilidade do método proposto, foram utilizados os microdados do ENEM, dispostos no *site* do INEP¹.
- **Métodos:** O uso de autoavaliação, combinada com avaliação por pares, confere ao aluno a oportunidade de exercitar seu senso crítico em relação a si mesmo e à coletividade. Por meio do histórico de avaliações, podem ser executadas análises de agrupamento que evidenciem padrões capazes de identificar grupos de desempenho. Uma vez caracterizados os grupos, pode-se realizar recomendação de itens de aprendizagem, com base no grupo que o sujeito melhor se adequar. O uso de modelos psicométricos demandam a caracterização dos itens de aprendizagem e

¹ www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem

Figura 2 – Estrutura da Pesquisa. Fonte: Elaborado pelo Autor.



oportunizam a estimativa do desempenho, o que possibilita traçar estratégias para maximizar o sucesso. O uso de técnicas de Processamento de Linguagem Natural permitiu a classificação dos itens de avaliação, quanto as competências estipuladas pelo quadro pedagógico disponibilizado no conjunto de dados disponibilizado pelo INEP. A classificação quanto à competência, confere aos itens uma característica complementar a ser considerada no processo de recomendação, caso haja interesse em filtrar itens pelo seu conteúdo, em razão de estratégias de ensino em específicas. Das informações disponibilizadas pelo INEP, constam as respostas dos examinados e suas notas. Com elas, técnicas de Aprendizagem de Máquina (*Machine Learning* — ML) puderam ser utilizadas em complementação à composição de modelos preditivos. Os bancos de itens caracterizados, oriundos dos microdados do INEP, puderam

ser usados no teste da recomendação para novos examinandos. Testes adaptativos demandam banco de itens padronizados; estabelecimento de estratégias de seleção de itens; e definição dos critérios de parada.

- **Estimar a Quantidade Mínimo de Instâncias de Treino:** A quantidade de dados históricos disponíveis influencia a qualidade dos modelos elaborados. Quantificando a quantidade mínima para uma predição razoável, é possível estimar o tempo necessário para compor um banco de itens padronizados. A variação da quantidade de itens e de sua qualidade, em termos psicométricos, implica na precisão da predição de desempenho.
- **Oportunidade de Pesquisa:** Os métodos e itens propostos visam apoiar os processos de avaliação em suas três configurações, a saber: avaliação diagnóstica — a qual visa identificar o nível de aprendizagem; a avaliação formativa — que após tratamento do diagnóstico feito na etapa anterior, tem em vista avaliar se as lacunas encontradas foram preenchidas; avaliação somativa — utilizada quando superadas as etapas anteriores, visa quantificar a proficiência para fins formais, como progressão de nível de aprendizagem.
- **Atores em Potencial:** Os atores deste contexto formam entre si um aglomerado de forças intervenientes, sendo a base da interação entre docentes e discentes, por meio do processo de ensino-aprendizagem. Mas do lado docente, há a Gestão Acadêmica, que determina rotinas e acompanha a execução do ensino. Já o corpo discente, tem os pais/responsáveis como grupo imediatamente correlato. Agrupando as duas searas, tem-se a comunidade acadêmica, com os demais profissionais envolvidos no processo, direto ou indiretamente, mas que são igualmente interessados no desempenho da instituição de ensino. Legisladores são atores interessados nos resultados para formulação de políticas públicas de ensino que sejam compatíveis com os anseios da sociedade. Por fim, os atores representado no grupo empresariado, interessa-se pela formação de colaboradores que possam integrar seus negócios ou consumidores, neste caso, instruídos a ponto de estar em atividade econômica.

1.2 Problema da Pesquisa

Com o propósito de ampliar o arcabouço científico de trabalhos que discutam a aplicação de modelos psicométricos na análise da aprendizagem, incluindo o tema de aprendizagem de máquina. Assumindo o cenário preliminar em que o escopo do universo do problema são: a necessidade de mensurar precisamente a avaliação de aprendizagem; o desejo de gerenciar com algum apoio tecnológico os dados acadêmicos para evitar falha humana por equívoco ou exaustão; a vontade de prover *feedback* personalizado sobre o

nível de ensino alcançado e também reduzir o esforço demandado do docente; a composição de banco de itens padronizados, o problema de pesquisa formulado foi:

- é possível estabelecer uma rota de avaliação personalizada, estruturado num formato de teste adaptativo computadorizado, apoiado por aprendizagem de máquina e modelos da Teoria de Resposta ao Item?

1.3 Objetivos

1.3.1 Objetivo Geral:

O objetivo principal deste trabalho é a proposição de um método personalização da avaliação, por meio de análise de agrupamento e modelos psicométricos da TRI, implementado num formato de teste adaptativo computadorizado.

1.3.2 Objetivos Específico:

1. Estudar formas de melhor representar componentes de habilidades por desempenho.
2. Identificar e analisar melhores formas de mapear perfis de alunos.
3. Compor uma estratégia de avaliação personalizada que opere com perfis de aprendizagem e modelos psicométricos.

1.4 Contribuições e Publicações

Agrupados por contribuições, os seguintes artigos foram publicados durante o desenvolvimento da presente tese:

I. Representação da aprendizagem por análise das distribuições estatísticas, análise de agrupamento, análise de correlações e distâncias vetoriais.

1. An Easy-to-Read Visual Approach to Deal With Peer Reviews and Self-Assessments in Virtual Learning Environments. In: Proceedings of the 2020 The 3rd International Conference on Big Data and Education (SILVA et al., 2020a).
2. Avaliação em Pares e Autoavaliação: Um Modelo Estatístico para Perfilação de Alunos. Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE) (OLIVEIRA, 2018).

II. Caracterização de perfis de aprendizagem por meio da pontuação de desempenho em itens de avaliação calibrados.

1. Towards a Tailored Hybrid Recommendation-Based System for Computerized Adaptive Testing Through Clustering and IRT. In: LANE, H. C.; ZVACEK, S.; UHOMOIBHI, J. (Ed.). Proceedings of the 12th International Conference on Computer Supported Education, CSEDU 2020 (SILVA et al., 2020a).
2. Lukewarm Starts for Computerized Adaptive Testing Based on Clustering and IRT. In: LANE, H. C.; ZVACEK, S.; UHOMOIBHI, J. (Ed.). Computer Supported Education. Cham: Springer International Publishing, 2021 (SILVA et al., 2021a).
3. Personalizing Evaluations to Aid Chemistry Learners. 2023. Paper presented at the 17th Annual International Technology, Education and Development Conference - INTED 2023 (OLIVEIRA; IZO; SILVA, 2023)

III. Multicaracterização de Itens de Avaliação e Personalização da Avaliação

1. Writing Proficiency Assessment: Regression Analysis of Item Response Theory Supported by Machine Learning Techniques. In: 2021 XLVII Latin American Computing Conference (CLEI)(SILVA et al., 2021b).
2. Recommendation Filtering à La Carte for Intelligent Tutoring Systems. In: BORRATTO, L. et al. (Ed.). Bias and Social Aspects in Search and Recommendation - First International Workshop, BIAS 2020 (SILVA et al., 2020c).
3. A Prototypical Semantic Annotator for A Tribuna Newspaper. In: LEMOS, D. L. da S. et al. (Ed.). Proceedings of the XIII Seminar on Ontology Research in Brazil and IV Doctoral and Masters Consortium on Ontologies (ONTOBRAS 2020) (OLIVEIRA et al., 2020).
4. Minimalist Fitted Bayesian Classifier-Based on Likelihood Estimations and Bag-of-Words. In: BERGET, G. et al. (Ed.). Linking Theory and Practice of Digital Libraries. Cham: Springer International Publishing, 2021 (BOURGUET; SILVA; OLIVEIRA, 2021).
5. Adapting NER (CRF+LG) for Many Textual Genres. In: IberLEF@SEPLN (PIROVANI et al., 2019).

1.5 Organização do Trabalho

O Capítulo 2 apresenta considerações sobre avaliação da aprendizagem e sobre representações pedagógicas da aprendizagem, sobre as quais, a análise computacional pode

auxilia na compreensão dos processos e fenômenos faceados. Ainda naquele capítulo, são apresentados os conjuntos de dados do Exame Nacional do Ensino Médio, utilizados na pesquisa.

O Capítulo 3 versa sobre Avaliação Apoiada por Computação, apresentando as fundamentações teóricas do tema, bem como o experimento de autoavaliação e avaliação em pares, com dados reais.

O Capítulo 4 apresenta a Caracterização da Aprendizagem por meio de notas artificiais, geradas a partir distribuições estatísticas com parâmetros controlados, a fim de reproduzir situações reais. Com base na caracterização de aprendizagem, apresentamos uma técnica de recomendação, recorrendo a modelos da Teoria de Resposta ao Item, com o modelo de 3 parâmetros.

O Capítulo 5 mostra a Avaliação de Proficiência em Redação pela utilização do Modelo de Resposta Gradual. Ademais, apresenta uma comparação entre técnicas de regressão e o modelo psicométrico.

O Capítulo 6 apresenta nossa proposta de teste adaptativo computadorizado, incrementado com as técnicas de análise de agrupamento e teoria da resposta ao item, implementado com os conceitos de agente inteligente.

O Capítulo 7 faz o fechamento da discussão apresentada com as considerações finais sobre avaliações inteligentes, bancos de itens padronizados e modelos preditivos. Complementam o capítulo as descrições das limitações da pesquisa e as perspectivas de estudos futuros.

2 Justificativa

A habilidade de expressão de um sujeito demonstra sua capacidade de compreensão da realidade. A demonstração do saber fazer é uma característica importantes a se observar no processo de avaliação (BAKER, 2001). A qual, sendo bem sucedida, permite inferir e balizar a qualidade final da aprendizagem. Contudo, para além da observação do saber fazer, não há uma plena concordância a respeito de quais outras informações devem ser consideradas para efeito de avaliação, bem como formas e métodos de avaliar (PERRENOUD, 1998). Ladeando tal questão, tem-se a divergência, mais conceitual do que pragmática, entre práticas educacionais eficazes. Cada sujeito constrói subjetivamente o significado de sua trajetória de aprendizagem, em termos práticos, isso ocorre a partir de suas experiências pregressas e as que lhe são propostas ao longo do processo de aprendizagem (GARCIA; GOMES, 2020). Desta maneira, surgem naturalmente trilhas individuais, mesmo que similares entre si em algum grau, tendem a ser um tanto peculiares por serem retratos de cada sujeito e suas idiossincrasias (ROMERO; MACHADO; WIVES, 2019). Práticas de *feedback* tem sido usada como critério de conscientização da meta-aprendizagem do aluno, bem como, fator de recomendação de objetos de aprendizagem. Quanto mais individualizado o *feedback* ou a recomendação, maior a chance dele ser considerado pelo aluno (OLIVEIRA; CIARELLI; OLIVEIRA, 2013).

No processo de ensino-aprendizagem, o instante do cálculo de notas de avaliação é um momento crítico, haja vista que lida com a mensuração da aprendizagem do aluno, que não é uma grandeza determinística. Nesta etapa ocorre a verificação do sucesso da aprendizagem pela avaliação da manifestação do aluno a respeito do objeto de aprendizagem tratado durante o processo de ensino. Assim, a observação da expressão do conhecimento do aluno é utilizada como fator de medida. Após o cálculo da nota, o docente tem a possibilidade de prover *feedback* a cada aluno, sobre quais entendimentos de seus conhecimentos apresentaram lacunas e, eventualmente, dialogar sobre possíveis formas de saná-las (PERRENOUD, 1998; SANTOS; BASSANI, 2020; FERREIRA et al., 2020). Outrossim, quando confrontado com avaliações de muitos alunos, um desafio sobressalente manifesta-se aos professores, pela necessidade de manipular uma massa de dados inteligente (VILLEGAS-CH et al., 2018), respeitando as características geradoras dos dados, bem como, assegurar a razoabilidade no cálculo das notas. Isto é, considerar as características peculiares de cada aluno. Por exemplo, as variabilidades socioeconômicas e cognitivas podem ocasionar resultados diferentes de aprendizagem (FERREIRA, 2019a).

Recursos para tratamento das notas tem sido implementado em Ambientes Virtuais de Aprendizagem (AVA) contemporâneos, algumas ferramentas nativas às plataformas e outras, integradas vias padrões de *software* (OLIVEIRA, 2018). O trabalho de Zampirolli

et al. (2020) mostra o fluxo completo de geração de questões automáticas *online*; bem como a correção das respostas do avaliado. Um estudo similar foi realizado por Romero, Machado e Wives (2019), no qual foi incrementada a análise dos questionários aplicados via AVA, utilizando modelos da Teoria de Resposta ao Item (TRI), demonstrando a possibilidade de prover mais informações ao docente, a respeito do instrumento de avaliação utilizado. Por meio da utilização de itens de avaliação baseados na TRI, Zaffalon et al. (2019) comparou as diferenças relativas ao modelo ELO, originalmente criado para avaliar jogadores de xadrez. Utilizando avaliações no formato de autoavaliações e avaliações por pares, os estudos de Silva et al. (2020a) e de Oliveira (2018) demonstraram o potencial de caracterização dos dados, utilizando correlação de *pearson* e cálculo de distância euclidiana para verificar similaridades dos desempenhos. As três principais razões para utilização da TRI são: teste de modelo; explicar a variação de resultados, em função das características do teste e da pessoa examinada; e definir uma pontuação para o teste, obtida por meio da mensuração do traço latente individual dos participantes (CHALMERS, 2012).

2.1 Análise da Avaliação e Predição da Aprendizagem

Um problema comum a todas as áreas do ensino é o esforço demandado do docente para tratar cada aluno de maneira personalizada, principalmente com turmas numerosas. Para enfrentar tal situação, propomos o emprego de técnicas de Inteligência Artificial, de Aprendizagem de Máquina e de Psicometria para incrementar a qualidade da aprendizagem, com vistas a também reduzir o esforço demandado do docente no processo avaliativo. A abordagem proposta inicia da premissa que os resultados dos alunos, concernentes às avaliações de objetos de aprendizagem, permitem modelar tanto os sujeitos avaliados, quanto os itens de avaliação. Possibilita também inferir a eficácia dos objetos de aprendizagem utilizados (BARRETO et al., 2020). A respeito da avaliação, o formato tradicional implica em apresentar itens de avaliação sobre determinados objetivos de aprendizagem, os quais se deseja aferir o conhecimento (DAVID; GREEN; SANTOS, 2020). Desse modo, a mensuração da percepção da aprendizagem é realizada segundo o volume de testes e a lógica avaliativa definida. Haja vista que almejamos computar processos de aprendizagem para mensuração cognitiva, contextualizados como avaliações de aprendizagem, consideramos razoável o emprego de técnicas da Psicometria para amparar metodologicamente a descrição pormenorizada dos resultados encontrados. Assim, recorreremos a modelos psicométricos da Teoria de Resposta ao Item. Com eles foi possível expandir as avaliações dos itens respondidos corretamente para além de medidas resumitivas tradicionais como média ou somatória, isto é, conseguimos caracterizar os itens de avaliação em termos de dificuldade, discriminação e propensão a chute. Foi possível também calcular a probabilidade de acerto por item.

Em relação ao aspecto comportamental da avaliação, consideramos o emprego

da combinação de avaliações por pares e autoavaliações, estimulando o senso crítico dos alunos avaliados. Ademais, tal formato avaliativo oportuniza a inter-disciplinaridade dos três domínios da *Taxonomia de Bloom* (COSTA; FERNANDES, 2020; OLIVEIRA; SILVA; NETO, 2020), por demandar esforço *cognitivo* pela avaliação de saberes; *psico-afetivo* em função da necessidade de considerar as respostas dos pares; e esforço *psicomotor* diante da necessidade de estipular notas para as respostas dos pares, tomando para si a perspectiva que o docente teria ao realizar a correção. Por meio de técnicas de análise de agrupamento das diversas informações geradas nos processos de ensino-aprendizagem (OLIVEIRA et al., 2017; SILVA et al., 2020b), pode-se evidenciar características e níveis de aprendizagem que subsidiem a composição de estratégias ensino customizadas.

Na verificação da abordagem proposta, foi recorrido a dados coletados de avaliações de alunos de curso de graduação da área de humanas em instituição de ensino pública federal; dados artificiais, gerados por meio distribuições estatísticas controladas em experimento; e dados oriundos do Exame Nacional do Ensino Médio (ENEM) entre os anos de 1998 e 2019. Para as três fontes de dados utilizadas foi possível agrupar os valores para representar o desempenho dos examinandos e identificar lacunas de aprendizagem, quando considerados os assuntos dentro de cada área de seleção avaliada. Na análise de agrupamento para identificação de perfis de aprendizagem, tratando as notas como vetores, foram utilizadas técnicas de correlação, cálculo da distância euclidiana e agrupamento pela técnica *k-means*, cujo número de grupos foi otimizado pelo cálculo das distâncias entre as instâncias (MARUTHO et al., 2018). Para calibração e teste dos modelos psicométricos, foram utilizadas as notas das bases mencionadas. A realização do experimento comparativo sobre as variáveis componentes do modelo e a quantidade de instâncias necessárias a uma predição de qualidade otimizada, foi utilizado o conjunto notas do Enem de 2019.

2.2 Breve Panorama Educacional Brasileiro e Exames Nacionais

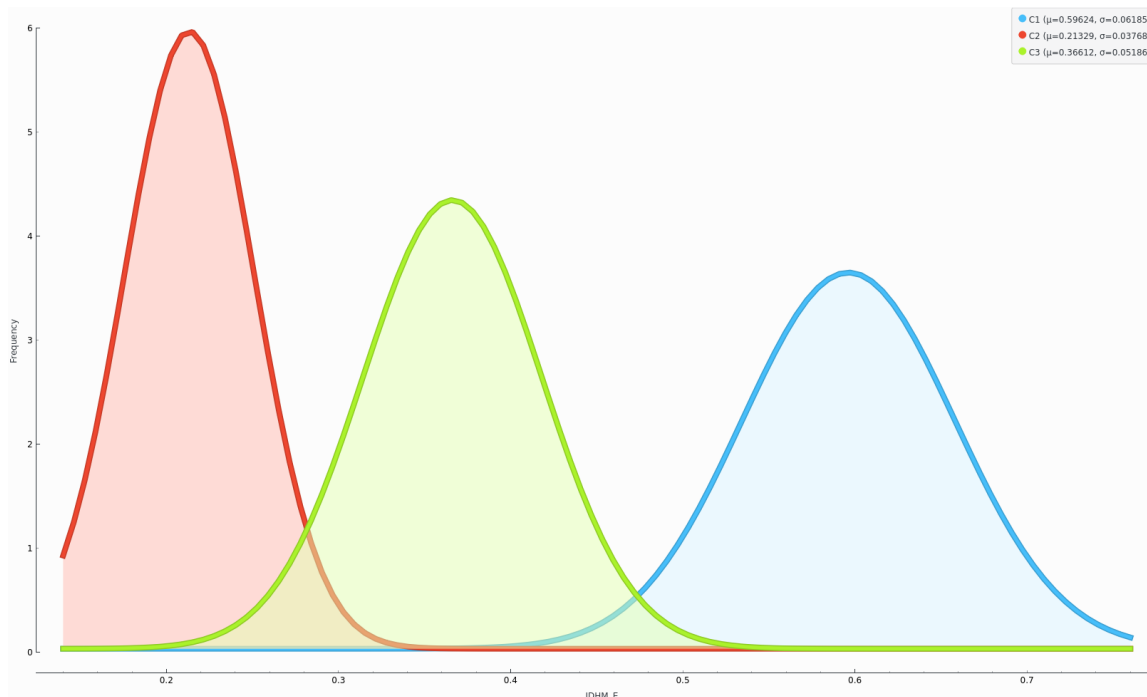
Tomando a educação brasileira como cenário de análise, sabidamente com contrastes agudos evidenciados nos mapeamentos educacionais, a exemplo do que pode ser visto no Índice de Desenvolvimento Humano — Vertente Educacional IDHM-E^{1,2}. Na Figura 3, apresentamos o gráfico de densidade das distribuições estatísticas geradas com os índices dos estados do Brasil. Os dados foram apurados nos períodos do Censo Nacional Brasileiro, isto é, 1991, 2000 e 2010 (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 1991; INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2000; INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2010).

¹ Ele é um índice sintético da dimensão Educação que é um dos 3 componentes do IDH. Seu cálculo é efetuado através da média geométrica do subíndice de frequência de crianças e jovens à escola, com peso de 2/3, e do subíndice de escolaridade da população adulta, com peso de 1/3.

² www.atlasbrasil.org.br. Acesso em 30/07/2022.

A referida figura sintetiza os valores em três grupos, representados pelos parâmetros de uma distribuição estatística normal. Na cor vermelha, o grupo de menor índice médio e desvio padrão, demonstrando ser o de menor variação entre todos. A distância da média do grupo inferior para o grupo superior (de cor azul), é de 0,38 unidades do índice.

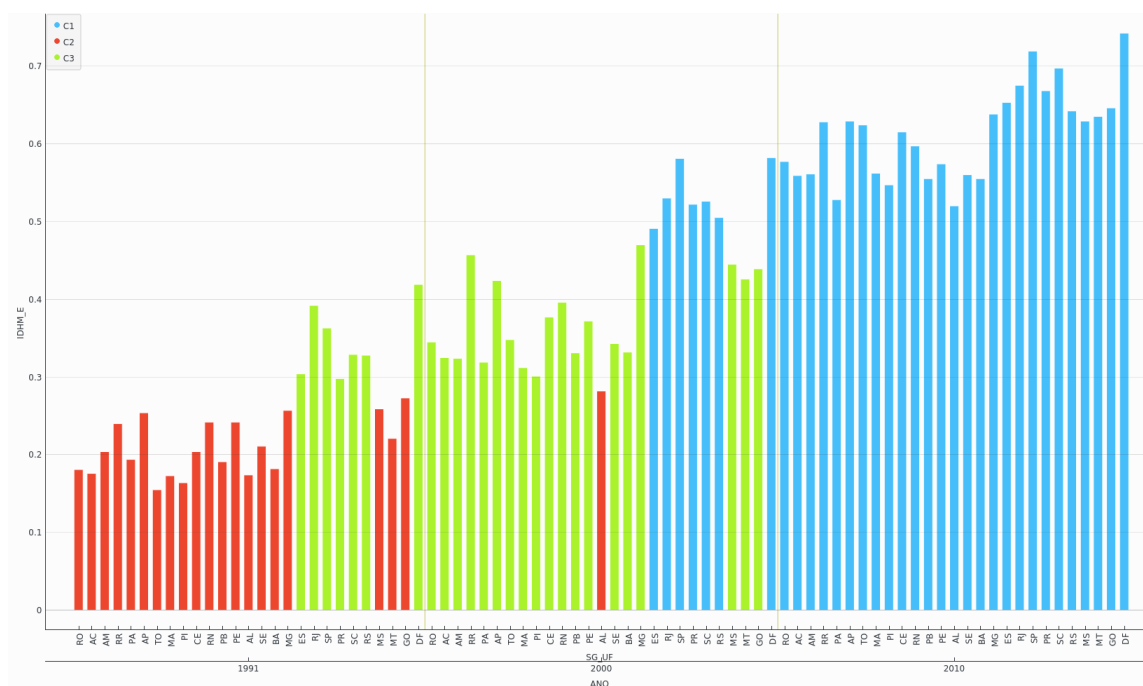
Figura 3 – Características dos Grupos.



Para o grupo intermediário, essa diferença cai menos da metade, sendo no valor de 0,15. Haja visto o período de apuração dos dados, na hipótese de um desenvolvimento educacional linear, seria uma expectativa plausível, a figuração de cada um dos estados da federação em cada um dos grupos. Caso fosse assim, seriam verdadeiras as associações: Grupo Inferior (C2 — vermelho) em 1991; Grupo Intermediário (C3 — verde) em 2000; e Grupo Superior (C1 — azul) em 2010. Todavia, não é o que se observa na Figura 4, cujo eixo vertical apresenta a escala do índice e o eixo horizontal apresenta os estados ordenados crescentemente por ano.

Vemos ainda na Figura 4 que em 1991 os estados das regiões sul e sudeste do Brasil, excetuando-se Minas Gerais, de antemão possuíam índices compatíveis com o Grupo Intermediário (C3 — verde). Situação idêntica ao que ocorreu com o Distrito Federal, o que evidenciou que tais entes prontamente compunham um grupo de maior frequência nos níveis educacionais. O mesmo ocorreu no ano 2000, porém, os referidos estados e o DF apresentavam índices da ordem do Grupo Superior (C1 — azul). Na contramão deste movimento, em 2000, o estado de Alagoas apresentava um índice da ordem do Grupo Inferior (C2 — vermelho), evidenciando algum atraso, comparado aos outros estados. Em 2010, todos os estados ficaram no grupo Superior (C1 — azul), contudo, Alagoas e Pará aparecendo com os menores índices, no valor pouco maior do que 0,5. No Distrito Federal

Figura 4 – Evolução Educacional por Ano.



e em São Paulo, o índice ultrapassou a marca de 0,7, sendo os maiores dentre todos os outros estados. A granulação e aprofundamento deste tipo de observação pode ser feita por meio de dados municipais, os quais também estão disponíveis na mesma fonte destes. Em âmbito governamental, existem iniciativas para mapear as informações sobre as estruturas educacionais, com os censos Escolar da Educação Básica e o Censo da Educação Superior. Há também os instrumentos de avaliação como o Sistema de Avaliação da Educação Básica (Saeb) e o Índice de Desenvolvimento da Educação Básica (Ideb), gerado a partir do Saeb e dos níveis de aprovação dos estudantes nas escolas. Como já citado anteriormente, há o Enem, que além de mecanismo de seleção para ingresso no ensino superior, tem função de indicador da qualidade do ensino pelo relacionamento dos sujeitos examinados com suas respectivas instituições de ensino. Numa perspectiva de incrementar os processos de aprendizagem a partir das evidências observadas, os resultados das provas nacionais podem ser insumos para duas importantes finalidades:

- i) *Caracterização dos Avaliados* — A primeira é identificar as lacunas de aprendizagem, granulando a informação em múltiplas camadas: ao nível nacional, regional, estadual, municipal e institucional — público ou privado. Somado a isso, é possível também a caracterização por gênero, etnia, faixa etária, dentre outros atributos socioeconômicos.
- ii) *Construção de Banco de Itens* — definir as provas como objeto de interesse e utilizar seus itens de avaliação para treinar sujeitos que serão submetidos ao teste efetivamente. Esse uso é naturalmente feito no processo de ensino, quando questões da prova são apresentadas aos alunos como exemplos da ocorrência da avaliação

de determinados temas. No entanto, esforços institucionais para apropriação dos itens ministrados e verificação continuada do público ensinado, antevendo quadros de desempenho, ainda não são populares.

Portanto, a mineração de informação do Enem serve para identificar grupos de desempenho e também, auxiliar na composição de bancos de itens de avaliação padronizados (SPENASSATO et al., 2016). Quanto mais informação sobre ambos os conjuntos, maior a possibilidade traçar estratégias para relacioná-los de maneira profícua (SILVA et al., 2021b).

2.3 Microdados do Enem

O emprego dos microdados do Enem como conjunto de teste, visou verificar as características dos dados, que pela origem psicométrica, assegura sua normalidade, isto é, dados distribuídos conforme as proporções esperadas nos intervalos interquartis. Dessa sorte, cada edição do Enem, na faixa de 2009 a 2019, pode ser considerada um banco de dados padronizado, haja vista que os itens de avaliação presentes nas provas são previamente validados pelo INEP, seguindo os critérios descritos nos documentos que acompanham seus microdados. A quantidade de inscritos, vista na Figura 5a, iniciando com cerca de 4 milhões em 2008 e chegando a quase 9 milhões em 2014, demonstram a amplitude territorial da avaliação. Pela diversidade de examinados e assuntos abordados dos conteúdos do ensino médio, os dados do exame também se mostram profícuos em termos de complexidade, variabilidade e propícios a pesquisadores voltados para dados educacionais. Ao tomar o ano de 2019 como referência, a Figura 5b lista os percentuais de participação por estados do Brasil. São Paulo com maior percentual de participantes, qual seja 14% e Roraima com o menor, menos de 1%. O conjunto de dados disponíveis por edição do ENEM, representa, informação territorial, sócio-econômica e também cognitiva.

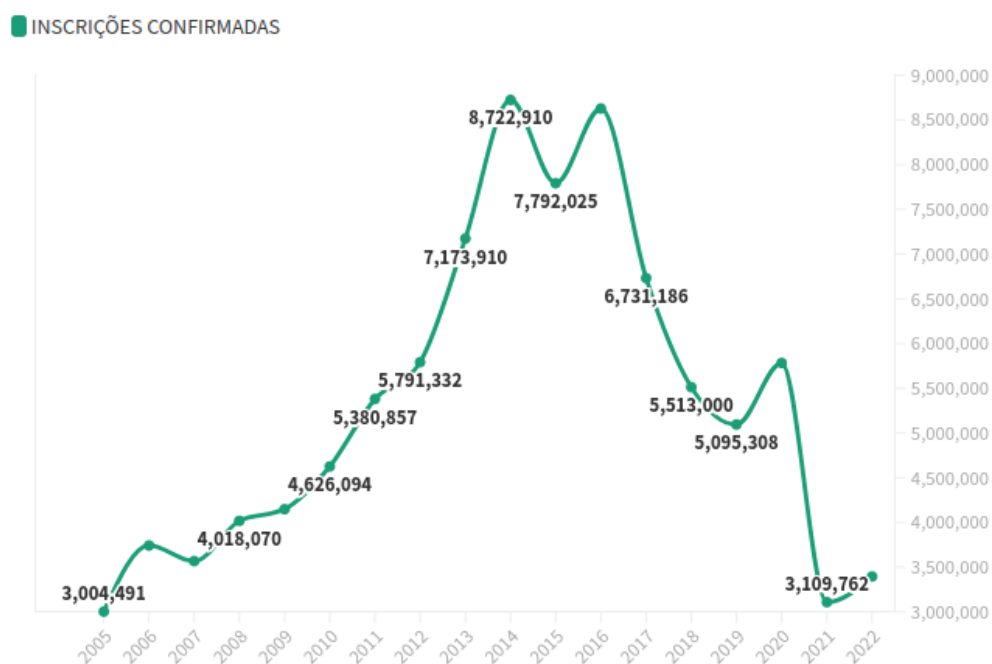
As notas de Ciências Naturais e Ciências Humanas podem ser conferidas nas Figuras 6a e 6b, respectivamente. Para ambos, as menores notas constam no ano de 2011. Para Ciências Naturais, as melhores notas ocorreram no ano de 2017. Para Ciências Humanas, a pontuação mais elevada foi atingida em 2018.

A Figura 7a exhibe distribuições das notas de Linguagem e Comunicação, cuja menor pontuação foi na edição de 2011, enquanto a maior pontuação apurada, foi em 2018. A Figura 7b mostra as notas de Matemática, com as menores e maiores notas nos mesmos anos das provas de Linguagem e Comunicação. Em ambas, nota-se também a uniformidade dos gráficos, em relação ao tamanho intervalo interquartil.

A Figura 8 demonstra o conteúdo dos microdados disponibilizados pelo governo brasileiro. Cada bloco da Figura 8 representa um conjunto semântico de dados, compostos

Figura 5 – Exame Nacional do Ensino Médio.

(a)



(b)

ENEM 2019 - Percentual de Participação por Estado

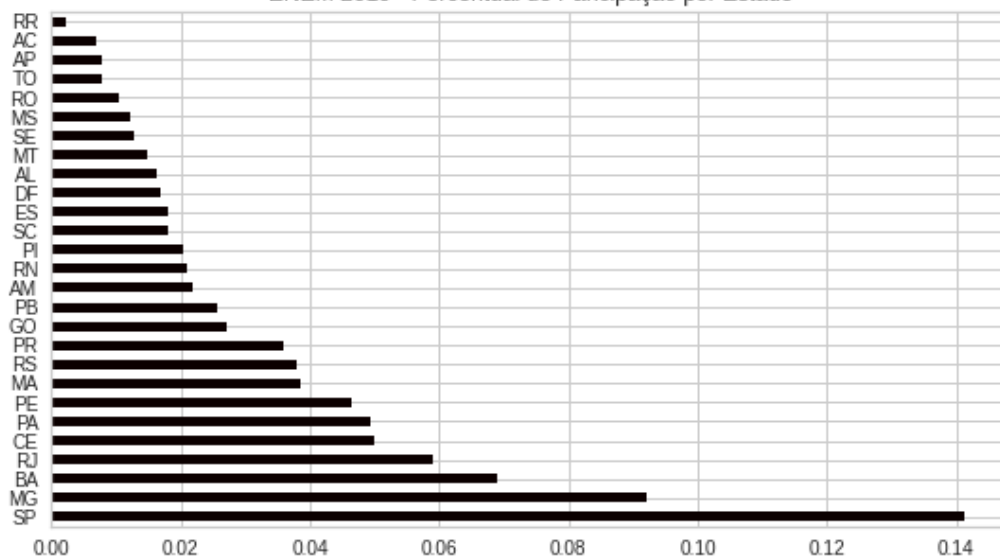
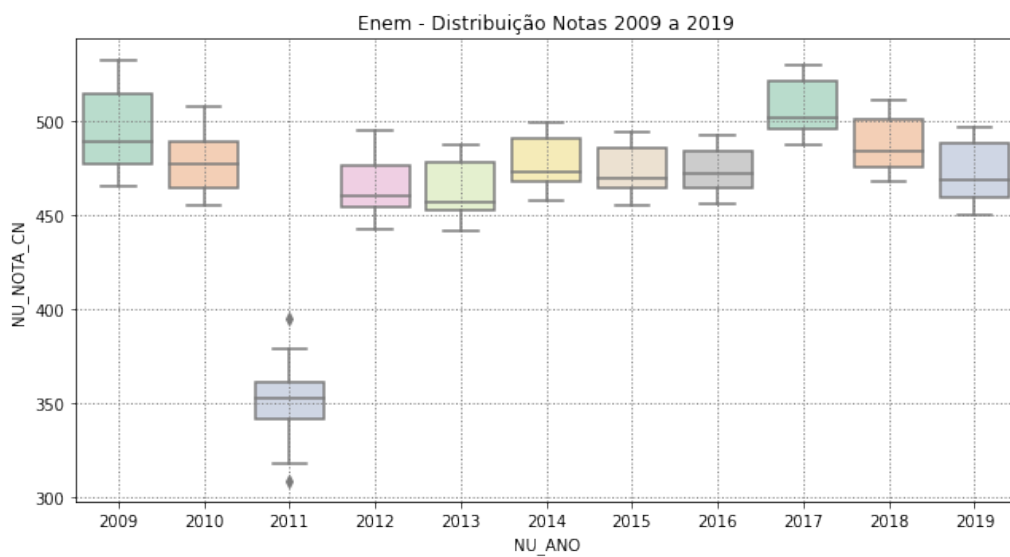


Figura 6 – Distribuição das Notas Objetivas — CN e CH

(a) Ciências Naturais



(b) Ciências Humanas

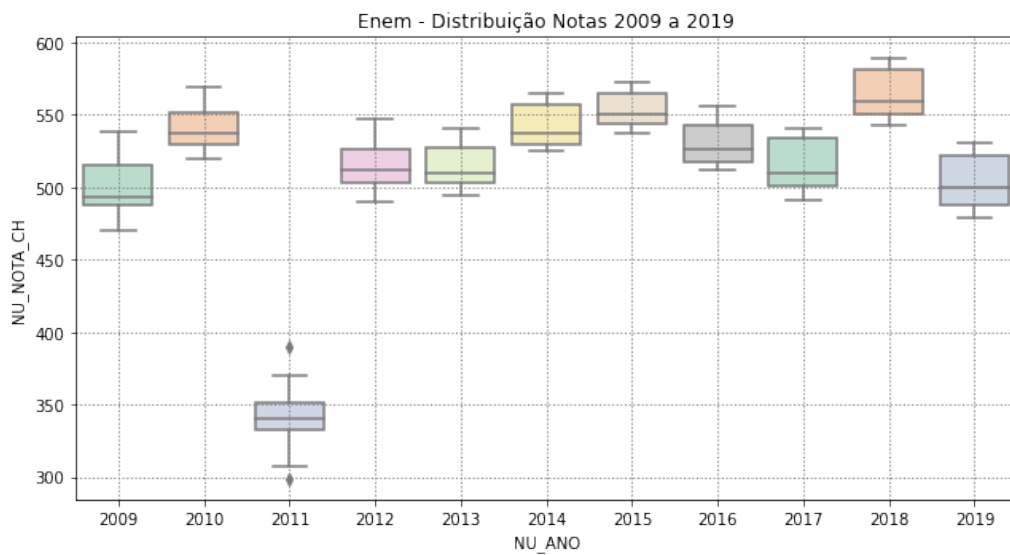
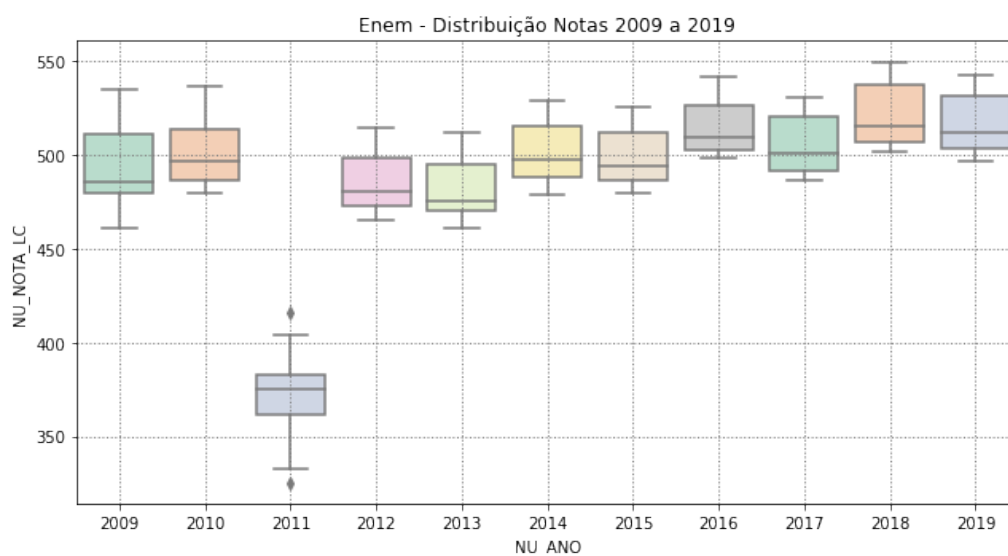


Figura 7 – Distribuição das Notas Objetivas - LC e MT

(a) Linguagem e Comunicação



(b) Matemática

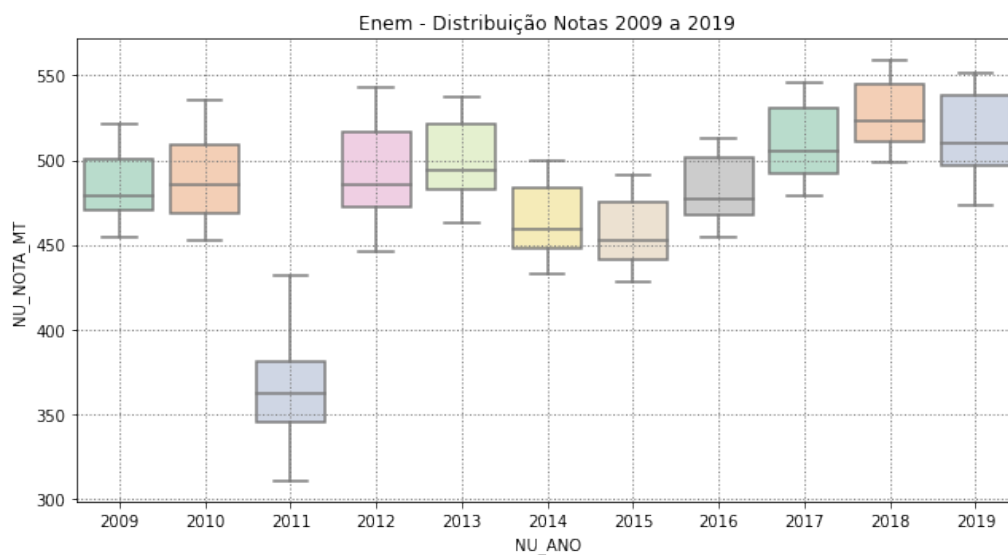
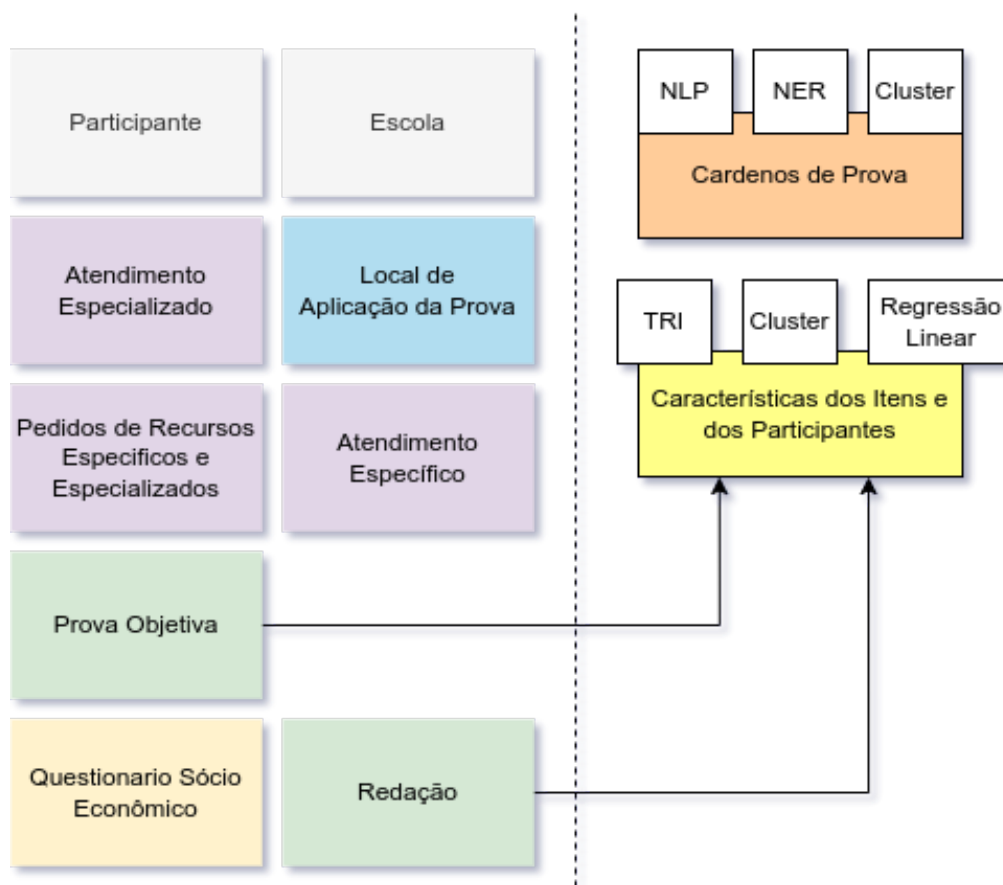


Figura 8 – Estrutura Microdados x Tendência Objetivas



por conjuntos de variáveis numéricas, alfanuméricas, binárias e também do tipo data (dia/mês/ano). Nas referências de metadados dos microdados, cada conjunto pode ser conferido em detalhes. Para esta pesquisa, focamos nos seguintes blocos utilizados: o bloco de Provas Objetivas continha os vetores de resposta dos participantes e as notas obtidas por área. O bloco de Redação, continha os valores alcançados por competência redacional avaliada; junto da nota final obtida na prova de redação da mesma edição do exame. E o bloco Questionário Socioeconômico continha questões relacionadas à configuração familiar, educacional e econômica do participante.

Para esses conjuntos, consideramos a utilização da Teoria de Resposta ao Item, Análise de Agrupamento e Modelos de Regressão, justificados adiante no texto. Com o bloco de caderno de provas, foi possível a utilização de processamento de linguagem natural (NLP), de extração de entidades nomeadas (NER) e o agrupamento de itens de avaliação por similaridade textual ou correspondência semântica.

Entre os anos de 1998 e 2008, as provas objetivas do Enem não eram separadas por área do conhecimento, mas sim, agrupadas por Competências transdisciplinares numeradas de 1 a 5. A partir de 2009, quando o Enem passou a ser instrumento de seleção para ingresso no ensino superior em instituições públicas e privadas, as provas objetivas foram

separadas por área do conhecimento, nomeadas Ciências Humanas, Ciências Naturais, Linguagem e Comunicação, Matemática. Outra mudança que ocorreu na reformulação do Enem, foi a nota da Redação, que passou a ser a somatória das Competências de Redação. Anteriormente, a nota era a média das competências. As competências das provas objetivas interdisciplinares diziam respeito aos seguintes tópicos:

- *Competência 1*: Dominar Linguagens
- *Competência 2*: Compreender Fenômenos
- *Competência 3*: Enfrentar Situações-Problema
- *Competência 4*: Construir Argumentações
- *Competência 5*: Elaborar Propostas

2.3.1 Prova de Redação

A prova dissertativa do Enem é avaliada em 5 (cinco) competências, aumentando gradativamente a complexidade das competências esperadas nos candidatos, a saber:

- i) Demonstrar domínio da modalidade formal de escrita da língua portuguesa;
- ii) Compreender a proposta de redação e aplicar conceitos das diversas áreas da conhecimentos para desenvolver o tema, dentro dos limites estruturais do texto dissertação-argumentativa em prosa.
- iii) Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.
- iv) Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção do argumento; e
- v) Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.

A avaliação de cada competência da dissertação é feita com a atribuição de uma nota que varia entre 0 e 200. O somatório das notas de cada competência integrará a nota final da dissertação, sendo a média aritmética do total de notas atribuídas pelo dois avaliadores. A pontuação máxima é de 1000 pontos se o aluno tiver totalidade na avaliação de todas as competências, e 0, se não pontuar em nenhuma ou for eliminado pelas causas previstas no edital Enem. As causas da eliminação automática na prova de redação com seus respectivos totais foram:

- i) Total de Redações Canceladas/Anuladas: 2.4K;
- ii) Cópia do Texto Motivador: 11.6K;
- iii) Texto em Branco 17.7K;
- iv) Fuga do Tema 15.5K;
- v) Fuga do Tipo Textual 1.6K;
- vi) Texto Insuficiente 3.1K; e
- vii) Parte de Texto Desconectada 2.1K.

No processo de pré-processamento dos dados, que consistia na extração do arquivo de notas e separação das colunas relacionadas às notas da redação, foram eliminados os dados faltantes. Assim, o experimento foi realizado com um conjunto de dados de 3.9M instâncias. A graduação de cada competência segue uma faixa de valores correspondente aos elementos avaliados. Como exemplo, listamos abaixo as pontuações (pontos) e os respectivos critérios de Competência 3:

200 pts: apresenta informações, fatos e opiniões relacionados ao tema proposto, de forma consistente e organizada, configurando autoria, em defesa de um ponto de vista;

160 pts: apresenta informações, fatos e opiniões relacionadas ao tema, de forma organizada, com evidências de autoria, em defesa de um ponto de vista;

120 pts: apresenta informações, fatos e opiniões relacionados ao tema, limitados aos argumentos dos textos motivadores e mal organizados, em defesa de um ponto de vista;

80 pts: apresenta informações, fatos e opiniões relacionadas ao tema, mas desorganizadas ou contraditórias e limitadas aos argumentos dos textos motivadores, em defesa de um ponto de vista;

40 pts: apresenta informações, fatos e opiniões pouco relacionadas ao tema ou inconsistentes e sem a defesa de um ponto de vista; e

0 pts: apresenta informações, fatos e opiniões não relacionadas ao tema e sem defender um ponto de vista.

O conjunto C consiste nas notas de cada aluno j nas cinco competências da prova de redação (Write Test — WT). Para os j -ésimo examinados os ensaios são corrigidos por dois avaliadores, sendo a nota final a média harmônica dos valores da avaliação, sendo:

- i) 1º Primeiro Revisor: $\overrightarrow{PR}_{i=1}^5 = \{c_{ij}\}$
- ii) 2º Segundo Revisor: $\overrightarrow{SR}_{i=1}^5 = \{c_{ij}\}$
- iii) Cálculo de WT: $\theta_j = \sum_{i=1}^5 \frac{\overrightarrow{PR}_{ij} + \overrightarrow{SR}_{ij}}{2}$
- iv) WT Escalada: $\theta_{sj} = \frac{\theta_j - \mu_\theta}{\sigma_\theta}$

Os valores resultantes da divisão das notas pelos avaliadores geram 11 possibilidades de respostas para a Competência avaliada desde os valores [0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200] para as Competitivas ocorrem C1, C3, C4 e C5. A competência C2 não apresentou notas com valor 0 ou 20. Para fins de comparação com a Competência estimada pelo modelo de nosso experimento, escalamos os valores da habilidade verdadeira, gerada pelo teste, para uma distribuição normal de características $N(\mu = 0, \sigma = 1)$, denotado pela variável theta (θ). A título de ilustração, a Figura 9a traz as médias de nota de redação por estado na avaliação de 2014, onde houve maior número de participantes de toda história do exame. A Figura 75b, mostra as médias nacionais da edição do Enem cujos vetores foram utilizados no modelo de predição de nota de avaliação discutido no Capítulo 5

2.3.2 Provas Objetivas

Juntamente dos microdados disponibilizados pelo INEP, uma planilha com os valores de referência de habilidades necessárias por item também é disposta. Nesta planilha constam códigos identificadores das provas. A seguir, listamos em tuplas: o código identificador do INEP, a área do conhecimento, quantidade de instâncias trabalhadas no Capítulo 5 desta pesquisa.

- 511 — Linguagem, Comunicação e suas Tecnologias — LC [3.860 de 983.302]
- 516 — Matemática, Informática e Suas Tecnologias — MT [2.721 de 930.661]
- 503 — Ciências Naturais — CN [3.922 de 929.506]
- 507 — Ciências Humanas — CH [3.860 de 983.328]

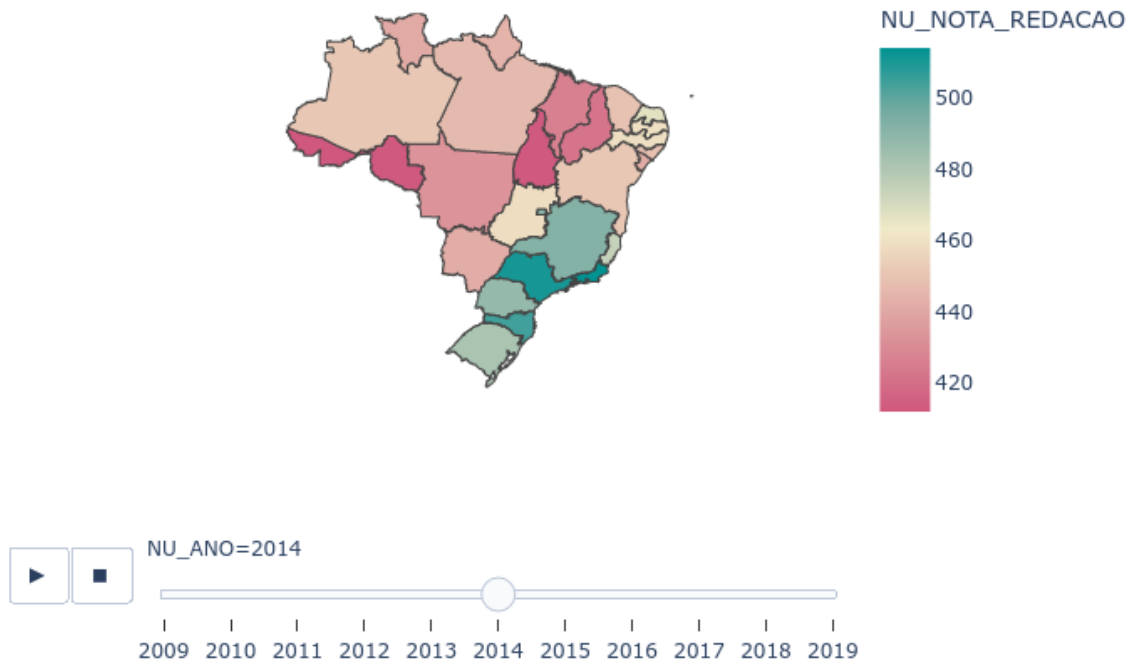
2.3.2.1 Conjunto do Estudo

Os experimentos relatados nos Capítulos 5 e 6, foram baseados na prova do Enem de 2019 que contou com 5.095.270 participantes. Os vetores foram selecionados descartando aqueles que continham valores nulos. Do restante do conjunto real, selecionamos aleatoriamente 30% dos dados, mas se mostrou um valor maior que o necessário em função da convergência dos valores de habilidade estimados e preditos. Isso ocorreu quando testamos com uma porção pequena de instâncias e incrementamos a quantidade nos valores de 100,

Figura 9 – Médias Nacionais das Notas de Redação por Estado

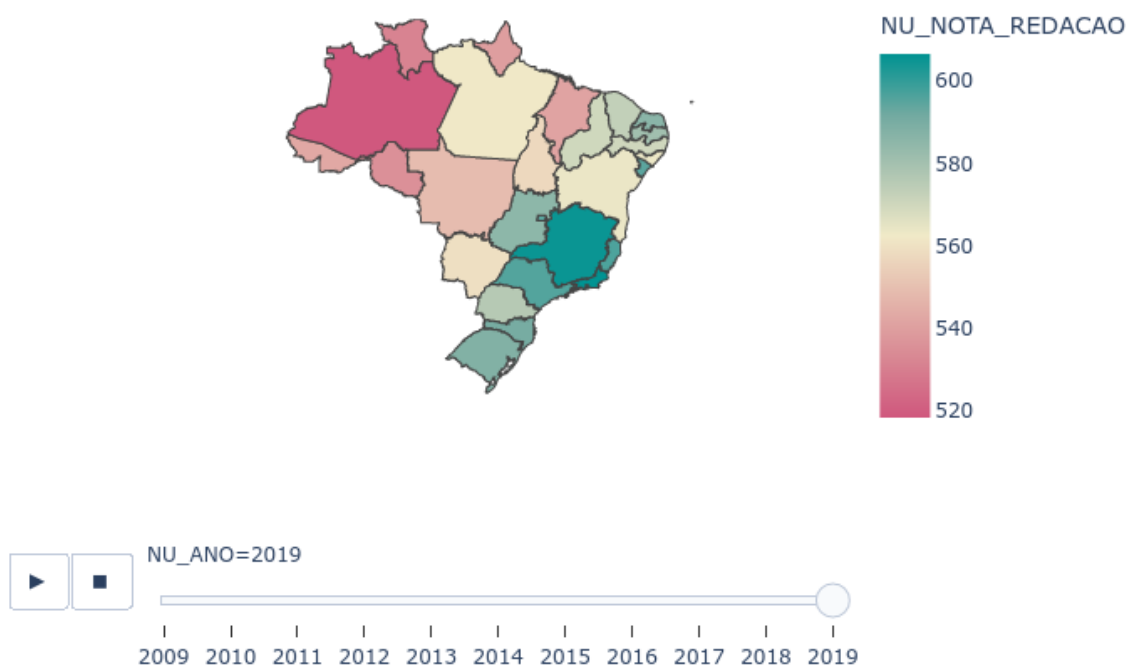
(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas



1000, 10000 e 100000. Com base no valor da nota por área, agrupamos os vetores em três grupos, relativos ao quartil inferior, intervalo interquartil e quartil superior. Destes, em igual quantidade, selecionamos aleatoriamente a quantidade enumeradas. A avaliação do CAT considerou sua formatação em termos de comprimento, variação da habilidade e grupo associado — todos discutidos em detalhes no Capítulo 6.

Ao empreender o processamento de linguagem natural, o critério de classificação foi a maior similaridade entre os textos descritivos dos Eixos Cognitivos e a descrição dos enunciados. Uma vez que os verbos dos eixos cognitivos seguem os objetivos educacionais propostos na Taxonomia de *Bloom*, a classificação proposta foi capaz de se ajustar a diferentes áreas do conhecimento. De mesmo modo, o critério de classificação utilizado para as competências das áreas de conhecimento foi o grau de similaridade textual entre as descrições disponibilizadas pelo INEP em seu conjunto de arquivos dos microdados e o enunciados das questões. O mesmo procedimento foi realizado com as especificações pedagógicas contantes nos microdados da edição de 2019.

2.3.3 Especificações Pedagógicas

Os microdados do Enem trazem consigo a documentação a respeito dos eixos cognitivos trabalhados na prova. A partir desses textos, é possível fazer associações com os enunciados dos itens de avaliação da prova, via NLP, e selecionar aqueles cujo conteúdo avaliativo seja pré-determinado, considerando o rol de opções que todos os itens de avaliação do Enem proporcionam. A seguir, é apresentada a lista de especificações, começando das competências gerais para as específicas:

- DL: *Domínio da Língua* — dominar a norma culta da Língua Portuguesa e usar as linguagens matemática, artística e científica e das línguas espanhola e inglesa.
- CF: *Construir e Formular* — construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas.
- SP: *Selecionar e Planejar* — selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema.
- CA: *Construir e Analisar* — relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.
- EP: *Elaborar e Propor* — recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural.

2.3.3.1 Competências e Referências de Linguagem e Comunicação

A lista de competências de aprendizagem de Linguagem e Comunicação informada pelo INEP totaliza 9 itens, exibidos a seguir:

- *CA-1: Aplicar as tecnologias da comunicação e da informação na escola, no trabalho e em outros contextos relevantes para sua vida*
- *CA-2: Conhecer e usar língua(s) estrangeira(s) moderna(s) como instrumento de acesso a informações e a outras culturas e grupos sociais*
- *CA-3: Compreender e usar a linguagem corporal como relevante para a própria vida, integradora social e formadora da identidade*
- *CA-4: Compreender a arte como saber cultural e estético gerador de significação e integrador da organização do mundo e da própria identidade*
- *CA-5: Analisar, interpretar e aplicar recursos expressivos das linguagens, relacionando textos com seus contextos, mediante a natureza, função, organização, estrutura das manifestações, de acordo com as condições de produção e recepção,*
- *CA-6: Compreender e usar os sistemas simbólicos das diferentes linguagens como meios de organização cognitiva da realidade pela constituição de significados, expressão, comunicação e informação.*
- *CA-7: Confrontar opiniões e pontos de vista sobre as diferentes linguagens e suas manifestações específicas*
- *CA-8: Compreender e usar a língua portuguesa como língua materna, geradora de significação e integradora da organização do mundo e da própria identidade.*
- *CA-9: Entender os princípios, a natureza, a função e o impacto das tecnologias da comunicação e da informação na sua vida pessoal e social, no desenvolvimento do conhecimento, associando-o aos conhecimentos científicos, às linguagens que lhes dão suporte, às demais tecnologias, aos processos de produção e aos problemas que se propõem solucionar*

A lista de referências conceituais de Linguagem e Comunicação informada pelo INEP totaliza oito tópicos mostrados a seguir:

- *LC-R1: estudo do texto as sequências discursivas e os gêneros textuais no sistema de comunicação e informação — modos de organização da composição textual; atividades de produção escrita e de leitura de textos gerados nas diferentes esferas sociais — públicas e privadas.*

- LC-R2: estudo das práticas corporais a linguagem corporal como integradora social e formadora de identidade — performance corporal e identidades juvenis; possibilidades de vivência crítica e emancipada do lazer; mitos e verdades sobre os corpos masculino e feminino na sociedade atual; exercício físico e saúde; o corpo e a expressão artística e cultural; o corpo no mundo dos símbolos e como produção da cultura; práticas corporais e autonomia; condicionamentos e esforços físicos; o esporte; a dança; as lutas; os jogos; as brincadeiras.
- LC-R3: produção e recepção de textos artísticos, interpretação e representação do mundo para o fortalecimento dos processos de identidade e cidadania — Artes Visuais: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade. Teatro: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Música: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Dança: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Conteúdos estruturantes das linguagens artísticas (Artes Visuais, Dança, Música, Teatro), elaborados a partir de suas estruturas morfológicas e sintáticas; inclusão, diversidade e multiculturalidade: a valorização da pluralidade expressada nas produções estéticas e artísticas das minorias sociais e dos portadores de necessidades especiais educacionais.
- LC-R4: estudo do texto literário relações entre produção literária e processo social, concepções artísticas, procedimentos de construção e recepção de textos — produção literária e processo social; processos de formação literária e de formação nacional; produção de textos literários, sua recepção e a constituição do patrimônio literário nacional; relações entre a dialética cosmopolitismo/localismo e a produção literária nacional; elementos de continuidade e ruptura entre os diversos momentos da literatura brasileira; associações entre concepções artísticas e procedimentos de construção do texto literário em seus gêneros (épico/narrativo, lírico e dramático) e formas diversas.; articulações entre os recursos expressivos e estruturais do texto literário e o processo social relacionado ao momento de sua produção; representação literária: natureza, função, organização e estrutura do texto literário; relações entre literatura, outras artes e outros saberes.
- LC-R5: estudo dos aspectos linguísticos em diferentes textos, recursos expressivos da língua, procedimentos de construção e recepção de textos — organização da macroestrutura semântica e a articulação entre ideias e proposições (relações lógico semânticas).
- LC-R6: estudo do texto argumentativo, seus gêneros e recursos linguísticos argumentação: tipo, gêneros e usos em língua portuguesa - formas de apresentação

de diferentes pontos de vista; organização e progressão textual; papéis sociais e comunicativos dos interlocutores, relação entre usos e propósitos comunicativos, função socio-comunicativa do gênero, aspectos da dimensão espaço-temporal em que se produz o texto.

- LC-R7: estudo dos aspectos linguísticos da língua portuguesa, usos da língua: norma culta e variação linguística — uso dos recursos linguísticos em relação ao contexto em que o texto é constituído: elementos de referência pessoal, temporal, espacial, registro linguístico, grau de formalidade, seleção lexical, tempos e modos verbais; uso dos recursos linguísticos em processo de coesão textual: elementos de articulação das sequências dos textos ou à construção da micro estrutura do texto.
- LC-R8: estudo dos gêneros digitais, tecnologia da comunicação e informação: impacto e função social — o texto literário típico da cultura de massa: o suporte textual em gêneros digitais; a caracterização dos interlocutores na comunicação tecnológica; os recursos linguísticos e os gêneros digitais; a função social das novas tecnologias.

2.3.3.2 Competências e Referências de Matemática

A lista de competências de aprendizagem de Matemática informada pelo INEP possui seis tópicos, mostrados a seguir:

- *CA-1: Construir significados para os números naturais, inteiros, racionais e reais*
- *CA-2: Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela*
- *CA-3: Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano*
- *CA-4: Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano*
- *CA-5: Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas,*
- *CA-6: Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.*

A lista de referências conceituais de Matemática informada pelo INEP totaliza cinco tópicos, sendo eles:

- MT-R1: Conhecimentos numéricos, operações em conjuntos numéricos (naturais, inteiros, racionais e reais), desigualdades, divisibilidade, fatoração, razões e proporções, porcentagem e juros, relações de dependência entre grandezas, sequências e progressões, princípios de contagem.
- MT-R2: Conhecimentos geométricos características das figuras geométricas planas e espaciais; grandezas, unidades de medida e escalas; comprimentos, áreas e volumes; ângulos; posições de retas; simetrias de figuras planas ou espaciais; congruência e semelhança de triângulos; teorema de Tales; relações métricas nos triângulos; circunferências; trigonometria do ângulo agudo.
- MT-R3: Conhecimentos de estatística e probabilidade, representação e análise de dados; medidas de tendência central (médias, moda e mediana); desvios e variância; noções de probabilidade.
- MT-R4: Conhecimentos algébricos gráficos e funções; funções algébricas dos 1.º e do 2.º graus, polinomiais, racionais, exponenciais e logarítmicas; equações e inequações; relações no ciclo trigonométrico e funções trigonométricas.
- MT-R5: Conhecimentos algébricos/geométricos, plano cartesiano; retas; circunferências; paralelismo e perpendicularidade, sistemas de equações.

2.3.3.3 Competências e Referências de Ciências Naturais

As competências são estão agrupadas em oito tópicos, como visto a seguir:

- CA-1: Compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade.
- CA-2: Identificar a presença e aplicar as tecnologias associadas.
- CA-3: Associar intervenções resultantes em degradação ou conservação ambiental a processos produtivos e sociais e a instrumentos ou ações científico-tecnológicos.
- CA-4: Compreender interações entre organismos e ambiente, em particular aquelas relacionadas à saúde humana, relacionando conhecimentos científicos, aspectos culturais e características individuais,
- CA-5: Entender métodos e procedimentos próprios das ciências naturais e aplicá-los em diferentes contextos.
- CA-6: Apropriar-se de conhecimentos da física para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

- CA-7: Apropriar-se de conhecimentos da química para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.
- CA-8: Apropriar-se de conhecimentos da biologia para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

E a lista de referências para Física possui vinte e um tópicos, sendo eles:

- CN-R1-F1: conhecimentos básicos e fundamentais Noções de ordem de grandeza.
- CN-R1-F2: conhecimentos básicos e fundamentais Notação Científica. Sistema Internacional de Unidades.
- CN-R1-F3: conhecimentos básicos e fundamentais Metodologia de investigação: a procura de regularidades e de sinais na interpretação física do mundo. Observações e mensurações: representação de grandezas físicas como grandezas mensuráveis.
- CN-R1-F4: conhecimentos básicos e fundamentais Ferramentas básicas: gráficos e vetores. Conceituação de grandezas vetoriais e escalares. Operações básicas com vetores.
- CN-R2-F8: O movimento, o equilíbrio e a descoberta de leis físicas Grandezas fundamentais da mecânica: tempo, espaço, velocidade e aceleração. Relação histórica entre força e movimento. Descrições do movimento e sua interpretação: quantificação do movimento e sua descrição matemática e gráfica. Casos especiais de movimentos e suas regularidades observáveis.
- CN-R2-F9: O movimento, o equilíbrio e a descoberta de leis físicas Conceito de inércia. Noção de sistemas de referência inerciais e não inerciais. Noção dinâmica de massa e quantidade de movimento (momento linear). Força e variação da quantidade de movimento. Leis de Newton.
- CN-R2-F10: O movimento, o equilíbrio e a descoberta de leis físicas, Centro de massa e a ideia de ponto material. Conceito de forças externas e internas. Lei da conservação da quantidade de movimento (momento linear) e teorema do impulso.
- CN-R2-F11: O movimento, o equilíbrio e a descoberta de leis físicas Momento de uma força (torque). Condições de equilíbrio estático de ponto material e de corpos rígidos. Força de atrito, força peso, força normal de contato e tração. Diagramas de forças. Identificação das forças que atuam nos movimentos circulares. Noção de força centrípeta e sua quantificação.
- CN-R2-F12: O movimento, o equilíbrio e a descoberta de leis físicas A hidrostática: aspectos históricos e variáveis relevantes. Empuxo. Princípios de Pascal, Arquimedes e Stevin: condições de flutuação, relação entre diferença de nível e pressão hidrostática.

- CN-R3-F13: Energia, trabalho e potência Conceituação de trabalho, energia e potência. Conceito de energia potencial e de energia cinética. Conservação de energia mecânica e dissipação de energia. Trabalho da força gravitacional e energia potencial gravitacional. Forças conservativas e dissipativas.
- CN-R4-F14: A Mecânica e o funcionamento do Universo Força peso. Aceleração gravitacional. Lei da Gravitação Universal. Leis de Kepler. Movimentos de corpos celestes. Influência na Terra: marés e variações climáticas. Concepções históricas sobre a origem do universo e sua evolução.
- CN-R5-F15: Fenômenos Elétricos e Magnéticos Carga elétrica e corrente elétrica. Lei de Coulomb. Campo elétrico e potencial elétrico. Linhas de campo. Superfícies equipotenciais. Poder das pontas. Blindagem.
- CN-R5-F16: Fenômenos Elétricos e Magnéticos Capacitores. Efeito Joule. Lei de Ohm. Resistência elétrica e resistividade. Relações entre grandezas elétricas: tensão, corrente, potência e energia. Circuitos elétricos simples. Correntes contínua e alternada. Medidores elétricos. Representação gráfica de circuitos. Símbolos convencionais. Potência e consumo de energia em dispositivos elétricos.
- CN-R5-F17: Fenômenos Elétricos e Magnéticos Campo magnético. Ímãs permanentes. Linhas de campo magnético. Campo magnético terrestre.
- CN-R6-F18: Oscilações, ondas, óptica e radiação, Feixes e frentes de ondas. Reflexão e refração. Óptica geométrica: lentes e espelhos. Formação de imagens. Instrumentos ópticos simples.
- CN-R6-F19: Oscilações, ondas, óptica e radiação Fenômenos ondulatórios. Pulsos e ondas. Período, frequência, ciclo. Propagação: relação entre velocidade, frequência e comprimento de onda. Ondas em diferentes meios de propagação.
- CN-R7-F20: O calor e os fenômenos térmicos Conceitos de calor e de temperatura. Escalas termométricas. Transferência de calor e equilíbrio térmico. Capacidade calorífica e calor específico. Condução do calor. Dilatação térmica. Mudanças de estado físico e calor latente de transformação.
- CN-R7-F21: O calor e os fenômenos térmicos Comportamento de Gases ideais. Máquinas térmicas. Ciclo de Carnot. Leis da Termodinâmica. Aplicações e fenômenos térmicos de uso cotidiano. Compreensão de fenômenos climáticos relacionados ao ciclo da água.

Para Química, são dezoito tópicos de referências, listados a seguir:

- CN-R1-Q1: transformações químicas. Evidências de transformações químicas. Interpretando transformações químicas.
- CN-R1-Q2: transformações químicas em Sistemas Gasosos: Lei dos gases. Equação geral dos gases ideais. Princípio de Avogadro, conceito de molécula; massa molar, volume molar dos gases. Teoria cinética dos gases. Misturas gasosas.
- CN-R1-Q3: Transformações Químicas Modelo corpuscular da matéria. Modelo atômico de Dalton. Natureza elétrica da matéria: Modelo Atômico de Thomson, Rutherford, Rutherford-Bohr. Átomos e sua estrutura. Número atômico, número de massa, isótopos, massa atômica. Elementos químicos e Tabela Periódica. Reações químicas.
- CN-R2-Q4: representação das transformações químicas Fórmulas químicas. Balanceamento de equações químicas. Aspectos quantitativos das transformações químicas. Leis ponderais das reações químicas. Determinação de fórmulas químicas.
- CN-R2-Q5: representação das transformações químicas Grandezas Químicas: massa, volume, mol, massa molar, constante de Avogadro. Cálculos estequiométricos.
- CN-R3-Q6: materiais, suas propriedades e usos - Propriedades de materiais. Estados físicos de materiais. Mudanças de estado. Misturas: tipos e métodos de separação. Substâncias químicas: classificação e características gerais.
- CN-R3-Q7: Materiais, suas propriedades e usos, Metais e Ligas metálicas. Ferro, cobre e alumínio. Ligações metálicas. Substâncias iônicas: características e propriedades. Substâncias iônicas do grupo: cloreto, carbonato, nitrato e sulfato. Ligação iônica. Substâncias moleculares: características e propriedades. Substâncias moleculares: H₂, O₂, N₂, Cl₂, NH₃, H₂O, HCl, CH₄. Ligação Covalente. Polaridade de moléculas. Forças intermoleculares. Relação entre estruturas, propriedade e aplicação das substâncias.
- CN-R4-Q8: água. Ocorrência e importância na vida animal e vegetal. Ligação, estrutura e propriedades. Sistemas em Solução Aquosa: Soluções verdadeiras, soluções coloidais e suspensões. Solubilidade. Concentração das soluções. Aspectos qualitativos das propriedades coligativas das soluções.
- CN-R4-Q9: água. Ácidos, Bases, Sais e Óxidos: definição, classificação, propriedades, formulação e nomenclatura. Conceitos de ácidos e base. Principais propriedades dos ácidos e bases: indicadores, condutibilidade elétrica, reação com metais, reação de neutralização.
- CN-R5-Q10: Transformações Químicas e Energia Transformações químicas e energia calorífica. Calor de reação. Entalpia. Equações termoquímicas. Lei de Hess. Trans-

formações químicas e energia elétrica. Reação de oxirredução. Potenciais padrão de redução. Pilha. Eletrólise. Leis de Faraday. Transformações nucleares.

- CN-R5-Q11: Transformações Químicas e Energia Conceitos fundamentais da radioatividade. Reações de fissão e fusão nuclear. Desintegração radioativa e radioisótopos.
- CN-R6-Q12: Dinâmica das Transformações Químicas Transformações Químicas e velocidade. Velocidade de reação. Energia de ativação. Fatores que alteram a velocidade de reação: concentração, pressão, temperatura e catalisador.
- CN-R7-Q13: Transformação Química e Equilíbrio Caracterização do sistema em equilíbrio. Constante de equilíbrio. Produto iônico da água, equilíbrio ácido-base e pH. Solubilidade dos sais e hidrólise. Fatores que alteram o sistema em equilíbrio. Aplicação da velocidade e do equilíbrio químico no cotidiano.
- CN-R8-Q14: Compostos de Carbono Características gerais dos compostos orgânicos. Principais funções orgânicas. Estrutura e propriedades de Hidrocarbonetos. Estrutura e propriedades de compostos orgânicos oxigenados. Fermentação. Estrutura e propriedades de compostos orgânicos nitrogenados. Macromoléculas naturais e sintéticas.
- CN-R8-Q15: Compostos de Carbono Noções básicas sobre polímeros. Amido, glicogênio e celulose. Borracha natural e sintética. Polietileno, poliestireno, PVC, Teflon, náilon. Óleos e gorduras, sabões e detergentes sintéticos. Proteínas e enzimas.
- CN-R9-Q16: Relações da Química com as Tecnologias, a Sociedade e o Meio Ambiente, Química no cotidiano. Química na agricultura e na saúde. Química nos alimentos. Química e ambiente.
- CN-R9-Q17: Relações da Química com as Tecnologias, a Sociedade e o Meio Ambiente, Indústria Química: obtenção e utilização do cloro, hidróxido de sódio, ácido sulfúrico, amônia e ácido nítrico. Mineração e Metalurgia. Poluição e tratamento de água. Poluição atmosférica. Contaminação e proteção do ambiente.
- CN-R10-Q18: energias químicas no cotidiano. Petróleo, gás natural e carvão. Madeira e hulha. Biomassa. Biocombustíveis. Impactos ambientais de combustíveis fósseis. Energia nuclear. Lixo atômico. Vantagens e desvantagens do uso de energia nuclear.

Por fim, para Biologia, o total de treze referências, sendo elas:

- CN-R1-B1: moléculas, células e tecidos. Estrutura e fisiologia celular: membrana, citoplasma e núcleo. Divisão celular. Aspectos bioquímicos das estruturas celulares. Aspectos gerais do metabolismo celular. Metabolismo energético: fotossíntese e

respiração. Codificação da informação genética. Síntese proteica. Diferenciação celular. Principais tecidos animais e vegetais.

- CN-R1-B2: moléculas, células e tecidos Origem e evolução das células. Noções sobre células-tronco, clonagem e tecnologia do DNA recombinante. Aplicações de biotecnologia na produção de alimentos, fármacos e componentes biológicos.
- CN-R1-B3: moléculas, células e tecidos Aplicações de tecnologias relacionadas ao DNA a investigações científicas, determinação da paternidade, investigação criminal e identificação de indivíduos. Aspectos éticos relacionados ao desenvolvimento biotecnológico. Biotecnologia e sustentabilidade.
- CN-R2-B4: hereditariedade e diversidade da vida. Princípios básicos que regem a transmissão de características hereditárias. Concepções pré-mendelianas sobre a hereditariedade. Aspectos genéticos do funcionamento do corpo humano. Antígenos e anticorpos. Grupos sanguíneos, transplantes e doenças autoimunes. Neoplasias e a influência de fatores ambientais.
- CN-R2-B5: hereditariedade e diversidade da vida. Mutações gênicas e cromossômicas. Aconselhamento genético. Fundamentos genéticos da evolução. Aspectos genéticos da formação e manutenção da diversidade biológica.
- CN-R3-B6: identidade dos seres vivos Níveis de organização dos seres vivos. Vírus, procariontes e eucariontes. Autótrofos e heterótrofos. Seres unicelulares e pluricelulares.
- CN-R3-B7: identidade dos seres vivos. Sistemática e as grandes linhas da evolução dos seres vivos. Tipos de ciclo de vida. Evolução e padrões anatômicos e fisiológicos observados nos seres vivos. Funções vitais dos seres vivos e sua relação com a adaptação desses organismos a diferentes ambientes.
- CN-R3-B8: identidade dos seres vivos. Embriologia, anatomia e fisiologia humana. Evolução humana. Biotecnologia e sistemática.
- CN-R4-B9: ecologia e ciências ambientais. Ecossistemas. Fatores bióticos e abióticos. Habitat e nicho ecológico. A comunidade biológica: teia alimentar, sucessão e comunidade clímax. Dinâmica de populações. Interações entre os seres vivos. Ciclos biogeoquímicos. Fluxo de energia no ecossistema. Biogeografia.
- CN-R4-B10: Ecologia e ciências ambientais Biomas brasileiros. Exploração e uso de recursos naturais. Problemas ambientais: mudanças climáticas, efeito estufa; desmatamento; erosão; poluição da água, do solo e do ar. Conservação e recuperação de ecossistemas. Conservação da biodiversidade.

- CN-R4-B11: Ecologia e ciências ambientais. Tecnologias ambientais. Noções de saneamento básico. Noções de legislação ambiental: água, florestas, unidades de conservação; biodiversidade.
- CN-R5-B12: Origem e evolução da vida A biologia como ciência: história, métodos, técnicas e experimentação. Hipóteses sobre a origem do Universo, da Terra e dos seres vivos. Teorias de evolução. Explicações pré-darwinistas para a modificação das espécies. A teoria evolutiva de Charles Darwin. Teoria sintética da evolução. Seleção artificial e seu impacto sobre ambientes naturais e sobre populações humanas.
- CN-R6-B13: Qualidade de vida das populações humanas Aspectos biológicos da pobreza e do desenvolvimento humano. Indicadores sociais, ambientais e econômicos. Índice de desenvolvimento humano. Principais doenças que afetam a população brasileira: caracterização, prevenção e profilaxia. Noções de primeiros socorros. Doenças sexualmente transmissíveis. Aspectos sociais da biologia: uso indevido de drogas; gravidez na adolescência; obesidade. Violência e segurança pública. Exercícios físicos e vida saudável. Aspectos biológicos do desenvolvimento sustentável. Legislação e cidadania.

2.3.3.4 Competências e Referências de Ciências Humanas

As competências são estão agrupadas em seis tópicos e as referências também em mesmo número.

- CA-1: Compreender os elementos culturais que constituem as identidades.
- CA-2: Compreender as transformações dos espaços geográficos como produto das relações socioeconômicas e culturais de poder.
- CA-3: Compreender a produção e o papel histórico das instituições sociais, políticas e econômicas, associando-as aos diferentes grupos, conflitos e movimentos sociais.
- CA-4: Entender as transformações técnicas e tecnológicas e seu impacto nos processos de produção, no desenvolvimento do conhecimento e na vida social.
- CA-5: Utilizar os conhecimentos históricos para compreender e valorizar os fundamentos da cidadania e da democracia, favorecendo uma atuação consciente do indivíduo na sociedade.
- CA-6: Compreender a sociedade e a natureza, reconhecendo suas interações no espaço em diferentes contextos históricos e geográficos.

A lista de referências é:

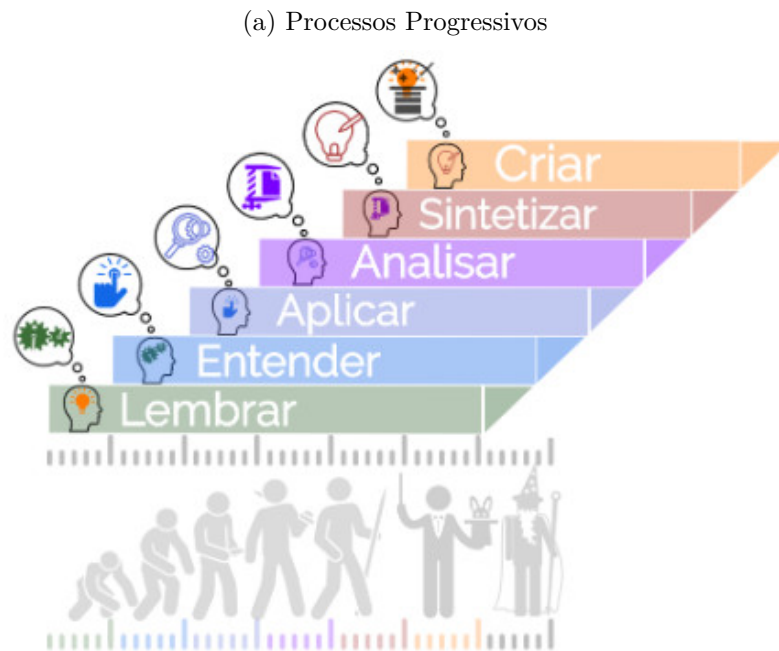
- CH-R1: Diversidade Cultural, Conflitos e Vida em Sociedade, Cultura Material e imaterial; patrimônio e diversidade cultural no Brasil. A Conquista da América. Conflitos entre europeus e indígenas na América colonial. A escravidão e formas de resistência indígena e africana na América. História cultural dos povos africanos. A luta dos negros no Brasil e o negro na formação da sociedade brasileira. o História dos povos indígenas e a formação sócio-cultural brasileira. Movimentos culturais no mundo ocidental e seus impactos na vida política e social.
- CH-R2: Formas de organização social, movimentos sociais, pensamento político e ação do Estado, Cidadania e democracia na Antiguidade; Estado e direitos do cidadão a partir da Idade Moderna; democracia direta, indireta e representativa. Revoluções sociais e políticas na Europa Moderna. Formação territorial brasileira; as regiões brasileiras; políticas de reordenamento territorial. As lutas pela conquista da independência política das colônias da América. Grupos sociais em conflito no Brasil imperial e a construção da nação. O desenvolvimento do pensamento liberal na sociedade capitalista e seus críticos nos séculos XIX e XX. Políticas de colonização, migração, imigração e emigração no Brasil nos séculos XIX e XX. A atuação dos grupos sociais e os grandes processos revolucionários do século XX: Revolução Bolchevique, Revolução Chinesa, Revolução Cubana. Geopolítica e conflitos entre os séculos XIX e XX: Imperialismo, a ocupação da Ásia e da África, as Guerras Mundiais e a Guerra Fria. Os sistemas totalitários na Europa do século XX: nazi-fascista, franquismo, salazarismo e stalinismo. Ditaduras políticas na América Latina: Estado Novo no Brasil e ditaduras na América. Conflitos político-culturais pós-Guerra Fria, reorganização política internacional e os organismos multilaterais nos séculos XX e XXI. A luta pela conquista de direitos pelos cidadãos: direitos civis, humanos, políticos e sociais. Direitos sociais nas constituições brasileiras. Políticas afirmativas. Vida urbana: redes e hierarquia nas cidades, pobreza e segregação espacial.
- CH-R3: Características e transformações das estruturas produtivas Diferentes formas de organização da produção: escravismo antigo, feudalismo, capitalismo, socialismo e suas diferentes experiências. Economia agro-exportadora brasileira: complexo açucareiro; a mineração no período colonial; a economia cafeeira; a borracha na Amazônia. Revolução Industrial: criação do sistema de fábrica na Europa e transformações no processo de produção. Formação do espaço urbano-industrial. Transformações na estrutura produtiva no século XX: o fordismo, o toyotismo, as novas técnicas de produção e seus impactos. A industrialização brasileira, a urbanização e as transformações sociais e trabalhistas. A globalização e as novas tecnologias de telecomunicação e suas consequências econômicas, políticas e sociais. Produção e transformação dos espaços agrários. Modernização da agricultura e estruturas agrárias tradicionais. O agronegócio, a agricultura familiar, os assalariados do campo e as lutas sociais no

campo. A relação campo-cidade.

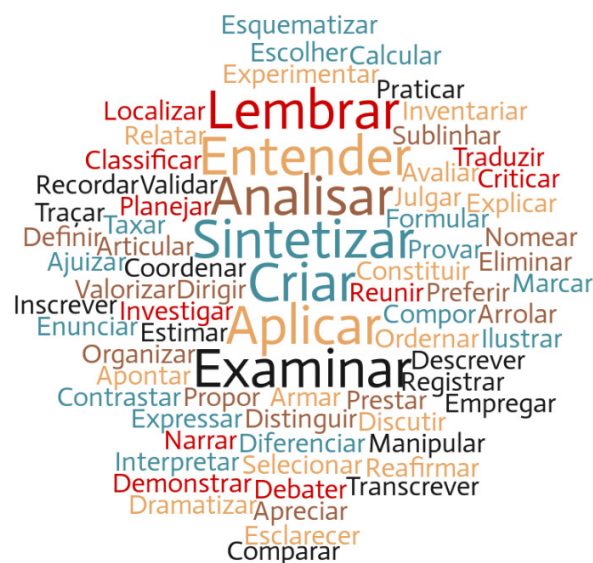
- CH-R4: Os domínios naturais e a relação do ser humano com o ambiente Relação homem-natureza, a apropriação dos recursos naturais pelas sociedades ao longo do tempo. Impacto ambiental das atividades econômicas no Brasil. Recursos minerais e energéticos: exploração e impactos. Recursos hídricos; bacias hidrográficas e seus aproveitamentos.
- CH-R5: Os domínios naturais e a relação do ser humano com o ambiente. As questões ambientais contemporâneas: mudança climática, ilhas de calor, efeito estufa, chuva ácida, a destruição da camada de ozônio. A nova ordem ambiental internacional; políticas territoriais ambientais; uso e conservação dos recursos naturais, unidades de conservação, corredores ecológicos, zoneamento ecológico e econômico. Origem e evolução do conceito de sustentabilidade. Estrutura interna da terra. Estruturas do solo e do relevo; agentes internos e externos modeladores do relevo. Situação geral da atmosfera e classificação climática. As características climáticas do território brasileiro. Os grandes domínios da vegetação no Brasil e no mundo.
- CH-R6: Representação espacial Projeções cartográficas; leitura de mapas temáticos, físicos e políticos; tecnologias modernas aplicadas à cartografia.

A Taxonomia de Bloom é um modelo de classificação de objetivos educacionais que descreve diferentes níveis de aprendizagem, desde o nível mais simples, como lembrar informações, até o nível mais complexo, como a criação de algo novo a partir do que foi aprendido (RAABE, 2018). Por outro lado, a Matriz de Referência do Enem é um documento que estabelece as habilidades e competências avaliadas na prova do Exame Nacional do Ensino Médio (Enem) no Brasil. Embora a Taxonomia de Bloom e a Matriz de Referência do Enem tenham finalidades diferentes, elas estão relacionadas de alguma forma. Isso ocorre porque a Matriz de Referência do Enem visa avaliar habilidades e competências relacionadas aos diferentes níveis da Taxonomia de Bloom. Por exemplo, as habilidades de lembrar informações e compreender textos são avaliadas no nível mais baixo da Taxonomia de Bloom, enquanto as habilidades de analisar, sintetizar e criar algo novo são avaliadas nos níveis mais altos da Taxonomia (ZUGE, 2020). Dessa forma, a Matriz de Referência do Enem utiliza a Taxonomia de Bloom como um referencial teórico para definir as habilidades e competências avaliadas na prova. Com isso, os itens da prova são elaborados para avaliar os diferentes níveis da Taxonomia de Bloom, permitindo assim uma avaliação mais ampla das habilidades e competências dos estudantes.

Figura 10 – Domínio Cognitivo. Fonte: Adaptado pelo Autor de (AMORIM, 2018).



(b) Termos Associados



2.4 Caracterização da Aprendizagem

A Taxonomia da Aprendizagem auxilia na denominação e descrição de processos de aprendizagem. A *Taxonomia da Aprendizagem de Bloom* (MARCHETI; BELHOT, 2010; OLIVEIRA; SILVA; NETO, 2020), descreve estágios de aprendizagem que podem ser usados como referência pedagógica, para representar níveis de conhecimento/aprendizagem. A referida taxonomia é organizada em três domínios, discutidos a seguir:

O *Domínio Cognitivo* reúne processos associados à psiquê como, por exemplo, processos mentais de identificação, compreensão, análise, crítica e criação. A Figura 10a

ilustra as etapas, do nível mais incipiente de conhecimento, na parte inferior dos degraus; até o nível de maior proficiência, localizado na parte superior. A Figura 10b reúne os termos associados a ações que contribuem para implementar os processos descritos nos degraus da Figura 10a. Assim, considerando avaliação da aprendizagem, questões que envolvem análise e síntese, demandam do aprendiz a identificação de conceitos e elementos principais, bem como a compreensão das eventuais relações existentes entre tais conceitos. São de ordem Cognitiva as atividades de fixação ou avaliativas que perguntem sobre o significado de determinados conceitos; a relação e diferenças entre conceitos, que busquem características de diferenciação e associação; que demandem construções, isto é, criações de construtos que sejam balizados pelos conceitos dos objetos de aprendizagem propostos.

O *Domínio Psico-afetivo* está caracterizado na Figura 11a, cujos degraus representam os processos psicológicos descritores de níveis de envolvimento, isto é, afeição do aprendiz em relação ao objeto de aprendizagem. A sequenciação de níveis de afetividade contribui para a compreensão de comportamentos dos aprendizes, bem como, para a consideração de ações que auxiliem no manejo das situações mapeadas. Caso inicialmente a *resposta afetiva* de um aprendiz a um determinado objeto de aprendizagem possa ser de *indiferença*. Enquanto é percebido tal conhecimento como a resposta a um problema que seja significativo para o aprendiz, a resposta afetiva pode progredir para o nível de *valorização* do saber. Nos níveis mais avançados, quando a *organização* afetiva do sujeito é tal que o manejo do conhecimento é considerado estruturado. Em última instância, o nível afetivo de *caraterização* é aquele onde o aluno passa a ser referência a respeito do conhecimento discutido, sendo, inclusive, passível de consulta pelos pares. A Figura 11b lista os termos associados a ações psico-afetivas. São exemplos atividades que discutam percepções e sensações a respeito dos objetos de aprendizagem, como, por exemplo: relato de experiência; estudo de casos; debate; listagem de influência do conhecimento nas práticas, dentre outras. A possibilidade de discutir os assuntos e associar as experiências, demonstram um caminho para materializar as razões da aprendizagem e suas vantagens.

Para representar o *Domínio Psicomotor*, a Figura 12a lista os processos relacionados à expressão física do conhecimento, em que se materializa as ações da consciência, com base na cognição e a afeição associada aos saberes. No nível inicial, os processos psicomotores visam a reprodução de práticas estabelecidas, com vistas a criar uma memória físico-muscular das ações elementares concernentes ao processo de aprendizagem. Ilustrativamente, considerando a aprendizagem de andar de bicicleta, ao nível de simulação, seria andar num equipamento com rodinhas ou numa bicicleta ergométrica. A partir da consciência dos movimentos de acionamento dos pedais e da memória física, progride-se para a reprodução em menor escala do ato, como pedalar num triciclo, no qual o conhecimento do equilíbrio sobre duas rodas ainda não é exigido. Sequencialmente o exercício do equilíbrio sobre duas rodas sugere a possibilidade de andar segurando a direção da bicicleta ou, até mesmo, sem segurar. São de ordem Psicomotora, atividades que peçam

Figura 11 – Domínio Psico-afetivo. Fonte: Adaptado pelo Autor de (AMORIM, 2018).

(a) Processos Progressivos



(b) Termos Associados

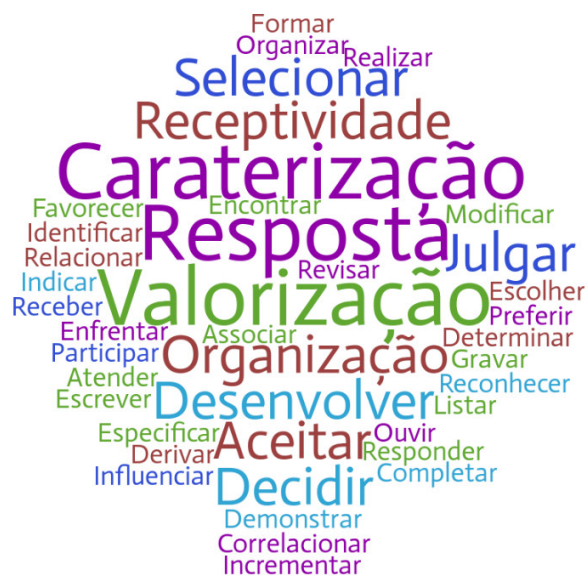
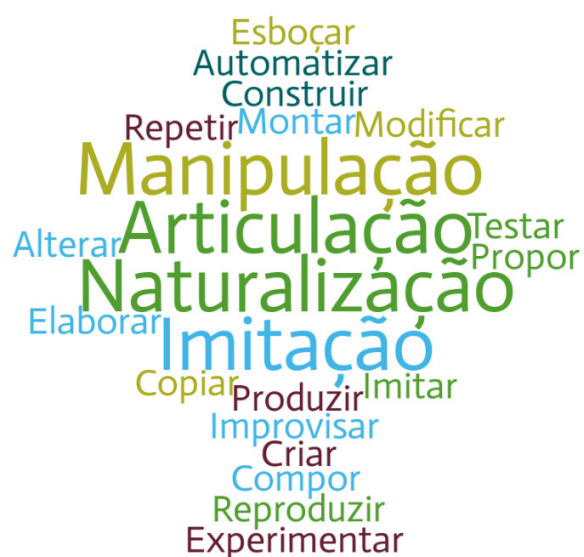


Figura 12 – Domínio Psicomotor. Fonte: Adaptado pelo Autor de (AMORIM, 2018).

(a) Processos Progressivos



(b) Termos Associados



para esboçar e reproduzir ações; e alterar, testar e improvisar ações. A Figura 12b lista os termos associados a ações psicomotoras. Com a prática recorrente, a expectativa é de naturalização das ações, isto é, realizando com tamanha consciência, que se assemelhe uma reação instintiva.

Sob o aspecto dos atores envolvidos em processos educacionais, é possível discutir procedimentos e informações que incrementem a qualidade da aprendizagem, com vistas a caracterizar a aprendizagem e propor sugestões de melhoria, corroborando com as conclusões de Fitzgerald (2021), De Jesus, Rodriguez e Costa Junior (2021). Alunos e professores são atores que ficam em evidência inicialmente, quando se considera contextos educacionais. Todavia, outros atores estão envolvidos, em maior ou menor proporção de forças intervenientes, na prática, educacional. Pais, gestores escolares, legisladores, comunidade e empresariado complementam o universo de ensino a qual cada unidade educacional pertence. Os dados gerados nos processos de ensino, são insumos para avaliação da qualidade alcançada (FITZGERALD, 2022). Por tais dados, os atores dialogam e estabelecem rumos para atingir as metas de aprendizagem (SILVA; BARBOSA; RIGO, 2021).

No Capítulo 3, uma abordagem de avaliação inteligente é demonstrada por meio de apoio computacional de um AVA e ferramentas para cálculo de estatísticas que ajudaram a representar melhor os resultados de uma avaliação por pares conjugada com uma autoavaliação.

3 Avaliação Inteligente de Aprendizagem Apoiada por Computação

No cotidiano educacional, de um lado, nota-se que os dados gerados pelos alunos são potencialmente informativos a respeito dos efeitos da aprendizagem (VILLEGAS-CH et al., 2018). De outro lado, a composição de estratégias didáticas individuais e provimento de *feedbacks*, são reconhecidas como tarefas árduas (CARLESS et al., 2011). Por isso, a utilização de ferramentas de apoio e o emprego de procedimentos estatísticos podem auxiliar o professor na apreciação dos dados gerados pelos alunos, a fim de melhorar a eficiência da experiência de aprendizagem (STRICKER; WEIBEL; WISSMATH, 2011; SCHANK; SAUNDERS, 2001). Neste capítulo, apresentamos procedimentos específicos para tentar evidenciar situações objetivas e subjetivas decorrentes do processo de ensino aprendizagem. Como exemplo, podemos citar a busca por identificar compreensões enviesadas dos saberes propostos e o mapeamento de alunos menos envolvidos nas atividades demandadas. Para tanto, nos valem do emprego de análise estatística das manifestações das percepções dos alunos. A proposta recorreu à avaliação nas modalidades de autoavaliação e avaliação por pares, inspirado nas obras de (OLIVEIRA et al., 2017; SPALENZA et al., 2018; OLIVEIRA, 2018; TSENG; TSAI, 2007; SITTHIWORACHART; JOY, 2008), em que os alunos são solicitados a expressar seu conhecimento a respeito de conteúdos ensinados. Como fatores positivos deste método, é possível citar: o estímulo à aprendizagem coletiva, por meio da avaliação por pares; a criação de um conjunto consistente de notas, em função do processamento estatístico das notas posteriormente; e o desenvolvimento do sentido crítico e vislumbre do horizonte cognitivo da turma.

Como resultado desta iniciativa, propomos uma visualização de fácil leitura, que pode ser integrada a um AVA, para analisar as relações entre as notas estipuladas pelo professor e as dos alunos. Observamos que as métricas sugeridas, coeficiente de similaridade de *pearson* e distância euclidiana, podem classificar os alunos de formas diferentes devido às suas semânticas e intervalos de medida próprios. Por isso, propusemos transformar a distância euclidiana, invertendo sua ordem de grandeza, passando a representar diferenciação para valores pequenos e proximidade em valores grandes; além disso, realizamos o procedimento de escalar os valores para o conjunto definido entre 0 e 1. Finalmente, foi realizada a agregação de cada métrica, para produzir uma única classificação. O escalonamento resultante do nosso método pode auxiliar o docente, seja selecionando possíveis candidatos para futuros monitores de turma ou focando em alunos que precisam estar mais envolvidos no processo de ensino-aprendizagem.

3.1 Trabalhos Relacionados

Um dos aspectos importantes do processo de ensino-aprendizagem é a fase de avaliação. Uma questão redundante encontrada por professores em sala de aula presencial ou virtual é a forma de lidar com muitos alunos (BRESLOW et al., 2013). Portanto, os professores costumam empregar autoavaliações para reduzir a carga de trabalho exigida durante a avaliação. Este fato desperta interesse dos pesquisadores refletido com o aumento substancial dos estudos sobre autoavaliações e revisões por pares na literatura. A utilização de diversas técnicas de avaliação amplia o escopo do processo de ensino-aprendizagem e permite individualizar o atendimento aos alunos (PERRENOUD, 1998).

A prática da avaliação pelos pares permite aos alunos descobrir novas estratégias para a resolução de problemas. Na autoavaliação, os alunos devem analisar suas próprias atividades, à luz do que foi explicado em aula, responsabilizando-se pelo seu envolvimento pessoal durante o processo de aprendizagem (SPILLER, 2012). Esta ação dá aos alunos a consciência sobre suas realizações. De acordo com Boud (2013), dois elementos são considerados essenciais para as autoavaliações: a construção de alguns padrões de desempenho esperados e novos valores construídos após a fase de correção. Naturalmente, uma questão importante relativa às estratégias de autoavaliação baseia-se na diferença entre as percepções dos alunos sobre seus próprios desempenhos e seus traços latentes reais (TOUSIGNANT; MARCHAIS, 2002). McMillan e Hearn (2008) argumentam que as autoavaliações podem promover motivações intrínsecas, esforços controlados, trabalhos orientados a objetivos e aprendizagens mais significativas. No trabalho de Nicol, Thomson e Breslin (2014), os autores afirmam que os *feedbacks* dos pares aumentam o engajamento no trabalho por meio de um processo introspectivo, dando maior autonomia aos alunos. Como visto em Spiller (2012), uma razão para usar uma revisão por pares é sua capacidade de apoiar a aprendizagem colaborativa por meio da troca de ideias sobre a percepção de um bom trabalho. Além disso, os alunos podem ajudar mutuamente, encontrando lacunas para compreender melhor o processo de aprendizagem. Enquanto as estratégias de autoavaliação promovem introspecções para os alunos em relação aos seus comportamentos, habilidades e práticas, as estratégias de avaliação por pares exigem uma análise das atividades dos pares.

No entanto, um problema que pode aparecer ao realizar as revisões por pares é a existência de possíveis desacordos entre as partes interessadas. Para resolver este problema, os autores do Wang et al. (2014) propõem uma abordagem para recompensar os alunos que dão notas justas e reforçar o apoio aos alunos que atribuem notas marginais. O método proposto por Shiba e Sugawara (2014) é baseado em redes de confiança geradas por avaliações mútuas realizadas pelos alunos. Na continuação do trabalho precursor de Stefani (1994), foram realizadas diversas experiências sobre autoavaliações e avaliações por pares. Sobre as autoavaliações, muitas vezes é realizada uma análise baseada em correlação entre

as notas dadas pelos próprios alunos e as notas dadas por um tutor ou professor (NGU; SHEPHERD; MAGIN, 1995; FALCHIKOV; GOLDFINCH, 2000; SITTHIWORACHART; JOY, 2004; CHALK; ADEBOYE, 2005; DENNY; LUXTON-REILLY; SIMON, 2009).

Frequentemente, as notas dadas por um professor podem ser potencialmente substituídas pelas notas produzidas pelos alunos, considerando uma leve tendência de pontuação mais baixa ou mais alta por parte dos alunos. Além disso, alguns modelos baseados em reconstrução de grau justo já foram contemplados na literatura (CHAPMAN; FIORE, 2000; DAVIES, 2000; CHO; SCHUNN, 2007; PARÉ; JOORDENS, 2008). Sobre a combinação de autoavaliações e avaliações por pares, duas estratégias são descritas por Oliveira e Spalenza (2017) para confrontar as notas dadas pelo professor e as notas sugeridas pelos alunos: os cálculos de distâncias e correlações. No artigo citado anteriormente, a correlação de Pearson (BUSSAB; MORETTIN, 2013) e a distância euclidiana foram usadas para apoiar a abordagem. Ao utilizar tal combinação, acreditamos produzir *feedbacks* mais rápidos para os alunos, dando-lhes mais cedo uma avaliação sobre seus próprios desempenhos durante o processo de aprendizagem, conforme preconizado por (PERRENOUD, 1998). Construímos nosso *framework* como uma extensão do Moodle e da arquitetura desenvolvida nos trabalhos de Oliveira et al. (2017), Spalenza et al. (2018), Oliveira (2018).

3.2 Arquitetura do Experimento

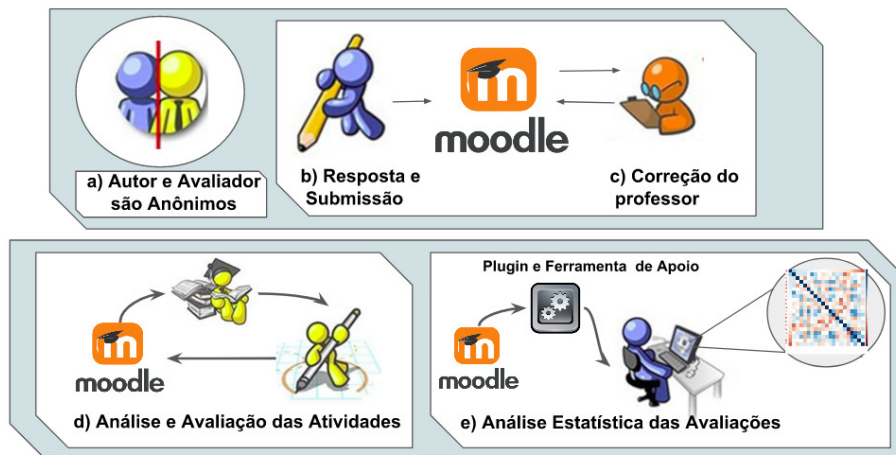
O objetivo de um AVA é fornecer suporte tecnológico durante o processo de ensino-aprendizagem — presencial ou à distância. Num AVA, é possível a organização de atividades, conteúdos e recursos. Um AVA como o Moodle oportuniza a integração de *softwares* de terceiros como um Sistema de Transferência de Dados criado e implementado por Spalenza et al. (2018). Por ele, é possível recuperar atividades realizadas pelos alunos no Moodle e enviar mensagens de retorno no campo de *feedback*, das atividades recuperadas. Para o nosso experimento, o Sistema de Transferência de Dados (SPALENZA et al., 2018) possibilitou extrair as respostas dos alunos e efetuar inserção de informação no Moodle. O experimento reportado foi realizado com alunos de graduação da área de humanas. A seguir, é apresentado, pormenorizadamente, o método de processamento utilizado e a saídas esperadas.

3.2.1 Método Proposto

A proposta de arquitetura implementada neste experimento foi uma extensão do trabalho de (OLIVEIRA; SPALENZA, 2017). Então, foram realizados os passos ilustrados na Figura 13, a qual é descrita a seguir.

- (a) **Impessoalização dos Dados:** com objetivo de assegurar maior isenção ao expe-

Figura 13 – Fluxo de Informação do Experimento. Fonte: (SILVA et al., 2020a)



rimento, foi realizada a impessoalização dos dados dos participantes, de modo a identificar cada resposta por um número identificador não associável ao indivíduo autor.

- (b) **Resposta e Submissão:** essa fase ocorreu no AVA de *Moodle*. Um grupo de vinte e três estudantes foram solicitados a responder atividades relativas a tópicos específicos do conteúdo de sala de aula.
- (c) **Correção do Professor:** O professor realizou a avaliação das atividades, estabelecendo notas para cada uma delas. Ao todo, sete atividades foram aplicadas em diferentes ocasiões, mas para o mesmo grupo de estudantes. Foram extraídos dois conjuntos de respostas para o experimento relatado.
- (d) **Avaliação em Pares e autoavaliação:** após terem sido expostos à correção das atividades supracitadas, os alunos foram solicitados a avaliar os seus pares, bem como proceder à sua autoavaliação definindo as suas próprias notas.
- (e) **Análise Estatística:** as notas dadas pelos alunos e pelo professor no processo de correção foram processadas estatisticamente, a fim de verificar a correlação e o distanciamento nos dados, com intuito de caracterizar numericamente as percepções encontradas.

Utilizando uma infraestrutura computacional, por meio das linguagens R e *Python*, as tarefas de correlação e classificação foram executadas. Assim, a análise partiu das três dimensões listadas abaixo:

- (1) **Correlação:** foi considerada a correlação entre as notas dadas pelo professor e as de cada aluno.

- (2) **Distanciamento:** foi considerada a distância entre as notas dadas pelo professor e as de cada aluno.
- (3) **Agregação de Correlação e Distanciamento:** A partir das métricas anteriores, foi gerada uma visualização mesclada permitindo a análise dos comportamentos dos alunos.

3.2.2 Processamento dos Dados

No pré-processamento dos dados, os alunos que não foram avaliados pelo professor foram removidos. O primeiro passo do processamento foi a importação das atividades do AVA, transformando as listas de notas pré-processadas em vetores n-dimensionais para cada atividade. Sua estrutura obedeceu a seguinte lógica: as listas de notas dadas pelo professor no vetor g^0 e as listas de notas produzidas por cada aluno (com um identificador j) por meio de avaliações por pares ou autoavaliações nos vetores g^j . Cada posição g_i^j nos vetores de notas refere-se à i -ésima série afetada para o aluno j após a avaliação.

A Figura 14a exibe os valores das notas apuradas em uma turma submetida ao processo de avaliação, autoavaliação e avaliação pelos pares. Os valores azulados, simbolizam notas próximas ao limite superior de 10 pontos. Os valores próximos a 5 pontos possuem coloração branca e os valores próximos a 0, estão na tonalidade vermelha. O eixo y possui as identificações impessoalizadas de 15 alunos, simbolizados pela sigla a_n , onde $n \in N$, tal que $N = \{1, 2, \dots, 15\}$. Ainda no eixo y, na última linha da Figura 14a, constam as notas atribuídas pelo professor. Adiante, passando para as etapas de análise dos dados por meio do cálculo das correlações e distâncias.

3.2.2.1 Cálculo das Correlações

Para medir os comportamentos dos alunos, uma correlação ρ é calculada usando o coeficiente de *pearson* descrito na Equação 3.1 (BUSSAB; MORETTIN, 2013). Essa métrica verifica a associação entre duas variáveis que podem ser interpretadas como notas dadas pelo professor e produzidas por cada aluno.

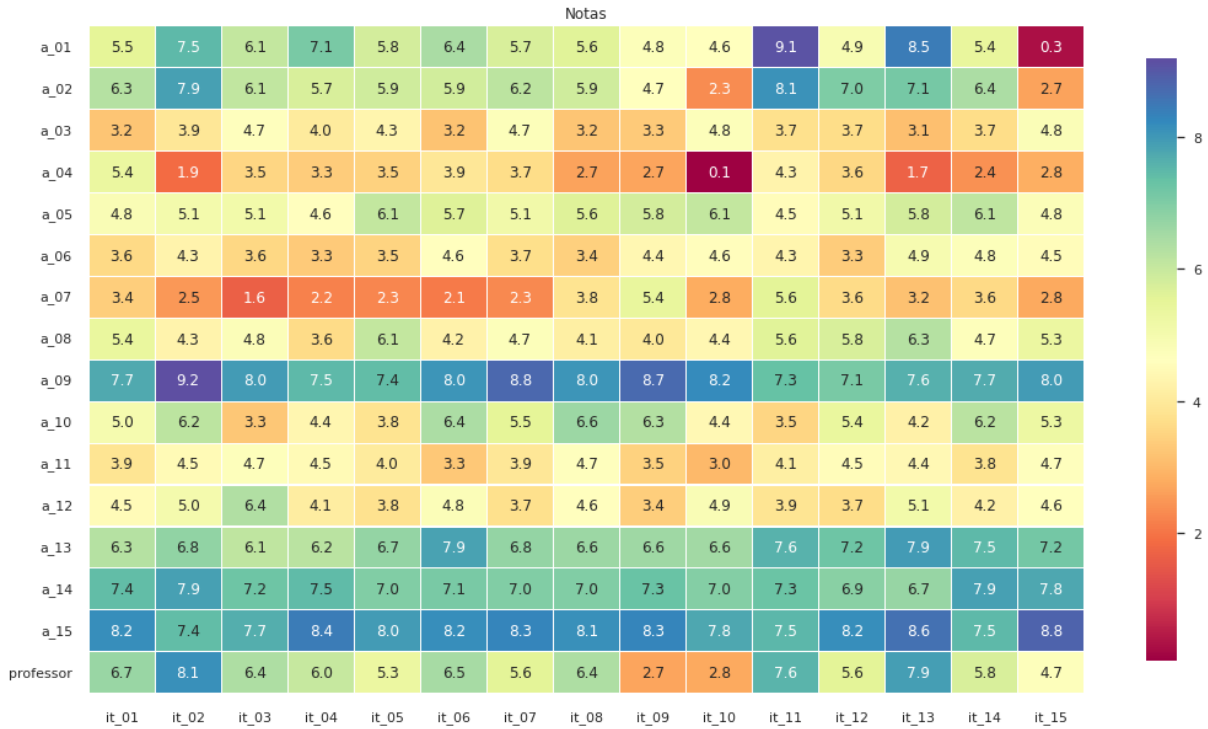
$$\rho_{i,j} = \frac{\sum_{k=1}^n (g_k^i - \bar{g}^i)(g_k^j - \bar{g}^j)}{\sqrt{\sum_{k=1}^n (g_k^i - \bar{g}^i)^2} \sqrt{\sum_{k=1}^n (g_k^j - \bar{g}^j)^2}} \quad (3.1)$$

onde:

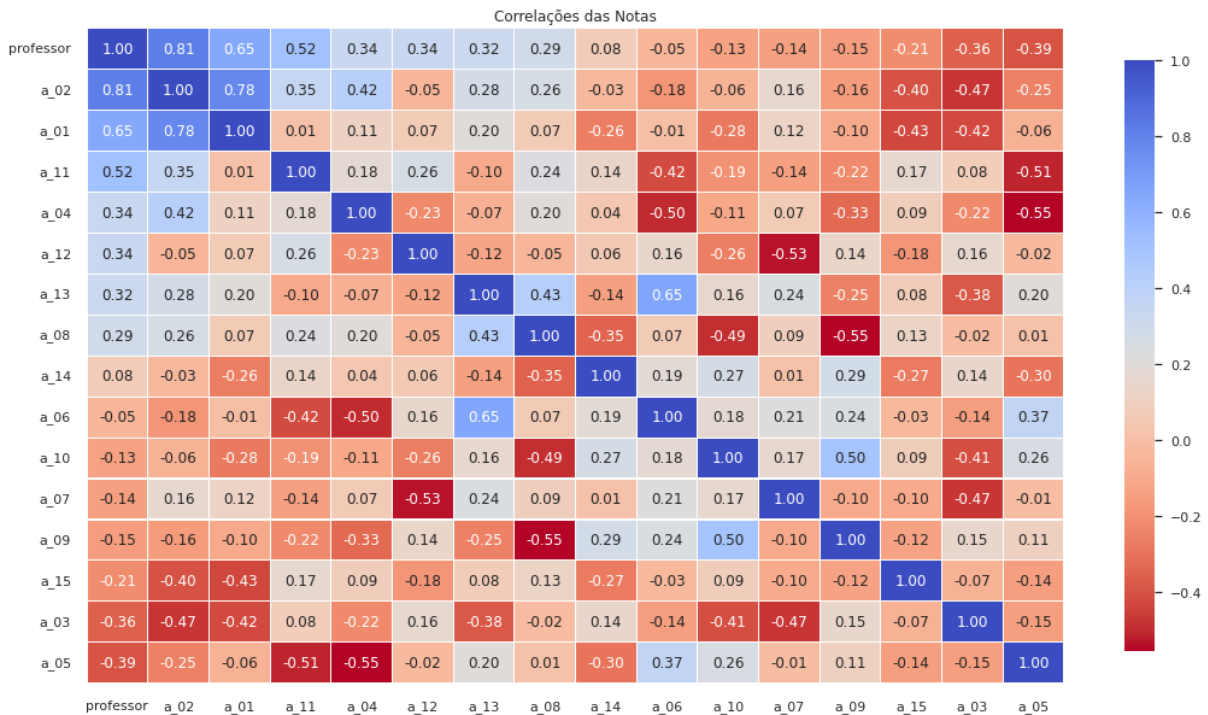
- n é o número de avaliados;
- g_k^i, g_k^j são as notas para cada atividade k ;
- $\bar{g}^k = \frac{1}{n} \sum_{i=1}^n g_i^k$ (média das notas).

Figura 14

(a) Notas Estipuladas



(b) Correlações de Pearson



A Figura 14b destaca as correlações variando de acordo com seguintes padrões: branco, quando os valores tendem a 0 (zero); azul, quando os valores tendem a 1 (um positivo); e vermelho, para valores tendendo a -1 (um negativo). Avaliando o grau de correlação, exibido na primeira linha da Figura 14b, notam-se valores acima de 0.5, os quais caracterizam forte correlação. No caso, são os alunos identificados por a_02 , a_01 e a_11 . Na mesma figura, as correlações inversas de maior intensidade ficaram por conta dos sujeitos a_05 e a_03 . O sujeito a_14 ficou com valor de correlação 0.08, simbolizando o menor nível de correlação entre todas as instâncias confrontadas com as notas estipuladas pelo professor.

Diante do interesse em combinar a métrica de correlação com distâncias euclidianas aferidas, os valores da correlação foram escalonados para o intervalo inferior 0 e o intervalo superior 1, seguindo a definição da Equação 3.2 (BUSSAB; MORETTIN, 2013).

$$S(\alpha_{i,j}) = \tau_{\mathbf{M}_\alpha} \frac{\alpha_{i,j} - \min(\mathbf{M}_\alpha)}{\max(\mathbf{M}_\alpha) - \min(\mathbf{M}_\alpha)} \quad (3.2)$$

Na Figura 15a, são exibidos os valores de correlação transformados, com a coloração representando azul para correlações diretas; e coloração vermelha, para correlações inversas. A matriz \mathbf{M}_ρ , resultante do cálculo das correlações, evidencia as relações entre as notas estipuladas pelo professor e as notas produzidas pelos alunos, ao realizarem suas autoavaliações e avaliação por pares.

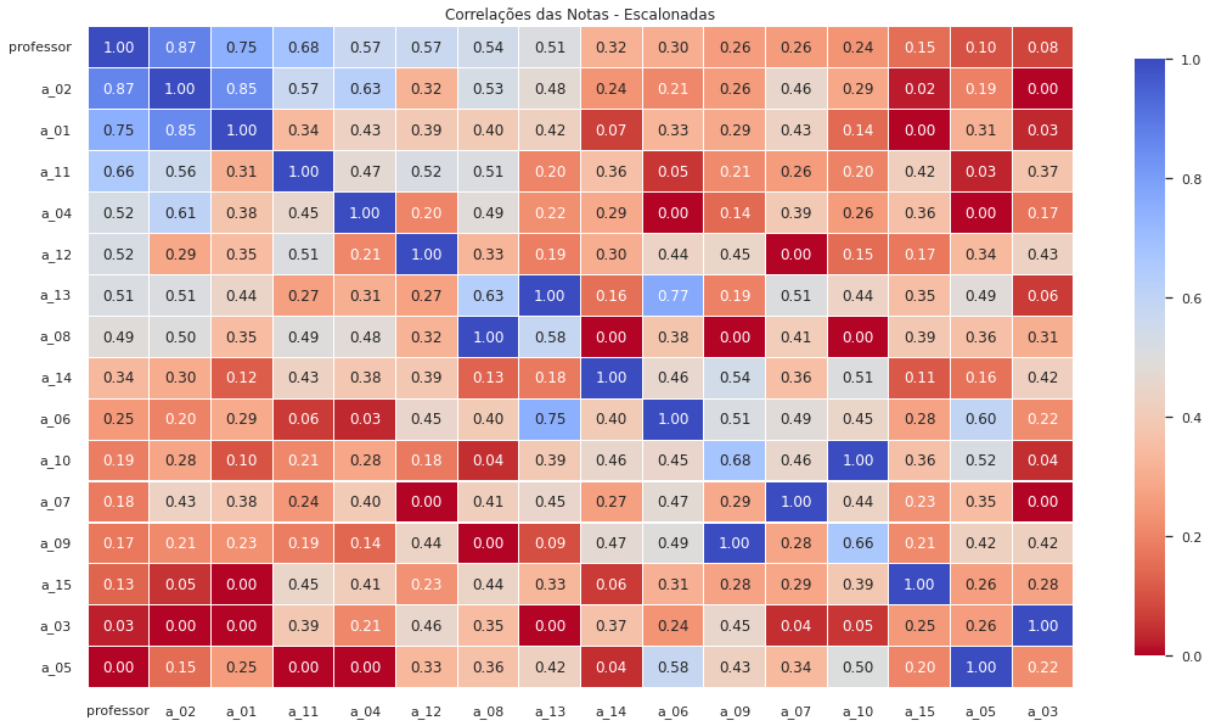
É visualmente evidente que existe um grau maior de correlação entre os pares de alunos a_06 , a_09 e a_10 e a_14 , variando entre 0.51 até 0.68; do que entre o professor e a_09 , cuja correlação escalonada ficou em 0.26. O padrão em vermelho, começando com o aluno a_14 e indo até o aluno a_03 , na região direita da matriz, mereceria uma investigação mais aprofundada (manual ou semi-automática) para verificar sua explicação, como, por exemplo, em razão de uma das seguintes causas: trabalho colaborativo, plágio, compreensão equivocada do assunto. Em face a situações em que os valores se apresentam com pouca intensidade devido a não participação dos alunos ou avaliação descompromissada, como visto na Figura 15a, caberia algum tipo de intervenção didática, para estimular os alunos a se envolverem mais no processo de ensino-aprendizagem.

3.2.2.2 Cálculo das Distâncias

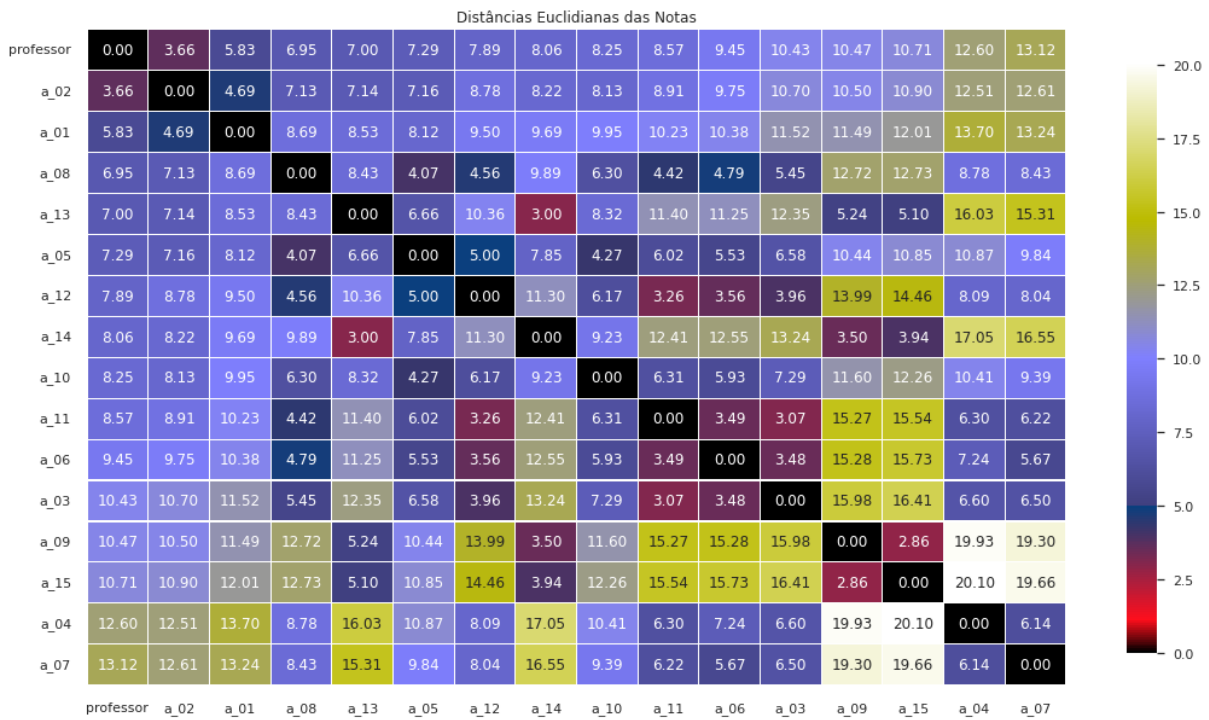
Dando sequência ao tratamento dos dados, tivemos em vista representar a outra possibilidade de medir o comportamento dos alunos, por meio da distância euclidiana entre vetores das notas avaliadas, como introduzida na Equação 3.3 (BUSSAB; MORETTIN, 2013).

Figura 15 – Caracterização pelas Correlações e Distâncias

(a) Correlação Escalonadas



(b) Distâncias



$$\delta_{i,j} = \sqrt{\sum_{k=1}^n (g_k^i - g_k^j)^2} \quad (3.3)$$

Nossa estrutura também gera uma matriz \mathbf{M}_δ resultante do cálculo das distâncias euclidianas entre as notas geradas nas avaliações. A Figura 15b, exibe os valores gerados. A coloração segue o seguinte princípio: valores menores nas tonalidades: pretas, vermelhas e azuis. Os valores próximos a 20 unidades de distância euclidiana, assumem cores entre o azul, amarelo e branco. As menores distâncias entre os alunos e professor, são dos sujeitos: a_02 , e a_01 . A maior distância, foi verificada para o indivíduo a_07 .

3.2.2.3 Classificação via Métricas Agregadas

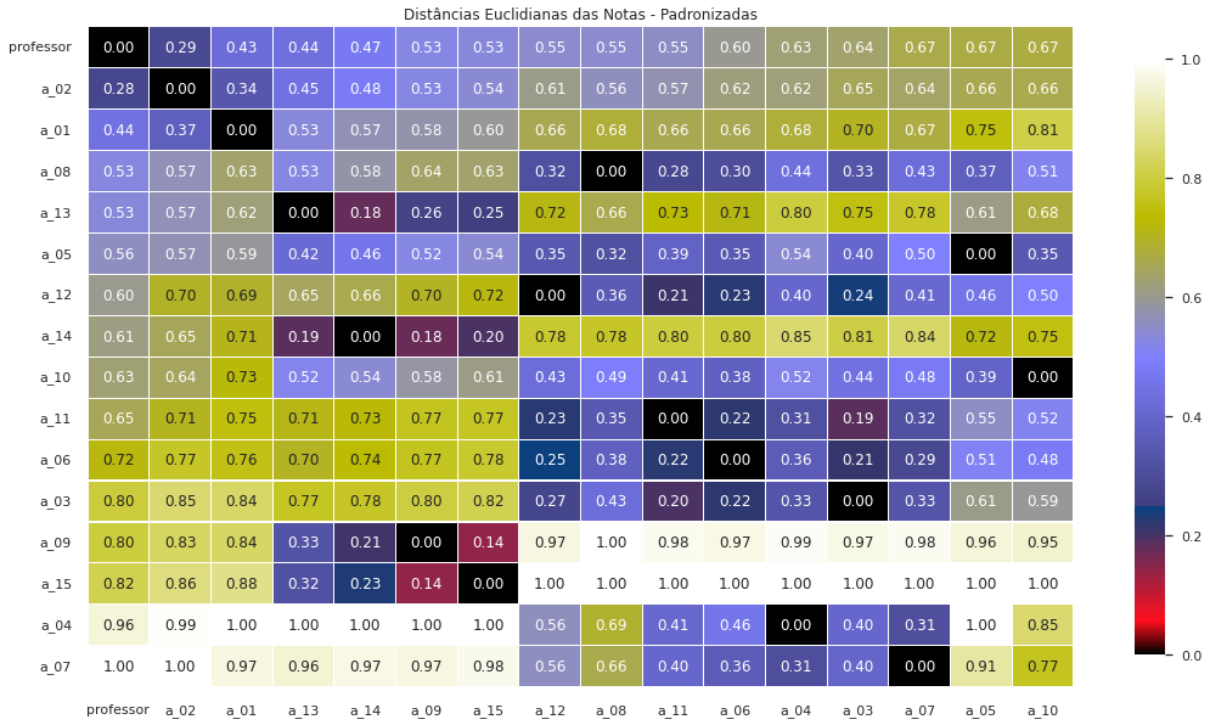
Embora seja possível optar por usar apenas uma métrica, seja com base em correlações, seja embasada em distâncias, o trabalho com ambas pode ser uma tarefa difícil para o professor. Podemos observar que classificações produzidos pelas duas matrizes das Figuras 15a e 16b são um pouco diferentes. Assim, nossa estrutura pode gerar uma matriz de mapa de calor mesclada, escalando os valores das correlações e distâncias em uma única métrica. A nova métrica tem um domínio definido entre 0 e 1. Os valores próximos a 1 representam os alunos de bom desempenho, enquanto os valores próximos a 0, evidenciam desempenho aquém do esperado. Para combinar as métricas de modo lógico, foi preciso calcular o inverso da métrica euclidiana ($\delta' = \frac{1}{\delta}$). O que resultou em um valor δ' , cuja semântica representa grande proximidade para números próximos a 1 (um) e baixa proximidade para valores próximos a 0 (zero). A normalização min-max altera os valores extremos de um conjunto de valores e organiza os outros valores no novo intervalo de domínio. Seja $\alpha_{i,j}$ uma métrica da matriz \mathbf{M}_α , $S(\alpha_{i,j})$ a métrica normalizada e $\tau_{\mathbf{M}_\alpha}$ (tem-se $0 \leq \tau_{\mathbf{M}_\alpha} \leq 1$) o limite superior ou limite desejado para \mathbf{M}_α : No nosso caso, como estamos interessados em mesclar apenas duas métricas, temos $\tau_{\mathbf{M}'_\delta} = 1 - \tau_{\mathbf{M}_\rho}$. Assim, definindo o limiar $\tau_{\mathbf{M}_\rho}$ entre 0 e 1, instanciamos dois pesos relativos às importâncias dadas para a correlação e a distância. A Figura 16a apresenta matriz de distâncias escalada. Enquanto a Figura 16b mostra a matriz de distância com os valores invertidos.

Por fim, a Figura 16b ilustra a composição final das métricas. Nela, a coloração verde, como mostra a primeira linha da segunda coluna, o valor 0.62 para o sujeito a_02 . As relações intermediárias, estão na cor vermelha. Com variação de 0.31 a 0.25. Os últimos quatro sujeito apresentaram a métrica mista inferior a 0.1. Seria esse o grupo de alerta, numa situação de avaliação diagnóstico.

A seguir, ilustramos os resultados por meio da classificação gerada com a métrica combinada, para os 4 melhores alunos, mostrando os parâmetros de regressão linear encontrados e as qualidades dos modelos avaliados.

Figura 16 – Caracterização Pelas Distâncias

(a) Padronizadas.



(b) Padronizadas Invertidas

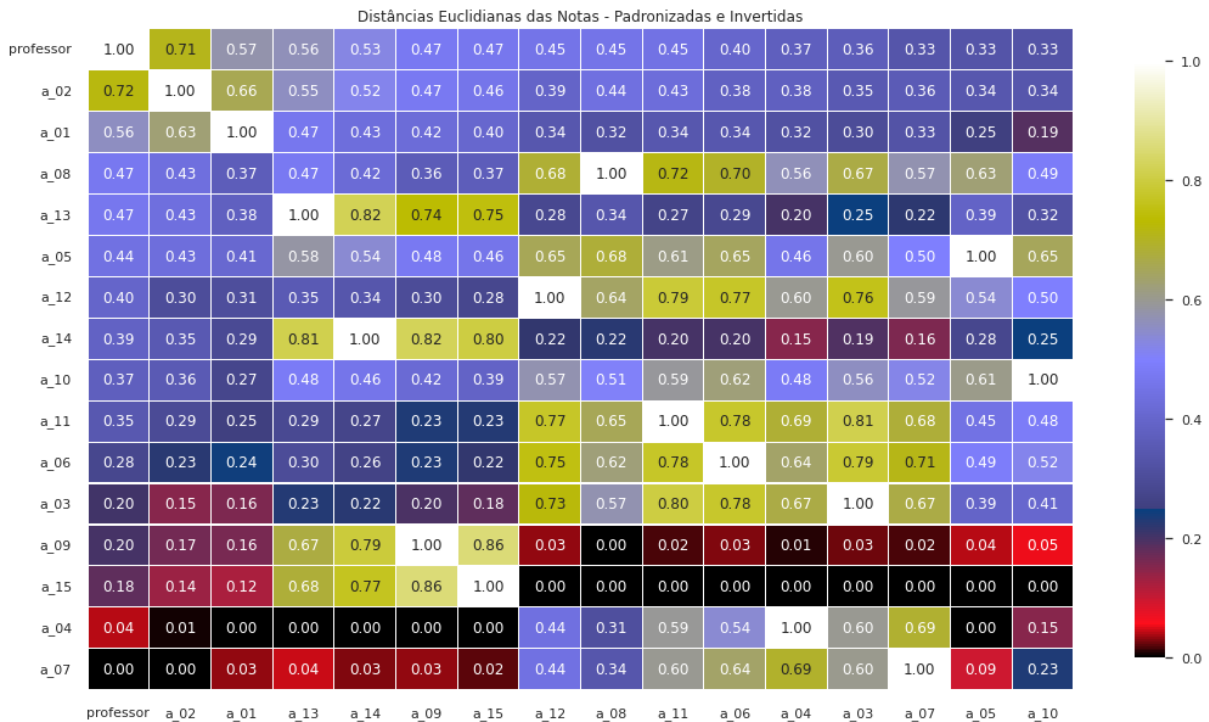
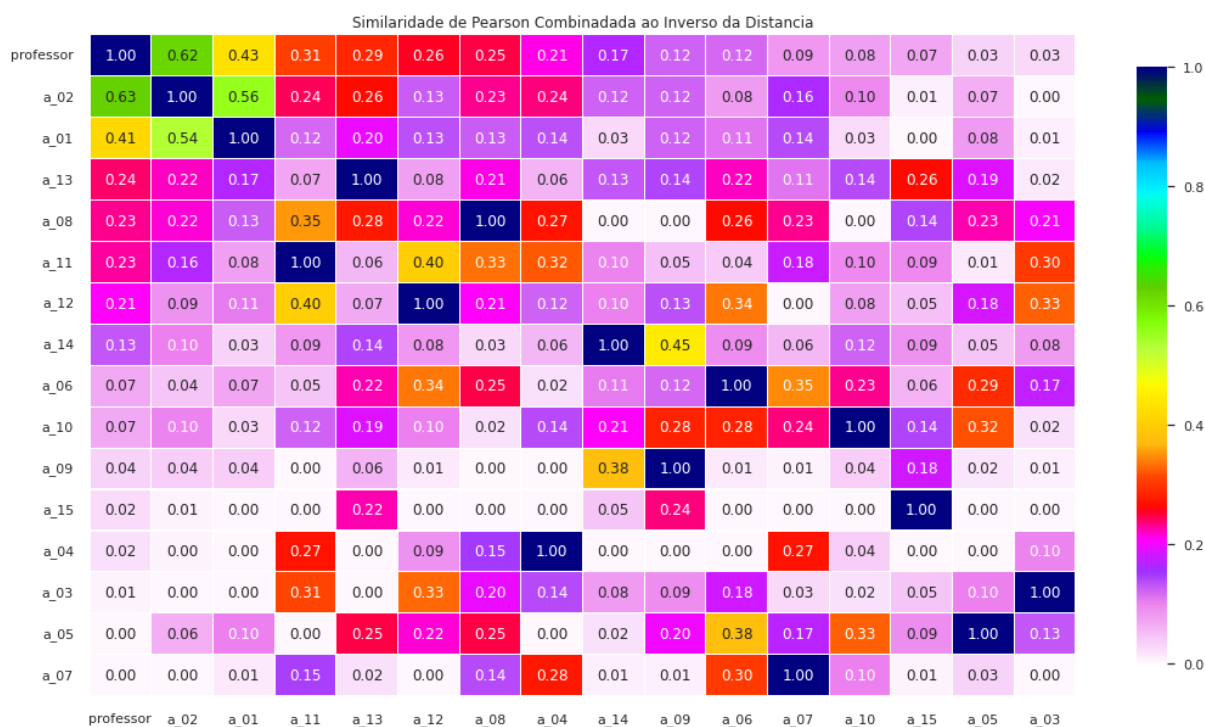


Figura 17 – Métricas Mescladas. Fonte: elaborado pelo autor.



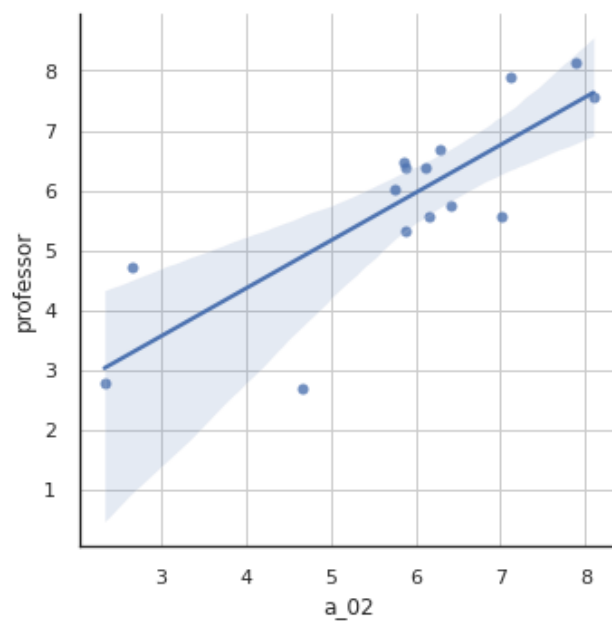
3.3 Resultados e Discussão

Por fim, a análise do modelo foi feita por meio de regressão linear. Assim, foram realizadas as regressões lineares para prever a nota que poderia ser atribuída pelo professor. A Tabela 1 apresenta os parâmetros das regressões dos três estudantes melhor classificados, a partir da métrica proposta. Pela lei dos grandes números, com um maior volume de atividades, as regressões ficariam melhor ajustadas. As métricas propostas têm a seguinte finalidade: o *T-estatístico* e o *p-value*, verificar a diferenciação da geração das notas, entre professor e aluno. E MSE (Mean Square Error) e (Root Mean Square Error), representam o valor residual dos modelo, médio e quadrático, respectivamente. O coeficiente angular, interceptor e R^2 , denotam as características dos modelos de regressão encontrados.

Nas Figuras 18a, 18b, 19a e 19b exibimos o gráfico de regressão dos alunos citados, os quais também podem ser observados no mapa de calor ilustrado pela Figura 17, na primeira linha, da esquerda para a direita, a partir da segunda até a quarta coluna. O parâmetro de regressão do sujeito *a_01* foi a que apresentou melhor nota inicial, com o interceptor de 2.88.

A utilização de ferramentas de regressão permitem ao docente estimar o sucesso do aluno e verificar a taxa de aprendizagem, tendo por base o coeficiente angular, a cada atividade realizada. Nesse sentido, o volume de dados histórico auxilia a composição de modelos melhores ajustados, incrementando a percepção do olhar docente sobre o processo de ensino-aprendizagem.

Figura 18 – Regressões do Primeiro e Segundo Colocados

(a) Melhor R^2 

(b) Menor Coeficiente Angular

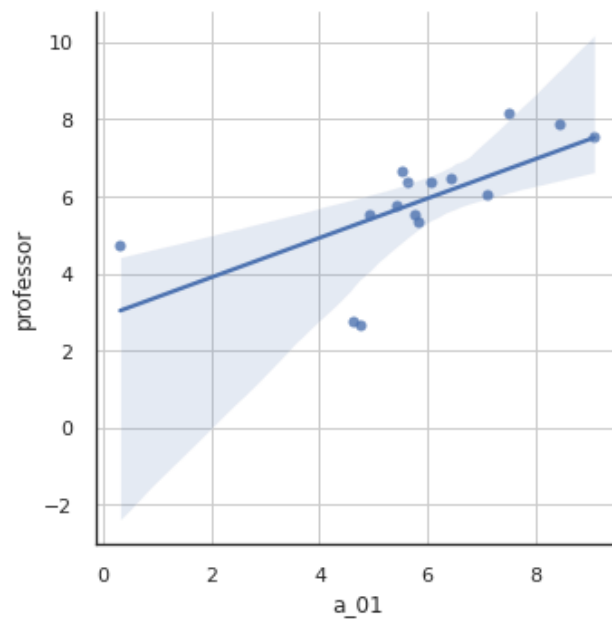
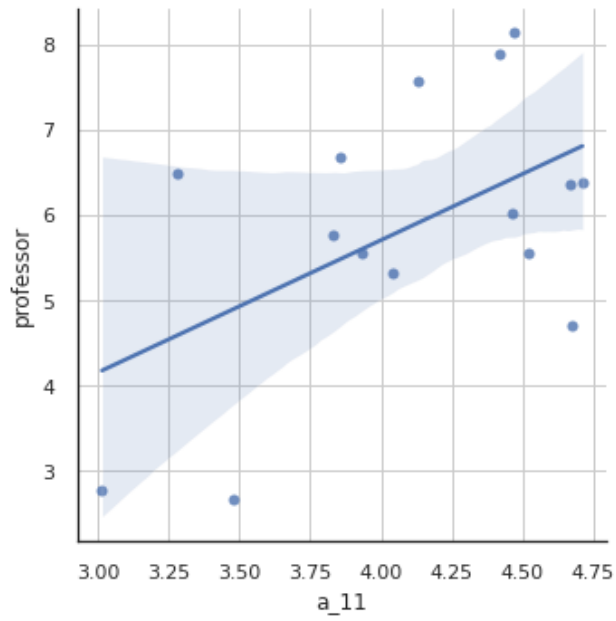


Figura 19 – Regressões do Terceiro e Quarto Colocados

(a) Menor Interceptor



(b) Pior Correlação

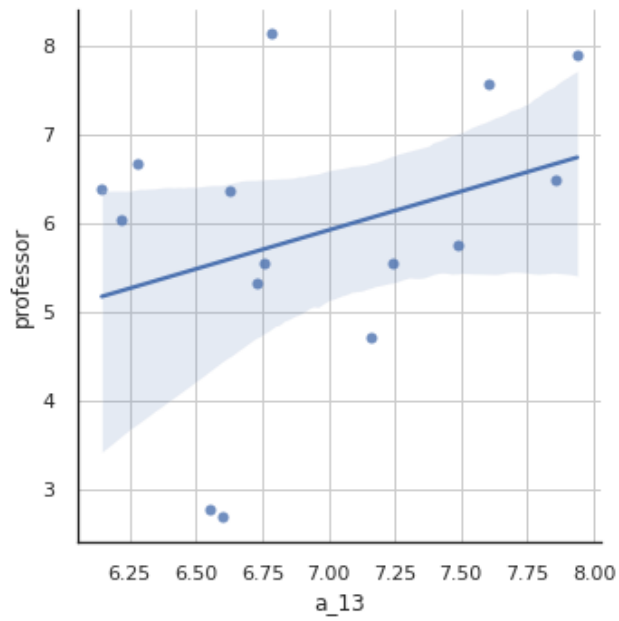


Tabela 1 – Parâmetros de Regressão. Fonte: elaborado pelo autor.

aluno	T-estatístico	p-value	mse	rmse	corr	coef. angular	interceptor	R^2
a_02	0,03	0,98	0,94	0,89	0,81	0,80	1,16	0,66
a_01	-0,05	0,96	1,51	2,27	0,65	0,51	2,88	0,42
a_11	-4,07	0,00	2,21	4,90	0,52	1,55	-0,51	0,27
a_13	2,45	0,02	1,81	3,27	0,32	0,87	-0,21	0,10

3.4 Conclusão

O suporte tecnológico das avaliações dos alunos é um artefato estimado para muitos professores. Comparado a uma execução manual de todas as etapas esperadas, o tempo e o esforço necessários são substancialmente reduzidos por esses suportes. Neste capítulo, discutimos o uso da análise estatística para evidenciar tendências desejáveis e indesejáveis durante o processo de ensino-aprendizagem. Propomos ampliar a fase de avaliação clássica, de modo a examinar senso crítico dos alunos diante da análise de sua resposta e dos seus pares. Essa estratégia foi usada em um conjunto de aulas de graduação onde o professor aplicou atividades a todos alunos antes de solicitar que avaliassem as respostas de seus colegas de classe. Esta prática tem se mostrado interessante para melhorar a capacidade de aprendizagem e participação dos alunos nas aulas. Este tipo de abordagem já foi implementado com sucesso em outros trabalhos ([LUXTON-REILLY, 2009](#); [OLIVEIRA, 2018](#)).

Nossa solução é singular ao propor um material visual de fácil leitura. Assim, com base nas métricas apresentadas, nosso *framework* gera três tipos de matrizes representadas em gráficos de calor. Como propostas futuras, pretende-se acompanhar o comportamento dos perfis dos alunos ao longo do tempo, coletando dados de forma contínua e verificando a transição dos alunos por diferentes perfis de grupos. No capítulo a seguir, discutimos a análise de agrupamento utilizando as notas dos alunos as ferramentas da TRI.

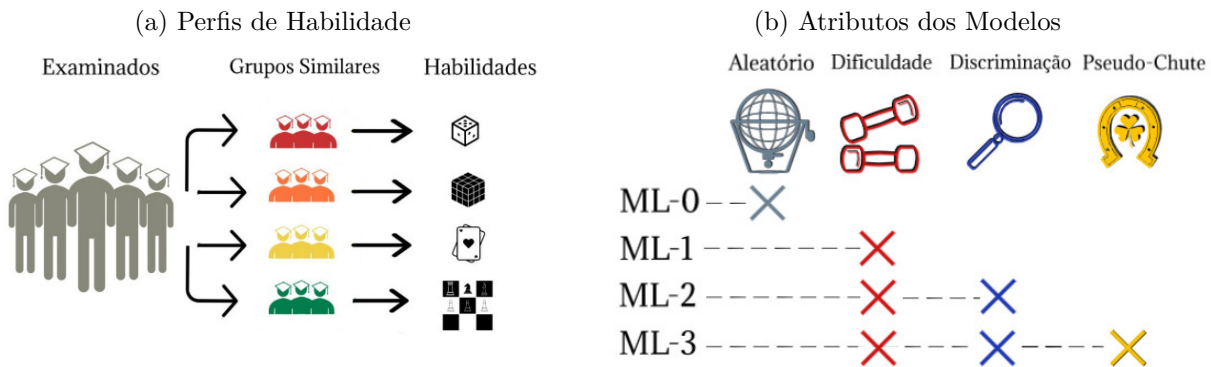
4 Evidenciação de Perfis de Aprendizagem e Caracterização de Atividades

A disponibilidade de vasto conteúdo digital, diversificado semanticamente, é um atributo marcante da contemporaneidade, observado em redes sociais como, por exemplo: *Facebook*, *Twitter*, *Linked In*. No entanto, tamanho conjunto de dados pode ser confuso para usuários que desejam selecionar itens específicos e descartar outros. Um sistema baseado em recomendações é uma solução interessante para resolver esse problema. Para equalizar a oferta e demanda digital, independentemente do tipo de objeto informacional veiculado, faz necessário em primeiro lugar caracterizar a demanda, em termos de interesses. Em segundo lugar, caracterizar a oferta, a respeito das características dos itens ofertados. Por último, definir um modelo de comportamento, que consiga prever predileções, com vistas a antecipar o interesse da demanda. Em se tratando de uma solução inteligente, é esperado que o modelo de comportamento possa melhorar sua predição com base nas informações coletadas ao longo do tempo (RUSSELL; NORVIG, 2009).

A título de ilustração, consideremos uma plataforma digital de comércio de música, em que os perfis dos usuários sejam definidos em função do histórico de consumo. Com esse conhecimento à priori, é possível verificar as características de consumo e predileção, uma vez que tais informações sejam armazenadas pela interface do serviço de música. Um histórico de consumo possibilita também a geração de grupo de usuários com similaridades de interesse. Para grupos cuja predileção é caracterizada, é possível recomendar itens que tenham propensão a aceitação. Por exemplo, se os usuários indicarem explicitamente uma preferência por um estilo de música chamado Música Popular Brasileira, o sistema de recomendação pode recomendar algumas músicas de um artista de renome como *Tom Jobim*. A tal abordagem, nomeia-se *filtragem baseada em conteúdo* (PAZZANI; BILLSUS, 2007). Complementarmente, existe a abordagem denominada *filtragem colaborativa* (HERLOCKER et al., 2004), em que o sistema reúne um grupo de usuários, em função de histórico de interesses pesquisados. Pelas predileções mapeadas no grupo, recomendações são efetuadas, agregando ao histórico os registros transacionais.

No campo dos dados educacionais, existem algumas situações semelhantes, especialmente quando os professores têm que selecionar itens de avaliação de acordo com alguns desempenhos esperados de seus alunos. Considerando as notas em andamento em uma disciplina, é possível agrupar alunos atuais e antigos pela semelhança de seus desempenhos em diferentes momentos do processo de ensino-aprendizagem. Por exemplo, na obra de Oliveira, Ciarelli e Oliveira (2013) é apresentado um sistema baseado em recomendações, o qual pode selecionar itens de avaliação compatíveis com objetivos educacionais coerentes,

Figura 20 – Esquemas Ilustrativos



do ponto de vista didático-pedagógico. Neste capítulo, propomos elevar tal implementação da técnica de recomendação de atividades a outro patamar, por meio da discussão a respeito de modelos preditivos, que assegurem a precisão matemática e estatística dos resultados. Bem como, apresentar uma forma de diferenciação de grupos de aprendizagem, como a ilustrado na Figura 20a, que possua mais camadas de interpretação do que apenas a mera consideração da pontuação dos acertos dos itens avaliados. Pela ideia ilustrada na Figura 20a, o nível de habilidade identificada nos alunos oportuniza a recomendação adequada de itens de avaliação, condizentes com as habilidades identificadas. Um grupo de alta desempenho, ilustrado na coloração verde, teria condições de manipular itens de avaliação complexo como um jogo de xadrez. Um grupo com desempenho inferior ao esperado, teria condições de lidar com jogos de menor complexidade.

Nossa abordagem tem condições de caracterizar as avaliações com modelos da psicometria e também pode agrupar os dados dos alunos de acordo com seu desempenho. Num contexto em que são agregados dados anteriores e verificadas trajetórias semelhantes, pode-se recomendar atividades que sejam compatíveis com desempenho do examinando, a fim de maximizar o desempenho, a partir da personalização da avaliação. Uma vez que visamos mensurar características subjetivas que possam representar níveis de aprendizagem, o conjunto de métodos e modelos da Teoria de Resposta ao Item (TRI), oportuniza um gradiente de possibilidades de aplicação e interpretação de itens de avaliação (PASQUALI, 2004). A caracterização dos itens de avaliação é suportada pela TRI, por meio do processo de estimação de parâmetros, realizado a partir das respostas dadas aos itens de avaliação elaborados e submetidos a teste. Assim, com o uso dos modelos da TRI, pode-se calcular a probabilidades de resposta correta de itens apreciados, em função do traço latente inferido do aluno. A TRI permite a análise qualitativa e quantitativa de itens para apoiar a construção de um caminho de avaliação (BAKER, 2001). A Figura 20b, denota, de maneira simplificada, a contribuição de cada modelo, em termos do significado dos parâmetros estimados. Quando não se considera características particulares de itens de avaliação, o critério de seleção torna-se aleatório, como um sorteio. Os modelos que levam em

consideração características como a dificuldade, o potencial de discriminação e a propensão a chute, são relacionados na Figura 20b, pelos ícones ilustrativos.

Desse modo, nossa abordagem híbrida, baseada em recomendações, lida tanto com as características dos alunos quanto com os atributos dos itens de avaliação. Com nosso método, podemos selecionar uma sequência ordenada de itens para construir um caminho de avaliação personalizado e individualizado para cada aluno. Se um aluno tem um certo traço latente contínuo, nossa abordagem pode encaminhar progressivamente esse aluno por um caminho de avaliação constante e coerente. Ao identificar seus pontos fracos e fortes, nossa sistemática recomenda as atividades mais adequadas para melhorar o desempenho do sujeito examinado. O mapeamento de características que propomos materializa a recomendação de Perrenoud (1998), que sugere a priorização de detecção de lacunas de aprendizagem e seu subsequente tratamento.

4.1 Trabalhos Relacionados

A taxonomia de aplicações de mineração de dados educacionais proposta por Bakhshinategh et al. (2018) designa a modelagem do aluno como uma tarefa essencial para realizar previsões de comportamento. Os modelos de alunos podem ser separados em duas grandes categorias: *centrados em especialistas*; e *centrados em dados* (MAYO; MITROVIC, 2001). Enquanto o primeiro modelo depende de especialistas para identificar as habilidades necessárias para resolver um problema, fornecendo a estrutura de resolução; o segundo modelo depende do uso dos dados produzidos no processo de ensino-aprendizagem, em especial, os dados de avaliação, para inferir as habilidades do aluno (JOHNS; MAHADEVAN; WOOLF, 2006), por meio da identificação de padrões. Fundamentados nessa premissa, podemos considerar que os rastros deixados pelas avaliações anteriores dos alunos representam dados balizadores a respeito da eficácia do processo ensino-aprendizagem. Ao observar os alunos em suas características multidimensionais durante o processo de ensino-aprendizagem, o professor pode identificar fatores intervenientes do desempenho, os quais possam incrementar ou prejudicar o progresso dos alunos (LIEBERMAN, 1990). Como as avaliações de aprendizagem visam medir variáveis não observadas diretamente, é relevante usar processos de avaliação padronizados e certificados para entender em detalhes o que é medido (BAKER, 2001). Acompanhando essa tendência, os trabalhos de Manouselis et al. (2011), Bakhshinategh et al. (2018), demonstram o poder informativo dos dados na orientação de rotas de ensino, para sanar lacunas de aprendizagem detectadas.

Na literatura, localiza-se substancial acervo de trabalhos a respeito de sistemas baseados em recomendações durante o processo de ensino-aprendizagem, com maior destaque para recomendação de conjuntos de recursos de aprendizagem. Por exemplo, no trabalho de Romero et al. (2007), o sistema utiliza técnicas de mineração da web para

recomendar *links* da *Internet*, para visitação e informação dos alunos. Lee et al. (2010) propõe uma arquitetura para recomendar conteúdos que possam reforçar áreas específicas em que um determinado aluno demonstre carência. No artigo do Shishehchi, Banihashem e Zin (2010), um sistema de recomendação semântica para *e-learning* é lançado por meio do qual, os alunos podem encontrar e escolher os materiais de aprendizagem adequados ao seu campo de interesse. O sistema descrito por Timms (2007) provê *feedbacks* a respeito de erros e dicas para apoiar os alunos confrontados com conjuntos de atividades. O sistema acima mencionado consegue determinar o nível de dicas que os alunos precisam durante a resolução de um problema.

Uma modalidade específica de avaliação personalizada é encontrada nos testes adaptativos computadorizados (*Computer Adaptive Test* — CAT), no qual, as etapas de: preparação do teste; aplicação; e término; oportunizam delinear o aluno examinado, com níveis de configuração variados (LEE et al., 2010). Os modelos da TRI representam um suporte aderente ao apoio da realização de CAT, pelo fato de permitir uma melhor compreensão dos resultados de uma determinada avaliação (LINDEN; HAMBLETON, 2013). Conforme afirmado por Roijers, Jeurig e Feelders (2012), a TRI pode ser usada para informar aos alunos sobre sua competência e aprendizado; e aos professores, informar sobre o progresso dos sujeitos examinados. De acordo com Sinharay, Johnson e Stern (2006), a verificação de modelos na TRI é uma área com bastante espaço para desenvolvimento. Este último trabalho examina o desempenho de uma série de medidas de discrepância para avaliar diferentes aspectos de ajuste de modelos da TRI e a criação de recomendações específicas. Enquanto os modelos usuais de TRI demandam assumir que o traço latente é preponderante e a população é homogênea, os modelos MixTRI consideram populações heterogêneas e traços latentes preponderantes por subpopulações (SEN; COHEN, 2019). Em seu trabalho, Johns, Mahadevan e Woolf (2006) propõe treinar modelos de TRI, para predição de desempenho, atingindo uma precisão de 72 % de probabilidade de sucesso na resposta de um item de avaliação de múltipla escolha. Em seu artigo, Wauters, Desmet e Noortgate (2010) exploram a possibilidade de projetar um sequenciamento adaptativo de itens combinando a dificuldade dos itens com o nível de conhecimento do aluno em sistemas tutores inteligentes. Lee e Cho (2015) propõe um método para selecionar itens e criar uma ficha de avaliação personalizada para testes adaptativos considerando tanto a habilidade quanto as características do aluno. Farida et al. (2011) propõe um método para gerar exercícios a partir da progressão do aluno observada através das informações coletadas. Na obra de Lee et al. (2010), observamos a descrição de um sistema de tutoria inteligente para aprendizado da língua inglesa, que fornece conteúdo adequado para níveis específicos de habilidade apoiados por TRI. Finalmente, na abordagem de Yeung (2019), a TRI foi acoplada à modelagem de rastreamento de conhecimento, isto é, determinar quando uma habilidade foi aprendida.

Estabelecer critérios de avaliação adequados para medir as variações de aprendiza-

gem encontradas é um desafio bem conhecido (PERRENOUD, 1998). Esse desafio aumenta proporcionalmente ao volume de alunos da turma. No entanto, apoiados por ambientes virtuais de aprendizagem, as aplicações de atividades e avaliações são facilitadas, permitindo usos mais eficientes das informações (SPALENZA et al., 2018). Em seus trabalhos, Oliveira, Ciarelli e Oliveira (2013), Oliveira (2018) usaram técnicas baseadas em agrupamento e classificação para identificar e tratar as lacunas individuais de aprendizagem, por meio do processamento de texto das respostas dos alunos.

4.2 Arquitetura do Experimento

Na computação, é comum gerar conjuntos de dados artificiais para testar novos protótipos (BOURGUET, 2017). O domínio do conhecimento do comportamento de distribuições estatísticas e seus parâmetros oportuniza a operação de bibliotecas computacionais como o *numpy* e *scipy*, desenvolvidas na linguagem *python*, que implementam códigos capazes de criar conjuntos de dados de tamanhos variados e características desejadas. Com dados artificiais, é possível focar o estudo na análise dos objetos matemáticos oriundos da modelagem do problema. Em termos práticos, a fim de modelar o comportamento de 1000 alunos que realizassem um teste contendo 45 itens, sendo a resposta dicotômica, isto é, certa ou errada.

Numa situação corriqueira de avaliação, a nota final dos avaliados seria a somatória de acertos. Pelas notas alcançadas, ter-se-ia uma perspectiva da aprendizagem dos avaliados a respeito dos conhecimentos aferidos. Contudo, a avaliação individual de acertos e erros é o que permite a identificação de lacunas de aprendizagem. No exemplo proposto, discutimos níveis de desempenho dos avaliados e características dos itens de avaliação.

4.2.1 Característica dos Dados

Para representar as respostas certas e erradas, de modo a compor uma matriz de valores que simbolizou as respostas de uma turma, construímos um conjunto de dados X , baseados de valores randômicos oriundos de 45 ensaios de Bernoulli por aluno sujeito simulado.

A expressão resumida de um ensaio de *Bernoulli* é $X \sim \text{Bernoulli}(p)$ usada para indicar que uma variável aleatória X tem distribuição Bernoulli com parâmetro p , onde $0 < p < 1$. Uma variável aleatória X de distribuição Bernoulli com probabilidade de sucesso p tem a função de massa de probabilidade denotada por:

$$f(x) = p^x(1 - p)^{1-x} \quad x = 0, 1.$$

para $0 < p < 1$.

Ensaio de Bernoulli, são experimentos com o resultados genericamente referidos como sucesso ($x = 1$) e fracasso ($x = 0$). A função da distribuição cumulativa $X \sim \text{Bernoulli}(p)$ é

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases}$$

A geração de valores randômicos teve os parâmetros controlados pela distribuição binomial, sendo $X \sim b(n, p)$, ver Equação 4.1 (BUSSAB; MORETTIN, 2013), com “n” sendo o número de eventos possível. O parâmetro p possui os seguintes comportamentos para o valor esperado e variâncias: $E(x) = p; Var(X) = p(1 - p)$.

$$X \sim p^x(1 - p)^{1-x}, x = 0, 1. \quad (4.1)$$

No conjunto de dados artificiais gerados, foi feita a mesclagem de três grupos de tipos de indivíduos, caracterizados seguinte conjunto de pares: $\{(0, 15; 27\%), (0, 6; 53\%), (0, 8; 21\%)\}$. O primeiro valor do par simboliza a probabilidade de sucesso no resposta ao item. O segundo valor, a proporção de indivíduos em relação ao conjunto. A Figura 21 apresenta o gráfico de barras com a média de acerto dos itens. Na Figura 22, é possível observar a distribuição formada pelas médias de acerto dos itens.

Figura 21 – Conjunto de Dados Artificiais. Fonte: elaborado pelo autor.

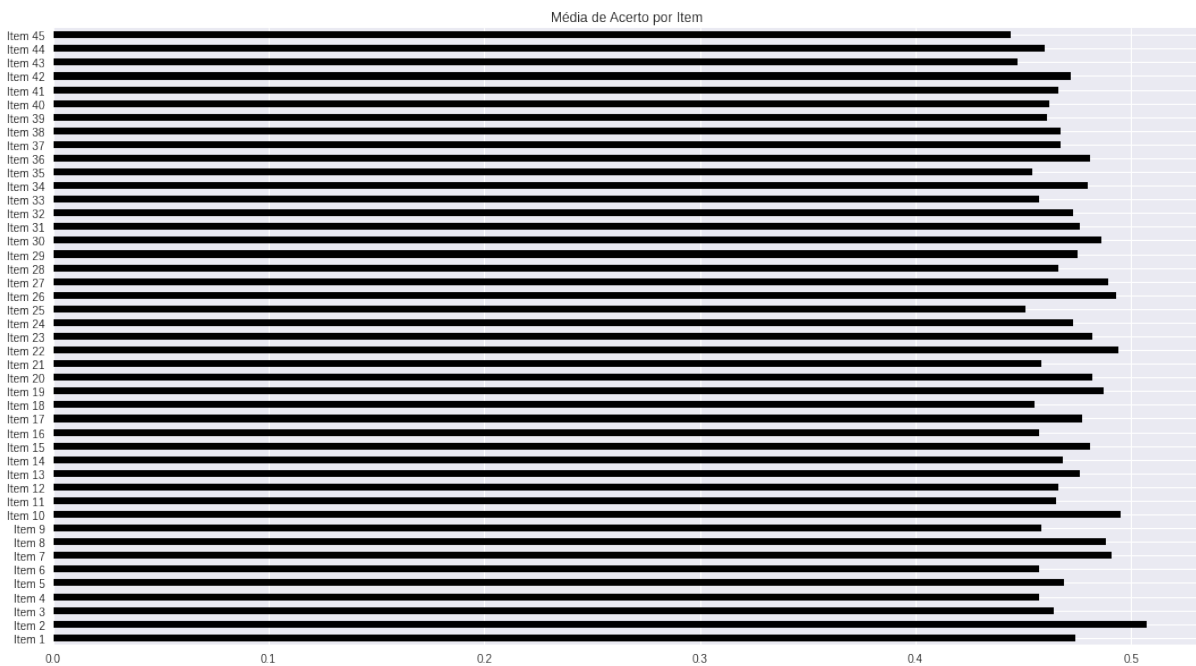
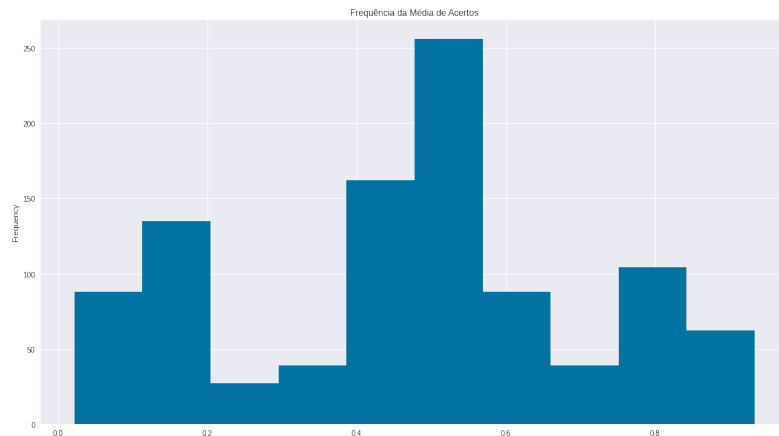


Figura 22 – Frequência da Média de Acertos. Fonte: elaborado pelo autor.



4.2.2 Identificação Inteligente de Perfis

Por meio de processos de análise de agrupamento, variáveis distintas podem ser interpretadas conjuntamente. Uma das técnicas mais populares é o método *k-means* que visa estabelecer uma heurística de otimização no espaço paramétrico das variáveis, de modo a agrupar em *k*-grupos as instâncias apresentadas, por meio de associação de cada uma aos pontos de referência, denominados *centroids*, estipulados randomicamente no conjunto. Para identificar a quantidade ótima de grupos, recorreremos ao método do arco, visa calcular a distância euclidiana intra-grupo e entre grupos ao longo de uma faixa estabelecida, retornando uma pontuação ponderada dos valores apurados (SPALENZA; PIROVANI; OLIVEIRA, 2019). O ponto em que há variação de crescimento se transforma para decrescimento da pontuação é considerado a região valores ótimos.

4.2.2.1 A Teoria de Resposta ao Item

A TRI (BAKER, 2001) tem sido considerado por muitos especialistas como um marco para a Psicometria moderna e uma extensão da Teoria Clássica dos Testes (CTT). Enquanto o CTT se baseia na proposição de que o erro de medição, uma variável aleatória latente, é um componente da pontuação de avaliação (TRAUB, 1997), a TRI considera a probabilidade de acertar ou errar determinados itens, dada a habilidade dos examinandos. Cada examinado possui alguma quantidade da habilidade subjacente (também chamada de traço latente) materializada como uma pontuação de habilidade (ou seja, um valor numérico denotado θ) em uma escala de classificação. A TRI defende que dependendo de um certo nível de habilidade, haverá naturalmente uma probabilidade denotada $P(\theta)$ com a qual um examinando responderá corretamente ao item. A curva em forma de S desta função é chamada de curva característica do item, e cada item tem a sua. A relação entre esta probabilidade e a pontuação de habilidade é modelada por uma função logística (ou seja, sigmoide). Tais funções relacionam a suposição natural de que a probabilidade será baixa para examinandos com habilidades fracas e alta para examinandos com grandes

habilidades, a probabilidade tende a zero nos níveis mais baixos de habilidade e tende a 1 no mais alto. A função depende de parâmetros que descrevem algumas propriedades de cada item.

4.2.2.2 Modelo de 1 Parâmetro

Também conhecido como Modelo de *Rasch*, expressão logística vista na Equação 4.2 é a FRI, estabelece para o sujeito x a probabilidade de sucesso $P(\theta_x)$. A variável contínua θ assume valores de Θ , com seu intervalo definido da seguinte forma $\Theta = [-3, \dots, 3]$.

$$P(S_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-1.7(\theta_j - b_i)}} \quad (4.2)$$

Ademais, Θ possui uma distribuição estatística normal, $\Theta \simeq N(0, 1)$. Os valores de $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, n$, onde S_{ij} representa a variável dicotômica do indivíduo j que diante da resposta correta do item i , assume valor 1; e para respostas incorretas, assume valor 0. O θ_j denota a habilidade do j -ésimo indivíduo. Na Equação 4.2, o parâmetro b_i é a dificuldade do item i , mensurado na mesma escala da habilidade. O valor 1,7 que multiplica o *logit* no modelo ajusta as curvas da função para ficar semelhante a uma ogiva normal.

4.2.2.3 Modelo de 2 Parâmetros

Acrescendo um nível a mais de informação sobre os itens, o Modelo Logístico de 2 Parâmetros alia a ideia de dificuldade do item ao conceito de discriminação. Para um item i , os valores negativos de discriminação significa que sujeitos de menor habilidade conseguem respondê-lo corretamente, todavia, não ocorre o mesmo com sujeitos de grande habilidade. Tal controvérsia faz com que consideremos itens assim inadequados.

$$P(S_{ij} = 1|\theta_j) = \frac{1}{1 + e^{-1.7a_i(\theta_j - b_i)}} \quad (4.3)$$

Usualmente, são considerados valores adequados aqueles no intervalo de 0 a 2. O ponto de inflexão na curva de característica possui inclinação proporcional ao valor da discriminação. A FRI do Modelo de 2 Parâmetros pode ser vista na Equação 4.3. Para um item i , os pressupostos do modelo *Rasch* são assumidos, acrescido do fator multiplicador a_i ao *logit* da Equação 4.3.

4.2.2.4 Modelo de 3 Parâmetros

O Modelo de 3 Parâmetros, visto na Equação 4.4, acrescenta a ideia de propensão ao chute — pseudo-chute. Nele, o estimador identifica a probabilidade de um sujeito

com baixa habilidade responder corretamente um item i . Situações desse tipo são vistas quando o parâmetro c_i difere de 0, tratados como acertos ao acaso, isto é, palpite ou pseudo-chute. Valores superiores a 0,35 são indesejáveis para este estimador, haja visto que a probabilidade acerto para sujeitos de baixa habilidade torna-se grande.

$$P(S_{ij} = 1|\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-1.7a_i(\theta_j - b_i)}} \quad (4.4)$$

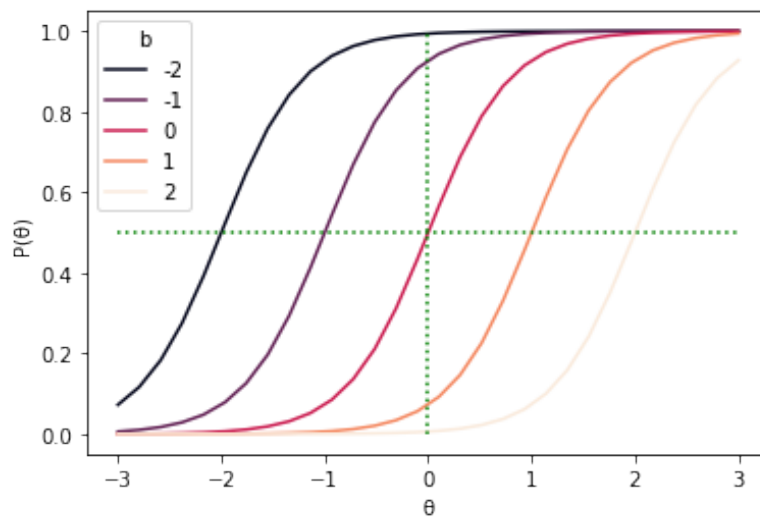
4.2.2.5 Parâmetros dos Modelos Preditores

Os modelos matemáticos providos pela TRI apresentam a relação probabilística entre um indivíduo acertar a resposta de um determinado item e sua habilidade correspondente. Assim, os modelos apresentados a seguir são de natureza dicotômica. A diferenciação dos modelos ocorre pelo conjunto de parâmetros associados, relativos à capacidade de discriminação (a_i), dificuldade do item (b_i) e propensão a chute (c_i). Computacionalmente, os parâmetros são feitos estimados num processo de otimização interativa que visa encontrar para a matriz de entrada das respostas, os coeficientes que melhor ajustem a função aos pontos dos vetores. A seguir, a descrição em mais detalhes de cada parâmetro.

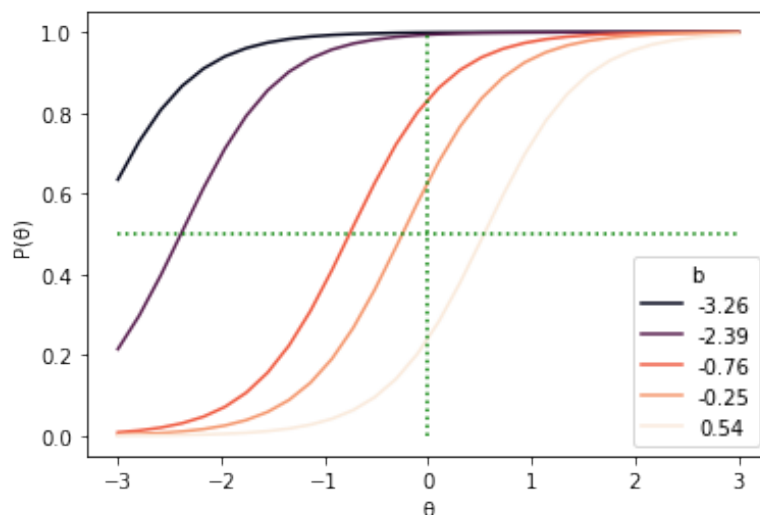
- *Parâmetro de Discriminação:* Uma segunda característica importante do item é seu poder discriminativo, ou seja, sua capacidade de diferenciar examinandos (ou seja, distinguir aqueles que têm sucesso daqueles que falham no item) em relação à sua pontuação de habilidade subjacente. A discriminação de um item i denotado a_i é a inclinação máxima da curva característica do item (ou seja, a inclinação da tangente geométrica que passa pelo ponto de inflexão da curva). Portanto, a inclinação pode ser mais ou menos inclinada: quanto mais íngreme a inclinação, mais o item é discriminativo.
- *Parâmetro de Dificuldade:* Independentemente do número de parâmetros utilizados para definir a curva característica de um item, a dificuldade do parâmetro está sempre presente. A dificuldade de um item i denotado (b_i) é a pontuação de habilidade que corresponde a uma probabilidade de sucesso igual a 0,5, para um sujeito de habilidade mediana, isto é, com θ igual a 0.
- *Parâmetro de Pseudo-Chute (Adivinhação):* Mesmo que o examinado não tenha nenhuma habilidade no escopo que está sendo avaliado, ele pode ter uma probabilidade não nula de responder corretamente ao item. Este é particularmente o caso de avaliações baseadas em escolhas múltiplas. A pseudo-adivinhação de um item i denotado c_i é a probabilidade de sucesso no item correspondente ao valor mínimo de habilidade subjacente.

Figura 23 – Curvas de Características

(a) Espaçamento Uniforme



(b) Espaçamento Aleatório



Com base nos parâmetros estimados para cada modelo, pode-se utilizar a Função de Resposta do Item (FRI), cujos valores descrevem a probabilidade de acerto para as habilidades informadas. Visualmente, ela é descrita pela chamada de Curva de Características. Sua importância está em apresentar visualmente a estimativa de acerto em função de uma determinada habilidade. A Figura 23a mostra um exemplo. O Eixo x possui os valores das habilidades de θ enquanto o eixo y, apresenta valores da probabilidade $P(\theta)$. De início, abstraindo os valores estimados para o parâmetro b , a Figura 23a lista curvas igualmente espaçadas no eixo x, considerando a referência de $P(\theta = 0) = 0.5$, na coloração vermelha, sendo este ponto o cruzamento das linhas pontilhadas em verde. Sucintamente, isso representa que um sujeito com habilidade mediana teria 50% de chances de acertar um item descrito por tal curva. No extremo esquerdo do gráfico, temos o item de maior probabilidade de acerto, em função da seguinte situação: $P(\theta = -2) = 0.5$, qual seja, um

sujeito com habilidade distante da média dois desvios padrões negativos teria 50% de chance de acerto. No outro extremo, o item mais desafiador, de cor marrom-claro, em que verificamos a seguinte situação: $P(\theta = 2) = 0.5$, no qual a probabilidade de acerto de 50% está prevista para sujeitos que estejam dois desvios padrões acima da média de acerto. Enquanto na Figura 23a vemos curvas igualmente espaçadas, na Figura 23b possui um espaçamento sem padrões aparentes. Utilizando as linhas verdes pontilhadas que cruzam o gráfico na posição $P(\theta = 0) = 0,5$, vemos que um sujeito com habilidade média teria aproximadamente 20% de sucesso no último item da direita, na coloração bege. Enquanto, os itens nos dois primeiros itens, a probabilidade seria de 100%. No terceiro item, na cor laranja, a chance seria superior a 80%. Para esse modelo, todos os itens possuem mesmo nível de discriminação e não se considera acertos ao acaso.

4.2.2.6 Função de Informação

A noção de precisão (ou informação) assume um papel essencial, ao indicar em particular em qual parte da escala de habilidade subjacente (ou seja, para qual(is) categoria(s) de examinandos) a precisão do item é o mais alto. Um item difícil fornecerá muito poucas informações sobre os examinandos com as habilidades mais fracas e, ao contrário, um item fácil não será um teste preciso para os examinandos com as habilidades mais fortes.

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)(1 - P_i(\theta))} \quad (4.5)$$

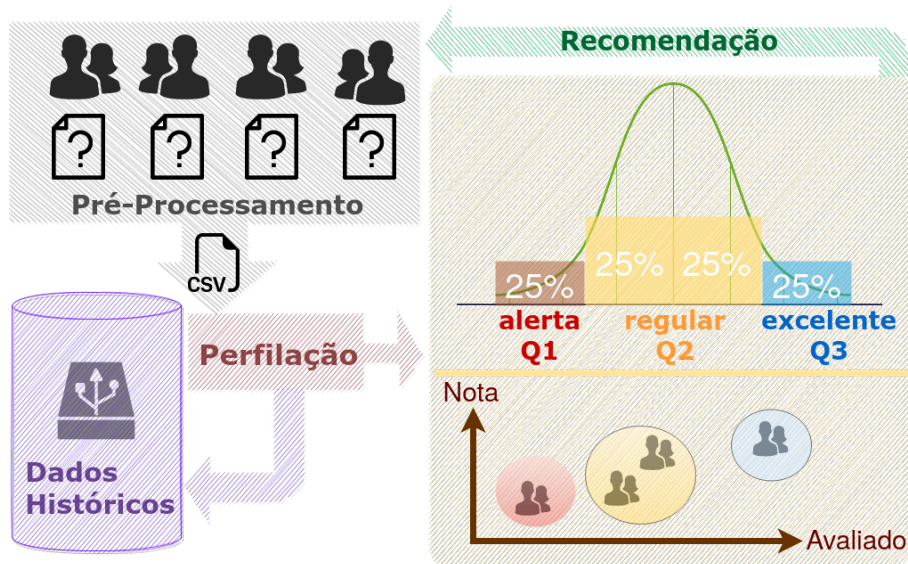
Para o item i cuja informação I_i se deseja saber para a habilidade θ , temos na Equação 4.5, para o Modelo de 3 Parâmetros. Caso se desejasse calcular a informação para o Modelo de 2 parâmetros, assumiria o valor de $c_i = 0$. De outra sorte, desejando-se calcular a informação para o Modelo de 1 Parâmetro, assumiria o valor de $a_i = 1$.

4.2.3 Descrição do Sistema

Argumentamos que ações passadas bem-sucedidas podem ser aplicadas a alunos semelhantes para estimular seus desenvolvimentos. Portanto, propomos selecionar itens com probabilidade controlada de sucesso que correspondam à capacidade estimada de um determinado aluno. Nosso sistema baseado em recomendações híbridas sob medida lida com as características dos alunos e as probabilidades de desempenho. A adoção de técnicas de mineração de dados e aprendizado de máquina ajuda a entender melhor os comportamentos de desempenho dos alunos. A Figura 24 representa o fluxo de trabalho global de nossa abordagem. Na etapa de processamento, as notas dos alunos representadas por um ícone de arquivo de planilha (de extensão .csv, denotando um arquivo numérico separado pelo caractere ';') são usadas como valores de entrada para a estimação dos

parâmetros preditores. A biblioteca *mirt*¹, implementada na linguagem de programação R foi utilizada para a estimação dos coeficientes da TRI, para os três modelos: *Rasch*, 2PL e 3PL. As demais análises e os gráficos, foram gerados com a linguagem *python*.

Figura 24 – Fluxo do Sistema de Recomendação. Fonte: elaborado pelo autor.



Para implementar o sistema de recomendação com uma abordagem de filtragem colaborativa, utilizou-se a técnica de agrupamento, a fim de agrupar vetores de notas de alunos semelhantes. Tratando a lista de notas de cada aluno como um vetor, um algoritmo de agrupamento consegue reunir os vetores mais semelhantes relacionados a um *centroide*. Por outro lado, para discriminar os itens de avaliação, apoiados pela TRI, nosso sistema pode gerar para cada item os valores dos parâmetros de dificuldade, discriminação e pseudo-chute. Portanto, o sistema pode agrupar alunos e atividades. Como frisado anteriormente, uma caracterização estatística dos valores encontrados, para conhecer quais sujeitos estão abaixo do intervalo interquartil, quais estão no intervalo e quais estão acima dele. Ilustrados na Figura 24 pelas regiões avermelhadas, amarelas e azuis, denotando situações de alerta, regulares e excelentes, respectivamente. Ao combinar as informações geradas pelo análise de agrupamento dos sujeitos e pela gerações de características dos itens, podemos compor um caminho de avaliações personalizado para cada aluno.

4.3 Resultados e Discussão

Nesta seção, descrevemos os resultados encontrados com os esforços de criar perfis de aprendizagem com base no histórico de respostas, por meio de um processo otimizado, de base estatística. A seguir, apresentamos as características dos itens de avaliação, utilizando os modelos da TRI citados. Comparamos a desempenho de algoritmos de aprendizagem

¹ philchalmers.github.io/mirt/

de máquina para atingir os mesmos resultados dos modelos psicométricos.

4.3.1 Perfis de Desempenho

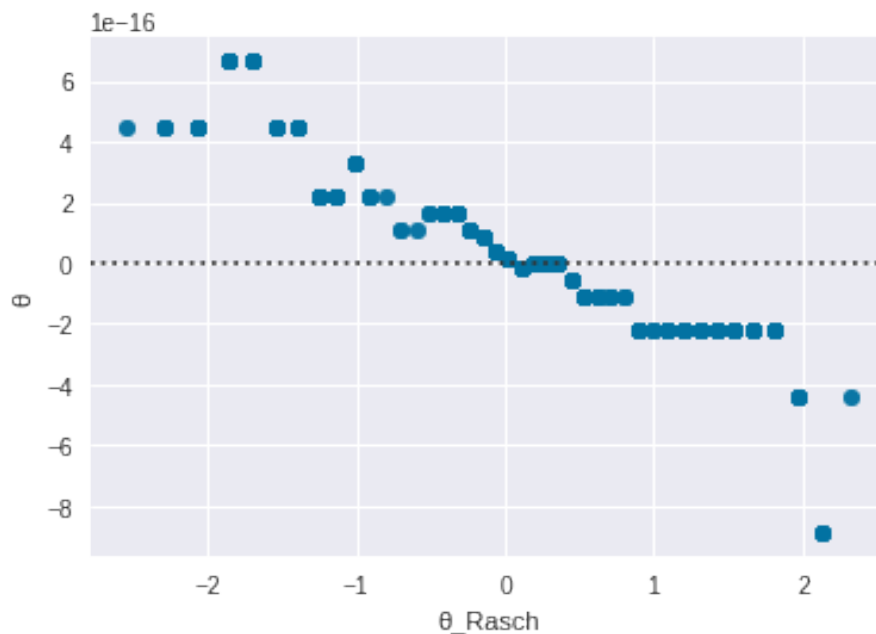
A Tabela 2 exibe as correlações encontradas em θ nas variáveis que denotam a habilidade. Com grande correspondência entre a média de acerto e ha habilidade estimada para o modelo Rasch.

Tabela 2 – Correlação das Habilidades

Modelo	μ_a	θ_{rasch}	θ_{2PL}	θ_{3PL}
μ_a	1,00	1,00	0,96	0,96
θ_{rasch}	1,00	1,00	0,96	0,96
θ_{2PL}	0,96	0,96	1,00	0,98
θ_{3PL}	0,96	0,96	0,98	1,00

Indo além da similaridade das variáveis estimadas, verificamos o gráfico de resíduo dos modelos, na Figura 25, em que é mostrado o valor da média de acerto no eixo vertical e os valores de θ . Nele, é observado a subestimação de valores negativos abaixo do primeiro desvio padrão negativo. E também verificamos a superestimação de valores acima do primeiro desvio padrão positivo.

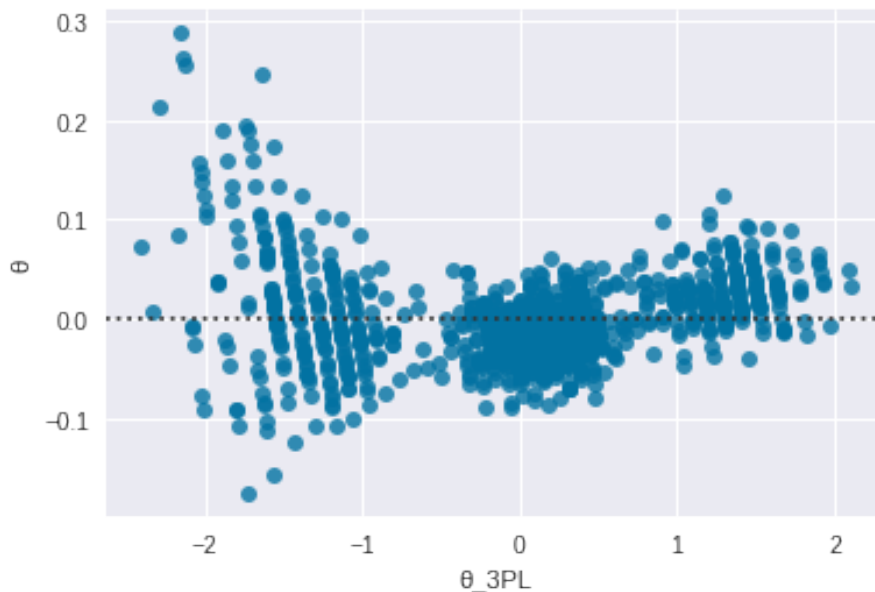
Figura 25 – Resíduos Habilidades Estimada Modelo Rasch e Média de Acerto Escalada. Fonte: elaborado pelo autor.



Na Figura 26, verifica-se o gráfico de resíduos da comparação da média de acertos denominada θ , no eixo vertical, em relação à habilidade estimada no modelo de 3PL. Neste

caso, temos uma maior divergência em relação a valores inferiores ao primeiro desvio padrão negativo.

Figura 26 – Resíduos Habilidades Estimada Modelo 3PL e Média de Acerto Escalada.
Fonte: elaborado pelo autor.



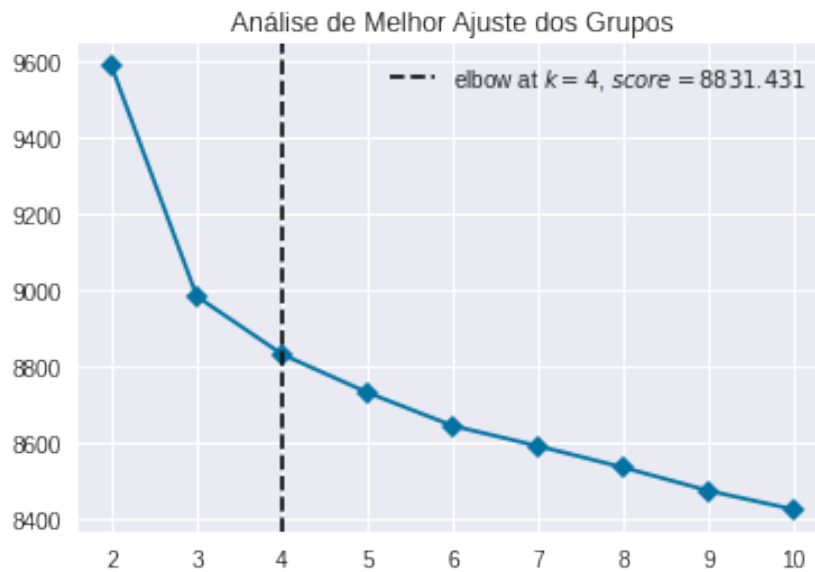
Verificando em similaridades por agrupamento, realizamos a análise variando a quantidade de grupos entre 2 e 10. A Figura 27 mostra o ponto da quantidade ótima de grupos, com valor de 4 e pontuação de 8831,43. A linha tracejada destaca o ponto ideal de grupos. Na prática, veremos que três grupos seriam suficientes para ter intervalos distintos entre os intervalos interquartis das distribuições das notas. Todavia, o padrão de sucesso dos grupos intermediários, variou a ponto de um dos grupos adentrar a região de valores negativos.

4.3.2 Grupos de Habilidades

Com base no agrupamento dos vetores, projetamos na Figura 28a o grupo de desempenho intermediário superior, em que os valores de θ ficaram com média em torno de 0,2 unidades e os intervalos interquartis no intervalo de 0,05 a 0,35, aproximadamente. A Figura 28b tem o intervalo interquartil adentrando a região de valores negativos, representando haver vetores cuja predominância de acerto representou menos que a média, todavia, ficaram ajustados no grupo intermediário.

A Figura 29 mostra o grupo inferior, com suas médias ficando todas abaixo de -1,25 unidades. Neste gráfico, observa-se também o não emparelhamento dos intervalos interquartis, evidenciando divergência em valores inferiores na estimação dos modelos. Por fim, a Figura 29 mostra as distribuições de habilidade do grupo com desempenho superior. A habilidade estimada para valores próximos ao limite superior, ocorre com

Figura 27 – Análise da Heurística do Melhor Número de Grupos. Fonte: elaborado pelo autor.



menos intensidade no modelo de 3PL, na coloração roxa.

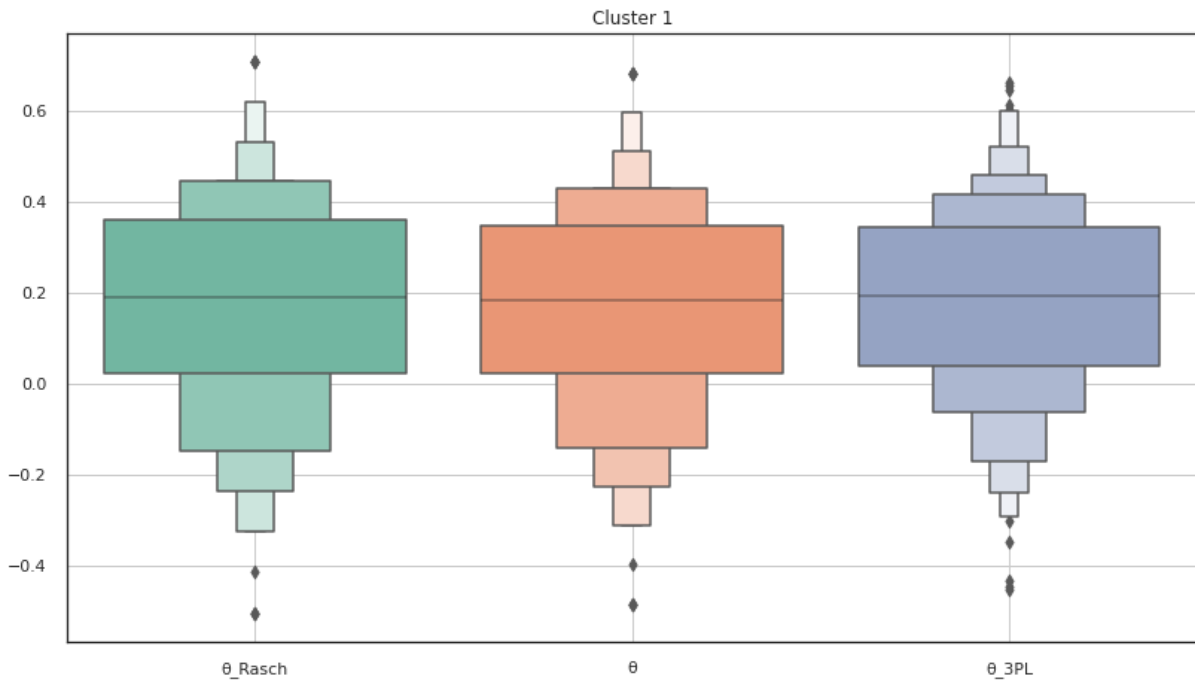
4.3.3 Caracterização dos Itens de Avaliação

As curvas de características do Modelo *Rasch* para o conjunto de itens proposto está na Figura 30a. Nela, nota-se as curvas de características bastante parecidas, dificultando a diferenciação entre os itens avaliados. A Figura 30b mostra as funções de informação dos itens, de mesmo modo, também visualmente demasiadamente similares. Para o modelo de três parâmetros da TRI, a Figura 31a apresenta diferenciação entre itens, a exemplo das curvas dos itens 7, 16 e 45, representados pelos quadrantes x6, x15 e x44. Neles, em função da probabilidade de palpite, os valores cortam o eixo vertical acima da origem. Nas funções de informação, listadas na Figura 31b, o nota-se o conjunto de itens que reúne menos informação em θ , dentre eles, o item 4, apresenta umas das distribuições menor ponto máximo de informação. De outro lado, o item 6, é o que mais reúne informação. A fim de verificar similaridades entre os itens de avaliação, também foi procedida a análise de agrupamento, chegando ao valor ideal de 4 grupos, como visto na Figura 32a. A fim de observar com maior riqueza de detalhes os coeficientes estimados e os agrupamentos em que ficaram assentados no processo de clusterização, apresentamos a Figura 32b. As colunas da esquerda para a direita referem-se a: coeficiente de discriminação, coeficiente de propensão a palpite, coeficiente de dificuldade do modelo de três parâmetros, coeficiente de dificuldade do modelo *Rasch*, média de acerto escalada para uma distribuição padronizada, e cluster definido para o item.

Analisando a dispersão dos coeficientes grupo a grupo, temos a Figura 33a em que os valores da média de acerto situam-se mais dispersos os demais coeficientes, todavia, ficando

Figura 28 – Desempenho por Cluster. Fonte: Fonte: elaborado pelo autor.

(a) Desempenho Intermediário Superior



(b) Desempenho Intermediário Inferior

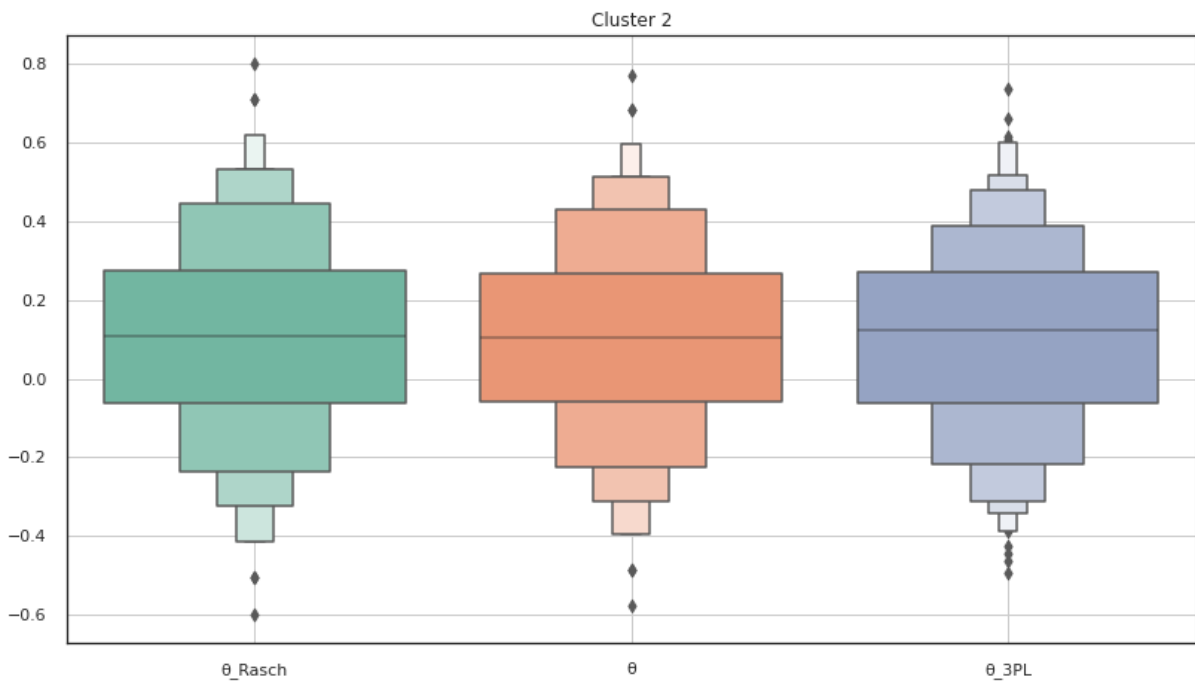
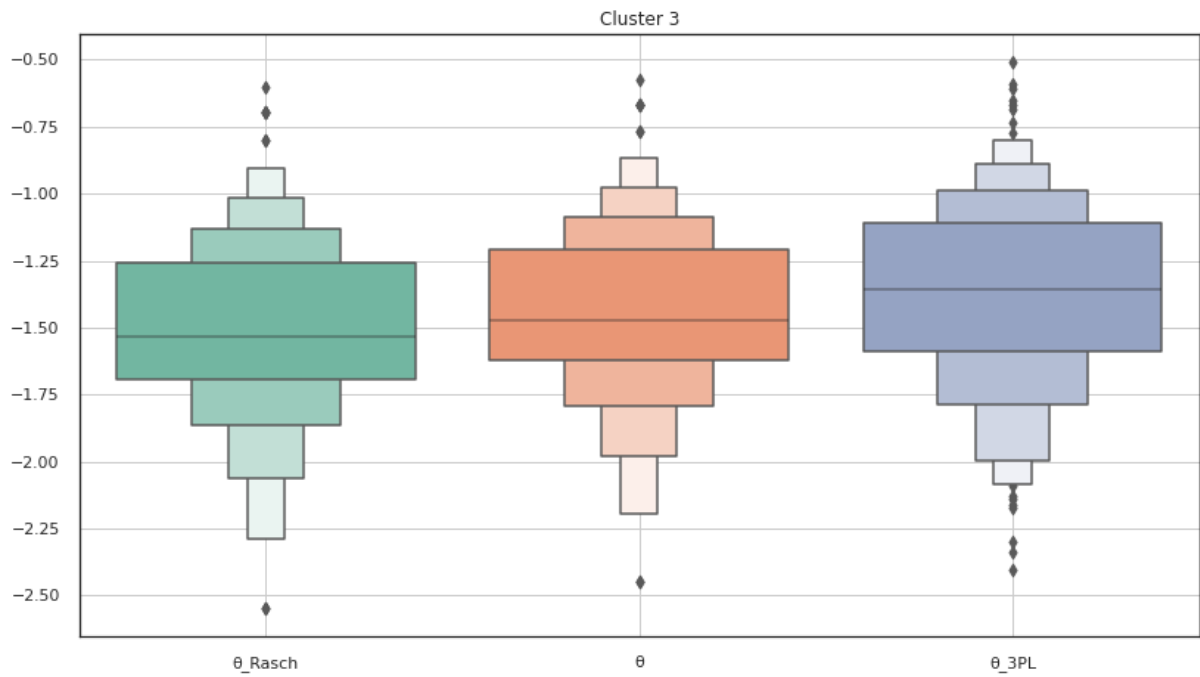


Figura 29 – Desempenho por Cluster. Fonte: Fonte: elaborado pelo autor.

(a) Desempenho Inferior



(b) Desempenho Superior

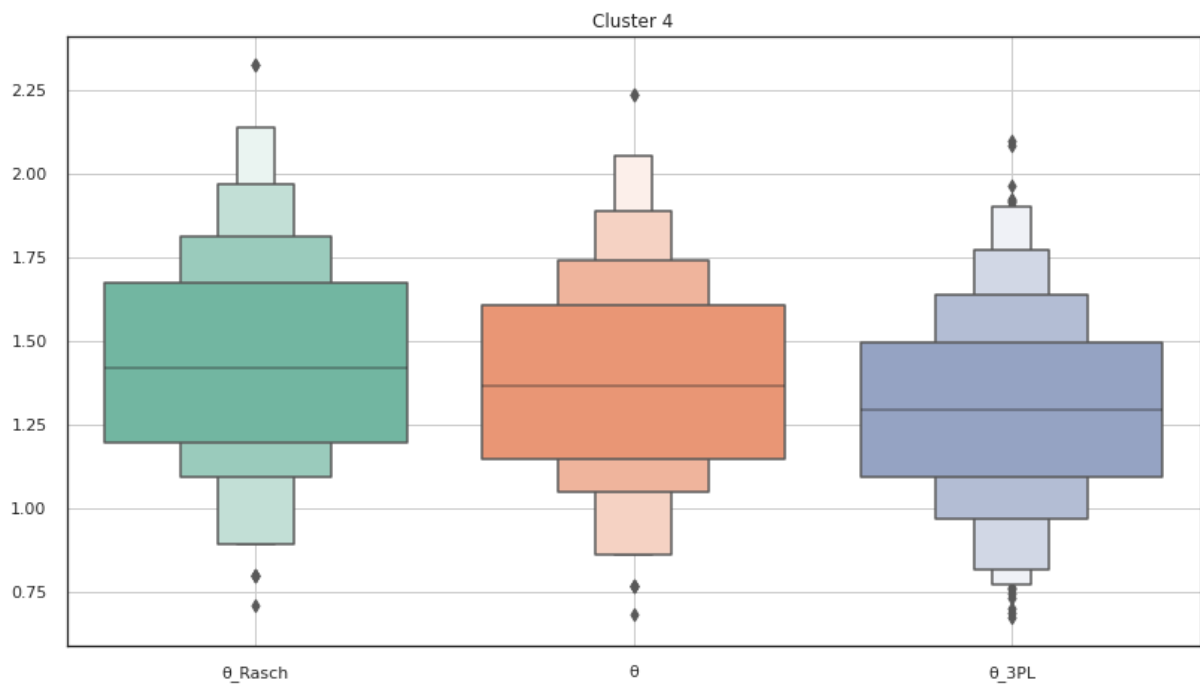


Figura 30 – Funções de Características e Funções de Informação

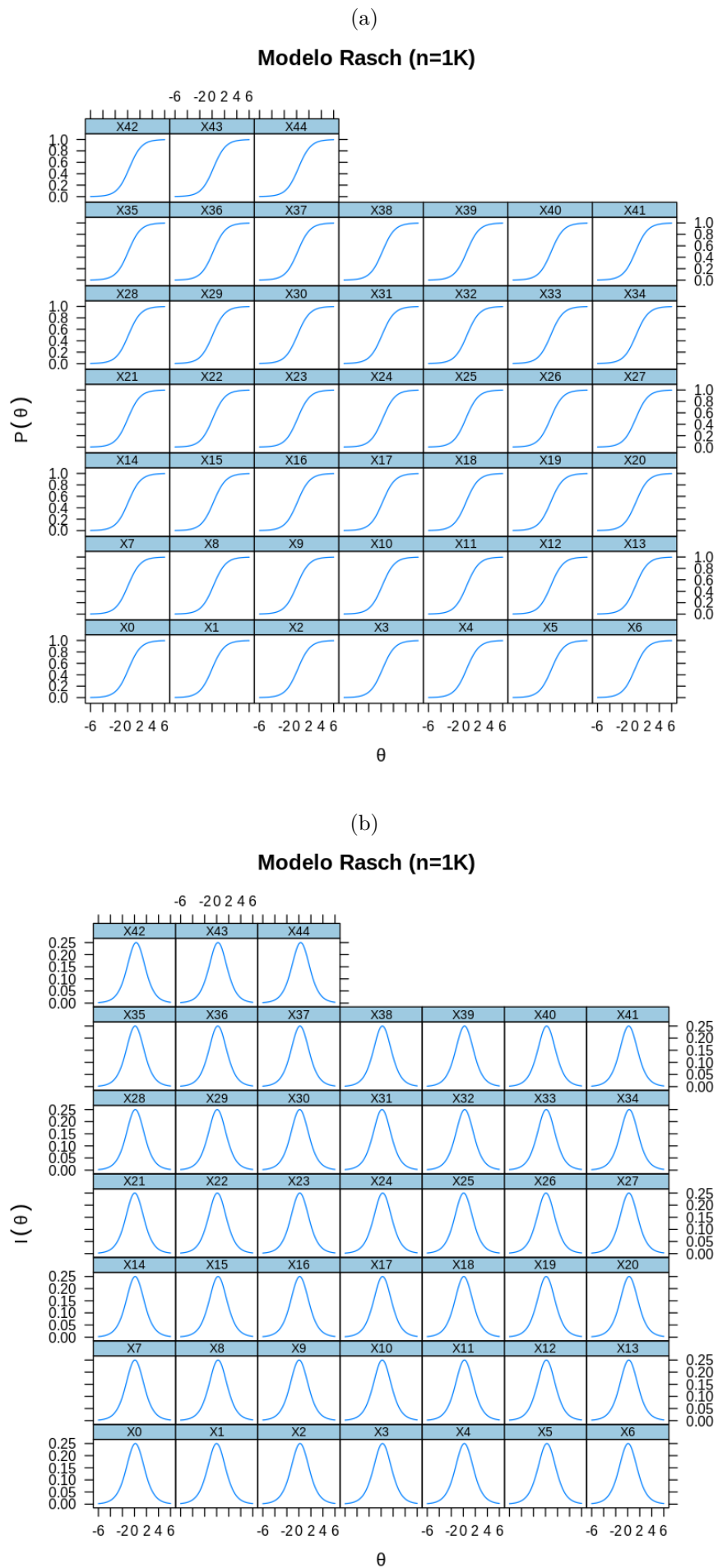


Figura 31 – Funções de Características e Funções de Informação

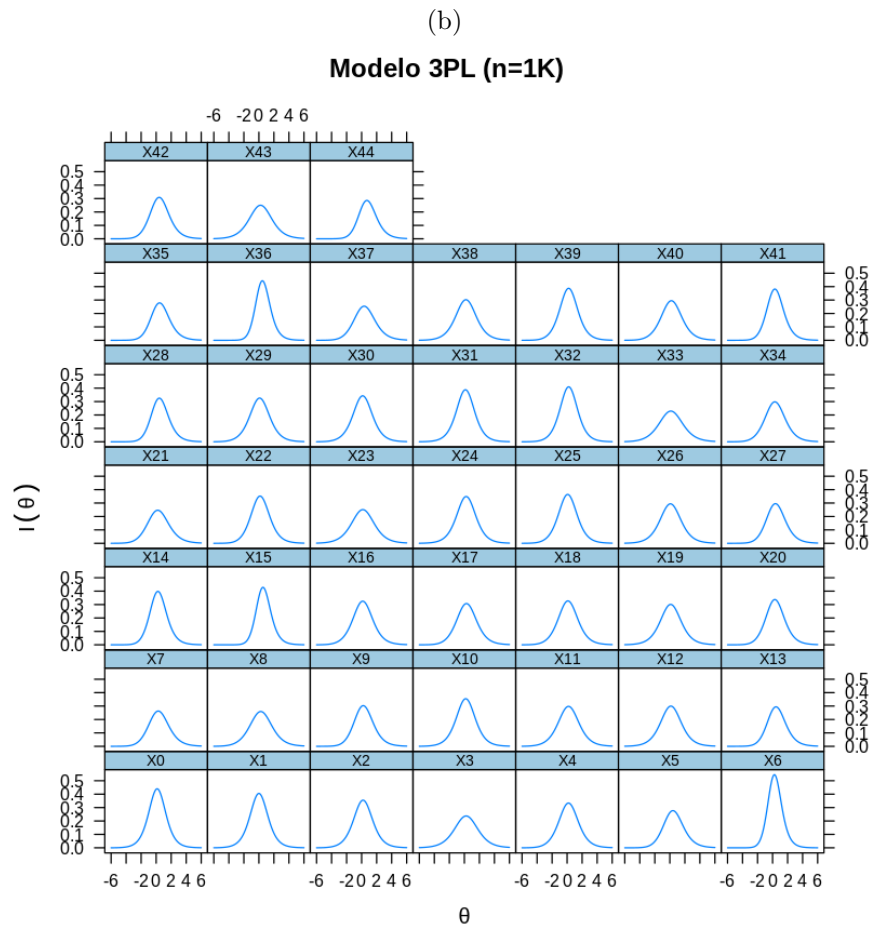
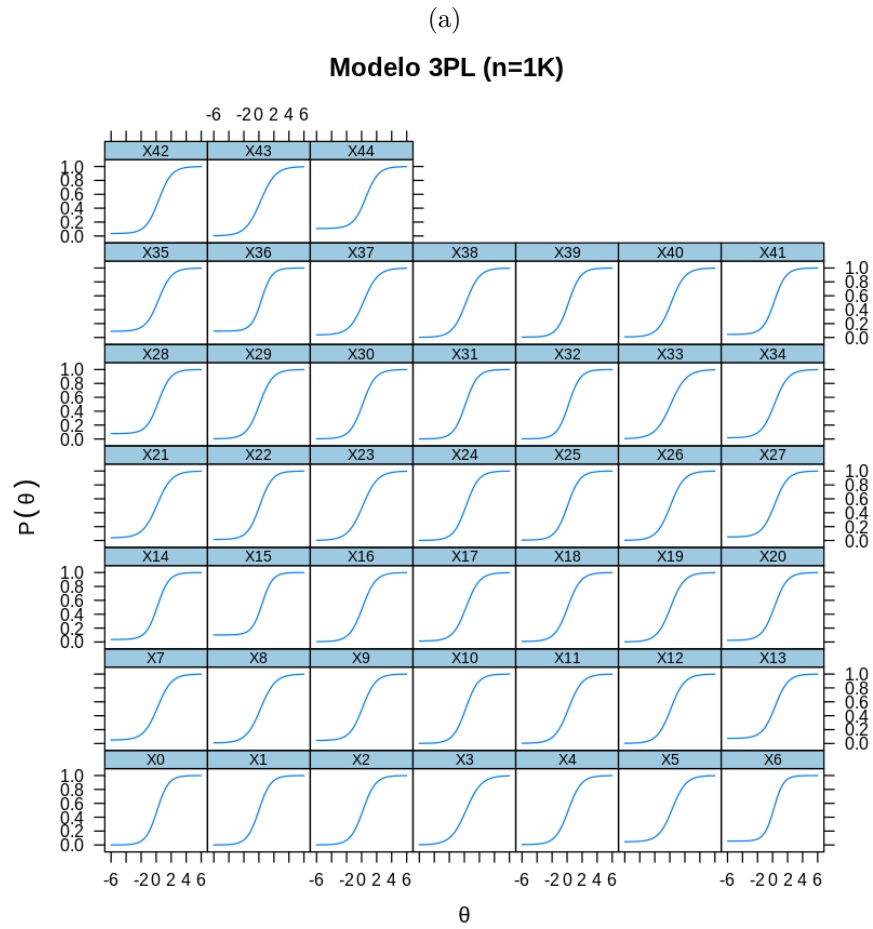
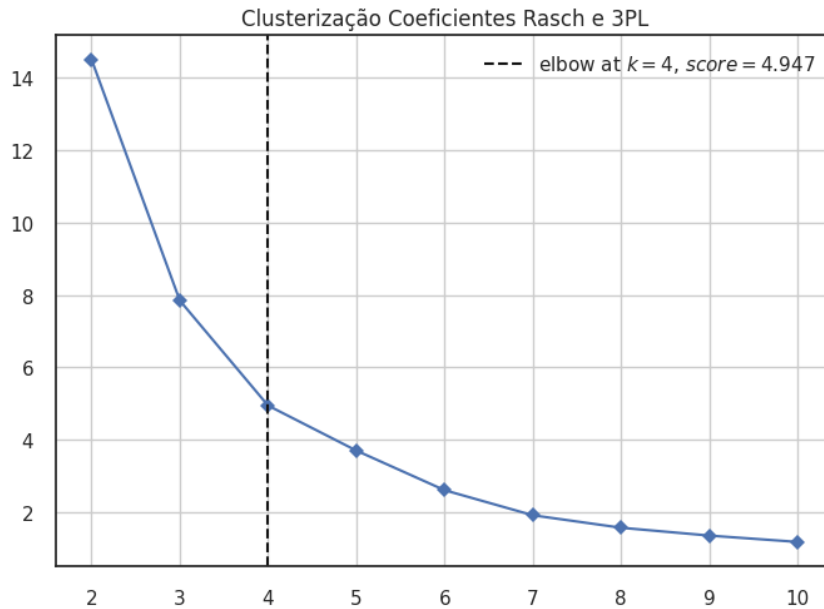


Figura 32 – Análise de Agrupamento dos Coeficientes da TRI

(a) Heurística de Otimização



(b) Coeficientes Estimados e Agrupados

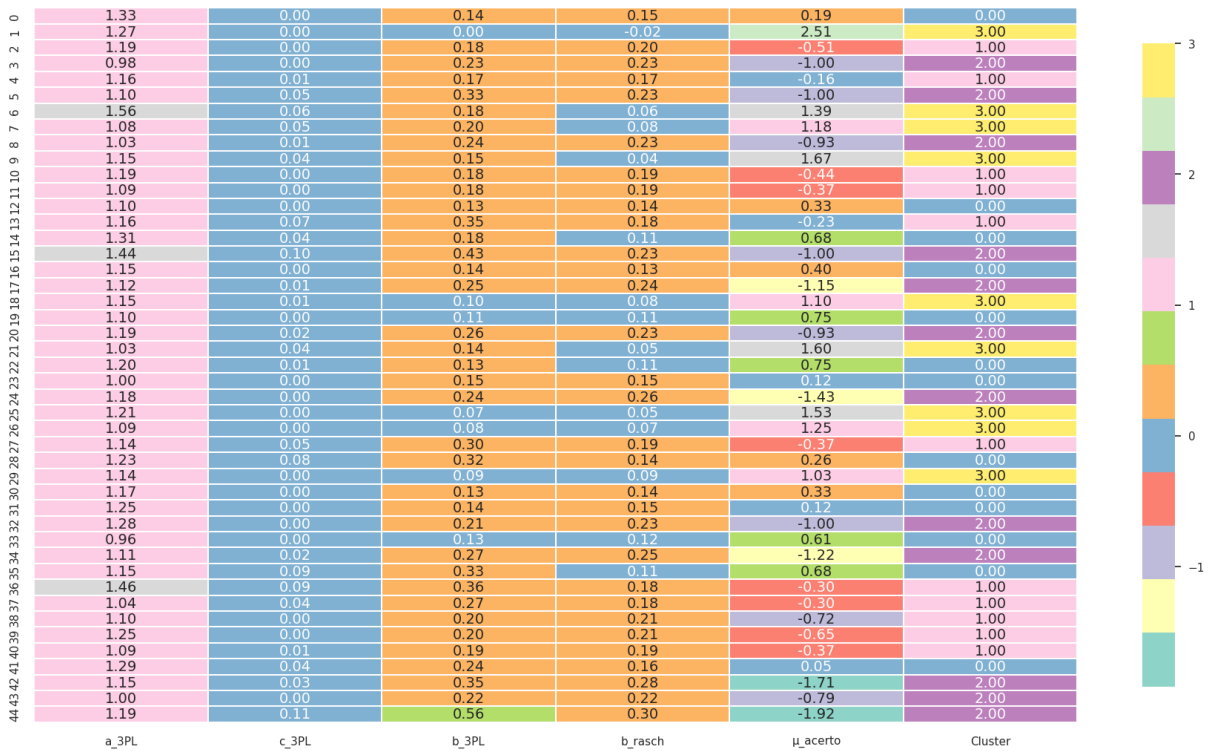
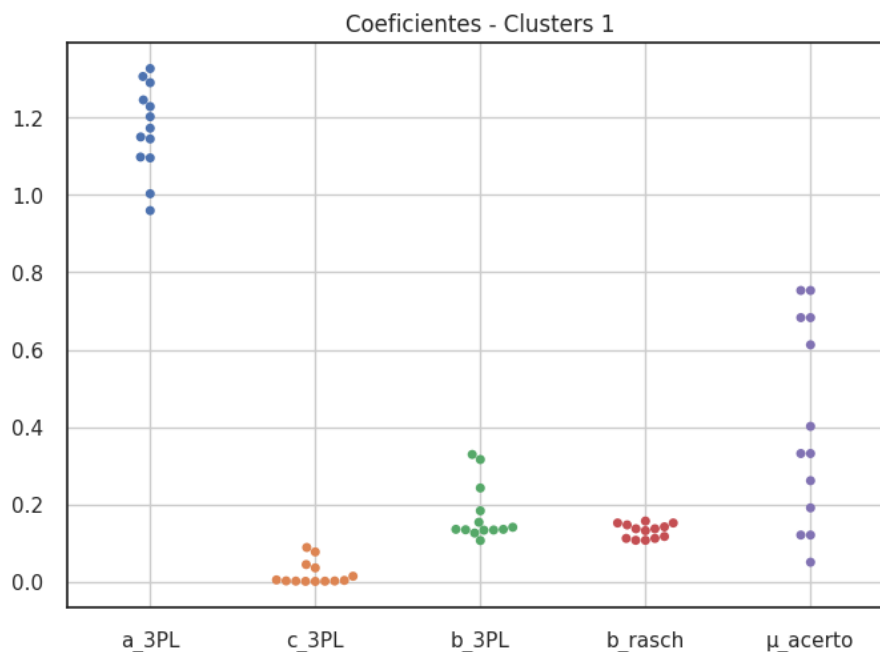
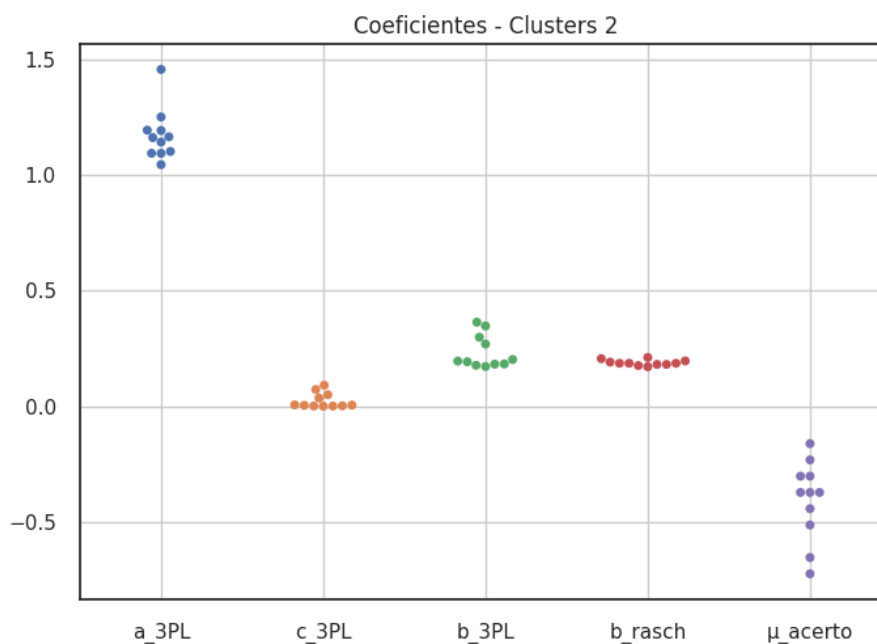


Figura 33 – Coeficientes TRI nos Grupos 1 e 2

(a)



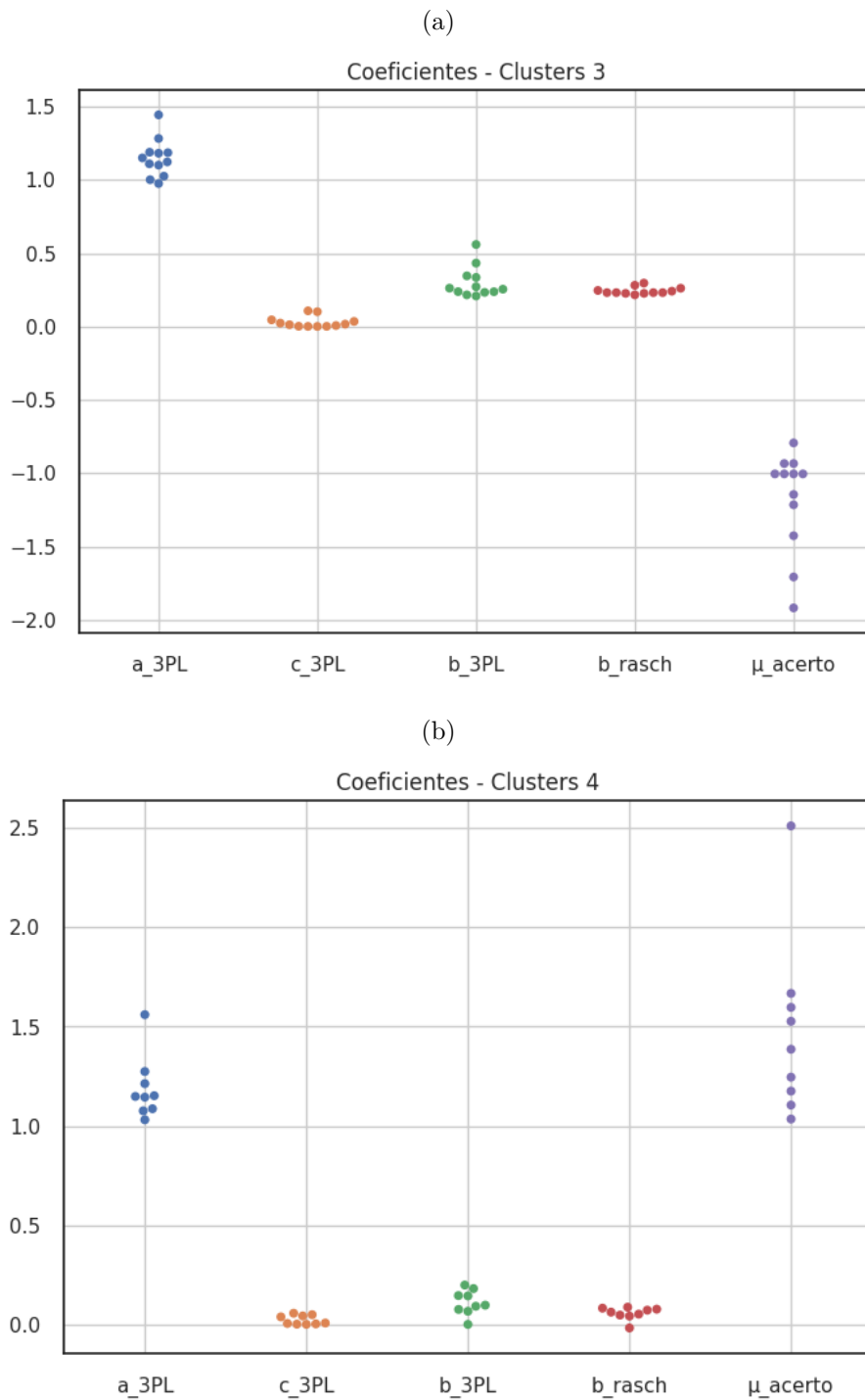
(b)



na região de valores positivos. Neste agrupamento também ocorrem valores expressivos de propensão a palpite, demonstrado pelo coeficiente c do modelo de três parâmetros. A Figura 33b apresenta valores de discriminação superiores à figura anterior. No entanto, sua média de acerto fica na região de valores negativos. Isso demonstra que itens de maior discriminação tendem a ser mais difíceis e ter menor proporção de acerto.

Na Figura 34a encontram-se os coeficientes de menor valor da média de acerto dos

Figura 34 – Coeficientes TRI nos Grupos 3 e 4



itens. Na Figura 34b localizam-se os itens de maior valor de média de acerto e menores valores de parâmetro de dificuldade em ambos modelos avaliados.

Recorrendo aos rótulos gerados no agrupamento para verificar quais itens são determinantes para enquadrar alunos entre um cluster e outro, elaboramos a árvore de decisão demonstrada na Figura 35. Na raiz da árvore, o item 31 ($a = 1,25; b = 0,14; c = 0$) aparece como fator principal de segregação entre os 4 clusters. No segundo nível da árvore,

observamos os itens 18 ($a = 1, 15; b = 0, 24; c = 0$) e 32 ($a = 1, 28; b = 0, 21; c = 0$).

Tomando de exemplo o quinto e último nível da árvore, temos o item 15 ($a = 1, 44; b = 0, 43; c = 0, 10$), que aparenta grande propensão a palpíte, dado seu valor de coeficiente e também ao seu elevado índice de dificuldade. Seguindo os caminhos da árvore, trilhas podem ser mapeadas. A seguir, comparamos a qualidade preditiva e a quantidade de características necessárias para equiparar os modelos da TRI com algoritmos de máquina de aprendizagem.

4.3.4 Comparação com Métodos de Aprendizagem de Máquina

O experimento comparativo descrito nessa seção utilizou a matriz de resposta dos alunos como entrada para os modelos regressores, variando a quantidade de itens (15, 30, e 45), sorteados randomicamente a cada rodada de execução. Foram realizadas 300 rodadas e o resultado agregado pelo valor médio das métricas calculadas. Os algoritmos utilizados foram: Regressão *Ridge*, Regressão Bayesiana *Ridge*, MPLPRegressor, Máquina de Suporte a Vetor (SVM), Gradiente Descendente Estocástico (SGD), Regressão Linear, Regressão Lars e Regressão Lasso. Tais algoritmos integram a biblioteca *Scikit-Learn*². O conjunto de dados foi dividido em duas porções, uma de treino e outra de teste. Os valores preditos foram truncados para 0, quando menores de 0,5 e para 1, em casos contrários. Com isso, fizemos a avaliação das métricas usuais em aprendizagem de máquina: acurácia, precisão, *recall* e *f1*.

A Figura 36 mostra a média de valores de classificação para as métricas: *t-estatístico*, *p-value*, correlação de *pearson*, *erro residual médio*, *erro residual quadrático*, *acurácia*, *precision*, *recall* e *f1*. Com 30 itens, as técnicas *SVM* e *Ridge* obtiveram os melhores valores para acurácia e precisão. Os piores valores foram das técnicas *Lars*, abaixo de 0,69. Com 15 itens, a técnica *SGD* alcançou grande similaridade ao estimar os valores de θ , obtendo uma correlação de 0,98, um erro médio de 0,19 e erro residual quadrático de 0,4, apresentando-se uma boa forma de poupar itens num teste que queira preservar a apreciação dos itens por parte dos alunos examinados. Por fim, os resultados dos piores algoritmos verificados nesta última abordagem pode ser vistos na Figura 36. Dentre eles, as piores desempenhos foram dos Métodos Lasso, *Ridge* e *Lars*, ocupando as últimas posições da classificação vista no gráfico.

4.4 Conclusão

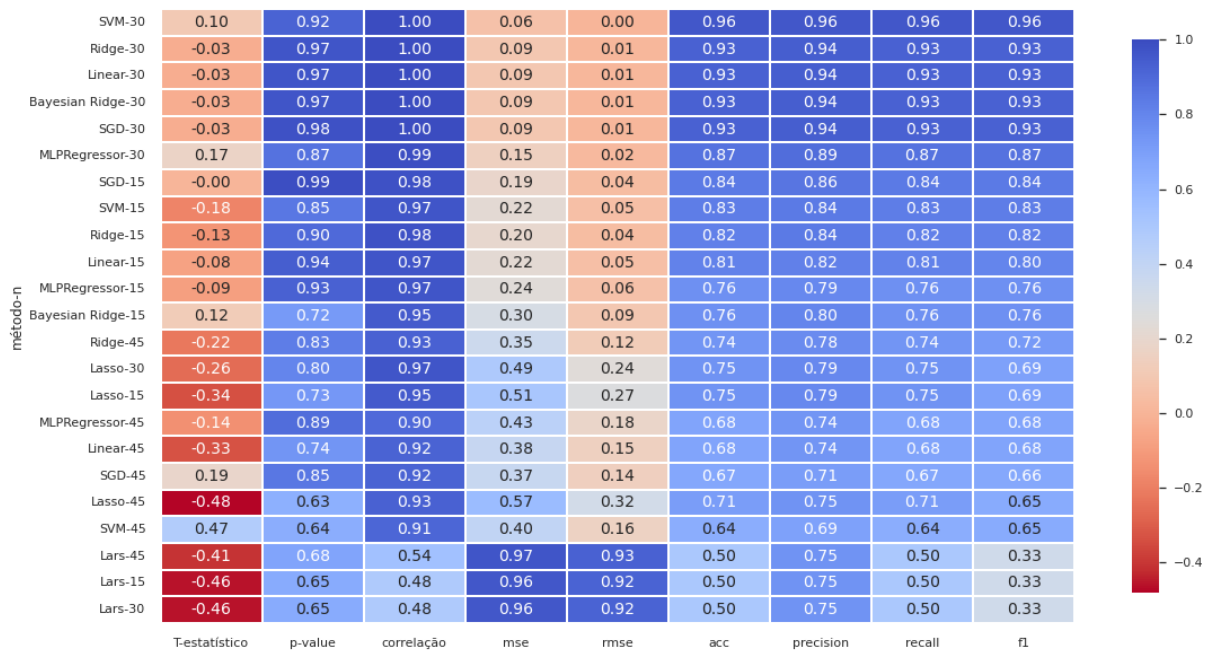
Todo mundo que ensina tem que gastar muito tempo criando exames, inspecionando e avaliando as atividades dos alunos para descobrir as lacunas de aprendizagem (MANGAROSKA; VESIN; GIANNAKOS, 2019). Os rastros deixados pelos ex-alunos em suas

² scikit-learn.org

avaliações representam dados estratégicos sobre a eficácia do processo ensino-aprendizagem. Alguns sistemas existentes (OLIVEIRA; CIARELLI; OLIVEIRA, 2013) podem recomendar atividades indicadas para perfis semelhantes que já receberam recomendações. No entanto, construir um caminho de avaliação adequado em que os níveis de conhecimento de cada aluno sejam frequentemente atualizados e contextualizados em todo o conjunto de itens disponíveis permaneceu uma questão em aberto.

Este capítulo propõe uma estratégia para selecionar as atividades adequadas por meio de um caminho de avaliação personalizado para cada aluno. Usamos dados históricos, ou seja, avaliações de alunos anteriores para construir um modelo estatístico para prever o sucesso futuro do aluno. Nosso sistema é suportado pelo uso comum de TRI e técnicas de agrupamento para gerar diferentes tipos de recomendações como filtros para selecionar atividades. A TRI garante o emprego de uma filtragem baseada em conteúdo relacionada às qualidades extrínsecas dos itens recomendados, enquanto as técnicas de agrupamento suportam a filtragem baseada em colaboração relacionada aos perfis intrínsecos dos alunos. Como sequência, discutimos a aplicação da TRI em testes de redação e comparamos com outras técnicas de regressão, para avaliar sua precisão e qualidade.

Figura 36 – Benchmark com Aprendizagem de Máquina. Fonte: elaborado pelo autor.



5 Avaliação de Proficiência em Redação por Análise de Regressão

No processo de ensino aprendizagem, a avaliação é um dos fatores de maior complexidade por envolver múltiplos aspectos a serem considerados e assegurados. A TRI, possui modelos matemáticos que permitem uma melhor compreensão do processo avaliativo (SHANMUGAM, 2019). Por exemplo, uma das ações que visam garantir a qualidade é a construção diligente de itens de avaliação. Isto porque é desejável que a avaliação consiga aferir nos examinandos o nível de habilidades (CURI; SILVA, 2019) e até mesmo, o nível cognitivo (SORREL et al., 2020; CURI et al., 2019). A identificação de itens de avaliação de qualidade depende de uma aferição criteriosa e amparada em métodos científicos para nortear a composição de estratégias de ensino eficazes (SILVA et al., 2020a). Outro ponto a ser considerado é o tipo de avaliação e o momento em que ela é aplicada no processo de ensino-aprendizagem (SILVA et al., 2020c). A aplicação de uma avaliação inadequada e num momento inoportuno, milita contra a qualidade do processo educacional. Para tanto, há que se distinguir tipos de avaliação e suas finalidades. Em projetos de ensino que não consideram tal separação, a identificação de problemas de aprendizagem tende a ser dificultada pelo viés das avaliações que usualmente tem intenção voltadas para fins de progressão acadêmica. Num cenário em que examinando está preocupado em atingir determinado patamar de pontuação, expressar o seu conhecimento de um modo mais amplo passa a ser tarefa secundária. Num contexto similar, o docente tende a abnegar o potencial informativo da avaliação.

O Exame Nacional do Ensino Médio, aplicado anualmente no Brasil desde o ano de 1998, é utilizado para fins de diagnóstico de aprendizagem e seleção para ingresso em cursos de nível superior, por exemplo. As edições de 2019 e 2020, contaram com um total de participantes de 5.09M e 5.89M, respectivamente¹. O custo da edição de 2020 foi da ordem de cerca de R\$ 682M. A prova realizada no ano de 2019 custou ao governo aproximadamente R\$ 537.7M. Logo, tem-se claro que é um valor expressivo que o governo investe para atender a uma grande parcela da população. Explorar formas de otimizar a realização da prova, bem como, seus dados históricos disponibilizados após a realização das avaliações, é uma oportunidade interessante para pesquisadores independentes.

Neste capítulo propomos o uso de TRI e de técnicas de aprendizagem de máquina para identificar perfis de aprendizagem e prever notas. Para demonstrar sua valia, fizemos um experimento utilizando os dados de redação do Enem. A contribuição deste estudo está em evidenciar o uso da teoria de resposta ao item, na prática, e aproximar

¹ <https://www.gov.br/inep/pt-br/assuntos/noticias/enem/divulgados-os-resultados-finais-do-exame>

seus resultados com técnicas de máquina de aprendizado. Com 10K instâncias e três características, das cinco disponíveis, para a calibração do modelo preditivo, alcançamos acurácia de 0,76. Recorrendo às cinco características, o modelo alcançou acurácia de 0,99. Ademais, a estimação de habilidade com o modelo de da teoria de resposta ao item, chegou à correlação de *pearson* de 0,92, e *erro quadrático residual* de 0,16, se comparado aos valores de habilidade informados no conjunto de dados de teste. Tais dados nos levam a crer na valia do uso da teoria de resposta ao item para incrementar o processo avaliativo de habilidades redacionais.

5.1 Trabalhos Relacionados

O experimento de [Neto et al. \(2020\)](#) tratou da correção automática de respostas textuais, por meio de similaridades. Conscientes da importância da tipificação das avaliações, outro ponto igualmente sensível é o tratamento matemático e estatístico das respostas coletadas nas avaliações. O estudo de [Oliveira et al. \(2021\)](#) demonstrou a possibilidade de correções automáticas de redações, seguindo o critério do Enem, com uma abordagem que considera o uso de entidades nomeadas associado a características básicas de produção de texto como ortografia e concordância. O estudo de [Smith, Chimedza e Bührmann \(2021\)](#) enfrenta o problema de dados faltantes na predição de evasão do ensino superior, demonstrando que a predição melhora com a técnica de imputação de dados baseada em *Random Forest*, a qual lida bem com grandes volumes de dados para decisão. Relativo à correção automática de redações, o levantamento realizado por [Costa, Oliveira e Castro Júnior \(2020\)](#) apresenta como métricas de avaliação a correlação de *pearson*, a média absoluta do erro. Além disso, o mesmo trabalho lista o método *Support Vector Machine* (SVM) como uma das principais técnicas de regressão utilizadas na literatura. O estudo de [Ferreira \(2019b\)](#) trata de fuga de temas, utilizando um conjunto de 2.1K textos de e a sistemática do Enem, com acurácia variando entre 0,81 e 0,96. Para além do tratamento do texto de resposta do aluno para avaliação da redação, há também emprego do processamento de texto para catalogação de itens de avaliação ([LIMA et al., 2021](#)).

As pesquisas de [Zaffalon et al. \(2019\)](#) descrevem a proximidade das estimativas de habilidade realizadas por TRI comparada as estimativas de uma adaptação do ELO, um modelo originário da classificação de jogadores de xadrez. Na educação, tal aplicação foi demonstrada com avaliação de atividades de programação ([ZAFFALON et al., 2020](#)). Ainda sobre ensino de programação, o Modelo de Resposta Gradual da TRI é utilizado para estimação e avaliação das dificuldades de desenvolvimento de aplicações mobile no estudo de [Da Cruz Alves et al. \(2020\)](#). Extrapolando seu uso dissociado, verifica-se na literatura a TRI combinada com outras técnicas, como a pesquisa de [Brito e Motta \(2014\)](#) que utilizou fez isso para proceder uma avaliação diagnóstica no ensino de matemática e recomendar jogos como itens de aprendizagem. Na área de matemática e com dados do

Exame Nacional do Ensino Médio — ENEM, [Spenassato et al. \(2016\)](#) demonstraram o uso de TRI integrado a um teste adaptativo computadorizado, cujos resultados evidenciaram a possibilidade de reduzir o comprimento de um teste, da precisão dos valores encontrados. Igualmente, encontramos a aplicação de TRI em ambientes virtuais de aprendizagem, como o estudo de [Romero, Machado e Wives \(2019\)](#) que apresenta um modelo para planejamento e criação de questionários adaptativos, neste tendo como suporte a análise histórica de questionários. Na revisão sistemática realizada por [Jatobá et al. \(2017\)](#), nota-se ampla utilização da TRI em testes adaptativos, bem como, no provimento de *feedback*, conscientizando o aluno de sua habilidade estimada. Ademais, o mesmo estudo de [Jatobá et al. \(2017\)](#) demonstra a majoração de modelos de TRI modelo dicotômicos, comparado ao modelo politômico de resposta gradual da TRI, bem como há predominância de utilização de dados reais nos experimentos, em comparação a dados artificiais. Anteriormente a isso, o estudo de [Henrique et al. \(2009\)](#) apresentou a utilização da TRI para criação de perfis de aprendizagem, baseados na frequência de acerto dos itens por faixa de categoria, para estimação de proficiência do uso de tecnologias. Mais recentemente, a pesquisa de [SANTOS et al., 2020](#), registrada no Comitê de Ética da Universidade Federal de Campina Grande, apresentou o emprego da TRI na identificação de habilidades de leitura em crianças para identificação de traços de dislexia na fase de alfabetização. A avaliação do modelo proposto por [Santos et al. \(2020\)](#) é a média e o desvio padrão da habilidade estimada, aliado ao coeficiente *Cronbach's Alpha*. A pesquisa de [Luo et al. \(2020\)](#) aponta um resultado curioso na aplicação de TRI na administração de itens para públicos de heterogeneidade variada, o resultado dos itens sofre variação. O que evidencia a necessidade de considerar a adaptabilidade a priori de itens conforme o perfil predito para o sujeito examinado. No trabalho de [NETO et al., 2020](#), é apresentada a correção automática de textos por meio de técnica *random forest* de *machine learning*, com um resultado bem próximo ao desempenho humano.

5.2 Arquitetura do Experimento

Nesta seção, apresentamos os materiais e métodos utilizados no experimento, bem como a estrutura utilizada. Partindo do contexto da fonte dos dados utilizados, da estrutura de avaliação da redação e da caracterização da aprendizagem sustentada pelos resultados de proficiência esperados na prova. Por fim, apresentamos a abordagem de previsão de notas de avaliação, utilizando técnicas de aprendizado de máquina.

5.2.1 Ferramentas e Métodos

Os recursos computacionais utilizados na realização deste experimento foram:

- *Google Colaboratory*²: foi usado o ambiente de experimento do Google, fazendo uso da linguagem de programação *Python*;
- *Scikit-Learn*: foram as bibliotecas de máquina de aprendizado e ferramentas estatísticas;
- *Georgia Tech Item Response Theory*: biblioteca de estimação de parâmetros por máxima verossimilhança da Teoria de Resposta ao Item³;

Os códigos gerados estão disponíveis para consulta⁴. A seguir, são descritas de modo sucinto, os conceitos empregados no experimento proposto.

5.2.1.1 Teoria de Resposta ao Item: Modelo de Resposta Gradual

A Teoria de Resposta ao Item, oriunda da psicometria, possui modelos matemáticos para tratar respostas com maior precisão em relação ao examinado e às características dos itens avaliativos (WU et al., 2020). Dentre seus modelos, há aqueles voltados para respostas dicotômicas (SILVA et al., 2020c; SILVA et al., 2020b), os quais tratam de características como dificuldade, discriminação e propensão ao palpite. Há também modelos voltados para respostas politômicas (ZHAN; WANG; LI, 2020), as quais buscam identificar o níveis de habilidade nos examinandos. Inspirado nos resultados de (CURI; SILVA, 2019), o Modelo de Resposta Gradual de (SAMEJIMA, 1997) foi selecionado para este experimento. Trata-se de uma generalização do modelo logístico de dois parâmetros em que as categorias de respostas podem ser ordenadas entre si, com intuito de obter mais informação do examinado, para além do acerto e erro.

Supondo que o escore das categorias de um item i são arranjados ordenadamente do menor para o maior, denotados por $k = 1, 2, \dots, m_i$, em que $m_i + 1$ é o número de categorias do i -ésimo item. Tem-se que a probabilidade de um indivíduo, em que, em que j selecionar uma particular categoria de resposta do item, i dada pela função expressa na Equação 5.1. com $i = 1, 2, \dots, I; j = 1, 2, \dots, n; ek = 0, 1, \dots, m_i$, em que:

$$P_{i,k}^+(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} \quad (5.1)$$

- I) $P_{i,k}^+(\theta_j)$: é a probabilidade de sucesso no item i , para a categoria k , do sujeito j ;
- II) $b_{i,k}$: é o parâmetro de dificuldade da k -ésima categoria do item i . Por definição, tem-se que: $b_{i,1} \leq b_{i,2} \leq \dots \leq b_{i,m_i}$;

² colab.research.google.com

³ eribeau.github.io/girth/

⁴ gitlab.com/rii_lcad/enem2019

- III) θ_j : denota o traço latente observado no j -ésimo sujeito, ora também chamado de habilidade ou proficiência;
- IV) a_i : é o parâmetro de discriminação (ou de inclinação) do item i ;
- V) D : é o fator de escala constante e igual a 1. Utiliza-se o valor 1,7 para gerar uma função logística semelhante a uma função ogiva normal;
- VI) $-Da_i(\theta_j - b_{i,k})$: logit do item i para a categoria k .

Deste modo, a probabilidade de um indivíduo j receber um escore k no item i é dada pela Equação 5.2 (SAMEJIMA, 1997).

$$P_{i,k}(\theta_j) = P_{i,k}^+(\theta_j) - P_{i,k+1}^+(\theta_j) \quad (5.2)$$

Considerando ainda que:

- I) $P_{i,0}^+(\theta_j) = 1$,
- II) $P_{i,m_i+1}^+(\theta_j) = 0$,
- III) $P_{i,0}(\theta_j) = P_{i,0}^+(\theta_j) - P_{i,1}^+(\theta_j)$,
- IV) $P_{i,0}(\theta_j) = 1 - P_{i,1}^+(\theta_j)$

Logo, $P_{i,m}(\theta_j) = P_{i,m}^+(\theta_j) - P_{i,m+1}^+(\theta_j)$, com $P_{i,m}(\theta_j) = P_{i,m}^+(\theta_j)$. Assim a Equação 5.2 pode ser reescrita do modo a seguir:

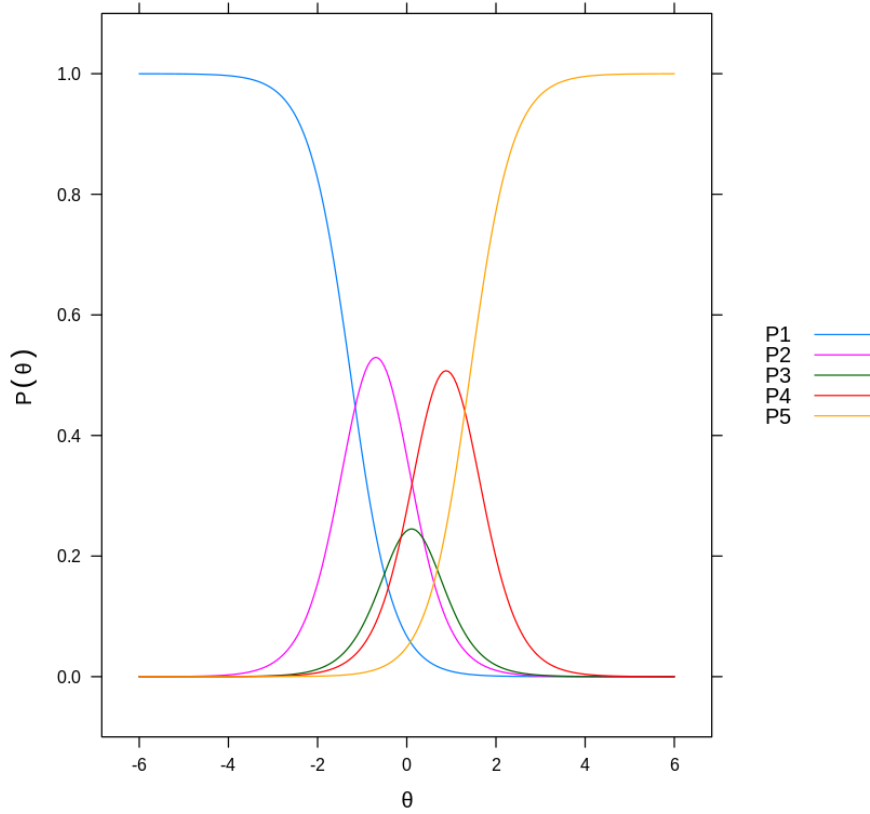
$$P_{i,k}(\theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k})}} - \frac{1}{1 + e^{-Da_i(\theta_j - b_{i,k+1})}} \quad (5.3)$$

Com um item de $m_i + 1$ categorias, m_i valores de dificuldade farão parte da composição do modelo e precisarão ser estimados, além do parâmetro de discriminação, responsável pela inclinação da curva de característica do item. O número de parâmetros a ser estimado é dado pela quantidade de categorias de resposta. Logo, para um teste com I itens, cada um com $m_i + 1$ categorias de resposta, teremos $[\sum_{i=1}^I m_i + I]$ parâmetros de itens a serem estimados. Além dos valores estimados para os parâmetros e probabilidade do sujeito no teste, os itens podem ser avaliados graficamente por suas curvas de característica e quantitativamente por sua função de informação.

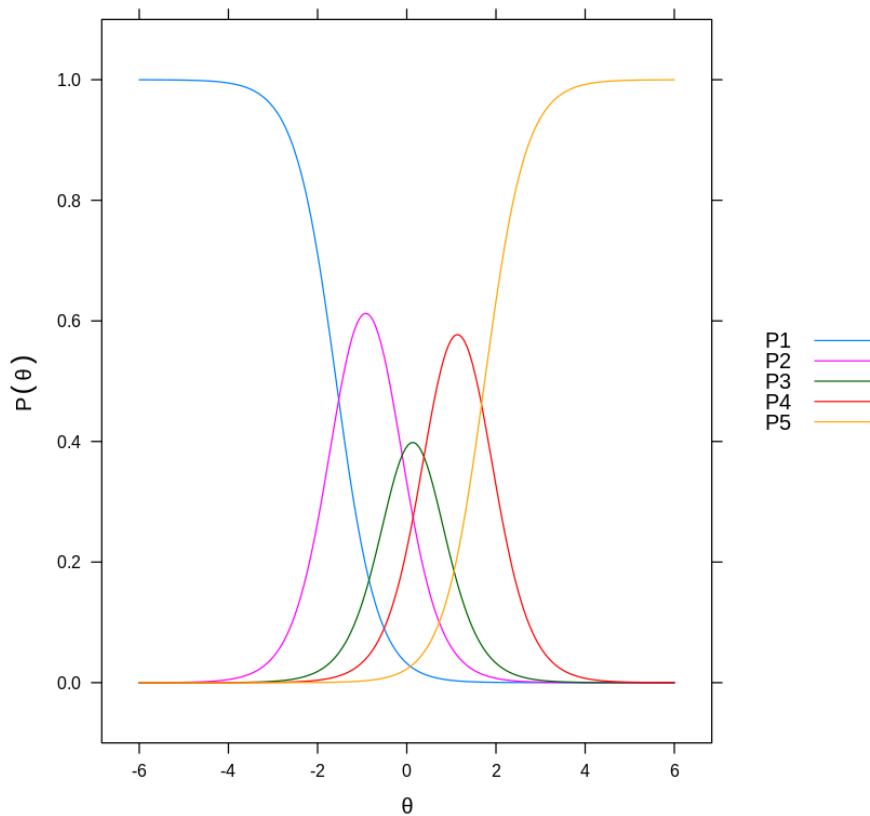
A curva de característica é a função que evidencia a relação existente entre a probabilidade de sucesso $P(\theta)$ e a habilidade do examinando, designada pela variável θ . Em outras palavras, pelo gráfico da curva de característica do item, vemos a relação $P(\theta = 1|\theta)$.

Figura 37 – Curvas de Características do Modelo de Resposta Gradual

(a) Exemplo 1

Probability Function for Item 1

(b) Exemplo 2

Probability Function for Item 10

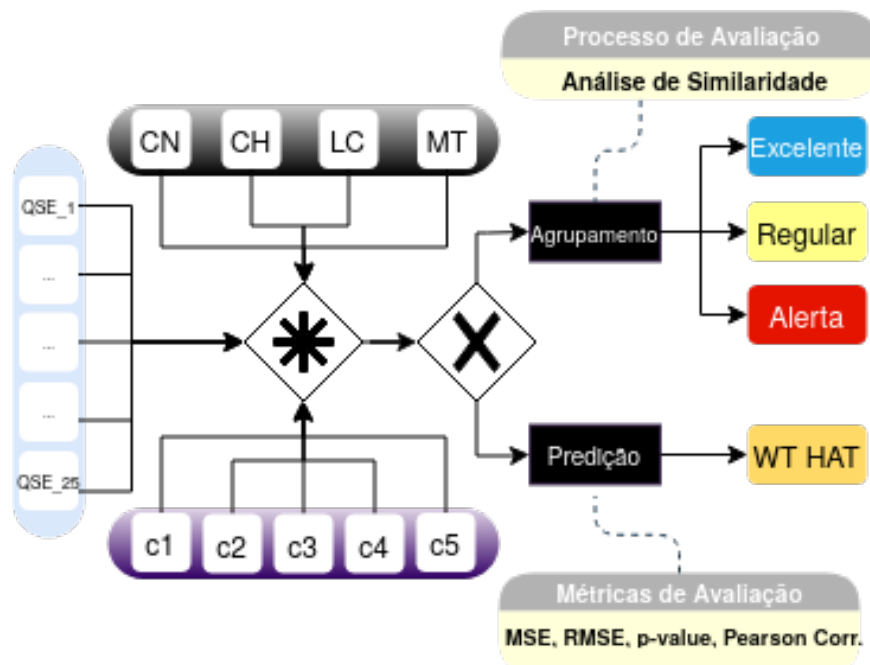
Na Figura 37a, é possível observar que o sujeito de $\theta = 0$, tem 30% de chance de selecionar P2 ou P4, nas cores rosa e vermelho, respectivamente. O mesmo sujeito, teria cerca de 25% de probabilidade selecionar P3. Logo, sujeitos de menor capacidade teriam maior probabilidade de selecionar os opções à esquerda de 0, enquanto sujeitos de maior capacidade teriam maior probabilidade de selecionar opções à direita de 0. Na Figura 37b, as probabilidades P2 e P4 tem pico próximo a 0.6, na figura anterior, as probabilidades intermediárias não alcançam a mesma marca.

A caracterização dos itens por meio da TRI oportuniza a avaliação qualitativa a partir dos valores quantitativos observados na estimação de parâmetros dos itens. Igualmente, o modelo logístico ocasiona a estimação de sucesso do examinando, permitindo o uso dos itens que melhor se adéquem à finalidade da avaliação.

5.2.2 Arquitetura Proposta

A Figura 38 mostra os subconjuntos de dados que são usados na arquitetura do experimento. O vetor cinza representa as pontuações dos testes objetivos, sendo Ciências Naturais (CN), Ciências Humanas (CH), Linguagem e Comunicação (LC) e Matemática (MT). O vetor de notas amarelas simboliza os pontos obtidos nas competições do Teste de

Figura 38 – Fluxo de Informação do Experimento. Fonte: (SILVA et al., 2021b)



Escrita. As competências variam de $c1$ a $c5$, aumentando o nível de habilidade exigido de cada um deles. Por fim, o vetor azul claro, com as respostas do questionário econômico e social. As questões têm os seguintes grupos temáticos: QSE1 e QSE2: níveis de instrução familiar; QSE3 e QSE4 ocupação econômica familiar; QSE5 a QSE9 habitação e rendimento; transporte veicular QSE10 e QSE11; padrão de conforto QSE12 a QSE20; QSE21 a QSE25

para comunicação e acesso à informação. Como forma de avaliar os modelos utilizados no experimento, propomos o uso da análise de similaridade para agrupamento. A avaliação da qualidade da predição é feita através da média do resíduo quadrático - mse; a raiz quadrada da média do resíduo quadrático - rmse; calculando o valor p; e pela *correlação de pearson*, similar ao visto no trabalho de (COSTA; OLIVEIRA; Castro Júnior, 2020). O fluxo proposto na Figura 38 demonstra a possibilidade de nossa abordagem em selecionar variáveis a partir das informações disponíveis sobre o teste e gerar o agrupamento por meio de métodos paramétricos ou não paramétricos. Alternativamente, também podemos prever a pontuação esperada no Teste de Escrita pelo examinando.

5.2.3 Caracterização da Aprendizagem

Uma vez que o objetivo deste experimento é identificar perfis de aprendizagem, foram traçadas duas alternativas para evidenciar a proficiência e em redação dos sujeitos examinados. Seguimos duas abordagens, sendo a primeira unidimensional, levando em consideração apenas a nota final da redação, já ponderada pelas competências dos examinados. Fazendo uso da técnica estatística de avaliação da dispersão por meio do intervalos interquartil, definimos três grupos: *Grupo Alerta* - notas do primeiro quartil (inferior); *Grupo Intermediário* - notas do quartil intermediário; e *Grupo Excelente* - notas do terceiro quartil (superior).

A segunda forma, multidimensional, utilizando a técnica de agrupamento *K-Means*, de modo a processar distâncias entre as notas das competências, considerando-as como vetores de cinco variáveis: $C_{i=1}^5 = \{c_{i,j}\}$. O vetor do examinando i é representado pela variável e_i , quanto as notas de cada competência são designadas pela variável c . Na Figura 39a o gráfico de *boxplot* mostra o comportamento das distribuições por grupo, sendo o vermelho, de número 2, o grupo inferior. O azul, índice 0 no eixo x, o grupo intermediário. E por fim, grupo verde, de índice 1, figurando como grupo superior. Na sequência, a Figura 39b apresenta os grupos gerados a partir do cálculo multidimensional, neste caso, nota-se as sobreposições de valores nas caldas de todas três distribuições, denotando melhor o mundo real por relacionar as dificuldades em comum dos vetores com valores próximos. Pelo resultado, demonstra-se equivocado analisar apenas a nota final por não representar a dimensionalidade dos dados e as competências avaliadas no Enem. Para efeito do experimento, consideramos a existência de três grupos. Mas analisando o conjunto de dados pela média de similaridade interna dos vetores por grupo, a exemplo do que fez (SYAKUR et al., 2018), notou-se que o equilíbrio ocorre com a divisão em 6 grupos. A Figura 40a mostra o experimento realizado com 100; 1K; 10K; e 100K instâncias. O melhor ajuste da similaridade interna dos agrupamentos ocorre com $k = 6$. A verificação da concordância do agrupamento unidimensional e multidimensional resultou nos seguintes valores *accuracy*: 0.536; *precision*: 0.651; *recall*: 0.536; e *f1*: 0.579. Por fim, trabalhando

Figura 39 – Distribuição de Densidades por Cluster: Métodos Inter-Quartil e K-Means.
Fonte: (SILVA et al., 2021b)

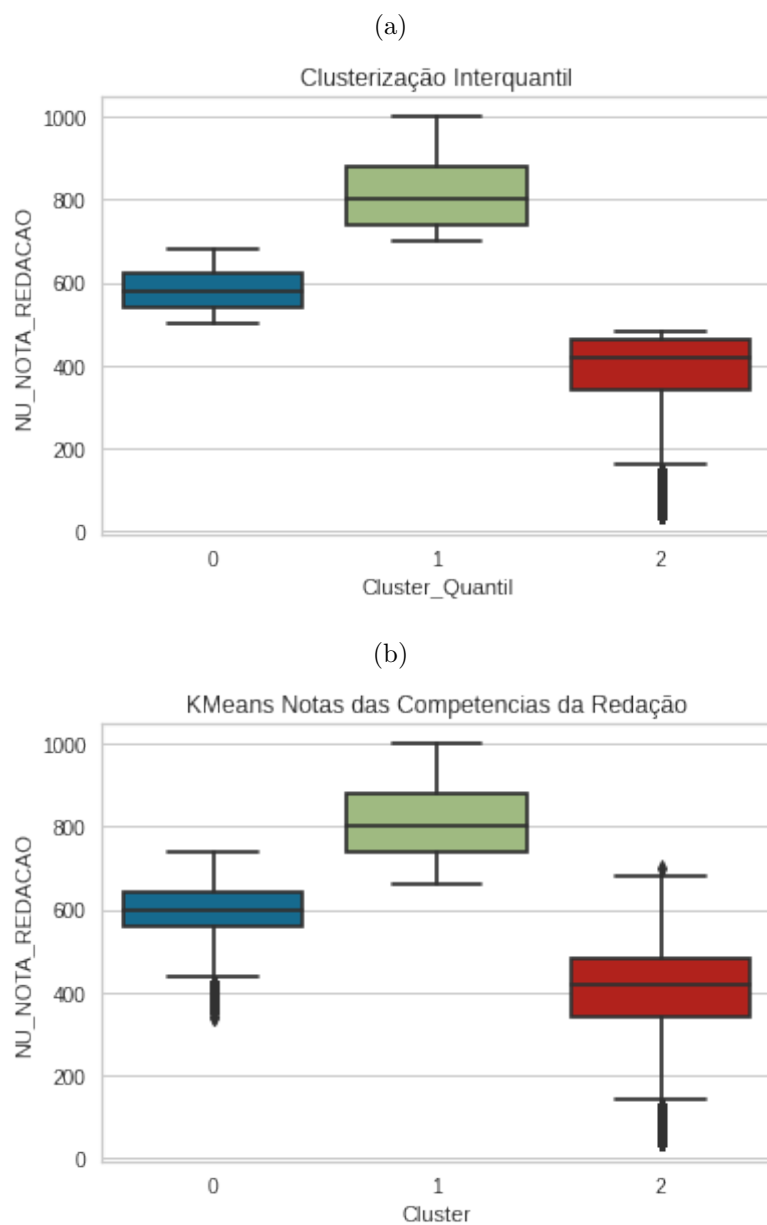
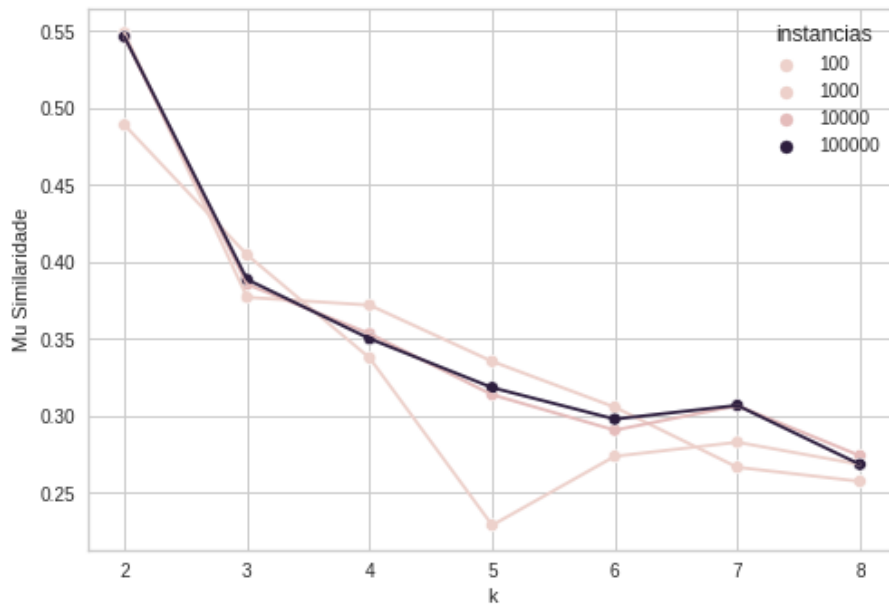
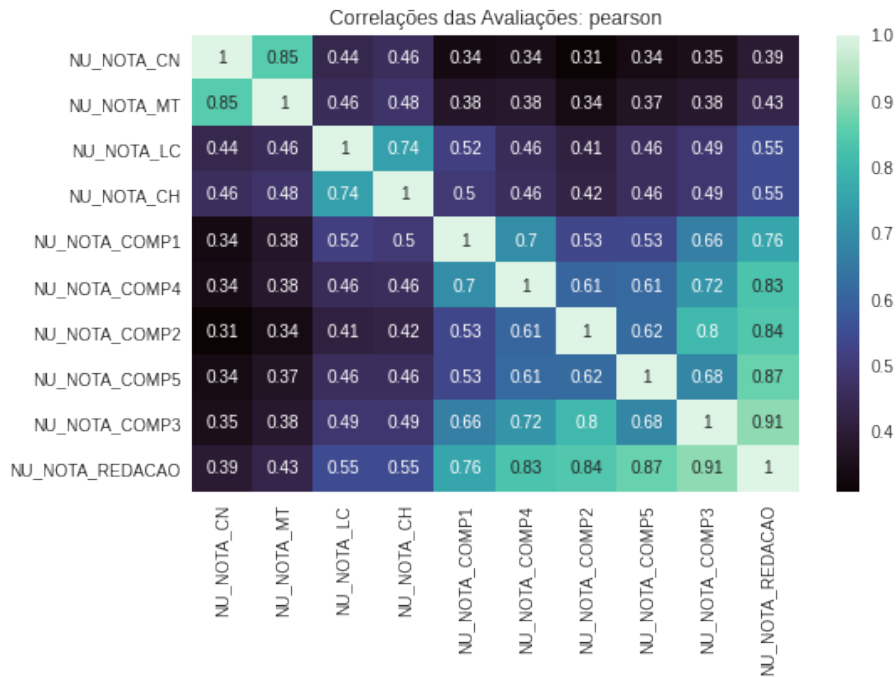


Figura 40 – Análise de Agrupamento e Correlações das Notas. Fonte: (SILVA et al., 2021b)

(a) Melhor Ajuste de K



(b) Correlação das Habilidades



com os valores das notas após agrupamento, foi buscado verificar existência de correlação, porém sem resultados melhores que a correlação das notas no valor absoluto. A Figura 40b apresenta os valores encontrados a partir do cálculo da correlação de *pearson*. Não foi identificada correlação significativa entre as notas objetivas e as notas da redação. Todavia, notou-se que as Competências 3 e 5 são as melhores correlacionadas à nota final da redação.

5.3 Resultados e Discussão

Os métodos e tecnologias aplicados aos dados dispostos no presente experimento buscaram caracterizar os itens de avaliação do Enem com os parâmetros do modelo de resposta gradual de Samejima, com seus valores e curvas de características. Para traçar o perfil dos examinados, lançamos mão da ação de agrupar os vetores de resposta das competências dos alunos para verificar a proporção dos grupos de aprendizagens de baixa, média e alta performance. Com isso, foi possível verificar a quantidade de instâncias e competências necessárias ao modelo para ter uma aproximação significativa da habilidade do examinado; e, por fim, avaliar a qualidade da predição gerada pelos modelos de máquina de aprendizagem comparados ao modelo de (SAMEJIMA, 1997). A seguir, descrevemos pormenorizadamente os resultados das ações elencadas.

5.3.1 Caracterização dos Itens

A Figura 41a mostra as curvas de características geradas com 500 vetores. A Figura 41b, representa a funções de informação com os parâmetros encontrados. A Figura 42a mostra as curvas de características geradas com 100 mil vetores. A Figura 41b, representa a funções de informação com os parâmetros encontrados. Ficando melhor acentuado que a estimativa com 500 vetores, todavia, mantendo o traçado como preconizado pelo modelo.

A Figura 43a mostra a densidade dos valores encontrados para os coeficientes, como esperado na TRI para Modelo de Resposta Gradual, os valores crescem de b1 para b10, ficando cada vez mais estreitos em torno da média, salvo para os coeficientes b5 e b7, que apresentaram maior intervalo interquartil, o que implica num maior desvio padrão. A Figura 43b mostra os valores dos coeficientes estimados agrupados por competências, demonstrando equilíbrio global entre as distribuições dos valores de dificuldade para os itens de avaliação, haja vista que os *boxplots* estão alinhados, com predominância do desvio padrão na competência 1(C1).

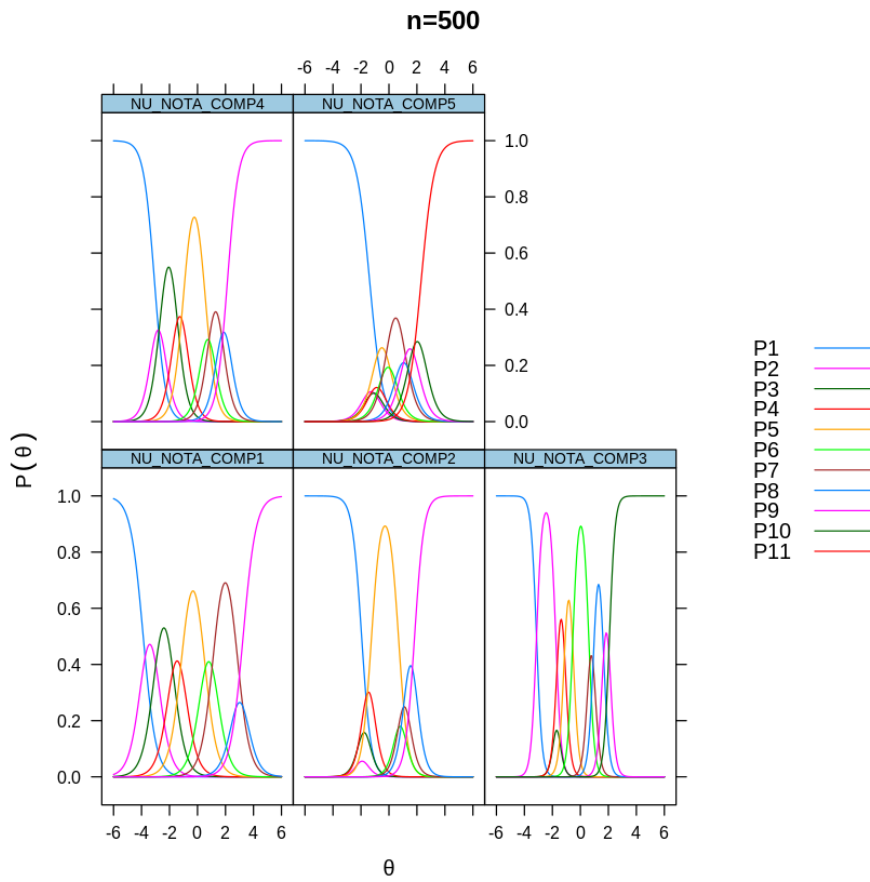
5.3.2 Avaliação do Tamanho da Porção de Instâncias para Treino e Quantidade de Características

Para discutir este objetivo, três tabelas são utilizadas. A Tabela 3 lista os 5 piores valores de correlação, sendo que varia de 0,71 com 100 instâncias e 1 Competência, listada na coluna *feat*. Os demais valores da tabela são ocupados pelo Modelo de *Samejima*. Os valores de resíduo quadrático são todos bem próximos, assim como a variação de *pearson*, todavia, o melhor valor de *p-value* ocorre com 1.5k instâncias de treino.

A Tabela 4 lista os 5 piores erros residuais médios, todos capitaneados pela técnica de Regressão *Least Angle Regression* (Lars). Nota-se ainda que apesar de ser a pior taxa de erro, os valores de *pearson* continuam elevados, sendo o menor 0,83. com apenas 100

Figura 41 – Modelo Samejima $n=500$. Fonte: (SILVA et al., 2021b).

(a) Curvas de Características



(b) Informação

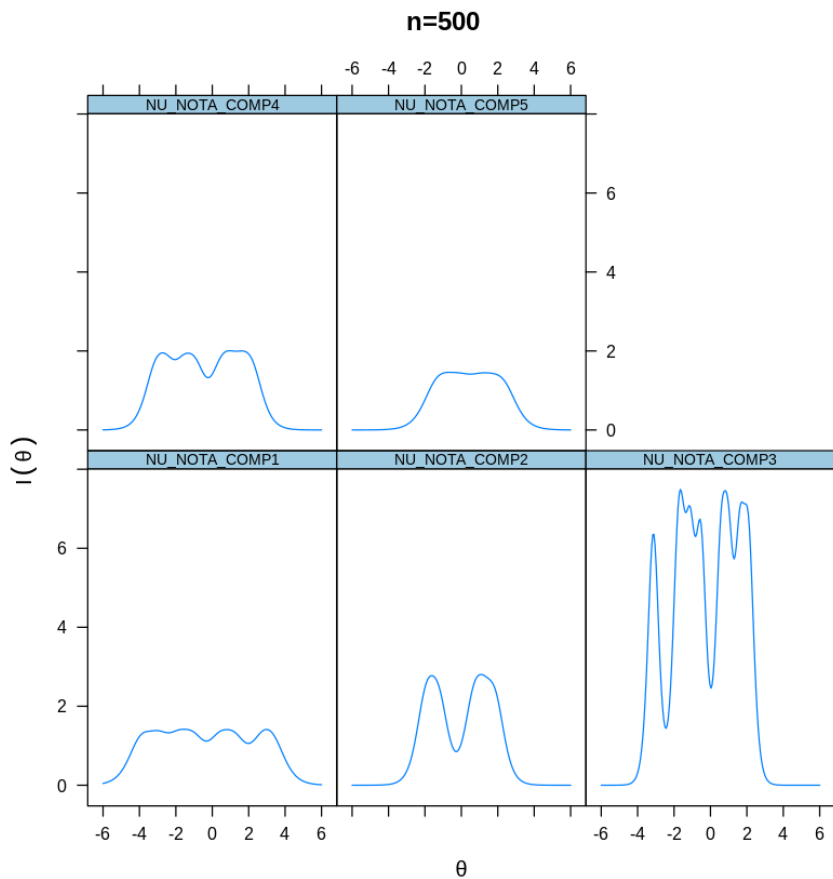


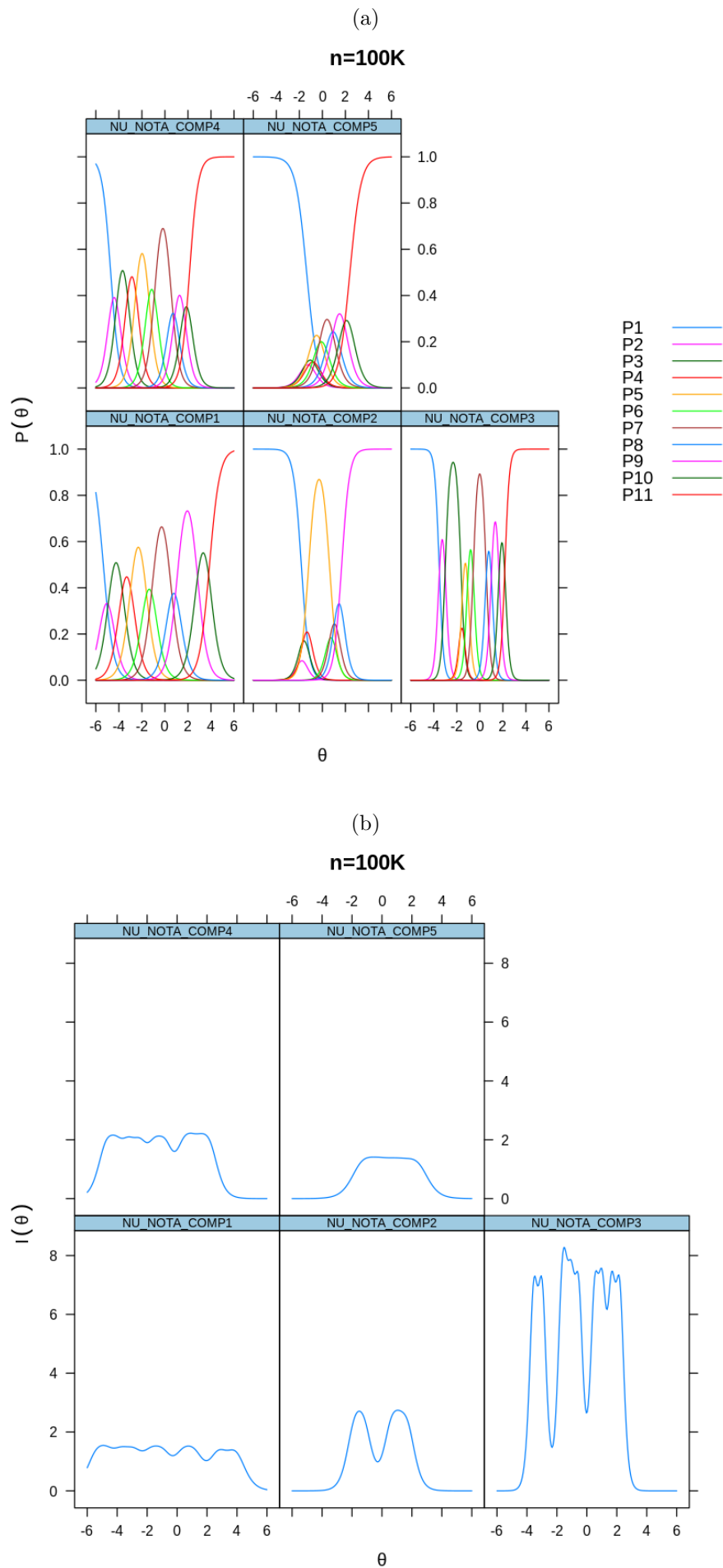
Figura 42 – Modelo Samejima $n=100k$. Fonte: (SILVA et al., 2021b).

Figura 43 – Distribuição de Densidades. Fonte: (SILVA et al., 2021b).

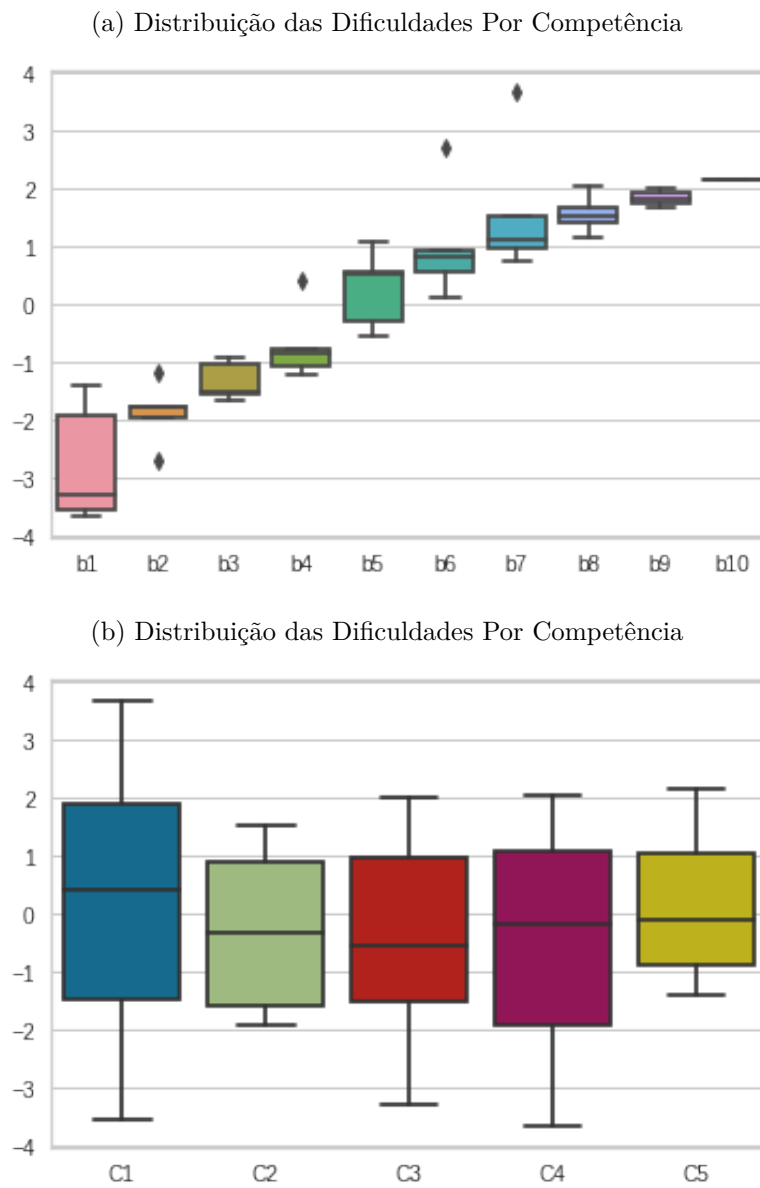


Tabela 3 – Top 5 Piores Valores de Correlação de Pearson. Fonte: (SILVA et al., 2021b)

Método	n	Feats	p-value	MSE	pearson	t
SVM_WT	100	1	0.62	0.65	0.71	0.01
IRT_WT	100	1	0.72	0.65	0.75	-0.40
IRT_WT	1000	1	0.24	0.62	0.79	-1.55
IRT_WT	1500	1	0.03	0.62	0.80	-2.52
IRT_WT	2000	1	0.31	0.60	0.80	-1.29

Tabela 4 – Top 5 Piores Valores de MSE. Fonte: (SILVA et al., 2021b)

Método	n	feats	p-value	MSE	pearson	t
Lars_WT	1000	2	0,48	0,86	0,89	0,63
Lars_WT	2000	2	0,63	0,85	0,89	0,58
Lars_WT	500	2	0,63	0,84	0,87	0,18
Lars_WT	1500	2	0,37	0,83	0,89	-0,33
Lars_WT	100	4	0,70	0,83	0,94	0,37

Tabela 5 – Top 5 Melhores P-values. Fonte: (SILVA et al., 2021b)

Método	n	feats	p-value	MSE	pearson	t-stat
IRT_WT	1500	1	0,03	0,62	0,80	-2,52
IRT_WT	2000	5	0,06	0,22	0,98	-2,02
IRT_WT	1500	4	0,14	0,28	0,96	-1,77
IRT_WT	2000	4	0,19	0,28	0,96	-1,68
IRT_WT	2000	2	0,21	0,41	0,91	-1,69

instâncias e 4 competências no modelo. Por sua vez, a Tabela 5 apresenta os melhores valores para p-value, indicando a confiabilidade estatística do método. O Modelo de Samejima com uma Competências e 1.5K instâncias foi capaz de alcançar um *p-value* de 0,03, porém com um MSE de 0,62. No entanto, a redução do MSE para 0,22 ocorre ao usar 2k instâncias e 4 Competências no Modelo. Essa versatilidade demonstra a possibilidade de usar mais ou menos competências nos modelos preditivos e estimar o resultado final do sujeito na realização do Enem.

Avaliação da Qualidade da Predição

A Tabela 6 lista todas as técnicas de aprendizagem de máquina utilizadas em nosso experimento, bem como o conjunto de dados de treino sendo os vetores de notas das Competências da redação (WT), as respostas do questionário sócio-econômico e o subconjunto do questionário sócio-econômico que trata especificamente de atividade econômica do examinando e dos seus pais ou responsável. De acordo com nossos resultados, listados na Tabela 6, as técnicas Regressão Bayesiana, Regressão Ridge, Regressão Linear ficaram empatadas, a Regressão Lasso e Técnica de SGE ficaram muito próxima deste primeiro grupo, variando na terceira casa decimal. O modelo de *Samejima* teve um comportamento semelhante à técnica de SVM. As técnicas utilizadas nos dados dos questionários sócio-econômicos se mostraram infrutíferas pelo resultados elevados de MSE e baixos de correlação de *pearson*.

Tabela 6 – Top 10 Melhores Valores de MSE. Fonte: (SILVA et al., 2021b)

Método	p-value	MSE	pearson	t
Bayesian Ridge_WT	0,853	0,260	0,943	-0,031
Ridge_WT	0,854	0,260	0,943	-0,031
Linear_WT	0,854	0,260	0,943	-0,031
Lasso_WT	0,853	0,261	0,943	-0,031
SGD_WT	0,853	0,264	0,943	-0,035
SVM_WT	0,772	0,307	0,932	0,182
TRI_WT	0,447	0,371	0,917	-0,796
Lars_WT	0,565	0,725	0,901	-0,113
Bayesian Ridge_esq_full	0,451	0,926	0,310	0,014
SVM_economic_acitivity	0,476	0,926	0,244	0,375

5.4 Conclusão

Todo mundo que ensina tem que gastar muito tempo criando exames, inspecionando e avaliando as atividades dos alunos para descobrir suas lacunas de aprendizagem (MANGAROSKA; VESIN; GIANNAKOS, 2019). As trilhas deixadas pelos alunos anteriores as avaliações representam dados estratégicos sobre a eficácia do processo de ensino-aprendizagem. Alguns sistemas existentes (OLIVEIRA; CIARELLI; OLIVEIRA, 2013) podem recomendar atividades indicadas para perfis semelhantes que já receberam recomendações. No entanto, construir um caminho de avaliação apropriado no qual os níveis de conhecimento de cada aluno sejam frequentemente atualizados e contextualizados em todo o conjunto de itens disponíveis permaneceu uma questão em aberto.

Este capítulo propôs uma estratégia criação de perfis e predição de notas para prover uma avaliação personalizado para cada aluno. Usamos dados históricos, ou seja, avaliações de alunos anteriores para construir um modelo estatístico a fim de prever o sucesso futuro do aluno. Nosso sistema é apoiado pelo uso de métodos TRI combinados com técnicas de TRI e técnicas de agrupamento para produzir diferentes tipos de recomendações como filtros para selecionar atividades. A TRI garante o emprego de uma filtragem baseada em conteúdo relacionada às qualidades extrínsecas dos itens recomendados, enquanto as técnicas de agrupamento apoiam a filtragem baseada em colaboração relacionada aos perfis intrínsecos dos alunos. Futuramente, pretendemos fornecer ao nosso sistema uma interface que suporte a seleção dos itens, seja utilizando as técnicas de agrupamento, seja aplicando os diferentes filtros apresentados neste trabalho.

6 Testes Adaptativos Computadorizados com Agentes Inteligentes

A organização metodológica dos processos de ensino-aprendizagem é um requisito primordial para assegurar qualidade e uniformidade aos resultados. A construção das inter-relações dos objetos de aprendizagem com os itens de avaliação é tarefa árdua e complexa para o educador. Parte da dificuldade ocorre em função da necessidade de reflexão sobre as melhores formas de transmitir os conceitos associados aos saberes almejados (ALVES; MEDEIROS; MELO, 2020). A precisa mensuração dos resultados alcançados serve de balizador das práticas, de modo a orientar o que pode ser melhorado. De modo geral, na construção de um projeto de curso, o esforço docentes para empreender a arquitetura da aprendizagem tende a ser conjugado com pares, a fim de fomentar a discussão e abordar pontos de vista variados, os quais componham um espectro de aprendizagem ajustado ao público pretendido (SILVA et al., 2020a). Perpassada a fase de concepção do projeto do curso, o plano de ensino de cada aula, representa um contrato didático-pedagógico de instância granular que assegura o alcance dos objetos de aprendizagem intencionados.

Diferentemente do apoio dos pares na elaboração de projetos de ensino, o labor docente tende a ser desassistido na elaboração do plano de ensino e demais produtos a ele associados, como por exemplo: apresentações em laminas; atividades de fixação do conteúdo; leituras recomendadas e itens de avaliação. De um lado, tal fato permite maior autonomia nas proposições, ocasionando flexibilidade aos processos, de modo a se ajustar aos aprendizes. De outro, há o risco de trabalhar com metodologias variadas, que dificultem a aferição dos seus resultados ao longo do tempo. Os ambientes virtuais de aprendizagem oportunizam a elaboração de acervos de conteúdo e de itens de avaliação. Apesar de oportuna, tal prática não é amplamente empreendida nos sistemas de ensino da rede pública, por exemplo. Logo, o retrabalho se torna praxe da atuação docente. A perspectiva de composição de acervos acadêmicos, que na academia já demonstra empenho no saneamento dos dados da produção acadêmica (BOURGUET; SILVA; OLIVEIRA, 2021), ainda não é abrangente com demais produtos dos processos de ensino-aprendizagem. A disponibilização digital dos produtos textuais do ensino poderiam ser melhor utilizadas com ferramentas de processamento de linguagem natural e enriquecimento semântico, como visto no trabalho de Oliveira et al. (2020), realizado com textos jornalísticos. Deste modo, percebe-se um rol amplo de possibilidades de manejo inteligente de avaliações de aprendizagem e seus itens de avaliação, materializados em ferramentas de apoio ao docente. Portanto, a redução do esforço docente e socialização as informações geradas na aprendizagem pode ser amparada por técnicas de Inteligência Artificial. São exemplos a

Figura 44 – Fluxograma de um CAT. Fonte: Elaborado pelo Autor.



caracterização de lacunas da aprendizagem, como narramos no Capítulo 3; a caracterização dos itens de aprendizagem, discutidos nos Capítulos 4 e 5.

A provisão de *feedback* sobre os níveis de aprendizagem atingidos e os perfis acadêmicos associados, pode ser automatizada via recursos computacionais. Com isso, atenuar o tempo gasto com esforços manuais e repetitivos, oportunizando ao docente um maior envolvimento com as problemáticas da aprendizagem. Outro ponto a ser destacado é a necessidade de estabelecer métricas de aferição da aprendizagem ajustados ao contexto de ensino e precisas — no que tange a métodos estatísticos e matemáticos. A implementação de medidas precisas, isto é, baseadas em modelos explicativos como aqueles apresentados na TRI, oportuniza evidenciar pontos de melhoria do processo de ensino, cuja suspeição ocorre de maneira intuitiva, mas que eventualmente, em função da carência de prova material, deslegitima o enfrentamento.

Conforme discutimos nos capítulos anteriores, a etapa de avaliação da aprendizagem é um momento crítico dos processos de ensino/aprendizagem. Por isso, reunir um arcabouço de conhecimento didático-pedagógico, amparado por métodos estatísticos/matemáticos tem grande valia para o educador dispor de maior precisão na mensuração dos resultados encontrados. No sentido de apresentar ferramentas, instruir sobre sua utilização e formalizar em procedimentos computacionais, apresentaremos neste capítulo uma proposta de automatização de avaliações nos moldes de um testes adaptativos computacionais (*Computer Adaptive Test - CAT*), como aqueles descritos nos trabalhos de [Silva et al. \(2020b\)](#), [Plajner \(2017\)](#). A Figura 44 ilustra as etapas de um CAT, com a habilidade do sujeito representada na variável θ , variando entre os estágios 0 até : k . A ideia geral é que início do teste seja com um nível habilidade definido, seja com histórica, seja um valor aleatório arbitrado. A seguir, um item de avaliação é selecionado.

A seleção pode levar em consideração a habilidade do sujeito de acordo com a arquitetura do teste. Então ocorre a exposição do item, para colher a resposta do avaliado. Diante do acerto ou erro, o nível de habilidade é atualizado e a regra de parada é avaliada.

De acordo com sua configuração, o teste é encerrado ou continua a seleção de itens e exposição ao sujeito avaliado.

A adoção de ferramentas inteligentes em plataformas de aprendizagem *online* coincide com a necessidade de melhorar os ambientes educacionais, proporcionando atendimento personalizado aos alunos (SILVA; CAVALCANTI, 2018). Para preencher essa lacuna, propomos um experimento com dados reais que evidencia os benefícios das técnicas de CAT, apoiados pelos modelos da TRI, a fim de quantificar e garantir a padronização dos resultados, bem como, a precisão da habilidade avaliada no sujeito examinado. Por isso, nos baseamos na valores estatísticos verificados nos resultados, ou seja, erro padrão de a estimativa e o grau de correlação com os dados reais.

6.1 Trabalhos Relacionados

O apoio computacional para identificar e estabelecer trilhas de aprendizagem que sejam convenientes aos níveis de habilidade encontrados a partir de dados coletados em AVA é um exemplo do tipo de trabalho que tem sido feito, em que, dentre uma porção de algoritmos de Aprendizagem de Máquina, um método trivial como o *k-means* alcança resultados estatisticamente significativos Monteverde et al. (2018). Em trabalhos mais específicos, a exemplo de Garcia e Gomes (2020), a classificação de currículo no ensino superior é avaliada, sob o prisma de trajetórias de aprendizagem, tomando como base um curso de Ciência da Computação, com mensuração das probabilidades de retenção, evasão e aprovação. Em contraste a técnicas simples como o *k-means*, o trabalho de Reyes et al. (2019), recorre a *transfer learning*, para prever sucesso acadêmico dos avaliados, exemplificando seu método com um conjunto de 3085 instâncias e alcançando uma acurácia no valor de 0,92, utilizando o algoritmo SVM. Também trabalhando com AVA, o estudo de Barbosa (2020), apresenta a integração de um *Plugin* ao Moodle, com intuito de reportar riscos de evasão visualmente, por meio da ferramenta *Google Charts*. Um trabalho similar, feito por Silva et al. (2020a), roteirizou a composição de materiais de aprendizagem associados as etapas do curso e caracterizados por processamento de linguagem natural, para fins de recomendação personalizada. No trabalho de Silva et al. (2020b), foi realizado um apanhado de publicações voltadas para análise de aprendizagem em abordagens em formato de jogo sério, cujo tema principal das discussões era o desempenho do aluno; e o método mais utilizado como análise era a análise de agrupamento. Problemas de classificação, visualização de dados e modelos de regressão ficando em segundo plano e a mineração de dados, em terceiro plano, considerando a frequência de assuntos no conjunto da revisão. O uso de agentes inteligentes em AVA possuem aplicações variadas.

Tão importante quanto reconhecer atributos dos níveis de aprendizagem distinto, é a necessidade de diferenciar itens de avaliação. Esse problema foi tratado por Júnior et al.

(2020) que apresentou uma arquitetura de recomendação de objetos de aprendizagem, os quais foram caracterizados com métodos de algoritmos bio-inspirado e web semântica. Seus melhores resultados foram com os métodos *Particle Swarm Optimization (JPSO)* e *algoritmo genético*. Na publicação de Parraga (2019), a engenharia de características para os objetos de aprendizagem baseia-se na coleta das informações de utilização dos AVA que as atividades estão hospedadas. O controle de acesso ao ambiente, como turnos escolhidos e tempo de permanência por objeto de aprendizagem, também são considerados na composição do modelo, não se restringindo meramente aos dados comportamentais de navegação nos conteúdos e realização das atividades. Algumas arquiteturas fazem uso conceitos de agentes inteligentes, delimitando o espaço de coleta de dados acadêmicos como espaço paramétrico do agente. As funções de otimização passam a ser associadas à função objetivo do agente, enquanto os sensores e atuadores são utilizados como interface das classes internas. No estudo de St-Hilaire et al. (2021), o uso do agente implica num ganho de aprendizagem de 70,43%, a um p – value de 0,038 — comparado a um AVA tradicional que não recorre a inteligência artificial. Os testes adaptativos computadorizados servem de apoio para automação da aplicação de avaliações. Aqueles apoiados por bancos de itens calibrados asseguram a precisão das métricas encontradas. A manutenção de banco de itens calibrados é um esforço que, sendo continuado, assegura uniformidade das habilidades estimadas. Contudo, a administração manual de itens de avaliação pode ser custosa em termos de carga horária e esforço cognitivo. Demanda num primeiro momento que a pessoa a manipular o item conheça do assunto para criticar sua construção e composição. Na sequência, se faz necessário aplicação em grupos de controle para estimar os parâmetros e verificar a concordância entre sujeitos distintos, mas de mesmo perfil. Sanada a composição do banco, a atenção deva ser voltada para o avaliado, conforme o tipo de teste planejado. O trabalho de Tian et al. (2021) discute a aplicação de identificação de trilhas de frustração para aprimorar o funcionamento do CAT, num curso de Introdução a Ciência da Computação.

6.2 Arquitetura do Experimento

Sucintamente, a intenção deste experimento é demonstrar a execução de um teste adaptativo computadorizado desde criação do banco de itens até a aplicação. Para isso, fizemos uso novamente do conjunto de dados do Enem de 2019, descrito com mais detalhes na Subseção 6.2.1. Foi utilizada a biblioteca *mirt* da linguagem R para estimar os parâmetros dos itens no modelo da TRI de 3PL para as questões objetivas. Esse procedimento ocorreu a partir dos vetores de resposta dos alunos, convertidos para valores binários.

Para a classificação textual dos enunciados, foram utilizados os arquivos das provas,

disponibilizados com os microdados¹ da edição de 2019. O texto foi extraído das provas e, utilizando a Linguagem de Programação Python e a biblioteca spaCy². Utilizamos um modelo básico para o idioma português brasileiro, isto é, sem treinamento específico para o corpus de áreas do conhecimento. Processamos também as definições das classes de competência por área de conhecimento. Tal etapa resultou na classificação dos conteúdos abordados em cada item. Por fim, a estrutura do CAT é discutida na Subseção 6.2.2, juntamente com a comparação da implementação de um Agente Inteligente que atue num contexto de CAT. A implementação do código de teste foi realizada com a linguagem *Python*.

6.2.1 Conjunto de Dados do Experimento

6.2.2 CAT - Abordagem Proposta

Usamos a notação de diagramas por raias para segregar recursos e papéis da execução de um cat. Na linha identificada por I/O (*input/output*) na Figura 45a, o ícone denominado Itens, representa o banco de itens de avaliação, os quais denotam perguntas de resposta de múltipla escolha, com apenas uma alternativa correta, das cinco apresentadas. A segunda raia, nomeada CAT — Fluxo Interno, apresenta a etapa de carregamento dos itens; representando a consulta ao conjunto de itens disponível. A atividade seguinte é a Definição do Método de Seleção do Item. Os métodos de seleção de itens são (BAKER; KIM, 2017): *sequencial*, isto é, pela ordem recuperada da consulta ao banco de itens; *orientado a ordenação* de algum atributo do item que esteja indexado; *orientado por conteúdo*, ou seja, de acordo com termos ou conceitos intencionados; orientado por grupo, selecionando itens que respondido por sujeitos pertencentes a um mesmo conjunto relacionado; e por fim, a seleção de itens *orientado ao resultado*, que por norma, amparada por modelos matemáticos, visa recuperar e apresentar itens que maximizem a chance de acerto do sujeito avaliado.

A Figura 45a, ainda na segunda raia, após a seleção do item, apresenta dois cenários, usando a variável θ para representar a habilidade do sujeito avaliado. Na cor vermelha, o losango ligado à linha da seta que segue para a etapa de Selecionar o Item, ilustra uma tomada de decisão, uma vez que é verificada habilidade do sujeito. Caso ela seja desconhecida, pode ser utilizado um critério de seleção de itens desejado, a exemplo dos que foram citados. De outra sorte, conhecendo-se o nível de habilidade do sujeito, pode-se selecionar um item que maximize o resultado. Sendo um ou outro caminho, o item é apresentado ao sujeito examinado, na terceira raia do diagrama da figura citada. Com a resposta do item, o fluxo segue para a verificação da pontuação de habilidade alcançada, que diante do acerto, pode ser incrementada; e em face do erro, pode ser penalizada. O

¹ Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>

² Disponível em: <https://spacy.io/>

passo seguinte é a verificação da regra de parada. As regras de parada são: invariação a habilidade; variação da habilidade num limiar determinado; quantidade de itens — chamado também de cumprimento/tamanho do teste; quantidade de itens disponíveis no banco de itens do teste. Por fim, após a verificação da regra de parada, caso não seja procedente, o teste continua, porém, com a habilidade do sujeito conhecida da última rodada, podendo ser atualizada a cada iteração do ciclo, até a conclusão. Na Figura 45b, consideramos o banco de itens e o histórico de resposta, na primeira raia do diagrama, como o Estado Inicial do Espaço paramétrico. Na segunda raia, da figura citada, temos a Função do Agente, no qual a estratégia seleção de itens, similar ao que foi citado para o cat tradicional, deve ser definida. A etapa de carregamento de itens é acompanhada do carregamento dos históricos. Isso pode ser feito por meio de modelos de análise de agrupamento ou os vetores de resposta, efetivamente. As etapas de seleção de item sem conhecimento da habilidade do sujeito e também com o conhecimento, seguem similar ao cat tradicional. A diferença está no tratamento conceitual da interface com o sujeito avaliado, haja vista que a apresentação do item ao sujeito configura como uma ação de atuação do agente; enquanto a coleta da resposta do sujeito, pode ser compreendida como uma ação do sensor, capturando sinais do espaço paramétrico em que o agente atua.

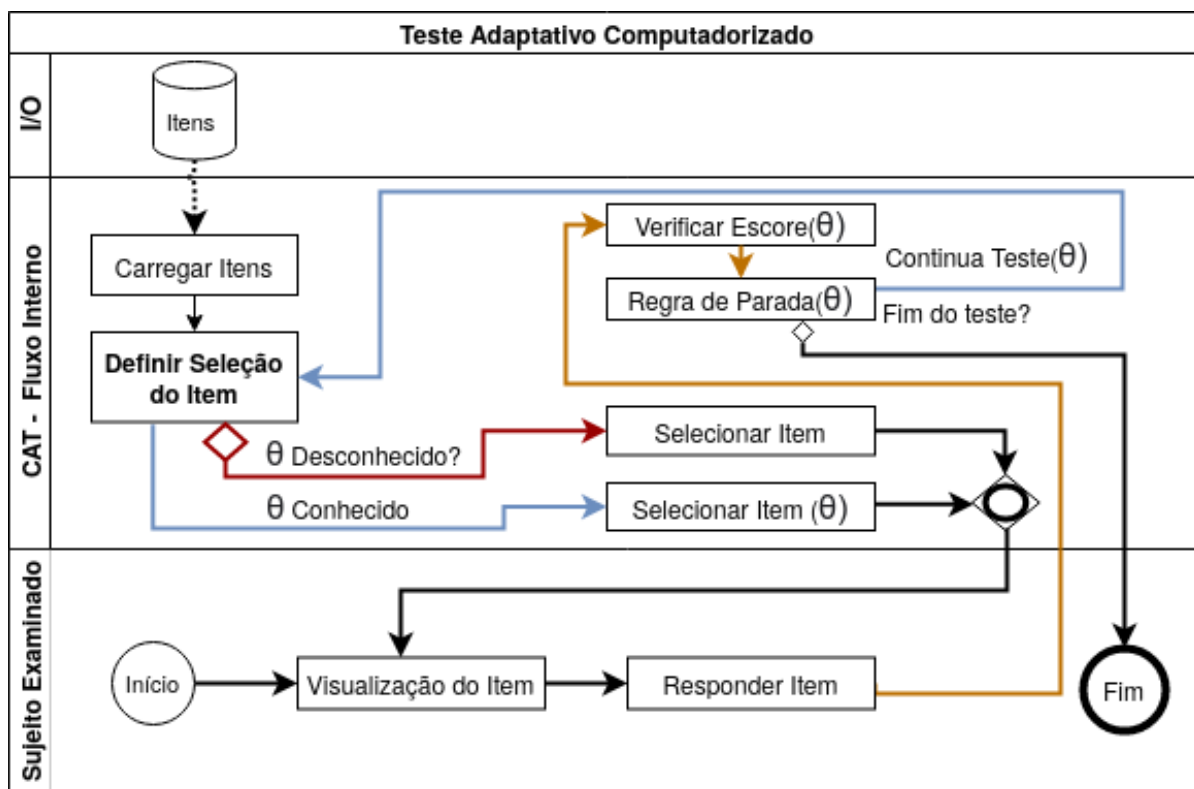
6.3 Resultados e Discussão

A Figura 46a mostra a capa da prova de LC e CH. Na Figura 46b consta a capa da prova de CN e MT. Após a aquisição dos dados, na etapa de pré-processamento, contabilizamos os dados faltantes, apresentados por eixos temáticos. Por meio dos vetores de resposta separados por tipo e código de prova, foi realizado a estimação dos coeficientes do Modelo 3PL da TRI. Com os enunciados das questões, procedemos a classificação por similaridade textual. Por fim, realizamos os testes do CAT, demonstrando a possibilidade de utilizar um agente inteligente que seja capaz de conjugar diferentes estratégias de avaliação.

A constatação inicial com o conjunto de dados selecionado foi a identificação dos dados faltantes, evidenciado pela Figura 47a, que exhibe o percentual do subgrupo de informações do participante. A variável tipo de ensino (Tipo de escola do Ensino Médio) é a que apresentou o maior percentual dentre todas, no valor de 40% de dados faltantes. As opções a esta variável faltante são: ensino regular, ensino especial, educação de jovens e adultos. A Figura 47b lista o subgrupo de informações da instituição vinculada ao participante, no caso, onde foi concluído ou onde está sendo cursado o ensino-médio. Nestes, o percentual de dados faltante é pouco inferior a 80% para as variáveis: tipo de funcionamento da escola, tipo de localização, tipo de dependência administrativa, informações de localização (estado e município) e o código INEP da escola. Nos dados da prova, isto é, respostas das questões objetivas e notas apuradas na relação, o cenário é o

Figura 45 – Esquema de Funcionamento do Teste Adaptativo

(a) Tradicional



(b) Abordagem baseada em Agente Inteligente

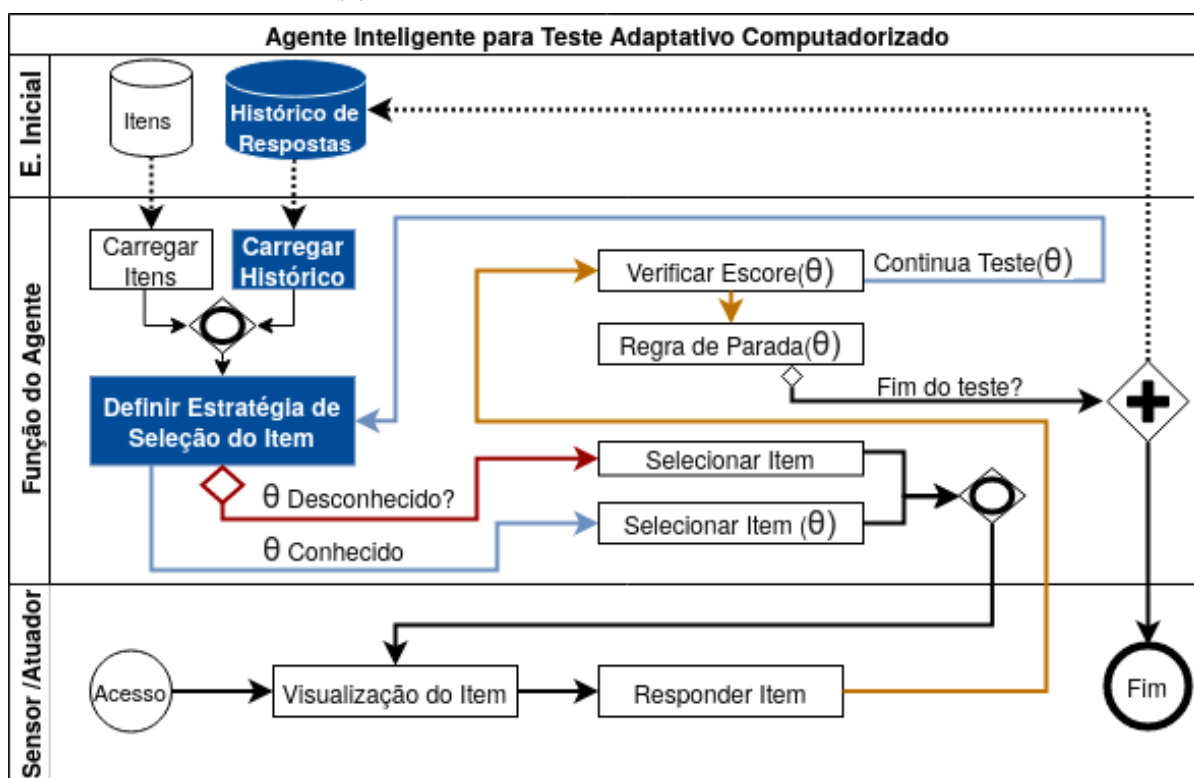
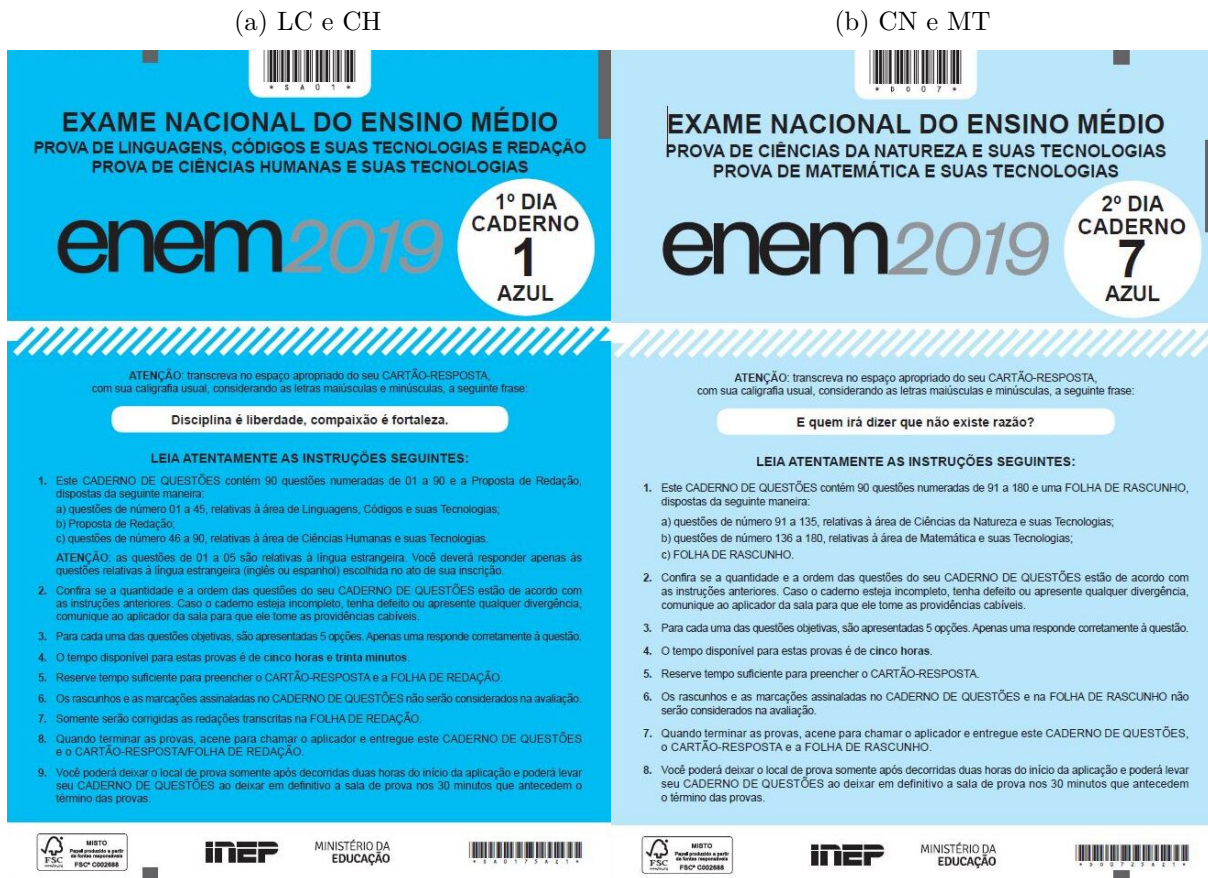


Figura 46 – Capas das Provas de 2019

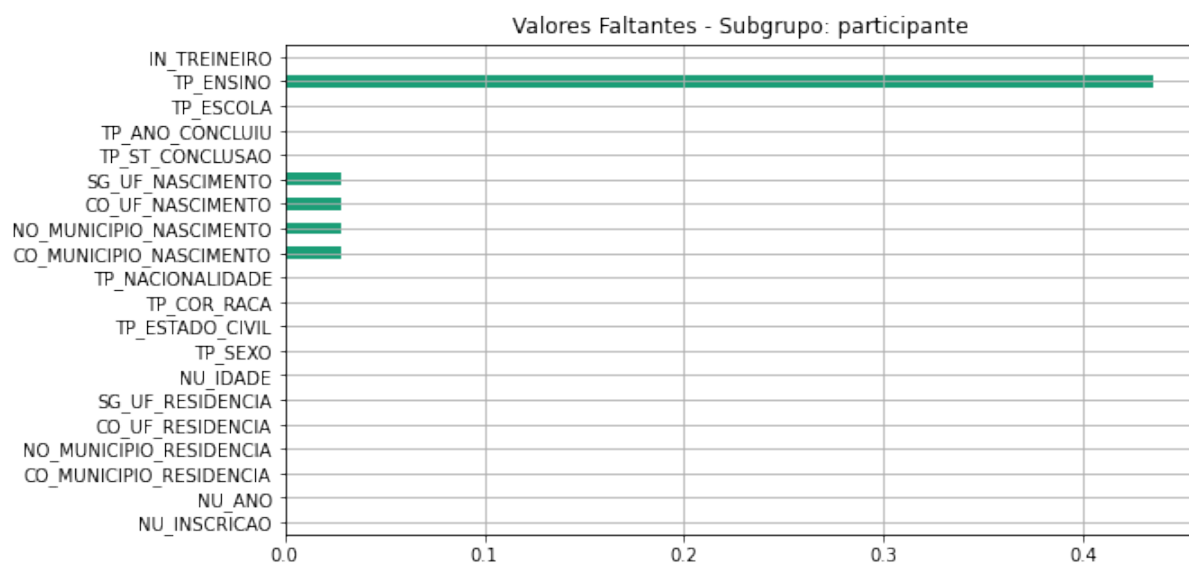


que está ilustrado na Figura 47a e 47b. Na primeira, os percentuais de dados faltantes para a prova de Matemática é similar ao da prova de Ciências Naturais, com valores próximos a 27%. O mesmo ocorre com as provas de Ciências Humanas e de Linguagem, com valores próximos a 23%. Na segunda imagem, com os percentuais concernentes à prova de redação, os valores ficaram também próximos a 23%. Com os dados completos disponíveis, procedemos a estimação dos parâmetros, os quais serão relatados a seguir.

O primeiro resultado oriundo da utilização do processamento de linguagem natural foi a contabilização dos tópicos dos eixos cognitivos por área do conhecimento. A Figura 49a mostra a contagem de tópicos dos eixos cognitivos identificados na avaliação de LC. Como nas demais figuras, 49b, 50a e 50b, há predomínio do tópico CF em todas as áreas. Em segundo lugar, vemos o tópico EP nas Figuras 49b e 50a. O mesmo ocorre, com o eixo DL, nas Figuras 49a e 50b. Apenas na área de conhecimento CN, Figura 50a, ocorreu o eixo CA, em menor proporção que os demais eixos na área. Adiante, por área no detalhamento do item, discutimos a frequência das competências presentes em cada item de avaliação.

Figura 47 – Dados Faltantes — Participante & Instituição

(a)



(b)

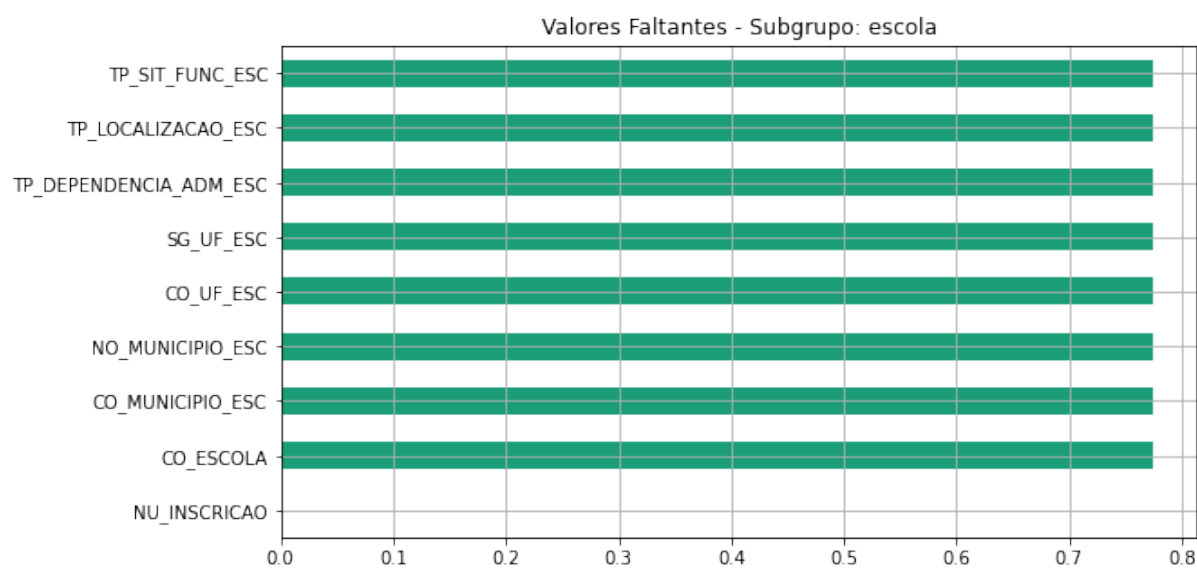
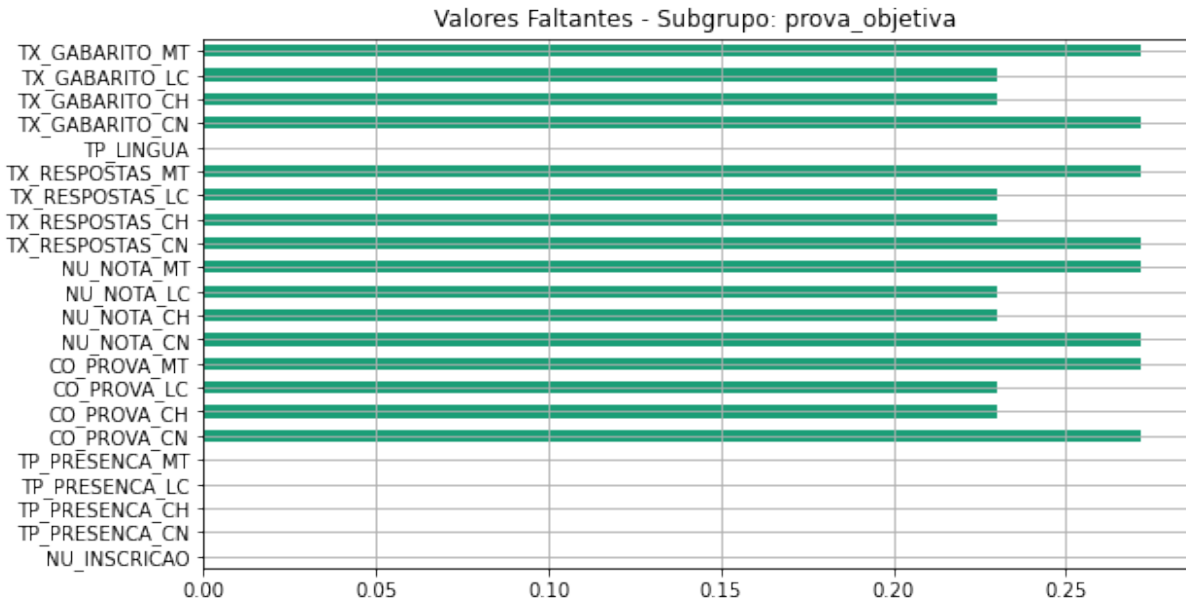


Figura 48 – Dados Faltantes — Objetivas & Redação

(a)



(b)

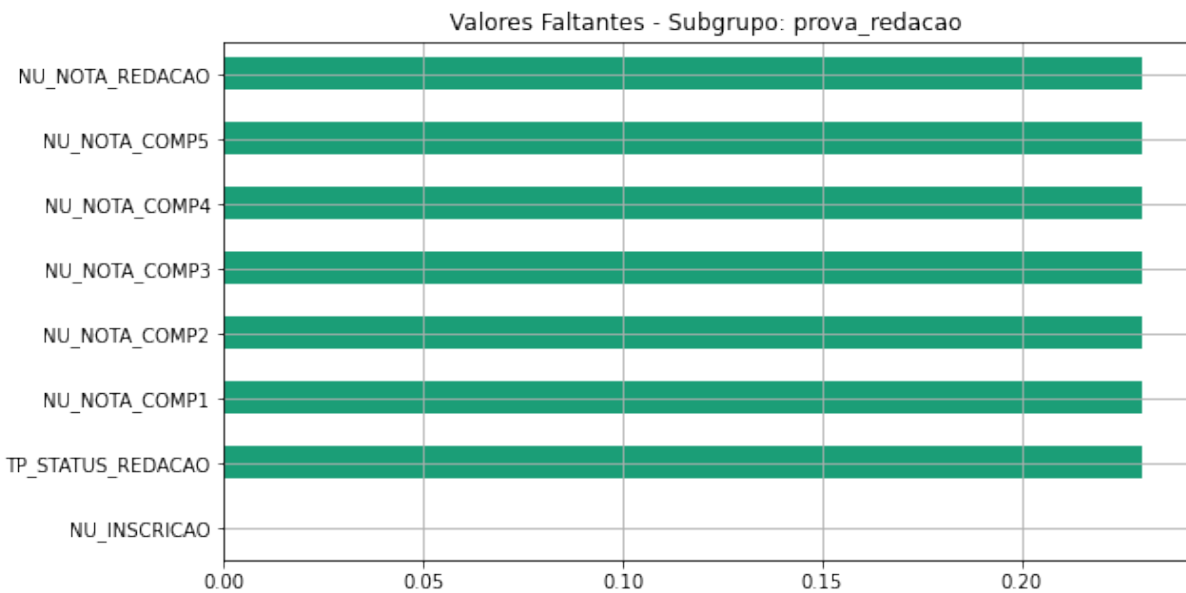
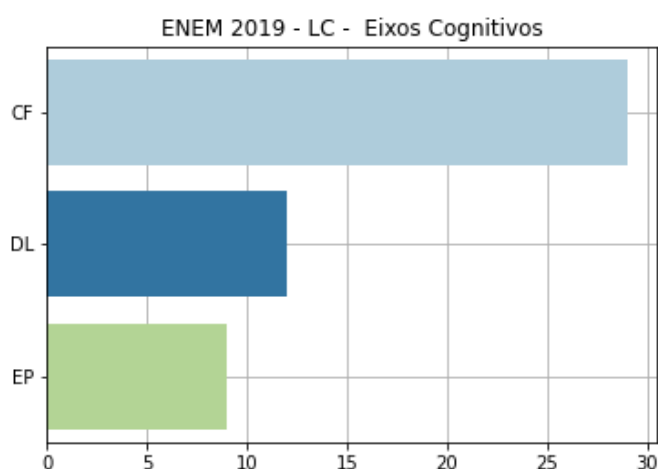
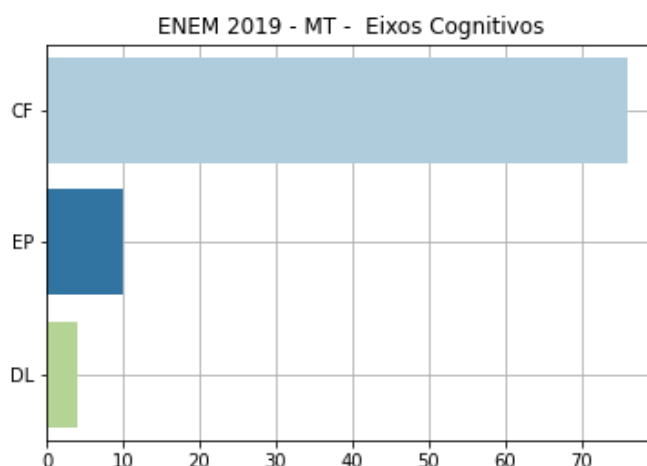


Figura 49 – Frequência de Eixos Cognitivos — LC e MT

(a)



(b)

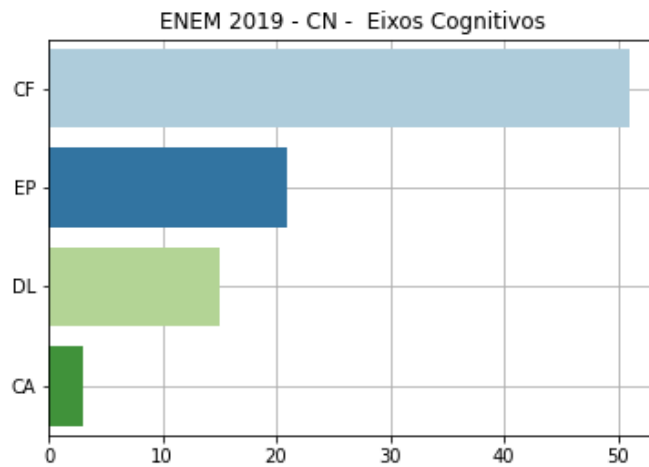


Itens de Avaliação Caracterizados

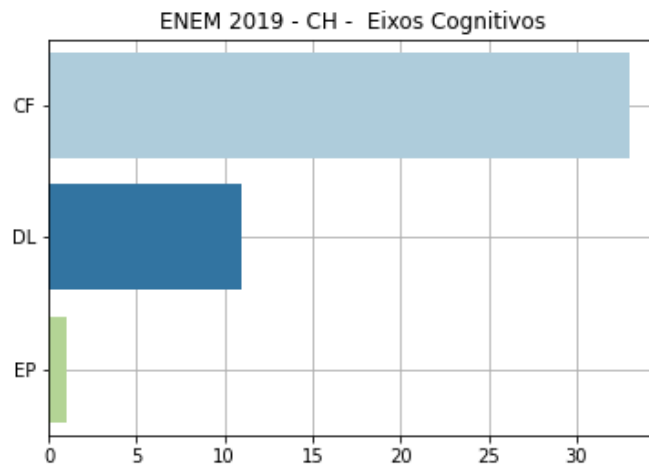
Foram utilizadas as funções de informação para caracterizar cada item estimado para o modelo de 3PL da TRI, vide Equações 4.4 e 4.5. Por meio das curvas de catarísticas, é possível diferenciar itens de qualidade daqueles que poderiam ser melhorados. No eixo y, para itens de qualidade com baixa propensão ao chute, espera-se a interseção próxima a 0, para valores inferiores ao segundo desvio padrão negativo. No eixo x, itens com dificuldade média tendem a comportamento definido na expressão $P(\theta = 0) = 0.5$, representando uma habilidade igual à média possuir 50% de probabilidade de sucesso. Nas funções de informação, são de interesse itens que reúnam maior quantidade de informação próximos ao valor médio e na região positiva dos desvios padrão de θ . Outro recurso utilizado na exibição dos parâmetros foram gráficos de barra em duas versões. A primeira, com os coeficientes de Discriminação (a) e Dificuldade (b), ordenados de maneira crescente

Figura 50 – Frequência de Eixos Cognitivos — CN e CH

(a)



(b)



em função do parâmetro a do Modelo de 3PL. A segunda forma utilizada, recorreu aos coeficientes de Discriminação (a) e Propensão a Chute (c), ordenados pelo mesmo critério da primeira. Por fim, também lançamos mão de gráficos mistos, combinando pontos em função das mesmas variáveis citadas para os gráficos de barra. Nestes tipos de gráficos, para auxiliar na descrição dos itens utilizamos a identificação de grupos por cores. A seguir, listamos por área do conhecimento, os objetos na ordem em que acabaram de ser elencados.

Parâmetros Estimados

A Figura 51a mostra os coeficientes estimados para a prova de LC, com legenda significando: a1, de cor azul, para discriminação; b, de cor laranja, para dificuldade; g

(*guess*), de cor vermelho, para propensão a chute; e b_{inep} , na cor roxa, representando o nível de habilidade exigido pelo item. Nas demais Figuras 51b, 52a e 52b essa mesma lógica se repete. Na Figura 51a, ocorrem dois parâmetros com valores acima de 30 e abaixo de -30, localizados nas posições 43 e 48. Vemos ainda na mesma imagem que os primeiros cinco itens são os de maior discriminação — nesta prova, denotam as questões de língua inglesa. Por fim, os itens em que os valores de b ficaram negativos são aqueles que mereceriam atenção na reconsideração do teste, com os agudos itens 5 e 49. A respeito da prova de MT, cujos coeficientes estimados observamos na Figura 51b, o item 10 foi o de maior dificuldade e o item 20, o de menor dificuldade. Ocorreram menos valores inadequados se comparado à prova de LC, por haver menos valores de b negativos.

A prova de CN, vista na Figura 52a, ocorrem três itens de bastante dificuldade, com os itens 27, 31 e 36. O Item 35 foi o de menor dificuldade. Entre os itens 15 a 18, observamos os parâmetros de discriminação e dificuldade bastante próximos. Na Figura 52b, estão os parâmetros da prova de CH. Nela, não ocorrem valores acima do terceiro desvio padrão positivo e nem abaixo do quarto desvio padrão negativo. Nesta prova, apenas 4 itens apresentaram valores negativos para discriminação. No mesmo gráfico, os valores com propensão a chute estão próximos ao item 10, 20 e 35.

Itens de Avaliação de Linguagem e Comunicação

Nas Figuras 53a e 53b, observamos as curvas de características e as funções de informação, respectivamente. Na primeira, vemos que as funções de probabilidade dos itens 38 e 50 são as que apresentam pior declive, com menor probabilidade de acerto nos valores de $\theta > 0$. Na segunda figura, os itens 49 e 47 são exemplos dos que reúnem maior quantidade de informação, dado o formato acentuado da curva normal vista no gráfico.

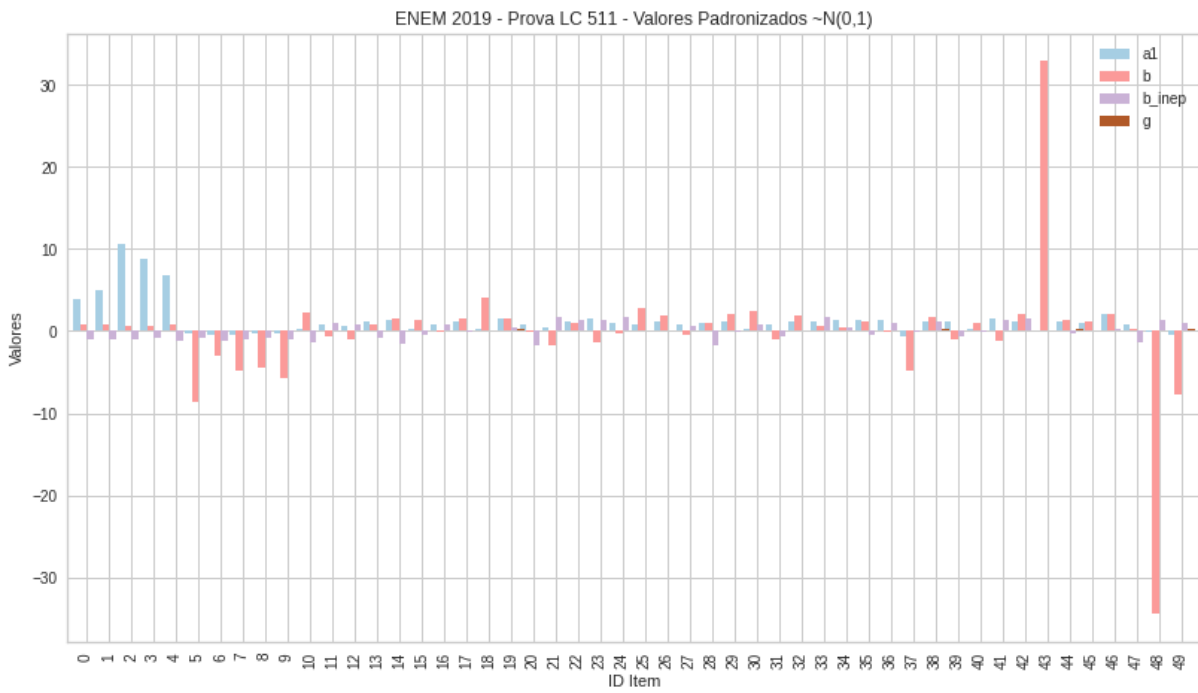
Conforme a classificação dos enunciados, na Figura 54a os conteúdos CA-2 e CA-5 foram os mais requisitados. Em relação às referências, na Figura 54b a menor ocorrência foi de LC-R5. A título de ilustração, os enunciados das questões classificadas pela TRI, observamos na A Figura 55a o item de maior valor de discriminação. Na Figura 55b, é apresentado item de maior valor de dificuldade. A resposta informada no gabarito está sinalizada com a seta em cor verde.

Itens de Avaliação de Matemática

Os itens de avaliação da Prova de Matemática estão dispostos na Figura 56a. Os itens estão numerados da linha inferior para a linha superior, da esquerda para a direita. No eixo x , os valores para θ e no eixo y , temos os valores para $P(\theta)$. Visualmente, identificamos a inadequação das curvas dos itens X10, X11, X21, X28, X40, X42. Todas com declive negativo ou formato linear. Do conjunto restante, nota-se grande propensão ao chute nos

Figura 51 – Parâmetros TRI das Provas de LC e MT de 2019

(a) Linguagem e Comunicação



(b)

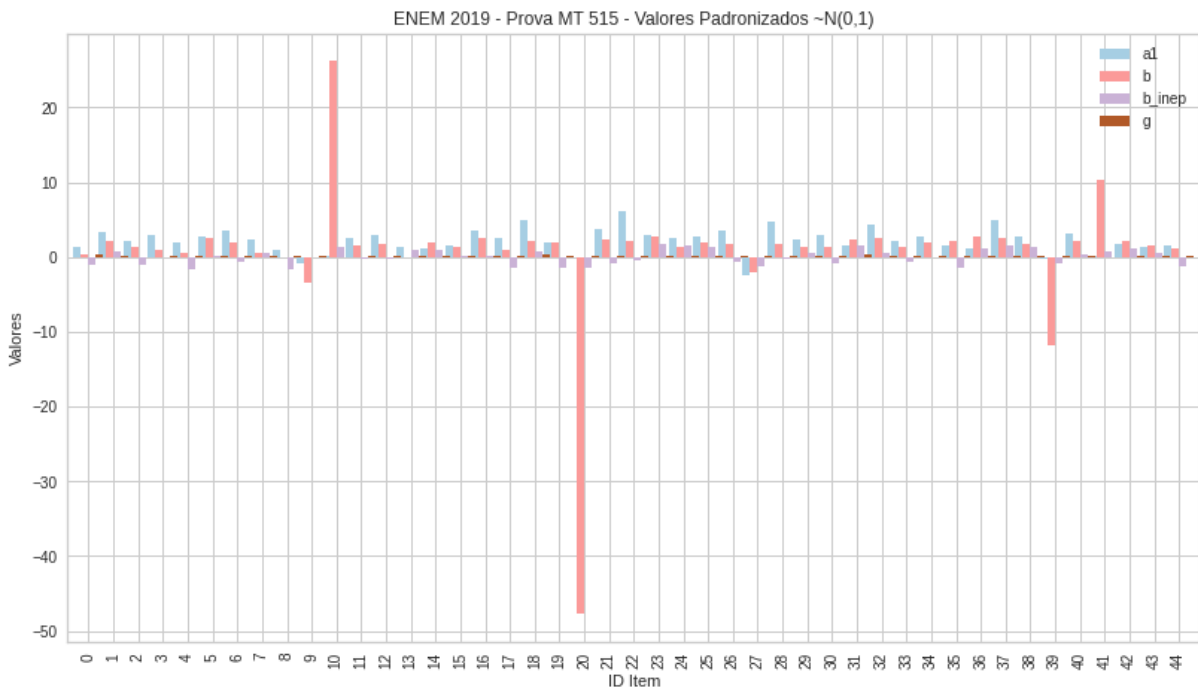
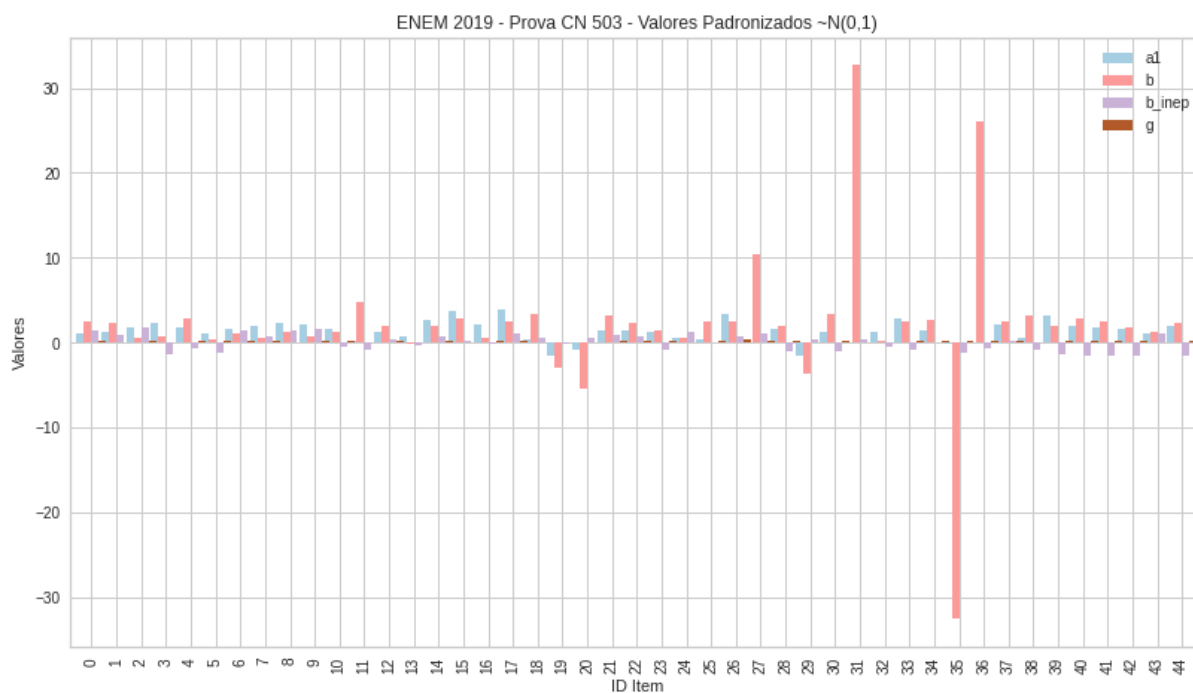


Figura 52 – Parâmetros TRI das Provas de CN e CH de 2019

(a)



(b)

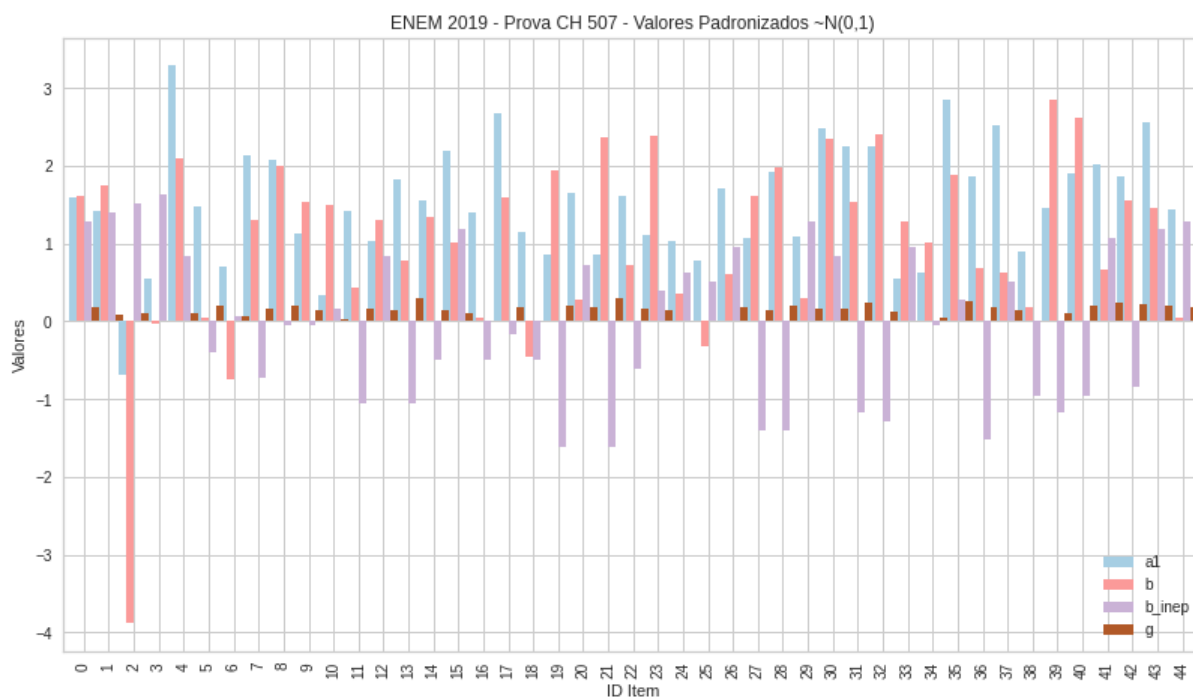
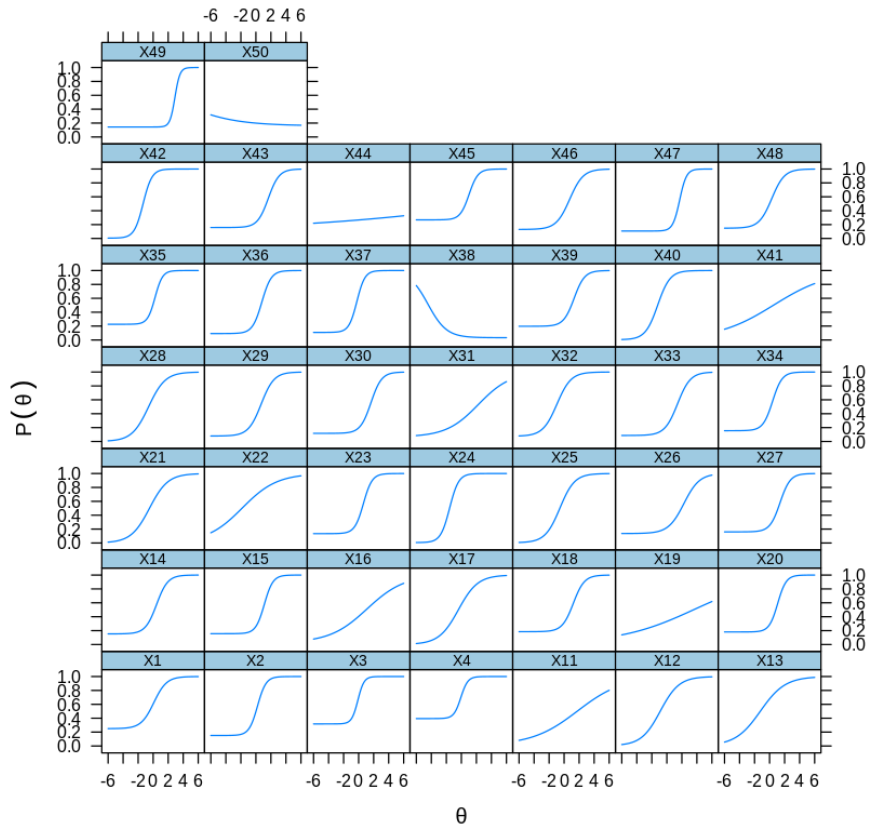


Figura 53 – Probabilidade de Sucesso e Informação dos Itens — LC

(a)

Enem - 2019 - LC - 511



(b)

Enem - 2019 - LC - 511

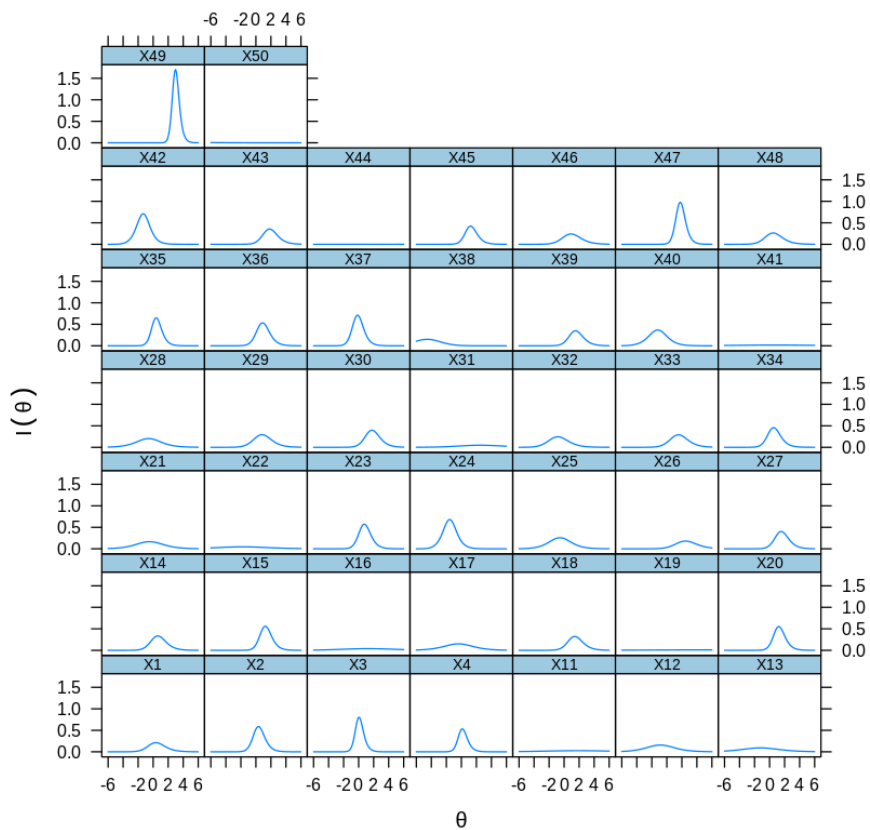
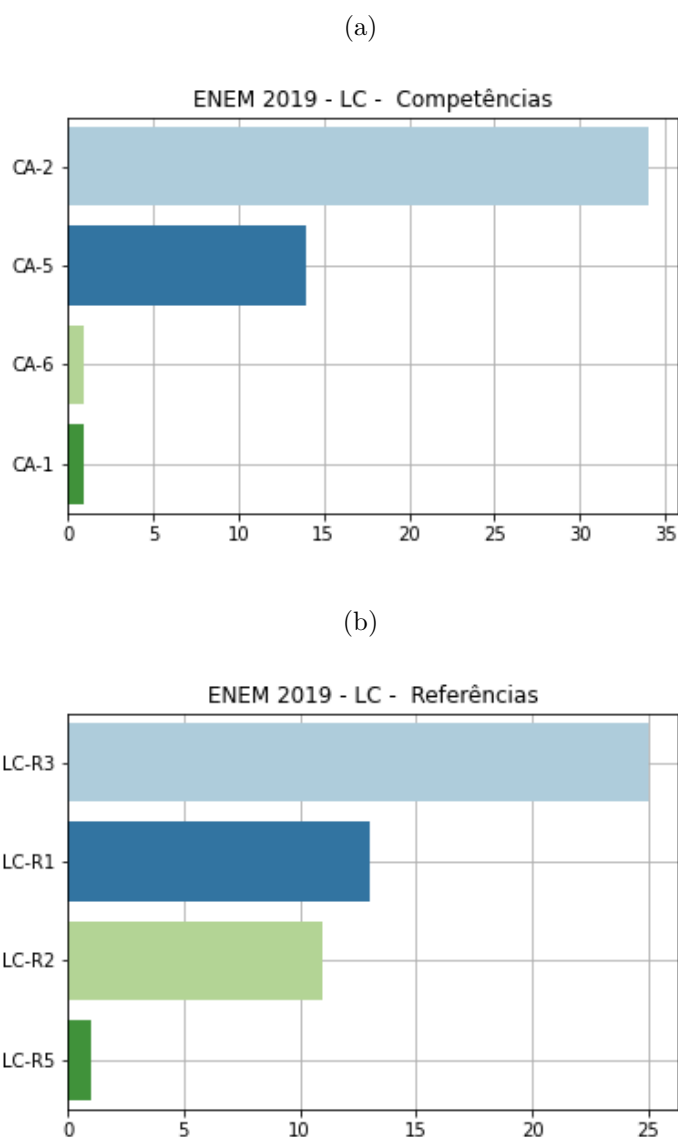


Figura 54 – Classificação Textual das Competências dos Itens de Linguagem e Comunicação



itens X01, X05, X09, X12, X13, X19 e X32. Dentre as funções de informação, visualizadas na Figura 56b, destacamos a ausência de informação nos itens no qual a discriminação visualmente comprovou inadequação do item, como, por exemplo, para os itens: X10, X11, X2. Com o item X28, vemos uma situação de acúmulo de informação na região negativa do terceiro desvio padrão. Na Figura 57a, são listadas as competências classificadas na prova, contendo apenas CA-1 e CA-2. As referências vistas na Figura 57b demonstra prevalência da MT-R3 e ínfima ocorrência de MT-R1. Na Figura 58a apresentamos o item de maior poder discriminativo. Na Figura 58b, é mostrado o item de maior dificuldade.

Figura 55 – Itens da Prova de LC

(a) Item 3 - Maior Valor de Discriminação

Questão 04

If children live with criticism, they learn to condemn.
 If children live with fear, they learn to be apprehensive.
 If children live with pity, they learn to feel sorry for themselves.
 If children live with ridicule, they learn to feel shy.
 If children live with tolerance, they learn patience.
 If children live with praise, they learn appreciation.
 If children live with acceptance, they learn to love.
 If children live with approval, they learn to like themselves.
 If children live with recognition, they learn it is good to have a goal.
 If children live with sharing, they learn generosity.
 If children live with fairness, they learn justice.
 If children live with kindness and consideration, they learn respect.
 If children live with friendliness, they learn the world is a nice place in which to live.

NOLTE, D. L. Disponível em: www.americanfamilytraditions.com. Acesso em: 30 jul. 2012.

Valores culturais de um povo revelam sua forma de ser, agir e pensar. Na concepção da autora, as diferentes formas de educar crianças nos Estados Unidos confirmam que as crianças

- A temem quem as amedronta.
- B aprendem com o que vivem.
- C amam aqueles que as aceitam.
- D são gentis quando respeitadas.
- E ridicularizam quem as intimida.

(b) Item 39 - Maior Valor de Dificuldade

Questão 39

Essa lua enlutada, esse desassossego
 A convulsão de dentro, ilharga
 Dentro da solidão, corpo morrendo
 Tudo isso te devo. E eram tão vastas
 As coisas planejadas, navios,
 Muralhas de marfim, palavras largas
 Consentimento sempre. E seria dezembro.
 Um cavalo de jade sob as águas
 Dupla transparência, fio suspenso
 Todas essas coisas na ponta dos teus dedos
 E tudo se desfez no pórtico do tempo
 Em lívido silêncio. Umas manhãs de vidro
 Vento, a alma esvaziada, um sol que não vejo
 Também isso te devo.

HILST, H. *Júbilo, memória, noviciado da paixão*. São Paulo: Cia. das Letras, 2018.

No poema, o eu lírico faz um inventário de estados passados espelhados no presente. Nesse processo, aflora o

- A cuidado em apagar da memória os restos do amor.
- B amadurecimento revestido de ironia e desapego.
- C mosaico de alegrias formado seletivamente.
- D desejo reprimido convertido em delírio.
- E arrependimento dos erros cometidos.

Itens de Avaliação de Ciências Naturais

Na prova de Ciências Naturais, os itens de curvas de maior inadequação foram os 12, 30, 32, 36 e 37, como pode ser visto na Figura 59a. Os itens que não reuniram muita informação neste teste foram os 12, 25, 28, 32, 36, 37 e 39, ilustrado pela Figura 59b. O resultado da classificação dos itens desta área do conhecimento pode ser visto na Figura 60a, em que vemos o contraste das competências CA-2 e CA-8. Na Figura 60b, o predomínio foi da referência CN-R5-F17, da área de Física. E a menor frequência foi da CN-R10-Q18, da área de Química. O item de maior coeficiente de discriminação desta prova pode ser visto na Figura 61a. Enquanto o item de maior índice de dificuldade está listado na Figura 61b.

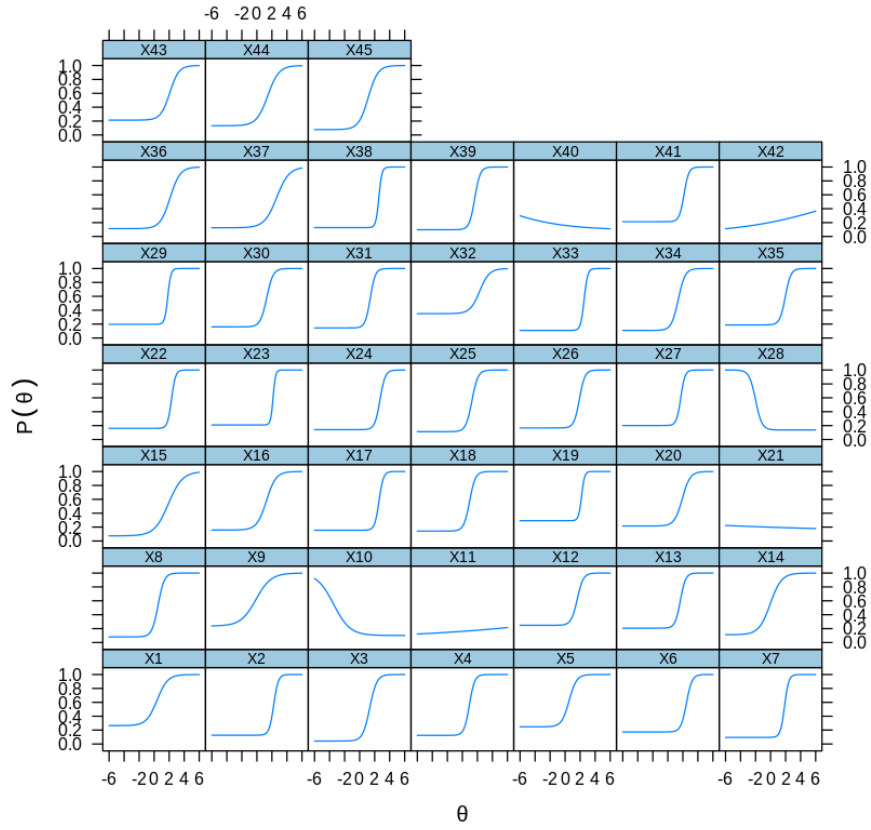
Itens de Avaliação de Ciências Humanas

A Figura 62a apresenta as curvas de característica dos itens da prova de Ciências Humanas. Nela, vemos a inadequação do item 3. Na Figura 62b, o item 5 ilustra uma grande quantidade de informação aglutinada em torno do segundo desvio padrão positivo. A classificação dos enunciados desta prova gerou a Figura 63a, com predomínio da competência CA-6, em contraste à competência CA-4, em menor frequência. A referência

Figura 56 – Probabilidade de Sucesso e Informação dos Itens - MT

(a)

Enem - 2019 - MT - 515



(b)

Enem - 2019 - MT - 515

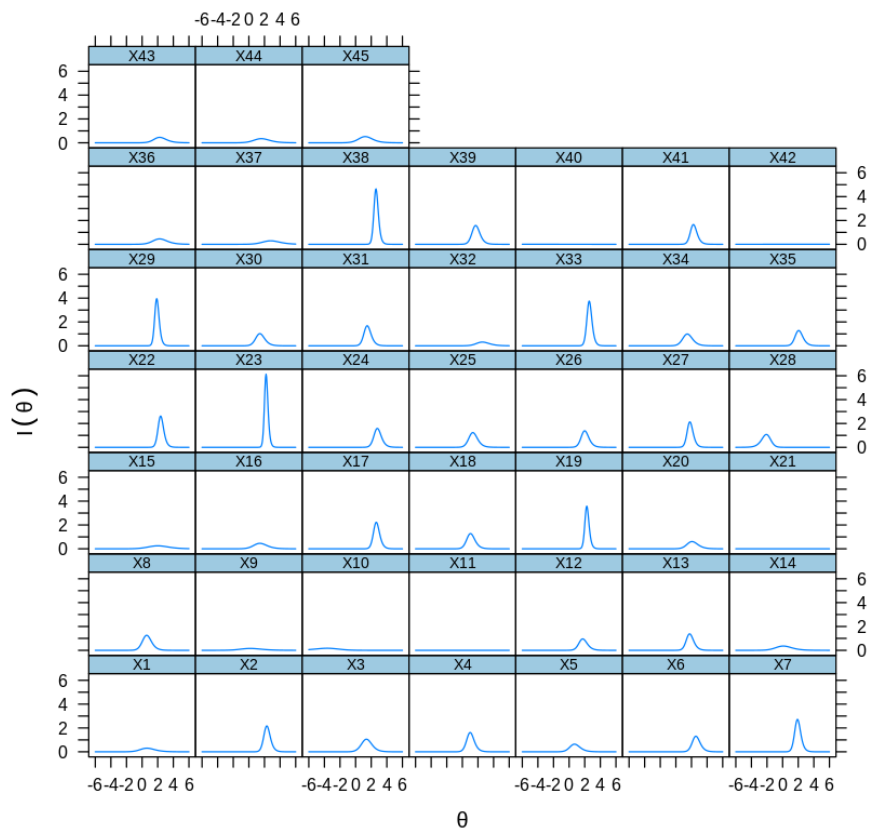
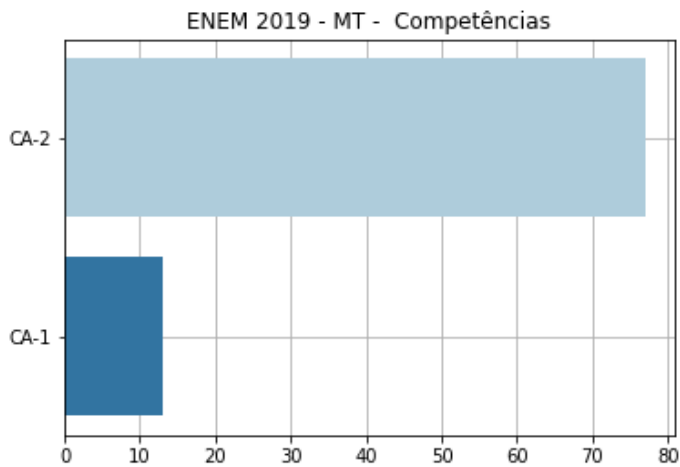
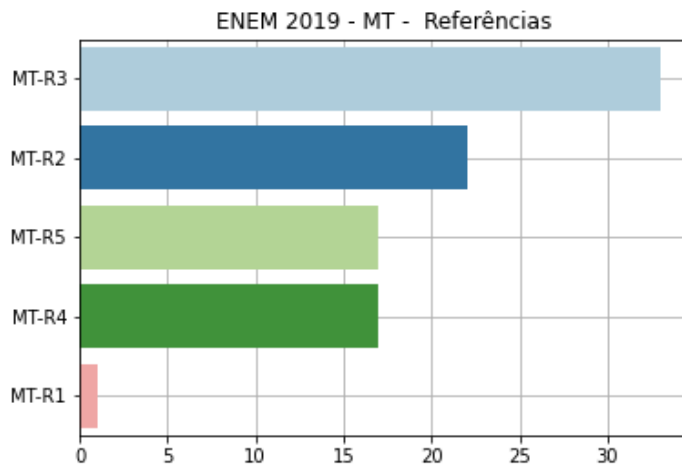


Figura 57 – Classificação Textual das Referências dos Itens de MT

(a)



(b)



de maior incidência foi a CH-R2, como pode ser visto na Figura 63b Na Figura 64a, o item de maior discriminação desta prova. E na Figura 64b o item de maior propensão ao palpite.

Comparação das Habilidades

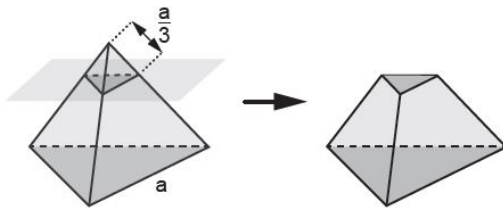
Nas Figuras 65a, 65b, 66a e 66b, listamos as distribuições estatísticas das habilidades. Na Figura 65a, vemos a comparação dos valores de habilidades estimados, denominados $\hat{\theta}$ com os valores de habilidade informados com os microdados do Enem, aqui nomeados de θ_{real} . Os valores das médias coincidem, bem como o intervalo interquartil. No limite inferior, as habilidades reais atingem patamares inferiores bem maiores que os valores estimados. De todas as provas, vemos que para MT a média estimada ficou um pouco

Figura 58 – Itens da Prova de MT

(a) Item 22 - Maior Valor de Discriminação

Questão 157

As luminárias para um laboratório de matemática serão fabricadas em forma de sólidos geométricos. Uma delas terá a forma de um tetraedro truncado. Esse sólido é gerado a partir de secções paralelas a cada uma das faces de um tetraedro regular. Para essa luminária, as secções serão feitas de maneira que, em cada corte, um terço das arestas sectionadas serão removidas. Uma dessas secções está indicada na figura.



Essa luminária terá por faces

- **A** 4 hexágonos regulares e 4 triângulos equiláteros.
B 2 hexágonos regulares e 4 triângulos equiláteros.
C 4 quadriláteros e 4 triângulos isósceles.
D 3 quadriláteros e 4 triângulos isósceles.
E 3 hexágonos regulares e 4 triângulos equiláteros.

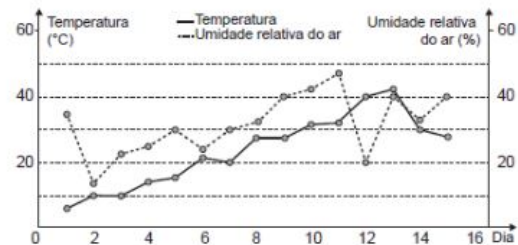
(b) Item 10 - Maior Valor de Dificuldade

Questão 146

O serviço de meteorologia de uma cidade emite relatórios diários com a previsão do tempo. De posse dessas informações, a prefeitura emite três tipos de alertas para a população:

- Alerta cinza: deverá ser emitido sempre que a previsão do tempo estimar que a temperatura será inferior a 10 °C, e a umidade relativa do ar for inferior a 40%;
- Alerta laranja: deverá ser emitido sempre que a previsão do tempo estimar que a temperatura deve variar entre 35 °C e 40 °C, e a umidade relativa do ar deve ficar abaixo de 30%;
- Alerta vermelho: deverá ser emitido sempre que a previsão do tempo estimar que a temperatura será superior a 40 °C, e a umidade relativa do ar for inferior a 25%.

Um resumo da previsão do tempo nessa cidade, para um período de 15 dias, foi apresentado no gráfico.



Decorridos os 15 dias de validade desse relatório, um funcionário percebeu que, no período a que se refere o gráfico, foram emitidos os seguintes alertas:

- Dia 1: alerta cinza;
- Dia 12: alerta laranja;
- Dia 13: alerta vermelho.

Em qual(is) desses dias o(s) aviso(s) foi(ram) emitido(s) corretamente?

- **A** 1
B 12
C 1 e 12
D 1 e 13
E 1, 12 e 13

abaixo da média real. Nas demais, os intervalos interquartis são coincidentes.

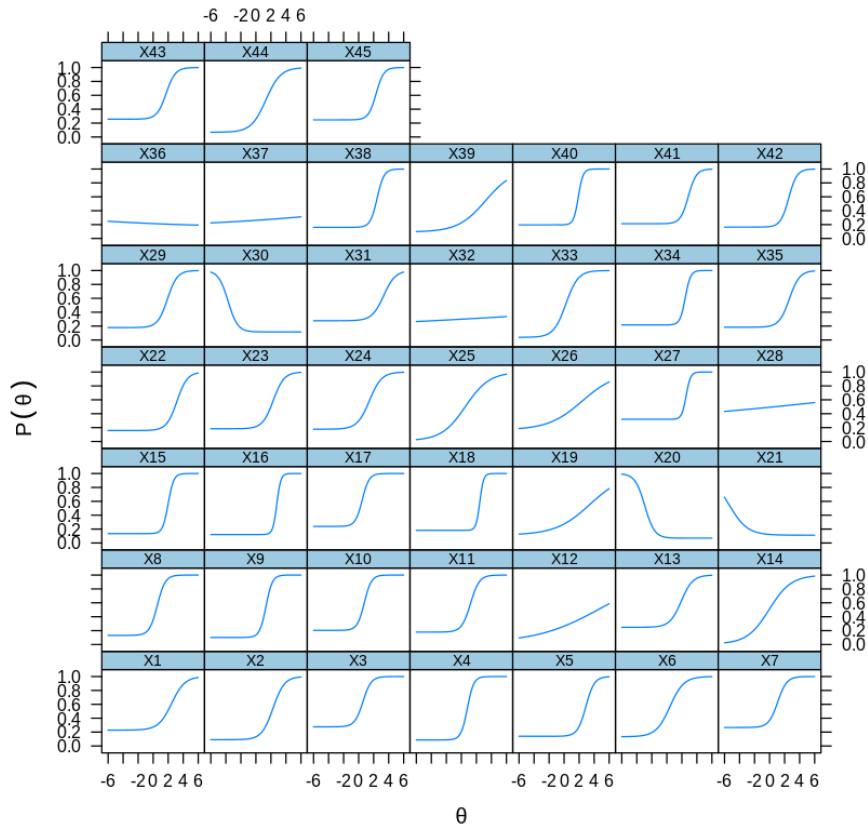
Outra forma de apresentar as habilidades encontradas foi por meio do gráfico de dispersão com densidade, identificado por grupos de valores similares. As Figura 67a, 67b, 68a e 68b apresentam os valores para as provas de LC, MT, CN e CH. Os grupos coloridos definem intervalos para os resultados esperados da habilidade real. Com exceção de MT, para as demais provas, os valores são de maior ajuste na região entre o primeiro desvio padrão positivo e negativo. Para valores abaixo de 1 desvio padrão negativo, nota-se maior dispersão entre os pontos.

Nos gráficos de dispersão de resíduos, listado nas Figuras 69a, 69b, 70a e 70b, confirmamos a maior distorção dos valores nos extremos inferiores.

Figura 59 – Probabilidade de Sucesso e Informação dos Itens — CN

(a)

Enem - 2019 - CN - 503



(b)

Enem - 2019 - CN - 503

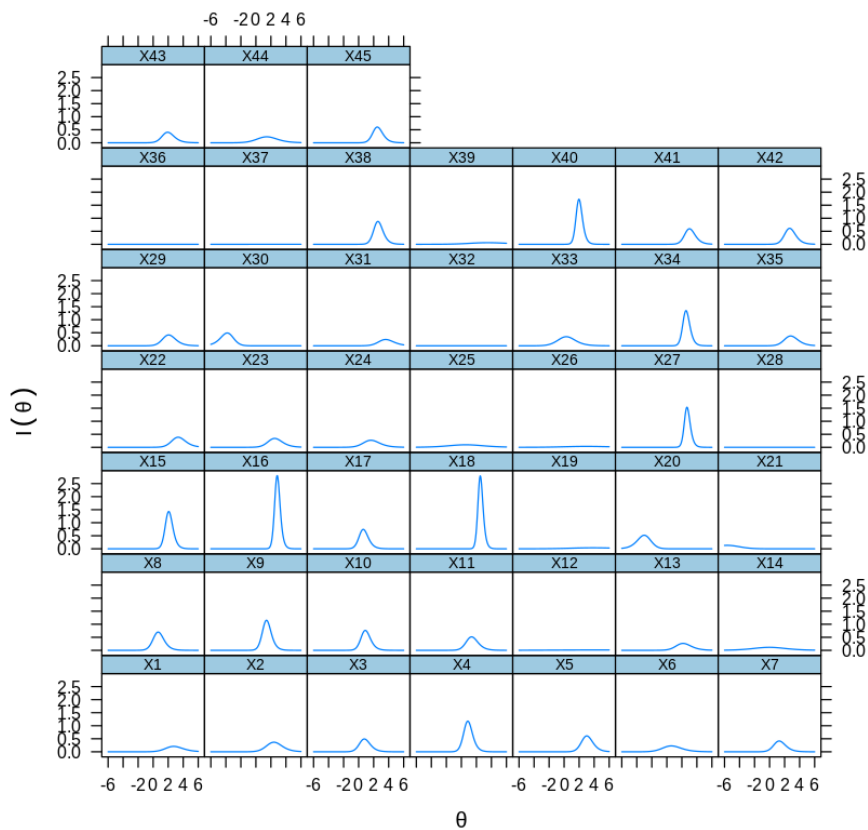
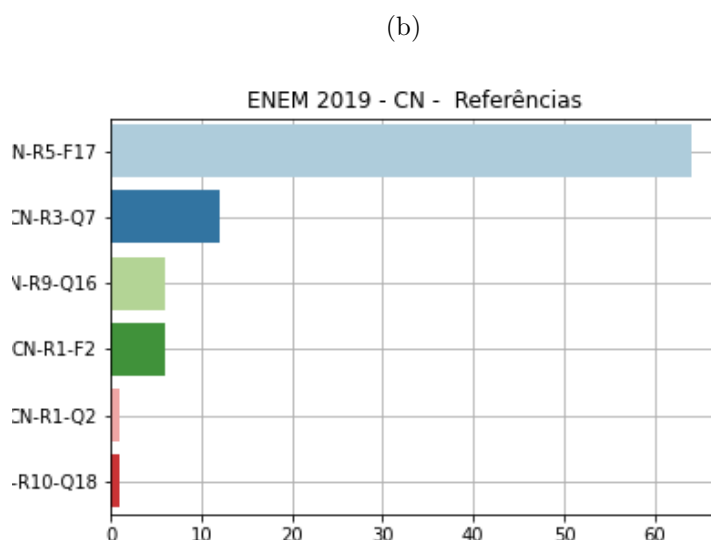
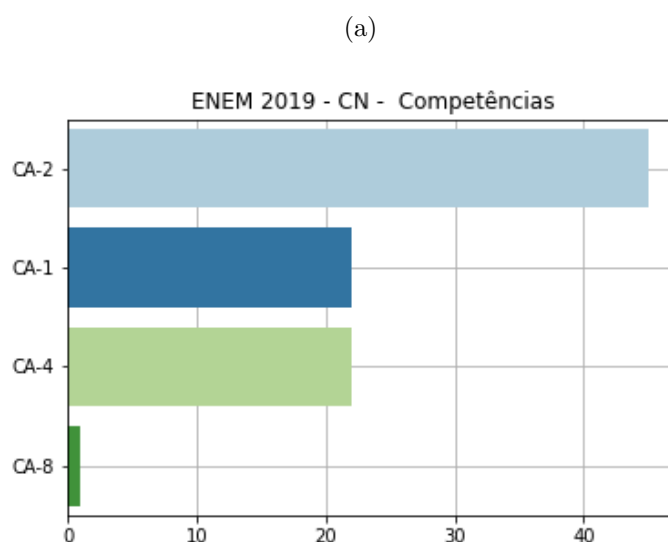


Figura 60 – Classificação Textual das Referências dos Itens de CN



Avaliação	mse	rmse	spearman	pearson	kendall	t-statistic	p-value
Prova CN	0.1940	0.0376	0.9811	0.9841	0.8987	1.399180e-10	1.0
Prova CH	0.2150	0.0462	0.9768	0.9912	0.9230	-9.123688e-11	1.0
Prova MT	0.2235	0.0499	0.9750	0.9657	0.8599	5.098831e-11	1.0
Prova LC	0.2410	0.0580	0.9709	0.9934	0.9325	-9.409642e-11	1.0

A seguir, o código ilustrativo da implementação do CAT.

6.3.1 CAT com Prova de MT

Na Figura 71a, os coeficientes (a) e (b), apresentado excepcionalmente neste tipo de gráfico, com a identificação (d), *difficulty*, do termo em inglês dificuldade. Naquela figura, na cor verde, vemos um par de itens com os maiores valores para discriminação e

Figura 61 – Itens da Prova de CN

(a) Item 15 - Maior Valor de Discriminação

Questão 108

Por terem camada de valência completa, alta energia de ionização e afinidade eletrônica praticamente nula, considerou-se por muito tempo que os gases nobres não formariam compostos químicos. Porém, em 1962, foi realizada com sucesso a reação entre o xenônio (camada de valência $5s^25p^6$) e o hexafluoreto de platina e, desde então, mais compostos novos de gases nobres vêm sendo sintetizados. Tais compostos demonstram que não se pode aceitar acriticamente a regra do octeto, na qual se considera que, numa ligação química, os átomos tendem a adquirir estabilidade assumindo a configuração eletrônica de gás nobre. Dentre os compostos conhecidos, um dos mais estáveis é o difluoreto de xenônio, no qual dois átomos do halogênio flúor (camada de valência $2s^22p^5$) se ligam covalentemente ao átomo de gás nobre para ficarem com oito elétrons de valência.

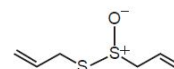
Ao se escrever a fórmula de Lewis do composto de xenônio citado, quantos elétrons na camada de valência haverá no átomo do gás nobre?

- A 6
- B 8
- C 10
- D 12
- E 14

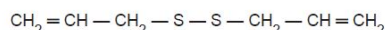
(b) Item 31 - Maior Valor de Dificuldade

Questão 122

O odor que permanece nas mãos após o contato com alho pode ser eliminado pela utilização de um "sabonete de aço inoxidável", constituído de aço inox (74%), cromo e níquel. A principal vantagem desse "sabonete" é que ele não se desgasta com o uso. Considere que a principal substância responsável pelo odor de alho é a alicina (estrutura I) e que, para que o odor seja eliminado, ela seja transformada na estrutura II.



Estrutura I



Estrutura II

Na conversão de I em II, o "sabonete" atuará como um

- A ácido.
- B redutor.
- C eletrólito.
- D tensoativo.
- E catalisador.

menores valores de dificuldade. Um segundo grupo, na coloração azul, representa o nível intermediário e o terceiro grupo, em vermelho, contem os itens mais difíceis, cujo intervalo sobrepõe o a distribuição de valores do grupo intermediário em grande medida.

Os gráficos de distribuição de probabilidade e também a direção da dispersão dos pontos na Figura 71a, aparentam uma eventual relação inversa das variáveis, em que a maior dificuldade implica num menor poder de discriminação. Na Figura 71b, observa-se a relação dos parâmetros de discriminação e propensão ao chute. Na coloração azul, o grupo 2, com valores acentuados para o coeficiente (a) e (c). O grupo de cor laranja também evidenciou itens com valores intermediários, porém para o coeficiente (c). Por fim, na coloração verde, os itens com os menores valores de discriminação. Na Figura 72a, apresentamos a variação da habilidade estimada ao longo CAT em duas situações, na primeira, com a exposição dos itens em sequência, com um visual bastante recortado, repleto de picos e vales. Na Figura 72b, a variação da habilidade apresenta um formato crescente, pois acompanha a chance de sucesso, em função do apoio do Modelo de 3PL e das características dos itens aplicados. Na Figura 73a, o resultado do CAT aplicado com variações de 15, 20 e 30 itens. Os valores das habilidades ficam agrupados em torno de 5 itens, na faixa de 15 a 20 e dispersos entre 20 e 30 itens. A variação de valores ocorre em função das características dos itens acertados. Itens com maior informação na região positiva dos desvios padrão da habilidade, representam maior pontuação. Na Figura 73b, o resultado do CAT por Grupos de Habilidade. Enquanto nos grupos de habilidade intermediária e superior, eram apresentados entre 15 e 30 itens, nos grupo de habilidade inferior, menos de 10 itens eram determinantes para a estimação de θ . Na Figura 74a, o

Figura 62 – Probabilidade de Sucesso e Informação dos Itens — CH

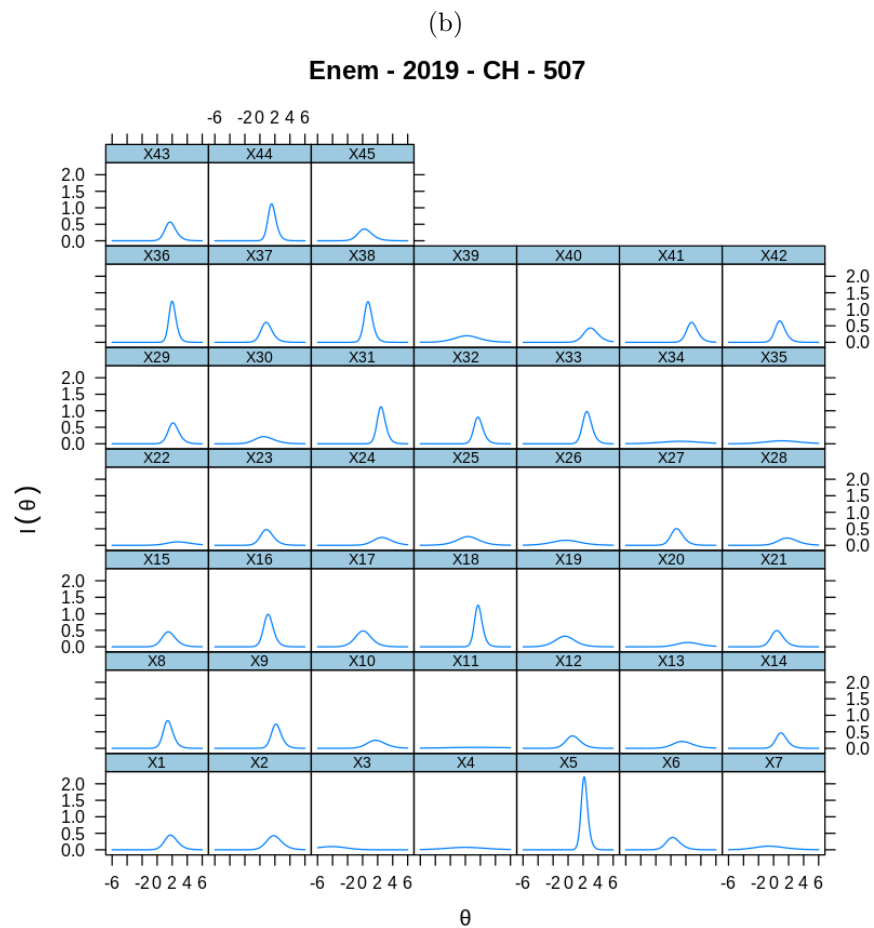
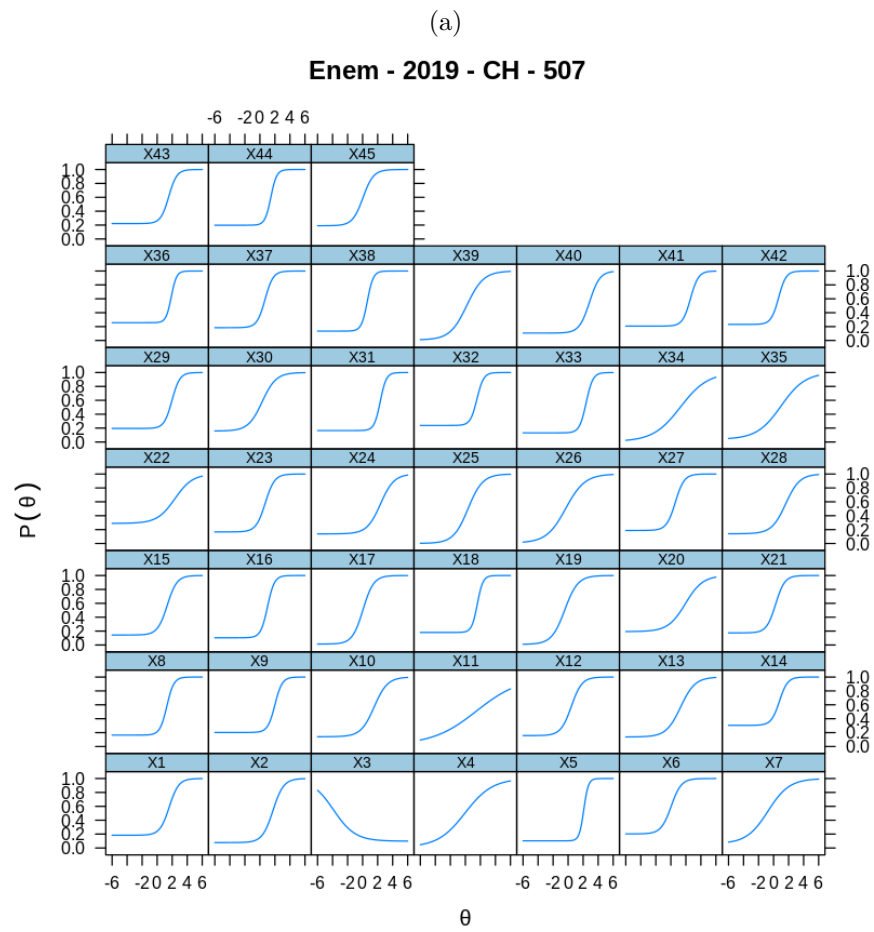
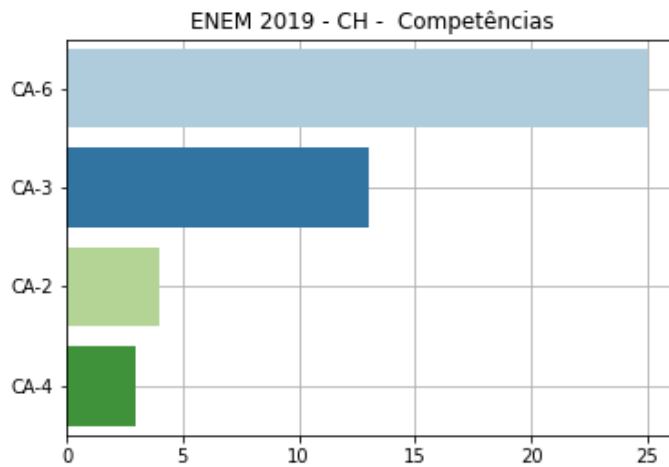
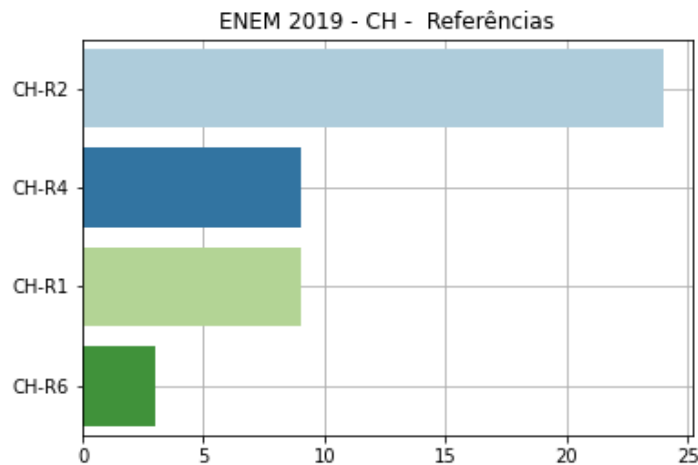


Figura 63 – Classificação Textual das Competências dos Itens de CH

(a)



(b)



CAT aplicado por quantidade de itens. Nota-se que os sujeitos com 15 itens, tem variação de habilidades do terceiro desvio padrão positivo até pouco abaixo de -1.5 . O teste com 30 itens apresentou maior possibilidade de alcance de habilidades em faixas de valores distintas.

Na Figura 74b, apresentamos o CAT com regra a variação da habilidade nos valores informados na legenda. A menor variação da habilidade alcançou maiores valores. Ao passo que a maior variação de θ não oportunizou uma maior pontuação. Como ilustrado, é possível explorar os itens conforme o interesse do avaliador. Além disso, por meio de técnicas computacionais, pode-se combinar as estratégias e, embarcado num formato de agente inteligente, a função objetivo pode assumir escopos combinados, segundo o espaço paramétrico disponível.

Figura 64 – Itens da Prova de CH

(a) Item 4 - Maior Valor de Discriminação

Questão 49

O Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (Ibama) está investigando o extermínio de abelhas por intoxicação por agrotóxicos em colmeias de São Paulo e Minas Gerais. Os estudos com inseticidas do tipo neonicotinoides devem estar concluídos no primeiro semestre de 2015. Trata-se de um problema de escala mundial, presente, inclusive, em países do chamado primeiro mundo, e que traz, como consequência, grave ameaça aos seres vivos do planeta, inclusive ao homem.

IBAMA. Polinizadores em risco de extinção são ameaça à vida do ser humano. Disponível em: www.mma.gov.br. Acesso em: 10 mar. 2014.

Qual solução para o problema apresentado garante a produtividade da agricultura moderna?

- A Preservação da área de mata ciliar.
- B Adoção da prática de adubação química.
- C Utilização da técnica de controle biológico.
- D Ampliação do modelo de monocultura tropical.
- E Intensificação da drenagem do solo de várzea.

(b) Item 14 - Maior Valor de Chute

Questão 59

A divisão política do mundo como apresentada na imagem seria possível caso o planeta fosse marcado pela estabilidade do(a)

- A ciclo hidrológico.
- B processo erosivo.
- C estrutura geológica.
- D índice pluviométrico.
- E pressão atmosférica.

6.4 Conclusão

Neste capítulo, realizamos experimentos para mostrar a vantagem de usar sistemas CAT para avaliar melhor e com mais precisão os alunos. O uso de um sistema automático também nos permite personalizar os processo de aprendizagem com *feedbacks* mais rápidos. Embora algumas estratégias alternativas estejam em disputa no literatura, a TRI é o modelo estatístico usual para referenciar o sistema CAT. Em nossos experimentos, nós adotamos o modelo da TRI de 3 parâmetros para mostrar duas aplicações do CAT e compararam seus resultados com as respectivas avaliações completas. Em ambas as aplicações, um número reduzido de itens alcançou praticamente a mesma capacidade calculada. Essa queda considerável nos números de itens para o teste implica uma enorme redução de custos que esta tecnologia pode trazer até o sistema educacional. Uma das dificuldades práticas para adotar uma tecnologia CAT por um, ou instituições, é o esforço para construir o banco calibrado de itens. Nossa proposta mostra como utilizar o Enem ou avaliações similares que disponham dos insumos básicos para a geração das características dos itens de avaliação.

Na literatura, verificamos abordagens de geração automática de itens de avaliação, por meio de técnicas de IA. Nossa contribuição num contexto deste seria de depurar os itens de qualidade, por meio da aplicação automática e análise dos coeficientes estimados e classificações textuais identificadas, sendo este o próximo passo pretendido por nossa linha de pesquisa. Além disso, outros tópicos incluídos direcionando nossos planos de pesquisa, como estudar modelos de TRI mais complexos para incorporá-los ao *plugin Moodle*.

Figura 65 – Distribuições das Habilidades de LC e MT

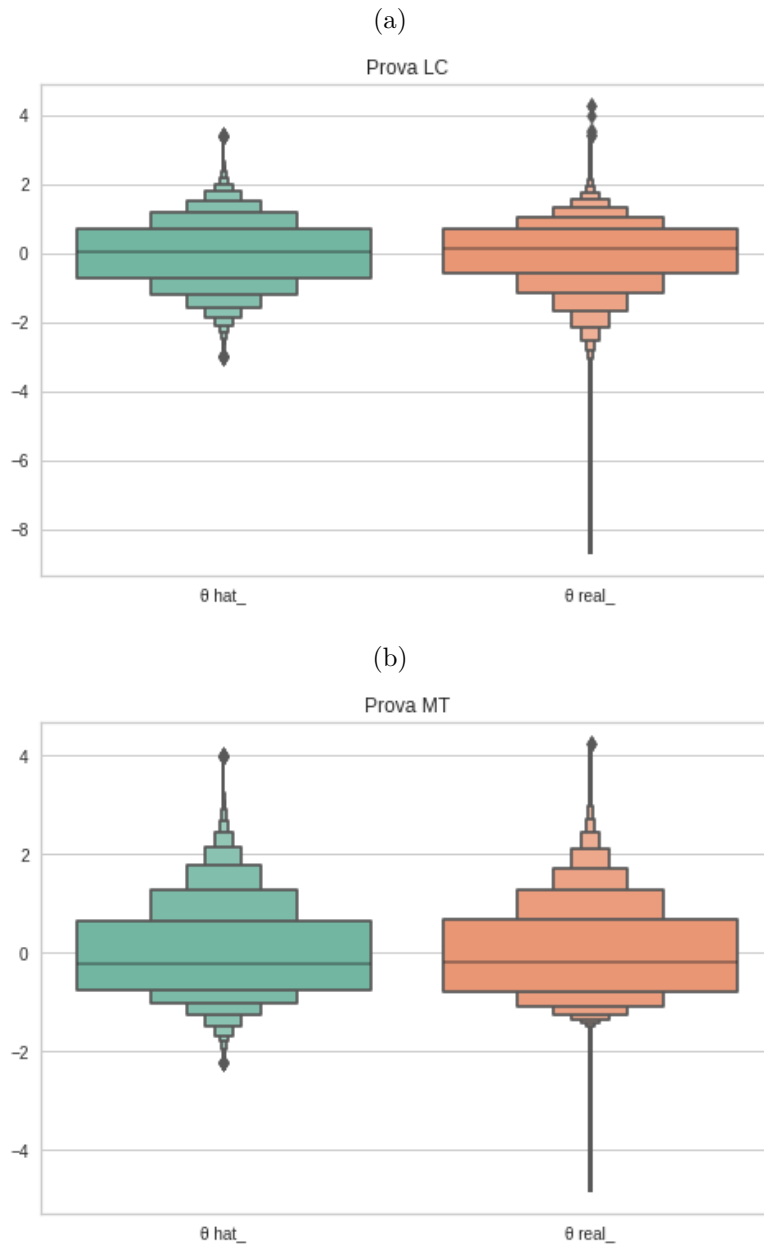


Figura 66 – Distribuições das Habilidades — CN e CH

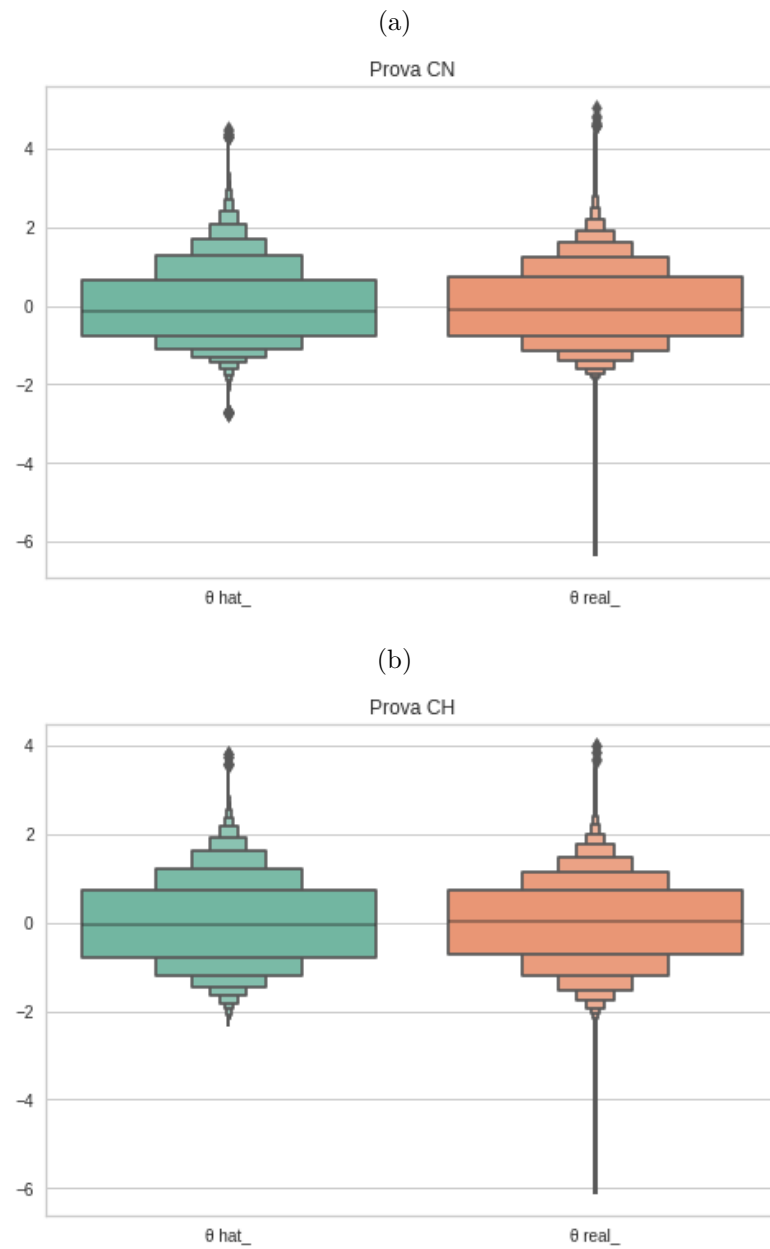


Figura 67 – Distribuições Conjuntas — LC e MT

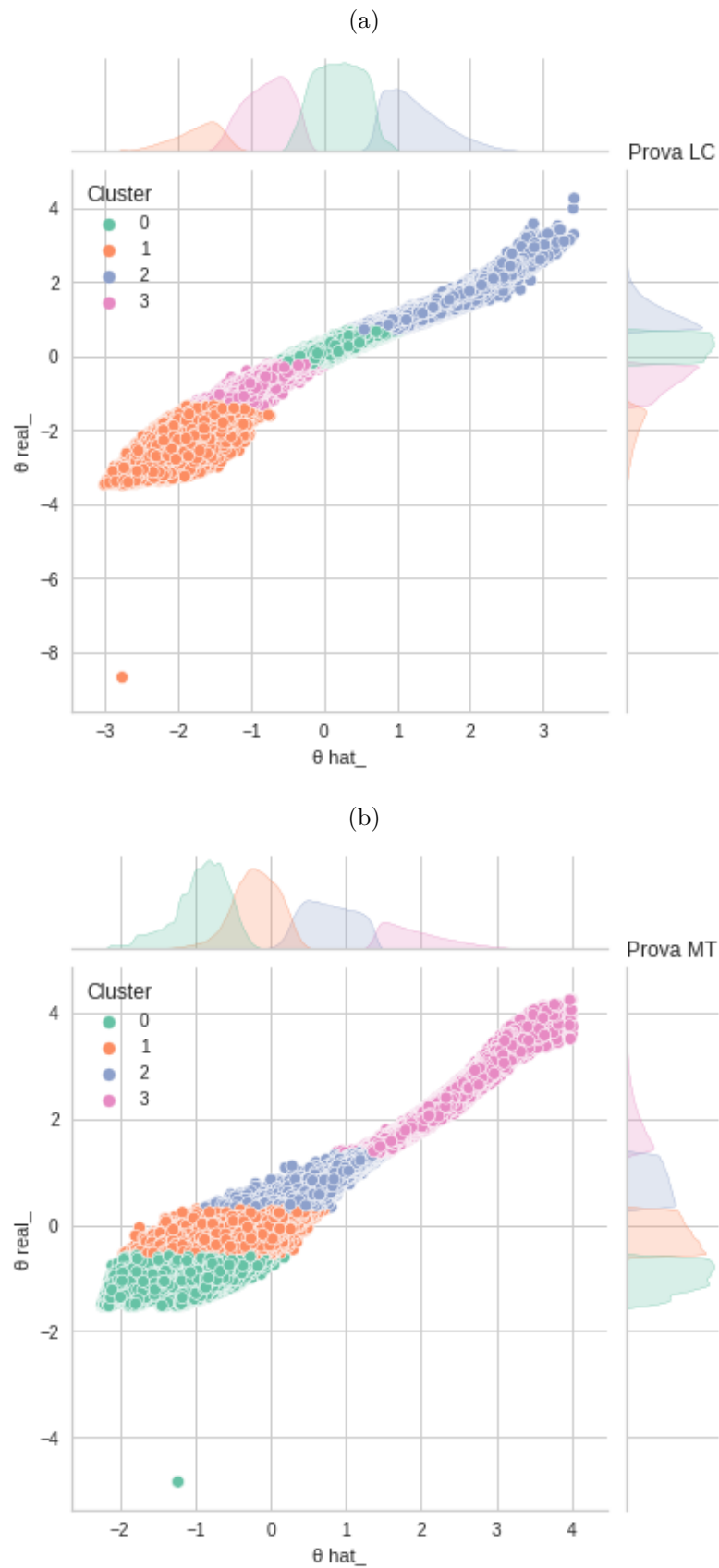


Figura 68 – Distribuições Conjuntas — CN e CH

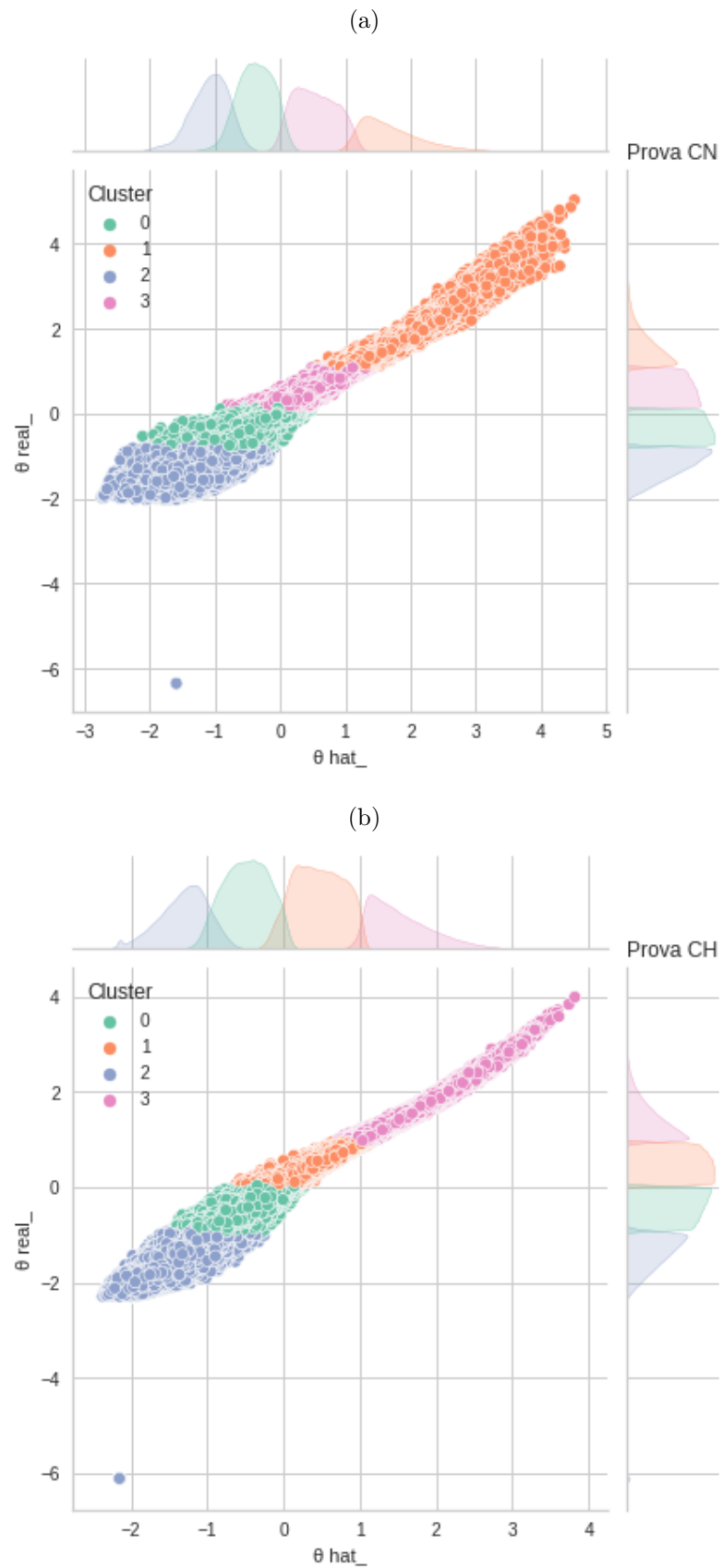


Figura 69 – Gráfico de Resíduos das Habilidades — LC e MT

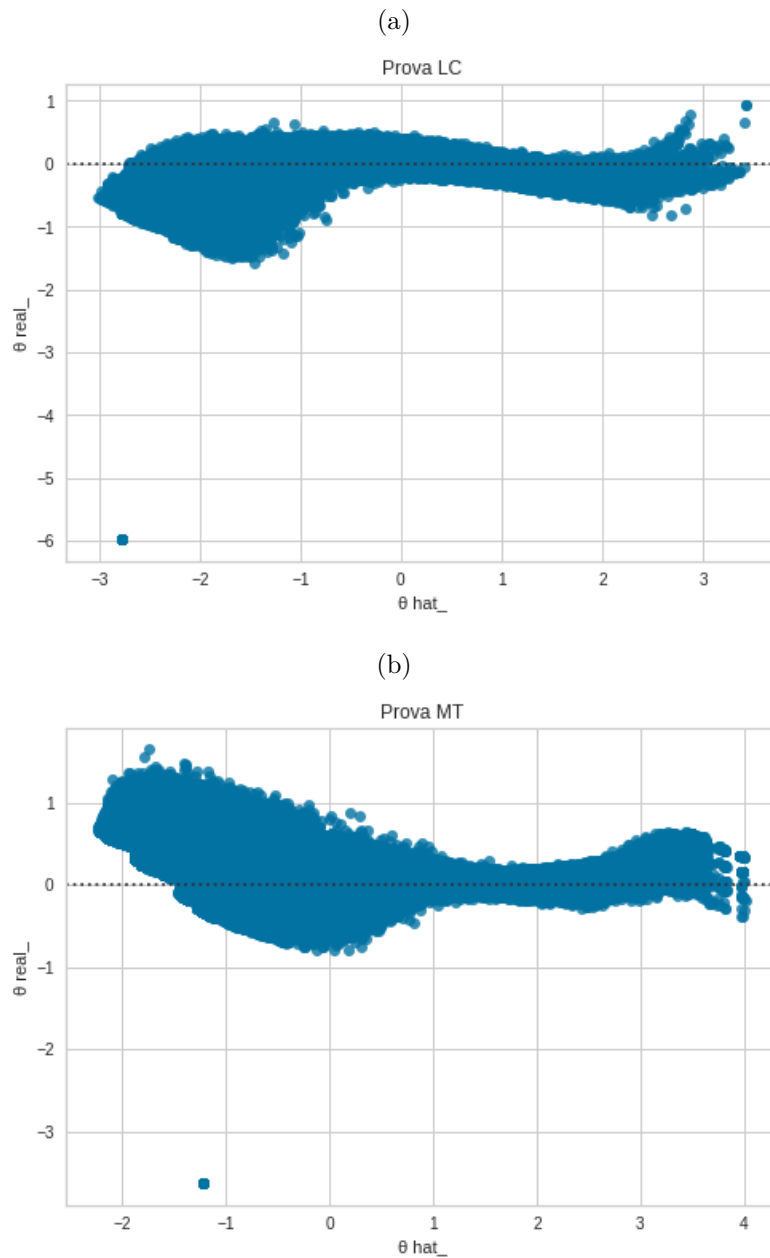


Figura 70 – Gráfico de Resíduos das Habilidades — CN e CH

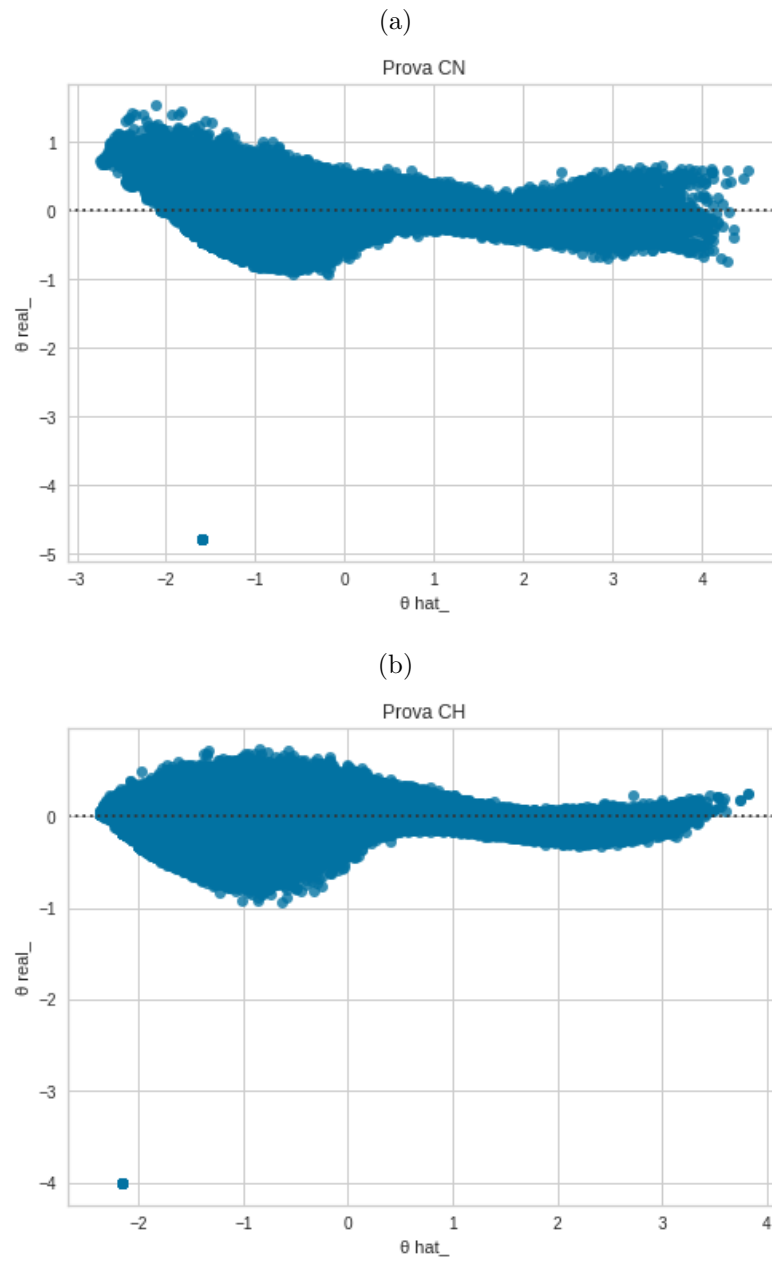
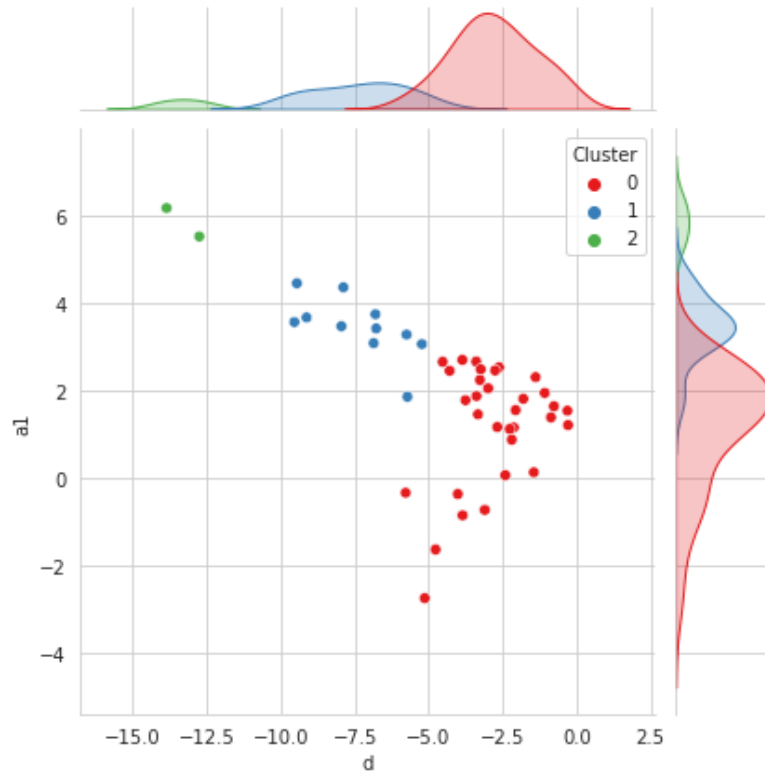


Figura 71 – Pares Ordenados e Densidades dos Parâmetros de MT

(a) Discriminação e Dificuldade



(b) Discriminação e Chute

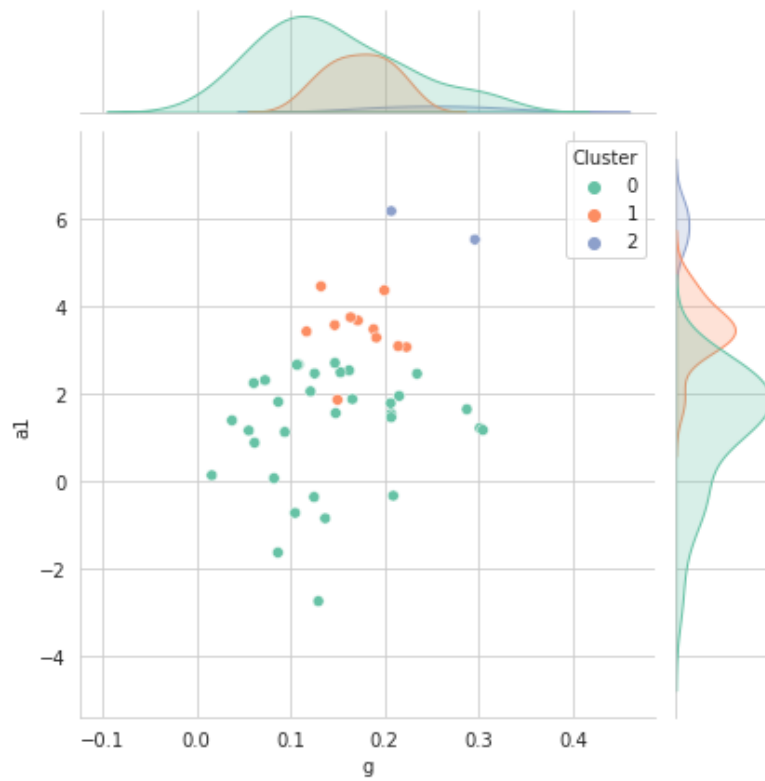
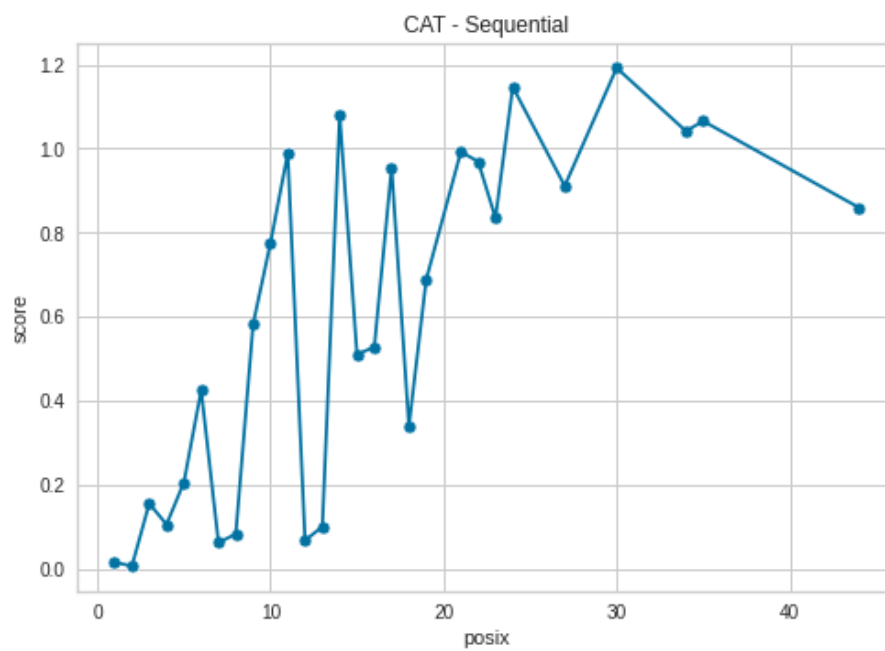


Figura 72 – Simulação do CAT

(a) Sequencial



(b) Otimizado

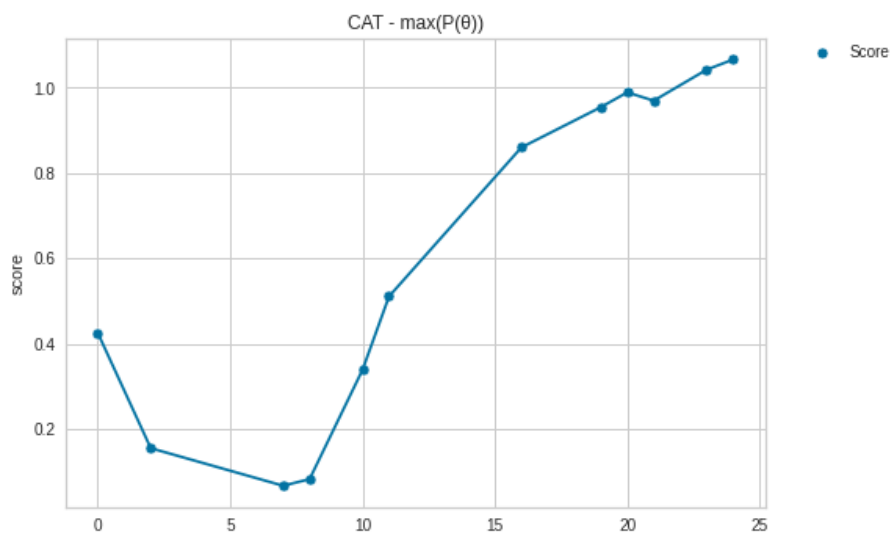
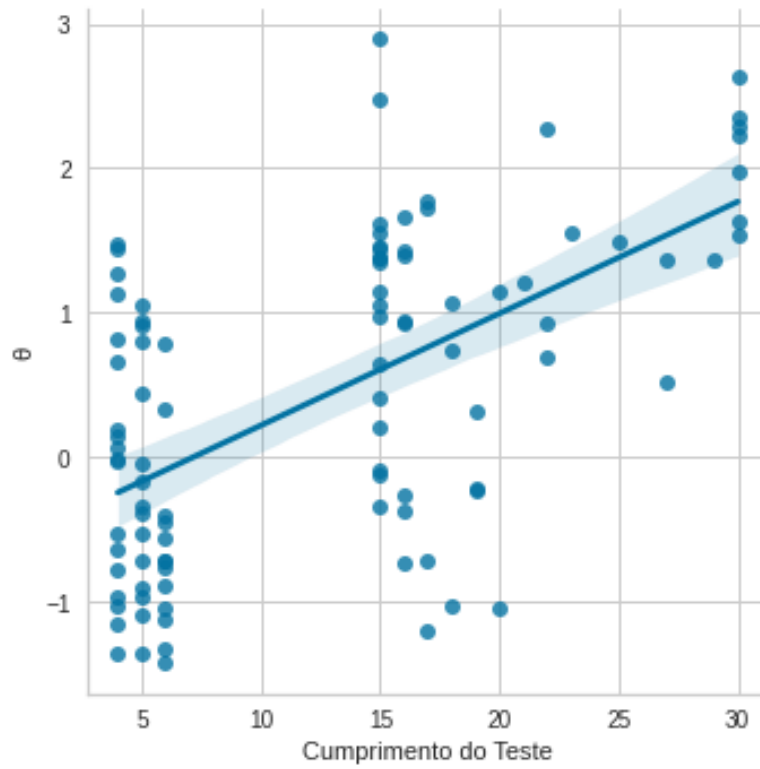


Figura 73 – Geral e Por Agrupamento

(a) Geral



(b) Otimizado por Grupo

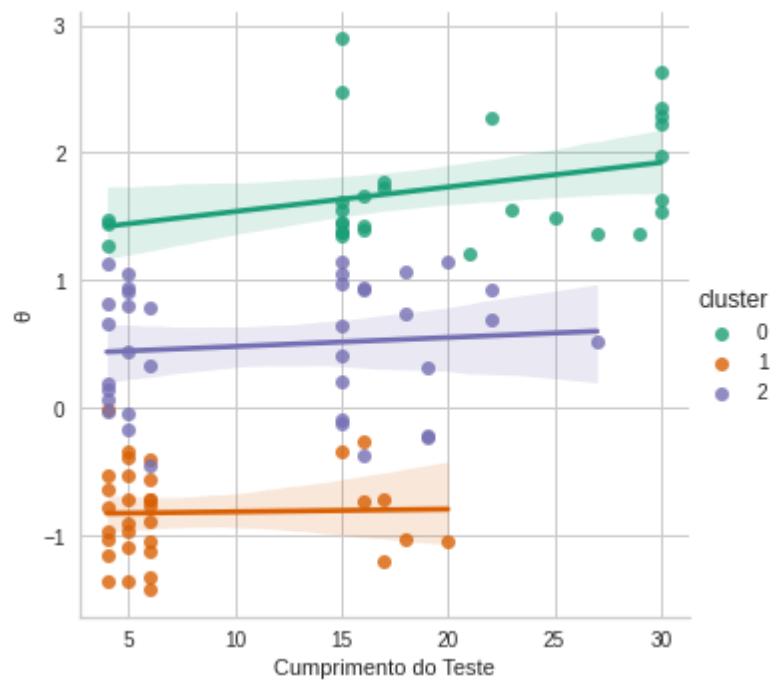
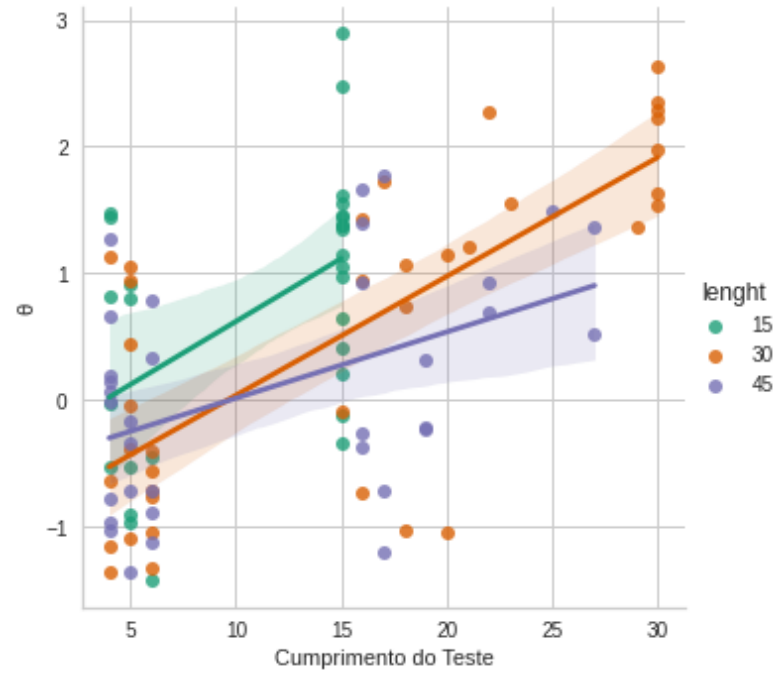
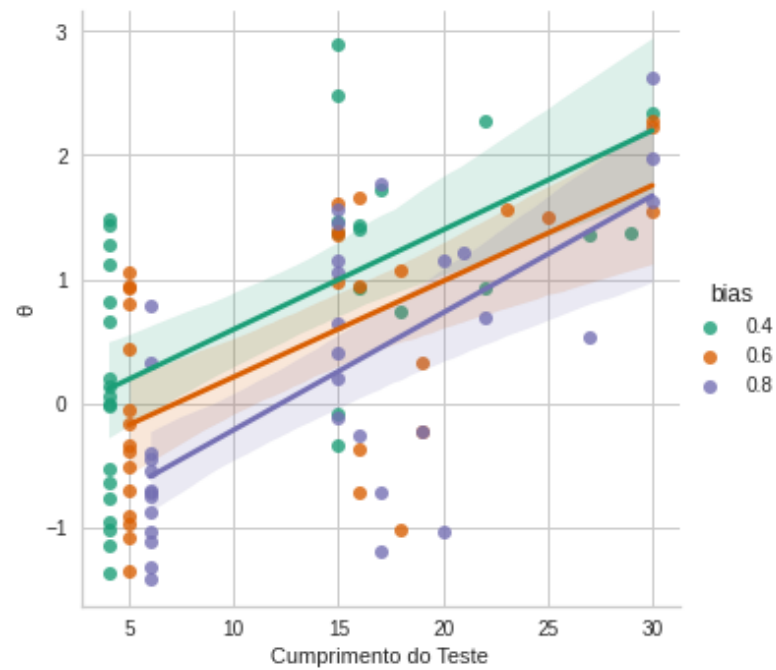


Figura 74 – Por Tamanho do Teste e Variação da Habilidade

(a) Por Quantidade de Itens

(b) Variação da θ 

7 Considerações Finais

O uso ferramentas computacionais, amparado por modelos estatísticos, pode auxiliar em grande medida o processo de ensino-aprendizagem. A discussão a respeito de melhores caracterizações pode ser popularizada por meio de bancos de itens formulados a partir de avaliações governamentais como o Enem. A utilização do Enem como ferramenta de seleção de acesso ao ensino superior, dentre outras possibilidades, oportuniza a geração de modelos preditivos de desempenho e tomando por base os valores seletivos do Sistema de Seleção Unificada que reúne em um sistema eletrônico gerido pelo MEC as vagas ofertadas por instituições públicas de ensino superior de todo o Brasil, sendo de maior proporção delas ofertada por instituições federais (universidades e institutos). Um panorama das inscrições pode ser visto na Figura ??, em que o auge foi visto no ano de 2014 e o menor menor patamar recente, ocorreu em 2021.

Na Figura ??, os quantitativos de participações, que apesar de decrescente, ainda representa milhares de estudante do país todo, haja visto que é instrumento utilizado no Sistema de Seleção Unificada (Sisu). Ele executa a seleção dos estudantes com base na nota do Exame Nacional do Ensino Médio (Enem). Até o limite da oferta das vagas, por curso e modalidade de concorrência, de acordo com as escolhas dos candidatos inscritos, eles são selecionados por ordem de maior classificação, em cada uma das duas edições anuais do Sisu. A classificação textual de itens de avaliação pode complementar o teste, por oportunizar mais características a serem consideradas ou descartadas, de acordo com objetivo educacional.

7.1 Contribuições da Pesquisa

A primeira contribuição deste trabalho foi formulação de um roteiro para identificar perfis de aprendizagem com base em critérios estatísticos e baseado numa heurística de otimização. O uso combinado de distâncias vetoriais para avaliar o rigor conferido a autoavaliação e a avaliação por pares. De mesmo modo, a utilização da similaridade entre vetores auxiliou a evidenciar padrões de performance. A explanação e aplicação dos Modelos da TRI proporcionaram rigor estatístico à seleção de itens para avaliação. Igualmente, os parâmetros estimados denotaram particularidades de cada prova analisada. A proposição da seleção automática de itens, num formato de CAT, com uma abordagem de agente inteligente, foi outra contribuição conceitual desta pesquisa, por aproximar da área da computação, transliterando termos ora usuais em testes computadorizados. Por fim, como estímulo á composição de bancos de itens baseados em provas nacionais, na Seção Anexo, listamos os gráficos das notas de Redação por estado brasileiro. Bem como,

as séries históricas agrupadas por região.

7.2 Trabalhos Futuros

A geração de características em objetos de aprendizagem é uma rota a ser perseguida na composição de uma abordagem holística da educação, mapeando lacunas de aprendizagem e recomendando objetos sob medida. Semelhante ao que é feito na TRI, com itens de avaliação, objetivamos complementar com modelos de seleção de conteúdo instrucional. Possíveis continuações desse trabalho incluem uma simplificação do modelo apresentado, necessitando de menos parâmetros para calibrar e procedimentos mais simples.

7.3 Conclusão

A utilização de arcabouço matemático estatístico na seara educacional ainda é escassa. A popularização de ferramentas para trabalhos complexos depende de discussão continuada e reforço das vantagens. A desmistificação de conceitos teóricos da psicometria pode também contribuir para o estímulo ao uso de técnicas mais precisas de mensuração de habilidades. A composição de banco de itens padronizados demanda tamanho esforço, semelhante ao que é visto na caracterização de perfis de aprendizagem. Métricas nacionais e internacionais de avaliação da educação estão à disposição da academia e este trabalho buscou ser mais um passo nessa construção continuada do conhecimento.

Referências

- ALVES, M.; MEDEIROS, F.; MELO, L. Levantamento do estado da arte sobre aprendizagem baseada em problemas na educação a distância e híbrida. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 61–71. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12762>>. Citado 2 vezes nas páginas 27 e 123.
- AMORIM, L. C. e Leandro Sanches e Lais Nascimento e Marlo Souza e R. Ontolo: Um modelo ontológico para avaliar o desempenho acadêmico na educação a distância. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 29, n. 1, p. 1898, 2018. ISSN 2316-6533. Disponível em: <<http://ojs.sector3.com.br/index.php/sbie/article/view/8183>>. Citado 4 vezes nas páginas 15, 62, 64 e 65.
- ASCARI, S.; GOTTARDO, E.; PIMENTEL, A. Identificação de intervenções tutoriais para ambientes virtuais de aprendizagem. In: . [S.l.: s.n.], 2020. p. 842–851. Citado na página 27.
- BAKER, F.; KIM, S.-H. *The Basics of Item Response Theory Using R*. [S.l.: s.n.], 2017. ISBN 978-3-319-54204-1. Citado na página 127.
- BAKER, F. B. *The basics of item response theory*. [S.l.]: Education Resources Information Center, 2001. Citado 4 vezes nas páginas 35, 82, 83 e 87.
- BAKSHINATEGH, B. et al. Educational Data Mining Applications and Tasks: A Survey of the Last 10 Years. *Education and Information Technologies*, Springer, v. 23, n. 1, p. 537–553, 2018. Citado na página 83.
- BARBOSA, M. de Souza Brito e Francisco Alencar de Medeiros e Ed Bezerra e A. R. Contribuições de um plugin do tipo report para a identificação do risco de evasão no ava moodle com base em visualização de dados. *Revista Brasileira de Informática na Educação*, v. 28, n. 0, p. 01–29, 2020. ISSN 2317-6121. Disponível em: <<http://ojs.sector3.com.br/index.php/rbie/article/view/v28p01>>. Citado na página 125.
- BARRETO, L. et al. Mineração de dados educacionais a partir da interação de alunos com uma plataforma educacional. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1052–1061. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12861>>. Citado na página 36.
- BOUD, D. *Enhancing learning through self-assessment*. Abingdon-on-Thames, United Kingdom: Routledge, 2013. Citado na página 68.
- BOURGUET, J. Purely synthetic and domain independent consistency-guaranteed populations in SHIQ(D). In: LOSSIO-VENTURA, J. A.; ALATRISTA-SALAS, H. (Ed.). *4th Annual International Symposium, SIMBig 2017, Lima, Peru, September 4-6, 2017*. [S.l.]: Springer, 2017. (Communications in Computer and Information Science, v. 795), p. 76–89. Citado na página 85.

- BOURGUET, J.-R.; SILVA, W.; OLIVEIRA, E. de. Minimalist fitted bayesian classifier-based on likelihood estimations and bag-of-words. In: BERGET, G. et al. (Ed.). *Linking Theory and Practice of Digital Libraries*. Cham: Springer International Publishing, 2021. p. 17–28. ISBN 978-3-030-86324-1. Citado 2 vezes nas páginas 33 e 123.
- BRESLOW, L. et al. Studying learning in the worldwide classroom: Research into edx's first mooc. *Research & Practice in Assessment*, Education Resources Information Center, v. 8, p. 13–25, 2013. Citado na página 68.
- BRITO, W. A.; MOTTA, C. L. Recomendação de Jogos na Aprendizagem da Matemática baseado na Análise Diagnóstica e Teoria de Resposta ao Item. In: *Anais do XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014)*. [S.l.: s.n.], 2014. v. 1, n. Cbie, p. 944. Citado na página 108.
- BUSSAB, W. O.; MORETTIN, P. A. *Basic statistics*. 8. ed. São Paulo: Saraiva, 2013. Citado 4 vezes nas páginas 69, 71, 73 e 86.
- CARLESS, D. et al. Developing sustainable feedback practices. *Studies in higher education*, Routledge, Abingdon-on-Thames, United Kingdom, v. 36, n. 4, p. 395–407, 2011. Citado na página 67.
- CHALK, B.; ADEBOYE, K. Peer assessment of program code: a comparison of two feedback instruments. In: *6th Annual Conference of the Higher Education Academy Subject Centre for Information and Computer Sciences*. Heslington, UK: Higher Education Academy, 2005. p. 106–110. Citado na página 69.
- CHALMERS, R. P. mirt: A multidimensional item response theory package for the r environment. *Journal of Statistical Software*, v. 48, n. 6, p. 1–29, 2012. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v048i06>>. Citado na página 36.
- CHAPMAN, O. L.; FIORE, M. A. Calibrated peer review™. *Journal of Interactive Instruction Development*, Learning Technology Institute, v. 12, n. 3, p. 11–15, 2000. Citado na página 69.
- CHO, K.; SCHUNN, C. D. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Comput. Educ.*, Elsevier Science Ltd., GBR, v. 48, n. 3, p. 409–426, abr. 2007. ISSN 0360-1315. Disponível em: <<https://doi.org/10.1016/j.compedu.2005.02.004>>. Citado na página 69.
- COSTA, L.; OLIVEIRA, E. H. T. de; Castro Júnior, A. Corretor Automático de Redações em Língua Portuguesa: um mapeamento sistemático de literatura. In: *SBIE - Simpósio Brasileiro de Informática na Educação*. [S.l.: s.n.], 2020. p. 1403–1412. Citado 2 vezes nas páginas 108 e 114.
- COSTA, N.; FERNANDES, M. Sequenciamento de ações pedagógicas por algoritmo genético utilizando taxonomia de bloom e assist. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1273–1282. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12883>>. Citado na página 37.
- COSTA, N. Torrezão da; JÚNIOR, C. P.; FERNANDES, M. Recomendação de ações pedagógicas utilizando planejamento automático e taxonomia digital de bloom. In: . [S.l.: s.n.], 2019. p. 1531. Citado na página 27.

- CURI, M. et al. Interpretable Variational Autoencoders for Cognitive Models. *Proceedings of the International Joint Conference on Neural Networks*, v. 2019-July, 2019. Citado na página 107.
- CURI, M.; SILVA, V. Academic English Proficiency Assessment Using a Computerized Adaptive Test. *TEMA (SÁCarlos)*, v. 20, 08 2019. ISSN 2179-8451. Citado 2 vezes nas páginas 107 e 110.
- Da Cruz Alves, N. et al. Análise do Nível de Dificuldade dos Conceitos de Design de Interface de Usuário usando a Teoria de Resposta ao Item. In: *XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)*. [S.l.: s.n.], 2020. p. 1563–1572. Citado na página 108.
- DAVID, P.; GREEN, J.; SANTOS, V. Contextos de aprendizagem para interações contingentes em fóruns de discussão online: uma investigação à luz da etnografia interacional. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 332–341. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12789>>. Citado na página 36.
- DAVID, T. N. e Gislaïne Coelho e Fernanda Campos e Regina Braga e Victor Stroele e J. Uso de proveniência de objetos de aprendizagem para identificação do estilo preferencial de aprendizagem. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 30, n. 1, p. 109, 2019. ISSN 2316-6533. Disponível em: <<http://ojs.sector3.com.br/index.php/sbie/article/view/8715>>. Citado na página 27.
- DAVIES, P. Computerized peer assessment. *Innovations in Education and Training International*, Routledge, v. 37, n. 4, p. 346–355, 2000. Disponível em: <<https://doi.org/10.1080/135580000750052955>>. Citado na página 69.
- De Jesus, H. O.; RODRIGUEZ, L. C.; Costa Junior, A. D. O. Predição de Evasão Escolar na Licenciatura em Computação. *Revista Brasileira de Informática na Educação*, v. 29, p. 255–272, 2021. ISSN 1414-5685. Citado na página 66.
- DENNY, P.; LUXTON-REILLY, A.; SIMON, B. Quality of student contributed questions using peerwise. In: *Proceedings of the Eleventh Australasian Conference on Computing Education, Wellington, New Zealand*. Sydney, Australia: Australian Computer Society, Inc., 2009. (ACE'09, v. 95), p. 55–63. Citado na página 69.
- FALCHIKOV, N.; GOLDFINCH, J. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, v. 70, n. 3, p. 287–322, 2000. Disponível em: <<https://doi.org/10.3102/00346543070003287>>. Citado na página 69.
- FARIDA, B. et al. Adaptive exercises generation using an automated evaluation and a domain ontology: the odala+ approach. *International Journal of Emerging Technologies in Learning (iJET)*, International Association of Online Engineering, v. 6, n. 2, p. 4–10, 2011. Citado na página 84.
- FERREIRA, F. et al. Trimelo: Uma proposta de modelo para avaliação multi-habilidades. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 102–111. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12766>>. Citado na página 35.

FERREIRA, F. S. e Yan Mendes e Bruno Dembogurski e Gustavo Semaan e Edelberto Silva e F. Evasão no ensino básico da rede pública municipal de juiz de fora: uma análise com mineração de dados. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 30, n. 1, p. 1371, 2019. ISSN 2316-6533. Disponível em: <<http://ojs.sector3.com.br/index.php/sbie/article/view/8869>>. Citado na página 35.

FERREIRA, R. Off-Topic Essay Detection : A comparative study on the Portuguese language. *Revista Brasileira de Informática na Educação - RBIE*, v. 27, n. 3, p. 177–190, 2019. Citado na página 108.

FITZGERALD, D. *Education data*. 2021. Disponível em: <<https://dataqualitycampaign.org/>>. Citado na página 66.

FITZGERALD, D. *Data Quality Campaign — dataqualitycampaign.org*. 2022. <<https://dataqualitycampaign.org/>>. [Accessed 12-Oct-2022]. Citado na página 66.

GARCIA, L. M.; GOMES, R. S. Visualização e análise da trajetória de aprendizagem realizada no currículo no ensino superior. In: . [S.l.: s.n.], 2020. p. 1593–1602. Citado 2 vezes nas páginas 35 e 125.

HENRIQUE, M. et al. Um Modelo para Geração de Perfis de Usuários Baseado em Técnicas de Psi- cometria. *Revista Brasileira de Informática na Educação – RBIE*, v. 17, n. 3, p. 25–37, 2009. Citado na página 109.

HERLOCKER, J. L. et al. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, ACM, v. 22, n. 1, p. 5–53, 2004. Citado na página 81.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo Demográfico 1991*. 1991. <<https://www.ibge.gov.br/estatisticas/sociais/populacao/1569-censo-demografico-1991.html>>. Acesso em: 10 mar. 2023. Citado na página 37.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo Demográfico 2000*. 2000. <<https://www.ibge.gov.br/estatisticas/sociais/populacao/8453-censo-demografico-2000.html>>. Acesso em: 10 mar. 2023. Citado na página 37.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo Demográfico 2010*. 2010. <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html>>. Acesso em: 10 mar. 2023. Citado na página 37.

JATOBÁ, V. et al. Testes Adaptativos Computadorizados baseados na Teoria de Resposta ao Item em Sistemas e-learning: Uma revisão sistemática da literatura. In: *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*. [S.l.: s.n.], 2017. v. 1, n. July 2018, p. 273. Citado na página 109.

JOHNS, J.; MAHADEVAN, S.; WOOLF, B. P. Estimating student proficiency using an item response theory model. In: IKEDA, M.; ASHLEY, K. D.; CHAN, T. (Ed.). *Intelligent Tutoring Systems, 8th International Conference, ITS 2006, Zhongli, Taiwan, June 26-30, 2006, Proceedings*. [S.l.]: Springer, 2006. (Lecture Notes in Computer Science, v. 4053), p. 473–480. Citado 2 vezes nas páginas 83 e 84.

Júnior, C. P. et al. Personalized recommendation of learning objects through

- bio-inspired algorithms and semantic web technologies: an experimental analysis. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1333–1342. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12889>>. Citado na página 126.
- LEE, Y.; CHO, J. Personalized item generation method for adaptive testing systems. *Multimedia Tools Appl.*, v. 74, n. 19, p. 8571–8591, 2015. Disponível em: <<https://doi.org/10.1007/s11042-013-1421-0>>. Citado na página 84.
- LEE, Y. et al. A personalized assessment system based on item response theory. In: LUO, X. et al. (Ed.). *Advances in Web-Based Learning - ICWL 2010 - 9th International Conference, Shanghai, China, December 8-10, 2010. Proceedings*. [S.l.]: Springer, 2010. (Lecture Notes in Computer Science, v. 6483), p. 381–386. Citado na página 84.
- LIEBERMAN, D. A. *Learning: Behavior and Cognition*. Belmont, CA: Wadsworth Publishing Company, 1990. (International Student Edition). Citado na página 83.
- LIMA, P. D. S. N. et al. Análise de conteúdo das provas do Enade para os alunos do curso de Bacharelado em Ciência da Computação. *Revista Brasileira de Informática na Educação*, v. 29, p. 385–413, 2021. ISSN 1414-5685. Citado 2 vezes nas páginas 27 e 108.
- LINDEN, W. J. van der; HAMBLETON, R. K. *Handbook of modern item response theory*. [S.l.]: Springer Science & Business Media, 2013. Citado na página 84.
- LUO, H. et al. Applying Item Response Theory Analysis to the Montreal Cognitive Assessment in a Low-Education Older Population. *Assessment*, v. 27, n. 7, p. 1416–1428, 2020. ISSN 15523489. Citado na página 109.
- LUXTON-REILLY, A. A systematic review of tools that support peer assessment. *Computer Science Education*, v. 19, n. 4, p. 209–232, 2009. Disponível em: <<https://doi.org/10.1080/08993400903384844>>. Citado na página 80.
- MANGAROSKA, K.; VESIN, B.; GIANNAKOS, M. N. Cross-platform analytics: A step towards personalization and adaptation in education. In: *Proceedings of the 9th International Conference on Learning Analytics & Knowledge, LAK 2019, Tempe, AZ, USA, March 4-8, 2019*. [S.l.]: ACM, 2019. p. 71–75. Citado 2 vezes nas páginas 103 e 122.
- MANOUSELIS, N. et al. Recommender systems in technology enhanced learning. In: *Recommender systems handbook*. [S.l.]: Springer, 2011. p. 387–415. Citado na página 83.
- MARCHETI, A.; BELHOT, R. Taxonomia de bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. *Gestão Produção*, v. 17, 01 2010. Citado na página 62.
- MARUTHO, D. et al. The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In: *2018 International Seminar on Application for Technology of Information and Communication*. [S.l.: s.n.], 2018. p. 533–538. Citado na página 37.
- MAYO, M.; MITROVIC, A. Optimising its behaviour with bayesian networks and decision theory. *International Artificial Intelligence Education Society*, Springer, v. 12, p. 124–153, 2001. Citado na página 83.

MCMILLAN, J. H.; HEARN, J. Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, Journal Storage, v. 87, n. 1, p. 40–49, 2008. Citado na página 68.

MONTEVERDE, I. et al. Framework conceitual para formação de grupos de alunos utilizando trilhas de aprendizagem em um ambiente virtual de aprendizagem. In: . [S.l.: s.n.], 2018. p. 1673. Citado na página 125.

NETO, S. et al. Avaliação automática de redações na língua portuguesa baseada na coleta de atributos e aprendizagem de máquina. In: . [S.l.: s.n.], 2020. p. 1162–1171. Citado 2 vezes nas páginas 108 e 109.

NGU, A. H.; SHEPHERD, J.; MAGIN, D. Engineering the 'peers' system: the development of a computer-assisted approach to peer assessment. *Research and Development in Higher Education*, Citeseer, v. 18, p. 582–587, 1995. Citado na página 69.

NICOL, D.; THOMSON, A.; BRESLIN, C. Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, Routledge, Abingdon-on-Thames, United Kingdom, v. 39, n. 1, p. 102–122, 2014. Citado na página 68.

OLIVEIRA, E. et al. The influence of ner on the essay grading. In: ABRAHAM, A. et al. (Ed.). *Intelligent Systems Design and Applications*. Cham: Springer International Publishing, 2021. p. 162–171. ISBN 978-3-030-49342-4. Citado na página 108.

OLIVEIRA, E. et al. A Framework for Peer Assessment in Programming Classes. In: SÁNCHEZ, J. (Ed.). *XXII Conferência Internacional sobre Informática na Educação – TISE*. Fortaleza, Brazil: <http://tise.cl/>, 2017. v. 13, p. 298–304. Disponível em: <http://www.tise.cl/volumen13/TISE2017/34.pdf>. Citado 3 vezes nas páginas 37, 67 e 69.

OLIVEIRA, E.; IZO, F.; SILVA, W. Personalizing Evaluations to Aid Chemistry. In: *INTED2023 Proceedings*. IATED, 2023. (17th International Technology, Education and Development Conference), p. 8183–8189. ISBN 978-84-09-49026-4. ISSN 2340-1079. Disponível em: <https://doi.org/10.21125/inted.2023.2223>. Citado na página 33.

OLIVEIRA, E.; SPALENZA, M. A. Self and Peer Assessment Strategies. In: *Anais do Computer on the Beach*. Florianópolis, Brazil: <https://computeronthebeach.com.br/>, 2017. p. 396–405. Disponível em: <https://siaiap32.univali.br/seer/index.php/acotb/article/view/10625>. Citado na página 69.

OLIVEIRA, E. de et al. A prototypical semantic annotator for A tribuna newspaper. In: LEMOS, D. L. da S. et al. (Ed.). *Proceedings of the XIII Seminar on Ontology Research in Brazil and IV Doctoral and Masters Consortium on Ontologies (ONTOBRAS 2020)*, Vitória, Brazil, November 23-26, 2020. CEUR-WS.org, 2020. (CEUR Workshop Proceedings, v. 2728), p. 22–34. Disponível em: <http://ceur-ws.org/Vol-2728/paper2.pdf>. Citado 2 vezes nas páginas 33 e 123.

OLIVEIRA, J. A. e Wesley Silva e Jessica Brito e E. Avaliação em pares e autoavaliação: Um modelo estatístico para perfilação de alunos. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 29, n. 1, p. 1653, 2018. ISSN 2316-6533. Disponível em:

<<https://br-ie.org/pub/index.php/sbie/article/view/8125>>. Citado 7 vezes nas páginas 32, 35, 36, 67, 69, 80 e 85.

OLIVEIRA, M. G.; CIARELLI, P. M.; OLIVEIRA, E. Recommendation of programming activities by multi-label classification for a formative assessment of students. *Expert Systems with Applications*, v. 40, n. 16, p. 6641–6651, 2013. ISSN 09574174. Citado 5 vezes nas páginas 35, 81, 85, 104 e 122.

OLIVEIRA, P.; SILVA, P.; NETO, P. R. Ontologia dos objetivos educacionais. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1183–1192. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12874>>. Citado 2 vezes nas páginas 37 e 62.

PARÉ, D. E.; JOORDENS, S. Peering into large lectures: examining peer and expert mark agreement using peerscholar, an online peer assessment tool. *J. Comp. Assisted Learning*, v. 24, n. 6, p. 526–540, 2008. Disponível em: <<https://doi.org/10.1111/j.1365-2729.2008.00290.x>>. Citado na página 69.

PARRAGA, J. G. e Milene Silveira e O. Processo de adaptação de objetos de aprendizagem com a inserção de técnicas de learning analytics. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 30, n. 1, p. 1681, 2019. ISSN 2316-6533. Disponível em: <<http://ojs.sector3.com.br/index.php/sbie/article/view/8901>>. Citado na página 126.

PASQUALI, L. *Psicometria – Teoria dos Testes na Psicologia e na Educação*. Petrópolis, RJ: Editora Vozes, 2004. 158–19 p. Citado na página 82.

PAZZANI, M. J.; BILLSUS, D. Content-based recommendation systems. In: BRUSILOVSKY, P.; KOBSA, A.; NEJDL, W. (Ed.). *The Adaptive Web, Methods and Strategies of Web Personalization*. [S.l.]: Springer, 2007. (Lecture Notes in Computer Science, v. 4321), p. 325–341. Citado na página 81.

PERRENOUD, P. *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages*. Bruxelles: De Boeck, 1998. Citado 5 vezes nas páginas 35, 68, 69, 83 e 85.

PIROVANI, J. P. C. et al. Adapting ner (crf+lg) for many textual genres. In: *IberLEF@SEPLN*. [S.l.: s.n.], 2019. Citado na página 33.

PLAJNER, M. Probabilistic models for computerized adaptive testing. *CoRR*, abs/1703.09794, 2017. Disponível em: <<http://arxiv.org/abs/1703.09794>>. Citado 2 vezes nas páginas 28 e 124.

RAABE, A. F. e Eliane Tomazoni e Rosana Paza e Rogério Perego e A. Bloom's taxonomy-based approach for assisting formulation and automatic short answer grading. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, v. 29, n. 1, p. 238, 2018. ISSN 2316-6533. Disponível em: <<http://ojs.sector3.com.br/index.php/sbie/article/view/7978>>. Citado na página 61.

REYES, D. D. L. et al. Predição de sucesso acadêmico de estudantes: uma análise sobre a demanda por uma abordagem baseada em transfer learning. *Revista Brasileira de Informática na Educação*, v. 27, p. 01, 01 2019. Citado na página 125.

ROIJERS, D. M.; JEURING, J.; FEELDERS, A. Probability estimation and a competence model for rule based e-tutoring systems. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. New York, NY, USA: Association for Computing Machinery, 2012. (LAK '12), p. 255–258. ISBN 9781450311113. Disponível em: <<https://doi.org/10.1145/2330601.2330663>>. Citado na página 84.

ROMERO, C. et al. Personalized links recommendation based on data mining in adaptive educational hypermedia systems. In: DUVAL, E.; KLAMMA, R.; WOLPERS, M. (Ed.). *Creating New Learning Experiences on a Global Scale, Second European Conference on Technology Enhanced Learning, EC-TEL 2007, Crete, Greece, September 17-20, 2007, Proceedings*. [S.l.]: Springer, 2007. (Lecture Notes in Computer Science, v. 4753), p. 292–306. Citado 2 vezes nas páginas 28 e 83.

ROMERO, O. O.; MACHADO, G. M.; WIVES, L. Analysis of questionnaires for Virtual Learning Environments based on Item Response Theory. In: *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019) Analysis*. [S.l.: s.n.], 2019. p. 509. ISSN 2316-6533. Citado 3 vezes nas páginas 35, 36 e 109.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: a modern approach*. 3. ed. [S.l.]: Pearson, 2009. Citado na página 81.

SAMEJIMA, F. Graded response model. In: _____. *Handbook of Modern Item Response Theory*. New York, NY: Springer New York, 1997. p. 85–100. ISBN 978-1-4757-2691-6. Disponível em: <https://doi.org/10.1007/978-1-4757-2691-6_5>. Citado 3 vezes nas páginas 110, 111 e 117.

SANTOS, G.; BASSANI, P. Articulando estudos nas áreas de design da aprendizagem, design thinking e canvas, para a proposição de currículos baseados em competências. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 192–201. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12775>>. Citado na página 35.

SANTOS, J. S. dos et al. Jogo para Mensuração de Habilidades Predictoras de Leitura : Construção e Análise. *Revista Brasileira de Informática na Educação – RBIE*, v. 28, n. 1, p. 260–272, 2020. Citado na página 109.

SCHANK, R.; SAUNDERS, H. Virtual learning: A revolutionary approach to building a highly skilled workforce. *Performance Improvement*, v. 40, n. 5, p. 39–41, 2001. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/pfi.4140400511>>. Citado na página 67.

SEN, S.; COHEN, A. S. Applications of mixture irt models: A literature review. *Measurement: Interdisciplinary Research and Perspectives*, Taylor & Francis, v. 17, n. 4, p. 177–191, 2019. Citado na página 84.

SHANMUGAM, R. *Handbook of Item Response Theory: Volume one, Models*. [S.l.: s.n.], 2019. 1–1 p. ISSN 0094-9655. ISBN 9781466514317. Citado na página 107.

SHIBA, Y.; SUGAWARA, T. Fair assessment of group work by mutual evaluation based on trust network. In: *IEEE Frontiers in Education Conference, FIE 2014, Proceedings, Madrid, Spain, October 22-25, 2014*. Washington, DC, USA: IEEE Computer Society, 2014. p. 1–7. Disponível em: <<https://doi.org/10.1109/FIE.2014.7044121>>. Citado na página 68.

- SHISHEHCHI, S.; BANIHASHEM, S.; ZIN, N. A. M. A proposed semantic recommendation system for e-learning: A rule and ontology based e-learning recommendation system. In: . [S.l.: s.n.], 2010. v. 1, p. 1 – 5. Citado na página 84.
- SILVA, F. Beletti da et al. Ferramenta de apoio ao planejamento educacional. 11 2020. Citado 2 vezes nas páginas 123 e 125.
- SILVA, G. et al. Discovery of research gaps in game learning analytics applications on learning processes. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. Porto Alegre, RS, Brasil: SBC, 2020. p. 1203–1212. ISSN 0000-0000. Disponível em: <<https://sol.sbc.org.br/index.php/sbie/article/view/12876>>. Citado na página 125.
- SILVA, L. da; BARBOSA, J.; RIGO, S. Análise de dados e serviços inteligentes aplicados na educação à distância: um mapeamento sistemático. *Revista Brasileira de Informática na Educação*, v. 29, p. 331–357, 04 2021. Citado na página 66.
- SILVA, L. S. e.; CAVALCANTI, E. R. Avaliação Experimental do Robomind no Ensino de Programação com Estudantes do Curso Técnico em Informática Integrado ao Ensino Médio. In: *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*. [S.l.: s.n.], 2018. v. 1, n. Cbie, p. 288. Citado na página 125.
- SILVA, W. et al. An easy-to-read visual approach to deal with peer reviews and self-assessments in virtual learning environments. In: *Proceedings of the 2020 The 3rd International Conference on Big Data and Education*. New York, NY, USA: Association for Computing Machinery, 2020. (ICBDE '20), p. 73–79. ISBN 9781450374989. Disponível em: <<https://doi.org/10.1145/3396452.3396461>>. Citado 6 vezes nas páginas 15, 32, 33, 36, 70 e 107.
- SILVA, W. et al. Towards a tailored hybrid recommendation-based system for computerized adaptive testing through clustering and IRT. In: LANE, H. C.; ZVACEK, S.; UHOMOIBHI, J. (Ed.). *Proceedings of the 12th International Conference on Computer Supported Education, CSEDU 2020, Prague, Czech Republic, May 2-4, 2020, Volume 1*. SCITEPRESS, 2020. p. 260–268. Disponível em: <<https://doi.org/10.5220/0009419902600268>>. Citado 3 vezes nas páginas 37, 110 e 124.
- SILVA, W. et al. Recommendation filtering à la carte for intelligent tutoring systems. In: BORATTO, L. et al. (Ed.). *Bias and Social Aspects in Search and Recommendation*. Cham: Springer International Publishing, 2020. p. 58–65. ISBN 978-3-030-52485-2. Citado 3 vezes nas páginas 33, 107 e 110.
- SILVA, W. et al. Lukewarm starts for computerized adaptive testing based on clustering and irt. In: LANE, H. C.; ZVACEK, S.; UHOMOIBHI, J. (Ed.). *Computer Supported Education*. Cham: Springer International Publishing, 2021. p. 287–301. ISBN 978-3-030-86439-2. Citado na página 33.
- SILVA, W. P. et al. Writing proficiency assessment: Regression analysis of item response theory supported by machine learning techniques. In: *2021 XLVII Latin American Computing Conference (CLEI)*. [S.l.: s.n.], 2021. p. 1–10. Citado 12 vezes nas páginas 16, 19, 33, 40, 113, 115, 116, 118, 119, 120, 121 e 122.
- SINHARAY, S.; JOHNSON, M. S.; STERN, H. S. Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, Sage Publications Sage CA: Thousand Oaks, CA, v. 30, n. 4, p. 298–321, 2006. Citado na página 84.

- SITTHIWORACHART, J.; JOY, M. Effective peer assessment for learning computer programming. In: *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, Leeds, United Kingdom*. New York, NY, USA: Association for Computing Machinery, 2004. (ITiCSE '04), p. 122–126. ISBN 1581138369. Disponível em: <<https://doi.org/10.1145/1007996.1008030>>. Citado na página 69.
- SITTHIWORACHART, J.; JOY, M. Computer support of effective peer assessment in an undergraduate programming class. *J. Comp. Assisted Learning*, v. 24, n. 3, p. 217–231, 2008. Disponível em: <<https://doi.org/10.1111/j.1365-2729.2007.00255.x>>. Citado na página 67.
- SMITH, B. I.; CHIMEDZA, C.; BÜHRMANN, J. H. Random forest missing data imputation methods: Implications for predicting at-risk students. In: ABRAHAM, A. et al. (Ed.). *Intelligent Systems Design and Applications*. Cham: Springer International Publishing, 2021. p. 298–308. ISBN 978-3-030-49342-4. Citado na página 108.
- SORREL, M. A. et al. Adapting cognitive diagnosis computerized adaptive testing item selection rules to traditional item response theory. *PLoS ONE*, v. 15, n. 1, p. 1–17, 2020. ISSN 19326203. Citado na página 107.
- SPALENZA, M. A. et al. Uma Ferramenta para Mineração de Dados Educacionais: Extração de Informação em Ambientes Virtuais de Aprendizagem. In: *Anais do Computer on the Beach*. Florianópolis, Brazil: <https://computeronthebeach.com.br/>, 2018. p. 741–750. Disponível em: <<https://siaiap32.univali.br/seer/index.php/acotb/article/view/12835>>. Citado 3 vezes nas páginas 67, 69 e 85.
- SPALENZA, M. A.; PIROVANI, J. P. C.; OLIVEIRA, E. de. Structures discovering for optimizing external clustering validation metrics. In: ABRAHAM, A. et al. (Ed.). *Intelligent Systems Design and Applications - 19th International Conference on Intelligent Systems Design and Applications (ISDA 2019), Auburn, WA, USA, December 3-5, 2019*. Springer, 2019. (Advances in Intelligent Systems and Computing, v. 1181), p. 150–161. Disponível em: <https://doi.org/10.1007/978-3-030-49342-4__15>. Citado na página 87.
- SPENASSATO, D. et al. Testes Adaptativos Computadorizados Aplicados em Avaliações Educacionais. *Revista Brasileira de Informática na Educação*, v. 24, n. 02, p. 1, 2016. ISSN 1414-5685. Citado 2 vezes nas páginas 40 e 109.
- SPILLER, D. *Assessment Matters: Self-Assessment and Peer Assessment*. Hamilton, New Zealand: The University of Waikato, 2012. Citado na página 68.
- ST-HILAIRE, F. et al. A comparative study of learning outcomes for online learning platforms. In: . [s.n.], 2021. The 2021 conference on Artificial Intelligence in Education, AIED 2021 ; Conference date: 14-06-2021 Through 18-06-2021. Disponível em: <<https://aied2021.science.uu.nl>>. Citado na página 126.
- STEFANI, L. A. Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education*, Taylor & Francis Group, v. 19, n. 1, p. 69–75, 1994. Citado na página 68.
- STRICKER, D.; WEIBEL, D.; WISSMATH, B. Efficient learning using a virtual learning environment in a university class. *Computers & Education*, v. 56, n. 2, p. 495–504, 2011. Disponível em: <<https://doi.org/10.1016/j.compedu.2010.09.012>>. Citado na página 67.

- SYAKUR, M. et al. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, v. 336, p. 012017, 04 2018. Citado na página 114.
- TIAN, X. et al. Modeling frustration trajectories and problem-solving behaviors in adaptive learning environments for introductory computer science. In: _____. [S.l.: s.n.], 2021. p. 355–360. ISBN 978-3-030-78269-6. Citado na página 126.
- TIMMS, M. J. Using item response theory (IRT) to select hints in an ITS. In: LUCKIN, R.; KOEDINGER, K. R.; GREER, J. E. (Ed.). *Artificial Intelligence in Education, Building Technology Rich Learning Contexts That Work, Proceedings of the 13th International Conference on Artificial Intelligence in Education, AIED 2007, July 9-13, 2007, Los Angeles, California, USA*. [S.l.]: IOS Press, 2007. (Frontiers in Artificial Intelligence and Applications, v. 158), p. 213–221. Citado na página 84.
- TOUSIGNANT, M.; MARCHAIS, J. E. D. Accuracy of Student Self-Assessment Ability Compared to Their Own Performance in a Problem-Based Learning Medical Program: A Correlation Study. *Advances in Health Sciences Education*, Kluwer Academic Publishers, v. 7, n. 1, p. 19–27, 2002. Disponível em: <<http://link.springer.com/10.1023/A:1014516206120>>. Citado na página 68.
- TRAUB, R. E. Classical test theory in historical perspective. *Educational Measurement*, NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, v. 16, p. 8–13, 1997. Citado na página 87.
- TSENG, S.; TSAI, C. On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, v. 49, n. 4, p. 1161–1174, 2007. Disponível em: <<https://doi.org/10.1016/j.compedu.2006.01.007>>. Citado na página 67.
- VILLEGAS-CH, W. et al. Big data, the next step in the evolution of educational data analysis. In: ROCHA, Á.; GUARDA, T. (Ed.). *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*. Cham: Springer International Publishing, 2018. p. 138–147. ISBN 978-3-319-73450-7. Citado 2 vezes nas páginas 35 e 67.
- WANG, Y. et al. A motivation model of peer assessment in programming language learning. *CoRR*, abs/1401.6113, p. 12, 2014. Disponível em: <<http://arxiv.org/abs/1401.6113>>. Citado na página 68.
- WAUTERS, K.; DESMET, P.; NOORTGATE, W. V. D. Adaptive item-based learning environments based on the item response theory: possibilities and challenges. *Journal of Computer Assisted Learning*, Wiley Online Library, v. 26, n. 6, p. 549–562, 2010. Citado na página 84.
- WU, M. et al. Variational item response theory: Fast, accurate, and expressive. In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*. [S.l.: s.n.], 2020. p. 257–268. ISSN 23318422. Citado na página 110.
- YEUNG, C. Deep-irt: Make deep learning based knowledge tracing explainable using item response theory. In: DESMARAIS, M. C. et al. (Ed.). *Proceedings of the 12th International Conference on Educational Data Mining, EDM'19, Montréal, Canada, July 2-5, 2019*. [S.l.]: International Educational Data Mining Society, 2019. Citado na página 84.

ZAFFALON, F. et al. Estudo Comparativo entre Modelos que Estimam a Habilidade dos Estudantes em Ambientes Virtuais de Programação. *Revista Brasileira de Informática na Educação*, v. 28, p. 776–795, 2020. ISSN 1414-5685. Citado na página 108.

ZAFFALON, F. et al. Um estudo comparativo entre dois modelos que estimam a habilidade dos estudantes: ELO e Teoria de Resposta ao Item. In: *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019) Um*. [S.l.: s.n.], 2019. p. 459. ISSN 2316-6533. Citado 2 vezes nas páginas 36 e 108.

ZAMPIROLI, F. D. A. et al. Online assessments with parametric questions and automatic corrections: an improvement for MCTest using Google Forms and Sheets. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020)*. [S.l.: s.n.], 2020. p. 51–60. Citado na página 36.

ZHAN, P.; WANG, W. C.; LI, X. A Partial Mastery, Higher-Order Latent Structural Model for Polytomous Attributes in Cognitive Diagnostic Assessments. *Journal of Classification*, Journal of Classification, v. 37, n. 2, p. 328–351, 2020. ISSN 14321343. Citado na página 110.

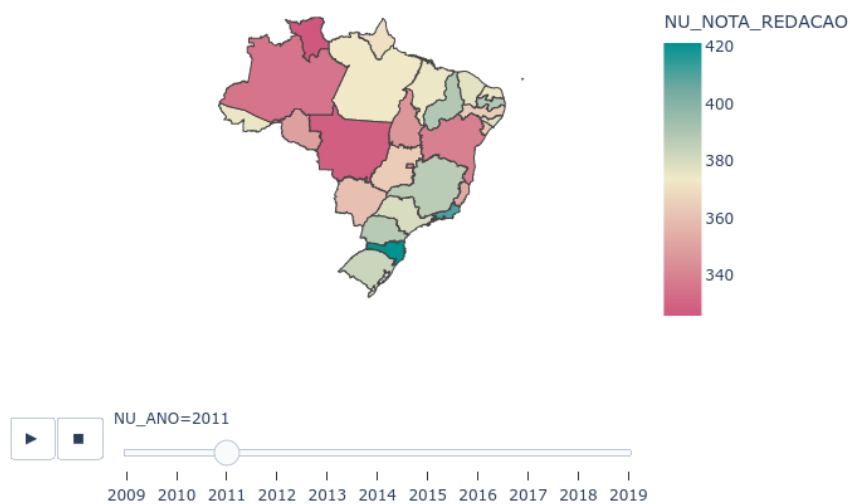
ZUGE, C. B. J. e Camila Andrade de Macedo e Vitor Santos Alencar e Robertino Santiago Junior e A. P. Abordagem metodológica para o ensino de arquitetura de computadores em ambientes não formais. *Revista Brasileira de Informática na Educação*, v. 28, n. 0, p. 335–358, 2020. ISSN 2317-6121. Disponível em: <<http://ojs.sector3.com.br/index.php/rbie/article/view/v28p335>>. Citado na página 61.

Apêndices

Figura 75 – Mapas de Contraste da Redação

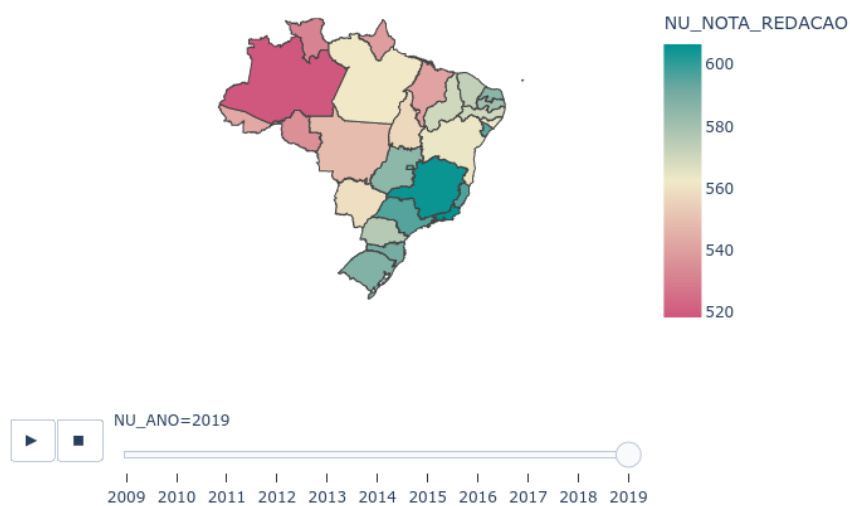
(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas



Enem - Mapeamento Histórico

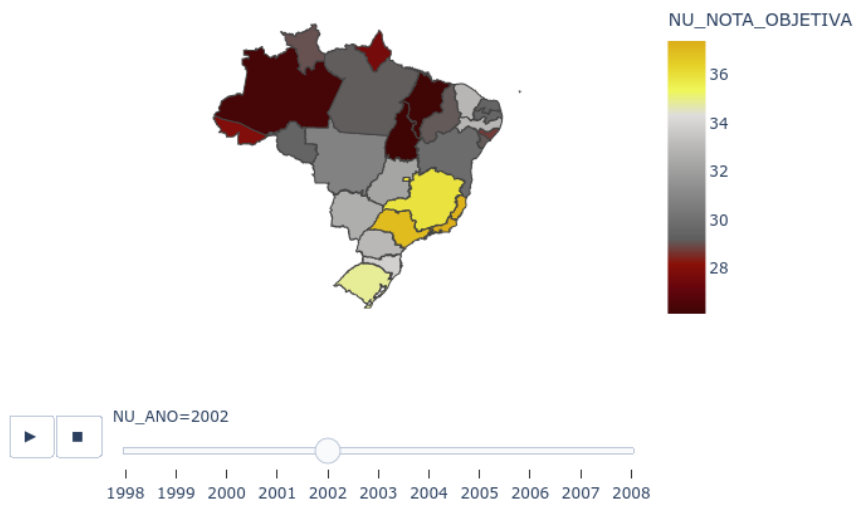
Código Prototipado

O código implementado para simulação do CAT foi construído na linguagem python, no Bloco Textual 7.3.

Figura 76 – Mapas de Contraste das Objetivas Interdisciplinares

(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas

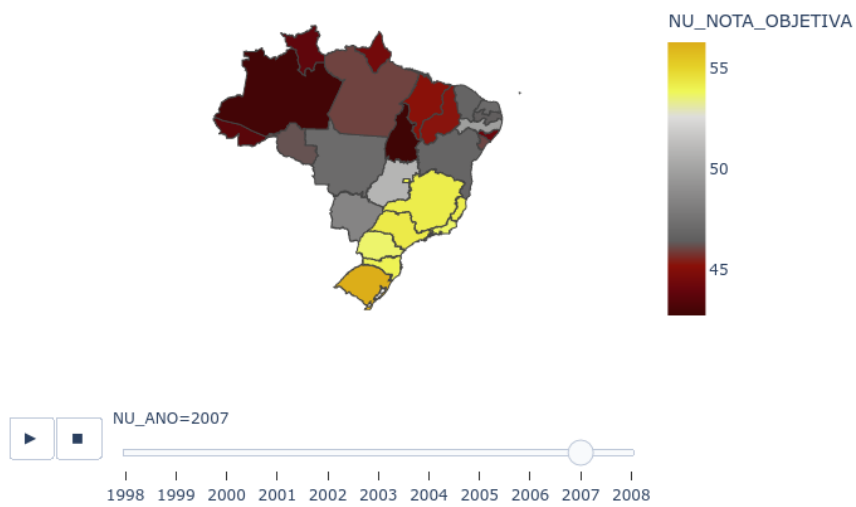
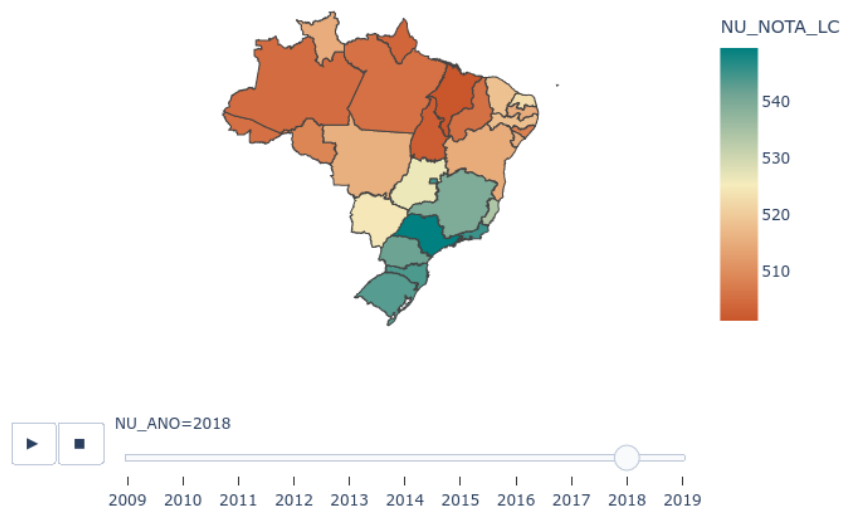


Figura 77 – Mapas de Contraste das Objetivas Interdisciplinares

(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas

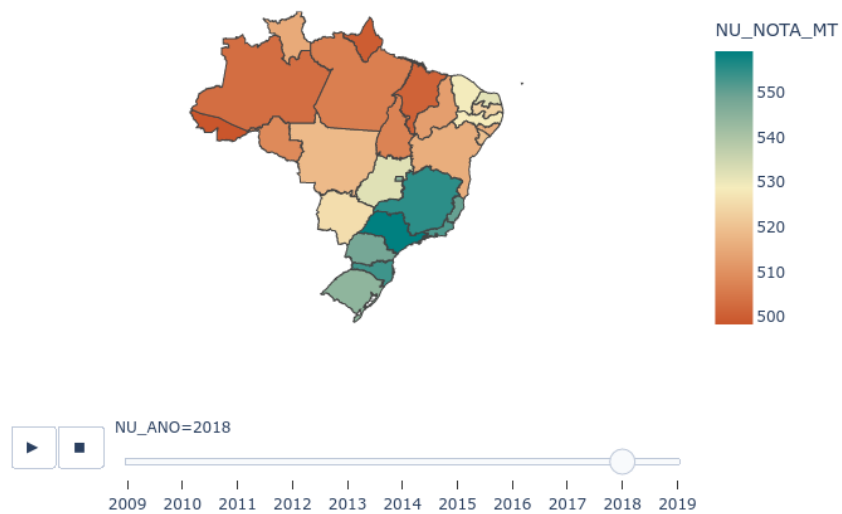
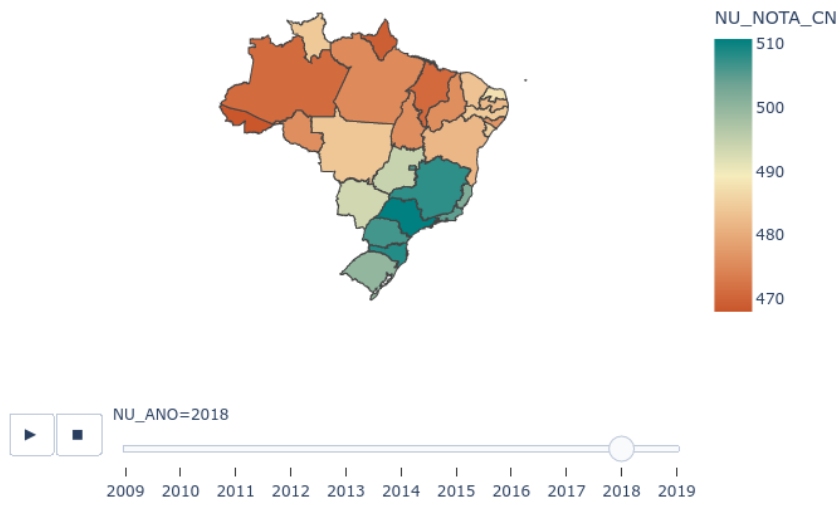


Figura 78 – Mapas de Contraste das Objetivos Interdisciplinares

(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas

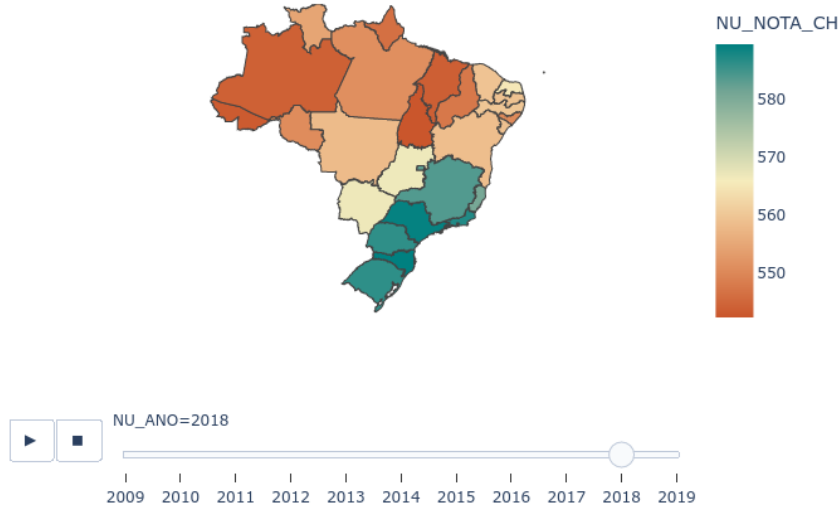
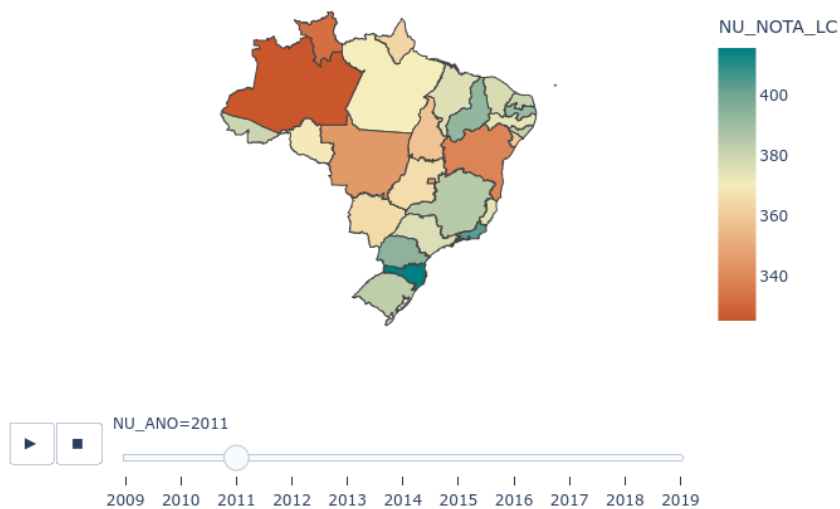


Figura 79 – Mapas de Contraste das Objetivos Interdisciplinares

(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas

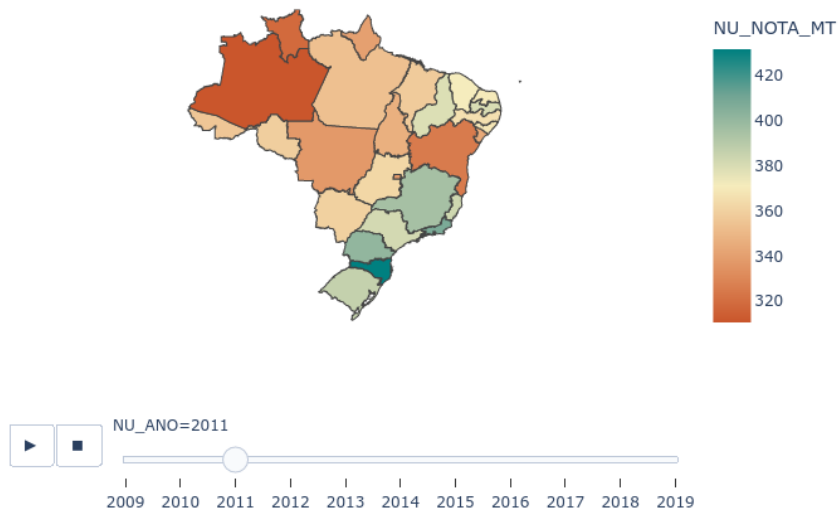
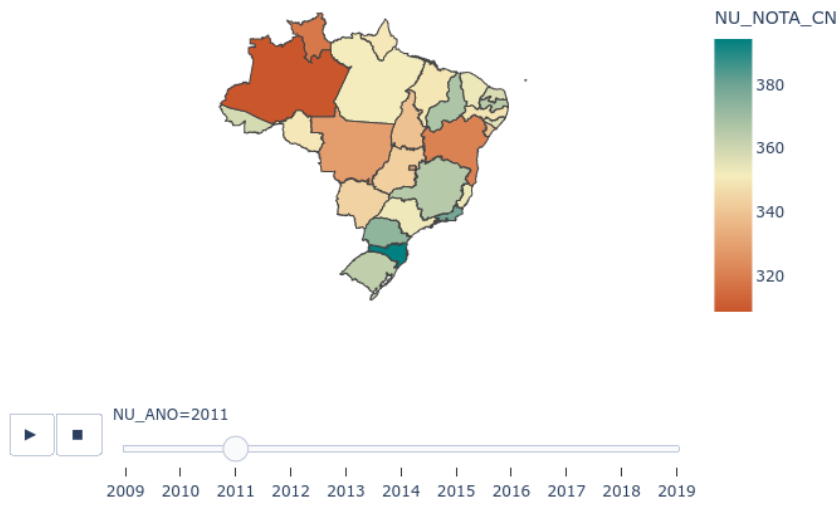


Figura 80 – Mapas de Contraste das Objetivos Interdisciplinares

(a)

Enem - Histórico de Notas



(b)

Enem - Histórico de Notas

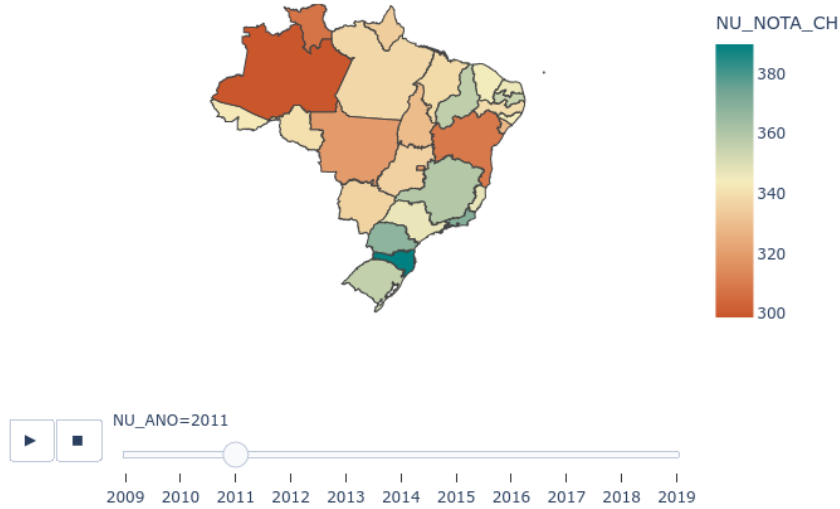
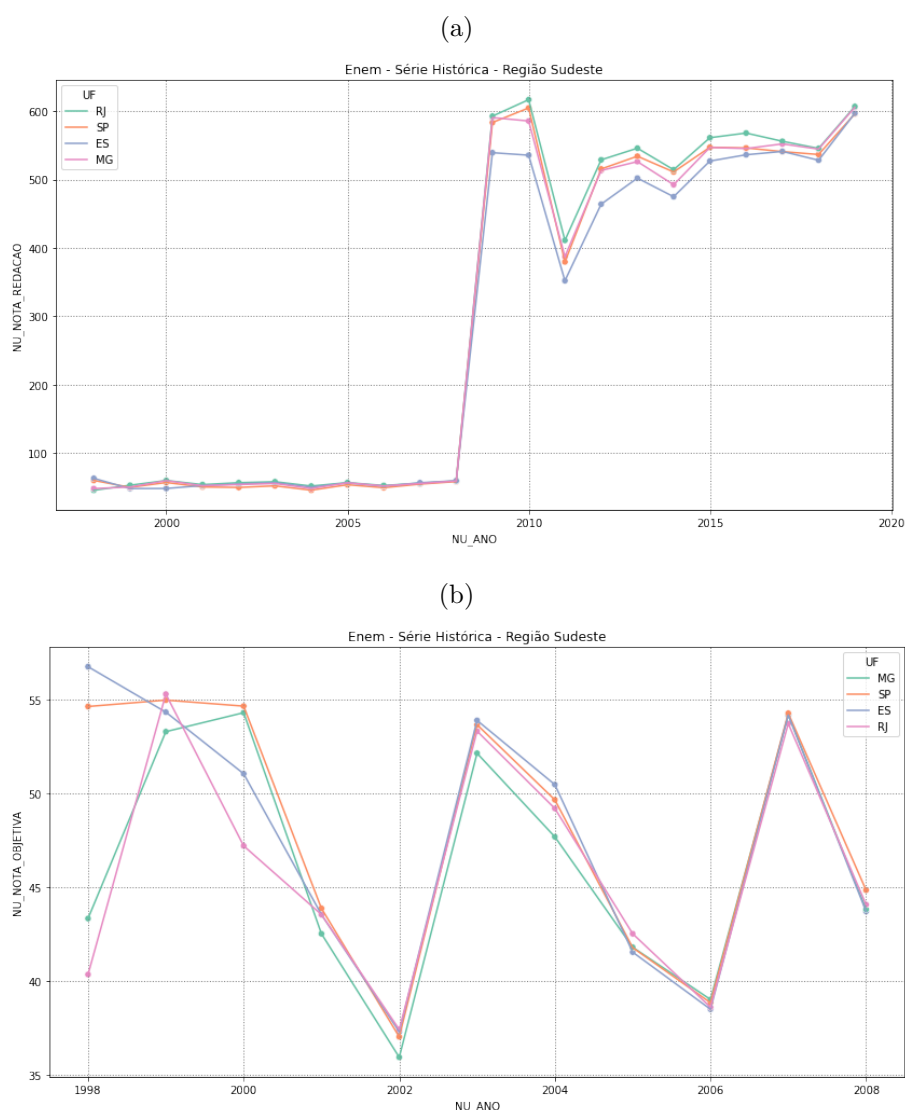


Figura 81 – Região Sudeste - Redação e Provas Interdisciplinares



Listagem 1 – CAT em Python

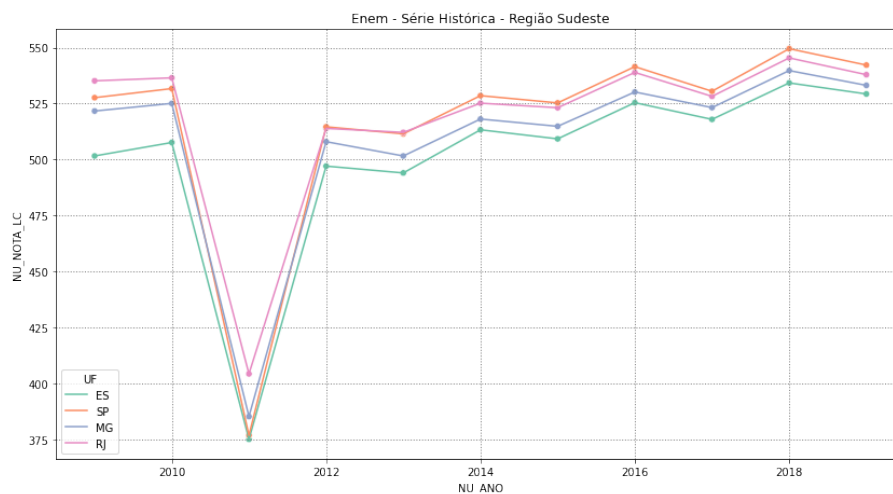
```

1 class cat_classico:
2     def __init__(self, cluster=3, parada = 44, limiar = .6):
3         if cluster != 3:
4             teste_examinando = vetor_resposta.query('Cluster =='+str(cluster))
5             .sample(1)
6         else:
7             teste_examinando = vetor_resposta.sample(1)
8             self.theta_real = teste_examinando['\u03B8'].values[0]
9             self.vetor_real = teste_examinando[range(45)]
10            self.coeficientes = coeficientes
11            self.historico = []
12            self.bd_item = coeficientes.copy()
13            self.theta = [0,0]
14            self.parada = parada

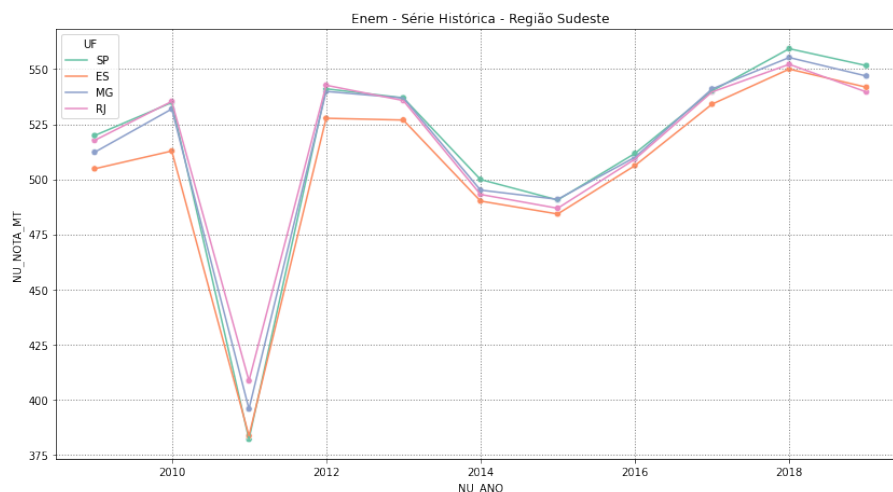
```

Figura 82 – Região Sudeste - Linguagem-Comunicação e Matemática

(a) LC



(b) MT



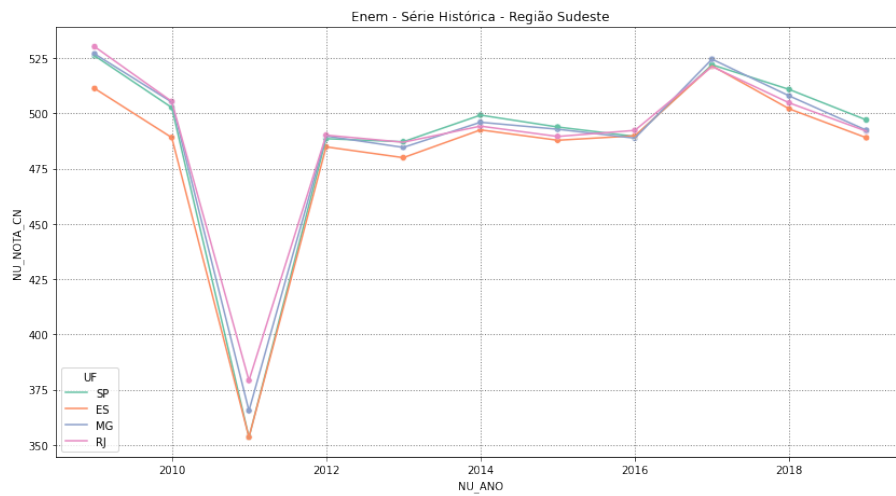
```

14 self.sequencia = []
15 self.limiar_parada = limiar
16
17 def verifica_sequencia (self,):
18     if len(self.sequencia) == 0:
19         valores = self.bd_item
20     else:
21         filtro = []
22         for todos in self.sequencia:
23             filtro.append(todos[0])
24         valores = self.bd_item[~self.bd_item.index.isin(filtro)]
25     _=[]
26     for it, jt in zip(valores.values,
27                     valores.index):
28         a = it[0]
29         b = it[-1]

```

Figura 83 – Região Sudeste - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH

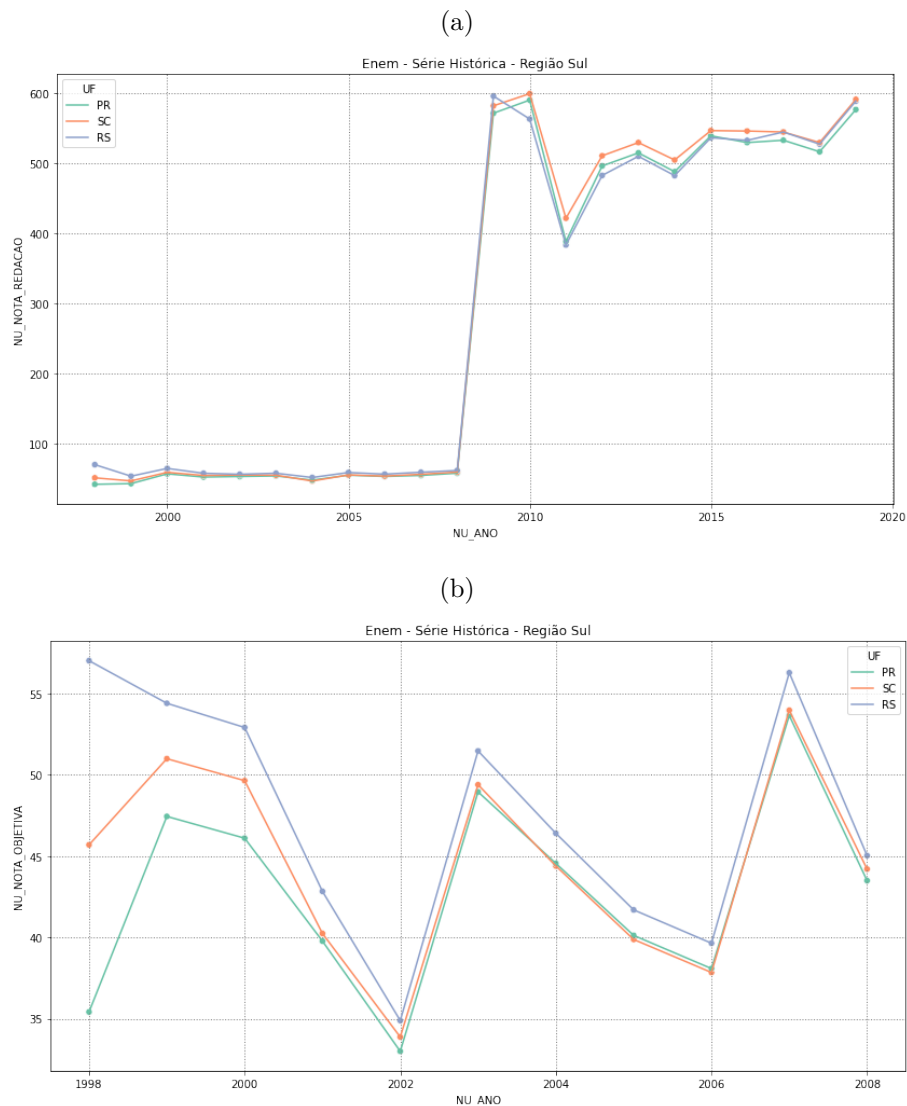


```

30     c = it[2]
31     prob_correct = irt_parameter_estimation.logistic3PLabc(a, b, c, np
    .mean(self.theta))
32     _ .append([jt, prob_correct])
33
34     df_aux = pd.DataFrame(_, columns=['posix', 'prob_correct']).
    sort_values(by='prob_correct',
35                ascending=False)
36
37     self.sequencia.append([df_aux.head(1)['posix'].values[0],
38                            df_aux.head(1)['prob_correct'].values[0]])
39
40 def verifica_sequencia_(self,):
41     if len(self.sequencia)==0:
42         valores = self.bd_item
43     else:

```

Figura 84 – Região Sul - Redação e Interdisciplinares



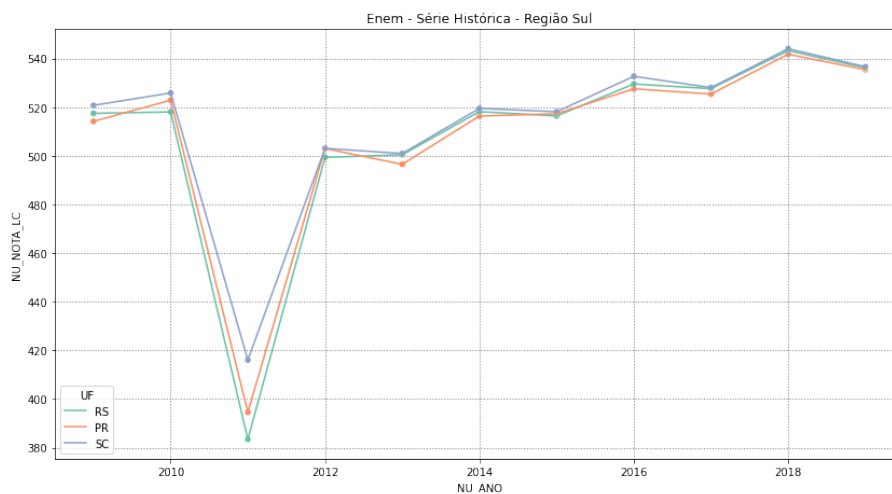
```

44     filtro = []
45     for todos in self.sequencia:
46         filtro.append(todos[0])
47     valores = self.bd_item[~self.bd_item.index.isin(filtro)]
48     _=[]
49     for it, jt in zip(valores.values,
50                     valores.index):
51         a = it[0]
52         b = it[-1]
53         c = it[2]
54         probb_correct = irt_parameter_estimation.logistic3PLabc(a, b, c, np
55     .mean(self.theta))
56         _.append([jt, probb_correct])
57
58     df_aux = pd.DataFrame(_, columns=['posix', 'probb_correct']).
59     sort_values(by='probb_correct',

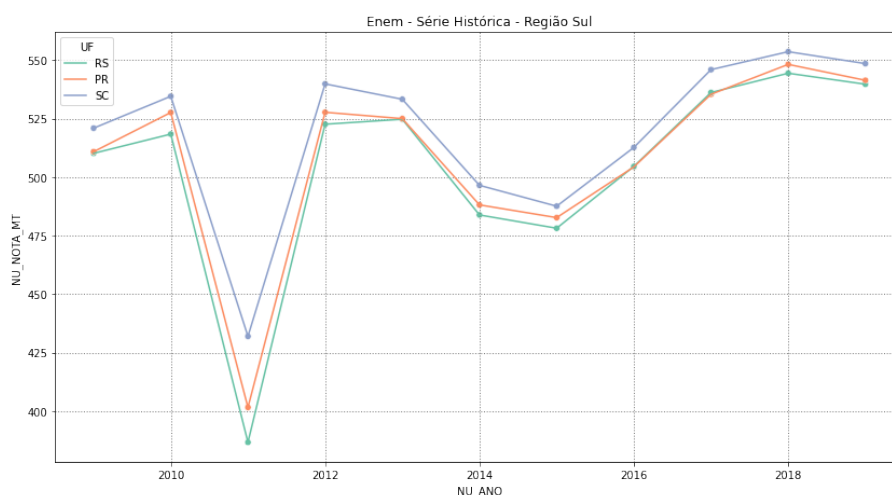
```

Figura 85 – Região Sul - Linguagem-Comunicação e Matemática

(a) LC



(b) MT



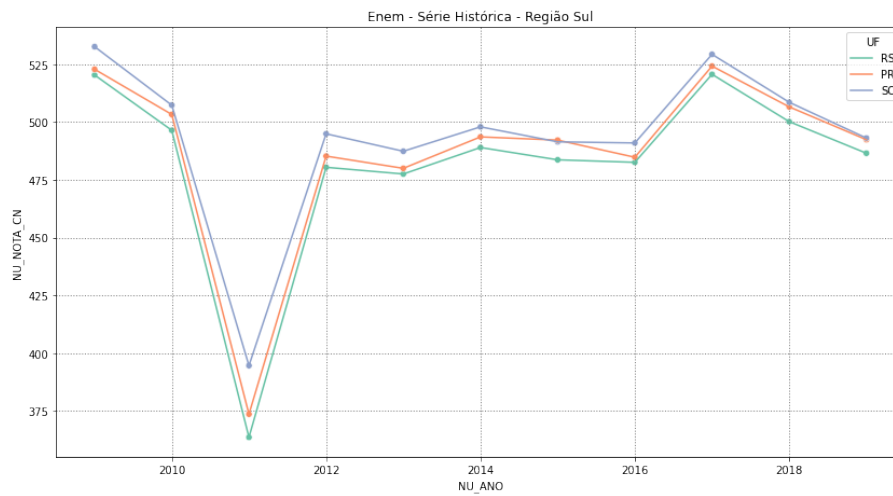
```

58         ascending=False)
59
60     self.sequencia.append([df_aux.head(1)['posix'].values[0],
61                            df_aux.head(1)['prob_correct'].values[0]])
62
63     def select_item(self):
64         if (len(self.theta) < self.parada) and abs(pd.Series(self.theta).std
65             ()) <= self.limiar_parada:
66             self.verifica_sequencia()
67             self.historico.append([
68                 self.sequencia[-1][0],
69                 self.sequencia[-1][1],
70                 self.vetor_real[self.sequencia[-1][0]].
71                 values[0],
72                 -logit(self.sequencia[-1][1]),
73                 ])

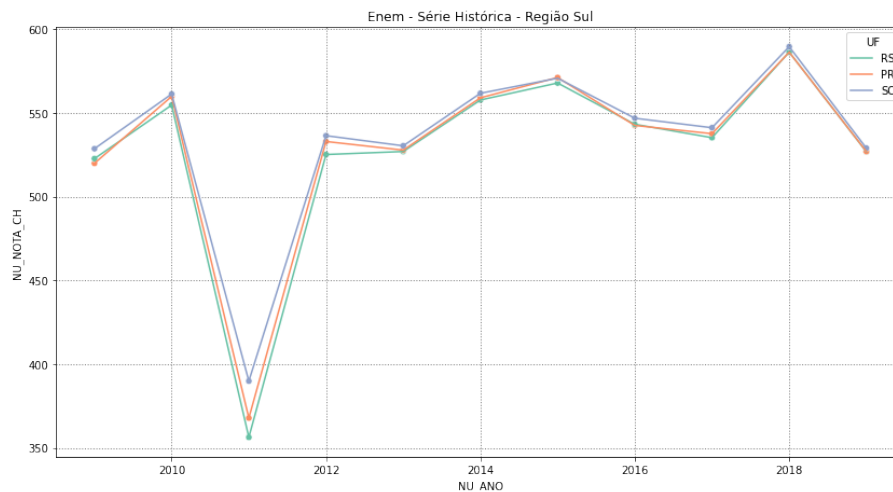
```


Figura 86 – Região Sul - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH



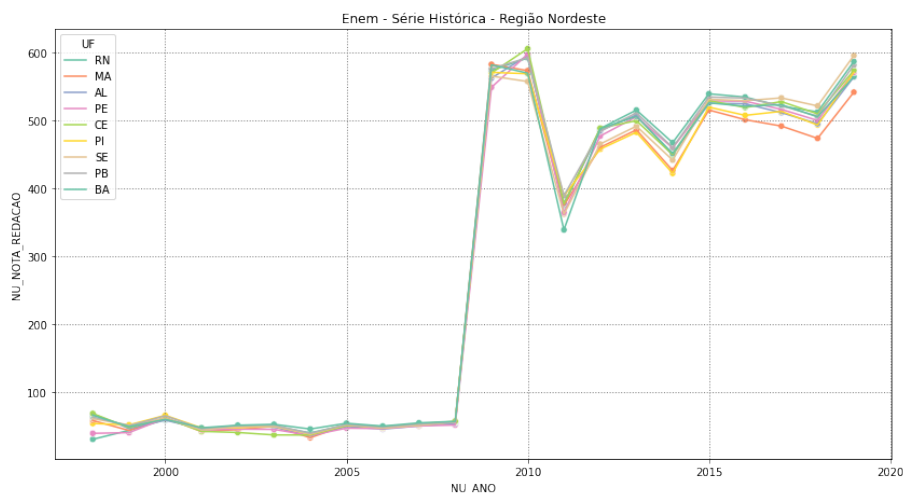
```

72     if self.vetor_real[self.sequencia[-1][0]].values[0] ==1:
73         self.atualiza_theta(tipo="cred")
74     else:
75         self.atualiza_theta(tipo="deb")
76
77     def atualiza_theta(self, tipo="cred" ):
78         if tipo == "cred":
79             self.theta.append(self.historico[-1][3])
80         else:
81             self.theta.append(self.theta[-1] - self.historico[-1][3])
82
83     def run(self, ):
84         for i in range(self.parada):
85             self.select_item()
86
87     # 0 alto, 2 medio, 1 baixo

```

Figura 87 – Região Nordeste - Redação e Interdisciplinares

(a)



(b)



```

88
89 cat = cat_novo(cluster=0,
90                 parada=45,
91                 limiar=.6)
92 cat.run()

```

Listagem 2 – CAT Baseado em Agente em Python

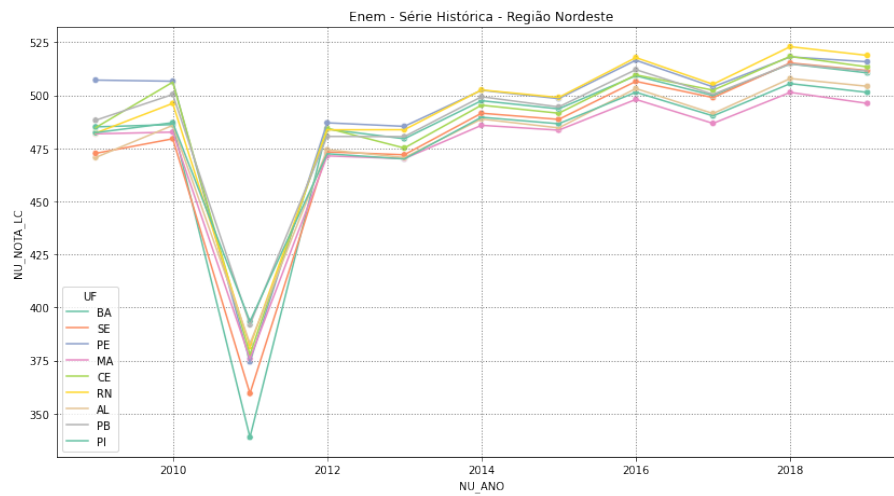
```

1 def run_cat(funcao_agente):
2     cat = cat_novo()
3     try:
4         while not bancoItens.fim_itens(Stop=True):
5             item = funcao_agente(bancoItens)
6             resp = apresenta_item(item)
7             atualiza_lista_itens(resp)
8     except KeyboardInterrupt:

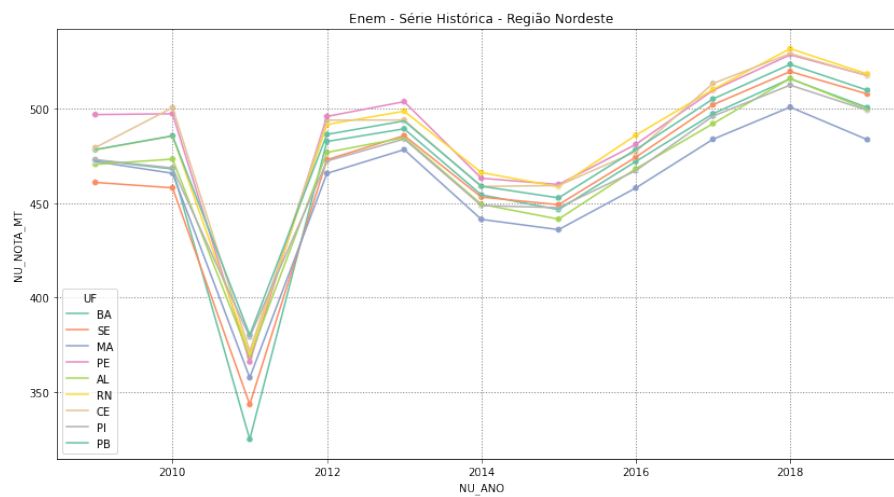
```

Figura 88 – Região Nordeste - Linguagem-Comunicação e Matemática

(a) LC



(b) MT



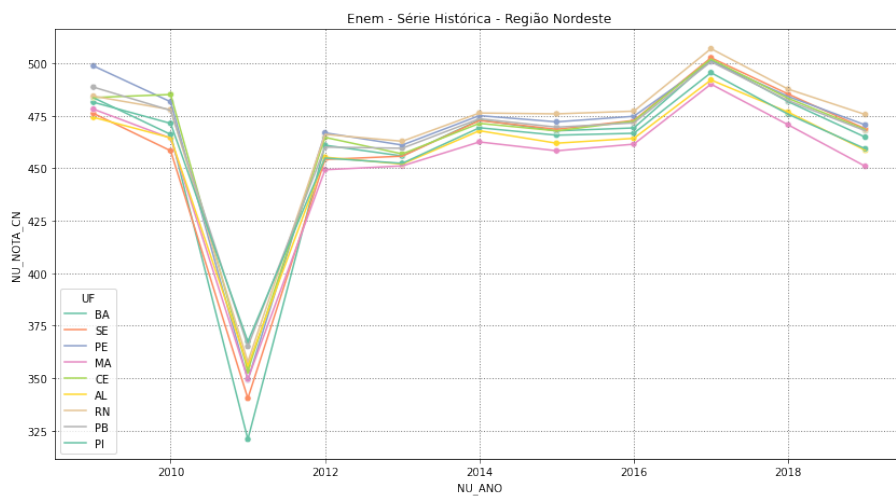
```

9     msg = "CAT Interrompido!"
10    return (None, msg, ItensRespondidos)
11    result = None
12
13    return (result, msg, board)

```

Figura 89 – Região Nordeste - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH

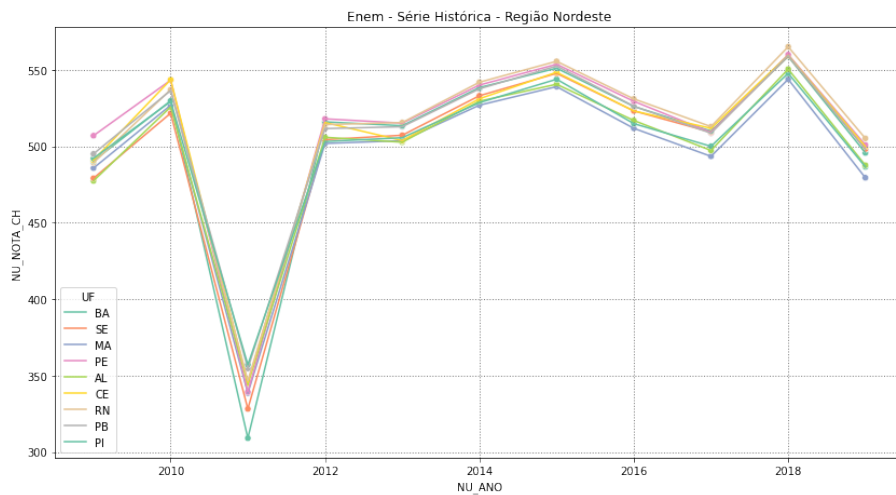
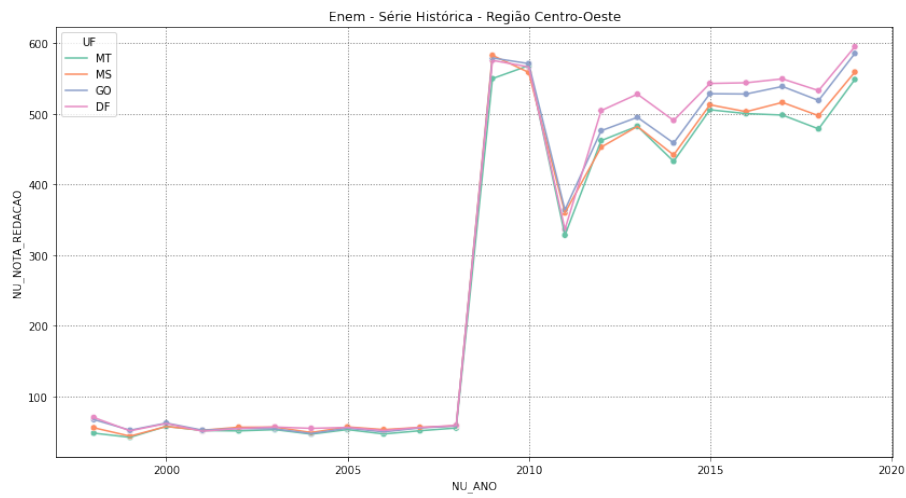


Figura 90 – Região Cento-Oeste - Redação e Interdisciplinares

(a)



(b)

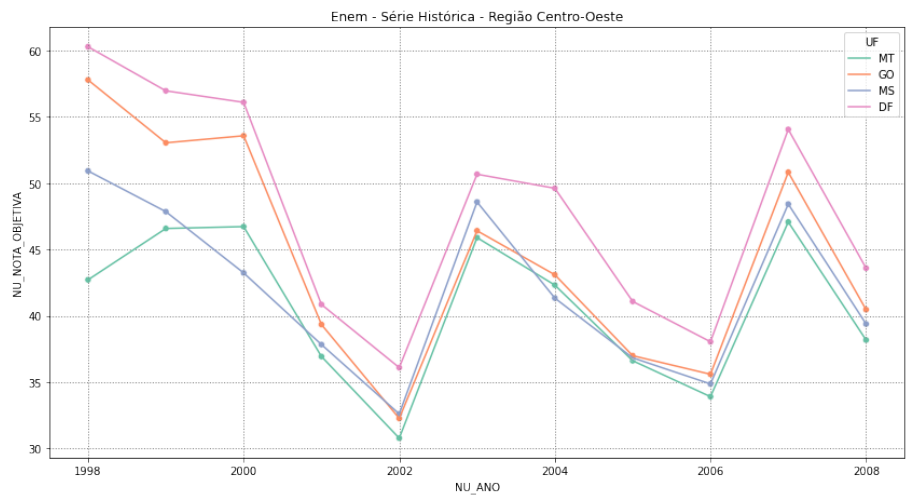
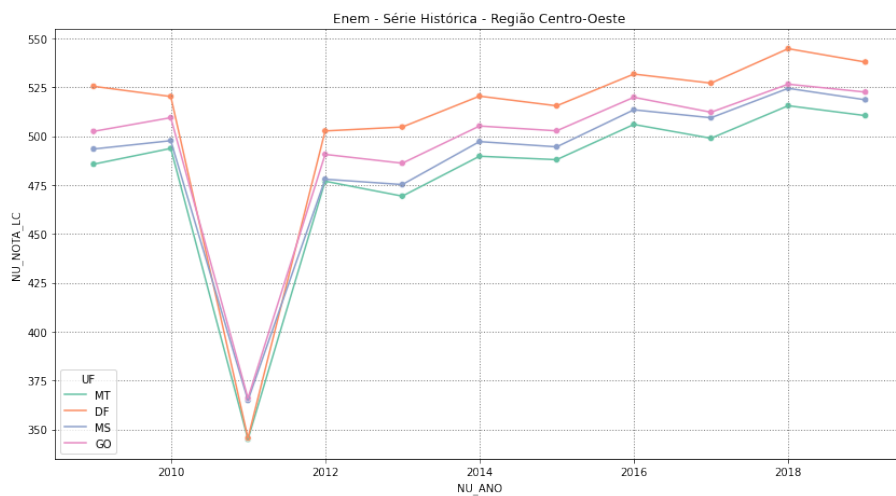


Figura 91 – Região Centro-Oeste - Linguagem-Comunicação e Matemática

(a) LC



(b) MT

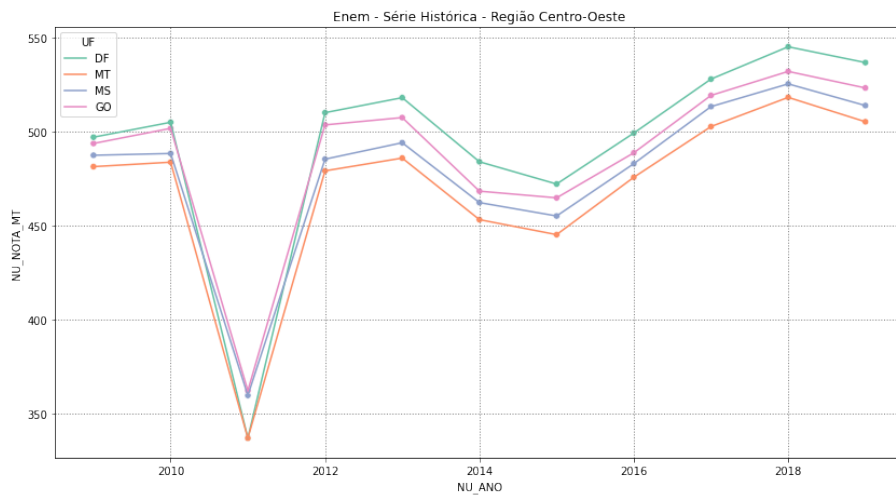
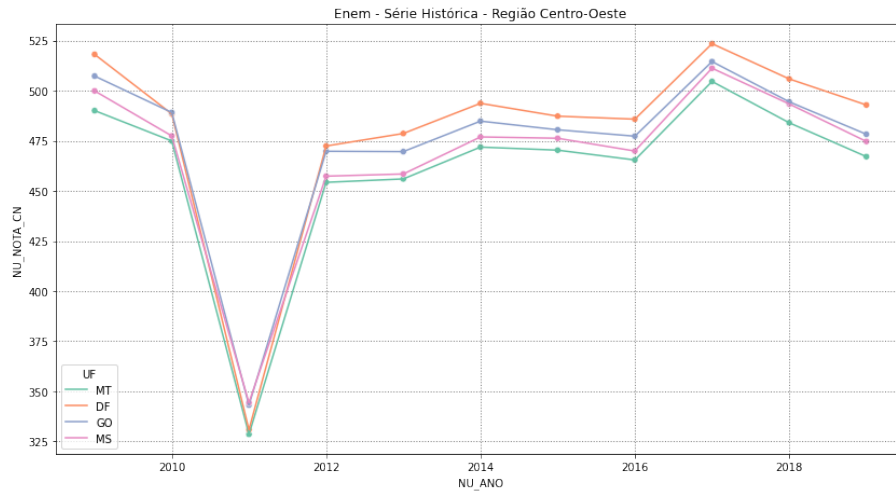


Figura 92 – Região Centro-Oeste - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH

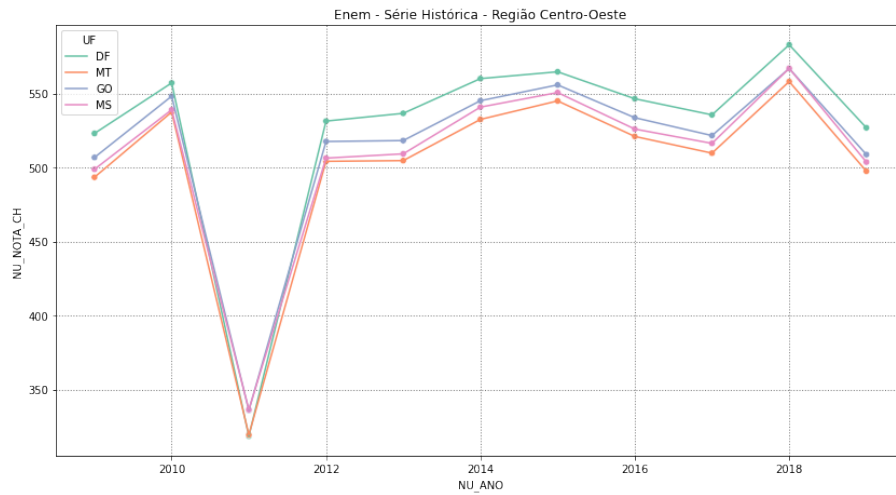
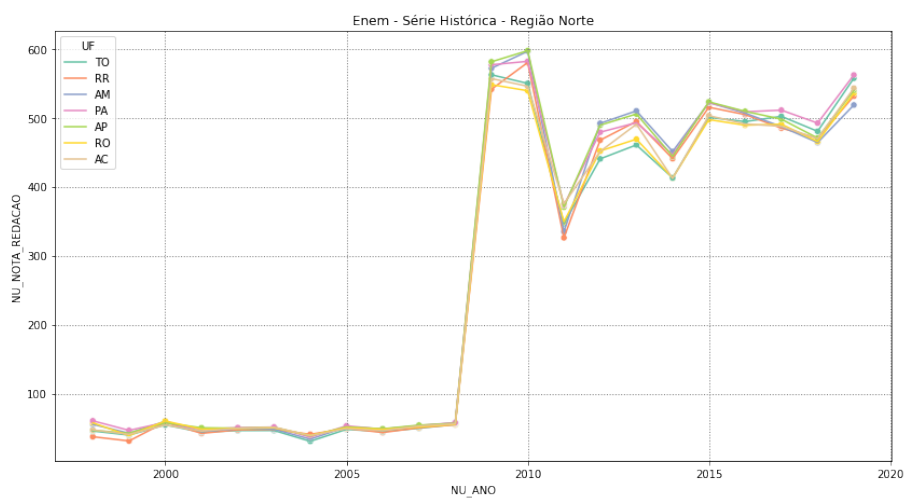


Figura 93 – Região Norte - Redação e Interdisciplinares

(a)

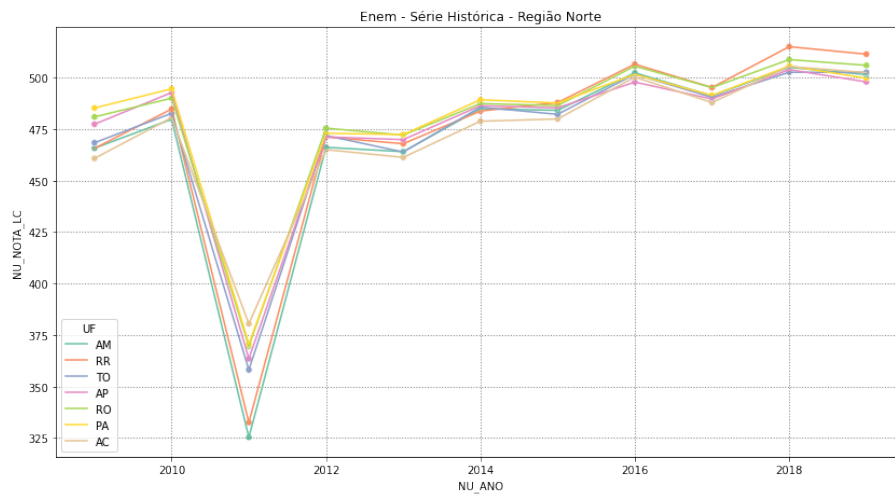


(b)



Figura 94 – Região Norte - Linguagem-Comunicação e Matemática

(a) LC



(b) MT

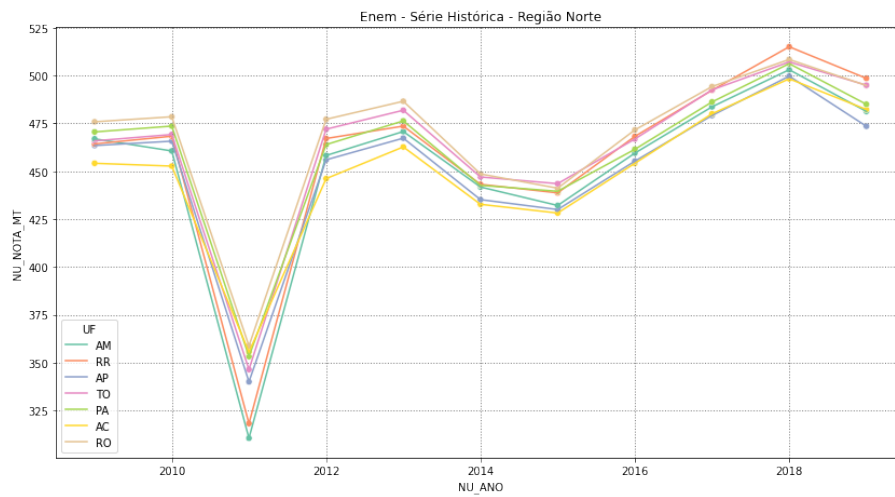
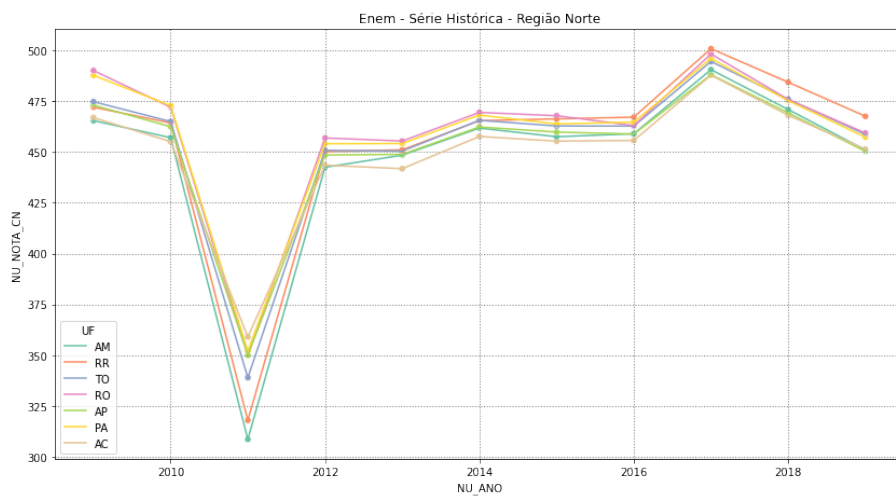


Figura 95 – Região Norte - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH

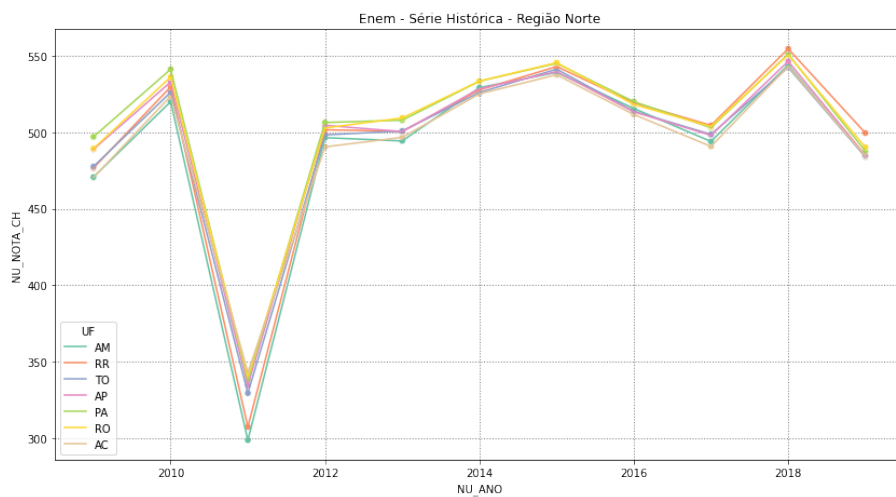
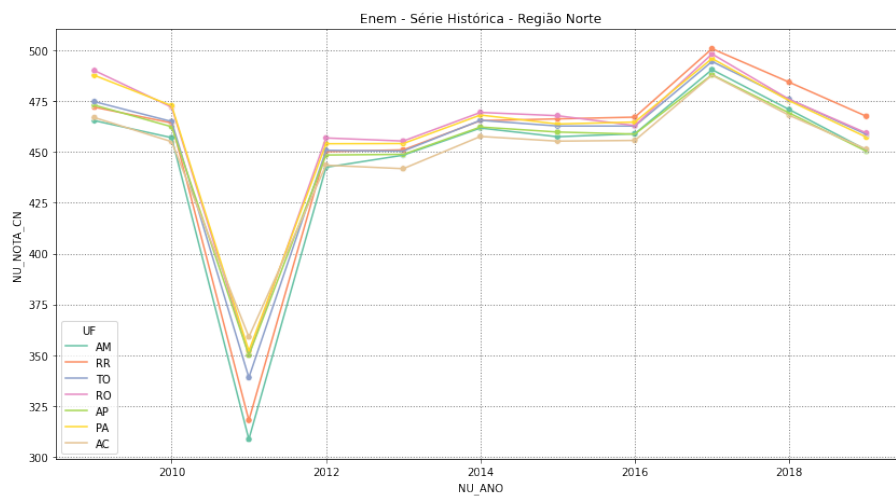


Figura 96 – Região Norte - Ciências Naturais e Ciências Humanas

(a) CN



(b) CH

