

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

BRUNO BASTOS STOLL

**FRAMEWORK PARA ANÁLISE E INTERVENÇÃO
NO PROCESSO DE APRENDIZADO**

Vitória - ES
2019

BRUNO BASTOS STOLL

FRAMEWORK PARA ANÁLISE E INTERVENÇÃO NO PROCESSO DE APRENDIZADO

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Informática, na área de concentração de Informática na Educação.

Orientador: Prof. Dr. Crediné Silva de Menezes

Vitória - ES
2019

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

S875f STOLL, BRUNO, 1984-
FRAMEWORK PARA ANÁLISE E INTERVENÇÃO NO
PROCESSO DE APRENDIZADO / BRUNO STOLL. - 2019.
132 f. : il.

Orientador: Crediné Silva de Menezes.
Coorientador: Davidson Cury.
Dissertação (Mestrado em Informática) - Universidade
Federal do Espírito Santo, Centro Tecnológico.

1. Análise do Aprendizado. 2. Mineração de Dados
Educacionais. 3. Sistemas de Recomendações. 4. Aprendizado de
Máquina. I. Silva de Menezes, Crediné. II. Cury, Davidson.
III. Universidade Federal do Espírito Santo. Centro Tecnológico.
IV. Título.

CDU: 004



FRAMEWORK PARA ANÁLISE E INTERVENÇÃO NO PROCESSO DE APRENDIZADO

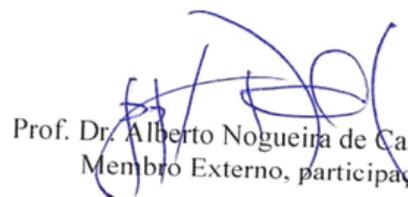
Bruno Bastos Stoll

Dissertação submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Mestre em Informática.

Aprovada em 19 de julho de 2019:


Prof. Dr. Credine Silva de Menezes
Orientador(a)


Prof. Dr. Davidson Cury
Membro Interno


Prof. Dr. Alberto Nogueira de Castro Junior
Membro Externo, participação remota

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória-ES, 19 de julho de 2019.

AGRADECIMENTOS

Gratidão é algo importante em todos os passos de nossas vidas. Em poucos momentos na vida conseguimos expressar nossa gratidão em nossas vitórias.

Em primeiro lugar gostaria de agradecer a Deus, por sempre estar em Tudo que faço e o quanto me abençoou em todas as áreas de minha vida.

Gostaria de agradecer à minha esposa, minha melhor amiga e companheira, por ser uma pessoa sensacional e esteve ao meu lado em todos os momentos me compreendendo.

À minha irmã e meu irmão, gostaria de agradecer por fazer parte de minha vida, sempre me apoiarem e estarmos juntos em todos os momentos importantes.

Também gostaria de agradecer à minha mãe Amélia, pois me ensinou todos os valores que tenho e tudo de bom que sou e tenho devo a ela.

Aos meus filhos, gostaria agradecer, pois sempre dedicarei todo meu amor por eles. E que, só por existirem, já me dá razão para viver e fazer as coisas certas, lhes servindo de exemplo.

Por fim, gostaria de agradecer ao Crediné e ao Dede, por me darem a oportunidade de aprender com vocês. Esta pesquisa mudou minha vida!

Muito obrigado a todos vocês!

RESUMO

O uso de tecnologias como a internet permite que produções e interações sejam armazenadas. Essas informações, quando analisadas, são fonte de conhecimento profundo sobre os usuários, suas preferências e necessidades. Assim, tal análise pode proporcionar novas perspectivas em várias áreas, tais como indústria, meio ambiente, saúde e educação. Quando aplicada na área da educação, a extração de conhecimento proporciona um entendimento de fatores dentro do processo educacional. Apesar de haver várias pesquisas nessa área, ainda existem grandes desafios. Um deles está relacionado à dificuldade em realizar uma análise detalhada de tais dados. Para o professor, analisar e monitorar os dados coletados das interações dos alunos é uma importante e complexa tarefa. Pois o volume de dados pode ser muito grande e, entre a coleta e a análise, geralmente há pouco tempo. Dentro desse contexto, delimitamos o objetivo na produção de uma solução computacional para análise de dados acadêmicos, acoplável a qualquer ambiente de aprendizado que colete dados digitalmente, na finalidade de apoiar professores na análise dessas informações. A ferramenta de análise de dados foi proposta para que seja flexível para ser aplicado em qualquer contexto, com os módulos de coleta, previsão, recomendações e visualização. Para validar a proposta, um protótipo foi implementado como parte da solução com os requisitos necessários para aplicação da prova de conceito. Os testes foram realizados com um conjunto público de dados. E os resultados indicam que o modelo conceitual proposto apresenta uma possibilidade de melhoria na análise de dados e intervenção no processo de aprendizado.

Palavras-chave: Análise do Aprendizado, Mineração de Dados Educacionais, Sistemas de Recomendações, Aprendizado de Máquina.

ABSTRACT

The use of technologies such as the Internet allows productions and interactions to be stored. This information, when analyzed, is a source of knowledge about users, their preferences and needs. Thus, such analysis is the new topics in several areas, such as industry, and the environment, health and education. When this is an area of education, an extraction of knowledge provides an understanding of various types of educational processes. Although there are several researches in this area, they were still great challenges. One is related to the difficulty in performing a detailed analysis of such data. For the teacher, analyzing and tracking the data is an important and complex task. Because the data volume can be very large and, between a collection and an analysis, it contains a small amount of time. This context, we delimit the production of computed computational analysis with analysis of data academic, connector for any client of knowledge computed data digital, with an data processing in data of data. The analysis data tool has been proposed to be flexible to be applied in any context, with the collection, forecast, recommendation and visualization module. In the month, external software is expected to be available for external tools. To validate the proposal, the prototype was implemented as part of the solution with the requirements necessary for the use of proof of concept. The testes were performed with a set of public data. And the results obtained with the conceptual model presents a possibility of improvement in data analysis and intervention in the learning process.

Keywords: Learning Analytics, Educational Data Mining, Recommendation Systems, Machine Learning.

LISTA DE FIGURAS

Figura 1.1 - Objetivo da arquitetura acoplável.....	19
Figura 2.1 - Processo LA.....	23
Figura 2.2 - Dimensões do LA.....	25
Figura 2.3 - Processo KDD.....	26
Figura 2.4 - Combinação das áreas da aplicação do EDM	27
Figura 2.5 - Árvore de decisão simples	29
Figura 2.6 - Distância dos vizinhos próximos	30
Figura 2.7 - Representação das distâncias do Hiperplano de um SVM	31
Figura 2.8 - Contornos de densidade (dois atributos)	32
Figura 2.9 - Organização das camadas de uma Rede Neural.....	33
Figura 2.10 - Papel desempenhado pelos neurônios das diferentes camadas de uma rede MLP	34
Figura 2.11 - Matriz de confusão.....	36
Figura 4.1 - Fases da Análise dos dados	47
Figura 4.2 - Modelo Conceitual do Sistema	48
Figura 4.3 - Arquitetura dos módulos de Previsão e Recomendação	52
Figura 4.4 - Modelo usado pelos módulos de previsão e recomendação	53
Figura 4.5 - Geração de modelos preditivos, treinamento e testes	54
Figura 5.1 - Atores / Perfil.....	59
Figura 5.2 - Diagrama de Casos de Uso	61
Figura 5.3 - Diagrama de Classes.....	63
Figura 5.4 - Diagrama de Componentes	65
Figura 5.5 - Metadados	66
Figura 5.6 - Caracterização das telas - perfil Analista	67
Figura 5.7 - Caracterização das telas - perfil Professor	68
Figura 5.8 - Caracterização das telas - perfil aluno	69
Figura 6.1 - Login do Sistema	72
Figura 6.2 - Analista	73
Figura 6.3 - Professor.....	73

Figura 6.4 - Aluno.....	73
Figura 6.5 - Definição da fonte de dados	74
Figura 6.6 - Selecionar tabela para importar	75
Figura 6.7 - Processo das Tarefas do PDI	76
Figura 6.8 - Fonte / Metadados / Limpa	76
Figura 6.9 - Fluxo de extração e tratamento Pentaho	78
Figura 6.10 - Gerar Caso	78
Figura 6.11 - Formulário de alteração de Metadados.....	81
Figura 6.12 - Formulário de Discretização	82
Figura 6.13 - Processo de discretização pelo módulo de Processamento	83
Figura 6.14 - Acurácia dos algoritmos preditivos.....	85
Figura 6.15 - Pesos das características	86
Figura 6.16 - Mensagem: (A) sugerida ao professor e (B) recebida pelo aluno	87
Figura 6.17 - Formulário de configuração de Visão.....	88
Figura 6.18 - Exibição da Visualização Previsão de desempenho por Disciplina	88
Figura 6.19 - Painel gerado com o Tableau	89
Figura 6.20 - Painel gerado com os recursos do sistema	90
Figura 6.21 - Configuração da Linha	91
Figura 6.22 - Associar visão a uma célula.....	91
Figura 6.23 - Formulário com o Painel do Professor associado	92
Figura 6.24 - Tela de configuração da Recomendação.....	92
Figura 6.25 - Formulário da Recomendação	93
Figura 6.26 - Recomendação recebida pelo Aluno	94
Figura 6.27 - Mensagem recebida pelo aluno usando o celular	94

LISTA DE TABELAS

Tabela 3.1 - Comparação entre as contribuições.....	44
Tabela 5.1 - Atores / Perfil.....	59
Tabela 5.2 - Casos de Usos.....	59
Tabela 5.3 - Classes mapeadas.....	62
Tabela 5.4 - Elementos do diagrama de componentes.....	64
Tabela 6.1 - Perfil de Usuários.....	71
Tabela 6.2 - Arquivos / Tabelas do OULA.....	75

LISTA DE SIGLAS

AVA - *Ambiente Virtual de Aprendizado*

DM - *Data Mining*

DL - *Deep Learning*

EAD - *Ensino à Distância*

EDM - *Educational Data Mining*

KDD - *Knowledge Discovery in Databases*

KNN - *K-Nearest Neighbors*

LA - *Learning Analytics*

ML - *Machine Learning*

MLP - *Multi-Layer Perceptron*

SR - *Sistemas de Recomendações*

SV - *Sistemas de Visualização*

SVM - *Support Vector Machine*

SUMÁRIO

1. INTRODUÇÃO	16
1.1. Contexto	16
1.2. Motivação e Justificativa.....	17
1.3. Objetivos	18
1.4. Contribuições	20
1.5. Questão da Pesquisa	20
1.6. Metodologia.....	21
1.7. Organização do Trabalho.....	21
2. FUNDAMENTAÇÃO TEÓRICA.....	22
2.1. Análise do Aprendizado.....	22
2.2. Mineração de Dados	25
2.3. Mineração de Dados Educacionais	26
2.4. Aprendizado de Máquina.....	28
2.4.1. Aprendizado Supervisionado.....	28
2.4.2. Aprendizado Não Supervisionado	28
2.4.3. Árvore de Decisão	28
2.4.4. K Vizinhos Próximos.....	30
2.4.5. Máquina de Suporte a Vetor.....	31
2.4.6. Naive Bayes	32
2.4.7. Redes Neurais.....	32
2.4.8. Perceptron em Multicamadas.....	33
2.4.9. Métricas sobre Análise preditiva.....	35
2.5. Sistemas de Recomendações.....	37
2.6. Sistemas de Visualização.....	38
3. TRABALHOS CORRELATOS.....	40
3.1. Identificar evasão de alunos.....	40
3.2. Previsão de desempenho	40
3.3. Visualização de dados.....	41
3.4. Ferramentas de análise de dados	41

3.5. Recomendações	42
3.6. Comparação entre as propostas	43
4. DESCRIÇÃO DA SOLUÇÃO.....	47
4.1. Coleta e Tratamento	49
4.1.1. Coleta de dados	49
4.1.2. Processamento de dados.....	50
4.1.3. Transformação	51
4.1.4. Reprocessamento	51
4.2. Previsão e Recomendação	52
4.2.1. Predição	53
4.2.2. Recomendação	55
4.2.3. Mensagens candidatas para avaliação dos Professores.....	56
4.2.4. Recomendações aos Alunos.....	57
4.3. Apresentação	57
4.3.1. Montar Visões e Painéis.....	57
5. PROVA DE CONCEITO.....	58
5.1. Casos de Uso.....	58
5.2. Classes.....	62
5.3. Diagrama de Componentes.....	64
5.4. Arquitetura Acoplável	66
5.5. Caracterização de Interfaces.....	66
5.6. Recursos do Protótipo	69
6. APLICAÇÃO DA PROVA DE CONCEITOS.....	71
6.1. Descrição da prova de conceito	71
6.2. Descrição do Dataset OULA	73
6.3. Carga e Tratamento	74
6.3.1. Importação dos Dados	74
6.3.2. Usando ferramentas externas para Integrar de Dados	76
6.3.3. Gerenciamento dos Metadados	79
6.3.4. Processamento	82
6.3.5. Alterações no banco de dados	83
6.4. Análise dos Dados	85

6.4.1. Previsão	85
6.4.1. Geração das mensagens de Recomendação	86
6.5. Visualização	87
6.5.1 Criar Visões.....	87
6.5.2. Painéis.....	89
6.5. Recomendações	92
6.6. Análise dos Resultados	95
7. CONCLUSÃO	96
7.1. Trabalhos futuros	96
7.2. Produções científicas	97
REFERÊNCIAS.....	98
APÊNDICE A - PROJETO DE SOFTWARE USANDO UML	106
APÊNDICE B - DESCRIÇÃO DO DATASET	129

1. Introdução

Este capítulo apresenta uma visão geral desta dissertação. São explicitados os argumentos que justificam o desenvolvimento do trabalho, o contexto e motivação, os objetivos, o método de pesquisa e a organização e estruturação da dissertação.

1.1. Contexto

Com o aumento do uso de tecnologias digitais, todas as produções e interações são sendo armazenadas. Quando aplicadas conjuntamente com a internet, torna-se possível o armazenamento e compartilhamento em alta escala de dados. E quando esses dados são analisados, poderão ser fonte de conhecimento das preferências e necessidades de usuários.

No contexto educacional, a análise sobre dados com o objetivo de extração de conhecimento proporciona um melhor entendimento e um profundo entendimento sobre fatores que influenciam o processo de aprendizado. O uso de recursos tecnológicos, como ambientes virtuais de aprendizado (AVA), favorece essa análise, pois os dados são coletas de forma automática, registrando todas interações realizadas entre os alunos e o ambiente. A maioria desses ambientes não dispõe de uma forma detalhada para analisar esses dados e intervenção no processo de aprendizado.

Para essa análise de dados existe um área de pesquisa chamada de Análise do Aprendizado, do inglês *Learning Analytics* (LA), que é o uso de dados e modelos para prever o progresso e o desempenho do aluno, e a capacidade de agir com base nessas informações (BROWN, 2012). O conceito LA é descrito em detalhes no capítulo 2. Fundamentação Teórica. Contudo, a maioria das pesquisas nessa área não combina os conceitos de coleta, previsão de desempenho, recomendações (intervenção no processo de aprendizado) e visualização. Dessa forma, faz-se necessário a criação de uma solução computacional, na finalidade de apoiar professores nas atividades.

1.2. Motivação e Justificativa

O uso de tecnologias da informação em cursos online é uma das grandes evoluções, pois favorece alunos que estudam quando e onde quiserem. No entanto, isso significa que a mediação pedagógica impõe condições e limitam possibilidades, surgindo novos desafios educacionais (CENSO, 2016). Um desses desafios é analisar o desempenho dos alunos e buscar desenvolver metodologias que favoreçam o aprendizado e aumentem o engajamento e a motivação, reduzindo o insucesso acadêmico.

Identificar fatores que influenciam seu aprendizado, prever resultados, visualizar dados e gerar recomendações é o estado da arte na análise de dados acadêmicos. Contudo, para que isso ocorra, as interações dos alunos devem ser periodicamente analisadas pelos professores através dos dados coletados, que por vezes, se torna uma tarefa impossível. Já que o volume de dados poderá ser muito grande.

Em AVA, todo o rastro de navegação é coletado automaticamente pelo próprio sistema através da interação de alunos com o seu conteúdo. Dessa forma, gerando grandes bases de dados com informações sobre o comportamento dos estudantes (BAKER, 2010). Analisar e monitorar esse volume de interações entre professor, alunos e conteúdo é uma importante e complexa tarefa do professor no processo de aprendizado online (MEDEIROS et al., 2013). Por meio do uso de análise desses dados é possível apontar os diferentes fatores que influenciam o comportamento do aluno e identificar aspectos sutis, muitas vezes imperceptíveis por professores (VERBERT *et al.*, 2013). Considerando que todos os alunos têm capacidades e ritmos de aprendizado diferentes e isso deverá ser levando em consideração pelos professores, que deverão direcionar o processo de aprendizado a cada um deles de forma personalizada. No entanto, professores têm dificuldades em identificar necessidades e individualidades que alunos estão tendo (FERGUSON, 2012). Essa intervenção poderá acontecer através do fornecimento de recomendações e *feedbacks* individualizados através de mensagens.

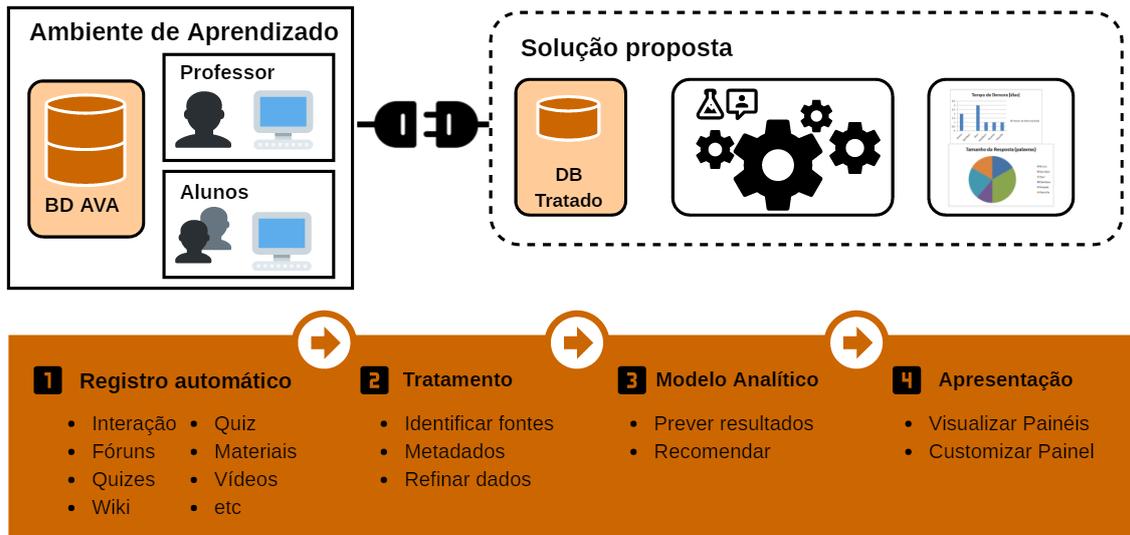
Vários trabalhos de pesquisa de informática aplicada à educação propõem diversas soluções computacionais para as áreas de previsão de desempenho, recomendações e visualização. Contudo, não há uma solução unificada que realize essas atividades e que seja acoplável a qualquer ambiente de aprendizado. Diante desse cenário, esta dissertação propõe um *framework* acoplável a qualquer ambiente de aprendizado e que realize as atividades de previsão de desempenho, geração de recomendações e visualização de dados através de coleta e tratamento de dados acadêmicos, a fim de facilitar professores nessas atividades de análise de dados e intervenção no processo de aprendizado.

Um *framework* é, em geral conceituado, como uma estrutura real ou conceitual destinada a servir de suporte ou guia para a construção de algo útil (PREE, 1994). O conceito de *framework* computacional para este trabalho é entendido como um modelo conceitual de elementos extensível de ferramentas de software caracterizados, que podem ser utilizadas para implementar uma solução computacional.

1.3. Objetivos

O objetivo geral desta dissertação é demonstrar que com o suporte computacional acoplável é possível apoiar professores nas atividades de identificar alunos em risco, gerar recomendações para eles e acompanhar o andamento da turma através de painéis de visualização de dados customizáveis. Dessa forma, é possível melhorar o processo de ensino. A solução computacional é direcionada para ambientes de aprendizado, no qual a coleta é favorecida por ambientes virtuais (AVA) porque os dados são coletados automaticamente, mas não se limitam só a esses ambientes.

Figura 1.1 - Objetivo da arquitetura acoplável



Fonte: próprio autor.

Conforme apresentado na Figura 1.1 - Objetivo da arquitetura acoplável, a solução computacional corresponde a um framework que, uma vez acoplado ao ambiente de aprendizado e devidamente configurado, será provedor de informações a professores lhes dando a oportunidade de intervenção no processo de aprendizado. O framework contém os módulos de coleta e tratamento, previsão de desempenho, recomendações e apresentação. Os objetivos específicos relacionados à pesquisa desta dissertação são:

- Propor uma ferramenta de análise de quaisquer dados acadêmicos;
- Gerar modelos preditivos para classificação de alunos em risco e seleção automática do algoritmo que tiver a melhor acurácia;
- Estabelecer uma forma inovadora de recomendações, usando uma métrica derivada do sistema preditivo; e
- Obter uma ferramenta que reduza o tempo entre a coleta e análise para apoio de tomada de decisões de professores.

1.4. Contribuições

Através da solução computacional proposta e da aplicação da prova de conceito, as contribuições são categorizadas como gerais, aos gestores, aos professores e alunos, conforme apresentado abaixo:

Contribuições gerais: disponibilizar uma ferramenta acoplável a qualquer ambiente de aprendizado para previsão de desempenho, recomendações e visualizações; para gerar as recomendações usamos técnicas de seleção de atributos, na finalidade de avaliar quais são os atributos mais relevantes para a previsão; uso de estatística (simples) para avaliar qual valor da característica é a mais adequada na mudança comportamental.

Aos gestores: avaliar taxas de evasão; indicadores sobre os materiais pedagógicos do curso.

Aos professores: ferramenta de diagnóstico; facilitar a visualização de dados sobre o andamento; facilitador na comunicação sobre as necessidades individualizadas dos alunos.

Aos alunos: apontar características que comprometam seu aprendizado; receber mensagens recomendações sobre comportamentos que não favorecem o sucesso acadêmico.

1.5. Questão da Pesquisa

As pesquisas na área da informática aplicada à educação não combinam as técnicas de coletar dados, prever resultados, gerar recomendações e visualizar dados. Dessa forma, faz-se necessário um modelo de arquitetura conceitual para essas atividades. Ou seja, a questão da pesquisa é:

“Como oferecer um suporte computacional acoplável a qualquer ambiente de aprendizado para previsão de desempenho, recomendações e visualizações?”

1.6. Metodologia

Para responder à questão da pesquisa, descrevemos um *framework* de um modelo conceitual para análise e intervenção no aprendizado. Esse modelo corresponde a um conjunto de elementos caracterizados pelas suas funcionalidades e que seja possível implementar em qualquer linguagem de programação. Projetamos a solução e implementamos um protótipo como parte da solução com os requisitos (mínimos) necessários, na finalidade de aplicar uma prova de conceitos. Por fim, realizamos experimentos para atestar a viabilidade do modelo proposto. Na pesquisa usamos a abordagem qualitativa, de natureza aplicada, com o objetivo descritivo e com procedimentos experimentais.

1.7. Organização do Trabalho

Este trabalho está organizado da seguinte forma: O Capítulo 2 apresenta a fundamentação teórica com os conceitos Análise do Aprendizado, Mineração de Dados Educacionais, Aprendizado de Máquina, Sistemas de Recomendações e Sistemas de Visualização. No Capítulo 3 são apresentados os trabalhos correlatos a esta dissertação, destacando as pesquisas relacionadas à previsão de desempenho de alunos, Visualização de dados, Ferramentas de LA, Recomendações e por fim, são comparados com a nossa pesquisa, destacando as diferenças e semelhanças. No Capítulo 4 é descrito um modelo conceitual para a construção de da ferramenta. No Capítulo 5 detalhamos o desenvolvimento do protótipo computacional da solução. E no Capítulo 6 apresentamos os experimentos da aplicação do protótipo com um conjunto de dados público. Por fim, no Capítulo 8 estão as conclusões desta dissertação, bem como a produção científica e os trabalhos futuros.

2. Fundamentação Teórica

As pesquisas voltadas à análise de dados acadêmicos vêm avançando nos últimos tempos e se consolidaram em duas áreas, que são chamadas Mineração de Dados Educacionais e Análise de Aprendizado. Ambas as áreas são voltadas ao desenvolvimento de métodos de descoberta de conhecimentos e padrões em bases de dados educacionais com o objetivo de melhorar o processo de aprendizado. Nesta seção são descritos os fundamentos teóricos utilizados nesta pesquisa.

2.1. Análise do Aprendizado

Há diferentes definições sobre LA. O LA é descrito como um campo emergente no qual ferramentas de análise de dados são usadas para melhorar a ensino e o aprendizado (ELIAS, 2011). A Sociedade de Pesquisa de Análise do Aprendizado, do inglês *Society for Learning Analytics Research (SOLAR)*¹, descreve como a medição, coleta, análise e publicação de dados sobre alunos e seus contextos, com o objetivo de compreender e aperfeiçoar o aprendizado no ambiente em que ele, ocorre. É o uso de dados e modelos para prever o progresso e o desempenho do aluno, e a capacidade de agir com base nessas informações (BROWN, 2012). A LA é descrita como a aplicação de *Web Analytics*, tendo em vista o perfil do aluno, um processo de coleta e análise de detalhes de interações individuais de alunos em atividades de aprendizado *online* com o objetivo de melhorar a pedagogia, capacitar a aprendizado ativa, atingir populações de estudantes de risco e avaliar fatores que afetam a conclusão e o sucesso no aprendizado (JOHNSON et al., 2016).

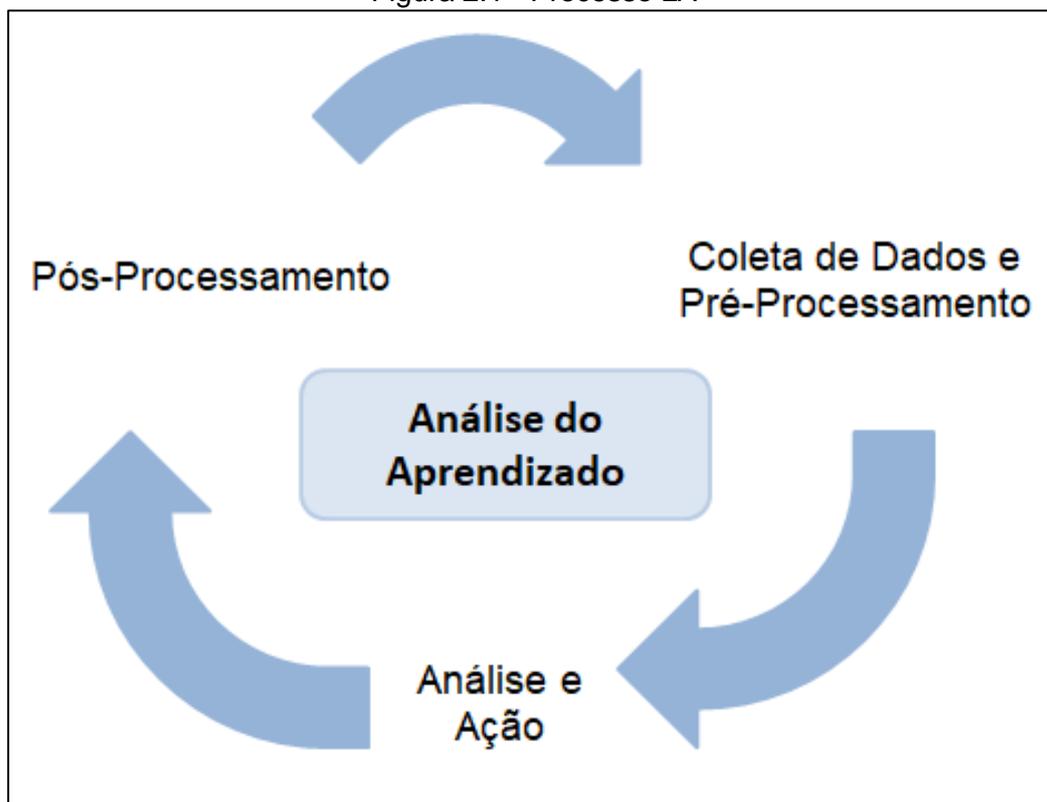
É difícil extrair valor das grandes quantidades de dados no aprendizado, pois quantidades significativas de atividades de alunos ocorrem externamente ao ambiente de aprendizado e, portanto, os registros são distribuídos em diversos locais diferentes, com diferentes padrões, proprietários e níveis de acesso (FERGUSON, 2012). O LA é focado na criação de sistemas capazes de ajustar e personalizar o conteúdo. Usando os dados de forma contínua, capturando, relatando, processando e agindo, assim minimizando o tempo entre

¹ <http://www.solaresearch.org/>

a captura e o uso dos dados. Faz uso de modelos analíticos para prever o comportamento, agir de acordo com as previsões. Para isso, os dados deverão ser periodicamente analisados e usados para tomar decisões, podendo haver melhorias nos processos educacionais (ELIAS, 2011).

Conforme ilustrado na Figura 2.1 - Processo LA, o processo geral de LA é geralmente um ciclo iterativo e é executado em três etapas principais: (1) coleta de dados e pré-processamento; (2) análise e ação; (3) e pós-processamento (CHATTI, 2012).

Figura 2.1 - Processo LA



Fonte: (CHATTI, 2012).

Coleta de dados e pré-processamento: Dados educacionais são a base do processo de LA.

Análises e ação: Com base nos dados pré-processados e no objetivo da análise, diferentes técnicas de LA podem ser aplicadas para explorar os dados, com a finalidade de descobrir padrões ocultos que podem ajudar a proporcionar uma experiência de aprendizado mais eficaz. Essas ações incluem

monitoramento, análise, previsão, intervenção, avaliação, adaptação, personalização, recomendação e reflexão.

Pós-processamento: Para a melhoria contínua do exercício de análise, o pós-processamento é fundamental. Ele pode envolver a compilação de novos dados de fontes de dados adicionais, refinando o conjunto de dados, determinando novos atributos necessários para a nova iteração, identificando novos indicadores/métricas, modificando as variáveis de análise ou escolhendo um novo método de análise.

O modelo de referência ilustrado na Figura 2.2 - Dimensões do LA, é descrito com base em quatro dimensões e identifica vários desafios e oportunidades de pesquisa na área (CHATTI, 2012). Este modelo possui as mesmas dimensões adotadas:

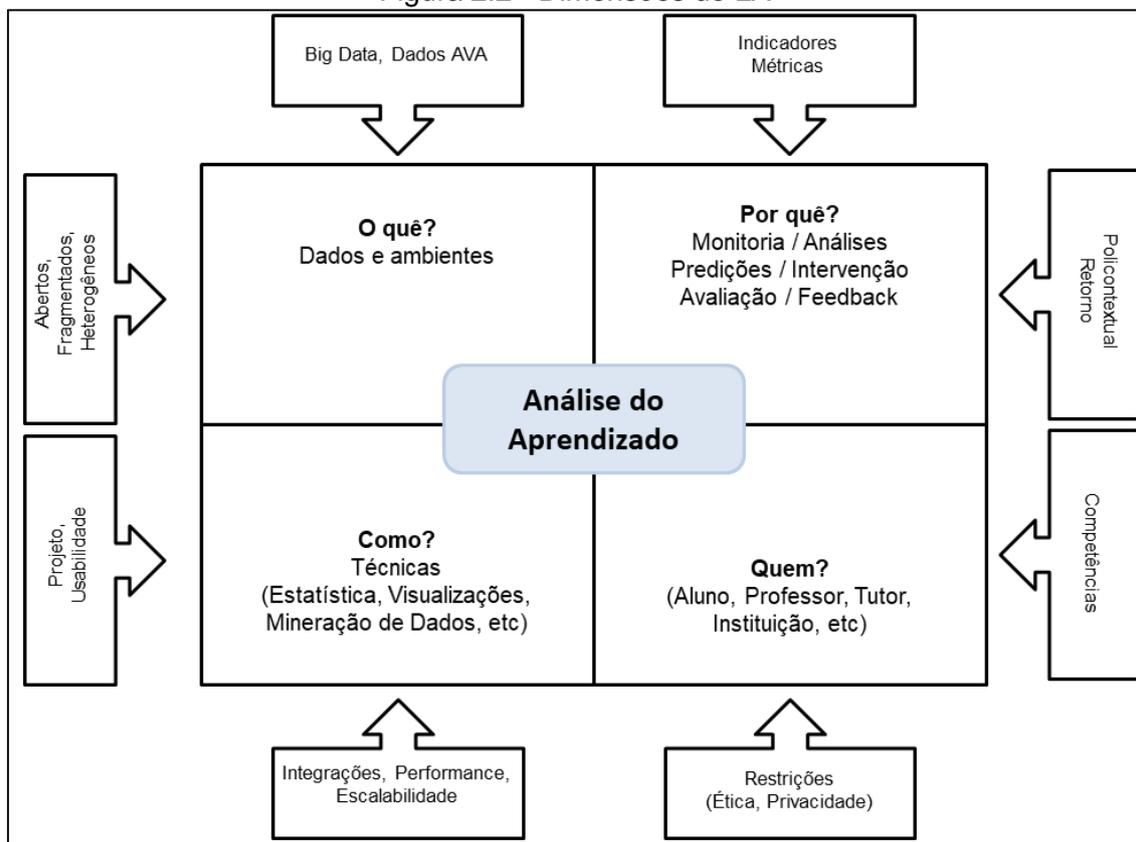
O quê? Esta dimensão refere-se aos tipos de dados coletados. Podem ser provenientes do ambiente de aprendizado utilizado, de fontes institucionais e também de redes sociais, como o Facebook e o Twitter.

Quem? Esta dimensão refere-se a quem os resultados da análise serão direcionados, podendo ser a alunos, professores, instituições educacionais (administradores), pesquisadores, projetistas de sistemas, etc. Cada um dos envolvidos tem suas próprias perspectivas, objetivos e expectativas em relação aos resultados.

Por quê? Esta dimensão está relacionada com os resultados seriam obtidos da análise realizada. Os objetivos são: Monitoramento e Análise; Predição e Intervenção; Tutoria e Mentoria; Avaliação e *Feedback*; Adaptação; Personalização e Recomendação; Reflexão.

Como? Esta dimensão refere-se às diferentes técnicas que podem ser empregadas para detectar padrões contidos nos dados e atingir os objetivos.

Figura 2.2 - Dimensões do LA



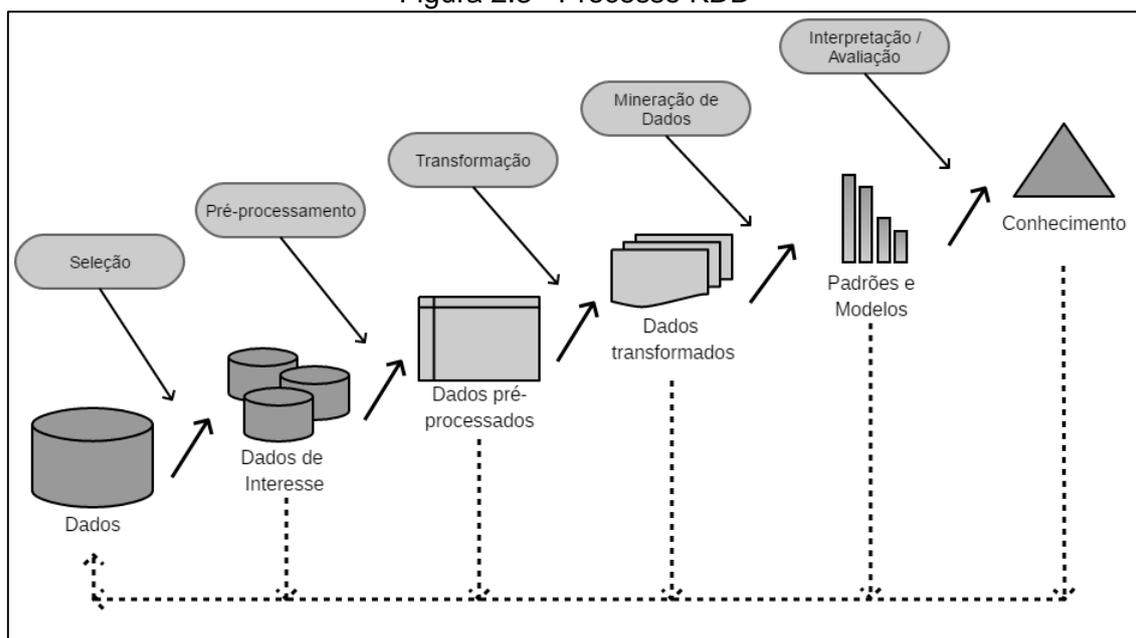
Fonte: (MOISSA, 2014; CHATTI, 2012).

2.2. Mineração de Dados

Mineração de dados é o processo de descoberta de conhecimentos interessantes e novos padrões, bem como modelos descritivos e preditivos de dados em larga escala. A mineração de dados é aplicação de algoritmos de aprendizado de máquina que permitem obter conhecimentos fundamentais a partir de grandes bases de dados. É um campo interdisciplinar que combina conceitos de áreas afins, como sistemas de banco de dados, estatística, aprendizado de máquina e reconhecimento de padrões. Na verdade, a mineração de dados faz parte de um processo maior de descoberta de conhecimento, que inclui tarefas de pré-processamento como extração de dados, limpeza de dados, junção de dados, redução de dados e construção de recursos, além das etapas de pós-processamento, como interpretação de modelos, confirmação e assim por diante. Esse processo de descoberta de conhecimento e mineração de dados tende a ser altamente iterativo e interativo (ZAKI et al. 2014). Por meio de uma

ampla variedade de campos, os dados estão sendo coletados e acumulados em um ritmo acelerado. Há uma necessidade crescente de geração de novas teorias e ferramentas computacionais, para ajudar os seres humanos a extraírem informações úteis (conhecimento) dos volumes de dados digitais de rápido crescimento. Essas teorias e ferramentas são os assuntos do campo da descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery in Databases* (KDD), (FAYYAD et al., 1996). A mineração de dados é uma parte integral do processo KDD. Conforme descrito na Figura 2.3 - Processo KDD, o KDD consiste em uma série de passos, desde a transformação, pré-processamento dos dados até o pós-processamento dos resultados da mineração (TAN et al., 2009). Observa-se que o processo iterativo acontece em cada etapa, o que significa que pode haver retorno a etapas anteriores para ajustes necessários.

Figura 2.3 - Processo KDD



Fonte: (FAYYAD et al., 1996).

2.3. Mineração de Dados Educacionais

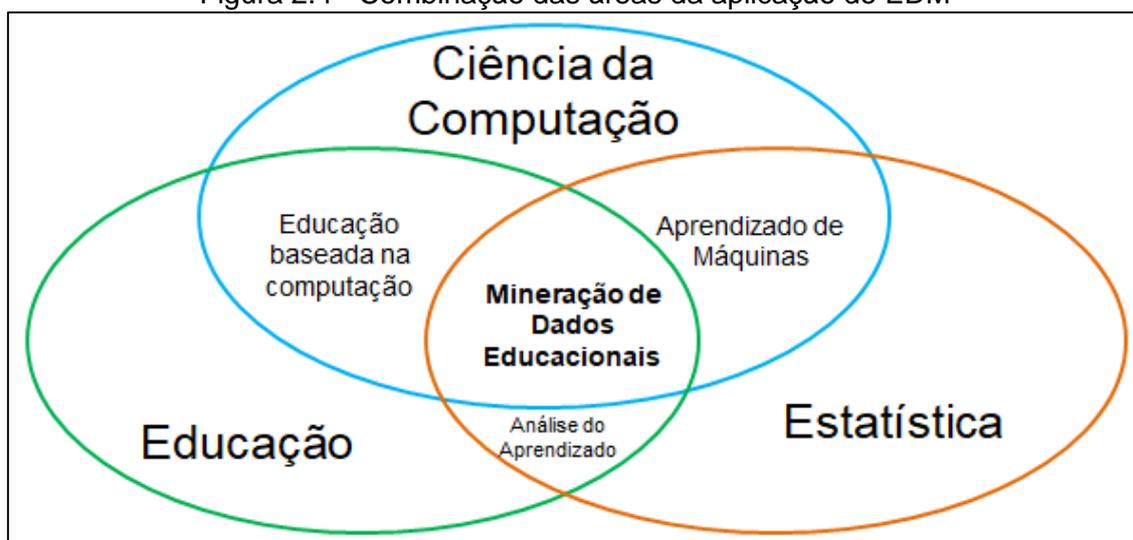
A Mineração de Dados Educacionais, em inglês *Educational Data Mining* (EDM), é a aplicação de técnicas de mineração de dados para tipos específicos de conjuntos de dados provenientes de ambientes educacionais para abordar questões educacionais importantes (ROMERO, 2007). Desta forma, pode-se

construir modelos analíticos que permitem descobrir padrões interessantes e tendências em informações de cada aluno. O processo de EDM converte os dados brutos de sistemas educacionais em informações úteis, que podem impactar na prática e na pesquisa educacional (ROMERO, 2010).

Mineração de dados educacionais é uma área interdisciplinar, incluindo áreas como a recuperação de informação, sistemas de recomendação, visualização de dados, análise de redes sociais (SNA), psicopedagogia, psicologia cognitiva, dentre outras. Em uma visão mais abrangente, definida por Romero e Ventura (2013), o EDM pode ser visualizado como a combinação das principais áreas: Ciência da Computação, Educação e Estatística, conforme apresentado na Figura 2.4 - Combinação das áreas da aplicação do EDM. Na intersecção dessas três áreas surge a relação com algumas subáreas estreitamente relacionadas com EDM, como a área de educação baseada em computador (E-learning), Mineração de Dados e Aprendizado de Máquina e Análise do Aprendizado (LA).

De todas as áreas mencionadas, a mais relacionada com EDM é LA, Em LA, as técnicas mais utilizadas são a aplicação de inferência estatística, visualização de dados, análise de redes sociais (SNA), análise de sentimento, análise de influência e análise de discurso.

Figura 2.4 - Combinação das áreas da aplicação do EDM



Fonte: (ROMERO, 2013).

2.4. Aprendizado de Máquina

O aprendizado de máquina, do inglês *Machine Learning* (ML), é uma subárea da Inteligência Artificial com o objetivo de desenvolver técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem-sucedida de um problema anterior (RICHERT, 2013). Existem alguns tipos de aprendizado de máquina como aprendizado supervisionado, e não supervisionado.

O objetivo do aprendizado de máquina é ensinar máquinas (software) a realizar tarefas, fornecendo-lhes alguns exemplos (como fazer ou não uma tarefa). Os pares seriam seus dados de treinamento e o conjunto de regras resultante (também chamado de modelo) poderia ser aplicado a futuros e-mails que ainda não vimos. Isso é aprendizado de máquina em sua forma mais simples (HEARTY, 2016).

2.4.1. Aprendizado Supervisionado

O supervisionado é a técnica de aprendizado de máquinas que se utiliza de experiências passadas para prever um valor. Em valores discretos utiliza técnicas de classificação que prevê rótulos de classes e em valores contínuos utiliza a técnica de regressão.

2.4.2. Aprendizado Não Supervisionado

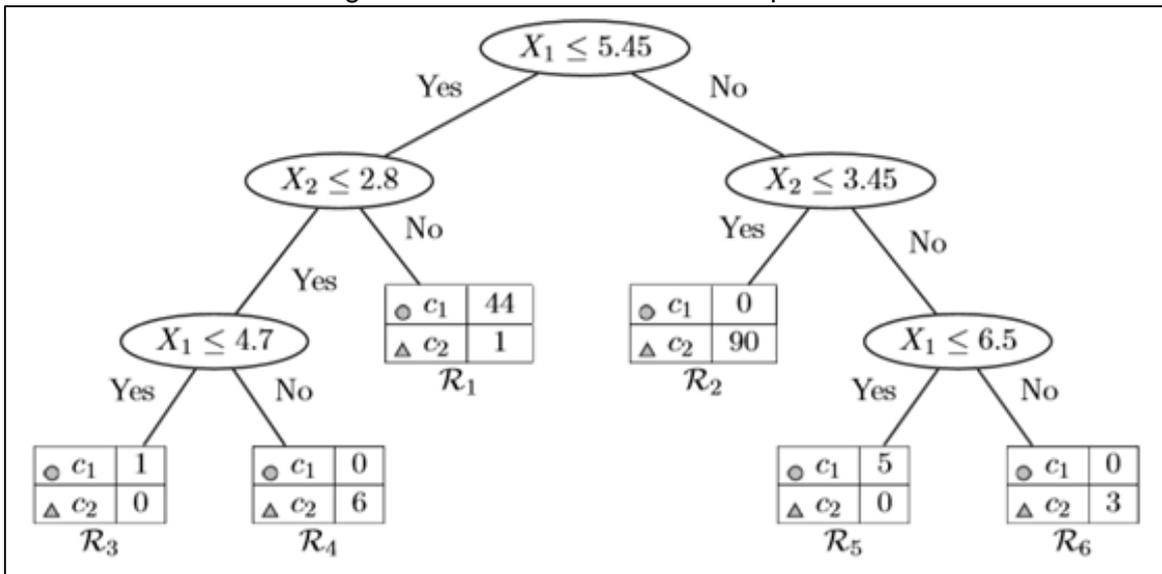
O aprendizado não supervisionado também é uma técnica de classificação onde é usado o agrupamento, do inglês *Cluster*, para classificar e agrupar classes similares de objetos.

2.4.3. Árvore de Decisão

O classificador Árvore de Decisão é um modelo de árvore recursivo baseado em partição para prever qual é o rótulo de uma instância, conforme apresentada na Figura 2.5 - Árvore de decisão simples. Uma árvore de decisão usa de paralelismo de eixos para dividir o espaço de dados em dois. Cada um desses

dois espaços de dados é dividido recursivamente até serem relativamente puros em termos das suas classes / rótulos. A hierarquia resultante de decisões divididas constitui o modelo de árvore de decisão, com os nós da folha rotulados com a classe majoritária entre os pontos nessas regiões. Para classificar uma nova instância, temos que avaliar recursivamente a qual ela pertence até chegarmos a um nó de folha na árvore de decisão para prever a qual seu rótulo (ZAKI et al. 2014).

Figura 2.5 - Árvore de decisão simples

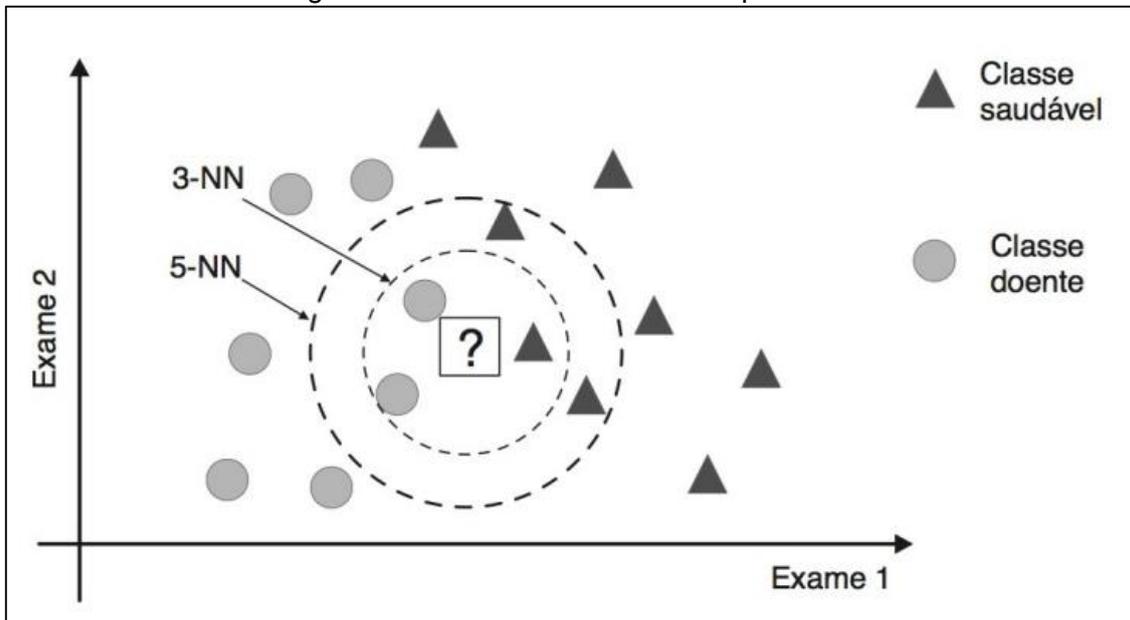


Fonte: (ZAKI et al. 2014).

2.4.4. K Vizinhos Próximos

O k Vizinhos Próximos, do inglês *K-Nearest Neighbors* (kNN) é um algoritmo de aprendizado de máquina que calcula a distância dos k (instâncias) vizinhos mais próximos, conforme apresentado na Figura 2.6 - Distância dos vizinhos próximos. Em que k é um parâmetro do algoritmo. Quando o valor de k é maior que 1, para cada ponto de teste, são obtidos k vizinhos. Cada vizinho vota em uma classe. As previsões dos diferentes vizinhos são agregadas de forma a classificar o ponto de teste. Essa agregação é efetuada de forma diferente em problemas de classificação e de regressão. Em problemas de classificação, em que a classe toma valores em um conjunto discreto, cada vizinho vota em uma classe (FACELI et al., 2011).

Figura 2.6 - Distância dos vizinhos próximos

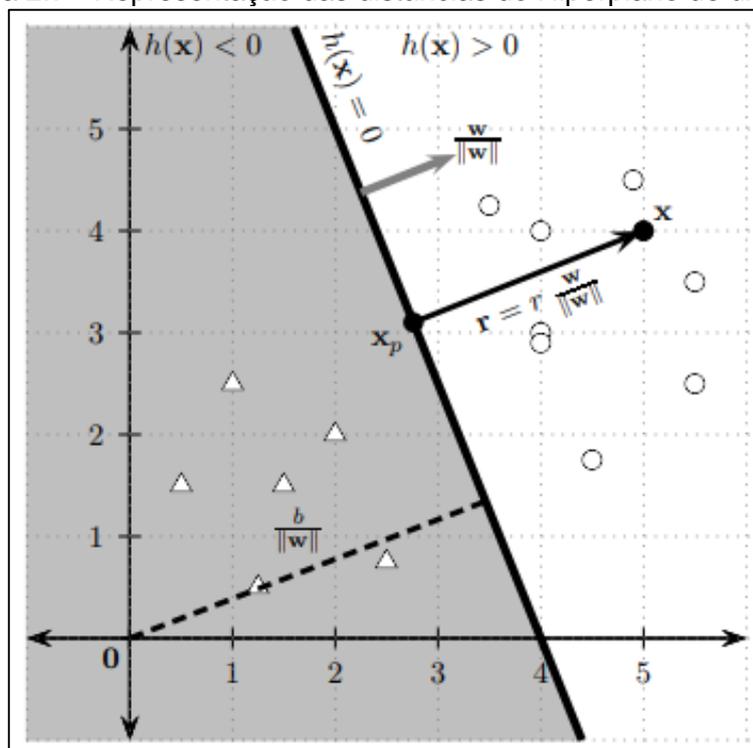


Fonte: (FACELI et al., 2011).

2.4.5. Máquina de Suporte a Vetor

O classificador Máquina de Suporte a Vetor, do inglês *Support Vector Machine* (SVM), é um dos classificadores mais eficientes para muitos domínios de problemas diferentes. O objetivo dos SVMs é para encontrar o hiperplano ideal que maximiza a margem de distância entre as classes, conforme apresentado na Figura 2.7 - Representação das distâncias do Hiperplano de um SVM. Através de seu kernel, SVM pode ser usado para encontrar limites não-lineares, no entanto, correspondem a um hiperplano linear em algum espaço “não-linear” de alta dimensão. O hiperplano é definido como o conjunto de todos os pontos representados em um gráfico, onde podemos obter o deslocamento quando os eixos são cruzados. Um hiperplano divide o espaço d-dimensional em dois meio-espaços. Diz-se que um conjunto de dados é linearmente separável se cada meio-espaço tiver pontos de uma única classe. Se o dataset de entrada é linearmente separável, então podemos encontrar um hiperplano de separação para todos os pontos rotulados (ZAKI et al. 2014).

Figura 2.7 - Representação das distâncias do Hiperplano de um SVM

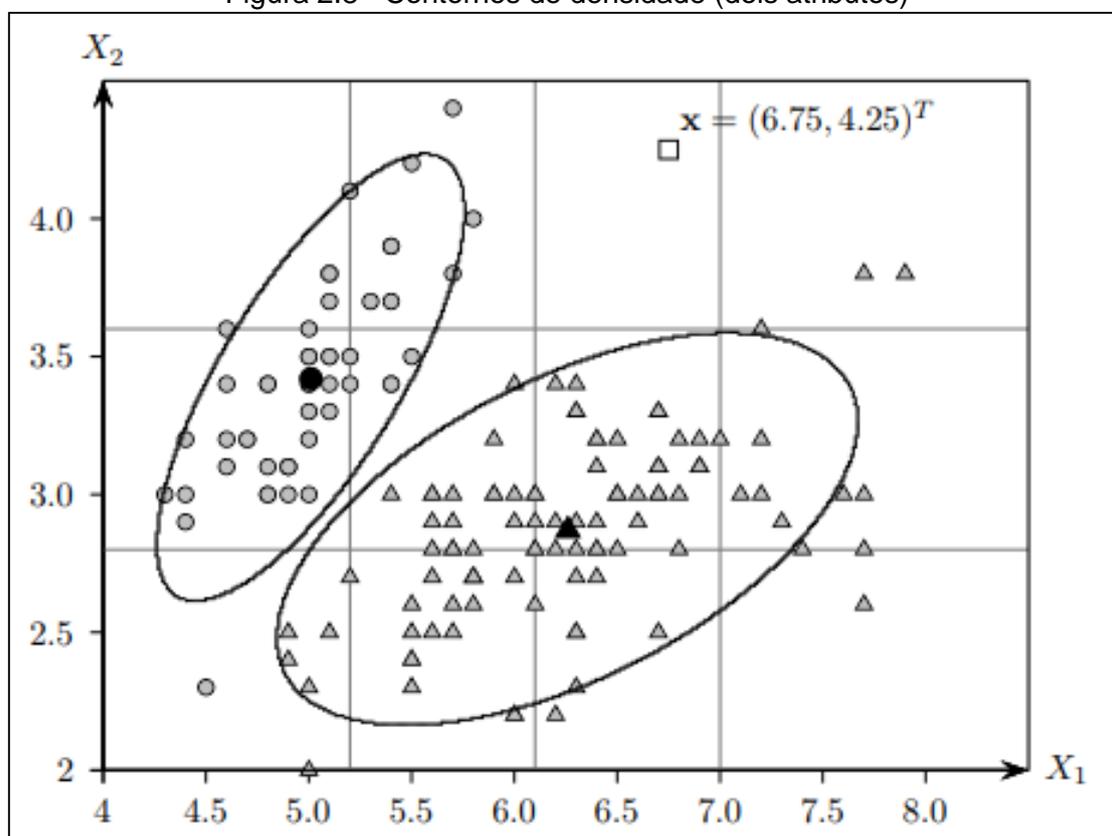


Fonte: (ZAKI et al. 2014).

2.4.6. Naive Bayes

O classificador de Bayes é uma abordagem de classificação probabilística, que usa o teorema de Bayes. A principal tarefa é estimar a função de densidade de probabilidade conjunta para cada classe, que é modelada através de uma distribuição normal multivariada, conforme apresentado na Figura 2.8 - Contornos de densidade (dois atributos). Uma limitação da abordagem de Bayes é o número de parâmetros a serem estimados, que é escalonado. O classificador de Bayes ingênuo faz a suposição simplificadora de que todos os atributos são independentes. No entanto, surpreendentemente eficaz para muitos conjuntos de dados (ZAKI et al. 2014).

Figura 2.8 - Contornos de densidade (dois atributos)



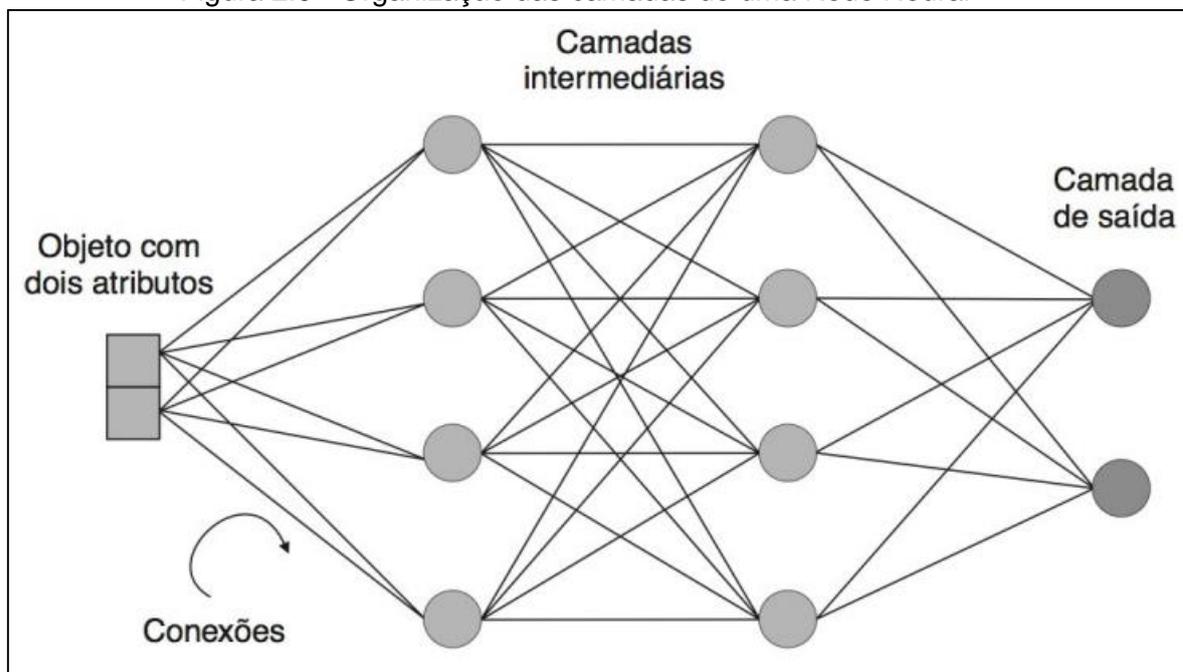
Fonte: (Zaki et al. 2014).

2.4.7. Redes Neurais

As redes neurais artificiais são inspiradas na estrutura e funcionamento do sistema nervoso, com o objetivo de simular a capacidade de aprendizado do cérebro humano na aquisição de conhecimento. A combinação de vários neurônios artificiais em sistemas neurais produz um elevado poder

computacional, uma vez que qualquer função possa ser representada por uma combinação de funções lógicas sendo modelada por uma rede formada por neurônios (FACELI et al. 2011). As redes neurais são uma classe de modelos matemáticos que treinam para produzir e otimizar uma definição para uma função (ou distribuição) sobre um conjunto de recursos de entrada. Logo, elas podem ser usadas para classificar, prever ou transformar suas entradas. Os algoritmos de redes neurais artificiais foram originalmente projetados a partir de estruturas neuronais biológicas. É orientado a grafos, onde os vértices são arranjadas em camadas. Cada vértice corresponde a uma unidade simplificada de um neurônio biológico. Conforme apresentado na Figura 2.9 - Organização das camadas de uma Rede Neural, ela é composta dos elementos: processo de aprendizado; conjunto de neurônios ou pesos; e funções de conectividade. A utilização desse conjunto de elementos nos permite aprimorar uma ampla gama de redes neurais (HEARTY, 2016).

Figura 2.9 - Organização das camadas de uma Rede Neural



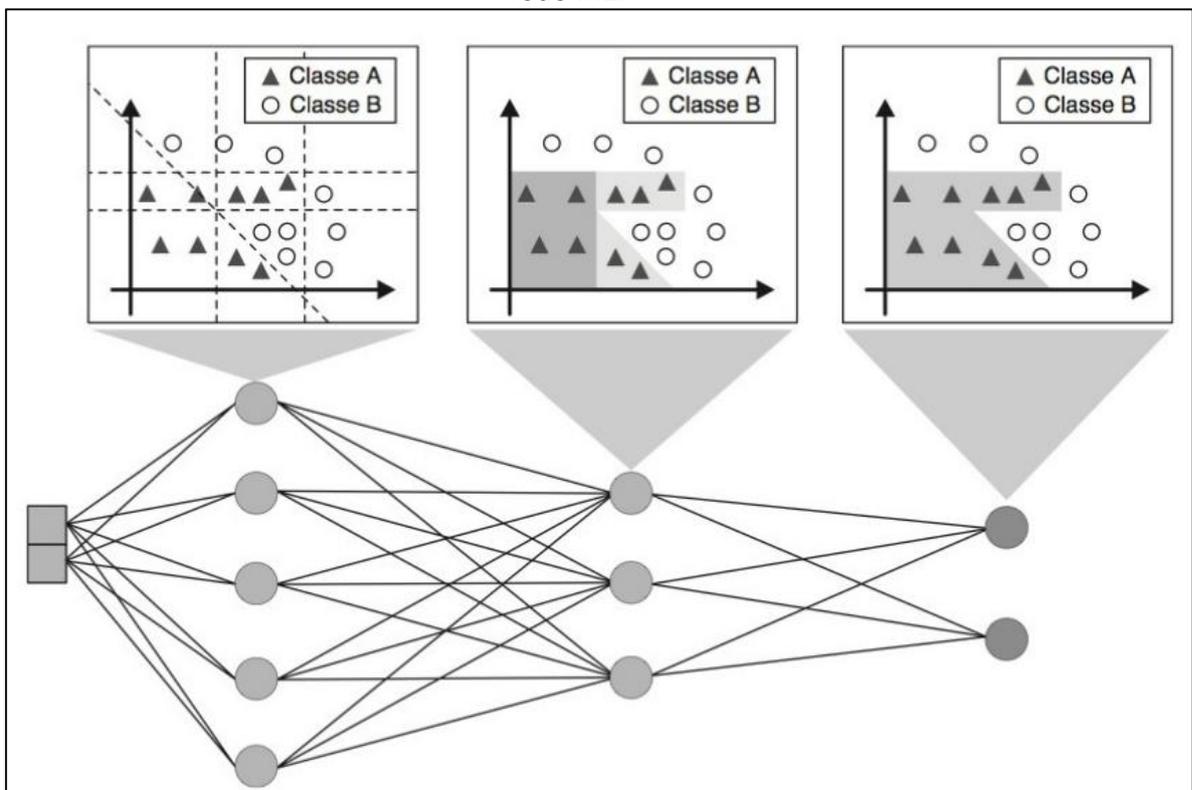
Fonte: (FACELI et al. 2011).

2.4.8. Perceptron em Multicamadas

O Perceptron em Multicamadas, em inglês *Multi-Layer Perceptron* (MLP), é uma rede neural que apresenta uma ou mais camadas intermediárias de neurônios e uma camada de saída. Nas camadas intermediárias são utilizadas funções de

ativação não lineares, como a função *sigmoidal*. Com funções de ativação lineares nos neurônios das camadas intermediárias, é equivalente a uma rede de uma só camada. Para resolver problemas não linearmente separáveis utilizando redes neurais artificiais, a alternativa mais utilizada é adicionar uma ou mais camadas intermediárias. Uma rede com uma camada intermediária pode implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função. A Figura 2.10 - Papel desempenhado pelos neurônios das diferentes camadas de uma rede MLP exemplifica o papel de cada neurônio para a definição das fronteiras. Na primeira camada, cada neurônio aprende uma função que define um hiperplano, o qual divide o espaço de entrada em duas partes. Cada neurônio da camada seguinte combina um grupo de hiperplanos definidos pelos neurônios da camada anterior, formando regiões convexas. Os neurônios da camada seguinte combinam um subconjunto das regiões convexas em regiões de formato arbitrário (FACELI et al. 2011).

Figura 2.10 - Papel desempenhado pelos neurônios das diferentes camadas de uma rede MLP



Fonte: (FACELI et al. 2011).

2.4.9. Métricas sobre Análise preditiva

Na aplicação de algoritmos de ML a problemas reais. Em geral, o conhecimento que se tem do domínio sendo investigado é provido unicamente pelo conjunto de dados de exemplos, a partir do qual a indução de um modelo preditivo/descritivo é então realizada. De maneira geral, pode-se afirmar que não existe técnica universal, ou seja, não é possível estabelecer, a priori, que uma técnica de aprendizado de máquina se sairá melhor na resolução de qualquer tipo de problema. Uma característica particular do domínio do aprendizado de máquina é a necessidade de experimentação. Pode ser necessário realizar ajustes em seus hiperparâmetros livres, o que leva à obtenção de múltiplos modelos para os mesmos dados. A validação de qualquer nova técnica geralmente envolve a realização de experimentos controlados, em que se demonstre a sua efetividade. Então, é recomendável seguir procedimentos que garantam a validade e a reprodutibilidade dos experimentos realizados. Essa avaliação experimental pode ser realizada segundo diferentes aspectos, tais como acurácia do modelo gerado, compreensibilidade do conhecimento extraído, tempo de aprendizado, requisitos de armazenamento do modelo, entre outros (FACELI et al. 2011).

Validação Cruzada

A validação cruzada é a técnica usada para validar o modelo preditivo, a partir do conjunto de dados. Nessa técnica, o conjunto de exemplos é dividido em dois subconjuntos de dados. Uma das partições é usada no treinamento de um modelo preditivo e a outra é usada para a validação do modelo (FACELI et al. 2011). Por simplicidade, seja um problema com duas classes. Usualmente, uma classe é denotada positiva (+) e a outra é denominada negativa (-). Temos então a matriz de confusão ilustrada na Figura 2.11 - Matriz de confusão, em que:

- **VP** corresponde ao número de verdadeiros positivos, ou seja, o número de exemplos da classe positiva classificados corretamente.
- **VN** corresponde ao número de verdadeiros negativos, ou seja, o número de exemplos da classe negativa classificados corretamente.

- **FP** corresponde ao número de falsos positivos, ou seja, o número de exemplos cuja classe verdadeira é negativa, mas que foram classificados incorretamente como pertencendo à classe positiva.
- **FN** corresponde ao número de falsos negativos, ou seja, o número de exemplos pertencentes originalmente à classe positiva, que foram incorretamente preditos como da classe negativa.

Figura 2.11 - Matriz de confusão

		Classe predita	
		+	-
Classe verdadeira	+	VP	FN
	-	FP	VN

Fonte: (FACELI et al. 2011).

Medidas de Desempenho

A partir da matriz de confusão, uma série de outras medidas de desempenho pode ser derivada. Entre elas, temos:

- Taxa de erro na classe positiva: proporção de classes positivas incorretamente classificadas, também conhecida como taxa de falsos negativos (**TFN**).

$$err_+(\hat{f}) = \frac{FN}{VP + FN}$$

- Taxa de erro na classe negativa: proporção de classes negativas incorretamente classificadas, também conhecida como taxa de falsos positivos (**TFP**).

$$err_-(\hat{f}) = \frac{FP}{FP + VN}$$

- Taxa de erro total: soma dos valores da diagonal secundária da matriz de confusão, dividida pela soma dos valores de todos os elementos da matriz.

$$ac(\hat{f}) = \frac{VP + VN}{n}$$

- Taxa de acerto ou acurácia total: calculada pela soma dos valores da diagonal principal da matriz de confusão, dividida pela soma dos valores de todos os elementos da matriz.

$$ac(\hat{f}) = \frac{VP + VN}{n}$$

- Precisão: proporção de exemplos positivos classificados corretamente entre todos aqueles preditos como positivos.

$$prec(\hat{f}) = \frac{VP}{VP + FP}$$

- Sensibilidade: corresponde à taxa de acerto na classe positiva. Também é chamada de taxa de verdadeiros positivos (**TVP**).

$$sens(\hat{f}) = rev(\hat{f}) = TVP(\hat{f}) = \frac{VP}{VP + FN}$$

- Especificidade: corresponde à taxa de acerto da classe negativa. Seu complemento corresponde à taxa **TFP**.

$$esp(\hat{f}) = \frac{VN}{VN + FP} = 1 - TFP(\hat{f})$$

As medidas anteriores podem ser facilmente generalizadas para problemas com mais de duas classes pela consideração de cada classe, como positiva em relação ao conjunto das demais classes (exceto a taxa de erro/acerto total, em que todas as classes são consideradas globalmente). Obtém-se um valor de desempenho para cada classe.

2.5. Sistemas de Recomendações

Com o aumento da quantidade de conteúdo disponível na internet é comum se deparar com uma diversidade grande de opções e temos dificuldade em escolher uma entre várias alternativas de conteúdo. Sistemas de recomendação (SR) vêm a auxiliar na solução desse problema. Segundo (COSTA et al. 2013), SR são técnicas que auxiliam usuários na busca e seleção de um conteúdo focado em seu perfil, funcionando literalmente como filtros de informações. O princípio técnico de um SR é o uso de um algoritmo para combinar dados de contexto e entrada para produzir sugestões aos usuários, visando a geração de

recomendações (ISINKAYE et al. 2015). Nesse caso são sistemas de filtragem de informação que lidam com o problema de sobrecarga de informação, filtrando dados de grande quantidade de informação de acordo com as preferências do usuário, interesse ou comportamento. Assim, tem a capacidade de prever se um determinado usuário prefere um item ou não com base no perfil do usuário (ISINKAYE et al., 2015).

No contexto educacional, na maioria dos casos, os sistemas de recomendações apresentam muitas limitações. Nessas situações, é necessário considerar outras questões além das preferências do professor ou do estudante em determinado recurso. Em particular, alguns requisitos deveriam ser considerados no projeto e desenvolvimento de um SR para ambientes educacionais online. Nesse sentido, os SR consistem em uma possível solução personalizada para cada indivíduo, baseando-se em suas características e/ou preferências. A utilização de Sistemas de Recomendação em AVA pode contribuir no processo de ensino e aprendizado, uma vez que os alunos poderão receber recomendações personalizadas de materiais didáticos, além de contribuir para explorar mais o acervo existente, permitindo maior interação entre os materiais didáticos, professores e alunos. (COSTA et al., 2013)

2.6. Sistemas de Visualização

O tempo tem um papel cada vez mais crítico no processo de tomada de decisão. Portanto, tornar a apresentação dos resultados visual, fácil de consumir, permitindo consultas, interação e exploração torna a tomada de decisão mais rápida e eficaz. Uma das melhores formas de explorar e tentar entender um grande conjunto de dados é pela visualização (COSTA, 2017). A visualização é a apresentação gráfica da informação, com o objetivo de proporcionar ao espectador uma compreensão qualitativa dos conteúdos informativos. É também o processo de transformar objetos, conceitos e números em uma forma que é visível aos olhos humanos. Visualização de dados auxilia a compreender padrões, tendências e relações entre grupos de dados.

Uma vantagem da exploração de dados de forma visual é que os usuários estão diretamente envolvidos no processo de mineração de dados. E encontrar

a informação valiosa escondida em grande quantidade de dados é uma tarefa difícil (Costa, 2017).

Para que a mineração de dados seja eficaz, é importante incluir o usuário no processo de exploração de dados e combinar a flexibilidade, a criatividade e o conhecimento geral do ser humano, com a capacidade de armazenamento e o poder computacional. A exploração de dados visuais visa integrar o ser humano no processo de exploração de dados, aplicando suas habilidades perceptivas aos grandes conjuntos de dados disponíveis nos sistemas de computador atuais (KEIM, 2002).

A visualização geralmente é parte de um processo maior, que pode ser a análise exploratória de dados, descoberta de conhecimento, apresentação de resultados de técnicas ou análise visual. Visualização e análise andam lado a lado, com o objetivo de construir um modelo que representa ou aproxima os dados. A visualização na exploração de dados é usada para transmitir informações, descobrir novos conhecimentos e identificar estruturas, padrões, anomalias, tendências e relacionamentos (KEIM, 2010).

3. Trabalhos Correlatos

Neste capítulo são descritos trabalhos de pesquisas relacionados à análise acadêmica, previsão de desempenho usando algoritmos de aprendizado de máquinas com abordagens de mineração de dados, aplicação de técnicas de recomendações e apresentação de dados. Buscamos também soluções que combinam as várias técnicas abordadas em nossa pesquisa, contudo as propostas na literatura são restritas e apresentam apenas experimentos ou um escopo fechado e reduzido. E percebemos que a maioria dos casos não combinam os diversos aspectos abordados nesta dissertação. Quando se trata de sistemas preditivos no contexto acadêmico as pesquisas (geralmente) não fazem abordagens na intervenção no processo de aprendizado. Quando fazem, auxiliam nas preferências dos estudantes e não em suas necessidades acadêmicas. Por fim, a maioria não combinava as técnicas apresentadas com um sistema de visualização.

3.1. Identificar evasão de alunos

Há pesquisas que usam técnicas de mineração de dados voltadas a identificar padrões que levam alunos a evadir, por meio de dados acadêmicos. A pesquisa (PASCOAL et al., 2016) propõe uma abordagem para auxiliar instituições nas tomadas de decisão para combater a evasão através de dados acadêmicos e socioeconômicos. O trabalho (OLIVEIRA JÚNIOR et al. 2015) propõe uma abordagem computacional para a identificação de padrões a serem utilizados na análise da evasão de estudantes em cursos presenciais de graduação, a fim de auxiliar os tomadores de decisão das instituições de ensino. E a pesquisa (LANES; ALCÂNTARA, 2018) visa identificar estudantes que apresentam risco de evasão a partir do seu primeiro ano no curso de graduação.

3.2. Previsão de desempenho

Há vários trabalhos e pesquisas voltados a previsão de desempenho de alunos através de técnicas de mineração de dados usando algoritmos de aprendizado de máquina. O trabalho (DE BRITO et al., 2014) propõe a previsão de

desempenho dos alunos no primeiro período do curso de Ciência da Computação da UFPB, através das suas notas de ingresso no vestibular. A pesquisa (GOTTARDO et al., 2012) propõe o uso de inferências sobre o desempenho de estudantes a partir de dados coletados em séries temporais, com o objetivo principal de investigar a viabilidade de obtenção destas informações em etapas iniciais de realização do curso, de forma a apoiar a tomada de ações proativas. Já o trabalho (VIEIRA; VEIGA, 2013) tem o foco em métodos de árvores de decisão para a previsão de desempenho acadêmico dos alunos. A pesquisa (GOVINDARAJAN et al., 2015) faz uso de dados acadêmicos na finalidade de reduzir taxas de insucesso de aprendizado.

3.3. Visualização de dados

Há vários trabalhos que propõem soluções voltadas à visualização de dados para apoiar os professores e alunos, gerando informações gráficas como *dashboards* para apoiar os professores na tomada de decisão. A pesquisa (VERBERT et al., 2013) apresenta um framework conceitual que ajuda professores e alunos a analisar dados sobre o aprendizado. Já o trabalho (DUVAL 2011) é voltado a técnicas de visualização provendo *dashboards* a professores e alunos usando ferramentas de coletar dados. E a pesquisa (KEIM, 2012) propõe uma classificação de técnicas de visualização combinada como parte da mineração de dados.

3.4. Ferramentas de análise de dados

Existem trabalhos que fazem uso de ferramentas ou soluções de análise de dados acadêmicos. O trabalho (ARNOLD; PISTILLI, 2012) é voltado a permitir professores a oportunidade de prover em tempo real feedback sobre seu aprendizado. A pesquisa (MARTIN; NDOYE, 2016) faz uso das ferramentas *Tableau* e *Many Eyes* para analisar dados acadêmicos. O trabalho (DYCKHOFF et al. 2012) apresenta detalhes teóricos sobre a implementação de uma ferramenta que permita professores explorarem e correlacionar o uso de objetos de aprendizado e o comportamento de alunos com resultados de avaliações com base em indicadores gráficos. Já a pesquisa (GRELLER; DRACHSLER, 2012) propõe um framework genérico dando suporte a práticas educacionais e discute

sobre as dimensões chave da análise do aprendizado no contexto educacional. E o trabalho (PADILHA; SOUZA, 2017) apresenta o uso das ferramentas *Gephi* e *RapidMiner* para análise de interações entre alunos e professores.

3.5. Recomendações

Nos últimos anos foram publicados alguns trabalhos voltados à recomendações no processo de aprendizado. No entanto, na maioria dos casos, as recomendações são baseadas nas preferências de ensino de professores e nas preferências de aprendizado dos alunos baseado em filtragem colaborativa. Na pesquisa (AGUIAR et. al.2018) é proposta uma estratégia de recomendação personalizada referente a Objetos de Aprendizado, com base nos Estilos de Aprendizado dos estudantes e na Filtragem Colaborativa baseada em Tendências combinadas por meio de um Algoritmo Genético. No trabalho (JÚNIOR; DORÇA, 2018) foi implementada uma proposta de recomendação de Objetos de Aprendizado usando algoritmos genéticos combinada com ontologias na modelagem de estudantes. Na pesquisa (BEZERRA et al., 2018) foi desenvolvido um SR híbrido integrado a uma ontologia que gerencia as informações de perfil do aprendiz e de conteúdos provenientes do Youtube e da Wikipédia. O trabalho (SANTANA et al., 2017) tem como objetivo extrair boas e más práticas pedagógicas a partir de processo de ensino usando Sistema de Tutor Inteligentes (STI), fornecendo recomendações a partir de respostas o sistema avalia quais tópicos de um tema específico precisa de reforço e recomenda os recursos na resolução das questões. A pesquisa (FERREIRA et al. 2018) estuda o avanço da aprendizado colaborativo utilizando o modelo proposto, chamado Big Five, que promove a recomendação de agrupamento usando como base os textos dos alunos. O trabalho (REZENDE et al. 2018) propõe um sistema, chamado *PERSONNA-TEC*, que é capaz de realizar recomendação de objetos de aprendizado através do plano didático do professor. A pesquisa (GOTARDO; CEREDA; JUNIOR, 2013) apresenta uma abordagem que usa algoritmos de aprendizado acoplados para integrar diferentes técnicas de aprendizado, mineração de dados e explorar um conjunto de dados educacionais oferecendo recomendações que possam indicar preferências do aluno que obtém melhores resultados.

3.6. Comparação entre as propostas

Os trabalhos acima referidos propõem técnicas de mineração de dados, técnicas de predição ou descrição, recomendações e visualização. Segundo (MOISSA; GASPARINI; KEMCZINSKI et al. 2014), em sua pesquisa, poucos trabalhos mencionam a intervenção e, no entanto, há uma grande quantidade de trabalhos que não possibilitam intervenção alguma. Conforme apresentado na Tabela 3.1 - Comparação entre as contribuições, a maioria dos trabalhos não combina as diversas técnicas abordadas pelo nosso trabalho, que são: Aprendizado de Máquina (ML), Sistemas de Visualização (SV), Sistemas de Recomendações (SR) e acesso a Múltiplas Bases de dados (MB). Também são destacadas as semelhanças e diferenças entre os trabalhos.

Os trabalhos foram agrupados conforme sua abordagem, como Previsão de desempenho, Visualização de dados acadêmicos, Recomendações e uso de Ferramentas de análise de dados. Pois, este trabalho consiste em um framework de um modelo conceitual caracterizado por essas funcionalidades. No qual, podemos destacar os trabalhos abaixo, pois auxiliaram na concepção da descrição da solução computacional.

- a) (GOVINDARAJAN et al., 2015), que faz uso de dados acadêmicos na finalidade de reduzir taxas de insucesso de aprendizado.
- b) (ARNOLD; PISTILLI, 2012), que é voltado a permitir professores proverem feedback aos seus sobre seu aprendizado.
- c) (KEIM, 2012) propõe uma classificação de técnicas de visualização combinada como parte da mineração de dados.
- d) (GOTARDO; CEREDA; JUNIOR, 2013) apresenta uma abordagem que usa algoritmos de aprendizado acoplados para integrar diferentes técnicas de aprendizado, mineração de dados e explorar um conjunto de dados educacionais oferecendo recomendações.

Tabela 3.1 - Comparação entre as contribuições

Trabalhos	ML	SV	SR	MB	Semelhanças	Diferenças
Evasão de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos (PASCOAL et al., 2016)	Sim	-	-	-	- Prevê o resultado de alunos	- É um experimento - O target só é a evasão escolar - Busca somente os fatores socioeconômicos, não dados acadêmicos
(TESE – DEVO CITAR?) Identificação de padrões para a análise da evasão em cursos de graduação usando mineração de dados educacionais (OLIVEIRA JÚNIOR et al. 2015)	Sim	Sim	-	-	- Prevê o resultado de alunos - Faz uso de seleção de atributos	- É um experimento - O target só é a evasão escolar - São experimentos usando WEKA, uma ferramenta pronta - A apresentação dos dados é por meio da geração do próprio WEKA
Predição do Desempenho do Aluno usando Sistemas de Recomendação e Acoplamento de Classificadores (GOTTARDO et al., 2012)	Sim	-	Sim	-	- Prevê o resultado do aluno - Usa técnicas de recomendações - Usa técnicas de discretização	- É um experimento - As técnicas de recomendação são usadas para montar um perfil do aluno, não apresentando recomendações sobre suas necessidades, mas sim voltadas aos seus interesses
Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina (DE BRITO et al., 2014)	Sim	Sim	-	-	- Prevê o resultado do aluno - Usa técnicas de visualização	- É um experimento - A técnica de visualização é somente para apresentar o resultado do experimento
Previsão de desempenho de estudantes em cursos EAD utilizando mineração de dados: uma estratégia baseada em séries temporais (GOTTARDO et al., 2012)	Sim	Sim	-	-	- Prevê o resultado do aluno - Usa técnicas de visualização	- É um experimento - A técnica de ML é a regressão - A técnica de visualização é somente para apresentar o resultado do experimento
Árvores de Decisão Aplicadas na Previsão de Desempenho de Alunos: Estado da Arte (VIEIRA; VEIGA, 2013)	Sim	-	-	-	- Uso de conceitos para previsão de desempenho do aluno - Uso da Árvore de Decisão	- É apenas um modelo conceitual para previsão de desempenho usando a Árvore de Decisão
Learning analytics dashboard application (VERBERT et al., 2013)	-	Sim	-	-	- Faz uso de SV - Solução para a educação com uso de painéis para alunos e professores	- É um estudo sobre 15 diferentes aplicações que fazem uso de técnicas de visualização para alunos e professores

Trabalhos	ML	SV	SR	MB	Semelhanças	Diferenças
Attention please!: learning analytics for visualization and recommendation (DUVAL 2011)	-	Sim	-	-	- Faz uso de SV	- Faz uso de ferramentas prontas como <i>wakooopa</i> e <i>tripit</i> para coletar e visualizar os dados
Course signals at Purdue: Using learning analytics to increase student success (ARNOLD; PISTILLI, 2012)	Sim	-	Sim	-	- Prevê o resultado do aluno - Usa técnicas de recomendações - Propõem uma solução	- A técnica de recomendação usando sistemas de cores: verde, amarelo e vermelho - Escopo fechado de uma universidade não abrindo maiores detalhes
Using learning analytics to assess student learning in online courses (MARTIN; NDOYE, 2016)	-	Sim	-	Sim	- Usa SV para professores e alunos	- É um experimento - Usa ferramentas prontas Tableau e Many Eyes - Não faz a separação dos dados por aluno, expondo dados de outros
Design and implementation of a learning analytics toolkit for teachers (DYCKHOFF et al. 2012)	Sim	Sim	Sim	-	- É discutido sobre detalhes teóricos de: previsão de desempenho (ML); SV; e SR	- É uma apresentação de detalhes teóricos sobre a implementação de uma ferramenta LA
Translating learning into numbers: A generic framework for learning analytics (GRELLER; DRACHSLER, 2012)	Sim	Sim	Sim	-	- É discutido sobre detalhes teóricos de: previsão de desempenho (ML); SV; e SR	- É uma apresentação de detalhes teóricos sobre a implementação de uma ferramenta LA
Uma Experiência do Uso de Ferramentas de Learning Analytics para Análise de Interações (PADILHA; SOUZA, 2017)	Sim	Sim	-	Sim	- Prevê o resultado do aluno - Usa SV para apresentar dados da interação dos alunos	- É um experimento - Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) - Gephi e RapidMiner
Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem (AGUIAR et. al.2018)	-	-	Sim		- Usa técnicas de recomendação	- Faz uso de objetos de aprendizado - Usa a semelhança entre itens/usuários, não faz uso de suas necessidades
Uma Abordagem para a Criação e Recomendação de Objetos de Aprendizagem usando um Algoritmo Genético, Tecnologias da Web Semântica e uma Ontologia (JÚNIOR; DORÇA, 2018)	-	-	Sim		- Usa técnicas de recomendação	- É um experimento - Faz uso de objetos de aprendizado - Usa a semelhança entre itens/usuários, não faz uso de suas necessidades - Usa algoritmos genéticos, enquanto usamos algoritmos classificadores

Trabalhos	ML	SV	SR	MB	Semelhanças	Diferenças
Um Sistema de Recomendação Híbrido Integrado a Ontologia em um Ambiente de Aprendizagem Ubíqua (BEZERRA et al., 2018)	-	Sim	Sim		- Usa técnicas de recomendação - Usa técnicas de visualização dos dados recomendados	- Usa ontologias - Usa a semelhança entre itens/usuários, não faz uso de suas necessidades
Extração e recomendação de boas e más práticas pedagógicas a partir de processos de ensino e aprendizagem usando um sistema tutor inteligente gamificado (SANTANA et al., 2017)	-	-	Sim		- os dados apresentados servem para melhora pedagógica	- Uso de Sistemas de Tutores Inteligentes - Usa técnicas de gamificação
Detecção automática de traços de personalidade e recomendação de agrupamento com o modelo Big Five (FERREIRA et al. 2018)	-	-	Sim	-	- Faz uso de algoritmos de aprendizado de máquina - Identifica um perfil do aluno através de dados educacionais	- É um experimento - Não há recomendações, nem visualizações - Os algoritmos classificadores são baseados em textos
Recomendação Baseada no Perfil e Contexto Tecnológico do Aluno (REZENDE et al. 2018)	-	-	Sim	Sim	- Faz uso de algoritmos de aprendizado de máquina - Propõem uma solução	- Usa ontologias - Faz uso de objetos de aprendizado - As recomendações são para objetos de aprendizagem - Usa a semelhança entre itens/usuários, não faz uso de suas necessidades
Predição do Desempenho do Aluno usando Sistemas de Recomendação e Acoplamento de Classificadores (GOTARDO; CEREDA; JUNIOR, 2013)	Sim	-	Sim		- Faz uso de algoritmos de aprendizado de máquina - Propõem uma solução para recomendações - Usa técnicas de visualização	- É um experimento - Usa a semelhança entre itens/usuários

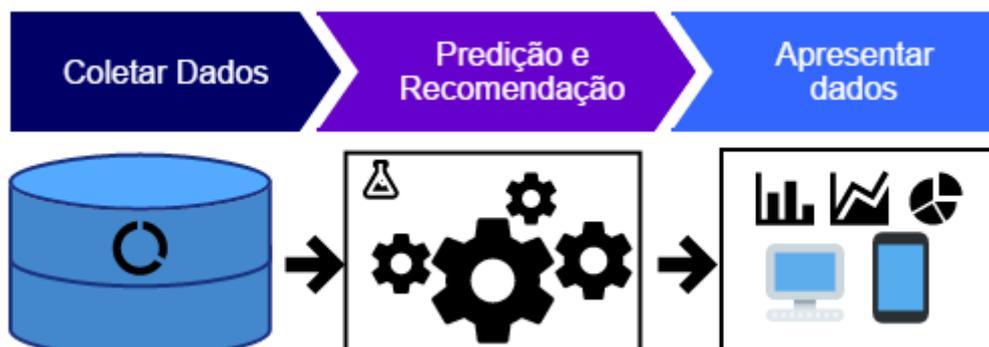
Fonte: próprio autor.

4. Descrição da Solução

Neste capítulo é descrito um modelo conceitual para a construção de uma ferramenta completa flexível e sensível ao contexto. Esta proposta consiste em uma ferramenta de análise de dados acadêmicos, que após ser devidamente configurada, será provedora de informações importantes aos alunos e professores. Essa solução permite ao professor: (a) obter informações sobre andamento do aprendizado dos alunos; (b) saber quais são os alunos que estão com dificuldades; (c) identificar quais são as principais características que estão comprometendo o desempenho acadêmico; e (d) efetuar eventuais intervenções no ensino de seus alunos através de um sistema de recomendações. Já os alunos terão informações sobre seu rendimento e receberão recomendações de seu professor no processo de aprendizagem.

Conforme apresentado na Figura 4.1 - Fases da Análise dos dados, a ferramenta é focada: em Coletar Dados Acadêmicos; Prever o resultado dos alunos e Gerar recomendações; e Apresentar painéis de visualização ao professor. No qual, um usuário com perfil administrativo (Analista) deverá configurar as fontes de dados, gerenciar os metadados e importar para o sistema. Após isso, os dados históricos de turmas anteriores servirão para treinamento dos modelos preditivos. Com base nesses modelos, os resultados dos alunos serão previstos e, para cada aluno classificado “em risco”, será gerada uma mensagem de recomendações. O professor terá a oportunidade de visualizar as mensagens e, caso necessário, ajustá-las antes de enviar ao aluno classificado “em risco”. Ao final, o usuário de perfil administrativo configurará painéis de visualização de dados que serão acessíveis pelo professor.

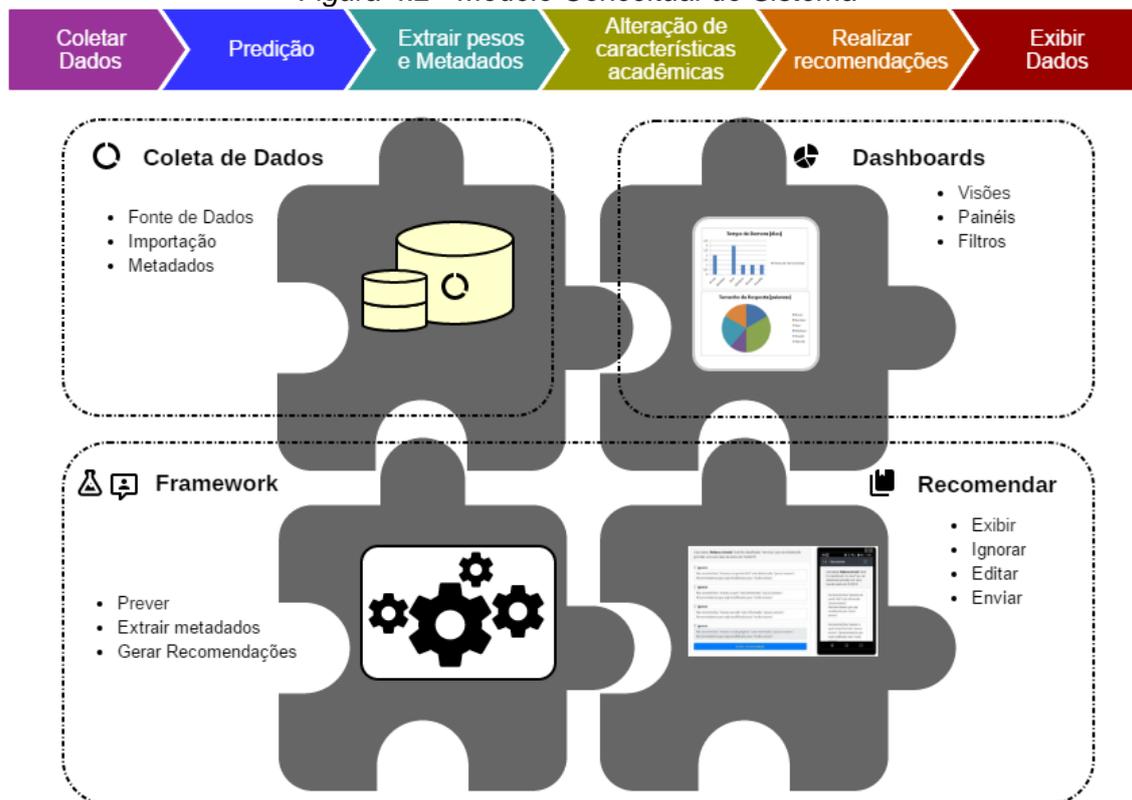
Figura 4.1 - Fases da Análise dos dados



Fonte: Próprio autor.

Os dados serão periodicamente reprocessados através da seleção, coleta e transformação dos dados minimizando o tempo de demora entre a captura e uso dessas informações. Segundo Clow (2012), o processo de LA é feito através de ciclos. Os dados deverão ser coletados e tratados para geração de métricas, análise ou visualizações e fechamento do ciclo através de feedback aos seus alunos por intermédio de intervenções apropriadas e personalizadas. A análise poderá ser incremental, ou seja, após o fechamento do ciclo o usuário poderá verificar pontos de melhoria. Tais como, adicionar outras informações de interações dos alunos, filtrar ou fazer diferentes tratamentos nos dados, criar novos modelos preditivos ou até adicionar outras métricas e gráficos nos painéis dos professores e dos alunos.

Figura 4.2 - Modelo Conceitual do Sistema



Fonte: Próprio autor

Conforme apresentado na Figura 4.2 - Modelo Conceitual do Sistema, cada etapa de dados conterá módulos do sistema. Na etapa Coleta e Tratamento de Dados contém os módulos de Coleta, Processamento e Transformação. Já na etapa de Análise, que corresponde ao framework (STOLL; CURY; DE

MENEZES, 2018), contém os módulos de Predição, Recomendação e geração das mensagens de recomendações. Sendo que a mensagem candidata é sugerida para o professor, que poderá avaliar sua pertinência antes de enviar, caso necessário, antes do envio. E na etapa de Apresentação contém os módulos de Visão e Painéis (STOLL et al. 2019).

Assim, apesar de todos os módulos projetados no sistema serem desenvolvidos, entendemos que ferramentas poderiam substituir as etapas de “*Coleta e Tratamento*” e “*Apresentação*”, já que existem várias ferramentas maduras no mercado que fazem isso. Sendo assim, é possível que a solução proposta possa trabalhar conjuntamente com outras.

4.1. Coleta e Tratamento

Na etapa de Carga e Tratamento ocorre a coleta, processamento e transformação dos dados. No módulo de Coleta de Dados será feita a carga inicial de dados, também será necessário catalogar as fontes de dados e gerenciar os metadados. Já no módulo de Processamento será possível alterar configurações de campos e valores nas tabelas importadas, limpar e transformar valores contínuos em valores discretos. No módulo de transformação será gerado um modelo de dados otimizado e eficiente para aplicar técnicas de aprendizado de máquina voltadas à predição.

4.1.1. Coleta de dados

No módulo de coleta de dados é possível coletar dados de vários ambientes através de importações para dentro da base de dados do próprio sistema. Então, as várias consultas executadas pelo sistema nos dados acadêmicos não prejudicarão a performance do banco de dados do AVA. O sistema poderá ter acesso aos dados através diversas fontes de dados, tais como:

- JSON: acrônimo de JavaScript Object Notation
- CSV: arquivo separado por vírgula
- Acesso direto ao banco de dados

Após a importação dos dados será possível fazer o gerenciamento dos metadados de cada uma das colunas das tabelas importadas. As tabelas vão

corresponder a um arquivo de cada fonte de dados. Se a conexão for direta pela base de dados, uma conexão estará vinculada a uma ou várias tabelas.

4.1.2. Processamento de dados

No módulo de processamento dos dados é possível realizar melhorias no formato e valores dos dados. Em alguns casos a redução de dimensionalidade e a discretização de dados podem melhorar a eficiência de algoritmos de aprendizado de máquina e apresentação de dados. O sistema deve permitir que o usuário realize os ajustes nos dados através de comandos SQLs ou através de uma funcionalidade, que realiza a categorização dos dados contínuos. As entradas são o nome da tabela e o nome da coluna. E a saída é um comando SQL que deverá substituir parte da consulta padrão. Conforme apresentado no SQL 2, no comando tem as máscaras `{{tabela}}` e `{{coluna}}`, no qual serão substituídas em tempo de execução pelos parâmetros. E os resultados da consulta são os quartis, demonstrado no SQL 1, que corresponde à expressão (*ExpCol*) que substitui o comportamento padrão do comando SQL de determinada tabela (`SELECT ExpCol1, ExpCol2[, ExpColN] FROM tabela`).

SQL 1 - Resultado com mascara para apresentar o dado discreto	
1.	CASE
2.	WHEN {{coluna}} <= {{RQ1}} THEN 'Q1'
3.	WHEN {{coluna}} <= {{RQ2}} THEN 'Q2'
4.	WHEN {{coluna}} <= {{RQ3}} THEN 'Q3'
5.	ELSE 'Q4'
6.	END

SQL 2 - Processamento para geração de quartis	
1.	SELECT *
2.	CASE WHEN coluna <= Q1 THEN 'Q1'
3.	WHEN coluna <= Q2 THEN 'Q2'
4.	WHEN coluna <= Q3 THEN 'Q3'
5.	ELSE 'Q4'
6.	END AS QUARTIL
7.	FROM (
8.	SELECT *,
9.	(SELECT COUNT(*)
10.	FROM (SELECT {{tpQuartil}} {{coluna}} FROM {{tabela}})) AS QTD,
11.	(SELECT {{tpQuartil}} CAST({{coluna}} AS FLOAT)
12.	FROM {{tabela}}
13.	ORDER BY 1
14.	LIMIT (((SELECT COUNT(*)
15.	FROM (SELECT {{tpQuartil}} {{coluna}}
16.	FROM {{tabela}})) / 4) * 1), 1) Q1,
17.	(SELECT {{tpQuartil}} CAST({{coluna}} AS FLOAT)
18.	FROM {{tabela}}
19.	ORDER BY 1
20.	LIMIT (((SELECT COUNT(*)
21.	FROM (SELECT {{tpQuartil}} {{coluna}}
22.	FROM {{tabela}})) / 4) * 2), 1) Q2,
23.	(SELECT {{tpQuartil}} CAST({{coluna}} AS FLOAT)
24.	FROM {{tabela}}
25.	ORDER BY 1
26.	LIMIT (((SELECT COUNT(*)
27.	FROM (SELECT {{tpQuartil}} {{coluna}}
28.	FROM {{tabela}})) / 4) * 3), 1) Q3
29.) AS T
30.	SELECT {{tpQuartil}} CAST({{coluna}} AS FLOAT) AS coluna
31.	FROM {{tabela}}
32.	ORDER BY 1
33.) AS T
34.	LIMIT 1

4.1.3. Transformação

No módulo de transformação, o usuário pode gerar modelos de dados mais otimizados e eficientes para algoritmos de máquinas. Também servindo de uma interface de fácil manipulação para geração de visões e painéis acessíveis para professores e alunos. A transformação será através de SQL, sendo que o usuário poderá escolher entre persistir os dados ou usar os SQL dinamicamente.

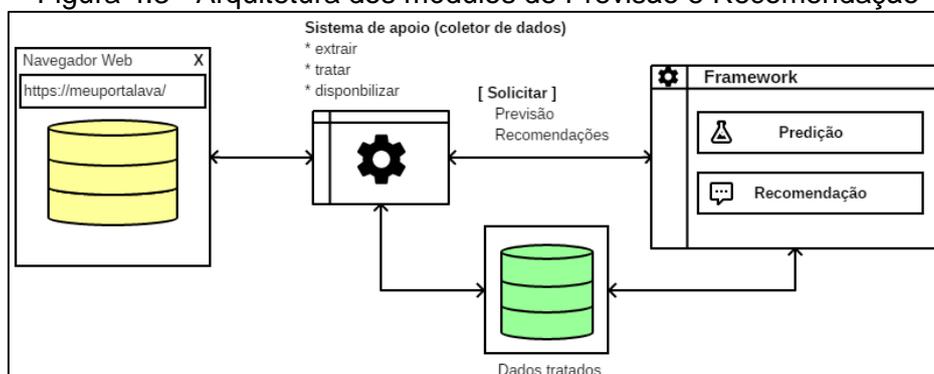
4.1.4. Reprocessamento

Este módulo é responsável pelo agendamento de extrações de dados do ambiente de aprendizado. Sendo assim, minimizará o tempo de demora entre a captura e uso dessas informações por professores. Todos os recursos de acesso aos dados deverão estar disponíveis durante o reprocessamento, dessa forma todos os dados se manterão atualizados. A periodicidade do reprocessamento fica a critério do usuário, que poderá escolher dias, semanas ou até meses.

4.2. Previsão e Recomendação

Na etapa de análise de dados ocorre o entendimento dos dados através de técnicas de predição e recomendações. Para esta fase, projetamos um framework que possui dois módulos. O primeiro faz a previsão de desempenho dos alunos e o segundo gera as mensagens de recomendações, que serão sugeridas aos professores. Conforme apresentado na Figura 4.3 - Arquitetura dos módulos de Previsão e Recomendação, ele recebe os dados acadêmicos previamente tratados, devolverá a previsão do desempenho dos alunos e sugere mensagens de recomendações.

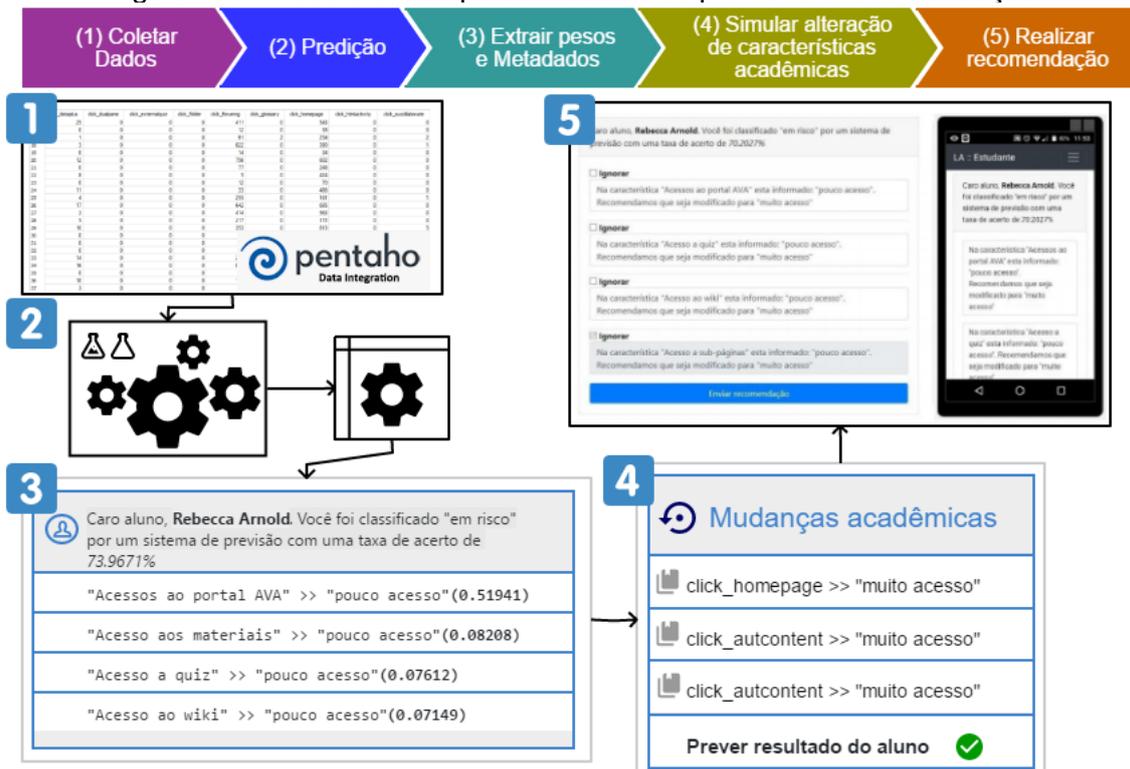
Figura 4.3 - Arquitetura dos módulos de Previsão e Recomendação



Fonte: Próprio autor.

Conforme apresentado na Figura 4.4 - Modelo usado pelos módulos de previsão e recomendação, o processo de previsão e recomendação ocorrerá através dos processos: (1) os dados deverão estar disponíveis e atualizados; (2) modelos preditivos são construídos através de algoritmos de aprendizado de máquina, o que tiver a melhor acurácia, realizará a previsão do desempenho dos alunos; (3) através de técnicas de engenharia de recursos (*features engineers*), serão extraídos os pesos das importâncias dos atributos; (4) o sistema fará as mudanças acadêmicas e reaplicará a técnica preditiva até que os alunos (antes classificados em risco) sejam classificados como sucesso acadêmico; (5) as mudanças comportamentais serão parte da mensagem de recomendação, para que professor avalie sua pertinência.

Figura 4.4 - Modelo usado pelos módulos de previsão e recomendação



Fonte: Próprio autor.

4.2.1. Predição

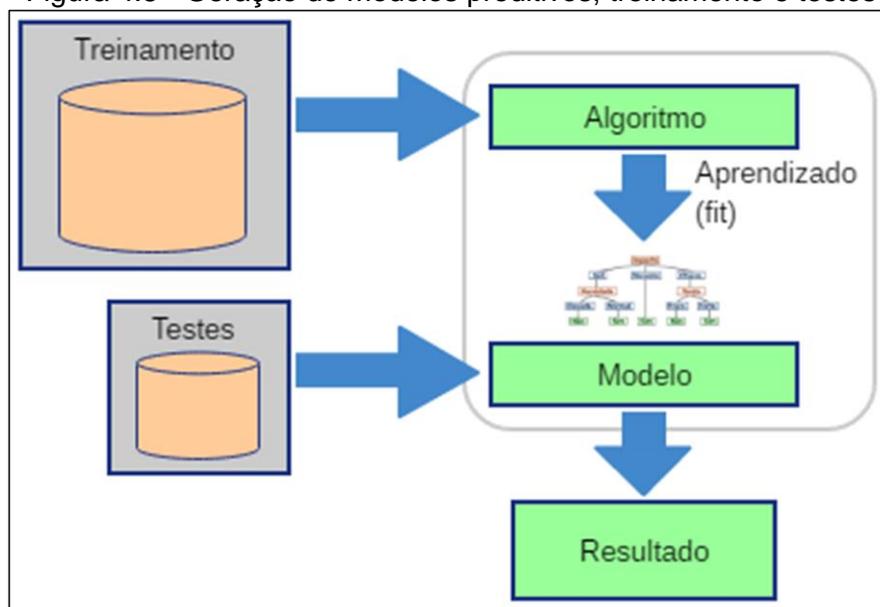
No módulo de Predição é usada a técnica preditiva da mineração de dados, com uso de algoritmos de aprendizado de máquina para obter conhecimentos fundamentais a partir das bases de dados acadêmicas. É realizada a previsão dos resultados acadêmicos dos alunos, através de algoritmos de aprendizado de máquina. Podendo classificar os alunos em “sucesso” ou “em risco” acadêmico. Na classificação “sucesso” acadêmico ocorre a previsão de que o aluno atenda aos propósitos educacionais de aprender e tirar notas boas. Já o resultado “em risco” ocorre quando o aluno dá sinais de que não está aprendendo, com probabilidade de não tirar notas boas, ou até mesmo desistir (evadir) de uma determinada disciplina ou do curso.

Para um problema específico, vários algoritmos podem ser apropriados, no entanto, um algoritmo pode ser mais adequado do que outros, mas nem sempre é possível saber com antecedência qual é a melhor opção (PAKALRA; OLPROD, 2018). Como o framework está preparado para ler diferentes estruturas de dados acadêmicos, usamos a seleção automática do algoritmo que tiver melhor acurácia, dentre as opções: Decision Tree, KNN, SVM, MLP e Naive

Bayes. Esses algoritmos foram selecionados, pois cada um tem uma abordagem diferente.

Conforme apresentado na Figura 4.5 - Geração de modelos preditivos, para montar os modelos preditivos serão usados os dados históricos de turmas anteriores para fazer o treinamento e teste dos algoritmos preditivos. Para avaliar a acurácia dos modelos usamos a técnica de validação cruzada, onde é feita a divisão dos dados já rotulados para treinamento e teste para verificar a quantidade de acertos (Carvalho et al. 2011). O algoritmo selecionado, que tiver a melhor acurácia, classificará instâncias ainda não rotuladas.

Figura 4.5 - Geração de modelos preditivos, treinamento e testes



Fonte: próprio autor.

As informações que são submetidas para montar os modelos preditivos têm diferentes importâncias. Uma técnica de engenharia de recursos (*features engineers*) é utilizada para saber quais são os atributos mais relevantes para classificação das instâncias (alunos).

Os algoritmos necessitam que os dados para treinamento e previsão sejam numéricos. Então, foi desenvolvido um mapeamento entre os valores (*mapping de-para*). No script 1 é apresentado um exemplo de *mapping* usado durante o pré-processamento do algoritmo de aprendizado de máquina.

Script 1 – Mapping preprocessing fit_transform

```

1. var mapping = {
2.   "click_dataplus": {
3.     "0": "Q1",
4.     "1": "Q2",
5.     "2": "Q3",
6.     "3": "Q4"
7.   },
8.   "click_homepage": {
9.     "0": "acesso moderado",
10.    "1": "muito acesso",
11.    "2": "pouco acesso"
12.  },
13.  "click_dualpane": {
14.    "0": "Q1",
15.    "1": "Q2",
16.    "2": "Q3",
17.    "3": "Q4"
18.  },
19.  "click_externalquiz": {
20.    "0": "Q1",
21.    "1": "Q2",
22.    "2": "Q3",
23.    "3": "Q4"
24.  }, (...)
25.  "final_result": {
26.    "0": "Falha",
27.    "1": "Sucesso"
28.  }
29. }

```

4.2.2. Recomendação

Existem as categorias de intervenções automáticas e a semiautomática. Na intervenção automática o sistema realiza ações pré-determinadas automaticamente com pouca ou nenhuma participação. Já as intervenções semiautomáticas são alertas e outros indicadores relevantes dos padrões de aprendizado enviados para os envolvidos no processo de ensino (BROWN, 2012). Optamos pela intervenção semiautomática para que o professor tenha a oportunidade de visualizar e alterar as mensagens antes de enviá-las ou simplesmente as ignorar caso não sejam apropriadas. Ou seja, o sistema gera mensagens candidatas de recomendações para que o professor julgue a pertinência delas antes do envio aos alunos. Também poderá alterar feedback com alguma informação que melhore e guie seus estudos.

As mensagens são geradas através das métricas derivadas do processamento do módulo de aprendizado de máquina, que são as características mais relevantes para a classificação das instâncias. Elas são extraídas através da técnica de engenharia de recursos (*features engineers*), chamada *feature_importances_*. Usamos cinco características mais relevantes inicialmente, podendo aumentar a quantidade, caso necessário. Cada

mensagem terá recomendações de mudanças no comportamento (características), que comprometem o desempenho acadêmico. Essas características correspondem aos dados coletados no ambiente de aprendizado, que podem ser desde informações sobre o uso do ambiente, como também dados demográficos. Contudo, usamos apenas dados sobre o uso do ambiente de aprendizado.

Algoritmo 1. Gerar mensagem - Características	
1.	Entrada: instanciaAluno
2.	Variáveis: conta, ResultadoAluno, msgRecomendação, listaCaracterística, característica, itemMsg;
3.	listaCaracterística, característica, itemMsg;
4.	Início
5.	ResultadoAluno = PreverResultado(instanciaAluno);
6.	Se ResultadoAluno igual 'sucesso' Então:
7.	sair();
8.	msgRecomendação = [];
9.	listaCaracterística = ListarCaracPorImportancia();
10.	Para característica em listaCaracterística:
11.	MudarCaracterística(instanciaAluno, característica);
12.	itemMsg = GerarItemMsg(instanciaAluno, característica);
13.	AdicionarMsg(msgRecomendação, itemMsg);
14.	conta = conta + 1;
15.	Se conta maior 5 Então:
16.	ResultadoAluno = PreverResultado(instanciaAluno);
17.	Se ResultadoAluno igual 'sucesso' Então:
18.	Parar
19.	
20.	Retornar msgRecomendação
21.	Fim

O algoritmo 1 demonstra que, para cada aluno classificado “em risco”, é realizada a validação das características (valores dos atributos). Nessa validação, caso o valor preenchido não corresponda ao da maioria dos estudantes classificados em “sucesso” acadêmico (estatística simples), o sistema altera o valor desse atributo. Usamos as cinco características mais relevantes. Após essas alterações, o sistema submete a instância (aluno) novamente à validação do algoritmo preditivo até o classificar em “sucesso” acadêmico, podendo incrementar a quantidade de características alteradas.

4.2.3. Mensagens candidatas para avaliação dos Professores

O professo visualizará somente informações de sua turma, listando apenas os alunos que estão matriculados em suas disciplinas. Ao clicar em cada aluno da lista, visualizará detalhes sobre o registro do aluno com suas informações e como o sistema o classificou. Se o aluno foi classificado “em risco”, o sistema sugere uma mensagem candidata com as alterações comportamentais para que o professo avalie sua pertinência antes do envio.

As mensagens candidatas geradas pelo sistema poderão ser visualizadas, alteradas e/ou ignoradas pelo professor. A mensagem candidata apresentada ao professor tem informações fixas como a taxa estimada de acerto, descrição e valor da característica. O professor também poderá visualizar informações técnicas sobre a classificação de cada aluno, bem como o peso de referência e informações estatísticas através de gráficos para cada atributo do modelo.

4.2.4. Recomendações aos Alunos

O aluno receberá as mensagens e ao abri-la, o professor saberá que ela foi lida pelo aluno. Com estas mensagens o aluno poderá ter uma reflexão sobre seu comportamento acadêmico e se está de acordo com a proposta acadêmica estabelecida pelo professor.

4.3. Apresentação

O sistema dispõe de um mecanismo de visualização de dados. O professor poderá visualizar graficamente informações sobre o andamento da turma. E o aluno poderá visualizar seus dados e as recomendações enviadas pelo professor. Será possível configurar várias visões e agrupadas em painéis que serão disponibilizados para os professores e alunos. As visões são configuráveis em: gráfico de pizza, gráfico de barras, gráfico de linha e listagem em grade (*Grid*) e informações em cartões (*Cards*).

4.3.1. Montar Visões e Painéis

No módulo de montagem de visão o administrador do sistema poderá gerar visões e associá-las a um painel de um professor ou de alunos. Esses painéis estarão acessíveis para o professor e o sistema deverá efetuar os filtros necessários para que só visualize os dados de suas turmas.

Cada uma dessas visões poderá corresponder uma forma específica de visualizar determinado dado. As formas de visualização desses dados poderão ser: Gráfico de Pizza; Gráfico Barras; Gráfico Linha; Informação em Cards; e Informação em Grid. Na montagem das visões, basta selecionar a tabela e as colunas. As colunas poderão estar vinculadas aos gráficos de formas diferentes, como Séries, Valores e Categorias, dependendo do tipo de gráfico.

5. Prova de Conceito

A arquitetura conceitual com descrições abstratas sobre a solução computacional é apresentada no capítulo anterior. Este capítulo tem por finalidade complementar essas informações com detalhes de projeto de software, descrevendo os elementos utilizados para o desenvolvimento. A solução foi projetada usando o Astah Community ², com os diagramas UML de Casos de Uso, de Classes e de Componentes. A descrição completa do projeto de software é apresentada no Apêndice A. Na seção 5.1 é apresentado o diagrama de caso de uso com as funcionalidades da solução. Na seção 5.2 é apresentado o diagrama de classes. Na seção 5.3 é apresentado o diagrama de componentes com informações sobre a arquitetura interna. Na seção 5.4 são apresentadas as caracterizações das interfaces com os usuários. E na seção 5.5 são apresentados os recursos utilizados na produção do protótipo.

5.1. Casos de Uso

Nesta seção, as funcionalidades do sistema são descritas em casos de uso através de um mapeamento comportamental do sistema. Na definição da solução de projeto de software, mapeamos três atores principais. Conforme apresentado na Figura 5.1 - Atores / Perfil, e descritos na Tabela 5.1 - Atores / Perfil os três atores são: Administrador; Aluno e Professor. Cada ator terá uma forma diferente de usar o sistema. O ator “Pessoa”, no qual todos os outros herdarão dele, corresponde a um perfil genérico que realiza atividades básicas e comuns no sistema.

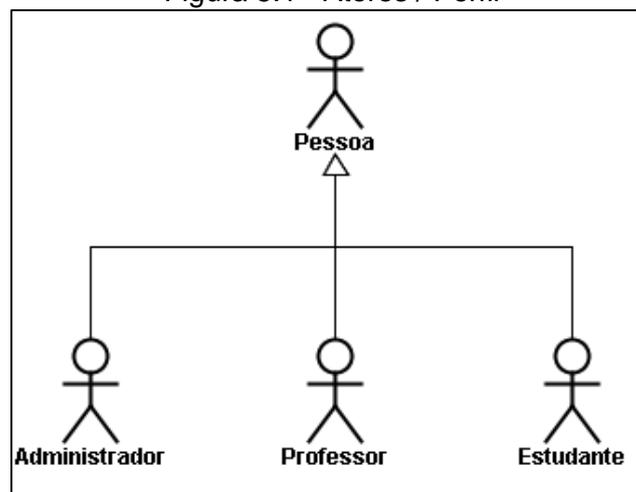
² <http://astah.net/editions/community>

Tabela 5.1 - Atores / Perfil

Nome	Descrição
Pessoa	Perfil genérico que atua de forma comum a todos os outros perfis de usuário.
Administrador	Perfil responsável por configurar o sistema.
Estudante	Perfil dos alunos que receberão as recomendações e que visualize seus dados através de painéis.
Professor	Perfil do professor que deverá julgar as mensagens de recomendações dos alunos e visualizar dados de sua turma através de painéis.

Fonte: Próprio autor.

Figura 5.1 - Atores / Perfil



Fonte: Próprio autor.

Para mapear os Casos de Uso, todos os perfis deverão herdar do ator Pessoa. Pois, haverá casos de uso que serão comuns a todos os atores do sistema. No diagrama de casos de uso, apresentado na Figura 5.2 - Diagrama de Casos de Uso, estão mapeadas as funcionalidades do sistema usando. E na Tabela 5.2 - Casos de Usos, estão descritos os casos de uso.

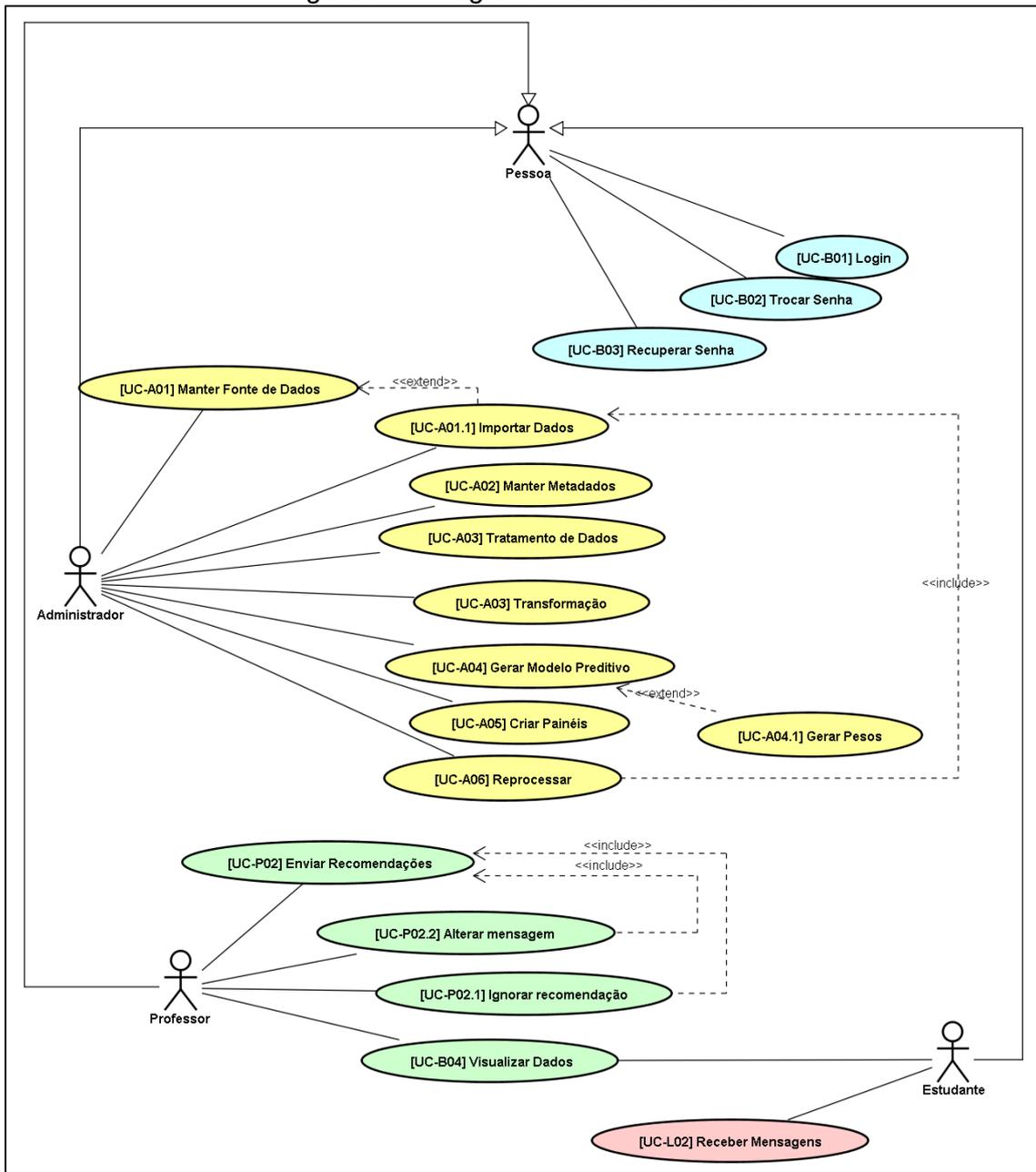
Tabela 5.2 - Casos de Usos

Nome	Descrição
[UC-A01.1] Importar Dados	O sistema deverá permitir que o usuário importe dados acadêmicos para dentro da base de dados.
[UC-A01] Manter Fonte de Dados	O sistema deverá permitir que o usuário as várias fontes de dados do AVA.
[UC-A02] Manter Metadados	O sistema deverá permitir que o usuário mantenha os Metadados.
[UC-A03] Processamento	O sistema deverá permitir que o usuário processe os dados.
[UC-A03] Transformação	O sistema deverá permitir que o usuário execute transformações eficientes e otimizadas nos dados.
[UC-A04] Gerar Modelo Preditivo	O sistema deverá permitir que o usuário gere modelos

Nome	Descrição
	preditivos nos dados.
[UC-A05] Criar Painéis	O sistema deverá permitir que o usuário crie painéis para professores.
[UC-A06] Reprocessar	O sistema deverá permitir que o usuário reprocessasse os dados acadêmicos.
[UC-B01] Login	O sistema deverá permitir que o usuário efetue login no sistema.
[UC-B02] Trocar Senha	O sistema deverá permitir que o usuário troque de senha.
[UC-B03] Recuperar Senha	O sistema deverá permitir que o usuário recupere a senha.
[UC-B04] Visualizar Dados	O sistema deverá permitir que professores e alunos visualizem os dados acadêmicos através dos painéis gerados. O professor somente terá acesso às suas turmas. E o aluno somente terá acesso aos seus dados acadêmicos.
[UC-L02] Receber Mensagens	O sistema deverá permitir que o aluno receba as recomendações dos professores.
[UC-P02.1] Ignorar recomendação	O sistema deverá permitir que o professor ignore recomendações.
[UC-P02.2] Alterar mensagem	O sistema deverá permitir que o professor altere a mensagem de recomendação.
[UC-P02] Enviar Recomendações	O sistema deverá permitir que o professor envie a mensagem de recomendações.

Fonte: Próprio autor.

Figura 5.2 - Diagrama de Casos de Uso



Fonte: Próprio autor.

5.2. Classes

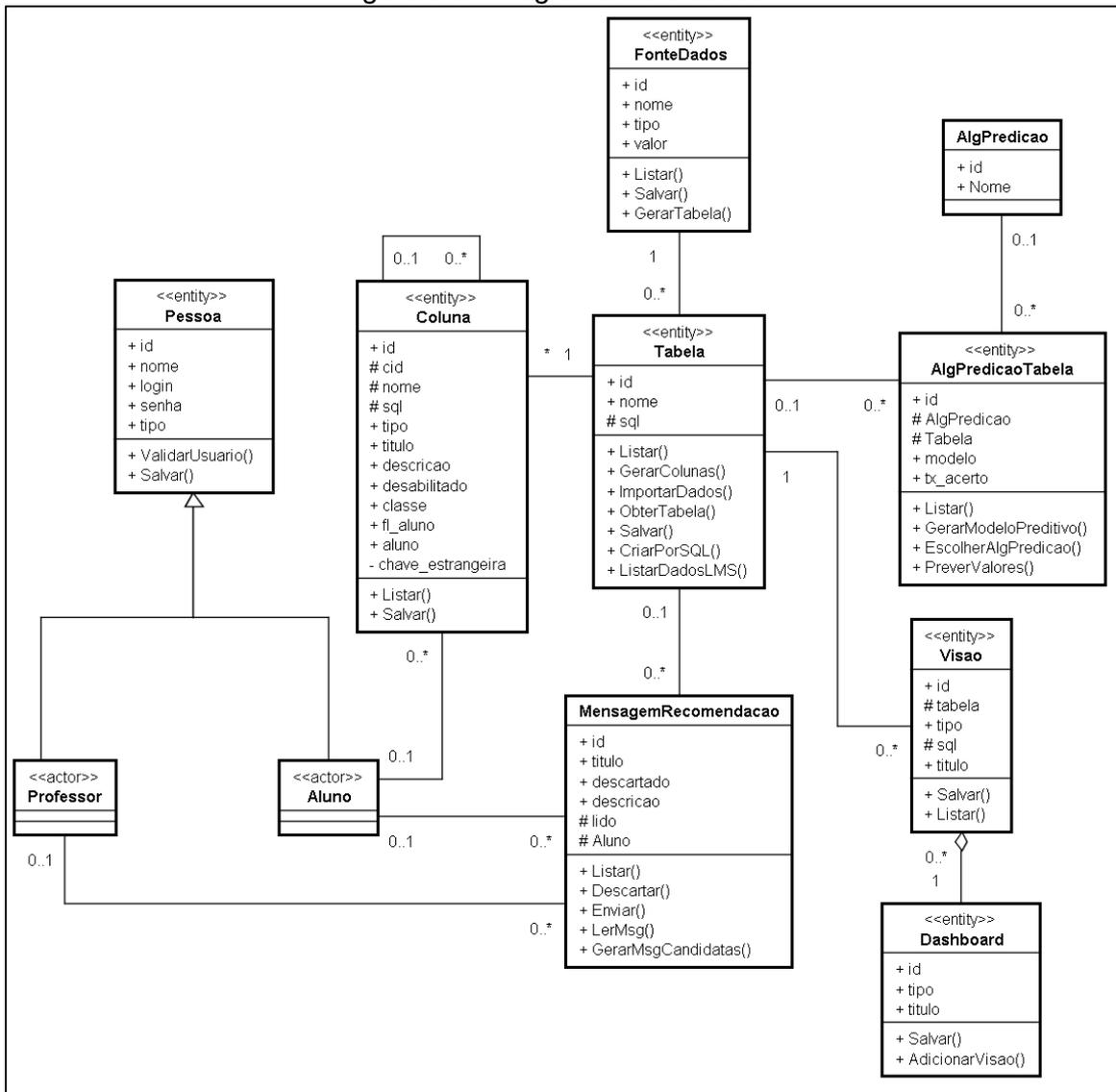
Nesta seção são descritos os componentes estruturais mapeados no sistema através das entidades de negócio. Na Tabela 5.3 - Classes mapeadas, contém as classes com suas descrições. Já na Figura 5.3 - Diagrama de Classes, contém o diagrama de classes.

Tabela 5.3 - Classes mapeadas

Nome	Tipo	Descrição
Administrador	Actor	Cls: Pessoa do tipo Administrador
AlgPredicao	Class	Tabela de domínio do analista. Algoritmos de Predição: * Decision Tree-based * K-nearest Neighbors * Support Vector Machine * Multi-layer Perceptron Classifier * Nayve Bayes
AlgPredicaoTabela	Entity	Algoritmos vinculados à tabela
Aluno	Actor	Cls: Pessoa do tipo Aluno
Coluna	Entity	Metadados - colunas das tabelas importadas
Dashboard	Entity	Painel do Professor ou Aluno
FonteDados	Entity	Fonte de dados proveniente dos dados do AVA
MensagemRecomendacao	Class	Mensagem gerada para recomendação aos alunos
Pessoa	Entity	Pessoa do sistema
Professor	Actor	Cls: Pessoa do tipo Professor
Tabela	Entity	Entidade: Tabela
Visao	Entity	Parte do painel (do aluno ou professor)

Fonte: Próprio autor.

Figura 5.3 - Diagrama de Classes



Fonte: Próprio autor.

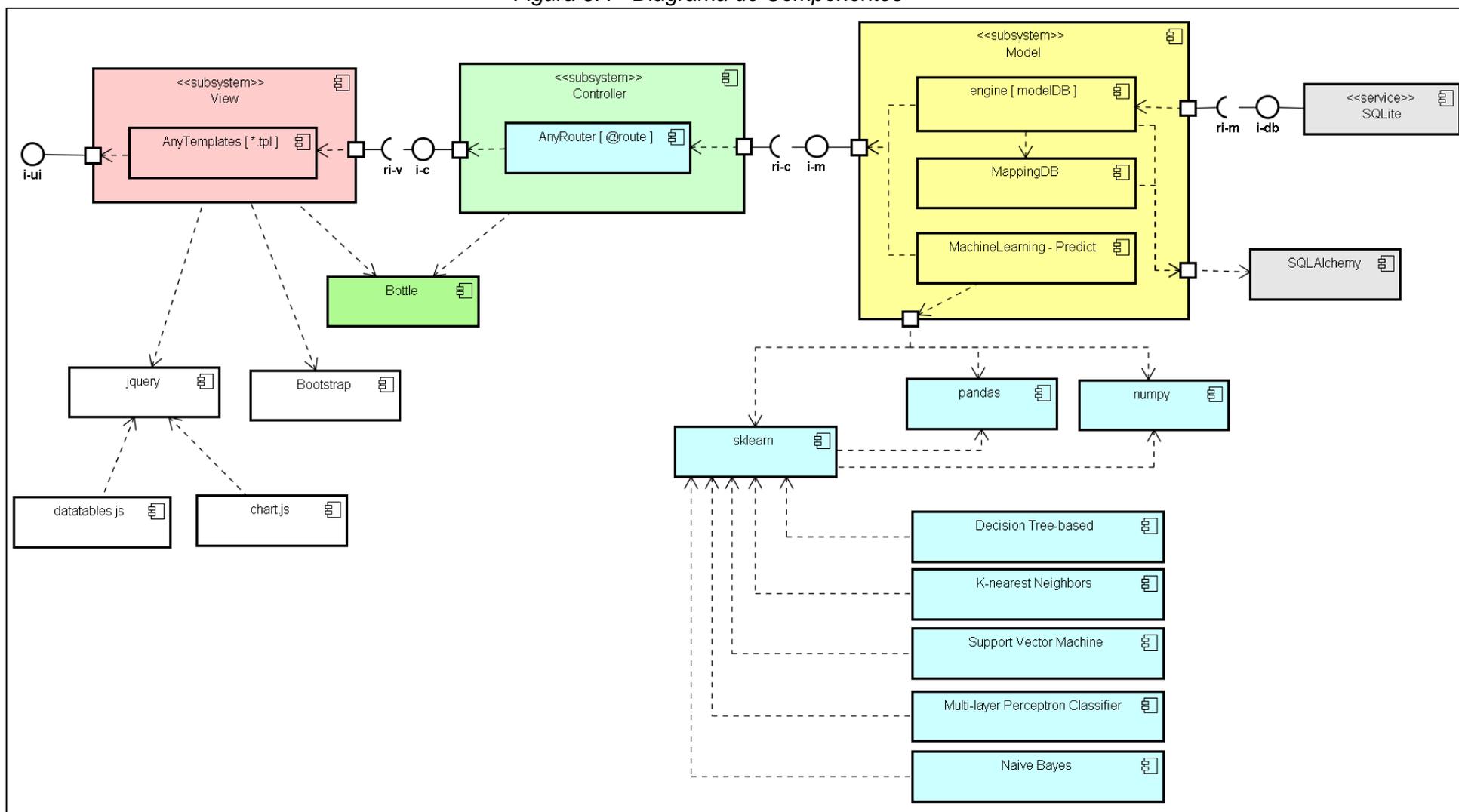
5.3. Diagrama de Componentes

No diagrama de componentes são descritos como os artefatos estão relacionados através dos subsistemas, bibliotecas, frameworks, dependências e interfaces. Conforme apresentado na Figura 5.4 - Diagrama de Componentes e listado na tabela Tabela 5.4 - Elementos do diagrama de componentes.

Tabela 5.4 - Elementos do diagrama de componentes

Nome	Type	Descrição
AnyRouter [@route]	Component	Rotas mapeadas pelo @route
AnyTemplates [*.tpl]	Component	Templates (Visão)
Bootstrap	Component	Biblioteca JS
Bottle	Component	Framework Web - Python
Controller	Component	Subsistema Controller
Decision Tree-based	Component	ML: Árvore de Decisão
K-nearest Neighbors	Component	ML: KNN
MachineLearning - Predict	Component	Módulo de Predição
MappingDB	Component	Mapeamento Entidade Relacionamento x Objeto
Model	Component	Subsistema Model Todas as regras de acesso a dados e ORM
Multi-layer Perceptron Classifier	Component	ML: MLP
Naive Bayes	Component	ML: Naive Bayes
SQLAlchemy	Component	Framework ORM de acesso ao banco de dados
SQLite	Component	Banco de Dados
Support Vector Machine	Component	ML: SVM
View	Component	Subsistema Visão
chart.js	Component	Biblioteca JS
datatables.js	Component	Biblioteca JS
engine [modelDB]	Component	Mecanismo de acesso ao Banco de Dados
i-c	Interface	Interface provida pelo Controller para a View
i-db	Interface	Interface provida pelo Banco de Dados para o Model
i-m	Interface	Interface provida pelo Model para o Controller
i-ui	Interface	Interface provida para o usuário
Jquery	Component	Biblioteca JS
numpy	Component	Biblioteca para manipulação de dados N-Dimensional
pandas	Component	Biblioteca de tratamento de dados usando DataFrames
ri-c	Interface	Interface requerida pelo Controller para o Model
ri-m	Interface	Interface requerida pelo Model para o Banco de Dados
ri-v	Interface	Interface requerida pela View para o Controller
sklearn	Component	Biblioteca de aprendizado de Máquina

Figura 5.4 - Diagrama de Componentes

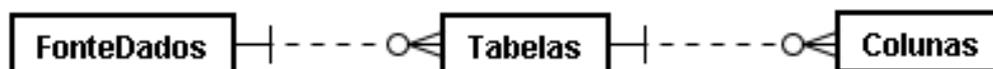


Fonte: Próprio autor.

5.4. Arquitetura Acoplável

Com o aumento das ofertas de recursos tecnológicos, faz-se necessário a construção de aplicações que possa usar e serem utilizadas. Isso facilita para que desenvolvedores criem novas partes (módulos) de um software e acoplem em determinada solução. Desta forma, a solução computacional desta dissertação **não só** foi projetada para ser usada como parte de um ambiente de aprendizado, **como também** utilizar ferramentas externas. Sendo assim, é possível utilizar de forma acoplada a ferramenta de integração chamada o *Pentaho Data Integration (PDI)* e a ferramenta de visualização de dados chamada de *Tableau*. Para o uso do Pentaho é necessário que ele integre diretamente na base de dados do sistema registrando os dados e metadados. Conforme apresentado na Figura 5.5 - Metadados, os metadados correspondem à fonte de dados, tabelas e colunas. Ele deverá, ao mesmo tempo em que importa os dados, registrar os metadados e criar os índices dos bancos.

Figura 5.5 - Metadados



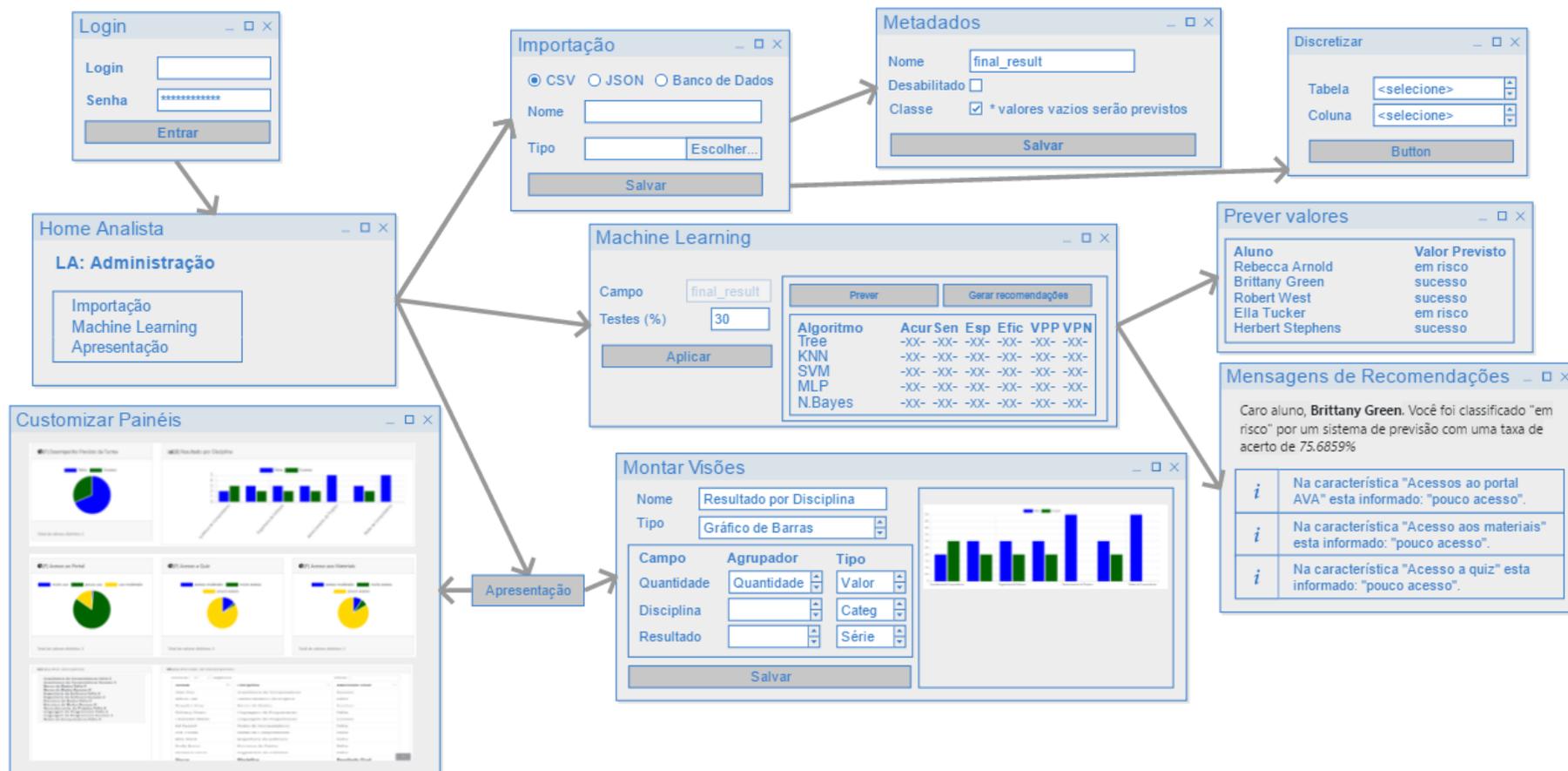
Fonte: Próprio autor.

5.5. Caracterização de Interfaces

A caracterização de interfaces foi projetada com a possibilidade de demonstrar recursos tecnológicos de características ainda não implementadas e exemplificando funcionalidades e fluxos. Foram configuradas as interfaces para os perfis de Analista, Professor e Aluno.

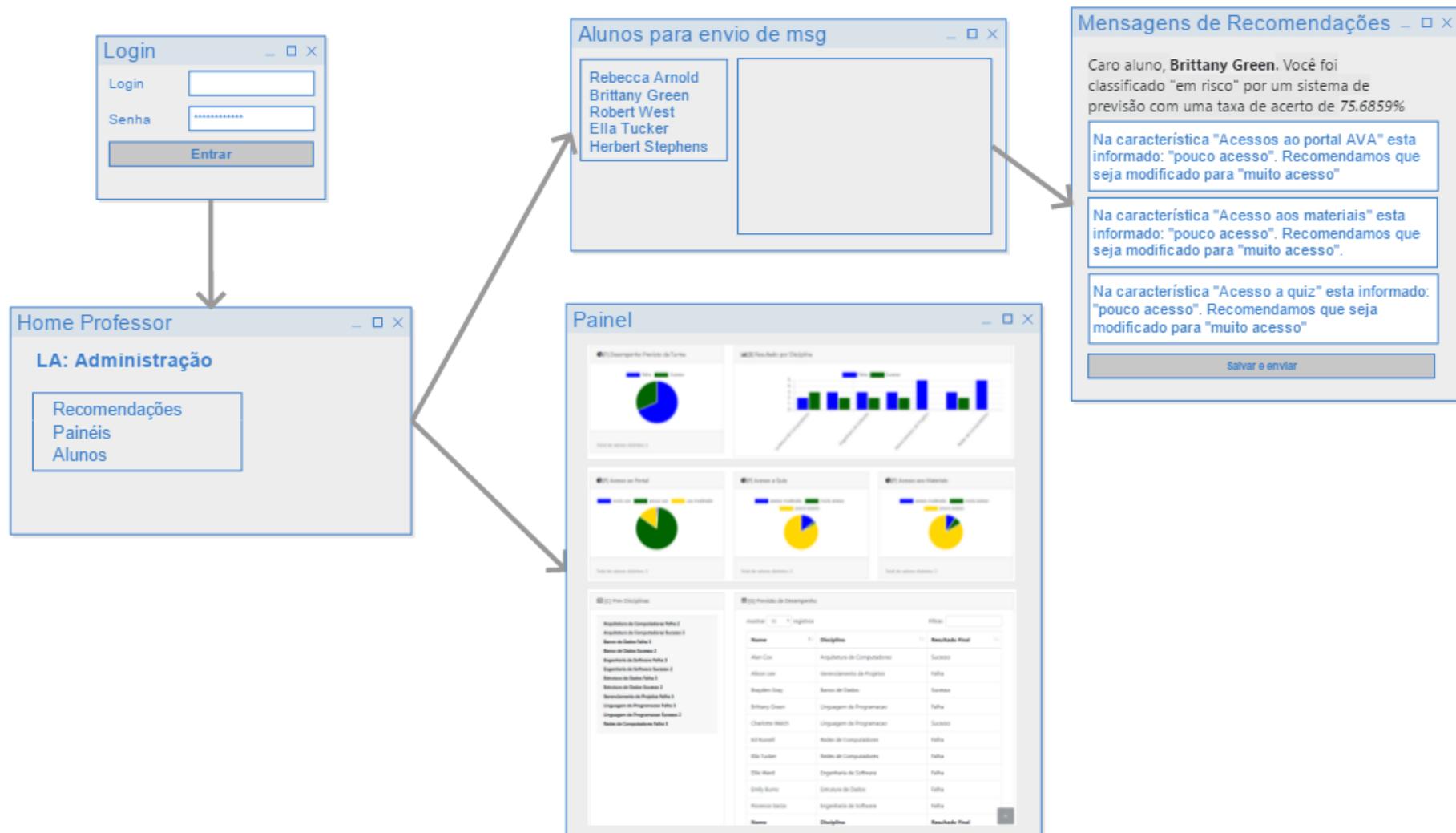
Conforme apresentado na Figura 5.6 - Caracterização das telas - perfil Analista, são descritas as funcionalidades do perfil Analista. As funcionalidades são: Importação de dados; gerenciamento de metadados; gerar modelos preditivos; prever resultado dos alunos; geração de mensagens; criação de visões; e customização de painéis.

Figura 5.6 - Caracterização das telas - perfil Analista



Fonte: próprio autor.

Figura 5.7 - Caracterização das telas - perfil Professor



Fonte: próprio autor.

Conforme apresentado na figura Figura 5.7 - Caracterização das telas - perfil Professor, o professor terá acesso à sua turma, selecionar cada aluno, verificar qual mensagem de recomendação sugerida pelo sistema, ajustar (conforme desejado), enviar ao aluno a mensagem e visualizar seus painéis. E, por fim, na Figura 5.8 - Caracterização das telas - perfil aluno, é demonstrada a visualização da recomendação recebida pelo aluno.

Figura 5.8 - Caracterização das telas - perfil aluno



Fonte: próprio autor.

5.6. Recursos do Protótipo

Os recursos tecnológicos para o desenvolvimento do protótipo são os descritos abaixo. Usamos a linguagem de programação Python ³. Usamos várias bibliotecas e frameworks. Na Figura 5.4 - Diagrama de Componentes, é apresentado o diagrama de componentes implementado em nosso protótipo, que descreve os subsistemas, comunicações, componentes, interfaces e dependências. No subsistema View contém os arquivos de templates (*.tpl) e usamos componentes visuais de Javascript e CSS,

³ <https://www.python.org/>

como JQuery⁴, Bootstrap⁵, Datatable.js e Chart.js. No subsistema Controller são mapeadas as rotas pelo framework web Bottle ⁶. Cada uma delas corresponde às requisições feitas ao sistema. No subsistema Model contém as entidades mapeadas pelo ORM SQLAlchemy ⁷, um framework de persistência de banco de dados. Contém um componente de software chamado Predict, que utiliza as bibliotecas: Pandas ⁸ para tratar dos dataframes; Numpy ⁹ para manipular dados genéricos e multidimensionais de forma eficiente; e o framework sklearn ¹⁰ para uso dos algoritmos de aprendizado de máquina classificados (Decision Tree, KNN, SVM, MLP e Naive Bayes). Usamos para armazenar os dados o banco de dados SQLite ¹¹. Assim, hospedamos o sistema no PythonAnywhere ¹², devido a facilidade de hospedagem da solução e instalação de bibliotecas Python.

⁴ <https://jquery.com/>

⁵ <https://getbootstrap.com/>

⁶ <https://bottlepy.org/>

⁷ <https://www.sqlalchemy.org/>

⁸ <https://pandas.pydata.org/>

⁹ <http://www.numpy.org/>

¹⁰ <https://scikit-learn.org/stable/>

¹¹ <https://www.sqlite.org/>

¹² <https://www.pythonanywhere.com/>

6. Aplicação da Prova de Conceitos

Este capítulo tem como objetivo comprovar que a implementação do modelo proposto é viável. Também descrevemos todos os passos necessários para criar um cenário de uso da ferramenta. A prototipagem fez parte do desenvolvimento da solução de software desta dissertação, caracterizando as funcionalidades finais do software.

6.1. Descrição da prova de conceito

Na finalidade de demonstrar que a proposta é viável, o protótipo funcional foi aplicado em um contexto de análise de dados utilizando o conjunto de dados OULA, descrito na seção 6.2. Os dados foram importados e ajustados, conforme descrito na seção 6.3. Usamos três perfis diferentes para acessar a aplicação, conforme apresentado na Tabela 6.1 - Perfil de Usuários. Os perfis de usuários usados são: Analista (bruno), Professor (stoll); e Aluno (260355). O perfil Analista corresponde a um perfil administrativo, que realiza as parametrizações iniciais e de ajustes quando necessário. Já o perfil Professor corresponde ao usuário que utilizará o sistema, na finalidade de analisar os dados, visualizar/ajustar recomendações geradas para os alunos. E, por fim, o perfil Aluno corresponde ao usuário que acessará o sistema visualizando as recomendações enviadas para ele.

Tabela 6.1 - Perfil de Usuários

Perfil	Usuário
Analista	bruno
Professor	stoll
Aluno	260355 (Rebecca Arnold)

Fonte: Próprio autor.

Para cada um dos usuários descritos acima foi utilizada a funcionalidade de Login, apresentada na Figura 6.1 - Login do Sistema. Em que o usuário deverá informar as credenciais para ter acesso ao sistema e o sistema deverá checar qual é o seu perfil.

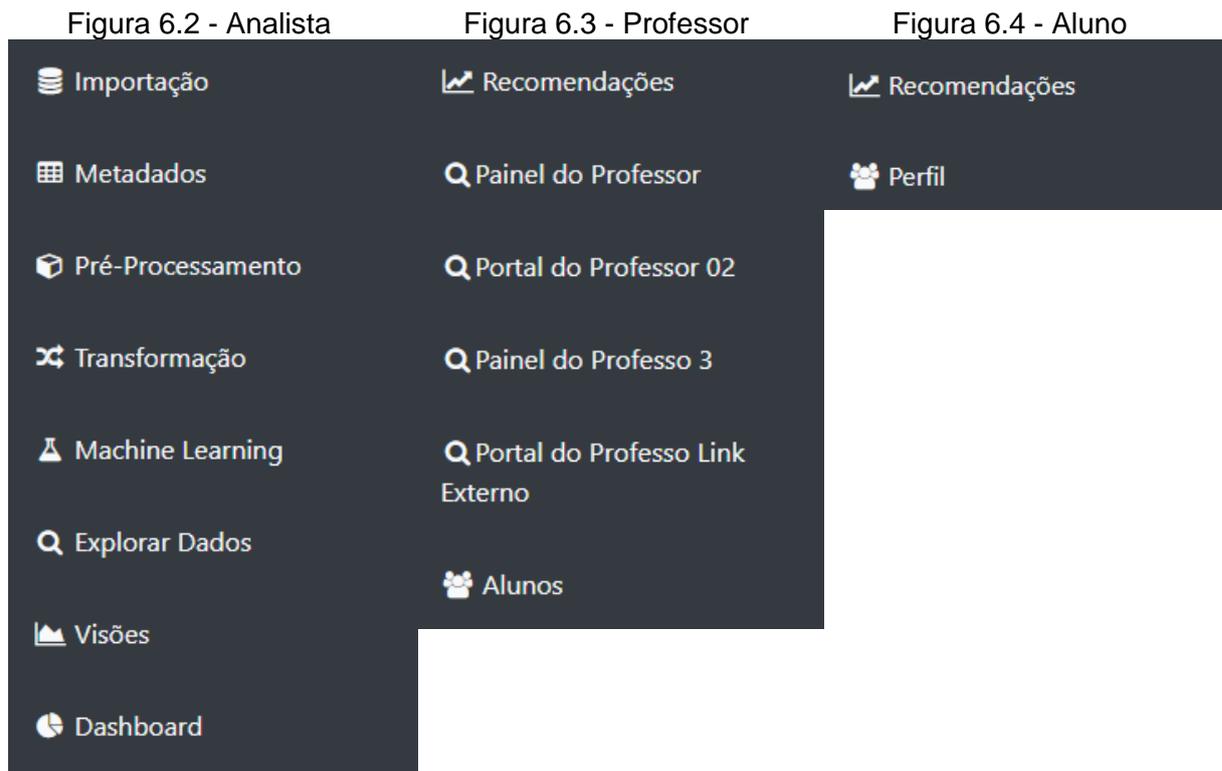
Figura 6.1 - Login do Sistema

A imagem mostra uma interface de login com o seguinte layout:

- Um cabeçalho com o texto "Login".
- Um campo de entrada rotulado "Login" com o placeholder "Informe login ou email".
- Um campo de entrada rotulado "Senha" com o placeholder "Entre com a senha".
- Um botão azul com o texto "Entrar".

Fonte: Próprio autor.

Ao entrar no sistema, caso o usuário seja do tipo Analista (Figura 6.2 - Analista) o sistema apresentará o menu com as opções: Importação; Metadados; Pré-Processamento; Transformações; Aprendizado de Máquina; Exploração; Visão; e Painel. Se o perfil do usuário for de Professor (Figura 6.3 - Professor) o sistema apresentará o menu com as opções: Recomendações; Lista de Painéis; e Alunos. E se o perfil do usuário for de Aluno (Figura 6.4 - Aluno) o sistema apresentará o menu com as opções: Recomendações e Perfil.



Fonte: Próprio autor

6.2. Descrição do Dataset OULA

O conjunto de dados OULA, *Open University Learning Analytics*, disponível no site *UCI Machine Learning Repository*¹³, foi disponibilizado para apoiar nas pesquisas relacionadas à análise de dados educacionais, contendo informações de cursos apresentados na Universidade Aberta (KUZILEK et al. 2017). O que torna o conjunto de dados único é o fato de conter dados demográficos e dados de fluxo de cliques agregados das interações dos alunos no AVA. Isso possibilita a análise do comportamento do aluno, representado por suas ações. O conjunto de dados contém as informações sobre 22 cursos, 32.593 alunos com seus resultados de avaliação e registros de suas interações com o AVA representados por resumos diários de cliques de alunos (10.655.280 registros de interação com o AVA). Esses dados são comumente usados em competições pela comunidade LAK, *Learning Analytics and*

¹³ <http://archive.ics.uci.edu/ml/>

Knowledge. No apêndice B contém todas as informações disponíveis do conjunto de dados.

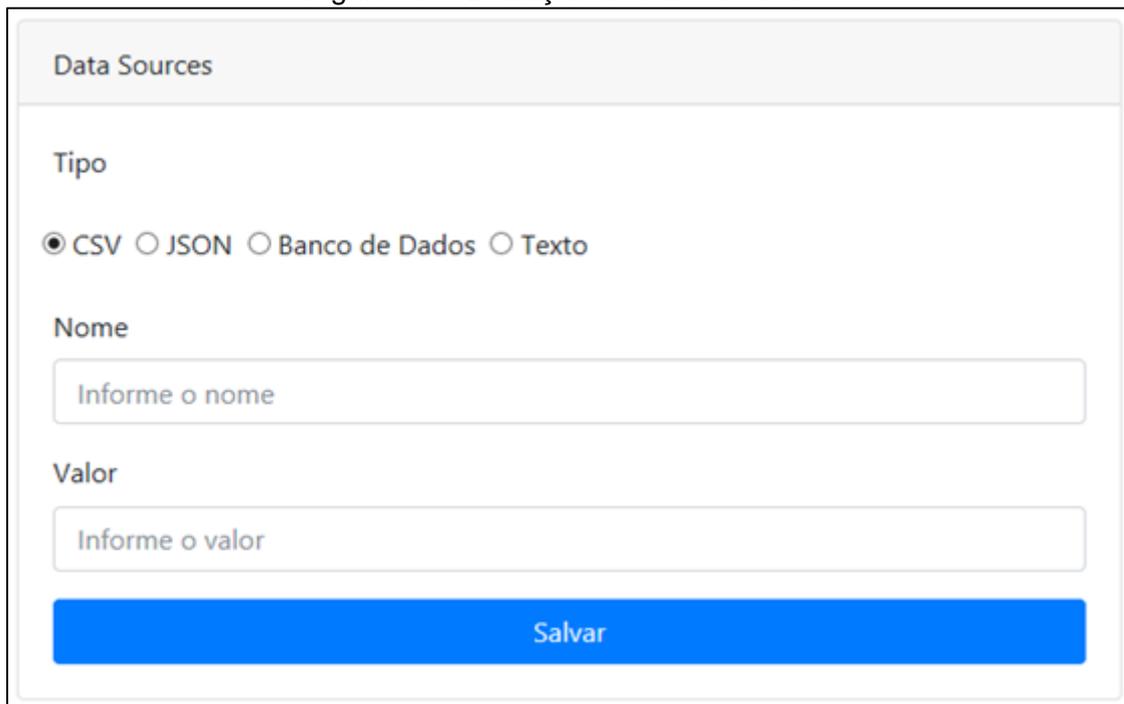
6.3. Carga e Tratamento

Com o perfil de usuário igual a Analista o sistema deverá apresentar a opção de Importação. Ao selecionar a opção o sistema apresentará a tela para que o usuário tenha a opção de registrar uma fonte de dados que proverá uma ou várias tabelas que poderão ser manipuladas dentro do sistema.

6.3.1. Importação dos Dados

Através do módulo de Coleta de Dados fizemos a importação dos arquivos disponíveis no conjunto de dados públicos. Na Tabela 6.2 - Arquivos / Tabelas do OULA estão descritas as informações sobre cada um dos arquivos importados. Além de informarmos ao sistema qual é o atributo classe para que o sistema possa prever os valores quando não estiver preenchido.

Figura 6.5 - Definição da fonte de dados



O formulário, intitulado "Data Sources", contém os seguintes campos e opções:

- Tipo:** Um grupo de botões de opção com as alternativas: CSV, JSON, Banco de Dados, e Texto.
- Nome:** Um campo de texto com o placeholder "Informe o nome".
- Valor:** Um campo de texto com o placeholder "Informe o valor".
- Salvar:** Um botão de ação de cor azul.

Fonte: Próprio autor.

Cada um dos arquivos CSV do conjunto de dados OULA foi importado usando a definição da fonte de dados, conforme Figura 6.5 - Definição da fonte de dados. Após isso, foi realizada a importação da tabela (arquivo) para o sistema, conforme Figura 6.6 - Selecionar tabela para importar.

Figura 6.6 - Selecionar tabela para importar

Fonte: Próprio autor.

Tabela 6.2 - Arquivos / Tabelas do OULA

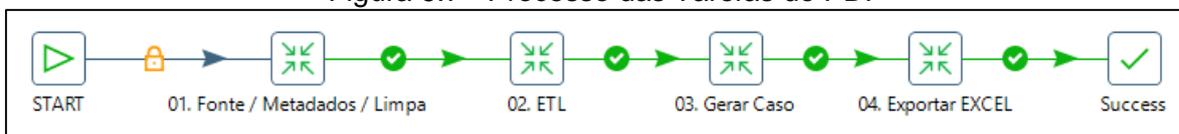
Tabela	Descrição
assessments.csv	Informações sobre avaliações dos módulos.
courses.csv	Lista de todas as disciplinas disponíveis.
studentAssessment.csv	Resultados das avaliações dos alunos.
studentInfo.csv	Informações demográficas sobre os alunos, juntamente com os seus resultados.
studentRegistration.csv	Registro do aluno no módulo.
studentVle.csv	Informações sobre as interações de cada aluno com os materiais no AVA.
vle.csv	Informações sobre os materiais disponíveis no AVA.

Fonte: Próprio autor.

6.3.2. Usando ferramentas externas para Integrar de Dados

Conforme apresentado na Figura 6.7 - Processo das Tarefas do PDI, a importação que foi realizada usando o Pentaho, com os seguintes passos: 01. Fonte / Metadados / Limpa; 02. ETL; 03. Gerar Caso; e 04. Exportar EXCEL.

Figura 6.7 - Processo das Tarefas do PDI



Fonte: Próprio autor.

No passo 01 é feita a geração da fonte de dados, registro dos metadados e limpeza do banco, caso não seja a primeira importação. O registro do metadado da tabela “*__dados_finais*” é registrado no SQL 3 - Gerar Tabela. No SQL 4 - Registrar metadados das colunas é apresentada a importação da coluna “*code_module*”. E no SQL - 5 Limpar dados contém o comando de limpar a tabela “*__dados_finais*”.

Figura 6.8 - Fonte / Metadados / Limpa



Fonte: Próprio autor.

SQL 3 - Gerar tabela

```
1. INSERT INTO tabela
2.     (id_fonte_dados, nome,
3.      descricao, sql_origem, sql_destino, sql_sem_hist)
4. SELECT 100, '__dados_finais', '', '', '', ''
5. WHERE NOT EXISTS (SELECT 1 FROM tabela WHERE nome = '__dados_finais');
```

SQL 5 - Limpar dados

```
1. DELETE
2. FROM __dados_finais;
```

SQL 4 - Registrar metadados das colunas

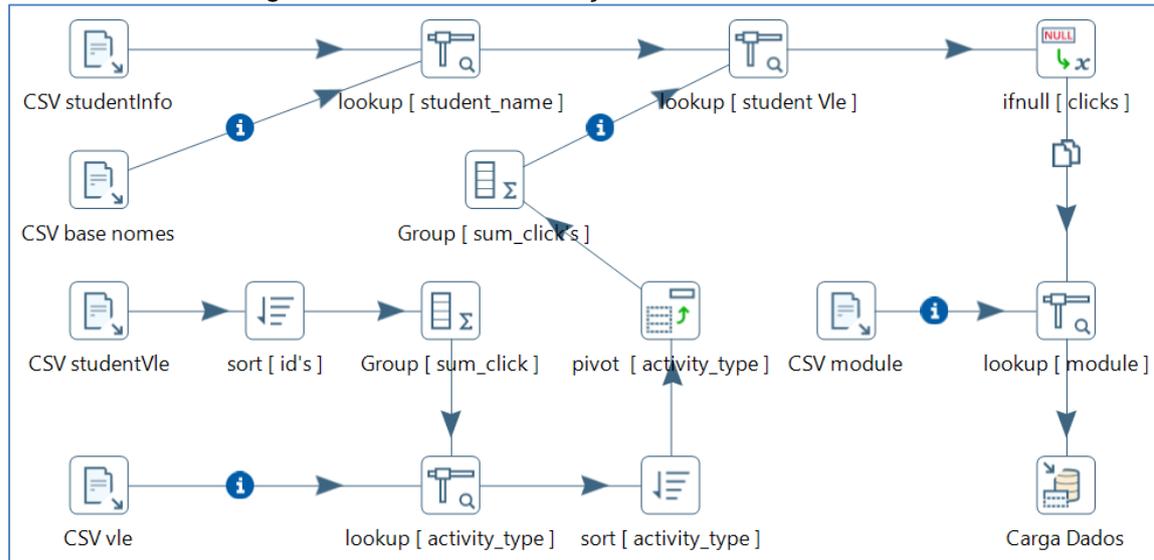
```

1.  INSERT OR IGNORE INTO coluna (id_coluna_cid, id_tabela, nome, sql, tipo, titulo,
2.  descricao, desabilitado, classe, fl_aluno, chave_estrangeira, id_tabela_fk,
3.  id_coluna_fk, val_aluno_risco)
4.      SELECT 1, (SELECT id FROM tabela WHERE nome = '__dados_finais' LIMIT 1),
5.      'code_module',
6.      'code_module',
7.      'T',
8.      'code_module',
9.      '',
10.     0,
11.     0,
12.     0,
13.     0,
14.     null,
15.     null,
16.     ''
17.  WHERE NOT EXISTS
18.     (SELECT '1'
19.      FROM coluna
20.      WHERE id_tabela = (SELECT id
21.                        FROM tabela
22.                        WHERE nome = '__dados_finais'
23.                        LIMIT 1)
24.      AND nome = 'code_module')

```

Na etapa 02 ETL é feita a importação da tabela “__dados_finais”, que é usada nos módulos de predição e recomendação. Conforme apresentado na Figura 6.9 - Fluxo de extração e tratamento Pentaho, buscamos os dados das tabelas “studentInfo”, “studentVle”, “vle”, dados da base de nomes fictício. Fizemos o *lookup* entre a tabela de nomes com os dados da “studentInfo”, agrupamos os dados da tabela “studentVle”, que são as informações de interações dos alunos com o AVA. Transformamos os dados dessas interações, disponíveis em linhas para serem apresentados em colunas (*Pivot*). Fizemos o *lookup* das duas combinações gerando a estrutura “lookup [student Vle]” e combinamos com a tabela dos módulos.

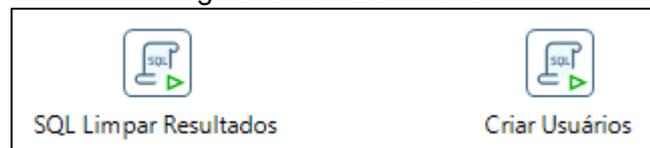
Figura 6.9 - Fluxo de extração e tratamento Pentaho



Fonte: Próprio autor.

Na etapa 03 Gerar Caso é feita a limpeza de alguns resultados para que compor o cenário de validação da ferramenta. Apagamos o resultado de cinco alunos aleatoriamente de cada um dos sete módulos, gerando um total de 35 alunos para compor as turmas do professor. Usamos uma base de nomes para atribuir aos 35 alunos. No SQL 6 - Criar casos para experimentos, contém o comando de criação do contexto de experimento. No SQL 7 - Gerar usuários, é apresentado o comando de criação de usuários para os alunos que compõem a turma.

Figura 6.10 - Gerar Caso



Fonte: Próprio autor.

SQL 7 - Gerar usuários	
1.	<code>INSERT INTO usuario (nome, login, senha, tipo)</code>
2.	<code>SELECT student_name, id_student, id_student, 'E'</code>
3.	<code>FROM __dados_finais</code>
4.	<code>WHERE final_result = '' OR final_result LIKE 'LA :: Prev%'</code>
5.	<code>AND NOT EXISTS (SELECT '1' FROM usuario WHERE usuario.login = id_student)</code>

SQL 6 - Criar casos para experimentos	
1.	UPDATE __dados_finais
2.	SET final_result = ''
3.	WHERE id_student IN ('447481', '2404408', '298034', '2649800', '260355')
4.	AND code_module = 'AAA'
5.	AND code_presentation = '2014J';
6.	
7.	UPDATE __dados_finais
8.	SET final_result = ''
9.	WHERE id_student IN ('633212', '608134', '631341', '627948', '679504')
10.	AND code_module = 'BBB'
11.	AND code_presentation LIKE '2014%';
12.	
13.	UPDATE __dados_finais
14.	SET final_result = ''
15.	WHERE id_student IN ('532020', '465331', '2559144', '645698', '554459')
16.	AND code_module = 'CCC'
17.	AND code_presentation LIKE '2014%';
18.	
19.	UPDATE __dados_finais
20.	SET final_result = ''
21.	WHERE id_student IN ('676865', '477180', '154073', '2611371', '2542116')
22.	AND code_module = 'DDD'
23.	AND code_presentation LIKE '2014%';
24.	
25.	UPDATE __dados_finais
26.	SET final_result = ''
27.	WHERE id_student IN ('479987', '1663815', '652606', '630047', '683728')
28.	AND code_module = 'EEE'
29.	AND code_presentation LIKE '2014%';
30.	
31.	UPDATE __dados_finais
32.	SET final_result = ''
33.	WHERE id_student IN ('602571', '633505', '697998', '629686')
34.	AND code_module = 'FFF'
35.	AND code_presentation LIKE '2014%';
36.	
37.	UPDATE __dados_finais
38.	SET final_result = ''
39.	WHERE id_student IN ('562064')
40.	AND code_module = 'FFF'
41.	AND code_presentation LIKE '2014%'
42.	AND num_of_prev_attempts = '1';
43.	
44.	UPDATE __dados_finais
45.	SET final_result = ''
46.	WHERE id_student IN ('621342', '674927', '614249', '651436', '513166')
47.	AND code_module = 'GGG'
48.	AND code_presentation LIKE '2014%';

6.3.3. Gerenciamento dos Metadados

Para o processo de treinamento de algoritmos de aprendizado de máquina é necessário identificar o atributo classe (classificador). Para isso, o sistema permite que o usuário selecione qual é o campo classe. O sistema também permite desabilitar

campos que não influenciam os algoritmos classificadores, tais como nome, identificador do aluno, data da matrícula, etc.

Após a importação foi necessário realizar o gerenciamento dos metadados de algumas colunas da tabela, conforme apresentado na Figura 6.11 - Formulário de alteração de Metadados, no qual é apresentado o formulário de edição das colunas. Desabilitamos as colunas para uso do módulo de predição: *code_module*, *code_presentation*, *id_student*, *age_band*, *gender*, *highest_education*, *imd_band*, *num_of_prev_attempts*, *region*, *student_name*, *descr_module* e *studied_credits*. Algumas colunas não influenciam no resultado do algoritmo, como colunas identificadoras (exemplo: *id_student*). Logo, as recomendações deverão buscar somente valores das colunas de interações com o ambiente de aprendizado, correspondendo ao comportamento acadêmico do aluno dentro do ambiente. Também informamos ao sistema que a coluna *final_result* é classe, ou seja, ela é usada para treinar e prever valores, quando vazia, os valores serão previstos. E por último, informamos que a coluna *id_student* é a coluna identificadora do aluno (matrícula). Ou seja, cada usuário do tipo Aluno terá um registro nessa tabela. Dessa forma, será possível identificar o registro do aluno e facilitando a geração das recomendações.

Figura 6.11 - Formulário de alteração de Metadados

Colunas

Tabela
__dados_finais

Nome
id_student

Tipo
Número

SQL
id_student

Desabilitar

Título
id_student

Descrição
Informe a descrição do campo

Classe (na predição - valores nulos serão previstos)

Valores aluno "em risco"
Informe o valor de detectar alunos 'em risco'

Identificação Aluno
 Chave Estrangeira

Tabela

Coluna

Salvar Cancelar

Fonte: Próprio autor.

6.3.4. Processamento

Categorizamos os dados, transformando valores contínuos da quantidade de cliques em valores discretos, com as descrições de “pouco uso”, “uso moderado” e “muito uso” conforme os quartis. Onde o primeiro quartil corresponde a “pouco uso”, o segundo e terceiro quartil corresponde a “uso moderado” e o quarto quartil corresponde a “muito uso”. Conforme apresentado na Figura 6.12 - Formulário de Discretização, o sistema permite que valores contínuos sejam transformados em valores discretos. Na Figura 6.13 - Processo de discretização pelo módulo de Processamento, bloco A, demonstra os valores contínuos antes da transformação, e bloco B, tem os dados do mesmo campo com informações após o processamento dos dados. Dessa forma, as mensagens de recomendações ficam de fácil entendimento para os alunos.

Figura 6.12 - Formulário de Discretização

Deseja Discretizar coluna? X

Nome
click_dataplus

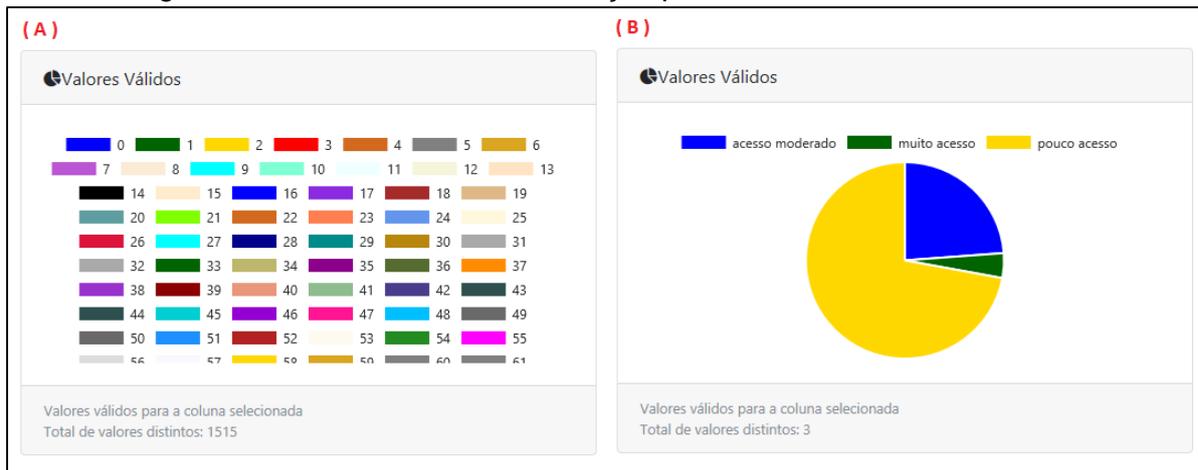
Expressão
CASE WHEN CAST(click_dataplus AS FLOAT) <= 22.0 THEN 'Q1' WHEN CAST(click_dataplus AS FLOAT) <= 44.0 THEN 'Q2' WHEN CAST(click_dataplus AS FLOAT) <= 67.0 THEN 'Q3' ELSE 'Q4' END

Quartil.Excl [[Quartil](#)]

Cancel Enviar

Fonte: Próprio autor.

Figura 6.13 - Processo de discretização pelo módulo de Processamento



Fonte: Próprio autor.

6.3.5. Alterações no banco de dados

Na finalidade de tornar o cenário de validação melhor, realizamos ajustes no banco de dados para criar a tabela de módulos e outra de nomes. Foram aplicadas as alterações no banco de dados utilizando a linguagem de manipulação de banco de dados SQL. O SQL 8 - Tabela de Módulo foi aplicada para gerar uma tabela chamada “__modulo”. No SQL 10 - Tabela de Nomes busca registros de uma base de nomes fictícios. Esses dados foram cruzados para geração da tabela “__dados_finais” do sistema, conforme apresentado no SQL 9.

```

SQL 8 - Tabela de Módulo
1. SELECT 'AAA' AS code_module,
2.     'Linguagem de Programação' AS descr_module
3. UNION
4. SELECT 'BBB' AS code_module,
5.     'Redes de Computadores' AS descr_module
6. UNION
7. SELECT 'CCC' AS code_module,
8.     'Banco de Dados' AS descr_module
9. UNION
10. SELECT 'DDD' AS code_module,
11.     'Estrutura de Dados' AS descr_module
12. UNION
13. SELECT 'EEE' AS code_module,
14.     'Engenharia de Software' AS descr_module
15. UNION
16. SELECT 'FFF' AS code_module,
17.     'Arquitetura de Computadores' AS descr_module
18. UNION
19. SELECT 'GGG' AS code_module,
20.     'Gerenciamento de Projetos' AS descr_module;

```

SQL 9 - Transformação de dados - “dados_finais” SUBSTITUIR SQL

```

1. CREATE TABLE [__dados_finais] AS
2. SELECT IFNULL((SELECT nm.NOME FROM __nomes_alunos AS nm WHERE
3.     nm.id_student = std.id_student LIMIT 1), 'ANONIMO') AS Nome,
4.     IFNULL(SUM(CASE WHEN vle.activity_type = 'dataplus'
5.         THEN stdVle.sum_click ELSE 0 END), 0) AS click_dataplus,
6.     IFNULL(SUM(CASE WHEN vle.activity_type = 'dualpane'
7.         THEN stdVle.sum_click ELSE 0 END), 0) AS click_dualpane,
8.     IFNULL(SUM(CASE WHEN vle.activity_type = 'externalquiz'
9.         THEN stdVle.sum_click ELSE 0 END), 0) AS click_extquiz,
10.    IFNULL(SUM(CASE WHEN vle.activity_type = 'folder'
11.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_folder,
12.    IFNULL(SUM(CASE WHEN vle.activity_type = 'forumng'
13.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_forumng,
14.    IFNULL(SUM(CASE WHEN vle.activity_type = 'glossary'
15.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_glossary,
16.    IFNULL(SUM(CASE WHEN vle.activity_type = 'homepage'
17.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_homepage,
18.    IFNULL(SUM(CASE WHEN vle.activity_type = 'htmlactivity'
19.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_htactivity,
20.    IFNULL(SUM(CASE WHEN vle.activity_type = 'oucollaborate'
21.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_oucollaborate,
22.    IFNULL(SUM(CASE WHEN vle.activity_type = 'oucontent'
23.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_oucontent,
24.    IFNULL(SUM(CASE WHEN vle.activity_type = 'ouelluminate'
25.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_ouluminate,
26.    IFNULL(SUM(CASE WHEN vle.activity_type = 'ouwiki'
27.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_ouwiki,
28.    IFNULL(SUM(CASE WHEN vle.activity_type = 'page'
29.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_page,
30.    IFNULL(SUM(CASE WHEN vle.activity_type = 'questionnaire'
31.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_questnaire,
32.    IFNULL(SUM(CASE WHEN vle.activity_type = 'quiz'
33.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_quiz,
34.    IFNULL(SUM(CASE WHEN vle.activity_type = 'repeatactivity'
35.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_repactivity,
36.    IFNULL(SUM(CASE WHEN vle.activity_type = 'resource'
37.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_resource,
38.    IFNULL(SUM(CASE WHEN vle.activity_type = 'sharedsubpage'
39.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_shdsubpage,
40.    IFNULL(SUM(CASE WHEN vle.activity_type = 'subpage'
41.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_subpage,
42.    IFNULL(SUM(CASE WHEN vle.activity_type = 'url'
43.        THEN stdVle.sum_click ELSE 0 END), 0) AS click_url,
44.    stdVle.id_student, stdVle.code_module, stdVle.code_presentation,
45.    md.descr_module AS descr_module,
46.    aval.TMA AS TMA, aval.CMA AS CMA, aval.Exam AS Exam,
47.    std.code_module AS code_module,
48.    std.code_presentation AS code_presentation,
49.    std.id_student, std.gender, std.region,
50.    std.highest_education AS highest_education,
51.    std.imd_band AS imd_band,
52.    std.age_band AS age_band,
53.    std.num_of_prev_attempts AS num_of_prev_attempts,
54.    std.studied_credits AS studied_credits,
55.    std.disability AS disability,
56.    std.final_result AS final_result
57. FROM __studentVle stdVle
58. INNER JOIN __vle AS vle ON vle.id_site = stdVle.id_site
59. INNER JOIN __studentInfo std ON std.id_student = stdVle.id_student
60.     AND std.code_module = stdVle.code_module
61.     AND std.code_presentation = stdVle.code_presentation
62. INNER JOIN __modulo_curso AS md ON std.code_module = md.code_module
63. GROUP BY stdVle.id_student, stdVle.code_module, stdVle.code_presentation

```

SQL 10 - Tabela de Nomes	
1.	SELECT 'SELECT ' '' First_NAME ' ' LAST_NAME '' AS NOME UNION'
2.	FROM minha_tabela_estudantes;
3.	(...)
4.	SELECT 'Kayla Jimenez' AS NOME UNION
5.	SELECT 'Rene Ramos' AS NOME UNION
6.	(...)
7.	SELECT 'Jenny Sullivan' AS NOME UNION
8.	SELECT 'Ellie Nguyen';

6.4. Análise dos Dados

Durante a fase de análise é realizada a previsão do desempenho dos alunos. E uma vez que possamos identificar os alunos em “risco acadêmico” serão geradas recomendações a eles.

6.4.1. Previsão

No módulo de Predição geramos os modelos de algoritmo de aprendizado de máquina, usando os dados transformados e tratados nas etapas anteriores. Conforme apresentado na Figura 6.14 - Acurácia dos algoritmos preditivos, os algoritmos são apresentados com as medidas de acurácia (Acur), sensibilidade (Sen), especificidade (Esp), eficiência (Efic), valor preditivo positivo (VPP) e valor preditivo negativo (VPN). O que teve a melhor acurácia foi a MLP, que foi selecionado para fazer a previsão de desempenho dos alunos, os classificando “em risco” ou “sucesso” acadêmico.

Figura 6.14 - Acurácia dos algoritmos preditivos

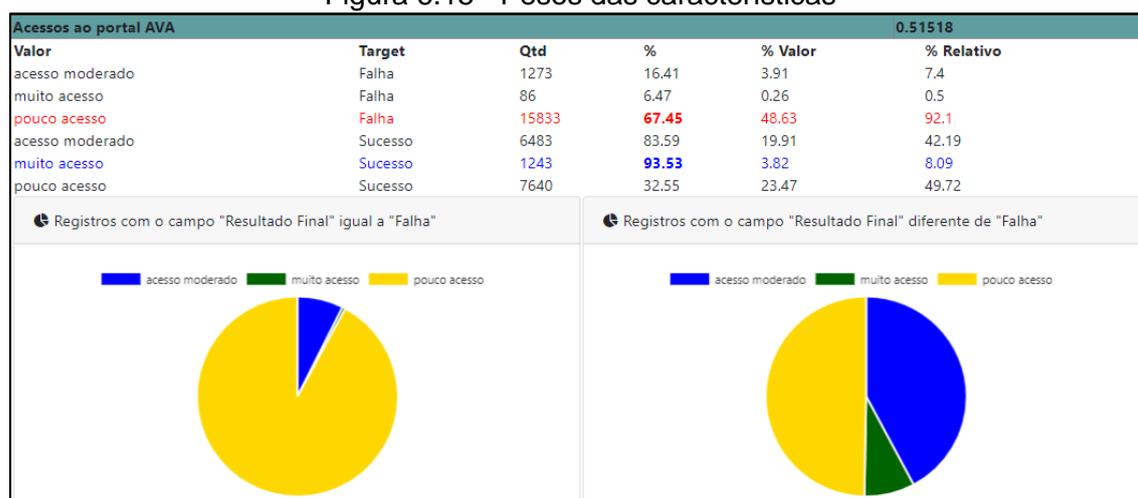
# Algoritmo	Acur	Sen	Esp	Efic	VPP	VPN
Decision Tree-based	71.4022	65.0	84.3	74.7	89.3	54.5
K-nearest Neighbors	68.3459	65.3	72.2	68.7	74.6	62.4
Support Vector Machine	73.8596	68.3	82.8	75.6	86.5	61.9
* Multi-layer Perceptron Classifier	73.9671	67.9	84.2	76.1	88.0	60.7
Naive Bayes	71.1104	64.6	84.5	74.6	89.6	53.6

* Algoritmo selecionado para prever os registros com a classe vazia

Fonte: Próprio autor.

A base da mensagem de recomendação é o valor da importância de cada característica processada pelo módulo de Predição. Conforme apresentado na Figura 6.15 - Pesos das características, o algoritmo preditivo apresentou a característica “Acessos ao portal AVA”, com uma importância de 0,51518 (valor de referência do algoritmo). A maior porcentagem de alunos com sucesso acadêmico teve “muito acesso” nessa característica. Com um total de 93,53% dos alunos com sucesso acadêmico e consecutivamente 6,47%, com insucesso acadêmico. Todas as características são apresentadas com diferentes importâncias e porcentagens.

Figura 6.15 - Pesos das características



Fonte: Próprio autor.

6.4.1. Geração das mensagens de Recomendação

Após a geração dos modelos analíticos é possível a previsão do desempenho dos alunos (que estão com o valor da coluna *final_result* vazia). Neste momento serão gerados, de forma automática, as mensagens de recomendação, e estará disponível ao professor. Conforme exemplificado na Figura 6.16 - Mensagem: (A) sugerida ao professor e (B) recebida pelo aluno, (A) o professor terá acesso a uma mensagem e após editá-la será enviada ao aluno (B).

Figura 6.16 - Mensagem: (A) sugerida ao professor e (B) recebida pelo aluno



Fonte: Próprio autor.

6.5. Visualização

Para facilitar a visualização do andamento de suas turmas, o professor poderá acompanhar através de painéis totalmente customizáveis. Esses painéis são compostos por pequenas visualizações, onde cada uma delas poderá ter diferentes configurações e buscar dados de diferentes formas.

6.5.1 Criar Visões

As visões são pequenos pedaços reaproveitáveis que compõem um painel. O usuário poderá criar a visão usando diferentes configurações. Conforme exemplificado na Figura 6.17 - Formulário de configuração de Visão, o usuário poderá selecionar os campos, configuração dos campos, o nome da visão e o tipo de visão. Os tipos de visão são: Gráfico de Pizza; Gráfico de Barras; Gráfico de Linhas; Informações em Cards; e Informações em Grades. Nós criamos as visões: [P] Acesso ao Portal (Gráfico de Pizza); [P] Acesso ao Fórum (Gráfico de Pizza); [P] Previsão de Desempenho (Gráfico de Pizza); [G] Lista de Nomes por Previsão de Desempenho

(Informação em Grid); [P] Acesso ao Quiz (Gráfico de Pizza); [P] Acesso ao AVA (Gráfico de Pizza); [P] Acesso ao Wiki (Gráfico de Pizza); [B] Previsão de desempenho por Disciplina (Gráfico Barras). Após a criação da visão é possível visualizá-la individualmente, conforme apresentado na Figura 6.18 - Exibição da Visualização Previsão de desempenho por Disciplina.

Figura 6.17 - Formulário de configuração de Visão

Definir Visão

[Voltar](#)

Tabela

- code_module
- code_presentation
- id_student
- Acesso dados Mais
- Acesso ao painel
- Acesso ao quiz
- Acesso ao diretório
- Acesso aos fóruns
- Acesso ao glossário
- Acesso ao AVA
- Acesso às atividades (HTML)
- Acesso ao portal de colaboração

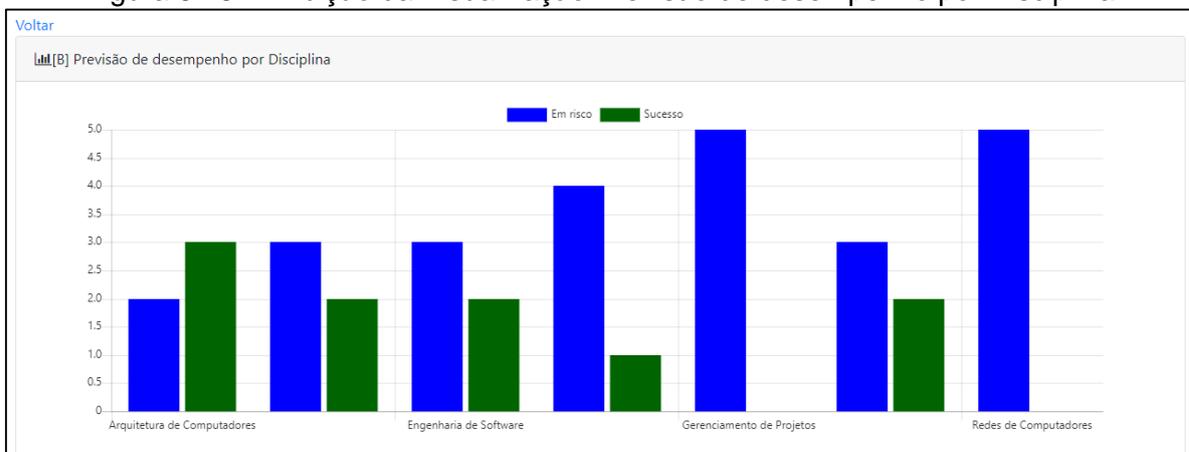
Nome

Tipo

Campo	Agrupador	Gráfico
id_student	<input type="text" value="Quantidade"/>	<input type="text" value="Valor"/>
click_forumng	<input type="text"/>	<input type="text" value="Serie"/>

Fonte: Próprio autor.

Figura 6.18 - Exibição da Visualização Previsão de desempenho por Disciplina

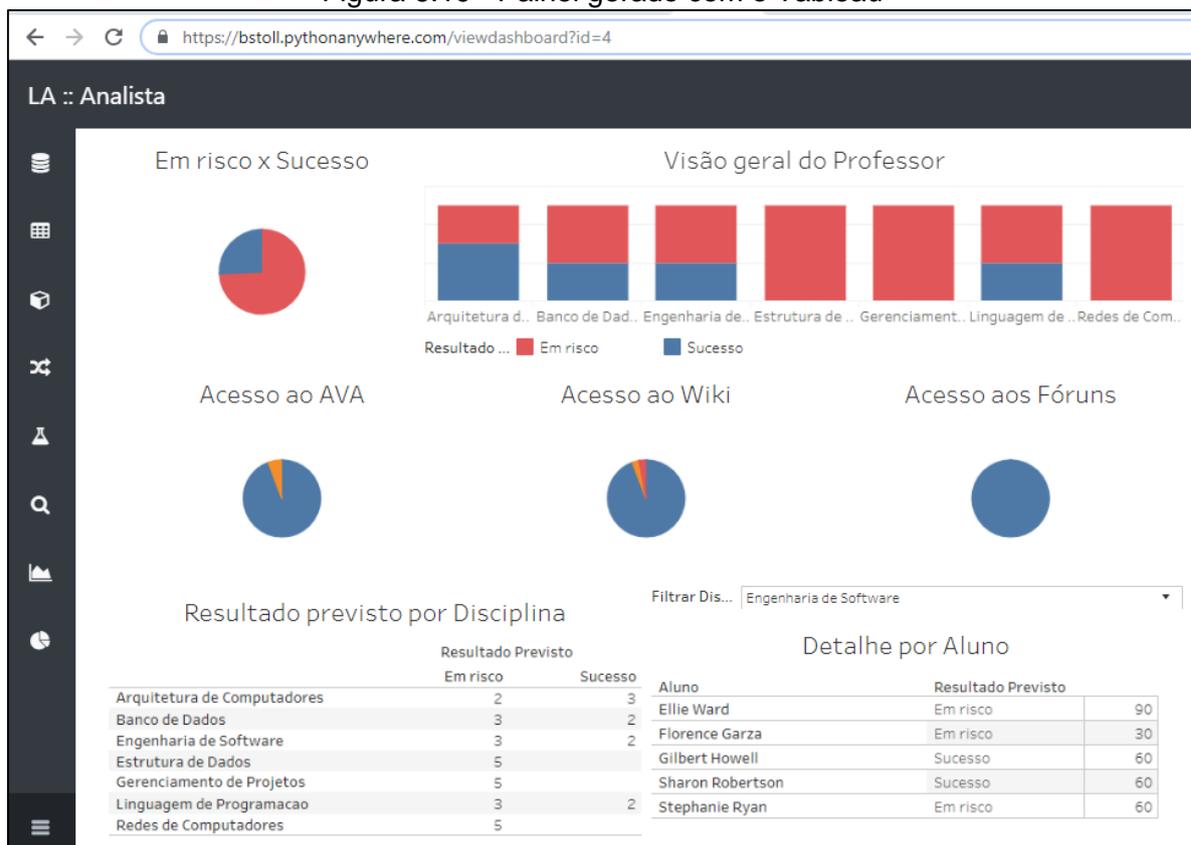


Fonte: Próprio autor.

6.5.2. Painéis

Conforme apresentado na Figura 6.20 - Painel gerado com os recursos do sistema e na Figura 6.19 - Painel gerado com o Tableau, o professor tem acesso a um painel com informações de sua turma. Nesse painel, que fizemos usando o sistema, é possível visualizar as informações: (a) quantidade de aluno por desempenho previsto das turmas; (b) previsão de resultado por disciplina; (c) acesso aos fóruns por mês; (d) quantidade de acessos ao portal; (e) quantidade de acessos aos quizzes; (f) quantidade de acessos aos materiais; (g) previsão de desempenho por disciplina; e (h) lista da previsão de desempenho.

Figura 6.19 - Painel gerado com o Tableau



Fonte: Próprio autor.

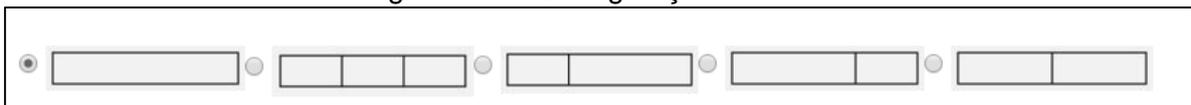
Figura 6.20 - Painel gerado com os recursos do sistema



Fonte: Próprio autor.

Na construção do painel é possível incluir, alterar ou remover qualquer Visão. O painel está organizado em linhas, conforme apresentado na Figura 6.21 - Configuração da Linha, cada um poderá ter uma, duas ou três células. Cada uma delas terá uma visão associada, conforme apresentado na Figura 6.22 - Associar visão a uma célula.

Figura 6.21 - Configuração da Linha



Fonte: Próprio autor.

Figura 6.22 - Associar visão a uma célula

Remover (1)	[P] Previsão de Desempenho	[B] Previsão de desempenho por Disciplina	
Remover (2)	[P] Acesso ao AVA	[P] Acesso ao Quiz	[P] Acesso ao Wiki
Remover (3)	[G] Lista de Nomes por Previsão de Desempenho	[P] Acesso ao Portal	

Fonte: Próprio autor.

Também é possível incorporar um relatório gerado na ferramenta Tableau. No qual, o usuário deverá informar o caminho do relatório, conforme apresentado na Figura 6.23 - Formulário com o Painel do Professor associado. A exibição do relatório associado e apresentado na Figura 6.19 - Painel gerado com o Tableau.

Figura 6.23 - Formulário com o Painel do Professor associado

Nome

Painel do Professor

Modelo

https://public.tableau.com/views/MESTRADO-PaineldoProfessor/PAINELDOPROFESSOR?:embed=y&:display_count=yes

Salvar

Fonte: Próprio autor.

6.5. Recomendações

Conforme apresentado na Figura 6.24 - Tela de configuração da Recomendação, quando um usuário do perfil professor (*stoll*) acessar o menu Recomendações com o perfil de professor, o sistema apresenta todos os alunos de suas turmas que foram classificados “em risco”. Ao clicar em cada um dos alunos o sistema apresenta as mensagens sugeridas pelo sistema. O professor pode editar cada uma das mensagens, conforme apresentado na Figura 6.25 - Formulário da Recomendação. Em nosso experimento, selecionamos a aluna “Rebecca Arnold”, ignoramos os três últimos itens da recomendação. E clicamos no botão enviar recomendação.

Figura 6.24 - Tela de configuração da Recomendação

Learning Analytis - Área do Professor

Recomendações aos Alunos

Alunos

- Rebecca Arnold
- Charlotte Welch
- Brittany Green
- Robert West
- Zoe Perez
- Ella Tucker
- Herbert Stephens
- Kimberly Lee
- Phoebe Gutierrez
- Ed Russell
- Lillian Harrison
- Joel Ramos
- Robert Bailey
- Brayden Gray

Caro aluno, **Rebecca Arnold**. Você foi classificado “em risco” por um sistema de previsão com uma taxa de acerto de 73.2914%

Ignorar

Na característica “Acesso ao AVA” esta informado: “(1) muito pouco acesso”. Recomendamos que seja modificado para “(4) muito acesso”

Ignorar

Na característica “Acesso aos conteúdos” esta informado: “(1) muito pouco acesso”.

Ignorar

Na característica “Acesso ao quiz” esta informado: “(1) muito pouco acesso”. Recomendamos que seja modificado para “(4) muito acesso”

Ignorar

Na característica “Acesso aos fóruns” esta informado: “(1) muito pouco acesso”. Recomendamos que seja modificado para “(4) muito acesso”

Ignorar

Fonte: Próprio autor.

Figura 6.25 - Formulário da Recomendação

Caro aluno, **Rebecca Arnold**. Você foi classificado "em risco" por um sistema de previsão com uma taxa de acerto de 73.2914%

Ignorar

Na característica "Acesso ao AVA" esta informado: "(1) muito pouco acesso".
Recomendamos que seja modificado para "(4) muito acesso"

Ignorar

Na característica "Acesso aos conteúdos" esta informado: "(1) muito pouco acesso".

Ignorar

Na característica "Acesso ao quiz" esta informado: "(1) muito pouco acesso".
Recomendamos que seja modificado para "(4) muito acesso"

Ignorar

Na característica "Acesso aos fóruns" esta informado: "(1) muito pouco acesso".
Recomendamos que seja modificado para "(4) muito acesso"

Ignorar

Na característica "Acesso ao Wiki" esta informado: "(1) muito pouco acesso".
Recomendamos que seja modificado para "(4) muito acesso"

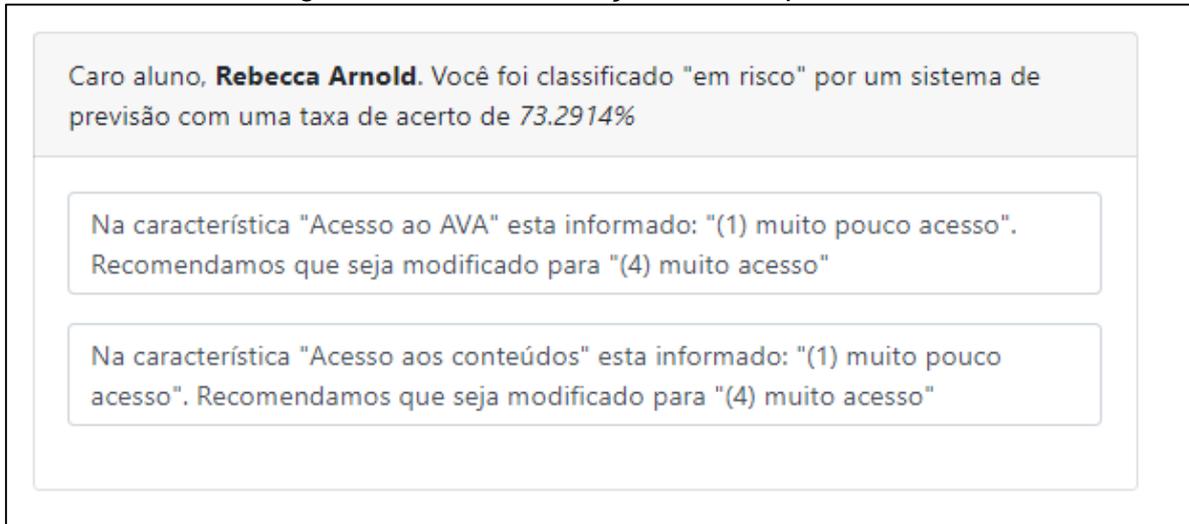
Enviar recomendação

Fonte: Próprio autor.

Após o envio, entramos no sistema com o usuário 260355 (Recebba Arnold) do perfil aluno e visualizamos a recomendação enviada pelo professor, conforme apresentado na Figura 6.26 - Recomendação recebida pelo Aluno, no qual foi feito o acesso usando o navegador Google Chrome. Na Figura 6.27 - Mensagem recebida pelo aluno usando o celular é exemplificado o acesso usando o dispositivo móvel

(Android na versão 6.0). Dessa forma, o aluno poderá ter reflexões pertinentes sobre seu comportamento através dessas mensagens de recomendação.

Figura 6.26 - Recomendação recebida pelo Aluno



Fonte: Próprio autor.

Figura 6.27 - Mensagem recebida pelo aluno usando o celular



Fonte: Próprio autor.

6.6. Análise dos Resultados

O objetivo geral desta dissertação é demonstrar que com o suporte computacional acoplável é possível apoiar professores em análise e intervenção no processo de aprendizado. Para isso, um protótipo funcional com os requisitos necessários foi implementado para validar esse objetivo. Nessa validação, foi realizada uma análise de dados usando três perfis diferentes: usuário *bruno*, com perfil de analista; usuário *stoll*, com perfil de professor; e o usuário 260355, com perfil de aluno. Com o perfil de analista foram realizadas as atividades de: importação do conjunto de dados; realização de alguns ajustes para caracterização do experimento; geração dos modelos preditivos; geração das recomendações; e configuração dos painéis. Já com o perfil de professor foram realizadas as atividades de: visualização dos painéis; visualização e ajustes nas mensagens de recomendação; e envio da recomendação para o aluno. Por fim, com o perfil de aluno foi realizada a atividade visualização das mensagens de recomendações.

A validação da proposta foi necessária para comprovar a eficácia de solução computacional. Ao aplicar a prova de conceitos, podemos observar que, apesar de nem todos os requisitos levantados foram implementados, a solução computacional proposta é satisfatória diante do objetivo geral e objetivos específicos.

7. Conclusão

O objetivo desta dissertação é demonstrar que com o suporte computacional acoplável é possível apoiar professores nas atividades análise e intervenção no processo de aprendizado. Desta forma, gerou-se a questão da pesquisa: “*Como oferecer um suporte tecnológico independente do ambiente para a acoplável a qualquer ambiente de aprendizado para previsão de desempenho, recomendações e visualizações?*”. Para isso, foi apresentado um modelo conceitual para a construção de uma ferramenta de análise detalhada de dados educacionais, flexível, sensível ao contexto e acoplável a qualquer ambiente. Para validar esse modelo, desenvolvemos um protótipo com os requisitos necessários para avaliar a viabilidade da construção de tal solução. Aplicamos a prova de conceito e os experimentos atestam a eficácia da ferramenta. Dessa forma, podemos concluir que o experimento comprova a eficiência do modelo conceitual proposto, respondendo à questão da pesquisa. Os primeiros resultados da ferramenta usando um conjunto de dados públicos indicam que a ferramenta gera mensagens, orientando com sugestões apropriadas aos alunos para evitar insucesso acadêmico.

Apesar das vantagens e contribuições apresentadas pela ferramenta proposta, algumas limitações foram identificadas: (a) dificuldade nos tratamentos iniciais dos dados; (b) dificuldade em encontrar dados reais de forma a apresentar o uso num contexto real; (c) problemas de desempenho; (d) dentre os vários experimentos realizados, não se pôde reaproveitar modelos analíticos entre os testes, devido a especificidade de cada caso.

7.1. Trabalhos futuros

Como trabalhos futuros, destacamos melhorias no modelo conceitual e contribuições relacionadas às áreas de pesquisa de aprendizado de máquina e recomendações baseadas em inteligência artificial. Podemos destacar: (a) pré-processamento de algoritmos preditivos; (b) ajustes automáticos de hiperparâmetros de algoritmos classificadores utilizando algoritmos genéticos; (c) usar algoritmos de aprendizado profundo, do inglês *Deep Learning (DL)*, em dados não estruturados para avaliar o

aprendizado e engajamento de alunos em aulas virtuais, com o Hangout como plataforma.

7.2. Produções científicas

No decorrer desta pesquisa, elaboramos artigos científicos que foi submetido e aprovado, listados à seguir:

- O artigo “*Framework para predições e recomendações em dados acadêmicos*” foi publicado na revista RELNOTE (2018). Esse artigo tem por objetivo apresentar o framework projetado com a concepção inicial da pesquisa.
- O artigo “*Análise de dados acadêmicos baseado em previsão, recomendação e visualização*” foi publicado na revista RELNOTE (2019). No qual, é a evolução do primeiro, contemplando painéis de visualização.

Referências

- _____. Censo EAD.BR: **Relatório Analítico da Aprendizagem a Distância no Brasil - 2017**. Associação Brasileira de Educação a Distância. 2017.
- _____. Censo EAD.BR: **Relatório Analítico da Aprendizagem a Distância no Brasil - 2018**. Associação Brasileira de Educação a Distância. 2018.
- AGUIAR, Janderson; FECHINE, Joseana; COSTA, Evandro. Recomendação de Objetos de Aprendizagem utilizando Filtragem Colaborativa baseada em Tendências e em Estilos de Aprendizagem. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1423.
- ARNOLD, Kimberly E.; PISTILLI, Matthew D. Course signals at Purdue: Using learning analytics to increase student success. In: **Proceedings of the 2nd international conference on learning analytics and knowledge**. ACM, 2012. p. 267-270.
- BAJAJ, Chandrajit; WILEY, C. FI John. Data visualization techniques. 1998.
- BARBOSA, Artur; SANTOS, Emanuele; PORDEUS, João Paulo. A machine learning approach to identify and prioritize college students at risk of dropping out. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2017. p. 1497.
- BEZERRA, Sidney Firmino et al. Um Sistema de Recomendação Híbrido Integrado a Ontologia em um Ambiente de Aprendizagem Ubíqua. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1253.
- BOLZAN, Willian; GIRAFFA, Lúcia MM. Estudo comparativo sobre sistemas tutores inteligentes multiagentes web. **Porto Alegre: FACIN-PUCRS**, 2002.
- BRANT-RIBEIRO, Taffarel; CATTELAN, Renan. Tamanho Ótimo de Amostra para Análise do Desempenho de Estudantes em Ambientes Educacionais Ubíquos. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2015. p. 31.
- BRASIL. Decreto nº 9.057, de 25 de maio de 2017. Regulamenta o art. 80 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional. **Diário Oficial da República Federativa do Brasil**. 2017.

- BROWN, John Seely. Growing up: Digital: How the web changes work, education, and the ways people learn. **Change: The Magazine of Higher Learning**, v. 32, n. 2, p. 11-20, 2000.
- BROWN, Malcolm. Learning analytics: Moving from concept to practice. **EDUCAUSE Learning Initiative**, v. 7, 2012.
- CAZELLA, Silvio Cesar et al. Desenvolvendo um Sistema de Recomendação de Objetos de Aprendizagem baseado em Competências para a Educação: relato de experiências. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2012.
- CHATTI, Mohamed Amine et al. A reference model for learning analytics. **International Journal of Technology Enhanced Learning**, v. 4, n. 5-6, p. 318-331, 2012.
- CLOW, Doug. The learning analytics cycle: closing the loop effectively. In: **Proceedings of the 2nd international conference on learning analytics and knowledge**. ACM, 2012. p. 134-138.
- COELHO, Orlando Bisacchi; SILVEIRA, Ismar. Deep Learning applied to Learning Analytics and Educational Data Mining: A Systematic Literature Review. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2017. p. 143.
- COSTA, Felipe Garcia da. Visualização de dados e sua importância na era do Big Data. 2017.
- COSTA, Evandro; AGUIAR, Janderson; MAGALHÃES, Jonathas. Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, 2013.
- DE BRITO, Daniel Miranda et al. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 882.
- DE MEDEIROS, Francisco Petrônio Alencar; GOMES, Alex Sandro. Monitoramento da experiência do usuário em ambientes colaborativos virtuais de aprendizagem: Um mapeamento sistemático. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2012.
- DE OLIVEIRA, Márcia G. et al. Análise de componentes latentes da aprendizagem de programação para mapeamento e classificação de perfis. In: **Brazilian Symposium**

- on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 134.
- DEFREITAS, Kyle; BERNARD, Margaret. A Framework for Flexible Educational Data Mining. In: **Proceedings of the International Conference on Data Mining (DMIN)**. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014. p. 1.
- DERMEVAL, Diego. Intelligent Authoring of Gamified Intelligent Tutoring System. In: **UMAP (Extended Proceedings)**. 2016.
- DETONI, Douglas; ARAUJO, Ricardo Matsumura; CECHINEL, Cristian. Predição de reprovação de alunos de educação a distância utilizando contagem de interações. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 896.
- DYCKHOFF, Anna Lea et al. Design and implementation of a learning analytics toolkit for teachers. **Journal of Educational Technology & Society**, v. 15, n. 3, 2012.
- DUVAL, Erik. Attention please!: learning analytics for visualization and recommendation. In: **Proceedings of the 1st international conference on learning analytics and knowledge**. ACM, 2011. p. 9-17.
- ELIAS, Tanya. Learning analytics: Definitions, Processes and Potential. **Learning**, p. 1-22, 2011.
- FACELI, Katti et al. Inteligência Artificial: Uma abordagem de aprendizado de máquina. 2011.
- FACHIN, Gleisy Regina Bories. Recuperação inteligente da informação e ontologias: um levantamento na área da Ciência da Informação. **BIBLOS**, v. 23, n. 1, p. 259-283, 2009.
- FALBO, Ricardo de Almeida; TRAVASSOS, Guilherme Horta. A integração de conhecimento em um ambiente de desenvolvimento de software. In: **II Congreso Argentino de Ciencias de la Computación**. 1996.
- FASSBINDER, Aracele; DELAMARO, Márcio Eduardo; BARBOSA, Ellen Francine. Construção e Uso de MOOCs: Uma Revisão Sistemática. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 332.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. ISSN 0738-4602.

- FERGUSON, Rebecca. Learning analytics: drivers, developments and challenges. **International Journal of Technology Enhanced Learning**, v. 4, n. 5/6, p. 304-317, 2012.
- FERGUSON, Rebecca; SHUM, Simon Buckingham. Social learning analytics: five approaches. In: **Proceedings of the 2nd international conference on learning analytics and knowledge**. ACM, 2012. p. 23-33.
- FERREIRA, Tais et al. Detecção automática de traços de personalidade e recomendação de agrupamento com o modelo Big Five. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1643.
- GAZZOLA, Murilo. Um Método para Avaliação Automática da Qualidade de Recursos Educacionais Abertos usando Deep Learning. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2017. p. 1477.
- GRELLER, Wolfgang; DRACHSLER, Hendrik. Translating learning into numbers: A generic framework for learning analytics. 2012.
- GIRAFFA, Lucia Maria Martins. Docentes analógicos e alunos da geração digital: desafios e possibilidades na escola do século XXI. **Re) invenção pedagógica**, p. 23-32, 2012.
- GOMES, Luiz Fernando. EAD no Brasil: perspectivas e desafios. **Avaliação: Revista da Avaliação da Educação Superior**, v. 18, n. 1, 2013.
- GONÇALVES, Enyo; VILELA, Jéssyka; BEZERRA, Jeandro. Análise Estatística de Notas e Interações em Cursos a Distância. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 71.
- GOTARDO, Reginaldo; CEREDA, Paulo Roberto Massa; JUNIOR, Estevam Rafael Hruschka. Predição do Desempenho do Aluno usando Sistemas de Recomendação e Acoplamento de Classificadores. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2013. p. 657.
- GOTTARDO, Ernani; KAESTNER, Celso; NORONHA, Robinson Vida. Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2012.

- GOVINDARAJAN, Kannan et al. Learning Analytics Solution for Reducing Learners. In: **2015 IEEE Seventh International Conference on Technology for Education (T4E)**. IEEE, 2015. p. 83-90.
- GUTERRES, João; SILVEIRA, Milene. Um panorama sobre Learning Analytics em Objetos de Aprendizagem. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1303.
- HEARTY, John. **Advanced Machine Learning with Python**. Packt Publishing Ltd, 2016.
- ISINKAYE, F. O.; FOLAJIMI, Y. O.; OJOKOH, B. A. Recommendation systems: Principles, methods and evaluation. **Egyptian Informatics Journal**, v. 16, n. 3, p. 261-273, 2015.
- JOHNSON, William B. **Pragmatic considerations in research, development, and implementation of intelligent tutoring systems**. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- JOHNSON, Larry et al. **NMC horizon report: 2016 higher education edition**. The New Media Consortium, 2016.
- JÚNIOR, Clarivando Belizário; DORÇA, Fabiano. Uma Abordagem para a Criação e Recomendação de Objetos de Aprendizagem usando um Algoritmo Genético, Tecnologias da Web Semântica e uma Ontologia. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1533.
- KEIM, Daniel A. Information visualization and visual data mining. **IEEE Transactions on Visualization & Computer Graphics**, n. 1, p. 1-8, 2002.
- LANG, C. et al. **The Handbook of Learning Analytics**. Society for Learning Analytics Research. 2017.
- LAN, Andrew S. et al. Behavior-based latent variable model for learner engagement. In: **Proceedings of International Conference on Educational Data Mining**. 2017. p. 64-71.
- LANES, Mariele; ALCÂNTARA, Cleber. Predição de Alunos com Risco de Evasão: estudo de caso usando mineração de dados. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1921.
- LUCKIN, Rose et al. **Intelligence unleashed: An argument for AI in education**. 2016.

- MASCHIO, Pedro et al. Um Panorama acerca da Mineração de Dados Educacionais no Brasil. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1936.
- MARTIN, Florence; NDOYE, Abdou. Using learning analytics to assess student learning in online courses. **Journal of University Teaching & Learning Practice**, v. 13, n. 3, p. 7, 2016.
- MEDEIROS, Francisco et al. Architecture for Social Interactions Monitoring in Collaborative Learning Environments as a Support for the Teacher's Awareness. In: **Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on**. IEEE, 2013. p. 123-127.
- MOISSA, Barbara; GASPARINI, Isabela; KEMCZINSKI, Avanilde. Learning Analytics: um mapeamento sistemático. **Nuevas Ideas en Informática Educativa TISE**, v. 2014, p. 283-290, 2014.
- MOISSA, Barbara; GASPARINI, Isabela; KEMCZINSKI, Avanilde. Educational Data Mining versus Learning Analytics: estamos reinventando a roda? Um mapeamento sistemático. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2015. p. 1167.
- NETTO, Carla; GUIDOTTI, Viviane; DOS SANTOS, Pricila Kohls. A evasão na EaD: investigando causas, propondo estratégias. In: **Congressos CLABES**. 2017.
- OLIVEIRA JÚNIOR, José Gonçalves de et al. **Identificação de padrões para a análise da evasão em cursos de graduação usando mineração de dados educacionais**. 2015. Dissertação de Mestrado. Universidade Tecnológica Federal do Paraná.
- OSTROW, Korinn S.; WANG, Yan; HEFFERNAN, Neil T. How Flexible Is Your Data? A Comparative Analysis of Scoring Methodologies across Learning Platforms in the Context of Group Differentiation. **Journal of Learning Analytics**, v. 4, n. 2, p. 91-112, 2017.
- PADILHA, Thereza; SOUZA, Italo. Uma Experiência do Uso de Ferramentas de Learning Analytics para Análise de Interações. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2017. p. 644.
- PASCOAL, Tulio et al. Evasao de estudantes universitários: diagnóstico a partir de dados acadêmicos e socioeconômicos. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2016. p. 926.

- PAZ, Fábio Josende; CAZELLA, Silvio César. Integrando Sistemas de Recomendação com Mineração de Dados Educacionais e Learning Analytics: Uma revisão sistemática da Literatura. **RENOTE**, v. 16, n. 1.
- POLSON, Martha C.; RICHARDSON, J. Jeffrey. **Foundations of intelligent tutoring systems**. Psychology Press, 2013.
- PREE, Wolfgang. Meta patterns—a means for capturing the essentials of reusable object-oriented design. In: **European Conference on Object-Oriented Programming**. Springer, Berlin, Heidelberg, 1994. p. 150-162.
- RAMOZZI-CHIAROTTINO, Z. O conhecimento como resultado da interação entre o organismo e o meio. **Psicologia e Epistemologia Genética de J. Piaget**. São Paulo: EPU, 1988.
- REZENDE, Paulo Alceu et al. Recomendação Baseada no Perfil e Contexto Tecnológico do Aluno. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1273.
- RICHERT, Willi. **Building machine learning systems with Python**. Packt Publishing Ltd, 2013.
- ROMERO, Cristobal et al. (Ed.). **Handbook of educational data mining**. CRC press, 2010.
- ROMERO, Cristobal; VENTURA, Sebastian. Educational data mining: A survey from 1995 to 2005. **Expert systems with applications**, v. 33, n. 1, p. 135-146, 2007.
- ROMERO, Cristóbal; VENTURA, Sebastián. Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 40, n. 6, p. 601-618, 2010.
- SALOMÃO VENÂNCIO, Ludmila; ERICHSEN NASSIF BORGES, Mônica. Cognição situada: fundamentos e relações com a Ciência da Informação. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, n. 22, 2006.
- SANTANA, Sivaldo Joaquim de et al. Extração e recomendação de boas e más práticas pedagógicas a partir de processos de ensino e aprendizagem usando um sistema tutor inteligente gamificado. 2017.
- SIEMENS, George et al. Open Learning Analytics: an integrated & modularized platform. 2011.
- SIEMENS, George; LONG, Phil. Penetrating the fog: Analytics in learning and education. **EDUCAUSE review**, v. 46, n. 5, p. 30, 2011.

- SIEMENS, George; D BAKER, Ryan SJ. Learning analytics and educational data mining: towards communication and collaboration. In: **Proceedings of the 2nd international conference on learning analytics and knowledge**. ACM, 2012. p. 252-254.
- SMITH, Alan Jay. The task of the referee. **IEEE Computer**, v. 23, n. 4, p. 65-71, 1990.
- STOLL, Bruno Bastos; CURY, Davidson; DE MENEZES, Crediné Silva. Framework para predições e recomendações em dados acadêmicos. **RENOTE**, v. 16, n. 2.
- STOLL, Bruno Bastos et al. Análise de dados acadêmicos baseado em previsão, recomendação e visualização. **RENOTE**, v. 17, n. 1, p. 286-295.
- VERBERT, Katrien et al. Learning analytics dashboard applications. **American Behavioral Scientist**, v. 57, n. 10, p. 1500-1509, 2013.
- VIEIRA, Marcos Alves; VEIGA, Ernesto Fonseca. Arvores de Decisao Aplicadas na Previsao de Desempenho de Alunos: Estado da Arte.
- WEBBER, Carine G. et al. Utilização de algoritmos de agrupamento na mineração de dados educacionais. **RENOTE**, v. 11, n. 1, 2013.
- WILLIAMS, Roy; KAROUSOU, Regina; MACKNESS, Jenny. Emergent learning and learning ecologies in Web 2.0. **The International Review of Research in Open and Distributed Learning**, v. 12, n. 3, p. 39-59, 2011.
- WISE, Alyssa Friend. Designing pedagogical interventions to support student use of learning analytics. In: **Proceedings of the fourth international conference on learning analytics and knowledge**. ACM, 2014. p. 203-211.
- ZAKI, Mohammed J.; MEIRA JR, Wagner; MEIRA, Wagner. **Data mining and analysis: fundamental concepts and algorithms**. Cambridge University Press, 2014.

Apêndice A - Projeto de Software usando UML

Neste Apêndice contém as informações do projeto de software com os artefatos: (a) requisitos do software; (b) casos de uso; (c) descrição dos casos de uso e fluxo de controle através de diagramas de sequência; (d) diagrama de classes e metadados das classes; (e) diagrama de componentes para apresentar a arquitetura do sistema, com três prismas diferentes, que são artefatos, componentes (arquitetura) e framework.

A00. Pacote de Requisitos

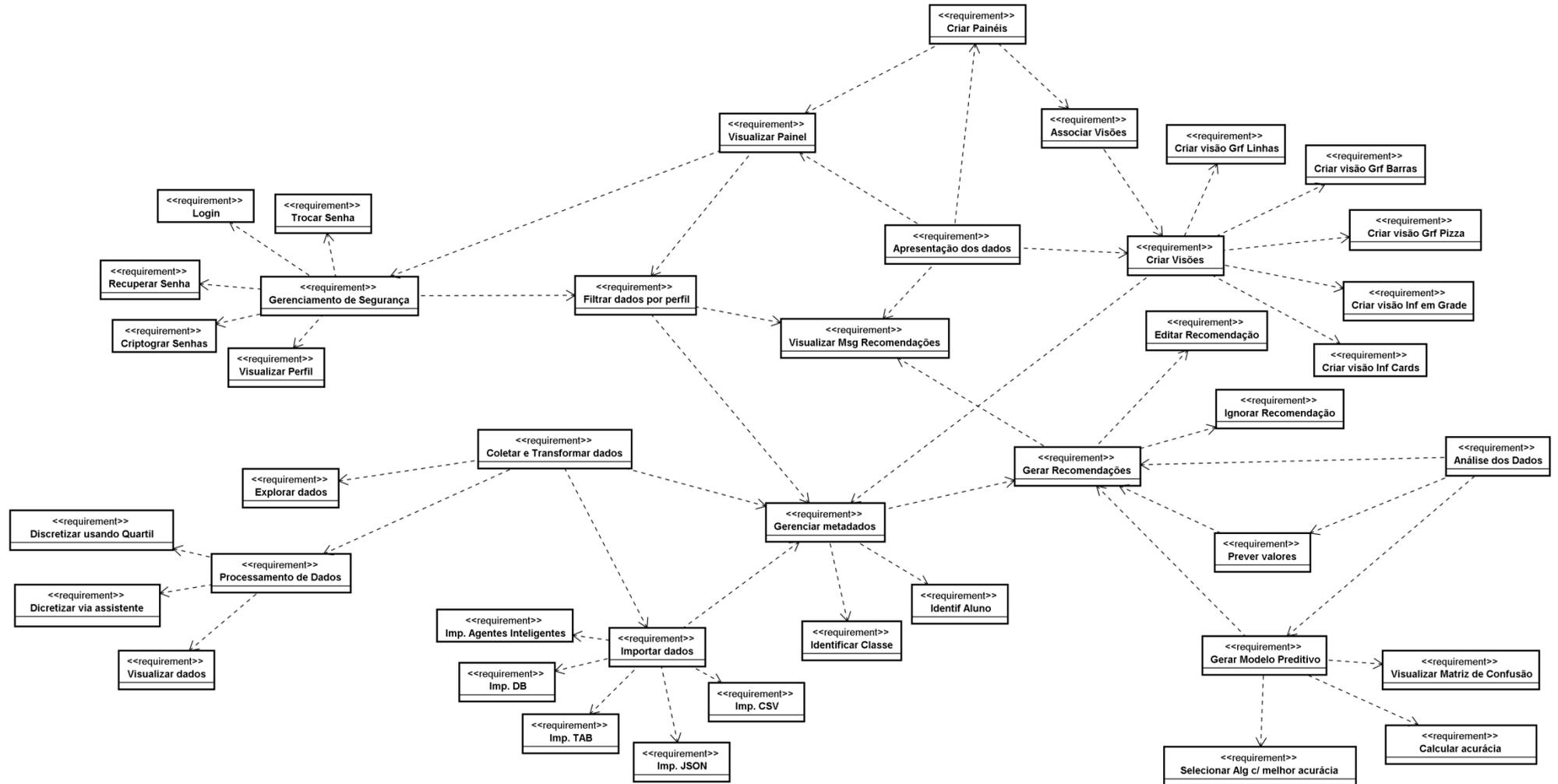
Quadro A01 - Requisitos do sistemas

Nome	Descrição
Análise dos Dados	Requisito Agrupado O sistema deverá permitir que os usuário possa analisar os dados acadêmicos.
Apresentação dos dados	Requisito Agrupado O sistema deverá permitir que seja visualizados os dados gerados ou importados.
Associar Visões	O sistema deverá permitir que sejam associadas as visões aos painéis.
Calcular acurácia	O sistema deverá permitir que sejam validadas as acurácias dos algoritmos.
Coletar e Transformar dados	Requisito Agrupado O sistema deverá permitir que sejam importados pelo sistema através das funcionalidades.
Criar Painéis	O sistema deverá permitir que sejam criados painéis de visualização das informações.
Criar Visões	O sistema deverá permitir que sejam criados visões dos dados.
Criar visão Grf Barras	
Criar visão Grf Linhas	
Criar visão Grf Pizza	
Criar visão Inf Cards	
Criar visão Inf em Grade	
Criptografar Senhas	Requisito não implementado.
Dicretizar via assistente	Requisito não implementado.

Nome	Descrição
Discretizar usando Quartil	O sistema deverá permitir a discretização dos dados usando quartil.
Editar Recomendação	O sistema deverá permitir que o usuário edite as mensagens de recomendações.
Explorar dados	O sistema deverá permitir que o usuário explore pelos dados.
Filtrar dados por perfil	O sistema deverá filtrar os dados confirma o perfil do usuário: A. Administrador P. Professor E. Estudante
Gerar Modelo Preditivo	O sistema deverá gerar os modelos preditivos: 1. Árvore de Decisão 2. KNN 3. SVM 4. MLP 5. Naive Bayes
Gerar Recomendações	O sistema deverá gerar as recomendações aos alunos classificados em risco.
Gerenciamento de Segurança	Requisito Agrupado O sistema deverá efetuar o gerenciamento de segurança
Gerenciar metadados	O sistema deverá permitir que seja feita a manutenção dos metadados das informações importadas
Identificar Aluno	O sistema deverá identificar se as informações correspondem ao estudante. Através dos metadados, será possível filtrar os dados para que o aluno somente visualize seus dados.
Identificar Classe	O sistema deverá identificar a coluna de classe. Ou seja, através desta informação será possível treinar os algoritmos e se estiver vazio os valores serão previstos pelo algoritmo de predição.
Ignorar Recomendação	O sistema deverá permitir ao professor ignorar determinada mensagem de recomendação.
Imp. Agentes Inteligentes	O sistema deverá permitir a importação automática via plugin Javascript.- Requisito não implementado
Imp. CSV	O sistema deverá permitir a importação de dados via arquivo CSV.
Imp. DB	O sistema deverá permitir a importação de dados via conexão de banco de dados.
Imp. JSON	O sistema deverá permitir a importação de dados via arquivo JSON.
Imp. TAB	O sistema deverá permitir a importação de dados via arquivo separado por TAB.
Importar dados	O sistema deverá permitir a importação de dados.
Entrar no Sistema	O sistema deverá permitir que o usuário entre no sistema através das credenciais login e senha.

Nome	Descrição
	O login poderá ser substituído pela matrícula, caso seja um estudante.
Prever valores	O sistema deverá fazer a previsão do desempenho do aluno através do algoritmo de predição usando o atributo/metadado identificado como classe no módulo de coleta de dados.
Processamento de Dados	O sistema deverá permitir a discretização de dados. Exemplo: Hora do dia: manhã/tarde/noite Idade (núm): bebê/jovem/adulto/idoso Temperatura (núm): frio/morno/quente
Recuperar Senha	O sistema deverá permitir que o usuário recupere a senha através de alguma identificação secreta.
Selecionar Alg c/ melhor acurácia	O sistema selecionará o algoritmo que tiver a melhor acurácia.
Trocar Senha	O sistema permitirá que o usuário troque a senha.
Visualizar Matriz de Confusão	O sistema deverá permitir que o usuário visualize a matriz de confusão gerada pelo algoritmo de predição.
Visualizar Msg Recomendações	O sistema deverá permitir que sejam visualizadas as mensagens de recomendação.
Visualizar Painel	O sistema deverá permitir que seja visualizado o painel (dashboard)
Visualizar Perfil	O sistema deverá permitir que o usuário visualize o próprio perfil.
Visualizar dados	O sistema permitirá que o usuário, ao selecionar a tabela e a coluna, visualize seus dados. Desta forma ficará fácil de identificar se a coluna tem valores discretos ou contínuos.

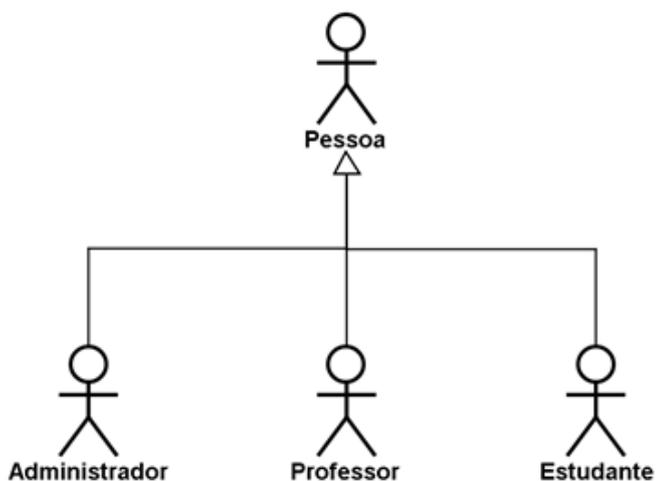
req [package] 00. Requisitos [Modelo]



A01. Casos de Uso

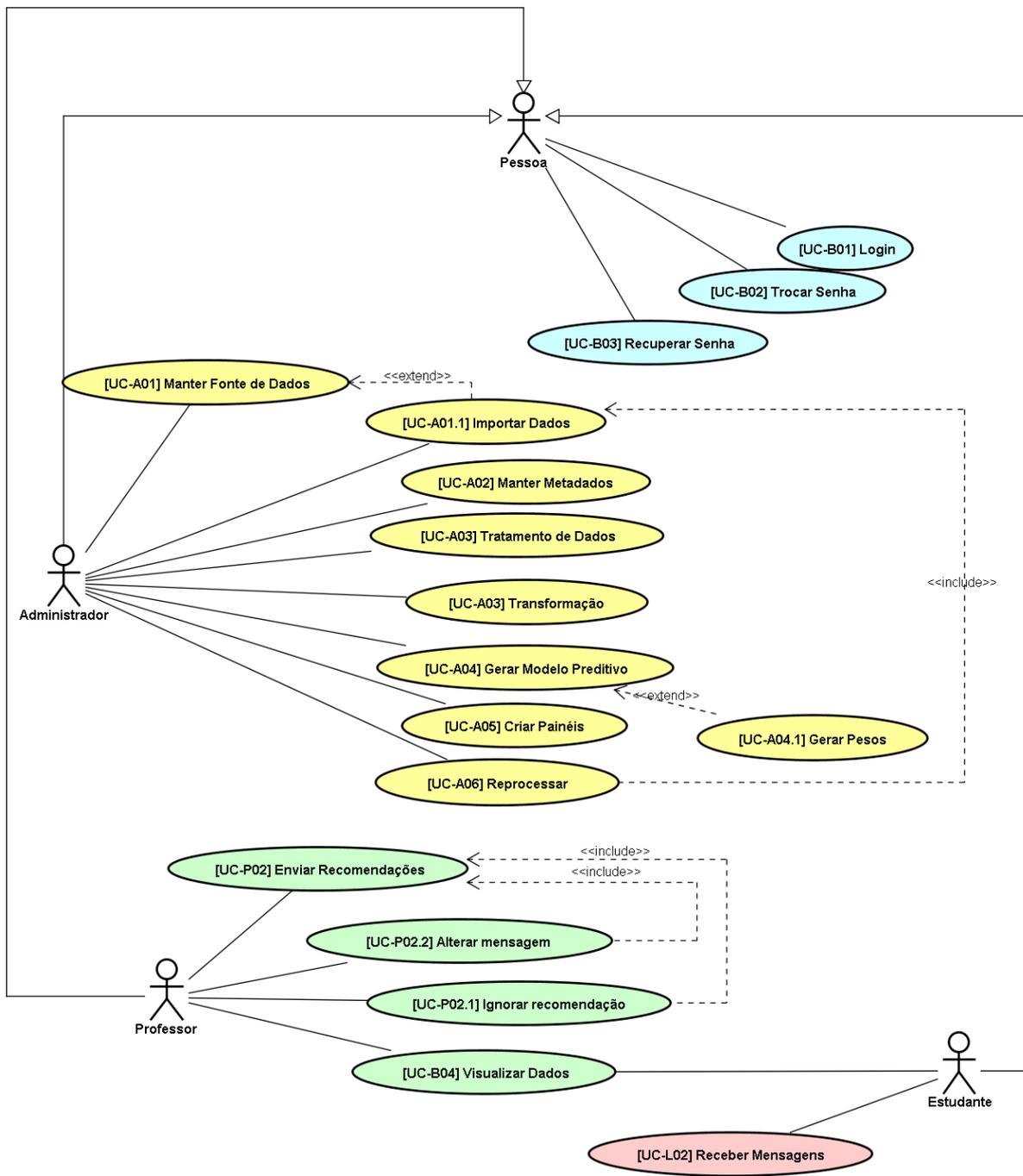
Pacote comportamental dos casos de uso com os fluxos representados através de diagramas de sequência.

A01.1 - Casos de Uso - Atores



Nome	Descrição
Administrador	Perfil responsável por configurar o sistema.
Estudante	Perfil dos alunos que receberão as recomendações e visualize seus dados através de painéis.
Pessoa	Perfil genérico que atua de forma comum a todos os outros perfis de usuário.
Professor	Perfil do professor que deverá julgar as mensagens de recomendações dos alunos e visualizar dados de sua turma através de painéis.

A01.2 - Casos de Uso - Básico

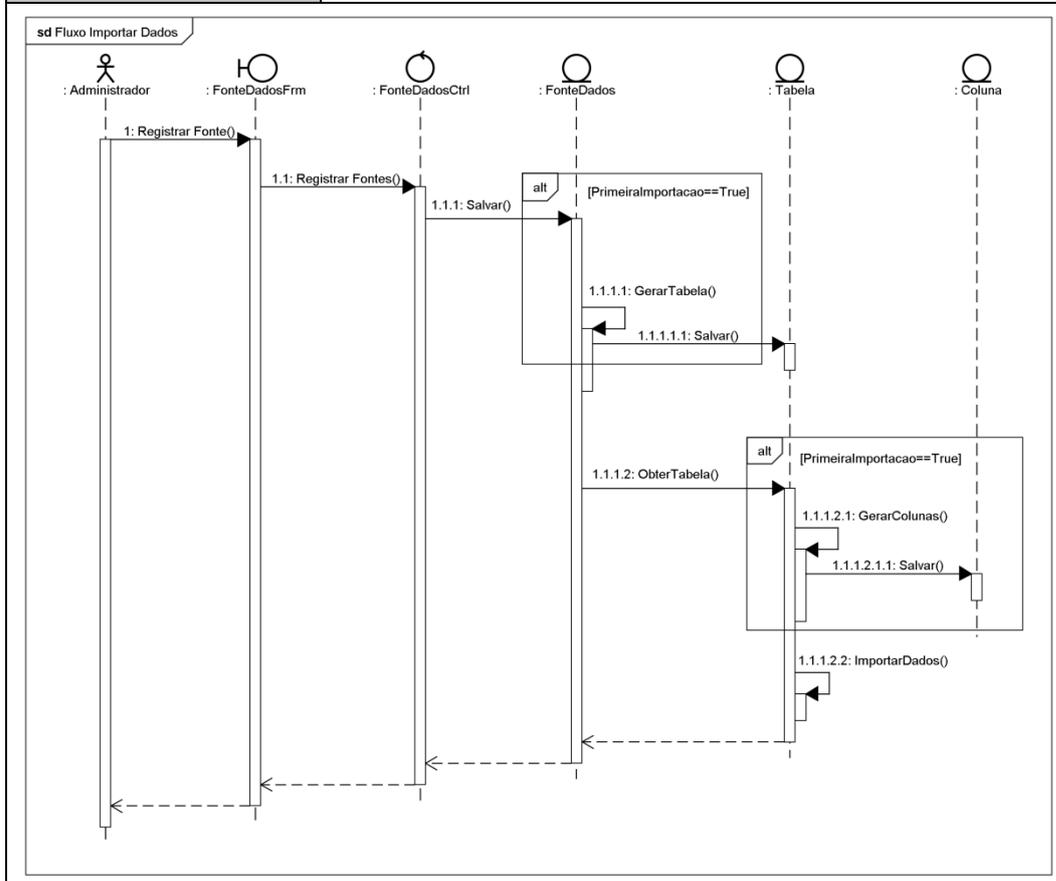


A01.3 - Descrição dos Casos de Uso

Caso de Uso	[UC-A01.1] Importar Dados
Atores(s)	Administrador
Descrição	O sistema deverá permitir que o usuário importe dados acadêmicos para dentro da base de dados.

Caso de Uso	[UC-P02] Enviar Recomendações
Atores(s)	Professor
Descrição	O sistema deverá permitir que o professor envie a mensagem de recomendações.

Caso de Uso	[UC-A01] Manter Fonte de Dados
Atores(s)	Administrador
Descrição	<ol style="list-style-type: none"> 1. O usuário deverá selecionar o tipo de fonte de dados, informar qual é o nome e a fonte de dados. E clicar na opção Registrar fonte de dados. 2. O sistema valida as informações. 3. Caso os dados estejam corretos o sistema salva o registro, caso contrário o sistema informa ao usuário. 4. O sistema salva os dados. 5. O sistema efetua a importação dos dados.

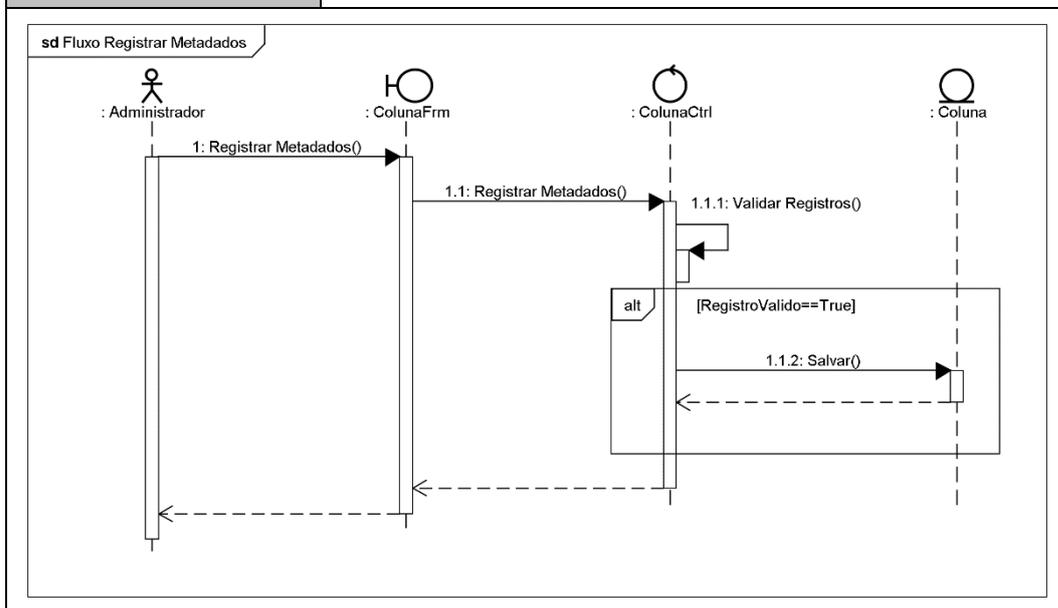


Caso de Uso	[UC-B02] Trocar Senha
Atores(s)	Pessoa
Descrição	O sistema deverá permitir que o usuário troque de senha.

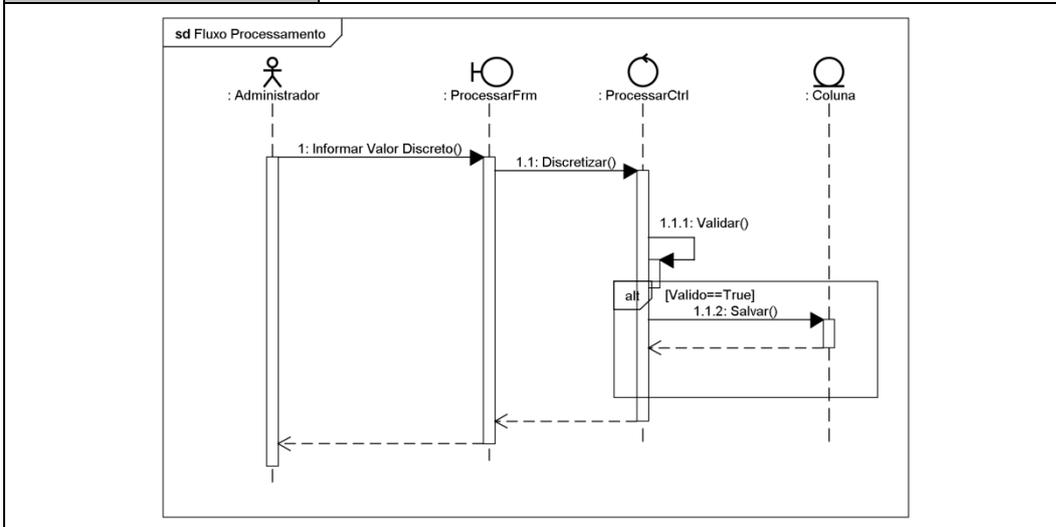
Caso de Uso	[UC-B03] Recuperar Senha
Atores(s)	Pessoa
Descrição	O sistema deverá permitir que o usuário recupere a senha.

Caso de Uso	[UC-B04] Visualizar Dados
Atores(s)	Estudante Pessoa
Descrição	O sistema deverá permitir que professores e alunos visualizem os dados acadêmicos através dos painéis gerados. O professor somente terá acesso às suas turmas. E o aluno somente terá acesso aos seus dados acadêmicos.

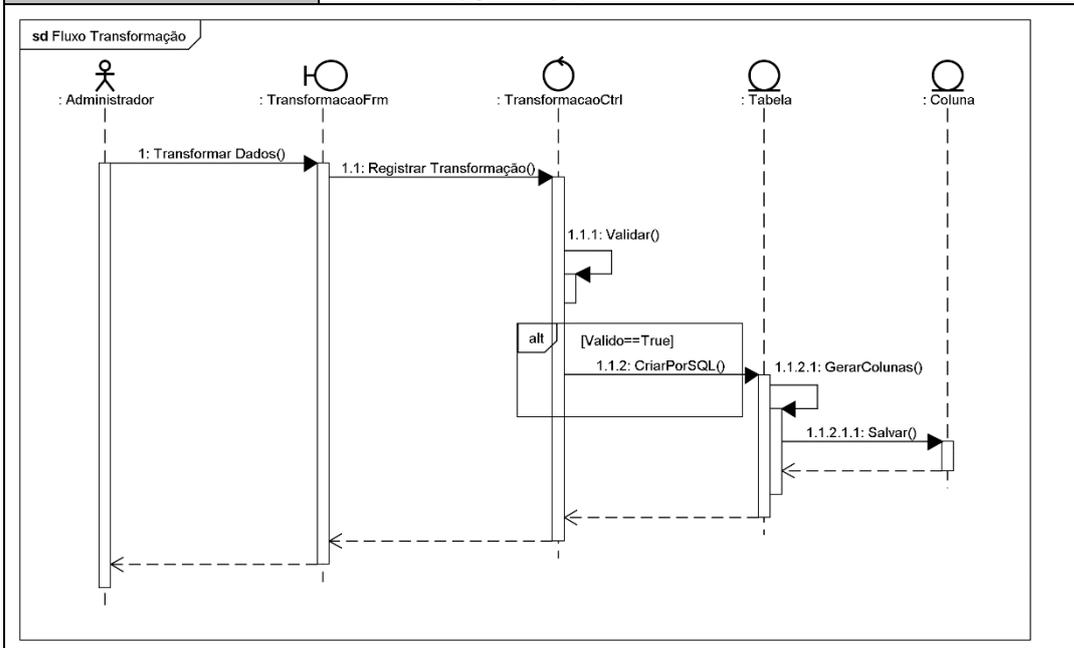
Caso de Uso	[UC-A02] Manter Metadados
Atores(s)	Administrador
Descrição	1. O usuário informa as alterações do metadado de uma determinada tabela. 2. O sistema valida os dados. 3. Caso os dados estejam corretos o sistema salva o registro, caso contrário o sistema informa ao usuário. 4. O sistema salva os dados.



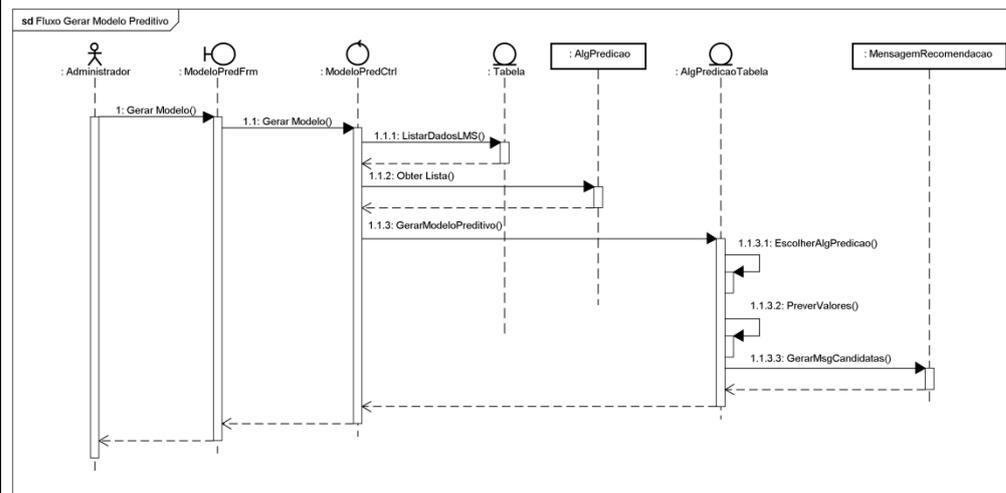
Caso de Uso	[UC-A03] Processamento
Atores(s)	Administrador
Descrição	<ol style="list-style-type: none"> 1. O usuário informa os parâmetros para discretizar a tabela. 2. O sistema valida os dados. 3. Caso os dados estejam corretos o sistema discretiza o campo, caso contrário o sistema informa ao usuário.



Caso de Uso	[UC-A03] Transformação
Atores(s)	Administrador
Descrição	<ol style="list-style-type: none"> 1. O usuário informa os dados para transformação. 2. O sistema valida as informações. 3. Caso os dados estejam corretos o sistema salva os dados, caso contrário o sistema informa ao usuário. 4. Caso a opção seja por SQL, então o sistema cria as tabelas



Caso de Uso	[UC-A04] Gerar Modelo Preditivo
Atores(s)	Administrador
Descrição	<ol style="list-style-type: none"> 1. O usuário informa os parâmetros para gerar o modelo preditivo. 2. O sistema lista os dados do AVA. 3. O sistema gera o modelo preditivo para os algoritmos. 4. O sistema gera as taxas de acurácia. 5. O sistema gera a tabela de xxx. 6. O sistema avalia a acurácia e seleciona o algoritmo com melhor acurácia. 7. O sistema prevê o resultado dos alunos. 8. O sistema gera as mensagens candidatas ao professor.

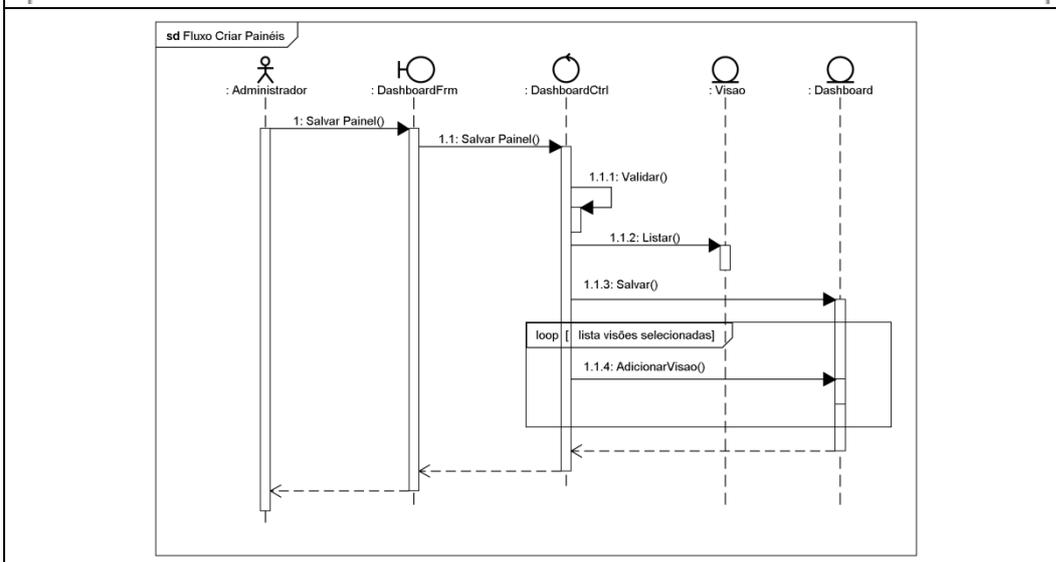
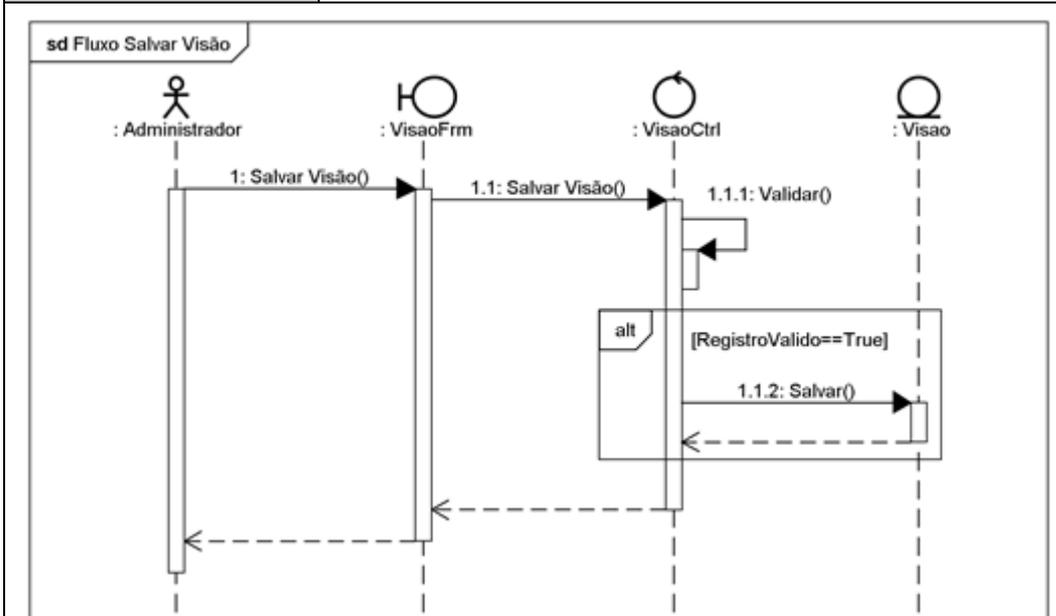


Caso de Uso	[UC-L02] Receber Mensagens
Atores(s)	Estudante
Descrição	O sistema deverá permitir que o aluno receba as recomendações dos professores
Caso de Uso	[UC-P02.1] Ignorar recomendação
Atores(s)	Professor
Descrição	O sistema deverá permitir que o professor ignore recomendações

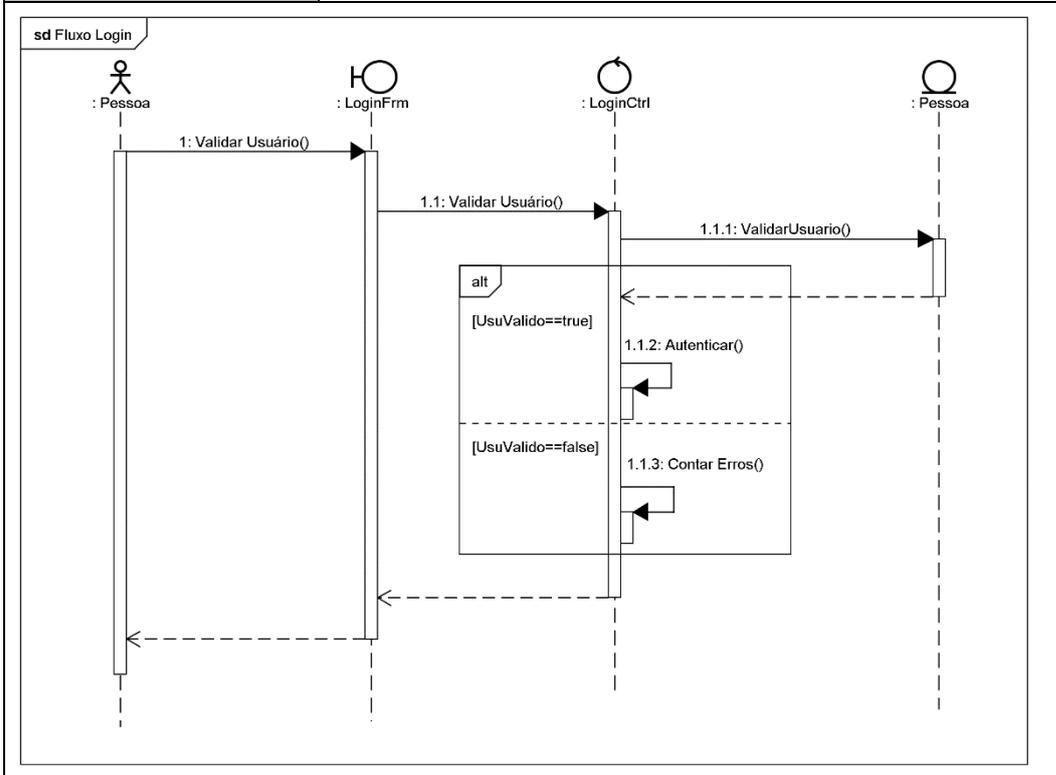
Caso de Uso	[UC-P02.2] Alterar mensagem
Atores(s)	Professor
Descrição	O sistema deverá permitir que o professor altere a mensagem de recomendação

Caso de Uso	[UC-A06] Reprocessar
Atores(s)	Administrador
Descrição	O sistema deverá permitir que o usuário reprocessasse os dados acadêmicos.

Caso de Uso	[UC-A05] Criar Painéis
Atores(s)	Administrador
Descrição	<p>A. Sub-fluxo Visão</p> <ol style="list-style-type: none"> 1. O usuário informa os dados do visão. 2. O sistema valida os dados. 3. Caso os dados estejam corretos o sistema salva o registro, caso contrário o sistema informa ao usuário. 4. O sistema salva os dados. <p>B. Sub-fluxo Painel</p> <ol style="list-style-type: none"> 1. O usuário informa os dados do painel. 2. O sistema valida os dados. 3. Caso os dados estejam corretos o sistema salva o registro, caso contrário o sistema informa ao usuário. 4. O sistema apresenta as visões ao administrador. 5. O usuário associa as visões ao painel.



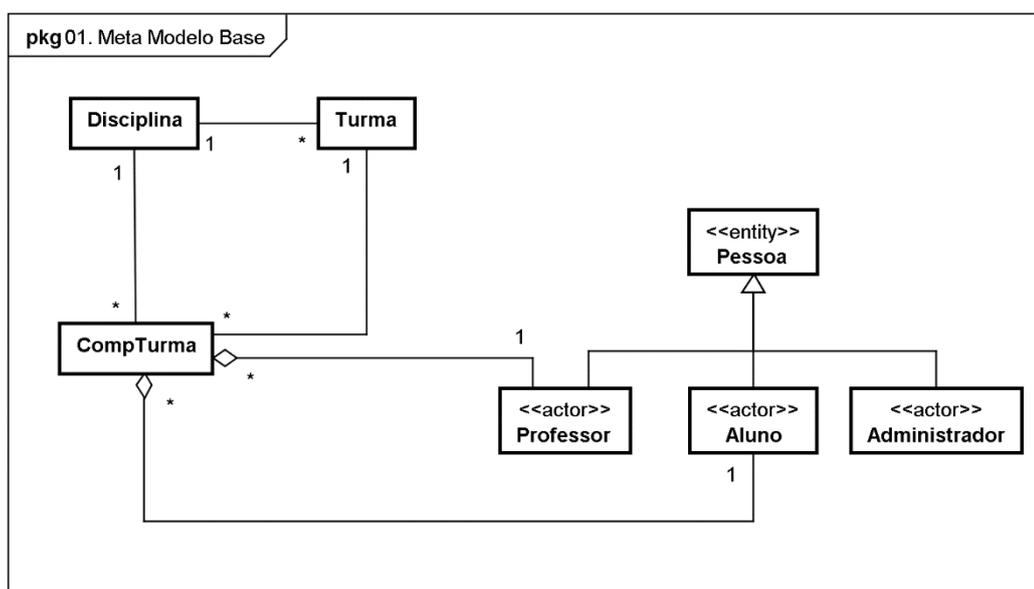
Caso de Uso	[UC-B01] Login
Atores(s)	Pessoa
Descrição	<ol style="list-style-type: none"> 1. O sistema apresenta a tela de login ao usuário. 2. O usuário informa as credenciais. 3. O sistema valida as informações. Caso as credenciais estejam inválidas o sistema informa ao usuário. 4. O sistema identifica o perfil do usuário (E: Estudante; P: Professor; A: Administrador).



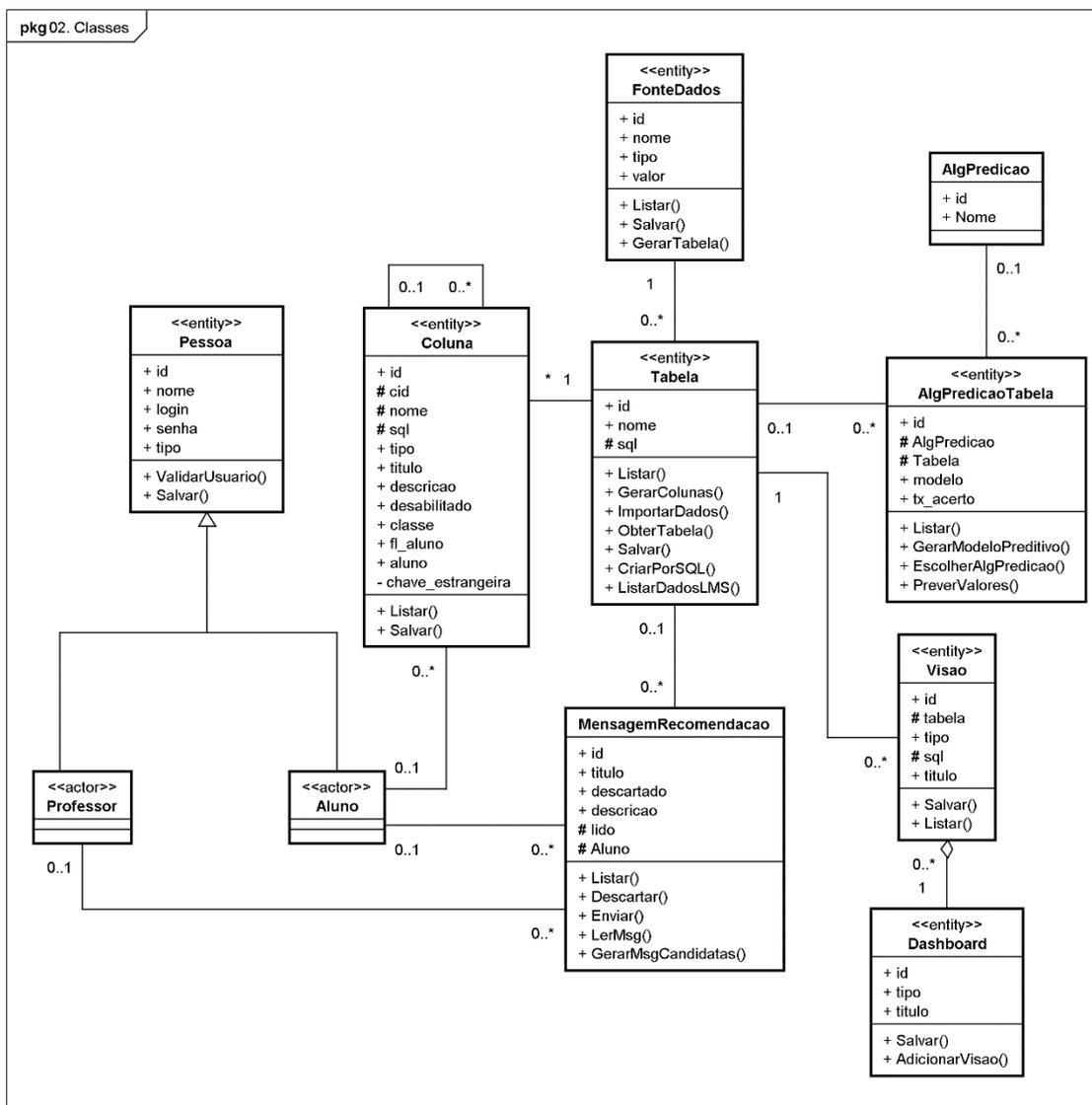
A02. Classes

Nome	Tipo	Descrição
Administrador	Actor	Cls: Pessoa do tipo Administrador
AlgPredicao	Class	Tabela de domínio do analista. Algoritmos de Predição: * Decision Tree-based * K-nearest Neighbors * Support Vector Machine * Multi-layer Perceptron Classifier * Nayve Bayes
AlgPredicaoTabela	Entity	Algoritmos vinculados à tabela
Aluno	Actor	Cls: Pessoa do tipo Aluno
Coluna	Entity	Metadados - colunas das tabelas importadas
Dashboard	Entity	Painel do Professor ou Aluno
FonteDados	Entity	Fonte de dados proveniente dos dados do AVA
MensagemRecomendacao	Class	Mensagem gerada para recomendação aos alunos
Pessoa	Entity	Pessoa do sistema
Professor	Actor	Cls: Pessoa do tipo Professor
Tabela	Entity	Entidade: Tabela
Visao	Entity	Parte do painel (do aluno ou professor)

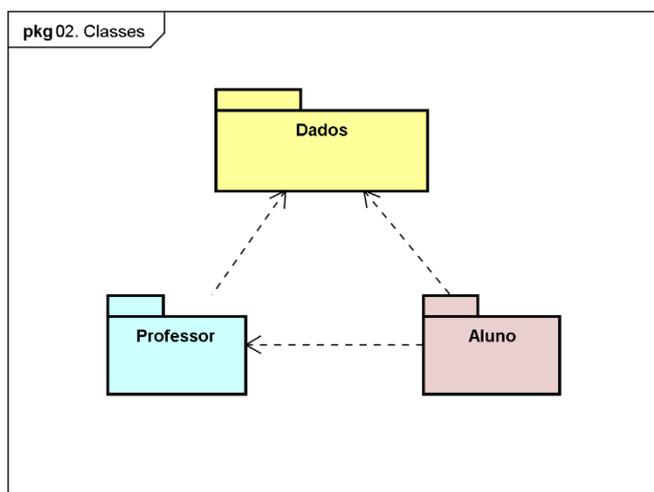
A02.1 - Diagramas de Classes - Base para análise de dados



A02.2 - Diagrama de Classes - Interno do sistema



A02.3 - Pacotes



A02.4 - Descrição das Classes

AlgPredicao [Class]	
Atributos	
Nome	Descrição
id	Identificador da entidade
Nome	Nome do algoritmo

AlgPredicaoTabela [Entity]	
Atributos	
Nome	Descrição
id	Identificação da entidade
AlgPredicao	Entidade: Algoritmo de Predição
Tabela	Entidade: tabela do sistema
modelo	Modelo Classificador
tx_acerto	Taxa de acertos
Operações	
Nome	Descrição
Listar	Listar dados
GerarModeloPreditivo	Gerar modelo preditivo
EscolherAlgPredicao	Escolher o algoritmo com melhor taxa de acerto
PreverValores	Prever modelos

Coluna [Entity]	
Atributos	
Nome	Descrição
id	Identificação da entidade
cid	Código da Coluna (externo)
nome	Nome da coluna
sql	SQL de processamento. Padrão: nome da coluna
tipo	T: texto; N: números; D: sequências temporais
titulo	Título (visível) da coluna
descricao	Descrição da coluna (tooltip)
desabilitado	Coluna desabilitada (True: Sim; False: Não)
classe	É classe. Se True então servirá de valor para treino. Quando Null este valor será previsto pelo algoritmo com melhor taxa de acertos.
fl_aluno	Identificador do aluno.
aluno	Entidade: Aluno
chave_estrangeira	Indicativo de chave estrangeira
Operações	
Nome	Descrição
Listar	Listar Colunas
Salvar	Salvar metadados da coluna da tabela

Dashboard [Entity]	
Atributos	
Nome	Descrição
id	Identificador da entidade
tipo	Tipo de Dashboard (A: Aluno; P: Professor)
titulo	Título do Painel
Operações	
Nome	Descrição
Salvar	Salvar dados
AdicionarVisao	Adicionar Visão

FonteDados [Entity]	
Atributos	
Nome	Descrição
id	Identificação da entidade
nome	Nome da fonte de dados
tipo	Tipos de fontes de dados: <ul style="list-style-type: none"> • JSON: acrônimo de JavaScript Object Notation • CSV: arquivo separado por vírgula • Banco de dados: conexão direta
valor	Valor da String de Conexão. Se for um arquivo deverá ser apresentado o fullpath do arquivo.
Operações	
Nome	Descrição
Listar	Listar fontes de dados
Salvar	Salvar fontes de dados
GerarTabela	Gerar Entidade(s) Tabela(s)

MensagemRecomendacao [Class]	
Atributos	
Nome	Descrição
id	Identificador da entidade
titulo	Título da mensagem
descartado	Descartado pelo professor
descricao	Corpo da Mensagem
lido	Lido pelo aluno
Aluno	Entidade: Aluno
Operações	
Nome	Descrição
Listar	Listar mensagens
Descartar	Descartar pelo professor
Enviar	Enviar para o aluno
LerMsg	Ler mensagem
GerarMsgCandidatas	Gerar mensagem candidata (através do algoritmo de predição)

Pessoa [Entity]	
Atributos	
Nome	Descrição
id	Identificador da entidade
Nome	Nome da pessoa
Login	Login para acessar o sistema
Senha	Senha para acessar o sistema
Tipo	Tipo de Pessoa. - P: Professor; - E: Estudante; - A: Analista
Operações	
Nome	Descrição
ValidarUsuario	Validar Usuário
Salvar	Salvar genérico

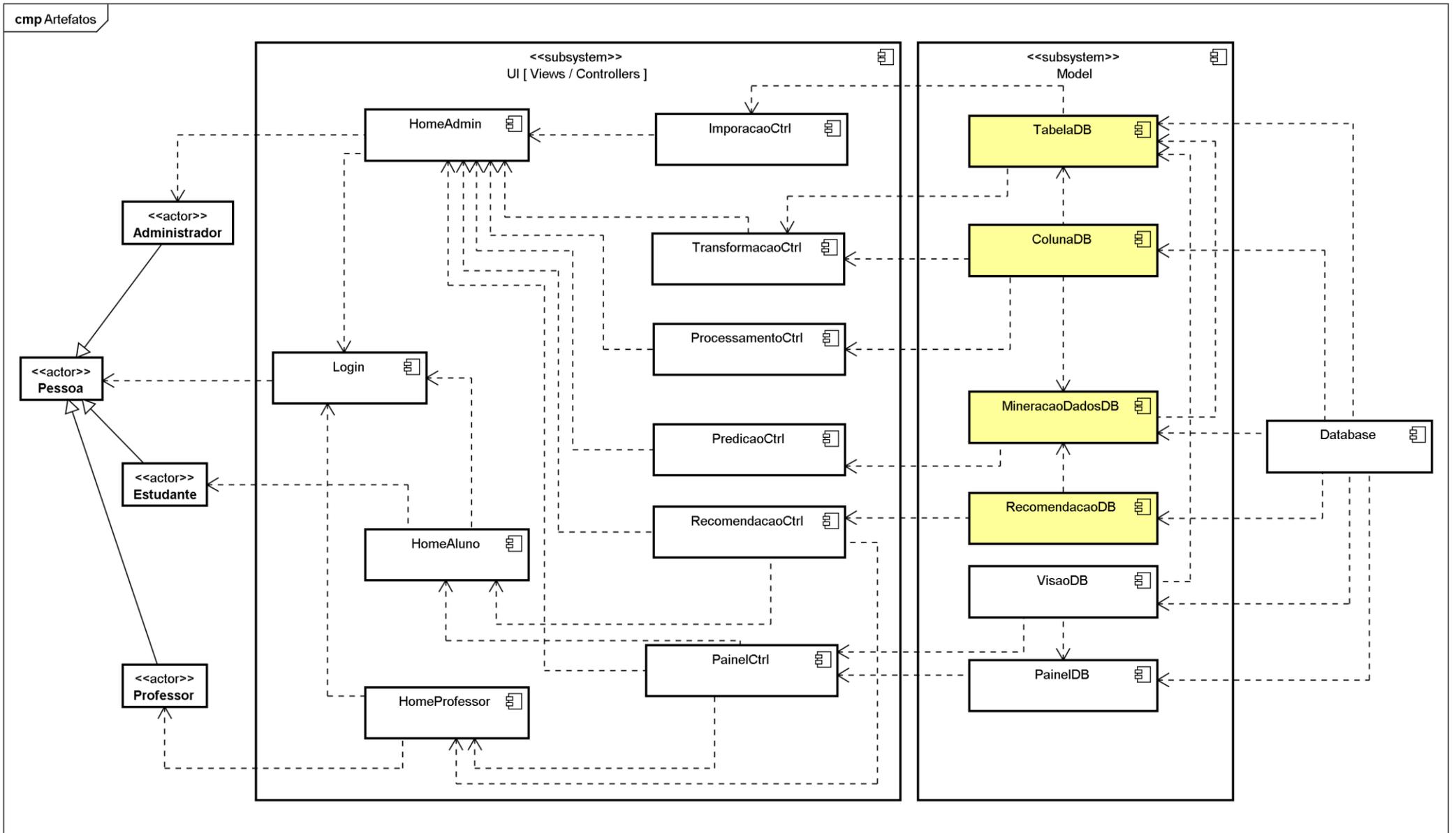
Tabela [Entity]	
Atributos	
Nome	Descrição
id	Identificador da entidade
nome	Nome da tabela
sql	SQL para gerar a tabela (quando feita pelo processo de transformação)
Operações	
Nome	Descrição
Listar	Listar tabelas
GerarColunas	Gerar colunas. Realizado no processo: - inicial de importação de dados; - transformação.
ImportarDados	Importar dados conforme parametro da fonte de dados
ObterTabela	Obter tabela
Salvar	Salvar registro da tabela
CriarPorSQL	Criar por SQL - Transformação
ListarDadosLMS	Listar dados do LMS (AVA)

Visao [Entity]	
Atributos	
Nome	Descrição
id	Identificador da entidade
tabela	Entidade: Tabela
tipo	Tipo de visão: - 1: Gráfico de Pizza - 2: Gráfico Barras - 3: Gráfico Linha - 4: Informação em Cards - 5: Informação em Grid
sql	SQL para gerar a visão
titulo	Título da Visão
Operações	
Nome	Descrição
Salvar	Salvar visão
Listar	Listar visões

A03. Diagramas de Componentes

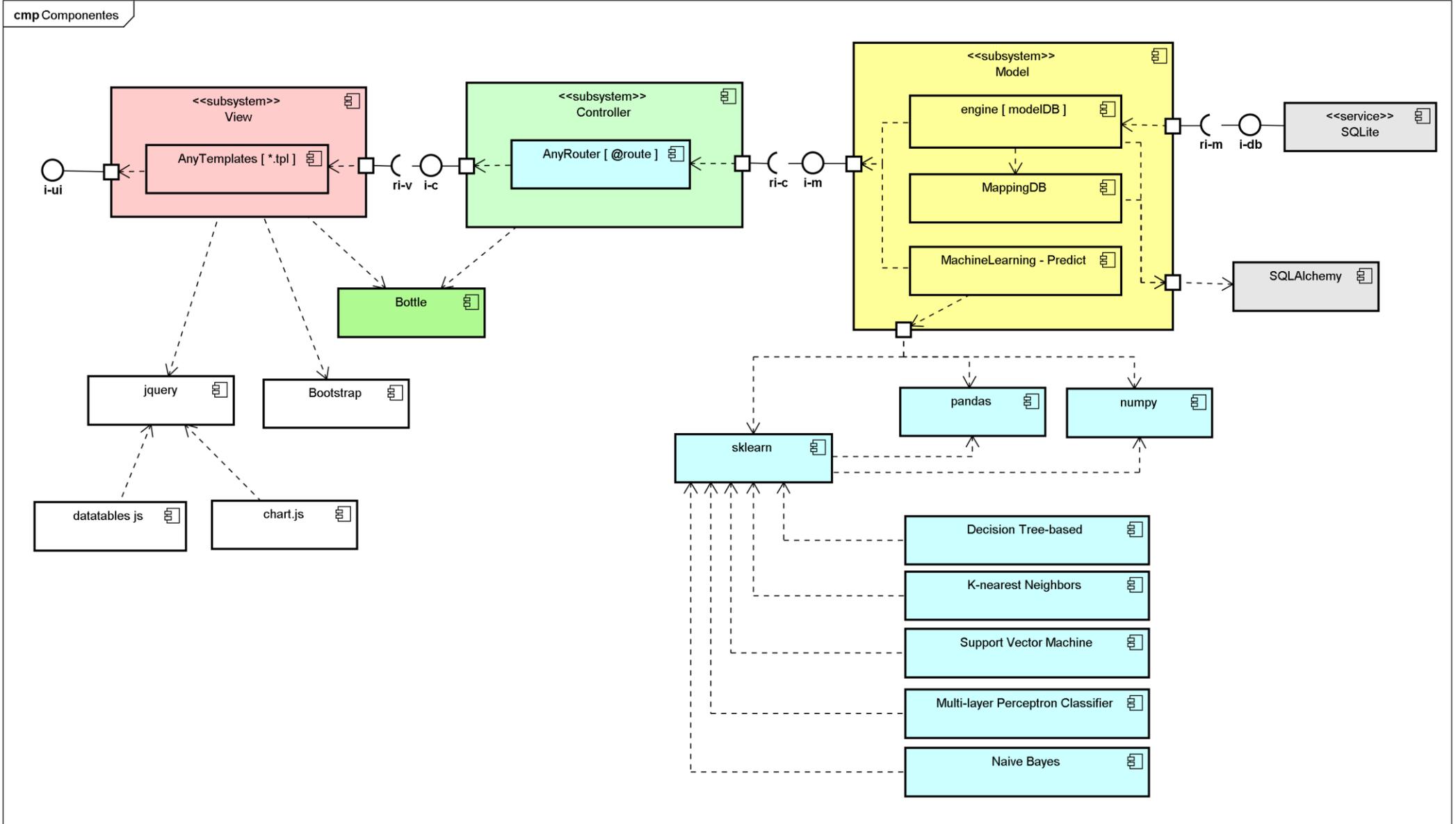
A03.1 - Artefatos

Nome	Descrição
ColunaDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade Coluna.
Database	Banco de Dados
HomeAdmin	Controller Home do Administrador
HomeAluno	Controller Home do Aluno
HomeProfessor	Controller Home do Professor
ImporacaoCtrl	Controller Importação / Fonte de Dados / Metadados
Login	Controller Login
MineracaoDadosDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade AlgPredicao; AlgPredicaoTabela; e MensagemRecomendacao.
Model	Pacote que contém todas as informações do modelo, parte estática do sistema e lógica de acesso aos dados.
PainelCtrl	Controller Painel e Visão
PainelDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade Painel
PredicaoCtrl	Controller Predição
ProcessamentoCtrl	Controller Pré-Processamento
RecomendacaoCtrl	Controller Recomendação
RecomendacaoDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade MensagemRecomendacao
TabelaDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade Tabela.
TransformacaoCtrl	Controller Transformação
UI [Views / Controllers]	Subsistema Visão e Controller - Mapeamento dos
VisaoDB	Mapeamento ORM e Lógica de acesso aos dados: Entidade Visao



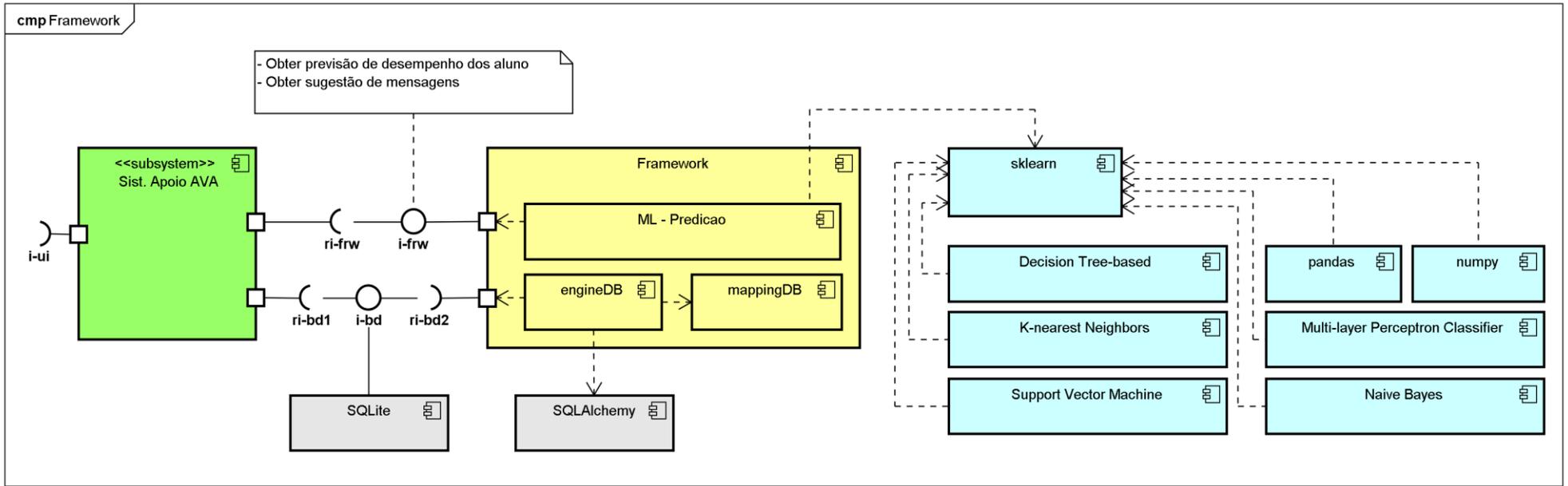
A03.2 - Componentes

Nome	Type	Descrição
AnyRouter [@route]	Component	Rotas mapeadas pelo @route
AnyTemplates [*.tpl]	Component	Templates (Visão)
Bootstrap	Component	Biblioteca JS
Bottle	Component	Framework Web - Python
Controller	Component	Subsistema Controller
Decision Tree-based	Component	ML: Árvore de Decisão
K-nearest Neighbors	Component	ML: KNN
MachineLearning - Predict	Component	Módulo de Predição
MappingDB	Component	Mapeamento Entidade Relacionamento x Objeto
Model	Component	Subsistema Model Todas as regras de acesso a dados e ORM
Multi-layer Perceptron Classifier	Component	ML: MLP
Naive Bayes	Component	ML: Naive Bayes
SQLAlchemy	Component	Framework ORM de acesso ao banco de dados
SQLite	Component	Banco de Dados
Support Vector Machine	Component	ML: SVM
View	Component	Subsistema Visão
chart.js	Component	Biblioteca JS
datatables.js	Component	Biblioteca JS
engine [modelDB]	Component	Mecanismo de acesso ao Banco de Dados
i-c	Interface	Interface provida pelo Controller para a View
i-db	Interface	Interface provida pelo Banco de Dados para o Model
i-m	Interface	Interface provida pelo Model para o Controller
i-ui	Interface	Interface provida para o usuário
jquery	Component	Biblioteca JS
numpy	Component	Biblioteca para manipulação de dados N-Dimensional
pandas	Component	Biblioteca de tratamento de dados usando DataFrames
ri-c	Interface	Interface requerida pelo Controller para o Model
ri-m	Interface	Interface requerida pelo Model para o Banco de Dados
ri-v	Interface	Interface requerida pela View para o Controller
sklearn	Component	Biblioteca de aprendizado de Máquina



A03.3 - Framework

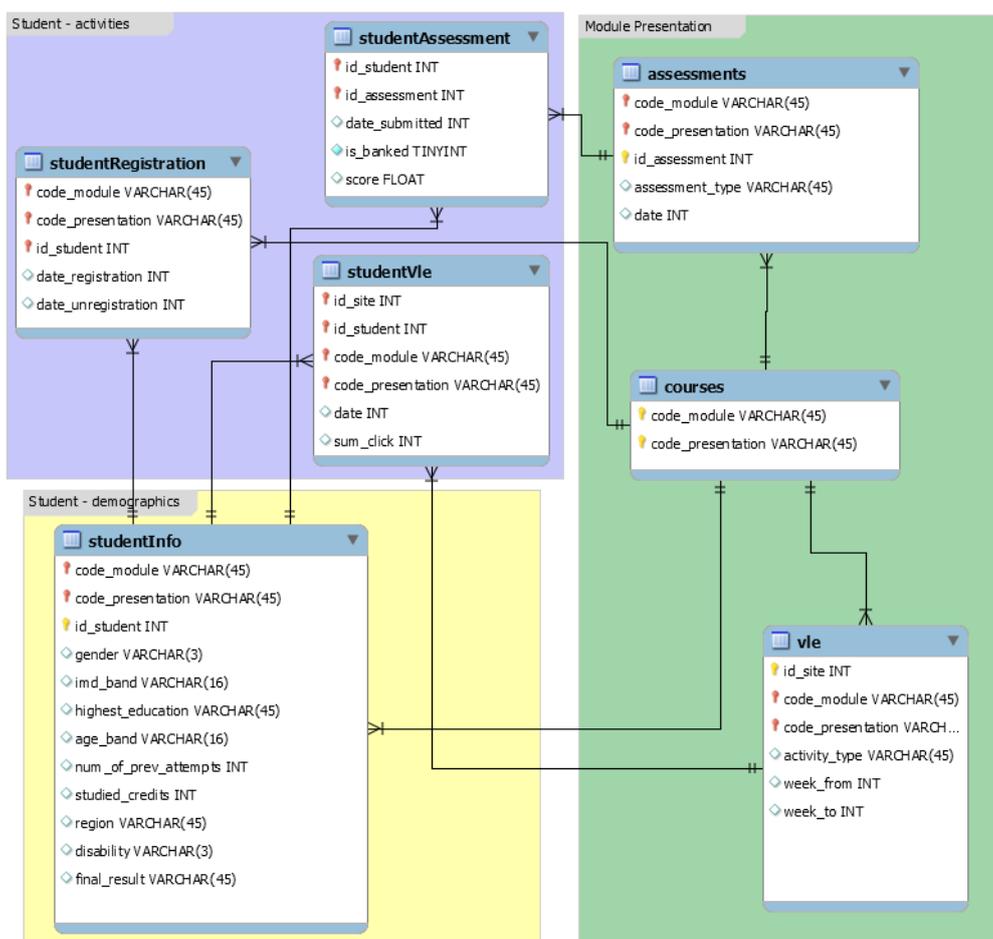
Nome	Type	Descrição
Decision Tree-based Framework	Component	ML: Árvore de Decisão
K-nearest Neighbors	Component	ML: KNN
ML - Predicao	Component	Módulo de Predição
Multi-layer Perceptron Classifier	Component	ML: MLP
Naive Bayes	Component	ML: Naive Bayes
SQLAlchemy	Component	Framework ORM de acesso ao banco de dados
SQLite	Component	Banco de Dados
Sist. Apoio AVA	Component	Sistema de apoio, responsável pela coleta dos dados e gerência dos metadados
Support Vector Machine	Component	ML: SVM
engineDB	Component	Mecanismo de acesso ao Banco de Dados
i-bd	Interface	Interface provida pelo Banco de Dados
i-frw	Interface	Interface provida pelo Framework para o SisApoio
i-ui	Interface	Interface provida para o usuário
mappingDB	Component	Mapeamento Entidade Relacionamento x Objeto
numpy	Component	Biblioteca para manipulação de dados N-Dimensional
pandas	Component	Biblioteca de tratamento de dados usando DataFrames
ri-bd1	Interface	Interface requerida pelo SisApoio para o Banco de Dados
ri-bd2	Interface	Interface requerida pelo Framework para o Banco de Dados
ri-frw	Interface	Interface requerida pelo SisApoio para o Framework
sklearn	Component	Biblioteca de aprendizado de Máquina



Apêndice B - Descrição do Dataset

Neste apêndice é descrito o conjunto de dados OULA, *Open University Learning Analytics*. Nele contém dados sobre cursos, alunos e suas interações com o *Virtual Learning Environment (VLE)* para sete cursos selecionados (chamados módulos). O conjunto de dados OULAD é uma coleção de dados tabulares de alunos dos anos de 2013 e 2014. Cada tabela contém informações diferentes, que podem ser vinculadas aos dados de outras tabelas usando colunas identificadoras. O conjunto de dados é orientado para o aluno, como ponto central. Os dados dos alunos incluem informações sobre suas informações demográficas e registros para os módulos. Para cada trio de apresentação do módulo do aluno, o conjunto de dados contém os resultados das avaliações dos alunos. As interações dos alunos com o VLE são registradas como um resumo de suas atividades diárias. O conjunto de dados contém 22 apresentações de módulo com 32.593 alunos.

Database schema



courses.csv

Arquivo contém a lista de todos os módulos disponíveis e suas apresentações. As colunas são:

- code_module – nome de código do módulo, que serve como o identificador.
- code_presentation – nome de código da apresentação. Consiste no ano e “B” para a apresentação a partir de fevereiro e “J” para a apresentação a partir de outubro.
- length – duração do módulo - apresentação em dias.

A estrutura das apresentações B e J pode diferir e, portanto, é uma boa prática analisar as apresentações B e J separadamente. No entanto, para algumas apresentações, a apresentação anterior B / J correspondente não existe e, portanto, a apresentação J deve ser usada para informar a apresentação B ou vice-versa. No conjunto de dados, este é o caso dos módulos CCC, EEE e GGG.

studentVle.csv

O arquivo studentVle.csv contém informações sobre as interações de cada aluno com os materiais no AVA. Este arquivo contém as seguintes colunas:

- code_module – um código de identificação para um módulo.
- code_presentation – o código de identificação da apresentação do módulo.
- id_student – um número de identificação único para o aluno.
- id_site – um número de identificação para o material VLE.
- date – a data da interação do aluno com o material medido como o número de dias desde o início do módulo-apresentação.
- sum_click – o número de vezes que um aluno interage com o material naquele dia.

vle.csv

O arquivo csv contém informações sobre os materiais disponíveis no VLE. Normalmente, são páginas em HTML, arquivos em pdf, etc. Os alunos têm acesso a esses materiais on-line e suas interações com os materiais são registradas. O arquivo vle.csv contém as seguintes colunas:

- id_site – um número de identificação do material.
- code_module – um código de identificação para o módulo.
- code_presentation – o código de identificação da apresentação.
- activity_type – a função associada ao material do módulo.
- week_from – a semana a partir da qual o material é planejado para ser usado.
- week_to – semana até que o material seja planejado para ser usado.

studentInfo.csv

Este arquivo contém informações demográficas sobre os alunos e seus resultados. Arquivo contém as seguintes colunas:

- code_module – um código de identificação para um módulo no qual o aluno está registrado.
- code_presentation – o código de identificação da apresentação durante o qual o aluno está registrado no módulo.
- id_student – um número de identificação único para o aluno.
- gender – o gênero do aluno.
- region – identifica a região geográfica, onde o estudante viveu enquanto fazia o módulo-apresentação.
- highest_education – maior nível de escolaridade do aluno na entrada para a apresentação do módulo.
- imd_band – especifica a faixa de Índice de Múltiplas locais onde o aluno morou durante a apresentação do módulo.
- age_band – band of the student's age.
- num_of_prev_attempts – o número de vezes que o aluno tentou este módulo.
- studied_credits – o número total de créditos para os módulos que o aluno está estudando atualmente.
- disability – indica se o aluno declarou uma deficiência.
- final_result – o resultado final do aluno no módulo de apresentação.

assessments.csv

Este arquivo contém informações sobre avaliações em apresentações de módulo. Normalmente, cada apresentação tem um número de avaliações seguidas pelo exame final. CSV contém colunas:

- code_module – código de identificação do módulo, ao qual pertence a avaliação.
- code_presentation – código de identificação da apresentação, ao qual pertence a avaliação.
- id_assessment – número de identificação da avaliação.
- assessment_type – tipo de avaliação. Existem três tipos de avaliações: Avaliação Marcada pelo Tutor (TMA), Avaliação Marcada por Computador (CMA) e Exame Final (Exame).
- date – informações sobre a data final da apresentação da avaliação calculada como o número de dias desde o início do módulo-apresentação. A data de início da apresentação tem o número 0 (zero).
- weight - peso da avaliação em%. Normalmente, os exames são tratados separadamente e têm o peso de 100%; a soma de todas as outras avaliações é de 100%.

Se a informação sobre a data do exame final estiver faltando, é no final da última semana de apresentação.

studentRegistration.csv

Este arquivo contém informações sobre o horário em que o aluno se registrou para a apresentação do módulo. Para os alunos que não registraram a data do cancelamento, também é registrado. Arquivo contém cinco colunas:

- code_module – um código de identificação para um módulo.
- code_presentation – o código de identificação da apresentação.
- id_student – um número de identificação único para o aluno.
- date_registration – a data do registro do aluno na apresentação do módulo, esse é o número de dias medidos em relação ao início da apresentação do módulo (por exemplo, o valor negativo -30 significa que o aluno se inscreveu na apresentação do módulo 30 dias antes de começar).
- date_unregistration – data de cancelamento do registro do aluno na apresentação do módulo, esse é o número de dias medidos em relação ao início da apresentação do módulo. Os alunos que completaram o curso têm este campo vazio. Os alunos que não foram registrados têm Retirada como o valor da coluna final_result no arquivo studentInfo.csv.

studentAssessment.csv

Este arquivo contém os resultados das avaliações dos alunos. Se o aluno não enviar a avaliação, nenhum resultado será registrado. As submissões do exame final estão faltando, se o resultado das avaliações não for armazenado no sistema. Este arquivo contém as seguintes colunas:

- id_assessment – o número de identificação da avaliação.
- id_student – um número de identificação único para o aluno.
- date_submitted – a data de envio do aluno, medida como o número de dias desde o início da apresentação do módulo.
- is_banked – um sinalizador de status indicando que o resultado da avaliação foi transferido de uma apresentação anterior.
- score – a pontuação do aluno nesta avaliação. O intervalo é de 0 a 100. A pontuação inferior a 40 é interpretada como Falha. As marcas estão na faixa de 0 a 100.