



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
DE COMPUTAÇÃO

ANDRÉ ALVES BEZERRA

PREDIÇÃO DE DESEMPENHO DE ESTUDANTES EM AMBIENTES
VIRTUAIS DE APRENDIZAGEM

SOBRAL

2017

ANDRÉ ALVES BEZERRA

PREDIÇÃO DE DESEMPENHO DE ESTUDANTES EM AMBIENTES VIRTUAIS DE
APRENDIZAGEM

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da Universidade Federal do Ceará, como requisito parcial para obtenção do título de mestre em Engenharia Elétrica e de Computação. Área de Concentração: Sistemas de Informação

Orientadora: Prof^a. Dr^a. Gisele Azevedo de Araújo Freitas

SOBRAL

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

B469p Bezerra, André Alves.

Predição de Desempenho de Estudantes em Ambientes Virtuais de Aprendizagem / André Alves
Bezerra. – 2017.
73 f. : il.

Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Sobral, Programa de Pós-Graduação
em Engenharia Elétrica e de Computação, Sobral, 2017.

Orientação: Profa. Dra. Gisele Azevedo de Araújo Freitas.

1. Ambientes Virtuais de Aprendizagem. 2. Modelos Preditivos. 3. Mineração de Dados Educacionais. I.
Título.

CDD 621.3

ANDRÉ ALVES BEZERRA

PREDIÇÃO DE DESEMPENHO DE ESTUDANTES EM AMBIENTES VIRTUAIS DE
APRENDIZAGEM

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da Universidade Federal do Ceará, como requisito parcial para obtenção do título de mestre em Engenharia Elétrica e de Computação. Área de Concentração: Sistemas de Informação

Aprovada em: 20 de Janeiro de 2017

BANCA EXAMINADORA

Prof^a. Dr^a. Gisele Azevedo de Araújo
Freitas (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Iális Cavalcante de Paula Júnior
Universidade Federal do Ceará (UFC)

Prof. Dr. José Cláudio do Nascimento
Universidade Federal do Ceará (UFC)

Prof. Dr. Carlos Estêvão Rolim Fernandes
Universidade Federal do Ceará (UFC)

À Thamar Ruth, por ser minha companheira
paciente, meu amor e melhor amiga durante
todo o desenvolvimento deste trabalho.

AGRADECIMENTOS

Agradeço a Deus pela vida e força que me deu para realizar o presente trabalho.

À minha família pelo apoio e orações.

À professora Gisele Azevedo de Araújo Freitas, pela paciência e orientação durante todas as atividades realizadas no Mestrado.

Ao professor Jarbas Joaci Mesquita de Sá Junior, pela didática e dedicação na disciplina de Reconhecimentos de Padrões. Os conhecimentos obtidos nesta disciplina contribuíram para que esta Dissertação fosse possível.

Aos colegas Rosângela, Darielson e Bruno, pelas trocas de experiências e sugestões durante a realização das disciplinas.

Aos amigos Éder, Anderson e Danilo, pelo incentivo contínuo.

Ao INTA, pela disponibilização dos dados utilizados na pesquisa deste trabalho.

RESUMO

Neste trabalho, investigamos a viabilidade do uso de técnicas de Mineração de Dados Educacionais baseadas na combinação de métodos de Redução da Dimensionalidade com Análise dos Componentes Principais (PCA), e Análise Discriminante Linear (LDA) para pré-processamento dos dados, e Redes Neurais Artificiais do tipo *Multilayer Perceptron* (MLP) para realizar inferências sobre o desempenho acadêmico de estudantes a partir de dados de bases de dados de Ambientes Virtuais de Aprendizagem (AVA). AVA são *softwares* educacionais baseados na web que disponibilizam conteúdos didáticos e avaliações *on-line*, e geram um grande volume de dados sobre interações dos estudantes no ambiente. O método que propomos para classificação dos estudantes, com os dados utilizados neste trabalho, obteve melhores resultados na predição do desempenho acadêmico de estudantes se comparado a outros métodos abordados na literatura.

Palavras-chave: Ambientes Virtuais de Aprendizagem. Redução da Dimensionalidade. Redes Neurais. Mineração de Dados Educacionais.

ABSTRACT

In this work, we investigate the using of Educational Data Mining (EDM) techniques based on Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) for data reduction and Multilayer Perceptron (MLP) to predict student's performance on Learning Environments Systems (LMS). LMS are web-based educational software that provides classes and online assignments, and generate a large volume of data on student's interactions in the environment. The method proposed, with the data used in this study, obtained better results in the prediction of the academic performance of students when compared to other methods approached in the literature.

Keywords: Learning Management Systems. Dimensionality Reduction. Neural Networks. Educacional Data Mining.

LISTA DE ILUSTRAÇÕES

Figura 1 – Visão geral do processo de KDD.	18
Figura 2 – Técnicas de mineração de dados aplicados à projeção de sistemas educacionais.	22
Figura 3 – Áreas relacionadas à Mineração de Dados Educacionais.	24
Figura 4 – Representação de um modelo preditivo.	25
Figura 5 – Taxonomia das Técnicas de Redução da Dimensionalidade	34
Figura 6 – Etapas do PCA.	36
Figura 7 – Etapas da LDA.	39
Figura 8 – Modelo de um neurônio artificial.	42
Figura 9 – Topologia genérica de uma MLP	43
Figura 10 – Método para classificação proposto neste trabalho.	44
Figura 11 – Diferentes resultados para uma predição de duas classes.	48
Figura 12 – Matriz de confusão com resultados de uma predição com três classes.	48
Figura 13 – Distribuição das notas dos 300 estudantes das 13 turmas selecionadas.	51
Figura 14 – Matriz de confusão da MLP no experimento 01.	55
Figura 15 – Fluxo do método proposto para o Experimento 02.	57
Figura 16 – Matriz de confusão da MLP no experimento 02.	58
Figura 17 – Fluxo do método proposto para o experimento 03.	59
Figura 18 – Médias das acurácias da MLP para diferentes números de componentes principais.	59
Figura 19 – Matriz de confusão da MLP no experimento 03.	60
Figura 20 – Fluxo do método proposto para o experimento 04.	61
Figura 21 – Matriz de confusão da MLP no experimento 04.	62
Figura 22 – Fluxo do método proposto para o experimento 05.	62
Figura 23 – Matriz de confusão da MLP no experimento 05.	63
Figura 24 – Acurácias da MLP dos Experimentos 01 a 05.	64
Figura 25 – Taxas de identificação dos estudantes com baixos níveis de desempenho nos experimentos.	64

LISTA DE TABELAS

Tabela 2 – Exemplos de tabelas da base de dados do <i>Moodle</i> sobre interações dos estudantes	17
Tabela 3 – Atributos do vetor de características do estudante no AVA	52
Tabela 4 – Distribuição dos dados dos estudantes selecionados.	52
Tabela 5 – Correlação entre os atributos do vetor de características e o atributo resultado	53
Tabela 6 – Distribuição dos estudantes em três classes	54
Tabela 7 – Medidas de Precisão, Revocação e <i>F-measure</i> da MLP no experimento 01.	56
Tabela 8 – Informações sobre os valores dos atributos dos dados originais e após a normalização.	57
Tabela 9 – Medidas de Precisão, Revocação e <i>F-measure</i> da MLP no experimento 02.	58
Tabela 10 – Medidas de Precisão, Revocação e <i>F-measure</i> da MLP no experimento 03.	60
Tabela 11 – Medidas de Precisão, Revocação e <i>F-measure</i> da MLP no experimento 04.	61
Tabela 12 – Medidas de Precisão, Revocação e <i>F-measure</i> da MLP no experimento 05.	63
Tabela 13 – Acurácia do método proposto comparado aos métodos utilizados em outros trabalhos.	65

LISTA DE ABREVIATURAS E SIGLAS

AVA Ambientes Virtuais de Aprendizagem

DM *Data Mining*

EAD Educação à Distância

EDM *Educational Data Mining*

KDD *Knowlegde Database Discovery*

LDA *Linear Discriminant Analysis*

MLP *Multilayer Perceptron*

PCA *Principal Components Analysis*

RNA Redes Neurais Artificiais

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Problemática	13
1.2	Objetivos deste trabalho	14
1.2.1	<i>Objetivo Geral</i>	14
1.2.2	<i>Objetivos Específicos</i>	15
1.3	Produção Científica	15
1.4	Organização do Trabalho	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Ambientes Virtuais de Aprendizagem	16
2.1.1	<i>Moodle</i>	17
2.2	Descoberta do Conhecimento e Mineração de Dados	18
2.3	Mineração de Dados Educacionais	20
2.3.1	<i>Definição</i>	20
2.3.2	<i>Objetivos da Mineração de Dados Educacionais</i>	22
2.3.3	<i>Áreas relacionadas à Mineração de Dados Educacionais</i>	24
2.4	Predição de Desempenho de Estudantes	25
3	TRABALHOS RELACIONADOS	28
4	PRÉ-PROCESSAMENTO DOS DADOS	32
4.1	Normalização dos Dados	33
4.1.1	<i>Normalização dos Dados com z-score</i>	33
4.2	Redução da Dimensionalidade	33
4.2.1	<i>Análise dos Componentes Principais (PCA)</i>	35
4.2.2	<i>Análise Discriminante Linear (LDA)</i>	37
5	MÉTODOS DE CLASSIFICAÇÃO	40
5.1	Redes Neurais Artificiais	40
5.1.1	<i>Perceptron</i>	42
5.1.2	<i>Multilayer Perceptron (MLP)</i>	43
5.1.2.1	<i>Treinamento de Redes MLP</i>	44
5.2	Métodos de Classificação Propostos	44
5.3	Estratégias de Validação	45

5.3.1	<i>Validação Cruzada</i>	45
5.3.2	<i>K-fold</i>	46
5.3.3	<i>Leave-one-out</i>	46
5.3.4	<i>Hold-out</i>	46
5.4	Matriz de Confusão	47
5.5	Medidas de Avaliação	48
5.5.1	<i>Precisão</i>	48
5.5.2	<i>Revocação</i>	49
5.5.3	<i>F-measure</i>	49
6	EXPERIMENTOS	50
6.1	Dados	50
6.2	Vetor de Características	51
6.3	Ambientes de Testes	54
6.4	Experimentos Realizados	54
6.4.1	<i>Experimento 01: Dados Originais</i>	55
6.4.2	<i>Experimento 02: Dados Normalizados com z-score</i>	57
6.4.3	<i>Experimento 03: Dados Normalizados com z-score e Reduzidos com PCA</i>	59
6.4.4	<i>Experimento 04: Dados Normalizados com z-score e Reduzidos com LDA</i>	60
6.4.5	<i>Experimento 05: z-score + PCA + LDA</i>	62
6.5	Análise dos Resultados	63
6.6	Comparação dos Métodos Propostos com outros Métodos de Classificação	65
7	CONCLUSÕES E TRABALHOS FUTUROS	67
	REFERÊNCIAS	69

1 INTRODUÇÃO

O emergente crescimento das tecnologias de informação e comunicação em instituições de ensino mudaram significativamente a maneira como educadores ensinam e estudantes aprendem. O uso da internet facilitou o acesso e expansão da educação através de *softwares* baseados na *web* e metodologias educacionais que compõem a Educação à Distância (EAD). Devido a capacidade destes *softwares* gerenciarem e disponibilizarem conteúdo, surgiu a possibilidade de transformar a modalidade do ensino presencial em uma modalidade em que todas as disciplinas e recursos para o aprendizado são fornecidos e acessíveis *on-line*. Dentre estes *softwares*, destacam-se os Ambientes Virtuais de Aprendizagem (AVA), por permitir que estudantes acessem o conteúdo *on-line* disponibilizado pelo educador, interaja com outros estudantes através de fóruns e *chats*, e tenha seu conhecimento avaliado a partir de tarefas, *quizzes* e etc. Adicionalmente, os AVA permitem aos educadores fornecer e gerenciar estes recursos de uma maneira relativamente fácil e de forma integrada.

Com a implantação de AVA em instituições de ensino, um grande volume de dados que descrevem o comportamento dos estudantes *on-line* foi disponibilizado. Paralelo a isto, técnicas de Mineração de Dados, do inglês *Data Mining* (DM), foram utilizadas na tentativa de identificar comportamentos e realizar previsões de desempenho dos estudantes a partir de dados gerados pelos AVA (GOTTARDO *et al.*, 2014; CONIJN *et al.*, 2016). DM é o uso de algoritmos para extração de padrões a partir de dados. O conhecimento obtido a partir dos dados pode ser aplicado para melhorias nos ambientes educacionais. A partir destes estudos deu-se origem a Mineração de Dados Educacionais, do inglês *Educational Data Mining* (EDM), com foco no desenvolvimento na exploração de informações coletadas em sistemas educacionais e fornecimento de melhores a estes sistemas, e a todos os envolvidos no processo de ensino-aprendizagem. (BAKER *et al.*, 2011).

1.1 Problemática

Os AVA têm a capacidade de coletar e armazenar um grande volume de dados sobre estudantes. Estes dados contém registros de acessos aos diversos recursos disponíveis no ambiente e as interações realizadas pelos estudantes, por exemplo, mensagens em

fóruns e *chats*, que são dados que, a partir destes podem se obter informações úteis sobre perspectivas e sentimentos dos estudantes durante a realização do curso (MERCERON; YACEF, 2005).

Apesar dos AVA disponibilizarem uma enorme base de dados sobre diversas ações dos estudantes, estes dados são pouco úteis se não forem utilizados em conjunto com ferramentas adequadas de análise e interpretação de dados. Tais limitações de informações podem trazer dificuldades no desenvolvimento e aperfeiçoamento de atividades de acompanhamento e monitoramento do processo de ensino-aprendizagem dos estudantes. Um processo automatizado de monitoramento e avaliação é um diferencial qualitativo nos cursos em um ambiente educacional EAD, pois o educador monitora eventos e obtém informações privilegiadas sobre o comportamento dos estudantes, e assim poderá tomar decisões adequadas e eficientes (ROMERO; VENTURA, 2010).

Processos automatizados para monitorar os estudantes poderiam acompanhar e verificar a utilização dos recursos de ensino disponibilizadas no curso, bem como avaliar o impacto destes no desempenho final dos estudantes.

O cenário supracitado, com ferramentas de acompanhamento automático de estudantes, mesmo que em partes, ainda está longe da realidade de muitas instituições de ensino. Desta forma, o estudo de técnicas de DM que manipulem grandes quantidades de dados gerados pelos AVA, e transforme estes dados em informações úteis, que apoiem as tomadas de decisão é de grande importância para a comunidade de pesquisa em Mineração de Dados Educacional, bem como as instituições de ensino que utilizam tais ambientes.

1.2 Objetivos deste trabalho

Diante do problema exposto, o presente trabalho tem os seguintes objetivos:

1.2.1 *Objetivo Geral*

Investigar a viabilidade da utilização de técnicas de Redução da Dimensionalidade e Redes Neurais Artificiais do tipo *Multilayer Perceptron* para predição do desempenho acadêmico a partir de informações das interações de estudantes em AVA.

1.2.2 *Objetivos Específicos*

- Obter informações provenientes de bases de dados de AVA que representem estudantes e que tenham correlação com o seu desempenho em cursos na modalidade EAD;
- Gerar modelos preditivos capazes de identificar estudantes com diferentes níveis de desempenho acadêmico;
- Avaliar o impacto de técnicas de pré-processamento dos dados nos processos de classificação de estudantes;

1.3 **Produção Científica**

- de Albuquerque, R. M.; **Bezerra A. A.**; de Souza, D. A.; do Nascimento, L. B. P; de Mesquita Sá Júnior, J. J.; do Nascimento, J. C. (2015, November). *Using neural networks to predict the future performance of students*. In 2015 International Symposium on Computers in Education (SIIE) (pp. 109-113). IEEE.

1.4 **Organização do Trabalho**

O restante desta dissertação esta organizada da seguinte forma:

No Capítulo 2 são introduzidos os conceitos de Mineração de Dados Educacionais, Ambientes Virtuais de Aprendizagem e Predição de Desempenho de Estudantes, essenciais para um bom entendimento deste trabalho. No Capítulo 3 são expostos trabalhos, que assim com este, utilizam técnicas de mineração de dados educacionais para predição do desempenho acadêmico de estudantes. O capítulo 4 apresenta os métodos de pré-processamento dos dados que serão aplicados para preparar os dados para o processo de classificação. No capítulo 5 apresenta-se a rede neural artificial que será utilizada para classificação dos estudantes quando ao seu desempenho acadêmico, e os métodos que avaliam a acurácia dos modelos preditivos obtidos. O Capítulo 6 descreve os experimentos realizados para avaliar a capacidade dos métodos propostos em identificar estudantes em diferentes níveis de desempenho e apresenta uma análise dos resultados dos experimentos. O Capítulo 7 conclui este trabalho e apresenta desafios futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta conceitos essenciais para o entendimento deste trabalho. A primeira seção aborda os Ambientes Virtuais de Aprendizagem (AVA), que são *softwares* baseados na *web* que disponibilizam recursos *on-line* para o ensino e geram grandes volumes de dados. A seção seguinte introduz a Mineração de Dados Educacionais, que é a aplicação de técnicas de mineração de dados em dados provenientes de ambientes de ensino. A terceira seção explana a Predição de Desempenho de Estudantes, que é o objetivo da utilização de técnicas de mineração de dados neste trabalho.

2.1 Ambientes Virtuais de Aprendizagem

A utilização de *softwares* tornou-se um método de apoio para atividades e processos de ensino-aprendizagem em ambientes educacionais. Dentre estes *softwares*, destacam-se os Ambientes Virtuais de Aprendizagem (AVA), que são ambientes digitais de ensino baseados na *web* que ofertam cursos *on-line* e podem oferecer uma grande variedade de canais e espaços de trabalho, facilitando assim o compartilhamento de informação e comunicação entre os participantes de cursos. Estes ambientes são tecnologias com capacidade de integrar e organizar processos educacionais mediados por computador, apoiar professores e promover o engajamento de alunos. Ferramentas disponíveis nesse tipo de ambiente permitem a comunicação entre os alunos e professores, ou entre os próprios alunos, de forma quase imediata, e permitem aos educadores distribuir informações para os alunos, produzir material de conteúdo, preparar atividades e testes, participar de discussões, gerenciar classes à distância, além de fornecer recursos para a aprendizagem colaborativa através de fóruns, *chats*, áreas de armazenamentos de arquivos, questionários, *quizzes*, serviços de notícias, etc (ROMERO *et al.*, 2008). A interação dos alunos com o AVA e com os conteúdos disponíveis torna-se fundamental para que os alunos possam organizar suas ideias e compartilhar seus conhecimentos (GALAFASSI *et al.*, 2014).

O uso de AVA e outros *softwares* no contexto acadêmico produz grandes repositórios de dados com informações sobre alunos, professores, disciplinas, dentre outras. A utilização de técnicas de *Data Mining* (DM) sobre estes dados permite a obtenção de informações úteis que apoiarão os trabalhos de educadores e demais envolvidos no processo de ensino-aprendizagem.

No decorrer desta dissertação, os conceitos de curso e disciplinas são os mesmos. Em AVA as disciplinas ofertadas podem ser tratadas como um curso.

2.1.1 Moodle

Moodle é uma plataforma de aprendizagem projetada para proporcionar a administradores de cursos, educadores e estudantes um sistema robusto de código aberto, seguro, multi-idiomas e integrado para criar ambientes de aprendizagem personalizados (DOUGIAMAS; TAYLOR, 2003). Este ambiente educacional disponibiliza ferramentas que apoiam educadores e estudantes, tais como, fóruns e questionários, além de um canal de comunicação entre educadores e alunos.

Moodle armazena todas as interações realizadas pelos usuários em forma de *logs*. Estes dados podem conter informações úteis que apoiem às tomadas de decisão quando utilizados com a técnica de mineração adequada. A tabela 2 mostra algumas tabelas da base de dados do *Moodle* que armazenam informações sobre interações dos estudantes no ambiente.

Tabela 2 – Exemplos de tabelas da base de dados do *Moodle* sobre interações dos estudantes

Nome	Descrição
<i>mdl_user</i>	Informações sobre todos os usuários
<i>mdl_user_students</i>	Informações sobre todos os estudantes
<i>mdl_log</i>	Registros de todas as ações do usuário
<i>mdl_assignment</i>	Informações sobre tarefas
<i>mdl_assignment_submissions</i>	Informações sobre tarefas submetidas
<i>mdl_forum</i>	Informações sobre todos os fóruns
<i>mdl_forum_posts</i>	Armazena todos as postagens do fórum
<i>mdl_forum_discussions</i>	Armazena todos os tópicos de discussões do fórum
<i>mdl_message</i>	Armazena todas as mensagens
<i>mdl_message_read</i>	Armazena todas as mensagens lidas
<i>mdl_quiz</i>	Informações sobre todos os quizzes
<i>mdl_quiz_attempts</i>	Armazena todas as tentativas de realizar o quiz
<i>mdl_quiz_grades</i>	Armazena a nota final em um quiz

Fonte – O autor

Moodle conta com 73.269 ambientes registrados em 233 países, sendo o Brasil o terceiro país que mais utiliza, mais de 10 milhões de cursos foram ofertados nesta plataforma a quase 95 milhões de usuários em todo o mundo¹.

¹ Fonte: <https://moodle.net/stats/>. Acessado em 04/nov/2016

As informações supracitadas mostram o quão o *Moodle* está difundido nas instituições de ensino do Brasil e no mundo, o que justifica estudos que tragam melhorias aos processos de ensino-aprendizagem em cursos que utilizam este ambiente.

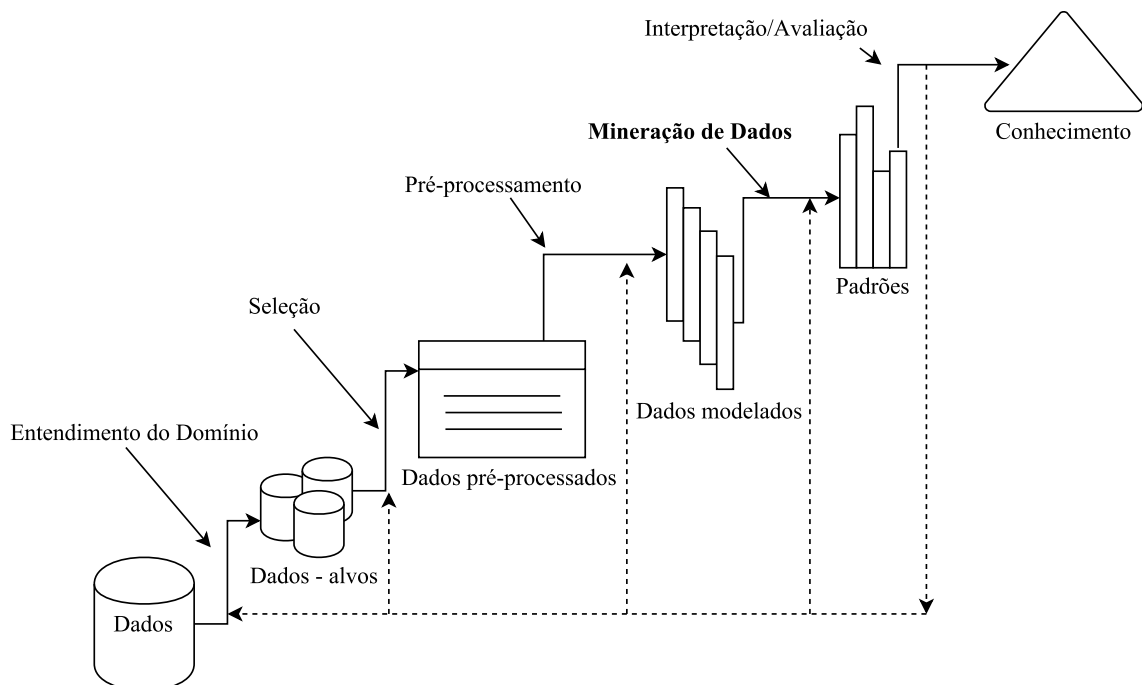
2.2 Descoberta do Conhecimento e Mineração de Dados

Knowledge Database Discovery (KDD) é um processo não-trivial de identificar padrões válidos, novos, potencialmente úteis e, por fim, compreensíveis, em dados (FAYYAD *et al.*, 1996b).

KDD não é trivial por não se tratar de um processo simples com cálculos diretos, como obter a média de alunos de uma turma, por exemplo. Trata-se de um processo com várias etapas e envolve a descoberta de correlações em grandes volumes de dados por meio de diversas técnicas de mineração de dados, bem como uma análise sobre estas correlações.

A **mineração de dados** é uma das várias etapas do processo de KDD, conforme mostrado na figura 1 e descritos a seguir:

Figura 1 – Visão geral do processo de KDD.



Fonte – (FAYYAD *et al.*, 1996b)

- **Entendimento do domínio:** Conhecimento prévio dos dados, do domínio da aplicação e dos objetivos do KDD, são importantes para obtenção de bons resultados.

- **Seleção dos dados:** Consiste na seleção de um subconjunto de dados que represente todos os dados disponíveis. Esta etapa é importante pois é computacionalmente inviável trabalhar com todos os dados, visto que estes possuem *terabytes* de informações.
- **Pré-processamento dos dados:** Nesta etapa é realizado uma limpeza nos dados por ser necessário a remoção de informações inconsistentes ou duplicadas. Também são verificados os casos onde há falta de informação.
- **Modelagem dos dados:** Com o objetivo de reduzir a dimensionalidade dos dados são aplicadas técnicas de seleção de atributos mais relevantes, assim os dados terão a maior representatividade com a menor quantidade de atributos possíveis.
- **Mineração de dados:** Considerada uma das mais complexas etapas do KDD, a mineração de dados também é um processo que contém sub-etapas. Nesta etapa é escolhido o algoritmo que será utilizado, de acordo com os dados e os objetivos do KDD.
- **Interpretação dos dados:** A análise dos dados obtidos pode ser feita através de métodos de visualização de informações. Caso os resultados não atendam as expectativas do que se espera após o processo de KDD, devem-se realizar algumas etapas anteriores novamente.
- **Validação dos resultados:** Os resultados obtidos na etapa de Mineração dos Dados devem ser validados com dados não utilizados pelos algoritmos. Os padrões descobertos serão válidos se atenderem um grau de certeza definido pelo responsável do processo de KDD.

Ainda segundo (FAYYAD *et al.*, 1996b), DM é um passo no processo do KDD que consiste na aplicação de análises de dados e algoritmos que, sob limitações de eficiência computacionais aceitáveis, produz uma relação particular de padrões a partir de dados.

Uma das tarefas mais importantes em DM é a classificação. Técnicas de classificação analisam um conjunto de dados com características e classes conhecidas e criam modelos capazes de classificar novas instâncias a partir de suas características. Essa tarefa é considerada preditiva, uma vez que são realizadas inferências sobre dados para se prever a classe de uma nova instância (MERSCHMANN, 2007). Esta é a abordagem empregada neste trabalho.

2.3 Mineração de Dados Educacionais

2.3.1 Definição

A Mineração de Dados Educacionais, do inglês *Educational Data Mining* (EDM), é a aplicação de técnicas de DM em dados educacionais. É uma área de pesquisa que emergiu nos últimos anos e é utilizada por pesquisadores de diversas áreas, tais como ciências da computação, educação, estatística, psicologia, sistemas de tutoria inteligente, na análise de grandes conjuntos de dados com objetivo de resolver problemas de pesquisas educacionais (BAKER; YACEF, 2009). EDM está preocupada com o desenvolvimento de métodos para explorar informações coletadas de ambientes educacionais, permitindo a compreensão dos alunos de forma mais eficaz e adequada, proporcionando melhores benefícios educacionais aos mesmos (BAKER *et al.*, 2011).

Em EDM, os dados provenientes de sistemas educacionais são convertidos em informações úteis que podem ter grandes impactos em pesquisas e práticas educacionais. O processo realizado não é muito diferente de outras aplicações de DM, tais como, negócios, genética, medicina, porque baseia-se nos mesmos passos do processo de mineração de dados em geral. De acordo com (GARCÍA *et al.*, 2011), estes passos são:

- **Passo I - Pré-processamento:** Os dados obtidos de sistemas educacionais têm que primeiro ser pré-processados e transformados em um formato apropriado para a mineração. Algumas das tarefas principais desta etapa são: limpeza, seleção de atributos e integração de dados.
- **Passo II - Mineração de Dados:** É a etapa central, e nela, técnicas de mineração são aplicados aos dados previamente pré-processados. Alguns exemplos de técnicas de mineração de dados são: visualização, regressão, classificação, agrupamentos, regras de associação, entre outros.
- **Passo III - Pós-processamento:** Nesta etapa, os resultados ou modelos obtidos são interpretados e usados em tomadas de decisão sobre o ambiente educacional.

As tarefas em um processo de EDM que são responsáveis pela análise de dados gerados a partir das interações entre estudantes nos ambientes de aprendizagem podem responder questões relacionadas a como melhorar o aprendizado do estudante (COSTA *et al.*, 2013). Assim, é possível compreendê-lo e como ele aprende, além de outros fatores que influenciam na aprendizagem. Também é possível identificar qual a abordagem instrucional

(aprendizagem individual ou colaborativa) proporciona melhores benefícios educacionais aos estudantes, se o aluno está desmotivado ou confuso. Com este conhecimento, é possível criar ambientes personalizados e métodos de ensino que ofereçam melhores condições de aprendizado (BAKER *et al.*, 2011).

Apesar das técnicas utilizadas em EDM serem provindas da DM, elas passam por adaptações devido as particularidades existentes em ambientes educacionais e seus dados. A seguir, é apresentada uma taxonomia, proposta por (BAKER *et al.*, 2011), que informa a categorização das sub-áreas de pesquisa da EDM.

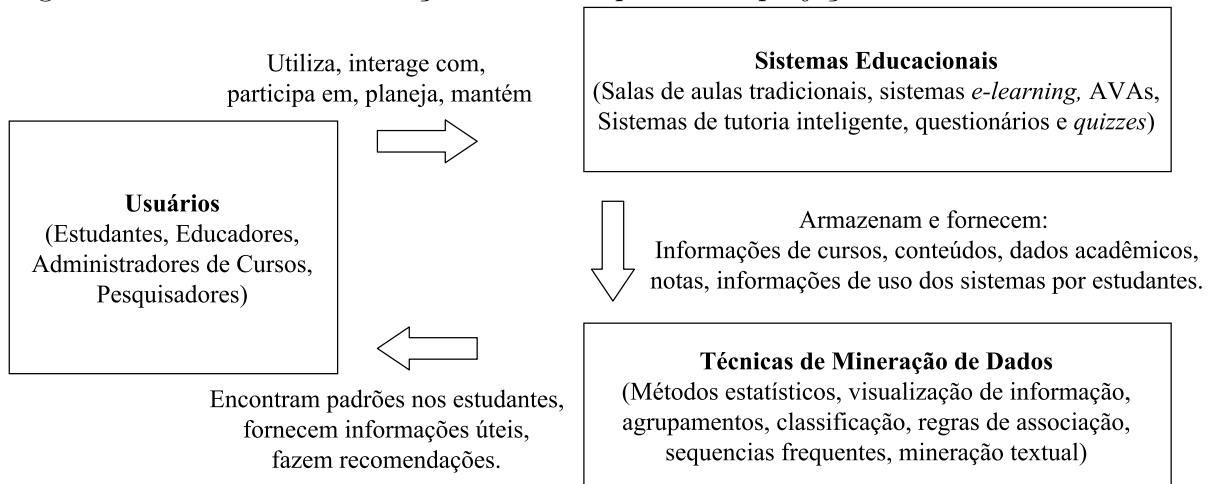
- Predição
 - Classificação
 - Regressão
 - Estimação de Densidade
- Agrupamento
- Mineração de Relações
 - Mineração de Regras de Associação
 - Mineração de Correlações
 - Mineração de Padrões Sequenciais
 - Mineração de Causas

Na **predição** são desenvolvidos modelos que farão inferências sobre dados a partir de informações de outros dados previamente analisados. No **agrupamento** os dados são dispostos em grupos que se formam de acordo com as similaridades de cada item. Na **mineração de relações** são descobertas relações entre variáveis contidas nos dados. A **mineração de regras de associação** procura gerar ou identificar regras do tipo $X \rightarrow Y$, ou seja, uma regra associa a ocorrência de um valor Y dado que ocorreu o valor X . Na **mineração de correlações** obtêm-se correlações lineares, que podem ser positivas ou negativas, entre duas variáveis. Na **mineração de causas**, a partir da análise dos padrões de covariância, verifica-se se um evento causa outro evento.

A aplicação de técnicas de mineração de dados em sistemas educacionais com o objetivo de melhorar o processo de ensino-aprendizagem é vista como uma técnica de avaliação formativa. De acordo com (ARRUABARRENA *et al.*, 2002), avaliação formativa é uma avaliação do programa educacional enquanto ele ainda está em progresso. Examinar como estudantes utilizam os sistemas educacionais é uma forma de avaliação formativa que

pode ajudar educadores na melhoria dos recursos de ensino disponíveis nestes sistemas. Técnicas de mineração de dados podem ser usadas para obter informações que auxiliem educadores na criação e manutenção de cursos em sistemas educacionais e tomadas de decisões.

Figura 2 – Técnicas de mineração de dados aplicados à projeção de sistemas educacionais.



Fonte – (ROMERO *et al.*, 2010)

A aplicação de EDM para projetar ou melhorar sistemas educacionais é um ciclo iterativo de formação de hipóteses, testes e ajustes. Como visto na figura 2, educadores e projetistas de ambientes educacionais planejam, constroem e mantêm estes sistemas educacionais. Estudantes utilizam o ambiente para o aprendizado. Estes ambientes geram informações sobre cursos, estudantes, uso do sistema e interações, assim diversas técnicas de mineração de dados podem ser aplicadas nestes dados objetivando descobrir informações úteis que tragam melhorias no processo de ensino-aprendizagem e na forma que os recursos são disponibilizados nos ambientes educacionais. O conhecimento obtido será utilizado não apenas por educadores ou administradores nas melhorias dos cursos, mas também por estudantes. Desta forma, a mineração de dados educacional pode ser orientado ao suporte de necessidades específicas de cada um dos envolvidos no processo de ensino-aprendizagem.

2.3.2 Objetivos da Mineração de Dados Educacionais

(ROMERO; VENTURA, 2013) propuseram a classificação dos objetivos da EDM de acordo com o usuário final (estudante, educador, administrador de curso e pesquisador) e o problema que será resolvido:

- **Estudantes:** Apoiar os estudantes no processo de ensino-aprendizagem, fornecer

feedback ou recomendações, responder as necessidades do estudante, melhorar o desempenho no aprendizado.

- **Educadores:** Entender o processo de aprendizagem dos alunos e como os métodos aplicados são recebidos pelos alunos, além de melhorar o desempenho no ensino.
- **Administradores de Curso:** Avaliar a melhor maneira de organizar os recursos (humanos e didáticos) disponíveis para a realização dos cursos.
- **Pesquisadores:** Desenvolver, comparar e avaliar técnicas de mineração de dados com o objetivo de recomendar a melhor técnica que traga informações úteis a cada problema ou tarefa específica dos ambientes educacionais. Estas informações apoiaram os trabalhos dos alunos, educadores e administradores de curso.

(ROMERO *et al.*, 2010) ainda destacam algumas aplicações da mineração de dados educacionais:

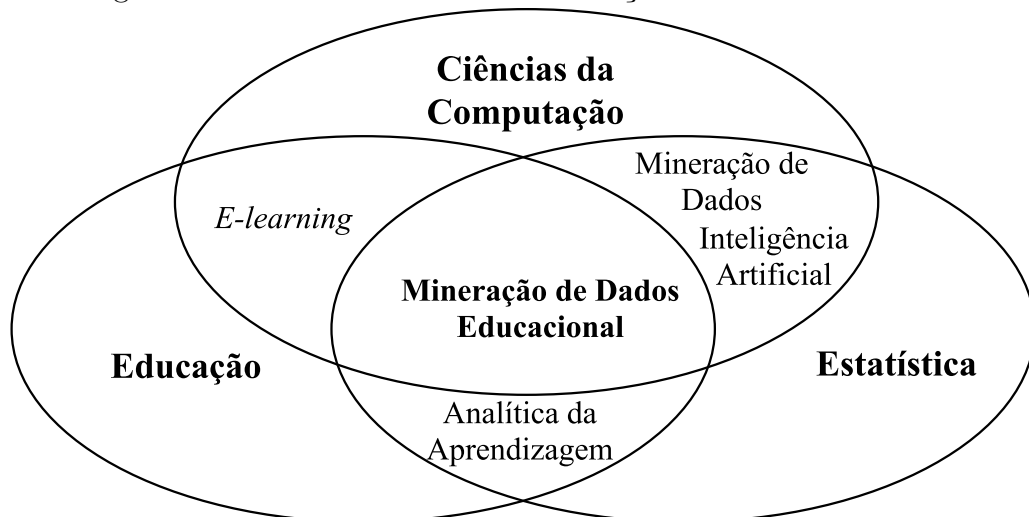
- **Comunicação entre as partes envolvidas:** O objetivo é ajudar administradores de cursos e educadores na análise das atividades realizadas pelos estudantes e nas informações de uso (acesso a fóruns, materiais didáticos, questionários) em cursos. As técnicas mais utilizadas para este propósito são a análise exploratória de dados através de análise estatística e técnicas de visualização de informações.
- **Manutenção e melhorias dos cursos:** Objetiva ajudar administradores de curso e educadores no que se pode melhorar em cursos (conteúdos, atividades, *links*, e etc.) utilizando, por exemplo, informações de uso do sistema educacional, ou seja, como e quais recursos foram utilizados pelos estudantes. As técnicas de mineração de dados mais comuns para este objetivo são as regras de associação, sequencias frequentes, classificação e agrupamentos.
- **Recomendações de conteúdos:** Recomenda a estudantes conteúdos que sejam mais apropriados para seu aprendizado. As técnicas mais comuns nesta aplicação são as regras de associação, sequencias frequentes, classificação e agrupamentos.
- **Predição de notas de estudantes e desempenho em cursos:** Obtém uma previsão de notas finais em cursos ou chances de um estudante abandonar o curso, baseados nos dados gerados por outros estudantes durante a realização de cursos. As técnicas utilizadas são classificação, agrupamentos e regras de associação.
- **Modelagem do perfil do usuário:** No domínio educacional, a modelagem do perfil do usuário pode detectar o estado e características do estudantes, tais como

satisfação, motivação, progresso da aprendizagem, preferências e outros fatores que impactam negativamente o desempenho destes estudantes (motivos que levam a erros em questionários, por exemplo). Além da classificação, agrupamentos, regras de associação, também são utilizadas técnicas de análises estatísticas e redes Bayesianas para este propósito.

2.3.3 Áreas relacionadas à Mineração de Dados Educacionais

EDM pode ser vista como um combinado de três áreas, conforme a figura 3: Ciência da computação, estatística e educação. A interseção destas três áreas também dá origem a outras áreas relacionadas a EDM, por exemplo, a Analítica da Aprendizagem, do inglês *learning analytics*, que é uma tecnologia educacional que coleta, mede, analisa e divulga dados de como os alunos se comportam nos AVA, com o objetivo de propor novos métodos de aprendizagem para estes alunos e demais envolvidos no processo de ensino-aprendizagem (ROMERO; VENTURA, 2013). EDM é uma área interdisciplinar e utiliza métodos e técnicas da estatística, aprendizado de máquina, mineração de dados, recuperação de informação, sistemas de recomendação, psicologia cognitiva, entre outros.

Figura 3 – Áreas relacionadas à Mineração de Dados Educacionais.

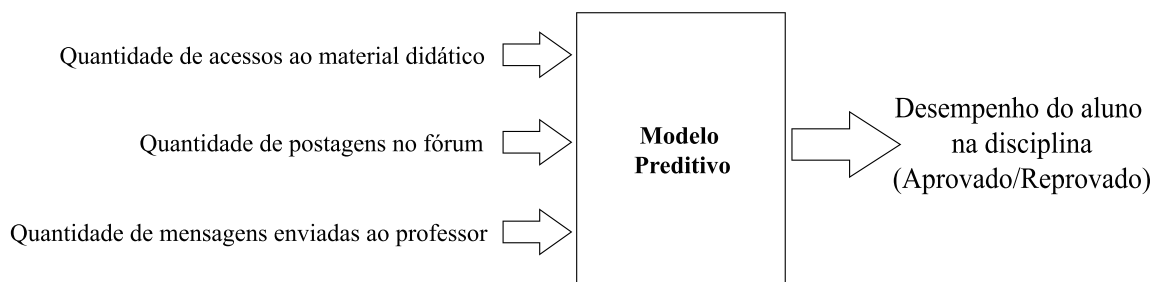


Fonte – (ROMERO; VENTURA, 2013)

2.4 Predição de Desempenho de Estudantes

Várias técnicas de mineração de dados podem ser utilizadas na criação de modelos que façam inferências sobre o desempenho de estudantes. Um modelo pode ser entendido como uma abstração que faz previsões sobre o futuro em informações do passado ou presente. Um modelo inclui um algoritmo de aprendizado de máquina que aprende sobre características de objetos em um conjunto de dados de treinamento com o objetivo de realizar previsões sobre outros objetos. Os modelos preditivos são basicamente divididos em duas sub-áreas: regressão e classificação de padrões. Na regressão analisa-se a relação entre duas variáveis com o objetivo de encontrar tendências entre estas variáveis. Na classificação de padrões, dados são atribuídos à classes de acordo com as informações dos dados previamente informados ao algoritmo de aprendizagem de máquina.

Figura 4 – Representação de um modelo preditivo.



Fonte – O autor

A figura 4 mostra um esquema para construção de um modelo preditivo, sendo as entradas de dados algumas informações do estudante. A partir da entrada de dados, o modelo pode inferir sobre estes dados e produzir uma previsão como saída. A previsão obtida pelo modelo, entre outras coisas, depende da qualidade e da quantidade de dados disponíveis. Um modelo é construído em função das informações disponíveis na base de dados e da seleção adequada de técnicas de mineração de dados.

O modelo gerado precisa ser validado para verificar se ele atende as condições impostas para solucionar o problema. Diferentes modelos podem solucionar o problema, escolhe-se aquele que apresenta melhores resultados na etapa de validação.

No processo de obtenção do modelo, duas fases são executadas: na primeira, um conjunto de dados de treinamento será utilizado pelo algoritmo para construir o modelo descritivo dos dados. Na segunda fase, novos dados, chamados dados de teste,

são utilizados com objetivo de testar o modelo gerado na primeira fase. A análise dos resultados é feita pelo responsável pela execução dos processos de KDD, afim de verificar se o modelo atende ao propósito. Caso um modelo não atenda as necessidades definidas, podem ser realizadas as seguintes ações:

- Executar o algoritmo novamente utilizando novos dados ou modificando os parâmetros;
- Utilizar outro algoritmo de aprendizado de máquina.

Neste trabalho, o objetivo da predição é estimar valores desconhecidos de uma variável que descreve o estudante. Na educação, este valor pode informar sobre desempenho acadêmico, conhecimento sobre determinado assunto, notas em avaliações, rendimentos em disciplinas, entre outras. Este valor pode ser numérico ou contínuo (regressão), ou, categórico ou discretos (classificação). A regressão procura a relação entre uma variável dependente e uma ou mais variáveis independentes. A classificação é um procedimento em que itens são individualmente colocados em grupos com base em informações quantitativas sobre uma ou mais características comuns aos itens e com base em um conjunto de treinamento com itens previamente rotulados, ou seja, os itens contêm informações sobre quais grupos pertencem. A predição de desempenho de estudantes é uma das técnicas mais abordadas em aplicações de DM na educação, e diferentes técnicas e modelos foram aplicados (redes neurais, classificadores Bayesianos, regressão e análise de correlação) (ROMERO; VENTURA, 2010).

Este trabalho utiliza Redes Neurais Artificiais (RNA) para obtenção de modelos façam inferências sobre os desempenhos de estudantes. Diversos tipos de redes neurais foram utilizadas para predição de notas finais de estudantes (redes *backpropagation* e *feedforward*) (GEDEON; TURNER, 1993), para predição da quantidade de erros cometidos por estudantes (redes *backpropagation* e *feedforward*) (WANG; MITROVIC, 2002), para predição de desempenhos em testes (redes *backpropagation* e *counter propagation*) (FAUSETT; ELWASIF, 1994), para predição de resultados em disciplinas (aprovação ou reprovação) a partir de *logs* do Moodle (*radial basis function*) (CALVO-FLORES *et al.*, 2006), e para predição do desempenho de candidatos a vagas em universidades (redes com topologia *multilayer perceptron*) (OLADOKUN *et al.*, 2008).

No próximo capítulo serão abordados outros trabalhos que utilizam algoritmos de mineração de dados para obtenção de informações úteis sobre estudantes a partir de

bases de datos de ambientes educacionais.

3 TRABALHOS RELACIONADOS

Diversos autores abordaram a Mineração de Dados Educacionais, do inglês *Educational Data Mining* (EDM), visando resolver problemas ou propor melhorias nos processos educacionais. (ROMERO *et al.*, 2013; FENG *et al.*, 2009) realizaram trabalhos que fornecem *feedbacks* para o apoio a educadores nas tomadas de decisões. (BARCELLOS *et al.*, 2007) propôs uma aplicação que recomenda artigos a estudantes com base nas informações de perfil destes estudantes. Outros autores trabalharam para identificar comportamentos indesejáveis nos estudantes, como falta de motivação (HERSHKOVITZ; NACHMIAS, 2008), predisposição à desistência (DEKKER *et al.*, 2009), reprovações (BRESFELEAN *et al.*, 2008) ou tentativas de burlar as avaliações (BAKER *et al.*, 2011). O trabalho (AYERS *et al.*, 2009) utiliza técnicas de agrupamento para proporcionar a estudantes um ambiente de ensino e conteúdos personalizados de acordo com o seu perfil e habilidades. (RIGO *et al.*, 2012) utiliza técnicas de *Data Mining* (DM) e *learning analytics* para estudar fatores que levam a evasão escolar.

Este trabalho objetiva a realização de predições de desempenho de estudantes. A técnica de DM utilizada é conhecida como classificação. O objetivo da classificação é fornecer modelos que possam inferir um aspecto particular dos dados através de alguma combinação de outros aspectos destes dados (WITTEN; FRANK, 2005). A previsão do desempenho, ou situação, acadêmica de alunos é um tema também muito estudado entre pesquisadores. Em (MANHÃES *et al.*, 2011), os autores utilizaram técnicas de mineração de dados para prever o risco de evasão de estudantes em cursos presenciais da Escola Politécnica da Universidade do Rio de Janeiro. Dez modelos diferentes foram testados, com acurácia média variando entre 75% e 80%, com os algoritmos de classificação *Multilayer Perceptron* e *RandomForest* apresentando os melhores desempenhos. O estudo mostrou que a partir das primeiras notas semestrais dos calouros é possível identificar com precisão de 80% a situação final do aluno.

Em (SANTOS *et al.*, 2012) utiliza-se dados de estudantes do curso de Engenharia da Computação da Universidade Federal do Pampa visando de contribuir para diminuição de problemas de estudantes que reprovam no início do curso. Com dados de ambientes virtuais de aprendizagem e técnicas de mineração de dados baseadas em agrupamentos com *K-means* e classificação com árvores de decisão *REPTree* e *J48* os autores criaram modelos capazes de identificar tendências a reprovação com acurácia em

torno de 69%. Os estudos mostraram que a medida que o aluno se dedica as atividades presenciais e semipresenciais, aumenta o seu desempenho nas avaliações somativas, e com isso, suas chances de aprovação.

Na tentativa de prever o desempenho acadêmico de estudantes na educação à distância, (GOTTARDO *et al.*, 2012) utilizam os algoritmos de classificação *RandomForest* e *Multilayer Perceptron* e obtêm precisão acima de 74%. Este trabalho diferencia-se ao tentar prever a reprovação de alunos e não seus desempenhos específicos, além de prover uma análise mais detalhada utilizando unicamente atividades agregadas. Este trabalho utiliza de uma grande variedade de atributos referentes ao uso dos Ambientes Virtuais de Aprendizagem (AVA) Moodle.

O trabalho (RODRIGUES *et al.*, 2013) utilizou dados do ambiente virtual de aprendizagem *Moodle* e séries temporais em um estudo da viabilidade do uso de um modelo de regressão linear para também prever o desempenho dos alunos. Diversos atributos foram utilizados, incluindo as médias das horas de acesso, quantidade de acessos aos materiais didáticos disponíveis, quantidade de interações ao longo das semanas em fóruns, vídeos, entre outros. O experimento foi realizado com dados da disciplina Jogos Educacionais da graduação em Licenciatura em Computação da Universidade Federal Rural de Pernambuco, e contou com a participação de 98 alunos.

Em (GOTARDO *et al.*, 2013), os autores realizam a tarefa de predição do desempenho do aluno utilizando Sistemas de Recomendação e Acoplamento de Classificadores. Os dados são relativos a um curso com 252 alunos. Os dados para treinamento dos classificadores foram extraídos da tabela de *logs* e separados por tipos, como por exemplo: tarefas, fóruns, *upload*, entre outros. Cada tipo transformou-se em um atributo, juntamente com a situação acadêmica do aluno. Os classificadores *J48* e *Naive Bayes* foram utilizados no trabalho citado. Segundo os autores, a abordagem utiliza um conjunto inicial de dados são aplicadas técnicas tradicionais de aprendizado de máquina que cooperam entre si, formando visões diferentes sobre o resultado. O melhor resultado foi aproximadamente 80% de predição, utilizando 40% do conjunto de dados, com o classificador *Naive Bayes*.

Em (DETONI *et al.*, 2014), utiliza-se Redes Bayesianas, *Multilayer Perceptron*, *J48* e *RandomForest* para prever a reprovação de alunos de educação à distância utilizando o registro de interações entre alunos, tutores e professores. Entre as técnicas utilizadas, os melhores resultados foram obtidas pelas Redes Bayesianas, com acurácia média de

80%. Os autores concluem que utilizar apenas a quantidade de interações dos alunos é viável para gerar previsões razoavelmente precisas, e que o modelo preditivo gerado é mais generalizável, podendo ser aplicado em outras situações onde é possível contar interações.

Em (SANTANA *et al.*, 2014), utiliza-se árvores de decisão *RandomForest* e *J48*, Redes Neurais *Multilayer Perceptron* e *Radial Basis Function*, além de *Support Vector Machine* e *K-nearest neighbors*, em conjunto com dados de interações em um curso ofertado na modalidade semipresencial no ambiente virtual de aprendizagem Moodle. Em um experimento, os autores dividiram os estudantes em três classes, conforme suas notas finais, e em outro experimento, os estudantes foram divididos em duas classes, aprovado e reprovado. Das técnicas de mineração utilizadas, a árvore de decisão *J48* obteve o melhor desempenho, alcançando uma acurácia média de 74%. Os autores concluem que é possível inferir sobre o desempenho do estudante a partir do perfil de uso do ambiente virtual de aprendizagem pelos estudantes.

Em (PEQUENO *et al.*, 2014), utiliza-se *Naive Bayes*, *Decision Table*, *OneR* e *J48*, além de técnicas de Estatística e Regras de Associação. Com dados de interações em fóruns de discussão, os autores investigaram padrões de comportamentos de estudantes e professores de cursos à distância da Universidade Federal do Ceará. Os autores abordaram dois tipos de ofertas de disciplinas, modulares e semestrais. Após experimentos, foram obtidas regras de associação com níveis de confiança de 93% e classificações com acurácia de 96% utilizando *Decision Table*. A análise dos resultados ainda mostra quais tipos de disciplinas, se modulares ou semestrais, são indicadas para cada área do conhecimento.

Um trabalho de doutorado descrito em (MANHÃES, 2015), propõe-se uma arquitetura que engloba todo o processo de descoberta de conhecimentos em dados educacionais. Foram utilizados dados coletados durante 16 anos sobre graduandos da Universidade Federal do Rio de Janeiro (UFRJ) para testar a arquitetura e os modelos propostos. Na abordagem proposta, os dados utilizados foram obtidos das bases de dados do sistema de gestão acadêmica da UFRJ. São utilizados diversos algoritmos de mineração de dados, com o objetivo de fazer previsões de desempenhos destes estudantes, entre eles *J48*, *MultiLayer Perceptron*, *Random Forest* e *Support Vector Machine (SVM)*. Os resultados dos experimentos mostram que a arquitetura proposta é capaz de prever o desempenho acadêmico dos graduandos a cada semestre letivo com precisão em torno de 80%. Além da previsão, o estudo também identificou as principais variáveis que distinguem

os estudantes que obtêm sucesso ou não na conclusão do curso de graduação.

Este trabalho diferencia-se dos demais supracitados por utilizar técnicas baseadas em *Principal Components Analysis* (PCA) e *Linear Discriminant Analysis* (LDA) na etapa de pré-processamento para redução dos dados que representam os estudantes. Para reduzir a quantidade de informações utilizadas para classificação, (GOTTARDO *et al.*, 2014) utiliza um subconjunto de características dos estudantes extraídas a partir da correlação destas com o resultado final como forma de reduzir os dados, porém esta abordagem não trouxe melhorias na acurácia do classificador *Multilayer Perceptron* (MLP) e *RandomForest*. Não encontramos na literatura a utilização do PCA em conjunto com a LDA para redução da dimensionalidade de dados que serão classificados com Redes Neurais Artificiais (RNA) do tipo MLP afim de predizer o desempenho acadêmico de estudantes, como proposto neste trabalho.

4 PRÉ-PROCESSAMENTO DOS DADOS

O pré-processamento dos dados é uma etapa essencial da Mineração de Dados, do inglês *Data Mining* (DM), que aborda vários problemas importantes do processo de Descoberta de Conhecimento em Dados, do inglês *Knowledge Database Discovery* (KDD). Os dados provenientes de bases de dados do mundo real geralmente têm inconsistências, erros, valores ausentes, ou simplesmente não estão adequados para a utilização nos processos de DM (GARCÍA *et al.*, 2015).

A etapa de pré-processamento consiste na preparação os dados para atenderem os requisitos de entrada de cada algoritmo de DM que serão utilizados nas próximas etapas de DM. São realizadas integração de dados, limpeza, normalização e transformação dos dados, além da redução dos dados. Tais técnicas visam melhorar a acurácia e eficiência de algoritmos de DM.

Dentre as diversas técnicas de pré-processamento de dados, destacamos:

- **Limpeza dos dados:** É comum que informações de bases de dados do mundo real sejam incompletas, com ruídos (*outliers*) e inconsistentes. As tarefas de limpeza dos dados visam identificar inconsistências nos dados, preencher ou remover amostras com falta de informações, além de tentar minimizar o efeito dos ruídos sobre os dados (HAN, 2001);
- **Integração de dados:** A atividade de mineração de dados geralmente requer integração de dados advindos de diversas bases de dados;
- **Redução de dados:** é uma técnica aplicada para obter uma amostra que represente todo o conjunto de dados. Esta amostra, apesar de menor, deve manter o máximo de representatividade possível dos dados originais. Um exemplo desta técnica é a redução da dimensionalidade, que visa diminuir a quantidade de atributos dos vetores de características mantendo a expressividade dos dados;
- **Normalização dos dados** pode ser aplicado quando é necessário mudar a escala dos números e assim diminuir a amplitude dos mesmos, fazendo com que grandes valores sejam convertidos para números nos intervalos de 0 e 1, ou -1 e 1, por exemplo (HAN, 2001).

Nenhuma técnica supracitada é mutualmente exclusiva, ou seja, elas podem ser abordadas em conjunto. Este trabalho utiliza técnicas de normalização de dados e de redução da dimensionalidade dos vetores de características com o objetivo de obter

melhores resultados nos processos de classificação.

4.1 Normalização dos Dados

A normalização é uma técnica de manipulação de dados que objetiva tornar a grandeza dos atributos dos vetores de características em uma mesma escala (HAYKIN *et al.*, 2009). As técnicas de normalização não devem alterar a forma que os dados estão distribuídos, assim, as informações implícitas de cada atributo são mantidas. É comum que os dados normalizados tenham intervalos entre 0 e 1, ou -1 e 1.

A normalização torna os dados mais adequados ao processo de treinamento de uma rede neural, pois evita a saturação dos neurônios, acelera o tempo de treinamento de uma rede neural, além de poder melhorar a acurácia e a eficiência de algoritmos de mineração que envolvem medidas de distância (HAN *et al.*, 2011; PRIDDY; KELLER, 2005).

4.1.1 Normalização dos Dados com *z-score*

Na normalização pelo *z-score* cada atributo \mathbf{x} de um vetor de características é normalizado baseado na sua média e desvio padrão, como definido na equação 4.1:

$$\mathbf{x} = \frac{(\mathbf{x} - \mu_x)}{\sigma_x} \quad (4.1)$$

em que μ_x é a média dos valores do atributos e σ_x é o desvio padrão.

4.2 Redução da Dimensionalidade

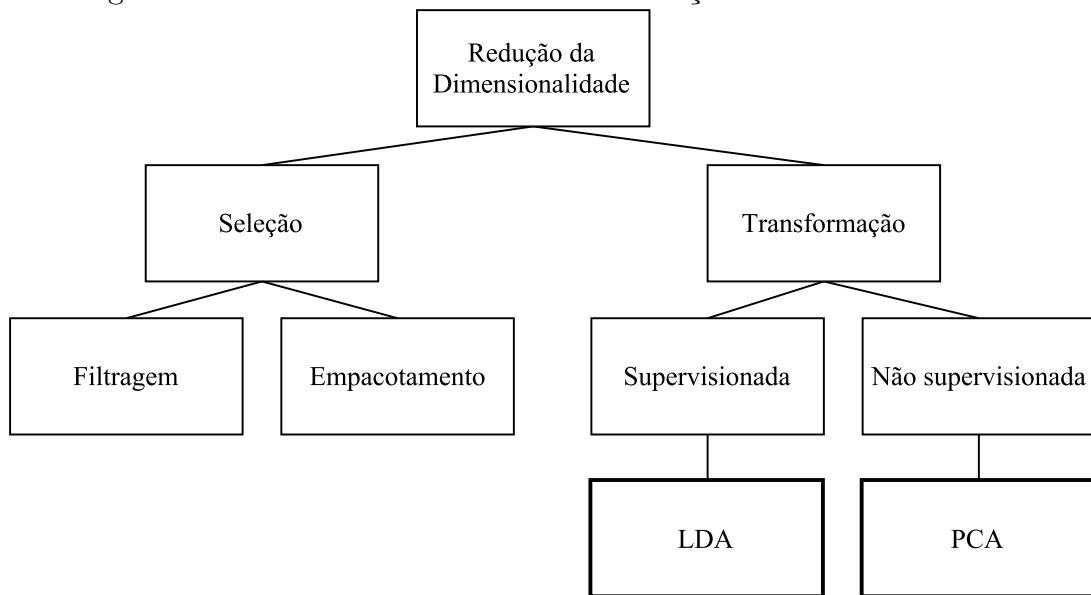
A Redução da Dimensionalidade consiste em reduzir o número de características utilizadas para representar um determinado dado. O conjunto de características reduzido deve ter o poder de representar os dados tal qual o conjunto dos dados originais. Vetores de características com a quantidade atributos reduzidos simplificam a representação do padrão, e desta forma, o processo de comparação será mais rápido e utilizará menos memória (JAIN *et al.*, 2000).

Em termos matemáticos, a redução da dimensionalidade pode ser interpretada matematicamente como: Seja $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ um conjunto de n vetores de um espaço

d -dimensional, $x_i \in \mathbb{R}^m$. A redução da dimensionalidade visa encontrar um conjunto correspondente de vetores $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$, de forma que $y_i \in \mathbb{R}^p$, sendo $p < m$ e \mathbf{Y} a maior representatividade possível de \mathbf{X} no espaço de menor dimensão.

A figura 5 mostra as diversas técnicas de redução de dimensionalidade, e estas podem ser classificadas em técnicas de seleção de características ou transformação de características.

Figura 5 – Taxonomia das Técnicas de Redução da Dimensionalidade



Fonte – O autor

Para reduzir as dimensões dos vetores de características, dois métodos de redução de dimensionalidade são empregados neste trabalho. O primeiro é a Análise dos Componentes Principais, do inglês *Principal Components Analysis* (PCA), que é uma técnica não-supervisionada utilizada para determinar vetores de características em um espaço projetado (MARTÍNEZ; KAK, 2001) e a Análise Discriminante Linear, do inglês *Linear Discriminant Analysis* (LDA), que é uma técnica de redução de dimensionalidade supervisionada, pois leva em consideração a informação da classe das amostras e visa minimizar a distância entre os elementos dentro da classe e maximizar a distância entre as classes (MARTÍNEZ; KAK, 2001).

A redução da dimensionalidade dos vetores de características implica nos seguintes benefícios:

- Melhora a qualidade dos dados através da remoção de atributos irrelevantes ou redundantes;

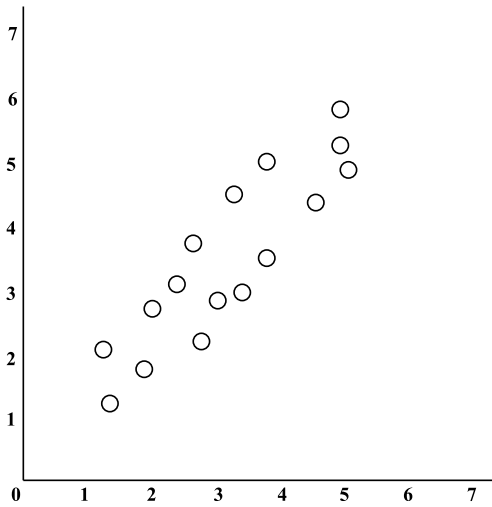
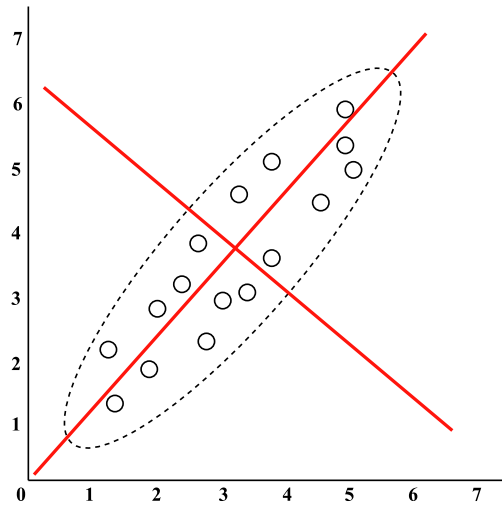
- Com menos dados para ser processado, há um aumento na velocidade do processamento e assim os algoritmos de DM podem aprender mais rápido;
- Aumenta o desempenho preditivo do algoritmo de DM, assim, o modelo generaliza melhor a partir dos dados;
- Tornam os resultados mais simples, para que estes sejam fáceis de entender e usar;

4.2.1 *Análise dos Componentes Principais (PCA)*

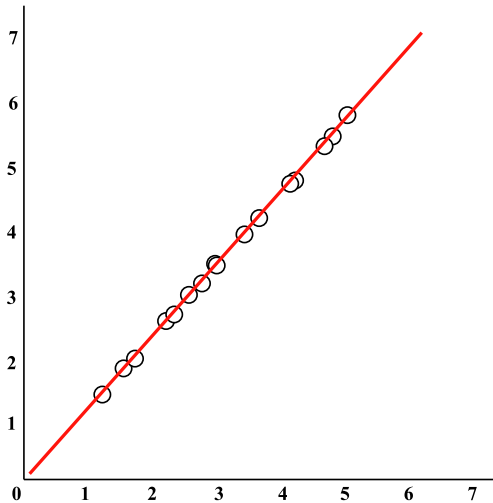
A Análise de Componentes Principais (PCA), é uma técnica de redução de dimensionalidade não-supervisionada, baseada na variância dos dados, e tem o propósito de derivar novas variáveis (em ordem decrescente de importância) que são combinações lineares das variáveis originais e que são descorrelacionadas (WEBB,). O PCA transforma um conjunto de variáveis do espaço \mathbb{R}^k em um outro conjunto no espaço \mathbb{R}^p contendo a máxima variância dos dados, e p é um valor menor que k . A matriz de transformação \mathbf{W} representa a transformação linear que mapeia os vetores de características originais $\mathbf{x}_i \in \mathbb{R}^k$ de um espaço k -dimensional para um vetor $\mathbf{y}_i \in \mathbb{R}^p$ com dimensões reduzidas de um subespaço p -dimensional.

Para ilustrar a PCA, utilizamos o conjunto de dados em \mathbb{R}^2 , dispostos no gráfico da figura 7(a). A figura 7(b) mostra a variabilidade dos dados no círculo preto. As linhas vermelhas representam as componentes principais, que são ortogonais em relação a outra, centrados na média e percorrem um caminho de maior variância dos dados em seus respectivos eixos. As figuras 7(c) e 7(d) mostram a projeção dos dados na primeira e segunda componentes, respectivamente. A primeira componente principal melhor representa a variância dos dados, de forma que os dados são transformados de \mathbb{R}^2 para \mathbb{R} com o mínimo de perda de informações.

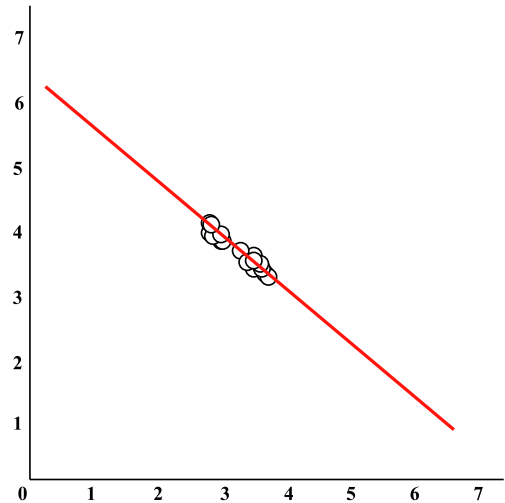
Figura 6 – Etapas do PCA.

(a) Dispersão dos dados em \mathbb{R}^2 

(b) Covariância e Componentes Principais dos Dados



(c) Projeção dos Dados na 1ª Componente Principal



(d) Projeção dos Dados na 2ª Componente Principal

Fonte – O autor

Matematicamente, os passos do PCA são descritos a seguir:

Passo 1: Calcular a média \mathbf{m} dos vetores de características:

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (4.2)$$

em que N é a quantidade de amostras.

Passo 2: Calcular a matriz de covariância \mathbf{S} :

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N (\mathbf{m}_i - \mathbf{x})(\mathbf{m}_i - \mathbf{x})^T \quad (4.3)$$

Passo 3: Resolver a decomposição:

$$\lambda_i \mathbf{e}_i = \mathbf{S} \mathbf{e}_i \quad (4.4)$$

em que λ_i é o autovalor associado ao autovetor \mathbf{e}_i .

Passo 4: Obter a matriz de transformação \mathbf{W} pela escolha dos p primeiros autovetores que foram ordenados de acordo com seus autovalores correspondentes λ_i em ordem decrescente:

$$\mathbf{W} = \{\mathbf{e}_i\}_{i=1}^p \quad (4.5)$$

Passo 5: No último passo, o vetor de características original \mathbf{x}_i pode ser transformado em um vetor de características \mathbf{y}_i com dimensão reduzida pela equação:

$$\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i \quad (4.6)$$

Entre as principais motivações para o uso do PCA estão:

- Proporcionar menor tempo de execução e uso de memória por algoritmos de aprendizado de máquina, devido sua capacidade de reduzir as dimensões dos vetores de características;
- Permitir a visualização dados multidimensionais em gráficos com 2 ou 3 dimensões.

4.2.2 *Análise Discriminante Linear (LDA)*

A Análise Discriminante Linear (LDA) é um método supervisionado de redução de dimensionalidade utilizado em problemas de reconhecimento de padrões. O objetivo da LDA é utilizar uma matriz de transformação \mathbf{W} , que seja capaz de maximizar a proporção da matriz da variabilidade inter-classes \mathbf{S}_B e diminuir a matriz da variabilidade intra-classes \mathbf{S}_w , em outras palavras, maximizar as distâncias entre as classes e minimizar a distância entre os objetos dentro de uma classe, e com a transformação linear, mapeia os vetores de características originais em vetores de um espaço com menor dimensão.

Para a redução da dimensionalidade, a LDA utiliza três matrizes. A primeira, matriz da variabilidade entre as classes é obtida pela equação:

$$\mathbf{S}_W = \sum_{i=1}^N \sum_{j=1}^{n_i} (\mathbf{x}_j^i - \mathbf{m}_i)(\mathbf{x}_j^i - \mathbf{m}_i)^T \quad (4.7)$$

em que $\mathbf{x}_j^i \in \mathbb{R}^k$ representa a j -ésima amostra da i -ésima classe e k é a dimensão do espaço de características; \mathbf{m}_i é a média da i -ésima classe; n_i é o numero de amostras na i -ésima classe; e N é a quantidade de classes. \mathbf{S}_W é a soma da matriz de covariância.

A segunda, matriz da variabilidade dos objetos dentro da classe é obtida pela equação:

$$\mathbf{S}_B = \sum_{i=1}^N n_i (\mathbf{m}_i - \mathbf{m}_{all})(\mathbf{m}_i - \mathbf{m}_{all})^T \quad (4.8)$$

em que \mathbf{m}_{all} é a média de todas as classes; \mathbf{S}_B é a soma das distâncias quadradas entre a média de cada classe e as médias de todas as classes.

Após obtidas as matrizes de variabilidade \mathbf{S}_W e \mathbf{S}_B , a terceira matriz \mathbf{W} pode ser obtida pela maximização do seguinte critério de Fisher:

$$\max_{\mathbf{W}} = \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}} \quad (4.9)$$

em que $\mathbf{W}^T \mathbf{S}_W \mathbf{W}$ e $\mathbf{W}^T \mathbf{S}_B \mathbf{W}$ são as novas matrizes da variabilidade inter-classes e variabilidade intra-classes no novo espaço de características, respectivamente.

Adicionalmente, o vetor \mathbf{y} com dimensões reduzidas pode ser obtido por $\mathbf{y} = \mathbf{W}^T \mathbf{x}$, sendo \mathbf{x} o vetor de características original.

De forma geral, \mathbf{W} pode ser obtida aplicando uma decomposição de autovalores de $\mathbf{S}_W^{-1} \mathbf{S}_B$, quando existe a matriz inversa de \mathbf{S}_W , pela equação a seguir:

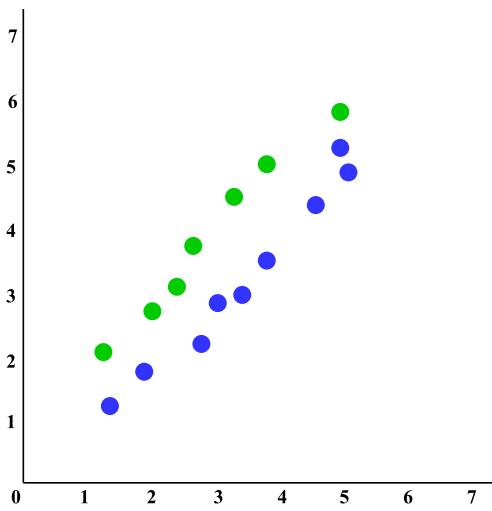
$$\mathbf{S}_B \mathbf{W} = \lambda \mathbf{S}_W \mathbf{W} \quad (4.10)$$

Os vetores de características com dimensões reduzidas após a transformação pela LDA terá o número de dimensões limitada a $N - 1$, em que N é o número de classes. Isto é, as colunas de \mathbf{W} são iguais aos autovetores com os $N - 1$ autovalores diferentes de zero da equação 4.10 (FUKUNAGA, 2013).

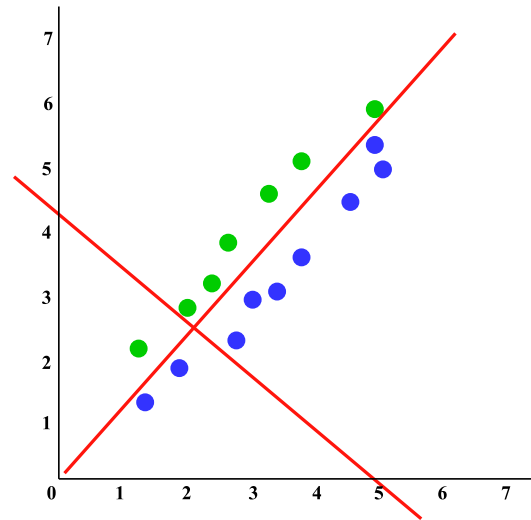
Para ilustrar a LDA, utilizamos um conjunto de dados em duas classes, mostrado na figura 8(a). A LDA ocorre com a projeção dos dados em um plano ortogonal ao hiperplano de separação, em que os novos eixos são denominados de eixos discriminantes. Os hiperplanos de separação dos dados e o plano ortogonal possuem o mesmo número de dimensão dos dados, desta forma, podemos encontrar um número de eixos discriminantes igual ao número de dimensões dos dados. Como os dados deste exemplo possuem duas classes, serão obtidos dois eixos discriminantes, como visto na figura 8(b). Nota-se que o primeiro eixo, na figura 8(c) representa melhor a separação dos dados, não levando em consideração a variância dos dados, como feito pelo PCA.

Em resumo, a normalização dos dados reduz a amplitude dos dados. A redução da dimensionalidade diminui a dimensão dos dados que representam os estudantes. As

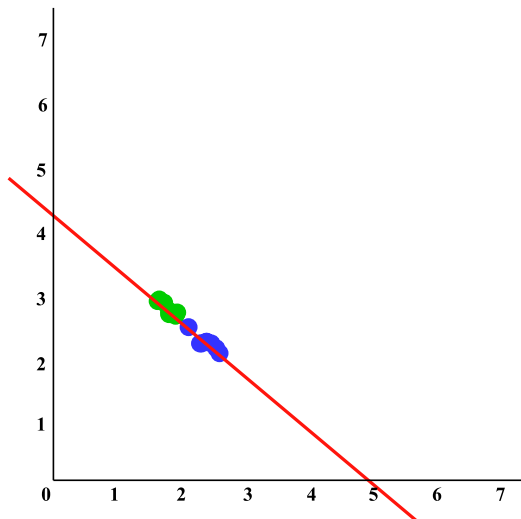
Figura 7 – Etapas da LDA.



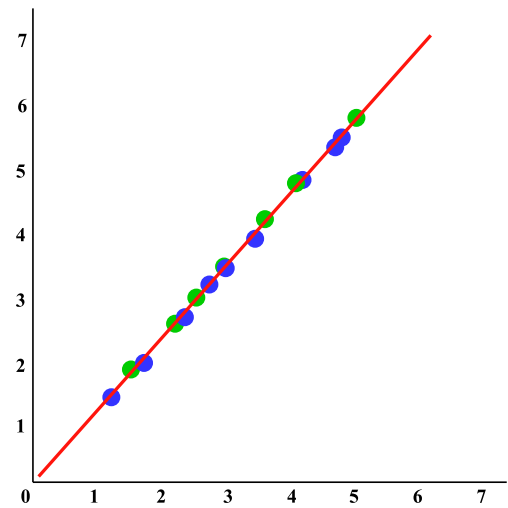
(a) Dados em 2 classes



(b) Eixos discriminantes



(c) Projeção no eixo com maior autovalor



(d) Projeção no eixo com menor autovalor

Fonte – O autor

técnicas apresentadas preparam os dados para etapa de classificação, que será descrita no próximo capítulo.

5 MÉTODOS DE CLASSIFICAÇÃO

Este capítulo introduz o método de reconhecimento de padrões baseados em classificação utilizado para predição do desempenho de estudantes em Ambientes Virtuais de Aprendizagem (AVA) e as estratégias aplicadas para validar os processos de classificação.

Reconhecimento de padrões é uma área do conhecimento no qual um dos objetivos é a classificação de objetos em determinadas classes (THEODORIDIS; KOUTROUMBAS, 2006). Reconhecimento de padrões visa construir uma representação mais simples de um conjunto de dados através de suas características mais relevantes, possibilitando a divisão destes dados em classes (DUDA *et al.*, 2012).

Em reconhecimento de padrões, cada amostra de um conjunto é representado por um **vetor** de características $v = [x_1, x_2, \dots, x_n]$, em que cada x_i para $(1 \leq i \leq n)$ é um variável que pode ser numérica (contínua ou discreta) e/ou simbólicas e equivale a uma característica ou atributo deste padrão. O espaço formado pelos vetores de características é denominado espaço de características e possui dimensão n (THEODORIDIS; KOUTROUMBAS, 1999). Uma **classe** é o conjunto de padrões com características em comum que, geralmente possuem um identificador simbólico ou numérico, denominado rótulo (do inglês *label*). O classificador fará a decisão de qual classe pertence o padrão de acordo com as informações contidas no vetor de características.

Neste trabalho, a classificação é realizada com Redes Neurais Artificiais (RNA) do tipo *Multilayer Perceptron* (MLP) e visa o reconhecimento de padrões nos vetores de características com informações sobre interações dos estudantes em AVA.

5.1 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são modelos computacionais projetados com base na estrutura do cérebro humano. Uma rede neural é um processador distribuído e paralelizado composto de unidades simples de processamento, denominadas neurônios, que têm uma tendência natural de armazenar conhecimento experimental e disponibiliza-lo para uso (HAYKIN, 1999). RNA possuem a capacidade de aprender por meio de um conjunto reduzido de exemplos e generalizar a informação aprendida, ou ainda, identificar padrões automaticamente de um conjuntos de dados complexos (BRAGA *et al.*, 2000).

RNA são semelhantes ao cérebro em dois aspectos:

- O conhecimento é adquirido pela rede por meio de um processo de aprendizagem;
- Os pesos de conexão entre neurônios, conhecidos como pesos sinápticos, são utilizados para armazenar o conhecimento adquirido.

Semelhante ao cérebro humano, as RNA passam inicialmente por uma fase de aprendizagem, em que uma amostra de um conjunto é apresentado para a rede, dos quais extrai as características necessárias para representar a informação fornecida. Com este aprendizado, a rede é capaz de reconhecer padrões nos demais objetos que não estavam na amostra inicial.

O primeiro modelo de neurônio artificial foi proposto por McCulloch e Pitts em 1943 (MCCULLOCH; PITTS, 1943), o mesmo é conhecido como neurônio de McColluch e Pitts e é o modelo de neurônio mais utilizado em diferentes arquitetura de RNA. A figura 8 mostra a estrutura de um neurônio artificial.

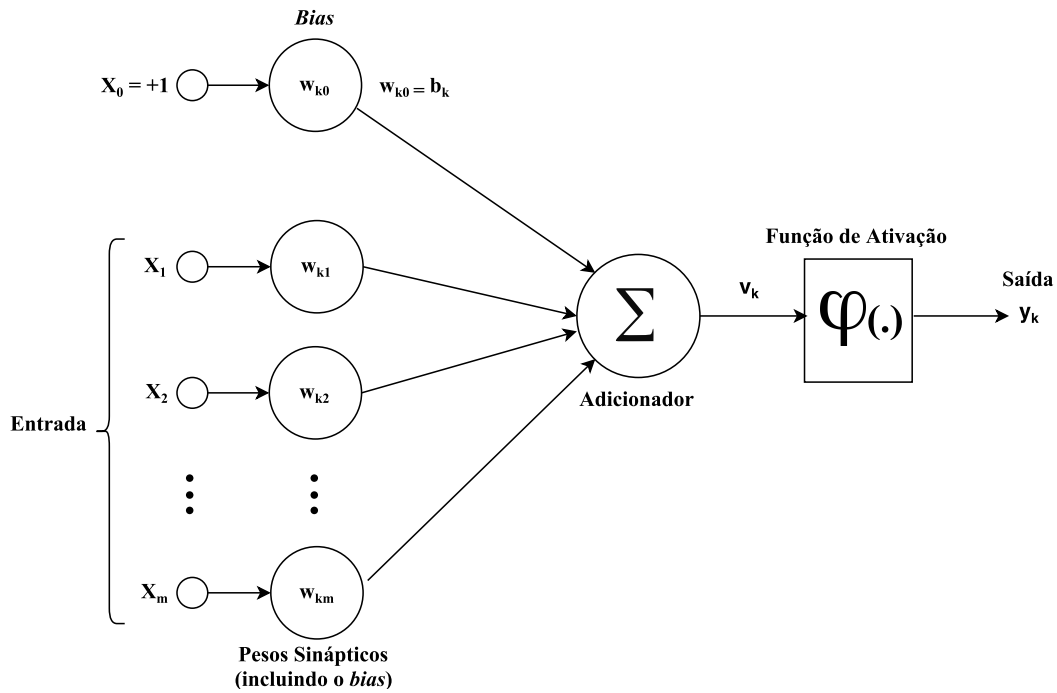
Os diversos tipos de redes neurais variam de acordo com a quantidade de camadas de neurônios, tipos de função de ativação, ligações entre os neurônios, porém, um item comum a todas elas: o neurônio. Um neurônio é a unidade de informação fundamental para a operação de uma rede neural e é composto pelos seguintes elementos:

- Um conjunto de **sinapses**, sendo que cada sinapse é caracterizada por um peso. Uma sinapse j , que está conectada a um neurônio k , recebe um sinal (valor) x_j como entrada, que por sua vez é multiplicado pelo peso sináptico W_{kj} .
- Um **adicionador** funciona como um combinador linear, pois, soma os valores de entrada já multiplicados pelos pesos sinápticos do neurônio.
- Uma **função de ativação** para limitar a amplitude da saída do neurônio. Geralmente, o intervalo de amplitude normalizada da saída de um neurônio está entre $[0, 1]$ ou $[-1, 1]$.

Um valor adicional, denominado *bias* b_k é utilizado para aumentar (*bias* com valor positivo) ou diminuir (*bias* com valor negativo) o resultado do cálculo do combinador linear na função de ativação. Nos neurônios artificiais os valores dos pesos podem ser negativos, se as sinapses forem inibitórias, ou positivos, se as sinapses forem excitatórias. A figura 8 exemplifica o modelo de um neurônio.

Um neurônio K pode ser matematicamente descrito pelas equações 5.1 e 5.2

Figura 8 – Modelo de um neurônio artificial.



Fonte – O autor

(HAYKIN, 1999):

$$v_k = \sum_{j=1}^m w_{kj} x_j \quad (5.1)$$

$$y_k = \varphi(v_k + b_k) \quad (5.2)$$

em que

- x_1, x_2, \dots, x_n é um vetor que contém os sinais de entrada;
- $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio K ;
- v_k é o resultado da combinação linear;
- b_k é o *bias*;
- $\varphi(\cdot)$ é a função de ativação;
- e y_k é o sinal de saída do neurônio.

5.1.1 Perceptron

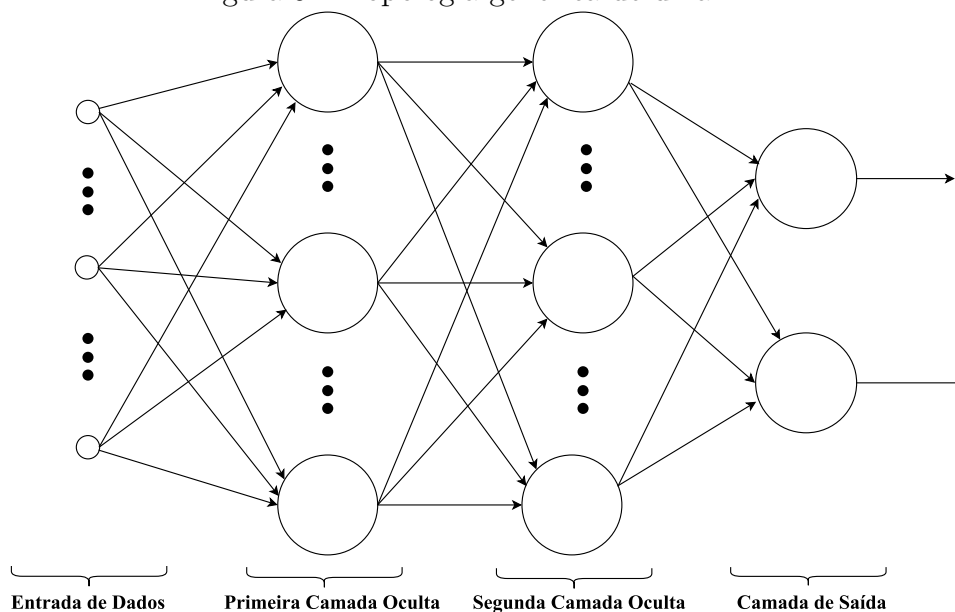
Perceptron é uma rede neural artificial com topologia simples, dispondo de uma única camada, a de saída, que por sua vez é composta por neurônios McCulloch e Pitts,

como mostrado na figura 8. A *perceptron* possui somente uma camada com propriedades adaptativas, a de saída. Esta peculiaridade tornou a rede conhecida como *perceptron* de única camada (BRAGA *et al.*, 2000). Esta RNA limita-se a resolver problemas linearmente separáveis.

5.1.2 *Multilayer Perceptron (MLP)*

As redes neurais artificiais do tipo *Multilayer Perceptron* (MLP), são compostas por um conjunto de unidades de entrada de dados que constituem uma camada de entrada, uma ou mais camadas intermediárias (ocultas) de neurônios, e uma camada de saída, também composta por neurônios. A motivação para utilizam de redes com esta estrutura é que elas aprendem, através das camadas intermediárias, tarefas complexas ao extrair progressivamente características significativas dos padrões de entrada (HAYKIN, 1999). Como visto, redes como a *perceptron*, que contém apenas uma camada, se limitam a resolver problemas linearmente separáveis. Por outro lado, as MLP podem ter em sua estrutura várias camadas intermediárias de neurônios, característica fundamental para esta possa separar elementos não separáveis por um hiperplano, ou seja, uma separação não-linear dos elementos. A figura 9 mostra uma MLP com duas camadas ocultas.

Figura 9 – Topologia genérica de uma MLP



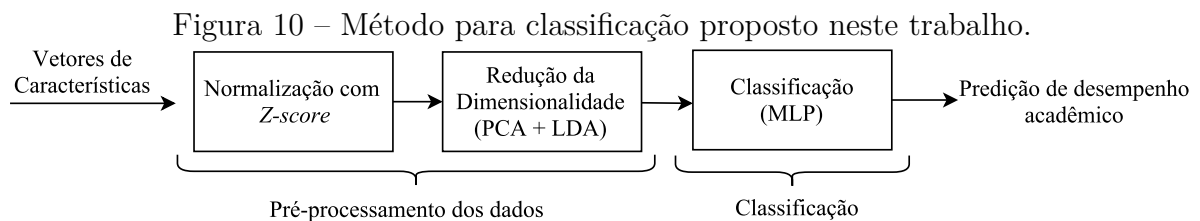
Fonte – O autor

5.1.2.1 Treinamento de Redes MLP

MLP utiliza o algoritmo *backpropagation* para treinamento da rede (LEUNG; HAYKIN, 1991). Este algoritmo é supervisionado e utiliza pares de entrada e saída (x , y_d) para, por meio dos mecanismos de correção de erros, ajustar os pesos da rede. O treinamento ocorre em duas fases, cada uma percorre um sentido diferente da rede. Estas fases são *forward* e *backward*. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada. A fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões (BRAGA *et al.*, 2000).

5.2 Métodos de Classificação Propostos

Este trabalho aborda normalização e redução dos dados na etapa de pré-processamento dos dados e redes neurais artificiais do tipo MLP para classificação dos estudantes quanto ao seu desempenho acadêmico.



Fonte – O autor

A normalização de dados é uma técnica aplicada quando os valores dos atributos têm grande amplitude, muitas vezes por estes usarem escalas diferentes. Neste trabalho, os valores estão em uma mesma escala, mas, com grande amplitude, como mostrado na tabela 4. Para obter melhores resultados na classificação dos estudantes com a MLP, conforme exposto na seção 4.1, os valores foram normalizados com *z-score*.

A redução da dimensionalidade foi abordada para transformação dos dados originais em dados mais compactos, e desta forma remover informações irrelevantes nos vetores de características, além de melhorar o processo de classificação com a MLP. Neste trabalho, utilizamos o *Principal Components Analysis* (PCA) e a *Linear Discriminant Analysis* (LDA) na etapa de pré-processamento dos dados. Ambas as técnicas são abordadas em conjunto em outro experimento.

Todos os processos de classificação foram realizados com a rede neural MLP e

avaliados com validação cruzada *K-fold*, com **K** igual a 10.

Este trabalho aborda os seguintes métodos:

- ***z-score* + PCA + MLP**: Os dados são normalizados com *z-score*, e reduzidos com o PCA;
- ***z-score* + LDA + MLP**: Os dados são normalizados com *z-score* e reduzidos com o LDA;
- ***z-score* + PCA + LDA + MLP**: Os dados são normalizados com *z-score*, reduzidos com o PCA e reduzidos novamente com o LDA;

A última etapa dos métodos descritos acima é a classificação com a rede neural MLP.

5.3 Estratégias de Validação

A acurácia preditiva informa a capacidade do modelo de classificação prever corretamente a classe de um objeto desconhecido. Esta estimativa permite a validação do modelo gerado, bem como a comparação de desempenhos entre diferentes classificadores sobre uma mesma base de teste.

5.3.1 Validação Cruzada

Neste trabalho, para obter a acurácia preditiva dos modelos de classificação, utilizou-se técnicas de validação cruzada, do inglês *cross-validation*, que são técnicas estatísticas úteis para determinar a capacidade de generalização do modelo a partir dos dados (DUDA *et al.*, 2001).

Para realizar a validação cruzada, a base de dados é dividida em duas partes, são elas:

- **Treinamento**: Utilizados para treinar o modelo. O treinamento permite que sejam observadas características dos objetos e as relações entre os dados de entrada e saída para criar uma representação deste dados.
- **Teste**: Geralmente é um conjunto menor que o utilizado no treinamento. O conjunto de teste permite a validação do modelo gerado.

A validação cruzada garante que os dados utilizados para treinamento e teste são diferentes. Garantindo assim, uma maior veracidade na capacidade de generalização do modelo preditivo.

A seguir, são descritas algumas variações da validação cruzada.

5.3.2 *K-fold*

A validação cruzada *K-fold* divide um conjunto de dados \mathbf{D} em K partes de tamanhos iguais, chamadas *folds* e representadas por $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_K$. Cada parte \mathbf{D}_i é, por sua vez, separada para ser o conjunto de teste, e as demais partes juntas serão o conjunto de treinamento $\mathbf{D} - \mathbf{D}_i$. Após treinar o modelo M_i com $\mathbf{D} - \mathbf{D}_i$, será avaliado seu desempenho com o conjunto de teste \mathbf{D}_i para obter a i -ésima estimativa θ_i . A acurácia será a média dos percentuais de acertos nas K iterações e pode ser estimado por:

$$\hat{\mu}_\theta = \frac{1}{K} \sum_{i=1}^K \theta_i \quad (5.3)$$

e a sua variância é dada por

$$\hat{\sigma}_\theta^2 = \frac{1}{K} \sum_{i=1}^K (\theta_i - \hat{\mu}_\theta)^2 \quad (5.4)$$

O algoritmo ?? descreve o pseudo-código para validação cruzada *K-fold*.

A literatura informa que K igual a 10 é o mais adequado para se obter uma boa estimativa da precisão do classificador (YANG; PEDERSEN, 1997). Uma desvantagem desta técnica é o custo computacional excessivo, uma vez que para avaliar o modelo são realizados K treinamentos (HAYKIN, 1999).

5.3.3 *Leave-one-out*

Leave-one-out é um método de validação cruzada *k-fold*, em que K é igual ao número de instâncias no conjunto de dados. Uma instância é retirada e as demais são utilizadas pelo método de aprendizagem para treinamento. O experimento é realizado K vezes, cada vez um dado diferente é usado para validação. A acurácia será a média dos acertos em todas as iterações. Este procedimento é interessante quando a quantidade de dados disponíveis é limitada.

5.3.4 *Hold-out*

Hold-out separa os dados em dois conjuntos, em muitos casos é comum separar um terço dos dados para teste e os dois terços restantes para treinamento (JIAWEI; KAMBER, 2001). Com o objetivo de evitar que a acurácia seja influenciada por uma

particularidade nos dados do conjunto de teste, é comum realizar o *hold-out* K vezes, gerando assim K pares (treinamento e teste) de conjunto de dados distintos. A acurácia é obtida a partir da média dos percentuais de acerto obtidos a partir de cada um dos K testes. Esta técnica é recomendada quando existe uma grande quantidade de dados.

A escolha do método para classificação pode se dá a partir da análise de outros fatores, além da acurácia preditiva (HAN, 2001), são eles:

- **Desempenho:** custo computacional para geração e aplicação do modelo.
- **Robustez:** capacidade do modelo de classificar corretamente instâncias a partir de dados com *outliers* ou valores de atributos desconhecidos.
- **Escalabilidade:** capacidade de construir modelos a partir de grandes quantidades de dados.
- **Interpretabilidade:** verificação de quão compreensível é o modelo ao usuário da aplicação.

5.4 Matriz de Confusão

Em um processo de classificação com apenas duas classes, por exemplo, *aprovado* ou *reprovado*, cada predição pode gerar resultados conforme a figura 11. Os verdadeiros positivos (V_P) e os verdadeiros negativos (V_N) são classificações corretas. Um falso positivo (F_P) ocorre quando o resultado foi predito incorretamente como *aprovado*. Um falso negativo (F_N) ocorrerá quando um resultado é predito incorretamente como *reprovado*. A taxa de verdadeiros positivos é V_P dividida pelo total de números positivos, que é $V_P + V_N$. A taxa de falsos negativos é F_N dividido pelo total de números dos negativos, $F_N + V_P$. A taxa de sucesso global, ou acurácia do classificador Ac , é o número de classificações corretas dividida pelo número total de classificações:

$$Ac = \frac{V_P + V_N}{V_P + V_N + F_P + F_N} \quad (5.5)$$

Em uma predição com mais de duas classes, o resultado pode ser informado em uma matriz de confusão com duas dimensões, com uma linha e uma coluna para cada classe. Cada célula da matriz mostra o número de acertos da predição em relação ao esperado. As linhas representam os resultados esperados e as colunas os valores preditos. Bons resultados são obtidos quando grandes números ficam na diagonal principal, indicando que

Figura 11 – Diferentes resultados para uma predição de duas classes.

		Classe Predita	
		Aprovado	Reprovado
Classe Correta	Aprovado	Verdadeiro Positivo	Falso Negativo
	Reprovado	Falso Positivo	Verdadeiro Negativo

Fonte – O autor

uma grande quantidade de predições das classes foram de acordo com as classes esperadas. A figura 12 mostra um exemplo de matriz de confusão para uma predição de três classes.

Figura 12 – Matriz de confusão com resultados de uma predição com três classes.

		Classes Previstas		
		A	B	C
Classes Corretas	A	84	15	1
	B	6	78	8
	C	0	10	59

Fonte – O autor

5.5 Medidas de Avaliação

5.5.1 Precisão

A precisão é definida pela razão entre os exemplos classificados corretamente, representado pelo parâmetro V_P (verdadeiros positivos), presente no numerador da equação 5.6, e todos aqueles que foram associados a uma determinada classe pelo modelo avaliado, representado pelo denominador da mesma equação, que calcula a soma entre V_P e F_P

(falsos positivos).

$$P = \frac{V_P}{V_P + F_P} \quad (5.6)$$

5.5.2 Revocação

A revocação é obtida por meio de todos os exemplos classificados corretamente, representado pelo parâmetro V_P presente no numerador da equação 5.7, e todos aqueles que deveriam ter sido associados a uma determinada classe pelo modelo avaliado, representado pelo denominador da mesma equação, que calcula a soma entre V_P e F_N (falsos negativos).

$$R = \frac{V_P}{V_P + F_N} \quad (5.7)$$

5.5.3 F-measure

F-measure é uma medida que calcula a média harmônica ponderada dos valores de precisão e revocação, representados por P e R , respectivamente, na equação 5.8.

$$F = 2 \cdot \frac{P \cdot R}{P + R} \quad (5.8)$$

6 EXPERIMENTOS

Este capítulo descreve a base de dados utilizada neste trabalho, os critérios para seleção de uma amostra dos dados, o processo de obtenção dos vetores de características dos estudantes e os ambientes para experimentos utilizados. Experimentos são realizados para verificar a capacidade dos métodos propostos predizer corretamente o desempenho acadêmico de estudantes. Por fim, os resultados são analisados e comparados com outros trabalhos da literatura.

6.1 Dados

Este trabalho utilizou dados reais gerados entre 2010 e 2012 nas bases de dados do AVA *Moodle*, versão 1.9, da instituição de ensino técnico INTEC (Instituto de Educação Técnico Profissional), localizada em Sobral, Ceará.

Por ser o objetivo deste trabalho utilizar métodos para predição de desempenho final de estudantes, não foram considerados estudantes desistentes, pois estes não tinham informações sobre seus resultados ou notas finais (essencial para o processo de classificação). A evasão de estudantes é foco de outros trabalhos e tem relação com outros fatores específicos, tais como, ambiente sócio-econômico, público-alvo dos cursos, número de disciplinas no semestre, tipo de admissão no curso (??MANHÃES *et al.*, 2011; ??; ??).

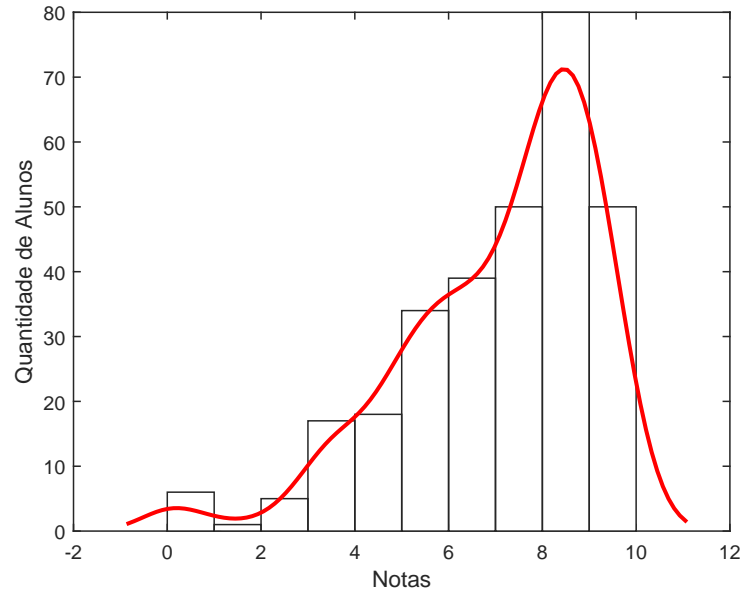
Dentre os diversos cursos (disciplinas) disponíveis na base de dados, seguiu-se os seguintes critérios para escolha de uma amostra para os experimentos:

- Maior quantidade de estudantes que concluíram a disciplina;
- Maior quantidade de oferta da disciplina para turmas diferentes;
- Disponibilidade do resultado de avaliações do conhecimento dos estudantes;
- Maior número de recursos do AVA *Moodle* utilizados.

Seguindo os critérios supracitados, foi escolhida uma disciplina com 62 turmas já encerradas. Estas turmas continham entre 5 a 40 concludentes. Foram selecionadas as 13 turmas com mais estudantes, totalizando 300 estudantes distintos.

A figura 13 mostra a distribuição das notas dos 300 estudantes.

Figura 13 – Distribuição das notas dos 300 estudantes das 13 turmas selecionadas.



Fonte – O autor

6.2 Vetor de Características

A fim de obter dados sobre interações realizadas pelos estudantes no *Moodle*, foi desenvolvido um *script* em SQL para consultar a base de dados. Alguns atributos obtidos com as consultas foram propostos por Gottardo (GOTTARDO *et al.*, 2014). Estes atributos são informações numéricas e representam os padrões de cada estudante. A tabela 3 lista todos os atributos que compõem o vetor de características que serão utilizados nos experimentos.

A tabela 4 mostra a distribuição dos valores numéricos de cada um dos atributos dos estudantes selecionados.

Alguns atributos da tabela 3 têm influência maior no resultado obtido pelo estudante na disciplina. Esta influência ocorre devido a correlação entre estes atributos. O coeficiente de correlação de *Pearson*, que é obtido pela divisão da covariância entre duas variáveis pelo produto de seus desvios padrão, mede a dependência linear entre duas variáveis (??). A tabela 5 mostra a correlação dos atributos, em ordem decrescente de correlação, que representam as interações dos estudantes no AVA e o atributo **resultado**, que é a nota final na disciplina. De acordo com (??), nenhum atributo utilizado neste trabalho têm correlação forte com o atributo **resultado**. Mesmo assim, todos eles têm correlação positiva, o que indica que o crescimento destes atributos contribui para o

Tabela 3 – Atributos do vetor de características do estudante no AVA

Atributo	Representação
1	<i>na</i> Número de acessos ao AVA
2	<i>npf</i> Número de postagens no fórum
3	<i>nch</i> Número de sessões de chat que o aluno participou
4	<i>nmec</i> Número de mensagens enviadas no chat
5	<i>nqr</i> Número de questões respondidas
6	<i>nqc</i> Número de questões respondidas corretamente
7	<i>nmr</i> Número de mensagens recebidas
8	<i>nme</i> Número de mensagens enviadas
9	<i>nrrpf</i> Número de respostas em postagens do professor no fórum
10	<i>nrrp</i> Número de mensagens recebidas do professor
11	<i>nmep</i> Número de mensagens enviadas ao professor
12	<i>naf</i> Número de acesso ao fórum do curso
13	<i>nat</i> Número de acesso aos tópicos de discussão do fórum do curso
14	<i>nbf</i> Número de buscas realizadas nos fóruns do curso
15	<i>naq</i> Número de acesso ao questionário
16	<i>namd</i> Número de acesso ao material didático
17	<i>nac</i> Número de acesso ao curso
18	resultado Desempenho do aluno no curso. Objetivo da predição.

Fonte – O autor

Tabela 4 – Distribuição dos dados dos estudantes selecionados.

Atributo	Menor Valor	Maior Valor	Média	Desvio Padrão
<i>na</i>	1	1274	249,83	206,13
<i>npf</i>	0	204	42,58	30,13
<i>nch</i>	0	12	2,14	2,4
<i>nmec</i>	0	256	17,5	29,25
<i>nqr</i>	0	464	223,84	138,64
<i>nqc</i>	0	15	6,23	2,98
<i>nmr</i>	42	678	138,16	75,33
<i>nme</i>	0	961	25,55	67,96
<i>nrrpf</i>	0	72	10,63	13,18
<i>nrrp</i>	0	480	88,04	58,62
<i>nmep</i>	0	541	12,34	35,24
<i>naf</i>	0	146	19,23	18,37
<i>nat</i>	0	183	19,15	20,94
<i>nbf</i>	0	61	1,24	4,35
<i>naq</i>	0	401	61,16	48,72
<i>namd</i>	0	338	20,76	28,31
<i>nac</i>	0	508	87,74	74,99
<i>resultado (nota final)</i>	0	10,0	7,08	2,07

Fonte – O autor

aumento da nota final do estudante. A tabela 5 mostra que os atributos que mais contribuem para o resultado são: número de questões respondidas, número de questões respondidas corretamente, número de postagens no fórum, número de sessões de chat que o aluno participou e número de acesso ao curso. Esta análise é importante pois na ausência de muitos atributos, a utilização de poucos atributos, mas com alta correlação com o resultado final e baixa correlação entre si, pode trazer resultados próximos aqueles que utilizam grandes quantidade de atributos, como identificado por (GOTTARDO *et al.*, 2014).

Tabela 5 – Correlação entre os atributos do vetor de características e o atributo resultado

Atributo	Coefficiente de correlação com o atributo <i>resultado</i>
<i>nqr</i>	0,3456
<i>nqc</i>	0,3426
<i>npf</i>	0,3214
<i>nch</i>	0,2798
<i>nac</i>	0,2514
<i>nat</i>	0,2403
<i>na</i>	0,2378
<i>nmec</i>	0,2366
<i>naf</i>	0,2143
<i>nmr</i>	0,1528
<i>namd</i>	0,1521
<i>nmrp</i>	0,1033
<i>nrfp</i>	0,0882
<i>nbf</i>	0,0558
<i>naq</i>	0,0537
<i>nme</i>	0,0324
<i>nmep</i>	0,0226

Fonte – O autor

O último atributo da tabela 3 refere-se ao resultado final do aluno no curso, esta informação é dada em valores contínuos, com valores entre 0,0 e 10,0, representando a nota final obtida na disciplina. Entretanto, estes valores foram discretizados para tornar possível a utilização de técnicas de classificação que exigem que os valores que representam uma classe sejam discretos, e adicionalmente, houvesse uma facilitação na interpretação dos resultados. O processo de discretização das notas finais dos estudantes ocorrem como descrito a seguir:

O desempenho acadêmico foi dividido em três níveis: **superior**, **intermediário** e **inferior**. Para isto, as notas foram discretizadas em três valores e os estudantes inseridos

nas classes **A**, **B** e **C** de acordo com suas notas. Os estudantes com as notas mais baixas e mais altas foram inseridos nas classes A e C, respectivamente. Os demais estudantes, com notas intermediárias, ficaram na classe B. O intervalo das notas foi definido a partir métricas de avaliação dos estudantes definidos pela instituição de ensino que forneceu os dados. Processo similar a este foi realizado por Gottardo (GOTTARDO *et al.*, 2012; GOTTARDO *et al.*, 2014), porém, adotando critérios diferentes para os intervalos das notas.

A tabela 6 mostra o intervalo de notas de cada uma das três classes.

Tabela 6 – Distribuição dos estudantes em três classes

Classe	Descrição	Intervalo de Notas	Números de Estudantes
C	Estudantes com desempenho superior	Notas maiores ou iguais a 6,0	219
B	Estudantes com desempenho intermediário	Notas menores que 6,0 e maiores ou igual a 3,0	69
A	Estudantes com desempenho inferior	Notas menores que 3,0	12

Fonte – O autor

6.3 Ambientes de Testes

Para fins de experimentos, utilizou-se a ferramenta *Weka*¹. *Weka* é um *software* de código aberto que fornece uma coleção de algoritmos de aprendizagem de máquina para diversas tarefas de mineração de dados, além de ferramentas para pré-processamento dos dados, classificação, regressão, agrupamentos, regras de associação e visualização de informações (??).

6.4 Experimentos Realizados

Para comprovar a validade dos métodos propostos e verificar a capacidade destes na predição do desempenho de estudantes em AVA, realizou-se os experimentos descritos a seguir.

¹ <http://www.cs.waikato.ac.nz/ml/weka/index.html> - Acessado em 07/nov/2016

6.4.1 Experimento 01: Dados Originais

Neste primeiro experimento, os dados que descrevem o comportamento dos estudantes nos AVA, vetores de características, não foram submetidos a nenhum tratamento para transformação ou redução dos dados. Desta forma, qualquer ruído ou informação irrelevante contido nos dados serão utilizados na etapa de treinamento da rede neural. O único procedimento realizado nesta etapa é a discretização das notas dos estudantes, conforme já informado na seção 6.2.

A rede neural MLP foi configurada com três camadas, camada de entrada, uma única camada oculta contendo 10 neurônios e uma camada de saída com 3 neurônios. O algoritmo de treinamento foi o *backpropagation* e as funções de ativação dos neurônios do tipo sigmoideal tangente hiperbólica. Cada treinamento da rede neural possui 500 épocas (ciclo de apresentação dos dados) e tem o objetivo de atingir o erro quadrático médio de 10^{-3} .

A figura 14 mostra a matriz de confusão da rede neural MLP na classificação dos estudantes. Os valores acima de cada coluna são os desempenhos (classes) esperados pela classificação, e os valores ao lado de cada linha, o desempenho predito pela rede neural, e a diagonal principal da matriz de confusão apresenta as classificações corretas. Nota-se que a rede neural MLP, neste experimento, não classificou corretamente nenhum estudante que pertence as classes **A** ou **B**, por outro lado, todos os estudantes da classe **C** foram classificados corretamente.

Figura 14 – Matriz de confusão da MLP no experimento 01.

	A	B	C
A	0	8.33% (1)	91.67% (11)
B	2.90% (2)	0	97.10% (67)
C	0	0	100.00% (219)

A tabela 7 mostra outras medidas do desempenho, precisão, revocação e *F-measure* da MLP neste experimento. A tabela permite uma análise detalhada do resultados da classificação por classe.

Tabela 7 – Medidas de Precisão, Revocação e *F-measure* da MLP no experimento 01.

Classe	Precisão	Revocação	<i>F-measure</i>
A	0,000	0,000	0,000
B	0,000	0,000	0,000
C	0,737	0,730	0,620
Média ponderada	0,538	0,730	0,620

O experimento 01, a MLP classifica corretamente 73% das amostras, desta forma, 219 estudantes no total de 300 foram classificados corretamente. Em alguns trabalhos também obteve-se acurácias próximas a 73% (SANTOS *et al.*, 2012; GOTTARDO *et al.*, 2012). Analisando apenas esta informação, pode-se concluir que a rede neural obteve resultados aceitáveis, porém esta conclusão pode induzir a erros, visto que neste experimento, o modelo gerado pela MLP não foi capaz de generalizar quanto aos estudantes com tendências a reprovação ou recuperação, classes A e B, respectivamente. Este resultado pode não ser muito útil, visto que existe a necessidade de identificar estudantes nos três níveis de desempenho. As justificativas para estes resultados podem ser a pouca quantidade de amostras para as classes A e B, e o desbalanceamento entre as classes, que ocorre, quando na etapa de treinamento são apresentadas uma quantidade excessiva de amostras de uma classe em relação a outras, diminuindo assim, a capacidade de generalização do classificador, que aprende muito sobre um padrão e pouco sobre outros. Este problema poderia ser resolvido adicionando novas amostras das classes A e B. Porém, como em todas as turmas da bases de dados utilizada contém um número maior de estudantes com perfil para classe C, a medida que fossem inseridos amostras das classes A e B destas turmas, também seriam inseridas da classe C. Optamos por não adicionar mais amostras.

Este experimento visa mostrar o impacto que os procedimentos descritos nos próximos experimentos trazem ao processo de classificação dos estudantes, reforçando a importância da etapa de pré-processamento dos dados e dos métodos propostos neste trabalho.

6.4.2 Experimento 02: Dados Normalizados com z -score

Para o experimento 02, os dados foram normalizados com z -score, para reduzir a amplitude dos valores dos atributos dos vetores de características. Apesar desta técnica ser comumente empregada em dados com diferentes escalas ou em um intervalo muito grande, e os valores numéricos dos atributos utilizados neste trabalho serem relativamente pequenos, esta técnica foi utilizada com o objetivo de trazer os benefícios descritos na seção 4.1 para o processo de classificação.

A figura 15 descreve o fluxo deste experimento, que consiste na normalização dos valores dos vetores de características antes do processo de classificação com a MLP.

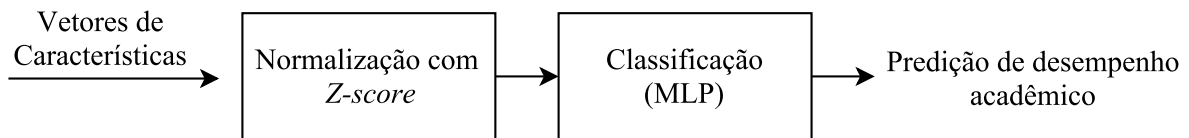


Figura 15 – Fluxo do método proposto para o Experimento 02.

Antes do processo de normalização dos dados com o z -score os valores numéricos dos atributos tinham amplitude informada na tabela 8. Esta diferença de valores ocorre devido os atributos conterem informações de diferentes interações dos estudantes no AVA, como mostrado na tabela 3, e algumas interações contém valores superiores a outras, por exemplo, o número de acessos ao AVA (**na**), que tem valor máximo igual a 1274, é bem maior que a quantidade máxima de questões respondidas corretamente por um estudante (**nqc**), que é igual a 15. Esta diferença de valores se repete em todas os vetores de características, pois é comum que o aluno tenha uma quantidade maior de acessos ao ambiente de virtual do que de questões respondidas corretamente em avaliações.

	Dados originais	Dados normalizados com z -score
Valor Máximo	1274	11,76
Valor Mínimo	0	-0,58
Média	59,88	0
Desvio Padrão	103,19	1,0

Tabela 8 – Informações sobre os valores dos atributos dos dados originais e após a normalização.

Neste experimento a rede neural MLP têm as mesmas configurações definidas para o experimento 01. Com os dados normalizados a MLP classificou corretamente 78,33%

dos estudantes, ou seja, 235 num total de 300. Além disto, a MLP identificou estudantes da classes A e B, acertando 41,67% e 46,38% em cada classe, respectivamente. Apesar de não ter havido erros para classe C no experimento 01, no experimento 02 a MLP errou em 9,59% dos estudantes. Este erro é aceitável, visto que é preferível para este estudo que o modelo tenha capacidade de prever o desempenho de diferentes tipos de estudantes, e não apenas um tipo. Estes resultados mostram o impacto que a normalização dos dados com *z-score* tem para o treinamento e acurácia da rede neural MLP. A matriz de confusão descrita na figura 16, mostra as classificações realizadas pela MLP. Nota-se que há uma distribuição maior de classificações corretas por classe.

	A	B	C
A	41.67% (5)	41.67% (5)	16.67% (2)
B	2.90% (2)	46.38% (32)	50.72% (35)
C	0	9.59% (21)	90.41% (198)

Figura 16 – Matriz de confusão da MLP no experimento 02.

Outras informações mais detalhadas dos resultados da rede neural MLP no processo de classificação no experimento 02 podem ser analisadas a partir da tabela 9.

Classe	Precisão	Revocação	<i>F-measure</i>
A	0,714	0,714	0,526
B	0,552	0,552	0,504
C	0,843	0,843	0,872
Média ponderada	0,771	0,783	0,774

Tabela 9 – Medidas de Precisão, Revocação e *F-measure* da MLP no experimento 02.

6.4.3 Experimento 03: Dados Normalizados com *z-score* e Reduzidos com PCA

O experimento 03 combina a técnica de normalização de dados *z-score* com redução da dimensionalidade baseada em PCA para obter melhores desempenhos no processamento e classificação com a rede neural MLP, como mostrado na figura 17.

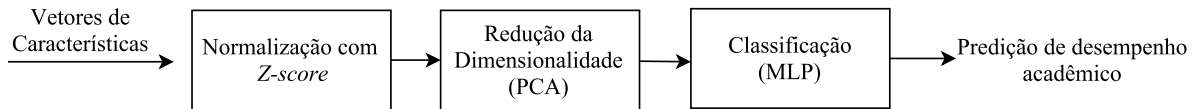


Figura 17 – Fluxo do método proposto para o experimento 03.

Inicialmente os vetores de características pertencem a um espaço \mathbb{R}^k , em que k é igual a 17. O objetivo desta abordagem, além de diminuir o tamanho dos vetores, é remover de informações irrelevantes contida neles, e assim obter melhor acurácia na classificação dos estudantes. Para estimar o número ótimo de componentes principais do PCA, foram realizados testes com componentes principais variando de 1 a 12, e para cada um destes, foram realizados experimentos com a MLP por 5 vezes. Um comparativo das médias das acurácias é informado na figura 18.

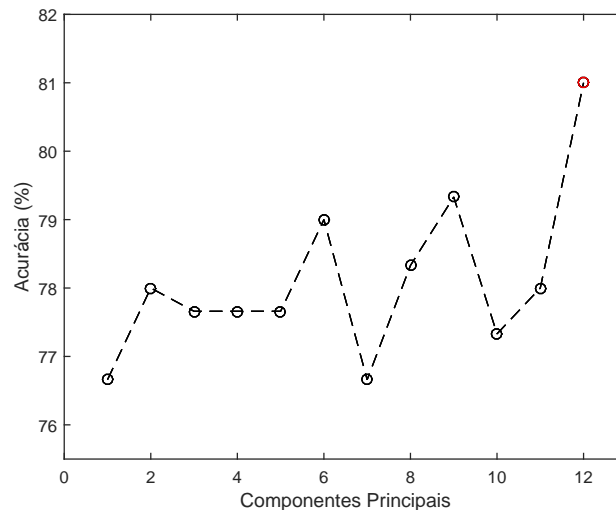


Figura 18 – Médias das acurácias da MLP para diferentes números de componentes principais.

Neste experimento, a rede neural MLP têm quatro camadas, camada de entrada, duas camadas ocultas com 10 neurônios cada, e uma camada de saída com 3 neurônios. As demais configurações são as mesmas dos experimentos anteriores.

Variando o número de componentes principais, a rede neural MLP obteve maior acurácia, 81%, quando utilizado 12 componentes principais, classificando corretamente 243 alunos. Neste experimento, a MLP conseguiu identificar estudantes em todos os níveis de desempenho, acertando 41,67%, 65,22% e 88,13%, para as classes A, B e C, respectivamente, como mostrado na matriz de confusão da figura 19. Novamente, a taxa de acertos da classe C diminuiu, mas houve um aumento para as classes A e B.

	A	B	C
A	41.67% (5)	50.00% (6)	8.33% (1)
B	5.80% (4)	65.22% (45)	28.99% (20)
C	0.91% (2)	10.96% (24)	88.13% (193)

Figura 19 – Matriz de confusão da MLP no experimento 03.

A tabela 10 informa as medidas de precisão, revocação e *F-measure* do processo de classificação no experimento 03.

Classe	Precisão	Revocação	<i>F-measure</i>
A	0,455	0,417	0,435
B	0,600	0,652	0,625
C	0,902	0,881	0,891
Média ponderada	0,815	0,810	0,812

Tabela 10 – Medidas de Precisão, Revocação e *F-measure* da MLP no experimento 03.

6.4.4 Experimento 04: Dados Normalizados com z-score e Reduzidos com LDA

Semelhante ao experimento 03, este experimento reduz a dimensionalidade dos dados na etapa de pré-processamento, neste caso, com LDA. O fluxo deste experimento é descrito na figura 20.

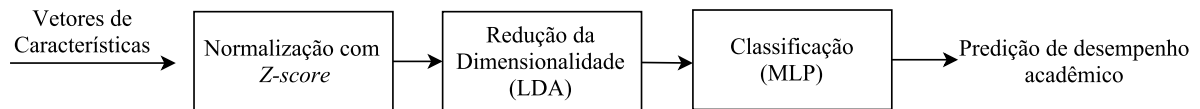


Figura 20 – Fluxo do método proposto para o experimento 04.

Novamente, os vetores de características pertencem a um espaço \mathbb{R}^k , em que k é igual a 17. Cada dimensão armazena uma informação das interações do estudante no AVA, conforme mostrado na tabela 3. Para obter a dimensão ótima dos novos vetores de características, foi realizado processo similar ao experimento 03. Para cada tamanho dos vetores, foram realizados os processos de classificação com a rede neural MLP. Ao utilizar apenas 1 atributo, a MLP classificou corretamente 81% dos estudantes, e com 2 atributos, 82%.

Neste experimento, a rede neural MLP, segue os mesmos parâmetros dos experimentos anteriores, porém com apenas três camadas, camada de entrada, uma camada oculta com 2 neurônios, e uma camada de saída com 3 neurônios. Utilizou-se neste experimento o valor 4 como semente para geração de números aleatórios para os pesos das camadas.

A figura 21 mostra a matriz de confusão da MLP quando utilizado os vetores de características com 2 dimensões. Com menos dados para serem processados, além de diminuir o tempo de treinamento da MLP, obteve-se uma acurácia de 82%. O total de estudantes classificados corretamente foram 246 num total de 300. A capacidade de generalizar da MLP também foi melhorada, permitindo que estudantes das classes A, B e C fossem classificados corretamente em 58,33%, 57,97%, 90,87% do casos, respectivamente. Este resultado superior ao experimento 03 pode ser atribuído ao fato da LDA reduzir as dimensões do vetor de forma supervisionada, ou seja, levando em consideração a classe que cada amostra pertence.

Outras análises dos resultados deste experimento podem feitas a partir da tabela 11.

Classe	Precisão	Revocação	<i>F-measure</i>
A	0,583	0,583	0,583
B	0,645	0,580	0,611
C	0,881	0,909	0,894
Média ponderada	0,815	0,820	0,817

Tabela 11 – Medidas de Precisão, Revocação e *F-measure* da MLP no experimento 04.

	A	B	C
A	58.33% (7)	25.00% (3)	16.67% (2)
B	5.80% (4)	57.97% (40)	36.23% (25)
C	0.46% (1)	8.68% (19)	90.87% (199)

Figura 21 – Matriz de confusão da MLP no experimento 04.

6.4.5 Experimento 05: *z-score* + PCA + LDA

O experimento 05 normaliza os dados com *z-score* e combina as duas técnicas de redução da dimensionalidade utilizadas nos experimentos anteriores. O fluxo é descrito na figura 22.

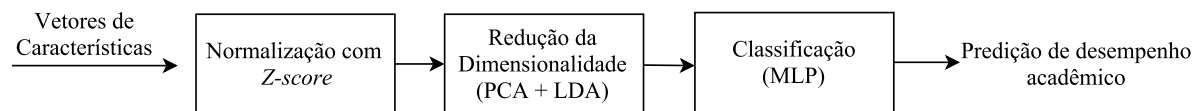


Figura 22 – Fluxo do método proposto para o experimento 05.

A rede neural MLP utiliza as mesmas configurações dos experimentos anteriores, variando a quantidade de camadas de neurônios. Utilizou-se uma camada de entrada, duas ocultas, com 2 e 10 neurônios, respectivamente, e uma camada de saída com 3 neurônios.

Após a normalização dos valores dos atributos, os vetores de características com dimensão igual a 17 são reduzidos com o PCA e utilizado os 12 componentes principais como entrada para a LDA, que reduz estes para 2 dimensões. Após estes procedimentos de pré-processamento dos dados, a rede neural MLP obteve acurácia de 82%, classificando corretamente 246 dos 300 estudantes. Este foi o experimento que melhor identificou estudantes com piores níveis de desempenho, aqueles contidos na classe A, acertando 66,67% destes, o que em um ambiente real de ensino-aprendizagem são os que mais precisam ser identificados, monitorados e orientados de forma que obtenham melhores resultados acadêmicos. A figura 23 mostra a matriz de confusão da MLP neste experimento.

Os valores de precisão, revocação e *F-measure* da MLP por classe neste experi-

	A	B	C
A	66.67% (8)	33.33% (4)	0
B	4.35% (3)	49.28% (34)	46.38% (32)
C	0.46% (1)	6.39% (14)	93.15% (204)

Figura 23 – Matriz de confusão da MLP no experimento 05.

mento são informados na tabela 12.

Classe	Precisão	Revocação	<i>F-measure</i>
A	0,667	0,667	0,667
B	0,654	0,493	0,562
C	0,864	0,932	0,897
Média ponderada	0,808	0,820	0,811

Tabela 12 – Medidas de Precisão, Revocação e *F-measure* da MLP no experimento 05.

6.5 Análise dos Resultados

Os resultados mostram que os métodos de pré-processamentos abordados podem melhorar a acurácia da rede neural MLP em torno de 9%, conforme mostrado na figura 24. Este aumento é importante quando visto que 66% estudantes com desempenho inferior, aqueles que mais precisam de acompanhamento, foram identificados após a transformação dos dados 25(a). A figura 25 mostra a evolução na identificação de estudantes com níveis de desempenho com a utilização dos métodos abordados.

Os três métodos propostos obtiveram acurácias com diferença em torno de 1%. Os métodos que utilizam PCA exigem mais camadas ocultas de neurônios e mais neurônios em cada camada, o que impacta diretamente no tempo de processamento da MLP, porém neste trabalho, foi a técnica de redução dos dados que melhor contribuiu para que a MLP identificasse estudantes com nível intermediário, conforme exposto na figura 25(b).

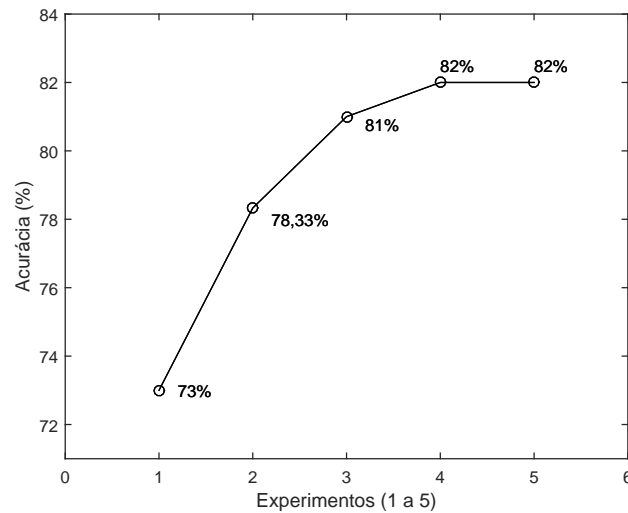
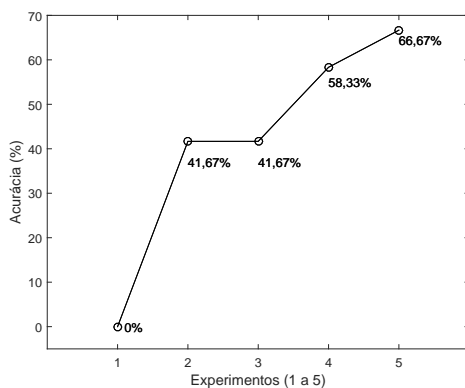
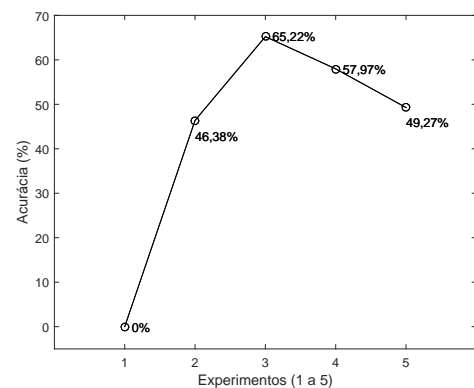


Figura 24 – Acurácias da MLP dos Experimentos 01 a 05.

A abordagem com o LDA é essencialmente importante quando necessário a visualização dos dados em um gráfico dispersão. Apesar do PCA também poder reduzir os dados para serem visualizados em 2 ou 3 dimensões, a abordagem de redução dos dados com LDA permitiu os melhores resultados no processo de classificação com MLP com 2 dimensões, enquanto a PCA foram necessários 12 dimensões, impossibilitando a visualização dos dados. Além disto, após a abordagem LDA, a MLP necessitou de menos camadas para obter resultados superiores aqueles obtidos com PCA, utilizando apenas uma camada oculta com 2 neurônios, diminuindo consideravelmente o tempo de processamento. Este estudo não visa uma análise aprofundada do tempo de processamento e outros recursos utilizados pela MLP para classificação.



(a) Estudantes com nível inferior



(b) Estudantes com nível intermediário

Figura 25 – Taxas de identificação dos estudantes com baixos níveis de desempenho nos experimentos.

Diante do exposto, concluímos que neste contexto, a LDA pode ser melhor abordagem que o PCA para redução da dimensionalidade. Não apenas pela acurácia preditiva obtida pelo modelo que utiliza a LDA, mas, também pela quantidade de recursos computacionais utilizados para classificação com a MLP após o pré-processamento dos dados. Esta análise é útil, visto que em um ambiente de ensino real, centenas ou milhares de estudantes deverão seus desempenhos monitorados.

6.6 Comparação dos Métodos Propostos com outros Métodos de Classificação

Para comparar o resultados obtidos com os métodos propostos neste trabalho com os resultados obtidos em outros trabalhos, foram realizados testes com os métodos de classificação utilizados em alguns trabalhos. Este estudo comparativo se deu como descrito nos parágrafos a seguir:

Foram selecionados alguns trabalhos que utilizaram métodos de classificação com algoritmos contidos na ferramenta *Weka*. Realizou-se o processo de classificação dos estudantes utilizando estes métodos, validação cruzada **K-fold** com **K** igual a 10, e os dados utilizados neste trabalho. O objetivo é verificar se o método proposto classifica melhor os estudantes quanto ao seu desempenho acadêmico em relação aos métodos utilizados nos outros trabalhos.

A tabela 13 lista alguns trabalhos que abordaram métodos de classificação para predição de desempenho acadêmico de estudantes, e os respectivos algoritmos utilizados.

Esta análise envolve os resultados obtidos nos experimentos 03, 04 e 05.

Tabela 13 – Acurácia do método proposto comparado aos métodos utilizados em outros trabalhos.

Método de Classificação	Trabalho que utiliza	Acurácia
Naive Bayes	(GOTARDO <i>et al.</i> , 2013)	62,33%
RBF	(SANTANA <i>et al.</i> , 2014)	73,33%
J48	(GOTARDO <i>et al.</i> , 2013; PEQUENO <i>et al.</i> , 2014)	74,66%
OneR	(PEQUENO <i>et al.</i> , 2014)	76,66%
MLP	(MANHÃES <i>et al.</i> , 2011; GOTARDO <i>et al.</i> , 2012; DETONI <i>et al.</i> , 2014; SANTANA <i>et al.</i> , 2014)	78,33%
RandomForest	(MANHÃES <i>et al.</i> , 2011; GOTARDO <i>et al.</i> , 2012; DETONI <i>et al.</i> , 2014; SANTANA <i>et al.</i> , 2014)	79%
PCA + MLP	Método Proposto	81%
LDA + MLP	Método Proposto	82%
PCA + LDA + MLP	Método Proposto	82%

A literatura informa que modelos preditivos que inferem sobre o desempenho de estudantes e que tem acurácia em média de 79% são considerados bons modelos (??). Os

melhores modelos obtêm acurácia em torno de 94%, mas estes utilizam grandes volumes de dados (uma amostra superior a 15.000 instâncias), este trabalho utilizou base de dados pequena, da qual foram extraídas informações de 300 estudantes. Desta forma, os métodos propostos neste trabalho, PCA + MLP, LDA + MLP e PCA + LDA + MLP obtêm acurácias, 81%, 82% e 82%, respectivamente, e estas são melhores que as acurácias obtidas por diversos métodos utilizados em outros trabalhos, podem ser considerados bons métodos para predição do desempenho de estudantes em AVA.

Concluimos que os métodos propostos tornam-se válidos para predição de desempenho de estudantes em ambientes virtuais de aprendizagem, não apenas pela taxa geral de acertos na classificação, mas pela identificação de estudantes pertencentes a classes com poucas amostras.

7 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho reafirma que as técnicas de Mineração de Dados, quando aplicada corretamente, podem trazer informações úteis para o acompanhamento de estudantes e melhorias dos processos de ensino-aprendizagem.

Uma abordagem baseada na redução da dimensionalidade com Análise dos Componentes Principais e Análise Discriminante Linear, e na combinação de ambas, objetivou a remoção de informações redundantes nos vetores de características, de forma que o vetor reduzido contenha a parte relevante das informações. Esta etapa de transformação e preparação dos dados dos estudantes para o processo de classificação com redes neurais de tipo *Multilayer Perceptron*, se mostrou um método eficiente para predição do desempenho acadêmico a partir de informações sobre o uso dos Ambientes Virtuais de Aprendizagem.

Os métodos propostos foram capazes de identificar estudantes com níveis mais baixos de desempenhos (inferior e intermediário), apesar da pouca quantidade de amostras. Em um conjunto formado por 300 estudantes de 13 turmas, apenas 12 tem nível inferior, ou seja, apenas 4% do total de estudantes. Esta pouca quantidade é justificada pelo fato de utilizarmos neste trabalho apenas informações de estudantes que tenham notas finais na disciplina. Eventuais casos de desistências não foram levados em conta. A normalização dos dados aumentou consideravelmente a capacidade do algoritmo de classificação na identificação dos estudantes com níveis inferiores, bem como os intermediários, classificando corretamente 41,67% e 46,38%, respectivamente, enquanto antes da normalização nenhum estudante destes níveis foi identificado. As taxas de acertos aumentaram para estudantes com níveis mais baixos quando os dados foram reduzidos com a abordagem PCA e LDA, e o melhor resultado foi obtido com a combinação de ambas as técnicas, chegando a identificar 66,67% dos estudantes com níveis mais baixo de desempenho. A identificação e acompanhamento destes estudantes permitirá que estes melhorem seus desempenhos, aumentando suas chances de aprovação no curso.

Este trabalho atendeu aos seguintes objetivos propostos: Foram levantadas informações que representam um estudante no AVA e estas tem correlação positiva com o resultado final na disciplina. Utilizou-se de técnicas de pré-processamento de dados e mineração de dados para criar modelos preditivos capazes de classificar, com boas acurácias, estudantes em três níveis de desempenho acadêmico. Avaliou-se a acurácia de diversos modelos preditivos a fim de informar quais deles obtêm melhores resultados e podem

contribuir com informações úteis para melhorias no processo de ensino-aprendizagem.

Para trabalhos futuros, existem os seguintes desafios: Investigar novas variáveis que informem sobre ações realizadas pelos estudantes em AVA. Combinar as técnicas de redução dos dados utilizados neste trabalho com outros métodos de classificação listado na tabela 13. Estudar a capacidade de generalização dos modelos quando aplicados a estudantes de diferentes cursos. Adaptar os modelos para que estes façam inferências em tempo real. Integrar os métodos propostos a ambientes virtuais de aprendizagem através do desenvolvimento de ferramentas que permitam a educadores e administradores de curso informar novos dados e analisar os resultados de forma fácil.

REFERÊNCIAS

- ARRUABARRENA, R.; PÉREZ, T. A.; LÓPEZ-CUADRADO, J.; GUTIÉRREZ, J.; VADILLO, J. On evaluating adaptive systems for education. In: SPRINGER. **International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems**. [S.l.], 2002. p. 363–367.
- AYERS, E.; NUGENT, R.; DEAN, N. A comparison of student skill knowledge estimates. **International Working Group on Educational Data Mining**, ERIC, 2009.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de dados educacionais: Oportunidades para o brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 02, p. 03, 2011.
- BAKER, R. S.; YACEF, K. The state of educational data mining in 2009: A review and future visions. **JEDM-Journal of Educational Data Mining**, v. 1, n. 1, p. 3–17, 2009.
- BARCELLOS, C. D.; MUSA, D. L.; BRANDÃO, A. L.; WARPECHOWSKI, M. Sistema de recomendação acadêmico para apoio a aprendizagem. **RENOTE**, v. 5, n. 2, 2007.
- BRAGA, A. d. P.; CARVALHO, A.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: Livros Técnicos e Científicos, 2000.
- BRESFELEAN, V. P.; BRESFELEAN, M.; GHISOIU, N.; COMES, C.-A. Determining students' academic failure profile founded on data mining methods. In: IEEE. **Information Technology Interfaces, 2008. ITI 2008. 30th International Conference on**. [S.l.], 2008. p. 317–322.
- CALVO-FLORES, M. D.; GALINDO, E. G.; JIMÉNEZ, M. P.; PINEIRO, O. P. Predicting students' marks from moodle logs using neural network models. **Current Developments in Technology-Assisted Education**, Citeseer, v. 1, p. 586–590, 2006.
- CONIJN, R.; SNIJDERS, C.; KLEINGELD, A.; MATZAT, U. Predicting student performance from lms data: A comparison of 17 blended courses using moodle lms. **IEEE Transactions on Learning Technologies**, IEEE, 2016.
- COSTA, E.; BAKER, R. S.; AMORIM, L.; MAGALHÃES, J.; MARINHO, T. Mineração de dados educacionais: Conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1–29, 2013.
- DEKKER, G.; PECHENIZKIY, M.; VLEESHOUWERS, J. Predicting students drop out: A case study. In: **Educational Data Mining 2009**. [S.l.: s.n.], 2009.
- DETONI, D.; MATSUMURA, R. A.; CECHINEL, C. Modelagem e predição da situação acadêmica de alunos de educação a distância. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. [S.l.: s.n.], 2014. v. 3, n. 1, p. 41.
- DOUGIAMAS, M.; TAYLOR, P. Moodle: Using learning communities to create an open source course management system. 2003.
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. 2nd. Edition. New York, 2001.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. [S.l.]: John Wiley & Sons, 2012.

FAUSETT, L.; ELWASIF, W. Predicting performance from test scores using backpropagation and counterpropagation. In: IEEE. **Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on**. [S.l.], 1994. v. 5, p. 3398–3402.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *et al.* Knowledge discovery and data mining: towards a unifying framework. In: **KDD**. [S.l.: s.n.], 1996. v. 96, p. 82–88.

FENG, M.; BECK, J. E.; HEFFERNAN, N. T. Using learning decomposition and bootstrapping with randomization to compare the impact of different educational interventions on learning. **International Working Group on Educational Data Mining**, ERIC, 2009.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. [S.l.]: Academic press, 2013.

GALAFASSI, F. P.; GLUZ, J. C.; GALAFASSI, C. Análise crítica das pesquisas recentes sobre as tecnologias de objetos de aprendizagem e ambientes virtuais de aprendizagem. **Revista Brasileira de Informática na Educação**, v. 21, n. 03, p. 100, 2014.

GARCÍA, E.; ROMERO, C.; VENTURA, S.; CASTRO, C. D. A collaborative educational association rule mining tool. **The Internet and Higher Education**, Elsevier, v. 14, n. 2, p. 77–88, 2011.

GARCÍA, S.; LUENGO, J.; HERRERA, F. **Data preprocessing in data mining**. [S.l.]: Springer, 2015.

GEDEON, T.; TURNER, S. Explaining student grades predicted by a neural network. In: IEEE. **Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference on**. [S.l.], 1993. v. 1, p. 609–612.

GOTARDO, R.; CEREDA, P. R. M.; JUNIOR, E. R. H. Predição do desempenho do aluno usando sistemas de recomendação e acoplamento de classificadores. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2013. v. 24, n. 1, p. 657.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R. V. Previsão de desempenho de estudantes em cursos ead utilizando mineração de dados: uma estratégia baseada em séries temporais. In: **anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2012. v. 23, n. 1.

GOTTARDO, E.; KAESTNER, C. A. A.; NORONHA, R. V. Estimativa de desempenho acadêmico de estudantes: Análise da aplicação de técnicas de mineração de dados em cursos a distância. **Revista Brasileira de Informática na Educação**, v. 22, n. 01, p. 45, 2014.

- HAN, J. **Micheline Kamber and Simon Fraser University “Data Mining Concepts and Techniques” Morgan Kaufmann Publishers.** [S.l.]: USA, 2001.
- HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques.** [S.l.]: Elsevier, 2011.
- HAYKIN, S. **Neural networks a comprehensive introduction.** [S.l.]: Prentice Hall, New Jersey, 1999.
- HAYKIN, S. S.; HAYKIN, S. S.; HAYKIN, S. S.; HAYKIN, S. S. **Neural networks and learning machines.** [S.l.]: Pearson Upper Saddle River, NJ, USA:, 2009. v. 3.
- HERSHKOVITZ, A.; NACHMIAS, R. Developing a log-based motivation measuring tool. In: **Educational Data Mining 2008.** [S.l.: s.n.], 2008.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 22, n. 1, p. 4–37, 2000.
- JIawei, H.; KAMBER, M. **Data mining: concepts and techniques. San Francisco, CA, itd: Morgan Kaufmann**, v. 5, 2001.
- LEUNG, H.; HAYKIN, S. The complex backpropagation algorithm. **IEEE Transactions on signal processing**, IEEE, v. 39, n. 9, p. 2101–2104, 1991.
- MANHÃES, L. M. B. **PREDIÇÃO DO DESEMPENHO ACADÊMICO DE GRADUANDOS UTILIZANDO MINERAÇÃO DE DADOS EDUCACIONAIS.** Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2015.
- MANHÃES, L. M. B.; CRUZ, S. M. S. da; COSTA, R. J. M.; ZAVALETA, J.; ZIMBRÃO, G. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: **Anais do Simpósio Brasileiro de Informática na Educação.** [S.l.: s.n.], 2011. v. 1, n. 1.
- MARTÍNEZ, A. M.; KAK, A. C. Pca versus lda. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 23, n. 2, p. 228–233, 2001.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MERCERON, A.; YACEF, K. Educational data mining: a case study. In: **AIED.** [S.l.: s.n.], 2005. p. 467–474.
- MERSCHMANN, L. de C. **Classificação probabilística baseada em análise de padrões.** Tese (Doutorado) — PhD Thesis, UFF-Universidade Federal Fluminense, Brazil, 2007, [www. ic. uff. br/plastino/TeseLuiz. pdf](http://www.ic.uff.br/plastino/TeseLuiz.pdf), 2007.
- OLADOKUN, V.; ADEBANJO, A.; CHARLES-OWABA, O. Predicting students’ academic performance using artificial neural network: A case study of an engineering course. **The Pacific Journal of Science and Technology**, v. 9, n. 1, p. 72–79, 2008.

- PEQUENO, H.; ÁVILA, R. L. de; CRUZ, E.; ALEXANDRE, M.; LIMA, E. T. de; CASTRO, M. F. de. Uma análise de interação em fóruns de ead. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2014. v. 25, n. 1, p. 1203.
- PRIDY, K. L.; KELLER, P. E. **Artificial neural networks: an introduction**. [S.l.]: SPIE Press, 2005. v. 68.
- RIGO, S. J.; CAZELLA, S. C.; CAMBRUZZI, W. Minerando dados educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. In: **Anais do Workshop de Desafios da Computação Aplicada à Educação**. [S.l.: s.n.], 2012. p. 168–177.
- RODRIGUES, R. L.; MEDEIROS, F. P. de; GOMES, A. S. Modelo de regressão linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2013. v. 24, n. 1, p. 607.
- ROMERO, C.; VENTURA, S. Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, Ieee, v. 40, n. 6, p. 601–618, 2010.
- ROMERO, C.; VENTURA, S. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 3, n. 1, p. 12–27, 2013.
- ROMERO, C.; VENTURA, S.; GARCÍA, E. Data mining in course management systems: Moodle case study and tutorial. **Computers & Education**, Elsevier, v. 51, n. 1, p. 368–384, 2008.
- ROMERO, C.; VENTURA, S.; PECHENIZKIY, M.; BAKER, R. S. **Handbook of educational data mining**. [S.l.]: CRC Press, 2010.
- ROMERO, C.; ZAFRA, A.; LUNA, J. M.; VENTURA, S. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. **Expert Systems**, Wiley Online Library, v. 30, n. 2, p. 162–172, 2013.
- SANTANA, L. C.; MACIEL, A. M.; RODRIGUES, R. L. Avaliação do perfil de uso no ambiente moodle utilizando técnicas de mineração de dados. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2014. v. 25, n. 1, p. 269.
- SANTOS, H.; CAMARGO, F.; CAMARGO, S. Minerando dados de ambientes virtuais de aprendizagem para predição de desempenho de estudantes. **Conferencias LACLO**, v. 3, n. 1, 2012.
- THEODORIDIS, S.; KOUTROUMBAS, K. Pattern recognition, academic press. **New York**, 1999.
- THEODORIDIS, S.; KOUTROUMBAS, K. Clustering: basic concepts. **Pattern Recognition**, p. 483–516, 2006.
- WANG, T.; MITROVIC, A. Using neural networks to predict student's performance. In: IEEE. **Computers in Education, 2002. Proceedings. International Conference on**. [S.l.], 2002. p. 969–973.

WEBB, A. Statistical pattern recognition. 2002. **John Willey & Sons**, v. 2.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2005.

YANG, Y.; PEDERSEN, J. O. A comparative study on feature selection in text categorization. In: **ICML**. [S.l.: s.n.], 1997. v. 97, p. 412–420.