

**Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Pós-Graduação em Ciência da Computação**

**Abordagem Avaliativa Multidimensional
para Previsão da Evasão do Discente
em Cursos *Online***

Alana Marques de Moraes

**CAMPINA GRANDE - PB
ABRIL – 2018**

Alana Marques de Moraes

Abordagem Avaliativa Multidimensional para Previsão da Evasão do Discente em Cursos *Online*

Tese apresentada à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande – Campus I, como parte dos requisitos regulamentar para obtenção do grau de Doutor em Ciência da Computação.

Orientadores:

Prof^a. Dr^a. Joseana Macêdo Fachine Régis de Araújo
Prof. Dr. Evandro de Barros Costa

**CAMPINA GRANDE - PB
ABRIL – 2018**

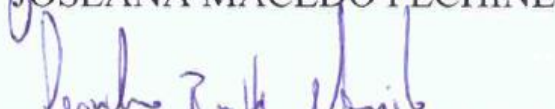
Alana Marques de Moraes

**Abordagem Avaliativa Multidimensional
para Previsão da Evasão do Discente em
Cursos *Online***

Tese apresentada ao Programa de Pós-Graduação
em Ciência da Computação da Universidade Federal
de Campina Grande.

Tese aprovada em: 20/04/2018


JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO


LEANDRO BALBY MARINHO


NAZARENO FERREIRA DE ANDRADE

EDSON PINHEIRO PIMENTEL

PATRICK HENRIQUE DA SILVA BRITO

Agradecimentos

Inicialmente, gostaria de agradecer a Deus e Nossa Senhora, por me guiarem pelos caminhos da determinação, por colocarem em meu caminho pessoas maravilhosas, pela minha família, pelas oportunidades que apareceram – e pelas que virão - na minha carreira profissional e por me permitirem chegar nesta etapa de minha vida.

Faço um profundo agradecimento aos meus pais Alberto Marques de Souza e Maria Nicolau de Moraes Souza pelos conselhos, pela força, pelos incentivos, pelo apoio financeiro e, principalmente, por todo amor recebido ao longo da minha vida. Às minhas irmãs Aline Marques de Moraes e Sarah Marques de Moraes pelo companheirismo, amor, incentivo e paciência ao longo do doutorado. Por fim, eu gostaria de agradecer a Wandemberg Mesquita pela sua paciência, compreensão, amor, apoio e por me ouvir quando eu precisei durante esses últimos anos.

Além disto, este trabalho não teria atingido o nível de qualidade atual se não fossem as observações enriquecedoras e contundentes dos meus queridos e competentes orientadores Dr^a. Joseana de Araújo e Dr. Evandro Costa. Agradeço ainda aos meus colegas do Laboratório de Inteligência Artificial (LIA), em especial a minha amiga Danielle Medeiros, por toda ajuda e apoio.

Por último, o agradecimento à CAPES pelo financiamento do doutorado e por possibilitar o doutorado sanduíche na Alemanha em 2015.

Resumo

A evasão discente foi apontada, por grande parte das instituições brasileiras de Educação a Distância (EaD), como um dos maiores obstáculos enfrentados pelos cursos *online* da atualidade. Reflexo disto, foram as taxas de evasão elevadas, entre 25 e 50% em todos os tipos de curso *online*, em relação à educação presencial em 2015 no Brasil. Além disto, a problemática não se restringe ao contexto brasileiro, pois diversas organizações internacionais relataram índices tão alarmantes quando os nacionais. É possível afirmar, de acordo com a literatura, que as estratégias adotadas na minimização dos índices de desistência do discente muitas vezes se limitam à avaliação pontual de um recurso pedagógico (fórum, bate-papo, *etc*). Há ainda uma carência de estudos que discutam sobre as interações educacionais do aluno evadido e considerem as incertezas avaliativas envolvidas no processo. Diversas metodologias podem ser utilizadas pelo docente para lidar com a subjetividade e incerteza associadas ao processo avaliativo, tais como: lógica *fuzzy*, redes bayesianas, *etc*. O presente estudo destacou a lógica *fuzzy* (ou difusa), pois esta técnica lida com a ambiguidade encontrada na avaliação presente na EaD. Esta modelagem permite que estados indeterminados possam ser tratados por dispositivos de controle e conceitos não quantificáveis possam ser avaliados pelo especialista. Assim, o presente estudo propôs uma abordagem avaliativa (*FuzzySD*) voltada ao docente para monitorar as turmas virtuais na predição e combate à evasão discente em cursos *online*. Para tanto, a pesquisa se subsidiou nas informações relacionadas às ações dos alunos durante o processo de ensino-aprendizagem. A análise realizada envolve fatores comportamentais do estudante com base em múltiplos critérios. Estes critérios comportamentais foram fundamentais na descrição das três métricas principais ao entendimento da evasão do aluno no contexto estudado, a saber: Autorregulação, Interação e Motivação do aluno. A avaliação com a *FuzzySD* foi realizada por meio da análise de precisão em recortes da amostra com base em dados de cursos *online* do IFPB Virtual. A precisão da abordagem, analisando as amostras teste, proporciona valores entre 65% e 95% de acertos. Por fim, a partir dos resultados, foram discutidos os desafios da aplicação de um módulo avaliativo multidimensional diante do contexto de cursos *online* brasileiros.

Palavras-chave: Inteligência Artificial, Avaliação multidimensional, Educação a Distância - Evasão, Mineração de Texto, Lógica *Fuzzy*, *FuzzySD*.

Abstract

The student's dropout in online courses was pointed out as one of the biggest issue in the area according to the educational organizations. The main result was the high dropout rates in these courses compared to face-to-face courses in 2015 in Brazil (between 25% and 50%). Moreover, this challenge belongs to the virtual learning worldwide. In this sense, the literature review showed that several approaches are adopted to minimize the student's dropout rates. However, they limited their focus on a learning unique resource (forum, chat, upload, download, submissions). Regarding these approaches, the current research detected an insufficiency of studies about the educational interactions of the dropout student and uncertainties in the assessment. Many methodologies are used by teachers to manage a subjectivity and uncertainty associated to the evaluation process, such as: fuzzy logic, Bayesian networks, Markov chains, etc. This thesis emphasized the fuzzy logic, because this technique solves the ambiguity found in the learning evaluation by the expert. We present an assessment framework (FuzzySD) to track the virtual classroom and to support the teacher to make a prediction of students' dropout in online courses. In this sense, we analyzed the information related to the students' actions during the teaching-learning process in online courses. It can be denominated as multidimensional analysis because we assessed multiple variables related to the student's behavior. Such behavioral variables were fundamental in the definition of three main metrics to the understanding of student avoidance, they are: Self-regulation, Interaction and Motivation. It is important to consider a presence of fuzzy logic to deal with the input metrics of the FuzzySD framework. The evaluation of FuzzySD was performed by means of precision analysis on sample cut-outs based on dataset from courses of "IFPB Virtual". The accuracy of the approach is analyzed as results of its own indicators between 65% and 95% of hits. Finally, the results discussed about the challenges of applying a multidimensional evaluative module to the context of Brazilian online courses.

Keywords: *Artificial Intelligence, fuzzy system, student multidimensional assessment, distance learning - dropout, online courses, FuzzySD.*

Lista de Figuras

Figura 2.1. Etapas da Revisão Sistemática.....	24
Figura 2.2. Artigos coletados na Revisão Sistemática.....	26
Figura 2.3. Linha de tempo da interação do aluno em um curso online.....	28
Figura 2.4. Arcabouço de predição da evasão.....	36
Figura 2.5. Visão geral das palavras chave mais comentadas em trabalhos com a abordagem multidimensional.....	48
Figura 3.1. Etapas da lógica <i>fuzzy</i>	52
Figura 3.2. Exemplo de árvore de decisão.....	54
Figura 3.3. Interconexões entre usuários de sistemas colaborativos.....	62
Figura 3.4. Teoria da autodeterminação, focada na Motivação.....	65
Figura 4.1. Visão Geral da Abordagem Multidimensional Avaliativo.....	72
Figura 4.2. Análise de correlação entre as variáveis discente.....	75
Figura 4.3. Abordagem <i>FuzzySD</i>	78
Figura 4.4. Gráfico de joelho da métrica Autorregulação.....	81
Figura 4.5. Gráfico de joelho da métrica Interação.....	81
Figura 4.6. Gráfico de Silhueta da métrica Autorregulação.....	82
Figura 4.7. Gráfico de Silhueta da métrica Interação.....	82
Figura 4.8. Agrupamentos encontrados a partir da métrica Autorregulação.....	83
Figura 4.9. Agrupamentos encontrados a partir da métrica Interação.....	84
Figura 4.10. Funções de pertinência da Autorregulação do aluno.....	91
Figura 4.11. Funções de pertinência da Interação do aluno.....	93
Figura 4.12. Funções de pertinência da Motivação do aluno.....	94
Figura 4.13. Função de pertinência Trapezoidal da Evasão do discente.....	98
Figura 4.14. Diagrama de pacotes da <i>FuzzySD</i>	105
Figura 5.1. Análise temporal na avaliação da <i>FuzzySD</i>	108
Figura 5.2. Funcionamento da lógica <i>fuzzy</i> na <i>FuzzySD</i>	113

Lista de Tabelas

Tabela 5.1. Cenários selecionados para a avaliação da <i>FuzzySD</i>	110
Tabela 5.2. Avaliação da <i>FuzzySD</i> nos cenários selecionados.....	115
Tabela 5.3. Avaliação das abordagens avaliativas em relação à <i>FuzzySD</i>	119

Lista de Quadros

Quadro 2.1. Principais Informações do Protocolo de Revisão Sistemática.	25
Quadro 2.2. Tipos de cursos e suas características.	25
Quadro 2.3. Trabalhos relacionados à predição da evasão.	38
Quadro 2.4. Fatores relacionados à evasão dos discentes de EaD.	41
Quadro 2.5. Quadro resumo dos métodos <i>fuzzy</i>	46
Quadro 2.6. Quadro resumo das avaliações multidimensionais.	46
Quadro 3.1. Fases e áreas da aprendizagem autorregulada.	61
Quadro 4.1. Variáveis dos discentes analisadas.	73
Quadro 4.2. Variáveis das discentes associadas às métricas de entrada.	77
Quadro 4.3. Exemplos de mensagens classificadas pelo professor.	87
Quadro 4.4. Valores das Variáveis Linguísticas.	89
Quadro 4.5. Funções de pertinência e variáveis linguísticas da Autorregulação.	91
Quadro 4.6. Funções de pertinência e variáveis linguísticas da Interação.	92
Quadro 4.7. Funções de pertinência e variáveis linguísticas da Motivação.	94
Quadro 4.8. Exemplo da Base de Regras obtidas no modelo.	96
Quadro 4.9. Funções de pertinência e variáveis linguísticas da métrica de saída Evasão.	99
Quadro 4.10. Matriz de confusão.	101
Quadro 4.11. Medidas calculadas a partir da Matriz de Confusão.	103
Quadro 5.1. Matrizes de confusão das amostras teste no IFPB Virtual.	114
Quadro 5.2. Quadro resumo dos modelos <i>fuzzy</i>	121

Lista de Abreviaturas

AVEA – Ambientes Virtuais de Ensino e Aprendizagem

CoG - Centro de Gravidade

EaD – Educação a Distância

FuzzySD – *Fuzzy Student Dropout*

IA – Inteligência Artificial

IFPB – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba

LDB - Lei de Diretrizes e Bases da Educação Nacional

MEC – Ministério da Educação

MOOC – *Massive Open Online Course*

RS – Revisão Sistemática

UAB - Universidade Aberta do Brasil

Sumário

1. Introdução.....	13
1.1 Motivação.....	16
1.2 Objetivos.....	19
1.3 Contribuição.....	20
1.4 Estrutura da Proposta.....	22
2. Estado da Arte.....	23
2.1 Evasão Discente.....	26
2.2 Avaliação do Discente no Combate à Evasão.....	29
2.2.1 Predição da Evasão.....	34
2.2.2 Métricas comportamentais discentes relacionadas à evasão.....	40
2.3 Avaliação Multidimensional.....	41
2.4 Considerações Finais.....	48
3. Fundamentação Teórica.....	50
3.1 Lógica <i>Fuzzy</i>	50
3.2 Árvores de Decisão.....	53
3.3 Mineração de Texto.....	55
3.4 Métricas Pedagógicas.....	57
3.4.1 Autorregulação.....	57
3.4.2 Interação.....	61
3.4.3 Motivação.....	64
3.5 Considerações Finais.....	66
4. Abordagem <i>FuzzySD</i>	67
4.1 Seleção do Contexto.....	67
4.2 Ferramentas utilizadas na Pesquisa.....	68
4.3 Proposição da Abordagem.....	69
4.3.1 Métricas de Entrada.....	70
4.3.1.1 Arquitetura Multidimensional da <i>FuzzySD</i>	71
4.3.1.2 Análise de Correlação das Entradas e a Evasão.....	73
4.3.1.3 <i>FuzzySD</i> no Contexto dos Cursos Online da UAB.....	75
4.4 Planejamento das Etapas da <i>FuzzySD</i>	78
4.4.1 Identificação das Variáveis <i>Fuzzy</i> Associadas à Abordagem.....	79
4.4.1.1 Mineração de Texto da Métrica Motivação.....	85
4.4.1.2 Cálculo do Tf-Idf.....	88
4.4.1.3 Dicionários Gerados.....	88

4.4.2 Etapa de Fuzzificação da <i>FuzzySD</i>	89
4.4.2.1 Definição das Funções de Pertinência Fuzzy	89
4.4.3 Etapa de Inferência	95
4.4.4 Etapa de Defuzzificação	97
4.5 Avaliação da <i>FuzzySD</i>	99
4.6 Arquitetura da <i>FuzzySD</i>	104
4.7 Considerações Finais	106
5. Apresentação e Análise dos Resultados da <i>FuzzySD</i>	107
5.1 Avaliação da Defuzzificação da Saída.....	107
5.2 Cenário do Estudo de Caso.....	109
5.3 Avaliação da Defuzzificação da Saída: Análise Temporal	111
5.3.1 Exemplo de Análise <i>Fuzzy</i>	111
5.3.2 Precisão dos Cortes Temporais	113
5.3.3. Analisando as Hipóteses.....	117
5.4 Outras Abordagens Avaliativas	117
5.4.1 Abordagens Avaliativas <i>Fuzzy</i>	120
5.5 Considerações Finais	122
6. Considerações Finais	123
6.1 Ameaças à Validade.....	124
6.2 Sugestões para Pesquisas Futuras	125
6.3 Publicações Relacionadas ao Doutorado	126
Anexo A- Carta de Anuência do Reitor do IFPB.....	129
Apêndice A – Dicionários (métrica Motivação) e Lista de <i>Stopwords</i>	130
A) Dicionário para Motivação Intrínseca	130
B) Dicionário para Motivação Extrínseca	130
C) Dicionário para Desmotivação.....	130
D) StopWords	131
Apêndice B – Exemplos de Árvores de Decisão para os Cursos de Letras e Segurança do Trabalho	132
Apêndice C – Arquivo entrada gerado no pré-processamento	137
Apêndice D – Entrada para a <i>FuzzySD</i>	138
Referências Bibliográficas.....	139

1. Introdução

A demanda por cursos *online* vem aumentando consideravelmente e representa uma tendência para a educação, além de se configurar como uma oportunidade de negócios. Neste sentido, a Educação a Distância (EaD) estimula a promoção de novos caminhos de aprendizagem para grandes contingentes populacionais, não mais de acordo somente com critérios quantitativos, mas com base na qualidade, flexibilidade, liberdade e ética no ensino.

Os primeiros modelos dessa geração de sistemas educacionais se desenvolveram em diversos lugares simultaneamente, com destaque para a experiência da Inglaterra durante a década de 70 (LITTO e FORMIGA, 2009; NUNES, 2009). Atualmente, o *Open University* registra mais de 900 cursos cadastrados, que foram acessados por quase 45 milhões de pessoas entre 2006-2016 (OPEN UNIVERSITY, 2017). Também nos últimos anos, alguns relatórios oficiais destacaram o crescimento e a diversidade dos cursos *online* no Brasil e no mundo (ABED, 2016; ALLEN e SEAMAN, 2016). Na perspectiva brasileira, o Censo EaD.BR destacou a expansão da EaD ao contabilizar mais de 5 milhões de alunos registrados entre 2014 e 2015 nas mais variadas áreas de conhecimento, níveis acadêmicos e tipos de cursos no Brasil (ABED, 2016). Outro índice que relatou este crescimento é o aumento no número de tutores e professores dedicados ao setor, que passou de 28,7 mil profissionais, em 2014, para 48,1 mil educadores, em 2015. Concomitantemente, cerca de 5,8 milhões de estudantes foram matriculados em pelo menos um curso *online* em 2014 nos Estados Unidos, representando um aumento de 3,9% em relação a 2013 (ALLEN e SEAMAN, 2016).

Além da democratização do ensino, a EaD apresenta notáveis vantagens sob o ponto de vista da eficiência, mesmo quando se contabiliza um grande contingente de alunos ou se observa o crescimento vertiginoso da demanda por matrículas. Nesta perspectiva, a expansão do acesso à EaD não deve estar relacionada apenas ao aumento do número de vagas em cursos *online*, mas ao acompanhamento adequado do aprendizado discente e ao

apoio aos processos decisórios do professor durante o processo de ensino.

Inicialmente, é fundamental definir alguns termos chave ao estudo com o intuito de eliminar possíveis dúvidas terminológicas nas discussões apresentadas neste documento. O primeiro ponto abordado é a conceituação da EaD na pesquisa, pois há uma diversidade de plataformas e arquiteturas de sistemas educacionais encontradas na literatura e mercado que se denominam como *online*. Este documento se subsidia no agrupamento apresentado no relatório de Allen e Seaman (2016). Os autores definem cursos *online* como aqueles em que pelo menos 80% do conteúdo educacional é transmitido via Ambiente Virtual de Ensino e Aprendizagem (AVEA). É importante destacar a diversidade de sistemas encontrados na literatura científica e a relevância dos *Massive Open Online Course* (MOOC) e das Universidades Abertas para tal popularização. Entre os sistemas utilizados, pode-se citar *Moodle*, *Coursera*, *eDX*, *Udemy*, *Rooda*, *Udacity*, *Blackboard* e *Amadeus*.

Normalmente, os cursos *online* utilizam alguns recursos didáticos, que são sistematicamente organizados pelo professor de acordo com as funcionalidades dos AVEA adotados. A maioria dos AVEA é composta por uma estrutura baseada em módulos para facilitar o processo de ensino-aprendizagem. Tais módulos contêm atividades e recursos pedagógicos direcionados a um subgrupo temático do curso. Entre os principais recursos pedagógicos, pode-se destacar: fóruns de discussão, bate-papos, envios de arquivos, questionários, *wikis* e objetos de aprendizagem. A diversidade de características destes recursos traz consigo diversas dimensões que podem ser analisadas de acordo com o objetivo da investigação. A multidimensionalidade dos recursos pedagógicos na pesquisa ora descrita é explorada com o intuito de melhor entender as interações dos alunos com todos os módulos de um curso *online*.

Neste sentido, ignorar as especificidades dos AVEA ou planejar inadequadamente os cursos *online* pode agravar problemas pedagógicos e acelerar o insucesso, que são identificados na literatura sob o formato de evasão e reprovação. A evasão é, certamente, um dos problemas mais comuns que aflige as instituições de ensino virtual, pois afeta os resultados dos sistemas educacionais. A busca por suas causas tem sido objetivo de muitas

pesquisas e investigações na literatura e é mais bem discutido no Capítulo 2 da presente tese (DE-LA-FUENTE-VALENTÍN, PARDO e KLOOS, 2013; ROSTAMINEZHAD *et al.*, 2013; TAMHANE *et al.* 2014).

No Brasil, a evasão é apontada por grande parte das instituições como um dos maiores obstáculos enfrentados nos diferentes tipos de cursos *online* (MEC, 2016). Reflexo disto, são as elevadas taxas de evasão na EaD (entre 25 e 50% em todas as modalidades de curso) em relação à educação presencial no Brasil (ABED, 2016). O Censo EaD.BR ainda aponta a falta de tempo para estudar ou participar do curso como o principal motivo para evasão nos cursos *online* do tipo regulamentados e livres em 2015 (ABED, 2016).

A problemática da evasão discente não se limita ao contexto brasileiro, pois diversas instituições internacionais relatam índices tão alarmantes quando os nacionais. Por exemplo, a Universidade Duke ofertou um curso de Bioeletricidade em 2012, por meio do AVEA *Coursera*, e obteve 12.175 estudantes inscritos. No entanto, apenas 7.761 estudantes assistiram a um vídeo, 3.658 interagiram com um questionário, 345 tentaram o exame final e, finalmente, apenas 313 obtiveram o certificado (YANG *et al.*, 2013). Apesar de mais de 10.000 alunos frequentemente se inscreverem para um único curso *online* nesses MOOC, a taxa de abandono geralmente é superior a 70%, e pode até alcançar 90% (BOYER e VEERAMACHANENI, 2015; ONAH, SINCLAIR e BOYATT, 2014; SINHA *et al.*, 2014). Outro exemplo relevante são as instituições asiáticas que relatam taxas de abandono escolar a distância entre 50% a 80% dos inscritos (CHOI *et al.*, 2013).

É sob esse enfoque que o presente documento define evasão discente, como a desistência do aluno de um curso *online*, ou seja, quando o discente não interage com as atividades do curso ou com os outros participantes e, conseqüentemente, não conclui requisitos para a finalização do curso ou de programas de estudo virtuais. Tais alunos são identificados como evadidos quando se encontram sob a seguinte condição: abstenção do discente por um períodoaaa o maior do que um quarto do tempo de execução do curso *online* (HALAWA, GREENE e MITCHELL, 2014).

Neste sentido, a presente pesquisa propôs colaborar com os desafios relacionados à avaliação e ao apoio ao docente na minimização dos índices de

evasão *na* EaD. O contexto selecionado é o da Universidade Aberta do Brasil (UAB) ou cursos *online* que possuam uma arquitetura semelhante. Os aspectos avaliativos relacionados ao combate desta problemática e suas lacunas na literatura são apresentados em detalhes no Capítulo 2 deste documento.

1.1 Motivação

Após a visão geral do contexto de atuação do presente estudo, uma série de motivações é elencada a fim de justificar a escolha deste cenário e da problemática da evasão discente. Sob esta óptica, são identificados três pontos fundamentais ao estudo: (i) índices elevados de evasão discente em cursos *online* no mundo; (ii) prejuízos causados pela evasão para as instituições e (iii) esforços do docente para avaliar e monitorar o desempenho dos alunos na EaD. Tais pontos são discutidos em detalhes na revisão de literatura apresentada no Capítulo 2.

Em relação à problemática da evasão, os pesquisadores almejam identificar e combater as causas da elevada evasão dos estudantes nos AVEA há mais de duas décadas. Uma questão que chama a atenção, segundo a pesquisa apresentada no Anuário Brasileiro Estatístico de Educação Aberta e a Distância (SANCHEZ, 2008), “[..] foi a clara constatação da evasão precoce. A quase totalidade dos alunos que deixaram os cursos o fazem logo no início, o que sugeriu mais atenção das instituições para esse relacionamento inicial”. Tal pesquisa afirmou que os motivos mais citados à evasão são “Financeiro”, “Falta de tempo”, “Não adaptação ao método” e “Achou que o método *online* era mais fácil”, que representam os mesmos dos dias atuais (MAURÍCIO, 2013). Os dois últimos itens da pesquisa da Sanchez (2008) tratam de uma questão de não adaptação à metodologia da EaD, destacando a consequência da frustração diante das perspectivas não atingidas pelo aluno virtual.

A evasão também representa um problema inerente aos sistemas de EaD, fato que alerta as instituições de ensino, pois desperdícios de recursos financeiros, sociais e humanos são gerados por este fato. Os desperdícios financeiros são muito relevantes nas decisões institucionais, principalmente em momentos de crise como são os atuais. Muitas vezes, relatórios financeiros são

analisados para verificar a viabilidade de um curso *online* em decisões relacionadas à expansão ou extinção destes. Por fim, é importante destacar que a sociedade perde muito com os investimentos mal aproveitados, uma vez que os estudantes ocupam as vagas e não finalizam seus cursos. Tais alunos representam profissionais que deixam de atuar na sociedade e influenciar positivamente na economia de sua região e do país.

Outra motivação que merece destaque, de acordo com a revisão sistemática apresentada no Capítulo 2, é o apoio limitado ao docente no monitoramento de suas turmas virtuais. Com o crescimento do número de alunos, gerenciar os processos pedagógicos visando à permanência e ao êxito dos discentes torna-se, cada vez mais, uma tarefa custosa ao docente, pois as ferramentas de monitoramento e avaliação formativa ainda são pouco difundidas nos AVEA. Assim, os dados gerados nas interações entre professores, estudantes e tutores são inúmeros e pouco explorados. Além disto, reuni-los e interpretá-los é uma atividade exaustiva e complexa.

Devido à dificuldade no processamento de grandes volumes de dados educacionais, o professor e o tutor, muitas vezes, se limitam à execução das análises de frequências das interações do aluno no curso *online* (MORAIS, ARAUJO e COSTA, 2014). Tal restrição pode ser atribuída à limitação de ferramentas de apoio à decisão e ao acompanhamento dos discentes nos ambientes de ensino de maneira geral.

Neste cenário, se evidenciam alguns estudos que objetivam minimizar tais índices utilizando técnicas de avaliação (formativa ou somativa) ao longo das interações dos alunos no curso *online* (DE ALMEIDA *et al.*, 2016). Contudo, a maioria das pesquisas identificadas lida com o insucesso do aluno por meio da investigação de métricas relacionadas à reprovação ou às notas, obtidas a partir de uma avaliação unidimensional ao final do curso. Estas atitudes restringem o processo de avaliação global do aluno como ator principal no processo de ensino-aprendizagem, pois os modelos avaliativos apresentados discutem uma percepção restrita da evolução do aluno ao longo do curso *online*. Além disto, a revisão da literatura apresenta uma realidade em que a maioria dos estudos executados foca na avaliação do aluno mediante suas interações com um tipo específico de recurso pedagógico. Este

tratamento pode ser equivocado durante a investigação de questões mais complexas como evasão, reprovação e avaliação do insucesso dos alunos, pelo seu caráter restritivo.

Algumas estratégias computacionais identificadas na literatura tentam minimizar a incerteza, normalmente presente durante a delimitação de conceitos avaliativos (Ruim, Bom, Baixo Desempenho, *etc.*) e na atribuição de valores numéricos a tais categorias por parte do professor. Outro aspecto relacionado à incerteza na avaliação são as análises subjetivas presentes em diversos momentos do processo de ensino-aprendizagem. Neste sentido, uma alternativa para a manipulação e controle do comportamento nebuloso das variáveis de interação envolvidas é a utilização de estratégias que envolvam lógica *fuzzy*, lógica bayesiana, redes neurais *etc.*

A presente pesquisa utiliza a técnica da lógica *fuzzy* para lidar com as variáveis nebulosas. Esta técnica tem o intuito de monitorar múltiplos aspectos educacionais para apoiar decisões sobre um fenômeno. Por isto, há diversos exemplos de sistemas *fuzzy* exitosos em diversas áreas da EaD, tais como: monitoramento (CHRYSAFIADI e VIRVOU, 2012), desempenho de classificação (HUDEC, 2016; RABBANY, TAKAFFOLI e ZAIANE, 2011), e sistemas de decisão (BAI e CHEN, 2008). Nesta abordagem, regras *fuzzy* podem ser obtidas e diretamente formuladas por especialistas na forma de regras linguísticas e o sistema *fuzzy* não sofre degradação do desempenho devido a parâmetros críticos definidos de uma forma não ideal.

Assim, diante da necessidade de desenvolver estudos que contribuam com a qualidade dos serviços prestados pelas instituições de EaD, de modo a analisar e medir o perfil de evasão do discente e, principalmente, atender as expectativas da comunidade, se define como problema de pesquisa a seguinte questão: “Como fatores comportamentais do aluno no AVEA podem ser uteis na previsão da evasão discente em momentos distintos de um curso *online*?”. Os fatores comportamentais, ou seja, as variáveis relacionadas às interações do aluno com o AVEA e sua turma, são definidos como elementos centrais no presente estudo, pela riqueza de informações pedagógicas contidas em tais dados. Para responder este questionamento, são utilizadas estratégias pedagógicas associadas a técnicas de IA, especificamente lógica *fuzzy*, a fim

de apoiar adequadamente o discente frente ao volume de dados educacionais manipulados. A lógica *fuzzy* foi definida como técnica base após o presente estudo identificar a natureza nebulosa dos dados analisados e as possibilidades em aberto ilustradas pela revisão da literatura apresentada no Capítulo 2.

Além disto, o presente trabalho propôs uma abordagem *fuzzy*, no qual uma das entradas é definida a partir de informações obtidas por meio de técnicas de mineração de texto. Normalmente, para extrair as informações sobre o curso, sobre as interações dos alunos ou sobre as intervenções do tomador de decisão, as pesquisas na área de Informática na Educação manipulam dois tipos de fonte de dados nos AVEA, a saber: registros das interações dos participantes (*logs* de interação) e consultas via banco de dados. Em relação às análises via *logs* de interação, identificou-se que boa parte das informações extraídas é coletada sob o formato de texto nos cursos *online*. Tais informações representam as postagens dos alunos, professores e tutores; as conversas via bate-papos entre os participantes; as mensagens de *feedback* providas ao longo do curso para o discente e respostas em questionários subjetivos. Neste sentido, uma das estratégias da presente tese consistiu em incorporar técnicas de Processamento de Linguagem Natural, por meio da mineração das mensagens, nas avaliações executadas. O intuito da mineração de texto é extrair de forma automática informações contidas nos textos produzidos pelo discente durante o curso. No Capítulo 3 deste documento são apresentadas em detalhes as definições e as técnicas utilizadas nesta fase.

1.2 Objetivos

Tendo em vista as motivações apresentadas na Seção anterior e a pergunta de pesquisa obtida diante da temática destacada, o objetivo geral da tese é propor e avaliar a aplicação de uma abordagem avaliativa multidimensional na previsão da evasão discente em um curso *online*, especificamente por meio do uso de uma arquitetura *fuzzy* com mineração de texto.

A presente tese prioriza analisar fatores relacionados à interação do discente com os recursos pedagógicos propostos pelo professor nestes cursos e se subsidia em técnicas de IA para lidar com os dados educacionais. Questões de interface e aspectos externos ao ambiente não são contemplados na investigação. Assim, o presente estudo desenvolve um modelo de avaliação e monitoramento dos discentes frente ao seu comportamento no ambiente, ou seja, é observado como o aluno interage na turma e suas ações dentro do curso *online*. Para avaliar se o objetivo geral é cumprido, o presente estudo tenciona executar uma avaliação baseada em dados temporais de cursos *online* da UAB.

Diante do exposto, alguns objetivos específicos são delineados para que o objetivo geral seja alcançado, descritos a seguir:

- a) Mapear fatores comportamentais dos discentes relevantes à evasão discente em cursos *online*.
- b) Investigar uma estratégia de apoio aos docentes frente ao combate aos elevados índices de evasão nos cursos *online*.
- c) Planejar como extrair informações uteis das estruturas disponíveis no AVEA atuais.
- d) Definir qual processo metodológico utilizar para minerar dados a partir dos textos produzidos pelos alunos no curso *online*.
- e) Planejar como avaliar a abordagem proposta por meio de uma análise temporal com dados de cursos da UAB.

1.3 Contribuição

Inicialmente, a principal contribuição da pesquisa ora descrita é a proposição de uma abordagem avaliativa que apoie os tomadores de decisão educacionais (especificamente os professores) e os apoiem frente à problemática evasão em cursos *online*. A contribuição teórica se deve ao fato deste trabalho ser uma alternativa frente a um cenário limitado de pesquisas que investigam aspectos de interação aluno-ambiente relacionados à evasão no contexto virtual. É identificado um quantitativo inferior a dez estudos sobre a temática com

características semelhantes a proposta nesta tese. A contribuição prática é a identificação de fatores comportamentais relacionadas à interação discente no AVEA que, de fato, possuem correlação com a problemática evasão. Tais fatores são fundamentais à proposição de uma abordagem avaliativa, factível à estrutura das instituições de EaD da atualidade, que apoie o docente na minimização das desistências de alunos nos cursos *online*.

No contexto computacional, a presente tese apresenta como contribuição a proposição de um método de avaliação multidimensional do aluno por meio da utilização de técnicas relacionadas à lógica *fuzzy* na modelagem das métricas do aluno no ambiente, que utilizem recursos de mineração dos textos de uma das entradas. Não foi encontrado na literatura, de acordo com as estratégias de busca utilizadas, trabalhos que envolvam a incerteza no combate à evasão discente em cursos *online*. Além disto, diante das análises encontradas para combater a evasão, durante a revisão de literatura, nenhuma delas combinou mineração de texto e lógica difusa em suas análises. Este trabalho julga importante tal combinação pois, boa parte das avaliações nos AVEA é muito subjetiva (muito bom, bom, ruim) e boa parte das informações, que é extraída dos cursos *online*, é obtida por meio de textos em fóruns, questionários, bate-papos e registros em arquivos de *log*.

Por fim, a melhoria no apoio ao docente surge como uma alternativa para uma execução adequada das estratégias pedagógicas de ensino-aprendizagem por parte dos tomadores de decisão nos AVEA. Tal fato pode ser primordial no acompanhamento mais adequado dos cursos e na minimização da desconfiança dos alunos, pois os índices de evasão representam um dos fatores responsáveis pela descrença destes discentes, frente aos cursos *online*. Segundo o Relatório Retratos da Sociedade Brasileira (CNI, 2014), a população se mostra relativamente dividida em relação à eficácia da EaD na sociedade, já que 43% dos entrevistados afirmaram que a EaD funciona na prática em contraposição à parcela de 34% dos entrevistados que não acreditavam na EaD. Além disto, há os outros 23% que não quiseram ou não souberam responder.

1.4 Estrutura da Proposta

Este documento está dividido em 6 capítulos. No Capítulo 1, é apresentada uma visão geral do documento que se segue, com descrição da motivação, contribuição e dos objetivos geral e específicos da pesquisa. No Capítulo 2, é apresentada uma revisão da literatura, por meio de uma revisão sistemática sobre, principalmente, a evasão discente em cursos *online*. Além disto, é apresentado o estado da arte e pesquisas importantes na fundamentação e execução do da pesquisa. No Capítulo 3 é conduzida uma discussão sobre conceitos chave ao entendimento da abordagem avaliativa proposta, a saber: Lógica *Fuzzy*, Mineração de Texto e Métricas Pedagógicas importantes ao modelo. Conseqüentemente, no Capítulo 4 é apresentada uma descrição da abordagem de avaliação multidimensional, denominada *FuzzySD*, baseada nas limitações detectadas na literatura. A abordagem tem como intuito prover apoio à avaliação do discente nos AVEA. Além disto, o capítulo contempla a execução de experimentos iniciais para validação de entradas e das regras *fuzzy*. No Capítulo 5, há a apresentação e discussão da avaliação executada e dos resultados da pesquisa. Por fim, no Capítulo 6 são apresentadas as considerações finais da presente tese e os possíveis trabalhos futuros resultantes. Ao final do documento, são encontrados os Anexos e Apêndices referentes ao estudo.

2. Estado da Arte

Após apresentar os apontamentos iniciais do estudo no capítulo anterior, o presente capítulo objetiva detalhar trabalhos relevantes na literatura que pesquisem e explorem a problemática da evasão discente nos AVEA.

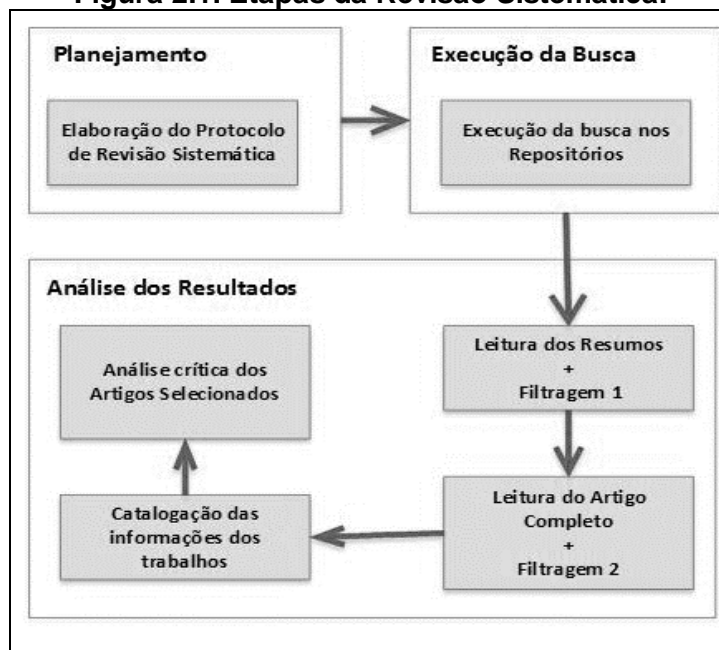
O presente estudo esmiuçou pesquisas e estudos relacionados ao combate à evasão discente em cursos *online*. Como consequência, algumas Revisões Sistemáticas (RS) foram realizadas ao longo do processo, mas o presente documento destaca a que focou sobre o mapeamento de trabalhos que investigam e propõem soluções sobre a evasão discente. Entende-se por RS de literatura a metodologia de busca científica que permite identificar, avaliar e interpretar pesquisas disponíveis e relevantes para uma determinada questão de pesquisa ou área temática (KITCHENHAM, 2004). Este método investigativo requer mais esforço do que investigações tradicionais (BIOLCHINI *et al.*, 2005). Sua grande vantagem é fornecer informações sobre os efeitos de algum fenômeno por meio de uma ampla gama de configurações e métodos empíricos. Esta metodologia aumenta a probabilidade de detecção de efeitos reais que estudos individuais são incapazes de detectar.

Foi realizada uma investigação de trabalhos publicados entre 2010 e 2017, e publicados em periódicos ou anais de eventos *online* de relevância, tais como: *IEEE Xplorer*, *Science Direct*, *Springer*, Periódicos Capes, e *ACM*.

A metodologia de RS executada percorreu três etapas principais: planejamento, execução da busca e análise dos resultados, ilustrada pela Figura 2.1. Na fase de planejamento, selecionou-se os repositórios de trabalhos científicos, definiu-se a melhor chave de busca a ser utilizada e elaborou-se os critérios de inclusão e exclusão para as filtagens dos trabalhos. Tais informações são fundamentais na elaboração do Protocolo de Revisão Sistemática que documenta todo o processo percorrido e guia as fases seguintes. A segunda etapa do processo contemplou a execução da busca nos repositórios. Nesta fase, ocorreu a primeira etapa de filtagens e leituras dos trabalhos. A partir deste momento do processo, os resumos foram lidos e, de acordo com os critérios de inclusão e exclusão, passaram, ou não, para a

próxima fase de análises (Filtragem 1). Os artigos puderam ser lidos em sua totalidade e passaram pela última filtragem (Filtragem 2). Por fim, os artigos foram finalmente catalogados e analisados com o auxílio da ferramenta *Evernote*.

Figura 2.1. Etapas da Revisão Sistemática.



Fonte: Próprio Autor.

A RS destacada neste documento tem como intuito investigar como as pesquisas acadêmicas buscaram minimizar os índices de evasão em cursos *online*. Neste sentido, a questão de pesquisa definida para esta investigação foi: "Como a literatura está combatendo os elevados índices de evasão discente na EaD ao redor do mundo?". O objetivo foi entender quais técnicas foram as mais exitosas, investigar os cenários de sucesso, entender quais os desafios e como contribuir para as pesquisas na área. As principais informações sobre a RS executada são apresentadas no Quadro 2.1.

Na Figura 2.2 são ilustrados os resultados obtidos de acordo com as bases de busca e as seleções feitas ao longo das filtragens. Neste sentido, identificou-se que a maioria dos trabalhos foi catalogada pelas bases *IEEE Xplorer* e *ACM*. Alguns trabalhos adicionais foram lidos fora do escopo da busca, por terem sido mencionados em publicações consideradas relevantes durante a RS. Por isso, alguns resultados apresentados podem estar fora do

período analisado pela busca.

Quadro 2.1. Principais Informações do Protocolo de Revisão Sistemática.

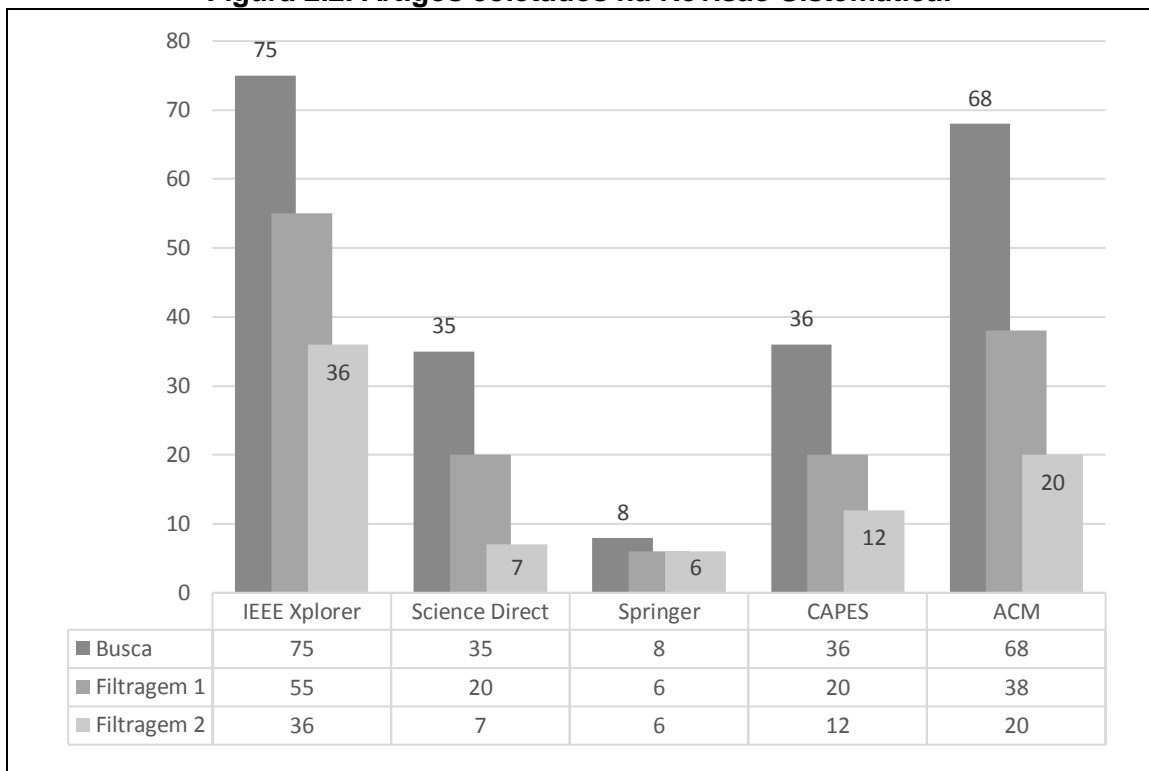
CRITÉRIO	REVISÃO SISTEMÁTICA EXECUTADA
Período da RS	Março de 2016 a agosto de 2017
Bases de Trabalhos científicos utilizadas	IEEE Xplorer, Science Direct, Springer
Expressão de busca	("Student" OR "Learner") AND ("dropout" OR "evasion") AND ("e-learning" OR "Online cooperative learning" OR "Distance education" OR "Interactive learning environment" OR "Intelligent tutoring system" OR "Online learning" OR "LMS" OR "MOOCs" OR "Massive Open Online Learning")
Anos das publicações selecionadas	Anos de 2010-2017
Bases de Trabalhos para a etapa da busca manual dos dados	Periódicos Capes, <i>Association for Computing Machinery</i> (ACM)
Expressões da busca manual	("Student" OR "Learner") AND ("dropout" OR "evasion") AND ("e-learning" OR "Online cooperative learning" OR "Distance education" OR "Interactive learning environment" OR "Intelligent tutoring system" OR "Online learning" OR "LMS" OR "MOOCs" OR "Massive Open Online Learning")
Ferramentas utilizadas	Evernote (https://evernote.com/intl/pt-br/) StArt (http://lapes.dc.ufscar.br/ferramentas/start-tool)
Número de Artigos coletados	222 artigos
Número de Artigos ao final das Filtragens	81 artigos

Fonte: Próprio Autor.

Quadro 2.2. Tipos de cursos e suas características.

Proporção de Conteúdo Entregue <i>Online</i>	Tipo de Curso	Descrição Típica
0%	Tradicional (Presencial)	Curso em que nenhuma tecnologia <i>online</i> é aplicada ao conteúdo.
1 – 29%	<i>Web</i> Facilitado	Curso que utiliza tecnologia baseada na <i>web</i> para facilitar o que é essencialmente um curso presencial. Pode utilizar um sistema de gerenciamento de cursos ou páginas da <i>web</i> para enviar o currículo e atribuições.
30 – 79%	Híbrido	Curso que combina a entrega <i>online</i> e presencial. Proporção substancial do conteúdo é entregue <i>online</i> , normalmente utiliza discussões <i>online</i> e, normalmente, tem um número reduzido de interações presenciais.
Mais de 80%	<i>Online</i>	Curso em que a maioria ou a totalidade do conteúdo é entregue por meio de um ambiente virtual educacional.

Fonte: Adaptado de Allen e Seaman (2016).

Figura 2.2. Artigos coletados na Revisão Sistemática.

Fonte: Próprio Autor.

Após definir as estratégias metodológicas percorridas para buscar e analisar trabalhos correlatos, é interessante relembrar quais as características dos cursos *online* analisados. Como comentado, o documento vigente se subsidia no agrupamento de cursos de Allen e Seaman (2016). Os autores definem que cursos *online* são aqueles em que pelo menos 80% do seu conteúdo foi transmitido por meio de um AVEA (ilustrado no Quadro 2.2).

2.1 Evasão Discente

Inicialmente, a revisão da literatura detectou que há uma série de definições referentes ao critério chave da pesquisa: a evasão discente. O Ministério da Educação (MEC) adota o conceito de evasão para cursos em geral como: “a saída definitiva do curso de origem sem conclusão, ou a diferença entre ingressantes e concluintes após uma geração completa” (MEC, 1997, p.19). Tal descrição deixa em aberto a possibilidade de combater ao longo do curso tal problemática, pois de acordo com a definição só é possível calcular a evasão ao final do semestre.

Quanto ao cenário de cursos *online*, a terminologia evasão discente possibilita diversas interpretações e já é utilizada em contextos distintos com significados diversos. Em alguns casos, considerou-se como evasão a desistência do curso pelo estudante, independentemente da quantidade de participações efetuadas (SANTOS e NETO, 2009; RODRIGUES, DE MEDEIROS e GOMES, 2013; BITTENCOURT e MERCADO, 2014); em outras situações diferenciou-se evasão de acordo com períodos médios para conclusão de curso e períodos anuais (ADACHI, 2009); bem como identificou-se como evasão as situações de desistência definitiva após determinado contato com o curso (FARIA, ALCANTARA e VASCO, 2008).

A definição mais ampla de evasão foi a falta de envolvimento dos discentes. Contudo, o envolvimento dos alunos pode ser difícil de quantificar como resultado (GUO, KIM e RUBIN, 2014), e várias medidas são usadas na literatura com este intuito. As definições de abandono dependem das metas pedagógicas e de engajamento do AVEA ou MOOC, e monitoramento dos sinais gerados pelos discentes. Duas formas distintas de engajamento influenciam as definições de evasão: a participação e a conclusão dos objetivos de aprendizagem.

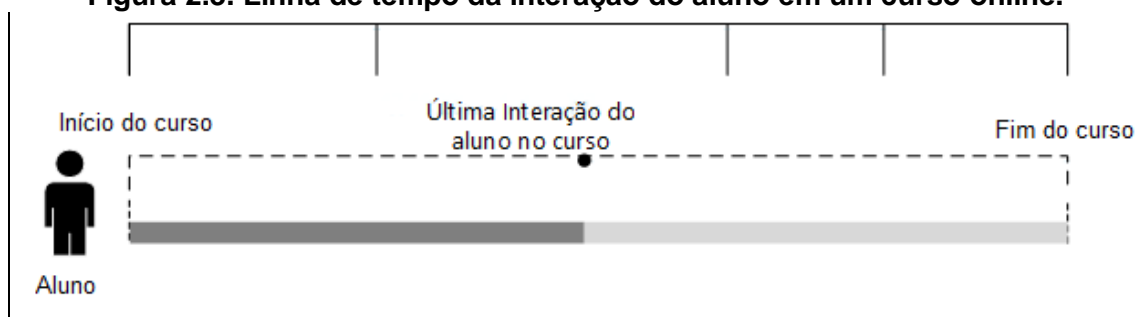
A falta de participação nos cursos *online* é definida como uma falta de interação com o curso *online* (BALAKRISHNAN e COETZEE, 2013; KLOFT, *et al.*, 2014), envio de tarefas e questionários (STEIN e ALLIONE, 2014), visualização de conteúdo de vídeo (SINHA *et al.*, 2014) ou participação em fóruns de discussão (YANG *et al.*, 2013, HE, 2012). Alguns pesquisadores se concentraram em investigar se o aluno alcançou os objetivos de aprendizagem do MOOC, definindo o abandono escolar como o momento em que o aluno não obtém um certificado ao final do curso (HE *et al.*, 2015) ou não consegue finalizar um determinado conjunto de módulos (DILLON *et al.*, 2016).

Em síntese, o presente estudo define evasão como: problema identificado quando o aluno permanece sem acessar e interagir com o AVEA por um período ininterrupto maior do que um quarto do tempo de execução do curso *online*, como mencionado na Introdução deste documento.

Na Figura 2.3 é ilustrada a interação do discente ao longo de um curso *online* por meio de uma barra cinza (clara e escura). A barra cinza escura

representa o período em que o usuário persistiu com as interações *online* e participou do curso. Por outro lado, o período da falta de interação do usuário com o AVEA é representado pela barra cinza claro. Se o usuário mantiver o período ausente maior do que um quarto do tempo do total do curso ele é considerado um aluno evadido. Caso as análises ocorram antes do final do curso, o tempo de interação considerado será o de atividade do curso *online*. Este tipo de análise pode ser aplicado em avaliações formativas.

Figura 2.3. Linha de tempo da interação do aluno em um curso online.



Fonte: Próprio Autor.

Após a análise inicial dos trabalhos, percebe-se que os sistemas educacionais estão cada vez mais adaptativos e se propõem a lidar com a individualidade dos alunos diante do processo de ensino-aprendizagem nas ferramentas de ensino virtual (WALTERS-ARCHIE, 2017; LEI *et al.*, 2016). Há ainda trabalhos que propõem melhorar a dinâmica pedagógica e personalizar o aprendizado por meio de soluções baseadas em técnicas de IA, que objetivam minimizar a distância transacional entre docentes e discentes, ou seja, diminuir o espaço cognitivo entre professor e aluno no ambiente educacional (MAMMAN *et al.*, 2017).

Este espaço psicológico e de comunicação existente entre os dois parceiros no empreendimento educacional é um espaço de potencial mal-entendido entre as entradas do instrutor e as do aluno, pois muitas vezes as diretrizes que guiam este processo não são claras para os participantes.

Macfadyen e Dawson (2010) afirmam que um dos desafios na área da EaD é desenvolver métodos capazes de prever o comportamento discente, de modo a possibilitar a intervenção de professores e tutores visando resgatar o estudante antes que seja reprovado ou desista do curso.

2.2 Avaliação do Discente no Combate à Evasão

Em sentido amplo, os métodos de avaliação da aprendizagem ocupam uma posição relevante no conjunto de práticas pedagógicas nos AVEA. Nesses ambientes, tais processos extrapolam a mecânica formal de, apenas, atribuir notas às interações dos discentes junto aos recursos pedagógicos e se tornam atividades, cada vez mais, complexas. Assim, o monitoramento das interações é uma das ações de maior representatividade neste contexto, pois minimizam as lacunas arquiteturas associadas à virtualização do processo de ensino-aprendizagem e as taxas de insucesso dos discentes em cursos *online* (GRAU-VALLDOSERA e MINGUILLÓN, 2011; BAXTER, 2012).

Na maior parte dos AVEA as medidas de insucesso são atribuídas a dois fenômenos principais: a evasão e a reprovação, como mencionado. De acordo com a literatura, entende-se o termo evasão *online* como a desistência do discente em interagir com as atividades do curso nos AVEA (CHEN e ZHANG, 2017). Ao mesmo tempo em que o fenômeno reprovação acontece quando o aluno não atinge o limiar mínimo de nota média para ser aprovado no curso e precisa fazê-lo novamente.

É relevante destacar que nesse contexto deficiente, em que estudantes iniciam, mas não terminam seus cursos *online*, geram-se desperdícios sociais, acadêmicos e econômicos, muitas vezes irreparáveis no cenário educacional (PARK e CHOI, 2009; LEEDS *et al.*, 2013). Assim, na presente Seção do documento objetiva-se apresentar trabalhos que tiveram o intuito de analisar e investigar esforços relacionados aos processos avaliativos intuindo a minimização dos índices de evasão dos alunos em cursos *online*.

Normalmente, o aluno da EaD precisa transpor o paradigma de que para aprender a presença do professor é algo fundamental no processo. Na visão de Peters (2003, p. 379), isso é expresso pelo fato dos alunos trabalharem em relativo isolamento em relação aos outros e todas as atividades para aquisição do conhecimento se iniciarem por eles mesmos, ou seja, os alunos se tornaram os principais envolvidos dentro do processo de ensino aprendizagem. Muitos discentes tenderam a evadir de um curso *online* por não estarem preparados a quebrar paradigmas educacionais. Neste

sentido, Peters (2003, p. 48) afirma que uma “mudança de paradigma” significaria que na educação certos modelos ou padrões não existiriam mais porque novos modelos e padrões que diferem dos antigos de modo marcante os substituíram.

Diante do exposto, um dos principais fatores problemáticos nos cursos *online* é a dificuldade apresentada pelos professores em apoiar o engajamento e motivação dos seus aprendizes ao longo das interações, já que o discente *online* é circundado por muitos elementos distrativos em relação aos presenciais (DETONI, ARAÚJO e CECHINEL, 2015; ABBAD, CARVALHO e ZERBINI, 2008; BELLONI, 2006). No contexto da EaD, é primordial que o aluno não só domine as ferramentas tecnológicas, mas que se disponha a fazer uso dessas e a tolerar alguns aspectos inerentes à EaD, tal como a típica limitação dos processos de interação professor-aluno e aluno-aluno.

Apesar do crescente interesse na EaD e do reconhecimento dos seus benefícios e vantagens, ainda restam graves hiatos na produção e sistematização dos conhecimentos em cursos *online*. A literatura revelou uma necessidade de análises e discussões que foquem em ações educacionais a distância, principalmente a respeito dos motivos influenciadores da evasão em cursos dessa natureza (VILARINHO e PARO, 2011; SALES, ABBAD e RODRIGUES, 2011).

Xenos *et al.* (2002) destacam, em seu estudo, que os cursos *online* são compostos por quatro categorias de discentes: (a) os estudantes que se inscrevem, mas nunca começam os seus estudos e não se registram no ano seguinte; (b) os estudantes que iniciam seus estudos e concluem com êxito algumas tarefas ou módulos, mas saem por várias razões; (c) os estudantes que não completam com sucesso alguns ou todos os módulos, mas continuam seus estudos e repetem estes módulos no ano seguinte; (d) os estudantes que concluem com êxito todos os módulos selecionados durante o primeiro ano de seus estudos. Deste modo, tal descrição se mostrou fundamental na identificação do perfil discente que compõe o público-alvo das análises da presente pesquisa, ou seja, os grupos (a) e (b).

Apesar da aparente simplicidade, o processo de aprendizagem é uma atividade ativa e neste sentido deve ser estimulada de modo contínuo, pois foi

pautado nas interações sociais, que são capazes de fornecer ao aprendiz maneiras de tornar mais eficazes seus esforços de adaptação e solução de problemas (VYGOTSKY, 2008). Além disto, a aprendizagem ativa tem como aliados a avaliação formativa e a autoavaliação frente aos desafios relacionados à diminuição dos índices de evasão de discentes no Brasil e no mundo.

Em relação ao fenômeno da evasão, o insucesso pode ocorrer por uma série de razões associadas a fatores distintos, mas a maioria das pesquisas que investigou sobre o tema se limitou a elencar fatores pessoais, profissionais e educacionais (ALLEN e SEAMAN, 2016; COSTA, 2013). Por exemplo, o estudo de Lee e Choi (2011) observou que as razões para a evasão puderam ser divididas, de modo geral, em fatores motivacionais internos (que influenciaram o desejo de um aluno a persistir no curso) e fatores externos. Neste mesmo sentido, Ameri *et al.* (2016) propuseram um arcabouço baseado em análises estatísticas para avaliar a evasão precoce por meio de métricas cadastradas durante a matrícula do aluno no curso, tais como: atributos demográficos, condições familiares, aspectos financeiros, *etc.*

A pesquisa de Sales, Abbad e Rodrigues (2011), executada na Empresa Brasileira de Pesquisa Agropecuária (Embrapa) entre 2008 a 2009, objetivou identificar variáveis preditivas de persistência e evasão em cursos *online* de treinamentos. Os dados foram selecionados por meio de questionários e avaliaram a percepção dos alunos sobre os fatores que facilitaram ou dificultaram sua permanência no curso. A amostra analisada era composta pelas respostas de 535 discentes concluintes e 48 discentes evadidos. Os autores realizaram análises de regressão logística para identificar a influência exercida por características ao longo do curso, do contexto e do perfil dos participantes dos treinamentos sobre a variável critério: situação do participante ao final dos cursos (concluinte ou evadido). Sales, Abbad e Rodrigues (2011) consideraram como evadidos os alunos que iniciaram o curso, mas não realizaram as atividades obrigatórias para conclusão e certificação. Os resultados mostraram que foram maiores as chances de concluir o treinamento para aqueles que expressaram autoavaliações mais favoráveis de disciplina para estudar e de interesse pelo curso e manifestaram

maior satisfação com o desempenho dos tutores. Os resultados desta pesquisa ressaltaram ainda que o uso adequado do recurso pedagógico denominado fórum de discussão e a dedicação de tempo para realização do curso e menos problemas familiares, possibilitaram a conclusão nos treinamentos. Para o estudo, dois pontos principais não influenciaram o fenômeno evasão durante as análises: a pouca diversidade entre a estrutura dos cursos analisados e as variáveis demográficas, tais como gênero e idade.

Outra investigação relevante foi executada por Choi *et al.* (2013) que explorou as motivações institucionais e características gerais do discente para a evasão no cenário da EaD na Coréia do Norte. No estudo, os autores objetivaram: 1) investigar a extensão e as causas do retorno discente para um curso *online* na Universidade Nacional da Coréia do Sul; e 2) sugerir ações para melhorar a retenção de estudantes, em geral, e aqueles com maiores riscos de abandono em particular.

A pesquisa de Choi *et al.* (2013) analisou as respostas de uma amostra de 1.353 discentes referentes aos cursos *online* de 2009, obtidas por meio da execução de um *survey*. Os resultados indicaram que a falta de *feedback* dos instrutores, a carga de trabalho pesada e as dificuldades de estudo à distância foram os principais motivos para o não retorno discente ao curso. As percepções dos alunos sobre o valor dos graus e suas idades, gênero e origens educacionais também foram considerados fatores significativos nas decisões de não se retornar ao curso. As soluções sugeridas para reduzir as evasões e recomeços incluíram: diminuição do número de horas de crédito exigidas por semestre; apoio social mais forte; uso de um sistema de inscrição mais flexível; e melhoria no uso da tecnologia e infraestrutura disponíveis para ajudar alunos e instrutores a construir comunidades de aprendizado mais fortes.

Diferentemente dos trabalhos anteriormente apresentados, nos quais as análises sobre a evasão só ocorrem *a posteriori* ao curso (avaliação somativa), foi identificada uma vertente de trabalhos que almeja realizar tais análises durante o andamento do curso *online* (avaliação formativa).

Considerando que a EaD permite uma medição do engajamento dos estudantes a partir da observação dos *logs* de interação do sistema e da

avaliação da intensidade de suas ações nas diferentes atividades das salas virtuais (BEER, CLARK e JONES, 2010), diversos trabalhos focam seus esforços na predição automática do desempenho acadêmico a partir da utilização de tais registros. Por exemplo, Manhães *et al.* (2011) executaram técnicas de mineração de dados para prever a evasão de estudantes em cursos presenciais da Escola Politécnica da Universidade do Rio de Janeiro. No experimento, dez modelos foram testados, com acurácia média entre 75% e 80%, com classificadores baseados em *Perceptron* de multicamadas e Florestas Aleatórias (*Random Forest*) proporcionando os melhores desempenhos. Ainda, Gottardo, Kaestner e Noronha (2012), apresentaram resultados de tentativas de prever a avaliação de desempenho de alunos de EaD. Os autores relataram taxas de precisão acima de 74%, com o uso de uma grande variedade de atributos.

Quanto às variáveis relacionadas à evasão no curso *online*, a literatura relata que os problemas relacionados ao desempenho do tutor, como: falta de apoio ao aluno, *feedback* insuficiente, falta de conhecimentos ou inabilidade para transmiti-los, são descritos como uma das causas para o abandono de cursos *online* (ABBAD, CARVALHO e ZERBINI, 2006). Apesar destes e de outros resultados, os estudos na área foram pouco conclusivos, de modo que assim se destacaram as investigações que se propuseram a avaliar o fenômeno da evasão, a exemplo deste relato.

Outro estudo que seguiu a mesma linha de investigação foi o trabalho de Ghamdi (2017), que objetivou identificar comportamentos verbais e não verbais de palestrantes de cursos *online* como fator potencial na redução do sentimento de isolamento percebido dos alunos e no apoio às experiências de aprendizagem dos alunos. Tais análises foram realizadas por meio de questionários qualitativos e quantitativos e se basearam em quatro grupos de variáveis de envolvimento de alunos: (1) participação em classe; (2) satisfação de comunicação; (3) aprendizagem afetiva; e (4) aprendizagem cognitiva. O estudo foi conduzido em uma universidade na Arábia Saudita e envolveu participantes matriculados em cursos de educação a distância. O principal resultado foi que os comportamentos de imediatismo baseados em textos facilitaram uma dupla influência: (a) um impacto ontológico, estimulando a

presença social do professor que apoiou a aprendizagem dos alunos e reduziu o senso de isolamento e separação, e (b) uma mudança epistemológica, por meio da qual os alunos estavam mais confiantes em se comunicar e participar ativamente para a construção do conhecimento.

Assim, limitar e minimizar o abandono é um dos principais objetivos e desafios das instituições que oferecem cursos nesta modalidade. Por isto, o primeiro passo para uma contribuição efetiva para este desafio é identificar quais métricas do discente estão diretamente relacionadas à interação do aluno com os AVEA e podem estar relacionadas a sua desistência.

Seguindo esta premissa, a presente tese pontua algumas pesquisas que identificam fatores comportamentais do discente relacionados ao fenômeno da evasão nos cursos *online*. Este é o principal aspecto investigado pelo estudo, pois são essas variáveis que descrevem o comportamento do aluno nos AVEA. Além disto, tais medidas podem ser ajustadas ao longo do processo de ensino-aprendizagem, por meio de mensagens *feedback* e de novas atividades pedagógicas pelo docente, de acordo com seus objetivos.

Como destaque inicial, é possível ressaltar o estudo de Murray *et al.* (2013) que investigou a correlação de fatores relacionados ao desempenho discente e percebeu que estudantes que apresentavam altas taxas de acesso aos conteúdos nos AVEA de um curso *online*, também obtinham as melhores notas e alcançavam os menores índices de insucesso. Além disto, os trabalhos de Kloft *et al.* (2014), Sinha *et al.* (2014) e Dickson (2005) identificaram a existência de correlação entre o número total de cliques dados por estudantes com seu desempenho e comportamento no curso *online*.

2.2.1 Predição da Evasão

Outra vertente presente nos resultados da RS executada, consiste na análise dos trabalhos que exploraram estratégias de predição da evasão. Normalmente, este processo é custoso e não é extensível a diferentes contextos, plataformas ou cursos com características diferentes. A maioria dos trabalhos selecionados executou estratégias de classificação binária das entradas, ou seja, se o discente evadiria ou não do curso. Para resolver esta

questão, uma série de técnicas foi destacada na literatura nos últimos anos (AMNUEYPORNSAKUL, BHAT e CHINPRUTTHIWONG, 2017; KLOFT *et al.*, 2014; MI e YEUNG, 2015; TAYLOR, VEERAMACHANENI e O'REILLY, 2014).

Diante da perspectiva da avaliação baseada na predição da evasão e da investigação sobre fatores comportamentais nos AVEA, destaca-se o artigo de Chen e Zhang (2017) que propôs um sistema supervisionado de predição da evasão discente, especificamente por meio da Florestas Aleatórias (*Random Forest*) para classificar os dados do MOOC. Para isto, foi utilizada uma base de dados históricos das interações dos alunos no curso de "*Data Structures and Algorithms*" da Universidade de Pequim, ministrado por meio do Coursera. Uma vez que diferentes recursos pedagógicos puderam ter benefícios distintos para diferentes cursos, foram criados dois componentes opcionais baseados no estudo estatístico de dados.

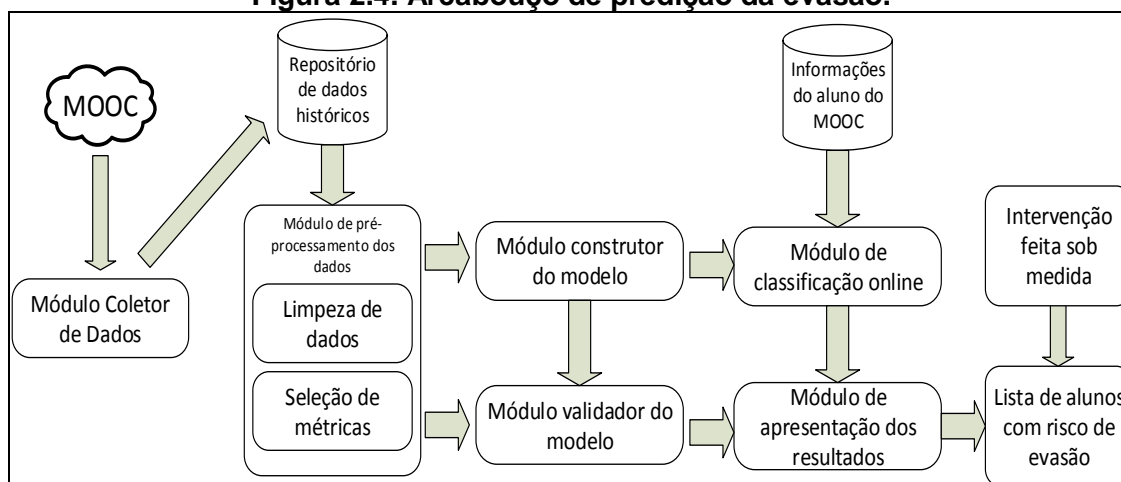
O estudo empírico mostrou que o sistema de previsão de abandono escolar alcançou alta eficácia na classificação dos estudantes. O sistema foi baseado em uma análise estatística das correlações entre os dados comportamentais dos alunos e a evasão. O trabalho de Chen e Zhang (2017) considerou que a evasão acontecia quando o aluno ficava sem interagir no ambiente por um período superior a 30% do tempo de atividade do curso, por exemplo: se o curso tivesse 10 semanas, caso o aluno não interagisse por 3 semanas sem intervalos, ele era considerado evadido.

Com base nos resultados do estudo estatístico e nos resultados do sistema de previsão de abandono, várias sugestões foram feitas para ajudar a melhorar o gerenciamento do curso na perspectiva da prevenção de abandono escolar, como oferecer aos alunos mais chance de tentar responder os questionários e tarefas, estender o período de realização de atribuições graduadas, incentivar os alunos a participar da discussão no fórum e projetar questionários em vídeo para dividir cada vídeo em pequenos fragmentos. Por fim, o sistema de predição identificou que o perfil de interação dos alunos persistentes e os alunos evadidos foram bem distintos entre si.

A pesquisa de Tang, Xie e Wong (2015) propôs prever a evasão ou persistência do aluno em cursos *online* da plataforma MITx e HarvardX por meio de técnicas de classificação de dados, especificamente árvores de

decisão (Figura 2.4). Dados históricos foram analisados e um modelo de classificação foi gerado automaticamente para identificar estudantes que potencialmente evadiriam. Este trabalho detectou que algumas intervenções apropriadas podem ser bem úteis no combate a esta problemática, tais como: lembretes por *email* e medições adicionais do desempenho. Além disto, este arcabouço se mostrou bem eficiente e proporcionou índices de predição próximos a 95% de acertos.

Figura 2.4. Arcabouço de predição da evasão.



Fonte: Adaptado de Tang, Xie e Wong (2015).

Outra pesquisa relevante é a de Wang, Yu e Miao (2014) que propôs um modelo de predição automática de evasão baseada na combinação de Redes Neurais Convolucionais (*Convolutional Neural Network*) e Redes Neurais Recorrentes (*Recurrent Neural Networks*) como classificadores. Este trabalho foca na predição da desistência discente nos 10 dias seguintes à análise. As principais conclusões obtidas pelos autores foram: (1) para problemas de classificação com base em registros de atividade, existem formas eficientes de extrair automaticamente recursos dos dados brutos, tais como: regressão logística e redes bayesianas; (2) a abordagem adotada pelo método proposto para extrair recursos se mostrou efetiva; e (3) o modelo de Rede Convolucional é um modelo eficiente para resolver o problema de previsão de abandono discente em MOOC. Desta forma, o método proposto economizou tempo e esforço humano, além de eliminar a incoerência potencial introduzida pelo processo manual. Os resultados experimentais em um grande conjunto de dados públicos demonstraram que o modelo proposto alcançou desempenho

comparável às abordagens que dependem da engenharia de *features* executada por especialistas.

Outra vertente de trabalho encontrada na literatura são os trabalhos que propuseram avaliar, além da evasão do aluno, a predição de conclusão do discente no curso. Nesse tipo de problema, em vez de prever se um estudante abandona um curso, estes predisseram se um aluno completará ou não um curso ou obterá o certificado correspondente. Por exemplo, He *et al.* (2015) usaram regressão logística para identificar se os discentes completariam ou não o curso em sua totalidade.

O trabalho de Nagrecha, Dillon e Chawla (2017) também utilizou árvores de decisão e regressão logística para prever a evasão na plataforma edX. Inicialmente, o estudo destacou que o problema do abandono em cursos *online* tem suas raízes no envolvimento dos alunos e não pode ser estudado de forma isolada. As características que englobaram elementos do envolvimento dos discentes no curso precisaram ser escolhidas de forma que estivessem disponíveis para todo o conjunto de alunos analisados. Etapas de pré-processamento como a Análise de Componentes Principais (*Principal Component Analysis* - PCA), que obscurecem a interpretação das entradas analisadas, não deveriam ser usadas de acordo com os autores. Dado que a evasão é uma problemática, as técnicas de re-escalamento ajudaram a aliviar seu efeito negativo sobre a previsão. Por meio de uma avaliação supervisionada longitudinal semana a semana, as previsões baseadas nos modelos de classificação forneceram aos pesquisadores informações aprofundadas sobre o motivo pelo qual um aluno foi susceptível a abandonar o curso *online*.

Em congruência com as constatações anteriores, destaca-se ainda o trabalho de Halawa *et al.* (2014). Neste estudo, os autores concordaram com a premissa do monitoramento contínuo de que a interação do aluno pode ser fundamental no combate à evasão. Contudo, o grande diferencial desta metodologia é que a análise se restringiu aos 10 primeiros dias do curso *online*, com o intuito de permitir intervenções dos docentes em situação de insucesso iminente. Este trabalho teve como objetivo apresentar funções de predição para identificar a evasão do aluno no ambiente, antes mesmo da sua

ocorrência, de acordo com as características de interação do aluno com o ambiente. O presente estudo propôs duas funções preditoras: uma para operar com alunos ativos e outra para analisar alunos que se ausentaram por um período ao longo do curso. Após a etapa de validação destes modelos, o primeiro tipo de predição conseguiu prever entre 40% e 50% das evasões. A segunda predição detectou 90% das desistências de alunos nos cursos observados. Para resumir as aplicações relacionadas à predição da evasão, um quadro resumo (Quadro 2.3) é apresentado.

Quadro 2.3. Trabalhos relacionados à predição da evasão.

TRABALHO	TÉCNICA DE PREDIÇÃO	OBJETIVO
Chaturvedi, Goldwasser e Daume (2014)	Cadeias de Markov (HMM), Modelos de Cadeias Globais, Regressão Logística e Árvores de Decisão	Predizer a intervenção do professor ou do tutor nos fóruns de discussão em um MOOC em relação à evasão discente.
Tang et al. (2015)	Árvores de decisão	Predizer a evasão ou persistência dos discentes em cursos <i>online</i> da plataforma Coursera.
Kloft et al. (2014)	<i>Support Vector Machine</i> (SVM)	Investigar as características dos discentes por meio de cliques, classificados por meio de uma SVM, para prever a evasão semana a semana.
Wang, Yu e Miao (2017)	Redes Neurais Convolucionárias e Redes Neurais Recorrentes	Propor um modelo de predição automática de evasão baseada na combinação de Redes Neurais Convolucionárias (<i>Convolutional Neural Network</i>) e de Redes Neurais Recorrentes (<i>Recurrent Neural Networks</i>).
Cheng e Zang (2017)	Florestas Aleatórias (<i>Random Forest</i>)	Propor um sistema não supervisionado de predição da evasão discente, especificamente por meio da Florestas Aleatórias para classificar os dados de MOOCs.
Nagrecha, Dillon e Chawla (2017)	Regressão logística e Árvores de decisão.	Fornecer aos pesquisadores informações aprofundadas sobre o motivo pelo qual um aluno é susceptível de abandonar.
Balakrishnan e Coetzee (2013)	Cadeias de Markov (HMM)	Identificar os padrões de comportamento dos alunos em relação à sua interação com o MOOC e sua propensão para permanecer no curso e para prever se um aluno provavelmente permanecerá no curso na semana seguinte.
Halawa et al. (2014)	Regressão logística	Propor duas funções preditoras: uma para operar com alunos ativos e outra para analisar alunos que se ausentam por um período de tempo ao longo do curso.

Continua

Continuação

TRABALHO	TÉCNICA DE PREDIÇÃO	OBJETIVO
Hernandez (2008)	Regressão logística	Prever a conclusão do curso <i>online</i> da faculdade comunitária com base nas percepções da faculdade, produzindo um modelo de previsão das taxas de conclusão do curso.
Boyer e Veeramachane ni (2015)	Regressão logística	Prever as paradas do aluno no curso, ou seja, prever quais alunos provavelmente deixarão de se envolver no curso <i>online</i> .
Mishra, Kumar e Gupta (2014)	Árvores de decisão (<i>Random Tree</i> e J48)	Utilizar técnicas de classificação para construir um modelo de previsão de desempenho baseado na integração social dos alunos, integração acadêmica e várias habilidades emocionais.

Fonte: Próprio Autor.

A literatura destaca que diversos trabalhos realizam análises multidimensionais, por meio da análise de arquivos de registros de atividades (os arquivos de *log* do AVEA). Tais registros permitem que os pesquisadores consigam ter acesso a partes de um curso *online* e obter informações relacionadas ao uso de recursos pedagógicos pontuais quando necessário, tais como: os vídeos, os bate-papos, o fórum de discussão e o curso *wiki* (AMNUEYPORNSAKUL, BHATE e CHINPRUTTHIWONG, 2014; KLOFT et al, 2014; SHARKEY e SANDERS, 2014; TAYLOR, VEERAMACHANENI e O'REILLY, 2014). Os registros das atividades podem ser separados em diferentes períodos (*i.e.*, por semana) e representados por suas características estatísticas. Normalmente, análises de frequência são as mais difundidas nos AVEA da atualidade. Contudo, cada vez mais, os tomadores de decisão e gestores educacionais demandam análises de dados educacionais que envolvam a extração de informações mais refinadas e adaptativas.

Com os recursos extraídos, muitos modelos clássicos de aprendizagem de máquinas são testados, incluindo *Support Vector Machine* (SVM) (AMNUEYPORNSAKUL, BHATE e CHINPRUTTHIWONG, 2014; KLOFT et al, 2014; SINHA et al., 2014), Árvores de Decisão (SHARKEY e SANDERS, 2014) e Regressão Logística (BOYER e VEERAMACHANENI, 2015; TAYLOR e VEERAMACHANENI, 2015).

A precisão da maioria dos trabalhos mencionado está entre 60 e 95%.

No entanto, a maioria das pesquisas anteriores sobre previsão de abandono do MOOC mediu a precisão dos testes no mesmo curso usado para treinamento, o que pode levar a estimativas de precisão excessivamente otimistas (WHITEHILL *et al.*, 2017). Por outro lado, Boyer e Veeramachaneni (2015) utilizaram modelos de aprendizagem de transferências para prever o abandono de um curso com dados de treinamento de outro curso.

2.2.2 Métricas comportamentais discentes relacionadas à evasão

Em um contexto mais amplo, há o estudo de Mezzari *et al.* (2013), que realizou um apanhado sobre os aspectos relacionados ao fenômeno da evasão na EaD. Este trabalho defende a participação dos alunos nos AVEA como indicativo eficiente na detecção de alunos em perigo de evasão. Inicialmente, a pesquisa validou alguns indicadores de desempenho, propostos por Araújo e Lucena (2005): tipo de participação, tipo de interação e tipo de comunicação. Tais índices permitiram a avaliação das interações e correlação dos resultados obtidos com o desempenho da equipe. Os contatos feitos com alunos que evadiram permitiram a discussão das causas mais comuns para a evasão. Assim, os indicadores investigados mostraram-se capazes de sugerir estratégias para detectar os estudantes na iminência da evasão e, com isto, minimizar tais números, mediante a introdução de um tutor mais motivador. Por fim, o trabalho destacou como estratégia para detecção da evasão do discente um processo de acompanhamento contínuo deste durante o curso.

Outro trabalho relevante para a presente análise é a pesquisa de Rostaminezhad (2013), que encontrou fatores relacionados ao complexo fenômeno da evasão nos AVEA. O estudo citado identificou uma série de critérios que afetaram direta ou indiretamente os alunos. Três critérios se destacam, de acordo com a literatura pesquisada: a satisfação do aluno como um fator de motivação, autorregulação e interação. Além disto, foi executado um estudo de caso com estudantes para investigar a importância destas métricas. A amostra coletada foi composta por 877 alunos de EaD da Universidade de Ciência e Tecnologia do Irã. Inicialmente, o estudo de caso sobre os discentes iranianos revelou que existe uma correlação forte entre a

autorregulação e o sucesso do aluno. Neste mesmo contexto, em relação à satisfação e interação dos discentes, mais pesquisas são necessárias, pois as correlações se mostraram fracas. Finalmente, para fins didáticos, no Quadro 2.4 estão elucidadas as principais métricas relacionadas à evasão dos discentes nos AVEA, que podem ser úteis na análise da evasão dos discentes em cursos *online*.

Quadro 2.4. Fatores relacionados à evasão dos discentes de EaD.

Métricas do Aluno de EaD	Trabalhos Relacionados
Motivação	Giesbers (2013) Park e Choi (2009) López-Pérez, Pérez-López e Rodríguez-Ariza (2011) Joo, Joung e Sim (2011) Rostaminezhad (2013)
Autorregulação	Rostaminezhad (2013) Cho e Shen (2013) Broadbent e Poon (2015)
Interação	Hernandez (2008) Mezzari <i>et al.</i> (2013) Rostaminezhad (2013)
Presença Social	Hernandez (2008) Liu, Gomez e Yen (2009).

Fonte: Próprio Autor.

2.3 Avaliação Multidimensional

Além das avaliações pontuais dos recursos pedagógicos dos AVEA, é importante mencionar também na literatura os estudos que priorizam o monitoramento multidimensional dos discentes. Normalmente, tais análises estão associadas à modelagem do conhecimento do professor frente ao processo avaliativo e ao entendimento da complexidade do aluno nas interações de ensino-aprendizagem virtuais.

Uma informação fundamental à discussão é a delimitação do termo multidimensionalidade para o presente trabalho, pois é importante diferenciá-lo de multivariado. Neste sentido, a pesquisa entende por avaliação multidimensional aquela que envolve diversos critérios relacionados ao processo de aprendizagem, sejam esses relacionados à interação ou ao desempenho do discente, ou seja, metodologias ou modelos que tentam se aproximar das análises globais, realizadas pelos docentes ao longo do

processo de ensino-aprendizagem. Normalmente, tais estudos pertencem a dois grupos principais: aqueles que utilizam metodologias avaliativas pertencentes a vários tipos recursos pedagógicos e aqueles que combinam critérios de diferentes atores dos AVEA, por exemplo: professores e alunos.

Na presente subSeção, o intuito principal é agrupar trabalhos que envolvam estudos e melhorias nas pesquisas relacionadas à avaliação multidimensional do discente nos cursos *online*. Além disto, a principal contribuição desta análise é a construção de um quadro resumo, com os principais estudos, e a discussão sobre as lacunas detectadas nesta temática.

No contexto das pesquisas identificadas na literatura sobre a temática multidimensional, uma das vertentes em que este tipo de análise se destaca é a parceria com a avaliação formativa. Assim, merece destaque o estudo de Daif e Rizkas (2013) que propôs uma modelagem de monitoramento multidimensional para analisar o desempenho do discente de EaD. Esta avaliação tem como objetivo principal minimizar as taxas de evasão de alunos por meio de melhor apoio no acompanhamento das turmas pelos docentes.

Os autores afirmam que os fatores principais para evasão são a presença e participação do aluno; incluindo a participação nas aulas, submissão de atividades, projetos e pontuações de exames e questionários. Além disto, o modelo de monitoramento proposto foi composto por uma série de funções: geração do evento educacional, registro de evento, recebimento de notificações pelo alunos, resposta e execução de ações pelos alunos, registro das atividades dos discentes, relatórios JIT, avaliação dos alunos, provimento de ajuda pelos instrutores, recebimento da ajuda pelos alunos, *feedback* do aluno e recebimento do *feedback* do aluno pelos instrutores.

A dinâmica provida por tais funções se mostrou fundamental no cálculo dos indicadores de monitoramento definidos pelos autores, a saber: comparecimento às aulas (*Class Attendance*), submissão de tarefas (*Tasks and Assignments submission*), realização de provas (*Exams*), respostas a questionários (*Quizzes*) e participação na aula (*Class Participation*). Estes cinco indicadores foram utilizados na composição da variável “índice” (*i*). Um aluno com $i = 1$, é um aluno que esteve presente em todas as aulas, exercícios, testes, apresentou todos os trabalhos e foi avaliado em todas as participações.

A grande limitação desta abordagem é desconsiderar as incertezas associadas ao processo avaliativo e não permitir ao docente utilizar recursos pedagógicos diferentes dos questionários e tarefas.

Dando continuidade à discussão sobre as pesquisas que abordam a temática avaliações multidimensionais aliada à avaliação formativa, evidencia-se também o estudo de Dias e Diniz (2013). Este artigo tem o intuito de estimar a qualidade da interação de alunos e professores no ambiente, propondo uma ferramenta a *FuzzyQol*. A adoção da lógica *fuzzy* é baseada na flexibilidade desta metodologia em relação à lógica clássica ao lidar com variáveis qualitativas e quantitativas conectadas. Segundo os autores, um AVEA tem, em média, 110 métricas que descrevem os usuários e podem ser usadas como entrada para este modelo. Contudo, a abordagem proposta utilizou 12 variáveis de entrada e 5 sistemas de inferência *fuzzy* ao longo do sistema de avaliação. Além disto, 120 regras IF/ELSE *fuzzy* foram concebidas para alimentar a etapa de inferência. Para a validação, o estudo utilizou dados de 51 semanas de interação em 5 cursos distintos da Universidade de Lisboa. O trabalho constatou que a avaliação poderia ser útil no monitoramento e *feedback* rápido aos alunos nas instituições de ensino.

Os sistemas de *feedback* são elementos fundamentais ao processo de ensino-aprendizagem e para avaliação formativa, é imprescindível comentar sobre a pesquisa de Kao, Lin e Chu (2012), que investiga esta relação. Além da avaliação formativa, este trabalho apresenta elementos da análise diagnóstica. Os autores propuseram uma avaliação multidimensional a fim de prover aos alunos interações personalizadas dos sistemas de *feedback*, apoiando o processo instrucional desses. Para isto, esta metodologia percorreu três etapas principais. Em primeiro lugar, vários fatores foram levados em consideração para fornecer diagnóstico mais flexível e completo ao docente. Em seguida, explorou-se a lógica *fuzzy* como o pensamento humano e julgamento para avaliar o sucesso da aprendizagem dos alunos. E, finalmente, de acordo com o diagnóstico, o *feedback* adaptativo proveu sugestão de correção e instrução ao discente. Os autores conduziram um experimento envolvendo 52 alunos matriculados em um curso de construção de compiladores em uma Universidade de Taiwan. Os resultados experimentais

mostraram que o método proposto pode ajudar significativamente os alunos a melhorar seu desempenho durante a aprendizagem.

Contrapondo à perspectiva dos trabalhos, anteriormente apresentados, que conectam os princípios da multidimensionalidade com a avaliação formativa, são mencionadas as pesquisas que aplicam tal abordagem diante da análise somativa. Neste sentido, destaca-se, inicialmente, os estudos publicados por Arias *et al.* (2012), Chen *et al.* (2011), Saleh e Kim (2009), Bai e Chen (2008) e Weon e Kim (2001). Estes artigos possuem duas intersecções principais: avaliação do desempenho do discente de maneira somativa e utilização da modelagem *fuzzy* para o presente objetivo. Inicialmente, Weon e Kim (2001) discutiram sobre um método de avaliação de desempenho do discente baseado nas variáveis difusas: a dificuldade, a importância e complexidade das questões. Contudo, este estudo apresentou alguns pontos negativos como a subjetividade na atribuição dos pesos e o fato de o contexto de validação ter sido executado em uma turma do 4º ano do ensino fundamental.

No ano de 2008, Bai e Chen (2008) tentaram solucionar a subjetividade associada ao trabalho anterior (WEON e KIM, 2001) por meio do uso da Matriz de Acertos e Matriz de Tempo, além da Matriz de Peso, da Matriz de Importância e da Matriz Complexidade. Novamente, essa abordagem utilizou as funções de pertinência para lidar com as seguintes variáveis linguísticas: “Baixo”, “Baixo-Médio”, “Médio”, “Médio-Alto” e “Alto”. No entanto, o método proposto pelo estudo ainda apresentou problemas de subjetividade, pois os resultados permaneceram dependentes dos diversos pesos atribuídos para cada fator de ponderação, os quais foram determinados pelo especialista.

Diante das limitações do estudo de Bai e Chen (2008), Saleh e Kim (2009) apresentaram uma alternativa melhorada ao método por meio de um sistema de avaliação *fuzzy*, considerando os mesmos critérios (importância, dificuldade e complexidade das questões) e baseando-se no método Mamdani na inferência *fuzzy* (MAMDANI e ASSILIAN, 1975) e Centro de Gravidade na defuzzificação. Com isto, a transparência e a razoabilidade do sistema de lógica *fuzzy* permitiram que as saídas fossem facilmente interpretadas e explicadas aos estudantes.

Outra proposta de melhoria ocorreu no trabalho de Chen *et al.* (2011). Os autores apresentaram um novo método para avaliar o sucesso da aprendizagem por meio da geração automática dos pesos dos atributos dos estudantes, sendo essas: taxa de precisão, taxa de tempo, dificuldade e complexidade, custo da resposta e importância. Assim, o método proposto foi mais sensível para definição do sucesso da aprendizagem dos alunos, pois este podia fornecer resultados mais ajustados e razoáveis à aprendizagem do que a avaliação proposta pelo método de Saleh e Kim (2009). Os autores perceberam que quando as dificuldades foram as mesmas e a média do desempenho dos discentes variou entre 0,84 e 0,78, o método de Saleh e Kim (2009) não foi sensível o suficiente para refletir a mudança do sucesso da aprendizagem dos alunos. Concomitantemente, os mesmos exemplos aplicados na nova validação demonstraram maior sensibilidade do estudo para refletir o desempenho dos alunos na aprendizagem do que o método de Saleh e Kim (2009).

Por fim, o último dos trabalhos citados no grupo é a pesquisa de doutorado de Arias *et al.* (2012), que se baseia na multidimensionalidade da avaliação do desempenho para lidar com a complexidade do processo. Neste contexto, o modelo proposto contemplou, novamente, as três dimensões citadas pelo método de Weon e Kim (2001): dificuldade, complexidade e importância das questões. Além disto, o modelo comentado é concebido com base na metodologia de Saleh e Kim (2009), adotando o mesmo método de fuzzificação (triangular), inferência (Mamdani) e defuzzificação (centro de gravidade).

Para finalizar as análises destas abordagens, no Quadro 2.5 é apresentada uma analogia entre estudos anteriormente citados e as etapas do processo de investigação *fuzzy*. As informações apresentadas na presente Seção resultam na construção do Quadro 2.6 que resumiu as informações mais importantes sobre as publicações multidimensionais discutidas e catalogadas na pesquisa.

Por fim, a RS sistemática ainda possibilitou uma análise sobre os termos mais mencionados no tocante aos trabalhos relacionados à temática de avaliações multidimensionais. Na Figura 2.5 são apresentadas as palavras-

chave mais utilizadas nos trabalhos que foram classificados com a temática. Os termos mais mencionados foram: *Student's evaluation*, *fuzzy grading systems* ou *Fuzzy reasoning* e *Response Accuracy* ou *Adaptive feedback*.

Quadro 2.5. Quadro resumo dos métodos fuzzy.

	Fuzzificação	Inferência	Defuzzificação
Weon e Kim (2001)	-	Funções de Avaliação	-
Bai e Chen (2008)	Método Triangular	Máximos e Pesos	Valores de Defuzzificação
Saleh e Kim (2009)	Método Triangular	Mamdani	Centro de Gravidade (CoG)
Arias et al. (2012)	Método Triangular	Mamdani	Centro de Gravidade (CoG)
Kao, Lin e Chu (2012)	Método Trapezoidal	Mamdani	Centro de Gravidade (CoG)
Dias e Diniz (2013)	Método Trapezoidal	Mamdani	Centro de Gravidade (CoG)
Chen et al. (2011)	Método Triangular	Mamdani	Centro de Gravidade (CoG)
Arias et al. (2013)	Método Triangular	Mamdani	Centro de Gravidade (CoG)

Fonte: Próprio Autor.

Quadro 2.6. Quadro resumo das avaliações multidimensionais.

Artigo	Objetivo	Método	Contexto	Métricas De Análise	Avaliação
Daif e Rizkas (2013)	Diminuir as taxas de evasão de cursos, ajudando instrutores a monitorar e incentivar seus alunos.	Geração de Relatórios JIT baseados no cálculo de métricas pré-definidas	Validado no curso <i>online</i> da Academia Árabe de Ciência, Tecnologia e de Transportes Marítimos.	5 índices de monitoramento: a) frequência, b) Tarefas e submissão, c) provas, d) Questionário, e) participação.	Formativa
Arias et al. (2012)	Adotar a multidimensionalidade na avaliação do desempenho do aluno	Lógica difusa (<i>Fuzzy</i>)	Validado na Escola Aletheia, rede privada de Campinas.	Três controladores <i>fuzzy</i> : Dificuldade, esforço e ajuste.	Somativa
Saleh e Kim (2009)	Propor um método de avaliação de roteiros de resposta dos alunos por meio de um sistema <i>fuzzy</i> .	Lógica difusa (<i>Fuzzy</i>)	Não foi validado neste estudo	Três controladores <i>fuzzy</i> : dificuldade, complexidade e importância.	Formativa

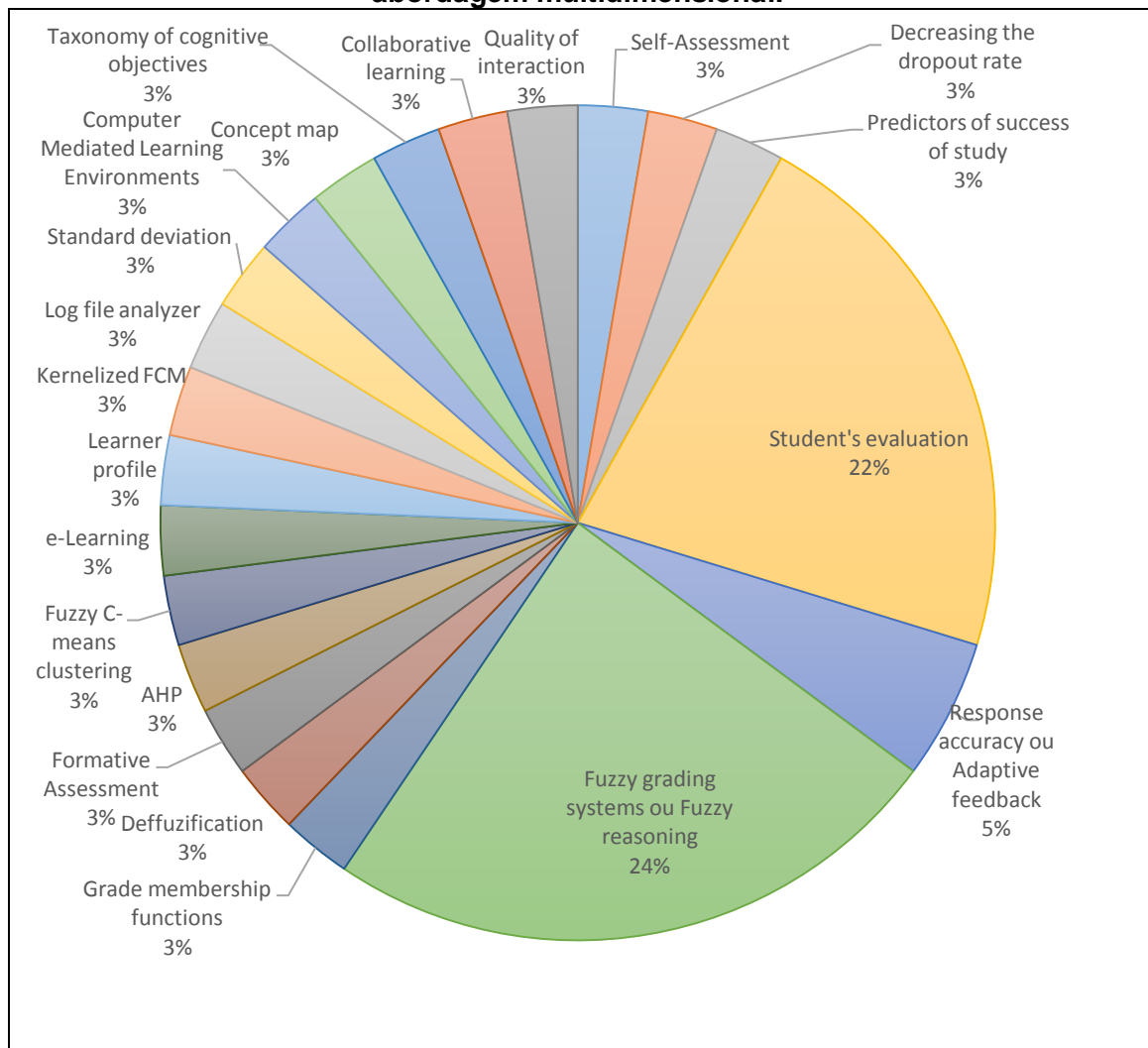
Continua

Continuação

Artigo	Objetivo	Método	Contexto	Métricas De Análise	Avaliação
Dias e Diniz (2013)	Estima a qualidade da interação de alunos e professores no ambiente, propôs uma ferramenta a <i>FuzzyQol</i>	Lógica difusa (<i>Fuzzy</i>)	Validado em 5 cursos de graduação da Universidade de Lisboa	Três controladores <i>Fuzzy</i> de entrada: View, Addition e Alterarion.	Formativa
Kao, Lin e Chu (2012)	Propor uma abordagem multidimensional de avaliação para prover personalização aos sistemas de <i>feedback</i>	Lógica difusa (<i>Fuzzy</i>)	Experimento envolvendo 52 alunos matriculados em um curso de construção de compiladores em uma universidade de Taiwan.	Quatro fatores para o diagnóstico: a) taxa de precisão b) dificuldade teste c) nível de confiança d) Tempo de resposta	Formativa e Diagnóstica
Hogo (2010)	Classifica os alunos em categorias específicas com base em seu perfil, são eles: regulares, trabalhadores, casuais, ruins e ausentes.	<i>Fuzzy C-means</i>	Duas turmas do segundo período de curso de Ciência da Computação da universidade de Saint Mary	a) acesso no campus (ou não) - valores binários 0 ou 1 b) tempo de acesso c) acessos feitos nos laboratórios presenciais c) Número de visitas (valores decimais). d) Downloads (valores decimais).	Formativa
Guixia e Liyan (2010)	Construir um sistema indexado de avaliação do efeito da aprendizagem	AHP e Lógica difusa	Curso de Inglês da Faculdade EaD	a) avaliação subjetiva b) avaliação objetiva c) registro <i>online</i>	Formativa
Weon e Kim (2001)	Apresenta um método de avaliação de desempenho do discente	Lógica difusa (<i>Fuzzy</i>)	Estudantes do 4º ano do ensino fundamental	Variáveis difusas: a dificuldade, a importância e complexidade das questões.	Somativa
Bai e Chen (2008)	Objetiva avaliar o desempenho dos alunos por meio de entradas difusas.	Lógica difusa (<i>Fuzzy</i>)	Curso <i>online</i> composto por 5 questões centrais nas interações.	Variáveis difusas: a dificuldade, a importância e complexidade das questões.	Somativa
Chen e Lin (2011)	Propôs um novo método para avaliar o sucesso da aprendizagem por meio da geração automática dos pesos dos atributos dos estudantes	Lógica difusa (<i>Fuzzy</i>)	--	taxa de precisão, taxa de tempo, dificuldade e complexidade, custo da resposta e importância.	Somativa

Fonte: Próprio Autor.

Figura 2.5. Visão geral das palavras chave mais comentadas em trabalhos com a abordagem multidimensional.



Fonte: Próprio Autor.

Os resultados das análises sobre as palavras chave foram apresentados na língua inglesa porque as chaves de busca na língua inglesa obtiveram os melhores resultados nesta RS. Há algumas pesquisas nacionais na lista de trabalhos selecionados, mas mesmo estes artigos apresentaram listas de palavras chave no idioma inglês ou foram publicados no idioma.

2.4 Considerações Finais

No presente capítulo, são descritas as principais características sobre os trabalhos relacionados à tese. Além disto, resalta-se que há diversas lacunas que precisam ser pesquisadas na área da avaliação em cursos *online*, principalmente no combate à evasão. Esta realidade é corroborada pelos

elevados índices de evasão em cursos desta modalidade no Brasil e no mundo. Outra conclusão obtida é a presença de pesquisas voltadas à avaliação dos discentes na EaD. Contudo, estes esforços ainda são muito concentrados na avaliação pontual dos recursos pedagógicos e do desempenho do aluno. Este tratamento pode ser equivocado ao se lidar com questões mais complexas como evasão e reprovação de alunos.

Neste sentido, o presente capítulo tem o intuito de guiar o leitor a identificar as potencialidades da avaliação multidimensional para combater as problemáticas destacadas em cursos *online*, principalmente a evasão. Assim, é importante finalizar esta discussão ressaltando os aspectos que mais se destacam para o presente documento. Inicialmente, a forte presença da inferência *fuzzy* nas avaliações multidimensionais pode ser encarada como um indício da força da incerteza associada a este processo. Outro fator importante é a inexistência de trabalhos que ajudem o professor a entender as saídas dos modelos propostos na literatura, normalmente é necessário um conhecimento específico e este aspecto que pode ser uma das causas para o uso insuficiente de ferramentas nos AVEA. Além disto, merece destaque a não identificação de sistemas de avaliação relacionados ao delineamento do perfil de evasão dos discentes baseados em lógica difusa, dada a incerteza associada a esta investigação.

Outra carência detectada, durante a revisão da literatura, foi de estudos que incorporassem extração de informações, a partir dos textos produzidos pelos alunos no combate à evasão. Análises como mineração de texto das mensagens poderiam prover entradas adequadas aos modelos sugeridos.

Há algumas abordagens avaliativas que poderiam ser aplicadas no contexto do IFPB Virtual, porém o presente trabalho quis investigar se as avaliações multidimensionais poderiam ser mais eficientes frente à problemática da evasão discente, comparando a precisão da abordagem proposta com as apresentadas na literatura.

3. Fundamentação Teórica

Neste capítulo, são apresentados os elementos conceituais mais relevantes ao entendimento do documento em sua totalidade. Alguns conceitos importantes são elucidados: lógica *fuzzy*, árvores de decisão, mineração de texto e aspectos pedagógicos relacionados à solução apresentada no presente documento.

3.1 Lógica *Fuzzy*

No Capítulo 2, a relevância dos trabalhos que lidam com a subjetividade nas avaliações é um aspecto importante a ser destacado. Outro aspecto relevante é a estratégia da aplicação da lógica *fuzzy* para lidar com tal cenário de incerteza. Esta e outras razões são motivadoras para incluir a lógica *fuzzy* na abordagem apresentada. Neste sentido, a lógica difusa torna-se adequada a sistemas adaptativos principalmente para servir de interface de representação do usuário, etapa imprescindível do mecanismo de adaptação e avaliação.

Ao contrário da lógica convencional, a lógica *fuzzy* utiliza a ideia de que tudo admite graus de pertinências, por exemplo: temperatura, pressão *etc.* Com isto, a lógica *fuzzy* tenta modelar o senso de palavras, tomada de decisão ou senso comum do ser humano. A teoria de conjunto *fuzzy*, também conhecido como nebuloso ou difuso, é um meio para especificar o quanto um objeto satisfaz uma descrição vaga. Por exemplo, considerando a proposição: “Maria é participativa na aula”. Isto é verdade se Maria participou dos exercícios e faltou as atividades em grupo? A maioria das pessoas responderia “sim” ou “não”, preferindo dizer um “talvez”. Neste caso, a situação destacada é um exemplo clássico de uma questão de incerteza. O termo linguístico “participativa” não se refere a uma demarcação nítida de objetos em duas classes – existem graus de participação.

Neste contexto, a lógica *fuzzy* surge como um método de raciocínio com expressões lógicas que descrevem a pertinência em conjuntos difusos. A ideia dos graus de pertinência possibilita agrupar os elementos de maneira

diferente da aplicada na lógica clássica, o que permite reinterpretar antigos conceitos, elaborados segundo esta lógica (RUSSELL e NORVIG, 2013). A presente teoria difusa é composta por diversas subteorias: teoria dos conjuntos *fuzzy*, sistemas dinâmicos *fuzzy*, agrupamentos *fuzzy*, tomada de decisão *fuzzy*, teoria das probabilidades *fuzzy*, etc.

A teoria base define que um conjunto *fuzzy* A definido no universo de discurso X é caracterizado por uma função de pertinência μ_A , que mapeia os elementos de X para o intervalo $[0,1]$.

$$\mu_A(x): X \rightarrow [0,1] \quad (1)$$

Quanto à representação de um conjunto *fuzzy*, se esse for discreto, pode-se enumerar seus elementos juntamente com seus graus de pertinência, na forma:

$$A = \frac{\sum \mu_A(x_i)}{x_i} \quad (2)$$

em que a somatória se refere à operação união (disjunção) e a notação $\mu_A(x_i)/x_i$ se refere ao elemento x_i que pertence ao conjunto *fuzzy* A com grau $\mu_A(x_i)$. Deste modo, um conjunto *fuzzy* A é uma coleção de pares:

$$A = \{(x, \mu(x)) \mid x \in X\} \quad (3)$$

O exemplo dado anteriormente utiliza faixas para categorizar as métricas em “Participativo” e “Não Participativo” para avaliar a participação do aluno. Estas denominações cujo valor é expresso qualitativamente por um termo linguístico e quantitativamente pela sua função de pertinência são denominadas variáveis linguísticas. A variável linguística é composta, portanto, por uma variável simbólica e por um valor numérico. Os termos linguísticos são dados para expressar conceitos e conhecimentos na comunicação humana, principalmente em muitas áreas em que é importante quantificar e qualificar tais dados (SIDDIQUE e ADELI, 2013; LIMA, PINHEIRO e SANTOS, 2014).

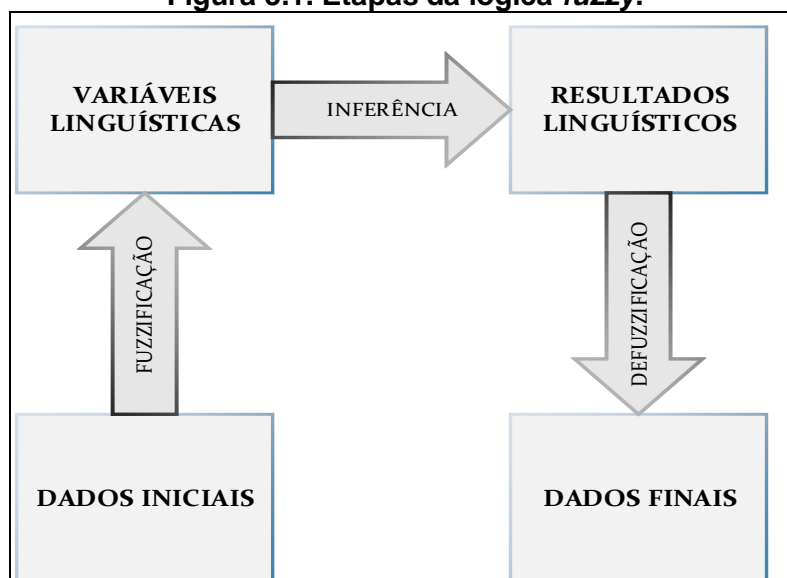
Para reconhecer uma transição com relativa dose de ambiguidade, dicotomias do tipo pertence - não pertence, verdadeiro – falso do conjunto clássico, o conjunto *fuzzy* determina que se considere a maior ou menor pertinência do valor da variável. Os valores da função de pertinência são

indicadores de tendências atribuídas subjetivamente por alguém, e dependem do contexto no qual estão inseridos.

Outro conceito essencial é o reconhecimento e caracterização das variáveis que compõem uma base de regras. Isto se estabelece a partir do conhecimento do especialista na forma de proposições do tipo “se E1 e/ou E2 então S1”, em que E1 e E2 referem-se a classificações linguísticas das variáveis de entrada e S1 a de saída (HUDEC, 2016).

De maneira geral, o sistema lógico *fuzzy* consiste em três operações principais: fuzzificação, inferência e defuzzificação. Na Figura 3.1, adaptada de Cox (1994), são ilustradas tais etapas. Inicialmente, na etapa de fuzzificação, o problema deve ser analisado e os dados de entrada transformados em variáveis linguísticas. Neste momento, é de extrema importância que todos os dados de incerteza sejam analisados e transformados em variáveis linguísticas. Após esta transformação, são determinadas também as funções de pertinência. No segundo momento, a etapa de inferência deve ser iniciada, considerando que na etapa anterior os dados de entrada já foram transformados em variáveis linguísticas. Nesta fase, são criadas as regras ou proposições por meio da associação das variáveis já definidas. Por fim, a etapa de defuzzificação envolve a conversão dos valores *fuzzy* em números reais, tendo assim um conjunto de saída matematicamente definido.

Figura 3.1. Etapas da lógica *fuzzy*.



Fonte: Adaptado de Cox (1994).

As funções de pertinência podem ter diferentes formas, dependendo do conceito que se deseja representar e do contexto em que foram utilizadas. Algumas características destas funções são: formato (triangular, trapezoidal, gaussiana, sigmoidal, *etc*); forma de obtenção (escolhidas pelos usuários baseadas em sua experiência ou por meio de um processo de otimização a partir de dados experimentais e/ou obtidas por simulação) e normalização (geralmente as funções de pertinência são definidas no intervalo $[0,1]$) (RUSSELL e NORVIG, 2013).

Durante avaliações subjetivas, a lógica *fuzzy* permite que as pesquisas lidem com uma grande variedade de informações vagas e incertas, as quais são traduzidas em expressões do tipo: a maioria, mais ou menos, talvez, *etc.*, ou seja, traduzindo a tentativa de modelar os termos e as decisões envolvidas na tarefa. Assim, estes fatores podem justificar a aplicação da técnica *fuzzy* em diversos trabalhos encontrados na revisão de literatura e apresentados no Capítulo 2.

3.2 Árvores de Decisão

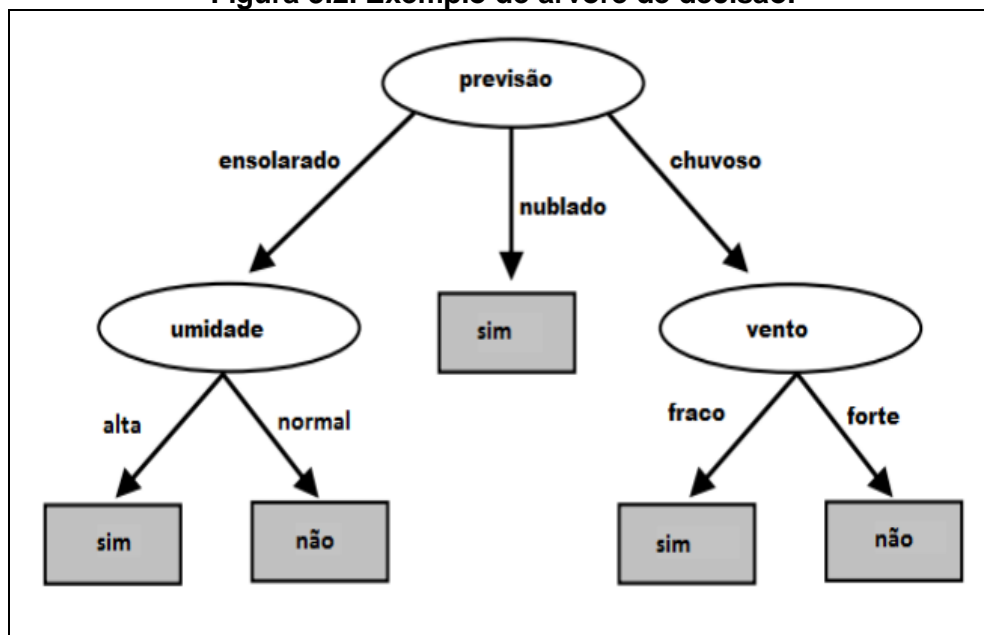
Existem muitos algoritmos de aprendizado que induzem classificadores. Nesta Seção, são descritos indutores que contribuem e priorizam o entendimento dos dados. Por exemplo, no desenvolvimento de sistemas especialistas é primordial que especialistas humanos possam verificar de modo rápido e fácil o conhecimento extraído e relacioná-lo ao seu próprio domínio de conhecimento.

Uma das técnicas utilizadas neste contexto são as árvores de decisão. Entende-se por árvores de decisão as estruturas formadas por um conjunto de nós de decisão, perguntas, que permitem a classificação de cada caso (RUSSEL e NORVIG, 2013). Uma árvore de decisão é definida recursivamente como (NILSSON, 2014):

- um nó folha que corresponde a uma classe ou
- um nó de decisão que contém um teste sobre algum atributo. Para cada resultado do teste existe uma aresta para uma subárvore.

Na Figura 3.2, é ilustrado uma árvore de decisão para a previsão de chuva. Cada elipse é um teste em um atributo para um conjunto de dados sobre o tempo e cada retângulo representa um resultado (se choverá ou não). Para diagnosticar (classificar) o tempo, basta começar pelo nó raiz, seguindo cada teste até que um nó folha seja alcançado.

Figura 3.2. Exemplo de árvore de decisão.



Fonte: Ken et al. (2010).

Outra relevância das árvores de decisão, é que sua facilidade de compreensão contribui para o entendimento do domínio considerado, fato que pode produzir um conhecimento novo para o especialista. O tempo de geração do modelo é menor em comparação ao de outras técnicas e, na maioria das vezes, não necessita de informações adicionais.

As árvores de decisão são modelos estatísticos, que podem ser usados com o objetivo de classificar objetos. Neste tipo de aplicação, a partir de um conjunto de instâncias, cujas classes são conhecidas, após a fase de treinamento, uma árvore de decisão é capaz de classificar novas instâncias. As árvores de decisão foram utilizadas para auxiliar na classificação e automatização no processo de geração de regras *fuzzy*.

Há diversas técnicas que podem ser utilizadas, sendo a pioneira delas a ID3. O algoritmo ID3, cuja criação se baseou em sistemas de inferência e em

conceitos de sistemas de aprendizagem, foi um dos primeiros de árvore de decisão. Logo após, foram elaborados diversos outros algoritmos, sendo os mais conhecidos o C4.5, o J48, o CART (*Classification and Regression Trees*) e o CHAID (*Chi Square Automatic Interaction Detection*), além de outros.

3.3 Mineração de Texto

A literatura investigada destaca a forte presença de recursos pedagógicos textuais nas atividades propostas ao longo dos cursos *online*, são eles: fóruns, bate-papos, questionários, glossário, *wiki etc.* Além disto, tais ferramentas são importantes na comunicação e agrupam diversas informações, dentre as quais opiniões e sentimentos expressos por seus usuários em simples conversas, *feedbacks* e mensagens.

Neste sentido, para entender melhor e extrair informações pertinentes dos textos, na pesquisa ora descrita foi utilizada a análise de polaridade, cujo principal objetivo é definir técnicas automáticas capazes de extrair informações subjetivas de textos no formato de linguagem natural, como opiniões e sentimentos. Além disto, a técnica mencionada almeja criar um conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão.

A identificação da polaridade dos textos (mensagem positiva ou negativa) é uma das áreas de pesquisa mais destacadas em Processamento de Linguagem Natural desde o início dos anos 2000 (LIU, 2010). Opiniões em sistemas computacionais, se devidamente extraídas e analisadas, permitem não só entender e explicar diversos fenômenos complexos, mas também os prevê (BENEVENUTO, RIBEIRO e ARAÚJO, 2015).

Há muitos contextos que têm utilizado técnicas de mineração de texto, por exemplo: (i) no auxílio a empresas do setor energético na análise dos impactos das ações empreendidas ao longo do tempo por meio de opiniões contidas nas mídias sociais (ALVES *et al.*, 2015); (ii) no mercado acionista, por meio da identificação do humor do mercado baseado nas opiniões de especialistas (O'HARE *et al.*, 2009); (iii) nas opiniões dos consumidores sobre produtos ou serviços (EIRINAKI, PISAL e SINGH, 2012; HU e LIU, 2004); (iv)

em aplicações de turismo, por meio da investigação dos comentários de viajantes (BJØRKELUND, BURNETT e NØRVK , 2012) e (v) na análise de políticos e de política (FANG *et al.*, 2012).

Há diversas técnicas sendo utilizadas para análise de polaridade. Nesta Seção, são destacadas duas abordagens importantes para o problema de extração da polaridade dos textos. Os métodos de detecção de polaridade em sentenças são divididos em dois grandes grupos: os baseados em aprendizagem de máquina (a partir de técnicas supervisionadas) e os métodos léxicos (a partir de técnicas não supervisionadas) (BENEVENUTO, RIBEIRO e ARAÚJO, 2015).

As estratégias baseadas em aprendizagem de máquina dependem de uma base de dados rotulada para a etapa de treinamento dos classificadores (NILSSON, 2014), o que pode ser um ponto negativo, devido ao custo da rotulação dos dados observados. Esta abordagem é considerada uma técnica supervisionada por exigir uma etapa de treinamento de um modelo com amostras previamente classificadas. Benevenuto, Ribeiro e Araújo (2015) destacaram que a execução da aprendizagem de máquina é composta por quatro etapas principais: 1 - obtenção de dados rotulados que (utilizados no treinamento e teste); 2 - definição das *features* (características) que permitam a distinção entre os dados; 3 - treinamento de um modelo computacional com um algoritmo de aprendizagem; 4 - aplicação do modelo. Apesar da necessidade de rotulação dos dados, as técnicas supervisionadas são mais populares na literatura (SHARMA e DEY, 2012; FELDMAN, 2013).

O outro grupo de métodos são os métodos léxicos que utilizam listas e dicionários de palavras associadas a grupos específicos. Esta abordagem se caracteriza como uma técnica não supervisionada e, diferentemente das supervisionadas, não necessitam de sentenças previamente rotuladas e treinamento para criação de um classificador. Apesar de não depender de dados rotulados para treinamento, a eficiência dos métodos léxicos está diretamente relacionada à generalização do vocabulário utilizado, para os diversos contextos existentes. Diversos trabalhos apresentam bons resultados com a aplicação de tal método (CHAVES *et al.*, 2012; SARMENTO *et al.*, 2009; TUMITAN e BECKER; 2013).

Um dicionário é usualmente composto por uma lista de palavras (*features*), contendo parte de todas as *features* do assunto abordado, relativo à língua adotada. Para cada *feature* identificada deve-se associar um valor que indique o grau de relação entre a palavra e os documentos. Há diversas técnicas que podem ser usadas, tais como: peso booleano, peso pela frequência, peso tf^1 - idf^2 , peso lrc^3 e peso baseado em entropia.

3.4 Métricas Pedagógicas

A EaD como uma prática educativa busca aproximar o saber do aprendiz do conhecimento almejado principalmente via estratégias de ensino construtivistas (VASCONCELOS *et al.*, 2015). Neste sentido, o AVEA permite que os saberes sejam concebidos pelo aprendiz em cada uma das situações com as quais ele encontrará durante o processo de aprendizagem. Assim, um dos pontos fortes do construtivismo é o fato da realidade ser abordada sob várias perspectivas possibilitando o entendimento do aprendiz sobre a temática. São por tais razões que a presente tese se subsidia em premissas da metodologia construtivista de ensino no planejamento de algumas etapas do processo, por exemplo: a definição das métricas de entrada da abordagem.

Em relação à revisão da literatura apresentada no Capítulo 2, observa-se que uma série de fatores comportamentais do discente é destacada como relevante no combate à evasão.

A partir dos conceitos apresentados, torna-se relevante descrever o que significa cada um destes conceitos frente às teorias da Psicologia Cognitiva e da Pedagogia, para facilitar o entendimento e a extração de tais informações nas etapas futuras. Dentre estes fatores, mereceram destaque: Autorregulação, Interação e Motivação.

3.4.1 Autorregulação

Diante da evolução do processo de ensino-aprendizagem, constata-se que o paradigma educacional mudou ao longo dos anos, de acordo com dois

¹ *tf* – term frequency

² *idf* – inverse document frequency

³ *lrc* – logarithm term component

aspectos estruturais: a exigência de um papel ativo do aluno na construção do conhecimento e o ambiente tecnológico que está associado às atividades pedagógicas (DE MATOS, 2014). No contexto atual, o processo de ensino-aprendizagem tem acentuado a construção do conhecimento por parte do aluno, realçando o seu papel ativo no controle do seu aprendizado. Para o desenvolvimento da autonomia discente, são necessários níveis elevados na autorregulação da aprendizagem, pois o aluno reflete sobre sua aprendizagem à medida que se envolve metacognitiva, motivacional e comportamentalmente no processo de aprendizagem (ZIMMERMAN, 2008). Assim, é possível afirmar que a EaD é sistematicamente organizada para propiciar a independência do aluno no estudo.

Em relação à tecnologia relacionada aos ambientes educacionais, este aspecto tem possibilitado o desenvolvimento de inúmeras ferramentas e aplicações interativas, o que torna fundamental ao aluno autorregular sua aprendizagem (AZEVEDO, 2011). A utilização de recursos pedagógicos que se tornem realmente eficientes deve permitir que o aluno assuma o controle do processo, por exemplo, definição dos objetivos, planificação da sua aprendizagem e monitoramento do seu progresso (WEERASINGHE, DU BOULAY e BISWAS, 2013), ou seja, os alunos devem ser convocados a autorregular sua aprendizagem.

É possível destacar algumas definições sobre a autorregulação da aprendizagem na literatura. Segundo Perrenoud (1999), a autorregulação faz menção às “capacidades do sujeito para gerenciar ele próprio seus projetos, progressos, estratégias diante das tarefas e desafios”. Esta característica está presente em todos os discentes em diferentes níveis de autorregulação. Além disto, é fundamental perceber que em processos de aprendizagem é importante que este seja elevado, o que certamente favorece uma autonomia progressiva no processo do aprender.

Neste mesmo sentido, Zimmerman (2000) corrobora com Perrenoud (1999) afirmando que a autorregulação da aprendizagem pode ser definida como “qualquer pensamento, sentimento ou ação desenvolvida e orientada pelos próprios alunos para a execução dos seus objetivos”. Enquanto isto, Silva *et al.* (2004) sugerem que a aprendizagem regulada pelo aluno resulta da

interação de conhecimentos, competências e motivações, que são aspectos fundamentais ao planejamento, organização e avaliação dos processos adotados.

Em um sentido amplo, a aprendizagem pode ser definida como autorregulada quando o aluno se sente livre para influenciar sobre o que, quando, onde e como aprender (WEINERT, 1982). Neste sentido, alunos autorreguladores da sua aprendizagem desenvolvem a consciência do que podem ou não realizar, do que entendem e do que não entendem de modo a encontrar recursos para atingir seus objetivos definidos (ROSÁRIO *et al.*, 2010).

Para realizar uma tarefa com sucesso, os alunos procuram a informação que precisam, percorrendo os passos necessários para a sua execução, mesmo quando encontram obstáculos (ZIMMERMAN, 2008). Os alunos autorreguladores da sua aprendizagem procuram desafios e superam obstáculos umas vezes com persistência outras vezes com resoluções inovadoras dos problemas. Na sua aprendizagem, estabelecem objetivos realistas, e recorrem a uma bateria de recursos, abordando as tarefas acadêmicas com confiança.

A aprendizagem autorregulada é encarada como um mecanismo para ajudar a explicar as diferenças de desempenho entre os alunos e como melhorar suas capacidades para resolução de problemas (SIEMENS, 2005). A autoaprendizagem está intimamente relacionada ao resultado acadêmico. Barnard-Brak, Lan e Paton (2010) observaram que os alunos com habilidades de aprendizagem mais autorreguladas têm resultados de aprendizagem acadêmica mais positivos em relação aos seus homólogos que não demonstram comportamentos de aprendizagem autorregulados.

A revisão de literatura mostra ainda que diversas pesquisas atestaram que discentes bastante autorregulados utilizam atividades de controle com mais frequência do que aqueles com pouca autorregulação (DIAS; LEITE, 2010; FRISON; MORAES, 2010; ROSÁRIO *et al.*, 2010; BERGAMIN *et al.*, 2012). Diante deste esforço de se autorregular, o discente naturalmente desenvolve comportamentos autorregulatórios. Tais comportamentos representam capacidades e estratégias que os alunos utilizam em função de

um objetivo individual para obter sucesso no processo pedagógico.

De forma resumida, diversos estudos na literatura indicam que as características do discente afetam diretamente a aprendizagem quando esta ocorre em um AVEA (DIAS e LEITE, 2010; BERGAMIN *et al.*, 2012; CAVANAUGH, LAMKIN e HU, 2012). Assim, os pesquisadores procuraram identificar quais variáveis dos discentes afetam os resultados da aprendizagem na EaD e a aprendizagem autorregulada tem sido considerada elemento chave para o sucesso nessa modalidade de ensino (PAVESI e ALLIPRANDINIA, 2015; BARNARD-BRAK, LAN e PATON, 2010; DIAS e LEITE, 2010; BERGAMIN *et al.*, 2012).

Com a intenção de descrever e analisar os processos existentes no processo de aprendizagem autorregulada, Pintrich (2004) citou quatro fases centrais: planejamento e ativação, monitorização, controle/regulação e avaliação. Apesar de serem fases de execução sequencial, estas não foram hierarquicamente organizadas, podendo ocorrer de forma simultânea e integrada ao longo de processo. Como pode ser observado no Quadro 3.1, em cada uma destas fases, há atividades de autorregulação nas áreas cognitiva, motivacional, comportamental e de contexto. Neste mesmo sentido, Zimmerman (2000) pontua a presença de três tipos de autorregulação: a comportamental, a ambiental e a interna.

É relevante destacar que, independente da abordagem, o presente trabalho foca na área comportamental da avaliação sob enfoque do contexto durante a extração de métricas relacionadas à autorregulação.

Sob a perspectiva da autorregulação, pode-se agrupar uma turma *online* em dois grandes grupos, a saber: os alunos autorregulados e os alunos não autorregulados. Entende-se por alunos autorregulados aqueles que são conscientes do processo de aprendizagem e que o controlam, de forma a selecionar métodos e estratégias mais eficientes. Estes estruturam melhor seu contexto de estudo, identificando situações em que necessitam de um apoio extra ou outras fontes de estudo, adaptando as estratégias de aprendizagem a seus objetivos pedagógicos no curso (RIBEIRO e SILVA, 2007).

Quadro 3.1. Fases e áreas da aprendizagem autorregulada.

Fases	Cognição	Motivação	Comportamento	Contexto
Fase 1 Planejamento e Ativação	Estabelecimento de metas. Ativação de conhecimento prévio relevante. Ativação de conhecimento metacognitivo.	Adoção de orientação a meta. Crenças de auto-eficácia. Ativação das crenças de valor da tarefa Ativação de interesse pessoal. Afetos/ emoções.	Planejamento do tempo e do esforço	Percepção da tarefa. Percepções do contexto
Fase 2 Monitorização	Consciência metacognitiva e auto-observação da cognição	Consciência e monitorização da motivação	Consciência e monitorização do esforço, uso do tempo, necessidade de ajuda. Auto-observação do comportamento	Monitorização das condições da tarefa e contextuais
Fase 3 Controle / Regulação	Seleção e uso de estratégias cognitivas e metacognitivas para a aprendizagem	Seleção e adaptação de estratégias para direcionar a motivação	Fortalecimento ou enfraquecimento do esforço	Alterações nos requisitos da tarefa Alterações no contexto
Fase 4 Reação e Reflexão	Julgamentos cognitivos Atribuições	Reações afetivas Atribuições	Alteração de comportamento: persistir, abandonar. Busca de ajuda	Avaliação da tarefa e do contexto

Fonte: Adaptado de Polydoro e Azzi (2009).

3.4.2 Interação

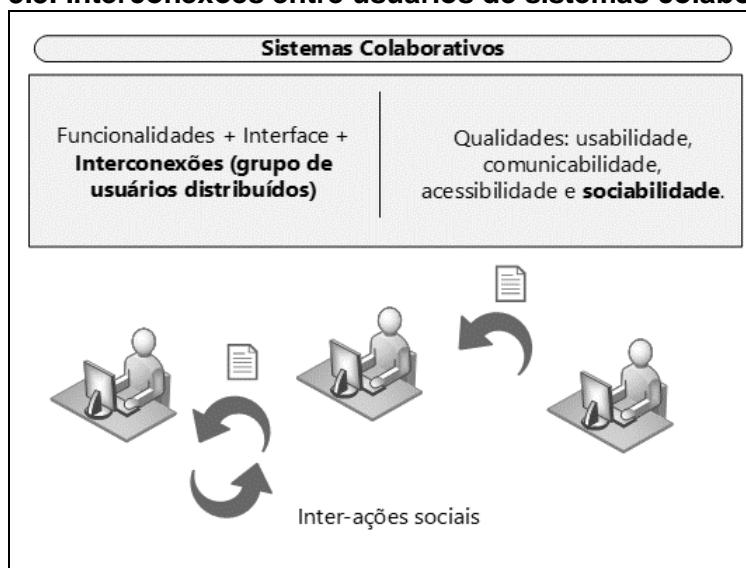
Apesar da aparente simplicidade, a interatividade é um conceito complexo, especialmente quando relacionado à área da Educação, além de representar um importante pilar para o entendimento da presente tese. No contexto da EaD, a complexidade é ainda maior, pois diversos agentes interagem de diferentes maneiras, por meio de inúmeras ferramentas e com expectativas e objetivos distintos. Neste sentido, é importante distinguir dois conceitos aparentemente semelhantes neste cenário: interação e interatividade. Inicialmente, a interação engloba o comportamento e as trocas entre indivíduos e grupos que se influenciam, nos casos em que há eventos recíprocos que requerem pelo menos dois objetos e duas ações (PRATES, 2012; LITTO e FORMIGA, 2009). A interatividade se refere aos atributos da tecnologia contemporânea utilizada na EaD, que permite conexão em tempo real. A

interação estaria, portanto, associada às pessoas enquanto a interatividade à aos canais de tecnologia. Não é objetivo da pesquisa discutir essa querela conceitual, mesmo porque a tendência é a de que essa se perpetue e seja periodicamente retomada quando necessário no documento.

A epistemologia genética de Piaget e o socioconstrutivismo de Vygotsky são considerados teorias interacionistas (PIAGET, 2013; DANIELS, 2016). Além disto, é por meio de interações que os seres humanos se desenvolvem e aprendem. Neste sentido, um caminho relevante para discutir a questão da interatividade na EaD é avaliar os níveis de interatividade de diferentes atividades realizadas em um curso.

Diversos ramos da computação podem ser envolvidos para uma análise das interações em uma comunidade *online*, tais como: comunicabilidade, usabilidade, acessibilidade e sociabilidade (Figura 3.3). A pesquisa foca suas análises nas questões relacionadas à comunicabilidade e qualidade da interação. Assim, a análise das interações aborda uma série de relações entre discentes, docentes e conteúdo nos AVEA, como: relacionamento aluno-professor, entre alunos e aluno-conteúdo (MOORE, 1989). Anderson (2003) ampliou a perspectiva de Moore, incluindo mais três tipos de interação: professor-professor, professor-conteúdo e conteúdo-conteúdo.

Figura 3.3. Interconexões entre usuários de sistemas colaborativos.



Fonte: Adaptado de Prates (2012).

De acordo com Moore (1989), a interação aluno-professor pode ser do tipo síncrona ou assíncrona e, normalmente, tem o intuito de prover motivação, por meio de *feedbacks*, e auxiliar o processo de aprendizado do discente. Yacci (2000) ressaltou a importância do *feedback* para a interação: sem o que ela chama de loop interativo, não se concretizaria a interatividade: “Estruturalmente a interatividade é um circuito de mensagens que flui de uma entidade originadora a uma entidade-alvo e, então retorna à entidade originadora”. Portanto, o retorno seria uma condição necessária para a interação. Uma pergunta do professor respondida pelo aluno não chega a ser interativa, pois não se detecta uma mensagem de *feedback* do professor. A interatividade deve ser investigada sob o ponto de vista do aluno na EaD. Yacci (2000) destacou os efeitos negativos do *feedback* atrasado por parte dos professores. Em geral, os alunos se desmotivam ao ler o que o professor escreveu, pois há um *timing* certo ou máximo para a resposta, depois do qual a interação não se completa. Quando o *feedback* demora muito, o objetivo original da mensagem já foi esquecido pelo aluno.

Outro tipo de interação comentada é a relação aluno-conteúdo, pois os AVEA permitem o desenvolvimento dos conteúdos e objetos de aprendizagem de diversas formas: som, texto, imagens, vídeo e realidade virtual. Além disto, o discente interage com o conteúdo de diversas maneiras: explorando, selecionando, controlando, construindo, respondendo, entre outras maneiras.

Por fim, o último tipo de interação tratado nesta Seção se refere ao relacionamento entre alunos. Como a relação aluno-professor, este relacionamento pode ser síncrono ou assíncrono, caracterizando o que se denomina aprendizado colaborativo e cooperativo, chamado de interação interpessoal (PRATES, 2012). Tais interações geram motivação e atenção, enquanto os alunos aguardam o *feedback* dos colegas, e diminui a sensação de isolamento do estudo à distância. Este tipo de interação estimula ainda o senso crítico e a capacidade de trabalhar em equipe. Há ainda AVEA que estimularam que tais interações permitissem a avaliação entre alunos (*peer review*).

É possível ainda investigar a interação sobre a ótica da Interação Humano Computador. Seguindo esta ótica, Prates (2012), por exemplo, afirma

que o sucesso de um sistema colaborativo depende da sua qualidade de uso, ou seja, “da qualidade da interação e interface oferecida ao usuário”. Para analisar tais pontos, o autor descreveu a qualidade de uso baseado na usabilidade, sociabilidade, comunicabilidade e acessibilidade. Enquanto a usabilidade está associada às características do sistema, a sociabilidade depende também dos membros do grupo e do uso que fazem do sistema (PRATES, 2012). O objetivo do estudo da interação, portanto, é analisar como a interação se relaciona com a evasão e como esses conceitos foram discutidos na EaD.

3.4.3 Motivação

Outra métrica discutida e utilizada no presente estudo é a Motivação em ambientes educacionais. A abordagem tratada é baseada na teoria da Autodeterminação, que tem sido discutida no campo da motivação da aprendizagem. Muitas pesquisas destacam que a motivação pode produzir efeitos positivos na aprendizagem e no desempenho dos estudantes (SCHUNK, 1991; WECHSLER, 2006). Os princípios da teoria da Autodeterminação apontam que as motivações discentes diferem e são influenciadas pelos contextos que subsidiam as necessidades psicológicas com manifestações distintas (LEAL, MIRANDA e CARMO, 2013). Tal fato torna a motivação dos discentes “um fenômeno complexo, que pode ser inferido mediante a observação do comportamento discente” (GUIMARÃES e BZUNECK, 2008, p.111).

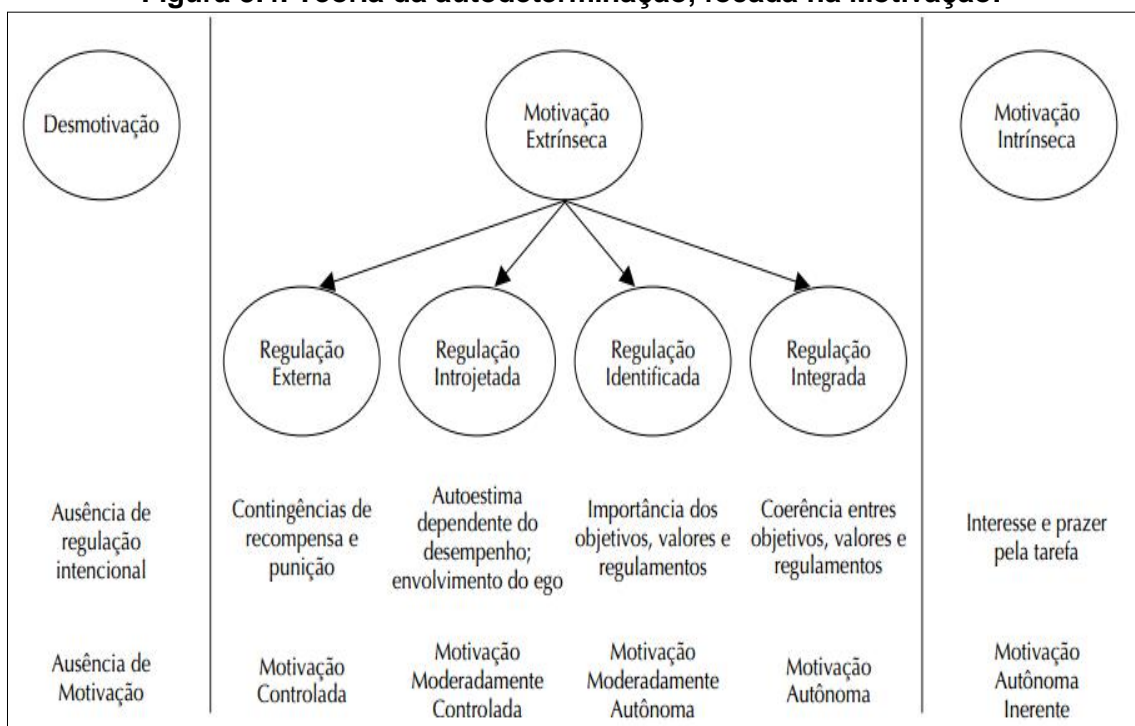
De acordo com a teoria da Autodeterminação mencionada, a Motivação de um discente está agrupada em três tipos principais: Desmotivação, Motivação Extrínseca e Motivação Intrínseca (Figura 3.4).

A Desmotivação engloba indivíduos que apresentam ausência de motivação, ou seja, o aluno não demonstra intenção nem comportamento proativo e se observa uma desvalorização da atividade, além de uma falta de percepção de controle pessoal (GUIMARÃES e BZUNECK, 2008).

O aluno com orientação motivacional extrínseca, busca a aprovação do professor e dos pares (outros alunos), uma necessidade de promoção

(MARTINELLI e SISTO, 2010; PAIVA e BORUCHOVITCH, 2010) ou ainda se envolve com tarefas apenas por acreditar que haverá alguma consequência (positiva ou negativa) (GOUDAS, DERMITZAKI e BAGIATIS, 2000; GUAY, BOGGIANO e VALLERAND, 2001). Nesta modalidade, a aprendizagem não é o objetivo principal. Este tipo de motivação pode ser subdividido em quatro grupos de regulação comportamental: regulação externa, regulação introjetada, regulação identificada e regulação integrada (LEAL, MIRANDA e CARMO, 2013).

Figura 3.4. Teoria da autodeterminação, focada na Motivação.



Fonte: Adaptado de Gagné e Deci (2005, p.336).

Entende-se por regulação externa quando o discente atua de modo a obter recompensas ou evitar punições. Ao falar de regulação introjetada, deve-se entender que o discente consegue administrar as consequências externas por meio dos resultados a pressões internas (culpa e ansiedade). A regulação identificada, que é o mais autônomo dos grupos, pois há alguma interiorização por parte do discente, mesmo que a razão para fazer alguma coisa seja de origem externa. Por fim há a regulação integrada, em que há coerência entre o comportamento, os objetivos e valores da pessoa. É a forma de Motivação Extrínseca mais autônoma, embora o foco ainda esteja nos benefícios pessoais

obtidos com a execução da atividade (GUIMARÃES e BZUNECK, 2008).

O último grupo definido pela teoria da Autodeterminação é a Motivação Intrínseca que é citada na literatura como um mediador importante na conquista da competência e no autoconceito discente. O discente motivado intrinsecamente em um determinado domínio do comportamento humano demonstra um esforço pessoal para executar a atividades propostas (PANSERA et al, 2016). A Motivação Intrínseca se refere à execução de uma determinada atividade pelo prazer que esta pode proporcionar, sem a necessidade de pressões externas, prêmios ou recompensa (CORPUS, MCCLINTIC-GILBERT e HAYENGA, 2009; MARTINELLI e SISTO, 2010; PAIVA e BORUCHOVITCH, 2010). O discente busca a atividade por esta ser interessante, envolvente e geradora de satisfação (MARTINELLI e SISTO, 2010; ZISIMOPOULOS e GALANAKI, 2009). Um discente com orientação motivacional intrínseca está mais impulsionado a estudar, a iniciar uma atividade e permanecer na atividade até o fim (PAIVA e BORUCHOVITCH, 2010).

3.5 Considerações Finais

Neste capítulo, é descrito o referencial teórico necessário para apoiar a pesquisa de tese em questão. São abordadas características da lógica *fuzzy* e da mineração dos textos. Tais técnicas são as principais na construção do *FuzzySD*, apresentado e discutido na Capítulo 4. Também foram apresentados métricas importantes da pedagogia para justificar as entradas utilizadas na abordagem, na estrutura *fuzzy* e na avaliação final da pesquisa.

4. Abordagem *FuzzySD*

Após a investigação sobre as perspectivas das pesquisas correlatas e a apresentação da fundamentação teórica, no presente capítulo é descrito o método científico percorrido durante o planejamento e execução o da pesquisa. Além disto, é discutida a metodologia adotada para construção da abordagem avaliativa multidimensional *FuzzySD*.

Na pesquisa é realizado um estudo, na esfera da avaliação e acompanhamento das interações pedagógicas dos estudantes, a fim de prover um modelo de monitoramento frente ao perfil de evasão dos discentes de EaD. Para tanto, a primeira atitude executada pela pesquisa é o planejamento da metodologia científica percorrida durante o processo. Entende-se por método científico a sucessão de passos estruturados e orientados no sentido de imprimir elevada probabilidade de precisão e validade aos resultados obtidos (GRESSLER, 2003).

Neste capítulo, são apresentadas as questões secundárias da tese, advindas da pergunta central da pesquisa: “Como e quais fatores comportamentais do aluno no AVEA impactam na evasão discente em cursos *online*?”. Para responder esta questão central, alguns experimentos são propostos e realizados ao longo do estudo.

Algumas etapas são descritas, a saber: seleção do contexto, instrumentação das análises executadas, etapas de concepção da *FuzzySD*, planejamento e implementação do modelo de avaliação e considerações finais sobre a abordagem. Tais etapas são delineadas com base no método científico apresentado por Wohlin (2012).

4.1 Seleção do Contexto

É fundamental considerar a descrição do universo como uma etapa importante para a pesquisa apresentada. A descrição do contexto-alvo analisado pela pesquisa contribuiu para melhor compreensão dos dados estudados e planejamento da avaliação dos resultados. No presente estudo, definiu-se

como contexto geral cursos *online* pertencentes à UAB ou com uma estrutura semelhante de curso. Contudo, para modelagem e avaliação tornou-se necessário definir um contexto de análise mais específico. Para tanto, foi protocolada uma parceria junto ao Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB), apresentada no Anexo A. Tal parceria permitiu acesso aos dados de alguns cursos *online* ministrados pela instituição, tais como: Licenciatura em Letras, Segurança do Trabalho e Ciência da Computação.

Os cursos analisados estavam concluídos e nenhum curso em andamento foi incorporado durante a investigação, pois não houve autorização da instituição parceira para tal interação.

4.2 Ferramentas utilizadas na Pesquisa

Para coleta, processamento de dados, análise dos experimentos e avaliação da *FuzzySD* tornou-se necessário utilizar uma série de ferramentas para apoiar o processo como um todo. Diversos *softwares* foram utilizados em fases distintas da pesquisa.

A fim de elucidar os instrumentos utilizados, a pesquisa foi dividida em sete etapas não sequenciais:

- a) coleta e processamento dos dados: *scripts* em Mysql⁴, R⁵, Python⁶ e Weka⁷;
- b) classificação dos dados: Python e Weka;
- c) análise de grupos (clusterização): *scripts* em R;
- d) modelagem *fuzzy*: *skfuzzy*⁸, Matlab⁹;
- e) aprendizagem não supervisionada: Python (dicionários);
- f) classificação das mensagens: Python (com a biblioteca NLTK¹⁰);

⁴ <https://www.mysql.com/>

⁵ <https://www.r-project.org/>

⁶ <https://www.python.org/>

⁷ <https://www.cs.waikato.ac.nz/ml/weka/>

⁸ <http://pythonhosted.org/scikit-fuzzy/>

⁹ <https://www.mathworks.com/products/matlab.html>

- g) construção da abordagem *FuzzySD*: Python;
- h) avaliação final: Weka e Python.

4.3 Proposição da Abordagem

Diante das informações obtidas após o levantamento bibliográfico e do esforço dos tomadores de decisão para avaliar os alunos na EaD, na presente pesquisa foi detectada a existência de problemáticas associadas ao tema que ainda precisavam ser pesquisadas e melhoradas no contexto da avaliação das turmas virtuais. Neste sentido, o trabalho propôs uma abordagem metodológico a fim de otimizar as análises, formativa ou somativa, baseando-se em uma análise multidimensional dos discentes nos AVEA.

A presente abordagem avaliativa tem como propósito analisar os indicadores multidimensionais associados às atividades pedagógicas do aluno no ambiente e, com isto, prover contribuições à avaliação dos discentes por meio do monitoramento de métricas que descrevam suas interações nos cursos *online*. Tal investigação pode ser útil no acompanhamento do perfil evasivo dos usuários no AVEA.

Lidar com a problemática da evasão tem sua relevância na EaD, pois esta problemática é um dos desafios detectados nos cursos *online* e atualmente não há uma solução efetiva na literatura mundial pesquisada, de acordo com a RS executada. Neste sentido, o assunto está inserido em um contexto complexo de causas e indicadores. Tal fato envolve direta ou indiretamente uma gama de outras questões da EaD, como: reprovação, desempenho, desmotivação, colaboração, tutoria, métodos avaliativos *etc.*

A estrutura padrão dos cursos ministrados nos AVEA envolve uma série de atividades pedagógicas apresentadas em grupo e distribuídas ao longo das aulas (ou módulos). Deste modo, é possível afirmar que cada grupo de atividades é composto por um conjunto de recursos pedagógicos, ou seja, materiais pedagógicos cadastrados pelo docente para nortear o discente no processo de ensino-aprendizagem virtual.

¹⁰ <http://www.nltk.org/>

A literatura anteriormente analisada demonstra que tais atividades desempenham um papel fundamental na avaliação e acompanhamento do processo de ensino-aprendizagem dos discentes nos AVEA (ROMERO *et al.*, 2013; HE, 2012).

Diante do panorama de elevados índices de desistência discente em cursos *online*, o presente estudo propôs uma abordagem avaliativa baseada na análise de múltiplos fatores comportamentais do estudante de EaD. A abordagem apresentada tem como entradas as informações obtidas por meio das atividades dos alunos do curso *online*.

A abordagem de avaliação apresentada pode ser executada diversas vezes ao longo do curso *online* (avaliação formativa) ou apenas ao seu final (avaliação somativa). Tal decisão fica a critério do docente e dos seus objetivos pedagógicos. A abordagem foi denominada Módulo de Avaliação Multidimensional *FuzzySD* (*Fuzzy Student Dropout*).

Em linhas gerais, a abordagem almejou, por meio de uma avaliação multidimensional das atividades discente nos cursos *online*, prover informações ao docente que a permita identificar a evasão discente no momento da avaliação. Esta ação teve como intuito principal minimizar os índices de evasão dos alunos participantes no contexto do estudo. Para tal, alguns caminhos foram percorridos durante a modelagem e construção da abordagem: definição das métricas de entrada da *FuzzySD*, discussão sobre as fases relacionadas à análise *fuzzy* e elaboração de uma estratégia de avaliação da abordagem.

4.3.1 Métricas de Entrada

Considerando os tipos de participação dos discentes em cursos *online*, os indicadores para avaliação das interações dos alunos devem ser amplos o suficiente para que seja possível contemplar os tipos e objetivos da modalidade a distância. O primeiro desafio explorado pelo estudo foi a seleção das variáveis de interação do discente com o AVEA, que fossem relevantes para medir a evasão discente. Como citado anteriormente, os *surveys*, de maneira geral, trataram de métricas relacionadas à evasão do aluno associando tal problemática aos critérios pessoais discentes e sobre a situação do aluno. Tais

critérios não podem ser utilizados pela presente abordagem, pois a abordagem avaliativa apresentada visa investigar fatores comportamentais do discente nos AVEA para relacioná-los à evasão.

De acordo com as pesquisas comentadas na revisão de literatura e com a discussão sobre as métricas pedagógicas no referencial teórico, algumas métricas de participação e interação do discente podem ser incorporadas como entradas da abordagem apresentada, pois tais critérios foram relevantes durante as análises relacionadas à evasão do discente. Diante destas métricas, a presente tese definiu três grupos como destaque: (i) variáveis relacionadas à Autorregulação do aluno; (ii) variáveis relacionadas à Interação e (iii) variáveis relacionadas à Motivação do aluno.

4.3.1.1 Arquitetura Multidimensional da FuzzySD

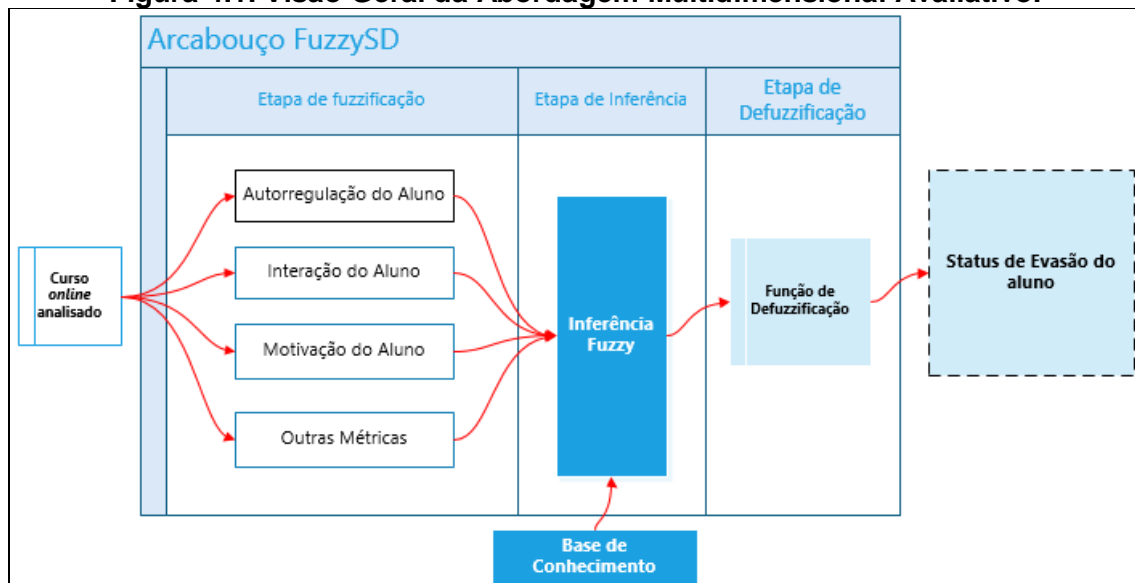
Cada curso tem suas peculiaridades e cada professor seus critérios de avaliação, que dependem da sua visão de mundo, de sociedade e da educação. Por isto, é necessário elencar alguns indicadores que são parâmetros para que o professor adapte seus objetivos à abordagem proposta. Após a definição do grupo de métricas relevantes à avaliação do discente, a estrutura da avaliação multidimensional foi definida de acordo com os objetivos da pesquisa.

Uma das dificuldades associadas aos processos avaliativos na educação é o aspecto subjetivo e a incerteza intrínseca a esta atividade. Com base nestes desafios, pesquisas apresentadas na revisão de literatura do presente documento propuseram e desenvolveram modelos e ferramentas de acompanhamento do desempenho baseados em conjuntos *fuzzy*, nos quais o professor, por exemplo, relaciona o grau de pertinência destes valores em intervalos de notas, assim como a base de regras. Seguindo a tendência apresentada pelos trabalhos correlatos, o presente documento detalha uma abordagem avaliativa baseada em premissas da lógica *fuzzy*.

Diante das métricas definidas na análise anterior, foi possível apresentar uma visão geral das entradas *fuzzy* e as etapas da utilização da lógica nebulosa na abordagem, ilustrado na Figura 4.1. Para fins didáticos, o sistema *fuzzy* foi detalhado de acordo com as três etapas fundamentais ao

processo: fuzzificação, inferência e defuzzificação.

Figura 4.1. Visão Geral da Abordagem Multidimensional Avaliatio.



Fonte: Próprio Autor.

A abordagem delineada deixou em aberto a possibilidade de inserção de novas métricas por meio do critério “Outros Métricas”. Tal estratégia foi adotada por identificar algumas especificidades de acordo com o contexto de avaliação adotado. A *FuzzySD* tenciona adaptar e validar a abordagem avaliativa a diferentes contextos de cursos *online*, seguindo o pré-requisito de seguir uma estrutura de ensino semelhante àquela definida pela UAB.

O passo seguinte foi analisar as variáveis de entrada selecionadas de acordo com as métricas definidas na arquitetura. Nesta perspectiva, percebeu-se que os AVEA atuais suportam a extração de informações relacionadas à Interação, à Autorregulação e à Motivação do aluno ao longo do curso. Contudo, elementos textuais ainda são pouco analisados nas ferramentas e podem prover informações relevantes para investigação da problemática da evasão discente.

Para definir quais características e entradas pertenceriam a cada uma das métricas selecionadas, foi preciso identificar se havia parâmetros que permitissem medir tais métricas no contexto analisado. Por isto, algumas análises de correlação foram executadas e discutidas nas próximas subseções.

4.3.1.2 Análise de Correlação das Entradas e a Evasão

Selecionadas as métricas de análise (Interação, Autorregulação e Motivação), o próximo passo, para ratificar a influência de tais informações diante da problemática da evasão discente, foi a execução de uma análise de correlação. Segundo o dicionário Aurélio (2002), “correlação significa relação mútua entre dois termos, qualidade de correlativo, correspondência. Correlacionar, significa estabelecer relação ou correlação entre; ter correlação”. Torna-se necessário muitas vezes, durante uma investigação de natureza quantitativa, avaliar o grau de relacionamento entre duas ou mais variáveis e, assim, descobrir com exatidão, o quanto uma variável interfere em outra.

O presente estudo analisou uma série de variáveis relacionadas às atividades dos discentes e suas interações nos cursos *online* (ilustradas no Quadro 4.1). Assim, foi avaliado se havia associação entre as variáveis selecionadas e o fenômeno da evasão.

Quadro 4.1. Variáveis dos discentes analisadas.

Variável	Campo na Tabela do curso	Natureza
1. Edição (ed)	Número de vezes que um recurso ou mensagem foi editada pelo discente	Quantitativa
2. Nota média - ScoreMean (sm)	Nota média na atividade	Quantitativa
3. Feedback (fe)	Número de <i>feedbacks</i> providos entre professor e aluno	Quantitativa
4. Fórum (fo)	Número de interações nos fóruns de discussões	Quantitativa
5. Bate-papo - Chat (ch)	Número de interações no bate-papo	Quantitativa
6. Visualizações (nv)	Número de visualizações dos recursos pedagógicos	Quantitativa
7. Tempo sem interação no ambiente (lac)	Número de dias sem acessar o ambiente.	Quantitativa
8. Status da Evasão (ds)	Situação de evasão do discente (dois valores possíveis: Evadido e Persistente)	Qualitativa

Fonte: Próprio Autor.

Todas as variáveis do aluno utilizadas no experimento foram mineradas por meio da extração de informações dos *logs* de interações dos alunos no curso *online* analisado. Nesta fase, *scripts* em Python e R foram programados

para extrair dados e analisar as correlações.

A evasão foi medida por meio da variável *lac* (tempo sem interação no ambiente), pois tal medida é responsável por retornar o número de dias em que o discente não interage com o curso *online*. A variável *lac* é utilizada para definir se o aluno está evadido ou persistente no curso *online*. Caso o número de dias contínuos, sem interagir com as atividades do curso, ultrapassasse um quarto do tempo total (25%) do curso, pode-se afirmar que o discente está evadido e a variável *status* da evasão (*ds*) assume o valor “Evadido”.

Para executar este experimento, foram utilizados dados das interações dos alunos do 1º período de Letras do IFPB do período 2015.2. A amostra analisada nesta análise de correlação contou com 148 alunos matriculados e uma taxa de evasão (de acordo com a definição de evasão adotada) de 35%.

As hipóteses formuladas na execução do experimento foram:

H_0 : Não existe correlação moderada ou forte entre a métrica e a evasão (*lac*) ($\rho < 0,3$)

H_1 : Existe correlação moderada ou forte entre a métrica e a evasão (*lac*) ($\rho \geq 0,3$)

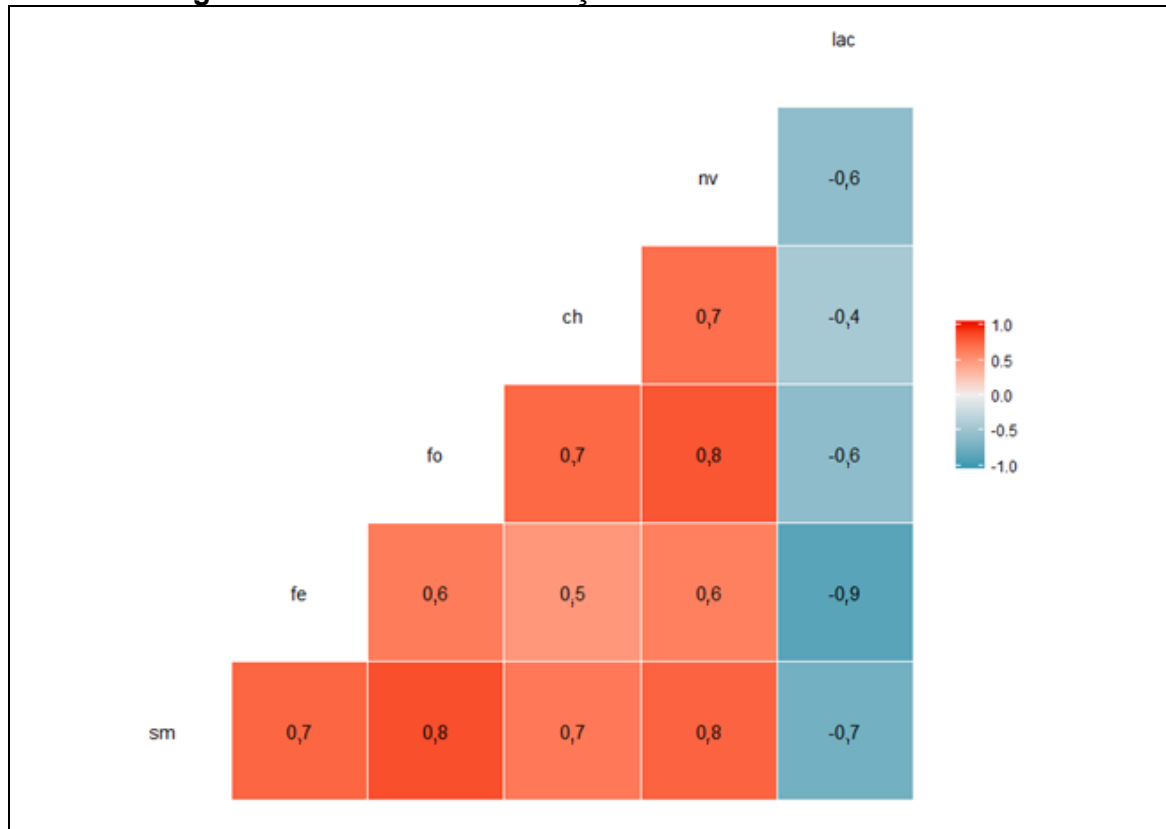
Os resultados da análise de correlação de Pearson são ilustrados na Figura 4.2, a partir dos quais é possível encontrar diversas variáveis relacionadas à problemática evasão no contexto analisado. A literatura relata que correlações moderadas são aquelas cujo coeficiente de correlação de Pearson (ρ) são superiores ao valor de 0,3 e inferior a 0,7. Além disto, Correlações fracas são inferiores a 0,3 e correlações fortes são superiores a 0,7 (SHEVLYAKOV e OJA, 2016).

Vale destacar, que houve correlação forte entre o tempo sem interação no ambiente (*lac*) e a nota média do aluno (*sm*), o número de *feedbacks* (*fe*), o número de interações em fóruns (*fo*) e o número de visualizações (*nv*).

Observou-se uma correlação moderada entre o último acesso (*lac*) e o número de mensagens bate-papos (*ch*) postadas. As edições de recursos e textos foram pouco executadas pelos discentes desta amostra e a correlação foi muito fraca para ser exibida na ilustração. O sinal negativo nas correlações

não interfere na análise só significam um comportamento inverso, por exemplo: quanto mais postagens em um fórum um aluno fizer, provavelmente, menor será o número de dias consecutivos sem acessar o ambiente.

Figura 4.2. Análise de correlação entre as variáveis discente.



Fonte: Próprio Autor.

Este experimento foi repetido para o segundo período do curso de Letras e o primeiro período do curso técnico de Segurança no Trabalho. Os resultados foram bem semelhantes e permitiram concluir que há métricas comportamentais do discente correlacionadas à problemática evasão em diferentes contextos. Tal fato possibilita o envolvimento de tais métricas em experimentos futuros, dado que, comprovadamente, há uma correlação com o fenômeno principal de estudo.

4.3.1.3 FuzzySD no Contexto dos Cursos Online da UAB

Após verificar as correlações de algumas variáveis relacionadas às atividades dos discentes no curso, na presente subSeção é detalhada cada uma das métricas da *FuzzySD* no cenário selecionado (cursos com a estrutura UAB).

É fundamental perceber que cada curso *online* tem suas singularidades

e especificidades, mas que há alguns pontos semelhantes que podem ser estudados também. Os cursos da UAB, por exemplo, possuem modelos de interações semelhantes, papéis de usuários bem definidos no sistema, organização de cursos em módulos de atividades e sistema de avaliação padrão. Assim, há um perfil de interações comum que pode ser explorado independente da diversidade dos cursos.

Os cursos *online* definidos pela UAB são compostos por um conjunto de recursos pedagógicos, agrupados em módulos (ou aulas). Tais recursos são explorados de acordo com os objetivos pedagógicos do docente e com as competências que precisam ser trabalhadas. Tal organização varia de acordo com os objetivos e estratégias pedagógicas do tomador de decisão. Para definir este aspecto, toma-se como base a estrutura dos módulos utilizados nos cursos *online* do IFPB (contexto de validação base).

A *FuzzySD* foi estruturada com base em três métricas principais: Autorregulação, Interação e Motivação. Em relação à presente abordagem, a característica multidimensional é claramente percebida ao se identificar a abordagem multifacetada da abordagem apresentada, ou seja, diversos aspectos relacionados ao processo de ensino-aprendizagem do discente estão envolvidos durante a análise das três métricas bases da abordagem.

Ao planejar a *FuzzySD* aplicada ao contexto dos cursos *online* do IFPB, foi necessário incluir outras variáveis comportamentais do discente para descrever tais métricas. Assim, o modelo proposto foi adaptado e novos critérios foram utilizados como entrada. Assim, o tripé de indicadores envolveu uma série de outras variáveis relacionadas ao comportamento do discente no contexto estudado.

As decisões sobre quais variáveis discente descrevem cada uma das métricas foram tomadas de acordo com a literatura de cada uma das métricas. Por exemplo, a métrica Autorregulação é descrita pelas variáveis média das notas, interações com o professor e edição de recursos. A seleção de tais variáveis motivou-se por uma série de razões. Inicialmente, a literatura mostrou que há uma correlação entre alunos autorregulados e o seu bom desempenho no curso, por isto explorar o uso do cálculo de médias do aluno. Outro fator destacado foi o de que a Autorregulação pode ser estimulada por professores e

tutores nos AVEA. Este estímulo, no caso do IFPB Virtual, ocorreu, principalmente, por meio de mensagens de *feedback*. O aluno reagir a *feedbacks* dados pelo professor é uma estratégia para se autorregular também.

Por fim, a última variável discente utilizada para descrever a Autorregulação foi a edição de recursos pedagógicos, pois esta atividade refletiu diretamente sobre o caráter reflexivo do discente no curso *online*. Edição de recursos é um indicativo de reflexão sobre as postagens e materiais compartilhados.

Seguindo tais investigações, tem-se a métrica Interação que, ao ser analisada, considerou as seguintes informações: número de visualizações das atividades, interação com outros alunos no ambiente e o número de dias sem acessar o ambiente. O objetivo desta métrica foi analisar as interações do discente em sua totalidade no curso. Por este motivo, foram contempladas a análise das visualizações de diversos tipos de recursos pedagógicos no ambiente. Outro tipo de interação que precisou ser incluída foram as interações entre pares ao longo do curso *online*. Por fim, a variável discente que verifica “o número de dias contínuos sem acessar o ambiente” foi incluída na análise da métrica Interação por representar a presença deste aluno nas atividades do curso analisado.

Em relação à métrica Motivação, exploram-se características extraídas de mensagens postadas nos fóruns de discussão. Para isto, foram utilizadas técnicas de mineração de texto com base nas mensagens nos fóruns de discussão. A decisão de utilizar as mensagens de fórum foi motivada pela forte correlação dessas com a evasão nos AVEA.

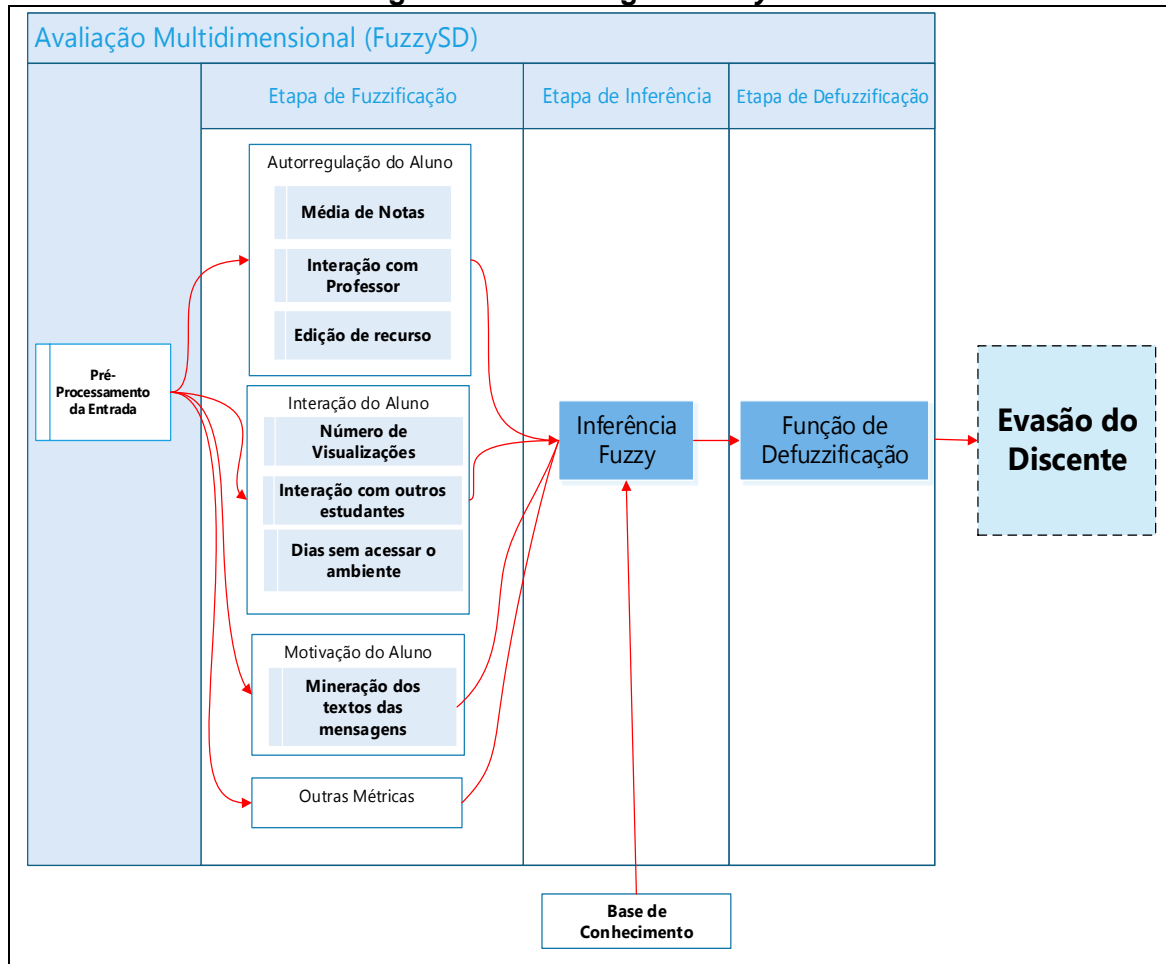
Quadro 4.2. Variáveis das discentes associadas às métricas de entrada.

Autorregulação	Interação	Motivação
Média das Notas	Número de visualizações das atividades	Mineração dos textos das mensagens (fórum de discussão)
Interação com o professor	Interação com outros alunos	
Edição de recursos pedagógicos	Dias sem acessar o sistema	

Fonte: Próprio autor.

No Quadro 4.2 são ilustradas as variáveis associadas a cada uma das métricas e na Figura 4.3 é apresentada a arquitetura geral da abordagem proposta.

Figura 4.3. Abordagem *FuzzySD*.



Fonte: Próprio autor.

4.4 Planejamento das Etapas da *FuzzySD*

Após a discussão sobre as métricas, o estudo focou seus esforços na modelagem da arquitetura *fuzzy* da abordagem. Assim, a presente subSeção tem como intuito nortear as decisões sobre as principais etapas da modelagem *fuzzy*, sendo essas: fuzzificação, inferência e defuzzificação, além da construção da base de regras importante para a fase da inferência.

4.4.1 Identificação das Variáveis *Fuzzy* Associadas à Abordagem

Como a abordagem *FuzzySD* foi modelada utilizando a estratégia *fuzzy*, alguns questionamentos podem incorrer na mente do leitor sobre esta decisão: as métricas selecionadas são variáveis nebulosas ou o que justifica a utilização da lógica *fuzzy* na modelagem do problema?

As respostas para tais questionamentos são discutidas e construídas com base na natureza nebulosa das métricas selecionadas como entradas na abordagem. Tal comportamento fez com que o tomador de decisão (professor, tutor ou gestor educacional) ao se referir às medidas de cada uma dessas se deparasse com termos como “muito motivado”, “pouco participativo”, “autorregulado”, “não autorregulado”, “indiferente”, *etc*, ao utilizar a linguagem natural para definir as saídas das métricas da abordagem. A lógica *fuzzy* lida com a natureza nebulosa de variáveis e permite tratar matematicamente tais valores. Outra relevância para a aplicação de *fuzzy* foi a possibilidade do discente utilizar informações intermediárias no planejamento da EaD. Por exemplo, ao avaliar a evasão especificamente o professor pode receber informações sobre a motivação do aluno e outras métricas como uma informação bônus e tomar decisões sobre aspectos intermediários.

Cada uma das métricas foi agrupada em categorias de acordo com as teorias pedagógicas mencionadas, por exemplo:

- (i) os alunos analisados, segundo a métrica Motivação, podem ser agrupados em três grupos principais: alunos motivados intrinsecamente, alunos motivados extrinsecamente e alunos desmotivados (mencionado na Seção 3.4.3);
- (ii) os alunos analisados, segundo a métrica Autorregulação, podem ser agrupados em dois grupos principais: aluno autorregulados e alunos não autorregulados; e
- (iii) os alunos analisados, segundo a métrica Interação, podem ser agrupados em três grupos principais: alunos não participativos, alunos pouco participativos e alunos participativos.

Tais agrupamentos são retomados nesta etapa do documento porque servem como base para a identificação das variáveis linguísticas associadas a cada uma das métricas analisadas. Os valores de uma variável linguística podem ser sentenças em uma linguagem especificada. A principal função das variáveis linguísticas é fornecer uma maneira sistemática para a caracterização aproximada de fenômenos complexos ou mal definidos (D'AMBROSIO, 2012). Em essência, a utilização do tipo de descrição linguística empregada por seres humanos, e não de variáveis quantificadas, permite o tratamento de sistemas que são muito complexos para serem analisados por meio de mecanismos matemáticos convencionais.

Neste sentido, o presente estudo, percebeu a necessidade de avaliar se, de fato, as categorias mencionadas representam o número adequado de variáveis linguísticas no contexto selecionado. Inicialmente, foi realizada uma entrevista não estruturada com um especialista, um professor do IFPB Virtual da área de Pedagogia, sobre a relevância das métricas e este afirmou que as categorias estavam em concordância com a realidade dos cursos.

Em um segundo momento, houve a necessidade de entender se os dados do cenário analisado também confirmariam tais categorias. Para investigar os indícios supracitados, foram realizadas análises de agrupamentos (clusterização) a fim de identificar quantos e quais são as categorias (*clusters*) ideais no cenário de cursos da UAB.

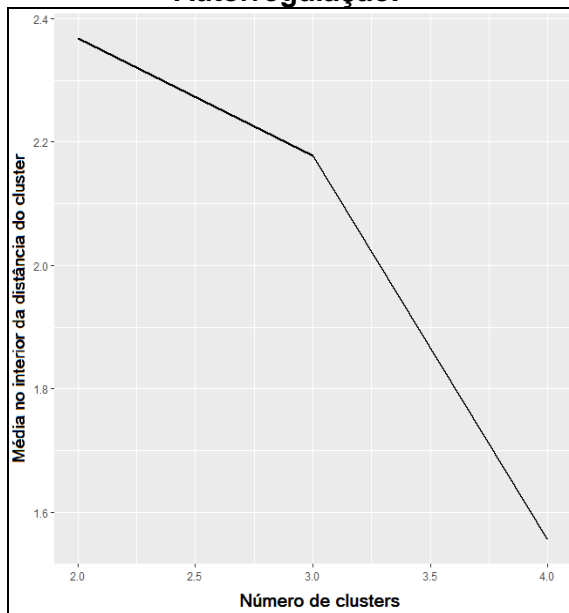
Para esta análise, foram selecionados dados de interações da turma do segundo período do curso técnico de Segurança do Trabalho no IFPB Virtual. A amostra analisada foi composta pelos registros de interações de 124 alunos durante o semestre. Neste experimento, foi aplicado o algoritmo *K-Means* (WU, 2012) sobre os dados, com os valores de $k = 3$ ou $k = 2$ (dependendo da métrica analisada), em que k representou o número de *clusters* analisados. Os métodos da silhueta e do joelho (ou cotovelo) foram utilizados para subsidiar a escolha do número adequado de *clusters*.

A técnica do joelho, que também pode ser denominada técnica do cotovelo, sugere *clusters* de modo que a soma total do espaço dentro de cada cluster (*total within cluster sum of squares - WSS*) seja minimizada. Para definir graficamente o número ideal de grupos, é necessário identificar a localização

de uma curva (joelho-cotovelo) na curva gerada, pois este ponto é considerado como um indicador do número apropriado de aglomerados (KASSAMBARA, 2017).

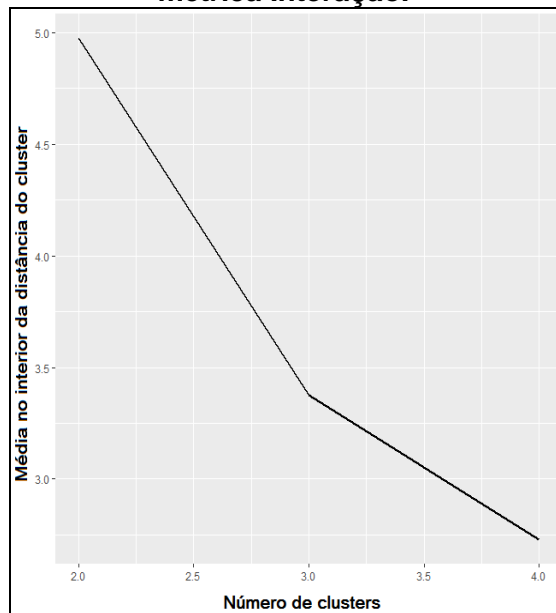
Ao analisar a técnica do joelho no presente experimento, houve um indicativo de que o número ideal de *clusters* para a amostra foi de três grupos para a métrica Interação e Autorregulação (Figura 4.4 e Figura 4.5). Como o especialista afirmou que dois grupos para a Autorregulação eram suficientes e a técnica do joelho mostrou três como o número adequado de grupos, a técnica do gráfico de silhueta serviu como uma comprovação final da amostra analisada.

Figura 4.4. Gráfico de joelho da métrica Autorregulação.



Fonte: Próprio Autor.

Figura 4.5. Gráfico de joelho da métrica Interação.

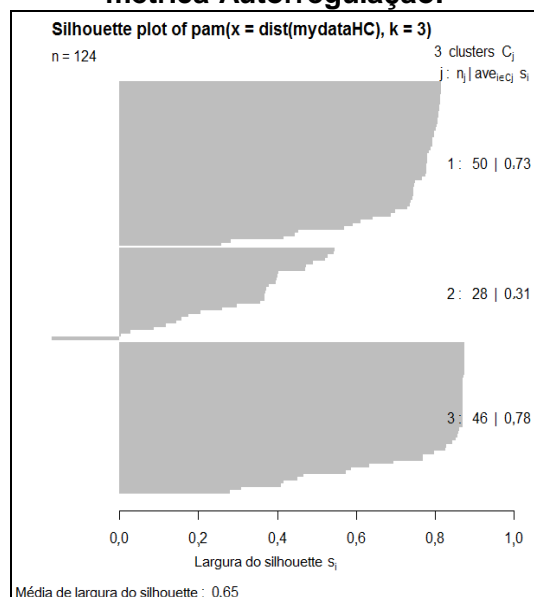


Fonte: Próprio Autor.

Como descrito, foi utilizado o gráfico da silhueta para avaliar os particionamentos como abordagem extra. Resumidamente, esta técnica mede a qualidade de um grupo, ou seja, determina o quão bem cada objeto está dentro do seu *cluster*. A largura de silhueta média alta indica um bom agrupamento (ROUSSEEUW, 1987). O método da silhueta média calcula a silhueta média de observações para diferentes valores de *k*. O número ótimo de *clusters* *k* é aquele que maximiza a silhueta média em uma variedade de valores possíveis para *k* (KAUFMAN e ROUSSEEUW, 2009). No gráfico de

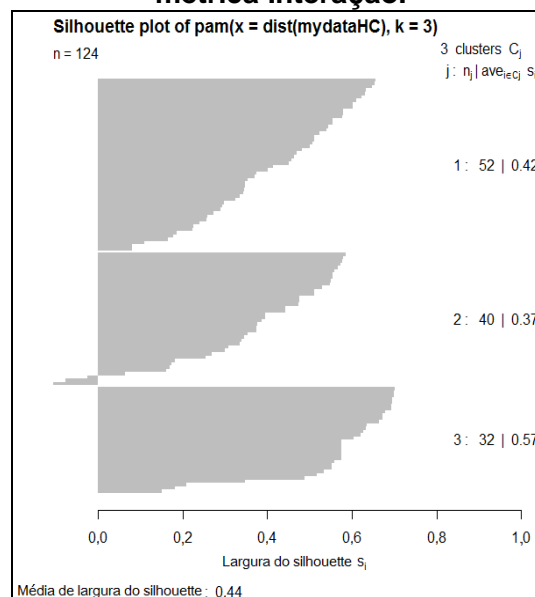
silhueta o eixo vertical representa os n objetos, enquanto o eixo horizontal representa o valor da silhueta para cada indivíduo. Na Figura 4.6 e na Figura 4.7 é apresentado o gráfico de silhueta para um conjunto de valores da métrica Autorregulação e Interação que foi subdividido em três *clusters*. O valor ótimo encontrado foi de três grupos para ambas as métricas.

Figura 4.6. Gráfico de Silhueta da métrica Autorregulação.



Fonte: Próprio Autor.

Figura 4.7. Gráfico de Silhueta da métrica Interação.



Fonte: Próprio Autor.

Após analisar o número de grupos, o estudo avaliou o comportamento dos grupos (*clusters*) encontrados. Inicialmente, investigou-se a métrica Autorregulação (Figura 4.8). Foi possível observar que houve a presença de três grupos bem distintos entre si. O grupo 1 se referiu aos estudantes que possuíam uma Autorregulação mediana, pois seus índices de Nota média e a interação com o professor foram equivalentes ao valor médio das interações entre o professor e aluno, além de possuírem notas médias próximas à média da turma.

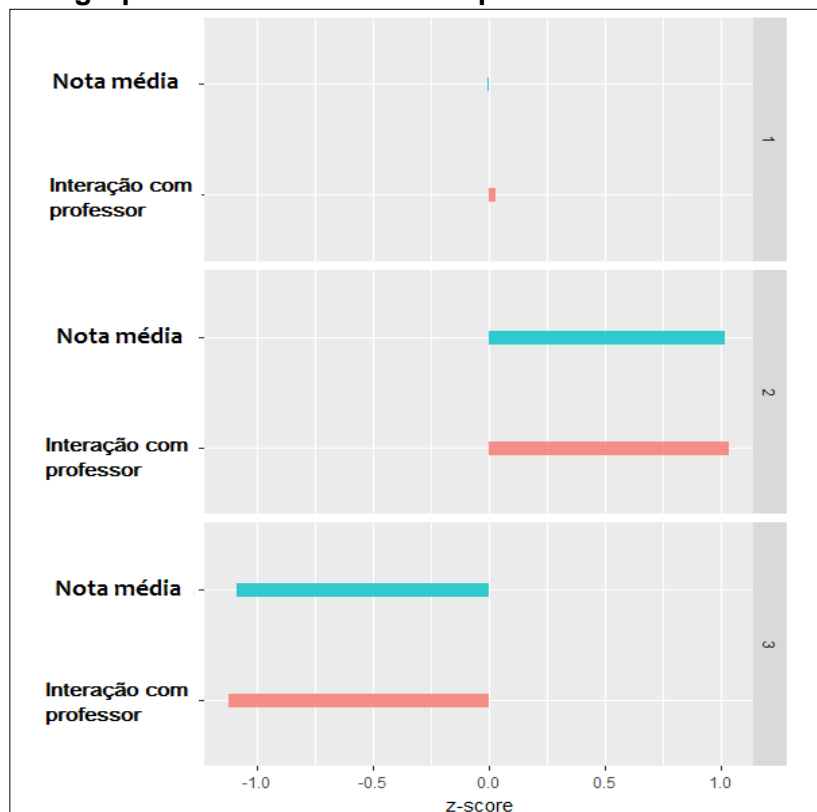
O grupo 2 agrupou os alunos mais bem autorregulados no curso, pois a nota média e o número de interação entre professores e alunos foram acima dos valores médios de tais variáveis no curso.

Por fim, o terceiro grupo apresentou o pior panorama em relação aos outros, pois os discentes obtiveram notas inferiores à média da turma e

interagiram abaixo do esperado com o professor.

Por tais fatos, pode-se denominar o 1º grupo: Alunos em processo de autorregulação, 2º grupo: Alunos autorregulados e o 3º grupo: Alunos não regulados.

Figura 4.8. Agrupamentos encontrados a partir da métrica Autorregulação.



Fonte: Próprio Autor.

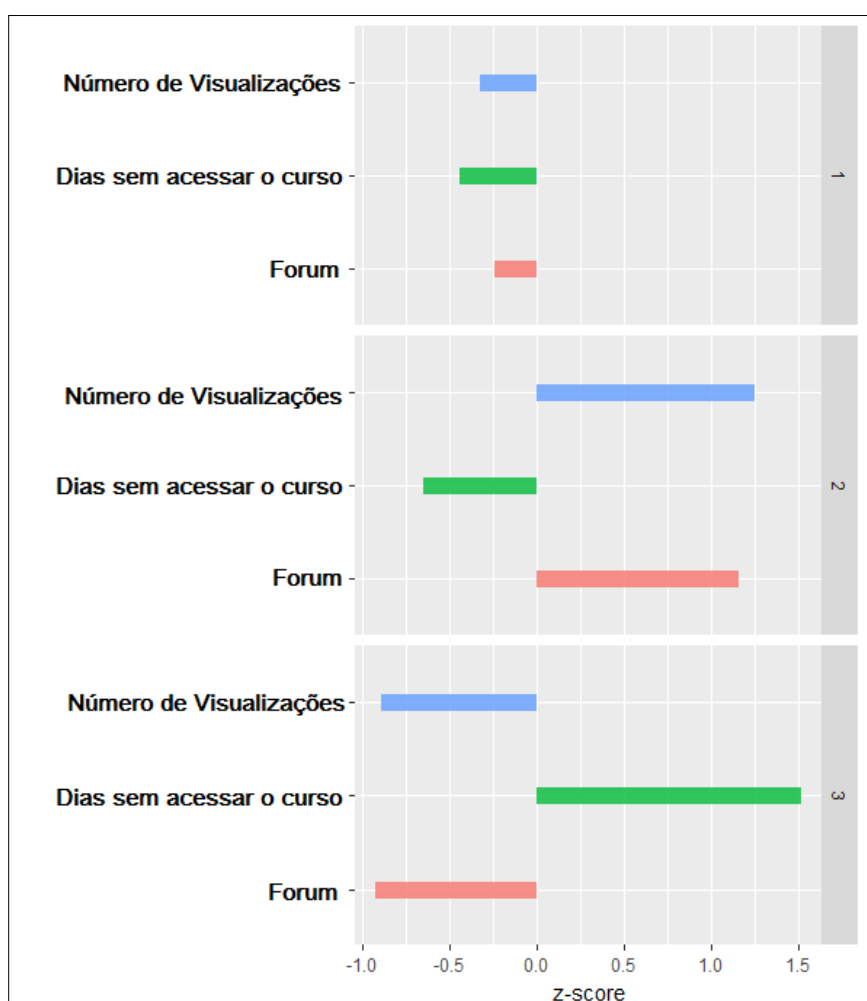
Em seguida, foi executada a análise da métrica Participação. Neste agrupamento, novamente foram definidos três grupos como mencionado (Figura 4.9). O grupo 1 representou aqueles alunos que sempre acessavam o ambiente, dado o índice *z-score* menor do que a média de dias sem acessar o curso *online* da amostra (representado pela barra verde). Contudo, estes alunos interagiram abaixo da média da amostra com os recursos pedagógicos, pois os números de visualizações e interações com os fóruns foram inferiores à média dessas variáveis na turma.

O grupo 2 foi composto por discentes participativos, pois apresentaram os menores índices de dias sem acessar o ambiente e os melhores índices de acesso e participação em fóruns da turma.

O grupo 3 representou os alunos que ficaram muito tempo sem acessar o ambiente e possuíam os registros muito inferiores à média da turma de número de visualizações e acessos ao fórum.

Entendendo melhor o comportamento dos grupos, foi possível denominar o grupo 1: Alunos Pouco Participativos, grupo 2: Alunos Participativos e o grupo 3: Alunos Não Participativos.

Figura 4.9. Agrupamentos encontrados a partir da métrica Interação.



Fonte: Próprio Autor.

A última métrica analisada foi a Motivação discente. Nesta etapa, a análise da Motivação foi executada por meio de uma classificação baseada na polaridade das mensagens de fóruns. A investigação realizada utilizou os agrupamentos citados na literatura e mencionados na Seção 3.4.3: Motivação Intrínseca, Motivação Extrínseca e Desmotivação. O objetivo da categorização da amostra pelo professor foi prover uma base de classificação para apoiar a

construção do dicionário de palavras que foi utilizado na fase de aprendizagem não supervisionada.

A categorização inicial se deu mensagem a mensagem pelo professor colaborador no experimento, diante da amostra de mensagens do primeiro período do curso de Letras. O docente do curso de Letras descreveu seus critérios de classificação como: “Caso o aluno não interaja ou responda coisas não compatíveis com a proposta da atividade, eu o classifico como Desmotivação. Caso o aluno questione sobre aspectos de nota, ou peça em algum momento para prorrogar os prazos da atividade ele demonstra uma Motivação Extrínseca, ou seja, o objetivo de ser aprovado na disciplina é o primordial. O último grupo Motivação Intrínseca deve ser detectado quando o aluno deixa claro que está se divertindo com a atividade, respondendo adequadamente e não questionando aspectos relacionados à nota ou pontuação da atividade.”.

4.4.1.1 Mineração de Texto da Métrica Motivação

Apesar de a área da mineração de texto abranger diversos tipos de estudos de classificação e subjetividade, como mencionado, a tarefa mais investigada se limitou a classificar a polaridade de textos em positiva, negativa e neutra.

Normalmente, o que é produzido em textos por humanos pode ser classificado em fato ou opinião segundo Liu (2010). Fatos podem ser entendidos como sentenças objetivas, desprovidas de sentimento, que transmitem alguma informação sobre um tema ou entidade. Em contrapartida, há a opinião de que estão presentes frequentemente em sentenças subjetivas que contêm algum aspecto de avaliação, impressão ou opinião.

Contudo, não é uma tarefa trivial estabelecer a qualidade exata de um classificador de texto, pois existe uma dificuldade em se identificar um conjunto de dados cujos rótulos (no experimento da métrica Motivação: desmotivação, motivação Intrínseca e motivação Extrínseca) sejam conhecidos e confiáveis. Naturalmente, alguns vieses acabam influenciando o processo e o comportamento do classificador (AVANÇO, 2015).

Neste experimento, foram selecionados dados do curso de Letras do IFPB Virtual referentes ao primeiro período (11.046 mensagens) de 2015.2.

Inicialmente, o professor classificou manualmente 1/3 das mensagens da amostra. No Quadro 4.3 são ilustrados exemplos de mensagens contidas na amostra classificadas pelo professor.

Neste caso, a mineração de texto executou uma investigação baseada na classificação dos textos, especificamente por meio da extração da polaridade das mensagens, pois para avaliar a desmotivação, lida-se com a identificação de termos negativos no discurso do discente.

Em relação à Motivação Extrínseca e a Intrínseca, é percebido que palavras positivas estão contidas em ambos os discursos. As diferenciações entre ambas são os objetivos das mensagens. Por um lado, as mensagens com Motivação Extrínseca apresentam elementos extras relacionados às preocupações discentes com a atividade em si, disciplina, nota, prazos de submissão *etc.* Finalmente, a motivação Intrínseca sempre deixa bem claro o prazer do aluno na execução da tarefa.

Após definir a estratégia de análise para minerar o texto, é possível de fato, iniciar o experimento para a construção do dicionário de dados relacionados utilizado para classificação não supervisionada das mensagens em três grupos: Desmotivação, Motivação Extrínseca e Motivação Intrínseca, como mencionado na Seção 3.4.3.

O próximo passo foi verificar as palavras mais comentadas em cada um dos tipos e, juntamente com o professor, elaborar um dicionário adequado ao contexto. Para isto, foi decidido utilizar a medida Tf-Idf (*term frequency and inverse document frequency*) para verificar quais as palavras mais referenciadas no texto e com isto construir, juntamente com um dos docentes do curso de Letras, o dicionário. Nesta fase do experimento, havia um terço da amostra de mensagens classificadas manualmente pelo professor que serviu de base para as análises.

A técnica Tf-Idf calcula um peso que representa a importância de um termo dentro de um documento. Isto é executado comparando a frequência de uso em um documento individual em oposição a todo o conjunto de dados (uma coleção de documentos). Este peso é uma medida estatística usada para avaliar a importância de uma palavra para um documento em uma coleção. A

importância aumenta proporcionalmente ao número de vezes que uma palavra aparece no documento, mas é compensada pela frequência da palavra. As variações do esquema de ponderação Tf-Idf são frequentemente utilizadas pelos motores de busca como uma ferramenta central na pontuação e classificação da relevância de um documento, dada uma consulta do usuário.

Quadro 4.3. Exemplos de mensagens classificadas pelo professor.

Motivação Extrínseca	Motivação Intrínseca	Desmotivação
<p>“No ambiente do Moodle foi postado um exercício que pede para elaborar um texto sobre as principais características do <i>Moodle</i> contudo a minha grande dúvida é no tocante as datas que foram postadas. Preciso saber a data real para o envio dessa atividade.”</p>	<p>“Olá amigos, gosto muito de cinema e tenho um bom acervo de filmes aqui. Acredito que a história de nosso país ainda é muito mal contada sobretudo de nosso Estado. Aspiro ainda o momento em que verei um cinema paraibano com requintes nacionais onde contaremos nossas histórias nossas literaturas e nossas conquistas desde os tempos coloniais.”</p>	<p>“Segue em anexo a propaganda de um computador.”</p>
<p>“Os processos de aprendizagens se dão por meio das relações sociais de modo geral em vários níveis desde as mais antigas até as mais atuais modernas como as que estão em nosso cotidiano as que estão nas instituições formais de ensino. Também nos educamos e aprendemos através do Cinema visto que o mesmo traz sempre em suas produções as criações tecnológicas nas quais se debate reflete sobre as mesmas. Portanto nos proporcionando um aprendizado sobre as tecnologias.”</p>	<p>“Boa tarde colegas do curso de Letras sou professora da rede estadual do RN. Sou apaixonada pela Língua Portuguesa e o curso de Letras para mim sempre foi um sonho a ser realizado tenho mais afinidade com a gramática leio pouco mas costume está em sintonia com os acontecimentos importantes.”</p>	<p>“Bom estou pensando em trancar o curso, pois está muito corrido, no entanto enquanto isso não acontece estou por aqui.”</p>
<p>“Boa noite professor, gostaria de saber se minha postagem está de acordo com a atividade passada.”</p>	<p>“Caro Professor: acredito que o compromisso e o foco no curso é primordial para um bom aprendizado e melhores perspectivas futuras.”</p>	<p>“Minha cidade não possui salas de cinema e perdi contato com esse tipo de entretenimento e aprendizado.”</p>

Fonte: Próprio Autor.

4.4.1.2 Cálculo do Tf-Idf

Normalmente, o peso de Tf-Idf é composto por dois cálculos principais: o primeiro é o da frequência normalizada (Tf), ou seja, o número de vezes que uma palavra é identificada em um documento, dividido pelo número total de palavras do documento; O segundo cálculo é a frequência do documento inverso (Idf), calculado como o logaritmo do número de documentos dividido pela quantidade de documentos no qual o termo específico aparece (MANNING, 2008).

- Tf: Frequência de prazo, verificou a frequência com que um termo ocorre em um documento. Uma vez que cada documento é diferente em comprimento, é possível que um termo apareça muito mais vezes em documentos longos do que os mais curtos. Assim, o termo frequência foi dividido pelo comprimento do documento (o número total de termos no documento) como forma de normalização:

$$Tf(t) = \frac{\text{Número de vezes que o termo } t \text{ aparece em um documento}}{\text{Número total de termos no documento}} \quad (4)$$

- Idf: Frequência inversa do documento, que mede o quão importante é um termo. Ao computar TF, todos os termos são considerados igualmente importantes. No entanto, se sabe que certos termos, como “de”, “para”, e “isto”, podem aparecer muitas vezes, mas têm pouca importância. Assim, precisa-se pesar os termos recorrentes, enquanto aumentasse os termos raros, ao computar o seguinte:

$$Idf(t) = \log \frac{\text{Número total de documentos}}{\text{Número de documentos com termo } t} \quad (5)$$

4.4.1.3 Dicionários Gerados

De acordo com a técnica Tf-Idf, foram elencadas as palavras mais citadas pelos alunos em cada um dos grupos de Motivação. Com base nestes dados e com o apoio do docente, foram elaborados dicionários utilizados para o cálculo da polaridade na fase da classificação. O docente adicionou palavras (*features*) que considerava importantes em cada um dos grupos. Foi permitida tal atitude,

pois normalmente podem existir palavras relacionadas ao conteúdo do fórum que podem nem ter sido comentadas e o professor teve a intenção de deixar registrado no dicionário. Um exemplo dos dicionários gerados e da lista de *stopwords* utilizadas estão apresentados no Apêndice A do presente documento.

4.4.2 Etapa de Fuzzificação da *FuzzySD*

A etapa de fuzzificação é responsável por identificar as variáveis difusas e conectá-las a suas respectivas variáveis linguísticas. As variáveis difusas foram pensadas e apresentadas no tópico anterior, sendo essas: Motivação, Autorregulação e Interação (apresentadas no Quadro 4.4).

Quadro 4.4. Valores das Variáveis Linguísticas.

MÉTRICAS		VARIÁVEIS LINGUÍSTICAS
Entradas	Autorregulação	<ul style="list-style-type: none"> • “Autorregulado”, • “Em Autorregulação”, • “Não Autorregulado”.
	Interação	<ul style="list-style-type: none"> • “Participativo”, • “Pouco Participativo”, • “Não Participativo”.
	Motivação	<ul style="list-style-type: none"> • “Desmotivado”, • “Motivado Extrinsecamente”, • “Motivado Intrínseca”.

Fonte: Próprio Autor.

4.4.2.1 Definição das Funções de Pertinência Fuzzy

A ação seguinte na etapa de fuzzificação é a definição das funções de pertinência das variáveis selecionadas. Não existem regras definitivas na metodologia *fuzzy* para a escolha de tais funções que representam o conhecimento de um especialista no assunto em questão ou informações extraídas de um banco de dados.

Deste modo, é executada uma transformação das variáveis de entrada do problema em valores *fuzzy*. Assim, para cada valor real de entrada é

aplicada uma função de pertinência que retorna o grau de pertinência da proposição. Este valor é necessariamente limitado entre 0 a 1. O grau de pertinência 0 significa que o valor não pertence ao conjunto, enquanto o grau de pertinência 1 indica que o valor é uma representação completa do conjunto.

Algumas justificativas e testes fizeram com que o estudo convergisse a definir a função de pertinência do tipo triangular e trapezoidal como a mais adequada para as entradas da abordagem apresentada dentre as quais: a aplicação em pesquisas correlatas sobre avaliação multidimensional e concordância por parte do docente consultor.

i) Autorregulação

Diante da métrica Autorregulação do aluno, a *FuzzySD* contabilizou para o cálculo desta métrica as seguintes variáveis: média das notas dos alunos no curso, o número de edições de recursos pedagógicos e o número de mensagens de *feedback* entre professor e aluno. Tais pontos foram justificados nas seções anteriores.

Inicialmente, as entradas foram normalizadas, pois tais valores apresentavam escalas distintas. De acordo com as variáveis que compõem a métrica Autorregulação do aluno, é possível descrever o universo amostral destes critérios por meio das Equações (6), (7) e (8):

$$\text{Média das notas do aluno (sm)} = \{0 \leq sm \leq 10 \mid sm \in \mathbb{R}\}; \quad (6)$$

$$\text{Número de edições de recursos (ed)} = \{ed \in \mathbb{N}, ed > 0\}; \quad (7)$$

$$\text{Número de mensagens de } \textit{feedback} \text{ (fe)} = \{fe \in \mathbb{N}, fe > 0\}. \quad (8)$$

Em seguida, foi calculada a média dos valores normalizados de cada uma das variáveis e geradas as funções de pertinência e os gráficos relacionados a cada uma das variáveis linguísticas (apresentados no Quadro 4.5 e na Figura 4.10). Nesta figura, as variáveis linguísticas foram apresentadas e diferenciadas por meio de cores distintas no gráfico (roxo, azul e vermelho). As linhas azuis representam o grupo de “Alunos Não Autorregulados”, as linhas azuis o grupo de “Alunos Em Autorregulação” e, por

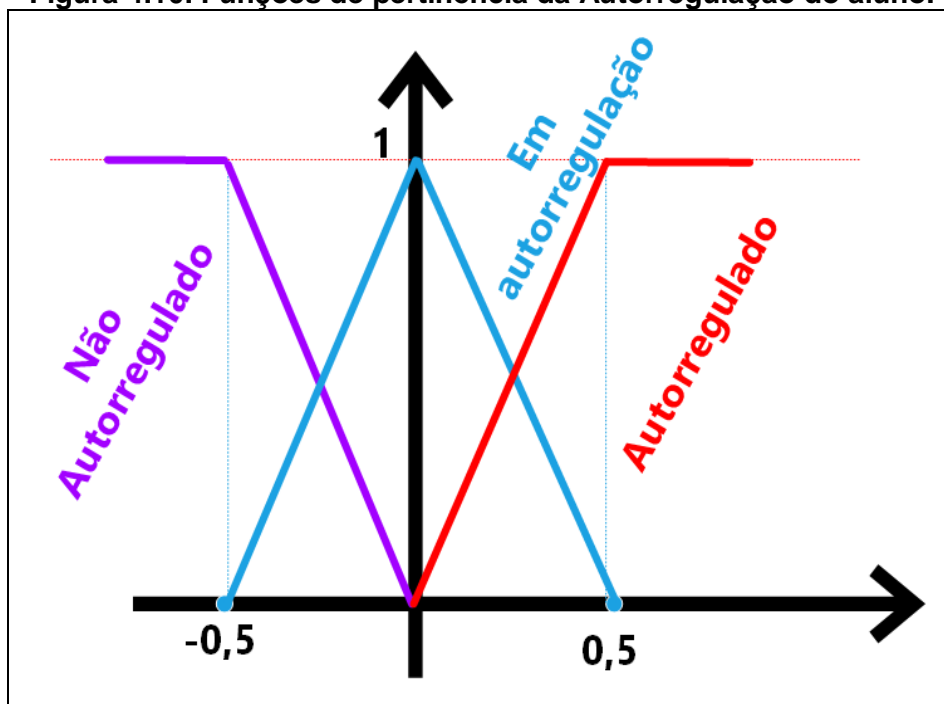
fim, em linhas vermelhas foi representado o grupo de “Alunos Autorregulados”.

Quadro 4.5. Funções de pertinência e variáveis linguísticas da Autorregulação.

VARIÁVEL LINGUÍSTICA	FUNÇÕES DE PERTINÊNCIA
Não autorregulado	$f_{\text{NãoAutorregulado}}(x) = \begin{cases} 1, & \text{se } x < -0,5; \\ -2x, & \text{se } -0,5 \leq x \leq 0; \\ 0, & \text{se } x > 0. \end{cases}$
Em autorregulação	$f_{\text{EmAutorregulação}}(x) = \begin{cases} 0, & \text{se } x < -0,5; \\ \frac{x+0,5}{0,5}, & \text{se } -0,5 \leq x \leq 0; \\ \frac{0,5-x}{0,5}, & \text{se } 0 < x \leq 0,5; \\ 0, & \text{se } x > 0,5. \end{cases}$
Autorregulado	$f_{\text{Autorregulado}}(x) = \begin{cases} 0, & \text{se } x < 0; \\ \frac{x-0,5}{0,5}, & \text{se } 0 \leq x \leq 0,5; \\ 1, & \text{se } x > 0,5. \end{cases}$

Fonte: Próprio Autor.

Figura 4.10. Funções de pertinência da Autorregulação do aluno.



Fonte: Próprio Autor.

ii) Interação

A etapa seguinte foi repetir as análises diante de todas as outras métricas definidas pela abordagem avaiativa. O segundo estágio consistiu em

calcular o critério Interação do Aluno de acordo com as variáveis que o compõem: número de visualizações, o número de mensagens trocadas entre alunos e o número de dias sem acessar o ambiente. No presente cenário, as mensagens trocadas entre os alunos foram coletadas exclusivamente de fóruns, pois os bate-papos foram pouco utilizados pelos professores e discentes. Por serem entradas quantitativas, é possível detalhar o universo amostral de tais informações por meio das Equações (9), (10) e (11).

$$\text{Número de Visualizações } (nv) = \{0 \leq nv \leq x \mid x \in \mathbb{N}\} \quad (9)$$

$$\text{Mensagens trocadas entre alunos } (ma) = \{0 \leq ma \leq x \mid x \in \mathbb{N}\} \quad (10)$$

$$\text{Dias sem acessar o ambiente } (lac) = \{0 \leq lac \leq x \mid x \in \mathbb{N}\} \quad (11)$$

A partir destas informações, é possível novamente gerar as funções de pertinência associadas às avaliações de cada uma das categorias sugeridas pelo especialista. Na Figura 4.11 e no Quadro 4.6 foram ilustradas as funções utilizadas e a visualização de tais informações.

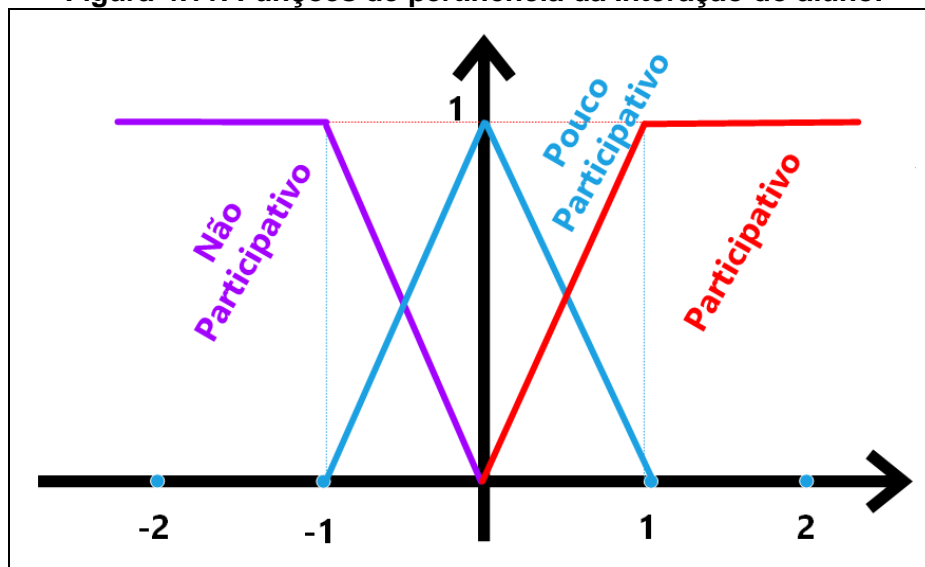
Quadro 4.6. Funções de pertinência e variáveis linguísticas da Interação.

VARIÁVEL LINGUÍSTICA	FUNÇÕES DE PERTINÊNCIA
Não Participativo	$f_{\text{NaoParticipativo}}(x) = \begin{cases} 1, & x < -1 \\ x, & -1 \leq x \leq 0 \\ 0, & x > 0 \end{cases}$
Pouco Participativo	$f_{\text{PoucoParticipativo}}(x) = \begin{cases} 0, & x < -1 \\ x + 1, & -1 \leq x \leq 0 \\ 1 - x, & 0 < x \leq 1 \\ 0, & x > 1 \end{cases}$
Participativo	$f_{\text{Participativo}}(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$

Fonte: Próprio Autor.

O docente consultor concordou com as categorias das métricas, mas as funções de pertinência precisavam ser validadas por este profissional por meio de um questionário executado nas próximas etapas.

Figura 4.11. Funções de pertinência da Interação do aluno.



Fonte: Próprio Autor.

iii) Motivação

A última métrica de entrada analisada foi a Motivação. Novamente, foi investigado o universo amostral da métrica por meio das entradas que a influenciaram. A Motivação foi medida por meio da variável polaridade das mensagens e se identificou que os valores do universo amostral pertenciam ao conjunto dos reais, variando entre valores positivos e negativos, de acordo com as mensagens do curso analisado, como ilustrado na Equação (12).

$$\text{Polaridade das mensagens } (pm) = \{ x \in \mathbb{R} \} \quad (12)$$

A medida da polaridade de cada um dos alunos foi normalizada e, de acordo com os resultados, as funções de pertinência foram propostas e testadas. Os resultados podem ser observados na Figura 4.12 e o Quadro 4.7.

Inicialmente, foi definido que valores negativos representariam os grupos de alunos Desmotivados no curso, mesmo sendo valores normalizados. Foram executados testes analisando as faixas de cada um dos grupos com a polaridade real calculada. Percebeu-se que o grupo de alunos Desmotivados correspondia à faixa de alunos entre $-\infty$ e 0.

De acordo com as funções de pertinência do grupo “Motivado Intrinsecamente” e “Motivado Extrinsecamente”, foi possível perceber que a Motivação não possui um domínio definido para x , pois tais valores dependem do perfil de interação dos participantes em cursos *online* analisados. O valor x

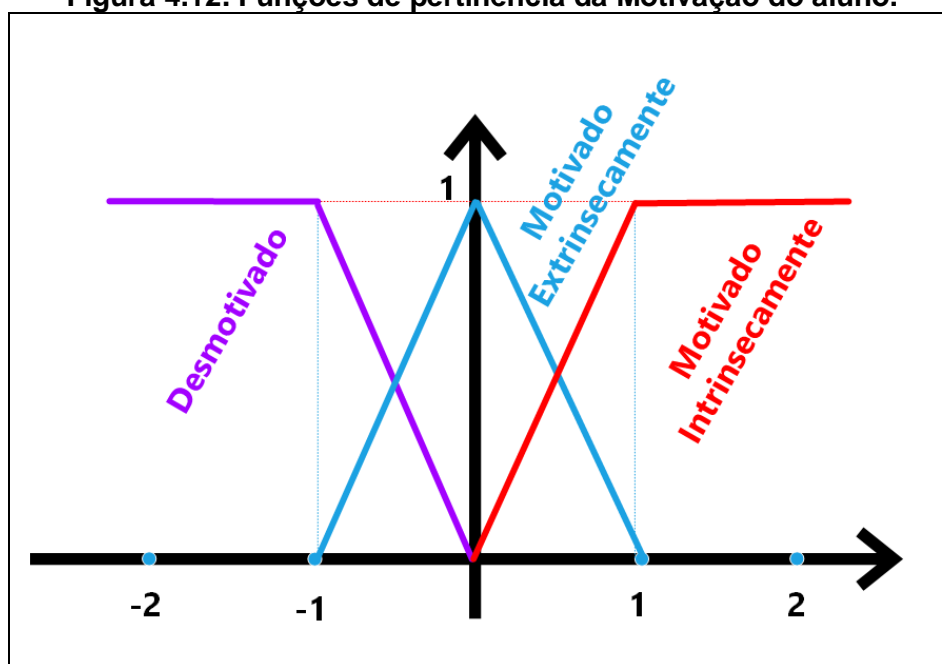
será descrito no estudo de caso apresentados no Capítulo 5.

Quadro 4.7. Funções de pertinência e variáveis linguísticas da Motivação.

VARIÁVEL LINGUÍSTICA	FUNÇÕES DE PERTINÊNCIA
Desmotivado	$f_{\text{Desmotivado}}(x) = \begin{cases} 1, & \text{se } x < -1; \\ -x, & \text{se } -1 \leq x \leq 0; \\ 0, & \text{se } x > 0. \end{cases}$
Motivado Extrinsecamente	$f_{\text{MotivadoExtrinsecamente}}(x) = \begin{cases} 0, & \text{se } x < -1; \\ x + 1, & \text{se } -1 \leq x < 0; \\ 1 - x, & \text{se } 0 \leq x \leq 1; \\ 0, & \text{se } x > 1. \end{cases}$
Motivado Intrinsecamente	$f_{\text{MotivadoIntrinsecamente}}(x) = \begin{cases} 0, & \text{se } x < 0; \\ x, & \text{se } 0 \leq x \leq 1; \\ 1, & \text{se } x > 1. \end{cases}$

Fonte: Próprio Autor.

Figura 4.12. Funções de pertinência da Motivação do aluno.



Fonte: Próprio Autor.

O professor consultor afirma que “motivação intrínseca é o tipo de motivação mais difícil de se atingir e pode ser interpretada como um tipo de motivação ideal em sistemas de aprendizagem”. Por este fato, foi considerado que o grupo de alunos “Motivados Extrinsecamente” correspondeu à faixa entre -1 e 1 e a “Motivados Intrinsecamente” à faixa entre 0 e $+\infty$. Tal estratégia foi

adotada para que os grupos não se sobrepusessem e fosse possível identificá-los.

4.4.3 Etapa de Inferência

A partir do método de inferência, ou seja, utilizando o conhecimento dos especialistas para desenvolver um raciocínio, modelam-se as incertezas dentro de um processo para tomada de decisão, fundamentado na existência de conexões entre causa e efeito, por meio de regras, naturais ou não, denominadas regras de inferência.

É importante ressaltar que tais regras condicionais podem ser inseridas no contexto da base de conhecimento. Caso um especialista julgue necessário adicionar alguma informação posterior à modelagem, regras como modificadores linguísticos, condições, proposições poderão ser inseridas no modelo a fim de tornar o sistema ainda mais robusto.

As regras foram definidas e depois examinadas de maneira distinta. No processo de inferência *fuzzy*, os princípios da lógica *fuzzy* foram usados para combinar as regras "SE-ENTÃO" existentes na base de regras com os dados de entrada. O resultado desta combinação representou uma região *fuzzy* de saída que se relaciona com a saída do processo. Os passos para o alcance dessa região *fuzzy* de saída foram constituídas por:

- definir as regras que estejam ativadas em um determinado momento;
- determinar a saída *fuzzy* de cada uma das regras ativadas; e
- combinar todas as saídas *fuzzy* calculadas.

A definição dos termos verbais, funções de pertinência e a inferência da base de regras, provenientes da teoria da lógica *fuzzy*, foram etapas executadas em sequência e resultaram na geração do modelo matemático da abordagem *fuzzy*.

A base de conhecimento é composta pela base de regras e pela base de dados. A base de dados armazena as definições das funções de pertinência, bem como as definições de normalização e discretização dos

universos de discurso. A base de regras possui o conjunto de regras de produção que apresentam a estrutura SE <premissa> ENTÃO <consequente>.

Como mencionado, o módulo de inferência processa os dados de entrada e as regras de produção inferindo as ações de controle nebulosas, de acordo com o estado do sistema. É necessário definir o conjunto de regras pertencentes à base de conhecimento. Com isto, há uma regra para cada termo linguístico de entrada, multiplicados entre si, resultando em 19 regras condicionais associadas aos seus respectivos termos de saída. Tais regras foram obtidas por meio de conversas com o professor especialista e, principalmente, por meio da aprendizagem de regras extraídas por meio de técnicas de classificação, especificamente as árvores de decisão.

As árvores de decisão do tipo Random Tree, NBTree, ADTree e J48 permitiram a identificação de regras que inicialmente o especialista não conseguiu perceber em uma análise pontual. As principais árvores de decisão obtidas nesta fase da pesquisa foram anexadas ao final do documento, no Apêndice B.

As regras condicionais seguem o mesmo formato do seguinte exemplo: **SE** Autorregulação é “Autorregulado” **E** Interação é “Participativo” **E** Motivação é “Motivado Intrinsecamente” **ENTÃO** Evasão é “Persistente”. No Quadro 4.8 são apresentadas algumas destas regras.

Quadro 4.8. Exemplo da Base de Regras obtidas no modelo.

Regra	Autorregulação	Interação	Motivação	Evasão
1	Autorregulado	Participativo	Motivado Intrinsecamente	Persistência
2	Autorregulado	Participativo	Motivado Extrinsecamente	Persistência
3	Autorregulado	Pouco Participativo	Motivado Intrinsecamente	Persistência
4	Autorregulado	Não Participativo	Desmotivado	Evasão
5	Autorregulado	Não Participativo	Motivado Intrinsecamente	Persistência
6	Em Autorregulação	Pouco Participativo	Motivado Extrinsecamente	Persistência
7	Em Autorregulação	Participativo	Motivado Extrinsecamente	Persistência

Continua

Continuação

Regra	Autorregulação	Interação	Motivação	Evasão
8	Em Autorregulação	Não Participativo	Motivado Extrínsecamente	Persistência
9	Em Autorregulação	Pouco Participativo	Motivado Intrínsecamente	Evasão
10	Em Autorregulação	Participativo	Desmotivado	Persistência
11	Em Autorregulação	Pouco Participativo	Desmotivado	Persistência
12	Não Autorregulado	Participativo	Desmotivado	Persistência
13	Não Autorregulado	Pouco Participativo	Desmotivado	Evasão
14	Não Autorregulado	Não Participativo	Motivado Extrínsecamente	Evasão
15	Não Autorregulado	Não Participativo	Motivado Intrínsecamente	Persistência
16	Não Autorregulado	Pouco Participativo	Motivado Intrínsecamente	Evasão
17	Não Autorregulado	Pouco Participativo	Motivado Extrínsecamente	Evasão
18	Não Autorregulado	Participativo	Motivado Extrínsecamente	Persistência
19	Não Autorregulado	Pouco Participativo	Desmotivado	Persistência

Fonte: Próprio Autor.

4.4.4 Etapa de Defuzzificação

A última etapa da análise *fuzzy* é a etapa de defuzzificação. Esta etapa é responsável por transformar os resultados *fuzzy* do processo em valores discretos. Há diversas técnicas que podem ser utilizadas nesta fase, tais como: o método dos máximos, da média dos máximos e o método do Centro de Gravidade (CoG), *etc* (HUDEC, 2016).

Na pesquisa foi utilizado o CoG, pois foi a metodologia mais utilizada no contexto dos trabalhos de avaliação multidimensionais apresentados no Capítulo 2 e foi a técnica que obteve os melhores resultados no domínio do estudo de caso realizado. Além disto, este método produziu resultados mais adequados, de acordo com a entrevista com o professor, em comparação com

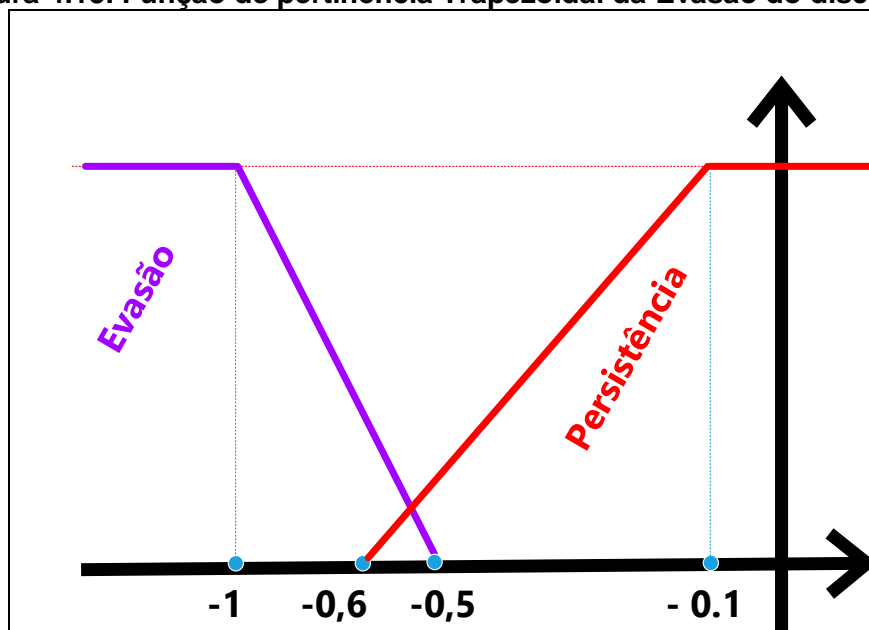
o método bissetor, que é também um método bastante utilizado na literatura.

Por meio do CoG, o valor numérico obtido representa o centro de gravidade da distribuição de possibilidade de saída do sistema *fuzzy*:

- determinar a abscissa do ponto centróide para cada saída ativada na inferência;
- calcular a área entre o grau de pertinência e o eixo x para cada saída ativada; e
- calcular a média ponderada dos pontos centróides pelas respectivas áreas.

Para finalizar o planejamento, o próximo passo foi definir as saídas *fuzzy* esperadas da variável Evasão, sendo essas: “Persistência” e “Evasão”, determinados pelas funções de pertinência trapezoidal (Figura 4.13 e Quadro 4.9). Alguns testes com exemplos e conversas com o especialista foram realizados nessa fase do planejamento da abordagem.

Figura 4.13. Função de pertinência Trapezoidal da Evasão do discente.



Fonte: Próprio autor.

Uma vez obtido o conjunto *fuzzy* de saída por meio do processo de inferência, no estágio de defuzzificação é efetuada a interpretação dessa informação. Isto se faz necessário pois, em aplicações práticas, geralmente são requeridas saídas precisas.

Quadro 4.9. Funções de pertinência e variáveis linguísticas da métrica de saída Evasão.

VARIÁVEL LINGUÍSTICA	FUNÇÕES DE PERTINÊNCIA
Evasão	$f_{\text{Evasão}}(x) = \begin{cases} 1; & x < -1 \\ \frac{-0,5-x}{0,5}; & -1 \leq x \leq -0,5 \\ 0; & x > -0,5 \end{cases}$
Persistência	$f_{\text{Persistência}}(x) = \begin{cases} 0; & x < -0,6 \\ \frac{x+0,6}{0,5}; & -0,6 \leq x \leq -0,1 \\ 1; & x > -0,1 \end{cases}$

Fonte: Próprio autor.

4.5 Avaliação da *FuzzySD*

Após definir a estrutura geral e as métricas relevantes na abordagem *FuzzySD*, foi planejada uma avaliação da abordagem apresentada em um contexto real de cursos a distância. Inicialmente, a análise piloto foi realizada com dados educacionais obtidos de alguns cursos online do IFPB Virtual. Tão importante quanto a concepção da abordagem avaliativa é a sua avaliação para identificar se realmente a *FuzzySD* se mostrou eficiente. Esta etapa é parte essencial na análise do presente modelo e foi realizada de forma apropriada ao desafio da validação de um sistema *fuzzy*.

É importante mencionar que o presente estudo se baseia nas análises experimentais apresentadas por Wohlin *et al.* (2000). Tal estratégia sugere que a realização de um estudo experimental, no âmbito da engenharia de *software*, deve executar as seguintes etapas: definição, planejamento, operação, análise e interpretação, validação e empacotamento. Deste modo, esta fase consiste na etapa de analisar se o estudo experimental proporcionou os resultados esperados. A avaliação de uma forma ampla é o processo de verificar para que serve e quanto serve um sistema computacional analisado (WAINER, 2007).

A avaliação da presente abordagem ocorreu seguindo duas estratégias principais: avaliação subjetiva e validação preditiva. A fase subjetiva foi executada por meio de uma abordagem *ad-hoc*, a partir da qual os revisores (neste caso, docentes de EaD) aplicaram seus conhecimentos específicos na verificação dos grupos, das regras *fuzzy* e do resultado da análise. Por

exemplo, todas as funções de pertinência das entradas e saída foram discutidas e ajustadas de acordo com o *feedback* dos especialistas. Estas conversas com professores e coordenadores da EaD ocorreram ao longo de todo o processo de planejamento e execução da abordagem, sempre por meio de entrevistas semiestruturadas. Os resultados das avaliações subjetivas foram relevantes para definir as métricas de entrada, os grupos linguísticos, as funções de pertinências e os grupos de regras *fuzzy*.

Concomitante, a validação preditiva ocorreu por meio da comparação de cenários e análise do desempenho do modelo apresentado. Este modelo de avaliação é usado para prever o comportamento do sistema, para isto comparações foram feitas entre a predição e o comportamento real (SARGENT, 2005).

Tipicamente em um processo de aprendizagem supervisionada em mineração de dados, após o pré-processamento e a formatação, os dados são fragmentados em dois subconjuntos, denominados base de treinamento e teste. Neste estudo, a amostra analisada foi segmentada de forma aleatória em treinamento e teste, respeitando a proporção 70% e 30% da amostra original, respectivamente.

Na primeira etapa, um algoritmo de indução de conhecimento é aplicado à base de treinamento. Com isso, obtém-se um modelo “treinado”, que de certa forma representa o conhecimento extraído. Na segunda etapa, o modelo obtido é aplicado ao fragmento da base de dados denominado amostra teste. Como a base de testes também é previamente rotulada, se pode medir a taxa de acerto do modelo, comparando o resultado obtido com a rotulação disponível na amostra teste.

No presente estudo, a amostra teste passou por uma validação cruzada (*k-fold cross-validation*), que consiste em dividir a base de dados em k partes (*folds*). Destas, $k-1$ partes são utilizadas para o treinamento e uma serve como base de testes. O processo é repetido k vezes, de forma que cada parte seja usada uma vez como conjunto de teste. Ao final, a correção total é calculada pela média dos resultados obtidos em cada etapa, obtendo-se assim uma estimativa da qualidade do modelo de conhecimento gerado e permitindo análises estatísticas (HAIR *et al.*, 2009).

Neste estudo, para avaliar o desempenho da abordagem FuzzySD e de outras técnicas, foram comparados os resultados entre as técnicas por meio de validação cruzada utilizando as bases de teste (*10 folds cross validation*).

Outro aspecto relevante, é que a avaliação do desempenho de um modelo de classificação geralmente envolve a análise da habilidade de previsão ou correta separação das classes. Uma ferramenta muito utilizada para essa atividade é a técnica conhecida como “*confusion matrix*” ou matriz de confusão. Nesta técnica, medidas sobre a qualidade da classificação da amostra teste são calculadas a partir de uma matriz que armazena os exemplos que foram classificados corretamente e incorretamente para cada classe, denominada matriz de confusão. Em problemas de classificação binária, uma classe é definida como positiva e a outra como negativa. Assim, os conceitos de FP, FN, VP e VN são descritos da seguinte forma:

- (i) Falsos Positivos (FP) são exemplos que são preditos à classe positiva, mas pertencem à classe negativa;
- (ii) Falsos Negativos (FN) são exemplos que são preditos como à classe negativa, mas pertencem à classe positiva;
- (iii) Verdadeiros Positivos (VP) são exemplos preditos corretamente como positivos e;
- (iv) Verdadeiros Negativos (VN) são exemplos que são preditos corretamente como negativos.

Para um problema de classificação binária, ou seja, que possui apenas duas classes, a matriz de confusão é mostrada no Quadro 4.10.

Quadro 4.10. Matriz de confusão.

Classe Verdadeira	Classe Predita		
	Positiva (Persistente)	Negativa (Evadido)	
Positiva (Persistente)	Verdadeiros Positivos (VP)	Falsos Positivos (FP)	(VP+FP)
Negativa (Evadido)	Falsos Negativos (FN)	Verdadeiros Negativos (VN)	(FN+VN)
	(VP+FN)	(FP+VN)	Total

Fonte: Próprio Autor.

Após ilustrar as partes de uma matriz de confusão, este trabalho apresenta tais estruturas de modo mais sucinto, respeitando o formato exemplificado pela matriz de confusão M a seguir:

$$M = \begin{array}{c|cc} & + & - \\ \hline + & VP & FP \\ - & FN & VN \end{array}$$

Nesta notação, as classes de ocorrências previstas pelas abordagens serão representadas pela primeira linha e as classes de ocorrências existentes no mundo real foram representadas pela primeira coluna. O exemplo, utilizou-se de um exemplo em que há duas classes possíveis para o modelo: + e -. Outra característica relevante desta estrutura é que a diagonal principal da matriz representa todas as ocorrências que foram corretamente classificadas pelo modelo.

Em relação à investigação sobre o desempenho da *FuzzySD*, uma atividade se destacou neste processo: a avaliação da precisão (ou acurácia) da abordagem após a defuzzificação. Entende-se por precisão a razão entre o número de acertos do modelo (acertos positivos = $VP + VN$) e o número total previsões realizadas (total previsões realizadas = $VP+VN+FP+FN$), como ilustrado na Equação (13). Tais valores deste cálculo são obtidos por meio das matrizes geradas nas análises.

$$precisão \text{ ou } acurácia = \frac{Acertos \text{ Positivos}}{Total \text{ de } previsões \text{ realizadas}} \quad (13)$$

Para saber o quão confiáveis são os dados provenientes da abordagem *fuzzy* executada, foi utilizada a técnica matriz de confusão. Além disto, é possível derivar medidas e, conseqüentemente, verificar erros oriundos do processo analisado. O Quadro 4.11 apresenta as principais medidas utilizadas.

Quadro 4.11. Medidas calculadas a partir da Matriz de Confusão.

MEDIDA AVALIATIVA	DESCRIÇÃO	CÁLCULO
Sensibilidade (S)	A capacidade do sistema em predizer corretamente a condição para casos que realmente a têm.	$S = \frac{\text{Acertos Positivos}}{\text{Total de Positivos}},$ onde “Total de Positivos” representa a soma dos “Verdadeiros Positivos” e de “Falsos Negativos”.
Especificidade (E)	A capacidade do sistema em predizer corretamente a ausência da condição para casos que realmente não a têm.	$E = \frac{\text{Acertos Negativos}}{\text{Total de Negativos}},$ onde “Total de Negativos” representa os “Acertos Negativos” somados aos “Falsos Positivos”.
Eficiência (Ef)	É a média aritmética da Sensibilidade e Especificidade. Na prática, a sensibilidade e a especificidade variam em direções opostas. Isto é, geralmente, quando um método é muito sensível a positivos, tende a gerar muitos falso-positivos, e vice-versa.	$Ef = \frac{(S + E)}{2}$

Fonte: Próprio autor.

A avaliação da sensibilidade, especificidade e eficiência também podem ser obtidos por meio de coeficientes de concordância derivados da matriz de confusão, sendo que estes podem ser expressos como concordância total ou para classes individuais (Quadro 4.11).

4.6 Arquitetura da *FuzzySD*

Outro aspecto pertinente a ser destacado é o funcionamento da *FuzzySD*, por meio de uma breve discussão sobre a arquitetura do projeto.

A abordagem avaliativa construída está estruturada conforme algumas diretrizes de padrões de projeto e o paradigma orientado a objetos. A linguagem base utilizada no desenvolvimento foi *Python*, versão 3.2, por ter um bom desempenho ao lidar com grandes quantidades de dados e pela variedade de bibliotecas gratuitas e abertas disponíveis, como mencionado na Seção 4.2.

A *FuzzySD* pode ser descrita como uma abordagem *offline*, pois suas análises acessaram os dados a partir de arquivos de *log* e registros em banco de dados local (exigência da instituição parceira). A *FuzzySD* avaliou dados históricos de cursos concluídos na sua avaliação.

A abordagem *FuzzySD* lida com informações extraídas do banco de dados dos cursos do IFPB Virtual. Inicialmente, tabelas resumo são organizadas curso por curso (arquivo apresentado no Apêndice C), organizando as ações dos alunos no ambiente e suas informações principais. Os dados são extraídos e pré-processados para analisar o significado de tais valores, de acordo com as métricas de entrada envolvidas (arquivo apresentado no Apêndice D). Dependendo das características do ambiente, o professor pode avaliar a motivação, autorregulação e interação dos alunos de forma direta e gerar apenas este arquivo.

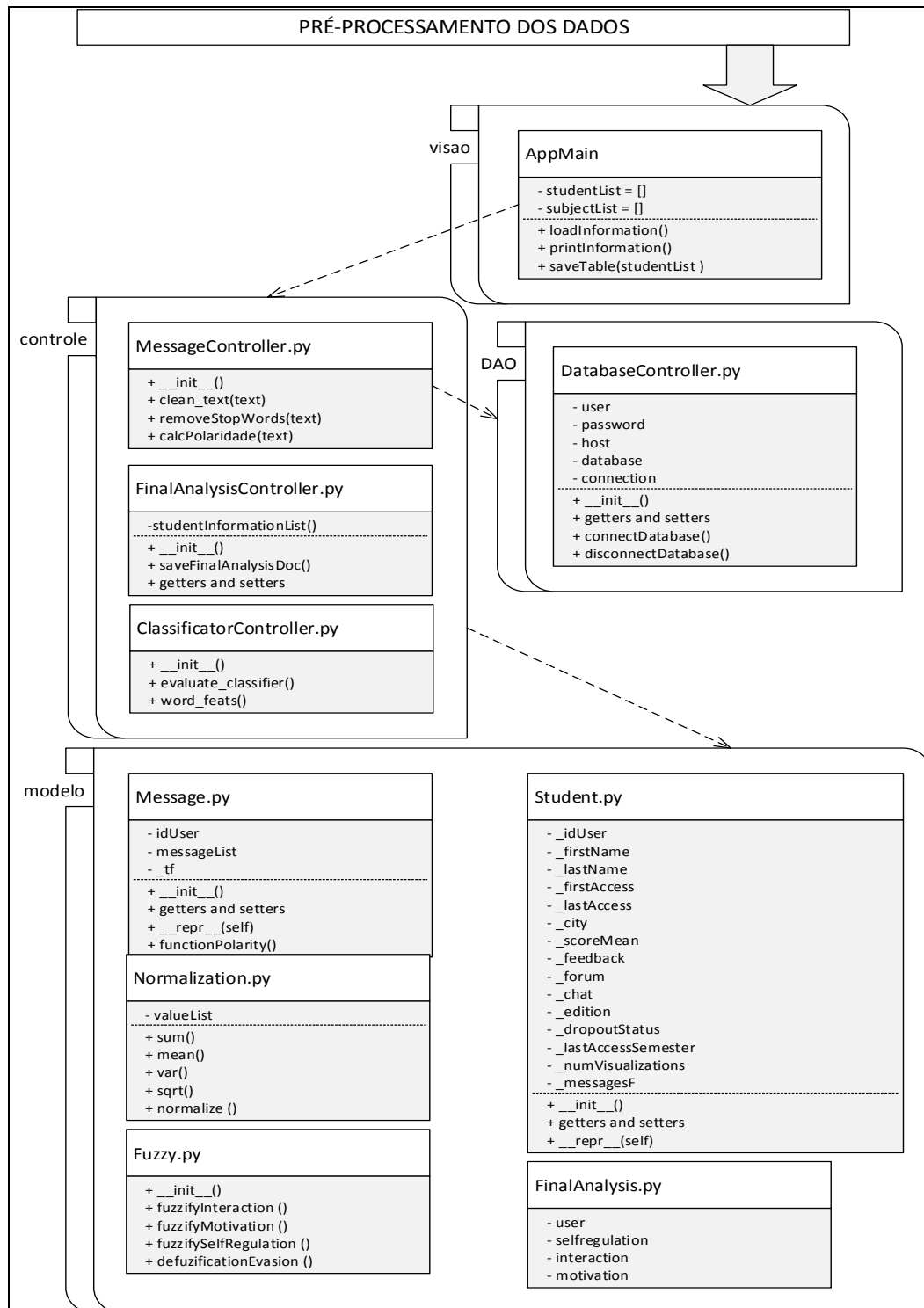
Com relação ao sistema *FuzzySD*, a abordagem foi concebida com base nos padrões GRASP (*General Responsibility Assignment Software Patterns*), MVC (Modelo, Visão e Controle) e o DAO (*Data Access Object*) para a comunicação com o banco de dados.

Na Figura 4.14, são ilustrados os principais pacotes, classes, métodos e arquitetura dos padrões utilizados. Neste diagrama, é importante ressaltar três classes: a *Fuzzy*, a *ClassifierController* e a *MessageController*.

A classe *Fuzzy*, presente no pacote modelo, é a responsável pelas funções associadas a lógica *fuzzy*, principalmente a fuzzificação das entradas e a defuzzificação. As outras duas classes controladoras destacadas (a

ClassifierController e a *MessageController*) são as responsáveis pela mineração dos textos das mensagens nos fóruns de discussão. Além disto, as estratégias de programação na arquitetura priorizaram o baixo acoplamento e a alta coesão da abordagem desenvolvida.

Figura 4.14. Diagrama de pacotes da FuzzySD.



Fonte: Próprio Autor.

4.7 Considerações Finais

O método científico e os resultados advindos da execução dos experimentos e da RS ao longo do estudo deram o subsídio para o planejamento e construção da abordagem avaliativa na identificação do perfil evasivo do aluno no ensino *online*, denominada *FuzzySD*. O propósito da pesquisa foi flexibilizar a análise de avaliações formativas e somativas, de acordo com a necessidade do docente no curso *online*.

No presente capítulo foram detalhadas as etapas percorridas e justificadas muitas decisões tomadas ao longo do processo. Esta discussão foi fundamental para o entendimento da abordagem apresentada e para a discussão dos resultados obtidos que serão apresentados no capítulo seguinte.

5. Apresentação e Análise dos Resultados da *FuzzySD*

Após definir as variáveis envolvidas no processo e a estrutura geral da abordagem no capítulo anterior, no presente capítulo é apresentada a arquitetura final e avaliada a eficiência da *FuzzySD*, diante do contexto de cursos *online* do IFPB Virtual.

5.1 Avaliação da Defuzzificação da Saída

A defuzzificação da saída foi investigada a partir de uma análise comparativa com a evasão real do sistema. Este tipo de investigação viabilizou os subsídios necessários para discutir sobre a precisão da abordagem frente circunstâncias reais de aprendizagem na EaD. De acordo com as definições explicitadas, este estudo considera como evadido os alunos que permanecem sem acessar e interagir com o AVEA um período, sem interrupções, superior a um quarto do tempo de execução do curso *online* analisado. Neste sentido, o presente estudo pretende avaliar se envolver fatores comportamentais do discente podem sinalizar sobre a evasão de modo tão eficiente quanto a avaliação temporal. Para estas análises um grupo de hipóteses foram postuladas para nortear o experimento, a saber:

H_{10} : Quem evade no curso é identificado pela abordagem (precisão $> 0,6$).

H_{11} : Quem evade no curso não é identificado pela abordagem (precisão $\leq 0,6$).

Outro aspecto relevante neste cenário é a investigação relacionada ao momento ideal para executar a abordagem avaliativa, ou seja, em que ponto do curso a abordagem *FuzzySD* retorna os melhores resultados (maior precisão). A abordagem foi construída para ser flexível quanto ao momento da análise, mas será que isto acontece em contextos reais de cursos *online*?

Estas hipóteses e o momento ideal de aplicação da *FuzzySD* foram avaliados por meio da medida precisão da abordagem em cada um dos cursos

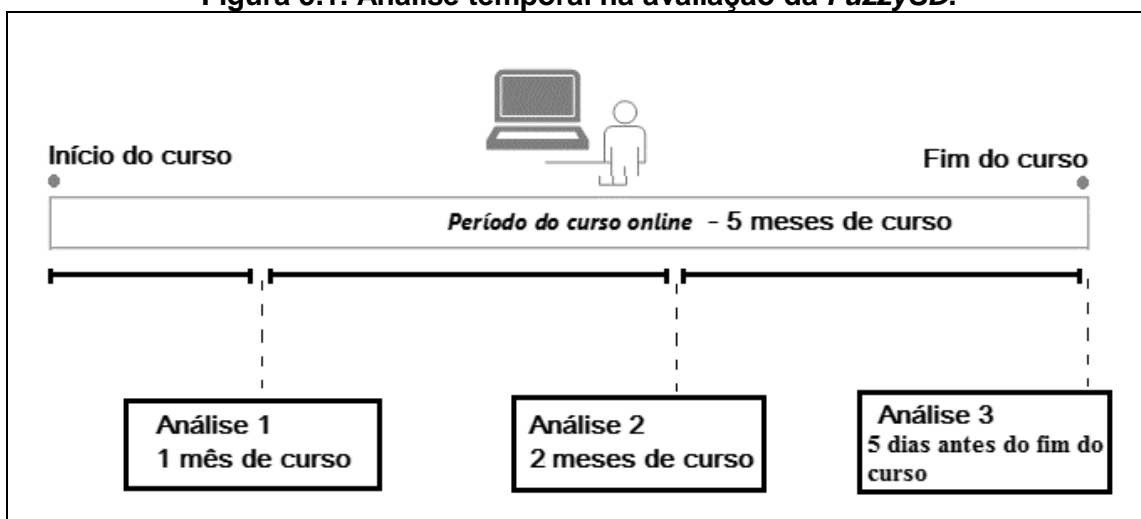
selecionados, mencionada na Seção 4.5. Os resultados detalhados foram apresentados e discutidos a seguir.

A precisão da abordagem foi avaliada em momentos distintos do curso *online*. Esta análise foi executada para confrontar algumas abordagens encontradas na literatura que focaram suas análises nos primeiros dias ou no meio do curso *online*. Para entender este comportamento, foram executadas análises em períodos diferentes nos cursos selecionados. Esta decisão pela análise temporal dos dados ocorreu devido à necessidade de verificar se, de fato, a precisão da *FuzzySD*, ao longo das semanas de curso, na identificação do perfil evasivo do discente.

As amostras selecionadas foram segmentadas de acordo com os períodos avaliativos do curso analisado, ou seja, o corte temporal analisado tem seu término após a identificação de um momento de avaliação no ambiente. Deste modo, foram definidos três cortes ao longo do semestre para avaliar sua precisão. O primeiro corte aconteceu um mês após o início do curso, o segundo corte após mais 2 meses e o terceiro corte após mais dois meses, que coincidiu com 5 dias antes da aplicação da última prova e o encerramento das atividades.

Na Figura 5.1, ilustra-se quando as análises foram feitas ao longo dos cursos *online*. É importante ressaltar que os cursos selecionados já tinham sido finalizados quando foram utilizados na avaliação da abordagem *FuzzySD*.

Figura 5.1. Análise temporal na avaliação da *FuzzySD*.



Fonte: Próprio Autor.

5.2 Cenário do Estudo de Caso

Para a pesquisa foram selecionados 4 períodos de dois cursos ministrados pelo IFPB Virtual, vinculado à UAB, de acordo com a aprovação e cooperação da instituição. Os cursos selecionados foram administrados por meio do AVEA Moodle. Estes cursos e períodos foram selecionados por dois motivos principais: (1) eram os cursos, dentro da plataforma do IFPB Virtual que tinham a maior quantidade de interações, alunos e atividades, e (2) representavam as turmas pioneiras dos seus respectivos cursos, fato que eliminava o risco de lidar com desbloqueio e transferência de alunos nas amostras (minimização das *filtragens* e *outliers*).

O primeiro curso analisado foi o curso de Letras, representado pela turma 2012 .2 do primeiro (148 alunos) e segundo períodos (65 alunos) (Tabela 5.1). No primeiro período, a turma se matriculou nas disciplinas Informática Básica, Fundamentos da Educação à Distância, Leitura e Produção de Texto I, Introdução à Linguística, História da educação brasileira, Introdução aos estudos literários e Inglês Instrumental. No segundo período, a turma se matriculou em Linguística I, Teoria Literária I, Filosofia da Educação, Literatura Brasileira I, Literatura e Ensino, Literatura Portuguesa I e Seminário Interdisciplinar I.

O segundo cenário analisado foi o curso de Segurança do Trabalho. Novamente, foi monitorada a turma 2012 .2, no primeiro (243 alunos) e no segundo (124 alunos) períodos. Os discentes matriculados cursaram as seguintes disciplinas no primeiro semestre: Introdução à Informática, Introdução à Higiene e Segurança do Trabalho, Inglês, Redação, Ética e Cidadania e Estatística Aplicada. No que se refere ao segundo período, foram oferecidas as disciplinas: Segurança do Trabalho I, Introdução à Administração, Psicologia do Trabalho, Metodologia da Pesquisa Científica, Primeiros Socorros e Desenho Técnico.

Inicialmente, é possível identificar que a amostra diminuiu quantitativamente entre o primeiro e o segundo períodos em ambos os cursos. Tal fato demonstra que neste cenário a evasão existe e deve ser combatida. Outro ponto relevante é que os alunos popularmente chamados de

“desbloqueados” não foram considerados na investigação, ou seja, aqueles alunos que são reprovados em alguma disciplina em um período e que acabam se matriculando novamente nessa nos períodos seguintes.

Todas as amostras analisadas foram segmentadas em treinamento e teste (como mencionado na Seção 4.5), respeitando, respectivamente, a proporção 70% e 30% da amostra. A amostra teste passou ainda por uma validação cruzada (10 *fold cross-tabulation*) e foi implementada em Python.

Tabela 5.1. Cenários selecionados para a avaliação da *FuzzySD*.

<i>Cenário Analisado</i>	Letras - 1º período	Letras - 2º período	Segurança do Trabalho - 1º período	Segurança do Trabalho - 2º período
<i>Tamanho da amostra</i>	148	65	243	124
<i>Período do curso</i>	2012.2	2013.1	2012.2	2013.1
<i>Disciplinas analisadas</i>	Informática Básica Fundamentos da Educação à Distância Leitura e Produção de Texto I Introdução à Linguística História da educação brasileira Introdução aos estudos literários Inglês Instrumental	Linguística I Teoria Literária I Filosofia da Educação Literatura Brasileira I Literatura e Ensino Literatura Portuguesa I Seminário Interdisciplinar I	Introdução a Informática Introdução à Higiene e Segurança do Trabalho Inglês Redação Ética e Cidadania Estatística Aplicada	Segurança do Trabalho I Introdução à Administração Psicologia do Trabalho Metodologia da Pesquisa Científica Primeiros Socorros Desenho Técnico
<i>Percentual final de evasão no curso</i>	35%	18%	49,5%	31%

Fonte: Próprio Autor.

O estudo foi conduzido de forma *offline*, o que significa que as análises são executadas *a posteriori* ao curso. Além disto, todos os participantes tiveram o mesmo ambiente e tempo para interagir com as atividades propostas. Deste modo, optou-se pela análise do semestre em sua totalidade, pois não havia possibilidade de o usuário sair apenas de uma disciplina e ser sinalizado como evadido pelo cálculo de evasão. Isto não seria possível, pois ao aluno acessar alguma das disciplinas no *Moodle* seria sinalizado como ativo no curso.

5.3 Avaliação da Defuzzificação da Saída: Análise Temporal

Após o detalhamento dos cenários envolvidos, é factível entender como ocorreu a análise sobre a precisão dos acertos da *FuzzySD*, como mencionado na Seção 4.5. O foco central desta avaliação consistiu em verificar a precisão da abordagem *FuzzySD* (nas amostras testes) e entender como ela se comportou em contextos de cursos *online* reais. Por meio da avaliação da defuzzificação da saída, este estudo comparativo foi realizado.

Inicialmente, é detalhado como o processo de fuzzificação e defuzzificação funciona na *FuzzySD*, por meio de um exemplo-base. Após isto, detalhes sobre os resultados da avaliação final são apresentados e discutidos.

5.3.1 Exemplo de Análise *Fuzzy*

Como mencionado, a *FuzzySD* utiliza para o cálculo da evasão *online* uma modelagem bidimensional do tipo Mamdani, utilizando as regras de inferência MIN-MAX associadas às regras condicionais para obtenção dos valores de saída. Além disto, executa-se o cálculo do centróide para defuzzificação e classificação resultante. As regras de inferência MIN-MAX têm o propósito de originar uma relação entre os termos das variáveis de entrada associadas aos termos da variável de saída, relacionado à tabela de regras, como uma multiplicação de matrizes, substituindo o produto pelo mínimo e a soma pelo máximo. Este método é proposto por Mamdani e Assilian (1975), além de ser amplamente utilizado em regras de inferência.

Para melhor elucidar o processo, é útil apresentar um exemplo-base na

discussão e, neste caso, utilizou-se um exemplo de um aluno do curso de Letras. Durante o curso selecionado, o aluno X interagiu até o momento da análise com alguns fóruns de discussão, editou respostas, conversou com colegas por meio do bate-papo, visualizou os materiais postados por colegas e professor, realizou o envio de atividades (*upload*), fez as atividades que serviam para pontuar e recebeu *feedbacks* dos professores em algumas interações. Numericamente, isso foi representado por 34 edições de recursos pedagógicos, 22 mensagens de *feedback* dos professores/tutores, 75 interações em fóruns, 140 mensagens postadas em fóruns, 1891 visualizações de recursos pedagógicos, nota média nas atividades 8,2 e nenhum dia ausente no ambiente. Além disto, o aluno produziu algumas mensagens nos fóruns de discussão que também foram analisadas.

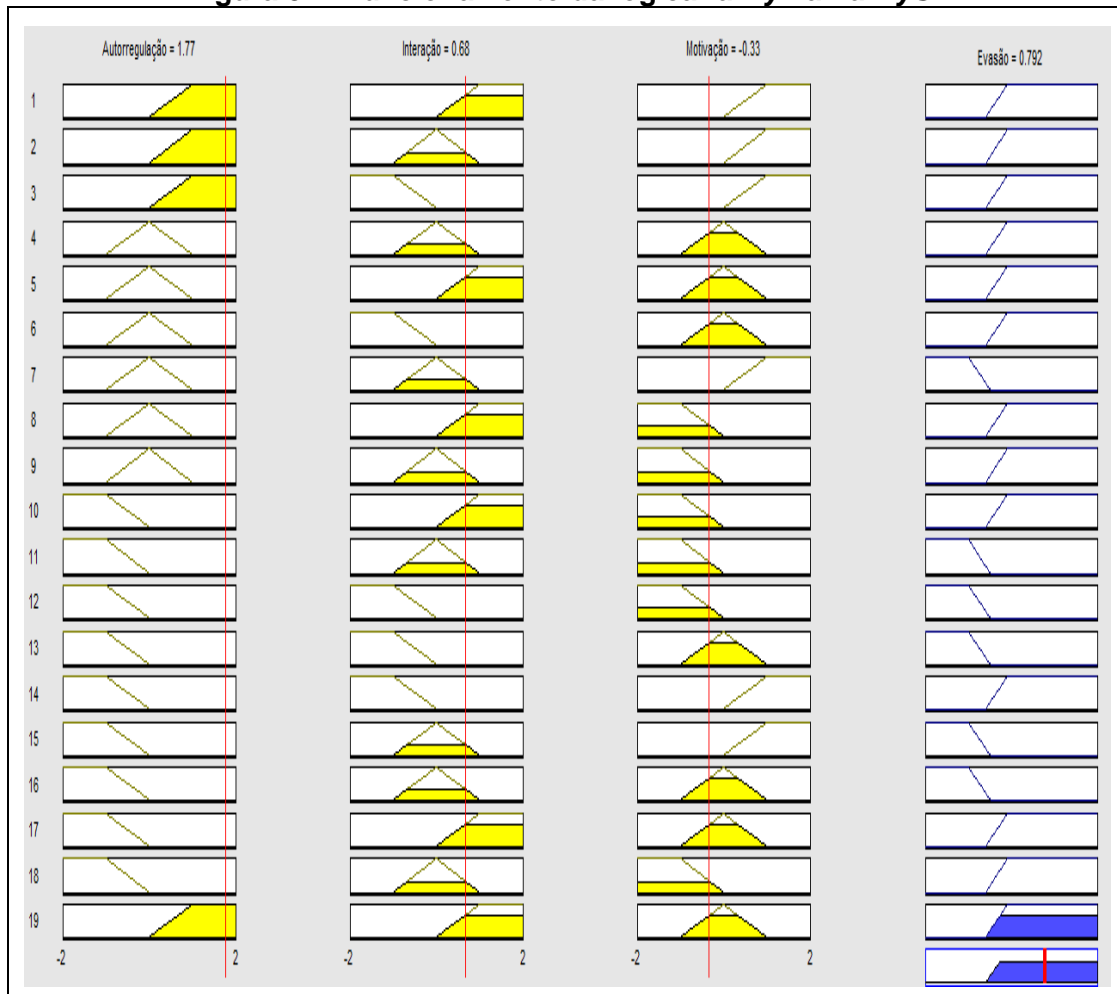
O próximo passo executado consistiu em normalizar tais valores a fim de corrigir as escalas diferenciadas entre as entradas. De acordo com tais informações, a próxima ação foi a identificação das variáveis de entrada *fuzzy* associadas às variáveis clássicas. A etapa de fuzzificação identificou que o aluno X do exemplo possuiu uma Autorregulação classificada como “Autorregulado”, uma Interação do tipo “Participativo” e a Motivação se enquadrou como “Motivado Extrinsicamente”; com os respectivos graus de pertinência: 1,77; 0,68 e -0,33.

Neste momento, as dezenove regras foram analisadas e testadas para permitir a combinação de regiões *fuzzy* de saída. Com isto, obteve-se uma região *fuzzy*, que representou a agregação das regras ativadas. Para esta agregação, na qual se utilizou a regra do máximo, obteve-se a região representada pela Figura 5.2. O gráfico apresentado foi gerado pela ferramenta *Matlab*, pois a *FuzzySD* não gera gráficos relacionados à fase de defuzzificação.

Após obter a região resultante da agregação das regras ativadas, aplicou-se um método de defuzzificação, que especifica um ponto na saída que melhor representa o conjunto *fuzzy*. O ponto localizado é um termo comum que combina a contribuição de cada regra ativada. A solução obtida pelo processo de defuzzificação é um número real que permite a identificação precisa da evasão do aluno. Neste exemplo, a Evasão correspondeu ao valor 0,792 que

representou uma situação de “Persistência” no curso *online*.

Figura 5.2. Funcionamento da lógica *fuzzy* na *FuzzySD*.



Fonte: Próprio Autor.

5.3.2 Precisão dos Cortes Temporais

As análises apresentadas no tópico anterior foram repetidas para todos os alunos presentes nos cursos *online* selecionados no estudo de caso observacional. As rodadas de avaliação foram executadas em três momentos distintos em cada um dos cursos (representado pelos 3 cortes).

Durante a avaliação da *FuzzySD*, calculou-se a evasão de todos os alunos da turma observando somente o comportamento dos discentes no curso, por meio dos registros nos arquivos de *logs*. Os resultados da abordagem foram comparados com a evasão real do sistema e a precisão da modelagem foi calculada.

Para isto, foram construídas as matrizes de confusão dos três cortes temporais selecionados (1º mês, 3º mês e por volta do 5º mês) dos dois cursos selecionados como cenário deste experimento. As matrizes de confusão, resultantes das amostras testes, podem ser observadas no Quadro 5.1. É possível observar que tais valores foram importantes para o cálculo da precisão e de outras métricas de avaliação da abordagem *fuzzy* a cada corte analisado (em seu treino e em seu teste). As análises relacionadas às amostras de treinamento foram executadas, mas apenas foram apresentadas as matrizes de confusão dos testes no Quadro 5.1. O símbolo “+” representa a classificação Persistente e o “-” a classificação Evadido.

Quadro 5.1. Matrizes de confusão das amostras teste no IFPB Virtual.

	Matriz de confusão do 1º corte	Matriz de confusão do 2º corte	Matriz de confusão do 3º corte
Letras (1º período)	$\begin{vmatrix} & + & - \\ + & 34 & 8 \\ - & 0 & 2 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 34 & 3 \\ - & 0 & 7 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 29 & 0 \\ - & 5 & 10 \end{vmatrix}$
Letras (2º período)	$\begin{vmatrix} & + & - \\ + & 8 & 1 \\ - & 6 & 5 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 8 & 0 \\ - & 6 & 6 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 13 & 0 \\ - & 1 & 6 \end{vmatrix}$
Segurança do Trabalho (1º período)	$\begin{vmatrix} & + & - \\ + & 26 & 11 \\ - & 9 & 24 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 29 & 6 \\ - & 6 & 29 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 28 & 0 \\ - & 7 & 35 \end{vmatrix}$
Segurança do Trabalho (2º período)	$\begin{vmatrix} & + & - \\ + & 24 & 7 \\ - & 3 & 3 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 22 & 1 \\ - & 5 & 9 \end{vmatrix}$	$\begin{vmatrix} & + & - \\ + & 24 & 0 \\ - & 3 & 10 \end{vmatrix}$

Fonte: Próprio Autor.

Na Tabela 5.2, são apresentados os principais resultados da avaliação, de acordo com as métricas precisão, sensibilidade, eficiência e especificidade. Tais métricas foram calculadas com base nas matrizes de confusão geradas. Em relação aos cortes temporais feitos, observou-se que a *FuzzySD* atingiu uma precisão máxima no curso de Letras (2º período) 95% no teste, no terceiro corte temporal, e no terceiro corte do curso de Segurança do Trabalho (2º

período) 91,9% no teste.

Tabela 5.2. Avaliação da *FuzzySD* nos cenários selecionados.

<i>Curso</i>	<i>Corte</i>	<i>Precisão</i>		<i>Sensibili- dade do teste</i>	<i>Eficiên- cia do teste</i>	<i>Especi- ficidade do teste</i>
		<i>Treina- mento</i>	<i>Teste</i>			
<i>Letras – 1º período</i>	1ª parte - (1º mês)	66,3%	81,8%	100%	60%	100%
	2ª parte - (2º e 3º meses)	84,6%	93,2%	100%	85%	70%
	3ª parte – (4º e 5º meses)	94,2%	88,6%	85,3%	92,6%	100%
<i>Letras – 2º período</i>	1ª parte - (1º mês)	46,7%	65%	57,1%	70,2%	83,3%
	2ª parte - (2º e 3º meses)	60%	70%	57,1%	78,6%	100%
	3ª parte – (4º e 5º meses)	97,8%	95%	92,9%	96,4%	100%
<i>Segurança do Trabalho - 1ª período</i>	1ª parte - (1º mês)	79,9%	71,4%	74,3%	71,4%	68,6%
	2ª parte - (2º e 3º meses)	79,9%	82,9%	82,9%	82,9%	82,9%
	3ª parte – (4º e 5º meses)	85,4%	90%	80%	90%	100%
<i>Segurança do Trabalho - 2ª período</i>	1ª parte - (1º mês)	72,4%	73%	88,9%	59,4%	30%
	2ª parte - (2º e 3º meses)	93,1%	83,8%	81,5%	85,7%	90%
	3ª parte – (4º e 5º meses)	92%	91,9%	88,9%	94,4%	100%

Fonte: Próprio Autor.

Outro aspecto relevante foi que a abordagem apresentou uma precisão similar às abordagens mencionadas na literatura (precisão > 60%). Isto representou um indicativo positivo para a abordagem, pois demonstrou que, apesar da complexidade do cenário e das características distintas entre cursos, houve um desempenho regular durante as predições. Contudo, ajustes na extração das informações e nas conversas com especialistas precisam ser aprimorados.

Os resultados mostraram também que a abordagem melhora ao longo do curso online, pois a precisão e a eficiência do teste aumentam ao longo dos cortes. Pode-se atribuir tal fenômeno ao aumento no número de informações providas pelos alunos até o final do curso online. A abordagem *FuzzySD* no

último corte temporal apresentou maior quantidade de informações para analisar e identificar a evasão do discente. Isto acontece até mesmo em cursos com evasão elevada, é o caso do primeiro período de Segurança do Trabalho (49,5%, segundo a Tabela 5.1).

Um comportamento relevante foi o desempenho mais eficiente da *FuzzySD* no terceiro corte temporal da amostra em ambos os cenários analisados. Isto indica que, em avaliações somativas, a abordagem pode ser uma perspectiva eficiente para o uso. Neste sentido, a *FuzzySD* poderia ser usada para entender as características dos grupos de discentes evadidos ou persistentes frente às entradas definidas (Autorregulação, Interação e Motivação). Os tomadores de decisão poderiam usar tais informações para refletir sobre estratégias pedagógicas adotadas e sobre o desempenho da turma *online*.

Outro ponto relevante, residiu no fato de que a falta de interação nos fóruns de discussão, durante o início do curso de Letras (primeiro e segundo período), pode ter afetado o desempenho da abordagem em relação à métrica Motivação. Normalmente, os alunos não interagem com mensagens relevantes em fóruns no primeiro mês de curso. Tal fato pode ter refletido nos piores índices nestes cortes da amostra analisada.

Alguns aspectos importantes não foram analisados pela abordagem, mas foram identificados como pontos que poderiam melhorar a qualidade da previsão do arcabouço, são eles:

- a) a falta de registros sobre os objetivos das atividades por parte dos professores, pois muitas vezes as informações eram recuperadas apenas da memória do docente;
- b) processo de consultar o coordenador da EaD para solucionar dúvidas quando o professor não pertencia mais ao corpo docente da instituição;
- c) o tempo de adaptação do aluno aos novos cursos, pois muitas vezes os professores da instituição comentaram a dificuldade da turma em acompanhar o ritmo das atividades e que isto acontecia com mais da metade para o final do curso;

- d) monitoramento das turmas *online* sem padronização dos tutores, apesar de existir um treinamento, e
- e) o mau uso de fóruns de discussão. Em alguns casos, os discentes fizeram propagandas de produtos naquele ambiente.

5.3.3. Analisando as Hipóteses

Após realizar a análise em dois cenários distintos, foi possível retomar e responder as hipóteses apresentadas no planejamento da avaliação da *FuzzySD*.

H_{10} : Quem evade no curso é identificado pela abordagem (precisão $> 0,6$).

H_{11} : Quem evade no curso não é identificado pela abordagem (precisão $\leq 0,6$).

De acordo com as análises prévias, foi possível aceitar a hipótese, pois a abordagem demonstra que tem um desempenho similar ao das abordagens apresentadas na literatura, pois sempre tem uma precisão maior do que 60%. Para domínios complexos, este desempenho se mostrou satisfatório dado a heterogeneidade das turmas *online*.

Identificou-se, ainda, que há momentos distintos para a atuação da *FuzzySD* retornar os melhores resultados da previsão da evasão em cursos *online*. No caso do primeiro período do curso de Letras, foi mais adequado utilizar a *FuzzySD* no segundo corte temporal (2 meses do início do curso), pois os resultados mostraram uma maior precisão nesta amostra (93,2%). Concomitantemente, no primeiro período do curso de Segurança do Trabalho, foi obtida uma taxa de acerto maior do que os outros extratos (90%). Isto significa que a *FuzzySD* pode ser utilizado ao longo do curso, pois as análises demonstraram que não há um ponto no qual o desempenho seja superior em todas as amostras.

5.4 Outras Abordagens Avaliativas

Após avaliar a precisão da abordagem, é importante investigar se a arquitetura *fuzzy* apresenta um desempenho superior em relação a outras metodologias

avaliativas discutidas na literatura. Para isto, é necessário realizar um estudo de caso com as mesmas entradas da abordagem *FuzzySD* utilizando outras técnicas, tais como: árvores de decisão, redes neurais e redes bayesianas. No tocante às árvores de decisão, foram selecionadas duas técnicas mencionadas: *J48* e *RandomForest*; uma técnica de rede neural: *Multilayer Perceptron*; e uma técnica de rede bayesiana: *Naive Bayes*.

Nesta investigação, foram utilizadas as mesmas entradas da abordagem *FuzzySD* (Autorregulação, Motivação e Interação) e mesmos dados relativos aos cursos *online* do estudo de caso final (Letras, 1º e 2º períodos, e Segurança do Trabalho, 1º e 2º períodos). Esta decisão permite uma comparação mais consistente, pois as análises envolvem os mesmos critérios e entradas, possibilitando o foco na técnica computacional.

Os dados de entrada seguiram o formato apresentado no Apêndice D. A etapa de pré-processamento foi realizada de modo único para o experimento que avaliou a precisão da *FuzzySD* e este que se propôs comparar abordagens avaliativas distintas.

Foram analisadas as precisões de cada uma das técnicas selecionadas nos três cortes temporais mencionados anteriormente (1º mês, 3º mês e por volta do 5º mês do curso). Estas amostras foram segmentadas, de acordo com a descrição da seção 4.5, em treinamento e teste, além da execução de uma validação cruzada (*10-fold cross validation*). Nesta etapa, os experimentos foram realizados por meio do uso da ferramenta gratuita *Weka*.

Após a execução das técnicas destacadas, as precisões foram agrupadas na Tabela 5.3 para facilitar a visualização dos resultados, os resultados da abordagem *FuzzySD* estão destacados. Neste sentido, é possível perceber que as entradas conseguiram descrever de forma eficiente (precisão > 70%) a maioria das classificações realizadas, o que permite afirmar que as entradas se mostraram como boas métricas para descrever a evasão discente em cursos *online*.

Ao comparar o desempenho das outras abordagens com a *FuzzySD*, é notável que o desempenho se mostrou ter maior acurácia na maioria dos cenários analisados, exceto em Letras 2º período – 2ª parte, Segurança do

Trabalho 1º período – 1ª parte e 2ª parte. Mesmo nas exceções, a FuzzySD apresentou uma precisão maior do que 70%, resultado satisfatório em relação ao que é encontrado na literatura.

A lógica fuzzy se mostrou consistente também em cenários nos quais a precisão foi inferior à 60% com outras técnicas. Por exemplo, a amostra de Letras 2º período – 1ª parte. Na amostra mencionada, somente a rede bayesiana e a FuzzySD proporcionaram uma precisão maior do que 50%. Contudo, a abordagem apresentada ainda proporcionou um desempenho melhor em relação à rede.

Tabela 5.3. Avaliação das abordagens avaliativas em relação à FuzzySD.

	Letras - 1º período			Letras - 2º período		
	1ª parte	2ª parte	3ª parte	1ª parte	2ª parte	3ª parte
Redes Neurais (Multilayer Perceptron)	70 %	75,5%	86,4%	46,6%	91,1%	74,8%
Árvores de decisão (J48)	70,7%	70,7%	86,4%	36,4%	73%	85,7%
Redes Bayesianas (Naive Bayes)	80%	85%	86,4%	58,5%	89,7%	76,5%
Árvores de decisão (Random Forest)	70,7%	81,5%	86,4%	42,3%	91,1%	91,1%
FuzzySD	81,8%	93,2%	88,6%	65 %	70%	95%
	Segurança do Trabalho – 1º período			Segurança do Trabalho – 2º período		
	1ª parte	2ª parte	3ª parte	1ª parte	2ª parte	3ª parte
Redes Neurais (Multilayer Perceptron)	78,7%	88,7%	88,8%	72,5%	73%	76,7%
Árvores de decisão (J48)	84,9%	86,6%	88,8%	58,7%	73%	85,7%
Redes Bayesianas (Naive Bayes)	80,1%	80,1%	84,4%	65,5%	73%	75,9%
Árvores de decisão (Random Forest)	80%	90%	88,8%	70,5%	76,3%	75,9%
FuzzySD	71,4%	82,9%	90%	73%	83,8%	91,9%

Fonte: Próprio Autor.

Tais resultados deram ao trabalho uma confiança maior sobre a escolha da metodologia *fuzzy* como base da abordagem apresentada. Os resultados mostraram que a acurácia da lógica *fuzzy* revelou-se, de fato, como a mais eficiente dentre as opções encontradas na literatura para este cenário.

5.4.1 Abordagens Avaliativas *Fuzzy*

Em se tratando de aplicações voltadas à investigação da evasão por meio de análises multidimensionais do discente, identificou-se 7 trabalhos relevantes na literatura pesquisada. No Quadro 5.2, são descritos os pontos principais destes trabalhos para realizar uma analogia com o estudo proposto na pesquisa ora descrita.

De acordo com o Quadro 5.2, percebe-se que os trabalhos correlatos à pesquisa focaram suas análises em alguns aspectos principais: avaliar desempenho e presença, qualidade da interação e aprendizagem. A presente pesquisa de tese propôs um diferencial que foi a análise de aspectos relacionados à evasão discente.

Outra perspectiva relevante foi que a base de avaliação das modelagens *fuzzy* é relativamente pequena na maioria dos trabalhos elencados, com exceção de Dias e Diniz (2013). Este ponto demonstrou que boa parte dos trabalhos deixou as avaliações e validações mais extensas para trabalhos futuros.

Os trabalhos de Bai e Chen (2008), Saleh e Kim (2008) e Chen e Lin (2009) focaram na discussão da modelagem matemática *fuzzy*, pois propuseram estratégias para automatizar a geração das funções de pertinência. Na pesquisa, foi decidido que as funções de pertinência seriam geradas por meio do apoio do docente especialista.

Outro aspecto importante a ser retomado é a precisão do modelo, pois a precisão da maioria dos trabalhos de predição sobre evasão existentes foi entre 50 e 90%. Contudo, tais modelos foram concebidos baseados em dados de treinamento e teste extraídos do mesmo curso, o que significa que não foram adequados para prever contextos de cursos distintos em tempo real, podendo ainda apresentar vieses na análise.

Quadro 5.2. Quadro resumo dos modelos *fuzzy*.

TRABALHO	ENTRADAS	OBJETIVO	MÉTODO	AMOSTRA
Bai e Chen (2008)	Dificuldade, complexidade e a importância das questões.	Avaliar desempenho do discente	Lógica <i>fuzzy</i>	10 alunos
Saleh e Kim (2009)	Dificuldade, a complexidade e a importância das questões.	Avaliar desempenho do discente	Lógica <i>fuzzy</i>	10 alunos
Arias et al. (2012)	Tempo de permanência no fórum, quantidade de interações, nota final e quantidade de citações ou referências do aluno no fórum por outros participantes	Avaliar desempenho do discente	Lógica <i>fuzzy</i>	10 alunos do quinto ano de matemática
Kao, Lin e Chu (2012)	Precisão, dificuldade no teste, nível de confiança e tempo de resposta.	Avaliar a aprendizagem no curso	Lógica <i>fuzzy</i>	52 alunos do curso de construção de compiladores
Dias e Diniz (2013)	Qualidade, ação e período de interação	Avaliar a qualidade das interações com o LMS	Lógica <i>fuzzy</i>	1037 alunos
Chen e Lin (2009)	Taxa de precisão, tempo, dificuldade, complexidade, custo da resposta e importância	Avaliar desempenho do discente	Lógica <i>fuzzy</i>	10 alunos
Albuquerque et al. (2017)	Tempo, quantidade de interações, quantidade de referências, notas	Avaliar a Presença Social em AVEA	Lógica <i>fuzzy</i> + Learning Vector	36 alunos
<u>FuzzySD</u>	Variáveis associadas às métricas: Motivação, Autorregulação e Interação	Avaliar a Evasão do discente	Lógica <i>fuzzy</i> + Mineração de Texto	Dois cenários: curso de letras (231 alunos) e curso de Segurança do Trabalho (358 alunos)

Fonte: Próprio Autor.

Na presente pesquisa, os vieses das análises foram minimizados ao se analisar dados históricos com características distintas: perfil de curso, período, características das atividades, *etc.* A *FuzzySD* surge então como uma alternativa de predição da evasão, pois as análises foram feitas em cenários

com perfis diferenciados entre si e pode ser estendida para outros contextos de EaD na UAB.

Uma vantagem destacada pelo docente consultor sobre o uso da *FuzzySD* foi que a abordagem *fuzzy* conseguiu prover ao professor um entendimento maior do processo avaliativo em relação às abordagens que oferecem saídas “prontas” (no nosso caso o risco de evasão ou persistência no curso). O docente que participou do processo de calibração das regras conseguiu entender e articular possíveis intervenções que deveriam ter sido feitas para tentar sanar ou restabelecer as motivações dos alunos que estava com o status Persistente, mas na iminência de atingir a Evasão. As intervenções não puderam ser feitas, pois, os dados os cursos analisados já estavam finalizados.

5.5 Considerações Finais

É importante destacar o papel central e desafiador da avaliação neste estudo, pois lidar com variáveis difusas exige uma investigação muito próxima do especialista da informação, desde o planejamento até a validação. Deste modo, entendeu-se que o tipo de avaliação mais adequado à EaD é a formativa, por fornecer subsídios para a melhoria contínua do processo de ensino-aprendizagem baseado nas ações de cooperação e nos projetos colaborativos. Neste mesmo contexto, é fundamental lembrar o papel essencial da avaliação continuada na alimentação de sistemas de retorno, como exemplificado no referencial teórico deste documento.

Contudo, a avaliação da *FuzzySD* demonstrou que alguns ajustes precisam ser executados para que a avaliação formativa seja amplamente utilizada pela abordagem. Este fato se deve à complexidade dos contextos de análise e às singularidades existentes entre cursos, principalmente nos textos produzidos pelo aluno. Incorporar técnicas mais avançadas de processamento de linguagem natural pode ser uma estratégia para tal aperfeiçoamento.

6. Considerações Finais

Este documento apresenta e fundamenta o processo de concepção, construção e avaliação de uma abordagem avaliativa multidimensional voltada à avaliação formativa (ou somativa) de discentes em AVEA, chamada *FuzzySD*. A pesquisa descrita se subsidia em estratégias multidimensionais para apresentar uma avaliação pedagogicamente dinâmica, semelhante aos processos de avaliação tradicionais.

Neste sentido, discorre-se sobre as principais características, a metodologia científica do processo, o planejamento da abordagem, os resultados e a avaliação da abordagem da pesquisa. Neste capítulo, são resumidas algumas análises finais sobre a pesquisa, destacando-se as suas principais realizações e contribuições para a área de avaliação na Informática na Educação. Além disto, algumas de suas limitações são apresentadas, culminando em uma lista de sugestões para estendê-las em pesquisas futuras.

Inicialmente, é importante retomar o problema inicial do trabalho apresentado no Capítulo 1: “Como e quais fatores comportamentais do aluno no AVEA impactam na evasão discente em cursos *online*?”. O trabalho de tese conseguiu apresentar e provar que há uma série de fatores comportamentais do discente que influenciam na sua evasão em um curso *online* (Capítulo 4, Seção 4.3.1). A pesquisa deixou em aberto na abordagem a possibilidade de inserir outros fatores, pois há uma diversidade de arquiteturas que podem ser exploradas e discutidas. Finalmente, respondendo ao questionamento inicial, a pesquisa identificou que há variáveis como edição de recursos, interações em fóruns de discussão, visualização dos *feedbacks* do professor, conversa com colegas, *etc.* que se correlacionaram diretamente com a problemática da evasão discente na arquitetura do Moodle.

A métrica da Interação também foi avaliada em um outro AVEA o LearnWeb, que é um sistema alemão de busca e compartilhamento colaborativo, que une diferentes serviços *online* sob a mesma perspectiva, tais como: *YouTube*, *Flickr*, *Bing*, *Vimeo* e *SlideShare*. Tais análises proporcionaram correlações semelhantes em cursos de Introdução à

Computação. Contudo, como as outras métricas não foram avaliadas neste mesmo ambiente, as discussões no documento focaram no AVEA Moodle.

Vale ressaltar as principais contribuições da pesquisa: apresentar uma abordagem *fuzzy* que incorpore como uma das entradas mineração de texto das mensagens, propor uma abordagem que apoie tomadores de decisão no combate à evasão discente, obter bons resultados em contextos distintos de atuação (dado que trabalhos correlatos foram concebidos e testados no mesmo contexto) e discutir sobre as funções de pertinência que foram planejadas e utilizadas no contexto do IFPB virtual.

A presente abordagem ainda se mostrou útil para apoiar tomadores de decisão de cursos *online* no monitoramento rápido de turmas. Muitas vezes o tempo de resposta rápida de uma análise desta categoria é fundamental para que atitudes possam ser providenciadas no combate à evasão discente.

6.1 Ameaças à Validade

Este trabalho de tese aborda problemas significativos relacionados ao combate à evasão discente em AVEA. Assim, destacam-se aspectos de sua viabilidade, elaborando-se modelos e ferramentas como parte da resposta às questões levantadas e objetivos estabelecidos no início da pesquisa, descritos no Capítulo 1. Neste sentido, investiu-se, por exemplo, em questões sobre reuso de conhecimento e de modelos descritos na literatura. Deste modo, buscou-se suporte técnico nas áreas de Inteligência Artificial e Engenharia de *Software*.

Contudo, durante esse processo algumas ameaças à validade precisam ser discutidas. A validade de um experimento está conectada ao nível de confiança que se tem no processo de investigação experimental como um todo, ou seja, o quão confiáveis são os elementos envolvidos no processo (TRAVASSOS, GUROV e AMARAL, 2002; WAINER, 2007). Dentre as ameaças à validade conhecidas na literatura, no presente estudo, pode-se destacar a identificação em sua estrutura de ameaças à validade de construção, interna e externa.

Entende-se por validade de construção quando o tratamento reflete a causa bem e o resultado reflete o efeito bem (TRAVASSOS, GUROV e

AMARAL, 2002). Neste grupo, a ameaça foi do tipo “Fatores humanos” porque os professores podem ter baseado suas respostas nos objetivos da pesquisa ou direcionado as respostas durante as entrevistas para se adequar ao que lembravam sobre os cursos ministrados nas amostras analisadas.

Quanto à validade interna, esta pode ser definida como: o relacionamento observado entre o tratamento e o resultado é causal, e não é resultado da influência de outro fator – não controlado ou medido (TRAVASSOS, GUROV e AMARAL, 2002). Neste caso, a “Instrumentação” foi o aspecto que influenciou nesta ameaça, pois a pesquisa focou em uma única arquitetura de AVEA.

A última ameaça à validade foi a externa, definida como as condições que limitam a habilidade de generalizar os resultados de um experimento para a prática industrial (TRAVASSOS, GUROV e AMARAL, 2002). Neste caso, a ameaça foi do tipo “Participantes”, pois pode ser que as amostras selecionadas não representem os alunos da UAB ou do IFPB Virtual, especificamente os dados mais recentes, pois a base se refere a uma versão de Moodle anterior.

6.2 Sugestões para Pesquisas Futuras

Após a análise crítica, é possível sugerir uma série de pesquisas futuras relacionadas à pesquisa desenvolvida, a destacar:

- a) utilizar a abordagem *FuzzySD* em outros AVEA: esta abordagem foi inicialmente planejada em outros ambientes (*LearnWeb*), mas as análises de mineração de texto e sua avaliação ainda podem ser estendidas para outros AVEAs;
- b) melhorar as estratégias de mineração de texto: é pertinente ampliar as técnicas de processamento de linguagem natural para que a métrica de Motivação não fique tão dependente do especialista e de documentação;
- c) incorporar técnicas de aprendizagem não supervisionada na concepção das funções de pertinência e regras *fuzzy*: apesar das análises de agrupamento apresentarem resultados

satisfatórios para a análise na geração das funções pertinência, é importante estimular a independência do sistema nesta fase do processo;

- d) incorporar análises longitudinais na construção da base de regras: apesar de ser uma tarefa bem genérica, seria adequado planejar bases de regras que fossem independentes da área de atuação do curso. Na presente abordagem, tal ação foi restrita porque cada curso tinha similaridades muito específicas, principalmente em relação à extração de conteúdos textuais;
- e) tornar a ferramenta *FuzzySD* web e desacoplar de um AVEA específico (no caso Moodle).
- f) Aplicar regressão ou cadeia de Markov, na construção do modelo e analisar o mesmo tripé de entradas para avaliar a evasão: será possível confrontar os resultados com os obtidos em outras pesquisas e com outras técnicas, bem como ampliar as entradas utilizadas.

6.3 Publicações Relacionadas ao Doutorado

O presente estudo resultou em diversas publicações acadêmicas em eventos científicos relevantes para o contexto da pesquisa. Além disto, dois artigos foram submetidos para revistas, os quais contemplam os resultados finais da pesquisa. Tais publicações demonstram a maturidade e evolução da pesquisa ao longo deste período, sendo essas em ordem temporal:

1. MORAIS, A. M. Planejamento de um Arcabouço Avaliativo para a Identificação do Perfil Evasivo do Discente: Estudo de Caso IFPB Virtual. Anais do XXIII Workshop de Informática na Escola (WIE 2017), v. 1, p. 993-1002, 2017.
2. MORAIS, A. M.; MEDEIROS, D. C.; TAIBI, D.; MARENZI, I. Formative Assessment supported by Analytics Techniques: a case study on teacher and students perceptions. CENTRAL EUROPE CEUR WORKSHOP PROCEEDINGS, v. 1850, p. 16-

- 36, 2016.
3. MORAIS, A. M.; MEDEIROS, D. C. Investigação Sobre O Perfil Interativo e Evasivo Do Discente Por Meio De Técnicas De Mineração De Dados Educacionais. In: 14^a Conferência Ibero Americana WWW/Internet 2016, 2016, Lisboa, Portugal. WWW/INTERNET 2016 e COMPUTAÇÃO APLICADA 2016. Lisboa: IADIS Press, 2016. p. 28-35.
 4. MORAIS, A. M.; MARENZI, I.; KANTZ, D. The LearnWeb formative assessment extension: Supporting awareness and reflection in blended courses. CEUR Workshop Proceedings, v. 1465, p. 97-103, 2015.
 5. MORAIS, A. M.; ARAUJO, J. M. F. R.; COSTA, E. B. Arcabouço Multidimensional de Avaliação dos Discentes baseado na Análise de Recursos Pedagógicos. In: XXXIV Congresso da Sociedade Brasileira de Computação - XXII Workshop sobre Educação em Computação, 2014, Brasília. Anais / XXXIV Congresso da Sociedade Brasileira de Computação. Porto Alegre: Sociedade Brasileira de Computação (SBC), 2014. p. 1430-1438.
 6. MORAIS, A. M.; ARAUJO, J. M. F. R. Educational Data Mining For Support E-Learning Teacher Based On Decision Tree. In: IADIS International Conference WWW/INTERNET 2013, 2013, Fort Worth, Texas, USA. Proceedings of the IADIS International Conference WWW/INTERNET 2013, 2013.
 7. MORAIS, A. M.; ARAUJO, J. M. F. R. Mineração de Dados Educacionais no Apoio ao Processo de Tomada de Decisão do Docente. In: XXXIII Congresso da Sociedade Brasileira de Computação, 2013, Maceió, Alagoas. Anais do XXXIII Congresso da Sociedade Brasileira de Computação, 2013. p. 478-483.
 8. MORAIS, A. M.; ARAUJO, J. M. F. R.; COSTA, E. B. Monitoring Student Performance Using Data Clustering and Predictive

Modelling. In: Proceedings of 2014 IEEE Frontiers in Education Conference. Madrid, Espanha. Proceedings of FIE 2014. p. 2492-2499. 2014.

9. MORAIS, A. M; ARAUJO, J. M. F. R.; COSTA, E. B. Perfil do Docente em Educação a Distância e suas Perspectivas Frente aos Processos Avaliativos. In: VII Anais do WAvalia – III CBIE. Dourados-MS. 2014.

Anexo A- Carta de Anuência do Reitor do IFPB



INSTITUTO FEDERAL DE
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
PARAÍBA

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA

CARTA DE ANUÊNCIA

Declaramos, para os devidos fins, que aceitaremos as pesquisadoras **Alana Marques de Moraes** e **Danielle Chaves de Medeiros**, pós-graduandas em Ciência da Computação pela Universidade Federal de Campina Grande (UFCG), nesta instituição para desenvolverem seus respectivos projetos de pesquisa, sob orientação da Prof. **Joseana Macêdo Fechine Régis de Araújo**, cujo objetivo é estudar e desenvolver métodos e metodologias para a avaliação de alunos em Educação a Distância, com o auxílio de técnicas de Inteligência Artificial, no IFPB.

Ressaltamos que os dados coletados serão mantidos em absoluto sigilo, de acordo com a Resolução do Conselho Nacional de Saúde (CNS/MS) 466/2012, que trata da Pesquisa envolvendo Seres Humanos. Salientamos ainda que tais dados só poderão ser utilizados tão somente na realização deste estudo. Assim, a aceitação está condicionada ao cumprimento das pesquisadoras aos requisitos da resolução citada anteriormente, comprometendo-se as pesquisadoras a utilizarem os dados e materiais coletados exclusivamente para os fins da pesquisa.

João Pessoa, em 18 de agosto de 2014.

A handwritten signature in black ink, appearing to read 'João Batista de Oliveira Silva', is written over a faint, circular stamp.

Prof. João Batista de Oliveira Silva
Reitor do IFPB

Apêndice A – Dicionários (métrica Motivação) e Lista de *Stopwords*

A) Dicionário para Motivação Intrínseca

'viciado', 'vidrado', 'addict', 'maravilhoso', 'maravilhosa', 'awesome', 'encanto', 'easy', 'fácil', 'facil', 'excite', 'excited', 'animado', 'love', 'loving', 'amor', 'amei', 'incredible', 'incrível', 'incrivel', 'interest', 'interest', 'interessante', 'like', 'gosto', 'adorei', 'gostei', 'perfect', 'perfeito', 'perfeita', ';)', ';D', 'nice', 'ótimo', 'otimo', 'ótima', 'otima', 'win', 'winner', 'winning', 'vencedor', 'vencer', 'crescimento', 'melhoria', 'cinema', 'ead', 'curso a distancia', 'curso à distância', 'cursos a distância', 'moodle', 'wow', 'omg'

B) Dicionário para Motivação Extrínseca

'score', 'nota', 'pontuacao', 'pontuação', 'cool', 'legal', 'subject', 'disciplina', 'atividades', 'aula', 'good', 'great', 'bom', 'ótimo', 'otimo', 'happy', 'feliz', 'satisfeito', 'satisfeita', 'fix', 'consertar', 'post', 'upload', 'enviar', 'enviei', 'postei', 'postar', ':)', '(:', ':-)', '(-:', '^', 'thank', 'thanks', 'obrigada', 'obrigado', 'brigada', 'brigado', 'issue', 'questão', 'questao', 'duvida', 'dúvida', 'desafio', 'tópico', 'topico', 'tópicos', 'tópicos', 'temática', 'cinema', 'ead', 'curso a distancia', 'curso à distância', 'cursos a distância', 'moodle'

C) Dicionário para Desmotivação

'broke', 'quebrado', 'não funciona', 'não funcionou', 'quebrada', 'bad', 'ruim', 'péssimo', 'bug', 'bugado', 'com problemas', 'cant', 'can\'t', 'não pode', 'impossível', 'impossivel', 'dont', 'don\'t', 'do not', 'does not', 'doesn\'t', 'no', 'não', 'n', 'nao', 'difficult', 'difícil', 'dificil', 'disater', 'desastroso', 'fail', 'falho', 'errado',

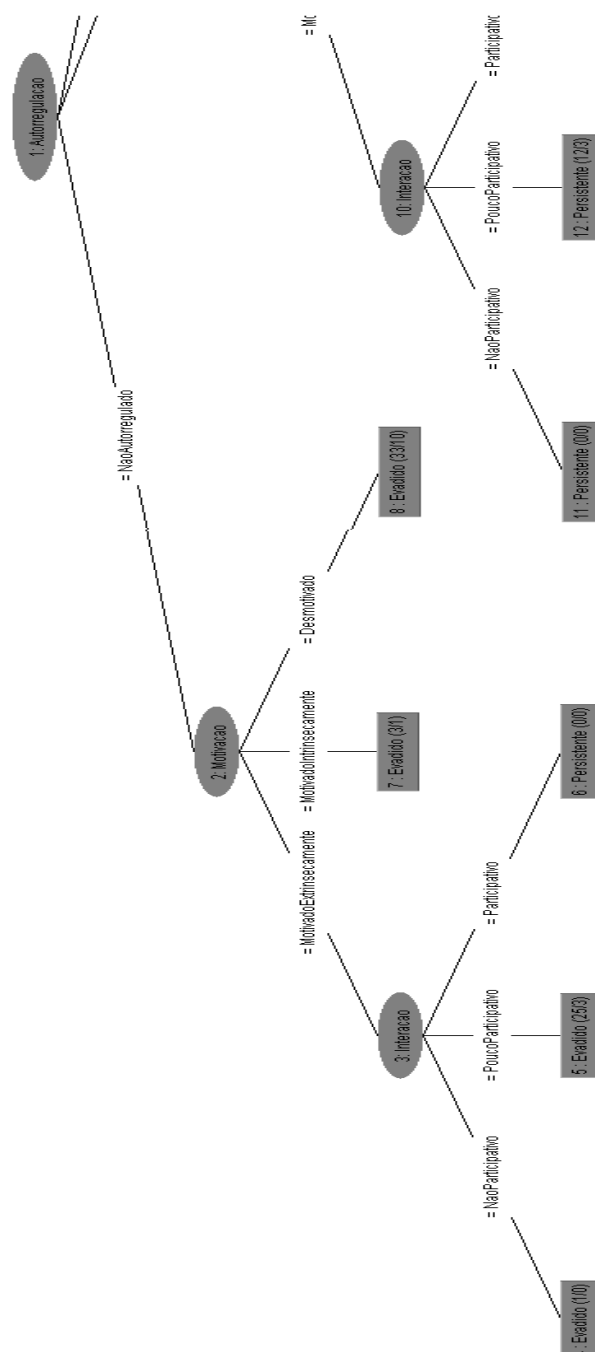
'errada', 'falso', 'expense', 'expensive', 'custoso', 'demorado', 'caro', ':(, ':)', ':/',
 'hate', 'odeio', 'serious', 'sério', 'shit', 'merda', 'slow', 'lento', 'terrible', 'terrível',
 'lose', 'perdi', 'trouble', 'problema', 'unhappy', 'infeliz', 'dropout', 'give up',
 'trancar', 'desistir', 'poor', 'pobre'

D) StopWords

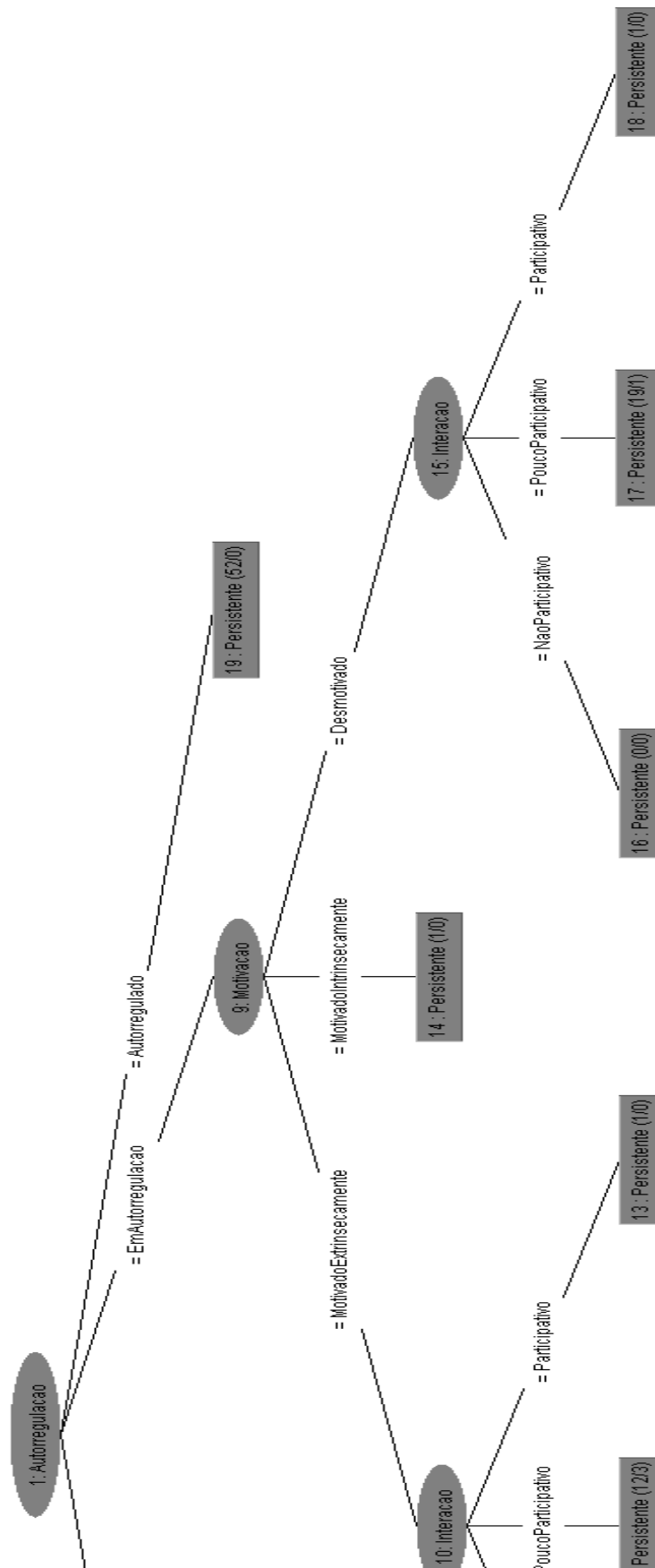
"0", "1", "2", "3", "4", "5", "6", "7", "8", "9", "a", "as", "es", "is", "os", "e", "i",
 "o", "u", "que", "do", "da", "de", "em", "um", "para", "com", "uma", "no", "se", "na",
 "no", "por", "pra", "mais", "dos", "como", " ", " ", ":", "!", "?", ".", "|", "mas", "foi", "ao",
 "ele", "das", "tem", "à", "seu", "sua", "ou", "quando", "muito", "há", "nos", "já",
 "está", "eu", "também", "só", "pelo", "pela", "até", "isso", "ela", "entre", "era",
 "depois", "sem", "mesmo", "aos", "ter", "seus", "quem", "nas", "me", "esse",
 "eles", "estão", "você", "tinha", "foram", "essa", "num", "nem", "suas", "meu",
 "às", "minha", "têm", "numa", "pelos", "elas", "havia", "seja", "qual", "ser", "será",
 "nós", "tenho", "lhe", "deles", "essas", "esses", "pelas", "este", "fosse", "dele",
 "tu", "te", "vocês", "vos", "lhes", "meus", "minhas", "teu", "tua", "teus", "tuas",
 "nosso", "nossa", "nossos", "nossas", "dela", "delas", "esta", "estes", "estas",
 "aquele", "aquela", "aqueles", "aquelas", "isto", "aquilo", "estou", "está",
 "estamos", "estão", "estive", "esteve", "estivemos", "estiveram", "estava",
 "estávamos", "estavam", "estivera", "estivéramos", "esteja", "estejamos",
 "estejam", "estivesse", "estivéssemos", "estivessem", "estiver", "estivermos",
 "estiverem", "hei", "há", "havesmos", "hã", "houve", "houvemos", "houveram",
 "houvera", "houvéramos", "haja", "hajamos", "hajam", "houvesse",
 "houvéssemos", "houvessem", "houver", "houvermos", "houverem", "houverei",
 "houverá", "houveremos", "houverão", "houveria", "houveríamos", "houveriam",
 "sou", "somos", "são", "era", "éramos", "eram", "fui", "foi", "fomos", "foram",
 "fora", "fôramos", "seja", "sejamos", "sejam", "fosse", "fôssemos", "fossem",
 "for", "formos", "forem", "serei", "será", "seremos", "serão", "seria", "seríamos",
 "seriam", "tenho", "tem", "temos", "tém", "tinha", "tínhamos", "tinham", "tive",
 "teve", "tivemos", "tiveram", "tivera", "tivéramos", "tenha", "tenhamos",
 "tenham", "tivesse", "tivéssemos", "tivessem", "tiver", "tivermos", "tiverem",
 "terei", "terá", "teremos", "terão", "teria", "teríamos", "teriam"

Apêndice B – Exemplos de Árvores de Decisão para os Cursos de Letras e Segurança do Trabalho

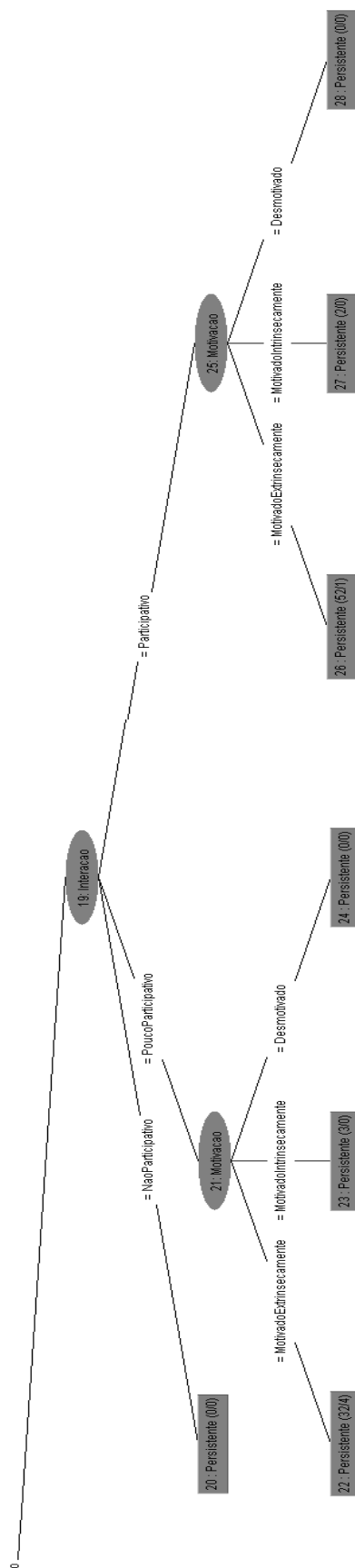
Árvore de decisão Random Tree do curso de Letras (Parte 1)



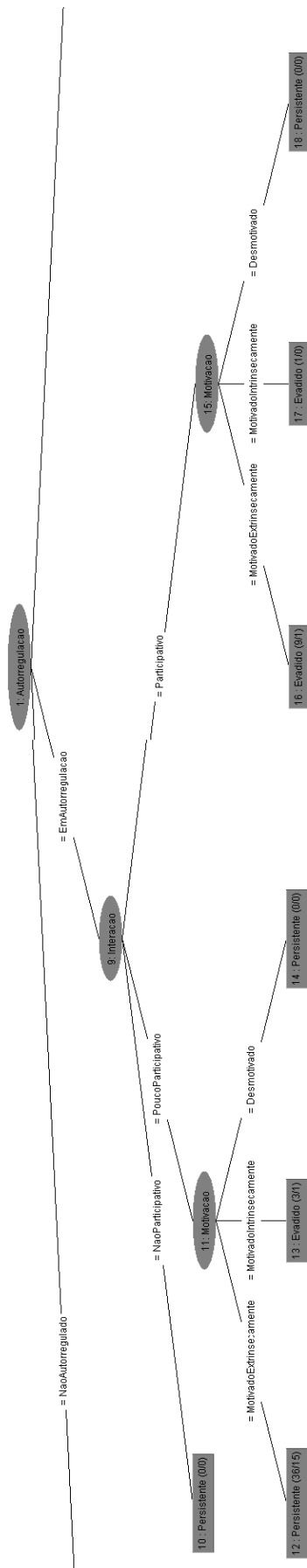
Árvore de decisão Random Tree do curso de Letras (Parte 2)



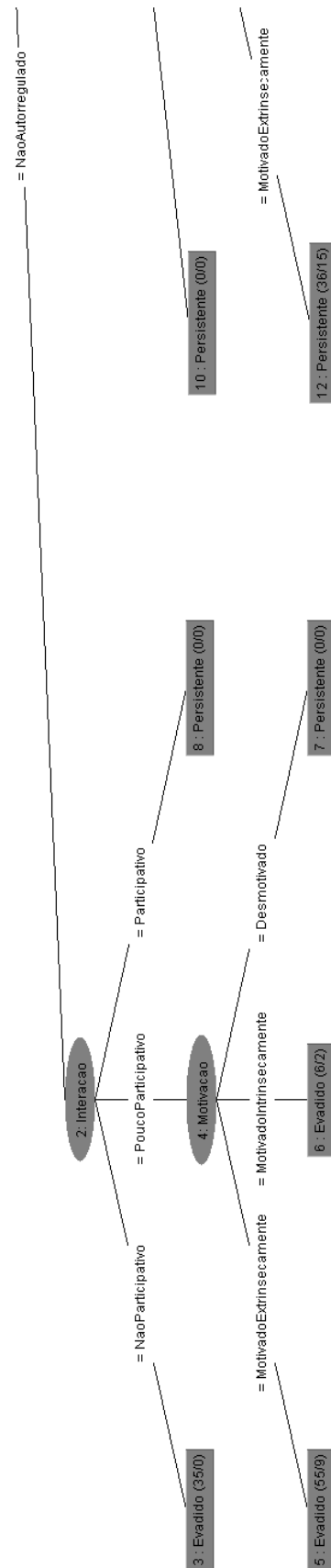
Árvore de decisão Random Tree do curso de Segurança do Trabalho (Parte 1)



Árvore de decisão Random Tree do curso de Segurança do Trabalho (Parte 2)



Árvore de decisão Random Tree do curso de Segurança do Trabalho (Parte 3)



Apêndice C – Arquivo entrada gerado no pré-processamento

idUser;FirstName;LastName;FirstAccess;LastAccess;City;Edition;ScoreMean;Feedback;Forum;Chat;NumVisualization;LastAccessCourse;DropoutStatus

830; Nome Aluno ;[ALLETPI*20122];20-11-2012;02-05-2013;Pedra Lavrada
;34;81,620053333;22;75;140;1891;0; Non-Dropout

831; Nome Aluno ;[ALLETPI*20122];19-11-2012;29-04-2013;Picuí
;8;12,519742000;9;0;2;185;6;Non-Dropout

833; Nome Aluno:[ALLETPI*20122];29-11-2012;29-11-
2012;Picuí;0;7,807742000;3;0;0;10;0;Dropout

834; Nome Aluno:[ALLETPI*20122];24-11-2012;14-05-
2013;Picuí;24;48,614860000;18;24;7;972;0;Non-Dropout

835; Nome Aluno:[ALLETPI*20122];24-11-2012;16-04-
2013;Picuí;7;37,635053333;17;0;1;343;0;Non-Dropout

836; Nome Aluno:[ALLETPI*20122];17-11-2012;03-05-2013;Jaçanã
RN;36;71,539743333;22;50;117;2242;0;Non-Dropout

837; Nome Aluno:[ALLETPI*20122];16-11-2012;11-12-
2012;Picuí;3;9,367742000;3;0;0;113;51;Dropout

838; Nome Aluno:[ALLETPI*20122];24-11-2012;15-12-
2012;Picuí;12;14,472904000;8;1;0;170;47;Dropout

839; Nome Aluno:[ALLETPI*20122];19-11-2012;14-05-
2013;Picuí;21;75,307091666;22;66;73;2478;0;Non-Dropout

840; Nome Aluno:[ALLETPI*20122];19-11-2012;22-03-
2013;Picuí;22;27,323171666;20;6;2;904;1;Non-Dropout

841; Nome Aluno:[ALLETPI*20122];18-11-2012;23-03-
2013;Picuí;12;28,291680000;19;19;75;1349;0;Non-Dropout

842; Nome Aluno:[ALLETPI*20122];19-11-2012;14-02-
2013;Picuí;5;12,826505000;19;8;1;569;1;Non-Dropout

843; Nome Aluno:[ALLETPI*20122];24-11-2012;28-04-
2013;Picuí;29;56,155946666;22;25;22;1412;0;Non-Dropout

Apêndice D – Entrada para a *FuzzySD*

Autorregulacao;Interacao;Motivacao;Evasao
 Autorregulado;Participativo;Desmotivado;Persistente
 NaoAutorregulado;PoucoParticipativo;MotivadoExtrinsecamente;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 Autorregulado;PoucoParticipativo;Desmotivado;Persistente
 EmAutorregulacao;PoucoParticipativo;Desmotivado;Persistente
 Autorregulado;Participativo;Desmotivado;Persistente
 NaoAutorregulado;PoucoParticipativo;MotivadoIntrinsecamente;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 Autorregulado;Participativo;Desmotivado;Persistente
 EmAutorregulacao;PoucoParticipativo;Desmotivado;Persistente
 EmAutorregulacao;PoucoParticipativo;Desmotivado;Persistente
 NaoAutorregulado;PoucoParticipativo;MotivadoExtrinsecamente;Evadido
 Autorregulado;PoucoParticipativo;MotivadoExtrinsecamente;Persistente
 NaoAutorregulado;PoucoParticipativo;MotivadoExtrinsecamente;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 Autorregulado;Participativo;Desmotivado;Persistente
 Autorregulado;Participativo;MotivadoExtrinsecamente;Persistente
 Autorregulado;Participativo;MotivadoExtrinsecamente;Persistente
 NaoAutorregulado;PoucoParticipativo;MotivadoExtrinsecamente;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 EmAutorregulacao;PoucoParticipativo;Desmotivado;Persistente
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 NaoAutorregulado;PoucoParticipativo;Desmotivado;Evadido
 Autorregulado;Participativo;Desmotivado;Persistente

Referências Bibliográficas

ABBAD, G.; CARVALHO, R. S.; ZERBINI, T. Evasão em curso via internet: Explorando variáveis explicativas. **ERA Eletrônica**, v.7, n.2, pp.1676-5648. doi:10.1590/S1676-56482006000200008. 2006.

ABED. Associação Brasileira de Educação a Distância (ABED). **Censo EAD.BR: relatório analítico da aprendizagem a distância no Brasil 2015**. ISBN 978-85-5972-158-4. Curitiba: InterSaberes, 2016.

ADACHI, A.A.C.T. Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais, **Dissertação de Mestrado**. Faculdade de Educação, UFMG, 2009.

ALBUQUERQUE, M., SALES, G.L., REBOUÇAS FILHO, P.; MEDEIROS, C. Avaliação da presencialidade em um fórum LV utilizando Lógica *Fuzzy*. In: Simpósio Brasileiro de Informática na Educação-SBIE. **Anais ...** v.28, n. 1, p. 1357. 2017.

ALLEN, I. E.; SEAMAN, J. **Online Report Card: Tracking Online Education in the United States**. Babson Survey Research Group. 2016.

ALVES, A.L.F.; Baptista, C.S.; de Andrade, L.H.; Paes, R.F.C. Uso de Técnicas de Análise de Sentimentos em Tweets relacionados ao Meio-Ambiente. **Anais VI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais - 2015** - Recife, PE. Pp. 73 – 82. 2015

AMERI, S.; FARD, M.J.; CHINNAM, R.B.; REDDY, C.K. Survival Analysis based Framework for Early Prediction of Student Dropouts. **Anais: 25th ACM International on Conference on Information and Knowledge Management. Indiana**. Pp. 903-912. 2016.

AMNUEYPORNSAKUL, B.; BHAT, S.; CHINPRUTTHIWONG, P. Predicting attrition along the way: The UIUC model. **Anais: EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs**. pp.: 55–59. 2014

ANDERSON, T. **Modes of interaction in distance education: Recent developments and research questions.** Handbook of distance education, pp.129-144. 2003

ARAUJO, L.; LUCENA, G. Comunidades virtuais de aprendizagem: novas dinâmicas de aprendizagem exigem novas formas de avaliação. **Anais: XVI Simpósio Brasileiro de Informática na Educação.** SBC/ UFJF, Juiz de Fora, MG. 2005.

ARIAS, R. A.; PANHAN, A. M.; BREDA, G. D.; ZARPELÃO, B. B.; MENDES, L. S. Avaliação Multidimensional Baseada em Lógica difusa para Ambientes de Ensino Mediado por Computador. **Revista Brasileira de Informática na Educação**, v.20, n.3, 74. 2013.

AURÉLIO, **O mini dicionário da língua portuguesa.** 4ª edição e ampliada do mini dicionário Aurélio. 7ª impressão – Rio Janeiro, 2002.

AVANÇO, L. V. **Sobre a normalização e classificação de polaridade de textos opinativos na web.** Dissertação de Mestrado. 102 pgs. ICMC/USP, 2015.

AZEVEDO, R.; JOHNSON, A.; CHAUNCEY, A.; GRAESSER, A. **Use of hypermedia to assess and convey self-regulated learning.** In B. Zimmerman, & D. Schunk (Eds.), Handbook of self-regulation of learning and performance (pp. 102-121). New York: Routledge. 2011.

BAI, S. M.; CHEN, S. M. Evaluating students' learning achievement using *fuzzy* membership functions and *fuzzy* rules. **Expert Systems with Applications**, v. 34, n.1, pp.399–410. 2008.

BALAKRISHNAN, G.; COETZEE, D. Predicting student retention in massive open *online* courses using hidden markov models. **Anais: Electrical Engineering and Computer Sciences University of California at Berkeley**, 2013.

BARNARD-BRAK, L.; LAN, W.Y.; PATON, V.O. Profiles in self-regulated learning in the *online* learning environment. **International Review of Research in Open and Distance Learning**, v.11, n.1, p.149-56, mar. 2010.

BAXTER, J. Who am I and what keeps me going? profiling the distance learning student in higher education. **International Review of Research in Open and Distance Learning**, v.13, n.4, pp.107-129. 2012.

BEER, C.; CLARK, K.; JONES, D. Indicators of engagement, presented at the Curriculum, technology & transformation for an unknown future. In: Steel, CH, Keppell, MJ, Gerbic, P. (eds). Curriculum, **Technology & Transformation for an Unknown Future**, pp. 75–86. 2010.

BELLONI, M. L. **Educação a distância**. São Paulo, SP: Autores Associados. 2006.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para Análise de Sentimentos em Mídias Sociais. **Anais do XXI Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)**, 2015, Manaus – AM. 2015

BERGAMIN, P.B.; ZISKA, S.; WERLEN, E.; SIEGENTHALER, E. The relationship between flexible and self-regulated learning in open and distance universities. **The International Review of Research in Open and Distance Learning**, v.13, n.2, p.101-23, 2012.

BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. **Systematic review in software engineering**. Technical report, RT–ES 679/05 System Engineering and Computer Science Dept., COOPE/UFRJ. 2005.

BITTENCOURT, I. M.; MERCADO, L.P.L. Evasão nos cursos na modalidade de educação a distância: estudo de caso do Curso Piloto de Administração da UFAL/UAB. Ensaio: **Avaliação e Políticas Públicas em Educação**. v.22, n.83, pp. 465-504, abril-junho/2014.

BJØRKEKELUND, E.; BURNETT, T.H.; NØRVÅG, K. A study of opinion mining and visualization of hotel reviews. **Anais 14th International Conference on**

Information Integration and Web-based Applications & Services (pp. 229-238). ACM. 2012.

BOYER, S.; VEERAMACHANENI, K. Transfer learning for predictive models in massive open *online* courses. **Artificial Intelligence in Education**, pp. 54–63. Springer, 2015.

BRASIL. Ministério da Educação. **Referenciais de Qualidade para Educação Superior a Distância**. Disponível em: <http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/636024> Acesso em: 05 de jun. 2017. ISBN 978-85-7863-055-3, 2016.

BROADBENT, J.; POON, W.L. Self-regulated learning strategies & academic achievement in *online* higher education learning environments: A systematic review. **The Internet and Higher Education**, v.27, pp.1-13. 2015

CAVANAUGH, T.; LAMKIN, M.L.; HU, H. Using a generalized checklist to improve student assignment submission times in an *online* course. **Journal of Asynchronous Learning Networks**, v.16, n.4, p.39-44, jun. 2012.

CHATURVEDI, S.; GOLDWASSER, D.; DAUME, H. Predicting instructor's intervention in mooc forums. **Anais: the 52nd Annual Meeting of the Association for Computational Linguistics**, pp. 1501–1511. ACL. 2014.

CHAVES, M.S; DE FREITAS, L.A.; SOUZA, M.; VIEIRA, R. Pirpo: An algorithm to deal with polarity in Portuguese *online* reviews from the accommodation sector. **Anais International Conference on Application of Natural Language to Information Systems**. pp. 296-301. Springer, Berlin, Heidelberg., 2012.

CHEN, S.; LI, Y. A Research of *Fuzzy* AHP Approach in Evaluating Distance Education System Alternatives. **Anais: Education Technology and Computer Science**, pp. 741-745. 2009.

CHEN, Y.; ZHANG, M. MOOC student dropout: pattern and prevention. **Anais: ACM Turing 50th Celebration Conference**. China. n.4, 2017.

CHENG, C. K.; PARÉ, D. E.; COLLIMORE, L. M.; JOORDENS, S. Assessing the effectiveness of a voluntary *online* discussion forum on improving students' course performance. **Computers & Education**, v.56, n.1, p.:253-261. 2011.

CHO, M.H.; SHEN, D. Self-regulation in *online* learning. **Distance education**, v.34, n.3, pp.290-301. 2013

CHOI, H.; LEE, Y.; JUNG, I.; LATCHEM, C. The extent of and reasons for non re-enrolment: A case of Korea National Open University. **Anais: The International Review of Research in Open and Distributed Learning**, v.14, n.4. 2013

CHRYSAFIADI, K.; VIRVOU, M. Evaluating the integration of *fuzzy* logic into the student model of a web-based learning environment. **Expert Systems with Applications**, 39(18), pp.13127-13134. 2012.

CNI – Confederação Nacional da Indústria. Pesquisa CNI-IBOPE: retratos da sociedade brasileira: Educação Profissional. Confederação Nacional da Indústria. – Brasília: CNI. 2014
COSTA, R. Educação a distância: um desafio atual. **Educação a Distância: elementos para pensar o ensino-aprendizagem contemporâneo**. Daniel Mill, Cristiano Maciel (Org.). Cuiabá: EdUFMT, 2013. 369p.

CORPUS, J. H.; MCCLINTIC-GILBERT, M. S.; HAYENGA, A. O. Within-year changes in children's intrinsic and extrinsic motivational orientations: contextual predictors and academic outcomes. **Contemporary Educational Psychology**, v.34, v.2, pp.154-166. 2009

COSTA, R. **Educação a distância: um desafio atual. Educação a Distância: elementos para pensar o ensino-aprendizagem contemporâneo**. Daniel Mill, Cristiano Maciel (Org.). Cuiabá: EdUFMT, 2013. 369p.

COX, E. **The fuzzy systems handbook: a practitioner's guide to building, using, and maintaining fuzzy systems**. New York: AP Professional, 1994.

DAIF, A.R.; RIZKAA, M.A. An Enhanced Model for Monitoring Learners' Performance in a Collaborative e-Learning Environment. **Anais: e-Learning and e-Technologies in Education (ICEEE)**. Lodz, pp. 313 – 317. 2013.

DANIELS, H. **Vygotsky and pedagogy**. Routledge, 2016

DE ALMEIDA, L.R.; DA COSTA, J.P.C.; DE SOUSA, R.T.; DE FREITAS, E.P.; CANEDO, E.D.; PRETTZ, J.; ZACARIAS, E.; DEL GALDO, G. Motivating attendee's participation in distance learning via an automatic messaging plugin for the moodle platform. **Anais Frontiers in Education Conference (FIE)**, 2016 IEEE (pp. 1-5). IEEE. 2016.

DE MATOS, D. P. **Estudos da língua e linguagem na Ead**. 2ed. João Pessoa: Editora da UFPB, 180 p. ISBN: 978-85-237-0955-6. 2014.

DE-LA-FUENTE-VALENTÍN, L.; PARDO, A.; KLOOS, C. D. Addressing drop-out and sustained effort issues with large practical groups using an automated delivery and assessment system. **Computers & Education**. v.61, pp. 33-42. 2013.

DETONI, D.; ARAÚJO, R.M.; CECHINEL, C. Modelagem e Predição de Reprovação de Acadêmicos de Cursos de Educação a Distância a partir da Contagem de Interações. **Revista Brasileira de Informática na Educação**. v.3, n.3. 2015.

DIAS, R.A.; LEITE, L.S. **Educação a distância: da legislação ao pedagógico**. 2.ed. Petrópolis: Vozes, 2010.

DIAS, S. B.; DINIZ, J. A. *FuzzyQoI* model: A *fuzzy* logic-based modelling of users' quality of interaction with a learning management system under blended learning. **Computers & Education**, n. 69, pp. 38-59. 2013.

DICKSON, W.P. Toward a deeper understanding of student performance in virtual high school courses: Using quantitative analyses and data visualization to inform decision making. **A synthesis of new research in K–12 online learning**, pp. 21-23, 2005.

DILLON, J.; BOSCH, N.; CHETLUR, M.; WANIGASEKARA, N.; AMBROSE, G.A.; SENGUPTA, B.; D'MELLO, S.K. Student emotion, co-occurrence, and dropout in a MOOC context. **Anais 9th International Conference on Educational Data Mining**, pp. 353–357, 2016.

EIRINAKI, M.; PISAL, S.; SINGH, J. Feature-based opinion mining and ranking. **Journal of Computer and System Sciences**, 78(4), pp.1175-1184. 2012

FANG, Y.; SI, L.; SOMASUNDARAM, N.; YU, Z. Mining contrastive opinions on political texts using cross-perspective topic model. **Anais ACM international conference on Web search and data mining - WSDM'12**. p. 63. ACM Press, New York, USA. 2012.

FARIA, M. O. F.; ALCANTARA, V. M.; VASCO, C. G. Índice e causa de evasão na modalidade a distância em cursos de graduação: estudo de caso. **Anais: VI Congresso Internacional de Educação Superior**, 2008.

FELDMAN, R. Techniques and applications for sentiment analysis. **Communications of the ACM v.56**, n.4, p.82. 2013

FRISON, J. M. B.; MORAES, M. A. C. As práticas de monitoria como possibilitadoras dos processos de autorregulação das aprendizagens discentes. **Poesis Pedagógica**, v. 8, n. 2, p. 144-158, 2010.

GAGNÉ, M.; DECI, E. L. Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26, 331–362. 2005.

GHAMDI, A.A. The influence of lecturer text-based immediacy on student engagement experiences and learning outcomes in distance education in Saudi Arabia. **Tese de doutorado**, Victoria University, 2017.

GIESBERS, B.; RIENTIES, B.; TEMPELAAR, D.; GIJSELAERS, W. Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing. **Computers in Human Behavior**. v.29, n.1, p.285-292. Janeiro 2013.

GOTTARDO, E.; KAESTNER, C.; NORONHA, R.V. Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados. In: XXXII Congresso da Sociedade Brasileira de Computação, **Anais ...** 2012.

GOUDAS, M., DERMITZAKI, I.; BAGIATIS, K. Predictors of students intrinsic motivation in school physical education. **Journal of Psychology of Education**, v.15, n.3, pp. 271-280. 2000.

GRAU-VALLDOSERA, J.; MINGUILLÓN, J. Redefining dropping out in *online* higher education: A case study from the UOC. **Anais: ACM International Conference Proceeding Series**, pp. 75- 80. 2011.

GRESSLER, L. A. **Introdução à pesquisa**. Edições Loyola. 2003.

GUAY, F.; BOGGIANO, A. K.; VALLERAND, R.J. Autonomy support, intrinsic motivation and perceived competence: conceptual and empirical linkages. **Personality and Social Psychology Bulletin**, v.27, n.6, p. 643-650. 2001.

GUIMARÃES, S. E. R.; BZUNECK, J. A. Propriedades psicométricas de um instrumento para avaliação da motivação de universitários. **Ciências & Cognição**, v.13, n.1, p.101-113. 2008.

GUIXIA, N.; LIYAN, H. Apply the formative assessment and *fuzzy* mathematics into college English *online* learning evaluation. **Anais: Networking and Digital Society (ICNDS), 2010 2nd International Conference on**. Wenzhou,China. 2010. p. 437-440. IEEE. 2010.

GUO, P.J., KIM, J.; RUBIN, R. How video production affects student engagement: An empirical study of mooc videos. **Anais: ACM conference on Learning@ scale conference**, pp.41–50. ACM, 2014.

HALAWA, S.; GREENE, D.; MITCHELL, J. Dropout Prediction in MOOCs using Learner Activity *Features*. **Anais: European MOOCs Stakeholders Summit (EMOOCs)**. Lausana, Suíça. p. 58-65, 2014.

HAIR, J. F.; WILLIAM C. B.; BARRY J. B.; ROLPH E. A.; RONALD L. T. **Análise multivariada de dados**. Bookman Editora, 2009.

HE, J.; BAILEY, J.; RUBINSTEIN, B. I.; ZHANG, R. Identifying at-risk students in massive open *online* courses. **Anais: AAI 2015**, pp.: 1749–1755, Austin Texas, USA, 2015.

HE, P. Evaluating students *online* discussion performance by using social network analysis. In: Information Technology: **Anais: Ninth International Conference on New Generations (ITNG), 2012**, p. 854-855. Baltimore, USA. IEEE, 2012.

HERNANDEZ, R.J. **Development and validation of an instrument to predict *online* student success using faculty perceptions**. 2008. Tese de Doutorado.

HOGO, M. A. Evaluation of e-learning systems based on *fuzzy* clustering models and statistical tools. **Expert systems with applications**, v.37, n.10, pp.6891-6903. 2010.

HU, M.; LIU, B. Mining and summarizing customer reviews. **Anais KDD 2004** (New York, New York, USA, 2004), ACM, ACM Press, p. 168. 2004

HUDEC, M., 2016. *Fuzzy Inference*. **Fuzziness in Information Systems** (pp. 101-137). Springer International Publishing.

JOO, Y. J.; JOUNG, S.; SIM, W. J. Structural relationships among internal locus of control, institutional support, flow, and learner persistence in cyber universities. **Computers in Human Behavior**, v. 27, pp. 714–722. 2011

KAO, Y. T.; LIN, Y. S.; CHU, C. P. A Multi-factor *Fuzzy Inference* and Concept Map Approach for Developing Diagnostic and Adaptive Remedial Learning Systems. **Procedia-Social and Behavioral Sciences**, n.64, pp. 65-74. 2012.

KASSAMBARA, A. **Practical guide to cluster analysis in r: unsupervised machine learning**. 1 ed. [S.L.]: STHDA. 186 p. 2017.

KAUFMAN, L.; ROUSSEEUW, P.J. **Finding groups in data: an introduction to cluster analysis** (Vol. 344). John Wiley & Sons. 2009.

KEN, F. *et al.* A Tree-Based Decision Model to Support Prediction of the Severity of Asthma Exacerbations in Children. **Journal of Medical Systems**, v.34, p.551-562, 2010.

KITCHENHAM, B. **Procedures for performing systematic reviews**. Keele, UK, Keele University. 2004. Disponível em: <<http://www.inf.ufsc.br/~awangenh/kitchenham.pdf>>. Acesso em: 05 de jun. 2014

KLOFT, M., STIEHLER, F., ZHENG, Z., PINKWART, N. Predicting MOOC dropout over weeks using machine learning methods. **Anais: EMNLP Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses**, pp.60–65. 2014.

KOHAVI, Ron; PROVOST, Foster. Glossary of terms. **Machine Learning**, v. 30, n. 2-3, p. 271-274, 1998.

LEAL, E. A.; MIRANDA, G.J.; CARMO, C.R.S. Teoria da autodeterminação: uma análise da motivação dos estudantes do curso de ciências contábeis. **Revista Contabilidade & Finanças-USP**, 24(62). 2013.

LEE, Y.; CHOI, J. A review of *online* course dropout research: implications for practice and future research. **Educational Technology Research and Development**, v. 59, n. 5, pp. 593-618. 2011.

LEEDS, E.; CAMPBELL, S.; BAKER, H.; ALI, R.; BRAWLEY, D.; CRISP, J. The impact of student retention strategies: An empirical study. **International Journal of Management in Education**, v.7, n.1-2, pp.22-43. 2013.

LEI, C.U.; OH, E.; LEUNG, E.; GONDA, D.; QI, X.; LEUNG, R.; KWOK, T.T.; YEUNG, Y.C.A.; LAU, R. Scale Out Teaching, Scale Up Learning: Professional development for e-teaching/learning. **Anais: Teaching, Assessment, and**

Learning for Engineering (TALE), 2016 IEEE International Conference on. pp. 265-270. IEEE. 2016.

LIMA, I.; PINHEIRO, C.; OLIVEIRA, F.C. **Inteligência Artificial**. 1ª Edição. ISBN: 9788535278088. Escala. 2014

LITTO, F.; FORMIGA, M. **Educação a distância: o estado da arte**. 5ª ed. São Paulo: Pearson, 2009. 461 p.

LIU, B. **Sentiment Analysis and Subjectivity**. Handbook of natural language processing, v.2, pp. 627-666. 2010

LIU, S. Y.; GOMEZ, J.; YEN, C. J. Community College *Online* Course Retention and Final Grade: Predictability of Social Presence. **Journal of Interactive Online Learning**, v.8, n. 2, pp.165-184. 2009.

LÓPEZ-PÉREZ, M. V.; PÉREZ-LÓPEZ, C. M.; RODRÍGUEZ-ARIZA, L. Blended learning in higher education: Students' perceptions and their relation to outcomes. **Computers & Education**, v.1, n.56, pp. 818–826. 2011.

MACFADYEN, L.P.; DAWSON, S. Mining LMS data to develop an “early warning system” for educators: A proof of concept. **Computers & Education**, v. 54, n. 2, p. 588-599, 2010.

MAMDANI, E.; ASSILLAN, S. An Experiment in Linguistic Synthesis with a *FuzzyLogic* Controller. **International Journal of Man-Machine Studies**, v.7, p.1–13, 1975.

MAMMAN, B.; YUSOF, A.; ABUHASSNA, H. M.; ALY, H.; AL-AHMADI, T.; ATAN, N. A.; KHAIR, F. **Design and Learning Strategies Applied in Mooc: A Meta-Analysis**. Sains Humanika, v. 9, p.1-4. 2017.

MANHÃES, L. M. B.; DA CRUZ, S. M. S.; COSTA, R. J. M.; ZAVALETA, J.; ZIMBRAO, G. Previsão de Estudantes com Risco de Evasão o Utilizando Técnicas de Mineração de Dados. **Anais do XXII SBIE**. 2011.

MANNING, C.D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**, first ed., Cambridge University Press, New York, 2008.

MARTINELLI, S. C.; SISTO, F. F. Motivação de estudantes: um estudo com crianças do ensino fundamental. **Avaliação Psicológica**, v.9, n.3, p.413-420. 2010.

MAURÍCIO, W. P. D. Evasão E Desistência Na Educação A Distância: Uma Problemática Em Discussão. In: XI Congresso Nacional De Educação – EDUCERE . **Anais...** Curitiba-PR. 2013, p. 5301-5312.

MEC. Referenciais de Qualidade para Educação Superior a Distância. Disponível em: <http://portal.inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/636024> Acesso em: 05 de jun. 2017. ISBN 978-85-7863-055-3, 2016.

MEC. Secretaria de Educação Superior / Ministério da Educação. Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Brasília, 1996/1997 Disponível em: <<http://www.dominiopublico.gov.br/download/texto/me001613.pdf>>. Acesso em: 22/11/2017. 1997

MEZZARI, A.; ROCKENBACH, L.; GORZIZA, B.; RIBAS, G.; FAVERO, R.; BULEGON, A. Estratégias para detecção precoce de propensão à evasão. **RIED**, v. 16, n. 2, pp 147-175. 2013. ISSN: 1138-2783.

MI, F; YEUNG, D.Y. Temporal models for predicting student dropout in massive open *online* courses. **Anais: 2015 IEEE International Conference on Data Mining Workshop (ICDMW)**. IEEE, 256–263. 2015

MISHRA, T.; KUMAR, D.; GUPTA, S. Mining Students' Data for Prediction Performance. **Anais: Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on**. v.8, n.9, pp.255-262. 2014

MOORE, M.G. **Three types of interaction**. 1989.

MORAIS, A. M., ARAUJO, JMFR, COSTA, EB. "Perfil do Docente em Educação a Distância e suas Perspectivas Frente aos Processos Avaliativos." **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. v. 3. n. 1. 2014.

MURRAY, M.; PÉREZ, J.; GEIST, D.; HEDRICK, A. Student Interaction with Content in *Online* and Hybrid Courses: Leading Horses to the Proverbial Water, **Informing Science**, v. 16, 2013.

NAGRECHA, S.; DILLON, J.Z.; CHAWLA, N.V. MOOC Dropout Prediction: Lessons Learned from Making Pipelines Interpretable. **Anais: 26th International Conference on World Wide Web Companion**. pp.: 351-359. 2017.

NILSSON, N.J. **Principles of artificial intelligence**. Morgan Kaufmann. 2014

NUNES, I. B. A história da Ead. **Educação a Distância: o estado da arte**. ABED. São Paulo, Pearson Education do Brasil, 2009. ISBN: 978-85-7605-197-8.

O'HARE, N.; DAVY, M.; BERMINGHAM, A.; FERGUSON, P.; SHERIDAN, P.; GURRIN, C.; SMEATON, A.F. Topic-dependent sentiment analysis of financial blogs. **Anais Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion** (pp. 9-16). ACM. 2009.

ONAH, D.F.O.; SINCLAIR, J.; BOYATT, R. Dropout rates of massive open *online* courses: behavioural patterns. **Anais: the Sixth International Conference on Education and New Learning Technologies**, pp. 5825–5834. 2014.

OPEN UNIVERSITY. **Anual report 2015-2016**. Disponível em: <<http://www.open.ac.uk/about/main/sites/www.open.ac.uk.about.main/files/files/ecms/web-content/about-annual-report-2015-2016.pdf>>. Acesso em: 01 mai. 2017.

PAIVA, M. L. M. F.; BORUCHOVITCH, E. Orientações motivacionais, crenças educacionais e desempenho escolar de estudantes do ensino fundamental. **Psicologia em Estudo**, 15(2), 381-389. 2010

PANSERA, S.M.; VALENTINI, N.C.; SANTAYANA DE SOUZA, M.; BERLEZE, A. Motivação intrínseca e extrínseca: diferenças no sexo e na idade. **Psicologia Escolar e Educacional**, São Paulo-SP, v. 20, n. 2, p. 313-320, mai./ago. 2016.

PARK, J.H.; CHOI, H. J. Factors Influencing Adult Learners' Decision to Drop Out or Persist in *Online* Learning. **Educational Technology & Society**, v.12, n.4, pp. 207–217. 2009.

PAVESI, M.A.; ALLIPRANDINI, P.M.Z. Autorregulação da Aprendizagem de Alunos de Cursos a Distância em Função do Sexo. **Revista de Ensino, Educação e Ciências Humanas**, 16(2), pp.100-108. 2015

PERRENOUD, P. **Avaliação: da excelência à regulação das aprendizagens: entre duas lógicas**. Porto Alegre: Artmed. 1999.

PETERS, O. **A educação a distância em transição: tendências e desafios**. Editora Unisinos, 2003.

PIAGET, J. **Principles of genetic epistemology: Selected works**. Vol. 7. Routledge, 2013.

PINTRICH, P. R. A conceptual framework for assessing motivation and self-regulated learning in college students. **Educational Psychology Review**, Florida, v. 16, n. 4, p. 385-406, 2004.

POLYDORO, S. A. J.; AZZI, R. G. Autorregulação da aprendizagem na perspectiva da teoria sociocognitiva: introduzindo modelos de investigação e intervenção. **Psicologia educativa**, São Paulo, n. 29, p. 75-94, 2009.

PRATES, R. O. Interação em sistemas colaborativos. In: PIMENTEL, Mariano; FUKS, Hugo. **Sistemas Colaborativos**. São Paulo-SP: Editora Campus, 2012. 264 e 293.

RABBANY, R.; TAKAFFOLI, M.; ZAIANE, O. Analyzing participation of students in *online* courses using social network analysis techniques. **Anais: Educational Data Mining**, pp. 21–30. 2011.

RIBEIRO, I. S.; SILVA, C. F. Autorregulação: diferenças em função do ano e área em alunos universitários. **Psicologia: Teoria e Pesquisa**, Brasília, v. 23, n. 4, p. 443- 448, out./dez. 2007.

RODRIGUES, R. L.; DE MEDEIROS, F. P. A.; GOMES, A. S. Modelo de Regressão Linear aplicador à previsão de desempenho de estudantes em ambiente de aprendizagem. **Anais do XXIV SBIE**. 2013

ROMERO, C.; LÓPEZ, M. I.; LUNA, J. M.; Ventura, S. Predicting students' final performance from participation in on-line discussion forums. **Computers & Education**, n. 68, pp. 458-472. 2013.

ROSÁRIO, P.; NÚÑEZ, J. C.; VALLE, A.; GONZÁLEZ-PIENDA, J.; LOURENÇO, A. Grade level, study time, and grade retention and their effects on motivation, self-regulated learning strategies, and mathematics achievement: a structural equation model. **European Journal of Psychology of Education**, 28(4), 1311- 1331. doi: 10.1007/s10212-012-0167-9. 2013.

ROSTAMINEZHAD, M. A.; MOZAYANI, N.; NOROZI, D.; IZIJ, M. Factors Related to E-learner Dropout: Case Study of IUST Elearning Center. **Procedia-Social and Behavioral Sciences**, v. 83 p. 522-527. 2013.

ROUSSEEUW, P.J. Silhouettes: A graphical and to the interpretation and validation of cluster analysis. **J. Comput. Appl. Math.**, 20, 53-65. 1987

RUSSELL, S.; NORVIG, P. **Inteligência Artificial** (3ª Edição). ISBN: 978-85-352-3701-6. Rio de Janeiro: Elsevier, 2013.

SALEH, I.; KIM, S. I. A *fuzzy* system for evaluating students' learning achievement. **Expert Systems with Applications**, v. 36, n.3, p. 6236-6243. 2009.

SALES, P. A. O.; ABBAD, G.; RODRIGUES, J. L. Variáveis preditivas de evasão e persistência em treinamentos a distância [CD-Rom]. Associação Nacional dos Programas de Pós-Graduação em Administração (Org.). **Anais do XXXV ENANPAD**. Rio de Janeiro: ANPAD. 2011.

SANCHEZ, F. (Coor). **Anuário brasileiro estatístico de educação aberta e a distância-ABRAEAD** -- 4. ed. -- São Paulo: Instituto Monitor, 2008.

SANTOS, E. M.; NETO, J. D. O. Evasão na Educação a Distância: identificando causas e propondo estratégias de prevenção. Universidade Metropolitana de Santos (Unimes). Núcleo de Educação a Distância - Unimes Virtual. Revista Científica de Educação a Distância. v.2, n.2, 2009.

SARGENT, R.G. Verification and validation of simulation models. In: 37th conference on Winter simulation. **Proceedings** ... pp. 130-143. winter simulation conference. 2005.

SARMENTO, L.; CARVALHO, P.; SILVA, M.J.; DE OLIVEIRA, E. Automatic creation of a reference corpus for political opinion mining in user-generated content. **Anais: International CIKM workshop on Topic-sentiment analysis for mass opinion** (pp. 29-36). ACM. 2009.

SCHUNK, D. H. Self-efficacy and academic motivation. **Educational Psychologist**, Greensboro, 3-4 (26), 207-231. 1991

SHARKEY, M.; SANDERS, R. A process for predicting mooc attrition. **Anais: EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs**, pages 50–54, 2014.

SHARMA, A.; DEY, S. A comparative study of feature selection and machine learning techniques for sentiment analysis. **Anais ACM research in applied computation symposium** (pp. 1-7). ACM. 2012.

SHEVLYAKOV, G.L.; OJA, H. **Robust correlation: Theory and applications** (Vol. 3). John Wiley & Sons. 2016.

SIDDIQUE, N.; ADELI, H. **Computational Intelligence: Synergies of fuzzy logic, neural networks and evolutionary computing**. ISBN: 9781118337844. Wiley. 2013.

SIEMENS, G. Connectivism: a learning theory for a digital age. **International Journal of Instructional Technology and Distance Learning**, v.2, n.1, 2005.

SILVA, A.; DUARTE, A.; SÁ, I.; SIMÃO, A. **Aprendizagem auto-regulada pelo estudante: perspectivas psicológicas e educacionais**. (pp. 11-39). Porto: Porto Editora. 2004.

SINHA, T.; JERMANN, P.; LI, N.; DILLENBOURG, P. Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions. **Anais Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses**. 2014.

STEIN, R.M.; ALLIONE, G. Mass attrition: An analysis of drop out from a principles of microeconomics MOOC. **Social Science Research Network**, pp. 1–19, 2014.

SUN, P.; TSAI, R.J.; FINGER, G; CHEN, Y.Y.; YEH, D. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. **Computers & Education**, v.50, n.4, p. 1183–1202. 2008.

TAMHANE, A.; IKBAL, S.; SENGUPTA, B.; DUGGIRALA, M.; APPLETON, J. Predicting student risks through longitudinal analysis. In: **Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining** (pp. 1544-1552). ACM. 2014.

TANG, J.K.T.; XIE, H.; WONG, T. A Big Data Framework for Early Identification of Dropout Students in MOOC. **Anais: International Conference on Technology in Education Technology in Education**. Technology-Mediated Proactive Learning. pp 127-132. 2015

TAYLOR, C.; VEERAMACHANENI, K.; O'REILLY, U.M. Likely to stop? predicting stopout in massive open *online* courses. **CoRR - ArXiv 2014**. arXiv preprint arXiv:1408.3382, v. 1408.3382, 2014

TRAVASSOS, G. H., GUROV, D.; AMARAL, E. A. G. "Introdução à Engenharia de Software Experimental". COPPE/UFRJ, Rio de Janeiro, **Relatório técnico: RT-ES-590/02**, <http://www.ufpa.br/cdesouza/teaching/topes/4-ES-Experimental.pdf>. 2002.

TUMITAN, D.; BECKER, K. Tracking Sentiment Evolution on User-Generated Content: A Case Study on the Brazilian Political Scene. **Anais SBBB** (Short Papers) (pp. 24-1). 2013

VASCONCELOS, Y.L.; YOSHITAKE, M.; DE FRANÇA, S.M.; CAVALCANTI, A.M.F.A. Construtivismo na educação a distância. **Revista de Ensino, Educação e Ciências Humanas**, 16(4), pp.338-348. 2015

VILARINHO, L.G.; PARO, E.M. Evasão de alunos na educação superior a distância: A experiência do módulo de acolhimento. **Voces y Silencios**. Bogotá. v.2, n.2. 2011.

VYGOTSKY, L. S. **Mind in society**: The development of higher psychological processes. Cambridge, MA: Harvard University Press. 2008.

WAINER, J. Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação. **Atualização em informática**, 1, pp.221-262. 2007.

WALTERS-ARCHIE, A. Academic support for *online* students in the English-speaking Caribbean at the University of the West Indies Open Campus. **Journal of Further and Higher Education**, p. 1-11. 2017.

WANG, W.; YU, H.; MIAO, C. Deep Model for Dropout Prediction in MOOCs. **Anais:2nd International Conference on Crowd Science and Engineering**. Pp.: 26-32. 2017

WECHSLER, S. M. **Manual estilos de pensar e criar**. São Paulo: LAMP/PUC. 2006.

WEERASINGHE, A.; DU BOULAY, B.; BISWAS, G. Workshop on Self-Regulated Learning in Educational Technologies (SRL@ ET): Supporting, Modeling, Evaluating, and Fostering Metacognition with Computer-Based Learning Environments. **Artificial Intelligence in Education** (pp. 956-956). Springer Berlin Heidelberg. 2013.

WEINERT, F. E. Selbstgesteuertes Lernen als Voraussetzung, Methode und Ziel des Unterrichts [Self-regulated learning as prerequisite, method and objective of instruction], **Unterrichtswissenschaft**, 10 (2), pp. 99–110. 1982

WEON, S.; KIM, J. Learning achievement evaluation strategy using *fuzzy* membership function. **Anais: Frontiers in Education Conference**, 2001. 31st Annual (Vol. 1, p. T3A-19). IEEE. 2001.

WHITEHILL, J.; MOHAN, K.; SEATON, D.; ROSEN, Y.; TINGLEY, D. MOOC Dropout Prediction: How to Measure Accuracy? **Anais: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale** (pp. 161-164). ACM. 2017.

WIVES, L. K.; BERCHT, M.; BASTOS, H. P. P. Análise manual e automática de pistas lexicais de presença social em chat educacional. In: Simpósio Brasileiro de Informática na Educação 2010. **Anais ...** 2010.

WOHLIN, C.; RUNESON, P.; HOST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. **Experimentation in Software Engineering**. Springer. 2012.

WU, J. **Advances in K-means Clustering: A Data Mining Thinking**. Springer. 2012.

XENOS, M.; PIERRAKEAS, C.; PINTELAS, P. A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. **Computers & Education**, v. 39, n.4, p.361-377. 2002.

YACCI, M. Interactivity demystified: A structural definition for distance education and intelligent CBT. **Educational Technology**, 40(4), pp.5-16. 2000

YANG, D.; SINHA, T.; ADAMSON, D.; ROSÉ, C. P. Turn on, tune in, drop out: Anticipating student dropouts in massive open *online* courses. **Anais: 2013 NIPS Data-Driven Education Workshop**. Nevada, USA. 2013. p. 1-8.

ZIMMERMAN, B. J. **Attaining self-regulation: a social cognitive perspective**. Em: M. Boekaerts; P. Pintrich e M. Zeidner (eds.). Handbook of Self-Regulation (13-39). New York: Academic Press. 2000.

ZIMMERMAN, B. J. Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. **American Educational Research Journal**, 45(1), 166-183. doi: 10.3102/0002831207312909. 2008.

ZISIMOPOULOS, D. A.; GALANAKI, E. P. Academic intrinsic motivation and perceived academic competence in greek elementary students with and without learning disabilities. **Learning Disabilities Research and Practice**, v.24, n.1, p. 33-43. 2009.

M827a

Morais, Alana Marques de.

Abordagem avaliativa multidimensional para previsão da evasão do discente em cursos online / Alana Marques de Moraes. – Campina Grande, 2018.

158 f.: il. color.

Tese (Doutorado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2018.

"Orientação: Profa. Dra. Joseana Macêdo Fechine Régis de Araújo, Prof. Dr. Evandro de Barros Costa".

Referências.

1. Inteligência Artificial. 2. Avaliação Multidimensional. 3. Educação à Distância - Evasão. 4. Mineração de Texto. 5. Lógica *Fuzzy*. 6. *FuzzySD*. I. Silva, Osvaldo Soares da. II. Maribondo, Juscelino de Farias. III. Título.

CDU 004.8(043)