

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Edilaine Santiago de Oliveira

**Uma Abordagem Baseada no Uso de Métricas do Modelo LV para
Identificação Automática de Estilos de Aprendizagem em Ambientes
Virtuais de Aprendizagem**

Fortaleza – CE

29 de abril de 2019

Edilaine Santiago de Oliveira

Uma Abordagem Baseada no Uso de Métricas do Modelo LV para Identificação Automática de Estilos de Aprendizagem em Ambientes Virtuais de Aprendizagem

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Computação Aplicada

Orientador: Dr. Gilvandenys Leite Sales

Coorientador: Dr. Amauri Holanda de Souza Júnior

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

29 de abril de 2019

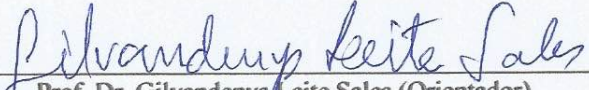
EDILAINE SANTIAGO DE OLIVEIRA

Uma Abordagem Baseada no Uso de Métricas do Modelo LV para Identificação Automática
de Estilos de Aprendizagem em Ambientes Virtuais de Aprendizagem

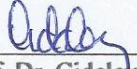
Dissertação submetida à Coordenação do Curso de Pós-graduação em Ciência da
Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como
requisito parcial para a obtenção do título de Mestre em Ciência da Computação, área de
concentração Ciência da Computação.

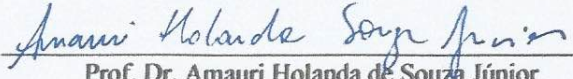
Aprovada em 06/02/2019

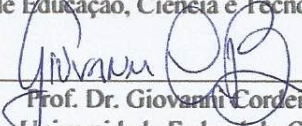
BANCA EXAMINADORA


Prof. Dr. Gilvandenys Leite Sales (Orientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE


Prof. Dr. Antônio Wendell de Oliveira Rodrigues
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - PPGCC/IFCE


Prof. Dr. Cidcley Teixeira de Souza
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - PPGCC/IFCE


Prof. Dr. Amauri Holanda de Souza Júnior
Instituto Federal de Educação, Ciência e Tecnologia do Ceará - PPGCC/IFCE


Prof. Dr. Giovanni Cordero Barroso
Universidade Federal do Ceará - UFC

*Dedico este trabalho à Deus, aos meus pais: Marcondes e Rozeneida e ao meu esposo
Ramires, por todo apoio e colaboração para a realização desta pesquisa*

Agradecimentos

Muitas pessoas colaboraram de alguma forma, para a realização deste trabalho. Aqui deixo meus sinceros agradecimentos:

Primeiramente agradeço à Deus, Senhor do universo, por todo seu amor e por tudo que faz por mim. Sem Ele nada sou.

Em segundo lugar agradeço aos meus pais: Marcondes Martins de Oliveira e Rozeneida da Silva Santiago Oliveira, por todos seus ensinamentos, por todo amor e por toda educação que me deram, em especial à minha mãe, pela sua dedicação, pelo seu apoio e por compreender minha ausência durante este período. À vocês, minha eterna gratidão.

Agradeço também ao meu esposo, Ramires Moreira, por sempre acreditar em mim e me ajudar a conquistar meus objetivos. Obrigada pelo seu amor e companheirismo. Amo muito você.

Ao meu orientador, professor Dr. Gilvandenys Leite Sales, por ter me aceitado como sua orientanda, me dando um voto de confiança. Obrigada por todas contribuições e oportunidades que me fizeram crescer na carreira acadêmica.

Agradeço ao meu co-orientador, professor Dr. Amauri Holanda de Souza Júnior, pelo tempo disponível em me ajudar, por todos os caminhos que me apresentou e por ter me mostrado uma direção a seguir.

À todos os professores do PPGCC, em especial aos que tive a honra de compartilhar alguns momentos: Dra. Ananda Freire, Dr. Cidley Souza, Dr. Nivando Bezerra, Dr. Geraldo Ramalho. Obrigada por me proporcionarem a oportunidade de aprender coisas novas e por toda dedicação que possuem como docentes.

Ao coordenador do PPGCC, Dr. Pedro Pedrosa, por sua disponibilidade e prontidão em nos ajudar e pela sua dedicação ao programa.

Agradeço à banca, por aceitar o convite e por todas as contribuições nesta pesquisa.

Aos meus colegas de turma, que me ajudaram e compartilharam comigo bons e maus momentos. Em especial a Laysa e a Pryscilla, amigas que o mestrado me trouxe.

À todos os meus amigos de trabalho pelo incentivo, em especial ao meu coordenador, Mateus Vieira, pela compreensão durante este período de pesquisa.

E a todos aqueles que, direta ou indiretamente, contribuíram para a realização deste sonho.

*“Ora, àquele que é poderoso para fazer tudo muito mais abundantemente além daquilo que pedimos ou pensamos, segundo o poder que em nós opera, à Ele seja a glória”
(Efésios 3:20-21)*

Resumo

Na Educação a Distância (EaD), a maioria dos Ambientes Virtuais de Aprendizagem (AVA) abordam o conteúdo da mesma maneira para todos os alunos. Esse tipo de abordagem pode não favorecer no desempenho destes, visto que cada um tem suas próprias preferências de aprendizagem. A entrega de conteúdos adaptáveis dentro do AVA, onde cada aluno terá acesso ao ambiente de acordo com o seu perfil, assim como métodos avaliativos diferenciados, são exemplos de melhorias que podem ser feitas através da identificação dos Estilos de Aprendizagem (EA). Diante deste contexto, delimita-se o objetivo de pesquisa deste trabalho como sendo um estudo comparativo referente à modelagem preditiva dos estilos de aprendizagem dos alunos através de uma abordagem com base nos atributos da literatura, gerados pelas interações dos alunos, e uma abordagem com base nos atributos gerados pelo modelo de avaliação *Learning Vectors* (Modelo LV), no ambiente virtual de aprendizagem MOODLE (*Modular Object-Oriented Dynamic Learning Environment*). O método utilizado nesta pesquisa foi baseado na metodologia KDD (*Knowledge Discovery in Databases*), passando pelas fases de entendimento do domínio, seleção e coleta dos dados, pré-processamento, formatação e modelagem dos dados e por último avaliação dos algoritmos de classificação. O modelo de EA utilizado foi o modelo de Felder e Silverman, que classifica o aluno em quatro dimensões (processamento, percepção, captação, organização). Já o processo de modelagem realizou-se por meio de dados extraídos da base de dados de um curso EaD. Como resultado da fase de modelagem, foi realizado o desenvolvimento, avaliação e comparativo de cinco modelos de classificação, por meio dos algoritmos: *K Nearest Neighbor* (KNN), Redes Bayesianas, *Multi-Layer Perceptron* (MLP), *Decision Tree* (DT) e *Support Vector Machine* (SVM). O modelo SVM apresentou-se como sendo o mais satisfatório, sendo capaz de prever o perfil dos alunos, com base nos atributos do Modelo LV, com uma taxa média de acurácia nas quatro dimensões de 81,00% e uma média da área sobre a curva ROC de 83,36%. Já na classificação com base nos atributos da literatura, o SVM obteve uma média de acurácia de 75,76% e média de área sobre a curva ROC de 76,85%. Os resultados mostram que o método proposto com base no Modelo LV, alcança valores satisfatórios que são comparáveis e até melhores que os métodos tradicionais da literatura, fornecendo assim uma alternativa válida para tal abordagem.

Palavras-chave: Estilos de Aprendizagem. Algoritmos de Classificação. Ambientes Virtuais de Aprendizagem. Educação a Distância. *Learning Vectors*.

Abstract

In Distance Education (DE), most Learning Management System (LMS) approach content in the same way for all students. This type of approach may not favor the performance this students, since each has its own learning preferences. The delivery of adaptive content within the LMS, where each student will have access to the environment according to his / her profile, as well as differentiated evaluation methods, are examples of improvements that can be made through the Identification of Learning Styles (LS). In this context, the research objective of this work is a comparative study regarding the predictive modeling of the learning styles of the students through an approach based on the attributes of the literature, generated by the interactions of the students, and in the attributes generated by the Learning Vectors (LV) Model within the learning management system MOODLE (Modular Object-Oriented Dynamic Learning Environment). The method used in this research was based on the KDD (Knowledge Discovery in Databases) methodology, through the phases of understanding the domain, selecting and collecting the data, preprocessing, formatting and modeling the data and finally evaluation of the classification algorithms. The LS model used was the model of Felder and Silverman, which classifies the student in four dimensions (processing, perception, capture, organization). The modeling process was accomplished using data extracted from the database of an DE course. As a result of the modeling phase, we developed, evaluated and compared five classification models, using K Nearest Neighbor (KNN), Bayesian Networks, Multi-Layer Perceptron (MLP), Decision Tree (DT) and Support Vector Machine (SVM). The SVM model was the most satisfactory, being able to predict the profile of the students, based on the attributes of the Model LV, with an average accuracy rate in the four dimensions of 81.00% and a mean of the area over the ROC curve of 83.36%. The classification based on the attributes of the literature, the SVM obtained an average accuracy of 75.76% and area average over the ROC curve of 76.85%. The results illustrate that the proposed method based on the Model LV achieves satisfactory values that are comparable and even better than the traditional literature methods, thus providing a valid alternative for such an approach.

Keywords: Learning Styles. Classification Algorithms. Learning Management System. Distance Education.

Lista de ilustrações

Figura 1 – Evolução do total de matrículas em cursos regulamentados totalmente a distância	26
Figura 2 – Escala do Estilo de Aprendizagem	35
Figura 3 – Estilos de Aprendizagem de Kolb	40
Figura 4 – Vetor-Aprendizagem	44
Figura 5 – Fator β - Positividade X Negatividade	45
Figura 6 – Vetor-Aprendizagem: (a) Nota mínima e (b) Nota máxima	45
Figura 7 – Modelo tecno-metodológico LV	46
Figura 8 – Ícones LV e menções qualitativas	47
Figura 9 – Hierarquia do aprendizado de acordo com o grau de supervisão	49
Figura 10 – Exemplo de classificação com KNN	51
Figura 11 – Rede Perceptron Simples	52
Figura 12 – Exemplo de uma MLP com duas camadas intermediárias	53
Figura 13 – SVM - Exemplo de Hiperplano Ótimo	54
Figura 14 – Exemplo de Árvore de Decisão	55
Figura 15 – Mapa conceitual dos Classificadores Descritos	56
Figura 16 – Etapas do Processo de Revisão	57
Figura 17 – Artigos selecionados por repositório	60
Figura 18 – Aplicação dos critérios de inclusão e exclusão	60
Figura 19 – Técnicas usadas para identificação dos estilos de aprendizagem	64
Figura 20 – KDD - <i>Knowledge Discovery in Databases</i>	69
Figura 21 – Preparação dos Dados	74
Figura 22 – Comparação da categorização da variável alvo	80
Figura 23 – Resultados dos Algoritmos de Classificação - Atributos da Literatura	85
Figura 24 – Resultados dos Algoritmos de Classificação - Atributos LV	86
Figura 25 – Médias das Acurácias Por Classificador	87
Figura 26 – Melhores Acurácias por Dimensão	87
Figura 27 – Grid Search - KNN - Atributos Literatura	113
Figura 28 – Grid Search - KNN - Atributos LV	115
Figura 29 – Grid Search - SVM - Atributos Literatura	116
Figura 30 – Grid Search - SVM - Atributos LV	117

Lista de tabelas

Tabela 1 – Dimensões do Modelo FSLSM	34
Tabela 2 – Resumo dos Modelos de Estilos de Aprendizagem	42
Tabela 3 – <i>Strings</i> de busca usadas nos repositórios	58
Tabela 4 – Critérios de Inclusão e Exclusão	59
Tabela 5 – Distribuição de artigos por ano	61
Tabela 6 – Países das pesquisas selecionadas	62
Tabela 7 – Modelos de Estilos de Aprendizagem	63
Tabela 8 – Algoritmos utilizados	65
Tabela 9 – Síntese dos trabalhos selecionados	66
Tabela 10 – Dados dos cursos de pós-graduação a distância - UFC Sobral	71
Tabela 11 – Padrões de comportamento relevantes para cada dimensão de Estilo de Aprendizagem do FSLSM	72
Tabela 12 – Ferramentas Educacionais e Atributos Extraídos	73
Tabela 13 – <i>Datasets</i> Formatados após Seleção de Atributos	75
Tabela 14 – Distribuição dos Estilos de Aprendizagem por Dimensão	79
Tabela 15 – Características <i>Datasets</i>	81
Tabela 16 – Resultados Classificação Atributos Literatura	82
Tabela 17 – Resultados Classificação Atributos LV	84

Lista de Siglas

ABED	Associação Brasileira de Educação a Distância
ACMDL	<i>Association for Computing Machinery Digital Library</i>
AUC	<i>Area Under the ROC Curve</i>
AVA	Ambientes Virtuais de Aprendizagem
CP	Coeficiente de Passo
CSV	<i>Comma-separated values</i>
DT	<i>Decision Tree</i>
EA	Estilos de Aprendizagem
EJA	Educação de Jovens e Adultos
FN	Falso Negativo
FP	Falso Positivo
FSLSM	<i>Felder and Silverman Learning Styles Model</i>
GRSLSS	<i>Grasha-Riechmann Student Learning Style Scales</i>
IA	Inteligência Artificial
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
KDD	<i>Knowledge Discovery in Databases</i>
KNN	<i>K Nearest Neighbor</i>
L-BFGS	<i>Limited-memory Broyden–Fletcher–Goldfarb–Shanno</i>
LSI	<i>Learning Styles Inventory</i>
LSQ	<i>Learning Styles Questionnaire</i>
LV	<i>Learning Vectors</i>
MBTI	<i>Myers-Briggs Type Indicator</i>
MLP	<i>Multi-Layer Perceptron</i>
MOOC	<i>Massive Open Online Course</i>

MOODLE *Modular Object-Oriented Dynamic Learning Environment*

NB *Naive Bayes*

PHP *PHP: Hypertext Preprocessor*

PS *Perceptron Simples*

RBF *Radial Basis Function*

RFE *Recursive Feature Elimination*

RNA *Rede Neural Artificial*

ROC *Receiver Operating Characteristics*

RSL *Revisão Sistemática da Literatura*

SEDUC *Secretaria de Educação*

SMOTE *Synthetic Minority Over-sampling Technique as presented*

SQL *Structured Query Language*

SVM *Support Vector Machine*

UFC *Universidade Federal do Ceará*

VARK *Visual, Auditivo, Leitura/Escrita e Sinestésico*

VN *Verdadeiro Negativo*

VP *Verdadeiro Positivo*

WEKA *Waikato Environment for Knowledge Analysis*

Lista de símbolos

α	Alfa
β	Beta
Δ	Delta
Σ	Somatório
θ	Teta
Δ_T	Variação Angular Total do vetor

Sumário

1	INTRODUÇÃO	25
1.1	Motivação	28
1.2	Descrição do Problema	29
1.3	Objetivos Geral e Específicos	30
1.4	Produção Científica	31
1.5	Estrutura da Dissertação	32
2	ESTILOS DE APRENDIZAGEM	33
2.1	O Modelo de Felder e Silverman	34
2.2	O Modelo de Myers-Briggs	36
2.3	O Modelo de Grasha-Riechmann	37
2.4	O Modelo de Honey-Mumford	38
2.5	O Modelo de VARK	39
2.6	O Modelo de Kolb	39
2.7	O Modelo de Dunn	41
3	MODELO LEARNING VECTORS (LV)	43
3.1	<i>Learning Vectors</i> e seus componentes	43
3.2	Avaliação usando <i>Learning Vectors</i>	46
3.3	Trabalhos Relacionados ao Modelo LV	48
4	RECONHECIMENTO DE PADRÕES	49
4.1	K Vizinhos Mais Próximos	50
4.2	Classificador Bayesiano	51
4.3	Rede Perceptron de Multicamadas	52
4.4	Máquina de Vetores de Suporte	54
4.5	Árvores de Decisão	55
5	REVISÃO SISTEMÁTICA DA LITERATURA	57
5.1	Processo de Revisão Sistemática	57
5.1.1	Planejamento	57
5.1.2	Execução	59
5.1.3	Análise e Interpretação dos Resultados	61
5.1.3.1	Informações Gerais dos Resultados	61
5.1.3.2	Respostas para as Questões de Pesquisa	62
5.1.4	Considerações sobre o capítulo	66

6	METODOLOGIA	69
6.1	Caracterização da Pesquisa e Entendimento do Domínio	70
6.2	Seleção e Coleta dos Dados	71
6.3	Pré-Processamento	73
6.4	Formatação dos Dados	75
6.5	Modelagem	76
6.6	Interpretação e Avaliação	77
7	RESULTADOS E DISCUSSÕES	79
7.1	Resultados sobre o entendimento do domínio	79
7.2	Construção dos modelos de classificação	80
7.2.1	Categorização da Variável Alvo	80
7.2.2	Balanceamento dos Dados	81
7.3	Validação dos modelos de classificação	82
7.4	Gráficos de Comparação dos Resultados da Validação	85
8	CONCLUSÃO	89
8.1	Dificuldades e Limitações Encontradas	89
8.2	Contribuições	90
8.3	Trabalhos Futuros	90
	REFERÊNCIAS	91
	APÊNDICES	101
	APÊNDICE A – CONSULTAS SQL	103
	APÊNDICE B – SCRIPTS PYTHON PARA ALGORITMOS DE CLASIFICAÇÃO	107
	APÊNDICE C – AJUSTES DOS MELHORES PARÂMETROS	113
	ANEXOS	119
	ANEXO A – QUESTIONÁRIO ILS	121

1 Introdução

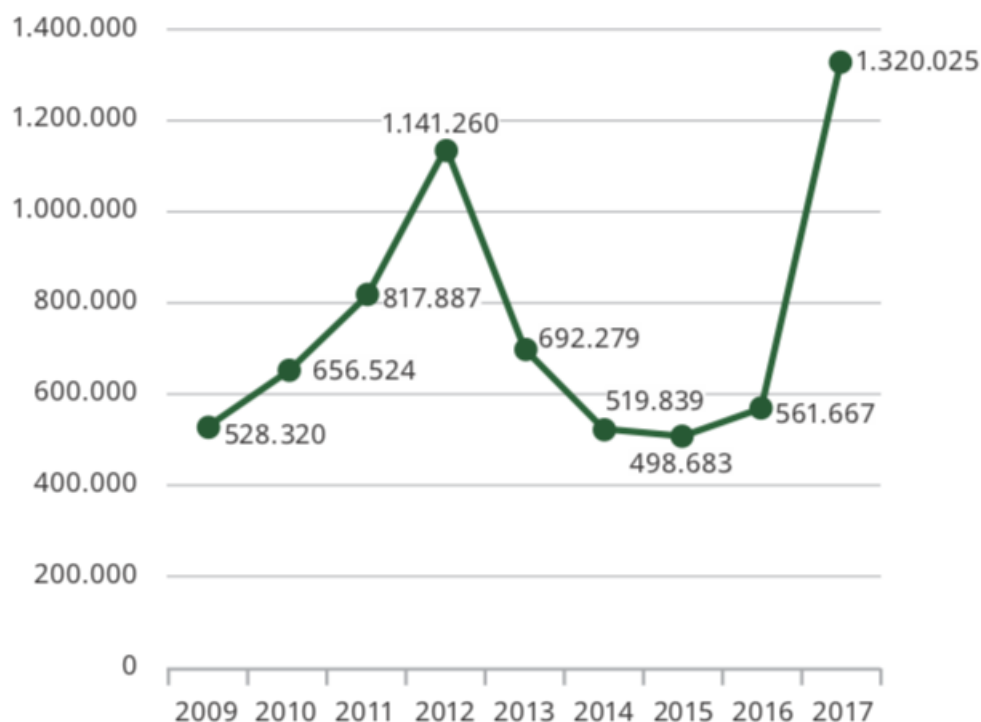
A disseminação e a evolução da tecnologia, que está cada vez mais presente em nosso meio, revolucionam a maneira como a educação é abordada. Esta revolução mudou a forma de aprender, tornando os indivíduos mais ativos na aprendizagem, buscando mais interatividade e participação (MACEDO, 2010). A modalidade de ensino de Educação a Distância (EaD) é uma prova de como a tecnologia afeta na educação, visto que a partir desta é possível abranger um maior número de pessoas interconectadas através da internet. O marco principal da EaD é a separação geográfica entre alunos e professores. Problemas que possam surgir devido a essa separação, provavelmente serão supridos pelas tecnologias (JOYE; YOUNG, 2012).

Essa modalidade pode ser classificada em dois tipos: integral ou híbrida, também conhecida como *Blended Learning*. No modo integral não existem encontros presenciais, ou seja, o curso é ofertado inteiramente a distância. Já no modo híbrido, combinam-se práticas pedagógicas do ensino presencial e do ensino a distância. Nesse modo, os alunos passam a conhecer e dominar os assuntos a partir de aulas online, já nas aulas presenciais terão aulas práticas, trabalhos em equipes, pesquisas, avaliações, entre outras atividades. Destaca-se como o principal benefício desta modalidade de ensino, a flexibilidade, já que o aluno poderá adaptar seu tempo de estudo de acordo com suas necessidades. Outras vantagens encontradas são: maior abrangência geográfica e o baixo custo, visto que diminui a demanda de infraestrutura física para as aulas.

Os recentes avanços tecnológicos têm propiciado um aumento no uso da EaD, que está apoiada por diversas tecnologias (GRANITO, 2008). Cada vez mais as instituições adotam essa abordagem de ensino. O crescimento no número de matrículas nesta modalidade comprova essa realidade.

Analisando o gráfico (FIGURA 1), elaborado pela Associação Brasileira de Educação a Distância (ABED), observa-se que os cursos regulamentados totalmente a distância contabilizaram 1.320.025 alunos. A série histórica revela uma ascensão rápida do número de alunos em cursos dessa modalidade entre os anos 2009 e 2012, quando foi criada a maioria das instituições formadoras em EaD, seguida de uma queda entre os anos de 2013 e 2015 e, desde então, uma nova ascensão, extremamente rápida, em 2017, com a flexibilização legal da oferta (ABED, 2018).

Figura 1 – Evolução do total de matrículas em cursos regulamentados totalmente a distância



Fonte: (ABED, 2018)

Segundo a ABED, as inovações contínuas em tecnologias e em abordagens pedagógicas estão entre as maiores dificuldades apontadas pelas instituições de ensino a distância no ano de 2016 (ABED, 2016). O uso de metodologias de ensino tradicionais na EaD pode afetar o rendimento acadêmico dos alunos, pois estas já não atendem às necessidades da era digital.

O ensino a distância requer um melhor planejamento e uma atualização dos métodos pedagógicos, pois a simples reprodução e disseminação do conteúdo aniquilam a qualidade dos cursos EaD, podendo causar insatisfação e evasão dos alunos (GRANITO, 2008). Logo, é necessário o uso de novos modelos didático-pedagógicos que busquem suprir as necessidades dos alunos, melhorando assim o aprendizado.

Outros desafios relacionados à EaD são as produções adequadas de conteúdos, a reformulação curricular e novos mecanismos de gerenciamento e avaliação (SALES; LEITE; JOYE, 2012). Essas mudanças precisam ser consideradas e implementadas para tentar resolver problemas maiores, como a evasão escolar.

Os recursos mais utilizados em cursos EaD são os Ambientes Virtuais de Aprendizagem (AVA), também conhecidos como sistemas de gerenciamento de aprendizagem. Esses

ambientes são *softwares* educativos que os usuários utilizam para interagir e construir conhecimentos (FILHO, 2014).

Um exemplo deste tipo de *software* é o Moodle (*Modular Object-Oriented Dynamic Learning Environment* - Ambiente de Aprendizagem Dinâmico Orientado a Objeto), uma plataforma *web* de ensino *open source*. Através desses ambientes virtuais é possível fazer o rastreamento de atividades, emitir relatórios, utilizar ferramentas síncronas e assíncronas, como chat, fórum, wikis, dentre outras que visam desenvolver a autonomia do aluno (SALES, 2010).

Uma das desvantagens desse tipo de ambiente é que a maioria aborda o conteúdo de maneira única para todos os alunos. Torna-se interessante que as formas de apresentação dos conteúdos dentro desses ambientes virtuais sejam adaptadas de acordo com as diferentes preferências do aluno, o que pode caracterizar diversos perfis. Dentre as técnicas que podem ajudar a detectar esses perfis, podemos citar a técnica de filtragem colaborativa, recomendação baseada em conteúdo e estilos de aprendizagem (EA) (FERREIRA, 2016).

Com relação aos Estilos de Aprendizagem, existem diversos autores na literatura que abordam diferentes tipos de modelos (KOLB, 1984) (DUNN; DUNN, 1974) (FELDER; SILVERMAN, 1988) (MYERS, 1962a). Cada um traz diferentes classificações e descrições dos perfis. Esses estilos são as preferências dos alunos na forma de perceber, captar, organizar, processar e compreender a informação (FELDER; SILVERMAN, 1988).

O uso de sistemas computacionais aliado aos conhecimentos educacionais, vem desempenhando um papel importante no ramo da EaD. Essa importância está principalmente relacionada no auxílio do controle à evasão e no aumento do desempenho acadêmico. Com isso, pesquisadores são motivados a investir tempo e esforços em pesquisas na área de aprendizagem de máquina aplicada em contextos educacionais (RAMOS, 2016) (RODRIGUES, 2016) (ROLIM; FERREIRA; COSTA, 2016).

Dessa forma, através do presente trabalho serão extraídos e analisados os dados de um ambiente virtual de aprendizagem. Estes dados serão explorados por diferentes técnicas de aprendizado de máquina, relacionadas à classificação, onde ocorre uma rotulação de alguns dados, a partir das suas similaridades com outros já classificados, para detecção automática dos estilos de aprendizagem.

Uma contextualização e a exposição dos fatos que constroem o problema de pesquisa deste trabalho foram discutidas nesta seção. A seguir, explica-se a motivação para o desenvolvimento desta dissertação.

1.1 Motivação

A identificação dos estilos de aprendizagem dentro de um AVA pode auxiliar na construção de sistemas personalizados que mesclam os benefícios dos sistemas de aprendizagem com uma abordagem adaptada para cada estilo. Essas plataformas de *e-learning* personalizadas buscam melhorar a satisfação e a motivação dos alunos ao longo do curso (HALAWA; SHEHAB; HAMED, 2015). Além disso, através dessa identificação o professor/tutor poderá fazer uso de metodologias de aprendizagem diferenciadas, assim como poderá melhorar suas formas de avaliação com base no perfil do aluno.

Em relação as formas de avaliação dentro de um AVA, o Modelo *Learning Vectors* (Modelo LV), proposto por Sales (SALES, 2010), se destaca por ser um modelo de suporte e gerenciamento à avaliação formativa concebido para automatizar o acompanhamento qualitativo/quantitativo de alunos. Este modelo atua sobre a interação entre os membros do grupo e suas ações sobre as ferramentas do ambiente de aprendizagem, como fórum, tarefa, *chat* e *wiki*.

Uma métrica pedagógica denominada Fator β , que se relaciona com a natureza qualitativa do modelo de avaliação LV, considera o desempenho do aluno em cada atividade. Este fator considera a razão entre resultados positivos e negativos das contribuições dos alunos nas atividades colaborativas ou individualmente desenvolvidas (SALES, 2010).

Apesar de possuir uma abordagem de avaliação formativa, aplicada no decorrer do processo de aprendizagem, o Modelo LV não leva em consideração os estilos de aprendizagem. O conhecimento desses estilos dos alunos pode ser um importante recurso em uma avaliação formativa que visa auxiliar a aprendizagem, ao alertar os professores da necessidade de planejar atividades de aprendizagem diversificadas (TREVELIN; NEIVA, 2011).

Métodos pedagógicos relacionados com a categorização dos estilos de aprendizagem, contribuem no processo de aprendizado do aluno melhorando o desempenho destes (HAIDER M.; CHAUDHARY, 2010), além de favorecer o desenvolvimento de materiais educacionais digitais mais adequados para cada perfil (SILVA, 2017). Logo, neste trabalho é utilizado um ambiente virtual de aprendizagem que implementa o modelo de avaliação LV de forma a extrair novos atributos, gerados a partir dele, que possam auxiliar na classificação dos estilos.

Dados gerados dentro dos ambientes virtuais de aprendizagem podem ser usados para uma classificação automática dos perfis dos alunos. As interações são coletadas a partir dos registros, no banco de dados, que armazenam informações dos usuários e do seu comportamento dentro do curso. Uma abordagem de detecção automática dos estilos de aprendizagem pode proporcionar melhores resultados no que se refere ao aproveitamento intelectual dos alunos (DORÇA, 2012) (FERREIRA et al., 2017) (GRAF,

2007) (SCHEUER; MCLAREN, 2012).

Assim, é possível utilizar essas informações para identificar os perfis de cada aluno e possivelmente propor adaptações no ambiente virtual de aprendizagem conforme perfil de cada um. Portanto, uma das motivações desta pesquisa é utilizar técnicas de classificação para compreender os comportamentos implícitos dos alunos, utilizando atributos do Modelo LV, ainda não avaliados para este propósito, de forma a propor uma outra abordagem que auxilie na identificação dos estilos de aprendizagem.

A motivação para esta pesquisa, que abrange alguns desafios da EaD, tais como, desempenho e interesse do aluno e necessidade de inovação tecnológica constante, foi discutida nessa seção. A seção a seguir trata da delimitação do problema de pesquisa e suas questões norteadoras.

1.2 Descrição do Problema

Gradativamente as instituições de ensino têm adotado o uso de cursos a distância. Esses cursos são gerenciados através de um AVA (GRAF, 2007). Seja dentro ou fora desses ambientes, uma difícil tarefa encontrada pelo professor é conseguir adaptar os métodos de aprendizagem para diferentes tipos de alunos (FELDER; BRENT, 2005). A falta de encontros presenciais provavelmente irá dificultar ainda mais a percepção de alguns aspectos comportamentais dos alunos.

Tanto em cursos presenciais ou em cursos a distância, essa dificuldade na adaptação dos métodos de ensino, pode ser um desafio. Em cursos a distância, porém, esse problema é maior devido ao alto número de alunos dentro das turmas, fazendo com que fique inviável o uso de abordagens de ensino diferenciadas de acordo com a necessidade de cada um.

Outro obstáculo relacionado a esse tipo de ambiente virtual é que eles tendem a ser centrados no curso ao invés do aluno (GRAF; LIST, 2005) e possuem uma limitação quanto ao uso de uma abordagem única para o ensino, conhecido como "*one-size-fits-all*", onde os mesmos recursos são apresentados de forma igual para todos (FELDER; BRENT, 2005). Esse tipo de abordagem pode não favorecer no desempenho dos alunos, visto que cada um tem preferências diferenciadas.

Abordagens de aulas mais diversificadas e equilibradas que tentam acomodar as diferentes necessidades dos alunos é o mais desejável que o professor pode fazer (FELDER; BRENT, 2005). Sabendo que cada aluno é diferente um do outro, ou seja, cada um tem seus pontos fortes e fracos, interesses, níveis de motivação e abordagens para estudar (HENZE; DOLOG; NEJDL, 2004), são necessárias novas mudanças.

A entrega de conteúdos adaptáveis dentro do AVA, no qual cada aluno terá acesso ao ambiente de acordo com o seu perfil, assim como métodos avaliativos diferenciados, são exemplos de melhorias que podem ser feitas através da identificação dos estilos de aprendizagem.

Estes estilos são definidos como as preferências dos alunos na forma de perceber, captar, organizar, processar e compreender a informação (FELDER; SILVERMAN, 1988). Recorrer aos estilos de aprendizagem demonstra uma preocupação em traçar perfis, melhorar práticas e oportunizar ações pedagógicas que estejam conectadas às preferências e a realidade discente (OTA; JÚNIOR; BARROS, 2017).

Os questionários são muitas vezes utilizadas para identificar os perfis dos alunos (FASIHUDDIN; SKINNER; ATHAUDA, 2016) (GRAF; KINSHUK; IVES, 2010) (NAKIĆ; GRAF; GRANIĆ, 2013), porém este método pode ser impreciso e demorado, pois depende diretamente da motivação dos alunos em responder, além do autoconhecimento a respeito de suas preferências (FERREIRA, 2016)

Assim, propostas utilizando técnicas de aprendizado de máquina alinhadas com teorias educacionais podem ajudar a classificar estilos de aprendizagem, de forma mais rápida em ambientes virtuais, levando em consideração os atributos gerados pelas interações dos alunos e os atributos gerados pelo Modelo LV.

Portanto, esse estudo remete à seguinte questão central de pesquisa a ser abordada:

Os indicadores quali-quantitativos gerados através do Modelo LV contribuem na identificação dos estilos de aprendizagem?

As três questões a seguir são norteadoras para a questão central dessa dissertação e são respondidas ao longo de sua execução:

- Quais os principais registros a serem analisados que podem ser extraídos da base de dados do ambiente virtual de aprendizagem que melhor classificam os estilos de aprendizagem dos alunos?
- Quais técnicas de aprendizado de máquina apresentam melhor desempenho em classificar os estilos de aprendizagem dos alunos?
- Qual a influência dos atributos do Modelo LV nos resultados da classificação dos estilos de aprendizagem dos alunos?

Nesta seção, apresentou-se a questão central e as questões norteadoras. Os objetivos geral e específicos são abordados a seguir.

1.3 Objetivos Geral e Específicos

Este trabalho tem como objetivo geral realizar um estudo comparativo referente à modelagem preditiva dos estilos de aprendizagem dos alunos através de uma abordagem

com base nos atributos da literatura, gerados pelas interações dos alunos, e nos atributos gerados pelo modelo de avaliação LV dentro de um ambiente virtual de aprendizagem. Como consequência, espera-se que esta proposta, baseada no Modelo LV, represente uma alternativa viável aos métodos de referência clássicos.

A seguir são apresentados os objetivos específicos que foram identificados e considerados como pré-requisitos para realização da pesquisa, conforme segue:

- Mapear fatores da literatura sobre estilos de aprendizagem com as variáveis comportamentais presentes no AVA;
- Construir as bases de dados utilizando os registros das variáveis comportamentais, baseadas na literatura e no Modelo LV, extraídas do AVA;
- Aplicar um questionário para os alunos de forma a validar a classificação feita pelos algoritmos;
- Construir uma modelagem capaz de prever o perfil de aprendizagem do aluno de acordo com as variáveis da literatura e do Modelo LV, extraídas do AVA;
- Comparar o desempenho dos algoritmos de classificação nas duas abordagens aplicadas, avaliando a eficácia dos atributos do Modelo LV na classificação dos estilos de aprendizagem.

A seguir uma descrição de trabalhos científicos aceitos e publicados durante o período da pesquisa.

1.4 Produção Científica

Durante o desenvolvimento desta dissertação, os seguintes artigos científicos foram aceitos e publicados, a saber:

- OLIVEIRA, E. S., SALES, G. L., PEREIRA, P. S., MOREIRA, R. N. Identificação Automática de Estilos de Aprendizagem: Uma Revisão Sistemática da Literatura. In: 26 Workshop sobre Educação em Computação (WEI) em XXXVIII Congresso da Sociedade Brasileira de Computação (CSBC), 2018, p. 102-114, Natal, Brasil.
- OLIVEIRA, E. S., SALES, G. L., SOUZA JUNIOR, A. H., MOREIRA, R. N. Identificação do Estilo de Aprendizagem utilizando o Modelo LV como auxílio para personalização de Sistemas Tutores Inteligentes. In: XXIX Simpósio Brasileiro de Informática na Educação (SBIE) em Congresso Brasileiro de Informática na Educação (CBIE), 2018, p. 1906-1910, Fortaleza, Brasil.

A seguir descreve-se a estrutura da dissertação, fazendo uma breve descrição dos conteúdos abordados nos capítulos subsequentes.

1.5 Estrutura da Dissertação

Este trabalho está dividido em oito capítulos. No Capítulo 2 apresenta-se a fundamentação teórica, que dá suporte para esta pesquisa. Neste capítulo, abordam-se alguns modelos de estilos de aprendizagem existentes na literatura.

No Capítulo 3 é apresentado o modelo de avaliação formativa, denominado: Modelo *Learning Vectors* (LV). Neste capítulo, expõem-se os componentes deste modelo, assim como a avaliação utilizada por ele e os trabalhos relacionados com o Modelo LV.

No Capítulo 4, são abordados os conceitos de reconhecimento de padrões e os classificadores de padrões utilizados: Máquina de Vetores de Suporte ou *Support Vector Machine* (SVM), Árvores de Decisão ou *Decision Tree* (DT), *Naive Bayes* (NB), *Perceptron* Multicamadas ou *Multilayer Perceptron* (MLP) e K-Vizinhos mais próximos ou *k-Nearest Neighbors* (KNN).

No Capítulo 5 descreve-se uma Revisão Sistemática da Literatura, acerca dos trabalhos relacionados sobre as estratégias e metodologias utilizadas para classificação automática de perfis de aprendizagem, assim como os modelos de estilos mais comuns.

No Capítulo 6 aborda-se a metodologia desta pesquisa, identificando cada uma das etapas seguidas, como, a caracterização do problema, o entendimento sobre os dados, os participantes envolvidos e o processo de classificação dos dados, de forma a alcançar todos os objetivos propostos.

No Capítulo 7 são descritos os resultados alcançados e as discussões acerca da comparação das abordagens implementadas. Por último no Capítulo 8 discorre-se sobre as considerações finais, limitações encontradas, contribuições e sugestões de trabalhos futuros.

2 Estilos de Aprendizagem

Neste capítulo inicia-se a seção referente à fundamentação teórica, na qual são apresentados os conceitos fundamentais dos temas de pesquisa abordados neste trabalho. A ênfase deste capítulo é resumir brevemente os principais modelos utilizados na literatura, para a identificação de estilos de aprendizagem.

Na Seção 2.1, discute-se o modelo de Felder e Silverman. Na seção 2.2 apresenta-se o modelo de Myers-Briggs, já na seção 2.3 apresenta-se o modelo de Grasha-Riechmann. Em seguida, o modelo de Honey-Mumford será discutido na seção 2.4. Na seção 2.5, apresenta-se o modelo VARK. Na seção 2.6, aborda-se o modelo de Kolb e por último, na seção 2.7, apresenta-se o modelo de Dunn.

O tema sobre estilos de aprendizagem é complexo e afetado por vários aspectos, com diferentes conceitos e pontos de vista, porém todos os conceitos concordam que os alunos têm diferentes maneiras nas quais preferem aprender (GRAF; KINSHUK; LIU, 2008).

Recentes trabalhos foram desenvolvidos utilizando o conceito de estilos de aprendizagem em diferentes abordagens, como jogos adaptativos, gamificação, predição de evasão e aprendizagem adaptativa (BUCKLEY; DOYLE, 2017), (HEIDRICH et al., 2018), (BALASUBRAMANIAN; ANOUNCIA, 2018), (SOFLANO; CONNOLLY; HAINEY, 2015). Não existe ainda uma definição única sobre este tema. Autores de diversos modelos de estilos de aprendizagem existentes na literatura, propõem diferentes descrições e classificações de tipos de aprendizagem.

Honey e Mumford definem os estilos de aprendizagem como uma descrição das atitudes e comportamentos que determinam o modo de aprendizagem preferido de um indivíduo (HONEY; MUMFORD, 1982). Já Felder e Silverman definem como pontos fortes e preferências na forma como os alunos aprendem e processam as informações (FELDER; SILVERMAN, 1988).

Nas próximas seções são apresentados vários modelos de estilo de aprendizagem comumente utilizados.

2.1 O Modelo de Felder e Silverman

No Modelo de Estilo de Aprendizagem de Felder e Silverman ou *Felder and Silverman Learning Styles Model* (FSLSM) (FELDER; SILVERMAN, 1988), os alunos são categorizados em quatro dimensões (TABELA 1). Estas dimensões mostram como os alunos preferem processar, perceber, captar e organizar a informação (GRAF; KINSHUK; LIU, 2009):

Tabela 1 – Dimensões do Modelo FSLSM

Dimensão	Categorias	
Processamento	Ativo	Reflexivo
Percepção	Sensorial	Intuitivo
Captação	Visual	Verbal
Organização	Sequencial	Global

Fonte: O Autor

- Ativo/Reflexivo: determina como os indivíduos preferem processar as informações. Alunos ativos preferem aprender fazendo, ou seja, valorizam atividades práticas. Gostam de realizar trabalhos em equipe, experimentações e discussões. Já o tipo reflexivo prefere trabalhos individuais, conceitos teóricos e aprender refletindo sobre algo.
- Sensorial/Intuitivo: refere-se a forma do indivíduo em perceber ou ter uma informação. Na dimensão de percepção o estilo sensorial prefere obter informações pelos seus sentidos e por meio de práticas (vendo, ouvindo, tocando), assim como prefere trabalhar com materiais concretos, fatos, dados experimentais e procedimentos padrões. O perfil intuitivo, no entanto, prefere abstrações, teorias, definições, inovações, desafios e criatividade.
- Visual/Verbal: está relacionada à forma de como os indivíduos preferem que as informações sejam apresentadas à eles. No modelo FSLSM, ela pode ser visual ou verbal. Perfis visuais preferem captar a informação através de imagens, gráficos, vídeos, enquanto perfis verbais preferem textos e áudios.
- Sequencial/Global: por último, tem a dimensão de organização, que indica como o indivíduo prefere progredir, podendo ser em sequência ou globalmente. Alunos sequenciais tendem a progredir de forma lógica e linear, organizando a informação a partir do detalhamento. No perfil global, os alunos preferem ter a visão geral do todo, aprendendo em saltos e com mais liberdade.

Esse modelo tornou-se um dos principais modelos de estilos de aprendizagem adotado em técnicas de classificação de aprendizes, como pode ser observado nos trabalhos relacionados. Escolheu-se este modelo para ser aplicado nesta pesquisa. O seu uso frequente

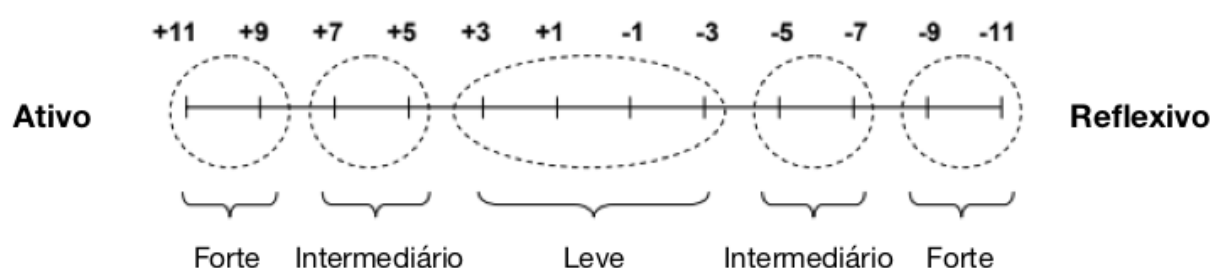
deve-se ao fato dele ter sido baseado em extensa experimentação, que avaliou e validou as dimensões em uma população de estudantes de engenharia (DORÇA, 2012).

Para detectar esses perfis, Felder e Soloman, desenvolveram o questionário *Index Learning Styles* (ILS)¹ que contém 44 questões objetivas, onde para cada dimensão existem 11 questões equivalentes, cada uma com duas opções de escolha (FELDER; SOLOMAN, 1997). Neste modelo os estilos não são convergentes, e sim pólos opostos que constituem uma dimensão. A identificação ocorre por meio de uma escala (FIGURA 2) que possui pesos diferentes.

O aluno é classificado nas quatro dimensões, onde em cada dimensão o mesmo pode ter preferência forte (9 ou 11), intermediária (5 ou 7) ou leve/neutra (1 ou 3) em um dos dois pólos da escala. Se a preferência do aluno for leve/neutra, ele se adapta bem em ambientes que abordam ambos os perfis, já com a preferência moderada ele aprenderá mais fácil se o ambiente de aprendizagem favorecer a sua dimensão. Com uma preferência forte para um estilo, o aluno pode apresentar dificuldades de aprendizagem se o ambiente não for favorável ao seu perfil.

Por exemplo: um aluno identificado com escala +11 na dimensão Ativo/Reflexivo, tem uma tendência a aprender mais facilmente trabalhando em equipe e com atividades práticas, enquanto um aluno com índice -9 nessa mesma dimensão, prefere aprender de maneira individual e com atividades mais teóricas. Embora o estudo tenha sido focado em estudantes de engenharia, o modelo pode ser útil para entender os vários estilos de aprendizagem.

Figura 2 – Escala do Estilo de Aprendizagem



Fonte: Adaptado de (GRAF, 2007)

Estes valores podem variar com o tempo, influenciados por estratégias vivenciadas, não podendo ser considerados características estáticas. Além disso, o FSLSM considera os estilos de aprendizagem como tendências, o que significa que os estudantes tendem a um estilo de aprendizagem específico, mas pode atuar em algumas situações de forma diferente (GRAF, 2007). Assim, os estilos de aprendizagem podem servir como indicadores de ações

¹ <https://www.engr.ncsu.edu/ils/>

a serem aplicadas durante o processo de aprendizagem de acordo com o perfil, assim como um indicador de pontos fortes e tendências que podem conduzir para dificuldades na aprendizagem (DORÇA, 2012).

2.2 O Modelo de Myers-Briggs

O modelo de Estilo de Aprendizagem criado por Myers e Briggs (MYERS, 1962b) está fundamentado na teoria dos tipos psicológicos proposta pelo psicólogo (JUNG, 1923). A essência da teoria é de que variações aparentemente aleatórias no comportamento são na verdade bastante ordenadas e consistentes, devido às diferenças básicas nas maneiras pelas quais os indivíduos preferem usar sua percepção e julgamento (MYERS-BRIGGS-FOUNDATION, 2018).

Com base nessa teoria, Myers e Briggs criaram o indicador do tipo de personalidade, chamado *Myers-Briggs Type Indicator* (MBTI) (MYERS, 1962b). Esse indicador avalia os tipos de personalidades e preferências através de quatro escalas com dois pólos opostos cada.

A primeira escala diz respeito aonde o indivíduo coloca mais sua atenção e obtém sua energia, podendo ser através de Extroversão ou Introversão (HALAWA; SHEHAB; HAMED, 2015). A segunda escala trata a maneira como o indivíduo assimila a informação e a percepção individual dos eventos circundantes, no qual pode ser do tipo Sensorial ou Intuitivo. A terceira escala trata a forma de como o indivíduo toma uma decisão, de forma Racional ou Sentimental.

Por último, a escala de como o indivíduo lida com o mundo externo. Algumas pessoas preferem coletar apenas dados suficientes para tomar uma decisão, outros estão sintonizados com as mudanças da situação, o que pode exigir uma mudança de estratégia ou mesmo mudar a meta (O'BRIEN; BERNOLD; AKROYD, 1998). Essas duas orientações finais são as atitudes de Julgadores ou Perceptivos.

- Extrovertido ou Introvertido: o perfil extrovertido prefere interagir com grupos de pessoas, prefere fazer ao invés de planejar, gosta de se envolver em eventos e ter várias atividades diferentes. Já o perfil introvertido prefere trabalhar com ideias e planejamentos de forma individual ou com poucas pessoas, refletindo mais do que agindo.
- Sensorial ou Intuitivo: o tipo sensorial usa os sentidos (vendo, ouvindo, tocando) para assimilar as informações. Gosta de ver o uso prático das coisas e prefere fatos e experiências do que definições e teorias. Já o tipo intuitivo prefere trabalhar com teorias abstratas e prefere aprender pensando sobre um problema do que com experiências práticas.
- Racional ou Sentimental: o perfil racional toma decisões de forma impessoal,

analisando os prós e os contras, sendo consistente e lógico nas decisões. Já o perfil sentimental prefere tomar decisões pesando o ponto de vista das pessoas envolvidas, se preocupa com valores e trabalha de forma mais diplomática.

- Julgadores ou Perceptivos: o perfil julgador trabalha de forma mais planejada e organizada, gosta de manter o controle do que precisa fazer. O perfil perceptivo, no entanto, trabalha de forma mais flexível e espontânea, ficando mais aberto à novas adaptações.

O indicador MBTI funciona como um instrumento para determinar os índices dos alunos em cada uma dessas dimensões. Esse indicador utiliza um questionário que possui 72 questões com respostas de sim ou não. A soma das respostas indica as preferências de aprendizagem de um aluno de acordo com o modelo proposto (FERREIRA, 2016).

2.3 O Modelo de Grasha-Riechmann

O modelo de Grasha-Riechmann (RIECHMANN; GRASHA, 1974) define os estilos de aprendizagem como diferentes formas que os alunos interagem com os colegas, com os professores e com os materiais de aula. Ou seja, trata dos relacionamentos interpessoais de cada indivíduo. Este modelo possui três dimensões bipolares, totalizando dessa forma seis estilos: participativo/inibido, competitivo/colaborativo, independente/dependente.

A primeira dimensão aborda a forma como o aluno se envolve na aula e suas atitudes no processo de aprendizagem. A segunda dimensão mede a interação do indivíduo com os outros. A terceira dimensão diz respeito à atitude do indivíduo na relação aluno-professor, identificando se ele prefere mais liberdade ou mais controle no ambiente de aprendizagem (EFRATI; LIMONGELLI; SCIARRONE, 2014). Cada uma das três dimensões possui uma escala de valores, onde o aluno tende para um determinado perfil, podendo essa tendência ser baixa, média ou alta (MEHDINEZHAD; AZARKHORDAD, 2016):

- Participativo/Inibido: o perfil participativo prefere leitura com discussões sobre o assunto e gosta de participar de atividades em grupos. Já o perfil inibido não gosta de participar de atividades com discussões e tende a ser anônimo.
- Competitivo/Colaborativo: indivíduos competitivos preferem que as atividades da aula sejam lideradas pelo professor. Gostam de se destacar e receber recompensas, trabalhando de forma individual. No entanto, perfis colaborativos gostam de compartilhar opiniões, conhecimentos e trabalhar em grupos.
- Independente/Dependente: o perfil independente tende a determinar seus objetivos e processos de aprendizagem sozinho, possui desejo de autonomia e pensa de forma independente. O perfil dependente prefere ter o professor como autoridade no processo de aprendizagem, gosta de receber instruções claras e possui pouca percepção de sua autonomia educacional.

Para medir esses estilos inicialmente utilizou-se um instrumento de 90 itens, chamado de *Grasha-Riechmann Student Learning Style Scales* (GRSLSS) (RIECHMANN; GRASHA, 1974). Esse instrumento divide 15 itens por estilo (participativo, inibido, competitivo, colaborativo, independente e dependente). Outra versão do GRSLSS foi criada por Grasha e Riechmann com apenas 60 itens (FERRARI et al., 1996). Para cada um dos seis estilos, existem 10 palavras que são somadas e divididas por 10, e a pontuação de cada estilo é obtida. A resposta está em 5 graus da escala Likert (MEHDINEZHAD; AZARKHORDAD, 2016).

2.4 O Modelo de Honey-Mumford

O modelo desenvolvido por Honey e Mumford (HONEY; MUMFORD, 1982) possui quatro estilos: reflexivo, teórico, pragmático e ativista. Este modelo é derivado diretamente da teoria de Kolb (KOLB, 1984). Para melhorar e aumentar o próprio aprendizado pessoal, cada aluno deve entender o seu estilo de aprendizagem e procurar oportunidades para aprender usando esse estilo. Os quatro estilos são listados a seguir (MOBBS, 2018):

- **Ativista:** são indivíduos que aprendem fazendo, se envolvendo completamente em novas experiências. Eles tendem a agir primeiro e refletir sobre a ação depois. Gostam de centralizar as atividades em torno de si. Possuem uma abordagem de aprendizagem aberta, onde se sentem entusiasmados com coisas novas. Gostam de atividades que envolvam debates, discussões em grupo, competição e solução de problemas.
- **Reflexivo:** indivíduos que aprendem mais refletindo e observando. Gostam de analisar os fatos de diferentes perspectivas e são cautelosos em tomar decisões. Dedicam tempo pensando antes de chegar em uma conclusão. Preferem atividades como, questionário de auto-análise, entrevistas e receber *feedbacks* de outros.
- **Pragmático:** indivíduos práticos, que são capazes de ver e entender a aplicação prática do que aprendem. Gostam de experimentar e testar novas ideias. Preferem atividades que envolvam estudo de caso, discussões e resoluções de problemas. Tendem a ser impacientes com discussões abertas.
- **Teórico:** gostam de entender o conceito por trás das ações e analisar novas informações. Preferem trabalhar com teorias, histórias, citações e estatísticas. Resolvem os problemas de maneira lógica, seguindo um passo-a-passo. Tendem a ser perfeccionistas e valorizam a racionalidade.

Para entender os estilos de aprendizagem, Honey e Mumford desenvolveram um Questionário de Estilos de Aprendizagem ou *Learning Styles Questionnaire* (LSQ) de auto avaliação que ajuda a identificar um dos quatro perfis proposto nesse modelo (HONEY; MUMFORD, 2000). Esse questionário possui 80 questões claras e relevantes onde o

indivíduo responde de forma positiva ou negativa cada pergunta.

2.5 O Modelo de VARK

Baseado nos modelos de programação neuro-linguística, Neil D. Fleming projetou o modelo VARK, e o lançou em 1987 (FLEMING; MILLS, 1992). O modelo VARK (Visual, Auditivo, Leitura/Escrita e Sinestésico) categoriza quatro modalidades de preferência sensorial sendo uma categoria do tipo multimodal (MOAZENI; POURMOHAMMADI, 2013).

- Visual: indivíduos que processam melhor as informações vendo e preferem receber informações das representações figurativamente, por exemplo, gráficos, fluxogramas, imagens e notações simbólicas.
- Auditivo: gostam de ouvir ou falar sobre as informações com outras pessoas, compreendem melhor ouvindo palestras e tutoriais.
- Leitura/Escrita: alunos desse perfil preferem ver palavras escritas, ler textos e fazer anotações.
- Sinestésico: eles têm preferência em aprender algo que tenha uma conexão com a realidade e adquirem mais conhecimento através da experiência simulada ou real.

Os multimodais são aqueles que têm preferências em mais de um modo. Para identificar o perfil do indivíduo utiliza-se o questionário VARK que contém 16 questões, na qual cada questão possui quatro alternativas e o respondente pode optar por marcar nenhuma, uma, duas, três ou quatro opções (FLEMING; MILLS, 2018). Os resultados não devem ser rigidamente aplicados, pois o questionário não tem a intenção de inserir os entrevistados em uma mentalidade de que foram "diagnosticados". Pelo contrário, ele é projetado para iniciar discussões e refletir sobre as preferências de aprendizagem. VARK está na categoria de instrução preferencial, pois lida com modos de percepção, centrando-se nas diferentes formas que se obtém e se repassam as informações (FLEMING, 2006).

2.6 O Modelo de Kolb

O modelo proposto por Kolb (KOLB, 1984) baseia-se na Teoria da Aprendizagem Experiencial, onde acredita-se na influência da experiência no processo de ensino-aprendizagem. Nesta teoria o processo de aprendizagem é concebido em um ciclo de quatro estágios de habilidades: experiência concreta, observação reflexiva, capacidade de conceituação abstrata e experimentação ativa (KOLB, 1981).

A experiência concreta é a base para realizar observações e reflexões. Essas observações são usadas para formar conceitos abstratos, que novamente atuam como base para

testar novas implementações através de experimentações. Testar implementações resulta em experiência concreta, fechando dessa forma o ciclo de aprendizado.

Para Kolb (KOLB, 1981) a maioria de nós desenvolve estilos de aprendizagem que enfatizam algumas habilidades de aprendizagem em detrimento de outras, como resultado de nossa hereditariedade, nossa experiência de vida e as necessidades do nosso ambiente atual. O modelo de Kolb considera duas dimensões que são pólos opostos: abstrato/concreto, ativo/reflexivo. Essas dimensões combinadas dão origem aos quatros estilos do modelo (FIGURA 3).

Figura 3 – Estilos de Aprendizagem de Kolb



Fonte: Adaptado de (KOLB, 1984)

O perfil acomodador possui como habilidades dominantes: experiência concreta e experimentação ativa. Seus pontos fortes concentram-se em realizar planos e experimentos, aplicar o conhecimento, praticar ações de forma ativa e se destacar em situações que exigem adaptação a circunstâncias imediatas. Já os assimiladores estão relacionados com a observação reflexiva e a capacidade de conceituação abstrata e portanto, são apostos aos acomodadores. Sua maior força reside na criação de modelos teóricos. Eles são bons no raciocínio indutivo e tendem a aprender por meio de observação e reflexão.

As habilidades dominantes dos convergentes são conceituação abstrata e experimentação ativa. Portanto, seus pontos fortes estão nas aplicações práticas das ideias, além de serem hábeis para resolver problemas e também nas tomadas de decisões. Já os indivíduos com maior experiência concreta e observação reflexiva são denominados divergentes. Eles são bons em ver situações concretas em muitas perspectivas diferentes, sendo assim mais criativos (KOLB, 1981).

O *Learning Styles Inventory* (LSI) é um questionário que foi desenvolvido para identificar os estilos de aprendizagem baseados no modelo de Kolb e está na versão 3.1 (KOLB; KOLB, 2005). Este questionário possui diversas versões. A primeira versão contém apenas nove itens e as demais contém doze, onde estes itens referem-se a sua maneira preferida de aprender.

Cada item possui quatro questões que devem ser classificadas atribuindo um grau de um a quatro para mensurar o quanto aquela questão descreve seu perfil de aprendizagem. Os resultados do LSI indicam as preferências dos indivíduos para os quatro modos, o que leva ao tipo preferido do Estilo de Aprendizagem de cada um (GRAF, 2007).

2.7 O Modelo de Dunn

O modelo de Dunn foi proposto em 1974 e depois aperfeiçoado e ampliado ao longo dos anos (DUNN; DUNN, 1974). Este modelo possui cinco variáveis: ambiental, sociológica, emocional, física e psicológica. Cada variável está relacionada aos estímulos que o indivíduo utiliza para aprender e quais elementos fazem parte desses estímulos.

A variável ambiental possui elementos como luz, som, temperatura. Já a variável sociológica trata da preferência do indivíduo em aprender sozinho, em dupla, em equipe, com uma autoridade ou de forma variada. A variável emocional consiste em fatores como motivação, responsabilidade e persistência.

A variável física é referente às preferências de percepção, como visual e auditiva, além dos fatores como ingestão de alimentos e horários do dia. A variável psicológica foi adicionada posteriormente ao modelo e inclui fatores referentes às preferências globais ou analíticas e preferências impulsivas ou reflexivas.

Para identificar o perfil do indivíduo, diversas versões de questionário foram desenvolvidas. O modelo do questionário se diferencia na aplicação para crianças e adultos. O questionário para crianças possui três versões e 104 questões, que empregam uma escala *Likert* de 3 ou 5 escolhas. A última versão para adultos, *Building Excellence Inventory*, possui 118 (GRAF, 2007) questões com escala de 5 escolhas. Como resultado, uma alta ou baixa preferência por cada fator é identificada.

A seguir, o resumo dos modelos de estilos de aprendizagem descritos neste capítulo (TABELA 2).

Tabela 2 – Resumo dos Modelos de Estilos de Aprendizagem

Modelo/Autor	Estilos/Dimensões
Felder e Silvermann	Ativo/Reflexivo, Sensorial/Intuitivo, Visual/Verbal, Sequencial/Global
Myers-Briggs	Extrovertido/Introvertido, Sensorial/Intuitivo, Racional/Sentimental, Julgadores/Perceptivos
Grasha-Riechmann	Participativo/Inibido, Competitivo/Colaborativo, Independente/Dependente
Honey e Mumford	Ativista, Reflexivo, Pragmático, Teórico
Fleming e Mills	Visual, Auditivo, Leitura e Escrita, Sinestésico
Kolb	Acomodador, Divergente, Assimilador, Convergente
Dunn	Ambiental, Sociológica, Emocional, Física, Psicológica

Fonte: O Autor

Apresenta-se no próximo capítulo um modelo de avaliação não-linear, que utiliza novas métricas pedagógicas qualitativas e subjetivas, de forma a oferecer suporte à necessidade constante do aluno de acompanhar e verificar seu desempenho.

3 Modelo Learning Vectors (LV)

Este capítulo apresenta um modelo de avaliação utilizado em ambientes virtuais de aprendizagens em curso EaD, intitulado como: Modelo LV. Entre os assuntos a serem explorados estão: o significado e conceito do modelo, os componentes que o constituem e, por fim, as suas abordagens de avaliação e os trabalhos relacionados.

O Modelo Learning Vectors (LV), proposto por Sales (SALES, 2010), se destaca por ser um modelo de suporte e gerenciamento à avaliação formativa concebido para automatizar o acompanhamento do desempenho de alunos da EaD dentro de um AVA nas dimensões qualitativa e quantitativa. A avaliação através dessas duas dimensões se torna mais completa, pois estas se complementam como partes de um todo, que em comum, visam traçar o perfil e o grau de aprendizagem atingido pelo aluno.

A construção desse modelo foi baseada no sistema dinâmico não-linear do modelo psicomatemático *Meta Learning*, desenvolvido por Losada (LOSADA, 1999). O Modelo LV também referencia-se em teorias interacionistas sistêmicas e em aprendizagem regulada, agregadas por uma avaliação formativa (SALES; BARROSO; SOARES, 2012).

3.1 *Learning Vectors* e seus componentes

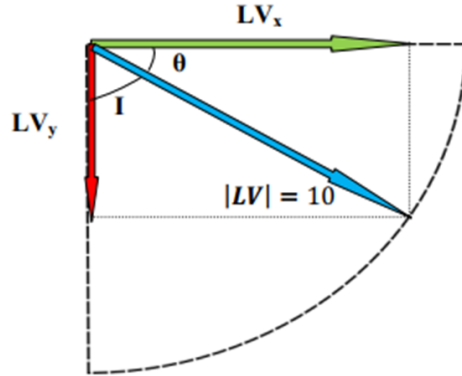
O LV atua sobre a interação entre os alunos e suas ações sobre os recursos presentes no ambiente de aprendizagem, como fórum, *chat*, quiz, tarefa e *wiki*. Este modelo também considera a frequência do aluno na participação dessas atividades. O desempenho dos alunos é representado pelos Vetores-Aprendizagem, que possuem módulo fixo de 10 unidades, com direções que iniciam no ângulo $\theta = -90^\circ$ com a horizontal (posição que representa a nota mínima), podendo ir até $\theta = 0^\circ$ (posição que representa a nota máxima), conforme se avaliam as atividades dos alunos ao longo do curso. O acompanhamento do desempenho no curso, feito pelo professor e pelo próprio aluno é realizado através de representações geométricas vetoriais (FIGURA 4).

O componente horizontal (LV_x), está relacionado com a positividade, onde o mesmo representa o desempenho em cada atividade e implica diretamente no rendimento e situação final do aluno, já a projeção vertical do LV (LV_y) relaciona-se à negatividade do seu desempenho (SALES; BARROSO; SOARES, 2012).

A equação 3.1 mostra como o valor do LV_x é obtido.

$$LVx = 10 * \cos[(-12\alpha + I)] \quad (3.1)$$

Figura 4 – Vetor-Aprendizagem



Fonte: (SALES, 2010)

A variável α representa o passo padrão de aprendizagem e possui valor de $7,5^\circ$, já a variável I contabiliza o parâmetro Variação Angular Total do vetor (Δ_T) (Equação 3.2). A direção dos LV varia discretamente de acordo com a variável I , dada em função do ângulo α e depende das avaliações atribuídas pelo professor às interações/atividades propostas ao aluno.

$$I = \frac{n}{2} * \alpha, \text{ com } \{n \in N | 0 \leq n \leq 24\} \quad (3.2)$$

Esses 25 possíveis valores discretos da variável I facilitaram a especificação dos vetores LV, já que estabelecem o limite de 90° para a variação, ficando assim o vetor LV limitado ao quarto quadrante (-90° a 0°). A seguir, a Equação 3.3 apresenta como é calculado o valor de LV_y , que representa a negatividade do desempenho do aluno.

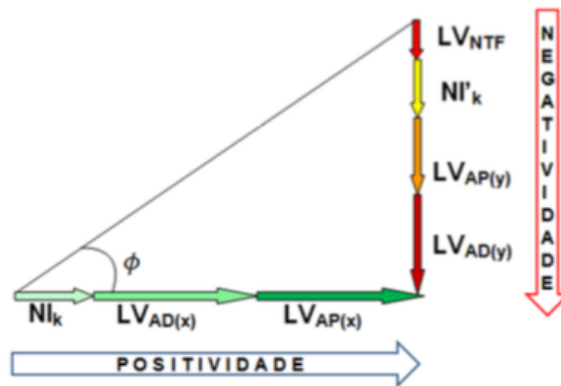
$$LV_y = 10 * \text{sen}[-12\alpha + I] \quad (3.3)$$

A razão entre os resultados positivos e negativos das contribuições dos alunos nas atividades colaborativas ou individualmente desenvolvidas, compõem uma métrica pedagógica não-linear denominada Fator β (Equação 3.4). Essa métrica se relaciona com a natureza qualitativa do modelo de avaliação LV, considerando o desempenho do aluno em cada atividade.

$$\beta = \frac{\text{Positividade}}{\text{Negatividade}} \quad (3.4)$$

As dimensões Positividade/Negatividade estão ilustradas a seguir (FIGURA 5):

Toma-se como contribuições positivas, o somatório das projeções horizontais dos LV fórum, tarefa, *chats*, *wikis*, sem qualquer ponderação, assim como o número de interações positivas. Já como contribuição negativa toma-se o somatório das projeções verticais dos

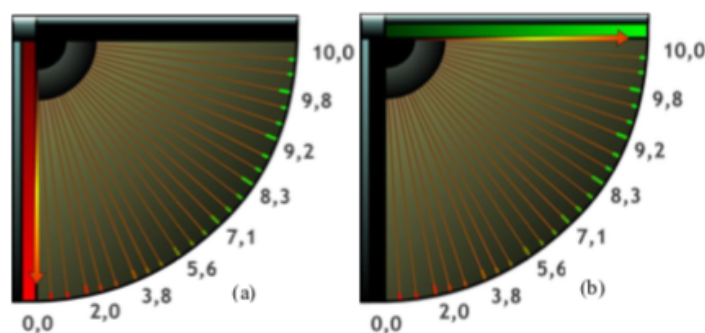
Figura 5 – Fator β - Positividade X Negatividade

Fonte: (SALES, 2010)

LV fórum, tarefa, *chats*, *wikis*, sem qualquer ponderação, assim como o número total de faltas e interações negativas.

O aluno também poderá acompanhar o seu desempenho consultando seus Vetores-Aprendizagem (FIGURA 6). A projeção vertical, representada pela faixa vermelha, relaciona-se com o fator negatividade. Já a projeção horizontal, representada pela faixa verde, relaciona-se com a positividade. O ponteiro laranja indica o nível que o aluno alcançou até o momento.

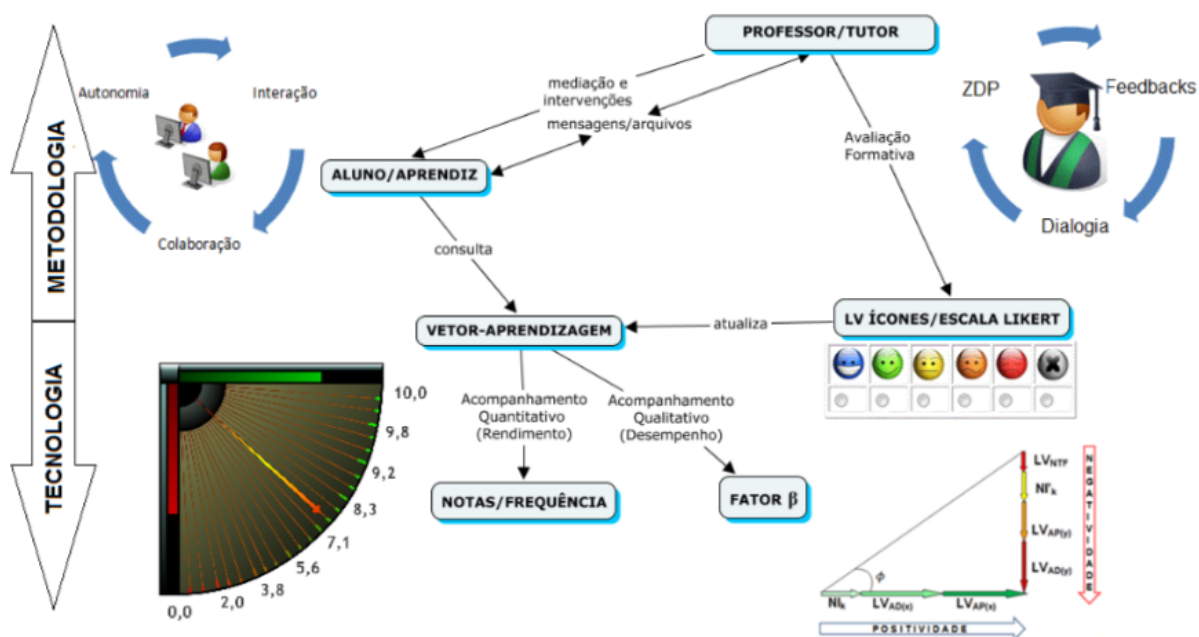
Figura 6 – Vetor-Aprendizagem: (a) Nota mínima e (b) Nota máxima



Fonte: (SALES, 2010)

Esse acompanhamento permite ao aluno realizar uma aprendizagem auto-regulada, observando em quais atividades ele precisa melhorar. Para o professor o modelo auxiliará na avaliação formativa, pois serve de indicador qualitativo e regulador da aprendizagem. De forma a se obter uma melhor compreensão do suporte à avaliação formativa proposta pelos LV, apresenta-se o modelo tecno-metodológico (FIGURA 7).

Figura 7 – Modelo tecno-metodológico LV



Fonte: (SALES, 2010)

Além de reduzir a carga de trabalho gerada para o professor na análise das informações e regulações da aprendizagem, esse modelo gera uma nova métrica pedagógica, que pode auxiliar na tomada de decisão do processo de avaliação e pode ajudar a identificar os perfis de aprendizagem dos alunos, além de indicar fatores de positividade e negatividade. Outra vantagem referente à este modelo deve-se ao fato dele influenciar a interação aluno-aluno, bem como a interação aluno-professor, diminuindo desta forma a distância transacional entre os mesmos.

3.2 Avaliação usando *Learning Vectors*

Para auxiliar o professor no momento da avaliação, uma escala de linguagem iconográfica, associada a menções qualitativas (Muito Bom, Bom, Regular, Fraco, Não Satisfatório e Neutro) é utilizada. Os LV Ícones (FIGURA 8), tais como os *emoticons*, bastante difundidos nas comunicações virtuais, podem ajudar a criar vínculos emotivos e sentimentais no processo de avaliação da aprendizagem. Recentemente foi realizada a atualização desses ícones para *gifs* ou ícones animados, buscando tornar a sala de aula virtual um espaço mais afetivo e motivador e a avaliação em um processo mais dinâmico (SALES et al., 2019).

Figura 8 – Ícones LV e menções qualitativas

	Muito Bom	Bom	Regular	Fraco	Não Satisfatório	Neutro
Menções Qualitativas						
Coeficiente do Passo (CP)	4	3	2	1	0	-

Fonte: (SALES; BARROSO; SOARES, 2012)

Coeficiente de Passo (CP) representa uma escala numérica associando valores diferentes para cada ícone LV, onde cada ícone avalia o nível de interação do aluno de forma subjetiva. Toda vez que o aluno faz o envio de alguma mensagem ou arquivo, utilizando-se das ferramentas de interação, aparece a Escala LV Ícone no AVA para o Professor/Tutor. Por outro lado, a interface permite que somente o aluno responsável pela mensagem ou arquivo enviado tenha acesso ao ícone atribuído pelo Professor/Tutor, o que garante sua privacidade (SALES; BARROSO; SOARES, 2012).

As categorias apresentadas relacionam-se ao grau com que o aluno apresenta suas reflexões frente ao conteúdo. O ícone Muito Bom é atribuído quando o aluno atinge reflexões profundas com questionamentos significativos ou síntese de ideias que geram novas interações. A categoria Bom indica que o aluno realizou boas reflexões e comentários significativos. Já a categoria Regular representa reflexões regulares que foram feitas de forma superficial sobre o tema proposto.

A avaliação com ícone Fraco condiz com reflexões vazias de conteúdo, que agregam pouco valor à discussão. A categoria Não Satisfatório representa respostas de alunos com posição passiva onde não contribuem para a discussão, limitando-se a concordâncias sem nenhuma relevância. A última categoria equivale ao ícone Neutro. Essa categoria corresponde às respostas que não se relacionam com o tema e consequentemente não agregam nota, porém permite computar a presença do aluno na atividade. Este ícone permite indicar interações não significativas para avaliação.

A implementação, testes e refinamentos do modelo foram feitos no ambiente virtual Moodle, que apresenta código-fonte aberto baseado em tecnologia *PHP: Hypertext Preprocessor* (PHP). Esta também foi a linguagem de programação utilizada na implementação dos LV. Este modelo foi avaliado e sua eficácia foi comprovada, através do *feedback* positivos de professores e alunos (SALES, 2010).

3.3 Trabalhos Relacionados ao Modelo LV

O Modelo LV foi utilizado para diversas aplicações em diferentes trabalhos, tanto na área da educação como na área corporativa e vem apresentando resultados satisfatórios. Por gerar dados referentes a avaliação das atividades de um ambiente virtual (fórum, *chat*, *wiki*, tarefa), esse modelo também pode ser útil para auxiliar na classificação de EA. Na pesquisa de Gonçalves ele foi aplicado ao fórum de perguntas e respostas do ambiente virtual *Moodle*, associado à metodologia ativa *Peer Instruction*, de forma a facilitar intervenções pedagógicas e motivação aos envolvidos (GONÇALVES; CUNHA; SALES, 2016) .

No trabalho de Cacaís e Sales, o modelo foi utilizado em um sistema colaborativo *gamificado* para avaliação de desempenho corporativo, denominado *Process Planning and Institutional Evaluation* (PIPA) (CACAIS; SALES, 2017). O Modelo LV auxiliou no acompanhamento da progressão dos profissionais, possibilitando um melhor *feedback* pelos seus supervisores.

Outro recente trabalho que abordou o Modelo LV foi o de Albuquerque. Nesta pesquisa foi proposto um modelo utilizando lógica *Fuzzy* para avaliar a presencialidade de alunos dentro de um fórum LV. O resultado do artigo mostrou que o modelo foi capaz de mensurar a presencialidade do aluno com perspectivas de ser aplicado a outras ferramentas de avaliação presentes nos AVAs (ALBUQUERQUE et al., 2017).

Já o trabalho de Pereira, teve como objetivo identificar grupos de alunos com o mesmo perfil de participação e interação em um ambiente de aprendizagem on-line, através de algoritmo de agrupamento hierárquico baseando-se no modelo de avaliação *Learning Vectors* (PEREIRA et al., 2018).

Seguindo a linha das pesquisas dos autores citados, este trabalho está fundamentado no modelo de estilos de aprendizagem de Felder e Silverman (FELDER; SILVERMAN, 1988) e no Modelo LV e seus Vetores-Aprendizagem (SALES, 2010). Portanto, esta pesquisa diferencia-se das demais ao utilizar o Modelo LV na identificação dos perfis dos alunos, aplicando-se diferentes técnicas de classificação de forma a comparar o desempenho de cada uma.

O capítulo a seguir apresenta e explica os classificadores que foram utilizados para identificação dos estilos de aprendizagem.

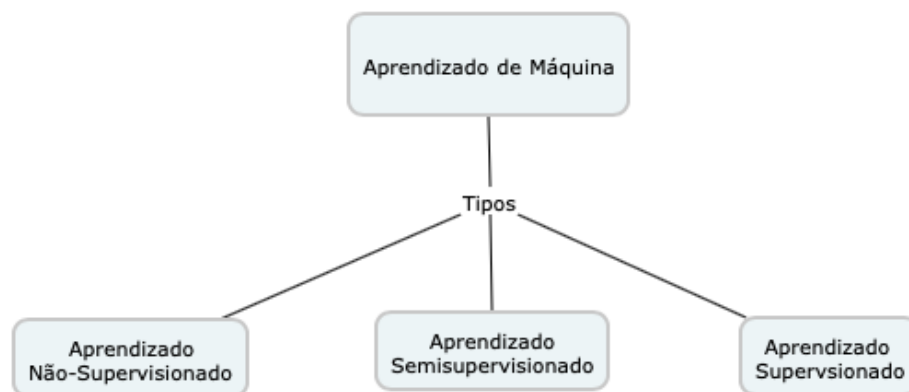
4 Reconhecimento de Padrões

A crescente demanda no uso de tecnologias nas mais diversas áreas, faz com que a exigência por automatização de processos repetitivos ou cansativos, envolvendo um certo grau de inteligência, se torne cada vez mais necessária. O aprendizado de máquina associa-se nesse campo de pesquisa e faz parte da grande área de Inteligência Artificial (IA) .

Segundo Coppin, Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos (COPPIN, 2010). A IA tem como objetivo principal prover uma máquina ou sistema com capacidade similar à dos seres humanos na realização de tarefas, ou seja, que pense, aja e tome decisões sólidas e racionais (RUSSEL STUART; NORVIG, 2013).

O aprendizado de máquina é um conceito relacionado à área de Inteligência Artificial cujo objetivo é desenvolver métodos computacionais capazes de tomar decisões com base em conhecimentos já existentes (MICHALSKI; CARBONELL; MITCHELL, 2014). Este conceito pode ser dividido em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semi-supervisionado (MONARD; BARANAUSKAS, 2003) (FIGURA 9). O reconhecimento de padrões é um subtópico desse campo de pesquisa e se enquadra na categoria de aprendizado supervisionado.

Figura 9 – Hierarquia do aprendizado de acordo com o grau de supervisão



Fonte: Adaptado de (MONARD; BARANAUSKAS, 2003)

O reconhecimento de padrões utiliza classificadores, que são elementos de software capazes de atribuir uma amostra, cuja classe é inicialmente desconhecida, a uma classe dentro de um conjunto de finitas possibilidades (NETO, 2006). Assim, o objetivo do aprendizado supervisionado é dirigir um mapeamento dos vetores x para os valores y , a fim de construir uma hipótese que permita prever valores de y para exemplos não rotulados, no qual x representa um vetor de atributos e y corresponde ao valor da classe (BATISTA; MONARD; CAROLINA, 2018). Portanto, neste capítulo, com o objetivo de facilitar a compreensão dos métodos propostos e analisar comparativamente o seu desempenho, são apresentados os classificadores para reconhecimento de padrões utilizados neste trabalho.

4.1 K Vizinhos Mais Próximos

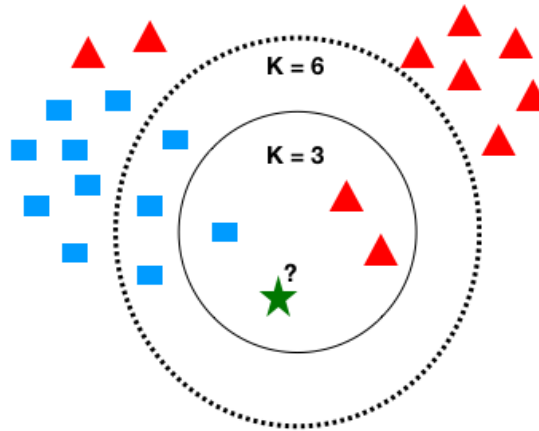
A técnica de classificação K Vizinhos Mais Próximos ou *K-Nearest Neighbor* (KNN) é um método baseado em instância, ou seja, considera-se que se dois exemplos do conjunto de dados são similares, eles então pertencem à mesma classe. Seja x um padrão a ser classificado, que não pertence ao conjunto de dados rotulados, a ideia do algoritmo dos k -vizinhos mais próximos consiste em atribuir a mesma classificação para o padrão x ao da classe mais representada nos k vizinhos mais próximos ou k ponto mais próximos (NETO, 2006).

Um exemplo desse algoritmo pode ser visto na FIGURA 10. Nesse exemplo a estrela verde representa o valor que precisa ser classificado. Os quadrados azuis representam a primeira classe e os triângulos vermelhos representam a segunda. Com valor de $k=3$, a classe escolhida será a segunda, já que os triângulos estão em maior quantidade. Com valor de $k=5$, porém, a classe escolhida será a primeira. Logo, observa-se a importância da escolha correta para o valor de k . O valor de k pode variar de uma base de dados para outra, assim é necessário a realização de testes com diferentes valores a fim de descobrir qual apresenta menor taxa de erros.

O algoritmo dos k -vizinhos mais próximos é facilmente adaptado para aproximar as funções objetivo de valor contínuo. Para conseguir isso, o algoritmo calcula o valor médio dos k exemplos de treinamento mais próximos, em vez de calcular seu valor mais comum (MITCHELL, 1997). Para calcular a distância entre os dois pontos são utilizadas algumas métricas mais comumente, entre elas: Euclidiana, Minkowsky, Mahalanobis, Hamming (SARMENTO, 2016). A técnica mais empregada no KNN é a distância Euclidiana, dada por:

$$d = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}, \quad (4.1)$$

Figura 10 – Exemplo de classificação com KNN



Fonte: Adaptado de (SARMENTO, 2016)

O KNN é um algoritmo simples que apresenta grande versatilidade. Suas aplicações centram-se principalmente sobre a classificação, ainda que também pode ser usado em regressão (SHAKHNAROVICH; DARRELL; INDYK, 2006). A principal motivação para utilização deste algoritmo está em que, padrões próximos no espaço de características possivelmente pertencem à mesma classe (NETO, 2006).

4.2 Classificador Bayesiano

A teoria da decisão bayesiana faz a suposição de que o problema de decisão é colocado em termos probabilísticos e que todos os valores de probabilidade relevantes são conhecidos (DUDA; HART; STORK, 2000). O classificador bayesiano é fundamentado no Teorema de Bayes (Equação 4.2). Neste teorema, calculam-se as probabilidades de um padrão pertencer a cada uma das possíveis classes, predizendo para aquele padrão, a classe mais provável, isto é, a classe que obteve a maior probabilidade a posteriori (THEODORIDIS; KOUTROUMBAS, 2009).

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)} \quad (4.2)$$

O classificador bayesiano, atribui, a um dado objeto, a classe que tem maior probabilidade de ser a verdadeira, dada as características desse objeto. Essa equação considera a relação entre uma probabilidade condicional e a sua inversa (DUDA; HART; STORK, 2000).

$P(A)$ e $P(B)$ são probabilidades a priori de A e B, respectivamente, ou seja, a probabilidade $P(A)$ não leva em consideração nenhuma informação a respeito de B e vice-versa. $P(B/A)$ e $P(A/B)$ são denominadas de probabilidades a posteriori de B condicional a A e de A condicional a B, respectivamente (FERREIRA, 2016).

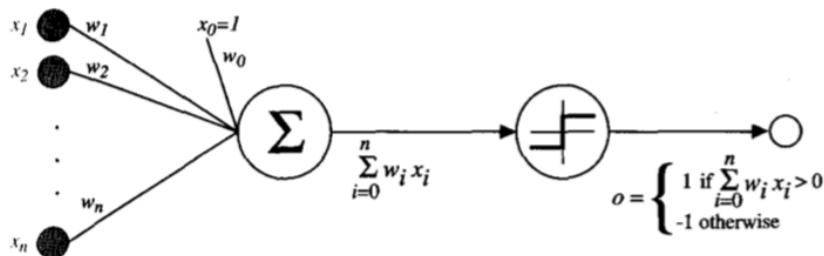
Como vantagem este classificador minimiza a probabilidade média de erro na classificação (DUDA; HART; STORK, 2000). No entanto, visto que a probabilidade estimada é baseada na frequência dos dados do treinamento, isto pode não ser confiável em regiões onde os dados de entrada se distribuem de forma esparsa (NETO, 2006).

4.3 Rede Perceptron de Multicamadas

Rede Perceptron de Multicamadas ou *Multilayer Perceptron* (MLP), é um exemplo de Rede Neural Artificial (RNA). Esse método se refere aos modelos computacionais baseados no sistema nervoso central dos animais. Por envolver elementos altamente interconectados este modelo se enquadra dentro do paradigma conexionista. A técnica da computação neural utiliza como modelo os procedimentos usados pelo cérebro humano na solução de problemas aplicando-os em sistemas computacionais (HAYKIN, 1998).

Uma rede *Perceptron* Simples (PS) é constituída de x sinais de entrada, associados a w pesos, uma função de ativação e uma saída (FIGURA 11). Em analogia com neurobiologia, esses pesos ou conexões são às vezes chamados de “sinapses” e o valor da conexão é “pesos sinápticos” (DUDA; HART; STORK, 2000). A simplicidade deste modelo está associado à sua condição de ser constituído de somente uma camada neural formada por somente um único neurônio.

Figura 11 – Rede Perceptron Simples

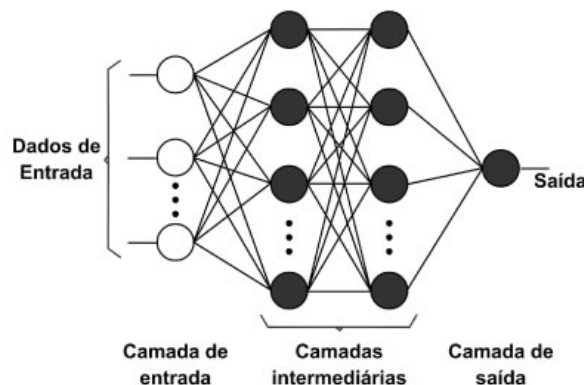


Fonte: (MITCHELL, 1997)

Com o objetivo de superar algumas limitações da PS, surgiu a MLP. A operação de uma MLP é baseada em unidades ou neurônios chamados *perceptrons*. Se vários neurônios do *perceptron* são usados em paralelo e conectados a um neurônio de saída, a rede MLP com uma camada oculta é obtida (JÚNIOR, 2014). Tipicamente, uma rede MLP (FIGURA 12) é composta de uma camada de entrada que recebe os vetores de atributos e uma ou mais camadas intermediárias, compostas por neurônios somadores com função de ativação não-linear e uma camada de saída, também composta por neurônios somadores (NETO, 2006).

A seguir, tem-se um exemplo de MLP com três entradas, duas camadas intermediárias com quatro neurônios e uma camada de saída com um neurônio, produzindo uma única informação de saída (FIGURA 12).

Figura 12 – Exemplo de uma MLP com duas camadas intermediárias



Fonte: (FIORIN et al., 2011)

A rede é totalmente conectada, ou seja, um neurônio em qualquer camada da rede está conectada a todos os neurônios da camada anterior e em cada ligação é associado um valor numérico chamado peso sináptico (HAYKIN, 1998).

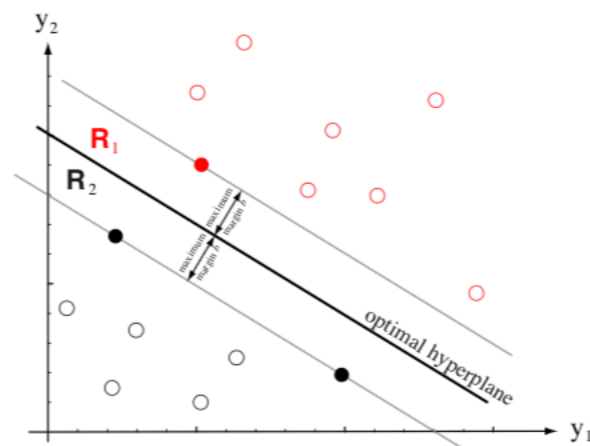
O treinamento dessa rede utiliza o algoritmo de retropropagação de erro, conhecido como *backpropagation*. Este algoritmo se baseia na aplicação repetida da regra da cadeia para calcular a influência de cada peso na rede com respectivo valor arbitrário de erro (SARMENTO, 2016). Dois fatores que precisam ser observados no uso de RNA, são o número de neurônios e a faixa de valores para as suas saídas, pois a escolha correta desses valores pode melhorar o desempenho da rede de forma substancial (NETO, 2006).

4.4 Máquina de Vetores de Suporte

Máquina de Vetores de Suporte ou *Support Vector Machine* (SVM), é uma máquina linear com algumas propriedades, que pode ser utilizada para classificação de padrões e regressão não-linear (HAYKIN, 1998). A técnica de SVM se enquadra no paradigma estatístico (VAPNIK, 1999), que possui como objetivo nesta tarefa de aprendizado, encontrar uma função ou modelo capaz de categorizar os dados em suas respectivas classes, através da exploração das dependências funcionais de um conjunto de dados por meio de modelos estatísticos (FERREIRA, 2016).

Os vetores de suporte (FIGURA 13), que são (igualmente) próximos do hiperplano, são as amostras de treinamento que definem o hiperplano de separação ideal também conhecido como *optimal hyperplane*, que separa todos os pontos de uma classe de dados dos de outra classe. (DUDA; HART; STORK, 2000). O objetivo principal da SVM é construir um hiperplano como a superfície de decisão de tal forma que a margem de separação entre exemplos positivos e negativos seja maximizada (HAYKIN, 1998). Assim quanto maior a margem, melhor a generalização do classificador.

Figura 13 – SVM - Exemplo de Hiperplano Ótimo



Fonte: (DUDA; HART; STORK, 2000)

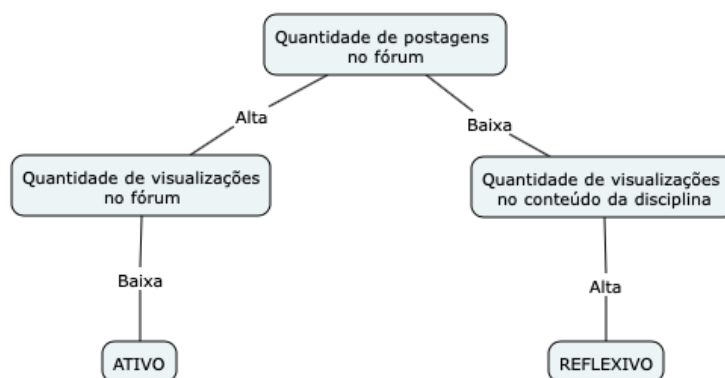
Inicialmente o SVM foi implementado para problemas de classificação binária. Diversos métodos foram criados de forma a estendê-lo para classificação com N classes. Um desses métodos é o *one-versus-all*, que constrói k modelos SVM onde k é o número de classes. Assim, o SVM é treinado com todos os exemplos da classe com rótulos positivos e todos os outros exemplos com rótulos negativos. Já o outro método denominado *one-versus-one* constrói $k(k-1)/2$ classificadores, no qual cada um é treinado através de todas as combinações de pares das k classes (HSU; LIN, 2002).

4.5 Árvores de Decisão

O modelo de Árvore de Decisão ou *Decision Tree*, (DT) é baseado no paradigma de aprendizado simbólico que fundamenta-se na construção de representações simbólicas, utilizando estruturas gráficas ou lógicas, de um conceito por meio da análise de exemplos e contraexemplos deste conceito (FERREIRA, 2016). Uma árvore de decisão é composta por uma sequência de perguntas. O primeiro nó da árvore fica no topo da árvore e é chamado nó raiz, onde se inicia o processo de classificação. Em seguida, ramificações sucessivas são conectadas a outros nós, essas conexões são chamadas links ou ramos. Continua-se neste processo até chegar a um nó folha, que possui um rótulo de categoria que será atribuído ao padrão de teste (DUDA; HART; STORK, 2000).

As Árvores de Decisão, classificam as instâncias ordenando-as na árvore da raiz para algum nó folha, onde será feita a classificação. Cada nó na árvore especifica um teste de algum atributo da instância e cada ramificação descendente desse nó corresponde a um dos valores possíveis para este atributo (MITCHELL, 1997). O número de ligações em cada nó é chamado fator de ramificação. Em árvores binárias o fator de ramificação de cada nó é igual a dois, em árvores n-árias o fator de ramificação é n (STANGE, 2010). A seguir, um exemplo de árvore de decisão, dentro do conceito de estilos de aprendizagem (FIGURA 14).

Figura 14 – Exemplo de Árvore de Decisão

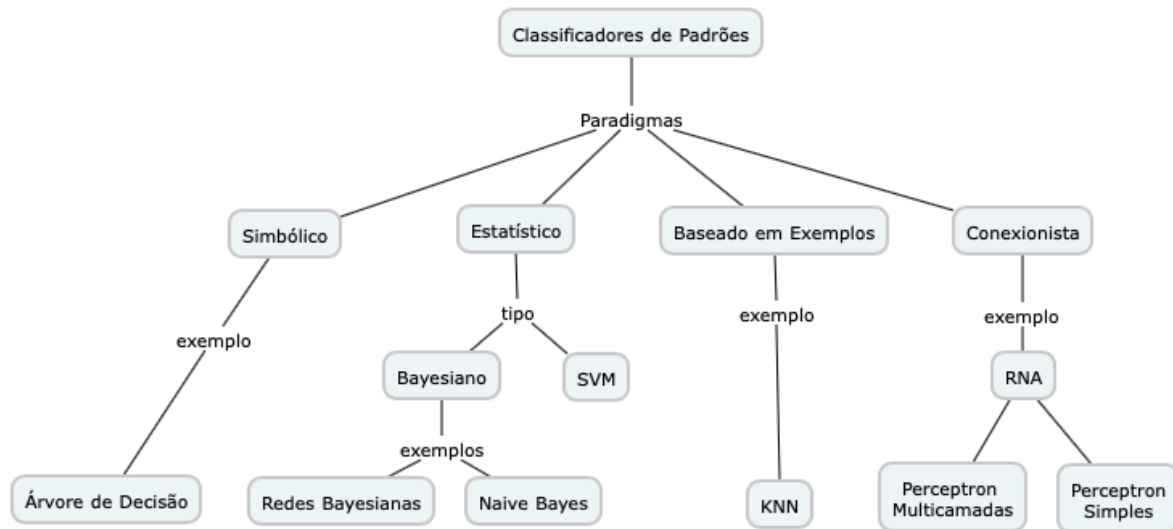


Fonte: O Autor

A partir desta DT, regras podem ser obtidas. Por exemplo, se a quantidade de postagens no fórum for alta e a quantidade de visualizações no fórum for baixa, provavelmente o aluno será do tipo ativo. Por outro lado, se a quantidade de postagens for baixa e a quantidade de visualizações no conteúdo da disciplina for alta, o aluno será classificado como reflexivo.

De forma a resumir os classificadores vistos neste capítulo, na FIGURA 15 é apresentado um mapa conceitual representando suas categorias.

Figura 15 – Mapa conceitual dos Classificadores Descritos



Fonte: O Autor

O próximo capítulo apresenta a Revisão Sistemática da Literatura elaborada de forma a sintetizar os trabalhos relacionados ao tema proposto nesta dissertação.

5 Revisão Sistemática da Literatura

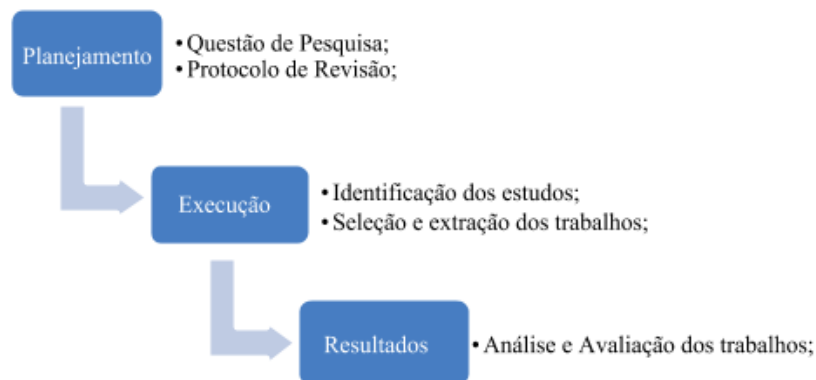
Neste capítulo realiza-se uma Revisão Sistemática da Literatura (RSL) a fim de identificar as estratégias e metodologias utilizadas para classificação automática de perfis de aprendizagem, assim como detectar os modelos de estilos mais comuns. O objetivo é avaliar e interpretar os estudos mais relevantes nessa área, identificando e analisando os trabalhos relacionados nos últimos anos.

5.1 Processo de Revisão Sistemática

5.1.1 Planejamento

De forma a atingir o objetivo desta revisão sistemática, as seguintes etapas foram realizadas (FIGURA 16), conforme indicado por Kitchenham (KITCHENHAM, 2004).

Figura 16 – Etapas do Processo de Revisão



Fonte: Adaptado de (KITCHENHAM, 2004)

Inicialmente foram definidas duas questões de pesquisa que esta RSL tem como objetivo responder:

- QP1: Quais modelos de estilos de aprendizagem na literatura são mais utilizados nos estudos realizados?
- QP2: Quais são os diferentes métodos usados para identificar estilos de aprendizagem de forma automática?

O protocolo de revisão foi elaborado de forma a responder às questões de pesquisa definidas. No mês de março de 2018 foram consultadas quatro bases de arquivos científicos para a etapa de seleção dos trabalhos: *Scopus*, *Association for Computing Machinery Digital Library* (ACMDL), *Web of Science* e *IEEEExplorer*.

Estas bases foram selecionadas devido aos mecanismos de buscas online que possuem, possibilitando o uso de filtros por ano e tipo de publicação, além de abranger pesquisas relacionadas ao tema da computação. As *strings* de busca utilizadas nestas bases (Tabela 3) foram elaboradas a partir de palavras-chaves como: *data mining methods*, *data mining techniques*, *machine learning*, *learning styles*, assim como seus sinônimos.

Tabela 3 – *Strings* de busca usadas nos repositórios

Repositório	String de Busca
Scopus	TITLE-ABS-KEY(("learning styles"OR "learning style") AND ("data mining techniques"OR "data mining methods"OR "data mining"OR "machine learning"))
ACM	Searched for+("learning style")+("data mining techniques" "data mining methods" data mining machine learning")
Web of Science	TS= (("learning style*") AND ("data mining technique*"OR "data mining method*"OR "data mining"OR "machine learning"))
IEEEExplorer	((((("learning styles") AND "data mining techniques") OR "data mining methods") OR "data mining") OR "machine learning")

Fonte: O Autor

Para que a pesquisa fosse mais específica e criteriosa, somente foram aceitos trabalhos que atendessem aos critérios de inclusão e exclusão definidos a seguir (Tabela 4).

Tabela 4 – Critérios de Inclusão e Exclusão

Critério	ID	Descrição
Inclusão	I-1	Trabalhos na área de Ciências da Computação
	I-2	Trabalhos escritos em inglês
	I-3	Trabalhos publicados entre 2014 e março de 2018
	I-4	Artigos revisados por pares
	I-5	Utiliza abordagens automáticas para classificação dos Estilos de Aprendizagem
Exclusão	E-1	Trabalhos Duplicados
	E-2	Trabalhos que não contém palavras-chaves no título ou resumo
	E-3	Número de páginas inferior a quatro
	E-4	Trabalhos não disponíveis integralmente na web e não acessíveis de forma gratuita

Fonte: O Autor

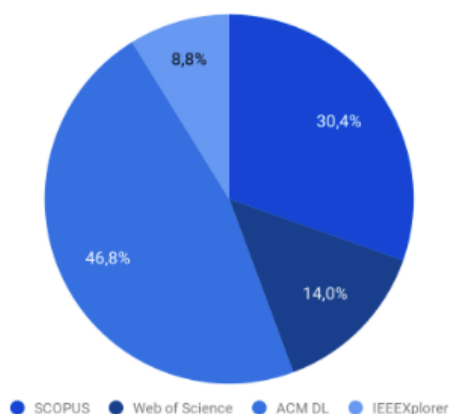
O objetivo principal nesta revisão é explorar artigos mais recentes para identificar quais técnicas de classificação e quais modelos de estilos de aprendizagem são mais utilizados atualmente, logo é importante ressaltar que apenas foram consideradas publicações realizadas nos últimos quatro anos. Ao todo foram selecionados cinco critérios de inclusão e quatro de exclusão, utilizados na etapa de seleção.

5.1.2 Execução

A etapa de execução realiza a identificação, seleção e extração dos trabalhos. Durante o processo de identificação, as *strings* de busca foram aplicadas nas bases de artigos científicos e o conjunto inicial foi de 566 artigos, distribuídos nos quatro repositórios (Figura 17). A maior parte dos artigos (77,2%) foi encontrada nas bases de dados da ACM DL e *Scopus*, onde quase metade (46,8%) dos artigos vieram da ACM DL. Os outros 22,8% dos trabalhos foram encontrados nos repositórios do *IEEE Explorer* e *Web of Science*, onde somente 8,8% vieram do *IEEE Explorer*.

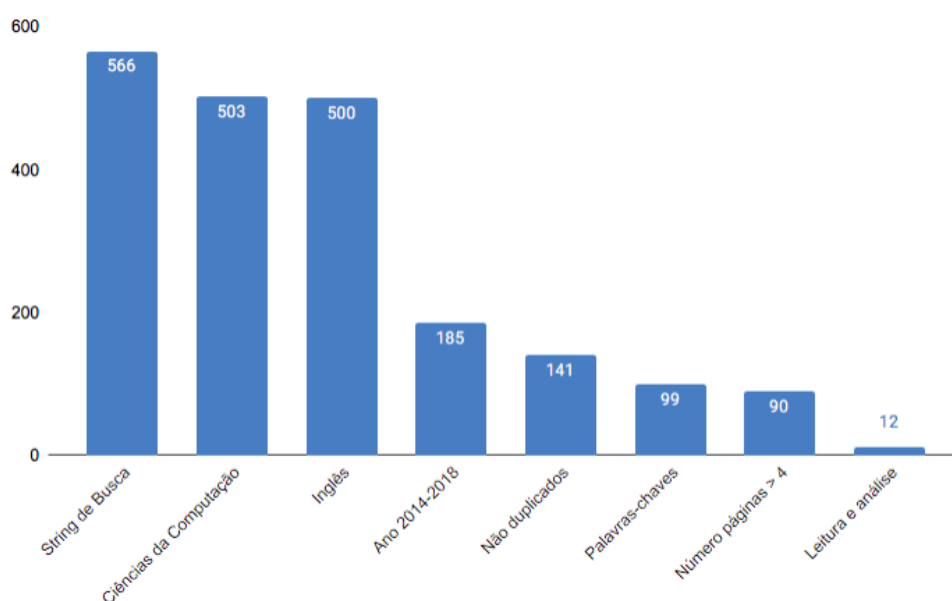
Na etapa de seleção ocorre a aplicação dos critérios de inclusão e exclusão. A fim de encontrar quais artigos utilizam abordagens automáticas para identificar os estilos de aprendizagem, foi realizada a leitura do título e resumo dos noventa artigos finais, onde apenas doze trabalhos foram escolhidos para uma análise completa (Figura 18). Estes trabalhos são analisados de forma mais detalhada na próxima seção a fim de responder às questões de pesquisa elaboradas.

Figura 17 – Artigos selecionados por repositório



Fonte: O Autor

Figura 18 – Aplicação dos critérios de inclusão e exclusão



Fonte: O Autor

Com a aplicação do primeiro critério de inclusão (somente artigos da área da computação), 63 artigos foram excluídos. Dos 503 artigos que ficaram, somente 3 não são da língua inglesa. Na aplicação do critério de inclusão 3 (trabalhos publicados entre 2014 e março de 2018), 185 artigos foram extraídos para aplicação dos outros critérios.

Após retirar os trabalhos duplicados, trabalhos que não contém palavras-chave no resumo ou título e trabalhos com número de páginas menor que quatro, foram selecionados 12 artigos para a etapa de análise e avaliação. Nesta etapa foi levado em consideração

também a qualidade de cada trabalho, assim como a qualidade da conferência/periódico onde os mesmos foram publicados.

Estes trabalhos foram selecionados para serem lidos por completo, de modo a avaliar se realmente descrevem a forma, com alto nível de detalhe, de como é feita a classificação dos estilos de aprendizagem nos ambientes virtuais de aprendizagem.

5.1.3 Análise e Interpretação dos Resultados

Para fornecer uma maneira mais simples de visualizar os resultados, estes foram separados em dois grupos: informações gerais e respostas das questões de pesquisa. Em primeiro lugar, será apresentada uma visão geral dos estudos relevantes, em seguida as respostas às questões de pesquisa.

5.1.3.1 Informações Gerais dos Resultados

Os documentos filtrados foram lidos e, em seguida, foram recuperadas muitas outras informações. Depois disso, armazenou-se o título, ano de publicação, local de publicação (periódico ou conferência), nome e localização do autor. Inicialmente considerou-se algumas informações gerais, como, distribuição de artigos por ano (Tabela 5).

Tabela 5 – Distribuição de artigos por ano

Ano	Artigo	Publicação
2014	(EFRATI; LIMONGELLI; SCIARRONE, 2014)	Conferência
	(MOHAMED; AHMAD; SHAMSUDDIN, 2014)	Conferência
2015	(HALAWA; SHEHAB; HAMED, 2015)	Conferência
	(HASSAN; HEGAZY, 2015)	Conferência
	(PAIREEKRENG; PREXAWANPRASUT, 2015)	Conferência
	(SWETA; LAL, 2015)	Conferência
2016	(ADEL; LATHAM; CROCKETT, 2016)	Conferência
	(HUNG; CHANG; LIN, 2016)	Periódico
	(LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016)	Periódico
	(POPESCU et al., 2016)	Conferência
2017	(BINH; DUY, 2017)	Conferência
	(ZHONG et al., 2017)	Periódico

Fonte: O Autor

Os seguintes padrões foram observados no resultado: o número de publicações varia de 2 a 4 artigos; 2015 e 2016 são os anos mais produtivos com 4 publicações cada. A maioria dos estudos (75%) foi publicada em anais de congressos, enquanto os trabalhos (HUNG;

CHANG; LIN, 2016; LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016; ZHONG et al., 2017), foram publicados em periódicos.

Já as informações sobre em quais países foram realizadas as pesquisas filtradas, são exibidas na Tabela 6. Uma tendência que se observa nos resultados é que, com exceção do Egito, todos os trabalhos são de países diferentes, o que pode apontar para existência de particularidades locais para o tema de perfis de aprendizagem.

Tabela 6 – Países das pesquisas selecionadas

Localização Autor	Artigo
China	(ZHONG et al., 2017)
Itália	(EFRATI; LIMONGELLI; SCIARRONE, 2014)
Egito	(HASSAN; HEGAZY, 2015), (HALAWA; SHEHAB; HAMED, 2015)
Tailândia	(PAIREEKRENG; PREXAWANPRASUT, 2015)
Malásia	(MOHAMED; AHMAD; SHAMSUDDIN, 2014)
Japão	(LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016)
Taiwan	(HUNG; CHANG; LIN, 2016)
Romênia e Estados Unidos	(POPESCU et al., 2016)
Vietnã	(BINH; DUY, 2017)
Reino Unido	(ADEL; LATHAM; CROCKETT, 2016)
Índia	(SWETA; LAL, 2015)

Fonte: O Autor

5.1.3.2 Respostas para as Questões de Pesquisa

Nesta seção apresentam-se os resultados agrupados para responder às duas questões de pesquisas levantadas no início deste documento. O processo foi iniciado com a leitura dos doze artigos selecionados. Inicialmente foi feita a identificação dos objetivos e das metodologias de cada um dos artigos, selecionando informações relevantes para responder cada uma das questões.

QP1: Quais modelos de estilos de aprendizagem na literatura são mais utilizados nos estudos realizados?

Diversos tipos de modelos de estilos de aprendizagem foram abordados nos trabalhos selecionados (TABELA 7). O modelo de Felder e Silverman foi o mais utilizado nas pesquisas, totalizando 66% dos trabalhos. Comparando-se o número de citações deste modelo com os demais, através do *Google Scholar*, também nota-se uma maior tendência em seu uso.

Tabela 7 – Modelos de Estilos de Aprendizagem

Modelo	Artigo
Felder-Silverman	(ADEL; LATHAM; CROCKETT, 2016), (BINH; DUY, 2017), (HUNG; CHANG; LIN, 2016), (LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016), (MOHAMED; AHMAD; SHAMSUDDIN, 2014), (PAIREEKRENG; PREXAWANPRASUT, 2015), (POPESCU et al., 2016), (SWETA; LAL, 2015)
Grasha-Riechmann	(EFRATI; LIMONGELLI; SCIARRONE, 2014)
Honey-Mumford	(ZHONG et al., 2017)
Myers-Briggs	(HALAWA; SHEHAB; HAMED, 2015)
VARK	(HASSAN; HEGAZY, 2015)

Fonte: O Autor

Dos oito trabalhos que abordam o modelo de Felder e Silverman, seis desses trabalhos (PAIREEKRENG; PREXAWANPRASUT, 2015; LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016; HUNG; CHANG; LIN, 2016; POPESCU et al., 2016; BINH; DUY, 2017; ADEL; LATHAM; CROCKETT, 2016), usam o questionário *Index Learning System* (ILS), elaborado por Felder e Soloman (FELDER; SOLOMAN, 1997) para identificar os estilos de aprendizagem dos alunos. Os autores fazem uso das respostas do questionário para comparar com a classificação feita pelas abordagens automáticas que utilizam os registros das interações dos alunos com o ambiente virtual ou dados pessoais extraídos da base de dados dos cursos.

O modelo de Grasha-Riechmann, aplicado como referência para classificar os estilos, também foi utilizado em um dos estudos, onde os arquivos de log da plataforma Moodle foram extraídos e utilizados para classificação dos estilos. No primeiro experimento os resultados mostraram resultados positivos entre os *clusters* e os estilos de aprendizagem (EFRATI; LIMONGELLI; SCIARRONE, 2014).

Outro modelo utilizado foi o de Honey e Mumford. Na pesquisa que abordou este modelo, foi utilizada mineração de dados educacionais para identificar os estilos de aprendizagem dos alunos com base nos comportamentos de aprendizagem destes em MOOCs (*Massive Open Online Course*) (ZHONG et al., 2017).

O modelo de Myers-Briggs também foi aplicado em um dos trabalhos selecionados. O trabalho proposto utilizou dados do envolvimento do aluno com o sistema *Moodle* e da rede social, *Facebook*. O objetivo do trabalho foi ajudar os alunos a se conscientizarem de sua personalidade, o que, por sua vez, os tornam mais eficientes em seus hábitos de estudo, assim como fornecer informações para professores, de forma a compreenderem melhor a personalidade de cada aluno (HALAWA; SHEHAB; HAMED, 2015).

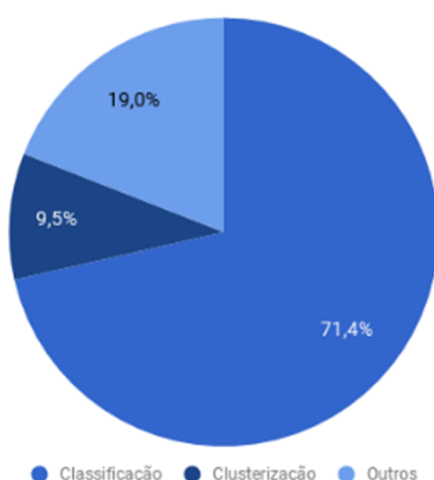
Por último temos o modelo de VARK, acrônimo para *Visual, Auditory, Read/Write, e Kinesthetic*. O trabalho utilizou duas bases de dados para fazer a classificação, onde a primeira refere-se aos dados do *Moodle* e a segunda base refere-se aos dados das redes

sociais, como *Facebook*, *Twitter* e *Google+* (HASSAN; HEGAZY, 2015).

QP2: Quais são os diferentes métodos usados para identificar estilos de aprendizagem de forma automática?

Inicialmente identificou-se a distribuição dos tipos de técnicas usadas nos trabalhos referentes à identificação dos estilos de aprendizagem (FIGURA 19). A maioria das pesquisas (71,4%) implementa a técnica de classificação. Esse método é um tipo de aprendizado de máquina supervisionado (DUDA; HART; STORK, 2000).

Figura 19 – Técnicas usadas para identificação dos estilos de aprendizagem



Fonte: O Autor

Ao todo, vinte algoritmos foram utilizados nesses trabalhos (TABELA 8). De todos esses, os algoritmos de Redes Bayesianas, Árvore de Decisão, *Naive Bayes* e Redes Neurais Artificiais (RNA), foram os mais referenciados nas pesquisas. As Redes Bayesianas foram aplicadas em cinco trabalhos e os demais em quatro. Apesar dessas técnicas não serem tão recentes, ainda são aplicáveis na classificação do perfil dos estudantes na atualidade.

Já os outros algoritmos foram usados em um ou dois trabalhos apenas. Dos doze artigos analisados, cinco deles (EFRATI; LIMONGELLI; SCIARRONE, 2014; POPESCU et al., 2016; HUNG; CHANG; LIN, 2016; BINH; DUY, 2017; SWETA; LAL, 2015), utilizaram somente um algoritmo, enquanto os demais utilizaram no mínimo dois (ZHONG et al., 2017) e no máximo dez (HALAWA; HAMED; SHEHAB, 2015).

Os algoritmos que obtiveram as melhores acurácias foram: RNA (100%), SVM (99,7%) (ZHONG et al., 2017), Bijetive Soft Set (98,4%) (MOHAMED; AHMAD; SHAM-SUDDIN, 2014), OneR (97,4%) (HALAWA; SHEHAB; HAMED, 2015) e IBK (97%) (HASSAN; HEGAZY, 2015).

Tabela 8 – Algoritmos utilizados

Algoritmo	Artigo	Acurácia
Bayesian Network	(HASSAN; HEGAZY, 2015), (LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016), (HALAWA; SHEHAB; HAMED, 2015), (ADEL; LATHAM; CROCKETT, 2016), (SWETA; LAL, 2015)	89,9%, 68,1%, 88,5%, 83,3%, -
Bijective Soft Set	(MOHAMED; AHMAD; SHAMSUDDIN, 2014)	98,4%
Decision Table	(HALAWA; SHEHAB; HAMED, 2015)	85,4%
Decision Tree/J48 Graft	(PAIREEKRENG; PREXAWANPRASUT, 2015), (LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016), (HALAWA; SHEHAB; HAMED, 2015), (ADEL; LATHAM; CROCKETT, 2016)	62,7%, 79,1%, 92,1%, 83,3%
Discriminant Function Analysis (DFA)	(POPESCU et al., 2016)	62,6%
Expectation Maximization (EM)	(EFRATI; LIMONGELLI; SCIARRONE, 2014)	-
FURIA	(MOHAMED; AHMAD; SHAMSUDDIN, 2014)	97,2%
Hyper Pipes	(HASSAN; HEGAZY, 2015)	92,3%
JRIP	(HALAWA; SHEHAB; HAMED, 2015), (HASSAN; HEGAZY, 2015)	63,5%, 90,3%
K-means	(ZHONG et al., 2017), (HUNG; CHANG; LIN, 2016)	56,5%, -
KNN/IBK	(HALAWA; SHEHAB; HAMED, 2015), (ADEL; LATHAM; CROCKETT, 2016), (HASSAN; HEGAZY, 2015)	90,6%, 62,9%, 97%
KStar	(HALAWA; SHEHAB; HAMED, 2015)	48,9%
Naive Bayes	(HALAWA; SHEHAB; HAMED, 2015), (PAIREEKRENG; PREXAWANPRASUT, 2015), (LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016), (ADEL; LATHAM; CROCKETT, 2016)	68,2%, 42,5%, 65,2%, 79,6%
OneR	(HALAWA; SHEHAB; HAMED, 2015)	97,4%
RandomTree/Random Forests	(LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016), (HALAWA; SHEHAB; HAMED, 2015), (ADEL; LATHAM; CROCKETT, 2016)	78,1%, 93,2%, 81,6%
Rede Neural/MLP	(ZHONG et al., 2017), (PAIREEKRENG; PREXAWANPRASUT, 2015), (BINH; DUY, 2017), (ADEL; LATHAM; CROCKETT, 2016)	100%, 88,3%, 80,6%, 69,9%
Rough Set	(MOHAMED; AHMAD; SHAMSUDDIN, 2014)	96,3%
Simple CART	(ADEL; LATHAM; CROCKETT, 2016)	83,3%
Simple Logistic	(HASSAN; HEGAZY, 2015)	93,8%
SVM	(ZHONG et al., 2017), (PAIREEKRENG; PREXAWANPRASUT, 2015)	99,7%, 25,5%

Fonte: O Autor

Metade dos trabalhos analisados (EFRATI; LIMONGELLI; SCIARRONE, 2014; HASSAN; HEGAZY, 2015; LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016; HALAWA; SHEHAB; HAMED, 2015; ADEL; LATHAM; CROCKETT, 2016; SWETA; LAL, 2015), utilizou a ferramenta *Waikato Environment for Knowledge Analysis* (WEKA) para aplicação e análise dos algoritmos.

Essa ferramenta possui uma coleção de algoritmos de aprendizado de máquina para mineração de dados, além de ser uma plataforma de código aberto baseada em Java desenvolvida na Universidade de Waikato, Nova Zelândia (LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016). Desses trabalhos que fizeram uso da ferramenta WEKA, cinco deles usaram o ambiente virtual Moodle, totalizando 42% dos doze trabalhos pesquisados.

Já os dados mais utilizados para fazer a identificação dos estilos de aprendizagem, foram dados coletados a partir dos registros gerados pelas interações dos alunos dentro dos AVAs (EFRATI; LIMONGELLI; SCIARRONE, 2014; ZHONG et al., 2017; HASSAN; HEGAZY, 2015; LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016; HUNG; CHANG; LIN, 2016; HALAWA; SHEHAB; HAMED, 2015; SWETA; LAL, 2015). Dois trabalhos (POPESCU et al., 2016; ADEL; LATHAM; CROCKETT, 2016) usaram análise de dados textuais, um utilizou somente dados do questionário ILS (BINH; DUY, 2017), outro usou dados pessoais (PAIREEKRENG; PREXAWANPRASUT, 2015) e o último usou dados simulados das interações no AVA (MOHAMED; AHMAD; SHAMSUDDIN, 2014).

Abaixo apresenta-se o resumo de todas as características extraídas dos trabalhos (TABELA 9).

Tabela 9 – Síntese dos trabalhos selecionados

Autor	Modelo	Algoritmo	Dados Utilizados	Qtde.Alunos	Ferramenta
(ADEL; LATHAM; CROCKETT, 2016)	Felder-Silverman	Naive Bayes, Bayes Network, J48, Simple CART, Random Forest, IBK e RBF	ILS e Registros do sistema de tutoria inteligente	75	WEKA
(BINH; DUY, 2017)	Felder-Silverman	MLP	ILS	316	Phyton
(EFRATI; LIMONGELLI; SCIARRONE, 2014)	Grasha-Riechmann	Expectation Maximization (EM)	Registros do ambiente virtual	-	WEKA
(HALAWA; SHEHAB; HAMED, 2015)	Myers-Briggs	NaiveBayes, BayesNet, Kstar, Random forest, J48, OneR, JRIP, KNN/IBK, RandomTree e Decision Table	Registros do ambiente virtual e Redes Sociais	240	WEKA
(HASSAN; HEGAZY, 2015)	VARK	Simple Logistic, IBK, Bayes Net, J48 Graft, Hyper Pipes, JRIP	Registros do ambiente virtual e Redes Sociais	131	WEKA
(HUNG; CHANG; LIN, 2016)	Felder-Silverman	K-means	ILS e Registros do ambiente virtual (jogo)	67	-
(LIYANAGE; GUNAWARDENA; HIRAKAWA, 2016)	Felder-Silverman	J48, Bayesian Network, Naive Bayes e Random Forests.	ILS e Registros do ambiente virtual	80	WEKA
(MOHAMED; AHMAD; SHAMSUDDIN, 2014)	Felder-Silverman	Bijective Soft Set, FURIA e Rough Set	Dados simulados de registros do ambiente virtual	-	-
(PAIREEKRENG; PREXAWANPRASUT, 2015)	Felder-Silverman	Decision Tree, Naive Bayes, RNA e SVM	ILS e Dados pessoais dos alunos	400	-
(POPESCU et al., 2016)	Felder-Silverman	Discriminant Function Analysis (DFA)	ILS e Análise de texto	66	Phyton
(SWETA; LAL, 2015)	Felder-Silverman	Bayesian Network	Registros do ambiente virtual	40	WEKA
(ZHONG et al., 2017)	Honey e Mumford	K-means, SVM, Rede Neural Artificial	ILS e Registros do ambiente virtual	183	-

Fonte: O Autor

5.1.4 Considerações sobre o capítulo

A análise dos trabalhos pesquisados e selecionados neste mapeamento sistemático, sobre diferentes abordagens automáticas na identificação dos estilos de aprendizagem de alunos, teve como objetivo identificar quais modelos e métodos mais utilizados para subsidiar a questão de pesquisa desta dissertação.

Foram relacionados trabalhos de 2014 a 2018 e questões de pesquisa foram definidas juntamente com um protocolo de revisão para orientar a busca feita em quatro bases de artigos. Pretende-se com esse estudo ampliar o estado da arte, contribuindo por evidenciar trabalhos de pesquisa voltados para identificação desses estilos.

Alguns pontos observados são de que a maioria dos trabalhos são de conferências, o modelo Felder e Silverman é o modelo mais utilizado e as técnicas mais utilizadas são do tipo classificação. As técnicas mais referenciadas foram: RNA, Árvores de decisão, Redes Bayesianas e Naive Bayes. Outro ponto que merece destaque nos resultados encontrados é a predominância de dois tipos de coleta de dados: as coletas baseadas em questionários e as coletas baseadas em registros do ambiente virtual de aprendizagem.

Com isso, além das técnicas/ferramentas utilizadas presentes nos estudos selecionados, esta pesquisa mostrou os avanços no que se refere ao uso de técnicas de mineração de dados educacionais para identificação dos perfis de aprendizagem. As fontes de dados usadas nos trabalhos citados são extraídas através de questionários ou logs de acessos de algum AVA, principalmente, o Moodle. Logo, esta pesquisa se difere das apresentadas por utilizar o Modelo LV, implementado no Moodle, para a extração de dados. Espera-se a partir desta revisão estimular a condução de mais pesquisas nesta área, de modo a potencializar o uso de novos algoritmos de classificação, além de estimular o desenvolvimento de ambientes virtuais personalizados com base nesses perfis.

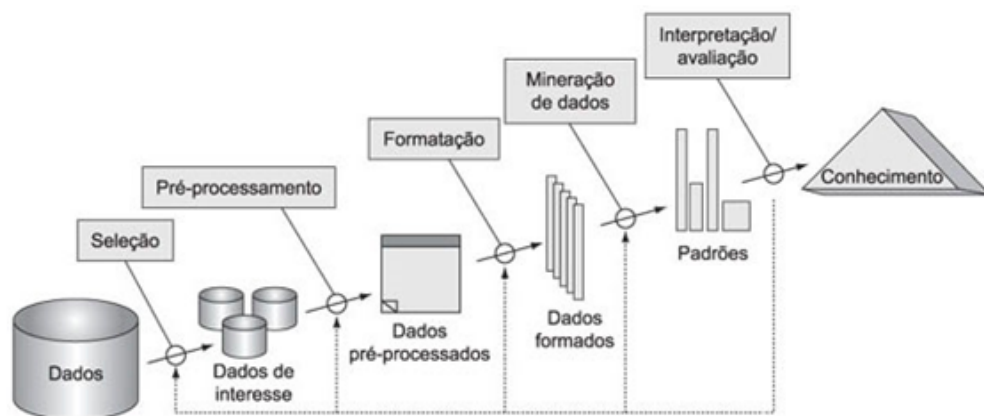
No próximo capítulo aborda-se o método de pesquisa aplicado neste trabalho, onde são esclarecidas as etapas de aquisição e preparação dos dados, bem como o processo de extração de características, experimentos e procedimento de análise.

6 Metodologia

Neste capítulo, apresentam-se as etapas do método de pesquisa utilizadas na realização deste trabalho. O subtópico 6.1 relata a caracterização e entendimento do domínio. Já no subtópico 6.2 apresenta-se a seleção e coleta dos dados, informando como e quais dados foram extraídos. No subtópico 6.3 descreve-se a etapa de pré-processamento, que trata dos métodos de seleção de características utilizados. Logo após, no subtópico 6.4 retrata-se a etapa de formatação dos dados, informando o método de balanceamento de dados. Já no subtópico 6.5 detalha-se a modelagem utilizada nos experimentos, apresentando os algoritmos escolhidos e as parametrizações configuradas. Por último, no subtópico 6.6, explica-se os procedimentos e as métricas usadas para interpretação e avaliação dos experimentos.

A metodologia utilizada nesta dissertação foi fundamentada no processo KDD (*Knowledge Discovery in Databases*) (FIGURA 20) e refere-se ao processo geral de descoberta de conhecimento útil a partir de dados. Este método é composto por várias etapas, entre elas: seleção, pré-processamento, formatação e mineração de dados, além da interpretação adequada dos resultados, que são essenciais para garantir que o conhecimento útil seja derivado dos dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). A utilização de uma metodologia de exploração de dados permite uma interpretação adequada dos resultados, pois exige que uma série de etapas seja seguida, antes da aplicação de algoritmos de aprendizado de máquina.

Figura 20 – KDD - *Knowledge Discovery in Databases*



Fonte: Adaptado de (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996)

6.1 Caracterização da Pesquisa e Entendimento do Domínio

Quanto à sua natureza, esta pesquisa pode ser classificada como aplicada, pois objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos (SILVA; MENEZES, 2005). Do ponto de vista da forma de abordagem do problema, ela pode ser classificada como quantitativa, pois utilizou-se métodos estatísticos para comparação e avaliação dos resultados.

Em relação aos objetivos ela está classificada como descritiva, dado que, visa descobrir a existência de associações entre variáveis (GIL, 2002). Com relação aos procedimentos técnicos, ela está classificada como bibliográfica e pesquisa de campo do tipo quantitativa-descritiva, que consiste na observação de fatos tal como ocorrem espontaneamente, na coleta de dados a eles referentes e no registro de variáveis que se presume relevantes, para analisá-los (LAKATOS; MARCONI, 2003).

Para o entendimento do domínio foram realizadas duas etapas: Revisão Sistemática da Literatura (descrita no Capítulo 3) e aplicação do questionário ILS, para a identificação dos perfis de aprendizagem dos estudantes (ANEXO A). Na primeira etapa foi realizado o entendimento da literatura sobre as problemáticas nos AVAs, no que se refere às adaptações de conteúdo e personalização de sistemas, de acordo com os estilos de aprendizagem e às diversas abordagens para sua identificação.

A segunda etapa foi a aplicação do questionário ILS (*Index Learning Styles*) do modelo de Felder e Silverman (FELDER; SOLOMAN, 1997). Como explicado no Capítulo 2, este questionário possui 44 questões, sendo 11 questões para cada dimensão. O questionário foi disponibilizado através de uma plataforma *web*¹, desenvolvida na linguagem Java² e JavaScript³. Já o Banco de Dados utilizado foi o *PostgreSQL*⁴.

O questionário foi aplicado aos alunos durante uma aula presencial, além de ter sido disponibilizado dentro do AVA onde o curso foi ofertado. As perguntas são objetivas e possuem apenas duas alternativas (a ou b). A resposta é calculada através da soma de cada item marcado nas quatro dimensões. Para cada uma das quatro escalas, subtrai-se o total menor do maior, mostrando o resultado da diferença (1 a 11) e a letra (a ou b) que teve o total maior. As respostas dos alunos foram calculadas e salvas automaticamente. O código fonte da plataforma do questionário encontra-se na plataforma de hospedagem *GitHub*⁵.

A resposta de cada aluno foi registrada e utilizada para rotular o conjunto de variáveis de acordo com o perfil identificado pelo questionário nas dimensões (processamento,

¹ <https://meuperfil.herokuapp.com>

² <https://java.com>

³ <https://www.javascript.com>

⁴ <https://www.postgresql.org>

⁵ <https://github.com/ramiresnas/questions>

percepção, captação e organização) do modelo FSLSM. Este modelo vem sendo aplicado em diversas pesquisas na comunidade científica, como pode-se observar na revisão sistemática do Capítulo 5.

Foram utilizados dados referentes à primeira disciplina de dois cursos de Especialização, um de Educação Ambiental e o outro de Educação de Jovens e Adultos (EJA), ofertados pela Universidade Federal do Ceará (UFC) em convênio com a Secretaria de Educação (SEDUC) de Sobral, Ceará. Os dois cursos operam na modalidade EaD, em regime semipresencial. A plataforma *Moodle*⁶ (versão 2.6) foi adotada como Ambiente Virtual de Aprendizagem (AVA). Esses dois cursos foram selecionados devido a quantidade de alunos matriculados e por utilizarem o *Plugin* de avaliação formativa do Modelo LV (SALES, 2010) como método de avaliação.

No total dez turmas foram analisadas, sendo cinco turmas no curso de Educação Ambiental e cinco turmas no curso de Educação de Jovens e Adultos. Ao todo, 112 alunos responderam ao questionário, de um total de 266 matriculados (TABELA 10). O próximo subtópico aborda a seleção e coleta dos dados a serem analisados nesta pesquisa.

Tabela 10 – Dados dos cursos de pós-graduação a distância - UFC Sobral

Disciplina	Período	Curso	Turma	Qtd. Alunos
Introdução ao Curso e ao Ambiente Virtual	04/09/17 à 22/10/17	Educação Ambiental	A	28
			B	26
			C	26
			D	26
			E	26
		Educação de Jovens e Adultos (EJA)	A	28
			B	26
			C	25
			D	25
			E	30
TOTAL				266

Fonte: O Autor

6.2 Seleção e Coleta dos Dados

Inicialmente foi realizado o entendimento sobre a arquitetura do sistema de Banco de Dados do *Moodle*⁷ e a documentação oficial do sistema foi utilizada. Além desta documentação, também utilizou-se a documentação do *plugin* LV para acesso aos dados da instituição. Em seguida foram identificadas e extraídas, por meio dos *scripts* de consultas *Structured Query Language* (SQL) (APÊNDICE A), as variáveis comportamentais armazenadas nos *logs* do Banco de Dados (BD) (MySQL Server⁸ 5.6) da plataforma *Moodle*,

⁶ www.helpclassonline.com.br

⁷ docs.moodle.org/dev/Database_Schema

⁸ <https://dev.mysql.com/downloads/mysql/>

levando em consideração os aspectos de anonimato e segurança das informações.

Esta identificação das variáveis levou em consideração os comportamentos dos alunos de acordo com as características definidas na literatura sobre a teoria de estilos de aprendizagem (GRAF; KINSHUK; LIU, 2009). As características equivalem ao conjunto de interações dentro de um AVA que são relevantes na identificação dos perfis, capazes de descrever as preferências de cada aluno, de acordo com os seus padrões de comportamento (TABELA 11). O “+” e “-” indicam, respectivamente, uma alta e uma baixa ocorrência do respectivo padrão do ponto de vista de cada dimensão.

Tabela 11 – Padrões de comportamento relevantes para cada dimensão de Estilo de Aprendizagem do FSLSM

Processamento		Percepção		Captação		Organização	
Ativo	Reflexivo	Sensorial	Intuitivo	Visual	Verbal	Sequencial	Global
forum_visualiz(-)	forum_visualiz(+)	quiz_revisoes(+)	quiz_revisoes (-)	conteudo_ visualiz (-)	conteudo_ visualiz (+)	visaogeral_ visualiz(-)	visaogeral_ visualiz(+)
forum_post(+)	forum_post(-)	conteudo_ visualiz(-)	conteudo_ visualiz(+)	forum_post(-)	forum_ visualiz(+)	visaogeral_ tempo(-)	visaogeral_ tempo(+)
conteudo_ visualiz(-)	conteudo_ visualiz(+)	exercic_ visualiz(+)	exemplos_ visualiz(-)	conteud-grafico_ visualiz(+)	forum_post(+)		

Fonte: Adaptado de (GRAF, 2007)

O fórum foi escolhido por ser utilizado como um recurso educacional importante na identificação dos estilos (GRAF; KINSHUK; LIU, 2009). Alunos ativos tendem a postar mais vezes nos fóruns do que os alunos reflexivos, que preferem participar de forma mais passiva lendo cuidadosamente e com mais frequência as postagens.

Na dimensão captação, o fórum também auxilia na identificação, no qual se espera que os alunos verbais usem o fórum mais frequentemente e com maior intensidade para discutir e se comunicar com os outros, do que os alunos visuais.

Já o recurso do *quiz* ajuda a identificar alunos na dimensão percepção (sensorial/intuitivo), em que alunos sensoriais tendem a revisar mais vezes as questões do *quiz* antes de enviá-lo, do que os alunos intuitivos. Quanto à identificação das preferências dos alunos por um estilo de aprendizagem global ou sequencial, leva-se em consideração o número de visualizações na página de visão geral do curso.

Além de atributos baseados na literatura, outros atributos também extraídos foram referentes ao Modelo LV (TABELA 12). A partir dessas observações, foram extraídos 21 atributos ao todo, para montar as bases de dados que serão utilizadas nos experimentos. Esses atributos foram extraídos a partir de várias ferramentas educacionais disponíveis no ambiente virtual, como: fórum, *quiz*, glossário, tarefa, *wiki* e outros.

Os resultados destas consultas foram inseridos em planilha eletrônica e salvos no formato .CSV (*Comma-separated values*), para posteriormente serem formatos e aplicados aos algoritmos de classificação.

Tabela 12 – Ferramentas Educacionais e Atributos Extraídos

Tipo	Ferramenta Educacional	Atributo Coletado	Descrição
Baseados na Literatura	Fórum LV	view_forum	Quantidade de visualizações nos fóruns LV
		post_forum	Quantidade de postagens nos fóruns LV
	Glossário	add_gloss	Quantidade de postagens no glossário
		view_gloss	Quantidade de visualizações no glossário
	Tarefa LV	arq_env_tarefa	Quantidade de arquivos enviados na tarefa LV
Baseados no Modelo Learning Vectors		view_tarefa	Quantidade de visualizações na tarefa LV
		review_quiz	Quantidade de revisões no quiz LV
	Quiz LV	media_tempo_quiz	Média de tempo para responder o quiz LV
		attempt_quiz	Quantidade de tentativas no quiz LV
	Material da Disciplina	view_conteudo	Quantidade de visualizações no material de aula
Baseados no Modelo LV	Visão Geral Disciplina	view_visao geral	Quantidade de visualizações no programa do curso
		num_avaliacao_muitobom	Quantidade de avaliações atribuídas como muito bom
		num_avaliacao_regular	Quantidade de avaliações atribuídas como regular
		num_avaliacao_fraco	Quantidade de avaliações atribuídas como fraca
		num_avaliacao_bom	Quantidade de avaliações atribuídas como bom
Baseados no Modelo LV		num_avaliacao_naosatisfatorio	Quantidade de avaliações atribuídas como não-satisfatórias
		fator_beta_disciplina	Nota do fator beta geral na disciplina
		fator_beta_forumlv	Nota do fator beta atribuída nos fóruns LV
		fator_beta_tarefavl	Nota do fator beta atribuída na tarefa LV
		fator_beta_quizlv	Nota do fator beta atribuída no quiz LV
Modelo Learning Vectors		fator_beta_wikilv	Nota do fator beta atribuída no Wiki LV

Fonte: O Autor

6.3 Pré-Processamento

Em posse das respostas do questionário ILS aplicado aos alunos (ANEXO A) e dos dados comportamentais dos alunos extraídos do Banco de Dados da plataforma *Moodle*, foi realizada uma junção desses dados. Após essa junção foi gerada uma tabela contendo as variáveis de ambas as bases de dados (FIGURA 21).

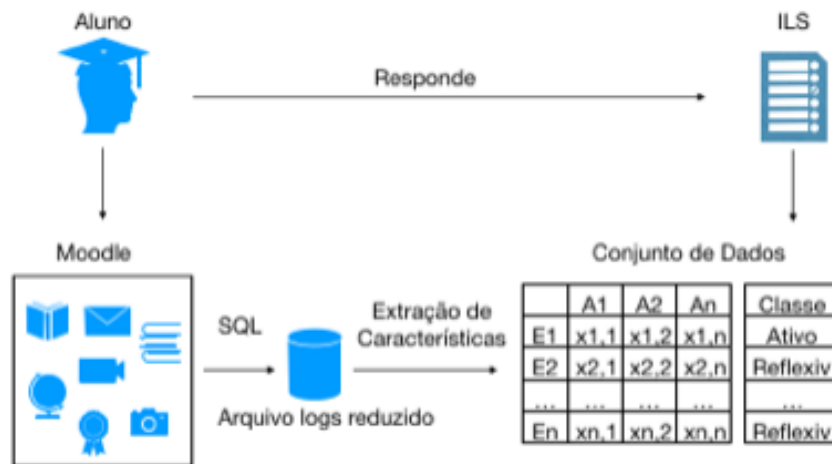
O questionário classificou o aluno em cada uma das dimensões. As respostas foram divididas de acordo com a sua escala, podendo cada perfil ser do tipo forte, moderado ou neutro. Como os experimentos foram realizados através de uma classificação binária, os perfis classificados como fortes e moderados foram considerados como sendo do mesmo perfil. Por exemplo, moderado reflexivo ou forte reflexivo, foi classificado como reflexivo, sem levar em consideração o grau do perfil. O perfil neutro não foi levado em consideração nos experimentos, pois esse perfil possui preferência entre ambas as categorias, não sendo possível assim, classificá-los em apenas uma.

A partir desta tabela gerada, foi feita uma separação dos atributos. Para cada uma das quatro dimensões do modelo FLSM, foi criado um *dataset* com atributos da literatura e outro somente com atributos relacionados ao Modelo LV.

Ainda nesta etapa, ocorreu o processo de limpeza dos dados e *feature selection* ou seleção de atributos, com o objetivo de eliminar atributos redundantes e irrelevantes. Esse processo visa escolher os melhores atributos de forma a melhorar a performance dos algoritmos, simplificar os modelos de predição e diminuir o custo computacional. Foram escolhidos e aplicados os métodos *Recursive Feature Elimination*⁹ (RFE), *Feature*

⁹ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE

Figura 21 – Preparação dos Dados



Fonte: O Autor

*Importance*¹⁰ e *SelectKBest*¹¹.

O método RFE funciona removendo recursivamente os atributos e construindo um modelo nos atributos que permanecem. Ele utiliza a precisão do modelo para identificar quais atributos (e combinação de atributos) contribuem mais para prever o atributo classe. Diferentes modelos podem ser aplicados com este método (GRANITTO et al., 2006) (BEDO; SANDERSON; KOWALCZYK, 2006). Neste trabalho utilizou-se o RFE com o algoritmo de regressão logística para selecionar os principais atributos.

Já o método *Feature Importance* pode ser aplicado através de algoritmos como *Random Forest* e *Extra Trees* (SUGUMARAN; MURALIDHARAN; RAMACHANDRAN, 2007). Este método atribui uma pontuação de importância para cada atributo, sendo que quanto maior for a pontuação, mais importante é aquele atributo. Neste trabalho utilizou-se o método com o algoritmo *Extra Trees*.

Outro método testado foi o *SelectKBest*. Neste método, testes estatísticos podem ser usados para selecionar os atributos que têm a relação mais forte com a variável de saída. Os resultados dos métodos que obtiveram o melhor desempenho e quais atributos foram escolhidos em cada dimensão, são apresentados na próxima seção.

¹⁰ https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances

¹¹ https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest

6.4 Formatação dos Dados

Nesta etapa, cada um dos oito *datasets* criados, foram reorganizados de acordo com os atributos selecionados pelos métodos na subseção anterior, para que possam ser interpretados na próxima etapa (TABELA 13). Nas dimensões de processamento, captação e organização, o método que fez a melhor seleção foi o RFE. Já na dimensão percepção o método escolhido foi o *Feature Importance*.

Tabela 13 – *Datasets* Formatados após Seleção de Atributos

Processamento-Literatura (Ativo/Reflexivo)	Percepção-Literatura (Sensorial/Intuitivo)	Captação-Literatura (Visual/Verbal)	Organização-Literatura (Sequencial/Global)
add_gloss	view_tarefa	add_gloss	add_gloss
arq_env_tarefa	view_forum	arq_env_tarefa	arq_env_tarefa
attempt_quiz	view_conteudo	review_quiz	view_tarefa
		view_visaogeral	
Processamento-LV (Ativo/Reflexivo)	Percepção-LV (Sensorial/Intuitivo)	Captação-LV (Visual/Verbal)	Organização-LV (Sequencial/Global)
fator_beta_tarefalv	fator_beta_disciplina	fator_beta_wikilv	fator_beta_forumlv
fator_beta_quizlv	num_avaliacao_muitobom	fator_beta_discip	num_avaliacao_regular
num_avaliacao_bom	num_avaliacao_bom	num_avaliacao_muitobom	num_avaliacao_naosatisfatorio
	num_avaliacao_fraco		

Fonte: O Autor

Após essa reorganização, todos os *datasets* passaram por um processo de balanceamento de dados, de forma a conseguir conjuntos de dados equilibrados e assim poder executar os algoritmos de classificação de forma mais eficiente. O desbalanceamento ocorre quando o número de observações pertencentes a uma classe é significativamente menor do que as pertencentes às outras classes. Nesse caso, o algoritmo de classificação pode acabar prevendo de forma tendenciosa e imprecisa.

Para resolver esse problema de desequilíbrio das classes aplicou-se uma técnica de reamostragem do tipo *Over-Sampling*, ou sobreamostragem, no qual aumenta-se o número de amostras na classe minoritária. O método de balanceamento aplicado foi o *Synthetic Minority Over-sampling Technique as presented* (SMOTE).¹²

Este é o método de *oversampling* mais popular que foi proposto para melhorar a sobreamostragem aleatória (LAST; DOUZAS; BAÇÃO, 2017). Ele usa um determinado número de vizinhos mais próximos para construir amostras sintéticas. Neste trabalho o número de vizinhos utilizados foi cinco. Esta técnica foi escolhida devido a pouca quantidade de dados para os experimentos, já que esse método não leva a perda de informação. Diversos trabalhos recentes utilizaram essa técnica em suas pesquisas, validando sua eficácia (GE; YUE; CHEN, 2017), (AHSAN; GOMES; DENTON, 2018), (ZENG et al., 2016).

¹² https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.over_sampling.SMOTE

6.5 Modelagem

Nesta fase define-se as técnicas de modelagem de dados, especificamente um conjunto de algoritmos de classificação, bem como seus ajustes de parâmetros. A classificação é uma tarefa preditiva, que define o valor de uma variável desconhecida (atributo classe) a partir de variáveis conhecidas.

Esta fase teve como objetivo prever os perfis de aprendizagem dos alunos, por meio de técnicas de aprendizado supervisionado, utilizando as variáveis comportamentais coletadas do AVA como preditoras dos estilos de aprendizagem dos alunos. O atributo classe equivale ao perfil do aluno que foi validado através da resposta do questionário ILS aplicado aos mesmos.

A base de treinamento e teste foi dividida através da técnica de validação cruzada (*Cross-Validation*) *Stratified K-Folds*¹³, com $k=5$. O subconjunto de teste equivale a 20% dos dados e o conjunto de treino equivale a 80%. Esta é uma variação do método *k-fold*, sendo que é preservada a porcentagem de amostras para cada classe. O valor k corresponde aos subconjuntos formados que são mutuamente exclusivos e de tamanhos semelhantes. A partir dessa divisão, um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para treino. Este processo é realizado k vezes alternando de forma circular e no final obtém-se o desempenho médio estimado a partir dos subconjuntos (*folds*).

Os classificadores: KNN, SVM, MLP, *MultinomialNB*, *Decision Tree*, foram implementados. Adotou-se diferentes tipos de classificadores, citados no Capítulo 4, com o objetivo de encontrar o melhor modelo para predição dos estilos de aprendizagem dos alunos. Esses algoritmos foram escolhidos com base na Revisão Sistemática da Literatura (Capítulo 5). Para cada classificador utilizou-se um total de trinta execuções.

De forma a melhorar a eficiência dos algoritmos, a técnica de Busca em Grade ou *Grid Search* foi aplicada nos classificadores KNN e SVM para escolha dos melhores parâmetros. O Classificador KNN foi parametrizado para usar valores de k variando de 1 a 15, de forma a encontrar o valor de k ideal. Já no SVM, foram variados os valores de C (1,10,20,30,40,50,60,70,80,90,100), de γ (1,0.1,0.001,0.0001) e tipo de kernel *Radial Basis Function* (RBF) ou Linear. Além desses classificadores, o *Decision Tree*, o *MultinomialNB* e *Perceptron* Multicamadas (MLP) também foram avaliados. O parâmetro para otimização de peso do algoritmo MLP foi setado para o valor *Limited-memory Broyden-Fletcher-Goldfarb-Shanno* (L-BFGS), pois este método é indicado para conjuntos de dados pequenos, já que os dados podem convergir mais rapidamente e ter um melhor desempenho (PEDREGOSA et al., 2011). BFGS de memória limitada ou L-BFGS é um algoritmo de otimização na família de métodos quase-Newton que aproxima o algoritmo de *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) usando uma quantidade limitada de memória

¹³ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold

de computador. Para os outros valores, utilizou-se parâmetros padrões.

Toda a fase de modelagem da classificação foi realizada por meio da construção de algoritmos utilizando a linguagem de programação *Python* com auxílio da biblioteca *scikit-learn* (PEDREGOSA et al., 2011). Além destas, a biblioteca de análise de dados *Pandas*¹⁴ foi utilizada para realizar a leitura e separação dos atributos (X) das classes (Y). O Apêndice B contém todos os *scripts* desenvolvidos para a construção dos modelos de classificação desta dissertação.

6.6 Interpretação e Avaliação

Para avaliar o desempenho dos classificadores bem como comparar a diferença entre as bases de dados, foram utilizadas as seguintes métricas de avaliação: acurácia, precisão, *recall*, F1, desvio padrão e AUC (*Area Under the ROC Curve*). De forma a entender cada uma dessas métricas, as seguintes terminologias foram usadas (RODRIGUES, 2016):

- Verdadeiro Positivo (VP): é relacionado ao número de instâncias positivas previstas corretamente;
- Verdadeiro Negativo (VN): é o número de exemplos negativos previstos corretamente pelo classificador;
- Falso Positivo (FP): é o número de exemplos negativos previstos erroneamente como positivos pelo classificador;
- Falso Negativo (FN): é o número de instâncias previstas erroneamente como negativos pelo classificador;

A partir destas terminologias, pode-se definir cada uma das métricas utilizadas neste trabalho:

- **Acurácia**, apresentada na Equação (6.1), é definida como sendo o número de instâncias corretas divididas pelo número total de instâncias (HALAWA; HAMED; SHEHAB, 2015);

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (6.1)$$

- **Precisão**, conforme indicado na Equação (6.2), determina o percentual de registros que são positivos no grupo que o classificador previu como classe positiva, ou seja, número total de verdadeiros positivos, dividido pelo número total de verdadeiros positivos mais os falsos positivos (KUMAR, 2013).

$$Precisao = \frac{VP}{VP + FP} \quad (6.2)$$

¹⁴ <http://pandas.pydata.org>

- **Recall**, como visto na Equação (6.3), é calculado através do número total de verdadeiros positivos, dividido pelo número total de verdadeiros positivos mais os falsos negativos (KUMAR, 2013);

$$Recall = \frac{VP}{VP + FN} \quad (6.3)$$

- **F1**, representa a média harmônica entre precisão e *recall* (HALAWA; HAMED; SHEHAB, 2015).

$$F1 = \frac{2 * Recall * Precisao}{Recall + Precisao} \quad (6.4)$$

A acurácia é a métrica mais simples, pois calcula-se apenas o número de acertos positivos dividido pelo número total de exemplos. Ela deve ser usada em *datasets* com a mesma proporção de exemplos para cada classe, pois em dados desbalanceados pode mostrar uma falsa impressão de bom desempenho. A precisão é uma boa medida para determinar quando os custos do Falso Positivo são altos. Já o *recall*, é a métrica que se utiliza para selecionar o melhor modelo quando há um alto custo associado ao Falso Negativo.

Outra métrica avaliada foi AUC. Nela, mede-se a área sob uma curva, denominada de curva ROC (*Receiver Operating Characteristics*), formada entre a taxa de exemplos positivos, que realmente são positivos, e a taxa de falsos positivos, ou seja, qual a probabilidade de o modelo classificar um exemplo positivo aleatório mais do que um exemplo negativo aleatório.

Essa métrica informa quanto o modelo é capaz de distinguir entre classes. Quanto maior a AUC, melhor o modelo em prever. O desvio padrão das acurácias nas trinta execuções também foi levado em consideração, para calcular a dispersão em torno da média encontrada. Ele serviu para expressar o quanto o conjunto de acurácias encontradas se desviou da média.

As métricas analisadas serviram para realizar a escolha do modelo de classificador que melhor se ajustou aos dados. Este capítulo dedicou-se à descrição da metodologia aplicada. No próximo capítulo são apresentados os resultados encontrados e as discussões a respeito dos experimentos realizados.

7 Resultados e Discussões

Neste capítulo, apresentam-se os resultados dos métodos expostos no capítulo anterior. São abordados os resultados referentes à construção dos modelos preditivos e suas respectivas validações.

7.1 Resultados sobre o entendimento do domínio

A primeira fase da realização desta dissertação, de acordo com a metodologia KDD (Capítulo 6), foi o entendimento do domínio. Esta fase foi realizada com o objetivo de compreender os estilos de aprendizagem dos alunos na modalidade de educação a distância. Em seguida, para ter uma visão exploratória sobre os perfis dos alunos, foi aplicado o instrumento de coleta ILS (*Index Learning Style*)(FELDER; SOLOMAN, 1997).

A distribuição das respostas coletadas no questionário mostra que mais da metade dos perfis, em cada uma das quatro dimensões, foram identificados como neutro (TABELA 14). Na dimensão processamento foram 70 alunos, já na dimensão percepção e captação, 68 alunos foram identificados como sendo do perfil neutro e por último, na dimensão organização, foram 67 alunos.

Tabela 14 – Distribuição dos Estilos de Aprendizagem por Dimensão

Dimensão	Categoria	Porcentagem
Processamento	Ativo	24,1%
	Reflexivo	13,4%
	Neutro	62,5%
Percepção	Sensorial	35,7%
	Intuitivo	3,6%
	Neutro	60,7%
Captação	Visual	22,3%
	Verbal	17,0%
	Neutro	60,7%
Organização	Sequencial	33,0%
	Global	7,1%
	Neutro	59,8%

Fonte: O Autor

O perfil neutro é um perfil que possui uma preferência leve/equilibrada em ambas as categorias da dimensão, ou seja, este tipo de perfil não necessita de um ambiente

personalizado de acordo com suas preferências, já que ele se adapta bem em ambos os ambientes.

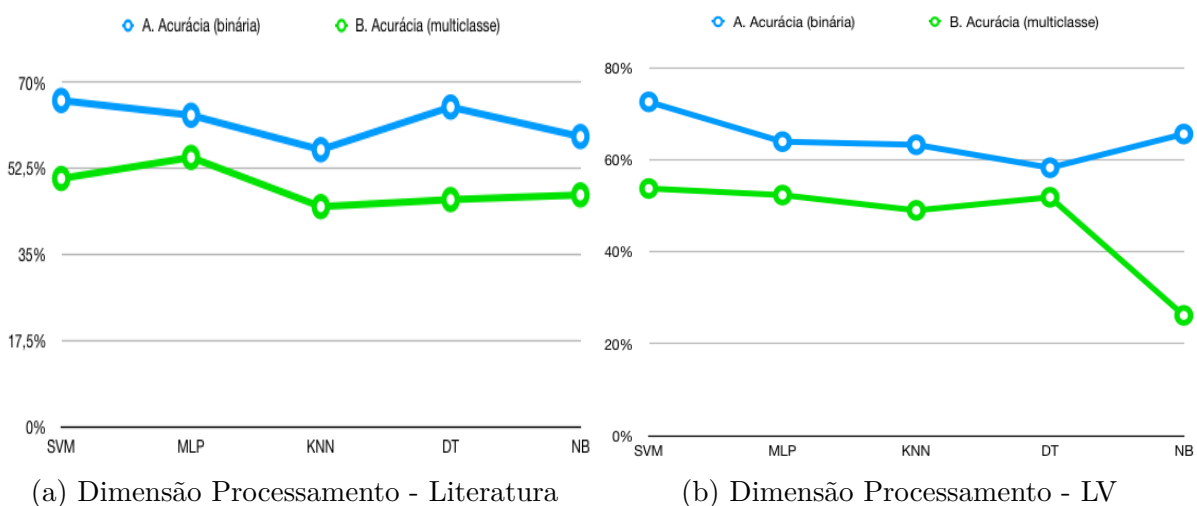
7.2 Construção dos modelos de classificação

7.2.1 Categorização da Variável Alvo

De forma a escolher a melhor maneira de classificar os dados, foi feito um experimento com a variável alvo categorizada de duas formas distintas: (A) Binária, (B) Multiclasse. Utilizou-se a dimensão processamento do modelo FSLSM para este experimento. Inicialmente executou-se a classificação binária. Foram consideradas apenas duas categorias da dimensão, com preferências fortes ou moderadas de Estilo de Aprendizagem. Já no segundo experimento, foram consideradas três classes, em que o perfil neutro também foi categorizado.

Os experimentos foram feitos com os cinco algoritmos de classificação selecionados nesta pesquisa (SVM, MLP, KNN, DT, NB), utilizando os parâmetros padrões para cada modelo (FIGURA 22). O objetivo desta fase foi entender o melhor limiar de categorização da variável alvo de acordo com o resultado da acurácia para cada um dos classificadores.

Figura 22 – Comparação da categorização da variável alvo



Fonte: O Autor

Percebe-se que a variável alvo categorizada de forma multiclasse obteve desempenho inferior em relação à classificação binária em ambos os *datasets* (FIGURA 22). Os resultados das acurácias na base de dados com atributos do Modelo LV apresentaram valores melhores do que na base com atributos da literatura. A categorização B(multiclasse), apresentou acurácias próximas de 50%, mostrando assim um desempenho muito baixo.

De forma a prosseguir com as análises, ficou estabelecido que as próximas fases do processo de modelagem teriam como padrão a categorização A, ou seja, que a variável alvo seria categorizada de forma binária, não levando em consideração os perfis do tipo neutro.

7.2.2 Balanceamento dos Dados

O desempenho dos classificadores propostos foi avaliado em oito conjuntos de dados reais. Os detalhes sobre cada conjunto de dados são apresentados de forma a distingui-los melhor (TABELA 15).

Tabela 15 – Características *Datasets*

<i>Datasets</i>	Atributos	Instâncias	Taxa de Desbalanceamento	Instâncias Balanceadas
Processamento-Literatura	3	42	0,3571	54
Processamento-LV	3	42	0,3571	54
Percepção-Literatura	3	44	0,0909	80
Percepção-LV	4	44	0,0909	80
Captação-Literatura	4	44	0,4318	50
Captação-LV	3	44	0,4318	50
Organização-Literatura	3	45	0,1777	74
Organização-LV	3	45	0,1777	74

Fonte: O Autor

A penúltima coluna da TABELA 15 apresenta a taxa de desbalanceamento ou razão de desequilíbrio. Essa taxa mede o grau de desequilíbrio de cada conjunto de dados e é definida pelo número de exemplos na classe minoritária dividido pelo número de exemplos da classe majoritária. Assim, quanto mais distante do valor 1, menos balanceados estão os dados.

Como explicado no Capítulo 6, os dados passaram pela técnica de balanceamento SMOTE, aumentando o número de instâncias da classe minoritária, de forma a equilibrar ambas as classes. A última coluna da TABELA 15 informa a quantidade de instâncias após o balanceamento.

Na dimensão processamento, por exemplo, 27 instâncias eram do perfil ativo e 15 do perfil reflexivo, totalizando 42 instâncias. Através do balanceamento, criou-se mais 12 amostras sintéticas do perfil reflexivo de forma a equilibrar as classes, totalizando assim 54 instâncias. Os *datasets* com menos dados são os que representam a dimensão Captação e Processamento do modelo FSLSM.

7.3 Validação dos modelos de classificação

Os parâmetros que obtiveram a melhor acurácia (APÊNDICE C) foram escolhidos para serem utilizados nos experimentos. Para todos os classificadores implementados, utilizou-se validação cruzada, com $k=5$ e um total de 30 execuções. A validação foi realizada a partir dos valores de Acurácia, Desvio Padrão, Precisão, *Recall*, F1 e AUC. Os resultados das classificações nas bases de dados das quatro dimensões utilizando atributos da Literatura, são apresentados na TABELA 16.

Tabela 16 – Resultados Classificação Atributos Literatura

Dimensão Processamento						
Método	Acurácia	Desvio Padrão	Precisão	<i>Recall</i>	F1	AUC
KNN	62,33%	0,1218	80,00%	28,66%	41,42%	64,82%
NB	59,00%	0,0533	67,77%	46,00%	49,04%	67,58%
MLP	64,51%	0,0876	78,32%	54,98%	58,86%	78,92%
SVM	69,99%	0,0988	86,66%	50,66%	60,99%	66,64%
DT	66,05%	0,0757	77,83%	55,89%	59,18%	76,50%
Dimensão Percepção						
Método	Acurácia	Desvio Padrão	Precisão	<i>Recall</i>	F1	AUC
KNN	86,25%	0,1075	81,64%	97,50%	88,30%	89,68%
NB	76,25%	0,0729	72,62%	85,00%	78,10%	81,56%
MLP	89,87%	0,0615	86,31%	97,33%	90,67%	90,91%
SVM	92,50%	0,0467	95,27%	89,99%	92,13%	95,31%
DT	91,62%	0,0769	88,91%	95,00%	91,55%	91,30%
Dimensão Captação						
Método	Acurácia	Desvio Padrão	Precisão	<i>Recall</i>	F1	AUC
KNN	61,99%	0,2039	61,96%	76,00%	66,73%	53,20%
NB	49,99%	0,1673	57,00%	40,00%	45,14%	57,60%
MLP	55,20%	0,1329	56,94%	64,93%	58,79%	58,16%
SVM	61,99%	0,0748	65,92%	67,99%	63,05%	63,20%
DT	55,99%	0,1414	52,27%	40,00%	39,98%	65,48%
Dimensão Organização						
Método	Acurácia	Desvio Padrão	Precisão	<i>Recall</i>	F1	AUC
KNN	72,85%	0,1270	71,03%	72,50%	71,03%	72,85%
NB	70,18%	0,1213	66,67%	80,71%	72,84%	72,24%
MLP	81,80%	0,1146	81,88%	82,33%	81,38%	83,60%
SVM	78,57%	0,1541	81,58%	75,71%	78,34%	82,23%
DT	83,65%	0,1105	82,92%	84,28%	83,50%	81,77%

Fonte: O Autor

Na dimensão Processamento, que classifica o aluno em ativo ou reflexivo, a melhor acurácia, precisão e F1 foram do classificador SVM, com valores de 69,99%, 86,66% e 60,99%, respectivamente. Este classificador foi executado utilizando os melhores parâmetros encontrados na busca em grade (kernel=RBF, C=60 e gamma=0,1).

Em segundo lugar, o modelo DT obteve uma acurácia de 66,05% e desvio padrão de 0,0757. O melhor valor da métrica AUC foi do classificador MLP, com valor de 78,92%. Já o pior resultado de acurácia foi do algoritmo NB, que obteve apenas 59,00% de acerto, porém com menor desvio padrão, no valor de 0,0533.

Na dimensão Percepção, onde o aluno pode ser classificado em sensitivo ou intuitivo, o SVM também foi o modelo com a melhor acurácia, precisão e F1, além de ter tido os melhores valores para desvio padrão e AUC. O valor da acurácia desse modelo foi de 92,50%, com desvio padrão de 0,0467.

Já os valores da Precisão, F1 e AUC foram, respectivamente: 95,27%, 92,13% e 95,31%. Os parâmetros utilizados neste classificador foram com valores de kernel=RBF, C=1 e gamma=0,1. O modelo DT também obteve o segundo lugar nesta base de dados, com uma acurácia de 91,62% e desvio padrão de 0,0769. O melhor valor da métrica *recall* foi do classificador KNN, com valor de 97,50%. Já os piores resultados de acurácia, precisão, *recall*, F1 e AUC foram do algoritmo NB.

A terceira dimensão avaliada foi a dimensão Captação, na qual o aluno pode ser classificado em visual ou verbal. Nesta dimensão o SVM e o KNN tiveram os melhores valores de acurácias, que foi de 61,99% em ambos classificadores. Já o melhor desvio padrão e precisão foram do algoritmo SVM, com valores de 0,0748 e 65,92%, respectivamente.

O KNN também foi o modelo com o melhor *recall* e F1, com valores de 76,00% e 66,73%, respectivamente. Os parâmetros utilizados no SVM tiveram os seguintes valores: kernel=linear, C=1 e gamma=1. Já o valor de K no algoritmo KNN foi igual a 5.

Por último, a dimensão Organização foi avaliada. Esta dimensão classifica os alunos em sequencial ou global. O algoritmo que obteve os melhores resultados de acurácia, desvio padrão, precisão, *recall* e F1 foi o DT, com valores de 83,65%, 0,1105, 82,92%, 84,28%, 83,50%, respectivamente. Já o melhor valor da métrica AUC foi de 83,60%, do algoritmo MLP, que obteve 81,80% de acurácia.

As outras quatro bases que foram analisadas são referentes aos atributos do Modelo LV (TABELA 17). Cada uma das dimensões foi avaliada, considerando todas as métricas de avaliação selecionadas, de forma a comparar os resultados dos classificadores e das bases posteriormente.

A primeira dimensão avaliada foi a dimensão Processamento (TABELA 17). Nesta dimensão o classificador SVM obteve os melhores valores em todas as métricas, com exceção da métrica precisão e AUC. Os valores obtidos foram: 72,33%, 0,1254, 73,99%, 73,30%. Esses valores equivalem, respectivamente, aos valores de acurácia, desvio padrão, *recall* e F1.

Já o algoritmo KNN, obteve a segunda melhor acurácia, com valor de 71,00%, a melhor precisão, com valor de 73,62% e melhor AUC (77,34%). Os parâmetros utilizados

no SVM foram, kernel=RBF, C=1, gamma=1. O valor de K escolhido no KNN através da busca em grade, foi igual a 3. Os piores resultados foram do algoritmo DT, que obteve acurácia de apenas 58,51% de acerto.

Na dimensão Percepção, com atributos do Modelo LV (TABELA 17), o SVM também obteve os melhores valores de acurácia (92,50%), precisão (91,11%) e F1 (92,94%). Este classificador foi parametrizado da seguinte forma: kernel=RBF, C=30, gamma=0,1. Em segundo lugar, os algoritmos MLP e DT obtiveram valores muito próximos de acurácia: 91,16% e 91,17%, respectivamente. Porém, o melhor desvio padrão (0,0584), *recall* (100,00%) e AUC (93,50%) foram do classificador MLP.

Tabela 17 – Resultados Classificação Atributos LV

Dimensão Processamento						
Método	Acurácia	Desvio Padrão	Precisão	Recall	F1	AUC
KNN	71,00%	0,1428	73,62%	70,67%	70,42%	77,34%
NB	65,67%	0,1562	71,09%	71,33%	66,49%	68,07%
MLP	66,71%	0,1526	69,21%	72,38%	69,50%	72,43%
SVM	72,33%	0,1254	73,43%	73,99%	73,30%	74,84%
DT	58,51%	0,2099	62,80%	58,67%	60,33%	63,49%
Dimensão Percepção						
Método	Acurácia	Desvio Padrão	Precisão	Recall	F1	AUC
KNN	87,50%	0,0968	82,66%	94,99%	88,36%	88,28%
NB	80,00%	0,0918	80,95%	82,50%	80,99%	86,25%
MLP	91,16%	0,0584	86,38%	100,00%	92,40%	93,50%
SVM	92,50%	0,1211	91,11%	94,99%	92,94%	91,25%
DT	91,17%	0,0737	88,17%	97,50%	91,61%	91,29%
Dimensão Captação						
Método	Acurácia	Desvio Padrão	Precisão	Recall	F1	AUC
KNN	69,99%	0,0894	71,87%	80,00%	73,16%	69,99%
NB	66,00%	0,1497	70,28%	64,00%	64,09%	67,20%
MLP	71,47%	0,1647	71,24%	76,00%	72,34%	72,99%
SVM	73,99%	0,1019	78,93%	76,00%	74,83%	80,00%
DT	63,00%	0,2180	64,48%	63,60%	65,37%	65,73%
Dimensão Organização						
Método	Acurácia	Desvio Padrão	Precisão	Recall	F1	AUC
KNN	75,53%	0,1173	70,49%	88,93%	78,56%	79,08%
NB	58,03%	0,0505	54,91%	91,78%	68,61%	81,00%
MLP	80,16%	0,0924	79,77%	87,32%	81,10%	83,63%
SVM	85,18%	0,0810	80,55%	94,64%	86,72%	87,36%
DT	73,07%	0,0760	78,32%	69,95%	71,64%	79,82%

Fonte: O Autor

A terceira dimensão avaliada foi a dimensão Captação. Nesta dimensão o SVM e o MLP tiveram os melhores valores de acurácias, que foram de 73,99% e 71,47%, respectivamente (TABELA 17). O SVM também obteve os melhores valores nas métricas

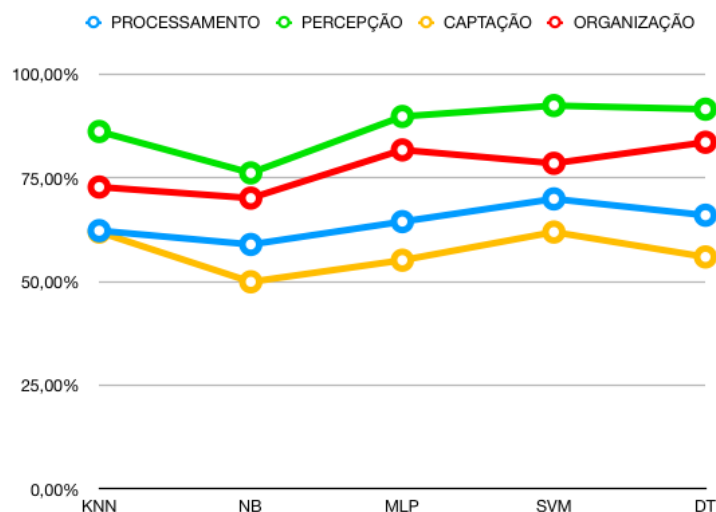
de precisão (78,93%), F1 (74,83%) e AUC (80,00%). Os valores dos parâmetros utilizados no SVM foram: kernel=RBF, C=80 e gamma=1. Já o classificador KNN obteve o menor desvio padrão (0,0894), com acurácia de 69,99% e o melhor *recall* (80,00%). O valor de K no KNN foi igual a 1. O pior desempenho foi do algoritmo DT, com acurácia de 63,00%.

Por último, a dimensão Organização foi avaliada. Os melhores valores de acurácia (85,18%), precisão (80,55%), *recall*(94,64%), F1 (86,72%) e AUC (87,36%), foram do classificador SVM (TABELA 17). Este algoritmo atingiu os melhores resultados em todas as métricas, com exceção do desvio padrão. Os valores dos parâmetros utilizados no SVM foram: kernel=RBF, C=40 e gamma=1. A segunda melhor acurácia foi do modelo MLP, que obteve 80,16% de acerto. Já a pior acurácia foi do classificador NB, com apenas 58,03% de acerto.

7.4 Gráficos de Comparação dos Resultados da Validação

Foi feita uma comparação dos resultados das acurácias de cada classificador. Inicialmente avaliou-se as bases de dados com atributos da literatura, em cada uma das dimensões (FIGURA 23). Observa-se que o SVM atingiu os melhores resultados em todas as dimensões, com exceção da dimensão Organização, onde a melhor acurácia foi do DT (83,65%).

Figura 23 – Resultados dos Algoritmos de Classificação - Atributos da Literatura



Fonte: O Autor

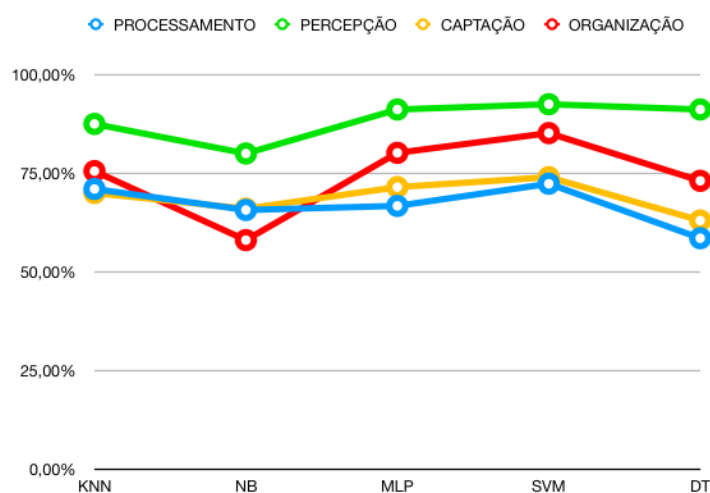
Já o algoritmo NB foi o que obteve as menores acurácias em todas as dimensões analisadas, mostrando desta forma que não é um modelo eficiente para a identificação de perfis de aprendizagem. A média de acurácias nas quatro dimensões, dos classificadores

KNN, NB, MLP, SVM e DT, nas bases de dados utilizando atributos da literatura, são respectivamente: 70,86%, 63,86%, 72,85%, 75,76% e 74,33%.

A dimensão Percepção foi a que obteve melhores acurácias em todos os classificadores avaliados. Já a dimensão Captação foi a que atingiu os menores valores. Este baixo desempenho pode ter sido ocasionado pelo fato de não terem tido atributos suficientes para identificar esta dimensão, visto que no curso analisado não tinham recursos educacionais relacionados à categoria visual desta dimensão, como vídeos e imagens.

Os resultados nas bases de dados utilizando atributos do Modelo LV também foram comparados (FIGURA 24). Observa-se que os algoritmos SVM e MLP obtiveram as melhores acurácias nas quatro dimensões analisadas. O SVM atingiu uma média de acurácias de 81,00% e o MLP 77,38% nas quatro dimensões.

Figura 24 – Resultados dos Algoritmos de Classificação - Atributos LV



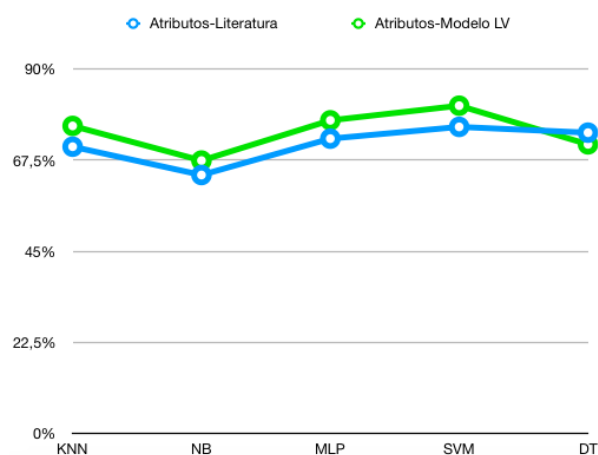
Fonte: O Autor

A dimensão com os melhores resultados, também foi a dimensão Percepção, porém as dimensões Processamento e Captação atingiram acurácias melhores, comparando-se com os resultados dos atributos da literatura (FIGURA 23). As acurácias nestas dimensões ficaram próximas de 75,00% no classificador SVM, usando atributos do modelo LV (FIGURA 24), já nas bases de dados dos atributos da literatura, ficaram entre 60,00% e 70,00%.

O classificador NB também não atingiu resultados satisfatórios com atributos do modelo LV, ficando com uma média de acurácias igual a 67,43%. Já os outros classificadores analisados (KNN, DT), tiveram valores aproximados, com uma média de acurácia de 76,00% e 71,44%, respectivamente.

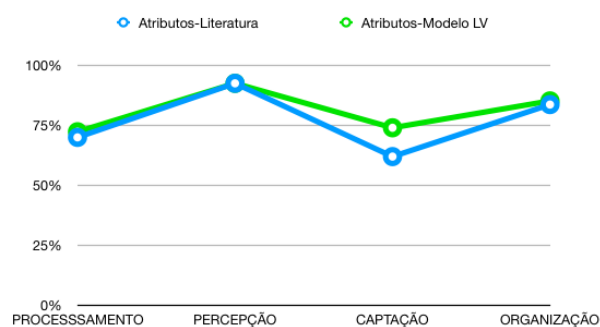
Pode-se observar que as médias das acurácias das quatro dimensões de cada conjunto de dados foram aproximadas, porém os experimentos realizados com as bases de dados contendo os atributos do Modelo LV, obtiveram desempenho superior em todos os classificadores, com exceção do DT (FIGURA 25). Isto indica a possibilidade dos atributos do Modelo LV em identificar os perfis de aprendizagem no modelo FSLSM.

Figura 25 – Médias das Acurácias Por Classificador



Fonte: O Autor

Figura 26 – Melhores Acurácias por Dimensão



Fonte: O Autor

Observando as melhores acurácias em cada uma das dimensões, percebe-se que os valores nas bases de dados com os atributos do Modelo LV e com atributos da Literatura, obtiveram resultados semelhantes (FIGURA 26). Nas dimensões Processamento, Captação e Organização, as acurácias das modelagens nas bases com atributos do Modelo LV se sobressaíram.

Nestas três dimensões, o modelo SVM foi o que obteve as melhores acurácias, sendo que o modelo KNN também atingiu o mesmo valor de acurácia na dimensão Captação, na base com atributos da Literatura. Já na dimensão Organização, o classificador DT foi o que conseguiu o melhor desempenho na base de dados com atributos da Literatura. Por último, na dimensão Percepção, o resultado da melhor acurácia foi igual (92,50%) em ambas as bases de dados, também com o classificador SVM.

Este capítulo buscou realizar a validação e o comparativo entre os resultados dos modelos de classificação que melhor se ajustassem às variáveis em estudo. Os resultados dos algoritmos foram apresentados considerando diferentes bases de dados, algumas com atributos da literatura e outras com atributos do Modelo LV, de forma a classificar os estilos de aprendizagem dos alunos. O classificador SVM apresentou-se como sendo o mais adequado dentre os cinco modelos avaliados.

A metodologia aplicada obteve avaliações positivas. Os resultados alcançados na fase de modelagem mostraram-se satisfatórios, indicando que o método proposto com base no Modelo LV, atinge valores compatíveis e até melhores que os métodos tradicionais da literatura, fornecendo assim uma alternativa válida para tal abordagem. Sendo assim, o Modelo LV também pode ser usado na identificação de perfis de aprendizagem, através de poucos atributos, que podem ser extraídos da tabela de *logs* do ambiente virtual com o *plugin* LV instalado.

No próximo capítulo são abordadas as conclusões deste trabalho, indicando as considerações finais, as limitações encontradas e os trabalhos futuros.

8 Conclusão

O último capítulo desta pesquisa apresenta as considerações finais, dificuldade e limitações, contribuições e trabalhos futuros. Neste trabalho buscou-se identificar os estilos de aprendizagem dos alunos com base nos atributos extraídos do banco de dados do *Moodle*, a partir de variáveis da literatura e do Modelo LV, utilizando diversos algoritmos de classificação. Buscou-se responder à seguinte questão de pesquisa: Os indicadores qualitativos gerados através do Modelo LV contribuem na identificação dos estilos de aprendizagem?

Para tal, os atributos referentes a este modelo foram extraídos da base de dados do AVA e em seguida modelados em diferentes técnicas de classificação, de forma a responder esta questão. Cinco algoritmos de classificação foram treinados e avaliados de forma a comparar o desempenho de cada um em classificar os estilos de aprendizagem utilizando atributos do Modelo LV e atributos usados na literatura.

Durante as fases e procedimentos desta dissertação, buscou-se desenvolver um modelo que permitisse realizar a previsão do Estilo de Aprendizagem de forma automática. Os achados desta pesquisa contribuíram de forma satisfatória, ao utilizar os atributos do Modelo LV na identificação dos perfis de aprendizagem. Dessa forma pode-se facilitar o acompanhamento personalizado do aluno, de acordo com suas preferências, seja na adaptação de ambientes virtuais de aprendizagem ou na recomendação de conteúdos educacionais personalizados.

8.1 Dificuldades e Limitações Encontradas

Uma das principais dificuldades, durante o desenvolvimento deste trabalho, foi a extração das bases de dados utilizadas na fase de modelagem. A complexidade da arquitetura do banco de dados *Moodle*, tornou a extração das bases de *logs* comportamentais uma das fases com maior dedicação de tempo e esforço. Outra limitação encontrada diz respeito a dependência de *logs* no ambiente virtual de aprendizagem. Alguns comportamentos, como quantidade de acesso em conteúdos gráficos, sequência de navegação nos conteúdos, tempo no fórum, considerados importantes para identificação do perfil, não são registrados no banco de dados, impossibilitando assim sua análise.

A quantidade de 112 alunos que responderam ao questionário ILS, também foi uma das limitações desta pesquisa. Isto ocasionou uma baixa quantidade de dados utilizados para os experimentos, o que pode ter comprometido os resultados. Também encontrou-se limitação quanto à quantidade e diversidade dos recursos educacionais disponibilizados

no ambiente virtual, dificultando a exploração dos comportamentos em diferentes perfis. Variáveis na dimensão captação, por exemplo, foram pouco exploradas, já que a disciplina não abordava atividades ou materiais para o perfil visual (vídeos, gráficos, imagens).

8.2 Contribuições

Com o desenvolvimento desta pesquisa, surgiram algumas contribuições de acordo com a execução da metodologia proposta. Algumas contribuições que merecem destaque são: (1) construção e disponibilização dos *scripts* para extração de variáveis em plataformas Moodle (APÊNDICE A); (2) geração de um conjunto de *scripts* em linguagem *Python* para construção dos classificadores (APÊNDICE B); e (3) construção de uma base de dados, composta de dados reais de alunos, relacionada a comportamentos que ajudam a identificar estilos de aprendizagem.

Outra contribuição foi a implementação de um sistema *web*, contendo o questionário ILS automatizado, onde as respostas são armazenadas e calculadas automaticamente, sem necessitar da intervenção de quem está respondendo as perguntas. Este formulário poderá contribuir com outras pesquisas que também utilizem o modelo FSLSM como objeto de estudo e está disponível na plataforma *GitHub*, conforme informado no Capítulo 6.

A principal contribuição foi no uso dos atributos gerados pelo Modelo LV para a identificação dos perfis de aprendizagem dos alunos. Estas métricas LV, são registradas na tabela de *logs* do banco de dados do AVA e mais fáceis de extrair, do que alguns outros atributos da literatura. A pesquisa contribui para o estado da arte, ao propor a identificação de estilos de aprendizagem utilizando a combinação do Modelo LV e do FSLSM, no processo de aprendizado de máquina com diferentes técnicas de classificação.

8.3 Trabalhos Futuros

Como possibilidades de trabalhos futuros, que podem ser explorados a partir desse estudo, tem-se:

- Validação do modelo de classificação com um maior número de dados;
- Avaliação do modelo em disciplinas customizadas com recursos educacionais diversos, que abordem todos os perfis;
- Criação de sistemas de aprendizagem adaptativos, de forma a criar ambientes personalizáveis de acordo com o perfil de cada aluno;
- Criação de um sistema de recomendação de objetos de aprendizagem, baseado no perfil do aluno;
- Realização da classificação multiclasse, incluindo as preferências neutras com abordagens mais precisas.

Referências

- ABED, A. B. de Educação a D. *Censo EAD.BR: Relatório Analítico da Aprendizagem a Distância no Brasil 2016*. Curitiba: InterSaberes, 2016. ISBN 978-85-5972-459-2. Citado na página 26.
- ABED, A. B. de Educação a D. *Censo EAD.BR: relatório analítico da aprendizagem a distância no Brasil 2017*. Curitiba: InterSaberes, 2018. ISBN 978-85-5972-806-4. Citado 2 vezes nas páginas 25 e 26.
- ADEL, N.; LATHAM, A.; CROCKETT, K. A. Towards socially intelligent automated tutors: Predicting learning style dimensions from conversational dialogue. In: *2016 Intl IEEE Conferences on Ubiquitous Intelligence Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld)*. [S.l.: s.n.], 2016. p. 315–320. Citado 5 vezes nas páginas 61, 62, 63, 65 e 66.
- AHSAN, M.; GOMES, R.; DENTON, A. Smote implementation on phishing data to enhance cybersecurity. In: *2018 IEEE International Conference on Electro/Information Technology (EIT)*. [S.l.: s.n.], 2018. p. 0531–0536. ISSN 2154-0373. Citado na página 75.
- ALBUQUERQUE, M. C. N. et al. Avaliação da presencialidade em um fórum lv utilizando lógica fuzzy. *XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, Recife- PE, 2017. Citado na página 48.
- BALASUBRAMANIAN, V.; ANOUNCIA, S. M. Learning style detection based on cognitive skills to support adaptive learning environment – a reinforcement approach. *Ain Shams Engineering Journal*, v. 9, n. 4, p. 895 – 907, 2018. ISSN 2090-4479. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2090447916300508>>. Citado na página 33.
- BATISTA, G.; MONARD, M.-C.; CAROLINA, M. Pré-processamento de dados em aprendizado de máquina supervisionado. 08 2018. Citado na página 50.
- BEDO, J.; SANDERSON, C.; KOWALCZYK, A. An efficient alternative to svm based recursive feature elimination with applications in natural language processing and bioinformatics. In: SATTAR, A.; KANG, B.-h. (Ed.). *AI 2006: Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 170–180. ISBN 978-3-540-49788-2. Citado na página 74.
- BINH, H. T.; DUY, B. T. Predicting students' performance based on learning style by using artificial neural networks. In: *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*. [S.l.: s.n.], 2017. p. 48–53. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.
- BUCKLEY, P.; DOYLE, E. Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. *Computers & Education*, v. 106, p. 43 – 55, 2017. ISSN 0360-1315. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S036013151630238X>>. Citado na página 33.

- CACAIS, M. G.; SALES, G. L. A collaborative system for corporate performance evaluation using gamification and the learning vectors model. In: *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*. [S.l.: s.n.], 2017. p. 484–489. Citado na página 48.
- COPPIN, B. *Inteligência Artificial*. [S.l.]: LTC, 2010. ISBN 9788521617297. Citado na página 49.
- DORÇA, F. A. *Uma abordagem estocástica baseada em aprendizagem por reforço para modelagem automática e dinâmica de estilos de aprendizagem de estudantes em sistemas adaptativos e inteligentes para educação a distância*. Tese (Doutorado) — Universidade Federal de Uberlândia, Uberlândia, 2012. Citado 3 vezes nas páginas 28, 35 e 36.
- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification (2Nd Edition)*. [S.l.]: Wiley-Interscience, 2000. ISBN 0471056693. Citado 5 vezes nas páginas 51, 52, 54, 55 e 64.
- DUNN, R.; DUNN, K. Learning style as a criterion for placement in alternative programs. *The Phi Delta Kappan*, Phi Delta Kappa International, v. 56, n. 4, p. 275–278, 1974. ISSN 00317217. Disponível em: <<http://www.jstor.org/stable/20297890>>. Citado 2 vezes nas páginas 27 e 41.
- EFRATI, V.; LIMONGELLI, C.; SCIARRONE, F. A data mining approach to the analysis of students' learning styles in an e-learning community: A case study. In: STEPHANIDIS, C.; ANTONA, M. (Ed.). *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge*. Cham: Springer International Publishing, 2014. p. 289–300. ISBN 978-3-319-07440-5. Citado 7 vezes nas páginas 37, 61, 62, 63, 64, 65 e 66.
- FASIHUDDIN, H.; SKINNER, G.; ATHAUDA, R. Using learning styles as a basis for creating adaptive open learning environments: An evaluation. *Int. J. Learn. Technol.*, Inderscience Publishers, Inderscience Publishers, Geneva, SWITZERLAND, v. 11, n. 3, p. 198–217, jan. 2016. ISSN 1477-8386. Disponível em: <<https://doi.org/10.1504/IJLT.2016.079034>>. Citado na página 30.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, p. 37–54, 1996. Citado na página 69.
- FELDER, R. M.; BRENT, R. Understanding student differences. *Journal of Engineering Education*, Blackwell Publishing Ltd, v. 94, n. 1, p. 57–72, 2005. ISSN 2168-9830. Disponível em: <<http://dx.doi.org/10.1002/j.2168-9830.2005.tb00829.x>>. Citado na página 29.
- FELDER, R. M.; SILVERMAN, L. K. Learning and teaching styles in engineering education. *ENGINEERING EDUCATION*, 1988. Citado 5 vezes nas páginas 27, 30, 33, 34 e 48.
- FELDER, R. M.; SOLOMAN, B. A. Index of learning styles questionnaire. North Carolina State University, 1997. Disponível em: <<http://www.engr.ncsu.edu/learningstyles/ilsweb.html>>. Citado 4 vezes nas páginas 35, 63, 70 e 79.
- FERRARI, J. R. et al. Psychometric properties of the revised grasha-riechmann student learning style scales. *Educational and Psychological Measurement*, v. 56, n. 1, p. 166–172,

1996. Disponível em: <<https://doi.org/10.1177/0013164496056001013>>. Citado na página 38.
- FERREIRA, L. D. *Técnicas de aprendizado de máquina aplicadas à identificação de perfis de aprendizado em um ambiente real de ensino*. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo, 2016. Citado 6 vezes nas páginas 27, 30, 37, 52, 54 e 55.
- FERREIRA, L. D. et al. Métodos de classificação aplicados à detecção automática de estilos de aprendizagem em um ambiente real de ensino. *Simpósio Brasileiro de Informática na Educação*, 2017. Citado na página 28.
- FILHO, A. A. T. *Análise da utilização de métodos ágeis no desenvolvimento de ambientes virtuais de aprendizagem: um estudo de caso do solar 2.0*. Dissertação (Dissertação de Mestrado) — Universidade Federal do Ceará, 2014. Citado na página 27.
- FIORIN, D. V. et al. Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. *Revista Brasileira de Ensino de Física*, Scielo, v. 33, p. 01 – 20, 03 2011. ISSN 1806-1117. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172011000100009&nrm=iso>. Citado na página 53.
- FLEMING, N. *Teaching and Learning Styles: VARK Strategies*. N.D. Fleming, 2006. Disponível em: <<https://books.google.com.br/books?id=SDTuXwAACAAJ>>. Citado na página 39.
- FLEMING, N.; MILLS, C. *The VARK Questionnaire*. 2018. Disponível em: <<http://vark-learn.com/the-vark-questionnaire/>>. Citado na página 39.
- FLEMING, N. D.; MILLS, C. Not another inventory, rather a catalyst for reflection. v. 11, 06 1992. Citado na página 39.
- GE, Y.; YUE, D.; CHEN, L. Prediction of wind turbine blades icing based on mbk-smote and random forest in imbalanced data set. In: *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. [S.l.: s.n.], 2017. p. 1–6. Citado na página 75.
- GIL, A. C. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas, 2002. ISBN 85-224-3169-8. Citado na página 70.
- GONÇALVES, A. J.; CUNHA, J. L. L.; SALES, G. L. Concepção do fórum per lv: avaliação formativa da aprendizagem. *Revista Novas Tecnologias na Educação*, 2016. ISSN 1679-1916. Citado na página 48.
- GRAF, S. *Adaptivity in Learning Management Systems Focussing on Learning Styles*. Tese (Doutorado) — Vienna University of Technology Faculty of Informatics, Vienna, 2007. Citado 4 vezes nas páginas 29, 35, 41 e 72.
- GRAF, S.; KINSHUK; IVES, C. A flexible mechanism for providing adaptivity based on learning styles in learning management systems. In: *2010 10th IEEE International Conference on Advanced Learning Technologies*. [S.l.: s.n.], 2010. p. 30–34. ISSN 2161-3761. Citado na página 30.
- GRAF, S.; KINSHUK; LIU, T. C. Identifying learning styles in learning management systems by using indications from students' behaviour. In: *2008 Eighth IEEE International Conference on Advanced Learning Technologies*. [S.l.: s.n.], 2008. p. 482–486. ISSN 2161-3761. Citado na página 33.

GRAF, S.; KINSHUK; LIU, T.-C. Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach. *Journal of Educational Technology and Society*, International Forum of Educational Technology and Society, v. 12, n. 4, p. 3–14, 2009. ISSN 11763647, 14364522. Disponível em: <<http://www.jstor.org/stable/jeductechsoci.12.4.3>>. Citado 2 vezes nas páginas 34 e 72.

GRAF, S.; LIST, B. An evaluation of open source e-learning platforms stressing adaptation issues. In: *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*. [S.l.: s.n.], 2005. p. 163–165. ISSN 2161-3761. Citado na página 29.

GRANITO, R. A. N. *Educação a distância e estilos de aprendizagem: elaboração de um protocolo de qualidade para ambientes virtuais de ensino*. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo, 2008. Citado 2 vezes nas páginas 25 e 26.

GRANITTO, P. M. et al. Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems*, v. 83, n. 2, p. 83 – 90, 2006. ISSN 0169-7439. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0169743906000232>>. Citado na página 74.

HAIDER M., S. A.; CHAUDHARY, B. An investigation of relationship between learning styles and performance of learners. *International Journal of Engineering Science and Technology*, 2010. Citado na página 28.

HALAWA, M. S.; HAMED, E. M. R.; SHEHAB, M. E. Personalized e-learning recommendation model based on psychological type and learning style models. In: *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*. [S.l.: s.n.], 2015. p. 578–584. Citado 3 vezes nas páginas 64, 77 e 78.

HALAWA, M. S.; SHEHAB, M. E.; HAMED, E. M. R. Predicting student personality based on a data-driven model from student behavior on lms and social networks. In: *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*. [S.l.: s.n.], 2015. p. 294–299. Citado 8 vezes nas páginas 28, 36, 61, 62, 63, 64, 65 e 66.

HASSAN, S.; HEGAZY, A. E. F. A model recommends best machine learning algorithm to classify learners based on their interactivity with moodle. In: *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*. [S.l.: s.n.], 2015. p. 49–54. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. ISBN 0132733501. Citado 3 vezes nas páginas 52, 53 e 54.

HEIDRICH, L. et al. Diagnosis of learner dropout based on learning styles for online distance learning. *Telematics and Informatics*, v. 35, n. 6, p. 1593 – 1606, 2018. ISSN 0736-5853. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0736585318302089>>. Citado na página 33.

HENZE, N.; DOLOG, P.; NEJDL, W. Reasoning and ontologies for personalized e-learning in the semantic web. *Journal of Educational Technology and Society*, International

Forum of Educational Technology and Society, v. 7, n. 4, p. 82–97, 2004. ISSN 11763647, 14364522. Disponível em: <<http://www.jstor.org/stable/jeductechsoci.7.4.82>>. Citado na página 29.

HONEY, P.; MUMFORD, A. *The Manual of Learning Styles*. Maidenhead: Peter Honey, 1982. 83 p. ISBN 09-508-4440-3. Citado 2 vezes nas páginas 33 e 38.

HONEY, P.; MUMFORD, A. The learning styles helper's guide. 01 2000. Citado na página 38.

HSU, C.-W.; LIN, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, v. 13, n. 2, p. 415–425, March 2002. ISSN 1045-9227. Citado na página 54.

HUNG, Y. H.; CHANG, R. I.; LIN, C. F. Hybrid learning style identification and developing adaptive problem-solving learning activities. *Computers in Human Behavior*, v. 55, p. 552 – 561, 2016. ISSN 0747-5632. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0747563215300145>>. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.

JOYE, C. R.; YOUNG, R. S. *Acolhimento em educação a distância*. [S.l.]: SETEC/IFCE, 2012. Citado na página 25.

JUNG, C. *Psychological Types: Or, The Psychology of Individuation*. Pantheon Books, 1923. (International library of psychology, philosophy and scientific method). Disponível em: <<https://books.google.com.br/books?id=XPKwAAAAIAAJ>>. Citado na página 36.

JÚNIOR, A. H. d. S. *Regional models and minimal learning machines for nonlinear dynamic system identification*. Tese (Doutorado) — Universidade Federal do Ceará, Ceará, 2014. Citado na página 53.

KITCHENHAM, B. *Procedures for Undertaking Systematic Reviews*. [S.l.: s.n.], 2004. Citado na página 57.

KOLB, A.; KOLB, D. The kolb learning style inventory—version 3.1 2005 technical specifications. 01 2005. Citado na página 41.

KOLB, D. Learning styles and disciplinary differences. v. 18, p. 151–164, 01 1981. Citado 2 vezes nas páginas 39 e 40.

KOLB, D. *Experiential Learning: Experience As The Source Of Learning And Development*. [S.l.: s.n.], 1984. v. 1. ISBN 0132952610. Citado 4 vezes nas páginas 27, 38, 39 e 40.

KUMAR, A. Classification model of prediction for placement of students. In: . [S.l.: s.n.], 2013. Citado 2 vezes nas páginas 77 e 78.

LAKATOS, E. M.; MARCONI, M. d. A. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003. ISBN 85-224-3397-6. Citado na página 70.

LAST, F.; DOUZAS, G.; BAÇÃO, F. Oversampling for imbalanced learning based on k-means and SMOTE. *CoRR*, abs/1711.00837, 2017. Disponível em: <<http://arxiv.org/abs/1711.00837>>. Citado na página 75.

LIYANAGE, M. P. P.; GUNAWARDENA, K. L.; HIRAKAWA, M. Detecting learning styles in learning management systems using data mining. *Journal of Information Processing*, v. 24, n. 4, p. 740–749, 2016. Citado 5 vezes nas páginas 61, 62, 63, 65 e 66.

LOSADA, M. The complex dynamics of high performance teams. *Mathematical and Computer Modelling*, v. 30, n. 9, p. 179 – 192, 1999. ISSN 0895-7177. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0895717799001892>>. Citado na página 43.

MACEDO, C. M. S. *Diretrizes para criação de objetos de aprendizagem acessíveis*. Tese (Doutorado) — Universidade Federal de Santa Catarina, Florianópolis, 2010. Citado na página 25.

MEHDINEZHAD, V.; AZARKHORDAD, F. Explaining the students' learning styles based on grasha-riechmann's student learning styles. v. 12, p. 72–79, 07 2016. Citado 2 vezes nas páginas 37 e 38.

MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. *Machine Learning: An Artificial Intelligence Approach*. [S.l.]: Morgan Kaufmann, 2014. ISBN 0934613095. Citado na página 49.

MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072. Citado 3 vezes nas páginas 50, 52 e 55.

MOAZENI, S.; POURMOHAMMADI, H. Smart teaching quantitative topics through the fark learning styles model. p. 1–7, 03 2013. Citado na página 39.

MOBBS, R. *Honey and Mumford*. 2018. Disponível em: <<https://www2.le.ac.uk/departments/doctoralcollege/training/eresources/teaching/theories/honey-mumford>>. Citado na página 38.

MOHAMED, H.; AHMAD, N. B. H.; SHAMSUDDIN, S. M. H. Bijective soft set classification of student's learning styles. In: *2014 8th. Malaysian Software Engineering Conference (MySEC)*. [S.l.: s.n.], 2014. p. 289–294. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: *Sistemas Inteligentes Fundamentos e Aplicações*. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114. ISBN 85-204-168. Citado na página 49.

MYERS-BRIGGS-FOUNDATION. *MBTI Basics*. 2018. Disponível em: <<https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics>>. Citado na página 36.

MYERS, I. B. *Manual: The Myers-Briggs Type Indicator*. [S.l.]: Consulting Psychologists Press, 1962. Citado na página 27.

MYERS, I. B. The myers-briggs type indicator manual. 01 1962. Citado na página 36.

NAKIĆ, J.; GRAF, S.; GRANIĆ, A. Exploring the adaptation to learning styles: The case of adaptive lesson module for moodle. In: HOLZINGER, A. et al. (Ed.). *Human Factors in Computing and Informatics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 534–550. ISBN 978-3-642-39062-3. Citado na página 30.

- NETO, A. R. R. *SINPATCO - Sistema Inteligente para Diagnóstico de Patologias da Coluna Vertebral*. Dissertação (Mestrado) — Universidade Federal do Ceará, UFC, Fortaleza, 2006. Citado 4 vezes nas páginas 50, 51, 52 e 53.
- O'BRIEN, T. P.; BERNOLD, L.; AKROYD, D. Myers-briggs type indicator and academic achievement in engineering education. v. 14, 01 1998. Citado na página 36.
- OTA, M. A.; JÚNIOR, C. F.; BARROS, D. Estilos de aprendizagem em ambientes virtuais: cenários de investigação na educação superior. *Educação, Formação e Tecnologias*, v. 10, n. 1, 2017. ISSN 1646-933X. Disponível em: <<http://eft.educom.pt/index.php/eft/article/view/586>>. Citado na página 30.
- PAIREEKRENG, W.; PREXAWANPRASUT, T. An integrated model for learning style classification in university students using data mining techniques. In: *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. [S.l.: s.n.], 2015. p. 1–5. Citado 5 vezes nas páginas 61, 62, 63, 65 e 66.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado 3 vezes nas páginas 76, 77 e 115.
- PEREIRA, P. S. et al. Análise de perfis de interação de alunos no ambiente virtual moodle via agrupamento hierárquico. *Simpósio Brasileiro de Informática na Educação*, v. 29, 2018. Citado na página 48.
- POPESCU, E. et al. Predicting student performance and differences in learning styles based on textual complexity indices applied on blog and microblog posts: A preliminary study. In: *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*. [S.l.: s.n.], 2016. p. 184–188. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.
- RAMOS, J. L. C. *Uma abordagem preditiva da evasão na educação a distância a partir dos construtos da distância transacional*. Tese (Doutorado) — Universidade Federal de Pernambuco, Pernambuco, 2016. Citado na página 27.
- RIECHMANN, S. W.; GRASHA, A. F. A rational approach to developing and assessing the construct validity of a student learning style scales instrument. v. 87, p. 213–223, 07 1974. Citado 2 vezes nas páginas 37 e 38.
- RODRIGUES, R. L. *Uma abordagem de mineração de dados educacionais para previsão de desempenho a partir de padrões comportamentais de autorregulação da aprendizagem*. Tese (Doutorado) — Universidade Federal de Pernambuco, Pernambuco, 2016. Citado 2 vezes nas páginas 27 e 77.
- ROLIM, V. B.; FERREIRA, R.; COSTA, E. Identificação automática de dúvidas em fóruns educacionais. *Simpósio Brasileiro de Informática na Educação*, 2016. Citado na página 27.
- RUSSEL STUART; NORVIG, P. *Inteligência artificial*. [S.l.]: Elsevier, 2013. ISBN 9788535237016. Citado na página 49.
- SALES, G. et al. Learning vectors model in formative assessment lms moodle: From emoticons to animated gifs. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, p. 64–76, 01 2019. Citado na página 46.

- SALES, G. L. *Learning Vectors (LV): um modelo de avaliação da aprendizagem em EaD online aplicando métricas não-lineares*. Tese (Doutorado) — Universidade Federal do Ceará, Ceará, 2010. Citado 9 vezes nas páginas 27, 28, 43, 44, 45, 46, 47, 48 e 71.
- SALES, G. L.; BARROSO, G.; SOARES, J. Learning vectors (lv): Um modelo de avaliação pro-cessual com mensuração não-linear da aprendizagem em ead online. *Revista Brasileira de Informática na Educação*, v. 20, 04 2012. Citado 2 vezes nas páginas 43 e 47.
- SALES, G. L.; LEITE, E. A. M.; JOYE, C. R. Gerenciamento da aprendizagem, evasão em ead online e possíveis soluções: Um estudo de caso no ifce. *Revista Novas Tecnologias na Educação*, 2012. Citado na página 26.
- SARMENTO, R. M. *Análise das densidades do tecido cerebral: nova abordagem para extração de características em imagens de tomografia computadorizada do crânio*. Dissertação (Dissertação de Mestrado) — Instituto Federal do Ceará, 2016. Citado 3 vezes nas páginas 50, 51 e 53.
- SCHEUER, O.; MCLAREN, B. M. Educational data mining. In: _____. *Encyclopedia of the Sciences of Learning*. Boston, MA: Springer US, 2012. p. 1075–1079. ISBN 978-1-4419-1428-6. Disponível em: <https://doi.org/10.1007/978-1-4419-1428-6_618>. Citado na página 29.
- SHAKHNAROVICH, G.; DARRELL, T.; INDYK, P. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)*. [S.l.: s.n.], 2006. Citado na página 51.
- SILVA, E. L. d.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. *UFSC*, Florianópolis, 2005. Citado na página 70.
- SILVA, Z. C. d. *Adaptação de apresentação de conteúdos de objeto de aprendizagem considerando estilos de aprendizagem*. Tese (Doutorado) — Universidade Federal do Paraná, Paraná, 2017. Citado na página 28.
- SOFLANO, M.; CONNOLLY, T. M.; HAINEY, T. An application of adaptive games-based learning based on learning style to teach sql. *Computers & Education*, v. 86, p. 192 – 211, 2015. ISSN 0360-1315. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0360131515000937>>. Citado na página 33.
- STANGE, J. J. N. R. L. Reconhecimento de padrões em classificadores – comparação de técnicas e aplicações. 4 *Workshop de Tecnologia Adaptativa – WTA-2010*, São Paulo-SP, 2010. Citado na página 55.
- SUGUMARAN, V.; MURALIDHARAN, V.; RAMACHANDRAN, K. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*, v. 21, n. 2, p. 930 – 942, 2007. ISSN 0888-3270. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0888327006001142>>. Citado na página 74.
- SWETA, S.; LAL, K. Web usages mining in automatic detection of learning style in personalized e-learning system. In: RAVI, V. et al. (Ed.). *Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO - 2015)*. Cham:

Springer International Publishing, 2015. p. 353–363. ISBN 978-3-319-27212-2. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.

THEODORIDIS, S.; KOUTROUMBAS, K. Chapter 2 - classifiers based on bayes decision theory. In: THEODORIDIS, S.; KOUTROUMBAS, K. (Ed.). *Pattern Recognition (Fourth Edition)*. Fourth edition. Boston: Academic Press, 2009. p. 13 – 89. ISBN 978-1-59749-272-0. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B9781597492720500049>>. Citado na página 51.

TREVELIN, A. T. C.; NEIVA, J. S. F. d. S. Estilos de aprendizagem e avaliação no ensino superior. *VI Workshop de Pós-Graduação e Pesquisa do Centro Paula Souza*, 2011. ISSN 2175-1897. Citado na página 28.

VAPNIK, V. N. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, v. 10, n. 5, p. 988–999, Sept 1999. ISSN 1045-9227. Citado na página 54.

ZENG, M. et al. Effective prediction of three common diseases by combining smote with totem links technique for imbalanced medical data. In: *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*. [S.l.: s.n.], 2016. p. 225–228. Citado na página 75.

ZHONG, S. et al. A computational investigation of learning behaviors in moocs. *Computer Applications in Engineering Education*, v. 25, n. 5, p. 693–705, 2017. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.21830>>. Citado 6 vezes nas páginas 61, 62, 63, 64, 65 e 66.

Apêndices

APÊNDICE A – CONSULTAS SQL

Apêndice contendo os *scripts* de consultas SQL, referentes à etapa de coleta de dados nas tabelas do Banco de Dados do ambiente virtual *Moodle*.

```
use moodle;
-- criando view usuario
create or replace view usuario as
(
SELECT u2.id as user_id , u2.firstname, u2.lastname,  c2.shortname curso,
c2.id as curso_id, c2.category as curso_category, cat2.id as category_id

FROM mdl_user u2
JOIN mdl_user_enrolments ue2 ON ue2.userid = u2.id
JOIN mdl_enrol e2 ON e2.id = ue2.enrolid
JOIN mdl_role_assignments ra2 ON ra2.userid = u2.id
JOIN mdl_context ct2 ON ct2.id = ra2.contextid AND ct2.contextlevel = 50
JOIN mdl_course c2 ON c2.id = ct2.instanceid AND e2.courseid = c2.id
JOIN mdl_role r2 ON r2.id = ra2.roleid AND r2.shortname = 'student'
JOIN mdl_course_categories cat2 on cat2.id  = c2.category and cat2.id in (21,22)
);

-- criando a view post
create or replace view posts_forumlv as (
select post.*, dis.course, dis.forumlv from mdl_forumlv_posts post
join mdl_forumlv_discussions dis on  dis.id = post.discussion
);

-- quantidade de posts enviadas nos foruns por disciplina
select u.firstname, u.lastname , u.curso, count(post.id) as total from
usuario as u
left outer join posts_forumlv post on post.userid = u.user_id and
post.course = u.curso_id
group by u.firstname, u.lastname , u.curso
order by u.curso , u.firstname;
```

```
-- quantidade de views
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.action in
('view forumlv')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view/edit wikilv por aluno/disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'wikilv' and log.action in ('view','edit')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view conteudos por disciplina (folder)
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'folder' and log.action in ('view')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view tarefaslv por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'tarefalsv' and log.action in ('view')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde upload tarefaslv por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'tarefalsv' and log.action in ('upload')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;
```

```
-- qtde review em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('review')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde attempt em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('attempt')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde close attemp em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('close attempt')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde continue attemp em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('continue attempt')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('view')
group by u.firstname, u.lastname,u.curso
```

```
order by u.curso, u.firstname;

-- qtde view summary em quiz por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'quiz' and log.action in ('view summary')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view visao geral do curso por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'url' -- and log.action in ('view summary')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde view glossary do curso por disciplina
select u.firstname, u.lastname, u.curso, count(log.action) total from
usuario u
left join mdl_log log on log.userid = u.user_id and log.module =
'glossary' and log.action in ('view')
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;

-- qtde add glossary do curso por disciplina
select u.firstname, u.lastname, u.curso, count(g.concept) total from
usuario u
left join mdl_glossary_entries g on g.userid = u.user_id
group by u.firstname, u.lastname,u.curso
order by u.curso, u.firstname;
select u.user_id, u.firstname,u.lastname, q.name,count(qa.attempt)
attempts from usuario u
left join mdl_quiz q on q.course = u.curso_id
left join mdl_quiz_attempts qa on qa.userid = u.user_id and q.id = qa.quiz
group by u.user_id, q.name
order by u.firstname,u.lastname;
```

APÊNDICE B – Scripts Python para Algoritmos de Classificação

Apêndice contendo os *scripts* de busca em grade e de classificação na linguagem *Python*, referentes aos seguintes algoritmos: SVM, KNN, Árvore de Decisão, *Multinomial NB* e MLP .

Script Classe Principal

```
import pandas as pd
from collections import Counter
import numpy as np
from sklearn.model_selection import cross_val_score
#from sklearn.cross_validation import cross_val_score
from sklearn import metrics
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import cross_val_predict
from sklearn_evaluation.plot import grid_search
from sklearn.metrics import confusion_matrix
import imblearn
from imblearn.over_sampling import SMOTE

df = pd.read_csv('atributos-processamento-literatura.csv')
X_df = df[['total_add_glossario', 'total_arq_env_taref',
'total_attempts_quiz']]
Y_df = df['processamento']

#SMOTE TECNINCA BALANCEAMENTO

sm = SMOTE(random_state=12)
X, Y = sm.fit_sample(X_df, Y_df)

#FUNCAO TREINO E PREDIÇÃO USANDO VALIDAÇÃO CRUZADA

def fit_and_predict(nome, modelo, X, Y):
    scores = cross_val_score(modelo, X, Y, cv = 5, scoring='accuracy')
    precisao = cross_val_score(modelo, X, Y, cv = 5, scoring='precision')
```

```

recall = cross_val_score(modelo, X, Y, cv = 5, scoring='recall')
f1 = cross_val_score(modelo, X, Y, cv = 5, scoring='f1')
auc = cross_val_score(modelo, X, Y, cv = 5, scoring='roc_auc')
taxa_acerto = scores.mean()
desviop = scores.std()
#taxa_acerto = np.mean(scores)
media_precisao = np.mean(precisao)
media_recall = np.mean(recall)
media_f1 = np.mean(f1)
media_auc = np.mean(auc)

return {'taxa_acerto':taxa_acerto,'desviop':desviop,
'media_precisao':media_precisao , 'media_recall':media_recall,
'media_f1':media_f1,
'media_auc':media_auc}

```

#FUNCAO OBTER VALORES MÉDIOS DAS MÉTRICAS DE AVALIAÇÃO

```

def media(results):
    mediag_recall = 0.0
    mediag_f1 = 0.0
    mediag_precisao = 0.0
    media_taxa_acerto = 0.0
    media_dp = 0.0
    mediag_auc = 0.0
    recals = []
    f1 = []
    precisoes = []
    taxas = []
    desvios = []
    auc = []
    for r in results:
        recals.append(r['media_recall'])
        f1.append(r['media_f1'])
        auc.append(r['media_auc'])
        precisoes.append(r['media_precisao'])
        taxas.append(r['taxa_acerto'])
        desvios.append(r['desviop'])
    mediag_recall = reduce( lambda x,y : x + y, recals ) / len(recals)

```

```

mediag_f1 = reduce( lambda x,y : x + y, f1 ) / len(f1)
mediag_precisao =
reduce ( lambda x,y: x + y, precisoes ) / len(precisoes)
media_taxa_acerto = reduce ( lambda x,y: x + y, taxas ) / len(taxas)
media_dp = reduce ( lambda x,y: x + y, desvios ) / len(desvios)
mediag_auc = reduce( lambda x,y : x + y, auc ) / len(auc)
return {'recall':mediag_recall, 'f1':mediag_f1,
'precisao' : mediag_precisao, 'taxa_acerto':media_taxa_acerto,
'desviop':media_dp, 'auc':mediag_auc}

```

Script SVM

```

from header import *
#SVM kernel rbf
from sklearn.svm import SVC
from sklearn.svm import LinearSVC

resultados = []
for i in range(0,30):
SVM = SVC(random_state = 0 ,C=60, gamma=0.1)
resultados.append(fit_and_predict("SVM", SVM, X, Y))
print media(resultados)

```

Script KNN

```

from header import *
#KNN
from sklearn.neighbors import KNeighborsClassifier
resultados = []
for i in range(0,30):
    knn = KNeighborsClassifier(n_neighbors=8)
    resultados.append(fit_and_predict("KNN", knn, X, Y))
print media(resultados)

```

Script Árvore de Decisão (*Decision Tree*)

```

from header import *

```

```
from sklearn import tree
resultados = []
for i in range(0,30):
    dt = tree.DecisionTreeClassifier()
    resultados.append(fit_and_predict("DecisionTreeClassifier",
    dt, X, Y))
print media(resultados)
```

Script Naive Bayes (*MultinomialNB*)

```
from header import *
from sklearn.naive_bayes import MultinomialNB
resultados = []
for i in range(0,30):
    modeloMultinomial = MultinomialNB()
    resultados.append(fit_and_predict("MultinomialNB",
    modeloMultinomial, X, Y))
print media(resultados)
```

Script MLP

```
from header import *
from sklearn.neural_network import MLPClassifier
resultados = []
i_range = range(0,30);
for i in i_range:
    mlp = MLPClassifier(solver='lbfgs')
    resultados.append(fit_and_predict("MLP", mlp, X, Y))
print media(resultados)
```

Script Busca em Grade KNN

```
from sklearn.grid_search import GridSearchCV
from sklearn.neighbors import KNeighborsClassifier
import matplotlib.pyplot as plt
k_range = list(range(1, 15))
```

```

param_grid = dict(n_neighbors=k_range)

for i in range(0,30):
    knn=KNeighborsClassifier()
    grid = GridSearchCV(knn, param_grid, cv=5, scoring='accuracy')
    grid.fit(X, Y)
grid_mean_scores = [result.mean_validation_score for result
in grid.grid_scores_]
print(grid.best_params_)
print(grid.best_score_)
plt.plot(k_range, grid_mean_scores)
plt.xlabel('Valor de K para KNN')
plt.ylabel('Taxa de acerto-Cross-Validation')
plt.show()

```

Script Busca em Grade SVM

```

from sklearn.svm import SVC
from sklearn.grid_search import GridSearchCV
from sklearn_evaluation.plot import grid_search
param_grid={'C':[1,10,20,30,40,50,60,70,80,90,100],
'gamma':[1,0.1,0.001,0.0001], 'kernel':['linear','rbf']}
SVM = SVC(random_state = 0)
grid_svm = GridSearchCV(SVM, param_grid, cv=5)
grid_svm.fit(X, Y)
grid_svm.grid_scores_
grid_scores = grid_svm.grid_scores_
print(grid_svm.best_params_)
print(grid_svm.best_score_)
to_vary = 'C'
grid_search(grid_scores, to_vary)
plt.xlabel('Valor de C')
plt.ylabel('Taxa de Acerto - Cross Validation')
plt.title('')
plt.legend(loc='best',prop={'size':7})
plt.show()

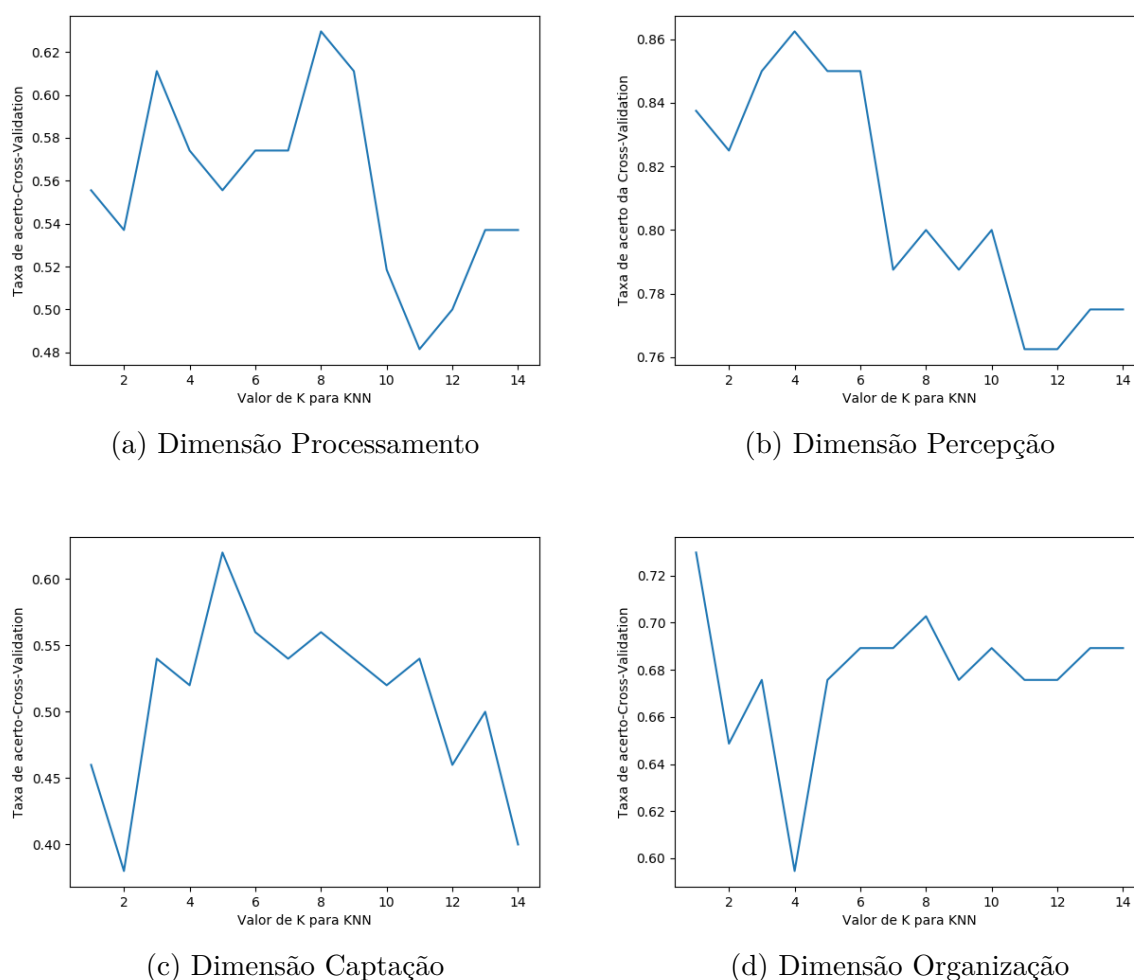
```


APÊNDICE C – Ajustes dos melhores parâmetros

Apêndice contendo os *scripts* os ajustes dos melhores parâmetros da busca em grade.

Os hiper-parâmetros dos modelos KNN e SVM foram escolhidos usando busca em grade e validação cruzada de 5 vezes. A seguir, descrevem-se os resultados da busca em grade no algoritmo KNN, de forma a encontrar o melhor número de vizinhos. Inicialmente executou-se a busca em grade nos atributos da literatura para cada uma das quatro dimensões (FIGURA 27).

Figura 27 – Grid Search - KNN - Atributos Literatura



Fonte: O Autor

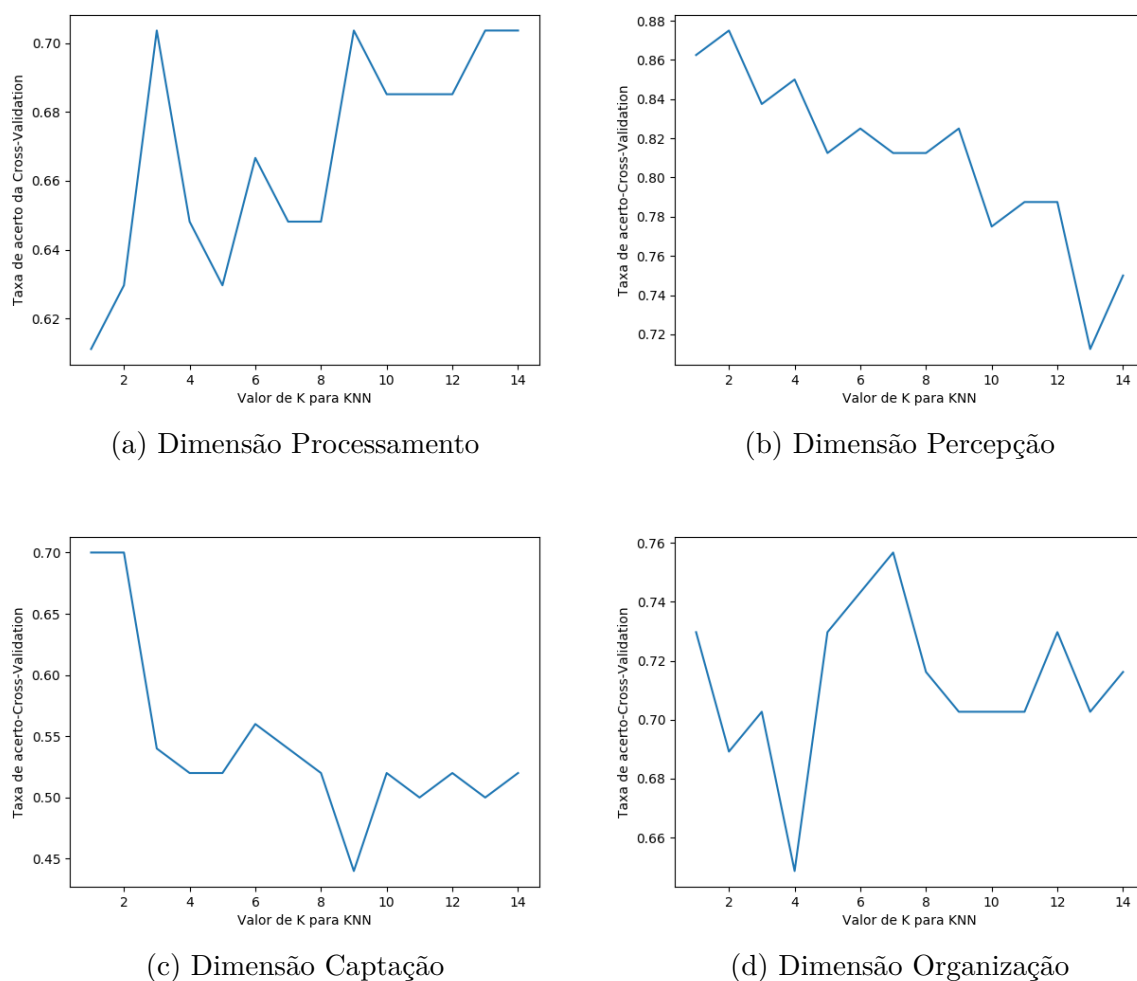
Os valores de K foram variados de 1 a 15. Os melhores valores de K encontrados foram: 8, 4, 5 e 1, respectivamente nas dimensões processamento, percepção, captação e organização. Levando em consideração as acurácias encontradas utilizando o valor padrão de K, com K=5, todas as dimensões, com exceção da dimensão captação, obtiveram acurácias melhores usando outros valores de K.

Essa melhora nos resultados, mostra a importância do uso da busca em grade. O pior valor na dimensão Processamento e na dimensão Percepção foi de K=11, já na dimensão captação o pior valor foi 2 e na dimensão organização foi 4, obtendo uma acurácia de apenas 59%.

Também foram testados os valores de K nas quatro dimensões usando os atributos do Modelo LV (FIGURA 28). Os melhores valores de K encontrados foram: 3, 2, 1 e 7, respectivamente nas dimensões processamento, percepção, captação e organização. Já os piores valores de K foram: processamento (1), percepção (13), captação (9) e organização (4).

Assim como os resultados encontrados na busca em grade com atributos da literatura, todas as dimensões obtiveram acurácias melhores usando valores de K diferentes do valor padrão, que é igual a cinco. Na dimensão captação, por exemplo, a acurácia obtida com K=5 foi de apenas 53%, já com K=1 a acurácia foi de 70%, apresentando uma melhora bastante considerável.

Figura 28 – Grid Search - KNN - Atributos LV



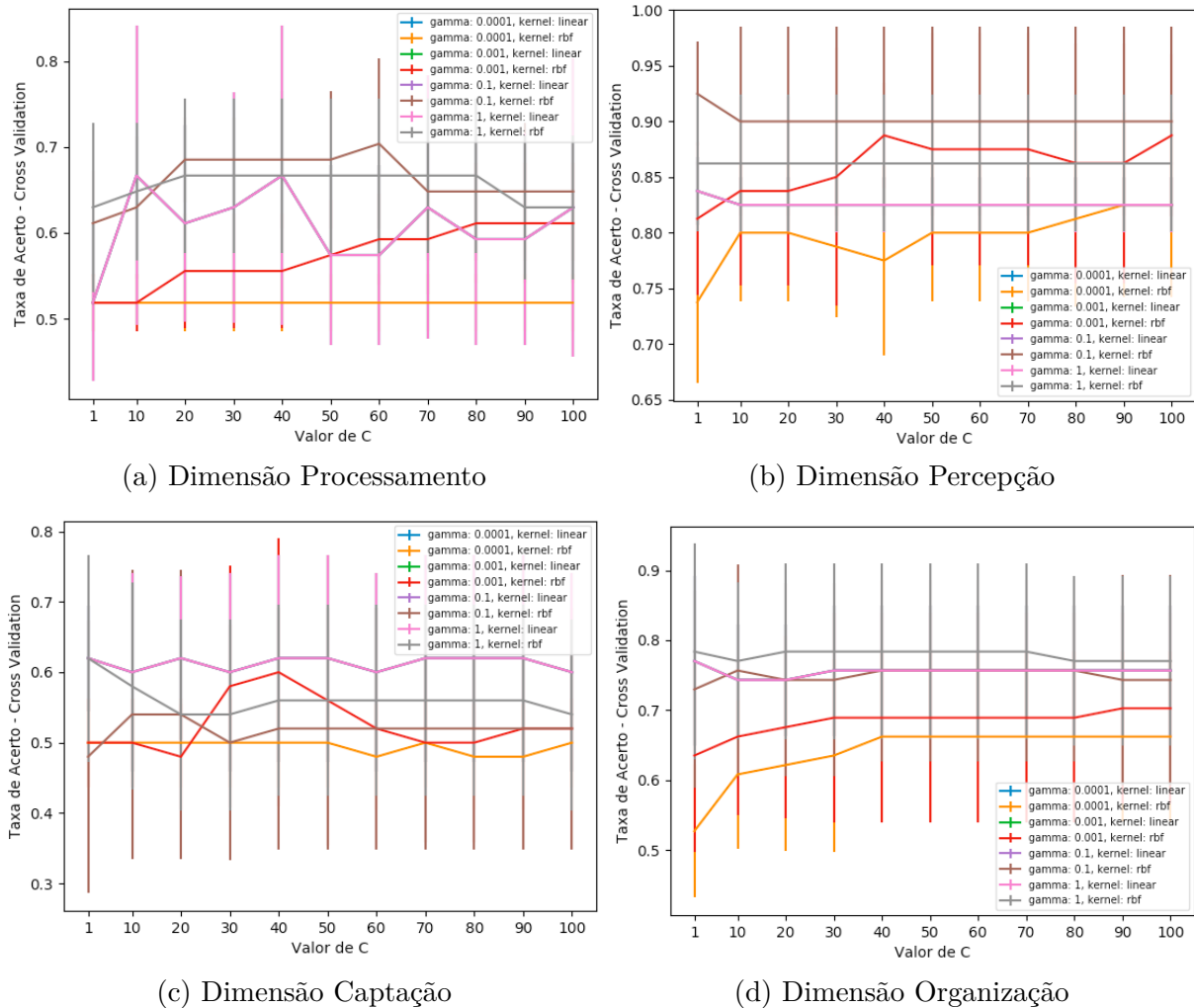
Fonte: O Autor

A busca em grade no algoritmo SVM, buscou encontrar os melhores valores de: C , γ e $kernel$. Os valores de C testados foram: 1,10,20,30,40,50,60,70,80,90,100. Um valor de C baixo torna a superfície de decisão suave, enquanto um C alto tem como objetivo classificar corretamente todos os exemplos de treinamento (PEDREGOSA et al., 2011). Já os valores de γ variados foram: 1,0.1,0.001,0.0001. Este parâmetro define quanta influência tem um único exemplo de treinamento. Quanto maior for, mais próximos outros exemplos devem ser afetados. Já o *kernel* foi parametrizado no tipo linear e no tipo RBF (*Radial Basis Function*).

Nos atributos da literatura esses parâmetros foram testados nas quatro dimensões (FIGURA 29). Em todas as dimensões, com exceção da dimensão de Captação, o *kernel* escolhido na busca em grade foi do tipo RBF. Nas dimensões Processamento e Percepção o melhor valor para γ foi 0,1, já nas dimensões Captação e Organização esse valor foi igual a 1. O melhor valor de C encontrado foi 1, menos na dimensão Processamento, que

foi igual a 60. O kernel RBF e o valor de $C=1$ são valores padrões setados no classificador SVM, já o valor de gamma é considerado apenas usando o kernel rbf e é calculado dividindo o número 1 pelo número de características.

Figura 29 – Grid Search - SVM - Atributos Literatura

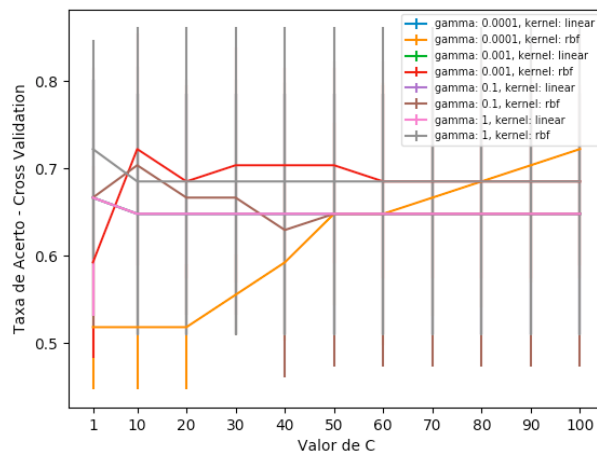


Fonte: O Autor

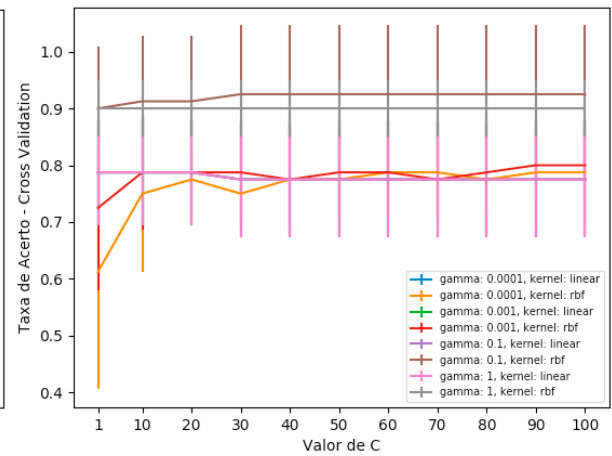
Também foi realizada a busca em grade nas quatro dimensões usando os atributos do Modelo LV (FIGURA 30). Assim como os resultados encontrados na busca em grade com atributos da literatura, todas as dimensões obtiveram melhores acurácias usando o *kernel* RBF. Já o valor de gamma escolhido em todas as dimensões, com exceção da dimensão Percepção, foi igual a 1.

Os valores de C foram variados e apenas na dimensão Processamento o melhor valor foi o valor padrão igual a 1. Nas outras dimensões os valores foram: 30, 80, 40, respectivamente nas dimensões Percepção, Captação e Organização. A melhor acurácia obtida foi na dimensão Percepção, com valor igual a 92,50%, utilizando kernel RBF, $C=30$, $\text{gamma}=0,1$.

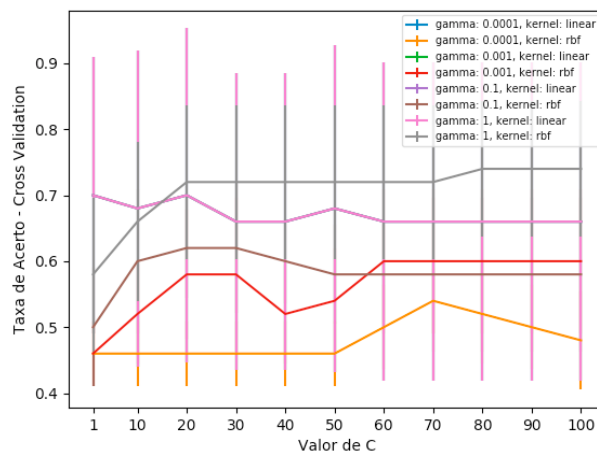
Figura 30 – Grid Search - SVM - Atributos LV



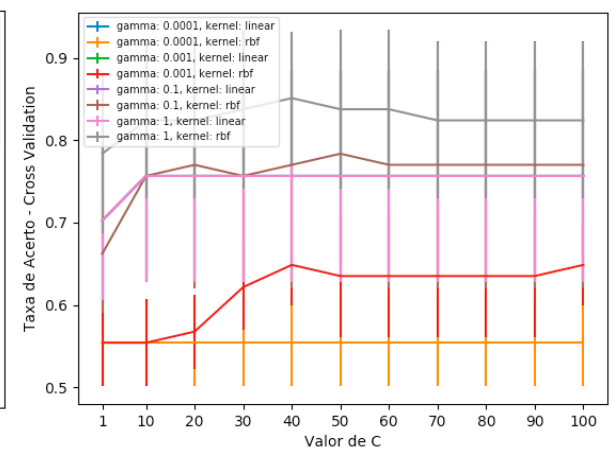
(a) Dimensão Processamento



(b) Dimensão Percepção



(c) Dimensão Captação



(d) Dimensão Organização

Fonte: O Autor

Anexos

ANEXO A – Questionário ILS

Questionário ILS

Prezado(a), este é o questionário ILS (*Index of Learning Styles*), proposto por Felder e Soloman (1991) para identificação do Estilo de Aprendizagem. Ele será utilizado como um dos instrumentos da pesquisa de campo – Uma Abordagem Baseada no Uso de Métricas do Modelo LV para Identificação Automática de Estilos de Aprendizagem em Ambientes Virtuais de Aprendizagem - elaborada pela mestrandia Edilaine Santiago de Oliveira, sob a orientação do Prof. Dr. Gilvandenys Leite Sales, no Instituto Federal do Ceará (IFCE). Logo, a sua participação é fundamental para o sucesso desta pesquisa. Desde já, agradeço a atenção e colaboração!

1 - Eu compreendo melhor alguma coisa depois de:

☐ experimentar

☐ refletir sobre ela

2 - Eu me considero:

☐ realista

☐ inovador(a)

3 - Quando eu penso sobre o que fiz ontem, é mais provável que aflorem:

☐ figuras

☐ palavras

4 - Eu tendo a:

☐ compreender os detalhes de um assunto, mas a estrutura geral pode ficar imprecisa.

☐ compreender a estrutura geral de um assunto, mas os detalhes podem ficar imprecisos.

5 - Quando estou aprendendo algum assunto novo, me ajuda:

☐ falar sobre ele.

☐ refletir sobre ele.

6 - Se eu fosse um professor, eu preferiria ensinar uma disciplina:

☐ que trate com fatos e situações reais.

☐ que trate com idéias e teorias.

7 - Eu prefiro obter novas informações através de:

() figuras, diagramas, gráficos ou mapas.

() instruções escritas ou informações verbais.

8 - Quando eu compreendo:

() todas as partes, consigo entender o todo.

() o todo, consigo ver como as partes se encaixam

9 - Em um grupo de estudo, trabalhando um material difícil, eu provavelmente:

() tomo a iniciativa e contribuo com idéias.

() assumo uma posição discreta e escuto.

10 - Acho mais fácil:

() aprender fatos.

() aprender conceitos.

11 - Em um livro com uma porção de figuras e desenhos, eu provavelmente:

() observo as figuras e desenhos cuidadosamente

() atento para o texto escrito.

12 - Quando resolvo problemas de matemática, eu:

() usualmente trabalho de maneira a resolver uma etapa de cada vez.

() frequentemente antevejo as soluções, mas tenho que me esforçar muito para conceber as etapas para chegar a elas.

13 - Nas disciplinas que cursei eu:

() em geral fiz amizade com muitos dos colegas.

() raramente fiz amizade com muitos dos colegas.

14 - Em literatura de não-ficção, eu prefiro:

() algo que me ensine fatos novos ou me indique como fazer alguma coisa.

() algo que me apresente novas idéias para pensar.

15 - Eu gosto de professores:

() que colocam uma porção de diagramas no quadro.

() que gastam bastante tempo explicando.

16 - Quando estou analisando uma história ou novela eu:

() penso nos incidentes e tento colocá-los juntos para identificar os temas.

☐tenho consciência dos temas quando termino a leitura e então tenho que voltar atrás para encontrar os incidentes que os confirmem.

17 - Quando inicio a resolução de um problema para casa, normalmente eu:

☐começo a trabalhar imediatamente na solução.

☐primeiro tento compreender completamente o problema.

18 - Prefiro a idéia do:

☐certo

☐teórico

19 - Relembro melhor:

☐o que vejo.

☐o que ouço.

20 - É mais importante para mim que o professor:

☐apresente a matéria em etapas sequenciais claras.

☐apresente um quadro geral e relacione a matéria com outros assuntos.

21 - Eu prefiro estudar:

☐em grupo.

☐sozinho(a).

22 - Eu costumo ser considerado(a):

☐cuidadoso(a) com os detalhes do meu trabalho.

☐criativo(a) na maneira de realizar meu trabalho.

23 - Quando busco orientação para chegar a um lugar desconhecido, eu prefiro:

☐um mapa

☐instruções por escrito

24 - Eu aprendo:

☐num ritmo bastante regular. Se estudar pesado, eu “chego lá”.

☐em saltos. Fico totalmente confuso(a) por algum tempo e, então, repentinamente eu tenho um “estalo”.

25 - Eu prefiro primeiro:

☐experimentar as coisas.

☐pensar sobre como é que eu vou fazer.

26 - Quando estou lendo como lazer, eu prefiro escritores que:

() explicitem claramente o que querem dizer.

() dizem as coisas de maneira criativa, interessante.

27 - Quando vejo um diagrama ou esquema em uma aula, relembro mais facilmente:

() a figura.

() o que o(a) professor(a) disse a respeito dela.

28 - Quando considero um conjunto de informações, provavelmente eu:

() presto mais atenção nos detalhes e não percebo o quadro geral.

() procuro compreender o quadro geral antes de atentar para os detalhes.

29 - Relembro mais facilmente:

() algo que fiz.

() algo sobre o que pensei bastante

30 - Quando tenho uma tarefa para executar, eu prefiro:

() dominar uma maneira para a execução da tarefa.

() encontrar novas maneiras para a execução da tarefa.

31 - Quando alguém está me mostrando dados, eu prefiro:

() diagramas e gráficos.

() texto resumizando os resultados.

32 - Quando escrevo um texto, eu prefiro trabalhar (pensar a respeito ou escrever):

() a parte inicial do texto e avançar ordenadamente.

() diferentes partes do texto e ordená-las depois.

33 - Quando tenho que trabalhar em um projeto em grupo, eu prefiro que se faça primeiro:

() um debate (brainstorming) em grupo, onde todos contribuem com idéias.

() um brainstorming individual, seguido de reunião do grupo para comparar idéias.

34 - Considero um elogio chamar alguém de:

() sensível

() imaginativo

35 - Das pessoas que conheço em uma festa, provavelmente eu me recordo melhor:

() de sua aparência.

()do que elas disseram de si mesmas.

36 - Quando estou aprendendo um assunto novo, eu prefiro:

()concentrar-me no assunto, aprendendo o máximo possível.

()tentar estabelecer conexões entre o assunto e outros com ele relacionados

37 - Mais provavelmente sou considerado(a):

()expansivo(a).

()reservado(a).

38 - Prefiro disciplinas que enfatizam:

()material concreto (fatos, dados).

()material abstrato (conceitos, teorias).

39 - Para entretenimento, eu prefiro:

()assistir televisão.

()ler um livro.

40 - Alguns professores iniciam suas preleções com um resumo do que irão cobrir.

Tais resumos são:

()de alguma utilidade para mim.

()muito úteis para mim

41 - A idéia de fazer o trabalho de casa em grupo, com a mesma nota para todos do grupo:

me agrada.

não me agrada.

42 - Quando estou fazendo cálculos longos:

tendo a repetir todos os passos e conferir meu trabalho cuidadosamente.

acho cansativo conferir o meu trabalho e tenho que me esforçar para fazê-lo.

43 - Tendo a descrever os lugares onde estive:

()com facilidade e com bom detalhamento.

()com dificuldade e sem detalhamento.

44 - Quando estou resolvendo problemas em grupo, mais provavelmente eu:

()penso nas etapas do processo de solução.

()penso nas possíveis consequências, ou sobre as aplicações da solução para uma ampla faixa de áreas.