



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

RONNIE CARLOS TAVARES NUNES

**UM MODELO DE PERFIL DE ALUNO VOLTADO A APLICAÇÕES DE
TÉCNICAS DE *LEARNING ANALYTICS***

FLORIANÓPOLIS

2019

Ronnie Carlos Tavares Nunes

**UM MODELO DE PERFIL DE ALUNO VOLTADO A APLICAÇÕES DE TÉCNICAS
DE *LEARNING ANALYTICS***

Dissertação submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia e Gestão do Conhecimento. Orientador: Prof. Dr. Alexandre Leopoldo Gonçalves.
Coorientador: Prof. Dr. João Artur de Souza.

FLORIANÓPOLIS

2019

Ficha de identificação da obra

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Nunes, Ronnie Carlos Tavares

Um modelo de perfil de aluno voltado a aplicações de técnicas de learning analytics / Ronnie Carlos Tavares Nunes ; orientador, Alexandre Leopoldo Gonçalves, coorientador, João Artur de Souza, 2019.

145 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2019.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Learning Analytics. 3. Perfil de Aluno. 4. Ambientes Virtuais de Aprendizagem. I. Gonçalves, Alexandre Leopoldo. II. Souza, João Artur de . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Ronnie Carlos Tavares Nunes

Um modelo de perfil de aluno voltado a aplicações de técnicas de *Learning Analytics*

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Márcio Vieira de Souza, Dr.
Universidade Federal de Santa Catarina

Prof. Fernando José Spanhol, Dr.
Universidade Federal de Santa Catarina

Prof. Juarez Bento da Silva, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia e Gestão do Conhecimento.

Prof. Dr. Roberto Carlos dos Santos Pacheco

Coordenador do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento

Prof. Dr. Alexandre Leopoldo Gonçalves

Orientador

Florianópolis, 28 de novembro de 2019.

Este trabalho é dedicado especialmente à minha querida mãe, cuja passagem para a outra vida deu-se em 08 de julho de 2019. Apesar de separados geograficamente, nosso contato sempre foi frequente, em todo o período de desenvolvimento da dissertação. A saudade e o sentimento de perda subsistirão em todos os momentos em que eu me lembrar deste período de estudos.

AGRADECIMENTOS

Agradeço ao meu orientador, Dr. Alexandre Leopoldo Gonçalves, pela compreensão, paciência, apoio, dedicação e orientação. Desde o período da Pós-Graduação em TIC aplicada a Segurança Pública e Direitos Humanos, realizada em Araranguá-SC, na qual também foi meu orientador, tem me acompanhado e contribuído de maneira decisiva para a minha formação. Meu respeito, admiração e eterna gratidão.

Ao Prof. Dr. João Artur de Souza, pela liberdade, autonomia e confiança depositada para a escolha e manipulação do tema. Aos demais professores que tive o prazer de conhecer durante todas as disciplinas que cursei. Aos professores Fernando José Spanhol, Márcio Vieira de Sousa e Juarez Bento da Silva, por aceitarem o convite para a banca de defesa. É uma honra tê-los como avaliadores deste trabalho.

Ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, pela oportunidade de participar desse prestigiado curso. A todos os demais professores que não tive contato, mais que certamente são essenciais para o programa e demais profissionais do PPGEKC.

“Education is not just about going to school and getting a degree. It's about widening your knowledge and absorbing the truth about life”.

*Shakuntala Devi (*1929-11-04 †2013-04-21)*

RESUMO

A análise das interações dos alunos com os ambientes virtuais de aprendizagem assumiu um papel relevante para decisões educacionais. A grande disponibilidade de cursos a distância permite o uso da tecnologia a fim de explorar os dados produzidos a partir dessas alterações. Pode assim, maximizar o aprendizado dos alunos, sugerindo atividades de acordo com o perfil de cada um. Entretanto, a utilização do perfil do aluno para análises mais abrangentes ainda é insipiente. Neste sentido, o presente trabalho propõe um modelo de dados de perfil de aluno voltado a aplicação de técnicas de *Learning Analytics* em Sistemas de Aprendizagem *Online*. O modelo, elaborado por meio do desenvolvimento de artefatos, teve como suporte a metodologia *Design Science Research*. Para a sua avaliação, utilizou-se uma base de dados de uma instituição de ensino que possui atividades ativas em um ambiente virtual de aprendizagem. A partir desses dados, foi possível a aplicação das técnicas escolhidas, obtendo-se informações relevantes para subsidiar os gestores no âmbito educacional. Análises estatísticas, análise de agrupamentos e sistemas de recomendação foram as técnicas aplicadas. De maneira geral, os resultados produzidos estão centrados na identificação e geração de grupos de perfis similares, considerando o estilo de aprendizagem e o tipo de personalidade dos alunos. Esta estratégia permitiu a obtenção de resultados promissores para a tomada de decisão no contexto educacional e com potencial para gerar uma contribuição efetiva para a área de *Learning Analytics*.

Palavras-chave: *Learning Analytics*, Perfil de Aluno, Ambientes Virtuais de Aprendizagem.

ABSTRACT

The analysis of students' interactions with virtual learning environments has assumed a relevant role for educational decisions. The wide availability of distance learning courses allows the use of technology to exploit the data produced from these interactions. It can thus maximize students' learning by suggesting activities according to their profile. However, using the student profile for broader analysis is still incipient. In this sense, the present work proposes a student profile data model, focused on the application of Learning Analytics techniques in Online Learning Systems. The model, created through the development of artifacts, was supported by the Design Science Research methodology. For its evaluation, it was used a database from an educational institution that has active activities in a virtual learning environment. From these data, it was possible to apply the chosen techniques, obtaining relevant information to support managers in the educational field. Statistical analyzes, cluster analysis and recommendation systems were the applied techniques. In general, the results produced focus on the identification and generation of similar profile groups, considering the students' learning style and personality type. This strategy allowed promising results for decision making in the educational context and with the potential to generate an effective contribution to the area of Learning Analytics.

Keywords: Learning Analytics, Student Profile, Virtual Learning Environment.

LISTA DE FIGURAS

Figura 1 - Procedimentos da metodologia DSR.	29
Figura 2 - Metodologia x estrutura do trabalho.	29
Figura 3 – As cinco gerações de EaD.	41
Figura 4 – Panorama de <i>Learning Analytics</i>	52
Figura 5 – Primeiro modelo de <i>Learning Analytics</i>	53
Figura 6 – O ciclo de LA.	54
Figura 7 – Modelo de referência para a LA.	55
Figura 8 – Fluxo interativo da LA.	56
Figura 9 – Modelo de LA combinado com <i>Teaching Inquiry</i>	57
Figura 10 - Exemplo de agrupamentos.	62
Figura 11 - Taxonomia de análise de agrupamento.	64
Figura 12 - Passos do algoritmo <i>K-means</i>	67
Figura 13 – Um modelo de sistema de recomendação.	70
Figura 14 – <i>Framework</i> da classificação funcional.	72
Figura 15 – Filtragem Colaborativa baseada no usuário.	76
Figura 16 – Filtragem Colaborativa baseada no item.	76
Figura 17 – Filtragem Colaborativa.	77
Figura 18 – Arquitetura de referência do ArCARE.	84
Figura 19 – Estrutura Analítica do Projeto (EAP).	91
Figura 20 – Modelo Conceitual.	94
Figura 21 – Modelo Lógico.	95
Figura 22 – Sexo dos alunos.	100
Figura 23 – Alunos por fase do curso.	101
Figura 24 – Tipos de curso oferecidos.	101
Figura 25 – Disciplinas por cada fase do curso.	102
Figura 26 – Alunos de acordo com o estilo cognitivo.	103
Figura 27 – Alunos com estilo de aprendizagem DSI.	103
Figura 28 – Alunos com estilo de aprendizagem DVV.	103
Figura 29 – Alunos com o estilo de aprendizagem DAR.	104
Figura 30 – Alunos com estilo de aprendizagem DSG.	104
Figura 31 – Exemplo de preenchimento para os dados de “PERSONALIDADE”.	105
Figura 32 – Quantidade de tipos de personalidade.	106

Figura 33 – Gráfico de contabilização dos tipos de personalidade.	106
Figura 34 – Planilha Estilo de Aprendizagem.	108
Figura 35 – Planilha de tipo de personalidade.	108
Figura 36 – Estrutura da tabela que representa os vetores dos alunos.	109
Figura 37 – Tabela “ <i>distances</i> ”.	109
Figura 38 – Grafo gerado com os dados de similaridade.	111
Figura 39 – Grafo de similaridade completo.	112
Figura 40 – Grafo do segundo cenário de recomendação.	123

LISTA DE QUADROS

Quadro 1 – Trabalhos no EGC com o termo: EaD.....	33
Quadro 2 – Trabalhos no EGC com o termo: MOOC.....	33
Quadro 3 – Trabalhos no EGC com o termo: AVA.....	34
Quadro 4 – Diferenças entre cMOOC e xMOOC.....	46
Quadro 5 – Dados disponíveis nos MOOCs.....	48
Quadro 6 – Exemplos de estratégias de LA.....	59
Quadro 7 – Exemplos de métodos utilizados na área de LA.....	59
Quadro 8 – Exemplos de técnicas utilizadas na área de LA.....	59
Quadro 9 – Métodos básicos de algoritmos de agrupamento.....	63
Quadro 10 – Procedimentos do algoritmo <i>K-means</i>	66
Quadro 11 – Aplicações dos Sistemas de Recomendação.....	78
Quadro 12 – Dados do estudante em rede social.....	87
Quadro 13 – Variáveis consideradas no aprendizado <i>online</i>	88
Quadro 14 – Representação da tabela “USUARIO”.....	96
Quadro 15 – Representação da tabela “ESTILO_APRENDIZAGEM”.....	96
Quadro 16 – Domínio dos atributos da tabela “ESTILO_APRENDIZAGEM”.....	96
Quadro 17 – Representação da tabela “AREA_INTERESSE”.....	97
Quadro 18 – Representação da tabela “AREA_INTERESSE_USUÁRIO”.....	97
Quadro 19 – Representação da tabela “PERSONALIDADE”.....	97
Quadro 20 – Representação da tabela “ATV_EXTRACURRICULAR”.....	98
Quadro 21 – Representação da tabela “ATV_EXTRACURRICULAR_USUARIO”.....	98
Quadro 22 – Instrução utilizada para gerar a tabela de similaridades.....	109
Quadro 23 – Comando SQL para importação dos nós.....	110
Quadro 24 – Comando SQL para importação das arestas.....	110
Quadro 25 – Comando SQL – Aluno alvo das recomendações.....	118
Quadro 26 – Comando SQL – Atividades dos outros alunos.....	118
Quadro 27 – Comando SQL – Atividades a serem recomendadas.....	118
Quadro 28 – Atividades a serem recomendadas ao aluno 47027.....	119
Quadro 29 – Algumas atividades recomendadas ao aluno 47420.....	119
Quadro 30 – Atividades recomendadas ao aluno 47554.....	120
Quadro 31 – Algumas atividades recomendadas ao aluno 46977.....	120
Quadro 32 – Atividades recomendadas ao aluno 47108.....	120

Quadro 33 – Algumas atividades recomendadas ao aluno 47106.....	121
Quadro 34 – Comando para criação da segunda tabela de distâncias.	122
Quadro 35 – Comando para importação dos nós no <i>Gephi</i> ®.....	122
Quadro 36 – Comando para importação das arestas no <i>Gephi</i> ®.....	122
Quadro 37 – Definição de alunos para o cenário.....	123
Quadro 38 – Exemplo de atividades recomendadas ao aluno 47440.	124
Quadro 39 – Exemplo de atividades recomendadas ao aluno 47420.	125
Quadro 40 – Exemplo de atividades recomendadas ao aluno 47868.	125

LISTA DE TABELAS

Tabela 1 – Indicadores de Qualidade de EaD.....	42
Tabela 2 – Definição de alunos para avaliação estatística.....	113
Tabela 3 – Acessos do grupo 1.....	113
Tabela 4 – Acessos do grupo 2.....	113
Tabela 5 – Acessos do grupo 3.....	113
Tabela 6 – Acessos do grupo 4.....	113
Tabela 7 – Acessos do grupo 5.....	113
Tabela 8 – Notas médias do grupo 1.....	114
Tabela 9 – Notas médias do grupo 2.....	114
Tabela 10 – Notas médias do grupo 3.....	114
Tabela 11 – Notas médias do grupo 4.....	114
Tabela 12 – Notas médias do grupo 5.....	114
Tabela 13 – Grupo 1.....	115
Tabela 14 – Grupo 2.....	115
Tabela 15 – Grupo 3.....	115
Tabela 16 – Grupo 4.....	115
Tabela 17 – Grupo 5.....	115
Tabela 18 – Definição de alunos para recomendação colaborativa.....	117

LISTA DE ABREVIATURAS E SIGLAS

AVA – Ambientes Virtuais de Aprendizagem
BDA – *Big Data Acadêmico*
BN – *Bayesian Networks*
BROAD-RS – *BROAD Recommendation System*
CALM – *Calculative Associative Logical Memorable*
CART – *Classification and Regression Tree*
CNE – Conselho Nacional de Educação
DSR – *Design Science Research*
EaD – Ensino a distância
EAP – Estrutura Analítica do Projeto
EDC – *Educational Data Clustering*
EUA – Estados Unidos da América
FSLSM – *Felder-Silverman Learning Style Model*
GPUs – *Graphics Processing Unit*
IEEE - *Institute of Electrical and Electronic Engineers*
IMS – *Instructional Management Systems*
KPI – *Key Performance Indicators*
LA – *Learning Analytics*
LIP – *Learner Information Package*
LMS – *Learning Management System*
MOOC – *Massive Open Online Course*
NB – *Naives Bayes*
NMF – *Negative Matrix Factorization*
PAPI - *Public and Private Information for Learners*
PLE – *Personal Learning Environment*
PMF – *Probability Mass Function*
PNE – Plano Nacional de Educação
PPGEGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento
QCL – Qualificações, Certificados e Licenças
SOLAR – Sociedade para a pesquisa em *Learning Analytics*
SPOC – *Small Private Online Course*
SQL – *Structured Query Language*

SR – Sistemas de Recomendação

SVD – *Singular Value Decomposition*

SVM – *Support-Vector Machine*

TA – *Teaching Analytics*

TEL – *Technology Enhanced Learning*

ULF – *Universal Learning Format*

UNESCO - Organização das Nações Unidas para a Educação, Ciência e Cultura

VARK – *Visual, Aural, Read-Write and Kinesthetic*

XML – *Extensible Markup Language*

LISTA DE SÍMBOLOS

® - Marca Registrada

SUMÁRIO

1	INTRODUÇÃO	20
1.1	DEFINIÇÃO DO PROBLEMA.....	22
1.2	OBJETIVOS.....	24
1.2.1	Objetivo geral	24
1.2.2	Objetivos específicos.....	24
1.3	JUSTIFICATIVA E RELEVÂNCIA DO TEMA.....	25
1.4	DELIMITAÇÕES DA PESQUISA.....	26
1.5	PROCEDIMENTOS METODOLÓGICOS	27
1.6	ADERÊNCIA AO PPGE GC.....	30
1.6.1	Identidade.....	31
1.6.2	Contexto Estrutural do EGC	32
1.6.3	Referências Factualis.....	32
1.7	ESTRUTURA DO TRABALHO	35
2	REFERENCIAL TEÓRICO	36
2.1	SISTEMAS DE APRENDIZAGEM <i>ONLINE</i>	36
2.1.1	Sistemas Educacionais	36
2.1.2	Sistemas Educacionais <i>Online</i>	37
2.2	EDUCAÇÃO A DISTÂNCIA	38
2.3	AMBIENTES VIRTUAIS DE APRENDIZAGEM - AVA.....	42
2.4	MOOCs	44
2.5	LEARNING ANALYTICS.....	48
2.5.1	Modelos.....	52
5.2	<i>Frameworks</i>	57
2.5.3	Estratégias, Métodos, Técnicas e Algoritmos	58
2.5.4	Análise de Agrupamento.....	60

2.5.4.1	Algoritmos por Particionamento	64
2.5.4.1.1	<i>K-means - Uma técnica baseada no centróide</i>	65
2.5.4.2	Algoritmos Hierárquicos	67
2.5.4.3	Modularidade.....	68
2.6	SISTEMAS DE RECOMENDAÇÃO.....	69
2.6.1	Abordagens de Recomendação	72
2.6.2	Baseada em Conteúdo	73
2.6.3	Filtragem Colaborativa	75
2.6.4	Abordagem Demográfica	79
2.6.5	Baseada em Conhecimento	79
2.6.6	Baseado na comunidade	80
2.6.7	Sistemas de Recomendação Híbridos	81
2.7	TRABALHOS RELACIONADOS	81
3	MODELO PROPOSTO	89
3.1	DESCRIÇÃO DO MODELO	89
3.2	MODELO CONCEITUAL E LÓGICO.....	93
4	APLICAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS	99
4.1	CENÁRIO DE APLICAÇÃO	99
4.2	ANÁLISE GERAL DOS DADOS.....	99
4.3	ANÁLISE BASEADA EM AGRUPAMENTO	106
4.4	PRIMEIRO CENÁRIO DE RECOMENDACÃO	116
4.5	SEGUNDO CENÁRIO DE RECOMENDACÃO	121
5	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	126
5.1	CONSIDERAÇÕES FINAIS	126
5.2	TRABALHOS FUTUROS.....	128
	REFERÊNCIAS	129

1 INTRODUÇÃO

Antes da *internet*, várias iniciativas de educação a distância faziam uso das tecnologias disponíveis a época, como por exemplo, a transmissão de programas de TV, áudios e vídeos gravados.

Com o surgimento da *internet*, este cenário sofreu grande mudança, e várias instituições passaram a oferecer cursos *online*. Assim, o professor além de ministrar o conteúdo, também passa a desenvolver, organizar e promover suporte ao aprendizado do aluno. Uma característica peculiar das plataformas *online* de ensino, é que uma grande quantidade de dados pode ser coletada, pois qualquer interação do aluno neste ambiente é registrada e deixa um rastro digital (LIÑÁN; PÉREZ, 2015).

A evolução dos sistemas educacionais *online* provocou um aumento considerável no volume de dados armazenados, mudando assim a aprendizagem tradicional, pois apresenta novas possibilidades aos alunos. Entretanto, também ocasiona novos desafios computacionais em função da sobrecarga de informações (WU; LU; ZHANG, 2015).

Assim, a tarefa de descobrir conhecimento relevante e tendências ocultas para subsidiar decisões, deve fazer o uso de técnicas adequadas para atingir esse propósito. A sobrecarga de informações traz aspectos negativos para o aluno, na medida em que torna mais difícil a escolha de conteúdo educacional relevante para suas atividades (TARUS; NIU; YOUSIF, 2017).

Devido a quantidade de informações disponíveis, métodos tradicionais, geralmente baseados na estatística descritiva, já não se mostram suficientes para trazer contribuições significativas para o processo. As plataformas de ensino *online* devem ser desenvolvidas para serem eficientes, levando em consideração o desempenho e as demandas individuais de cada aluno, no sentido de oferecer conteúdo educacional relevante (ANSARI et al., 2017).

Recentemente, técnicas de várias áreas têm sido aplicadas com o intuito de relevar padrões nos dados gerados pela participação de alunos em ambientes de *E-Learning*. Essas técnicas são chamadas de *Learning Analytics* (LA).

O campo de LA surgiu no cenário de educação em virtude do uso da tecnologia, propondo análises para a otimização do aprendizado do aluno (ELLIS; HAN; PARDO, 2017). Busca promover, dentre outros, o engajamento e retenção do aluno, sugerindo recomendações e intervenções. Para isto, realiza a análise dos dados das interações dos alunos com o ambiente de *E-Learning* através de diferentes técnicas estatísticas e de mineração de dados (NA; TASIR, 2017).

A mineração de dados educacionais é uma das técnicas de LA que trata do processo de obtenção de dados úteis a partir de grandes e complexos conjuntos de dados educacionais, no intuito de melhorar a percepção de como os estudantes aprendem e, a partir disso, melhorar os resultados educacionais (ROMERO; VENTURA, 2013).

Learning Analytics representa uma intersecção de diferentes disciplinas acadêmicas (como pedagogia, psicologia, ciência de dados, ciência da computação). Ela sintetiza diversas técnicas dessas áreas para promover ações personalizadas no que tange ao entendimento do aprendizado dos alunos. Dentre as áreas, ressalta-se a análise acadêmica, análise de ações, mineração de dados educacionais, sistemas de recomendação e o aprendizado adaptativo personalizado. Geralmente, as técnicas aplicadas são o aprendizado de máquina, ou inteligência artificial, recuperação de informação, estatística e visualização de informação (CHATTI et al., 2013; FERGUSON, 2012).

Outros termos guardam relação muito próxima da definição de LA. Por exemplo, a Análise Acadêmica é a identificação de padrões significativos nos dados educacionais para informar sobre possíveis problemas acadêmicos, como por exemplo, retenção e taxa de sucesso, a fim de proporcionar o uso de estratégias adequadas de intervenção (CAMPBELL; DEBLOIS; OBLINGER, 2007).

Todos os conceitos anteriores se referem a processar grande quantidade de dados educacionais, mas somente a LA enfatiza a otimização do processo de aprendizado e dos ambientes educacionais em tempo real (IFENTHALER; WIDANAPATHIRANA, 2014).

Neste cenário, o perfil do aluno é uma importante fonte de informações. Drán e Amandi (2008) descrevem que inúmeros modelos de perfil de aluno são compostos de apenas dois tipos de perfis: o perfil do aluno e o perfil do grupo. Assim sendo, propuseram uma estrutura de perfil de aluno composto por três componentes: o perfil individual, o perfil do grupo e o perfil colaborativo, sendo que este último possui informações a respeito das habilidades colaborativas de cada aluno e o contexto onde a habilidade aparece.

Organizações internacionais também procuram padronizar as informações de alunos nos ambientes *online* de aprendizagem. IEEE PAPI, IMS LIP, EDUPERSON e ULF são padrões que surgiram neste sentido (BREMIGARTNER, 2017). No contexto dos sistemas de recomendações educacionais, as informações do perfil do aluno são decisivas, porque influenciam diretamente no resultado e na acurácia das recomendações (AMINI et al., 2014; LUNA et al., 2015).

O uso de ontologias para definição do perfil do aluno também aparece na literatura. Por exemplo, Shao (2017) utilizou um método para descrever as preferências, interesses e

hobbies de alunos em ambientes de *E-Learning* através da utilização de ontologia de domínio, aplicando a lógica difusa para prover serviços personalizados de aprendizagem.

Em Bernardino (2017) é desenvolvido um perfil de educando que coleta as preferências e interações destes com o ambiente de aprendizagem *online*, com o objetivo de realizar recomendações de conteúdos relevantes. Para isso, foram utilizadas abordagens tradicionais da área de Sistemas de Recomendação, além de técnicas estatísticas.

Portanto, LA não envolve apenas a análise das interações dos alunos com o ambiente *online* de aprendizagem, mas também a análise das características particulares, preferências, estilos de aprendizagem e outras informações obtidas a partir do perfil do aluno.

1.1 DEFINIÇÃO DO PROBLEMA

Os avanços tecnológicos estão transformando a maneira como interagimos com o mundo. Nas escolas, com a utilização de computadores e dispositivos digitais, busca-se promover a inclusão digital, assim como, promover metodologias de ensino mais conectadas com a experiência extraclasse dos alunos. Entretanto, a dificuldade em utilizar as tecnologias digitais nos processos pedagógicos ainda existe, entre outros motivos, devido à falta de preparo dos professores, ainda que essa utilização represente inovação (FIUZA; MOCELIN, 2016).

A educação a distância representa a transformação da disponibilidade de oportunidade educacional, e leva a reflexão de como aprendemos e força a revisão do pensamento sobre como ensinamos. Os preconceitos contra o aprendizado que ocorre fora da sala de aula ou ambiente de ensino vêm sendo abandonados, muito em virtude da expansão e desenvolvimento de novas tecnologias. Assim, o principal desafio dos professores é determinar qual informação é a mais importante, e não mais a quantidade (MOORE; KEARSLEY, 2013).

Em virtude das instituições de ensino superior estarem coletando todos os dados disponíveis sobre seus estudantes e registrando-os em base de dados cada vez mais acessíveis e mais complexas, pode-se afirmar que estamos entrando em uma nova era, onde estes dados objetivam impactar positivamente no desempenho do estudante, agilizando processos e promovendo uma utilização mais efetiva dos recursos (MATSEBULA; MNKANDLA, 2017).

Os sistemas educacionais *online* representaram um avanço quando comparado ao sistema de ensino tradicional, pois são capazes de difundir o acesso à informação e à educação, onde cada aluno conta com recursos de interatividade de apoio, como fóruns e bate-papo. Entretanto, esses sistemas educacionais *online* entregam o mesmo conteúdo a todos os alunos baseado no sistema de ensino tradicional, e não consideram as características de cada aluno individualmente (IMRAN et al., 2016).

Com a enorme disponibilidade de sistemas educacionais *online*, é indispensável que estes sistemas se adaptem a grande variação de fatores pessoais e sociais dos alunos, pois estes fatores afetam diretamente os resultados da aprendizagem (SHAWKY; BADAWI, 2018).

Esses sistemas desempenham um papel importante, especialmente em países em desenvolvimento, devido as restrições na estrutura educacional, e muitas vezes representam uma inovação. Porém, a alta taxa de evasão dos cursos *online* requer um esforço adicional dos educadores para motivar a audiência (LOBO; ANSARI, 2015).

Neste cenário, LA utiliza técnicas e métodos buscando promover melhorias tanto no processo de ensino como de aprendizagem. Isto ocorre através de uma avaliação crítica dos dados e na geração de padrões que podem identificar os hábitos educacionais do estudante, promovendo suporte à tomada de decisão, facilitando avaliações realísticas e promovendo supervisão pessoal no progresso do aluno (PEÑA-AYALA, 2017).

Cada estudante tem suas características individuais, preferências e atitudes que podem ser consideradas para enriquecer o seu perfil, de modo que esses dados se agreguem a outros e formem um conjunto de dados com informações significativas.

Uma dessas características é a personalidade. Geralmente estes dados são coletados através de questionários. Contudo, também é possível utilizar métodos não explícitos para modelar a personalidade do aluno através da análise dos vestígios que são deixados nos ambientes de *E-Learning* (TLILI et al., 2017).

Outra forma de enriquecer o perfil do aluno é através da coleta de dados oriundos das plataformas de mídias sociais e utilizar estes dados para avaliar a melhoria das notas dos estudantes nos sistemas de aprendizagem *online*.

Neste sentido, Al-Mamoori, Shehab e Fakharany (2018), utilizam ontologias para criar um perfil de aluno específico para cada aluno, que permite o desenvolvimento de um sistema de aprendizado adaptativo. A proposta, segundo os autores, é capaz de fornecer ao aluno e ao professor conteúdo personalizado e mais significativo, sendo possível identificar o estilo de aprendizado do estudante, suas preferências e sua personalidade. Entretanto, Dutt, Ismail e Herawan (2017) defendem que é necessário a realização de mais pesquisas relacionadas à modelagem do perfil do aluno, principalmente quanto a sua utilização em tarefas de aprendizado de máquina.

Apesar da análise dos dados e dos perfis dos estudantes terem o potencial de aumentar o entendimento das suas necessidades educacionais, a fim de influenciar positivamente o aprendizado, há que se observar ainda as questões éticas e dilemas envolvidos. Como exemplo,

citam-se as condições para consentimento de acesso aos dados, a ocultação de dados de identificação, e a vulnerabilidade dos dados (SLADE; PRINSLOO, 2013).

A pesquisa de Kraivixien, Wongwanich e Sujiva (2014) relatou evidências de que os professores têm dificuldade no entendimento dos requisitos e implementação do perfil do estudante, assim como, dificuldade em relacionar os dados dos estudantes com a preparação dos objetivos da instrução e com o próprio ensino, mostrando um comportamento moderado na utilização e definição desses perfis. Para melhorar a efetividade do processo de ensino, os autores defendem que os professores podem utilizar as informações do perfil do aluno para gerenciar a instrução, escolher o método pedagógico mais adequado e o conteúdo que mais se adequa aos estudantes que têm perfis semelhantes, assim como organizar a sala de aula.

Assim, percebe-se que o perfil do aluno merece atenção na sua correta definição visando contribuir de maneira significativa nos resultados da aplicação das técnicas de LA.

A partir dos elementos identificados, esta dissertação objetiva responder a seguinte pergunta de pesquisa: **“Como definir o perfil de alunos a partir de Sistemas de Aprendizagem *Online* visando possibilitar a aplicação de técnicas de análise de dados?”**

1.2 OBJETIVOS

A seguir são declarados o objetivo geral e os objetivos específicos deste trabalho.

1.2.1 Objetivo geral

Propor um modelo de perfil de aluno voltado à aplicação de técnicas de *Learning Analytics* em Sistemas de Aprendizagem *Online*.

1.2.2 Objetivos específicos

- Identificar os principais elementos presentes na literatura que auxiliem na definição de um perfil de aluno;
- Identificar informações úteis em Ambientes de Aprendizagem *Online* para a aplicação de métodos e técnicas de *Learning Analytics* e Sistemas de Recomendação;
- Definir um modelo de dados de perfil de aluno que contemple uma estrutura lógica que seja adequada ao armazenamento das informações necessárias para as análises;
- Avaliar o modelo de modo a demonstrar o seu potencial para suportar a aplicação de técnicas de análise de dados e extração de informações úteis à tomada de decisão.

1.3 JUSTIFICATIVA E RELEVÂNCIA DO TEMA

Segundo a UNESCO (2017), em seu relatório chamado *Global Education Monitoring Report*, os governos estão cada vez mais interessados na coleta de dados sobre o resultado da aprendizagem dos estudantes. Estas informações, sendo de boa qualidade, objetivam possibilitar aos gestores em áreas educacionais, tanto a nível nacional, estadual e municipal, a tomada de decisões baseadas em evidências, independente de interesses políticos.

Ainda segundo o relatório, ao nível de sistema, os resultados da avaliação ajudam a monitorar a adequação aos padrões mínimos exigidos. No nível individual, podem influenciar na tomada de decisão relativa aos processos de admissão e progressão do estudante. No nível institucional, os dados destinam-se à comparação das escolas entre si, com o objetivo de identificar pontos fracos e propor melhorias.

Entretanto, o relatório assevera que os dados individuais de aprendizagem dos alunos são utilizados de formas diferentes nos países e que a coleta sistemática de dados individuais sobre os resultados da aprendizagem individual devem se tornar mais comuns (UNESCO, 2017, p. 154).

No Japão, priorizaram-se historicamente os exames vestibulares, e somente em 2007 foi introduzido um exame de larga escala nacional. Porém, os exames não permitem uma comparação da evolução do progresso do aluno de tempos em tempos. Na Inglaterra, os resultados são comparados a padrões nacionais, através de um mecanismo de avaliação dos alunos e um sistema de avaliação externa. A análise dos dados é realizada com o objetivo de preparar os inspetores antes das visitas escolares; estabelecer metas; propiciar informação para os pais e sobretudo, identificar os alunos que necessitam de apoio e acompanhamento, além de monitorar o desempenho para finalidade de prestação de contas (UNESCO, 2017).

A avaliação dos professores também é realizada por diversos métodos, dependendo do contexto social e político existentes. Mas o relatório atesta que nenhuma estratégia de medição isolada é capaz de capturar toda a gama de desempenho do professor ou a composição de qualidades importantes para um ensino eficaz. Além disso, diretores, pais e alunos valorizam as diferentes capacidades e conhecimentos dos professores, e têm diferentes percepções e graus de objetividade sobre o ensino de alta qualidade (UNESCO, 2017, p. 92).

O relatório aponta que os professores necessitam desenvolver habilidades para a avaliação do desempenho dos alunos e para analisar os dados dos sistemas educacionais. Todavia, eles não se sentem preparados para a tarefa. Em um estudo conduzido nos EUA, dois terços dos

professores consideram não ter facilidade para manipular os dados a fim de melhorar o processo de ensino e muitos consideraram a quantidade de dados excessiva.

De fato, com a evolução dos sistemas educacionais, o excesso de informações disponíveis e o crescente volume de dados armazenados nas bases traz uma sobrecarga de informações. Todavia, esses dados podem ser analisados para a obtenção de informação útil e relevante. Também é possível moldar os perfis de cada aluno e sugerir atividades adequadas às suas necessidades (WU; LU; ZHANG, 2015).

Quando se realiza a busca por estudos a respeito das técnicas de LA, percebe-se que poucos estudos têm utilizado o perfil do aluno em conjunto com os dados educacionais para a descoberta de conhecimento, deixando assim uma lacuna que pode ser explorada.

Segundo Lotsari et al. (2014), LA envolve a coleta dos dados, que são derivados dos estudantes e do ambiente *online* de ensino que eles participam. Assim, o perfil do aluno é importante porque concentra os dados particulares de cada aluno, que podem ser utilizados posteriormente para diversos tipos de análises estatísticas e técnicas de LA.

1.4 DELIMITAÇÕES DA PESQUISA

A delimitação é uma fase essencial para o desenvolvimento do estudo. Algumas dimensões devem ser traçadas para uma efetiva delimitação: conceitual, escopo, temporal e geográfica.

Na dimensão conceitual, este estudo aborda os ambientes *online* de aprendizagem, especificamente na concepção de um modelo de perfil de aluno com uma quantidade e qualidade de informações relevantes para aplicação de técnicas de LA. Assim, conceitua-se ambientes *online* de aprendizagem, *E-Learning* e *Learning Analytics*, através de uma revisão sistemática da literatura para subsidiar a evolução do estudo.

Quanto ao escopo, pretende-se restringir e direcionar o estudo para um modelo de perfil de aluno válido para ambientes *online* de aprendizagem. Não significa que o modelo não possa ser aplicado em outras situações de aprendizagem. Entretanto, somente nos ambientes *online* de aprendizagem são capturados dados diversos das interações dos alunos com o próprio ambiente, resultando em uma gama de informações muito mais ampla.

Não há pretensão no desenvolvimento de um *software* como produto. Porém, o modelo de perfil de aluno poderia ser incorporado ao desenvolvimento de qualquer ambiente *online* de aprendizagem para proporcionar informações relevantes, análises e recomendações baseadas no perfil do aluno. Portanto, o foco reside em utilizar as informações obtidas através da análise

da literatura para estruturar um modelo de perfil de aluno que consiga entregar informações úteis para a aplicação de técnicas de LA.

Por fim, a revisão realizada considera o ambiente de Ensino a Distância (EaD) como um ambiente *online* de aprendizagem, objetivando definir a constituição do perfil de aluno com a quantidade de informações necessárias e suficientes para aplicação de técnicas de análise de dados.

Com relação às dimensões temporal e geográfica, os dados utilizados são de alunos que frequentaram cursos suportados por um ambiente *online* de aprendizagem no ano de 2017. Na dimensão geográfica, o enfoque reside em uma organização de ensino, localizada no sul do Brasil. Além dos elementos vinculados a delimitação, aspectos com relação ao alcance do estudo também devem ser considerados. É possível estender as análises a outros cenários, mediante uma intervenção para se adequar o modelo de perfil de aluno.

As delimitações apresentadas, quanto ao alcance do estudo, não o restringem ao ponto de ser aplicável a somente um caso específico. Ao contrário, este estudo pode ser válido, útil e aplicável em cenários variados. Destaca-se que, em alguns casos, é necessário realizar adequações ao modelo proposto, as quais podem variar de acordo com o objetivo que se pretende atingir com a análise do perfil de aluno proposto.

1.5 PROCEDIMENTOS METODOLÓGICOS

A definição dos procedimentos é importante uma vez que permite a descrição e classificação da pesquisa em diversos aspectos. Sob o ponto de vista da natureza, este trabalho é caracterizado como uma pesquisa aplicada, pois seu principal interesse consiste em solucionar problemas do dia a dia (DRESCH; LACERDA; JÚNIOR, 2015).

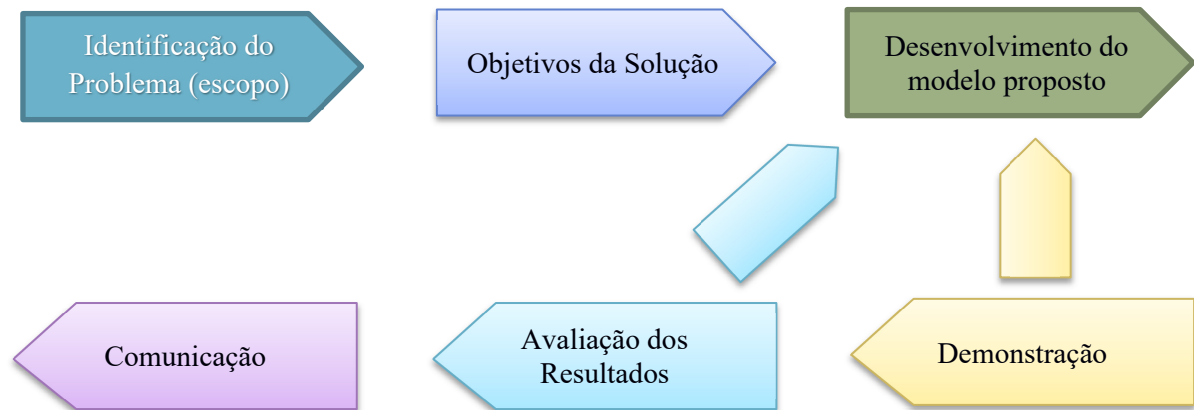
Também pode ser classificado como uma pesquisa tecnológica, pois tem como objetivo a criação de um artefato tecnológico visando uma solução para um problema existente. O objeto da pesquisa tecnológica é conhecimento prescritivo, pois as pesquisas são orientadas à solução de problemas, uma vez que se constitui, em função de um propósito, um objeto com propriedades desejadas e fabricado conforme especificado no projeto (CUPANI, 2011).

Para se alcançar os objetivos, adotou-se a *Design Science Research* (DSR). Essa metodologia possibilita a criação de artefatos para solucionar problemas do mundo real, evoluindo o conhecimento puramente acadêmico e disciplinar, oriundo das ciências tradicionais, para um conhecimento transdisciplinar, com alcance e relevância também fora da academia (DRESCH; LACERDA; JÚNIOR, 2015). Para alcançar os objetivos desta pesquisa, o trabalho foi realizado da seguinte forma:

- Identificação da Problemática: Foram relacionados o conceito de *Learning Analytics* e características do perfil de aluno, a fim de demonstrar a relevância do perfil do aluno para as análises neste campo de estudo;
- Definição dos objetivos: Os objetivos promovem o direcionamento da pesquisa, no sentido de se analisar as soluções para o problema apresentado;
- Desenvolvimento: Constitui na proposição do modelo para armazenar as informações sobre o perfil do aluno e demais dados estatísticos de acesso ao ambiente online de aprendizagem;
- Avaliação: Consiste em popular o modelo com os dados oriundos da plataforma de ensino e a realização de simulações e avaliações, visando demonstrar a viabilidade do modelo;
- Análise: Representa a análise dos resultados obtidos por meio do modelo proposto, de acordo com os objetivos propostos;
- Considerações finais: Apresenta as conclusões e perspectivas de trabalhos futuros.

A metodologia DSR de Peffers (2007), foi a escolhida para operacionalização da pesquisa (Figura 1). A primeira etapa consiste na identificação do problema e sua motivação, considerando a relevância e importância, além da aplicabilidade da solução. Na segunda etapa, são definidos os objetivos ou resultados esperados. A terceira etapa consiste em projetar e desenvolver o artefato que dará suporte à solução do problema. A quarta atividade representa a demonstração da viabilidade do artefato, quando eventualmente pode haver a necessidade de novas alterações, retornando assim à etapa anterior. A quinta atividade consiste na avaliação, ou seja, comparar os resultados obtidos com os requisitos definidos na segunda etapa. Aqui também, caso haja necessidade, deve-se retornar à etapa anterior para promover modificações. Finalmente, chega-se a fase de comunicação, onde é demonstrado, além do rigor com que a pesquisa foi realizada, a sua eficácia (DRESCH; LACERDA; JÚNIOR, 2015).

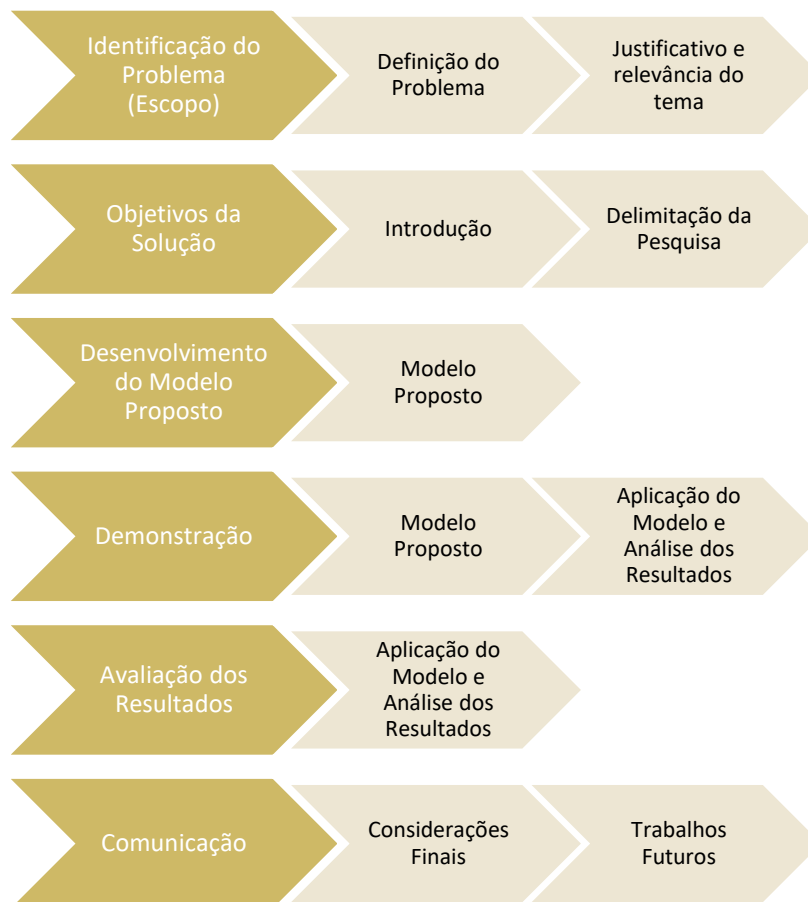
Figura 1 - Procedimentos da metodologia DSR.



Fonte: Elaborada pelo autor (2019).

A Figura 2 busca representar a equivalência dos tópicos da presente dissertação com a metodologia DSR.

Figura 2 - Metodologia x estrutura do trabalho.



Fonte: Elaborada pelo autor (2019).

Durante a realização das primeiras pesquisas, foi definido o escopo e a delimitação dos assuntos abordados, com a anuência do orientador. Buscando o objetivo de se estabelecer um protocolo de pesquisa, foram definidas as palavras-chave *Learning Analytics*, perfil de aluno, ambientes online de aprendizagem, *student profile*, *online learning environments*. A definição da busca é importante para que o procedimento de revisão sistemática da literatura seja reproduzível. Desta forma, as buscas foram realizadas nas seguintes bases: *Scielo*, *Web of Science*, *ACM*, *Scopus*, *IEEE* e *Springer Link*.

Após a revisão da literatura concluída, definiu-se o modelo proposto para essa dissertação. O objetivo do modelo proposto é oferecer aos *stakeholders* informações confiáveis e relevantes para auxiliar no entendimento do processo de aprendizagem, assim como, a análise, a avaliação e a seleção de recursos educacionais que mais se adequam aos perfis dos alunos.

A partir da definição do modelo proposto, o mesmo foi criado em um banco de dados e populado com o objetivo de atestar a sua viabilidade, com os dados oriundos de uma plataforma de ensino a distância.

Na etapa seguinte, procurou-se definir os cenários e a metodologia para utilização do modelo proposto, a partir da análise dos dados educacionais para aplicação das técnicas de LA escolhidas. Nesta etapa, foram utilizadas as técnicas de análise de agrupamentos e sistemas de recomendação.

Com o cenário definido, executou-se a aplicação das técnicas selecionadas nos dados do perfil do aluno e avaliou-se os resultados obtidos. O objetivo proposto para essa etapa consistiu na análise das informações que podem ser extraídas do modelo proposto e que auxiliem tanto os gestores educacionais na tomada de decisão, quanto aos alunos, através de recomendações pertinentes aos seus perfis.

A etapa seguinte foi atingida com escrita da dissertação.

1.6 ADERÊNCIA AO PPGE GC

Com o objetivo de evidenciar a aderência da dissertação ao objeto de pesquisa do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGE GC), definiu-se três itens para serem abordados: a identidade da dissertação, o contexto estrutural no EGC e as referências factuais.

1.6.1 Identidade

A presente dissertação está relacionada à área de concentração de Engenharia do Conhecimento, na linha de pesquisa: Teoria e Prática em Engenharia do Conhecimento, cujo objetivo é estudar metodologias e técnicas da Engenharia do Conhecimento e da Inteligência Computacional e suas relações com a gestão e com a mídia e conhecimento (EGC, 2018a).

A aderência deste trabalho ao objeto de pesquisa do Programa de Pós-Graduação¹ em Engenharia e Gestão do Conhecimento pode ser reforçada a partir da análise de seu viés interdisciplinar:

O programa articula conexões interdisciplinares destacando três ênfases de formação e pesquisa: (1) explicitação, emulação e modelagem do conhecimento, englobando a criação, a descoberta, a aquisição, a formalização, a codificação, o armazenamento, a distribuição e uso de conhecimento (Engenharia); (2) planejamento e alinhamento coletivo e organizacional do conhecimento compostos de sub processos como integração, avaliação, auditoria, retenção-descarte, criação-inovação, propriedade e evolução do conhecimento (Gestão); e (3) difusão, comunicação e compartilhamento do conhecimento que abrange a preservação, disseminação, transferência, socialização e acesso ao conhecimento (Mídia) (EGC, 2018b).

Pacheco, Tosta e Freire (2010) descrevem a trajetória interdisciplinar do PPGEKC como um processo em construção. Esse processo lida com o paradoxo de estar situado dentro de um departamento disciplinar, mas onde as disciplinas se conhecem, se respeitam e ainda podem convergir para “um todo que um dia já foi naturalmente unido”, resultando na integração do conhecimento.

O modelo proposto está baseado sobre o processo de análise de informações educacionais resultantes da interação dos alunos com o ambiente *online* de ensino, assim como de outras informações julgadas relevantes que compõem o seu perfil. Conhecimento é conteúdo ou processo resultante de interações sócio técnicas entre agentes humanos e tecnológicos que geram valor científico, social, econômico ou cultural (EGC, 2018b). Assim, o conhecimento neste trabalho se faz presente nas análises das informações educacionais, que objetivam subsidiar os gestores com informações úteis e relevantes para o suporte a decisões gerenciais pedagógicas.

Esta proposta tem como objetivo definir um modelo de perfil de aluno para aplicação de métodos e técnicas da Engenharia do Conhecimento a fim de possibilitar a análise de dados educacionais em ambientes de aprendizagem *online*.

¹ Disponível em: <http://www.egc.ufsc.br/pos-graduacao/programa/>

1.6.2 Contexto Estrutural do EGC

A Engenharia do Conhecimento tem suas origens na área da Inteligência Artificial. Assim, o objetivo dela guarda similaridade com o objetivo da Engenharia de *Software*: tornar o processo de desenvolver sistemas baseados em conhecimento em uma disciplina de engenharia, tendo como objetivo pesquisar métodos e técnicas para a extração e classificação do conhecimento (STUDER; BENJAMINS; FENSEL, 1998).

De acordo com Schreiber et al., (2000), a Engenharia do Conhecimento envolve a aplicação de métodos, técnicas e ferramentas voltadas a modelagem, classificação e representação do conhecimento, sendo reconhecida como uma atividade de modelagem, oferecendo o ferramental necessário para formalizar e explicitar o conhecimento.

Com relação ao escopo deste trabalho, o aspecto que a contextualiza na área de Engenharia do Conhecimento reside no fato do método possuir como objetivo a materialização dos macroprocessos de explicitação, formalização, disseminação e gestão do conhecimento.

O presente trabalho está centrado na linha de pesquisa Teoria e Prática em Engenharia do Conhecimento, que busca, por ferramental computacional, apoiar a processos de aquisição, classificação e representação do conhecimento, permitindo que esses processos possam ser geridos e utilizados pelas três áreas do programa: engenharia, gestão e mídia do conhecimento.

Neste sentido, a área de *Learning Analytics* envolve a utilização de ferramentas para promover a melhoria da eficiência das práticas de ensino, que poderiam ficar limitadas caso fossem abordadas em um programa unicamente disciplinar.

1.6.3 Referências Factuais

O presente trabalho possui como foco principal as áreas de Sistemas de Aprendizagem *Online* e *Learning Analytics*, notadamente relacionados aos perfis de estudantes. A seguir, serão apresentados os trabalhos desenvolvidos no programa que estão relacionados ao foco do trabalho.

No domínio de Sistemas de Aprendizagem *Online*, relacionados ao perfil do aluno, foi encontrado o seguinte trabalho:

- AMORIM, João Schorne de. O perfil do aluno na educação a distância: um estudo sobre a inclusão digital na Polícia Militar de Santa Catarina. Dissertação, 2012.

Na área de *Learning Analytics*, não foram encontrados trabalhos. No intuito de ampliar

a busca, também foi utilizado o termo “Mineração de Dados Educacionais”, “Ambientes de Aprendizagem *Online*”, sendo que nenhum trabalho foi encontrado. Modificando os termos de pesquisa para “AVA”, “MOOC” e “EaD”, os seguintes trabalhos (Quadros 1 a 3) foram encontrados nos últimos cinco anos (2015-2019):

Quadro 1 – Trabalhos no EGC com o termo: EaD

Tipo	Título	Autor(a)	Ano
Tese	Möbius: um modelo para Polos EaD.	DIANA, Juliana Bordinhão.	2019
Tese	Gestão em educação a distância: um <i>framework</i> baseado em boas práticas.	NUNES, Carolina Schmitt.	2017
Tese	Processos flexíveis para a produção de materiais didáticos para a Educação a Distância: recomendações pautadas na perspectiva interdisciplinar.	BLEICHER, Sabrina.	2015
Tese	Diretrizes para práticas de gestão do conhecimento na educação a distância.	ALRCON, Dafne Fonseca.	2015
Dissertação	Servqual como instrumento de gestão da Qualidade da Informação em ambiente EaD.	BENTANCOURT, Silvia Maria Puentes.	2015
Tese	Influência da Competência Empreendedora dos Coordenadores nos Indicadores de Desempenho dos Polos Ead.	WOLF, Sérgio Machado.	2014

Fonte: Elaborado pelo autor (2019).

Quadro 2 – Trabalhos no EGC com o termo: MOOC

Tipo	Título	Autor(a)	Ano
Tese	Framework conceitual de elementos da mídia digital MOOC: tecnologia pedagógica com ênfase em videoaulas.	LOPES, Luciana Dornbusch.	2019
Tese	A colaboração e inclusão de cursos <i>online</i> abertos e massivos (MOOCs).	KUNTZ, Viviane Helena.	2019
Dissertação	Acessibilidade e usabilidade em plataformas MOOC: indicadores de boas práticas.	NASCIMENTO, Heluiza Ormelez de Almeida.	2019
Dissertação	Avaliação da Qualidade da Informação de cursos massivos: um estudo de caso do Telelab.	BIAGIOTTI, Breno de Almeida.	2016
Dissertação	Avaliação e seleção de plataforma para cursos <i>online</i> abertos e massivos em instituições de ensino superior.	EBONE, Denise Santin.	2015

Fonte: Elaborado pelo autor (2019).

Quadro 3 – Trabalhos no EGC com o termo: AVA

Tipo	Título	Autor(a)	Ano
Dissertação	Adaptação como mídia para o conhecimento: uma análise de ambientes virtuais de aprendizagem utilizados em disciplinas de graduação.	SIMON, Rangel Machado.	2017
Tese	Aplicação de um modelo adaptativo de tutores inteligentes para disseminação do conhecimento em ambientes virtuais de ensino-aprendizagem.	PALOMINO, Cecilia Estela Giuffra.	2017
Tese	Criação de Valores em Comunidades de Prática: Um Framework para um Ambiente Virtual de Ensino e Aprendizagem Bilíngue.	PIVETTA, Elisa Maria.	2016

Fonte: Elaborado pelo autor (2019).

Apesar de não ter relação direta com o presente trabalho, foi encontrado o trabalho abaixo que propõe um modelo para a extração automática de perfis de especialistas a partir de documentos não estruturados. Isto poderia ser adaptado para realizar a extração automática do perfil de alunos, visando possibilitar a aplicação de técnicas de análise de dados no contexto de LA:

- TAXWEILER, Rudger Nowasky do Nascimento. Um modelo para a extração de perfil de especialista aplicado às ferramentas de *Expertise Location* e apoio à gestão do conhecimento. Dissertação, 2016.

Dessa maneira, esta dissertação busca contribuir com aspectos que guardam alguma relação com estes trabalhos, procurando descrever um perfil de aluno que seja adequado para a aplicação de técnicas de *Learning Analytics*.

Diante das referências apresentadas, a presente dissertação é aderente ao PPGEGC por propor um modelo de perfil de aluno, utilizando conhecimentos de diversas áreas, visando subsidiar os gestores com informações relevantes para auxiliá-los na tomada de decisão no contexto educacional. Além disso, o modelo de perfil viabiliza uma avaliação mais personalizada, possibilitando a recomendação de atividades, cursos, entre outros, aos alunos. Compreende-se que o presente trabalho está de acordo com a área de concentração de Engenharia do Conhecimento, pois objetiva promover suporte a área de Gestão do Conhecimento, e possui os trabalhos anteriormente elencados que abordam temáticas semelhantes.

1.7 ESTRUTURA DO TRABALHO

O presente trabalho é composto de 5 capítulos descritos a seguir:

- O primeiro capítulo apresenta o tema e os objetivos desta pesquisa, suas delimitações e aderência ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento;
- O segundo capítulo é composto do referencial teórico, em que são abordados os assuntos principais relacionados a pesquisa, sendo eles, Ambientes de Aprendizagem *Online*, *Learning Analytics* e Sistemas de Recomendação;
- O terceiro capítulo apresenta o modelo proposto por meio de uma descrição detalhada sobre as etapas e estruturas que o compõem;
- O quarto capítulo apresenta a proposição de avaliação do modelo e a discussão dos resultados alcançados por meio de um cenário de estudo;
- O quinto e último capítulo apresenta as considerações finais da dissertação e as sugestões de trabalhos futuros. Por fim, são disponibilizadas as referências utilizadas na pesquisa.

2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo apresentar o referencial teórico dos temas que são imprescindíveis para o desenvolvimento deste trabalho. A fundamentação teórica apresenta os principais conceitos relacionados ao desenvolvimento do método.

Neste capítulo, abordam-se os sistemas de aprendizagem online, técnicas de *Learning Analytics*, perfis de recomendação, assim como o perfil do aluno nestes ambientes.

2.1 SISTEMAS DE APRENDIZAGEM ONLINE

2.1.1 Sistemas Educacionais

A gestão de um sistema, presencial ou a distância, envolve as atividades de coordenação, orientação, execução e acompanhamento das atividades para atingir os objetivos esculpidos no planejamento e escopo do projeto (LENZI, 2010). Desta forma,

“Aqueles que se dedicam ao ensino a distância não estão isentos dos mesmos desafios encontrados no sistema tradicional de ensino. Devem dotar-se das mesmas competências comuns a quaisquer gestores, mas o cenário no qual exercem suas atividades é bastante diferente (RUMBLE, 2003, p. 14).”

Um sistema educacional é um *software* ou programa de computador, desenvolvido por meio de linguagem de programação, que auxilia no processo de ensino e aprendizagem. Podem ser divididos nas seguintes categorias: exercício e prática, simuladores, jogos educacionais, resolução de problemas, pacotes utilitários, ferramentas de referências e tutoriais (SANTOS, 2016).

Entretanto, para ser educativo, um *software* deve ter sido concebido com base na teoria da aprendizagem, com vistas à construção do conhecimento e à colaboração entre os atores envolvidos, através de um ambiente interativo (RIBEIRO, 2016).

Para Bertoletti et al. (2003), um sistema educacional pode ser visto como um programa utilizado com uma finalidade educacional. Percebe-se que a utilização de recursos tecnológicos no processo de ensino é cada vez mais necessária para permitir que os alunos construam seus saberes a partir de interações em um mundo de pluralidades, sem limitações geográficas e culturais (DE OLIVEIRA; MOURA; SOUSA, 2015).

Uma vez que o indivíduo pode escolher qual assunto ele deseja incorporar, no seu próprio ritmo, dentro de suas possibilidades, o *software* educacional representa uma vantagem na complementação da formação do indivíduo.

Um dos fatores que contribuíram para a evolução dos ambientes e *softwares* educacionais foi a popularização das tecnologias e aplicativos relacionados ao uso de tecnologias da informação e comunicação, assim como, mais recentemente, a Web Semântica, pois permitem um ambiente adaptável, com possibilidade de compartilhamento e reuso da informação, assim como estabelecem novas formas de colaboração entre os alunos (KAMAL; MOHAMED; ALI, 2016; ZEM-LOPES et al., 2014).

Com a evolução para os sistemas *online* de aprendizagem, novas possibilidades foram criadas para os alunos não só utilizarem a tecnologia para a recuperação de informações, mas para a produção criativa, para a comunicação e o compartilhamento de informações, tornando-os parte ativa que contribui no conteúdo online (NACU; MARTIN; PINKARD, 2018).

2.1.2 Sistemas Educacionais *Online*

De acordo com Dias e Gomes (2004), os sistemas educacionais online são aplicações para a *web* que disponibilizam um conjunto de funcionalidades, com o objetivo de criar e gerenciar um espaço onde os alunos têm acesso ao conteúdo do curso e à interação com os professores e demais alunos. Além da perspectiva pedagógica, também integra funcionalidades na perspectiva administrativa, como gestão de turmas e calendários, alocação de professores e gestão de planos de formação.

A quantidade significativa de informações disponibilizadas por estas plataformas permite que elas sejam utilizadas para pesquisar padrões que podem ajudar aos alunos e professores em seus objetivos educacionais. Para isso, uma das alternativas é fazer uso de ferramentas de *Big Data* (CANTABELLA et al., 2019; KIM et al., 2016; LI; ZHAI, 2018; LOGICA; MAGDALENA, 2015; MARTINS; MIGUÉIS; FONSECA, 2018; MATSEBULA; MNKANDLA, 2017; POONSIRIVONG; JITTAWRIAYNUKON, 2018; SCHWERDTLE; BONNAMY, 2017; UDUPI; SHARMA; JHA, 2016).

Entretanto, há de se observar que modelos de EaD que desconsideram as necessidades e dificuldades dos alunos, visando apenas a formação em massa, através da transmissão vertical de conhecimento, pouco contribuem para uma educação inclusiva e para o sucesso da aprendizagem (RABELLO; PEIXOTO, 2011).

Uma das vertentes que podem ser consideradas para a tarefa de se obter dicas de como os estudantes aprendem e obter informações para o futuro desenvolvimento dos sistemas, é baseada em seu próprio uso atual. Assim, a análise dos arquivos de *log* gerados pelos sistemas educacionais *online* constituem uma importante fonte de informações para a melhoria dos sistemas (HANSEN et al., 2017).

As funcionalidades de um sistema educacional *online* podem ser bastante simples ou elaboradas. Assim, para o desenvolvimento de um sistema que possua os recursos de análise de documentos, gráficos, perguntas e respostas, notas, materiais de cursos, provas e vídeos, algumas considerações devem ser levadas em conta, como a estabilidade e eficiência do sistema, nas centenas ou milhares de operações de reprodução de vídeo, *download* de arquivos, etc. (SIKE; WEIGUO, 2017).

2.2 EDUCAÇÃO A DISTÂNCIA

A evolução para uma educação mediada por tecnologias digitais, disponível através de diversos tipos de conectividade provocou um aumento de metodologias educacionais, dentre as quais se destaca a Educação a Distância (EaD) (SILVA; DIANA; SPANHOL, 2015).

Segundo Moore e Kearsley (2013), educação a distância é o aprendizado planejado que ocorre em um lugar diferente do ensino tradicional, requerendo comunicação por meio de tecnologias e uma organização institucional especial.

Dentre os termos geralmente utilizados para refletir o mesmo tema, encontram-se *E-Learning*, ensino *online* e aprendizado distribuído. Em alguns países, os termos utilizados são: educação aberta; aprendizado aberto; e aprendizado aberto e a distância (*Open and Distance Learning – ODL*).

Em uma dimensão ampla, EaD “antes de ser a distância, é educação, com todas as implicações desse fato” (NOVAK, 2010, p. 18).

A modalidade de educação a distância apresenta como principal característica a separação física e geográfica entre alunos e professores, sendo que, quando ofertada e mediada pela *internet*, as interações podem ser estabelecidas independentemente dos participantes estarem conectados ao mesmo tempo na *web* (MACHADO, 2013).

Litto e Formiga (2009) consideram que o termo Educação a Distância, cuja abreviatura é EaD, um termo mais genérico e amplo, sendo historicamente mais usado. Este termo ainda considera um fator de fundamental importância em qualquer processo de formação: o ensino, diretamente relacionado com a atuação docente.

Saba (2003) define que a modelagem da educação a distância representa um comportamento complexo, adaptativo e não linear de vários agentes, que necessitam de ferramentas tecnológicas capazes de lidar com a tarefa.

O primeiro formato de educação a distância, utilizado principalmente por escolas americanas, foi o estudo por correspondência. Entretanto, não há consenso na literatura sobre a origem da educação a distância.

Segundo Landim (1997), no ano de 1728 a Gazeta da Cidade oferecia cursos por correspondência. Litto e Formiga (2009), afirmam que um dos cursos oferecidos era o curso de taquigrafia, ministrado pelo Professor *Caulleb Phillips*. Segundo o professor, os alunos poderiam aprender em casa as lições que eram ministradas em sua escola. Mas o autor afirma que a educação a distância pode ter surgido muito antes, na época em que Platão escrevia para Dionísio ou talvez quando o Apóstolo Paulo escrevia suas epístolas para os primeiros cristãos.

Outras iniciativas mundiais presentes na literatura também merecem citação. Na Suécia, em 1833, foi oferecido um curso por correspondência; em 1883, a Universidade do Estado de Nova Iorque iniciou a concessão de diplomas de bacharéis por correspondência; o curso de Hebreu oferecido pela Universidade de Chicago em 1881; em 1889, os cursos a distância do *Queen's College*, do Canadá, visaram a diminuição de custos, tendo em vista o tamanho continental do país (LANDIM, 1997; LOYOLLA e PRATTES, 1998).

No século 19, a utilização do ensino a distância é relacionada com a Era Industrial, pois a descoberta de novas tecnologias e o crescimento do conhecimento científico necessitavam de mão de obra qualificada para enfrentar os desafios econômicos e sociais que surgiram (COSTA; FRANCO, 2006).

No Brasil, a primeira menção oficial ao termo Educação a Distância, na legislação educacional do País, foi com a publicação da LDB 9.394/96, sendo que o Art. 80 foi específico para tratar de questões referentes a essa modalidade de ensino. O artigo descreve que “O Poder Público incentivará o desenvolvimento e a veiculação de programas de ensino a distância, em todos os níveis e modalidades de ensino e de educação continuada” (BRASIL, 1996, p. 27833).

Em 10/03/2016, o MEC aprovou o parecer número 564/2015 (BRASIL, 2015), da Câmara de Educação Superior do Conselho Nacional de Educação, que propôs as Diretrizes e Normas Nacionais para a oferta de Programas e Cursos de Educação Superior na Modalidade a Distância. Instituída pela Resolução nº 1, de 11 de março de 2016 (BRASIL, 2016), consolidou um novo marco regulatório da educação a distância em ensino superior no Brasil. Neste então novo instrumento legal, o conceito de EaD foi ampliado em relação ao contido no Decreto 5.622/2005, que regulamenta a Lei de Diretrizes Básicas em relação a educação a distância no Brasil.

Segundo Holmberg (2003), umas das descrições mais detalhadas sobre as características da educação a distância foi realizada por Keegan (2013):

- A quase permanente separação entre professor e aluno durante a duração do processo de aprendizado (uma das características que distingue a educação a distância daquela realizada face a face);
- A influência da organização de ensino no planejamento e preparação dos materiais para o aprendizado e na oferta de serviços de suporte para os estudantes;
- O uso de diversos tipos de mídias, como impressões, áudio, vídeo ou computadores;
- A disponibilização da possibilidade da comunicação bidirecional;
- A quase permanente ausência de um grupo de estudos durante o processo de aprendizado, pois as pessoas assistem as aulas individualmente, não em grupos, mas com a possibilidade de encontros ocasionais.

Bates (2005) assevera a existência de três gerações da educação a distância. A primeira foi caracterizada pelo uso de uma única tecnologia, sem interação com a instituição de ensino. Os recursos educacionais que geralmente foram utilizados nesta geração foram o rádio e a televisão. Mas a principal forma era a educação por correspondência.

A segunda geração utilizou mídias impressas e de rádio, sendo que os materiais produzidos eram específicos para o uso na educação a distância. Tinham um ótimo *design*, a produção e distribuição eram centralizadas, a transmissão da informação era unilateral, faziam uso de grandes sistemas burocráticos e tinham resultados economicamente viáveis. Era considerado um método industrial por essência, já que utilizava métodos de produção em massa de produtos padronizados. Um exemplo que pode ser citado é a *British Open University*.

A terceira geração, também descrita como baseada no conhecimento, utilizou mídias de comunicação bidirecional, como *internet* e videoconferência, que permitem não só a interação entre o professor e o estudante, mas principalmente entre os estudantes. Possui como característica um pequeno grupo de desenvolvimento, com serviços administrativos baseados na *web*. Proporciona cursos customizados, produção rápida com pouco investimento inicial. Como exemplo, podem-se citar as universidades convencionais que também atuam na educação a distância, ou seja, trabalhando a vertente do ensino presencial e *online*.

Clark e Mayer (2016), definem *E-Learning* como atividades de ensino fornecidas por um dispositivo digital, como um computador ou telefone celular, que tem como objetivo promover suporte ao aprendizado.

Já Moore e Kearsley (2008), defendem a existência de cinco gerações. A primeira seria equivalente ao estudo por correspondência, que ocorreu em virtude da descoberta da escrita. A segunda envolve a utilização do rádio e televisão. A terceira envolve a utilização das mesmas

tecnologias, mas sob outro prisma: o da universidade aberta. A quarta geração é caracterizada pela utilização de plataformas midiáticas e a teleconferência. Por fim, a quinta geração, que é baseada na utilização da *internet*, possibilitando a construção de ambientes virtuais de aprendizagem assíncronos (Figura 3).

Figura 3 – As cinco gerações de EaD.



Fonte: Moore e Kearsley (2008) – Adaptado pelo autor.

Segundo dados da ABED – Associação Brasileira de Educação a Distância (Tabela 1), em seu censo 2017-2018, os indicadores de qualidade de EaD considerados mais relevantes pelos respondentes, de acordo com a escala de *Likert* de 1-5, são os seguintes:

Tabela 1 – Indicadores de Qualidade de EaD.

Indicador de qualidade	Média de concordância
Conteúdos corretos e atualizados	4,84
Professores qualificados	4,82
Atendimento e necessidades dos alunos	4,78
Tutores qualificados	4,78
Metodologias eficazes	4,76
Tecnologia confiável para a docência	4,69
Tecnologia confiável para a gestão	4,69
Conteúdos atraentes	4,68
Infraestrutura em geral	4,59
Metodologias inovadoras	4,53
Tutores persistentes	4,5
Tecnologia inovadora para a docência	4,47
Conteúdos variados	4,46
Gestão inovadora	4,44
Tecnologia inovadora para a gestão	4,44
Baixas taxas de evasão	4,02

Fonte: ABED (2017).

2.3 AMBIENTES VIRTUAIS DE APRENDIZAGEM - AVA

O desenvolvimento e a evolução da *internet* possibilitaram a criação de plataformas de ensino, ou seja, espaços virtuais que organizam a prática pedagógica em uma sala virtual, impondo ao professor grande responsabilidade na organização pedagógica.

Essas plataformas, usualmente são denominadas AVA (Ambientes Virtuais de Aprendizagem) ou AVEAs (Ambientes Virtuais de Ensino e Aprendizagem). O segundo termo é considerado mais genérico, pois valoriza o papel do professor no planejamento, implementação, acompanhamento e avaliação das atividades didáticas mediadas por esses ambientes (MACHADO, 2013).

Essas plataformas geralmente são implementadas na *web*, e consistem de um conjunto de ferramentas, assíncronas ou não, que permitem o desenvolvimento de ambientes colaborativos de aprendizagem, proporcionando atividades educacionais flexíveis e autônomas (GARCÍA-ÁLVAREZ; NOVO-CORTI; VARELA-CANDAMIO, 2018; AL GHAMDI; SAMARJI; WATT, 2016).

Esses sistemas computacionais, também podem ser utilizados como suporte e apoio na modalidade de ensino presencial, no sentido de complementar o conteúdo trabalhado em sala de aula, pois agregam diversos recursos e ferramentas que permitem a interação e a comunicação entre alunos e professores separados geograficamente e temporalmente.

Spalenza et al. (2018), consideram que os AVAs modelam as informações para auxiliar na evolução da disciplina, permitindo a organização das atividades, proporcionando amplo acesso ao conteúdo, devido a alta disponibilidade deste tipo de sistema. Os autores ainda ressaltam a relevância da coleta de dados educacionais, coadunando com o objetivo desta dissertação.

Apesar da capilaridade dos AVAs, proporcionada pelo uso da tecnologia, um dos problemas frequentemente apresentados é a alta taxa de evasão, não sendo possível a identificação de uma única causa (ABED, 2017).

Com relação a usabilidade e design, alguns aspectos devem ser considerados, como por exemplo, os aspectos de usabilidade e *design* de interfaces (DE ARAÚJO LIMA; MERINO; TRISKA, 2018).

Pereira, A. (2007) resalta as seguintes características presentes nos AVAs:

- O acesso ao ambiente do AVA é realizado por meio da identificação do usuário;
- Usuários não cadastrados possuem visão limitada do conteúdo, ou seja, existem espaços privativos para compartilhamento de arquivos e/ou mensagens;
- Os passos do estudante na plataforma são acompanhados através do registro do progresso das atividades e páginas consultadas;
- É possível disponibilizar matérias em momentos específicos dentro da plataforma, através dos recursos disponibilizados, como o recurso calendário;
- Avaliação formativa;
- Há formas variadas de comunicação e interação, como chat, blog, fórum, etc.;
- Há possibilidade de administrar outros recursos, como FAQ e sistema de busca.

Os AVA mais conhecidos são o Moodle[®], o EdX[®], o GoogleClassroom[®], o AMADEUS[®] [Gomes et al. 2009], o TelEduc[®] [Rocha 2002] e o MORFEU[®] [Menezes et al. 2008], e tiveram papel fundamental na intensificação do EaD e dos MOOC's [Dahlstrom et al. 2014].

Estes ambientes constituem a fonte primária de informações sobre o aluno, pois todas as suas interações podem ser monitoradas e armazenadas, servindo como subsídio para a análise

e descoberta de conhecimento. Este conhecimento, então, pode ser aplicado para a melhoria do processo de ensino através de técnicas e métodos apropriados.

2.4 MOOCs

Mattar (2013) informa que os MOOCs são uma tentativa de ampliação do modelo conectivista de aprendizagem para larga escala e os define como cursos *online*, abertos e massivos. Entretanto, o autor assevera que o conceito padece de maiores considerações, em função da diversidade de atributos que podem compor este modelo de negócio.

Por exemplo, o conceito de ser “aberto” tem problemas, pois em vários desses cursos, caso o aluno queira receber um certificado ao final, precisa pagar. Um curso sem pré-requisito, não significa necessariamente que é aberto, pois necessita de inscrição e pode ou não utilizar uma plataforma fechada. O conceito de curso massivo também é impreciso, na medida em que não se consegue estabelecer com clareza qual a quantidade de alunos um curso deveria ter para se considerado massivo. O fato é que, conforme Mattar (2013) indica, essas atividades de aprendizagem informal competem pelo tempo com outras atividades do dia a dia do estudante.

Segundo Ma, Lee e Kuo (2013), o MOOC pretende fornecer acesso aberto, baseado em um modelo de educação a distância, promovendo uma participação interativa em larga escala. Johnson et al. (2013), afirmam que os atuais modelos de MOOCs reproduzem amplamente os formatos de aulas expositivas, não incluindo as noções de abertura e conectivismo propostas por Siemens e Downes (2008).

Dentre os atributos que são consenso na literatura a fim de caracterizar MOOCs, pode-se citar os seguintes (GARDNER; BROOKS, 2018):

Massivos, abertos e online – São os atributos mais representativos associados aos MOOCs. O termo massivo indica a existência de uma quantidade maior de estudantes, mesmo quando comparado às maiores salas de aula tradicionais. Geralmente, o time de instrutores que promovem suporte aos cursos é pequeno, o que resulta em um número maior de estudantes por professor, quando comparado com a educação superior tradicional ou cursos de *E-Learning*. São abertos, porque estão disponíveis para qualquer estudante, sendo geralmente públicos e gratuitos, apesar do modelo de negócios ter percebido a dificuldade em se manter os cursos livres e gratuitos. São *online*, ou seja, baseados na *internet*, onde todos os materiais dos cursos são acessados via *web* através de um computador ou outro dispositivo com um navegador ou uma aplicação específica para a plataforma.

Sem reprovação ou notas – A participação no curso tem a opção de ser livre de certificação, créditos ou programas de graduação, ou com penalidades para o mau desempenho

ou reprovação no curso. Portanto, os participantes ou tem pouco ou nenhum investimento para completar o curso, o que torna a tarefa de prever o sucesso ou não de um aluno uma tarefa bastante difícil. Nesta concepção, os cursos online pagos ou que objetivam o acúmulo de créditos não são considerados MOOCs. Apesar de existir diversos outros ambientes que oferecem cursos sem reprovação ou notas, como tutoriais, essas plataformas não oferecem as mesmas funcionalidades de um MOOC.

Assíncronos – Apesar de possuírem uma data inicial e final, o aluno tem total liberdade para completar as tarefas, na ordem que desejar e a qualquer tempo. Os autores afirmam que um modelo totalmente assíncrono passou a ser utilizado em algumas plataformas, indicando uma mudança no modelo de negócio, pois a inscrição deixa de ser por curso e passa a ser por assinatura.

Heterogêneo – A heterogeneidade revela-se tanto no aspecto demográfico como na variedade de intenções. O histórico dos estudantes pode variar desde pessoas com diploma na área abordada pelo curso, pessoas que trabalham diretamente na área ou pessoas que não possuem nenhuma formação, com diferentes idades e de países diferentes. Entretanto, a simples tentativa de obter dados demográficos é difícil, pois fica restrita a resposta de um formulário não obrigatório, que possui baixo nível de respostas, o que resulta na impossibilidade de utilização destes dados pelos sistemas preditivos, que então focam no comportamento *online* dos alunos.

Mattar (2013) e Kaushik (2015) reportam que o primeiro MOOC foi oferecido por Siemens e Downes em 2008, com o tema *Connectivism and Connective Knowledge*, tendo cerca de 2.400 inscritos, na Universidade de Manitoba, Canadá. Na ocasião, o termo foi cunhado por Dave Cormier para descrever esse curso (YUAN; POWELL, 2013). Já no primeiro curso foi identificado o alto número de desistências, associado a falta de tempo, barreiras de linguagem, diferença de fusos horários e dificuldades com o uso da tecnologia.

Quando os cursos se assemelham às atividades desenvolvidas em um curso presencial, e o professor ainda exerce o papel de disseminador do conhecimento, possuindo assim um caráter instrucionista, são considerados xMOOCs (BOND; LEIBOWITZ, 2013). Desta forma o professor é quem cria o conteúdo a ser ministrado, assim como as atividades, em forma de etapas a serem percorridas, com debates incentivados, mas direcionados.

Caso os cursos sejam baseados na teoria de Siemens (2005), são considerados cMOOCs (baseados no conectivismo), pois os alunos são convidados a reportar conteúdos externos que possam enriquecer o debate, caracterizando o conceito de redes. São coautores do

conteúdo. Por vez, quando a tecnologia é utilizada apenas como suporte aos cursos presenciais é considerado *Blended Learning* (BASTOS; BIAGIOTTI, 2014).

Segundo Yuan e Powell (2013), quando comparados aos cursos *online* tradicionalmente ofertados, os MOOCs possuem duas características: o acesso livre, onde qualquer pessoa pode participar sem o pagamento de taxas; e a escalabilidade, pois os cursos são construídos para suportar um número indefinido de participantes. Entretanto, a depender do provedor do MOOC, alguns podem ser “abertos”, mas não “massivos” e, outros podem ser “massivos”, mas não “abertos”.

Outra questão que deve ser levada em consideração é em relação ao papel do professor neste tipo de plataforma. Segundo Bastos e Biagiotti (2014), a didática de ensino no MOOC diverge das metodologias tradicionais, havendo a necessidade dos professores se adequarem para o novo contexto educacional. Com o fito de destacar algumas características que diferem os cMOOCs dos xMOOCs, o Quadro 4 é apresentado:

Quadro 4 – Diferenças entre cMOOC e xMOOC.

cMOOC	xMOOC
<ul style="list-style-type: none"> • Conectivista – redes. • Maior interação. • Enfatiza o “sistema social e técnico de aprendizado em que a voz do professor não é uma ponte essencial, mas um nó em uma rede geral”. • Criação / exploração de área temática em ambiente “atelier”. • Produtos exclusivos criados pelos alunos consistem em posts, imagens, diagramas, vídeos. • Fóruns de discussão, grupos Diigo, Twitter e outras redes sociais são componentes-chave. • Facilitador agrega, revisa, resume e reflete sobre a atividade no boletim diário / semanal. • Plataforma e ferramenta de colaboração “<i>Bootstrapped</i>”. • Focado na criação de conhecimento. 	<ul style="list-style-type: none"> • Cognitivo-comportamentalista. • Interação limitada. • Atividades semanais pré-determinadas, conduzidas por instrutores, estruturadas e sequenciadas. • Consistem em vídeos curtos e baseados em conteúdo, leituras e conjuntos de problemas. • Questionários contidos (auto graduados), avaliações por pares. • A participação no fórum de discussão é um componente opcional no xMOOC. • Entregue por meio de provedor de plataforma de terceiros (por exemplo, Coursera, edX). • Focado no consumo de conhecimento.

Fonte: Kaushik (2015).

Segundo Shah (2018), os cinco maiores provedores de MOOCs são: Coursera[®], com 30 milhões de usuários cadastrados; EDx[®], com 14 milhões; Xuetang[®], com 9,3 milhões de usuários registrados; Udacity[®], com 8 milhões de usuários e FutureLearn[®], com 7,1 milhões de usuários.

Um dos desafios ao se tentar cursar um MOOCs é com relação a escolha de qual curso deve-se realizar, em virtude da grande diversidade de cursos disponíveis. Com o intuito de auxiliar nesta questão, uma das abordagens é utilizar os sistemas de recomendação.

Todavia é preciso lembrar que, a teoria do comportamento planejado, incorpora os conceitos centrais das ciências sociais e comportamentais, de maneira que se permite a previsão e entendimento de comportamentos particulares em cenários específicos (AJZEN, 1991), sendo portanto, uma teoria precursora deste cenário.

Symeonidis e Malakoudis (2019), propuseram um modelo de matriz de fatoração multidimensional combinado com a filtragem colaborativa, uma das técnicas algorítmicas dos sistemas de recomendação, para realizar previsões de classificação. Assim, o algoritmo SVD++ (*Singular Value Decomposition*) foi estendido para possibilitar a coleta de informações de fontes externas, como suas habilidades, cursos frequentados e características para realizar previsões de classificação. Como resultado, recomendações de cursos são disponibilizados aos alunos, de modo que as habilidades tratadas nos cursos ofereçam o suporte necessário para eles alcançarem o emprego almejado. Pang et al. (2017), também abordam a filtragem colaborativa para recomendação de cursos em MOOCs.

Uma arquitetura de um sistema de recomendação baseado na *web*, que tem a finalidade de encontrar os cursos e materiais de acordo com os interesses dos alunos, foi proposta por Campos, dos Santos e Oliveira (2018). O estudo explorou o conceito e uso de *Linked Open Data*, modelando uma arquitetura de um sistema de recomendação baseado na *web*, capaz de integrar diferentes plataformas MOOCs para recomendar cursos, módulos ou partes de cursos e materiais relevantes dos interesses dos alunos.

Os sistemas de recomendação também podem ser utilizados como suporte ao Big-Data (HOU et al., 2018); como protótipos de recomendação para plataformas específicas (GOPE; JAIN, 2017); com algoritmos de agrupamento e mineração baseada no conteúdo para recomendar cursos (CHAKRABORTY; THOPUGUNTA; BANSAL, 2018); com sistemas para recomendar recursos educacionais abertos baseados em MOOCs (HAJRI; BOURDA; POPINEAU, 2018); ou com a implementação de algoritmo de recomendação de cursos com base no interesse do usuário, perfis demográficos e relação de pré-requisito do curso, usando a estratégia de filtragem colaborativa (JING; TANG, 2017). Cita-se ainda, a utilização de redes neurais para a detecção dinâmica dos estilos de aprendizagem para recomendação de conteúdos (HOU et al., 2018; YU; JIANG; WARSCHAUER, 2018).

Outra abordagem, é a utilização de métodos estatísticos para realizar a previsão da performance do estudante de MOOC, através da utilização dos dados das atividades educacionais (XING et al., 2015).

Vigentini et al. (2017), ainda destacam os tipos de dados encontrados nos MOOCs, categorizando-os de acordo com a Quadro 5. Os tipos dados são comumente válidos tanto para MOOCs quanto para plataformas de ensino virtual, mas a granularidade dos detalhes pode variar. Os autores citam o caso, por exemplo, de dados demográficos, que possuem uma riqueza de detalhes nas instituições que oferecem cursos por créditos, mas já nos MOOCs as informações a respeito dos participantes são limitadas e esparsas. Quando se analisa as informações de *logs* das atividades *online*, os MOOCs se destacam, já que os *logs* são muito mais sofisticados do que os cursos onde as instituições utilizam a plataforma *online* apenas como uma extensão dos cursos presenciais.

Quadro 5 – Dados disponíveis nos MOOCs.

Dados demográficos: dados demográficos gerais dos alunos, incluindo idade, sexo, idioma, nível de escolaridade e localização. Dados demográficos são comumente adquiridos durante o processo de registro, e dados demográficos adicionais podem ser adquiridos através de pesquisas.	Dados de desempenho: desempenho do aluno com base em avaliações formativas e somativas. Isso é coletado a partir de trabalhos de casa, questionários e exames, mas também inclui resultados de pesquisas / atividades pré-curso destinadas a diagnosticar o conhecimento do aluno antes de fazer o curso.
Dados da atividade: como e quando os alunos estão usando os recursos, como assistir a vídeos, ler material, enviar trabalhos de casa, responder questionários ou usar o fórum de discussão. A maioria das plataformas divide o uso por conteúdo e tipo de mídia (por exemplo, exibições de página, exibições de atribuição, exibições de livro-texto, exibições de vídeo).	Dados de <i>feedback</i> : <i>feedback</i> dos alunos provenientes de pesquisas e comentários em fóruns. Além disso, informações sobre os objetivos de aprendizado do aluno e motivação e uso pretendido do material.

Fonte: Vigentini et al. (2017).

2.5 LEARNING ANALYTICS

Quando se fala em Tecnologia Aprimorada de Aprendizagem (*Technology Enhanced Learning* – TEL) consideram-se várias tecnologias digitais projetadas para melhorar o aprendizado (BALACHEFF et al., 2009).

Assim, temos os *Personal Learning Environment* (PLE) (ATTWELL, 2007), Sistemas Educacionais de Hipermedia Adaptativa (BRUSILOVSKY; MILLÁN, 2007), Ambientes Interativos de Aprendizagem (SCAIFE et al., 1997), *Learning Management Systems* (LMS) (WATSON; WATSON, 2007), Ambientes Virtuais de Aprendizagem (DILLENBOURG et al., 2002), Simulações de Aprendizagem Imersiva (KENNEDY et al.,

2013), *Intelligent Tutoring Systems - ITS* (ALEVEN et al., 2008) e *Massive Open Online Course* (MOOC) (FOURNIER; KOP; DURAND, 2014).

Todos esses sistemas interagem com os alunos para auxiliá-los em seu aprendizado e essas interações deixam dados de rastreamento, também conhecidos como dados de registro ou *logs* (GAŠEVIĆ et al., 2016). Esses dados podem ser estudados com o objetivo de otimizar o aprendizado dos estudantes, sendo este campo denominado de *Learning Analytics* (LA).

A quantidade de dados disponíveis, através da proliferação digital, determina o surgimento da LA (LARUSSON; WHITE, 2014) e a ausência de barreiras técnicas para processamento tornam os sistemas de aprendizagem baseados em computador ideais para este tipo de pesquisa (OCHOA; WORSLEY, 2016).

LA é uma área interdisciplinar, que se utiliza de métodos da mineração de dados educacionais e outras áreas afins, para dar suporte a avaliação do aprendizado. Busca o entendimento do processo de ensinar e aprender através da medição, coleta, análise e relato de dados sobre os alunos e seus contextos, com a finalidade de entender e otimizar o aprendizado (PECHENIZKIY; GASEVIC, 2015).

LA estuda o desenvolvimento de métodos para lidar com base de dados educacionais, a fim de suportar o processo de aprendizado (CHATTI et al., 2014), analisando os rastros deixados pelos estudantes na utilização das ferramentas digitais, como os Sistemas de Gerenciamento de Aprendizado. Dessa forma, LA é composta pelas técnicas necessárias para a medição, coleta, análise e relato de dados sobre alunos e seus contextos, cujo objetivo principal é otimização do aprendizado e dos ambientes educacionais online em que ele ocorre (FERGUSON, 2012; LOCKYER; HEATHCOTE; DAWSON, 2013; VAN HARMELEN; WORKMAN, 2012).

A LA permite que os professores colem e analisem os dados de aprendizado do aluno para exibir o estado de aprendizado atual, o progresso da aprendizagem e a abrangência dos objetivos de aprendizado alcançados (DYCKHOFF et al., 2012).

Wise e Shaffer (2015) caracterizam a LA como uma habilidade em descobrir padrões e associações ao longo do tempo. Brooks e Thayer (2016) consideram a LA como uma área de interesse de análise diretamente relacionada com a experiência e aprendizado do estudante, tendo como objetivo melhorar o desempenho acadêmico e serviços relacionados.

LA utiliza-se de diferentes técnicas de mineração de dados e outros métodos de análise para prover informações sobre o que ocorre com o aluno em um LMS, mediante a análise de dados baseados em evidências (NA; TASIR, 2017). De modo geral, LA não é uma área nova de pesquisa, pois representa a intersecção de várias disciplinas acadêmicas, como pedagogia,

psicologia, ciência da computação e suas técnicas (aprendizado de máquina, inteligência artificial, sistemas de recomendação, estatística). Portanto, objetiva o monitoramento, análise, predição, intervenção, tutoria, mentoria, avaliação, adaptação, personalização, recomendação, consciência, reflexão no processo de aprendizagem *online*.

O monitoramento e a análise podem ter como objetivo potencializar o desempenho do estudante, identificando aqueles que estão em risco de reprovação. Neste sentido, Dodge, Whitmer e Frazee (2015) utilizaram a análise de regressão combinada com a análise de dados demográficos para verificar o impacto das intervenções realizadas, sugerindo formas de melhorar o desempenho.

Um modelo de intervenção é sugerido por Wu, Huang e Zou (2015), onde é levado em consideração a identificação do status do aprendizado do estudante, a correspondência e cálculo da estratégia de intervenção, a aplicação e análise da eficácia, através de uma estrutura em loop baseada em pesquisas na área.

Com o objetivo de fomentar o processo de ensino através do *feedback*, Clow (2012) articulou um ciclo de LA para fechar o ciclo de *feedback*, através de quatro passos conectados: a geração de dados dos alunos; a utilização desses dados para produzir métricas, análises e visualizações; e o fechamento do loop, que é representado pelo *feedback* do produto para os estudantes através de um ou mais tipos de intervenções.

Já Wise (2014), estudou o *design* das intervenções de LA, com o objetivo de melhorar a participação e dedicação dos estudantes. A autora define uma intervenção de LA como um quadro de atividades pelas quais as ferramentas de análise, dados e relatórios analíticos são obtidos e utilizados.

As técnicas avançadas de LA também são utilizadas para prever o desempenho do estudante (DAUD et al., 2017). Segundo os autores, a habilidade de prever o desempenho dos estudantes pode ter benefícios, quando se analisa as ações possíveis nos sistemas educacionais. Assim, vários modelos de classificação são aplicados para prever se um estudante terá condições de completar ou não o curso.

A potencialidade das técnicas de LA e sua utilização para melhorar os processos de ensino profissional deve ser mais bem explorada, inclusive com iniciativas em outras áreas, como na área de ciências humanas (HARFIELD, 2014).

Neste sentido, *Social Learning Analytics* é um termo utilizado para os métodos e técnicas que são empregadas para a customização e análise das interações que ocorrem nos fóruns de discussão online (ADRAOUI et al., 2017).

Bodily e Verbert (2017) analisaram sistemas que produzem relatórios de análise de aprendizado, na forma de *dashboards*, que rastreiam dados de aprendizagem e relatam diretamente aos alunos.

Percebe-se que a LA utiliza diferentes técnicas para converter dados em informação útil. Alguns estudos merecem citação: a análise de aprendizagem através de Big Data (MATSEBULA; MNKANDLA, 2017); a associação com jogos educacionais (TLILI et al., 2017); a análise dos tipos de intervenções na aprendizagem e seus efeitos (NA; TASIR, 2017); a utilização de algoritmo de agrupamento em análise de aprendizagem (BHARARA; SABITHA; BANSAL, 2017); recomendações sensíveis ao contexto (ALMUTAIRI; SIDIROPOULOS; KARYPIS, 2017); LA como indicador comportamental em ambientes virtuais (PHILLIPS et al., 2012); uso do estado da arte em memória computacional para implementação de framework de *Learning Analytics* (LAVETI et al., 2017) e, LA para construir e realizar atualizações frequentes do perfil cognitivo do aluno (MOTHUKURI et al., 2017).

LA é geralmente orientada pelos dados, mas também pode utilizar conhecimentos prévios de bases teóricas. Neste sentido, Simon (2017) propôs uma “estrutura de 2 níveis” que considera a LA como um processo de aprendizagem de nível 2 que inclui cinco componentes: partes interessadas, objetivos, dados, abordagens técnicas e *feedback*.

Com base nessa estrutura, uma amostra de trabalhos relacionados a LA foi analisada pelo autor para mostrar como esses trabalhos a relacionam com o conhecimento prévio (*a priori knowledge* - APK) e conclui que o conhecimento prévio que promove suporte ao aprendizado do aluno é a teoria da aprendizagem. Entretanto, pode haver a necessidade de utilização de outras teorias, a depender dos objetivos almejados pela aplicação das técnicas de LA.

Com certa frequência é possível encontrar resultados relevantes apenas para um determinado subgrupo de alunos e, a partir disso, tenta-se generalizar para todos. Para amenizar este problema, Lockyer et al. (2013) e Wise e Shaffer (2015) propõem também a utilização de conhecimento prévio, ou seja, o conhecimento que o aluno já possui antes de iniciar outra atividade de aprendizado, para lidar com este problema.

Alguns estudos propõem uma combinação das análises de aprendizagem e métodos de pesquisa convencionais, como proposto por Gilmore (2014). Um panorama conceitual da área de LA (Figura 4) é apresentado por Peña-Ayala, Cárdenas-Robledo e Sossa (2017), onde a LA é dividida em três domínios principais: Um olhar sobre a LA; Ambientes de Aprendizagem; Aplicações de LA.

Figura 4 – Panorama de *Learning Analytics*.

Learning Analytics	Um olhar sobre LA	Domínios Subjacentes	Sistema de Informação baseados em computador Computadores na Educação Sistemas baseados em conhecimento
		Linhas relacionadas a LA	Análises Análise Educacional
		Natureza da LA	Práticas Definição de LA Ambientes de Aprendizagem e recursos Stakeholders e níveis
	Ambiente de Aprendizagem	Paradigmas de Aprendizagem	Aprendizado autorregulado Metacognição Aprendizado colaborativo Aprendizado em contexto social Processamento de linguagem natural, discurso, conversação Diversos paradigmas de aprendizado
		Linha de base da LA	Modelos Estruturas Stratégias, métodos, técnicas e algoritmos Diversos paradigmas de aprendizagem
		Panorama legal	Governança Ética Privacidade Proteção de Dados Termos Legais Tópicos Diversos
		Aplicações	Funcionalidades
	Suporte ao Aprendizado		Internalização Externalização
	Suporte ao usuário		Suporte ao professor Visualização Gerenciamento do tempo Outros
	Interação entre pares		Aprendizado colaborativo suportado por computador LA Social Processamento de linguagem natural

Fonte: Peña-Ayala; Cárdenas-Robledo e Sossa (2017) – Adaptado e traduzido pelo autor.

2.5.1 Modelos

Diversos modelos são encontrados na literatura que representam a aplicação das técnicas de LA. Nesta seção, serão exemplificadas algumas pesquisas que abordam modelos desenvolvidos na área.

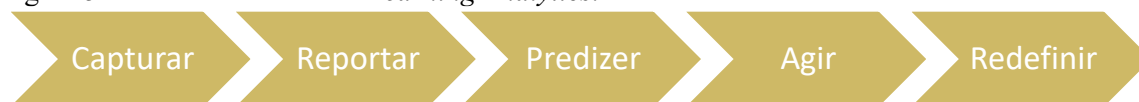
- Modelo probabilístico para identificação de fatores de abandono de estudos de ciência da computação através de redes bayesianas (LACAVE; MOLINA; CRUZ-LEMUS, 2018);
- Modelo para detectar automaticamente os estilos de aprendizagem dos alunos a partir dos seus dados de mídia pessoais, acadêmicos e sociais e fazer

recomendações para alunos, professores, educadores e administradores para a melhoria geral dos resultados de aprendizagem (JENA, 2018);

- Modelo de coleta e estruturação de dados de ação de ensino-aprendizagem baseado na teoria da atividade, projetado para identificar atividades com base em uma série de ações do aprendiz ao longo do tempo (KIM; MOON, 2018);
- Modelo de classificação de regressão logística para prever o desempenho de aprendizagem dos alunos, que calcula a relação entre as características do comportamento de aprendizagem e o efeito de aprendizagem em dados de SPOC – *Small Private Online Course* (RAN et al., 2018);
- Modelo de valor de negócio de *Big Data* Acadêmico (BDA) e de LA que explica os benefícios potenciais do BDA e o resultado ao nível de negócio que pode ser obtido, desenvolvendo tais capacidades analíticas nas instituições de ensino superior (CHAURASIA et al., 2018);
- Modelo probabilístico de tópico orientado para emoção, com potencial para descobrir as emoções dos alunos em seus comentários, melhorar a experiência de aprendizado online e identificar alunos em risco (LIU et al., 2018).

O primeiro modelo proposto (CAMPBELL; DEBLOIS; OBLINGER, 2007) é caracterizado por um conjunto de perguntas, agrupadas em cinco etapas, conforme a Figura 5:

Figura 5 – Primeiro modelo de *Learning Analytics*.

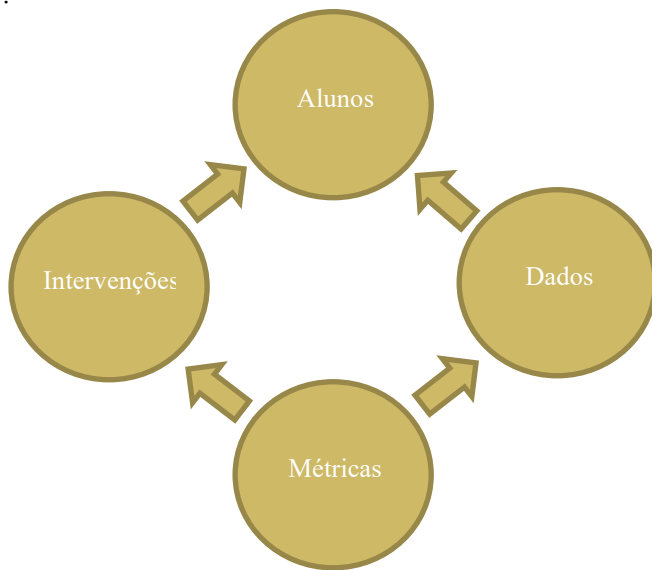


Fonte: Campbell, Deblois e Oblinger (2007) – Traduzido pelo autor.

Na primeira etapa devem ser decididos quais dados serão coletados, qual a frequência da coleta e onde os dados serão armazenados, assim como, a definição do formato e da política de segurança; na segunda etapa são produzidos os relatórios. Então, são definidos que tipo de informação será reportado, pra quem, e qual a frequência; A terceira etapa corresponde a atividade de aplicar os algoritmos de previsão e verificar a precisão, os dados de entrada e quais previsões devem ser reportadas para os *stakeholders*; A próxima etapa consiste na ação, ou seja, quais ações devem ser executadas no ambiente de ensino para otimizar o processo de ensino; e por fim, o último passo consiste no refinamento dos resultados obtidos nos passos anteriores.

Clow (2012) propôs outro modelo baseado no modelo de Campbell, Deblois e Oblinger, estabelecendo um ciclo de análise de aprendizagem fundamentada (Figura 6).

Figura 6 – O ciclo de LA.

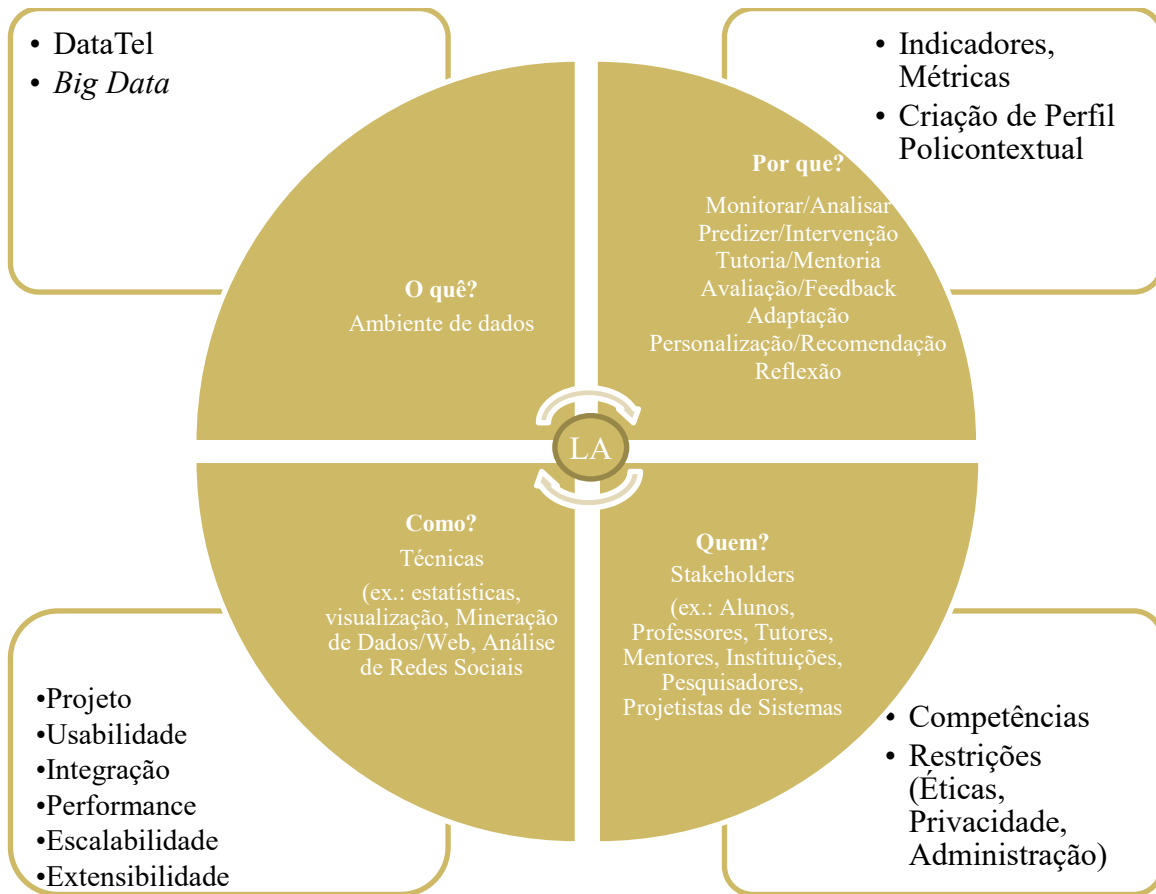


Fonte: Adaptado de Clow (2012).

Os estudantes são os participantes do curso; os dados são aqueles gerados e capturados durante o processo de aprendizado; as métricas são necessárias para fornecer informações sobre a aprendizagem e, o último passo, as intervenções, é onde acontece o para melhorar todo o aprendizado.

Segundo Chatti et al. (2013), o modelo de LA envolve quatro dimensões (Figura 7): que tipo de dados o sistema reúne, gerencia e usa para a análise? Quem é o alvo da análise? Por que o sistema analisa os dados coletados? Como o sistema executa a análise dos dados coletados?

Figura 7 – Modelo de referência para a LA.

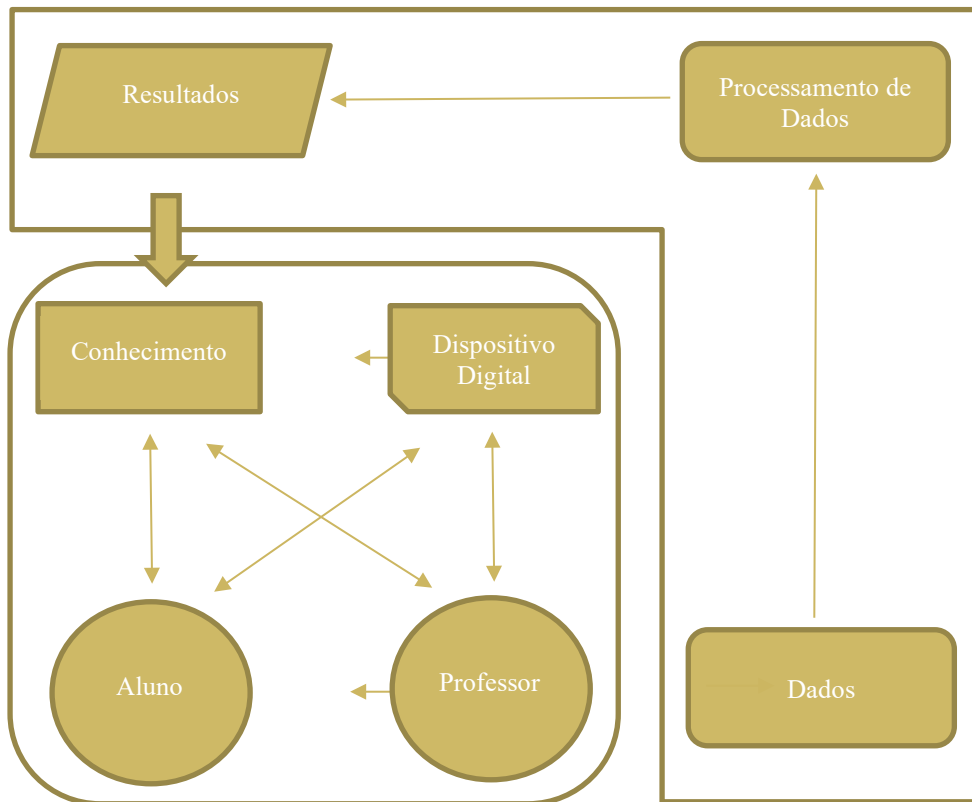


Fonte: Chatti et al. (2013) – Traduzido e adaptado pelo autor.

Convém ressaltar a última dimensão, que consiste nas técnicas para detectar padrões escondidos em conjuntos de dados educacionais: estatísticas, visualização de informação, mineração de dados, classificação, agrupamento, regras de associação e análise de redes sociais.

O processo de LA segue um fluxo interativo, conforme exposto na Figura 8. Deste modo, os dados são coletados do processo de aprendizagem; são processados; os resultados do processo são avaliados, interpretados e reintroduzidos no processo. Segundo Simon (2017), a maior parte dos estudos abordam LA de acordo com este ciclo (CHATTI et al., 2013; CLOW, 2012; DYCKHOFF et al., 2012; KHALIL; EBNER, 2015).

Figura 8 – Fluxo interativo da LA.



Fonte: Simon (2017) – Adaptado pelo autor.

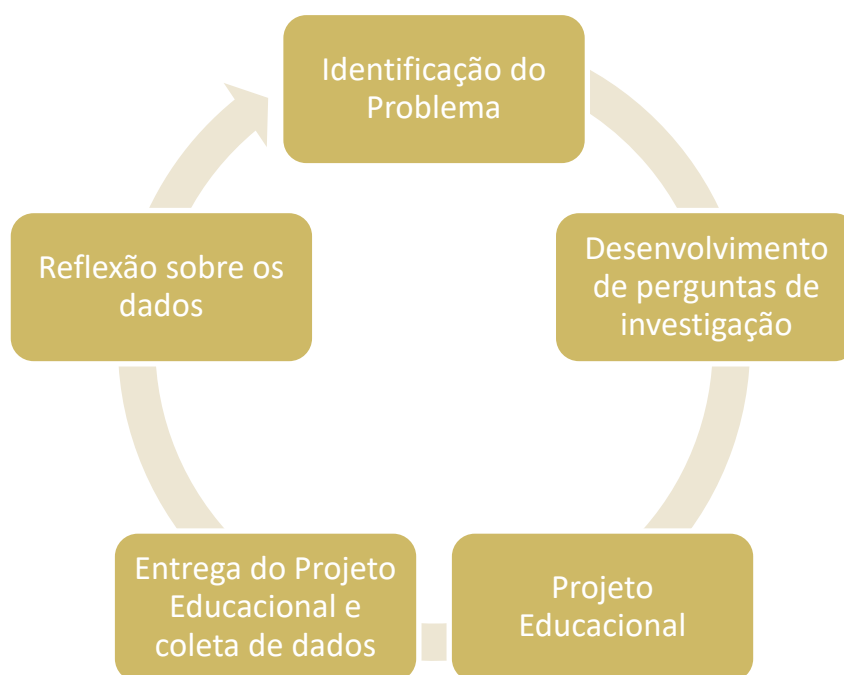
De acordo com Sciarrone (2018), um outro campo de pesquisa tem sido proposto com o objetivo de complementar a sinergia com a LA, chamado de modelo de questionamento do professor (*Teacher Inquiry Model*). Assim Gauthier (2013) e Prieto et al. (2011) definem que "*Teaching Analytics*" refere-se aos métodos e ferramentas digitais para ajudar os professores a analisar e melhorar os projetos educacionais antes da entrega. Além disso, desenvolvimentos mais recentes sobre "*Teaching Analytics*" também suportam a análise de como o professor oferece os projetos educacionais.

Tendo em vista essa sinergia, Sergis e Sampson (2017), propuseram um novo modelo representando a integração entre LA e TA (*Teaching Learning*) (Figura 9), por considerarem mais adequado para apoiar o *Teaching Inquiry*.

O modelo tem seis etapas. Na primeira etapa, chamada de identificação do problema, o professor identifica algumas questões, a fim de melhorar o processo de aprendizagem. Na segunda etapa, o projeto da entrega educacional e coleta de dados, o professor desenvolve perguntas da pesquisa; o que é necessário investigar, e que dados educacionais devem ser captados. A terceira etapa, o projeto educacional, é onde o professor desenvolve o projeto educacional a ser entregue a fim começar a pesquisa. Na etapa 4, que consiste na entrega do projeto educacional e coleta de dados, o professor entrega o projeto educacional aos alunos e

recolhe os dados educacionais. Na etapa 5, a análise dos dados educacionais, o professor analisa os dados educacionais a fim de obter *insights* para responder à pergunta definida pela pesquisa. Finalmente, na etapa 6, que é a reflexão sobre os dados, o professor revisa o processo de aprendizagem, a fim de responder à pergunta de pesquisa definida.

Figura 9 – Modelo de LA combinado com *Teaching Inquiry*.



Fonte: Sergis e Sampson (2017) – Adaptado e traduzido pelo autor.

2.5.2 Frameworks

Frameworks procuram orientar o desenvolvimento de uma abordagem em LA, representando um arcabouço conceitual composto de um conjunto de conceitos. A seguir, são comentados alguns estudos que propõem *frameworks* de LA.

- *Framework* integrado de LA para a percepção da qualidade na educação superior que aborda a utilização dos indicadores-chave de desempenho relacionados à qualidade na educação superior (VAROUCHAS; SICILIA; SÁNCHEZ-ALONSO, 2018): a pesquisa teve como finalidade discutir os *Key Performance Indicators* (KPIs) relacionados à qualidade na educação superior. Para isto, realizou, entre outras contribuições, uma pesquisa aplicada de LA que agrupou 20 fatores de qualidade em três dimensões da LA: conteúdo, processo e engajamento;
- *Community Inquiry Framework* sob a perspectiva de LA: é um framework teórico para explicar o processo de aprendizado e experiência online, através da correlação

e regressão linear, analisando a relação entre os comportamentos online dos alunos. Utilizou dados de questionários e dados do comportamento de aprendizagem, que foram coletados de uma plataforma de aprendizagem. Este estudo identificou quatro tipos de comportamentos de aprendizagem como preditores significativos para o desempenho de aprendizagem nos MOOCS (XU; LUO; TAN, 2018);

- *Framework* baseado em *Big Data* para analisar o comportamento de alunos nos sistemas de gerenciamento de aprendizagem (LMS) (CANTABELLA et al., 2019);
- Deep LA: visão geral dos *softwares* livres de *Deep Learning* que podem ser utilizados para implementar *frameworks* de *Deep Learning* que podem ser usados para implementar técnicas de *Deep Learning Analytics* e acelerar o processo de treinamento usando *Graphics Processing Units* (GPUs) (PASSALIS; TEFAS, 2019);
- *Framework* para integrar a análise de aprendizagem em jogos sérios para pessoas com deficiências intelectuais (NGUYEN; GARDNER; SHERIDAN, 2018);
- Como construir e distribuir um sistema educacional que pode ser utilizado como um *framework* para LA na Finlândia (LAAKSO et al., 2018);
- Desenvolvimento de ferramentas de análise preditiva para fornecer aos *stakeholders* informações automatizadas e oportunas para avaliar o desempenho do aluno em relação ao resultado de seu sucesso e para informar decisões pedagógicas ou estratégias de intervenção (HE et al., 2018).

2.5.3 Estratégias, Métodos, Técnicas e Algoritmos

Com o intuito de se ter um panorama das estratégias, métodos, técnicas e algoritmos utilizados no campo da LA, os quadros 6, 7 e 8, retirados do trabalho de Penã-Ayala, Cárdenas-Robledo e Sossa (2017), reúnem uma amostra de artigos que abordam algumas técnicas. A partir disso, para o presente trabalho, adotar-se-á duas técnicas de LA para análise dos dados educacionais: análise de agrupamentos e sistemas de recomendação.

Quadro 6 – Exemplos de estratégias de LA.

Nome / Autor	Perfil
Muñoz-Merino et al. (2015)	Eficácia precisa: analisa a eficácia dos alunos com recursos educacionais e atividades no MOOC.
Kim et al. (2016)	Facilitador: encoraja os potenciais alunos com baixo desempenho a participarem durante um curso. Estruturado: orienta os aprendizes iniciantes em um estágio inicial de discussão por meio de solicitações de discussão.
Chiu e Fujita (2014)	Análise estatística do discurso: modelagem de processos de discussão online para enfrentar diversas dificuldades analíticas envolvendo dados, resultados e variáveis explicativas.
Knight e Littleton (2015)	Funções de discurso, foco, conteúdo e exemplo de análise de aprendizagem centrada no discurso: contém um conjunto de técnicas adequadas dedicadas a quatro instâncias de relações entre uma função específica, foco e conteúdo.

Fonte: Penã-Ayala, Cárdenas-Robledo e Sossa (2017) – Traduzido pelo autor.

Quadro 7 – Exemplos de métodos utilizados na área de LA.

Nome / Autor	Perfil
Fidalgo-Blanco et al. (2015)	Modelo de treinamento abrangente da competência do trabalho em equipe: integra ferramentas de vários LMS e facilita o registro de interações do usuário, bem como o acesso à evidência de trabalho em equipe.
Kim et al. (2016)	Extração de variáveis de proxy: descreve os principais fatores considerados na discussão on-line assíncrona.
Chiu e Fujita (2014)	Análise estatística do discurso: modelagem de processos de discussão online para enfrentar diversas dificuldades analíticas envolvendo dados, resultados e variáveis explicativas.
Knight e Littleton (2015)	Ciência tradicional e baseada em dados: compara os pontos de vista de ambos os métodos, bem como seus passos com base em estratégias indutivas e dedutivas.
(Siadaty, Gašević e Hatala (2016)	Estratégias motivacionais para o questionário de aprendizagem: avaliar a orientação motivacional dos alunos e o uso de diferentes estratégias de aprendizagem. Inventário de estratégia de aprendizagem e estudo: avalia as estratégias de aprendizagem que os estudantes universitários relatam que estão usando.

Fonte: Penã-Ayala, Cárdenas-Robledo e Sossa (2017) – Traduzido pelo autor.

Quadro 8 – Exemplos de técnicas utilizadas na área de LA

Nome / Autor	Perfil
Fidalgo-Blanco et al. (2015)	Trabalho em equipe virtual: inclui fóruns, wikis, WhatsApp... em um ambiente de aprendizado misto, onde as equipes também têm reuniões pessoalmente.
Muñoz-Merino et al. (2015)	Validação de especialistas: obtém diversas perspectivas dos resultados de acordo com o contexto da aplicação.
Kim et al. (2016)	Floresta aleatória: apresenta amostragem aleatória e estratégias de conjunto.
Wise e Shaffer (2015)	Segmentação de sequência: organiza sequências de ações do usuário a cada vez que um usuário conclui uma ação de “avaliação”.
Knight e Littleton (2015)	Funções do discurso, foco, conteúdo e exemplo de LA centrada no discurso: contém um conjunto de técnicas adequadas dedicadas a quatro instâncias de relações entre uma função específica, foco e conteúdo.
Kovanović et al. (2016)	Estimativa de tempo na tarefa: calcula o tempo gasto na leitura de discussões, tempo na tarefa a partir de dados de rastreamento e dados de auto relato sobre a quantidade de tempo que os alunos passaram usando o sistema.

Fonte: Penã-Ayala, Cárdenas-Robledo e Sossa (2017) – Traduzido pelo autor.

Um exemplo de aplicação de LA fora da academia são as capacitações profissionais realizadas na área de segurança pública, voltada aos agentes da lei. Historicamente, essas ações capacitivas possuem dois enfoques: uma tradicional, baseada na trajetória histórica herdada do exército, caracterizada por uma prática mecanicista e baseada numa ideologia bélica, apenas reprodutora de informações; e outra vertente voltada a incorporação da educação em direitos humanos, a partir do policiamento comunitário, onde os policiais participam de constantes atividades de ensino-aprendizagem e exercem um espírito crítico sobre a atividade desenvolvida, a fim de promover a humanização e fomentar o sentimento de segurança (SILVA; SILVA, 2015).

Neste cenário, as atividades de capacitação e formação profissional voltadas a esse grupo necessitam de abordagens tecnológicas para subsidiar as ações que visam otimizar a mudança de paradigma requerida pela sociedade. Exemplo de iniciativa é a utilização de ambientes virtuais de treinamento, como retrata Bertram et al. (2011) e Li (2008).

2.5.4 Análise de Agrupamento

A organização de dados em grupos semelhantes é um dos modos fundamentais de compreensão e aprendizado. A análise de agrupamento consiste no estudo formal de métodos e algoritmos para classificar ou agrupar objetos de acordo com suas características intrínsecas medidas ou percebidas (JAIN, 2010). A classificação de objetos similares em grupos consiste numa importante atividade humana (KAUFMAN; ROUSSEEUW, 2009).

Muitas técnicas estatísticas são propostas para analisar dados, tais como análise de variância, regressão linear, análise discriminante, análise de correlação canônica, dimensionamento multidimensional, análise fatorial, análise de componentes principais e análise de agrupamentos (JAIN, 2010).

Quando se analisam dados, os procedimentos podem ser exploratórios ou descritivos, e confirmatórios ou inferenciais. Exploratório ou descritivo significa que o investigador não possui modelos ou hipóteses predefinidos, mas quer entender as características gerais ou estrutura dos dados; confirmatório ou inferencial significa que o investigador quer confirmar a validade de uma hipótese, modelo ou um conjunto de suposições (TUKEY, 1977). Os agrupamentos representam um procedimento-chave para ambas as abordagens (JAIN; MURTY; FLYNN, 1999).

A análise de agrupamentos é um processo de formação de grupos (*clusters*) de objetos semelhantes de um dado conjunto de entradas. Objetos pertencentes ao mesmo cluster são

considerados "similares" uns aos outros, enquanto objetos em dois *clusters* diferentes são "dissimilares" (OSIŃSKI, 2013).

Os agrupamentos são um dos métodos de análise de dados mais utilizados nas aplicações práticas em diversas áreas de pesquisa (AHMAD; STARKEY, 2017). Xie et al. (2011) ressaltam as vertentes de mineração de dados, reconhecimento de padrões, suporte a decisão e aprendizado de máquina. Vega-Pons e Ruiz-Shulcloper (2011) explicam que a análise de agrupamentos é essencial em qualquer campo de pesquisa que envolva a análise ou o processamento de dados multivariados, tais como: mineração de dados, taxonomias, recuperação de informação, segmentação de imagens, classificação de padrões, etc.

De acordo com Karaboga e Ozturko (2011), o objetivo da análise de agrupamentos é agrupar dados em *clusters* de tal forma que as semelhanças entre membros de dentro do mesmo *cluster* sejam máximas, enquanto as semelhanças entre membros de diferentes *clusters* sejam mínimas.

Neste sentido, representa uma tarefa básica para a exploração de um conjunto de dados, sendo aplicada em uma ampla variedade de disciplinas, como medicina, psicologia, biologia, sociologia, reconhecimento de padrões e processamento de imagens (SU; CHOU, 2001; DHARMARAJAN; VELMURUGAN, 2013).

Jain (2010) assevera que a análise de agrupamentos de dados tem sido utilizada para três propósitos:

- Estrutura subjacente: obter informações sobre dados, gerar hipóteses, detectar anomalias e identificar características relevantes;
- Classificação Natural: identificar o grau de semelhança entre formas ou organismos (relação filogenética);
- Compressão: ser um método para organizar os dados e representa-los através de protótipos de *clusters*.

Segundo Wu, Xia e Le (2017), em comparação com a classificação supervisionada, a análise de *cluster* é uma aprendizagem não supervisionada, o que significa que os números e a estrutura dos agrupamentos são desconhecidos antecipadamente. Objetos no mesmo grupo têm um alto grau de similaridade, enquanto objetos entre diferentes *clusters* têm um alto grau de dissimilaridade.

No mesmo sentido, Goldschmidt, Passos e Bezerra (2015), afirmam que diferente da tarefa de classificação, em que os registros estão relacionados a rótulos predefinidos, na análise de agrupamentos os objetos considerados como entrada não possuem rótulos. Por este motivo,

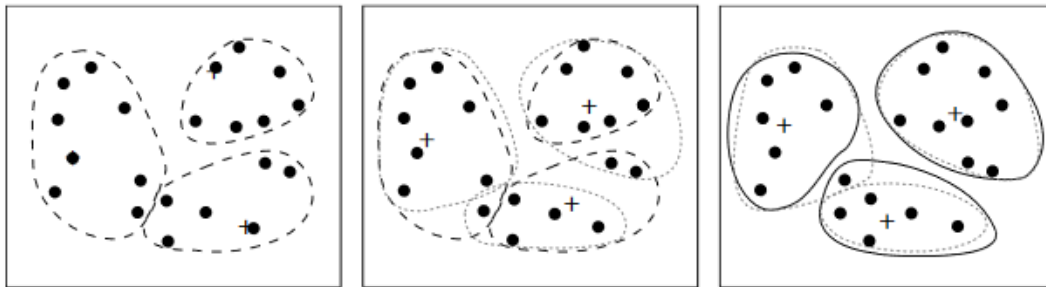
ela é considerada como uma tarefa de aprendizado não supervisionado, não possuindo classificação prévia dos dados para posterior agrupamento (KONCHADY, 2006).

Osiński (2003) defende a ideia de que a análise de agrupamentos se originou no campo da estatística, onde foi aplicada a dados numéricos. Entretanto, a Ciência da Computação e a Mineração de Dados estenderam a noção a outros tipos de dados, como texto ou multimídia. Assim, a análise de agrupamentos tem sido abordada em muitos contextos e por pesquisadores em muitas disciplinas, devido a sua grande utilidade na análise exploratória de dados (DHARMARAJAN; VELMURUGAN, 2013).

Yu et al. (2018) apontam que a análise de agrupamentos fornece um meio comum de identificar a estrutura em dados complexos e, por esse motivo, há um crescente interesse na utilização desse método como uma ferramenta para a análise de grandes conjuntos de dados, como declarado em Prayaga et al. (2019).

A análise de agrupamentos pode ser utilizada em diferentes situações. Organizar documentos com outros relacionados ao seu conteúdo, agrupar os resultados de uma busca textual, gerar conjuntos de palavras com base em sua coocorrência, são alguns exemplos de aplicações. A Figura 10 apresenta um exemplo de agrupamentos, sendo que o centro de cada agrupamento é representado pelo símbolo de “+”.

Figura 10 - Exemplo de agrupamentos.



Fonte: Han, Pei e Kamber (2011).

Existem diversos métodos para realizar a análise de agrupamentos em dados, sendo que cada um pode gerar agrupamentos diferentes para o mesmo conjunto de dados (HAN; PEI; KAMBER, 2011). Estes métodos distinguem-se pelo tipo de resultado a ser produzido e pelas diferentes formas em definir a proximidade entre um indivíduo e um grupo já formado, ou entre dois grupos quaisquer (VASCONCELOS et al., 2007).

De acordo com Jain, Murty e Flynn (1999), a atividade típica de agrupamento envolve as seguintes etapas:

- Representação de padrões;

- Definição de uma medida de proximidade de padrão apropriada para o domínio de dados;
- Agrupamento;
- Abstração de dados;
- Avaliação da produção.

Os métodos de análise de agrupamentos podem ser divididos em três grupos básicos. O primeiro é o dos algoritmos baseados em distâncias, o segundo é o dos baseados em distribuições de probabilidades e, por fim, o grupo de algoritmos de agrupamento baseados em densidade (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Já Han, Pei e Kamber (2011), informam que é difícil fornecer uma categorização nítida dos métodos de agrupamento, pois essas categorias podem se sobrepor, de modo que um método pode ter recursos de várias outras categorias. Os autores realizaram um sumário que oferece uma visão geral dos métodos básicos de agrupamentos (Quadro 9).

Quadro 9 – Métodos básicos de algoritmos de agrupamento.

Método	Características Gerais
Particionamento	- Encontra aglomerados de forma esférica mutuamente exclusivos; - Baseado em distância; - Pode usar média ou <i>medoid</i> (etc.) para representar o centro do cluster; - Efetivo para conjuntos de dados de tamanho pequeno a médio.
Hierárquicos	- <i>Cluster</i> é uma decomposição hierárquica (ou seja, vários níveis); - Não é possível corrigir mesclagens ou divisões erradas; - Pode incorporar outras técnicas como <i>microcluster</i> ou considerar objetos ligados (<i>linkages</i>).
Baseados em Densidade	- Pode encontrar <i>clusters</i> de forma arbitrária; - <i>Clusters</i> são regiões densas de objetos no espaço que são separados por regiões de baixa densidade; - Densidade de <i>cluster</i> : cada ponto deve ter um número mínimo de pontos dentro de sua "vizinhança"; - Pode filtrar <i>outliers</i> (observação que pode estar longe da observação central de determinado agrupamento).
Métodos baseados em grade	- Utiliza uma estrutura de dados de grade de multirresolução; - Baixo tempo de processamento (normalmente este tempo é independente do número de objetos de dados, mas ainda dependente do tamanho da grade).

Fonte: Han, Pei e Kamber (2011) – Adaptado pelo autor.

De forma similar, Fahad et al. (2014) indicam que para categorizar os algoritmos de agrupamento, pode-se considerar a forma dos agrupamentos gerados. Então, os algoritmos podem ser classificados como agrupamento hierárquico, agrupamento por particionamento, agrupamento baseado em densidade, agrupamento baseado em modelo e agrupamento baseado em grades. A Figura 11 ilustra a taxonomia.

Figura 11 - Taxonomia de análise de agrupamento.



Fonte: Fahad et al. (2014) – Adaptado pelo autor.

Segundo Dutt, Ismail e Herawan (2017), quando a abordagem de agrupamento é utilizada para analisar dados provenientes de sistemas educacionais, ela é chamada de Agrupamento de Dados Educacionais, ou em inglês, *Educational Data Clustering* (EDC).

2.5.4.1 Algoritmos por Particionamento

Os algoritmos por particionamento dividem o conjunto de dados em k grupos. Desta forma, estes algoritmos escolhem k objetos como sendo os centros dos k grupos, e depois os objetos são divididos entre os k grupos de acordo com a medida de similaridade adotada, de maneira que cada objeto fique no grupo que forneça a menor distância entre o objeto e o centro do referido grupo (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Os métodos de particionamento tentam agrupar diretamente objetos de dados, dividindo os dados em *clusters* homogêneos. Primeiro, os centros de *cluster* são atribuídos arbitrariamente, então, os pontos de dados são atribuídos ao centro de *cluster* mais próximo baseado na distância ou similaridade (BEN AYED; BEN HALIMA; ALIM, 2015).

Segundo Han, Pei e Kamber (2011), os *clusters* são formados para otimizar um critério de particionamento, como uma função de dissimilaridade baseada na distância, para que os objetos dentro de um grupo sejam “semelhantes” uns aos outros e “dissimilares” para objetos em outros grupos em termos dos atributos do conjunto de dados.

Fahad et al. (2014) apontam que os agrupamentos particionados por esses métodos devem preencher os seguintes requisitos:

- Cada grupo deve conter pelo menos um objeto, e;
- Cada objeto deve pertencer exatamente a um grupo.

Os métodos por particionamento fornecem uma partição do conjunto de dados em um número prefixado de *clusters*. Cada *cluster* é representado pelo seu vetor centróide (vetor que representa o centro do agrupamento). Então, um algoritmo tenta minimizar uma função de critério por meio de um processo iterativo, no qual todos os centróides são atualizados na tentativa de melhorar a qualidade dos *clusters* finais (PACIFICO; LUDERMIR, 2014).

2.5.4.1.1 *K-means* - uma técnica baseada no centróide

Ayed, Halima e Alimi (2014) informam que o algoritmo *K-means* é o algoritmo de análise de agrupamentos mais utilizado desde 1965, principalmente devido a sua simplicidade de implementação e sua eficácia.

O algoritmo *K-means* considera que os registros do conjunto de dados correspondem a pontos no espaço, onde cada atributo corresponde a uma dimensão deste espaço. Além disso, o *K-means* possui um parâmetro de entrada, k , que corresponde a quantidade de grupos a serem identificados pelo algoritmo (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

O algoritmo *K-means* trabalha em geral, com a distância Euclidiana para lidar com dados de atributo numérico. Utiliza assim, uma distância de correspondência simples para lidar com dados puros de classificação (JINYIN et al., 2017).

O Quadro 10 retrata o procedimento realizado pelo algoritmo *K-means* e suas etapas:

Quadro 10 – Procedimentos do algoritmo *K-means*.

Algoritmo: *k-means*. O algoritmo *k-means* realiza o particionamento, sendo que o centro de cada cluster é representado pelo valor médio dos objetos no cluster.

Entrada:

- k : o número de clusters,
- D : um conjunto de dados contendo n objetos.

Saída: um conjunto de k clusters.

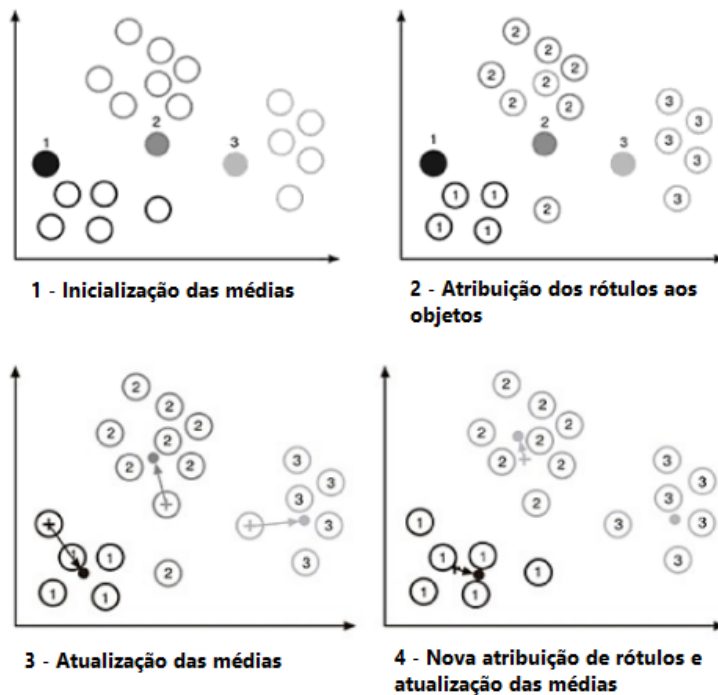
Método:

- (1) escolher arbitrariamente k objetos de D como os centros iniciais do cluster;
- (2) repetir
- (3) (re) atribuir cada objeto ao cluster ao qual o objeto é o mais similar, com base no valor médio dos objetos no cluster;
- (4) atualizar as médias do cluster, ou seja, calcula o valor médio dos objetos para cada cluster;
- (5) até que não haja mudança;

Fonte: Han, Pei e Kamber (2011).

Um exemplo de aplicação do *K-means* é realizado por Goldschmidt, Passos e Bezerra (2015), na Figura 12. Considerando um arquivo com 20 registros de dados, o *K-means* é aplicado, considerando-se $k=3$. Na primeira imagem, os centroides são selecionados aleatoriamente. Então, cada ponto restante é alocado a algum grupo, em função de sua distância em relação a cada um dos centroides (imagem 2). Após isso, os centroides são atualizados (imagem 3). A seguir, ocorre uma nova realocação de pontos (imagem 4), então o processo continua até a convergência, ou seja, até que os pontos não se alterem mais de um agrupamento para outro.

Figura 12 - Passos do algoritmo *K-means*.



Fonte: Goldschmidt, Passos e Bezerra (2015) – Adaptado pelo autor.

Han, Pei e Kamber (2011) afirmam que o método *K-means* não é adequado para descobrir *clusters* com formas não convexas ou *clusters* de tamanhos muito diferentes. Além disso, é sensível a pontos de dados de ruído e *outliers*, porque um pequeno número desses dados pode influenciar substancialmente o valor médio.

Desta forma, uma abordagem para tornar o método *K-means* mais eficiente em grandes conjuntos de dados é usar um conjunto de amostras de bom tamanho no agrupamento. Outra é empregar uma abordagem de filtragem que usa um índice de dados hierárquicos espaciais para economizar custos computacionais. Uma terceira abordagem explora a ideia de *micro clustering*, que primeiro agrupa os objetos próximos em “microagrupamentos” e, em seguida, realiza o agrupamento de k grupos a partir dos micros agrupamentos.

2.5.4.2 Algoritmos Hierárquicos

Han, Pei e Kamber (2011), definem que um método de agrupamento hierárquico funciona agrupando objetos de dados em uma hierarquia ou “árvore” de *clusters*. Segundo os autores, representar objetos de dados na forma de uma hierarquia é útil para resumo e visualização de dados.

Essa decomposição hierárquica é representada por um dendrograma, ou seja, uma árvore que iterativamente divide o conjunto de dados em subconjuntos menores até que cada

subconjunto consista de somente um objeto, gerando uma decomposição hierárquica do conjunto de dados (GOLDSCHMIDT, PASSOS; BEZERRA, 2015).

Os métodos de agrupamento hierárquico fornecem uma série de partições do conjunto de dados com base em processos iterativos. Os algoritmos hierárquicos são divididos em duas categorias: métodos aglomerativos e métodos divisivos (PATEL; THAKRAL, 2016), dependendo se a decomposição hierárquica é formada de maneira *bottom-up* (mesclagem) ou *top-down* (divisão) (HAN; PEI; KAMBER, 2011).

Os métodos aglomerativos iniciam com cada padrão formando seu próprio agrupamento, ou seja, cada um dos n objetos é colocado em seu próprio grupo. Então calcula-se a distância entre os grupos e se obtém uma hierarquia, sucessivamente, mesclando os *clusters* (ZHOU; XU; LIU, 2016).

Já nos métodos divisivos, o processo começa com todos os objetos em um único grupo. Então, um grupo é selecionado e dividido em dois grupos menores. Este processo persiste até que se tenha n grupos, ou até que uma condição de término seja cumprida (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Nos algoritmos hierárquicos a maneira de medir a distância entre grupos pode ser classificada como (AYED, HALIMA E ALIMI, 2014):

- Ligação simples, do inglês *Single Link* (distância mínima);
- Ligação média, do inglês *Average Link* (distância média);
- Ligação completa, do inglês *Complete Link* (distância máxima);
- Método de Ward, também conhecido como “Mínima Variância”. Assim, os métodos hierárquicos requerem uma matriz que contém as métricas de distância entre os agrupamentos em cada estágio do algoritmo, chamada matriz de similaridade.

2.5.4.3 Modularidade

A modularidade é uma medida de vizinhança, que mede a estrutura de uma rede, de forma a mensurar a força da divisão de cada ponto (vértices/nós) em grupos. Ela mede a qualidade da divisão da rede (NEWMAN; GIRVAN, 2004). Logo, redes com alta modularidade possuem conexões densas entre os nós das comunidades e baixa densidade entre os nós de diferentes comunidades.

Segundo Newman e Girvan (2004), a medida é baseada em uma medida de associação proposto em um artigo anterior do próprio Newman, chamado “*Mixing patterns in networks*”

(NEWMAN, 2003). O autor define a medida de modularidade pela Equação 1, sendo que $\|e\|$ define a soma dos elementos da matriz “ e ”, soma esta que mede a fração de arestas na rede que conecta nodos de mesmo tipo, menos o valor esperado da mesma quantidade em uma rede com as mesmas divisões de comunidade, mas com conexões aleatórias entre os vértices. Seu valor varia de $Q=0$ (quando o número de arestas não é melhor do que um valor aleatório) e $Q=1$ (quando indica redes com estruturas de comunidade fortes).

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tr } e - \|e^2\|, \quad (1)$$

Os algoritmos baseados na otimização da modularidade são fáceis de implementar, são populares, possuem tempo de execução relativamente menor e são ideais para *Big Data* e redes sociais *online* (KARATAS; SAHIN, 2018).

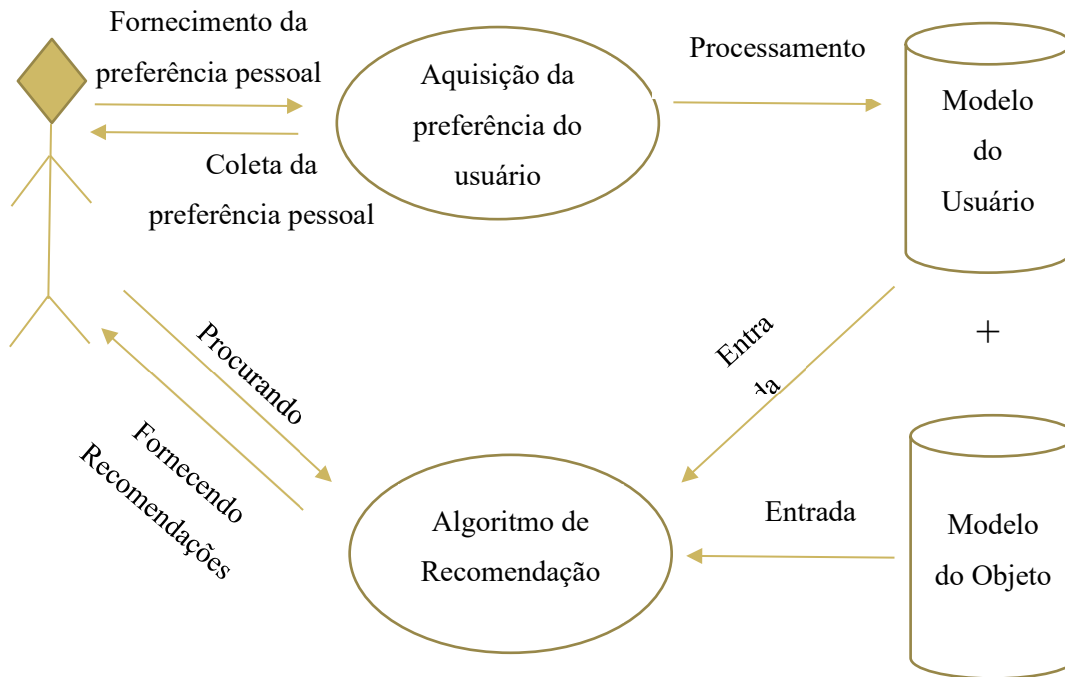
Na prática, o algoritmo proposto por Girvan e Nesman é um método iterativo de particionamento com a finalidade de remover progressivamente as arestas com o maior “*betweenness*” (número de caminhos mais curtos entre dois pares de nós que percorrem determinada aresta) até a rede se partir em componentes. Para se saber qual é a melhor divisão, os autores propõem justamente observar a modularidade máxima (BOCCALETTI et al., 2007).

2.6 SISTEMAS DE RECOMENDAÇÃO

Na década de 1990 surgiram as primeiras pesquisas na área de Sistemas de Recomendação (SR), a partir de estudos de recuperação e filtragem de informações (PARK et al., 2012; WANG; WU, 2011). Segundo os autores, a personalização de aplicações que utilizam tecnologias da informação e comunicação caracterizadas pelos SR têm tido muita atenção, especialmente em virtude dos grandes bancos de dados.

Chen, R. et al. (2018) informam que um SR é composto por três partes: usuário, recurso de item e algoritmo de recomendação, conforme apresentado na Figura 13.

Figura 13 – Um modelo de sistema de recomendação.



Fonte: Chen, R. et al. (2018) – Adaptado pelo autor.

Os SR estão sendo pesquisados e implantados extensivamente em várias áreas de aplicação, incluindo comércio eletrônico (YAO; JIANG, 2018), recomendação de filmes (CHEN, J. et al., 2018), educação (SYED et al., 2018), turismo (ISHANKA; YUKAWA, 2018), redes sociais (FAN et al., 2018) e serviços em geral (ZHU et al., 2018b). Algoritmos de recomendação são amplamente discutidos na literatura e em vários estudos (CHAVAN; MUKHOPADHYAY, 2017; CHEN, R. et al., 2018; JALILI et al., 2018; YANG; CHEN, J. et al., 2018; ZHU et al., 2018a).

Também têm sido aplicados na recomendação personalizada de itens em um grande volume de dados, que sejam de interesse ao perfil do usuário (FIGUEROA et al., 2015). Segundo os autores, um item pode se enquadrar em diferentes categorias. Assim, um item pode ser uma música, um lugar, uma notícia, um livro, um filme ou um evento qualquer.

A recomendação dos itens é realizada a partir de abordagens e técnicas, relacionada ao algoritmo utilizado na recomendação (RICCI et al., 2015).

Os SR foram desenvolvidos para ajudar os usuários a encontrar o que pode interessá-los e os proprietários de empresas a vender seus produtos com mais eficiência. Eles encontraram muita atenção na academia e na indústria. Um algoritmo de recomendação leva em conta as interações item-usuário, ou seja, o histórico de classificação (ou compra) de usuários, os itens e suas informações contextuais, se disponíveis (JALILI et al., 2018).

Os SR são ferramentas e técnicas que fornecem sugestões sobre itens que podem ser úteis para os requisitos de um usuário. Estes itens são recomendados e podem ser caracterizados pela complexidade, valor ou utilidade (TARUS; NIU; YOUSIF, 2017) e tem como objetivo auxiliar os usuários em suas decisões, fornecendo recomendações de fácil acesso e alta qualidade, tornando-se uma tecnologia poderosa e de impacto significativo (JANNACH et al., 2010).

Uma das vertentes dos SR auxiliam os usuários a localizarem conteúdos, produtos ou serviços, através da análise e sugestões dos próprios ou de outros usuários, utilizando a tecnologia analítica para calcular a probabilidade de recomendação (PARK et al., 2012).

De modo geral, podem ter como tarefa, a previsão ou recomendação. Previsão significa decidir se um usuário gostaria ou não de um item específico, prevendo determinada classificação para um novo item. Por outro lado, a recomendação refere-se a tarefa de recomendar uma lista de itens que o usuário provavelmente gostaria (JALILI et al., 2018).

Segundo Noor e Khan (2016), os SR personalizam a experiência do usuário, pois permitem recomendações direcionadas ao usuário de acordo com o perfil, conteúdo, informação e *feedback*.

São vistos como ferramentas de *softwares* baseadas em aprendizado de máquina e em técnicas de recuperação de informações que fornecem sugestões de itens úteis de acordo com o interesse do usuário (AAMIR; BHUSRY, 2015; SYED et al., 2018).

Possuem a capacidade de prever preferências e interesses de um usuário através da análise e comportamento de usuários com perfis semelhantes. As preferências são fornecidas conforme os usuários interagem com os sistemas, sendo que essas interações fornecem as informações que são coletadas e utilizadas para construir o perfil (BOBADILLA et al., 2013).

O perfil do usuário nestes sistemas é visto como uma diretriz para a recomendação de objetos (PINTER et al., 2012). Assim, para obter informações voltadas a prestação de recomendações personalizadas, o SR deve desenvolver e manter um perfil de usuário (JANNACH et al., 2010).

Podemos coletar as informações para a recomendação de duas maneiras: a forma explícita ou a forma implícita. Na forma explícita as informações podem ser coletadas através de questionários ou cadastros de informações, respondidos pelo usuário. Na forma implícita o sistema rastreia as atividades dos usuários e coleta as informações (BOBADILLA et al., 2013; PINTER et al., 2012).

Na área educacional, a abordagem tem sido aplicada em ambientes educacionais, no intuito de personalizar esses ambientes, utilizando os dados dos alunos, materiais e processos de aprendizagem (TARUS; NIU; YOUSIF, 2017).

Nos sistemas educacionais as recomendações mais utilizadas são em relação aos materiais e conteúdo de aprendizagem, assim como a recomendação de atividades online, exercícios, mensagens ou questionários. Os SR no contexto educacional são considerados *softwares* inteligentes que recomendam objetos a partir de ações dos próprios educandos (SALEHI; POURZAFERANI; RAZAVI, 2013).

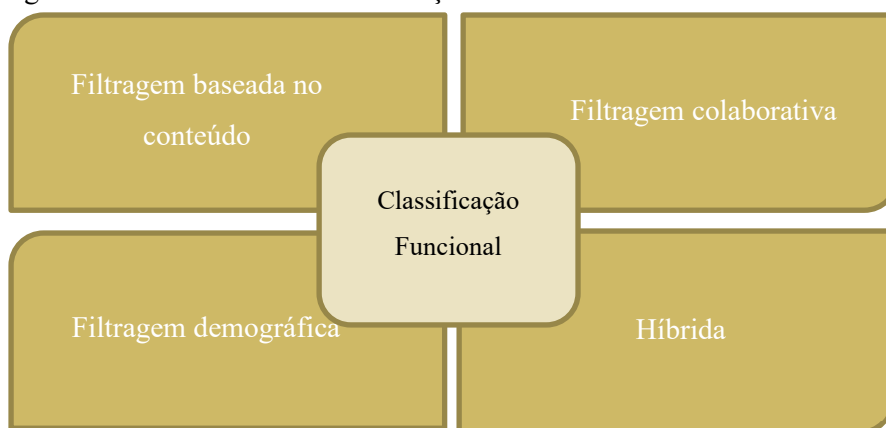
Desta forma, a recomendação auxilia os alunos no processo de aprendizagem sugerindo caminhos de aprendizagem e materiais relacionados com os seus objetivos e interesses, tornando a aprendizagem mais eficaz e o sistema educacional mais eficiente (PINTER et al., 2012).

Em virtude da existência de diversos métodos para se chegar a recomendação de um item, se faz necessário examinar as diferentes abordagens existentes para se avaliar qual seria mais adequada em cada situação.

2.6.1 Abordagens de Recomendação

Dentre os algoritmos de recomendação clássicos comumente usados, destacam-se os algoritmos de filtragem colaborativa, (que podem ser divididos em filtragem colaborativa baseada em memória e filtragem colaborativa baseada em modelo), algoritmo de recomendação baseado em conteúdo, filtragem demográfica e algoritmo de recomendação híbrida (CAO et al., 2018; JALILI et al., 2018). A Figura 14 apresenta um *framework* da classificação funcional dos SR.

Figura 14 – *Framework* da classificação funcional.



Fonte: Aamir e Bhusry (2015) – Adaptado pelo autor.

Entretanto, Burke (2007) faz a distinção de seis tipos de classes de recomendação:

- Baseada em conteúdo: o sistema aprende a recomendar itens semelhantes aos que o usuário gostou no passado. A similaridade dos itens é calculada com base nos recursos associados aos itens comparados;
- Filtragem colaborativa: recomenda ao usuário ativo os itens que outros usuários com gostos parecidos gostaram no passado. A similaridade no gosto de dois usuários é calculada com base na similaridade no histórico de classificação dos usuários;
- Demográfica: esse tipo de sistema recomenda itens com base no perfil demográfico do usuário. A suposição é que diferentes recomendações devem ser geradas para diferentes nichos demográficos;
- Baseada em conhecimento: sistemas baseados em conhecimento recomendam itens com base em conhecimento de domínio específico, sobre como determinados recursos de itens atendem às necessidades e preferências dos usuários e, em última análise, como o item é útil para o usuário;
- Baseado na comunidade: esse tipo de sistema recomenda itens com base nas preferências dos amigos dos usuários. Esta técnica segue a epigrama "Diga-me quem são seus amigos e eu lhe direi quem você é";
- Híbrido: estes SR são baseados em uma combinação das técnicas acima mencionadas.

2.6.2 Baseada em Conteúdo

Esta abordagem tem por objetivo explorar informações derivadas de textos ou descrições de itens, previamente avaliados no passado por um usuário e recomendar novos documentos ou itens relacionados ao perfil deste usuário (AAMIR; BHUSRY, 2015). Portanto, é baseada no conceito de que itens com atributos semelhantes serão classificados de forma semelhante (BOBADILLA et al., 2013; JANNACH et al., 2010).

Desta forma, a recomendação baseada em conteúdo recomenda itens similares aos que o usuário preferiu no passado, através da extração das palavras-chaves dos itens, que pode ser realizada automaticamente ou manualmente. Os perfis de usuários podem ser construídos explicitamente, questionando seus interesses, ou implicitamente analisando as classificações do usuário (NIEMANN; WOLPERS, 2015).

No mesmo sentido aponta Reddy et al. (2019), pois afirmam que esse tipo de recomendação faz o uso dos interesses particulares de um usuário e tenta corresponder o perfil de um usuário aos atributos avaliados pelos vários objetos de conteúdo a serem recomendados.

De forma geral, o processo de recomendação é composto por três etapas. Na primeira etapa, os termos podem ser alocados automaticamente ou manualmente. Quando são locados automaticamente, é necessário executar um método que possa extrair esses termos dos itens. Na segunda etapa, os termos devem ser representados de tal forma que tanto o perfil do usuário quanto os itens possam ser agrupados de forma significativa. Já na terceira fase, um algoritmo de aprendizado deve ser selecionado para aprender o perfil do usuário com base nos itens e fazer recomendações com base nesse perfil de usuário (AAMIR; BHUSRY, 2015).

Segundo Aamir e Bhusry (2015), a abordagem possui as seguintes vantagens:

- O método examina apenas os itens e o perfil do usuário para realizar recomendações;
- O método pode esclarecer o usuário sobre os recursos nos quais a recomendação é baseada;
- Novos itens podem ser recomendados antes de serem avaliados por um grande número de usuários.

As desvantagens da abordagem baseada em conteúdo são:

- Se o conteúdo não incluir uma quantidade adequada de informações para diferenciar os itens com precisão, as recomendações recebidas não serão precisas;
- O sistema pode não fornecer recomendações confiáveis quando não há quantidade adequada de informações para construir um bom perfil de um usuário;
- Ocorrência de eventos ao acaso de maneira benéfica (*Serendipity*): o sistema recomendará ou sugerirá os itens que tiverem altas preferências quando forem comparados com o perfil do usuário; é por isso que as recomendações são realizadas para o usuário contendo apenas os itens ou produtos semelhantes aos que já foram classificados.

Como exemplo deste tipo de abordagem, citam-se as recomendações de filmes pela classificação de gênero. O sistema pode recomendar filmes do mesmo gênero em que o usuário classificou positivamente anteriormente. Outro exemplo é a utilização de pesquisas na web

realizadas anteriormente pelos usuários, para recomendar páginas específicas de mesmo conteúdo. (JANNACH et al., 2010; RICCI et al., 2015).

Conclui-se que, para fazer recomendações, estes sistemas necessitam de técnicas apropriadas para representar os itens, construir o perfil do usuário e medir a semelhança do item com o perfil (JANNACH et al., 2010; RICCI et al., 2015).

Algoritmos genéticos, *feedback* de relevância, classificador bayesiano e redes neurais estão entre as técnicas de aprendizagem para o aprendizado de um perfil de usuário (AAMIR; BHUSRY, 2015).

Lu et al. (2015) descrevem técnicas de recomendações heurísticas, usando métodos tradicionais de recuperação de informações, como a medida de similaridade do cosseno. Bobadilla et al. (2013) cita métodos heurísticos ou algoritmos de classificação, tais como: indução de regras, métodos dos vizinhos mais próximos, classificadores lineares e métodos probabilísticos.

2.6.3 Filtragem Colaborativa

Esta técnica pode ser considerada a origem dos SR, uma das mais aplicadas e utilizadas, tendo como princípio recomendar itens quando os usuários são semelhantes. Desta forma os usuários podem ajudar uns aos outros através de seus perfis e classificações (WANG; WU, 2011).

Segundo Chen, R. et al. (2018) e Reddy et al. (2019), os sistemas de filtragem colaborativa tradicionais têm duas abordagens: abordagens baseadas em memória e abordagens baseadas em modelos.

Abordagens baseadas em memória analisam continuamente os dados do usuário para fazer recomendações. À medida que utilizam as classificações dos usuários, eles gradualmente melhoram a precisão ao longo do tempo. Eles são independentes de domínio e não requerem análise de conteúdo. Segundo Ning, Desrosiers e Karypis (2015), essa abordagem ainda é subdividida em duas formas: baseada no usuário e baseada no item (Figura 15 e 16).

Abordagens baseadas em modelos desenvolvem um modelo do comportamento de um usuário e, em seguida, usam determinados parâmetros para prever o comportamento futuro. O uso de algoritmos baseados em particionamento também leva a uma melhor escalabilidade e precisão (BHATT; PREMAL; GAUDANI, 2014).

Figura 15 – Filtragem Colaborativa baseada no usuário.



Fonte: Aamir e Bhusry (2015) – Adaptado pelo autor.

Figura 16 – Filtragem Colaborativa baseada no item.



Fonte: Aamir e Bhusry (2015) – Adaptado pelo autor.

Desta forma, o principal objetivo da filtragem colaborativa é explorar informações sobre o comportamento ou opiniões de uma comunidade de usuários no passado e prever quais itens os usuários irão gostar ou se interessar no presente, com base na semelhança dos históricos e classificações destes usuários (JANNACH et al., 2010; (CHEN, R. et al., 2018).

Segundo Chavan e Mukhopadhyay (2017), a filtragem colaborativa é uma tecnologia popular devido a sua capacidade de analisar os dados do usuário. O modelo de recomendação é composto por três módulos: entrada, algoritmo de recomendação e saída.

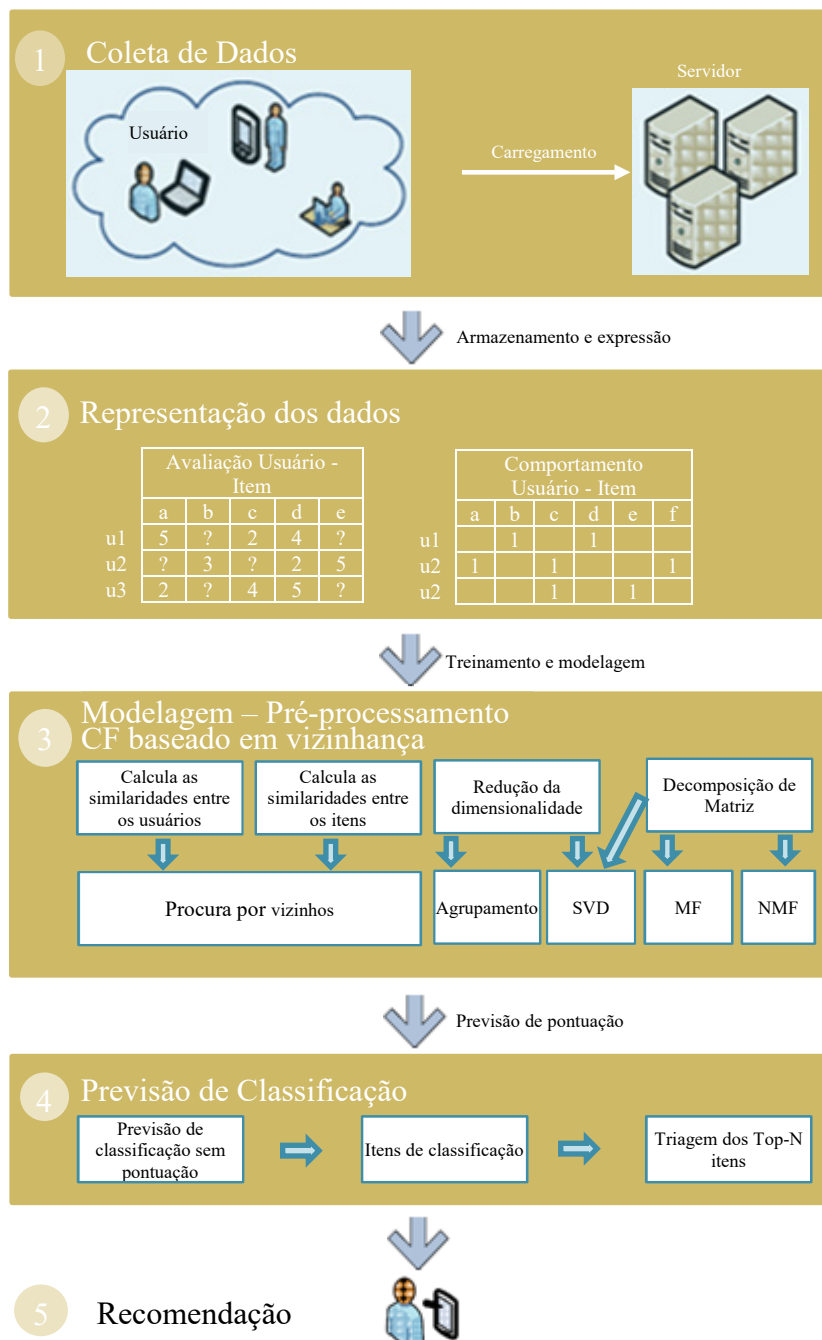
Esta abordagem é considerada colaborativa por criar ou classificar grupos de usuários que compartilham perfis, comportamentos ou classificações similares para recomendar produtos, serviços ou conteúdo, de acordo com as avaliações geradas pelo grupo ao qual um usuário pertence. É baseada no cálculo de semelhança entre usuários ou itens (CAPUANO et al., 2014; CARRER-NETO et al., 2012).

Segundo Jannach et al. (2010), as abordagens colaborativas tradicionais utilizam uma matriz de avaliações de itens dos usuários como entrada e normalmente produzem, como saída, uma previsão do tipo numérica, indicando se o usuário gostará de um determinado item ou não, bem como, uma lista de itens recomendados (Figura 17).

No entanto, Chen, R. et al. (2018) advertem que com o advento do *Big Data*, a abordagem da filtragem colaborativa começou a sofrer de várias deficiências, como problemas de dispersão de dados, inicialização a frio (*cold start*) e escalabilidade, que afetam seriamente a qualidade das recomendações.

Para resolver os problemas mencionados, muitas técnicas de mineração de dados e de aprendizado de máquina, como *Clustering*, *Singular Value Decomposition (SVD)*, *Probability Mass Function (PMF)* e *Non-Negative Matrix Factorization (NMF)* são propostas para melhorar o desempenho dos sistemas de recomendação. Na era do *Big Data*, para resolver os problemas de dispersão de dados e *cold start*, fatores sociais (interesses pessoais, influência interpessoal, *status* social) são considerados recentemente relevantes para melhorar o desempenho (CHEN, R. et al., 2018).

Figura 17 – Filtragem Colaborativa.



Fonte: Chen, R. et al. (2018) – Traduzido e adaptado pelo autor.

A abordagem de filtragem colaborativa apresenta, entre outras, as seguintes vantagens e desvantagens (AAMIR; BHUSRY, 2015):

- Vantagens:
 - Não necessita de uma representação de itens em termos de recursos, sendo baseado apenas no julgamento da comunidade de usuários participantes;
 - Não necessita de envolvimento humano. Portanto, a escalabilidade do banco de dados de itens é elevada;
 - Para realizar as previsões, podem utilizar recomendações cruzadas, pois geralmente essas recomendações são diferentes para cada usuário. Portanto, não requer nenhum tipo de conhecimento de domínio, economizando tempo.
- Desvantagens:
 - O item, para ser recomendado a um usuário, deve ser classificado por outro(s) usuário(s) ou correlacionado com outros itens semelhantes do banco de dados;
 - Apesar de haver a disponibilidade de grandes bancos de dados para muitos aplicativos de *e-commerce*, os usuários ativos não classificam todos os itens, ou seja, existe esparsidade podendo conduzir a resultados muito escassos;
 - A abordagem é demorada e dispendiosa em virtude da diversidade dos dados.

Este tipo de recomendação tem sido aplicado em sistemas de varejo *online*, recomendando produtos conforme compras ou classificações de usuários em diversas áreas comerciais. O Quadro 11 apresenta algumas aplicações dos sistemas de recomendação em várias áreas.

Quadro 11 – Aplicações dos Sistemas de Recomendação.

Categoria	Sistema de Recomendação
Vídeo	Netflix®, Hulu®, Youtube®, Youku®, iQiyi®, Tudou®, Ku6®
Notícias	Google News®, Ifeng®, Toutiao®, NetEase News®, Digg®
Música	Yahoo! Music®, Pandora®, Douban Music®, QQ Music®, Google Play®, Last.fm®
Redes Sociais	Facebook®, Twitter®, Sina Weibo®, QQ®, LinkedIn®
<i>E-Business</i>	Amazon®, eBay®, Taobao®, JD®, Suning®

Fonte: Chen, R. et al. (2018).

Por ter sido explorado por muitos anos, vários algoritmos e técnicas foram propostas e avaliadas com sucesso. Uma das técnicas mais populares é a baseada nos vizinhos mais

próximos, devido a sua simplicidade, eficiência e capacidade de produzir recomendações precisas e personalizadas (JANNACH et al., 2010).

Wang e Wu (2011) citam o método de mineração de dados denominado regras de associação, para encontrar associações úteis que podem ser usadas como base para recomendações. Burke (2007) destaca as técnicas de aprendizagem, incluindo redes neurais, indexação semântica latente (*LSI – Latent Semantic Indexing*) e redes bayesianas. Koren e Beel (2015) apresentam os modelos de fatoração de matriz e medidas de similaridade.

Jannach et al. (2010) também descrevem as técnicas de filtragem colaborativa, em especial a recomendação do vizinho mais próximo, que identifica outros usuários que possuem preferências semelhantes às do usuário atual.

Esta técnica pode ser desenvolvida utilizando o coeficiente de *Pearson*, no qual traz valor 1 para forte correlação positiva, e valor -1 para forte correlação negativa. Outras métricas também podem ser utilizadas para medir a proximidade entre usuários, tais como, semelhança de cosseno ajustada, coeficiente de correlação de classificação de *Spearman* e a medida de diferença quadrática média.

Entretanto, a medida de similaridade do cosseno é considerada a métrica-padrão, por demonstrar resultados precisos em diversas pesquisas (KHRIBI; JEMNI; NASRAOUI, 2008; JANNACH et al., 2010; RODRÍGUEZ; DUQUE; OVALLE, 2016; TAKANO; LI, 2010).

2.6.4 Abordagem Demográfica

Neste tipo de abordagem, o sistema recomenda itens com base no perfil demográfico do usuário. Desta forma, a suposição é que diferentes recomendações devem ser geradas para diferentes nichos demográficos (BOBADILLA et al., 2013). Diversos sites adotam soluções de personalização simples e eficazes com base em dados demográficos. Por exemplo, os usuários são enviados para determinados sites com base em seu idioma ou país. As sugestões também podem ser personalizadas de acordo com a idade do usuário. Embora essas abordagens tenham sido bastante populares na literatura de marketing, tem havido relativamente pouca pesquisa adequada de sistemas de recomendação sobre sistemas demográficos (RICCI; ROKACH; SHAPIRA, 2015).

2.6.5 Baseada em Conhecimento

Aqui, são recomendados itens com base no conhecimento de domínio específico sobre como determinados recursos de itens atendem as necessidades e preferências dos usuários, ou seja, como o item é útil para o usuário. Sistemas de recomendação baseados em conhecimento

tem se baseado na análise de casos passados similares (CHEN; PU, 2012). Nesses sistemas, uma função de similaridade estima o quanto a necessidade do usuário combina com as recomendações. Assim, a pontuação de similaridade pode ser interpretada diretamente como a utilidade da recomendação para o usuário (RICCI; ROKACH; SHAPIRA, 2015).

Sistemas baseados em restrições representam outro tipo de sistemas de recomendação baseados em conhecimento.

Com relação ao conhecimento usado, ambos os sistemas (baseados em casos ou restrições) são semelhantes: os requisitos do usuário são coletados; são propostas automaticamente soluções para requisitos inconsistentes em situações onde nenhuma solução pode ser encontrada; e os resultados da recomendação são explicitados. A principal diferença está na forma como as soluções são calculadas. Sistemas de recomendação baseados em casos determinam recomendações com base em métricas de similaridade, ao passo que os sistemas de recomendação baseados em restrições exploram predominantemente bases de conhecimento predefinidas que possuem regras explícitas sobre como relacionar os requisitos do cliente com os recursos do item.

No início da implementação, os sistemas baseados em conhecimento tendem a funcionar melhor do que outros, mas se não estiverem equipados com componentes de aprendizado, eles podem ser superados por outros métodos que podem explorar e utilizar os logs da interação humano / computador (RICCI; ROKACH; SHAPIRA, 2015).

2.6.6 Baseado na comunidade

A recomendação neste tipo de sistema é baseada nas preferências dos amigos dos usuários. (ARAZY; KUMAR; SHAPIRA, 2009). Segundo Sinha e Swearingen (2001), as pessoas tendem a confiar mais em recomendações de seus amigos do que em recomendações de indivíduos semelhantes, mas desconhecidos. Desta forma, os sistemas de recomendação social é uma designação utilizada para sistemas baseados na comunidade, que utilizam as redes sociais abertas para recomendação de itens (GOLBECK, 2006).

Essa abordagem modela e adquire informações sobre as relações sociais dos usuários e as preferências dos amigos do usuário, sendo que a recomendação é baseada em classificações fornecidas pelos amigos do usuário. Por meio da extração de informações constantes nas redes sociais, objetiva-se identificar e realizar a inferência de características, preferências e interesses educacionais dos usuários, através de técnicas de extração de texto e web semântica (PEREIRA, C. et al., 2014). Conclui-se, então, que esta abordagem de sistema de recomendação vem acompanhando o crescimento das redes sociais, permitindo, portanto, uma estratégia simples e

abrangente para aquisição de dados relacionados às relações sociais dos usuários (RICCI; ROKACH; SHAPIRA, 2015).

2.6.7 Sistemas de Recomendação Híbridos

Esta abordagem baseia-se na combinação das técnicas anteriormente comentadas. Um sistema híbrido combina técnicas de recomendação diferentes para que as vantagens de uma abordagem corrijam as desvantagens da outra. Por exemplo, os métodos de filtragem colaborativa sofrem de problemas de novos itens, isto é, eles não podem recomendar itens que não têm classificações. Isso não limita as abordagens baseadas em conteúdo, já que a previsão para novos itens é baseada em suas descrições (características) que normalmente estão facilmente disponíveis (RICCI; ROKACH; SHAPIRA, 2015).

Dadas duas (ou mais) técnicas básicas de recomendação, várias maneiras foram propostas para combiná-las para criar um novo sistema híbrido. Como já mencionamos, o contexto do usuário quando ela está buscando uma recomendação pode ser usado para personalizar melhor a saída do sistema. Por exemplo, em um contexto temporal, as recomendações de férias no inverno devem ser muito diferentes das fornecidas no verão. Ou uma recomendação de restaurante para um sábado à noite com seus amigos deve ser diferente daquela sugerida para um almoço de dia de trabalho com colegas de trabalho.

2.7 TRABALHOS RELACIONADOS

Dentre os trabalhos encontrados na literatura que lidam com algum tipo de perfil de aluno, ou perfil de usuário, buscou-se incluir aqueles que guardam algum tipo de relação com o tema tratado nesta dissertação.

Além disso, o modelo aqui proposto visa evoluir o modelo de Bernardino (2017), no sentido de incluir outras informações ou características que possam ou melhorar a acurácia da recomendação ou permitir a aplicação de outras técnicas de LA que possam subsidiar a tomada de decisões no contexto educacional.

Em seu modelo, a autora define um modelo de perfil do educando a partir de suas preferências e interações, com o objetivo de permitir a recomendação de conteúdos em sistemas educacionais *online*. A estrutura principal do modelo contemplou quatro módulos: Identificação, Aprendizagem, Educacional e Histórico.

O módulo “Identificação” contemplou as principais informações pessoais do usuário, onde constam dados como o nome, nome de contato, data de nascimento, e-mail, sexo, estado civil, escolaridade e idioma.

O módulo “Aprendizagem” guarda as informações referentes ao estilo de aprendizado do usuário em relação as suas preferências no processo de ensino. Dentre os itens que compõem este módulo, tem-se o estilo cognitivo, o estilo de aprendizagem, acessibilidade e áreas de interesse.

O estilo cognitivo contemplou cinco modalidades de aprendizagem, conforme o modelo VARK (visuais, auditivas, leitura/escrita, cinestésica e multimodais). O estilo de aprendizagem utiliza o modelo FSLSM – *Felder-Silverman Learning Style Model* (FELDER; SILVERMAN, 1998) com quatro dimensões: sensorial/intuitivo, visual/verbal, ativo/reflexivo e sequencial/global. A acessibilidade é classificada em auditiva, visual, auditiva/visual ou nenhuma. As áreas de interesse estão relacionadas com os interesses dos educandos em campos de estudo.

O módulo “Educativo” armazenou os cursos, disciplinas e atividades, em que os alunos estão matriculados. As atividades são compostas por conteúdos, apostilas, fórum, aulas virtuais, avaliações ou outras atividades relacionadas com o curso e disciplina que poderão estar cadastradas no sistema.

O módulo “Histórico” representou as informações e interações que os alunos podem realizar no sistema educacional, como dados de acesso, links acessados, atividades acessadas, tempo de acesso e classificação. A classificação é o *feedback* do usuário referente a atividade acessada, que pode ser classificada como “Curtiu” e “Não Curtiu”. Os dados de acesso, neste modelo, contemplaram o tipo de dispositivo que o educando utilizou no acesso ao sistema, tipo de navegador (browser), sistema operacional e data. Nos links acessados foram contemplados os menus e atividades que o aluno acessou no sistema.

Nas atividades acessadas foi previsto a descrição da atividade e a hora de início e fim, resultando no tempo de acesso e a classificação que o aluno informou referente a atividade acessada.

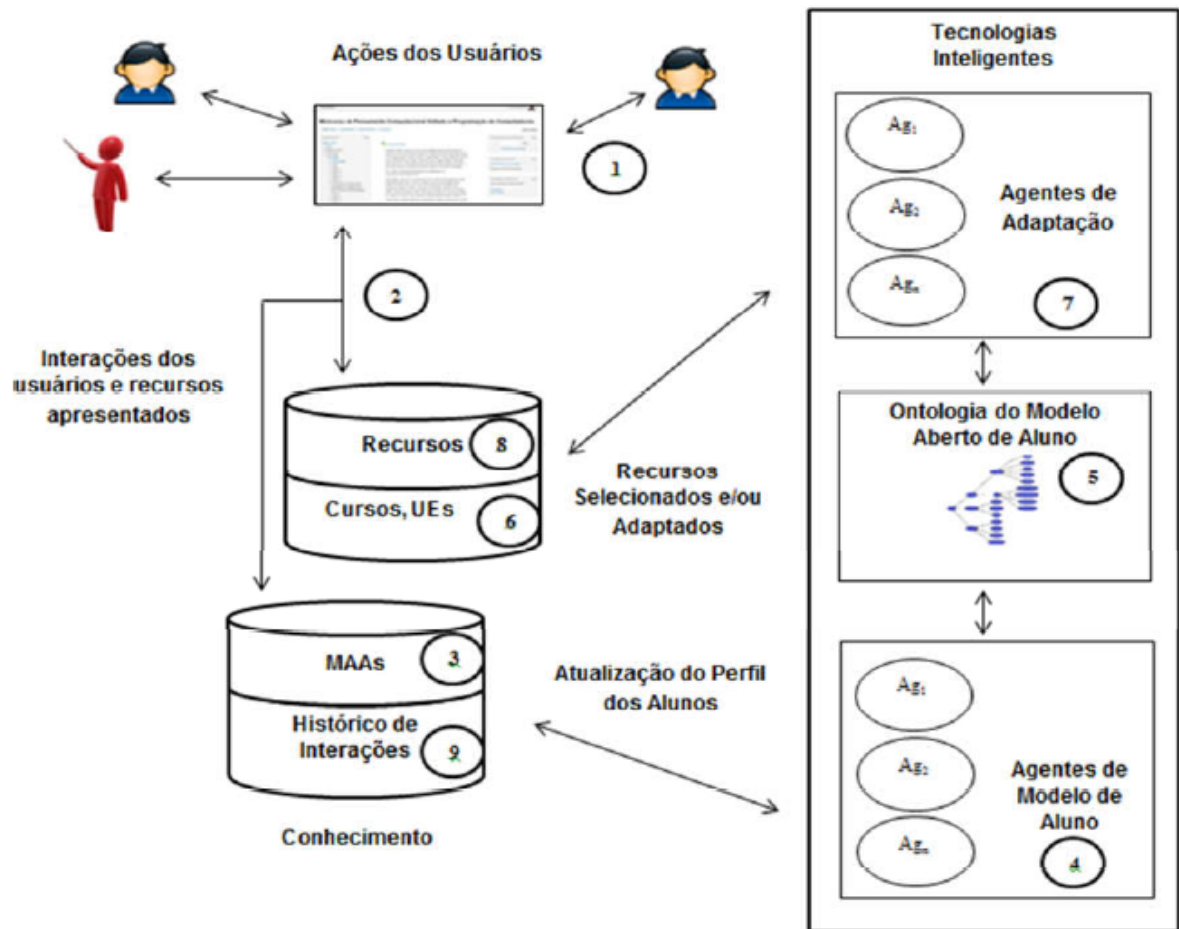
Segundo a autora, as informações a serem utilizadas pelo modelo proposto podem ser coletadas de forma implícita ou explícita, dependendo da estrutura existente nos sistemas de gestão educacional dos cenários de aplicação. O módulo “Identificação” pode selecionar todos os dados do cadastro do educando de forma implícita, caso exista no sistema de gestão, do contrário, deverá solicitar o preenchimento das informações de forma explícita. No módulo “Aprendizagem” é solicitado ao aluno que responda às suas preferências de aprendizagem

tornando a captura das informações explícita. Já o módulo “Educativo” pode selecionar todas as informações de forma implícita quando existir no sistema de gestão educacional. Já o módulo “Histórico” realiza a captura das informações de forma implícita e dinâmica, conforme o aluno navega no sistema educacional *online*.

Em sua tese de doutorado, Bremgartner (2017) baseia-se em um *framework* chamado de ArCARE (Arcabouço Conceitual de Adaptação de Recursos Educacionais), apresentando uma estratégia que permite a criação de *Smart Learning Environments* (SLEs) e adaptação de recursos para alunos em AVAs no decorrer do curso, com base no Construtivismo de Piaget. Este *framework* utiliza a tecnologia de sistema multiagente, que manipula uma ontologia de modelo aberto de aluno composto de várias características do aluno, de acordo com suas interações com o AVA e seu engajamento no curso, tais como competências, habilidades, desempenho nas atividades, frequência, preferências e estilos de aprendizagem (Figura 18).

Desta forma, estes recursos adaptados no AVA podem se originar em uma recomendação de colegas, para ajudar alunos com dificuldades em alguma atividade, bem como também a recomendação de propostas de atividades colaborativas, com o objetivo de o aluno ter uma aprendizagem mais efetiva do conteúdo de um determinado curso.

Figura 18 – Arquitetura de referência do ArCARE.



Fonte: Bremgartner (2017).

A solução descrita na sua tese também se baseia na aprendizagem colaborativa pela interação entre aprendizes, na qual um método é empregado com o propósito de construir seu conhecimento de maneira mais significativa em um AVA. Além disso, o modelo de cada aluno é alterado dinamicamente durante o curso, por meio das interações do aluno com o AVA. Este modelo é apresentado para o aluno, utilizando-se LA, com o objetivo de o aluno conhecer seu estado atual no curso, permitindo uma aprendizagem autorregulada.

Segundo o autor, as informações que fazem parte do perfil do aluno podem se originar em diversas fontes: o próprio aluno, a interação do aluno com o ambiente, o professor ou o administrador do sistema. Desta forma, podem ser utilizadas as informações mais relevantes que irão constituir em um modelo de aluno para algum objetivo específico no sistema (adaptação do sistema), como preferências, conhecimento, objetivos e histórico de navegação.

O autor analisou o modelo de dados para informações sobre alunos, definidos por organizações internacionais, como aqueles definidos pelo *Institute of Electrical and Electronics Engineers (IEEE PAPI)* e *Instructional Management Systems (IMS LIP)*, EDUPERSON e

Universal Learning Format (ULF), dentre outras, que trabalham na padronização de informações de EaD, definindo modelos de dados para informações sobre os alunos.

O modelo escolhido foi o IMS LIP, que possui esquemas XML para implementação do modelo e possui uma divisão em 11 categorias:

- *Identification*: contém os dados demográficos e biográficos do aluno;
- *Accessibility*: contém dados sobre a acessibilidade da informação ao usuário, idioma, preferências (físicas e tecnológicas);
- *Goal*: descreve os objetivos e desejos do aluno;
- QCL: nesta categoria são descritas as qualificações, certificados e licenças (ou diplomas) do aluno;
- *Activity*: qualquer atividade relacionada com estudo, já concluída ou em andamento, é referenciada nesta categoria;
- *Competency*: descreve as habilidades do aluno, experiências e conhecimentos já adquiridos. Essas habilidades podem ser associadas com base no histórico de realizações do aluno, podendo ser obtidas com dados das categorias QCL e *Activity*;
- *Interest*: contém os interesses do aluno;
- *Affiliation*: possui informação sobre membros de organizações profissionais;
- *Transcript*: contém um resumo por instituições, das realizações acadêmicas do aluno;
- *Security key*: contém as informações de segurança e privacidade relacionadas ao aluno;
- *Relationship*: contém os relacionamentos definidos entre o usuário e sua identificação, acessibilidade, qualificações, competências, objetivos, atividades, interesses, chaves de segurança e afiliações.

Entretanto, o autor ressalta as deficiências do modelo, como a presença de dados conflitantes que se sobrepõem e provocam dúvidas quando se vai realizar a classificação dos dados nas diversas categorias, por exemplo QCL (Qualificações, Certificações e Licenças), *Activity* (Atividades), *Transcript* (Transcrições) e *Competency* (Competências).

Além disso, um dos principais problemas do LIP é o fato dele ser muito extenso, pois descreve muitas categorias de dados que normalmente não são utilizadas em um único sistema. Portanto, cada sistema utiliza as categorias que são consideradas mais relevantes para o fim específico a que se destina.

Rajenthiran, Shaharaneer e Jamil (2017) usaram a abordagem de mineração de dados, através da análise de agrupamento, para estabelecer um perfil dos alunos que participam de atividades extracurriculares. Os autores citam a participação e o envolvimento dos alunos em esportes, debates, clubes, publicações escolares, conselhos estudantis assim como outros eventos sociais como fundamentais para melhorar a performance acadêmica dos alunos. Outras atividades extracurriculares que podem ser citadas: esportes, atividades musicais, debate, modelo, arte, música, drama, debate e discussão, concurso de declamação, concurso de escrita de histórias, concurso de redação, artesanato de arte, competição de recitação, decoração de revistas de parede, escrita de artigos para revista escolar, músicas folclóricas, dança folclórica, show de flores, decoração escolar, fazer escultura, competição de vestido de fantasia, preparação de gráfico e modelos, fazer álbum, fotografia, modelagem de argila, fabricação de brinquedos, fabricação de sabão, fazer cesta, exposições organizacionais, celebração de festival, etc.

Assim, selecionaram algumas variáveis para o processo de agrupamento, como: gênero, renda familiar e experiência acadêmica.

Para o trabalho aqui proposto, vamos algumas dessas informações de atividades extracurriculares para compor o perfil do aluno. Assim, sendo, a tabela de atividades extracurriculares será composta pelos seguintes campos: id_atividade, descrição, data de início e data final da atividade.

Jha; Shorko (2018) em seu estudo desenvolveram uma abordagem algorítmica para o perfil do estudante na seleção de disciplinas e profissões. Foi proposto um algoritmo calculativo lógico-memorável (CALM) que contempla o pensamento calculativo, a associação, a percepção lógica e os recursos de memorização, capaz de projetar uma estrutura de perfis de estudantes que leva a uma melhoria na qualidade do processo de aprendizagem.

Desta forma, os autores defendem que os traços de personalidade têm uma contribuição significativa ao aprendizado. Neste sentido, utilizaram oito traços de personalidade para correlacionar a temática de atuação profissional e carreira.

Esses oito traços serão também utilizados para complementar o perfil do aluno proposto, através de uma nova tabela. O primeiro traço é extrovertido, ou seja, alguém que seja extremamente extrovertido e sociável. O segundo traço é introvertido, ou que se afasta da realidade. O terceiro traço é sensitivo, ou emocionalmente sentimental.

O quarto traço é intuitivo, ou rigorosamente analítico. A quinta característica é pensador ou intelectualmente estimulado. O quinto traço é sentimental, ou aquele que é especulativo e

vulnerável. Perceptível é o sétimo, com a característica de saber lidar com detalhes. O último traço é crítico ou julgador, que se caracteriza pela habilidade em fazer julgamentos racionais.

Há de se observar que os autores chegam a esses traços de personalidade através da utilização de questionários que geram uma pontuação. Ao final, de acordo com a pontuação obtida, o perfil é atribuído ao aluno. Entretanto, não utilizaremos o questionário, por não ser o objetivo desta dissertação o estabelecimento dos traços de personalidade e sim, o quanto essa informação pode influenciar nas análises de aprendizagem.

Com o intuito de prever a performance do estudante, através da análise dos dados de seu perfil, DAUD et al. (2017) utilizaram as informações dos gastos familiares e informações pessoais dos alunos coletando dados em diferentes universidades do Paquistão. Através da utilização de técnicas de LA, como modelos de classificação discriminativa e generativa, eles objetivaram prever se um estudante iria conseguir terminar ou não o seu curso.

Para tal intento, utilizaram os modelos de classificação *Support Vector Machine* (SVM), C4.5, árvores de classificação e regressão (CART), Redes Bayesianas (BN) e *Naive Bayes* (NB).

Os autores concluem, de forma geral, que LA baseada em recursos personalizados pode melhorar a previsão do desempenho dos alunos e que as descobertas feitas requerem pesquisas adicionais para a incorporação de novas informações, como talentos, habilidades e competências pessoais de diferentes fontes on-line.

Al-Mamoori, Shehab e Fakharany (2018), realizaram a concatenação de duas ontologias bidimensionais, que utilizam como dados as atividades realizadas pelos estudantes no LMS e o comportamento *online* em uma rede social para aprimorar o seu perfil. Segundo os autores, o modelo fornece informações vitais aos educadores proporcionando uma melhor compreensão da personalidade de cada aluno.

Dentre os dados coletados da rede social, destacam-se aqueles apostos no Quadro 12.

Quadro 12 – Dados do estudante em rede social.

Atributos	Explicação
Pid P	Id do <i>post</i> principal
Id_Student_P	Id da postagem do usuário
Student_Name_P	Nome do usuário
Time_Stamp_P	Tempo gasto visitando diferentes páginas
Shares_P	Número de compartilhamentos
Likes_P	Número de <i>likes</i>

Fonte: Al-Mamoori, Shehab, e Fakharany (2018) – Adaptado pelo autor.

No artigo de Tan et al. (2018), os autores utilizaram LA para investigar a experiência de aprendizado de estudantes internacionais em um MOOC, no idioma inglês, criado pela Universidade de *Edinburgh*. Através da utilização do algoritmo de análise de cluster *K-means*,

os autores classificaram os estudantes em quatro categorias majoritárias e identificaram três tipos de comportamento de aprendizado como significantes para prever a performance dos estudantes.

Os autores asseveram que a literatura existente tem retratado o perfil do estudante através das experiências acadêmicas, gênero, idade, da percepção própria da aprendizagem *online*, nível de participação nas atividades propostas e da habilidade na linguagem. As variáveis consideradas estão representadas no Quadro 13.

Quadro 13 – Variáveis consideradas no aprendizado *online*.

Variáveis		Origem das variáveis
Navegação	Arquivos de Vídeo	Número de visualizações do vídeo
	Arquivo	Número de download de arquivos pdf ou cliques
	Página	Quantas vezes o aluno visualizou a página do curso
Discussão	Fórum	Quantas vezes o aluno visualizou o fórum do curso
	Criação de <i>post</i>	Número de posts criado pelo aluno
	Novo	Número de alunos que responderam as postagens temáticas
	Resposta	Número de alunos que responderam posts de outros alunos
Nível do inglês		O nível é determinado pelo grau de CET 4 e 6
Comparecimento na videoconferência		Se os alunos participaram da videoconferência
Performance acadêmica		A pontuação média do teste da segunda semana

Fonte: Tan et al. (2018) – Adaptado pelo autor.

De acordo com a atitude e performance acadêmica, as quatro categorias de alunos foram nomeadas da seguinte forma: alunos competentes e ativos, alunos competentes e inativos, alunos incompetentes e inativos, e alunos espectadores.

Rezende et al. (2018) evoluíram a ontologia chamada PERSONNA, para modelagem do perfil e do contexto do aluno em um ambiente de *E-learning*, para o que chamaram de PERSONNA-TEC. Esta ontologia, integrada com a arquitetura BROAD-RS (BROAD Recommendation System), se mostra capaz de realizar a recomendação de objetos de aprendizagem sensíveis ao contexto, facilitando o processo de recomendação de objetos de aprendizagem aderentes aos dispositivos utilizados pelo aluno.

De Melo, Dantas e Fernandes (2017) propuseram um modelo do estudante baseado em emoções e perfis de personalidade com a finalidade de selecionar as estratégias pedagógicas adequadas a cada perfil afetivo. A validação foi realizada através da utilização das técnicas de redes neurais artificiais e árvores de decisão. Os resultados obtidos foram considerados precisos para integração em AVAs, assim como a otimização das estratégias pedagógicas adequadas a cada perfil.

3 MODELO PROPOSTO

Nesta seção, serão descritas as principais características e informações referentes ao modelo do perfil, sendo apresentada também a estrutura e o modelo lógico. Para os fins aqui propostos, perfil de usuário e perfil de aluno assumem o mesmo significado.

3.1 DESCRIÇÃO DO MODELO

Com o fito de subsidiar a construção do modelo, os trabalhos correlatos na área foram examinados, visando identificar os atributos mais significantes que possam otimizar o resultado da aplicação de técnicas de LA. Desta forma, as técnicas mais utilizadas foram identificadas, assim como os tipos de dados mais utilizados em cada abordagem.

De acordo com a revisão da literatura, pôde-se observar que os dados que compõem o perfil do usuário ou aluno originam-se de diversas fontes. Quando os dados são coletados através de pesquisas realizadas com os usuários, a coleta é explícita. Quando os dados são oriundos da interação dos alunos com o ambiente de aprendizagem, ou outros meios, a coleta é implícita.

Os dados coletados podem ser utilizados para uma determinada finalidade ou objetivo, a depender da necessidade do *stakeholder*. Dentre os objetivos pesquisados, encontram-se a personalização de conteúdo, a personalização baseada no contexto, melhorar a compreensão da personalidade de cada aluno e predição do desempenho acadêmico. A metodologia pedagógica também se beneficia da aplicação das técnicas de LA, na medida que as recomendações geradas podem inferir se o aluno tem necessidades específicas e, a partir disso, proporcionar elementos para se avaliar possíveis mudanças.

Cada vez mais, o perfil de alunos torna-se fundamental para a aplicação de técnicas de LA visando subsidiar a tomada de decisão. O perfil pode consistir de vários tipos diferentes de informações, como os conhecimentos anteriores do aluno, o estilo de aprendizagem, os pontos proeminentes, as fraquezas, o ambiente de aprendizagem preferido, a progressão acadêmica, entre outros. Neste trabalho, o foco consiste em dois tipos de informação:

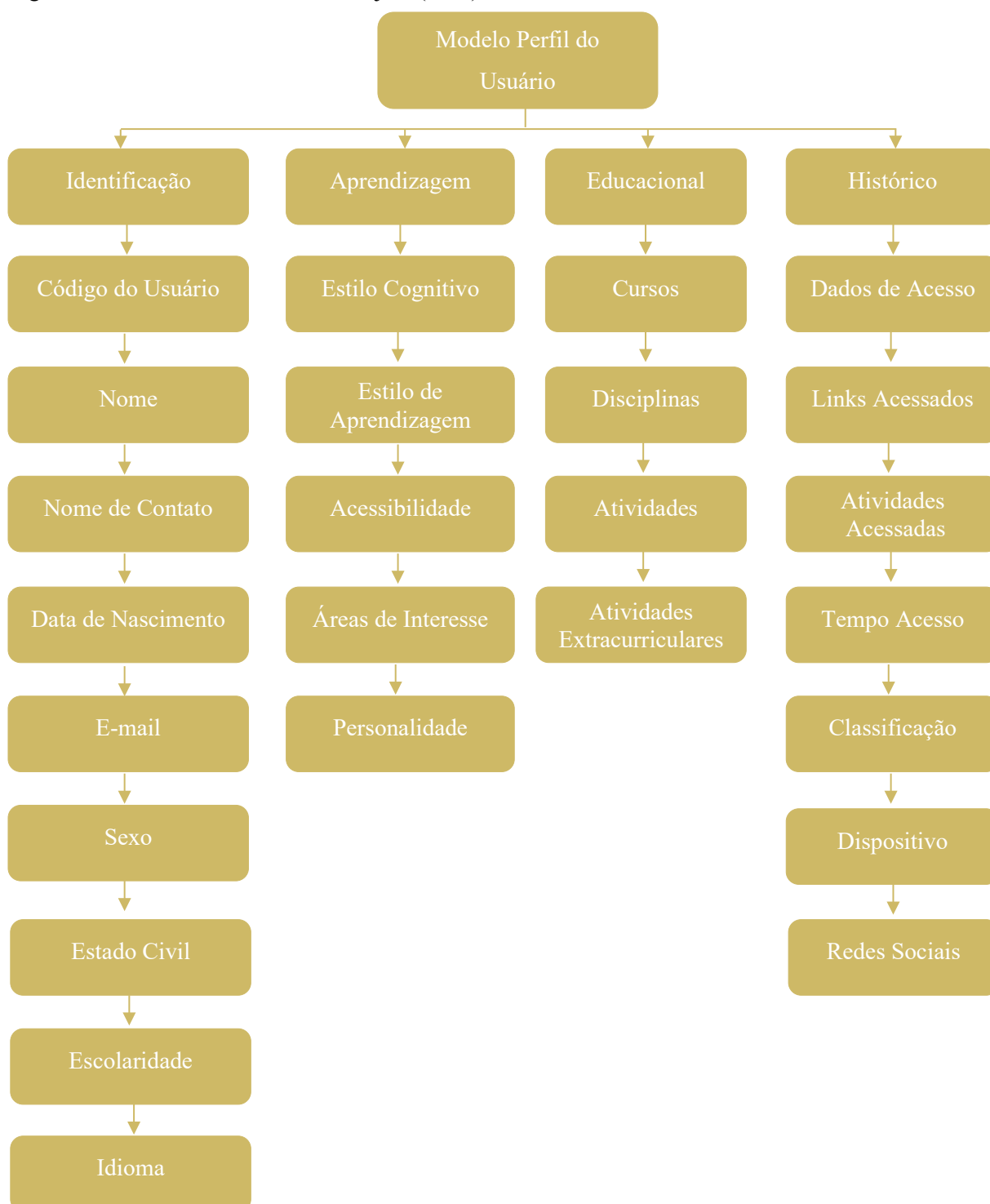
1. Um histórico das interações do usuário com o ambiente de ensino. Isso pode incluir os conteúdos que foram lidos, as atividades realizadas, os exercícios que foram completados, o tempo de realização de atividades *online* e várias outras informações que constam nos arquivos de interação do aluno com o ambiente virtual;

2. Informações sobre o estilo de aprendizagem e sobre o tipo de personalidade do aluno. O estilo de aprendizagem permite designar conteúdos e atividades direcionadas, que podem aumentar o desempenho acadêmico. Por outro lado, o tipo de personalidade, é utilizado para recomendar atividades a outros alunos com o mesmo perfil, baseado na filtragem colaborativa. Assim, através desta técnica de LA, pode-se recuperar os n itens com maior probabilidade de interesse para o aluno, de acordo com sua personalidade.

O perfil aqui definido possui dados estáticos e dinâmicos. Os dados dinâmicos são aqueles oriundos da interação com o ambiente de aprendizagem *online*, ou da resolução de alguma atividade *online*. Os dados estáticos são aqueles obtidos através das formas tradicionais, como questionários.

A Figura 19 representa a Estrutura Analítica do Projeto, com a descrição do modelo proposto. O modelo proposto neste trabalho teve como base o modelo de Bernardino (2017). Para tal, novas perspectivas de informações para análise foram incluídas, dentre elas o tipo de personalidade dos alunos, as atividades extracurriculares e o histórico de acesso às redes sociais. Todavia, a quantidade de módulos permaneceu inalterada, constituindo-se de quatro módulos: Identificação, Aprendizagem, Educacional e Histórico.

Figura 19 – Estrutura Analítica do Projeto (EAP).



Fonte: Elaborada pelo autor (2019).

As informações do módulo “Identificação”, representam os dados pessoais do usuário, como código do usuário, nome, nome de contato, data de nascimento, e-mail, sexo, estado civil, escolaridade e idioma.

O módulo “Aprendizagem” objetiva representar o estilo de aprendizagem do aluno em relação as suas preferências no processo de ensino. Assim, as preferências podem ser classificadas pelo estilo cognitivo, estilo de aprendizagem, acessibilidade e áreas de interesse.

O “Estilo Cognitivo” é determinado através do modelo VARK, sendo composto pelas dimensões: visual, auditiva, leitura/escrita, cinestésico e multimodal. Para a definição do estilo de aprendizagem, utilizou-se o modelo FSLSM, com quatro dimensões: sensorial/intuitivo, visual/verbal, ativo/reflexivo e sequencial/global.

A “Acessibilidade” é classificada em auditiva, visual, auditiva/visual ou nenhuma. As áreas de interesse estão relacionadas com os interesses dos alunos em cada campo de estudo.

As “Atividades Extracurriculares” podem ser obtidas através de questionário. Portanto, a aquisição destes dados é explícita, não sendo obtida através dos vestígios de navegação dos alunos no ambiente *online* de ensino.

Os “Traços de Personalidade” podem ser obtidos através de um questionário, apresentado no primeiro acesso ao ambiente *online* de ensino. Entretanto, neste trabalho, os dados foram gerados de forma aleatória, com o intuito de simular a coleta destes dados.

O módulo “Educaional” é utilizado para armazenar as informações referentes aos cursos, disciplinas e atividades que estarão cadastradas no sistema.

A “Área de Histórico” é utilizada para armazenar as interações do aluno com o sistema, como dados de acesso, os links que determinado aluno acessou, as atividades acessadas, o tempo de acesso e a classificação (feedback do aluno quanto a atividade acessada).

Os dados referentes ao dispositivo utilizado pelo aluno podem ser coletados automaticamente e incluídos no banco de dados. Outras informações também podem ser consideradas relevantes para realizar recomendações de objetivos de aprendizagem de acordo com o contexto do aluno. Desta forma, se o aluno estiver utilizando um dispositivo com pouca capacidade gráfica, e sinal de *internet* ruim, pode-se sugerir objetos de aprendizagem compatíveis à tela do dispositivo utilizado pelo aluno.

Dados referentes ao comportamento nas redes sociais podem ser obtidos de forma implícita, com ferramentas adequadas para extração das informações. Quando incluído no perfil dos alunos, também podem ser utilizadas para aprimorar o perfil e subsidiar decisões educacionais.

3.2 MODELO CONCEITUAL E LÓGICO

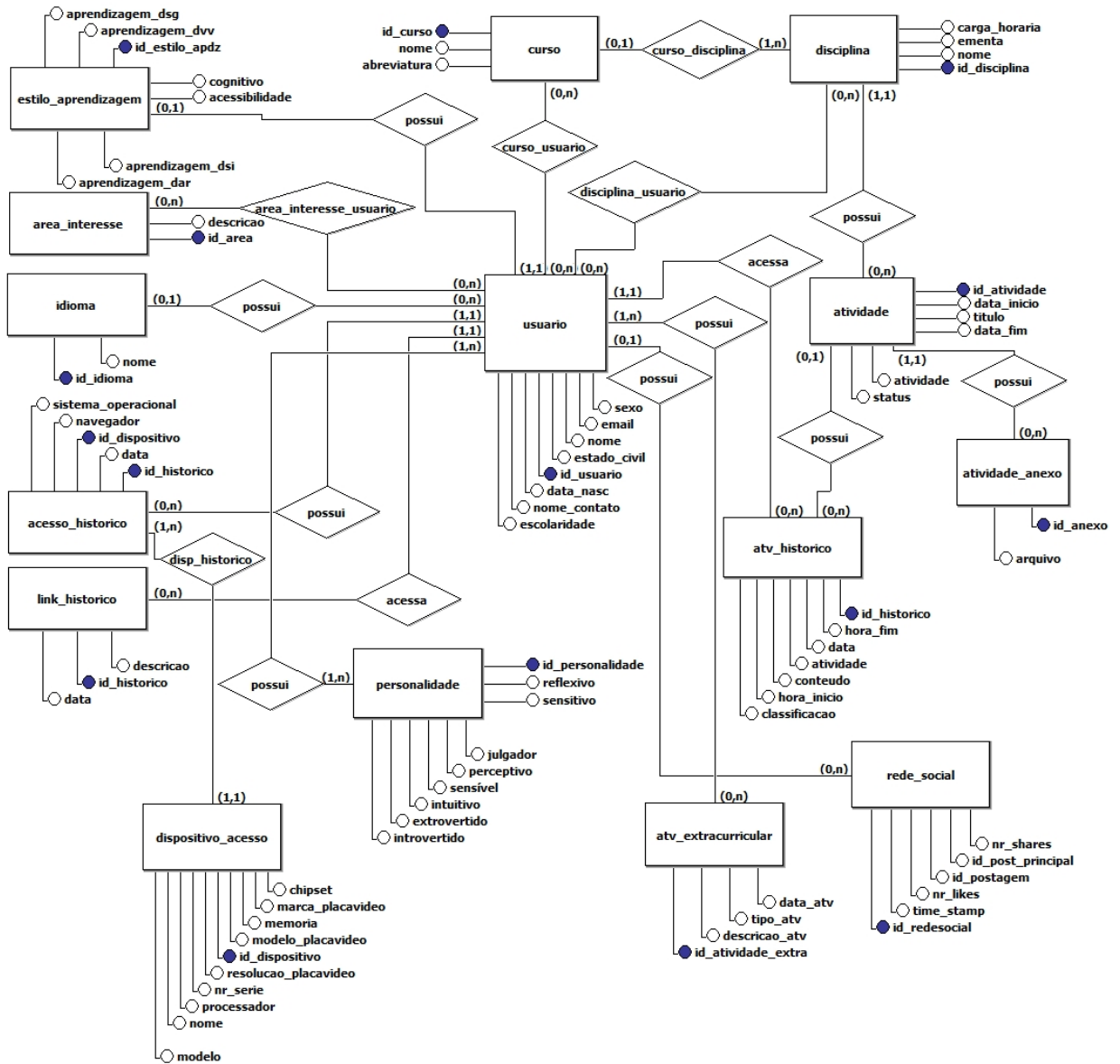
O modelo conceitual (Figura 20) consiste de uma estrutura em blocos representativa de todas as relações entre cada componente da estrutura, independente de tecnologia, onde são identificados as tabelas e os relacionamentos entre elas. É formado pelas tabelas, que podem ser coisas, pessoas, lugares, organizações; pelos relacionamentos entre as tabelas, que podem ter diferentes graus (binário, ternário, n-ário); pela cardinalidade, que representa a quantidade de vezes que um elemento de um conjunto de tabelas pode, em um determinado instante, estar associado em um dado relacionamento, a outros elementos de outras tabelas; e pelos atributos, que são as informações que caracterizam as tabelas e os relacionamentos.

O modelo lógico (Figura 21), envolve a análise dos dados que futuramente serão armazenados para que se defina o tipo de cada um dos atributos de cada tabela. Portanto, ele é independente do modelo físico. Desta forma, esse modelo retrata as ligações entre as tabelas do banco de dados, as chaves primárias e outras informações relacionadas à normalização, devendo ainda informar, para determinada coluna, o tipo de dado, o seu tamanho e se o dado é opcional ou não.

O processo de derivação do modelo conceitual geralmente é realizado em várias etapas (ELMASRI; NAVATHE, 2011), sendo:

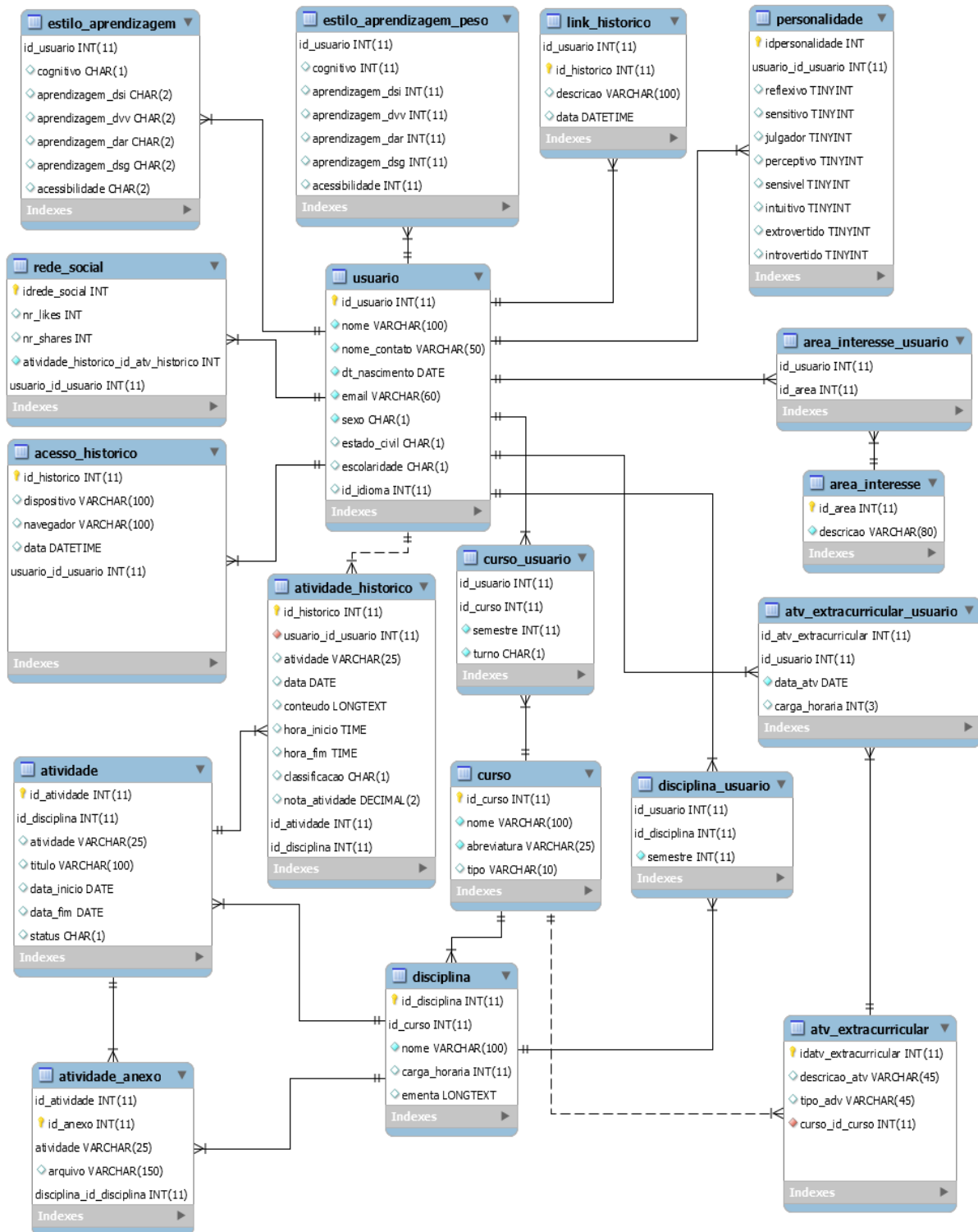
- Mapeamento dos tipos de entidades regulares;
- Mapeamento dos tipos de entidades fracas;
- Mapeamento dos tipos de relacionamentos binários 1:1;
- Mapeamento de tipos de relacionamentos binários 1:N;
- Mapeamento de tipos de relacionamentos binários N:M;
- Mapeamento de atributos multivalorados;
- Mapeamento de autorrelacionamento 1:N;
- Mapeamento de autorrelacionamento N:N;
- Mapeamento de tipos de relacionamento n -ário (qualquer relacionamento maior que o binário);
- Mapeamento de especialização / generalização.

Figura 20 – Modelo Conceitual.



Fonte: Elaborada pelo autor (2019).

Figura 21 – Modelo Lógico.



Fonte: Elaborada pelo autor (2019).

A tabela USUARIO (Quadro 14) é uma das mais importantes do modelo proposto, pois relaciona-se com diversas tabelas. Dessa forma, a representação da tabela com um exemplo de instância é fundamental para se identificar o domínio dos atributos.

Quadro 14 – Representação da tabela “USUARIO”.

Atributo	Tipo	Domínio do Atributo	Instância
id_usuario	int	-	1
nome	varchar(100)	-	José da Silva
nome_contato	varchar(50)	-	José
dt_nascimento	date	-	16/06/1974
email	varchar(50)	-	jose.silva@email.net
sexo	char(1)	M,F	M
estado_civil	char(1)	S, C, E, V, U	C
escolaridade	char(1)	F, M, G, P, D	D

Fonte: Bernardino (2017).

A próxima tabela, denominada ESTILO_APRENDIZAGEM (Quadro 15), também não sofreu alterações com relação ao modelo anteriormente criado, e foi obtida a partir do módulo Aprendizagem da EAP. No Quadro 16 representa-se a descrição do domínio dos atributos.

Quadro 15 – Representação da tabela “ESTILO APRENDIZAGEM”.

Atributo	Tipo	Domínio do Atributo	Instância
id_usuario	int	-	50
cognitivo	char(1)	A, C, L, M, V	A
aprendizagem_dsi	char(2)	I, S, SI	I
aprendizagem_dvv	char(2)	E, V, VV	E
aprendizagem_dar	char(2)	A, R, AR	A
aprendizagem_dsg	char(2)	G, S, SG	G
acessibilidade	char(2)	A, V, AV, N	A

Fonte: Bernardino (2017).

Quadro 16 – Domínio dos atributos da tabela “ESTILO APRENDIZAGEM”.

Atributo	Domínio do Atributo
cognitivo	A - Auditivo, C - Cinestésico, L - Leitor/Escrita, M - Multimodal, V – Visual
aprendizagem_dsi	I - Intuitivo, S - Sensorial, SI - Sensorial/Intuitivo
aprendizagem_dvv	E - Verbal, V - Visual, VV - Visual/Verbal
aprendizagem_dar	A - Ativo, R - Reflexivo, AR - Ativo/Reflexivo
aprendizagem_dsg	G - Global, S - Sequencial e SG - Sequencial/Global
acessibilidade	A - Auditiva, V - Visual, AV - Auditiva/Visual, N - Nenhuma

Fonte: Bernardino (2017).

As áreas de interesse foram representadas pelas tabelas “AREA_INTERESSE” e “AREA_INTERESSE_USUARIO”, conforme os Quadros 17 e 18.

Quadro 17 – Representação da tabela “AREA INTERESSE”.

Atributo	Tipo	Domínio do Atributo	Instância
id area	int	-	1
descricao	varchar(80)	-	Gestão de Projetos

Fonte: Bernardino (2017).

Quadro 18 – Representação da tabela “AREA INTERESSE USUÁRIO”.

Atributo	Tipo	Domínio do Atributo	Instância
id usuario	int	-	20
id area	int	-	5

Fonte: Bernardino (2017).

A tabela “PERSONALIDADE” representa um dos objetos adicionados ao modelo, com o intuito de utilizá-la na aplicação de técnicas de LA. A representação desta tabela consta no Quadro 19.

Quadro 19 – Representação da tabela “PERSONALIDADE”.

Atributo	Tipo	Domínio do Atributo	Instância
id personalidade	int	-	2
reflexivo	boolean	0 ou 1	1
sensitivo	boolean	0 ou 1	0
julgador	boolean	0 ou 1	0
perceptivo	boolean	0 ou 1	0
sensível	boolean	0 ou 1	1
intuitivo	boolean	0 ou 1	0
extrovertido	boolean	0 ou 1	0
introvertido	boolean	0 ou 1	0

Fonte: Elaborado pelo autor (2019).

No viés educacional da EAP, temos a tabela CURSOS, que em virtude do relacionamento N:N com a tabela USUARIO, gerou a tabela CURSO_USUARIO, que armazena todos os usuários que possuem matrícula no curso. Além destas, também foram criadas as tabelas DISCIPLINA, CURSO, DISCIPLINA_USUARIO. Esta última contém as informações dos usuários relacionados às disciplinas no curso.

A última tabela do módulo Educacional é ATV_EXTRACURRICULAR (Quadro 20), responsável por armazenar as atividades extracurriculares de um determinado curso. A tabela ATV_EXTRACURRICULAR_USUARIO, foi a tabela resultante do relacionamento entre as tabelas USUARIO e ATV_EXTRACURRICULAR, sendo responsável por armazenar as informações das atividades extracurriculares de um determinado usuário (Quadro 21).

Quadro 20 – Representação da tabela “ATV EXTRACURRICULAR”.

Atributo	Tipo	Domínio do Atributo	Instância
id_atv_extracurricular	int	-	1
id_curso	int	-	200
tipo_atv	varchar(60)	Hobbies; Esportes; Idiomas; Olimpíadas do conhecimento; Voluntariado; Palestra; Simpósio; Curso; Seminário.	Esportes
descrição	varchar(60)	-	Futebol estudantil

Fonte: Elaborado pelo autor (2019).

Quadro 21 – Representação da tabela “ATV EXTRACURRICULAR_USUARIO”.

Atributo	Tipo	Domínio do Atributo	Instância
id_atv_extracurricular	int(11)	-	1002
id_usuario	int(11)	-	25764
data_atv	date	-	19/04/2019
tipo_atv	varchar(45)	-	Simpósio
descrição	varchar(45)	-	4º Simpósio de Novas Tecnologias

Fonte: Elaborado pelo autor (2019).

O módulo Histórico da EAP é formado pelas informações concernentes aos dados de acesso, os *links* acessados durante a interação com o ambiente de ensino, assim como as atividades acessadas e as notas em cada atividade, o tempo de acesso, a classificação, o dispositivo utilizado, e por último, as informações das postagens sobre as atividades realizadas nas redes sociais.

A tabela ATIVIDADE_HISTORICO é onde são armazenadas as informações do histórico de atividades de um determinado usuário, e suas notas em cada uma das atividades avaliativas. A tabela ACESSO_HISTORICO coleta os dados do dispositivo e do navegador utilizado para realizar o ambiente virtual. A tabela ATIVIDADE, por sua vez, contém as atividades de uma determinada disciplina. Tem como relacionamentos as tabelas ATIVIDADE_HISTORICO, DISCIPLINA e ATIVIDADE_ANEXO. Esta última tabela armazena os anexos de cada atividade, para uma determinada disciplina.

O próximo capítulo trata dos cenários de aplicação e discute os resultados obtidos com a utilização das técnicas de LA, a fim de demonstrar a viabilidade do modelo proposto.

4 APLICAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS

Neste capítulo é estabelecido o cenário de aplicação do modelo, as análises iniciais para entendimento do cenário e, a partir disso, a aplicação de métodos de LA para a geração de análises mais abrangentes que possam auxiliar na tomada de decisão no contexto educacional.

4.1 CENÁRIO DE APLICAÇÃO

Com o objetivo de se realizar análises gerais que possam auxiliar no entendimento e potencialidade do modelo, foram utilizados os dados de uma instituição de ensino que possui atividades ativas em um ambiente virtual de aprendizagem. Dentre os cursos cujos dados foram importados, encontram-se curso técnicos e cursos livres.

Os dados referentes aos alunos, cursos, disciplinas, áreas de interesse e acessibilidade, foram importados da plataforma. Os dados referentes a personalidade, ao estilo cognitivo e aprendizagem, foram obtidos de forma explícita, não sendo, portanto, obtidos da plataforma de ensino a distância.

Infelizmente não consta no modelo os dados de data-hora de término de uma determinada atividade, mas apenas a data-hora inicial. Com isso, informações referentes ao tempo que o aluno ficou conectado a plataforma não estão disponíveis.

4.2 ANÁLISE GERAL DOS DADOS

A partir do modelo proposto e da base de dados, várias informações podem ser obtidas. Dentre elas, a descrição de como os dados estão registrados, através de contagens, bem como avaliações estatísticas básicas. Estas primeiras avaliações permitem um entendimento inicial do potencial do modelo. Para esta etapa, foram realizadas consultas diretamente no banco de dados, através da linguagem SQL (*Structured Query Language*).

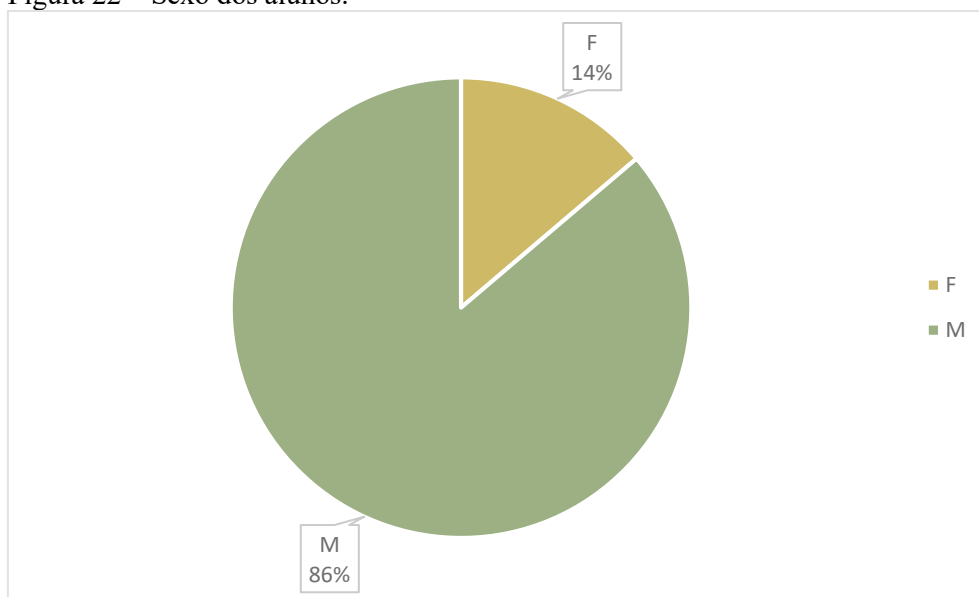
O banco de dados possui 348 registros, que correspondem ao número de alunos matriculados nos cursos. Ao todo, são 59 cursos cadastrados, entre cursos de graduação, extensão e técnico. O número total de disciplinas, considerando todos os cursos é de 135. Os eventos de acesso ao sistema também são contabilizados e estão cadastrados 418.543 eventos de *login*, sendo possível verificar: a data-hora de acesso; o tipo de dispositivo utilizado, por exemplo, “Android®” ou o “iPhone®”; o navegador e qual sistema operacional o usuário estava utilizando no momento do acesso.

Vinte e três áreas de interesse estão registradas na base de dados. As atividades que foram executadas em cada disciplina contêm 1004 registros. Cada atividade pode ter ainda um ou vários anexos, totalizando 1310 registros. Existe ainda, na tabela de histórico das atividades, 421.887 registros, que representam o registro das atividades de cada usuário no sistema, que pode ter acessado um conteúdo, um serviço, um recado, o calendário, um fórum, uma anotação virtual ou ter realizado uma avaliação no LMS. As atividades avaliativas possuem nota registrada no banco de dados.

Como atividades extracurriculares, foram recuperadas as informações de 58 registros, com as respectivas cargas horárias.

Dos alunos cadastrados na plataforma a maioria é do sexo masculino (Figura 22), representando 86,21% de todos os alunos, enquanto o sexo feminino representa 13,79%.

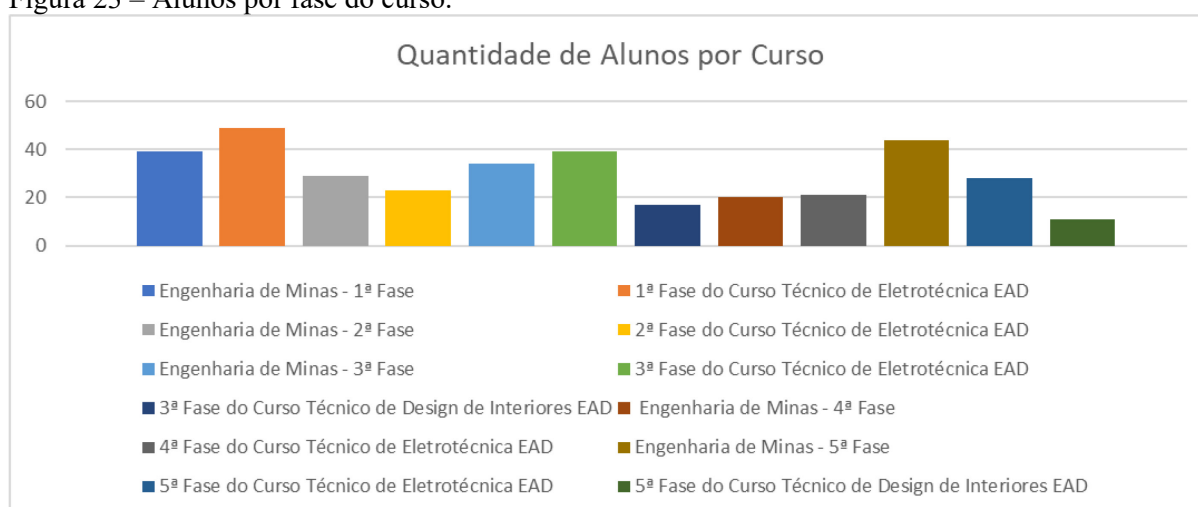
Figura 22 – Sexo dos alunos.



Fonte: Elaborada pelo autor (2019).

Apesar de 59 cursos estarem cadastrados na plataforma de ensino, somente 12 possuem alunos matriculados. Os cursos foram divididos em fases no ambiente EaD (Figura 23).

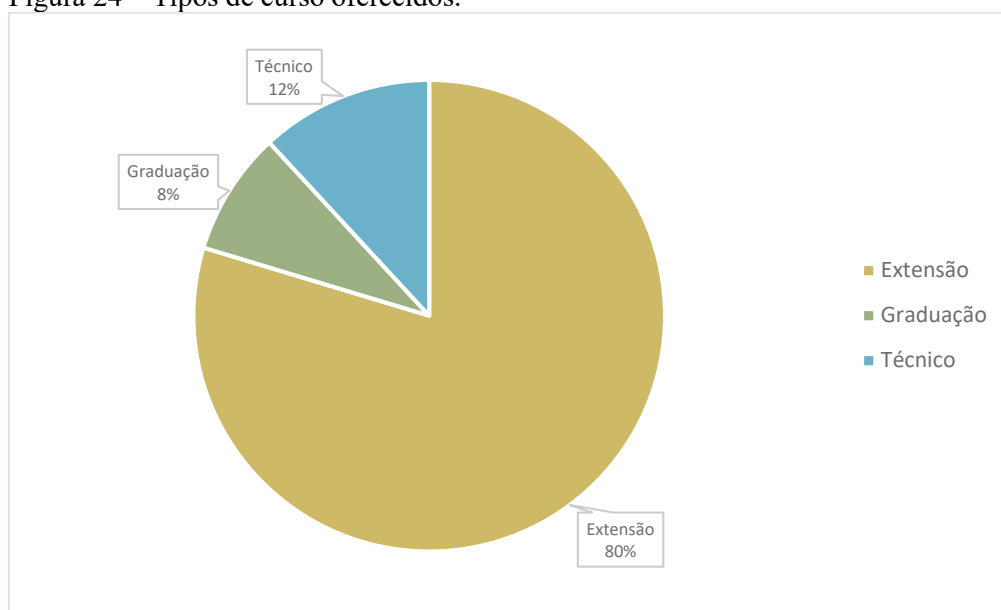
Figura 23 – Alunos por fase do curso.



Fonte: Elaborada pelo autor (2019).

Dentre os cursos oferecidos através da plataforma (Figura 24), a grande maioria trata-se de cursos de extensão (80%), sendo que os cursos de graduação respondem somente por 8% dos cursos disponíveis, enquanto os cursos técnicos respondem por 12%.

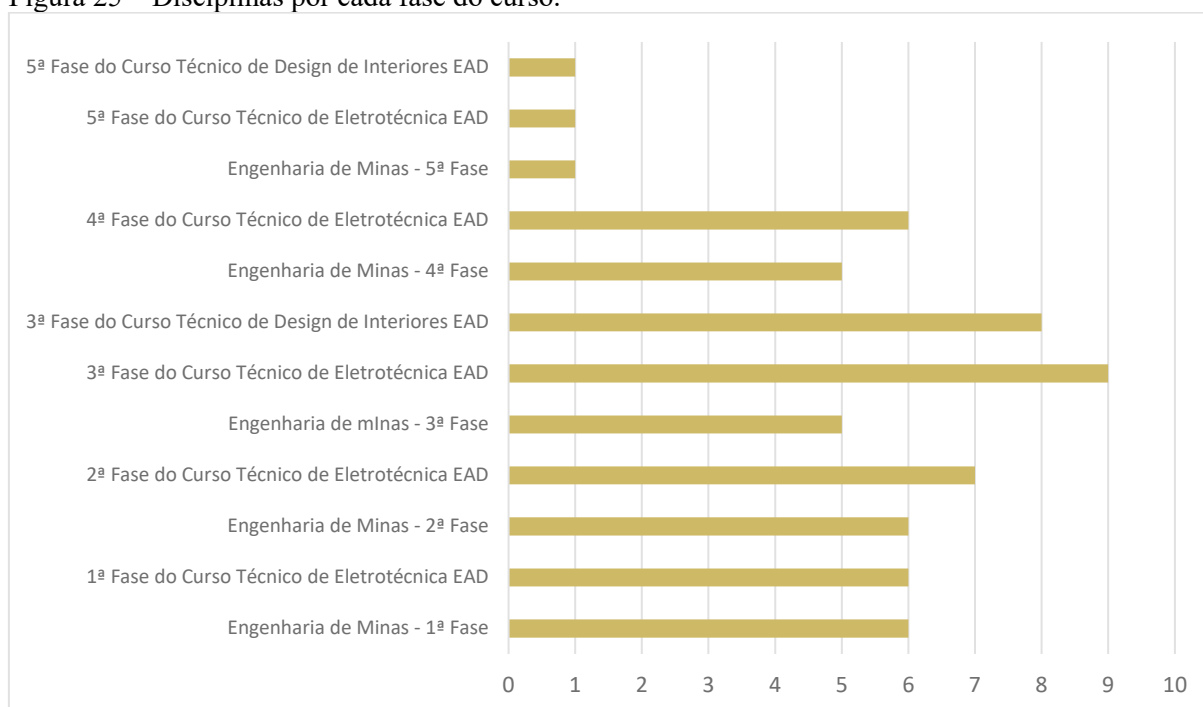
Figura 24 – Tipos de curso oferecidos.



Fonte: Elaborada pelo autor (2019).

Cada curso pode conter uma ou várias disciplinas. Considerando somente os cursos nos quais há alunos matriculados, chegamos ao gráfico da Figura 25, que indica a quantidade de disciplinas por cada fase do curso.

Figura 25 – Disciplinas por cada fase do curso.



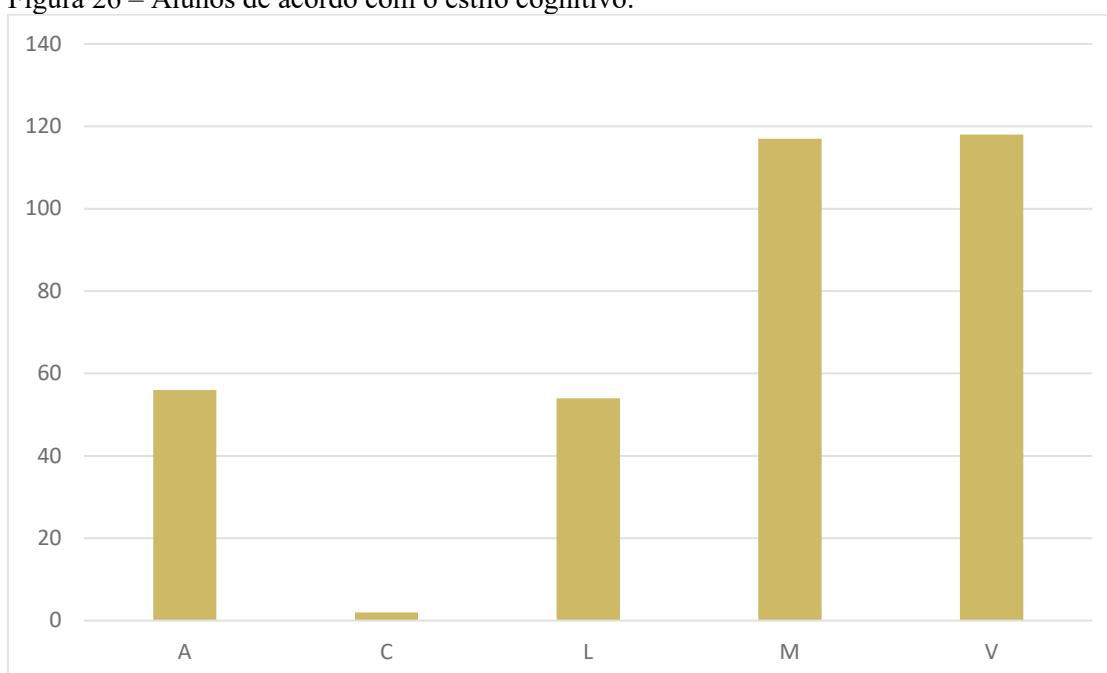
Fonte: Elaborada pelo autor (2019).

A carga horária é diferente para cada um dos doze cursos cadastrados e, considerando apenas os cursos que possuem alunos matriculados, nove possuem a carga horária de 360 horas, sendo que somente o curso de Engenharia de Minas – 2ª fase, possui 456 horas aula e a 2ª fase do curso Técnico de Eletrotécnica EaD possui 155 horas aula.

Considerando 1.553 registros de usuários cadastrados em disciplinas, temos uma média de 6 disciplinas cursadas por cada aluno. A 3ª fase do curso técnico de eletrotécnica é a que apresenta um número maior de disciplinas (9), enquanto as 5ª fases dos cursos técnicos de Design de Interiores, Eletrotécnica e Engenharia de Minas apresentam somente uma disciplina. Na média, cada curso EaD apresenta a quantidade de 6 disciplinas.

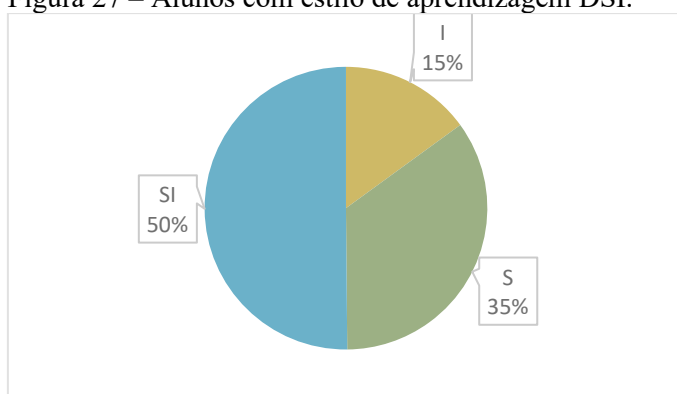
Os estilos cognitivos de cada aluno, de acordo com a metodologia utilizada, oriundos de uma coleta explícita, resultaram nas Figuras 26 a 30. Dos 348 alunos ativos na plataforma, 56 foram classificados como auditivos, 2 como cinestésicos, 54 como leitor/escrita, 117 como multimodais e 118 alunos foram classificados como visuais.

Figura 26 – Alunos de acordo com o estilo cognitivo.



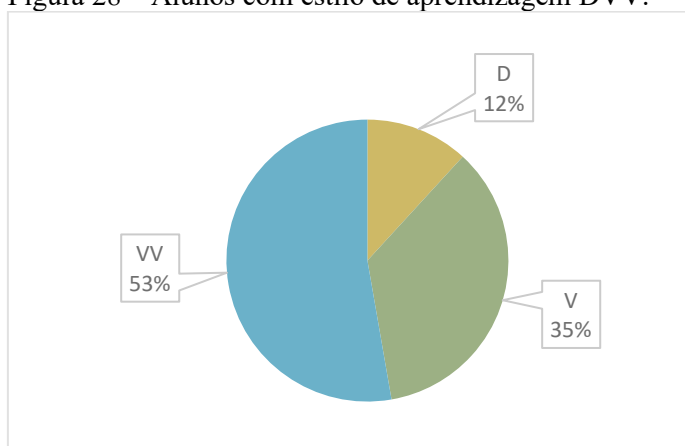
Fonte: Elaborada pelo autor (2019).

Figura 27 – Alunos com estilo de aprendizagem DSI.



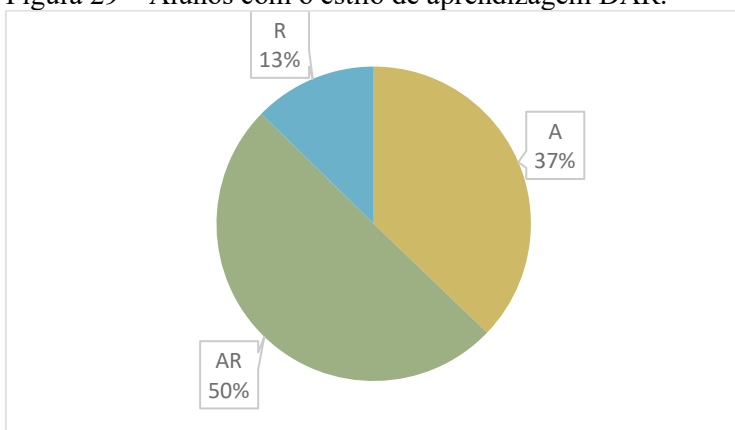
Fonte: Elaborada pelo autor (2019).

Figura 28 – Alunos com estilo de aprendizagem DVV.



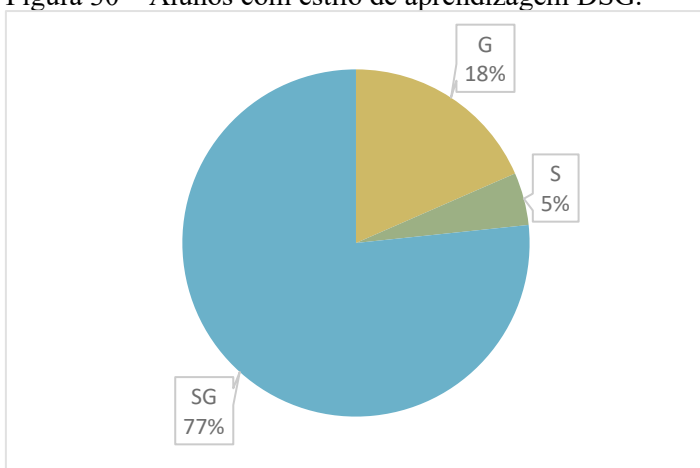
Fonte: Elaborada pelo autor (2019).

Figura 29 – Alunos com o estilo de aprendizagem DAR.



Fonte: Elaborada pelo autor (2019).

Figura 30 – Alunos com estilo de aprendizagem DSG.



Fonte: Elaborada pelo autor (2019).

Com relação aos estilos cognitivos, percebe-se o domínio dos alunos visuais e multimodais. O tipo cinestésico, para este conjunto de dados, foi o que apresentou o menor valor (apenas 2 alunos).

No estilo de aprendizagem_dsi, alunos sensoriais/intuitivos correspondem a exatamente 50%, do total, sendo que os alunos sensoriais correspondem a 35% e alunos intuitivos representam 15%.

No estilo de aprendizagem_dvv, a maioria dos alunos são visuais/verbais (com 53%), seguido por alunos visuais (35%) e alunos verbais (12%).

Já no estilo aprendizagem_dar, ativo/reflexivo resultou no valor de 50%, seguido por ativo, com 37% e finalmente o tipo reflexivo, com apenas 13% dos alunos.

Por fim, no estilo aprendizagem_dsg, 77% dos alunos pesquisados pertencem ao estilo sequencial/global, enquanto 18% são globais e 5% são sequenciais.

Com relação a personalidade do aluno, os dados foram gerados com valores binários, onde o valor 1 representa a existência da personalidade e o valor 0 representa a ausência da personalidade (Figura 31). Como as personalidades não são excludentes entre si, um determinado aluno pode apresentar qualquer uma das possibilidades.

Figura 31 – Exemplo de preenchimento para os dados de “PERSONALIDADE”.

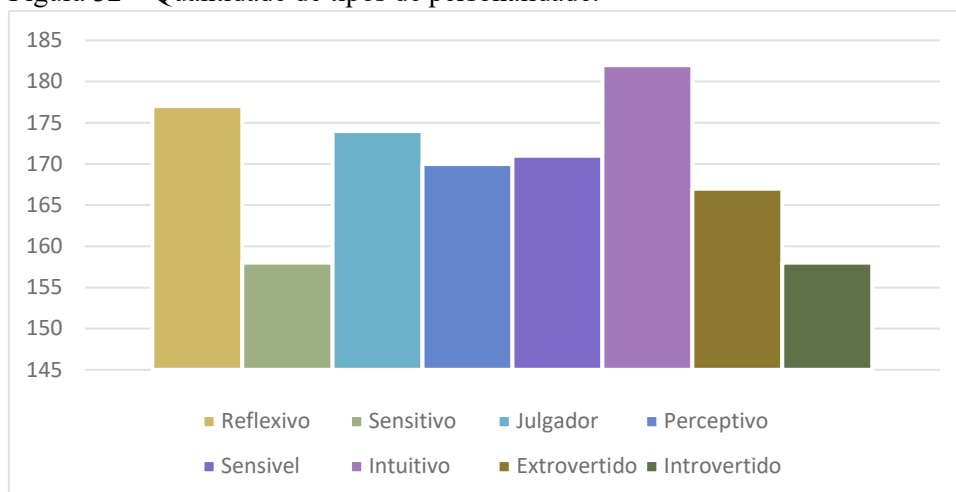
	idpersonalidade	id_usuario	reflexivo	sensitivo	julgador	perceptivo	sensível	intuitivo	extrovertido	introvertido
▶	2	2109	0	1	1	0	1	0	0	1
	3	2255	0	1	0	0	1	0	0	1
	4	2836	1	1	0	0	1	1	1	0
	5	3412	1	1	0	1	1	0	0	1
	6	23057	0	0	1	1	1	0	0	1
	7	24621	0	1	1	1	1	1	0	1
	8	26358	1	0	0	0	0	1	1	0
	9	26753	1	1	1	0	1	0	0	1
	10	27857	1	1	0	1	0	0	1	0
	11	28002	1	0	0	0	1	1	0	1
	12	29098	0	0	1	0	0	1	1	0

Fonte: Elaborada pelo autor (2019).

Dos 348 alunos constantes no banco de dados, 39 alunos (11,2%) possuem de 1 a 2 personalidades; 83 alunos (23,8%) possuem 3 personalidades; 116 alunos (33,3%) possuem 4 personalidades cadastradas; 82 alunos (23,5%) possuem 5 personalidades e 26 alunos (7,4%) possuem 6 ou mais. Dois alunos (0,6%) não declararam nenhum dos tipos de personalidade apresentados.

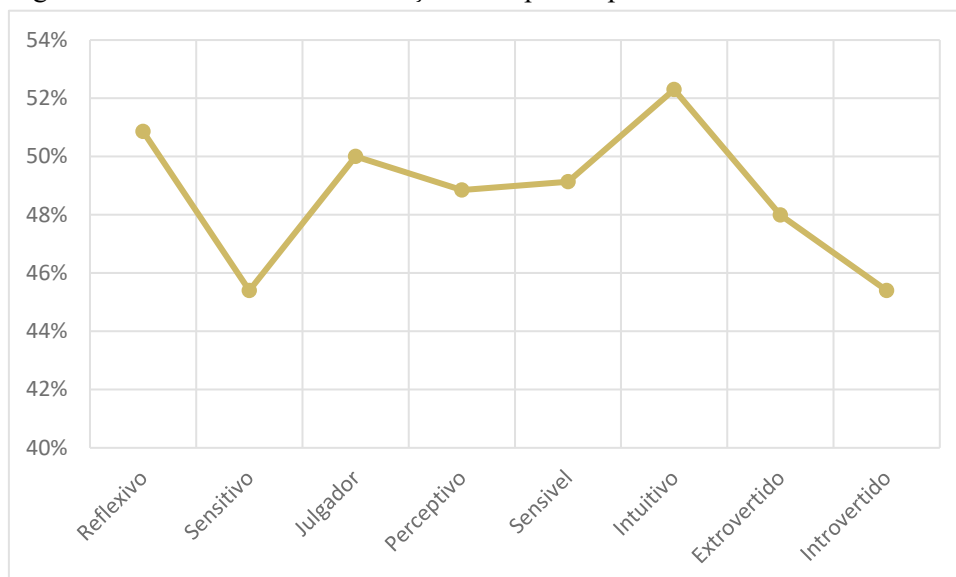
No gráfico da Figura 32, pode-se observar a quantidade de vezes que cada tipo de personalidade aparece no perfil do aluno. Observa-se que o tipo de personalidade “intuitivo” é o mais referido pelos alunos, enquanto que “introvertido” e “sensitivo” são os tipos menos referidos. Já no gráfico da Figura 33, são mostrados os percentuais de cada tipo de personalidade. Assim, corroborando o gráfico anterior, 52% dos alunos possuem a personalidade “intuitivo”, enquanto 45% dos alunos possuem tanto a personalidade “sensitivo”, como “introvertido”.

Figura 32 – Quantidade de tipos de personalidade.



Fonte: Elaborada pelo autor (2019).

Figura 33 – Gráfico de contabilização dos tipos de personalidade.



Fonte: Elaborada pelo autor (2019).

4.3 ANÁLISE BASEADA EM AGRUPAMENTO

Para esta primeira avaliação do modelo utilizou-se o conceito de análise de agrupamento, sendo inicialmente necessária a geração de uma matriz de similaridades entre os alunos considerando os estilos de aprendizagem.

Para a obtenção da similaridade entre os alunos, foi utilizado o cálculo da similaridade do cosseno, uma das técnicas mais utilizadas e eficazes para essa tarefa. A fórmula utilizada para o cálculo é a apresentada na Equação 2 a seguir.

$$\cos \theta = \frac{\sum_{i=1}^n (t_i \times q_i)}{\sqrt{\sum_{k=1}^n (t_k)^2} \times \sqrt{\sum_{j=1}^n (q_j)^2}} \quad (2)$$

Para se adequar os dados em um formato propício para os cálculos, algumas alterações nos dados foram necessárias, especialmente a transformação em valores binários, 0s ou 1s.

A primeira intervenção foi realizada na tabela de estilo de aprendizagem. Neste caso, foi criada uma planilha binária contendo todos os atributos possíveis de cada campo (Figura 34). Por exemplo, se o estilo cognitivo do aluno foi definido como sendo “L” (leitura/escrita, com o domínio do atributo sendo “A”, “C”, “L”, “M”, “V” no banco de dados), este recebe o valor 1, enquanto todos os demais recebem o valor “0”. De maneira similar, esta estratégia foi também aplicada ao campo aprendizagem_dsi. Considerando o domínio do campo como I - Intuitivo, S - Sensorial, SI - Sensorial/Intuitivo, caso o aluno seja sensorial, este atributo recebe o valor “1” e todos os outros o valor “0”.

Desta forma, seguindo este raciocínio, todos os demais campos foram preenchidos.

A mesma conduta foi adotada para as informações sobre personalidade. Entretanto, neste caso, os valores não são exclusivos, já que um aluno pode apresentar um ou vários traços de personalidade (Figura 35). Quando um aluno apresenta determinado traço de personalidade este recebe o valor “1” e quando não possui, recebe o valor “0”. Por fim, os dados referentes a personalidade foram agregados a planilha de estilo de aprendizado, resultando em uma única planilha com todos os dados.

Figura 34 – Planilha Estilo de Aprendizagem.

ESTILO APRENDIZAGEM																							
id_usuario	cognitivo					aprendizagem_dsi				aprendizagem_dvv			aprendizagem_dar			aprendizagem_dsg			acessibilidade				
	A	C	L	M	V	I	S	SI	E	V	VV	A	R	AR	G	S	SG	A	V	AV	N		
2109	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
2255	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	1		
2836	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	
3412	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
23057	0	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
24621	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1		
26358	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1		
26753	0	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
27857	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
28002	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1		
29098	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
29571	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
30589	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1		
31739	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	1	0	0	0	0	1		
32060	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1		
32138	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	1		
33556	0	0	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1		
35506	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	1		
36012	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1		
36487	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	1		
36531	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	1		
36532	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1		
36873	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1		
37350	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1		
37816	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	1		
38229	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1		
38379	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1		

Fonte: Elaborada pelo autor (2019).

Figura 35 – Planilha de tipo de personalidade.

PERSONALIDADE							
reflexivo	sensitivo	jugador	perceptivo	sensível	intuitivo	extrovertido	introverso
0	1	1	0	1	0	0	1
0	1	0	0	1	0	0	1
1	1	0	0	1	1	1	0
1	1	0	1	1	0	0	1
0	0	1	1	1	0	0	1
0	1	1	1	1	1	1	0
1	0	0	0	0	1	1	0
1	1	1	0	1	0	0	1
1	1	0	1	0	0	1	0
1	0	0	0	1	1	0	1
0	0	1	0	0	1	1	0
1	1	1	1	0	0	0	1

Fonte: Elaborada pelo autor (2019).

Em seguida, foi criada uma tabela chamada “*vectors*” com os vetores para cada aluno da planilha (Figura 36). O primeiro campo desta tabela, representa o Id do usuário (ou aluno). O segundo campo, é um identificador criado que representa cada uma das instâncias do mesmo aluno, sendo, portanto, um campo incremental. O terceiro campo representa os valores de cada atributo da tabela. Neste sentido, cada aluno possuirá várias linhas representando cada dimensão da planilha e a composição das linhas formam o vetor do respectivo aluno.

Figura 36 – Estrutura da tabela que representa os vetores dos alunos.

vid	did	dimension
2109	1	0
2109	2	0
2109	3	1
2109	4	0
2109	5	0
2109	6	0
2109	7	1

Fonte: Elaborada pelo autor (2019).

Uma vez pronto o conteúdo dos vetores, foi executada a criação de uma tabela com uma consulta associada para criar a tabela de distâncias (Quadro 22). Esta tabela, que representa a matriz de distâncias, foi gerada com três campos: o primeiro é o identificador do aluno a ser comparado (vid1); o segundo é o identificador do segundo aluno ao qual o primeiro aluno está relacionado (vid2); e o terceiro campo é o valor da similaridade do cosseno entre os alunos (vid1 e vid2) (Figura 37).

Quadro 22 – Instrução utilizada para gerar a tabela de similaridades.

```
create table distances as
select v1.vid as vid1,v2. vid as vid2, SUM(v1.dimension*v2. dimension)/
(SQRT(SUM(v1.dimension*v1.dimension))*SQRT(SUM(v2.dimension*v2.dimension))) as
cosine from vectors v1 inner join vectors v2 on v1.did=v2.did and v1.vid<>v2.vid
where v1.vid<v2.vid
group by v1.vid,v2.vid
order by v1.vid
```

Fonte: Elaborado pelo autor (2019).

Figura 37 – Tabela “distances”.

	vid1	vid2	cosine
▶	2109	2255	0.7378647873726218
	2109	2836	0.5720775535473553
	2109	3412	0.6674238124719146
	2109	23057	0.7999999999999998
	2109	24621	0.7302967433402214
	2109	26358	0.6324555320336759
	2109	26753	0.7627700713964739
	2109	27857	0.5999999999999999
	2109	28002	0.5999999999999999
	2109	29098	0.5270462766947299
	2109	29571	0.6674238124719146

Fonte: Elaborada pelo autor (2019).

De modo geral, o cálculo da similaridade foi efetuado entre um aluno e todos os demais, e assim sucessivamente, de modo que a tabela “distances” contenha todas as possibilidades distintas de similaridade, o que resultou em 59.685 registros.

O próximo passo consistiu em importar os dados para o *software Gephi*®, a fim de analisar o grafo gerado pelos dados e, a partir disso, obter informações sobre os nodos (alunos) mais relevantes e os grupos de alunos com maior similaridade. Esses dados, uma vez analisados, permitem estabelecer inferências a respeito das similaridades entre os alunos considerando os estilos de aprendizado e o tipo de personalidade, e juntos com as outras informações constantes no banco de dados, pode ser expandido para análises diversas.

A importação para o *Gephi*® foi realizada estabelecendo-se como parâmetro uma similaridade entre os alunos maior do que 0,9 (considerando um intervalo entre 0 e 1). Em testes realizados, valores inferiores a 0,9 produziram redes com muitos nodos e altamente conectados, não permitindo análises conclusivas sobre o relacionamento entre os alunos. Na importação dos dados foi definida uma consulta para este cenário, conforme Quadros 23 e 24:

Quadro 23 – Comando SQL para importação dos nós.

```
select vid as id from (  
  select distinct vid1 as vid from distances where cosine > 0.9 union  
  select distinct vid2 as vid from distances where cosine > 0.9) as res
```

Fonte: Elaborado pelo autor (2019).

Quadro 24 – Comando SQL para importação das arestas.

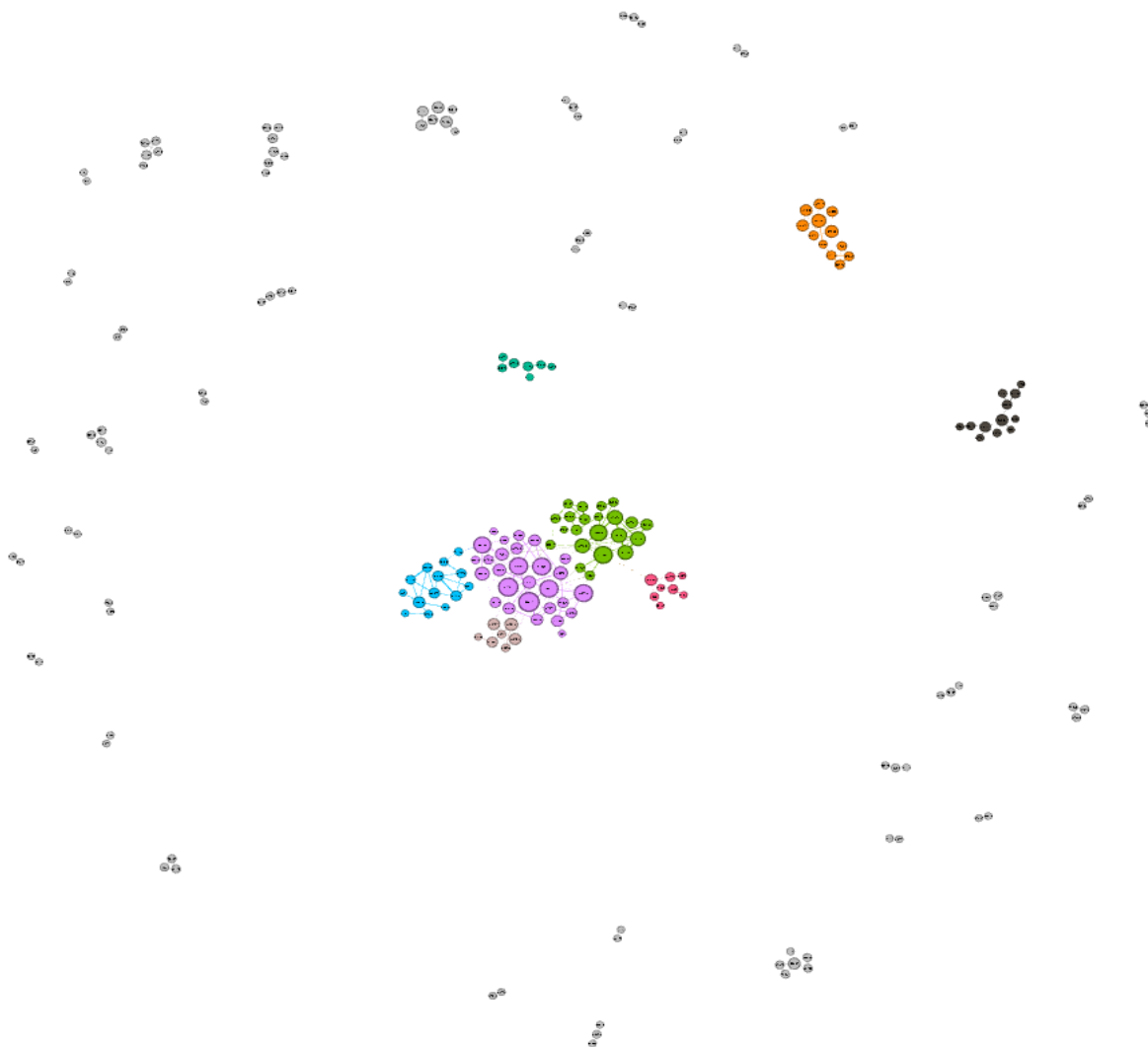
```
select vid1 as source, vid2 as target, cosine as weight from distances where cosine > 0.9
```

Fonte: Elaborado pelo autor (2019).

Após a importação, para realizar a distribuição dos nós evitando sobreposições foi utilizado o algoritmo Force Atlas 2. Este algoritmo permite que o gráfico seja reorganizado considerando conceitos de física, como a atração e a repulsão, produzindo distâncias proporcionais e organizadas entre os nodos. Possui uma velocidade única de convergência adaptativa que permite maior eficiência na convergência.

O componente principal da rede gerada (Figura 38), que representa todos os nodos conectados, possui mais de 39 componentes. O diâmetro da rede ficou em 15 elementos (distância entre os dois nodos mais distantes). Também foram gerados grupos esparsos, que possuem similaridade parecida, mas sem ligação ao grupo central. A Figura 39 apresenta toda a rede criada, incluindo os nós não conectados, formando componentes menores.

Figura 39 – Grafo de similaridade completo.



Fonte: Elaborada pelo autor (2019).

Cada conjunto de cores representa um grupo de alunos que possuem elevada similaridade. A constituição desses grupos é obtida através do algoritmo de modularidade descrito na seção 2.5.4.3. Já o tamanho de um nó em particular é determinado pelo seu grau, ou seja, a quantidade de conexões diretas com outros nós. Quanto maior o tamanho do nó maior é a sua relevância.

Sendo assim, a rede gerada permite localizar os alunos mais relevantes em cada grupo (*cluster*) por estarem mais conectados a outros alunos. No intuito de analisar sistematicamente os alunos mais importantes na rede em cada grupo, foram selecionados cinco outros alunos que possuem as maiores similaridades, considerando os vetores de estilo de aprendizado e tipo de personalidade (Tabela 2).

Tabela 2 – Definição de alunos para avaliação estatística.

Grupo	Cor	Identificador do Aluno Chave	Alunos mais próximos
1	Roxa	47027	- 47071, 47427, 47053, 47066, 46938
2	Azul	46574	- 47020, 46838, 45969, 47748, 47896
3	Verde	47494	- 46590, 47554, 45459, 45388, 47064
4	Cinza	46736	- 45387, 45315, 45329, 45484, 43583
5	Vermelho	45448	- 46964, 46970, 45269, 45289, 47106

Fonte: Elaborada pelo autor (2019).

A partir das informações da rede, considerando os nós mais relevantes de cada grupo, realizou-se um conjunto de análises. Essas análises tiveram o objetivo de verificar comportamentos similares em relação a diversos aspectos monitorados através da plataforma de aprendizagem *online*. A primeira avaliação foi realizada em relação à regularidade de acesso durante todo o curso, resultando nas Tabelas 3 a 7. Foram contabilizando todos os acessos, mesmo aqueles nos quais não foram realizadas nenhuma atividade.

Tabela 3 – Acessos do grupo 1.

Id Aluno	Qtde
47027	886
47071	864
47427	2.176
47053	1.501
47066	1.153
46938	826

Fonte: Elaborada pelo autor (2019).

Tabela 6 – Acessos do grupo 4.

Id Aluno	Qtde
46736	256
45387	63
45315	58
45329	10
45484	2
43583	601

Fonte: Elaborada pelo autor (2019).

Tabela 4 – Acessos do grupo 2.

Id Aluno	Qtde
46574	2341
47020	606
46838	33
45969	862
47748	2976
47896	931

Fonte: Elaborada pelo autor (2019).

Tabela 7 – Acessos do grupo 5

Id Aluno	Qtde
45448	5
46964	882
46970	945
45269	0
45289	3
47106	935

Fonte: Elaborada pelo autor (2019).

Tabela 5 – Acessos do grupo 3.

Id Aluno	Qtde
47494	2042
46590	1498
47554	1941
46708	637
45388	10
47064	271

Fonte: Elaborada pelo autor (2019).

Percebe-se que pessoas que têm o perfil semelhante, não necessariamente possuem a mesma frequência de acesso, que pode ser um indicativo de que o fato de o perfil ser semelhante, não tem relação com esta a variável analisada. Percebe-se no grupo 5 que apesar de um dos alunos ainda não ter acessado nenhuma atividade, ainda assim o seu perfil é compatível com os demais integrantes do grupo.

Outra análise realizada refere-se às notas de cada aluno. Neste sentido, foi realizado o cálculo da média geral da nota dos alunos analisados para cada um dos grupos, gerando as Tabelas 8 a 12.

Tabela 8 – Notas médias do grupo 1.

Id Aluno	Valor
47027	8,16
47071	8,11
47427	7,99
47053	8,02
47066	7,87
46938	8,22

Fonte: Elaborada pelo autor (2019).

Tabela 11 – Notas médias do grupo 4.

Id Aluno	Valor
46736	8,01
45387	-
45315	-
45329	-
45484	-
43583	7,84

Fonte: Elaborada pelo autor (2019).

Tabela 9 – Notas médias do grupo 2.

Id Aluno	Valor
46574	8,07
47020	8,23
46838	-
45969	8,09
47748	7,96
47896	7,98

Fonte: Elaborada pelo autor (2019).

Tabela 12 – Notas médias do grupo 5.

Id Aluno	Valor
45448	-
46964	8,02
46970	8,07
45269	-
45289	-
47106	8,10

Fonte: Elaborada pelo autor (2019).

Tabela 10 – Notas médias do grupo 3.

Id Aluno	Valor
47494	7,90
46590	8,14
47554	8,01
46708	7,83
45388	-
47064	8,01

Fonte: Elaborada pelo autor (2019).

Nesta análise, os alunos que possuem perfil semelhante acabaram apresentando o mesmo padrão de notas, com baixíssima disparidade. A diferença da média de nota mais baixa (7,84) para a mais alta (8,23) foi de apenas 0,39 décimos. É preciso ressaltar que é necessária uma análise com um conjunto de dados mais amplo para a definição de uma hipótese que

poderia ser ou não comprovada através da análise proposta. Os alunos que não apresentam notas médias, são aqueles alunos que ainda não executaram nenhuma atividade avaliada no ambiente *online* de aprendizagem.

Outra vertente de análise diz respeito a participação dos alunos nos fóruns propostos nas disciplinas. A socialização das dificuldades encontradas nas atividades propostas, assim como outros comentários, pode servir ao propósito de complementar a análise do perfil do aluno. Assim, nas tabelas 13 a 17, constam a quantidade de participações dos alunos com alta similaridade, nos fóruns das diversas disciplinas dos cursos.

Tabela 13 – Grupo 1.

Id Aluno	Acessos aos fóruns
47027	280
47071	376
47427	674
47053	646
47066	437
46938	286

Fonte: Elaborada pelo autor (2019).

Tabela 14 – Grupo 2.

Id Aluno	Acessos aos fóruns
46574	1180
47020	247
46838	33
45969	421
47748	1417
47896	277

Fonte: Elaborada pelo autor (2019).

Tabela 15 – Grupo 3.

Id Aluno	Acessos aos fóruns
47494	800
46590	797
47554	729
46708	126
45388	-
47064	56

Fonte: Elaborada pelo autor (2019).

Tabela 16 – Grupo 4.

Id Aluno	Acessos aos fóruns
46736	60
45387	-
45315	-
45329	-
45484	-
43583	120

Fonte: Elaborada pelo autor (2019).

Tabela 17 – Grupo 5.

Id Aluno	Acessos aos fóruns
45448	-
46964	322
46970	104
45269	-
45289	-
47106	252

Fonte: Elaborada pelo autor (2019).

Os grupos 1, 2 e 3 apresentam a maior quantidade de alunos que efetivamente acessaram os fóruns. Entretanto, o simples acesso ao fórum não representa participação efetiva na atividade. O grupo 2 abriga os dois alunos com mais números de acesso: 1.414 e 1.180. Avaliando o número de acessos do aluno que ficou na terceira posição, do grupo 3 (800), vemos

uma diferença de 617 para o primeiro colocado, representando um percentual de 43,42%. Desta forma, não há como formular uma hipótese de que alunos com perfis semelhantes apresentarão comportamento semelhante também nos fóruns, já que a quantidade de acesso pode sofrer diversas interferências não avaliadas, como indisponibilidade da internet ou falta de comprometimento do aluno com as atividades propostas.

Infelizmente, o banco de dados não possui a informação da data-hora que o aluno saiu do ambiente, apenas a data-hora inicial, impossibilitando análises temporais, como o tempo que o aluno gastou em determinada atividade.

Com relação às áreas de interesse, o banco de dados não contém uma ampla gama de informações, pois apenas 23 áreas foram cadastradas. Além disso, nem todos os alunos possuem a informação gravada. No grupo 1, apenas o aluno 47427 possui áreas de interesse cadastradas (Caracterização Tecnológica de Matérias Primas Mineraias, Gestão Ambiental na Mineração, Higiene e Segurança na Mineração, Tratamento de Minérios e Gestão de Projetos). No grupo 2, nenhum aluno possui esta informação. No grupo 3, apenas os alunos 47494 e 47554 forneceram esta informação. Neste grupo, são comuns aos dois alunos as áreas de interesse Caracterização Tecnológica de Matérias Primas Mineraias, Gestão Ambiental na Mineração e Gestão de Projetos. No grupo 4 e 5, nenhum aluno apresenta os dados.

4.4 PRIMEIRO CENÁRIO DE RECOMENDAÇÃO

Como complemento às análises baseadas na tarefa de agrupamento realizadas na seção anterior, nesta, e na próxima seção, serão apresentados cenários para LA que envolvam recomendações aos alunos.

A proposta deste primeiro cenário de recomendação consiste na sugestão de atividades para um aluno, de acordo com o seu perfil de personalidade, utilizando para isso, princípios da abordagem de filtragem colaborativa. Assim, as atividades realizadas por outros alunos que possuem a personalidade similar com a do aluno foco da recomendação, serão recomendadas a este aluno.

Para se realizar tal tarefa, definiu-se qual o nível da similaridade desejada para se realizar a recomendação. Assim, o mesmo padrão utilizado na unidade anterior também será utilizado aqui, ou seja, usuários que possuem um nível de similaridade acima de 0,9, o que pode ser entendido como 90%.

A abordagem utilizada, conforme mencionado anteriormente, é a recomendação colaborativa, que basicamente, consiste em recomendar atividades ainda não realizadas por

determinado aluno, considerando as atividades que colegas com perfis compatíveis já executaram.

Foi escolhido um aluno de cada grupo para ser o alvo da recomendação (Tabela 18), juntamente com outros alunos pertencentes ao mesmo grupo e que estão estudando no mesmo curso. As atividades já executadas pelos alunos escolhidos, ainda não realizadas pelo aluno alvo da recomendação serão então recomendadas.

Apesar de alguns dos grupos possuírem alunos de cursos diferentes com alta similaridade de perfil, foram considerados apenas alunos que estão realizando o mesmo curso, já que não faria sentido recomendar atividades de outro curso.

Para que tal análise seja realizada, optou-se pela diferença de conjuntos, concretizada através de consulta, sendo que o resultado gerado foi um grupo de atividades que constam no conjunto de atividades dos alunos relacionados que ainda não foram realizadas pelo aluno que é o alvo da análise.

Tabela 18 – Definição de alunos para recomendação colaborativa.

Grupo	Cor	Id_Aluno Alvo da Recomendação	Alunos com alto grau de similaridade	Id_Curso
1	Roxa	47027	- 46938, 47071, 47066, 47033, 46913	2141
2	Azul	47420	- 47748, 47850, 47853, 47896, 47905	1156
3	Verde	47554	- 47494, 47212	1141
		46977	- 47001, 47750	1156
		47108	- 47064, 47063	2141
4	Cinza	45484	- 45387, 45315	5141
5	Vermelho	47106	- 46970	2156
		45460	- 45289	5141
		45422	- 45269	5175

Fonte: Elaborada pelo autor (2019).

Para se obter as atividades que os alunos com alto grau de similaridade do perfil realizaram no ambiente de ensino, nas quais o aluno alvo da recomendação não realizou, foram utilizadas as consultas elencadas no Quadro 25, Quadro 26 e Quadro 27, substituindo os campos `id_usuario` e `id_curso` conforme a necessidade. A primeira consulta seleciona as atividades já realizadas pelo aluno alvo da recomendação, considerando que as atividades são diferentes entre si. A segunda consulta, elenca todas as atividades feitas pelos outros alunos com alto grau de similaridade considerando o perfil de ambos.

Quadro 25 – Comando SQL – Aluno alvo das recomendações.

```

SELECT DISTINCT atividade.id_atividade, atividade.titulo, disciplina.nome, curso.id_curso,
curso.nome
FROM ((atividade INNER JOIN atividade_historico ON atividade.id_atividade =
atividade_historico.id_atividade) INNER JOIN disciplina ON atividade.id_disciplina =
disciplina.id_disciplina) INNER JOIN curso ON disciplina.id_curso = curso.id_curso
WHERE (((atividade_historico.id_usuario)=47420) AND ((curso.id_curso)=1156));

```

Fonte: Elaborado pelo autor (2019).

Quadro 26 – Comando SQL – Atividades dos outros alunos.

```

SELECT DISTINCT atividade.id_atividade, atividade.titulo, disciplina.nome, curso.id_curso,
curso.nome
FROM ((atividade INNER JOIN atividade_historico ON atividade.id_atividade =
atividade_historico.id_atividade) INNER JOIN disciplina ON atividade.id_disciplina =
disciplina.id_disciplina) INNER JOIN curso ON disciplina.id_curso = curso.id_curso
WHERE (((curso.id_curso)=1156) AND ((atividade_historico.id_usuario) In
(47748,47850,47853,47896,47905)));

```

Fonte: Elaborado pelo autor (2019).

Quadro 27 – Comando SQL – Atividades a serem recomendadas.

```

SELECT DISTINCT [Aluno Alvo_2].id_atividade, [Aluno Alvo_2].titulo, [Aluno
Alvo_2].disciplina.nome, [Aluno Alvo_2].id_curso, [Aluno Alvo_2].curso.nome
FROM [Aluno Alvo_2] LEFT JOIN [Aluno Alvo] ON [Aluno Alvo_2].id_atividade = [Aluno
Alvo].id_atividade

```

Fonte: Elaborado pelo autor (2019).

A terceira consulta retorna os registros da segunda consulta, excluindo aqueles registros que estão na primeira consulta, ou seja, é uma consulta diferencial, de modo que o resultado contenha todos os registros que serão recomendados ao aluno alvo. A recomendação colaborativa consiste basicamente nisto: analisar grupos de pessoas com interesses ou perfis semelhantes e recomendar as atividades que o grupo já realizou para aquelas pessoas que ainda não realizaram estas atividades. Assim, quanto maior o grupo e a semelhança dos perfis, melhor será a probabilidade de a recomendação ser adequada.

Na análise do grupo 1, os alunos pertencentes ao curso de Engenharia de Minas com Id 46938, 47071, 47066, 47033, 46913 realizaram quatorze atividades que o aluno 47027 não o fez (Quadro 28). Dessa forma, essas quatorze atividades podem ser recomendadas para o aluno 47027, nas disciplinas de Topografia I, Perfuração e Desmonte, Prospeção e Máquinas e Equipamentos.

Quadro 28 – Atividades a serem recomendadas ao aluno 47027.

Id_Atv	Título	Disciplina
6434	UNIDADE 3	Topografia I
6439	UNIDADE 4	Topografia I
6564	Desmonte de Rochas com Explosivos - Perfuração, Tipos de Explosivos e Acessórios mais comuns.	Perfuração e Desmonte
6755	UNIDADE 5	Prospecção
6757	UNIDADE 6	Prospecção
6758	UNIDADE 7	Prospecção
6760	UNIDADE 8	Prospecção
48475	Avaliação <i>online</i> - Questão 03	Máquinas e Equipamentos
49548	Fórum temático - Desmonte de rochas com o uso de explosivos	Perfuração e Desmonte

Fonte: Elaborado pelo autor (2019).

O Quadro 29, possui parte dos resultados da consulta para a análise do grupo 2 (alunos da 1ª fase do Curso Técnico de Eletrotécnica EaD). O processo de recomendação executado, informou que o grupo de alunos 47748, 47850, 47853, 47896, 47905 realizou 93 atividades que o aluno com Id 47420 não realizou. Portanto, essas atividades foram recomendadas ao aluno.

Quadro 29 – Algumas atividades recomendadas ao aluno 47420.

id_atv	Título	Disciplina
139	Revisão dos conteúdos	Redação Técnica I
152	Web conferência Cálculo Técnico	Cálculo Técnico I
156	Eletrostática e Eletrodinâmica	Eletricidade I
160	Resolução do Trabalho	Eletricidade I
163	Resolução Trabalho - Parte 2	Eletricidade I
164	Medições Elétricas - Resolução dos Exercícios	Eletricidade I
169	Avaliação <i>Online</i> - Solução	Eletricidade I
186	Vídeo Conferência	Instalações Elétricas I

Fonte: Elaborado pelo autor (2019).

Partindo para o grupo verde, os alunos que possuem alta similaridade e pertencem ao mesmo curso e mesma fase, foram selecionados, resultando em apenas três alunos em cada fase do curso. Então as atividades realizadas por dois deles, que não foram executadas pelo aluno alvo, foram recomendadas, conforme Quadros 30, 31 e 32. Observa-se que este grupo de similaridade possui alunos de três cursos diferentes, sendo que o Quadro 30 retrata os alunos da 1ª fase do curso de Engenharia de Minas, o Quadro 31 os alunos da 1ª fase do Curso Técnico de Eletrotécnica EaD e o Quadro 32, os alunos da 2ª fase de Engenharia de Minas.

Quadro 30 – Atividades recomendadas ao aluno 47554.

id_atv	Título	Disciplina
173	Webconferência	Cartografia Geral
5524	Aula 2 – Unidade 3: MS Word – Tópicos 1 e 2	Informática Básica
6376	Aula 6 – Química	Ciências Aplicadas
52737	Avaliação <i>online</i>	Mecânica de Rochas

Fonte: Elaborado pelo autor (2019).

O Quadro 30 retrata algumas das atividades recomendadas ao aluno com id 46977. Neste caso foram geradas o total de 84 recomendações para o aluno.

Quadro 31 – Algumas atividades recomendadas ao aluno 46977.

id_atv	Título	Disciplina
520	Fórum temático / Avaliação <i>online</i> / Encontro presencial	Segurança e Normas Técnicas em Eletricidade I
2035	Avaliação On-line	Informática Aplicada I
2054	Entrega de Trabalho	Redação Técnica I
2068	Avaliação <i>online</i>	Redação Técnica I
2094	Trabalho <i>online</i> (Unidade 1, 2 e 3)	Cálculo Técnico I
6550	Capacitores	Eletricidade I

Fonte: Elaborado pelo autor (2019).

Para o aluno 47108 foram recomendadas três atividades. Este aluno realizou, segundo a base de dados, o total de 68 atividades e os outros dois alunos juntos, apenas 36. Mas dentro dessas 36, três não foram realizadas pelo aluno alvo.

Quadro 32 – Atividades recomendadas ao aluno 47108.

id_atv	Título	Disciplina
189	Web conferência	Perfuração e Desmonte
6434	Unidade 3	Topografia I
6440	Unidade 4	Topografia I

Fonte: Elaborado pelo autor (2019).

Partindo para os alunos do grupo cinza, cujo aluno alvo é o 45484, não foi possível gerar as recomendações de atividades, pois além deste aluno não ter realizado ainda nenhuma atividade, seus colegas de curso que possuem similaridade de perfil também não executaram nenhuma atividade. Possivelmente, no momento de coleta dos dados, os alunos da 5ª fase do curso de Engenharia de Minas ainda não tinham realizado nenhuma atividade na plataforma de ensino.

No menor dos grupos, o vermelho, somente grupos de dois alunos pertencem ao mesmo curso. Desta forma, a recomendação pode não ser tão efetiva, apesar da similaridade de

perfil. Ainda assim, 15 atividades são recomendadas para o aluno 47106. Algumas delas, constam no Quadro 33.

Quadro 33 – Algumas atividades recomendadas ao aluno 47106.

id_atv	Título	Disciplina
197	Videoconferência Informática Aplicada II - Autocad	Informática Aplicada II
5553	Unidade 3 - Tópico 5 - Medidas de controle do risco elétrico	Segurança e Normas Técnicas em Eletricidade II
5556	Unidade 3 - Tópico 7 - Documentação de instalações elétricas	Segurança e Normas Técnicas em Eletricidade II
6056	Detecção de Incêndio	Instalações Elétricas II
6325	Carta Comercial	Redação Técnica II
6326	Ofício	Redação Técnica II
6327	Circular	Redação Técnica II
6328	Currículo	Redação Técnica II
6740	Exercícios Propostos	Eletricidade II
7080	Circuitos Indutivos	Eletricidade II
7104	Aula 5 - Unidade 2: Tópicos 5 a 8	Informática Aplicada II

Fonte: Elaborado pelo autor (2019).

Assim como no grupo cinza, os alunos selecionados do grupo vermelho, da 5ª fase do curso de Engenharia de Minas (curso id 5141) e os alunos da 5ª Fase do Curso Técnico de Design de Interiores EaD (curso 5175), não realizaram nenhuma atividade no ambiente virtual de aprendizagem, inviabilizando a geração de recomendações baseadas na filtragem colaborativa. Porém, observa-se a potencialidade do modelo proposto no intuito de recomendar atividades semelhantes a alunos que possuem perfis semelhantes, aumentando possivelmente, a efetividade de ações de avaliação de aprendizagem.

4.5 SEGUNDO CENÁRIO DE RECOMENDACÃO

Diferentemente do primeiro cenário, este segundo cenário de análise foi realizado no intuito de verificar a similaridade de alunos dentro de um mesmo curso, com o intuito de ofertar conteúdo pertinente e relacionado ao curso.

Neste cenário, uma atividade que não tenha sido realizada por determinado aluno, mas já foi realizada por seu colega de curso que possui um perfil semelhante, será recomendada. Caso todos os alunos devam realizar todas as atividades propostas na plataforma, a recomendação pode ser utilizada para avaliação do ritmo de estudos.

Mas se for implementado a possibilidade de se ofertar conteúdo personalizado, baseado em alguma característica do aluno ou do contexto, tal cenário se torna mais interessante.

A partir disso, uma seleção foi realizada na tabela de similaridades, considerando apenas alunos do mesmo curso. Para isso, foram eleitos os alunos da 1ª Fase do Curso Técnico de Eletrotécnica EaD, código 1156.

Para a criação da segunda tabela de distâncias considerando somente os alunos daquele curso, foi utilizada a instrução SQL constante no Quadro 34, gerando 1.176 possibilidades. Porém, a tabela contém todas as similaridades entre os alunos do curso, sendo necessário então, estabelecer o parâmetro de corte adequado para gerar o grafo e estabelecer o grau de influência de cada aluno. A definição deste parâmetro é empírica, e depende da quantidade de informações que cada valor retorna. Assim sendo, foi estabelecido que seriam considerados para a importação no *Gephi*®, aqueles alunos que apresentassem uma similaridade entre si superior a 0,7 ou 70%, selecionados através das instruções SQL constantes nos Quadros 35 e 36.

Quadro 34 – Comando para criação da segunda tabela de distâncias.

```
create table distances2 as
select v1.vid as vid1, v2.vid as vid2, SUM(v1.dimension*v2. dimension)/
(SQRT(SUM(v1.dimension*v1.dimension))*SQRT(SUM(v2.dimension*v2.dimension))) as cosine
from alunos_curso_1156 v1 inner join alunos_curso_1156 v2 on v1.did=v2.did and v1.vid<>v2.vid
where v1.vid<v2.vid group by v1.vid, v2.vid order by v1.vid
```

Fonte: Elaborado pelo autor (2019).

Quadro 35 – Comando para importação dos nós no *Gephi*®.

```
select vid as id from (select distinct vid1 as vid from distance2 where cosine > 0.7 union
select distinct vid2 as vid from distance2 where cosine > 0.7) as res
```

Fonte: O autor.

Quadro 36 – Comando para importação das arestas no *Gephi*®.

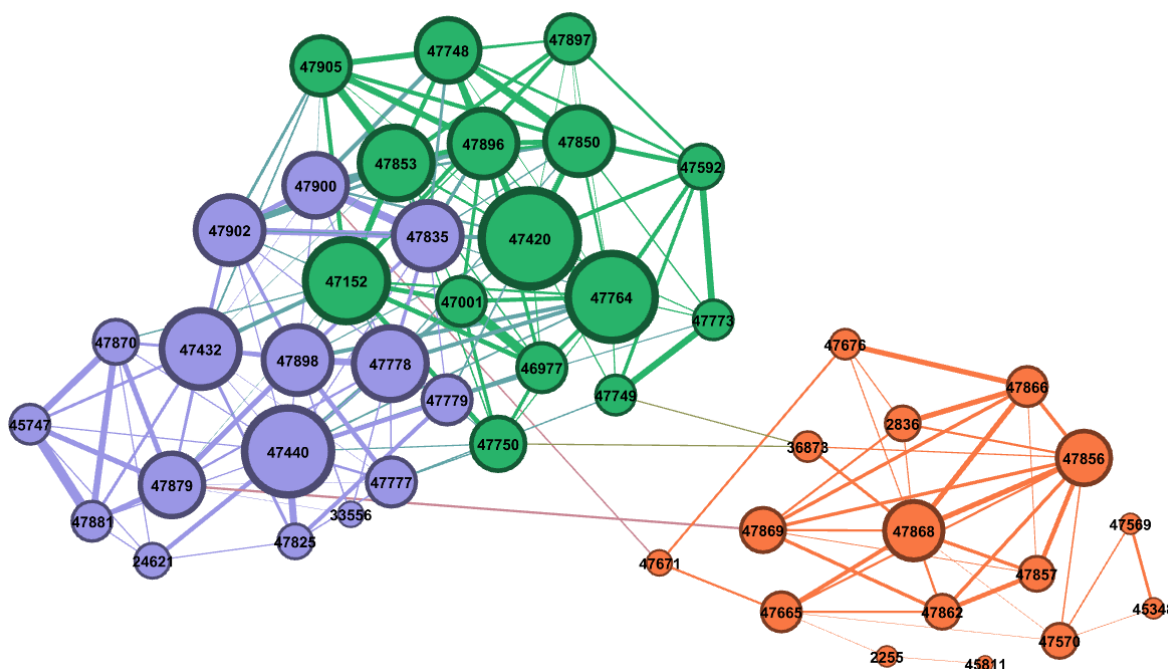
```
select vid as id from (select distinct vid1 as vid from distance2 where cosine > 0.7 union select
distinct vid2 as vid from distance2 where cosine > 0.7) as res
```

Fonte: Elaborado pelo autor (2019).

Uma vez estabelecido o parâmetro a ser utilizado como ponto de corte, os dados foram importados para a definição dos grupos, de acordo com a similaridade dos perfis, resultando em um grafo com três grupos principais (Figura 40), que foram coloridos na cor verde, roxa e laranja. Cada nó representa um aluno e quanto maior o nó, maior a representatividade de

características no perfil semelhantes a outros alunos. Portanto, também neste cenário, cinco alunos de cada grupo foram escolhidos para a recomendação (Quadro 37). O aluno com maior grau de cada grupo será o originário das recomendações.

Figura 40 – Grafo do segundo cenário de recomendação.



Fonte: Elaborada pelo autor (2019).

Quadro 37 – Definição de alunos para o cenário.

Grupo	Cor	Id Aluno_Chave	Alunos com grau próximo
1	Verde	47420	- 47152, 47764, 47853, 47896, 47850
2	Roxa	47440	- 47432, 47898, 47778, 47835, 47902
3	Laranja	47868	- 47869, 47856, 47866, 47665, 47857

Fonte: Elaborado pelo autor (2019).

As mesmas consultas constantes nos Quadros 25, 26 e 27 foram reutilizadas para extrair da base de dados as atividades realizadas por cada aluno.

Na análise do grupo 1, a primeira consulta (Quadro 25) retornou 16 atividades realizadas pelo aluno alvo (aluno 47420) no ambiente de aprendizagem. A segunda consulta (Quadro 26), retornou 112 atividades realizadas por todos os outros alunos do grupo que apresentam a similaridade no perfil. A terceira consulta (Quadro 27) retornou as atividades que os outros alunos realizaram, excluindo as atividades já realizadas pelo aluno alvo. Desta forma,

96 atividades são sugeridas para recomendação. O Quadro 38 apresenta algumas atividades recomendadas de acordo com o procedimento adotado.

Quadro 38 – Exemplo de atividades recomendadas ao aluno 47440.

Id_Ativ	Título	Disciplina
133	Web conferência Excel	Informática Aplicada I
139	Revisão dos conteúdos	Redação Técnica I
160	Resolução do trabalho	Eletricidade I
164	Medições elétricas - resolução dos exercícios	Eletricidade I
5727	Evolução da comunicação empresarial	Redação Técnica I
5728	Quesitos de um trabalho técnico	Redação Técnica I
5996	Aula 4 - Unidade 4 - tópico 1, 2, 3 e 4	Cálculo Técnico I
6543	Medições elétricas	Eletricidade I
6765	Magnetismo e eletromagnetismo	Eletricidade I
6931	Ligações usuais e suas representações - II	Instalações Elétricas I

Fonte: Elaborado pelo autor (2019).

Repetindo as consultas para cada aluno alvo e para cada grupo de alunos com perfis semelhantes temos as outras recomendações de atividades.

No grupo 2, as consultas indicaram que o aluno 47.440 realizou 82 atividades no curso *online*, contra 111 atividades do grupo. Logo, 29 atividades são recomendadas pra esse aluno, algumas delas exemplificadas no Quadro 39.

Para o grupo 3, o resultado apresentou 66 atividades já realizadas pelo aluno 47868 na plataforma, e o grupo de alunos com grau próximo realizou 112 atividades, resultando em uma recomendação com 46 atividades, algumas também constantes no Quadro 40.

Quadro 39 – Exemplo de atividades recomendadas ao aluno 47420.

Id_Atv	Título	Disciplina
139	Revisão dos conteúdos	Redação Técnica I
152	Web Conferência Cálculo Técnico	Cálculo Técnico I
156	Eletrostática e Eletrodinâmica	Eletricidade I
160	Resolução do Trabalho	Eletricidade I
164	Medições Elétricas - Resolução dos Exercícios	Eletricidade I
507	Avaliação <i>Online</i>	Instalações Elétricas I
514	Orientações Finais/Avaliação <i>online</i>	Instalações Elétricas I
520	Fórum temático / Avaliação <i>online</i> / Encontro presencial	Segurança e Normas Técnicas em Eletricidade I
2054	Entrega de Trabalho	Redação Técnica I
2068	Avaliação on Line	Redação Técnica I

Fonte: Elaborado pelo autor (2019).

Quadro 40 – Exemplo de atividades recomendadas ao aluno 47868.

Id_Atv	Título	Disciplina
157	Associação de Resistores	Eletricidade I
160	Resolução do Trabalho	Eletricidade I
520	Fórum temático / Avaliação <i>online</i> / Encontro presencial	Segurança e Normas Técnicas em Eletricidade I
2078	Avaliação On-line	Informática Aplicada I
5537	Aula 4 - MS Word: Unidade 3 - Tópico 5	Informática Aplicada I
5643	Fundamentos da eletricidade	Instalações Elétricas I
5727	Evolução da Comunicação Empresarial	Redação Técnica I
6584	Indutores	Eletricidade I
6591	Sistema elétrico brasileiro	Instalações Elétricas I
6593	Símbolos gráficos para instalações elétricas	Instalações Elétricas I

Fonte: Elaborado pelo autor (2019).

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste capítulo são apresentadas as considerações finais referentes ao desenvolvimento desta pesquisa, as contribuições produzidas e, por fim, as perspectivas de trabalhos futuros.

5.1 CONSIDERAÇÕES FINAIS

O objetivo do trabalho foi propor um modelo de perfil de aluno adequado para a aplicação de técnicas de LA.

A fim de se avaliar o modelo proposto, técnicas de LA foram utilizadas para extrair e analisar informações relevantes para propiciar a análise dos dados educacionais a partir de ambientes de aprendizagem *online*.

Além de informações básicas sobre o aluno e suas atividades realizadas no ambiente de ensino *online*, coletadas de forma automática, outras informações julgadas importantes por outros trabalhos na literatura foram inseridas no modelo proposto, utilizando uma estrutura lógica capaz de armazenar de forma eficaz os dados selecionados para as análises. Alguns desses dados inexistentes no banco de dados original foram inseridos com valores aleatórios, objetivando simular uma coleta explícita para o modelo.

Após a revisão sistemática da literatura, foram estabelecidos os procedimentos metodológicos do trabalho, resultando na proposição de uma visão macro através de uma estrutura analítica do projeto, em que as partes desta estrutura são vistas como módulos que permitem a composição do modelo.

O referencial teórico orbitou nas áreas dos sistemas educacionais, sistemas educacionais *online*, educação a distância, ambientes virtuais de aprendizagem, MOOC's, *Learning Analytics*, Análise de Agrupamento e Sistemas de Recomendação.

O modelo proposto iniciou pela visão conceitual, apresentando as entidades e seus relacionamentos. A etapa seguinte constituiu na confecção do modelo lógico com as tabelas, colunas e domínios.

Os dados utilizados neste trabalho foram providos por uma instituição de ensino, que oferta em sua grade curricular cursos técnicos a distância. Como já mencionado, dados sobre a personalidade de cada aluno não estavam disponíveis na base de dados, desde modo, foram informados aleatoriamente.

Após a carga dos dados e o devido pré-processamento, análises considerando estatística descritiva foram realizadas, e técnicas de *Learning Analytics* foram aplicadas. Portanto, através de consultas estruturadas no banco de dados, as informações foram extraídas,

analisadas e gráficos foram gerados, de modo a proporcionar uma visão geral das características de cada informação disponível.

Dentre as informações analisadas, considerando o modelo, tem-se o sexo, a quantidade de aluno em cada fase do curso, os tipos de cursos oferecidos, a quantidade de disciplinas em cada fase do curso, a quantidade de alunos com determinado estilo cognitivo e com estilos de aprendizagem diferentes, e o tipo de personalidade de cada aluno. Estas últimas características, analisadas sob o aspecto de LA, através do cálculo da medida de similaridade do cosseno, proporcionaram, juntamente com o estilo de aprendizagem, a recomendação de conteúdo relevante aos alunos.

No contexto de recomendações, identificou-se que a abordagem colaborativa, de acordo com a pesquisa na literatura, é especialmente útil quando já se possui uma grande quantidade de dados disponíveis para análise. Desta forma, seria interessante também utilizar outras abordagens, ou mesmo uma abordagem mista para se analisar a base de dados de ambientes *online* de aprendizagem que ainda possuem poucas interações dos alunos.

A análise de agrupamentos foi especialmente útil para se catalogar, dentro de um mesmo grupo, alunos com perfil semelhantes, utilizando para isto o algoritmo de modularidade. Vale mencionar que os grafos apresentados possuem alguns integrantes que fazem parte de dois grupos diferentes. Isso demonstra que os grupos criados não são mutuamente excludentes, pois um aluno pode apresentar características comuns aos dois grupos.

As atividades recomendadas aos alunos foram obtidas através de consultas a base de dados e depois aplicando uma consulta diferencial, para fins de acompanhamento didático. Em uma implementação real, um método em uma linguagem de programação ou serviço, poderiam ser criados para automaticamente realizar essa tarefa e apresentar ao aluno as atividades recomendadas.

Dois cenários de recomendação foram propostos e analisados, sendo que os resultados encontrados permitiram evidenciar a potencialidade do modelo para disponibilizar conteúdo relevante para os alunos, de acordo com a personalidade de cada um. A aplicação de cada uma das técnicas com outros filtros, poderia ser utilizada para segmentar os dados de acordo com a turma ou com o curso. A partir das análises realizadas sobre o modelo proposto, no sentido de produzir informações úteis à tomada de decisão, pode-se afirmar que o objetivo geral deste trabalho foi alcançado.

5.2 TRABALHOS FUTUROS

Ao longo desta pesquisa algumas áreas para estudo futuros foram identificadas. A realização de uma análise de outras características dos alunos que podem ser utilizadas para reuni-los em grupos relevantes de acordo com o objetivo educacional se mostra uma temática interessante. Por outro lado, uma rotina de análise das interações do aluno com o ambiente educacional *online* poderia utilizar dados específicos para procurar interações do aluno em outras plataformas digitais, enriquecendo as informações para a composição de seu perfil de aprendizado, como as redes sociais e participações em fóruns.

A agregação de outras características do aluno no modelo pode subsidiar e complementar as informações para análise. Assim, seria possível o agrupamento dos alunos de acordo com o estilo de aprendizagem, de acordo a personalidade, de acordo com o desempenho na execução das atividades propostas na plataforma ou outros critérios que sejam considerados relevantes.

Dados relativos as atividades extracurriculares, dados do dispositivo utilizado pelo aluno ao acessar o ambiente *online* de ensino, e de redes sociais, também podem ser relevantes para subsidiar decisões educacionais e podem ser considerados para análises futuras.

Um outro aspecto que poderia ser trabalhado futuramente é a utilização de outros tipos de abordagens de recomendação, como a recomendação baseada em conteúdo, baseada na comunidade, a demográfica ou mesmo a abordagem híbrida, onde duas ou mais técnicas são combinadas visando aprimorar a qualidade das recomendações. Assim, os resultados da aplicação das diferentes abordagens poderiam ser avaliados, de modo a se verificar qual delas são mais relevantes para o conjunto de dados existente.

Outra possibilidade seria avaliar o perfil do aluno de modo temporal, ou seja, considerando intervalos de tempo a partir das informações geradas pelos alunos. Deste modo, um modelo suportado por aprendizado de máquina, considerando informações como os acessos e atividades executadas por determinado aluno na plataforma de ensino, poderia ser estabelecido a fim de prever possíveis desistências de cursos.

REFERÊNCIAS

- AAMIR, Mohammad; BHUSRY, Mamta. Recommendation system: state of the art approach. **International Journal of Computer Applications**, v. 120, n. 12, p. 25–32, 2015.
- ADRAOUI, Meriem et al. Social learning analytics to describe the learner's interaction in online discussion forum in Moodle. In: **2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)**. IEEE, 2017. p. 1-6.
- AJZEN, Icek. The theory of planned behavior. **Organizational Behavior and Human Decision Processes**, v. 50, n. 2, p. 179–211, 1991.
- AHMAD, Aliyu Usman; STARKEY, Andrew. Application of feature selection methods for automated clustering analysis: a review on synthetic datasets. **Neural Computing and Applications**, v. 29, n. 7, p. 317-328, 2018.
- ALEVEN, Vincent et al. Intelligent tutoring systems for III-defined domains. In: **9th International Conference on Intelligent Tutoring Systems**, 2008.
- AL GHAMDI, A; SAMARJI, Ahmad; WATT, Anthony. Essential considerations in distance education in KSA: teacher immediacy in a virtual teaching and learning environment. **International Journal of Information and Education Technology**, v. 6, n. 1, p. 17–22, 2016.
- AL-MAMOORI, Heba AA; SHEHAB, Mohamed Eleman; EL FAKHARANY, Essam. Student profile in e-learning environment based on two-dimensional ontologies. In: **International Conference on Advanced Machine Learning Technologies and Applications**. Springer, Cham, 2018. p 301-310.
- ALMUTAIRI, Faisal M.; SIDIROPOULOS, Nicholas D.; KARYPIS, George. Context-aware recommendation-based Learning Analytics using tensor and coupled matrix factorization. **IEEE Journal on Selected Topics in Signal Processing**, v. 11, n. 5, p. 729–741, 2017.
- AMINI, Bahram et al. Capturing scholar’s knowledge from heterogeneous resources for profiling in recommender systems. **Expert Systems with Applications**, v. 41, n. 17, p. 7945-7957, 2014.
- ANSARI, M. H. et al. CodERS: A hybrid recommender system for an E-learning system. **Proceedings - 2016 2nd International Conference of Signal Processing and Intelligent Systems, ICSPIS 2016**, p. 14–15, 2017.
- ARAZY, Ofer; KUMAR, Nanda; SHAPIRA, Bracha. Improving social recommender systems. **IT Professional**, v. 11, n. 4, p. 38–44, 2009.
- ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA. **Censo EaD.BR: Relatório Analítico da Aprendizagem a Distância no Brasil**, 2017. Disponível em: <http://www.abed.org.br/site/pt/midiateca/censo_ead/1554/2018/10/censoeadbr2017/2018>. Acesso em: 21 out. 2018.

ATTWELL, Graham et al. Personal learning environments - the future of eLearning. **Elearning papers**, v. 2, n. 1, p. 1–8, 2007.

AYED, Abdelkarim Ben; HALIMA, Mohamed Ben; ALIMI, Adel M. Survey on clustering methods: towards fuzzy clustering for big data. In: **6th International Conference on Soft Computing and Pattern Recognition (SoCPaR)**. IEEE, 2014, p. 331–336.

BALACHEFF, N. et al. **Technology-enhanced learning**. Berlin: Springer, 2009.

BASTOS, Rogério Cid; BIAGIOTTI, Breno. MOOCs: uma alternativa para a democratização do ensino. **RENOTE**, v. 12, n. 1, p. 1–9, 2014.

BATES, AW. Tony. **Technology, e-learning and distance education**. Routledge, 2005.

AYED, Abdelkarim Ben; HALIMA, Mohamed Ben; ALIMI, Adel M. Survey on clustering methods: Towards fuzzy clustering for big data. In: **2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)**. IEEE, 2014. p. 331-336.

BERNARDINO, Giana da Silva. **Um modelo de perfil de educando baseado em sistemas educacionais aplicado ao contexto de sistemas de recomendação**. 2017. Dissertação (Mestrado em Tecnologias da Informação e Comunicação) - Universidade Federal de Santa Catarina, Araranguá, 2017.

BERTOLETTI, Ana Carolina et al. Educar pela pesquisa – uma abordagem para o desenvolvimento e utilização de softwares educacionais. **RENOTE**, v. 1, n. 2, 2003.

BERTRAM, Johanna; MOSKALIUK, Johannes; CRESS, Ulrike. Virtual police: acquiring Knowledge-in-use in Virtual Training Environments. In: 2011 IEEE International Symposium on VR Innovation. IEEE, 2011. p. 341-342.

BHARARA, Sanyam; SABITHA, Sai; BANSAL, Abhay. Application of learning analytics using clustering data mining for students' disposition analysis. **Education and Information Technologies**, v. 23, n. 2, 2017.

BHATT, Bhumika; PREMAL, J. Patel; GAUDANI, Hetal. A review paper on machine learning based recommendation system, 2014.

BOBADILLA, Jesús et al. Recommender systems survey. **Knowledge-Based Systems**, v. 46, p. 109–132, 2013.

BOCCALETTI, S. et al. Detecting complex network modularity by dynamical clustering. **Physical Review E**, v. 75, n. 4, p. 1–4, 2007.

BODILY, Robert; VERBERT, Katrien. Review of research on student-facing learning analytics dashboards and educational recommender systems. **IEEE Transactions on Learning Technologies**, v. 10, n. 4, p. 405-418, 2017.

BOND, Paul; LEIBOWITZ, Faye. MOOCs and serials. **Serials Review**, v. 39, n. 4, p. 258–260, 2013.

BRASIL. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. **Resolução nº 1, de 11 de março de 2016**. Define as diretrizes e normas nacionais

para a oferta de programas e cursos de educação superior na modalidade a distância. Brasília: Diário Oficial da União, 11/mar, 2016. Disponível em: <<http://portal.mec.gov.br/docman/marco-2016-pdf/35541-res-cne-ces-001-14032016-pdf/file>>. Acesso em: 2 dez. 2018.

_____. Ministério da Educação. Conselho Nacional de Educação. Câmara de Educação Superior. Parecer CNE/CES n. 564/2015. Disponível em: <www.abed.org.br/arquivos/parecer_cne_ces_564_15.pdf>. Acesso em: 2 dez. 2018.

_____. **Lei nº 9.394, de 20 de dezembro de 1996.** Estabelece as diretrizes e bases da educação nacional. Brasília, DF, dez 1996. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/19394.htm>. Acesso em 26 jul. 2019.

BREMGARTNER, Vitor. **Arcabouço conceitual de adaptação de recursos educacionais.** 2017. Tese (Doutorado em Informática) - Instituto de Computação, Universidade Federal do Amazonas, 2017.

BROOKS, D. Christopher; THAYER, Terri-Lynn. B. Institutional analytics in higher education. In: **Research report.** Educause Center for Analysis and Research (ECAR), 2016.

BRUSILOVSKY, P.; MILLÁN, E. User models for adaptive hypermedia and adaptive educational systems. In: **The adaptive web.** Springer, 2007. p. 3–53.

BURKE, R. Hybrid web recommender systems. In: **The adaptive web.** Springer, Berlin, 2007. p. 377–408.

CAMPBELL, John P.; DEBLOIS, Peter B.; OBLINGER, Diana G. Academic analytics: a new tool for a new era. **EDUCASE Review**, v. 42 n. 4, p. 40, 2007.

CAMPOS, Rodrigo; DOS SANTOS, Rodrigo Pereira; OLIVEIRA, Jonice. Web-based recommendation system architecture for knowledge reuse in MOOCS ecosystems. In: **2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science (IRI).** IEEE 2018. p. 193–200.

CANTABELLA, Magdalena et al. Analysis of student behavior in learning management systems through a Big Data framework. **Future Generation Computer Systems**, v. 90, p. 262–272, 2019.

CAO, Langcai et al. Design and implementation of writing recommendation system based on hybrid recommendation. **IEEE Access**, v.6, p. 72506-72513, 2018.

CHAKRABORTY, Jaydeep; THOPUGUNTA, Gurusrikar; BANSAL, Srividya. Data extraction and integration for scholar recommendation system. In: **2018 IEEE 12th International Conference on Semantic Computing (ICSC).** IEEE, 2018. p. 397–402.

CHATTI, Mohamed Amine et al. A reference model for learning analytics. **International Journal of Technology Enhanced Learning**, v. 4, n. 5-6, p. 318–331, 2013.

_____, Mohamed Amine et al. Learning Analytics: challenges and future research directions. **Eleed**, v. 10, n. 1, p. 1–17, 2014.

- CHAURASIA, Sushil S. et al. Big data academic and learning analytics: connecting the dots for academic excellence in higher education. **International Journal of Educational Management**, v. 32, n. 6, p. 1099–1117, 2018.
- CHAVAN, Reshma; MUKHOPADHYAY, Debajyoti. A comparative study of recommendation algorithms in e-commerce. In: **2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)**. IEEE, 2017. p. 163-168.
- CHEN, Jie. et al. An implicit information based movie recommendation strategy. In: **2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)**. IEEE, 2018. p. 405–410.
- CHEN, Li.; PU, Pearl. Critiquing-based recommenders: survey and emerging trends. **User Modeling and User-Adapted Interaction**, v. 22, n. 1–2, p. 125–150, 2012.
- CHEN, Rui. et al. A survey of collaborative filtering-based recommender systems: from traditional methods to hybrid methods based on social networks. **IEEE Access**, v. 6, p. 64301–64320, 2018.
- CHIU, Ming Ming; FUJITA, Nobuko. Statistical discourse analysis: a method for modeling online discussion processes. **Journal of Learning Analytics**, v. 1, n. 3, p. 61–83, 2014.
- CLARK, Ruth C.; MAYER, Richard E. **E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning**. 3^a ed. John Wiley & Sons, 2016.
- CLOW, Doug. The Learning Analytics cycle - closing the loop effectively. **Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12**, p. 134, 2012.
- COSTA, Luciano Andreatta Carvalho da; FRANCO, Sérgio Roberto Kieling. Aprendizagem colaborativa na Educação a Distância: aspectos teóricos, estratégias pedagógicas e experiências realizadas. **RENOTE: revista Novas Tecnologias na Educação**. Porto Alegre, RS, v. 4, n. 2, p. 1–10, 2006.
- CUPANI, Alberto. **Filosofia da tecnologia: um convite**. Florianópolis: Ed. da UFSC, 2011.
- DAHLSTROM, Eden; BROOKS, D. Christopher; BICHSEL, Jacqueline. The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives. 2014.
- DAUD, Ali. et al. Predicting student performance using advanced Learning Analytics. In: **Proceedings of the 26th International Conference on World Wide Web Companion**. International World Wide Web Conferences Steering Committee, 2017. p. 415-421.
- DE ARAÚJO LIMA, Gean Flávio; MERINO, Eugênio Adres Diaz; TRISKA, Ricardo. Métodos mais usados para avaliações de Ambientes Virtuais de Aprendizagem (AVAs). **Human Factors in Design**, v. 7, n. 13, p. 132–147, 2018.
- DE MELO, Sara; DANTAS, Adilmar Coelho; FERNANDES, Márcia. Modelo do estudante

baseado em emoções e perfis de personalidade para recomendação de estratégias pedagógicas personalizadas. In: **XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)**, v. 1, 2017. p. 967.

DE OLIVEIRA, Cláudio; MOURA, Samuel Pedrosa; SOUSA, Edinaldo Ribeiro de. TIC'S na educação: A utilização das tecnologias da informação e comunicação na aprendizagem do aluno. **Pedagogia em Ação**, v. 7, n. 1, p. 75–95, 2015.

DHARMARAJAN, A.; VELMURUGAN, T. Applications of partition based clustering algorithms: A survey. In: **2013 IEEE International Conference on Computational Intelligence and Computing Research**. IEEE, 2013. p. 1-5.

DIAS, Ana Augusta; GOMES, Maria João. E-learning para e-formadores. **Anais do eLES'04-eLearning no Ensino Superior**, 2004.

DILLENBOURG, Pierre; SCHNEIDER, Daniel; SYNTETA, Paraskevi. Virtual learning environments. In: **3rd Hellenic Conference" Information & Communication Technologies in Education"**. Kastaniotis Editions, Greece, 2002. p. 3-18.

DODGE, Bernie; WHITMER, John; FRAZEE, James P. Improving undergraduate student achievement in large blended courses through data-driven interventions. **Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15**. ACM, 2015 p. 412–413.

DRESCH, Aline; LACERDA, Daniel Pacheco; JÚNIOR, José Antônio Valle Antunes. **Design Science Research: método de pesquisa para avanço da ciência e tecnologia**. Bookman Editora, 2015.

DURÁN, Elena; AMANDI, Analía. Collaborative student profile to support assistance in CSCL environment. In: **Proceedings of the 2008 Euro American Conference on Telematics and Information Systems**. ACM, 2008. p. 38.

DUTT, Ashish; ISMAIL, Maizatul Akmar; HERAWAN, Tutut. A systematic review on educational data mining. **IEEE Access**, v.5, p. 15991, 2017.

DYCKHOFF, Anna Lea et al. Design and implementation of a learning analytics toolkit for teachers. **Journal of Educational Technology & Society**, v. 15, n. 3, p. 58-76, 2012.

EGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. **Linhas de Pesquisa**. UFSC, Florianópolis. Disponível em: < <http://www.egc.ufsc.br/pesquisas/linhas-de-pesquisa/>>. Acesso em: 26 ago. 2018a.

EGC – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. **Programa**. UFSC, Florianópolis. Disponível em: < <http://www.egc.ufsc.br/pos-graduacao/programa/>>. Acesso em: 26 ago. 2018b.

ELLIS, Robert A.; HAN, Feifei; PARDO, Abelardo. Improving Learning Analytics—combining observational and self-report data on student learning. **Journal of Educational Technology & Society**, v. 20, n. 3, p. 158-169, 2017.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Database systems**. Pearson Education Boston, v. 9. MA, 2011.

FAHAD, Adil et al. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. **IEEE transactions on emerging topics in computing**, v. 2, n. 3, p. 267-279, 2014.

FAN, Chenxi et al. Social Network Mining for Recommendation of Friends Based on Music Interests. In: **2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)**. IEEE, 2018. p. 833-840.

FELDER, Richard M; SILVERMAN, Linda K. Learning and Teaching Styles in Engineering Education. **Engineering Education**, v. 78, n. 7, p. 674–681, 1998.

FERGUSON, Rebecca. Learning analytics: drivers, developments and challenges. **International Journal of Technology Enhanced Learning**, v. 4, n. 5/6, p. 304–317, 2012.

FIDALGO-BLANCO, Ángel et al. Using Learning Analytics to improve teamwork assessment. **Computers in Human Behavior journal**, v. 47, p. 149–156, 2015.

FIGUEROA, Cristhian et al. A systematic literature review of Linked Data-based recommender systems. **Concurrency and Computation: Practice and Experience**, v. 27, n. 17, p. 4659–4684, 2015.

FIUZA, Patricia; MOCELIN, Roberta Ribas. Systematic review of literature: the contributions to the learning process by digital technologies and pedagogical architectures. In: **New Advances in Information Systems and Technologies**. Springer, 2016. p. 225–234.

FOURNIER, Helene; KOP, Rita; DURAND, Guillaume. Challenges to research in MOOCs. **MERLOT Journal of Online Learning and Teaching**, v. 10, n. 1, p. 1–15, 2014.

GARCÍA-ÁLVAREZ, María Teresa.; NOVO-CORTI, Isabel; VARELA-CANDAMIO, Laura. The effects of social networks on the assessment of virtual learning environments: A study for social sciences degrees. **Telematics and Informatics**, v. 35, n. 4, p. 1005–1017, 2018.

GARDNER, Josh.; BROOKS, Christopher. Student success prediction in MOOCs. **User Modeling and User-Adapted Interaction**, v. 28, n. 2, p. 127–203, 2018.

GAŠEVIĆ, Dragan. et al. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. **Internet and Higher Education**, v. 28, p. 68–84, 2016.

GAUTHIER, Geneviève. Using teaching analytics to inform assessment practices in technology mediated problem solving tasks. **IWTA@ LAK**, 2013.

GILMORE, Dawn Marie. Goffman's front stage and backstage behaviors in online education. **Journal of Learning Analytics**, v. 1, n. 3, p. 187–190, 2014.

GOLBECK, Jennifer. Generating predictive movie recommendations from trust in social networks. In: **International Conference on Trust Management - LNCS**. Springer, Berlin, Heidelberg, 2006. p. 93-104.

- GOLDSCHMIDT, Ronaldo; PASSOS, Eduardo; BEZERRA, E. Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. **Rio de Janeiro-RJ: Elsevier**, p. 56-60, 2015.
- GOMES, Alex Sandro et al. Amadeus: Novo modelo de sistema de gestão de aprendizagem. **Revista Brasileira de Aprendizagem Aberta e a Distância**, v. 8, 2009.
- GOPE, Jyotirmoy; JAIN, Sanjay Kumar. A learning styles based recommender system prototype for edX courses. In: **2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)**. IEEE, 2017. p. 414-419
- HAJRI, Hiba; BOURDA, Yolaine; POPINEAU, Fabrice. A system to recommend open educational resources during an online course. **CSEDU (1)**. 2018. p. 99–109.
- HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data mining: concepts and techniques**. Elsevier, 2011.
- HANSEN, Christian et al. Sequence modelling for analysing student interaction with educational systems. **arXiv preprint arXiv:1708.04164**, 2017.
- HARFIELD, Timothy D. Teaching the unteachable: on the compatibility of learning analytics and humane education. In: **Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, LAK 2014**. ACM, 2014. p. 241–245.
- HE, Lingjun et al. Predictive analytics machinery for STEM student success studies. **Applied Artificial Intelligence**, v. 32, n. 4, p. 361–387, 2018.
- HOLMBERG, Borje. A theory of distance education based on empathy. **Handbook of distance education**, p. 79–86, 2003.
- HOU, Yifan et al. Course recommendation of MOOC with big data support: a contextual online learning approach. In: **IEEE ENFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)**. IEEE, 2018. p. 106–111.
- IFENTHALER, Dirk; WIDANAPATHIRANA, Chathuranga. Development and validation of a learning analytics framework: Two case studies using support vector machines. **Technology, Knowledge and Learning**, v. 19, n. 1-2, p. 221-240, 2014.
- IMRAN, Hazra et al. PLORS: a personalized learning object recommender system. **Vietnam Journal of Computer Science**, v. 3, n. 1, p. 3–13, 2016.
- ISHANKA, UA Piumi; YUKAWA, Takashi. User emotion and personality in context-aware travel destination recommendation. In: **2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)**, IEEE, 2018. p. 13–18.
- JAIN, Anil K.; MURTY, M. Narasimha; FLYNN, Patrick J. Data clustering: a review. **ACM computing surveys (CSUR)**, v. 31, n. 3, p. 264-323, 1999.
- _____, Anil K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010.
- JALILI, Mahdi et al. Evaluating collaborative filtering recommender algorithms: a survey.

IEEE Access, v. 6, p. 74003-74024, 2018.

JANNACH, Dietmar et al. **Recommender systems: an introduction**. Cambridge University Press, 2010.

JENA, R. K. Predicting students ' learning style using learning analytics: a case study of business management students from India. **Behaviour & Information Technology**, v. 37, n. 10–11, p. 978–992, 2018.

JHA, Sunil Kr; SHORKO, Sefan. A novel CALM algorithm in student profiling. **Computer Applications in Engineering Education**, v. 26, n. 4, p. 841–851, 2018.

JING, Xia; TANG, Jie. Guess you like: course recommendation in MOOCs. In: **Proceedings of the International Conference on Web Intelligence**. ACM, 2017. p. 783–789.

JINYIN, Chen. et al. A novel cluster center fast determination clustering algorithm. **Applied Soft Computing Journal**, v. 57, p. 539–555, 2017.

JOHNSON, Larry et al. **NMC Horizon Report: 2013 Higher Education Edition**. Austin: The New Media Consortium, 2013.

KAMAL, Rtili Mohammed; MOHAMED, Khaldi; ALI, Dahmani. Improvement of complementary pedagogical resources indexing based on pedagogical warehouse for recommendation system CEHL. In: **2016 3rd International Conference on Systems of Collaboration (SysCo)**. IEE, 2016. p. 1-5.

KARABOGA, Dervis; OZTURK, Celal. A novel clustering approach: Artificial Bee Colony (ABC) algorithm. **Applied soft computing**, v. 11, n. 1, p. 652-657, 2011.

KARATAS, Arzum; SAHIN, Serap. A comparative study of modularity-based community detection methods for online social networks. In A. Tarhan and Murat E. (Eds.), paper presented at the 12th Turkish National Software Engineering Symposium, UYMS 2018; Istanbul; Turkey, 2018.

KAUFMAN, Leonard; ROUSSEEUW, Peter J. Finding groups in data: an introduction to cluster analysis. Jhon Wiley & Sons, 2009.

KAUSHIK, Anna. Massive open online course (MOOC): a new online learning approach to LIS professionals. **Library Waves-A Biannual Peer Reviewed Journal**, v. 1, n. 2, p. 110-118, 2015.

KEEGAN, Desmond. **Foundations of distance education**. 2nd. ed. New York: Routledge, 2013.

KENNEDY, Gregor et al. Mining interactions in immersive learning environments for real-time student feedback. **Australasian Journal of Educational Technology**, v. 29, n. 2, 2013.

KHALIL, Mohammad; EBNER, Martin. Learning analytics: principles and constraints. In: **EdMedia+ Innovate Learning**. Association for the Advancement of Computing in Education (AACE), 2015. p. 1789-1799.

KHRIBI, Mohamed Koutheair; JEMNI, Mohamed; NASRAOUI, Olf. Automatic

recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In: **2008 Eighth IEEE International Conference on Advanced Learning Technologies**. IEEE, 2008. p. 241-245.

KIM, Dongho et al. Toward evidence-based learning analytics: using proxy variables to improve asynchronous online discussion environments. **Internet and Higher Education**, v. 30, p. 30–43, 2016.

KIM, Kyungrog; MOON, Namme. A model for collecting and analyzing action data in a learning process based on activity theory. **Soft Computing**, v. 22, n. 20, p. 6671–6681, 2018.

KNIGHT, Simon; LITTLETON, Karen. Discourse-centric learning analytics: mapping the terrain. **Jornal of Learning Analytics**, v. 2, n. 1, p. 185–209, 2015.

KONCHADY, Manu. **Text mining application programming**. Massachusetts: Charles River Media, 2006.

KOVANOVIĆ, Vitomir et al. Does time-on-task estimation matter? Implications for the validity of learning analytics findings. **Jornal of Learning Analytics**, v. 2, n. 3, p. 81–110, 2016.

KOREN, Yehuda; BELL, Robert. Advances in collaborative filtering. In: **Recommender systems handbook**. Springer, Boston, MA, 2015. p. 77-118.

KRAIVIXIEN, Dhitiporn; WONGWANICH, Suwimon; SUJIVA, Siridej. Concepts of setting and utilizing student profile. **Procedia-Social and Behavioral Sciences**, v. 116, p. 1689-1695, 2014.

LAAKSO, Mikko-Jussi et al. Designing and creating a framework for learning analytics in Finland. In: **2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)**. IEEE, 2018 p. 0695-0700.

LACAVE, Carmen; MOLINA, Ana I.; CRUZ-LEMUS, José A. Learning Analytics to identify dropout factors of computer science studies through bayesian networks. **Behaviour & Information Technology**, v. 37, n. 10–11, p. 993–1007, 2018.

LANDIM, CMMPF. Educação a distância: algumas considerações. **Rio de Janeiro**, 1997.

LARUSSON, Johanmm Ari; WHITE, Brandon (Ed.). **Learning analytics: from research to practice**. v. 13, Springer, 2014.

LAVETI, Ramesh Naidu et al. implementation of learning analytics framework for MOOCs using state-of-the-art in-memory computing. **2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)**. IEEE, 2017. n. 1, p. 1–6.

LENZI, Greicy Kelli Spanhol. **Diretrizes para a gestão de projetos de cursos de capacitação na modalidade de educação a distância**. 2010.

LI, Shulong. Experiences of Simulation Training with Computer Technology in Police College. In: **2008 The 9th International Conference for Young Computer Scientists**. IEEE, 2008. p. 2604-2608.

LI, Yi.; ZHAI, Xiaoning. Review and prospect of modern education using Big Data. **Procedia Computer Science**, v. 129, p. 341–347, 2018.

LIÑÁN, Laura Calvet; PÉREZ, Ángel Alejandro Juan. Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. **International Journal of Educational Technology in Higher Education**, v. 12, n. 3, p. 98-112, 2015.

LITTO, Fredric M.; FORMIGA, Marcos. **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009.

LIU, Zhi et al. An emotion oriented topic modeling approach to discover what students are concerned about in course forums. **2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)**. IEEE, 2018. p. 170–172.

LOBO, Vivian Brian; ANSARI, Nazneen. Multimedia enabled virtual classroom for distance education. In: **2015 International Conference on Green Computing and Internet of Things (ICGCIoT)**. IEEE, 2015. p. 1598-1603.

LOCKYER, Lori; HEATHCOTE, Elizabeth; DAWSON, Shane. Informing pedagogical action: aligning Learning Analytics with learning design. **American Behavioral Scientist**, v. 57, n. 10, p. 1439–1459, 2013.

LOGICA, Banica; MAGDALENA, Radulescu. Using big data in the academic environment. **Procedia Economics and Finance**, v. 33, n. 15, p. 277–286, 2015.

LOTSARI, Elvira et al. A learning analytics methodology for student profiling. In: **Hellenic Conference on Artificial Intelligence**. Springer, Cham, 2014. p. 300-312.

LUNA, Vladimir et al. An ontology-based approach for representing the interaction process between user profile and its context for collaborative learning environments. **Computers in Human Behavior**, v. 51, p. 1387-1394, 2015.

MA, Joseph D.; LEE, Kelly C.; KUO, Grace M. A massive open online course on pharmacogenomics: Not just disruptive innovation but a possible solution. **Pharmacogenomics**, v. 14, n. 10, p. 1125–1127, 2013.

MACHADO, Viviane Guidotti. **Aula Virtual: implicações e desafios docentes considerando o cenário de educação apoiada por tecnologias digitais**. 2013. Dissertação (Mestrado em Educação) - Pontificia Universidade Católica do Rio Grande do Sul, Rio Grande do Sul, 2013.

MARTINS, Maria P. G.; MIGUÉIS, Vera L.; FONSECA, D. S. B. Educational data mining: a literature review. In: **2018 13th Iberian Conference on Information Systems and Technologies (CISTI)**. IEEE, 2018. p. 1-6.

MATSEBULA, Felize; MNKANDLA, Ernest. A Big Data architecture for Learning Analytics in higher education. In: **2017 IEEE AFRICON**. IEEE, 2017. p. 951–956.

MATTAR, João. Aprendizagem em ambientes virtuais: teorias, conectivismo e MOOCs. **São Paulo: TECCOGS-PUC/SP**, v. 7, p. 21–40, 2013.

MENEZES, Crediné Silva et al. MOrFEU–Multi-organizador flexível de espaços virtuais para

apoiar a inovação pedagógica em EaD. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2008. p. 451-460.

MOORE, Michael; KEARSLEY, Greg. **Educação a distância: uma visão integrada**. Tradução: Roberto Galman. São Paulo: Cengage Learning, 2008.

_____, Michael; KEARSLEY, Greg. **Educação a distância: sistemas de aprendizagem on-line**. São Paulo: Cengage Learning, 2013.

MOTHUKURI, Uday Kumar et al. Improvisation of learning experience using Learning Analytics in eLearning. In: **2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)**. IEEE, 2017. p. 1-6. 2017.

MUÑOZ-MERINO, Pedro J. et al. Precise effectiveness strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs. **Computers in Human Behavior**, v. 47, p. 108–118, 2015.

NA, K. S.; TASIR, Z. A Systematic review of Learning Analytics intervention contributing to student success in online learning. **2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)**. IEEE, 2017. p. 62–68.

NACU, Denise; MARTIN, Caitlin K.; PINKARD, Nichole. Designing for 21st century learning online: a heuristic method to enable educator learning support roles. **Educational Technology Research and Development**, v. 66, n. 4, p. 1029–1049, 2018.

NEWMAN, Mark E. J. Mixing patterns in networks. **Physical Review E**, v. 67, n. 2, p. 026126, 2003.

_____, Mark E. J.; GIRVAN, Michelle. Finding and evaluating community structure in networks. **Physical Review E**, v. 69, n. 2, p. 026113, 2004.

NGUYEN, Andy; GARDNER, Lesley A.; SHERIDAN, Don. A framework for applying learning analytics in serious games for people with intellectual disabilities. **British Journal of Educational Technology**, v. 49, n. 4, p. 673–689, 2018.

NING, Xia; DESROSIERS, Christian; KARYPIS, George. A comprehensive survey of neighborhood-based recommendation methods. In: **Recommender systems handbook**. Springer, Boston, MA, 2015. p. 37-76.

NOVAK, Silvestre. **Educação a distância e racionalidade comunicativa: a construção do entendimento na comunidade virtual de aprendizagem**. 2010. Tese (Doutorado em Educação) - Faculdade de Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

OCHOA, Xavier; WORSLEY, Marcelo. Augmenting Learning Analytics with multimodal sensory data. **Journal of Learning Analytics**, v. 3, n. 2, p. 213–219, 2016.

OSIŃSKI, Stanisław. An algorithm for clustering of web search results. **Master, Poznań University of Technology**, Poland, 2003.

PACHECO, Roberto Carlos dos Santos; TOSTA, Kelly Cristina Benetti Tonani; FREIRE,

Patricia de Sá. Interdisciplinaridade vista como um processo complexo de construção do conhecimento: uma análise do Programa de Pós-Graduação EGC/UFSC. **Revista Brasileira de Pós-Graduação**, v. 7, n. 12, 2010.

PACIFICO, Luciano Demetrio Santos; LUDERMIR, Teresa Bernarda. A group search optimization method for data clustering. In: **2014 Brazilian Conference on Intelligent Systems**. IEEE, 2014. p. 342-347.

PANG, Yanxia et al. Collaborative filtering recommendation for MOOC application. **Computer Applications in Engineering Education**, v. 25, n. 1, p. 120–128, 2017.

PARK, Deuk Hee et al. A literature review and classification of recommender systems research. **Expert Systems with Applications**, v. 39, n. 11, p. 10059-10072, 2012.

PASSALIS, Nikolaos; TEFAS, Anastasios Deep learning analytics. In: **Machine Learning Paradigms**. Springer, Cham, 2019, p. 339–370.

PATEL, K. M. Archana; THAKRAL, Prateek. The best clustering algorithms in data mining. In: **2016 International Conference on Communication and Signal Processing (ICCSP)**. IEEE, 2016. p. 2042–2046.

PECHENIZKIY, Mykola; GASEVIC, Dragan. Introduction into sparks of the Learning Analytics future. **Journal of Learning Analytics**, v. 1, n. 3, p. 145–149, 2015.

PEFFERS, Ken et al. A design science research methodology for information systems research. **Journal of management information systems**, v. 24, n. 3, p. 45-77, 2007.

PEÑA-AYALA, Alejandro (Ed.). **Learning analytics: fundamentals, applications and trends: a view of the current state of the art of enhance e-Learning**. Springer, 2017.

_____, Alejandro; CÁRDENAS-ROBLEDO, Leonor Adriana; SOSSA, Humberto. A landscape of Learning Analytics: an exercise to highlight the nature of an emergent field. In: **Learning Analytics: Fundamentals, Applications, and Trends**. Springer, Cham, 2017. p. 65-112.

PEREIRA, Alice Theresinha Cybis; SCHMITT, Valdenise; DIAS, M. R. A. C. Ambientes virtuais de aprendizagem. **AVA-Ambientes Virtuais de Aprendizagem em Diferentes Contextos**. Rio de Janeiro: Editora Ciência Moderna Ltda, p. 6-7, 2007.

PEREIRA, Crystiam Kelle et al. Extração de características de perfil e de contexto em redes sociais para recomendação de recursos educacionais. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 506.

PHILLIPS, Rob et al. Exploring Learning Analytics as indicators of study behaviour. In: **EdMedia+ Innovate Learning**. Association for the Advancement of Computing in Education (AACE), 2012, p. 2861–2867.

POONSIRIVONG, Korn; JITTAWRIAYNUKON, Chanintorn. Big data analytics using association rules in eLearning. In: **2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA)**. IEEE, 2018. p. 14–18.

- PRAYAGA, Lakshmi et al. Clustering and predicting driving violations using web-enabled big data techniques. In: **Cognitive Informatics and Soft Computing**. Springer, Singapore, 2019. p. 495-503.
- PRIETO, Luis P. et al. Recurrent routines: Analyzing and supporting orchestration in technology-enhanced primary classrooms. **Computers & Education**, v. 57, n. 1, p. 1214–1227, 2011.
- RABELLO, Cíntia Regina Lacerda; PEIXOTO, Maurício Abreu Pinto. Educação a distância : do ensino por correspondência à aprendizagem virtual - garantia de sucesso? **Revista EducaOnline**, v. 5, n. 1, p. 101–121, 2011.
- RAJENTHRAN, Hemabegai A./P.; SHAHARANEE, Izwan Nizal Mohd; JAMIL, Jastini Mohd. Student profiling on university co-curricular activities using cluster analysis. In: **AIP Conference Proceedings**. AIP Publishing, 2017. p. 050035.
- RAN, Juan et al. Logistic Regression Analysis on Learning Behavior and Learning Effect Based on SPOC Data. In: **2018 13th International Conference on Computer Science & Education (ICCSE)**. IEEE, 2018. p. 1-5.
- REDDY, S. R. S. et al. Content-based movie recommendation system using genre correlation. In: **Smart Intelligent Computing and Applications**. Springer, Singapore, 2019. p. 391-397.
- REZENDE, Paulo Alceu et al. Recomendação baseada no perfil e contexto tecnológico do aluno. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2018. p. 1273.
- RIBEIRO, Idália Sofia Ferreira. **A integração das tecnologias da informação e comunicação na educação especial: software educativo**. 2016. Dissertação (Mestrado em Ciências da Educação) Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal, 2016.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. Recommender systems: introduction and challenges. In: **Recommender System Handbook Second Edition**. p. 1–34, 2015.
- ROCHA, Heloíza. Ambiente TelEduc para educação a distância baseada na web. **Educação a Distância: fundamentos e práticas**. Campinas, SP: UNICAMP/NIED, 2002.
- RODRÍGUEZ, P. A.; DUQUE, N. D.; OVALLE, D. A. Método híbrido de recomendación adaptativa de objetos de aprendizaje basado en perfiles de usuario. **Formacion Universitaria**, v. 9, n. 4, p. 83–94, 2016.
- ROMERO, Cristobal; VENTURA, Sebastian. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 3, n. 1, p. 12-27, 2013.
- RUMBLE, G. **A gestão dos sistemas de ensino a distância**. Brasília, DF: UnB; UNESCO, 2003.
- SABA, Farhad. Distance education theory, methodology, and epistemology: A pragmatic

paradigm. **Handbook of distance education**, v. 1, p. 3–20, 2003.

SANTOS, Thaína. Saldanha dos. **Softwares educacionais e a formação de professores: as percepções de licenciandos sobre a informática na educação**. 2016. Dissertação (Mestrado em Educação, Ciências e Matemática) - Faculdade de Física, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2016.

SCAIFE, Michael et al. Designing for or designing with? Informant design for interactive learning environments. In: **Proceedings of the ACM SIGCHI Conference on Human factors in computing systems**. ACM, 1997. p. 343-350.

SCHREIBER, August. Th et al. **Knowledge engineering and management: the CommonKADS methodology**. MIT press, 2000.

SCHWERDTLE, Patricia.; BONNAMY, James. Big data in nurse education. **Nurse Education Today**, v. 51, p. 114–116, 2017.

SCIARRONE, Filippo. Machine learning and Learning Analytics: integrating data with learning. In: **2018 7th International Conference on Information Technology Based Higher Education and Training (ITHET)**. IEEW, 2018. p. 1-5.

SERGIS, Stylianos; SAMPSON, Demetrios G. Teaching and learning analytics to support teacher inquiry: A systematic literature review. In: **Learning analytics: Fundamentals, applications, and trends**. Springer, Cham, 2017. p. 25-63.

SHAH, D. By the numbers: MOOCS in 2017. Class Central [Blog post]. 2018

SHAO, Jingfeng. Research on Fuzzy Ontology E-learning Based on User Profile. In: **Proceedings of the 2017 International Conference on E-Education, E-Business and E-Technology**. ACM, 2017. p. 46-49.

SHAWKY, Doaa; BADAWI, Ashraf. A reinforcement learning-based adaptive learning system. In: **International Conference on Advanced Machine Learning Technologies and Applications**. Springer, Cham, 2018. p. 221-231.

SIADATY, Melody; GASEVIC, Dragan; HATALA, Marek. Trace-based micro-analytic measurement of self-regulated learning processes. **Journal of Learning Analytics**, v. 3, n. 1, p. 183–214-183–214, 2016.

SIEMENS, George. Connectivism: Learning as network-creation. **ASTD Learning News**, v. 10, n. 1, p. 1-28, 2005.

_____, George; DOWNES, Stephen. Connectivism & connective knowledge. **Universidade de Manitoba**, 2008.

SIKE, Ren; WEIGUO, Lin. Architecture design of online education system. In: **2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)**. IEEE, 2017. p. 689-692.

SILVA, Andreza Regina Lopes da; DIANA, Juliana Bordinhão; SPANHOL, Fernando José. Perspectivas da educação a distância: um olhar a partir da criatividade. In: VANZIN, Tarcísio; SILVA, Allan Jones Andreza; SILVA, Luciano Nascimento. **EDUCAÇÃO POLICIAL**

MILITAR E A CONSTRUÇÃO DE UMA SEGURANÇA CIDADÃ NA PARAÍBA. **Revista de Direito**, v. 8, n. 02, p. 25-49, 2016.

SIMON, Jean. A priori knowledge in Learning Analytics. In: **Learning Analytics: Fundamentals, Applications, and Trends**. Springer, Cham, 2017. p. 199–227.

SINHA, Rashmi; Swearingen, Kirsten. Comparing recommendations made by online systems and friends. In: DELOS. 2001.

SLADE, Sharon; PRINSLOO, Paul. Learning analytics: Ethical issues and dilemmas. **American Behavioral Scientist**, v. 57, n. 10, p. 1510-1529, 2013.

SPALENZA, Marcos A. et al. Uma ferramenta para mineração de dados educacionais: extração de informação em ambientes virtuais de aprendizagem. **Anais do Computer on the Beach**, p. 741–750, 2018.

STUDER, Rudi; BENJAMINS, V. Richard; FENSEL, Dieter. Knowledge engineering: principles and methods. **Data & Knowledge Engineering**, v. 25, n. 1–2, p. 161–197, 1998.

SYED, Thoufeeq Ahmed et al. Building a dataset for personalized learning recommendation system. In: **2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)**. IEEE, 2018. p. 2080–2086.

SYMEONIDIS, Panagiotis; MALAKOUDIS, Dimitrios. Multi-modal matrix factorization with side information for recommending massive open online courses. **Expert Systems with Applications**, v. 118, p. 261–271, 2019.

TAKANO, Kosuke; LI, Kin Fun. An adaptive e-learning recommender based on user's web-browsing behavior. In: **2010 International Conference on P2P, Parallel, Grid, Cloud and Internet Computing**. IEEE, 2010. p. 123-131.

TAN, Yueying et al. Learning profiles, behaviors and outcomes: investigating international students' learning experience in an english MOOC. In: **2018 International Symposium on Educational Technology (ISET)**. IEEE, 2018. p. 214–218.

TARUS, John K.; NIU, Zhendong; YOUSIF, Abdallah. A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. **Future Generation Computer Systems**, v. 72, p. 37–48, 2017.

TLILI, Ahmed et al. A smart educational game to model personality using learning analytics. In: **2017 IEEE 17th International conference on advanced learning technologies (ICALT)**. IEEE, 2017. p. 131-135.

TUKEY, J. W. Exploratory data analysis. Addison-Wesley Publishing Company, 1977.

UDUPI, Prakash Kumar; SHARMA, Nisha; JHA, S. K. Educational data mining and big data framework for e-learning environment. In: **2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)**. IEEE, 2016. p. 258-261.

ULBRICHT, Vania Ribas; BATISTA, Cláudia Regina (Eds.). **Criatividade e Inovação na**

Educação. São Paulo: Pimenta Cultural, p. 266, 2015.

UNESCO. **Global education monitoring report**. 2017. Disponível em: <<https://en.unesco.org/gem-report/report/2017/accountability-education>>. Acesso em 20 dez. 2017.

VAN HARMELEN, Mark; WORKMAN, David. Analytics for learning and teaching. **CETIS Analytics Series**, v. 1, n. 3, p. 1–40, 2012.

VAROUCHAS, Emmanouil; SICILIA, Miguel-Angel; SÁNCHEZ-ALONSO, Salvador. Towards an integrated learning analytics framework for quality perceptions in higher education: a 3-tier content, process, engagement model for key performance indicators. **Behaviour & Information Technology**, v. 37, n. 10-11, p. 1129-1141, 2018.

VASCONCELOS, E. S.; CRUZ, C. D., BHERING, L. L.; RESENDE JÚNIOR, M. F. R. Método alternativo para análise de agrupamento. **Pesq. agropec. bras.**, Brasília, v. 42, nº 10, 2007.

VEGA-PONS, Sandro; RUIZ-SHULCLOPER, José. A survey of clustering ensemble algorithms. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 25, n. 03, p. 337-372, 2011.

VIGENTINI, Lorenzo et al. Overcoming the MOOC data deluge with Learning Analytic dashboards. In: **Learning Analytics: Fundamentals, Applications, and Trends**. Springer, Cham, 2017. p. 171-198.

WANG, Shu-Lin; WU, Chun-Yi. Application of context-aware and personalized recommendation to implement an adaptive ubiquitous learning system. **Expert Systems with applications**, v. 38, n. 9, p. 10831-10838, 2011.

WATSON, William; WATSON, Sunnie Lee. An argument for clarity: what are learning management systems, what are they not, and what should they become. 2007.

WISE, Alyssa Friend. Designing pedagogical interventions to support student use of learning analytics. In: **Proceedings of the fourth international conference on learning analytics and knowledge**. ACM, 2014. p. 203-211.

_____, Alyssa Friend; SHAFFER, David Williamson. Why theory matters more than ever in the age of big data. **Journal of Learning Analytics**, v. 2, n. 2, p. 5-13, 2015.

WU, Dianshuang; LU, Jie; ZHANG, Guangquan. A fuzzy tree matching-based personalized e-learning recommender system. **IEEE transactions on fuzzy systems**, v. 23, n. 6, p. 2412-2426, 2015.

WU, Fati; HUANG, Luoying; ZOU, Rui. The design of intervention model and strategy based on the behavior data of learners: A learning analytics perspective. In: **International Conference on Hybrid Learning and Continuing Education**. Springer, Cham, 2015. p. 294-301.

WU, Yiwei; XIA, Xiaoling; LE, Jiajin. Evaluation of students' achievements based on factor analysis and cluster analysis. In: **2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)**. IEEE, 2017. p. 820-823.

XIE, Juanying et al. An Efficient Global K-means Clustering Algorithm. **JCP**, v. 6, n. 2, p. 271-279, 2011.

XING, Wanli et al. Group learning assessment: developing a theory-informed analytics. **Journal of Educational Technology and Society**, v. 18, n. 2, p. 110–128, 2015.

XU, Songer; LUO, Heng; TAN, Yueying. Re-examining the community of inquiry framework from the perspective of Learning Analytics. In: **2018 13th International Conference on Computer Science & Education (ICCSE)**. IEEE, 2018. p. 1-5.

YANG, Chen et al. A nearest neighbor based personal rank algorithm for collaborator recommendation. In: **2018 15th International Conference on Service Systems and Service Management (ICSSSM)**. IEEE, 2018. p. 1-5.

YAO, Jie; JIANG, Wenjun. Utilizing the bidirectional effect of evolutive trust-rating for recommendation in E-Commerce. In: **2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)**. IEEE, 2018. p. 1015–1022.

YU, Renzhe; JIANG, Daokun; WARSCHAUER, Mark. Representing and predicting student navigational pathways in online college courses. In: **Proceedings of the Fifth Annual ACM Conference on Learning at Scale**. ACM, 2018. p. 44.

YU, Shyr-Shen et al. Two improved k-means algorithms. **Applied Soft Computing**, v. 68, p. 747-755, 2018.

YUAN, Li; POWELL, S. J. MOOCs and open education: implications for higher education. 2013.

ZEM-LOPES, Aparecida M.; PEDRO, Laís Z.; ISOTANI, Seiji. Qualidade de softwares educacionais baseados na web (semântica): um mapeamento sistemático. **RENOTE**, v. 12, n. 1, 2014.

ZHOU, Shibing; XU, Zhenyuan; LIU, Fei. Method for determining the optimal number of clusters based on agglomerative hierarchical clustering. **IEEE transactions on neural networks and learning systems**, v. 28, n. 12, p. 3007-3017, 2016.

ZHU, Haiping et al. A Cross-Curriculum Video Recommendation Algorithm Based on a Video-Associated Knowledge Map. **IEEE Access**, v. 6, p. 57562-57571, 2018a.

ZHU, Qiliang et al. FMSR: A fairness-aware mobile service recommendation method. In: **2018 IEEE International Conference on Web Services (ICWS)**. IEEE, 2018b. p. 171–178.