

**INSTITUTO
FEDERAL**
Paraíba

Instituto Federal de Educação, Ciência e Tecnologia da Paraíba - Campus

João Pessoa

Programa de Pós-Graduação em Tecnologia da Informação

Johnny Yuri Solano Marinho

**ActivePlan: Uma Abordagem para Formação de Grupos de
Estudantes com Perfis Heterogêneos para Uso de Práticas
Ativas de Aprendizagem**

DISSERTAÇÃO DE MESTRADO

**JOÃO PESSOA
2022**

Johnny Yuri Solano Marinho

**ActivePlan: Uma Abordagem para Formação de Grupos de
Estudantes com Perfis Heterogêneos para Uso de Práticas
Ativas de Aprendizagem**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Tecnologia da Informação, pelo Programa de Pós Graduação em Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB.

Orientadora: Profa. Dra. Damires Souza

João Pessoa
2022

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Nilo Peçanha do IFPB, *campus* João Pessoa.

M338a Marinho, Johnny Yuri Solano.

Activeplan : uma abordagem para formação de grupos de estudantes com perfis heterogêneos para uso de práticas ativas de atividades / Johnny Yuri Solano Marinho. – 2022.

108 f. : il.

Dissertação (Mestrado em Tecnologia da Informação) – Instituto Federal de Educação da Paraíba / Programa de Pós-Graduação em Tecnologia da Informação, 2022.

Orientação : Prof^a D.ra Damires Yluska de Souza Fernandes.

1. Aprendizagem ativa. 2. Grupos heterogêneos de estudantes. 3. Taxonomia de Bloom. 4. Suporte docente. 5. Aprendizagem de máquinas não supervisionado. I. Título.

CDU 37.015.3(043)

Lucrecia Camilo de Lima
Bibliotecária – CRB 15/132

Johnny Yuri Solano Marinho

ActivePlan: Uma Abordagem para Formação de Grupos de Estudantes com Perfis Heterogêneos para Uso de Práticas Ativas de Aprendizagem

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Tecnologia da Informação, pelo Programa de Pós Graduação em Tecnologia da Informação do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB.

Aprovado em 27 de Maio de 2022.

BANCA EXAMINADORA:

Documento assinado digitalmente



FRANCISCO PETRONIO ALENCAR DE MEDEIROS

Data: 19/07/2022 10:45:20-0300

Verifique em <https://verificador.itl.br>

Prof. Dr. Francisco Petrônio Alencar de Medeiros – IFPB
Avaliador

Prof. Dr. Lafayette Batista Melo – IFPB
Avaliador

Prof. Dr. Patrícia Cabral de Azevedo Restelli Tedesco – UFPE
Avaliador Externo

DAMIRES YLUSKA DE SOUZA
FERNANDES:69208581420

Assinado de forma digital por DAMIRES
YLUSKA DE SOUZA FERNANDES:69208581420
Dados: 2022.07.19 13:53:49 -03'00'

Prof. Dr. Damires Yluska de Souza Fernandes – IFPB
(Orientadora)

Visto e permitida a impressão
João Pessoa

Prof. Dr. Francisco Petrônio A. de Medeiros (Coordenador PPGTI)

Ninguém caminha sem aprender a caminhar, sem aprender a fazer o caminho caminhando, refazendo e retocando o sonho pelo qual se pôs a caminhar.

Paulo Freire

AGRADECIMENTOS

Agradeço a toda minha família, meus amigos e colegas. A todos os professores do Programa de Mestrado em Tecnologia da Informação - PPGTI do IFPB, em especial à minha professora e orientadora Dr. Damires Yluska. de Souza Fernandes e ao professor e coordenador do curso Francisco Petrônio A. de Medeiros. O presente trabalho foi realizado com apoio do Instituto Federal da Paraíba (IFPB). O autor agradece o apoio por meio dos dados disponibilizados por Souza et al., (2020) utilizados na experimentação da abordagem.

RESUMO

Diante do atual panorama educacional brasileiro, de altos índices de reprovação e de ainda recorrentes práticas tradicionais associadas à pedagogia bancária, alternativas baseadas em práticas ativas de aprendizagem têm se revelado poderosas ferramentas de engajamento, interatividade e até mesmo de melhoria do rendimento escolar de estudantes. Isso ocorre também em cenários onde uma turma de estudantes se caracteriza por sua heterogeneidade de perfis, seja em termos de aspectos cognitivos, interacionais ou outros. Nessa perspectiva, este trabalho propõe uma abordagem, denominada ActivePlan, que objetiva identificar os perfis cognitivos e interacionais de estudantes em uma turma e, com base neles, formar grupos de estudantes com perfis heterogêneos de modo que os grupos possam ser utilizados em Práticas Ativas de Aprendizagem (neste trabalho, denominadas de PAAs). A ideia geral da abordagem é: (i) coletar dados de turmas de estudantes e enriquecê-los com alguns indicadores conforme uma PAA em questão; (ii) agrupar estudantes por similaridade de perfis, e utilizar as categorias de perfis similares identificadas para formar novos grupos, desta vez, com perfis heterogêneos de estudantes com vistas à aplicação de uma determinada PAA. A abordagem usa um método de aprendizado de máquina não supervisionado e propõe um algoritmo adaptado para a formação de grupos heterogêneos de estudantes para PAAs. Uma ferramenta foi implementada para a abordagem ActivePlan. A abordagem foi avaliada experimentalmente com base em dados reais de estudantes em turmas do ensino técnico integrado ao médio, considerando as PAAs denominadas Trezentos e Aprendizagem baseada em Equipes. Por meio dos perfis dos estudantes e do enriquecimento baseado em seus níveis de desempenho, grupos com perfis mistos de estudantes compatíveis com a prática Trezentos e com a prática Aprendizagem baseada em Equipes foram gerados. O valor obtido para a métrica de avaliação “diversidade” para os grupos formados indica uma composição, de fato, heterogênea no tocante aos perfis esperados de estudantes às práticas consideradas.

Palavras-Chave: Aprendizagem ativa; Grupos heterogêneos de estudantes; Suporte docente; Taxonomia de Bloom; Aprendizado de máquina não supervisionado; Agrupamento.

ABSTRACT

The Brazilian educational scenario is still characterized by high failure rates and, sometimes, recurrent traditional practices associated with what has been called “banking pedagogy”. To help matters, active learning practices have been considered. They have been accomplished as tools for enhancing engagement, interactivity and also school performance of students. Some active learning practices are designed to be used with heterogeneous student classes. In this light, this work proposes an approach, called ActivePlan, which aims to identify the cognitive and interactional profiles of students in a given class and, based on these profiles, to form heterogeneous groups of students. The general idea of the approach is organized by means of three aspects: (i) to collect data from a class of students and enrich them with some indicators according to a PAA at hand; (ii) to cluster students at first by similar cognitive and interactional profiles and (iii) to use the identified similar profile categories to form new groups, this time, with heterogeneous student profiles in order to apply a particular PAA. The approach makes use of an unsupervised machine learning method and proposes an algorithm adapted for the formation of heterogeneous groups of students for PAAs. A tool was implemented for the ActivePlan approach. The approach was experimentally evaluated based on real data from students in classes of integrated technical education to high school, considering the PAAs called Three Hundred and Team-based Learning. Through student profiles and enrichment based on their performance levels, groups with mixed profiles of students compatible with the Three Hundred practice and the Team-based Learning practice were generated. The value obtained for the evaluation metric “diversity” for the groups formed indicates actually heterogeneous compositions regarding expected profiles of students for the considered PAAs.

Keywords: Active learning; Heterogeneous groups; Teaching support; Bloom's taxonomy; Unsupervised Machine learning; Clustering.

LISTA DE FIGURAS

Figura 1 – Alguns benefícios do uso de metodologias ativas	23
Figura 2 – Etapas do processo CRISP-EDM.....	34
Figura 3 – Métodos hierárquicos e exemplo de dendrograma	38
Figura 4 – Exemplo de representação do método demográfico.....	39
Figura 5 – K-means com k igual a 3 (3 clusters)	40
Figura 6 – SSE, método Elbow e a quantidade de clusters.....	41
Figura 7 – Exemplo de Média do Coeficiente de silhueta.....	44
Figura 8 – Instâncias em um espaço bidimensional.....	45
Figura 9 – Matriz de distância.....	45
Figura 10 –Visão geral da proposta do Framework Conceitual para Formação de Grupos ..	46
Figura 11 –Formação de clusters e o nível de aprendizagem a partir de exercícios e da TB...48	
Figura 12 –Processo de formação de grupos heterogêneos.....	49
Figura 13 –Processo realizado em duas etapas para formação de grupos heterogêneos.....	50
Figura 14–Gráfico de densidade das médias dos atributos em cada cluster do curso de Pedagogia.	52
Figura 15 –Exemplo de questão rotulada ao nível “um” da TB.....	56
Figura 16 – Atividades da Etapa 2 da Abordagem ActivePlan.....	57
Figura 17– Algorithm_1 - Activeplan_main	63
Figura 18– Algorithm_2 - Algoritmo de geração de grupos para PAA.....	64
Figura 19 –Diagrama de atividades da ferramenta ActivePlan para o perfil professor.....	67
Figura 20– Fragmento do conjunto de dados de estudantes	68
Figura 21a–Tela principal da interface da ActivePlan.....	68
Figura 21b–Opções para o módulo do professor.....	68
Figura 22a–Número de clusters ideal mediante Elbow - Experimento 1.1.....	74
Figura 22b–Quantitativo de estudantes por cluster - Experimento 1.1.....	74
Figura 23–Métrica Silhouette - Experimento 1.1.....	75
Figura 24a–Número de clusters ideal mediante Elbow - Experimento 1.2.....	76
Figura 24b–Quantitativo de estudantes por cluster - Experimento 1.2.....	76
Figura 25–Métrica Silhouette - Experimento 1.2.....	77
Figura 26–Resultado da clusterização de estudantes associados ao desempenho - Experimento	81

Figura 27–Resultados da métrica de diversidade para cada grupo quanto à PAA Trezentos - Experimento 2.1.....	81
Figura 28–Resultados da métrica de diversidade para cada grupo quanto à PAA ABE - Experimento 2.1.....	83
Figura 29–Resultado da clusterização de estudantes associados ao desempenho. Experimento2.7.....	83
Figura 30–Resultados da métrica de diversidade para cada grupo quanto à PAA Trezentos. Experimento 2.2.....	83
Figura 31–Resultados da métrica de diversidade para cada grupo quanto à PAA ABE. Experimento 2.2.....	85

LISTA DE TABELAS

Tabela 1 – Exemplo de formação de grupos heterogêneos para a PAA Trezentos.....	30
Tabela 2 – Exemplo de identificação de perfil do cluster conforme a média dos atributos....	60
Tabela 3 – Metadados estruturais do dataset original juntamente com metadados obtidos após enriquecimento.....	70
Tabela 4 – Metadados estruturais do dataset original juntamente com médias de notas anteriores (1o ano) de turmas do segundo e terceiro ano.....	72
Tabela 5 – Amostra de resultados da atribuição de um estudante de cada cluster - experimento 1.1.....	75
Tabela 6 – Amostra de resultados da atribuição de um estudante de cada cluster - experimento 1.2.....	77
Tabela 7 – Ranking dos clusters a partir das médias de cada atributo - experimento 2.1.....	79
Tabela 8 – Ranking dos clusters a partir das médias de cada atributo - experimento 2.2.....	84

LISTA DE QUADROS

Quadro 1 – Competências desejáveis aos profissionais docentes conforme DCNs.....	18
Quadro 2 – Método tradicional e a utilização de método ativo a partir de uma perspectiva dialógica.....	25
Quadro 3 – Ações realizadas para aplicação das PAAs Trezentos e ABE.....	30
Quadro 4 – Resumo dos níveis da Taxonomia de Bloom quanto ao domínio cognitivo.....	33
Quadro 5 – Síntese dos estudos relacionados à pesquisa.....	53
Quadro 6 – Síntese dos atributos Considerados nos estudos.....	54
Quadro 7 – Amostra de estudantes em grupos heterogêneos turmas de primeiro ano.....	89
Quadro 8 – Amostra de estudantes em grupos heterogêneos turmas do segundo e terceiro ano	89

LISTA DE ABREVIATURAS E SIGLAS

MDE Mineração de Dados Educacionais

PAA Práticas ativas de aprendizagem

ML Machine learning

MD Mineração de dados

DM data mining

AM Aprendizado de máquina

TB Taxonomia de Bloom

CRISP-EDM Cross Industry Standard Process for Data Mining

CRISP-DM Cross Industry Standard Process for Educational Data Mining

KDD Knowledge Discovery in Database

AVA Ambiente Virtual de Aprendizagem

MOODLE Modular Object-Oriented Dynamic Learning Environment

SUAP Sistema Unificado da Administração Pública

SPAЕ Sistema de Processo Administrativo Eletrônico

UML Unified Modeling Language

Sumário

1 INTRODUÇÃO	17
1.1 MOTIVAÇÃO	17
1.2 QUESTÕES DE PESQUISA E PROPOSTA DE SOLUÇÃO	21
1.3 OBJETIVOS.....	21
1.4 ESCOPO DO ESTUDO.....	22
1.5 ORGANIZAÇÃO DA DISSERTAÇÃO	22
2 REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS.....	23
2.1 PRÁTICAS ATIVAS DE APRENDIZAGEM.....	23
2.1.1. <i>Método tradicional x Método ativo</i>	24
2.1.2 <i>Formação de grupos de estudantes e PAA</i>	27
2.1.2.1 <i>Estratégias para formação de grupos em ambiente escolar</i>	28
2.1.2.2 <i>Formação de grupos para Práticas ativas de aprendizagem (PAAs)</i>	30
2.2 TAXONOMIA DE BLOOM.....	32
2.3 MINERAÇÃO DE DADOS EDUCACIONAIS.....	33
2.3.1 <i>Modelo de processo para EDM</i>	34
2.3.2 <i>Aprendizado de Máquina</i>	36
2.3.3 <i>Método de Agrupamento</i>	38
2.4 AVALIAÇÃO DE GRUPOS DE ESTUDANTES.....	42
2.4.1. <i>Avaliação de grupos homogêneos</i>	43
2.4.2. <i>Avaliação de grupos heterogêneos</i>	44
2.5 TRABALHOS RELACIONADOS.....	46
2.5.1. <i>Framework Conceitual para Formação de Grupos de Alunos utilizando Trilhas de Aprendizagem em um Ambiente Virtual de Aprendizagem</i>	46
2.5.2. <i>Mineração de dados na identificação de grupos de estudantes com dificuldades de aprendizagem no ensino de programação</i>	47
2.5.3. <i>Comparative Analysis of Clustering Algorithms and Moodle Plugin for Creation of Student Heterogeneous Groups in Online University Courses</i>	49
2.5.4. <i>Using machine learning techniques to support group formation in an online collaborative learning environment</i>	50
2.5.5. <i>Identificação de perfis de interação de estudantes de educação a distância por meio de técnicas de agrupamentos</i>	51
2.5.6. <i>Síntese dos trabalhos relacionados</i>	53
3 ABORDAGEM ACTIVEPLAN	56
3.1 COLETA E PREPARAÇÃO DOS DADOS DOS ESTUDANTES.....	56
3.1.1. <i>Indicadores associados ao nível da TB e ao CRE</i>	57
3.2 FORMAÇÃO DE GRUPOS COM PERFIS HETEROGÊNEOS DE ESTUDANTES PARA PAAS.....	58
3.2.1. <i>Geração de grupos similares por meio do K-Means</i>	59
3.2.2. <i>Associação dos clusters a perfis de estudantes</i>	59
3.2.3. <i>PAA e quantidade de membros por grupos</i>	62
3.2.4. <i>Geração de grupos heterogêneos para PAA</i>	63
3.3 ACTIVEPLAN VERSUS TRABALHOS RELACIONADOS.....	66
4 IMPLEMENTAÇÃO E EXPERIMENTOS	67
4.1 FERRAMENTA ACTIVEPLAN.....	67
4.2 FONTES E CONJUNTOS DE DADOS	70

4.2.1 Conjunto de dados 1 - Estudantes de turmas de primeiro ano pertencentes à disciplina de Informática Básica	70
4.2.2 Conjunto de dados 2 - Estudantes do segundo e terceiro ano	72
4.3 EXPERIMENTOS.....	73
4.3.1 Experimento 1- Clusters de estudantes.....	74
4.3.1.1 Cenário 1.1 - Avaliação de Clusters de estudantes para turmas de primeiro ano	74
4.3.1.2 Cenário 1.2 - Clusters de estudantes das turmas de segundo e terceiro ano	77
4.3.2 Experimento 2 - Grupos heterogêneos de estudantes para PAAs.....	79
4.3.2.1 Experimento 2.1 - Grupos heterogêneos de estudantes resultantes da turma de primeiro ano	79
4.3.2.2 Experimento 2.2 - Grupos heterogêneos de estudantes resultantes da turma de segundo e terceiro ano	84
4.3.3 Discussão geral sobre os resultados dos experimentos	88
5 CONSIDERAÇÕES E TRABALHOS FUTUROS	90
REFERÊNCIAS BIBLIOGRÁFICAS	91
APÊNDICE A – TABELA COMPARATIVA DOS TRABALHOS RELACIONADOS E A ABORDAGEM PROPOSTA.....	102
APÊNDICE B – AVALIAÇÃO DE MÉTODOS DE CLUSTERIZAÇÃO.....	104
ANEXO A – ALGORITMO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS (MAINA ET AL., 2017).....	105
ANEXO B – AVAS E GERAÇÃO DE GRUPOS DE ESTUDANTES.....	106

1 INTRODUÇÃO

Este capítulo mostra o contexto da pesquisa, apresentando a motivação, justificativa, questões de pesquisa, objetivos do trabalho e a estrutura deste documento.

1.1 Motivação

O professor é peça-chave para a modificação, para essa melhoria, seria entender o contexto no qual os professores estão inseridos, bem como as práticas pertinentes à atividade docente, a saber:

Bons professores são as peças-chave na mudança educacional. Os professores têm muito mais liberdade e opções do que parece. A educação não evolui com professores mal preparados. Muitos começam a lecionar sem uma formação adequada, principalmente do ponto de vista pedagógico. Conhecem o conteúdo, mas não sabem como gerenciar uma classe, como motivar diferentes alunos, ou que dinâmicas utilizar para facilitar a aprendizagem (MORAN, 2012)

Moran (2012) afirma que os professores começam a lecionar, muitas vezes, sem o devido preparo e, portanto, em muitos casos, desconhecem as estratégias pedagógicas que podem auxiliá-los na condução das atividades docentes. Oliveira (2004), Abonízio (2012), Santos (2021), Farias et al. (2020), Barros et al. (2021) e Moura et al. (2019) também relatam as condições de trabalho e a precarização quanto à atuação desse profissional. Na prática, muitos professores necessitam de jornadas múltiplas em várias escolas, muitas vezes, sem vínculo efetivo, o que agrava ainda mais todo este quadro.

Nesse panorama, o cenário da educação brasileira está repleto de problemas e desafios. Freire (1997), Blikstein (2012) e Pachane et al. (2004) revelam e evidenciam a realidade na qual docentes acabam ainda muito voltados a práticas tradicionais e, isso resulta, muitas vezes, em insatisfação por parte dos estudantes. Tal fato remete, em parte, à abordagem metodológica aplicada em sala de aula. Comumente, nesse cenário, presencia-se ainda o ensino uniforme e a denominada pedagogia bancária mencionada por Freire (1997). O termo “pedagogia bancária” remete à prática docente centrada no conteúdo, quando ocorre a transmissão e depósito de informações e conhecimento no estudante (Freire, 1997). Existem, nesse cenário, inúmeros desafios, por exemplo, a necessidade de condução do ensino conforme às diferentes características ou perfis dos estudantes presentes em uma turma.

Outro problema refere-se à sobrecarga de trabalho oriunda da prática docente e do processo de ensino e aprendizagem, em especial, na condução de turmas com quantidade expressiva de estudantes (e.g., a complexidade de lidar com turmas com quantitativo de 250 ou mais estudantes) (FRAGELLI, 2019). Turmas grandes podem ocasionar ainda mais o isolamento dos estudantes, algumas vezes com exclusão de parcela dos estudantes como, por exemplo, estudantes com ritmos de aprendizagem mais baixos. Estudantes com esse perfil apresentam defasagem e necessitam de auxílio mais efetivo e direto do docente para alcançar os objetivos educacionais esperados.

Neste sentido, sabe-se que os estudantes aprendem de formas diferentes e em ritmos distintos (MORAN, 2015). A heterogeneidade nos perfis dos estudantes, também apontada por Perrenoud (2000) e Perrenoud (2015), em concordância com Moran (2015), revela a necessidade de uma abordagem diferenciada de enfrentamento nas práticas de ensino e aprendizagem. Compactuando com essa perspectiva, em 2019, as Diretrizes Curriculares Nacionais (DCNs) definiram competências desejáveis a profissionais docentes. Dentre as competências, no que tange à heterogeneidade de estudantes e à prática docente, destacam-se (Quadro 1): (i) a capacidade do professor de demonstrar conhecimento sobre os estudantes e como eles aprendem; e (ii) o planejamento de ações de ensino que resultem em aprendizagem efetiva.

Quadro 1 - Competências desejáveis aos profissionais docentes conforme DCNs

Capacidade do professor de demonstrar conhecimento sobre os estudantes e como eles aprendem	Planejamento de ações de ensino que resultem em aprendizagem efetiva
<ul style="list-style-type: none"> ● Adotar um repertório adequado de estratégias de ensino e atividades didáticas orientadas para uma aprendizagem ativa e centrada no estudante. ● Articular estratégias e conhecimentos que permitam aos estudantes desenvolver as competências necessárias, bem como desenvolver as habilidades de níveis cognitivos superiores. ● Conhecer os contextos de vida dos estudantes, reconhecer suas identidades e elaborar estratégias para contextualizar o processo de aprendizagem. ● Aplicar estratégias de ensino diferenciadas que promovam a aprendizagem dos estudantes com diferentes necessidades e deficiências, levando em conta seus diversos contextos culturais, socioeconômicos e linguísticos. 	<ul style="list-style-type: none"> ● Adotar um repertório diversificado de estratégias didático pedagógicas considerando a heterogeneidade dos estudantes (contexto, características e conhecimentos prévios).

Fonte: Adaptado de Brasil (2019).

Os pontos listados no Quadro 1 remetem à necessidade do reconhecimento do contexto do estudante, bem como recomendam a utilização de estratégias de aprendizagem

centradas no estudante. Nesse sentido, os pontos mostrados corroboram com a possibilidade de adoção de metodologias ativas de aprendizagem. Para isso, é importante auxiliar o docente nessa condução e, algumas vezes, no reconhecimento da heterogeneidade dos perfis dos estudantes de uma turma de modo que isso possa ser usado nas estratégias metodológicas.

Um dos autores que discorrem sobre isso é Perrenoud (2000) que menciona a necessidade de inventabilidade didática do docente em cenários heterogêneos. Nesse sentido, o contexto do estudante e a heterogeneidade presente em sala de aula podem ser aproveitados mediante a adoção de metodologias ativas de aprendizagem (Mazur, 2015). Algumas metodologias ativas como, por exemplo, Sala de Aula Invertida (Bacich et al., 2018), Instrução por Pares (Araujo e Mazur, 2015), Trezentos (Fragelli, 2019), entre outras práticas, aparecem como propostas diferenciadas para dinamizar o ambiente escolar e auxiliar o docente na condução do ensino em turmas com perfis heterogêneos de estudantes. Neste trabalho, utiliza-se o termo *Práticas Ativas de Aprendizagem (PAA)* na referência a métodos, técnicas, abordagens, estratégias de ensino centradas nos estudantes com o intuito de auxiliar a condução da prática docente diante de um cenário heterogêneo de turma.

Para isso, Moran (2015) remete à importância do reconhecimento prévio dos perfis de estudantes com base em seus conhecimentos e competências já existentes. Soma-se a isso a ideia de interação social e de mediação como ponto central do processo educativo (Vigotsky, 1984). Portanto, percebe-se a importância da identificação dos perfis dos estudantes, o potencial com relação à adoção de PAAs e a necessidade de formação de grupos para interação discente como meio de auxiliar nas práticas docentes.

Nessa perspectiva, em tempos recentes, muitas dessas práticas têm sido apoiadas ou mesmo implementadas por meio de Tecnologias Digitais de Informação e Comunicação (TDIC). Em um contexto de ensino sempre desafiador, seja em modo presencial ou remoto, para que os professores desempenhem seu papel de agente transformador, ferramentas são cada vez mais necessárias. O ferramental por meio de TDICs cresceu devido à modalidade de Educação a Distância (EaD) e, mais recentemente, devido à necessidade de reorganização do ensino para um modelo remoto, especialmente por causa da pandemia da COVID-19.

Importante lembrar que a Educação a Distância (EaD) mediatizada por TDICs vem sendo utilizada de forma crescente por instituições públicas e privadas em cursos diversos de graduação e pós-graduação (RODRIGUES, 2020). Particularmente, em relação a cursos de graduação, segundo dados do Censo da Educação Superior (INEP, 2020), no cenário

educacional brasileiro, percebe-se que gradativamente o número de ingressos e de matrículas em cursos EaD no ensino superior vem crescendo na última década.

Com a realidade trazida pela pandemia de COVID-19 no Brasil, as discussões sobre EaD e ensino remoto têm ocupado a cena e recebido maior destaque na área da educação. Nesse panorama, é importante registrar a diferença entre EaD e atividades remotas pela internet. Na EaD, é definido e aplicado um modelo subjacente de educação, desde o planejamento até a execução de um determinado curso ou disciplina. Esse modelo ampara as escolhas pedagógicas e organiza os processos de ensino e de aprendizagem. Assim, concepções teóricas, fundamentos metodológicos e especificidades são definidos e sustentam, teórica e praticamente, a modalidade EaD (Rodrigues, 2020).

O ensino remoto, por sua vez, vem sendo buscado ultimamente e pode ser definido como um formato de escolarização caracterizado pelas condições de distanciamento entre professor e estudante utilizando-se de tecnologias digitais (Morais et al., 2020). O ensino remoto diferencia-se da EaD, pelas razões apresentadas anteriormente, ou seja, na EaD, existe um modelo subjacente que ampara a definição de papéis por meio de atores (professor, tutores), com concepção e uso de materiais específicos pensados para um dado curso.

Seja em modalidade EAD, seja em modalidade remota ou mesmo híbrida de ensino, o ferramental de apoio ao docente inclui o uso de um Ambiente Virtual de Aprendizagem (AVA) (Morais et al., 2020). Por meio de AVAs, é possível coletar dados de estudantes incluindo dados de interações, o que torna esses ambientes propícios à identificação de perfis de estudantes.

A situação provocada pela pandemia da COVID-19 revelou ainda mais a necessidade de repensar o modelo de ensino-aprendizagem para estratégias cada vez mais ativas e com base em modelos de ensino que utilizem formatos remotos ou híbridos. Para Lima et al. (2022), essa situação pode, dado seu caráter inovador, flexível e com uma boa relação custo-benefício, tornar os alunos protagonistas do seu processo de ensino-aprendizagem, influenciando positivamente em seu desempenho acadêmico. Ou seja, há uma tendência que muitas das atividades de ensino possam ser tanto presenciais quanto remotas, dependendo da necessidade e das estratégias em pauta (SCHIEHL e GASPARINI, 2017), mas isso pode ser fortemente amparado a partir de práticas ativas de aprendizagem, de acordo com Moran (2015), e pode auxiliar no processo de aprendizagem.

1.2 Questões de pesquisa e proposta de solução

Frente ao panorama apresentado, originam-se as questões de pesquisa que norteiam este trabalho:

Q1: Diante dos diferentes perfis cognitivos e interacionais dos estudantes em uma turma, seja em ensino remoto ou híbrido, como formar ou agrupar perfis similares de estudantes?

Q2: A partir dos grupos de estudantes com perfis similares identificados, como formar novos grupos com perfis heterogêneos de estudantes e associar esses grupos a PAAs?

A partir das questões de pesquisa norteadoras, este estudo propõe uma abordagem, denominada ActivePlan, para formação de grupos de estudantes com perfis heterogêneos que possam ser associados a algumas PAAs. O perfil dos estudantes leva em conta aspectos cognitivos e interacionais.

1.3 Objetivos

Este trabalho tem como objetivo geral propor e desenvolver uma abordagem computacional (ActivePlan) para suporte ao docente na formação de grupos de estudantes com perfis heterogêneos com vistas a utilização de algumas PAA em um contexto de ensino remoto.

O trabalho conta com os seguintes objetivos específicos:

- Propor e desenvolver estratégia para identificação e agrupamento de perfis de estudantes a partir de dados coletados em AVAs ou em outros sistemas acadêmicos;
- Propor e desenvolver estratégia para formação de grupos de estudantes com perfis heterogêneos para adoção de PAA;
- Desenvolver ferramenta que auxilie na automatização da abordagem proposta e forneça: (i) visualização e identificação dos diferentes perfis de estudantes, e (ii) sugestão de organização de grupos para uma PAA a ser escolhida pelo professor;
- Coletar e selecionar dados relativos a perfis de estudantes em atividade de ensino remoto e utilizá-los para avaliar a abordagem desenvolvida;
- Avaliar os resultados da abordagem proposta e desenvolvida, de modo a responder às questões de pesquisa elencadas.

1.4 Escopo do estudo

O estudo está inserido no âmbito da Informática na Educação, visando-se, prioritariamente, contribuir a partir da elaboração de uma abordagem computacional para o suporte docente na utilização de PAA. Ressalta-se que o estudo não engloba a busca pela otimização, comparação ou eficiência de algoritmos para o processamento dos agrupamentos gerados. Porém, apresenta uma avaliação dos resultados obtidos com o desenvolvimento da abordagem e geração de grupos.

1.5 Organização da dissertação

Os capítulos subsequentes estão organizados da seguinte maneira: no Capítulo 2, abordam-se o referencial teórico e os trabalhos relacionados a esta dissertação.

O Capítulo 3 destina-se à apresentação e descrição da abordagem proposta.

No Capítulo 4, descrevem-se e discutem-se os resultados alcançados na implementação e na avaliação experimental.

Finalmente, o Capítulo 5 exhibe as conclusões e tece comentários acerca de trabalhos futuros.

O Apêndice A mostra um quadro comparativo entre os trabalhos relacionados e a abordagem ActivePlan. O Apêndice B apresenta um quadro comparativo entre os algoritmos de clusterização testados para uso pela ActivePlan.

O Anexo AI apresenta o algoritmo para formação de grupos heterogêneos proposto por (Maina et al., 2017), e o Anexo B exhibe um quadro indicando alguns mecanismos de formação de grupos para AVAs.

2 Referencial teórico e trabalhos relacionados

Nesta seção, são introduzidos conceitos e descritos alguns estudos relacionados à temática deste trabalho. O aporte teórico utilizado envolve, predominantemente, aspectos da área de Educação e da Ciência da Computação voltados ao entendimento desta dissertação.

2.1 Práticas ativas de aprendizagem

A aprendizagem ativa sempre esteve presente em nossas vidas desde o momento em que nascemos. Segundo Bacich et al., (2018) o ser humano está em constante aprendizado. Vivenciam-se a todo instante experiências que resultam no aprendizado. Tal fato processa-se na tentativa de solucionar problemas reais dos mais simples aos mais complexos que relacionam-se com escopos diversos como aqueles de cunho social, profissional ou pessoal. Assim ressalta-se que:

Num sentido amplo, toda a aprendizagem é ativa em algum grau, porque exige do aprendiz e do docente formas diferentes de movimentação interna e externa, de motivação, seleção, interpretação, comparação, avaliação, aplicação. (MORAN, 2018)

Nessa perspectiva, a aprendizagem ativa, conforme Moran (2018), remete à exigência e participação do aprendiz como protagonista do aprendizado. Para Berbel (2012), no tocante à sala de aula e ao processo de ensino e aprendizagem, as estratégias metodológicas adotadas pelo docente mostram-se relevantes e podem despertar, por exemplo, a curiosidade e interesse do estudante, podendo contribuir para que o discente tenha um perfil mais ativo.

São apontados alguns benefícios relacionados à aplicação de metodologias ativas na Figura 1, conforme Venturini e Silva (2018, p.69).

Figura 1 -Alguns benefícios do uso de metodologias ativas



Fonte: Elaborado pelo autor com base em Venturini e Silva (2018, p. 69)

Eles pontuam alguns benefícios no processo de ensino e aprendizagem com a adoção de metodologias ativas, como: Maior retenção de informação ao longo do tempo, mais engajamento, melhor relacionamento com colegas, melhor expressão oral, melhoria de hábitos de estudo dos estudantes, melhora de desempenho do estudante, desenvolvimento de habilidades para resolver problemas reais e maior satisfação em assistir às aulas.

Neste trabalho utilizaremos a definição de metodologia ativa de aprendizagem segundo Pereira (2012) e Bacich et al. (2018, p.15) que a definem como:

- “Qualquer processo de organização da aprendizagem, estratégias didáticas, que incluam, efetivamente, o estudante no centro do processo” (Pereira, 2012,p.6).
- “Estratégias de ensino centradas na participação efetiva dos estudantes na construção do processo de aprendizagem, de forma flexível, interligada e híbrida” (Bacich et al., 2018, p.15).

De modo geral, nesta dissertação, usa-se o termo Prática Ativa de Aprendizagem (PAA) para referenciar métodos, técnicas, abordagens, estratégias de ensino centradas nos estudantes com o intuito de auxiliar a condução da prática docente diante de um cenário heterogêneo de turma.

Para melhor embasamento da temática associada a práticas ativas de aprendizagem, a Seção 2.1.1 discorre brevemente sobre o método tradicional e o método ativo de aprendizagem. A Seção 2.1.2 aborda a formação de grupos de estudantes e PAA.

2.1.1. Método tradicional x Método ativo

O denominado método de ensino tradicional insere o professor no centro do processo de ensino aprendizagem (Berbel, 2012). Nesse método o professor é o detentor da informação e do conhecimento que transfere informações aos estudantes. Estes, por sua vez, normalmente, possuem um papel passivo nesse processo. Tal cenário mostrou-se condizente há algumas décadas já que a informação era algo de difícil acesso (Morán, 2015). Nesse sentido, na concepção de Freire (1997), evidenciam-se problemas no ensino tradicional e na recorrente aplicação da pedagogia bancária em vez da adoção de uma perspectiva dialógica. O Quadro 2 sintetiza algumas das diferenças entre o método tradicional e o método ativo a partir de uma perspectiva dialógica.

Quadro 2 - Método tradicional e a utilização de método ativo a partir de uma perspectiva dialógica.

Característica	Metodologia tradicional	Metodologia ativa a partir de perspectiva dialógica
Papel do estudante no processo de ensino e aprendizagem	<ul style="list-style-type: none"> ● Passivo na maior parte do processo. 	<ul style="list-style-type: none"> ● O estudante é levado a participar do processo de ensino e aprendizagem. ● Centrado no estudante. ● Considera que todo estudante pode contribuir com o processo de ensino e aprendizagem.
Papel do professor no processo de ensino e aprendizagem	<ul style="list-style-type: none"> ● É considerado o único detentor do saber. ● Centraliza todo processo de ensino e aprendizagem 	<ul style="list-style-type: none"> ● Mediador e articulador do processo de ensino e aprendizagem.
Recursos	<ul style="list-style-type: none"> ● Geralmente utiliza-se de lousa e slides 	<ul style="list-style-type: none"> ● Vasta utilização de TICs^{5*}
Prática docente	<ul style="list-style-type: none"> ● Mais rígida e inflexível; ● Educação Bancária; ● Transmissão de conhecimentos e ampliação de informações; ● Depósitos de conteúdos, que são objetos do professor; ● Não há estímulo para criação; ● Educação verticalizada; ● Considera-se a realidade do professor; ● Pode não haver mudanças de hábitos e comportamentos; ● Proporciona menor autonomia ao estudante. 	<ul style="list-style-type: none"> ● Flexibilidade nos espaços e tempo; ● Educação problematizadora; ● Construção de conhecimentos e competências; ● Estudante portador de um saber, objeto de uma ação educativa e sujeito da própria educação; ● Aprende-se por meio de uma prática reflexiva; seres ativos; ● Educação baseada no diálogo; ● Considera-se a realidade do educando; mudanças duradouras de hábitos e comportamentos; ● Proporciona maior autonomia ao estudante; ● Geralmente inclui atividades, desafios e tarefas realizadas em grupos.
Processo avaliativo	<ul style="list-style-type: none"> ● Avaliação no final da disciplina 	<ul style="list-style-type: none"> ● Geralmente ocorre no formato contínuo ao longo da disciplina; considera a evolução do estudante durante o processo.
Benefícios	<ul style="list-style-type: none"> ● Mais controle do cronograma da disciplina, visto que não existe uma preocupação contínua com a aprendizagem. ● Amplia informações e conhecimentos já existentes; ● Produz aquisição de conhecimento. 	<ul style="list-style-type: none"> ● Construção coletiva do conhecimento; ● Proporciona ao estudante visão crítica e reflexiva da realidade; ● Capacita o estudante para tomada de decisões; ● Promove mais autonomia no estudante; ● Promove mais engajamento no estudante; ● Proporciona feedbacks mais rapidamente, devido à mediação e ao acompanhamento dos estudantes.

Fonte: Adaptado de Figueiredo et al. (2010)

Conforme sintetizado no Quadro 2, tendo alguns sido citados na seção anterior, existem muitos benefícios na aplicação de metodologias ativas em sala de aula no tocante à promoção de engajamento, interação, autonomia do estudante e feedbacks mais rápidos para o professor. Outra vantagem da adoção de metodologias ativas, segundo Silberman (1996), relaciona-se à retenção da maior quantidade de conteúdo e de informação do assunto abordado. Deste modo, os estudantes compreendem mais por mais tempo, o que acarreta em maior satisfação e prazer no processo de ensino e aprendizagem. Além das vantagens e benefícios da aprendizagem ativa citados, percebe-se que, quando aliada ao reconhecimento do contexto no qual os estudantes estão inseridos, a:

Aprendizagem mais profunda requer espaços de prática frequentes (aprender fazendo) e de ambientes ricos em oportunidades. Por isso, é importante o estímulo multissensorial e a valorização dos conhecimentos prévios dos estudantes para “ancorar” os novos conhecimentos. (BACICH et al., 2018, p.?).

Deste modo, Bacich et al. (2018) e Moran (2015) estão consonantes na definição da importância, no reconhecimento dos conhecimentos prévios dos estudantes e na aplicação das metodologias ativas em ambientes escolares. As metodologias ativas enfatizam a participação e promovem engajamento do estudante no processo de ensino e aprendizagem. Isso pode ocorrer mediante mudança de algumas práticas tradicionais na condução da aula.

Neste sentido, o papel desempenhado pelo professor revela-se fundamental no processo de ensino e aprendizagem também em metodologias ativas (Morán, 2015). Este exerce um novo papel de mediador e articulador na condução do processo de ensino e aprendizagem. Nesse panorama, algumas estratégias revelam-se pertinentes para o sucesso da aprendizagem, as quais o professor pode utilizar, por exemplo:

a criação de desafios, atividades, jogos que realmente trazem as competências necessárias para cada etapa, que solicitam informações pertinentes, que oferecem recompensas estimulantes, que combinam percursos pessoais com participação significativa em grupos, que se inserem em plataformas adaptativas, que reconhecem cada aluno e, ao mesmo tempo, aprendem com a interação, tudo isso utilizando as tecnologias adequadas. (MORÁN, 2015).

Nesse contexto, analogamente a Moran (2012), BONWELL (1991, apud BARBOSA, 2013) menciona que o professor pode amparar-se constantemente em estratégias que buscam a utilização da aprendizagem ativa em sala de aula. Listam-se algumas estratégias, a saber:

- Discussão de temas e tópicos de interesse para a formação profissional, dependendo da modalidade de ensino;
- Trabalho em equipe com tarefas que exigem colaboração de todos;
- Estudo de casos relacionados com áreas de formação específica;
- Debates sobre temas da atualidade;
- Geração de ideias (brainstorming) para buscar a solução de um problema;
- Produção de mapas conceituais para esclarecer e aprofundar conceitos e ideias;

- Modelagem e simulação de processos e sistemas típicos da área de formação;
- Criação de sites ou redes sociais visando a aprendizagem cooperativa;
- Elaboração de questões de pesquisa na área científica e tecnológica.

Dessa forma, apesar de tantas estratégias ativas de aprendizagem existentes, ressalta-se ainda a adoção de práticas tradicionais caracterizadas pela “exclusividade da ação intelectual do professor e a representação do livro didático como fontes exclusivas do saber na sala de aula” (PEREIRA, 2012, p.6). A definição de Pereira (2012) e a de Bacich et al. (2018, p.15) enfatizam a necessidade de postura ativa do estudante no processo de ensino e aprendizagem. Pereira (2012) acrescenta o fator de oposição à metodologia tradicional que representa a unilateralidade do professor e exclusividade da utilização do livro. Nesta perspectiva, a definição de Pereira (2012) vislumbra a ideia das inúmeras possibilidades de aprendizagem, além da forma tradicional.

Revelam-se algumas dessas possibilidades a partir de tendências de PAAs como (Navarro, 2017): Solução de problemas, Sala de aula invertida e Instrução por pares. Apesar de muitos autores definirem simplesmente como uma metodologia, elas evoluíram ao ponto de inspirar a criação de outras metodologias ativas como, por exemplo, a instrução por pares inspirou o surgimento do método Trezentos.

2.1.2 Formação de grupos de estudantes e PAA

Para (Moran et al., 2015; Navarro, 2017 e Wommer, 2020) uma das características predominantes das metodologias ativas é a participação e interação dos estudantes de forma colaborativa e em grupo. Nesta perspectiva, existem inúmeras estratégias descritas na literatura de como formar grupos de estudantes para utilização em ambientes escolares. Algumas caracterizam-se pela praticidade e pecam na qualidade do grupo formado, quando não se verifica agregação de ganhos significativos no processo de ensino e aprendizagem. Por outro lado, há estratégias de formação de grupos que podem trazer mais benefícios no processo de ensino e aprendizagem, mais especificamente no que tange à colaboração e à interação dos estudantes. A Seção 2.1.4.1 apresenta algumas estratégias de âmbito global, comumente utilizadas em sala de aula para formação de grupos, enquanto a Seção 2.1.4.2 descreve algumas estratégias promissoras, focadas em PAA, para utilização em ambiente escolar.

2.1.2.1 Estratégias para formação de grupos em ambiente escolar

No que tange a estratégias gerais para Formação de Grupos de estudantes, frequentemente, o professor ou tutor o faz criando grupos da seguinte forma (Ounnas et al., 2009; HENRY, 2013; Maqtary et al. , 2019):

1. **Formados aleatoriamente:** Essa estratégia prevê a atribuição aleatória dos estudantes aos grupos, não existindo critério específico definido. Esta abordagem é geralmente usada para formar grupos informais temporários. É considerada prática, já que não exige do professor ou do tutor a identificação de perfis de estudante nem a atribuição mais coesa e eficiente dos membros de um grupo. Entretanto, um dos principais problemas da estratégia é o fato de criar grupos desbalanceados (Considerando perfis dos estudantes) já que os estudantes são alocados aleatoriamente aos grupos.
2. **Formados pelos próprios estudantes:** Também conhecido como estratégia auto-selecionada. Nesta, a formação de grupos de estudantes é delegada aos estudantes e tende a ocorrer mediante vínculos afetivos (irmão, namorados, primos, amigos muito próximos), o que compromete a composição diversificada dos membros de um grupo já que se baseia no interesse dos estudantes (Bollela et al., 2014). Dessa forma existe alta probabilidade da formação dos grupos ocorrer de forma desbalanceada.
3. **Formados a partir de seleção especificada pelo Professor/Tutor:** Nessa estratégia o professor ou tutor atribui os estudantes aos grupos. Dessa forma, pode-se aferir que é uma técnica mais assertiva de formação de grupos no que se refere ao balanceamento (grupo mais coeso), quando comparado às demais estratégias de formação de grupo. Quando a estratégia é realizada manualmente, gera-se mais sobrecarga de trabalho extra ao professor/tutor para organização e formação dos grupos. Tal fato agrava-se proporcionalmente à quantidade de estudantes da turma, tornando-se até mesmo inviável para muitos professores quando em turmas muito grandes.

Na estratégia de formação de grupos selecionados pelo professor ou tutor, comumente, necessita-se da identificação dos perfis de estudantes pelo professor/tutor visando-se alcançar os objetivos educacionais definidos em seu planejamento para a atividade desenvolvida. Dependendo do propósito, o professor pode selecionar grupos de três maneiras de acordo com (Ounnas et al., 2009; HENRY, 2013; Maqtary et al., 2019): similaridade, heterogeneidade (diversidade) ou ambas características dos perfis de estudantes.

No que tange à formação de grupos compostos de estudantes com **características/perfis similares**, Oliveira et al. (2016, p. 968, apud Oakley, Felder e Brent 2004) ressaltam que grupos formados somente de aprendizes com facilidade de aprendizagem tendem a gerar aprendizados solitários pois os estudantes tendem a realizar suas tarefas individualmente, ou seja, com pouca interação. Oliveira (op. cit.) menciona que, para equipes ou grupos compostos somente com perfis de aprendizes com dificuldades de aprendizagem, há uma tendência ao reforço de erros e conceitos distorcidos.

Analogamente grupos formados apenas com estudantes tímidos, ou mesmo a estratégia de formar grupos somente com perfis de estudantes extrovertidos tende a gerar grupos que apresentam comportamento de dispersão, com excesso de comunicação oriundo da interação dos membros do grupo. Outra alternativa na formatação referente à composição de grupos considera **características/perfis heterogêneos de estudantes**.

Neste sentido, diferentes perfis de estudantes interagem e colaboram entre si, o que para (Vygotsky,1984; Perrenoud, 2015; Moran 2015; Fragelli 2019) pode possibilitar um ambiente rico para aprendizagem, propiciando a interação e colaboração entre os estudantes.

Neste aspecto de formação de grupo, para Rosso et al. (2016), a utilização de grupos compostos de estudantes com características/perfis heterogêneos:

inicialmente a diversidade da equipe tende a inibir a expressão individual e a se confiar mais em um ou dois de seus membros, com o passar do tempo, à medida que o grupo se comunica mais e melhor, cada um fica mais à vontade para se expor e se expressar, e passa a confiar no potencial de cada membro para o melhor desempenho da equipe (Rosso et al., 2016, p. 604)

Rosso et al. (2016) acrescentam que os grupos compostos de perfis diversificados formados propiciam um ambiente rico de aprendizagem devido às diferentes percepções dos membros do grupo. Neste sentido, frequentemente, presencia-se o confronto de ideias mediante a análise, discussão ou reflexão até os membros chegarem em um consenso diante de uma questão/problema.

Por fim, os denominado de **grupos mistos**, ou seja, o grupo é composto de estudantes que são similares em algumas características e dissimilares em outras (Maqtary et al.,2019). Diante das estratégias de formação de grupos citadas, comumente os AVA dispõem de mecanismos para geração de grupos. Alguns mecanismos são mostrados no Anexo B. Por meio desses exemplos, percebe-se uma vasta presença de mecanismos de formação de grupos de estudantes, contudo não presencia-se formação de grupos voltados para PAA.

2.1.2.2 Formação de grupos para Práticas ativas de aprendizagem (PAAs)

Este trabalho foca em práticas ativas de aprendizagem que podem ser beneficiadas com a formação de grupos de estudantes com perfis heterogêneos. Neste contexto, existem na literatura, algumas PAAs que foram planejadas e pensadas para trabalhos com grupos de estudantes com perfis heterogêneos, a exemplo a metodologia Trezentos (TR). Outras PAA também apontam a importância que a composição de cada grupo seja formada por perfis heterogêneos, a exemplo das metodologias Testes conceituais (TC) e Aprendizagem baseada em equipe (ABE), do inglês Team-Based learning - TBL.

Em especial, o estudo realizado neste trabalho concentra-se, inicialmente, na aplicação das definições organizadas por meio da abordagem ActivePlan às PAAs denominadas Trezentos e ABE. Por esta razão, estas duas PAAs são descritas a seguir.

O Quadro 3 sintetiza o passo a passo das PAA Trezentos e ABE. Lista-se o passo a passo das PAA Trezentos utilizada no estudo de Fragelli (2019) e Castro et al. (2021), e da PAA ABE (Alcantara et al., 2020).

Quadro 3 –Ações realizadas para aplicação das PAAs Trezentos e ABE

Trezentos	ABE
1) Avaliam-se os estudantes mediante atividade;	1) Divide-se a turma em equipes baseando-se na heterogeneidade da turma;
2) Formam-se grupos de estudantes heterogêneos conforme o desempenho na atividade;	2) Realiza-se a garantia de Preparo mediante pré-leitura
3) Destina-se um período de estudo em grupo para refatoração da atividade;	3) Aplica-se um teste individual de garantia de aprendizagem;
4) Realiza-se a análise e avaliação do desempenho do grupo;	4) Aplica-se o teste coletivo;
5) Inicia-se novamente o processo, ocorre a reaplicação do método com formação de novos grupos.	5) Destina-se um período de tempo para apelos ou recursos e apresentação aos outros grupos;
	6) Intermedia-se mediante uma intervenção conclusiva e explicativa do professor;
	7) Avalia-se em pares.

Fonte: elaborado pelo autor com base em Fragelli (2019) e Alcantara et al. (2020).

Percebe-se no Quando 3 que ambas as metodologias apresentam foco inicial na formação dos grupos de forma heterogenia. A PAA Trezentos estabelece a composição de grupos baseada em um critério de rendimento, ou seja, considerando estudantes de bom rendimento e de baixo rendimento. Nesta estratégia, realiza-se a definição da quantidade de grupos dividindo-se o número total de estudantes da turma por cinco, arredondando esse número para baixo (Fragelli, 2019). Nessa prática, deve existir pelo menos um líder (ajudante) no grupo, cuja consideração para esse papel ocorre a partir da identificação de um

perfil de estudante com bom rendimento. Dessa forma, pelo menos 20% do total de estudantes deve atingir o mínimo satisfatório de rendimento para exercer o referido papel, conforme exemplo mostrado na Tabela 1. Quanto ao quantitativo de estudantes, considera-se o tamanho da turma, da seguinte forma: (i) Em turmas grandes, com mais de 20 estudantes, utiliza-se um quantitativo de 5 a 6 membros por grupo; (ii) No caso de turmas com até 20 estudantes, emprega-se um quantitativo de 3 a 4 membros por grupo. No exemplo apresentado na Tabela 1, foram considerados quatro grupos, onde cada grupo é composto de cinco membros. Em cada grupo, há um perfil de líder na sua composição, assim como existem perfis de membros ajudantes (Estudantes que possuem mais desempenho) e ajudados (Estudantes que possuem menos desempenho). Para a PAA Trezentos a estratégia desenvolvida sugere a atribuição do papel de ajudante aos estudantes que obtiverem desempenho satisfatório no aspecto cognitivo e o papel de ajudado aos que obtiverem menor desempenho. O desempenho é estimado por meio de atividade realizada no início da aplicação da PAA conforme exposto na Tabela 1.

Tabela 1 - Exemplo de formação de grupos heterogêneos para a PAA Trezentos

Estudante	Desempenho	Grupo	Papel
Cora Coralina	9,70	1	Ajudante (líder)
José de Alencar	6,40	1	Ajudante
Carlos Drummond	5,30	1	Ajudado
Akinwande Soyinka	3,10	1	Ajudado
Nikola Tesla	0,20	1	Ajudado
Toni Morrison	9,70	2	Ajudante (líder)
Machado de Assis	8,30	2	Ajudante
Castro Alves	5,80	2	Ajudado
Leonardo da Vinci	3,30	2	Ajudado
Gertrude Elion	2,30	2	Ajudado
Michael Faraday	9,50	3	Ajudante (líder)
Isaac Newton	8,40	3	Ajudante
Aristóteles	5,90	3	Ajudado
Thomas Alva Edison	4,10	3	Ajudado
Arquimedes	2,60	3	Ajudado
Galileu Galilei	9,30	4	Ajudante (líder)
René Descartes	8,80	4	Ajudante
James Maxwell	6,20	4	Ajudado
Albert Einstein	4,10	4	Ajudado
Marie Curie	2,70	4	Ajudado

Fonte: (Fragelli, 2019)

A Tabela 1 exemplifica a formação de grupos em uma turma sintética (com estudantes de nomes e desempenho fictícios) composta de vinte estudantes. Neste exemplo, percebe-se a necessidade da heterogeneidade de perfis com relação ao desempenho dos estudantes, assim como da identificação de papéis de ajudante e ajudados visando à colaboração mútua entre os estudantes em um grupo. Ou seja, grupos heterogêneos ou mistos são esperados na aplicação

da PAA Trezentos. Em particular, grupos formados com estudantes que tenham bom, médio e baixo desempenhos são esperados para que a colaboração ocorra e todos possam se beneficiar.

A PAA Aprendizagem Baseada em Equipes (ABE), no que tange à utilização de equipes em ambiente escolar (De Oliveira et al., 2016; Bollela et al. 2014), indica que o docente deve criar as equipes (também denominadas como grupos) compostas, idealmente, por cinco a sete estudantes (membros). O objetivo é formar equipes compostas de membros com perfis distintos, ou seja, a composição do grupo precisa ser heterogênea.

De acordo com Bollela et al. (2014), o primeiro passo para organização de alguma atividade como, por exemplo, utilizando-se ABE é a formação de grupos de estudantes com perfis diversos. Michaelsen et al. (2008) discorre que este primeiro passo é um dos princípios fundamentais para aplicação da ABE. Algumas características e perfis de estudantes podem ser levados em consideração para elaboração de equipes. Para Oliveira et al. (2016), pode-se considerar os conhecimentos, experiências pessoais e/ou profissionais, interesses, entre outros fatores, pois tais características favorecem o surgimento de equipes com níveis semelhantes de interatividade.

2.2 Taxonomia de Bloom

Mediante a Taxonomia de Bloom (TB), desenvolvida por Bloom et al., (1956), pode-se buscar a verificação do nível cognitivo de um estudante. Uma das possíveis aplicações da TB remete à ideia de que o professor pode sistematizar a aquisição de uma competência e habilidade mais simples para embasar, posteriormente, a aquisição de uma habilidade mais complexa (Cabral, 2019). Para possibilitar isso, a TB estrutura-se em 6 níveis, que estão organizados hierarquicamente, conforme apresentados no Quadro 4.

Os níveis da TB se relacionam com os conhecimentos e competências do estudante de acordo com os objetivos educacionais definidos pelo docente. Os objetivos educacionais podem ter como base a natureza da disciplina, do curso, assim como outros aspectos que o docente julgar relevantes. Ressalta-se que o relacionamento hierarquizado compreende que cada nível apresenta-se de forma gradativa. Ou seja, para estudantes alcançarem níveis mais abstratos e profundos, sugere-se alcançar primeiramente os níveis mais básicos.

Quadro 4: Resumo dos níveis da Taxonomia de Bloom quanto ao domínio cognitivo

Nível	Descrição	Amostra de verbos
1-Lembrar	Recordar ou reconhecer informações, ideias, e princípios na forma (aproximada) em que foram aprendidos.	Escreva, liste, rotule, nomeie, identifique, cite e defina.
2-Entender	Traduz, compreende ou interpreta informação com base em conhecimento prévio.	Explique, traduza, ordene, diferencie, resuma, parafraseie, descreva e ilustre.
3-Aplicar	Seleciona, transfere e usa dados e princípios para completar um problema ou tarefa com um mínimo de supervisão.	Use, desenvolva, compute, resolva, demonstre, aplique e construa.
4-Analisar	Distingue, classifica e relaciona pressupostos, hipóteses, evidências ou estruturas de uma declaração ou questão.	Analise, classifique, categorize, deduza e separe.
5-Avaliar	Aprecia, avalia ou crítica com base em padrões e critérios específicos.	Julgue, argumente, compare, contraste, recomende, critique e justifique.
6-Criar	Cria, integra e combina ideias num produto, plano ou proposta, novos para ele.	Crie, proponha, formule, modifique, planeje, elabore hipótese(s), invente e projete

Fonte: (LIMA, 2009) com adaptações.

Destaca-se que a TB vem recebendo revisões ao longo do tempo. Os estudos de Anderson et al., (2001) e Krathwohl et al., (2002) reavaliaram a TB e alteraram o posicionamento dos dois últimos níveis. Os níveis síntese (nível 5) e avaliação (nível 6) da taxonomia original de Bloom et al., (1956) foram substituídos pelos verbos “criar” no lugar de “avaliação” e “avaliar” no lugar de “síntese”. Então, a partir de verbos, como os descritos na Quadro 4, pode-se elaborar ou identificar quais atividades, objetivos educacionais são pertinentes a um determinado plano de aula (Costa et al., 2019; de França et al., 2013). Corroborando com esta perspectiva, Galhardi et al., (2013) discutem que a TB, embora formulada na década de 50, tem sido revisitada por pesquisadores que reconhecem nela mais do que um instrumento para a avaliação do processo ensino-aprendizagem, mas também uma ferramenta útil e eficaz no planejamento e implementação de aulas, assim como, na organização e criação de estratégias de ensino (Galhardi; Azevedo, 2013).

2.3 Mineração de Dados Educacionais

A Mineração de Dados Educacionais (do inglês, Educational Data Mining, ou EDM) possui foco na utilização de algoritmos de Mineração de Dados (do inglês, Data Mining, ou DM) voltados para tratar problemas oriundos do domínio educacional (Baker et al., 2011). Considera-se que a EDM é uma subárea da MD e, assim, permeia áreas como (Romero et al., 2013): educação, estatística, ciência da computação e a própria MD. Nesta perspectiva, Maschio et al. (2018) revelam um cenário cada vez mais crescente de estudos e pesquisas no Brasil no tocante à MDE.

A EDM pode utilizar-se de algoritmos de aprendizado de máquina para obter resultados a partir dos dados explorados em ambientes educacionais. Conforme Dutt et al. (2017), alguns exemplos de estudos de aplicação de EDM estão focados em assistir a solução de problemas como: retenção de alunos, desempenho escolar, detecção de comportamentos, evasão discente, etc. Esta seção introduz a área de Mineração de Dados Educacionais e descreve alguns dos métodos mais utilizados para fins de descoberta de conhecimento no contexto educacional.

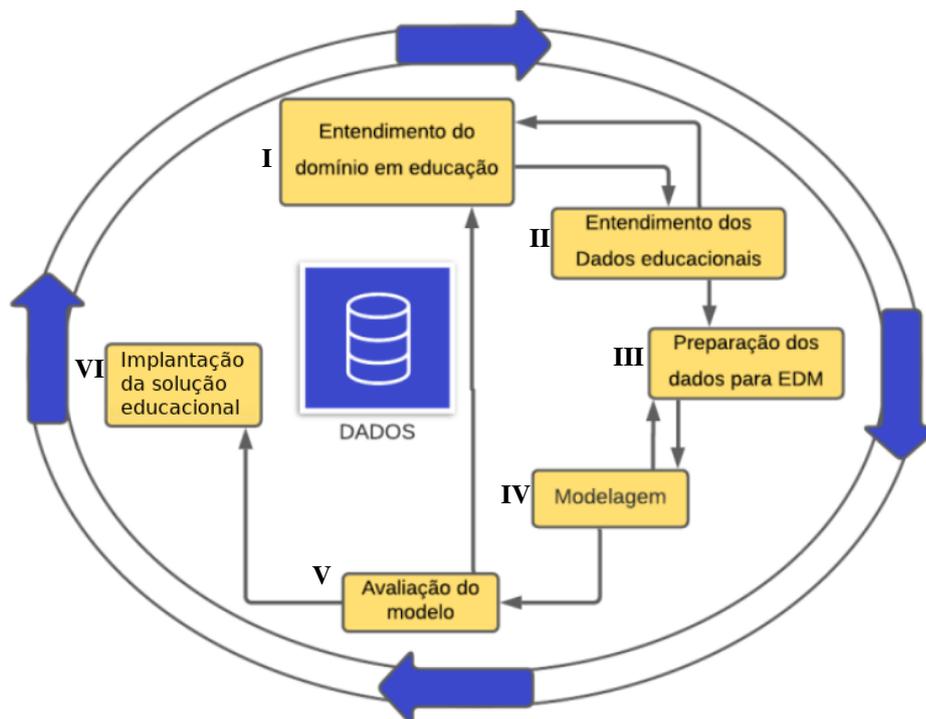
2.3.1 Modelo de processo para EDM

Para favorecer os percursos metodológicos necessários ao desenvolvimento de projetos de EDM, assim como em projetos gerais em MD, surgiram alguns modelos de processos. Os modelos de processos buscam prover etapas que permitam o trabalho com os dados de modo que se possa extrair ou descobrir conhecimento de modo sistematizado. O primeiro modelo de processo criado foi o Knowledge Discovery in Databases (KDD) (Fayyad et al., 1996a).

O KDD é um processo que se divide em cinco etapas que compreendem a escolha e seleção dos dados até a descoberta e interpretação do conhecimento propriamente dita. Análogo ao processo KDD, existe o processo Cross Industry Standard Process for Data Mining (CRISP-DM), descrito por Chapman (2000). Martínez-Plumed et al. (2019) discorrem que o CRISP-DM surgiu há mais de 20 anos motivado pela necessidade de facilitar e sintetizar o processo de projetos de mineração de dados e descoberta de conhecimento na indústria.

Existem inúmeras variações do CRISP-DM, a exemplo do CRISP-EDM, proposto por RAMOS et al. (2020), para o contexto educacional. Neste, busca-se a adaptação das etapas do processo original, almejando-se a descoberta do conhecimento no domínio educacional, o processo CRISP-EDM inicia-se na Etapa I (Entendimento do domínio educacional) e segue até alcançar a Etapa VI (Implantação da solução educacional) que é a última etapa conforme exposto na Figura 2.

Figura 2 - Etapas do processo CRISP-EDM.



Fonte: Elaborado pelos autores com base em RAMOS et al. (2020)

O CRISP-EDM unifica o modelo já consolidado CRISP-DM de modo aplicado a problemas dirigidos a dados educacionais. As seis etapas do processo CRISP-DM no contexto da EDM (no domínio educacional) foram adaptadas por RAMOS et al. (2020), conforme mostra a Figura 2. Dessa forma, baseando-se em Chapman (2000), as etapas originais foram estendidas para as seguintes: Entendimento do domínio educacional, Entendimento dos dados educacionais, Preparação dos dados para EDM, Modelagem, Avaliação do modelo e Implantação da solução educacional. Descrevem-se as etapas conforme o estudo de Ramos et al. (2020) a seguir:

A primeira etapa, denominada Entendimento do Domínio Educacional, visa entender e elucidar o problema, questões e requisitos intrínsecos ao domínio educacional. Definem-se, nesta etapa, os objetivos e perspectivas para o projeto de EDM. Ressalta-se que os projetos de EDM caracterizam-se pelo fator multidisciplinar (Romero et al., 2010). Como ilustração, em um projeto de EDM é natural a presença de especialistas como professores, estudantes, técnicos e pedagogos. Especialistas do domínio ajudam na compreensão do problema no âmbito educacional visando-se desenvolver possíveis soluções.

Em seguida, na segunda etapa, conhecida como Entendimento dos Dados Educacionais, posterior à definição dos objetivos e compreensão do domínio educacional, verifica-se quais

dados são relevantes ao problema de mineração que podem ser utilizados nesse processo. Inclui-se nesta etapa a coleta e o entendimento dos dados considerando fatores como natureza, fontes de dados e características ou atributos necessários ao desenvolvimento da solução.

A partir da terceira etapa, denominada Preparação dos Dados para EDM, verifica-se a necessidade de atividades de integração, limpeza, derivação de atributos, entre outras que irão ser executadas com o objetivo de obter-se um conjunto de dados refinado ou transformado, adequado à etapa de Modelagem de aprendizado de máquina.

A quarta etapa, conhecida como Modelagem, destina-se à criação do modelo de Aprendizado de Máquina (AM). Para isso, métodos de aprendizado supervisionado e/ou não supervisionado são normalmente empregados. Aprendizado de máquina é tópico da “Seção 2.2.2”.

A quinta etapa, nomeada de Avaliação do modelo de AM, destina-se à avaliação e aprovação do modelo obtido. Nesta, avaliam-se se os requisitos definidos no início do processo foram atendidos. Examina-se também se o modelo obtido é eficaz diante dos objetivos definidos na primeira etapa. Nesta etapa, pode-se, por exemplo, rever e analisar o processo realizado e definir os próximos passos inclusive a volta a etapas anteriores tendo em vista que todo o processo CRISP-EDM é iterativo.

A última etapa, denominada de Implantação, possui foco na organização e estruturação para uso do modelo pelo usuário ou pela aplicação que irá consumi-lo. Depois de finalizada a última etapa, comumente o processo CRISP-EDM, ciclicamente, pode voltar para a primeira etapa ou para outras fases anteriores, e um novo ciclo pode ser realizado. Pode-se, por exemplo, adicionar novas questões a serem respondidas dentro do domínio educacional e aprimorar o modelo de aprendizado já criado. Ramos et al. (2020) mencionam que o processo CRISP-EDM não é rígido, ou seja, sua sequência de fases não é rigorosa.

Utiliza-se o modelo de processo CRISP-EDM neste trabalho como base metodológica para a criação e execução da abordagem proposta. Isso se deve à sua capacidade de padronização e organização de atividades destinadas a projetos e problemas em EDM.

2.3.2 Aprendizado de Máquina

O Aprendizado de Máquina (AM) fornece algumas bases técnicas para a Mineração de Dados. Por meio de algoritmos de AM, é possível obter conhecimento valioso (Mitchel,

1997). O aprendizado de máquina é, em geral, subdividido em aprendizado supervisionado (preditivo) e aprendizado não supervisionado (descritivo) (Harrington, 2012). Os métodos supervisionados utilizam-se de dados previamente conhecidos e rotulados, enquanto naqueles não supervisionados não existem rótulos pré-estabelecidos. Métodos supervisionados são muito empregados para prover tarefas de classificação de dados, assim como estimar resultados inclusive no panorama de dados educacionais. Para Alpaydin (2010), mediante a técnica supervisionada denominada regressão, pode-se a partir de um conjunto de dados de entrada, considerando-se atributos relevantes para determinado domínio de aplicação, elaborar um modelo de AM obtendo-se uma variável numérica de saída (resultante do processo). Por exemplo, em Silva et al. (2019), utiliza-se regressão linear com atributos de entrada referentes a dados socioeconômicos (quantidade de pessoas que moram na residência, incentivo dos pais às tarefas escolares, área onde o estudante reside, dentre outras) para produção da estimativa do desempenho do estudante. O resultado nesse tipo de tarefa é dado a partir de uma variável de saída numérica (valor contínuo).

A classificação supervisionada é realizada com base em atributos de um conjunto de dados objetivando-se a geração da variável de saída (resultante) com valor nominal (e.g., verdadeiro ou falso; aprovado, reprovado ou retido) (Harrington et al., 2012). Evidencia-se um exemplo da aplicação de classificação supervisionada em Costa et al. (2015). Eles processaram dados de entrada pertinentes ao desempenho acadêmico de estudantes que ingressaram por cotas e, ao final, obtiveram atributos nominais indicando a probabilidade de fracasso ou sucesso escolar. No aprendizado não supervisionado, enquadram-se, por exemplo, tarefas como associação e agrupamento. Essas tarefas provêm o processamento e mapeamento de dados de entrada, mesmo sem conhecimento prévio e sem a presença de um supervisor, de modo a gerar análises descritivas (Alpaydin, 2010).

Em tarefas de regras de associação, a busca é por padrões de relacionamentos fortes entre os dados. Objetiva-se, na associação, encontrar subconjuntos de itens que ocorrem de forma simultânea e frequente em uma fração mínima e previamente estabelecida do conjunto de dados (Agrawal et al., 1993). Alguns exemplos de aplicação de associação podem ser visualizados em transações de cestas de compras (e.g. quem compra produto x compra, também, produto y), vídeos assistidos em alguma plataforma (e.g., quem aluga ou compra determinado filme x pode acabar comprando também filme y), comportamentos de estudantes em AVA (e.g., estudantes que podem apresentar dúvidas em determinado conteúdo também precisam de auxílio). Para melhor compreensão do processo EDM utilizando-se de regras de

associação, Baker e Isotani (2011) descrevem de modo resumido um exemplo dentro da perspectiva da EDM:

Ao analisar um conjunto de dados seria possível identificar uma regra que faz a associação entre a variável “objetivo do aluno”, uma variável binária que pode ter os valores alcançado ou não alcançado, e uma outra variável binária “pedir ajudar ao professor” que pode ter os valores sim ou não. Neste contexto, se o aluno tem como objetivo aprender geometria, mas está com dificuldade (i.e., a variável objetivo do aluno tem valor não alcançado), então é provável que ele peça ajuda do professor (i.e., a variável pedir ajuda ao professor tem valor positivo). (BAKER e ISOTANI, 2011, p.6)

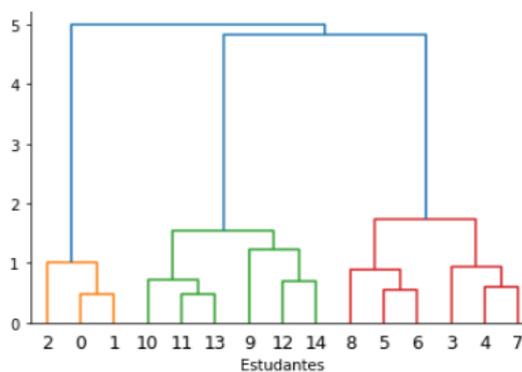
Neste contexto, o autor complementa a busca ao gerar e identificar regras do tipo se-então (do inglês, *if-then*) que possibilitem realizar a associação do valor obtido de um atributo ao valor de outro atributo. Dessa forma, Baker e Isotani (op. cit.) menciona que no caso de uma condição satisfeita, ou seja, verdadeira (e.g., valor 1 atribuído ao atributo Y) e uma regra associe essa condição ao valor de uma outra atributo X, então pode-se inferir o valor do atributo X. Diante dos exemplos citados percebe-se inúmeras possibilidades de utilização das regras de associação no âmbito da EDM. Considerando a importância da técnica de agrupamento, também não supervisionada, para este trabalho, a próxima seção descreve-a em mais detalhes.

2.3.3 Método de Agrupamento

O agrupamento está enquadrado como uma tarefa de aprendizado não-supervisionado. Define-se que o método de agrupamento (do inglês, *clustering*) agrupa itens semelhantes, entretanto, para isso, não utiliza rótulos de classes pré-definidas (Harrington, 2012). No agrupamento, busca-se a geração de grupos de instâncias observando-se similaridades entre elas (com base em seus atributos) (Han et al., 2011). Por exemplo, utiliza-se, em Correia et al. (2012), um algoritmo de agrupamento visando à formação de turmas mais homogêneas para a aplicação de recuperação paralela. Para isso, utiliza-se um conjunto de dados de entrada composto de atributos como dificuldade de interpretação, dificuldade de raciocínio lógico, dificuldades de argumentação, dentre outras. O objetivo é a geração de grupos (clusters) de estudantes similares de acordo com os atributos usados. Nesse contexto, um cluster apresenta uma coleção de instâncias similares intra-cluster (interna ao cluster) e dissimilares a outras instâncias pertencentes a outros clusters (inter-clusters). Dutt et al., (2017) apontam que, mediante a tarefa de agrupamento, possibilita-se melhorias em sistemas educacionais. Dessa forma existem algumas áreas de pesquisa na Informática na Educação que utilizam-se da clusterização para este fim: aprendizagem colaborativa, tutores inteligentes, estilos de

aprendizagem dentre outras. Em geral, as técnicas de agrupamento são categorizadas em (Han, 2011): Métodos Hierárquicos; Métodos Particionais ou partitivos e Métodos baseados em Densidade. Algoritmos de agrupamento hierárquicos utilizam as semelhanças das instâncias, sem requerer configurações pré definidas (Alpaydin, 2010), como, por exemplo, a necessidade de números de clusters definidos. Métodos hierárquicos basicamente apresentam clusters em forma de árvore, comumente representado por dendrogramas, como exemplificado na Figura 3 (Han et al., 2011).

Figura 3 - Métodos hierárquicos e exemplo de dendrograma



Fonte: Elaborado pelo autor

A Figura 3 apresenta, em um nível mais abrangente, 3 clusters. Para isso, no eixo horizontal estão os estudantes e no vertical a distância utilizada na identificação da similaridade entre os estudantes. Os algoritmos hierárquicos podem ser aglomerativos (abordagem de baixo para cima, bottom up) ou divisivos (abordagem de cima para baixo, top-down) (Madhulatha, 2012). A geração de clusters hierárquicos aglomerativos inicia a partir de clusters menores (um cluster para cada instância) e, a partir daí, geram-se clusters cada vez maiores a cada iteração, incrementalmente, até formar um único cluster abarcando todas as instâncias. O algoritmo Ward (aglomerativo), por exemplo, foi aplicado no estudo de Ramos et al. (2020) para análise e compreensão da escolha do número de clusters adequado. Outro exemplo diz respeito ao algoritmo AGNES (Agglomerative Nesting), também do tipo aglomerativo, que foi aplicado em Pereira et al. (2018) para identificação da participação e interação de estudantes.

Já os métodos hierárquicos divisivos partem de um único grande cluster que engloba todas as instâncias e, a partir disso, vai fragmentando o cluster corrente e gerando outros clusters menores em cada iteração até que cada instância represente um cluster. Divisive ANalysis (DIANA) é um exemplo de algoritmo hierárquico divisivo (Onan, 2020).

Métodos de agrupamento demográficos, por sua vez, baseiam-se na densidade de grupos de objetos com base na noção de densidade (Han, 2011). De forma geral, identificam-se as áreas no espaço amostral onde existam concentrações de instâncias, e cada uma dessas áreas forma um novo cluster. Não requer a inserção do número de cluster, pode-se configurar parâmetros para geração dos grupos (e.g., raio e número mínimo de instâncias por grupo). O método é robusto ao identificar outliers (ruído), que são os pontos que ficam no final do processo de agrupamento. A Figura 4 apresenta um exemplo de representação do método demográfico.

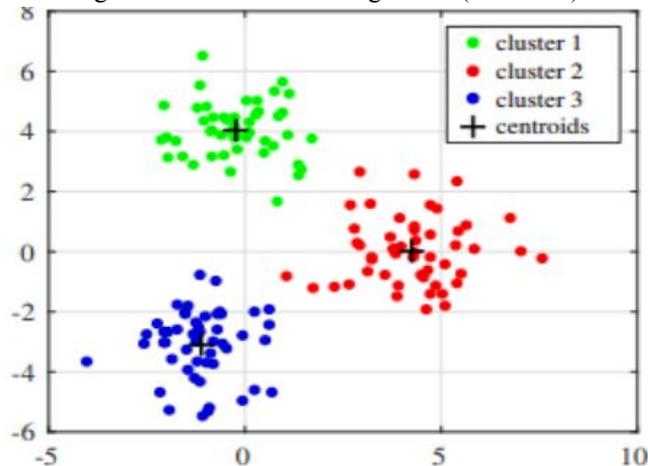
Figura 4 - Exemplo de representação do método demográfico



Fonte: (GANDHI et al., 2021, p.63)

A Figura 4 expõe dois grupos de pontos com alta densidade. Percebe-se também a presença de alguns pontos distantes dos dois grupos que possivelmente formam um terceiro grupo (de outliers). De acordo com Ester et al., (1996) e Gandhi et al., (2021), um exemplo de algoritmo que se enquadra nessa categoria é o DBSCAN (Density-based spatial clustering of applications with noise). Os métodos de agrupamento particionais ou partitivos formam uma terceira categoria e objetivam segmentar um conjunto de dados em partições a partir de núcleos, chamados de centróides (k), conforme ilustrado na Figura 5.

Figura 5 - K-means com k igual a 3 (3 clusters)



Fonte: (SOUZA, 2020, p. 58)

O K-means é um dos principais algoritmos de agrupamento partitivo, muito utilizado

devido à sua simplicidade (Souza, 2020). Revela-se o algoritmo mais utilizado para soluções que utilizam agrupamento em EDM (Dutt et al., 2017). O algoritmo k-means faz uso de uma estratégia particional e funciona, resumidamente, da seguinte forma:

1. Previamente define-se o número de clusters que será utilizado no processamento do algoritmo (k).
2. Distribui-se randomicamente os locais onde os centróides de cada cluster serão alocados. Esses centróides são utilizados como ponto de referência para alocação das demais instâncias dos clusters que serão incluídas de acordo com o cálculo da proximidade entre elas e o centróide. O cálculo da proximidade é realizado por meio, por exemplo, da Distância Euclidiana, Cosseno ou Manhattan.
3. O terceiro passo do algoritmo é recalcular os centróides e reorganizar os elementos pertencentes aos clusters até chegar em uma condição de parada com base na imutabilidade das distâncias calculadas em relação às instâncias.

Apesar de utilizada a distância euclidiana nos exemplos, outras métricas de distância podem ser utilizadas. “Normalmente, algum tipo de medida de distância é usado para decidir quão semelhantes são as instâncias” (RAMOS, et al., 2016, p. 3756.). Outros exemplos de medidas de cálculo de distância, conforme Akay e Yuksel (2019) são: Minkowski, Single Linkage, Manhattan, Linkage e o Two-step.

No que tange à definição do número de clusters pelos métodos de agrupamento, existem várias maneiras de definição do número de clusters no processo de MD que podem, em geral, ocorrer automaticamente ou ser requisitado pelo algoritmo de agrupamento. Nos métodos partitivos, pode-se, por exemplo, realizar tal definição mediante regras de negócio definidas no início do processo de mineração de dados. Han et al., (2011) mencionam outra forma muito popular para definição do número de clusters, que é utilizando-se do método elbow (também denominado de método do cotovelo). Busca-se com o método Elbow analisar matematicamente a variação dos valores gerados a partir de cada número de clusters. A partir disso, obtém-se o melhor número de grupos (o mais provável) que pode ser utilizado, mesmo sem conhecer a resposta previamente. De acordo com Syakur et al. (2017), utiliza-se o cálculo da SSE (Sum of Square Error ou Soma do Erro Quadrado, em português) mediante o método Elbow objetivando-se a avaliação e determinação da quantidade ideal de clusters. Um exemplo desse processo é exibido na Figura 6.

Figura 6. SSE, método Elbow e a quantidade de clusters



Fonte: Adaptado de (SYSKUR et al. , 2017, P. 5)

Na Figura 6 são exibidos os valores do SSE e a quantidade de clusters que variam de dois a oito. Neste caso, identifica-se a quantidade ideal de clusters para o conjunto de dados analisados, que resulta em três clusters. Ou seja, com três clusters atende-se ao critério do cotovelo, definido como o momento analisado no método Elbow que aponta a quantidade ideal de clusters conforme estudo de (Crispim et al., 2019; Hackeling, 2017). Dessa forma pode-se considerar o SSE e o método elbow na identificação da quantidade de clusters ideal, como no exemplo exposto.

2.4 Avaliação de grupos de estudantes

Um aspecto de fundamental importância na tentativa de criar grupos de estudantes consiste na determinação do quão próximo seus perfis apresentam-se uns dos outros ou o quão distante eles estão. Neste trabalho, busca-se, principalmente, gerar grupos com perfis de estudantes heterogêneos com fins de aplicação de PAAs. Mas, considerando os objetivos definidos para este estudo, é preciso medir a qualidade dos grupos gerados numa etapa preliminar, voltada à identificação de perfis similares, assim como na fase final, com os grupos formados de modo heterogêneo. Para isso, algumas métricas de avaliação consideradas na literatura (Oliveira et al., 2022; Kelvin et al., 2021; Moubayed 2020; Macedo et al., 2020) foram identificadas. Dentre as possibilidades encontradas, duas métricas foram escolhidas para as avaliações neste trabalho, a saber: (I) Silhueta e (II) Diversidade. Elas são descritas a seguir.

2.4.1. Avaliação de grupos homogêneos

No que tange à avaliação de grupos gerados por meio de métodos de AM baseados em agrupamento, uma das medidas mais utilizadas refere-se ao coeficiente de silhueta - do inglês silhouette - (Rousseeuw, 1987). Conforme o estudo de Rousseeuw (1987), para aplicação do coeficiente de silhueta, necessita-se de: (i) conjuntos de clusters gerados (mediante algum método de agrupamento), e (ii) conjunto de distâncias entre as instâncias, que pode ser mensurada, por exemplo, usando a distância euclidiana. A partir do coeficiente de silhueta é gerado um valor numérico para descrever o quão bem uma instância está alocada a seu cluster.

Neste sentido, obtém-se valores dentro do intervalo de [-1; +1]. Um valor alto (mais próximo de 1) indica uma configuração de cluster mais apropriada, e valores próximos a zero revelam a alocação inconclusiva (instância poderia estar tanto no seu cluster quanto em algum outro cluster) e, perto de -1, indica que a alocação foi realizada de forma não adequada. Pode-se calcular o índice de silhueta conforme a Fórmula (1)

$$S(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (1)$$

A Fórmula (1) apresenta o cálculo da S(i), que representa o valor resultante da silhueta correspondente de uma instância. Os atributos a(i) e b(i) destinam-se, respectivamente, ao valor da média da distância da instância em relação às demais instâncias de seu cluster (intracluster) e da dissimilaridade em relação a outros clusters (intercluster).

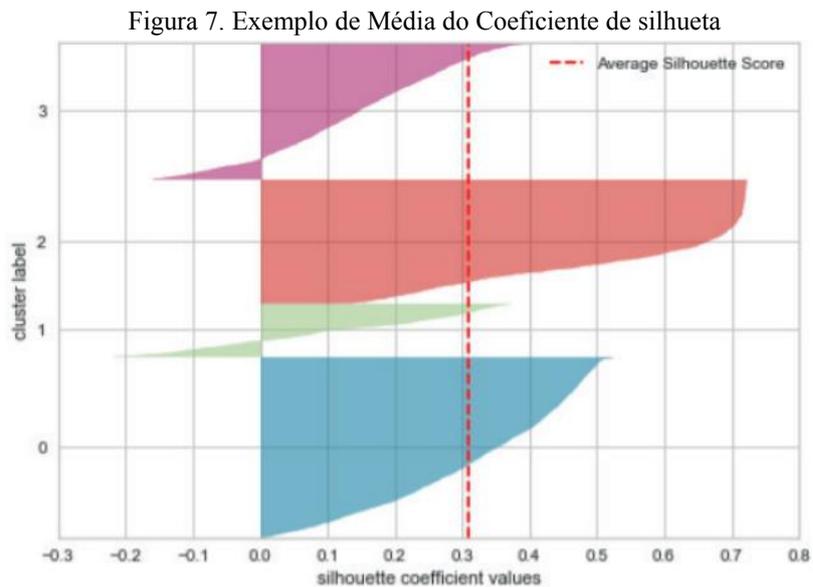
Desta forma, mediante o índice de silhueta, pode-se observar a qualidade dos clusters baseada na proximidade entre as instâncias de um determinado cluster e na proximidade ou distância dessas instâncias aos demais clusters. A fórmula para cálculo da média dos índices Silhouette de um determinado cluster é mostrada na Fórmula (2).

$$\frac{\sum_1^n s(i)}{n} \quad (2)$$

A Fórmula (2) corresponde ao somatório dos valores dos índices S(i) das instâncias dividido por n, onde n refere-se ao número de instâncias pertencentes ao cluster, o que demonstra a relação entre os atributos nos seus respectivos clusters. Também é possível uma

análise holística de todas as instâncias do dataset conforme o cálculo da média mostrada na Fórmula (2). Neste caso considera-se todos valores de $S(i)$ no qual n representa a quantidade de instâncias de todo dataset.

A Figura 7 expõe um exemplo de média de todos índices $S(i)$ e apresenta a média dos índices de silhueta de 3,1, conforme linha tracejada vertical. Conforme exposto na Figura 7 o eixo vertical refere-se aos quatro clusters (0..3) utilizados no estudo, e o eixo horizontal destina-se aos valores do coeficiente de silhueta.



Fonte: (MACEDO, 2020, p.330)

2.4.2. Avaliação de grupos heterogêneos

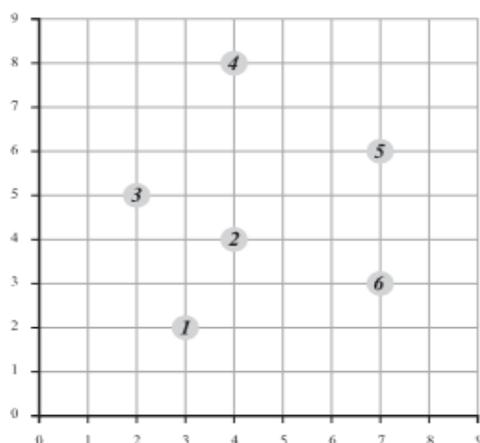
Por outro lado, ao focar na análise da heterogeneidade envolvida diante de grupos compostos de perfis de estudantes distintos, é preciso mensurar a diversidade entre as instâncias em cada grupo de modo que esta seja a maior possível. Assim a métrica diversidade de cada grupo pode ser calculada mediante a Fórmula (3), a qual refere-se à quantidade de instâncias de um grupo, d representa a distância euclidiana entre os elementos do grupo, e i e j indicam os índices dos elementos de cada instância (Kuo et al., 1993).

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij} \quad (3)$$

Tal métrica possibilita analisar a heterogeneidade individualmente de cada um dos grupos formados. Observa-se, assim, que a partir dessa métrica, pode-se verificar se a diversidade entre as instâncias dos grupos formados foi maximizada. Objetivando-se ilustrar a

aplicação da métrica, um exemplo de sua utilização é apresentada no estudo de Martí, Gallego e Duarte (2010). No exemplo exposto na Figura 8 e na Figura 9 apresentam-se seis instâncias e as distâncias correspondentes no ambiente bidimensional.

Figura 8 – Instâncias em um espaço bidimensional



Fonte: MARTÍ, GALLEGO e DUARTE (2010)

Percebe-se, visualmente, na Figura 8 a distância elevada da “instância 4” para a “instância 1”, também verifica-se a proximidade da “instancia 1” com a “distância 2”. Na Figura 9 são apresentados os cálculos das distâncias euclidianas correspondentes a cada instância em relação às demais instâncias, ou seja, a distância euclidiana da “instância 1” para “instancia 2” é 2.24, já a distancia da “instância 4” para a “instancia 1” é 6.08. Neste sentido, conforme a Formula (3), o valor resultante da diversidade de um conjunto de dados é diretamente proporcional ao somatório das distância calculadas, ou seja, a diversidade é maior se as distâncias entre as instancias são maiores, e a diversidade é menor conforme as distâncias são menores.

Figura 9 - Matriz de distância

	1	2	3	4	5	6
1	-	2.24	3.16	6.08	5.66	4.12
2	2.24	-	2.24	4.00	3.61	3.16
3	3.16	2.24	-	3.61	5.10	5.39
4	6.08	4.00	3.61	-	3.61	5.83
5	5.66	3.61	5.10	3.61	-	3.00
6	4.12	3.16	5.39	5.83	3.00	-

Fonte: MARTÍ, GALLEGO e DUARTE (2010)

A partir da métrica de diversidade pode-se obter os indicadores da heterogeneidade em um conjunto de dados. Nesta dissertação, utiliza-se a métrica de diversidade para avaliação dos grupos formados para PAA. Dessa forma, pode-se verificar o quão distantes são

os membros de um grupo e conseqüentemente o quão heterogêneo é o grupo.

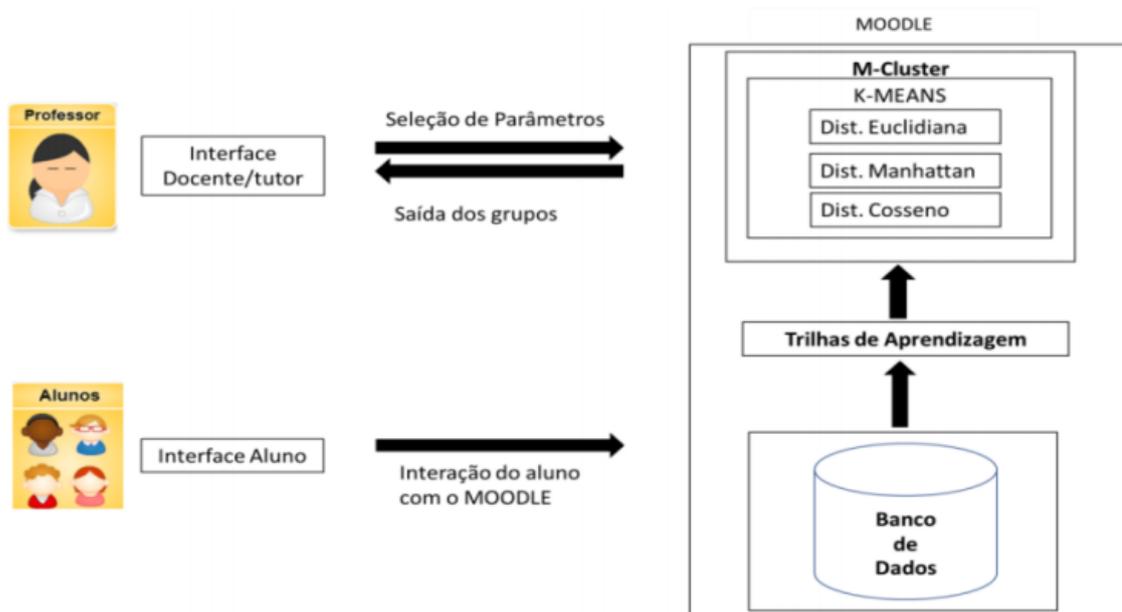
2.5 Trabalhos relacionados

Apresentam-se nesta seção alguns trabalhos considerados relacionados à proposta de abordagem apresentada nesta dissertação. Os trabalhos foram selecionados a partir de uma busca *ad-hoc* da literatura e objetivam a formação de grupos de estudantes para suporte à prática docente (Ramos et al. 2020; França et al. 2013; Monteverde et al., 2018; Nalli et al., 2021; Maina et al., 2017). Os trabalhos envolvem a formação de grupos com perfis tanto similares quanto grupos com perfis distintos. Ao final, sintetizam-se informações relevantes em um quadro para discussão.

2.5.1. Framework Conceitual para Formação de Grupos de Alunos utilizando Trilhas de Aprendizagem em um Ambiente Virtual de Aprendizagem

A proposta de framework elaborada por Monteverde et al. (2018) objetiva a formação de grupos de estudantes para atividades de aprendizagem colaborativa em um AVA, denominado MOODLE (Modular Object-Oriented Dynamic Learning Environment). Para isso, foram coletados dados oriundos da interação de estudantes de duas turmas de curso a distância da Universidade Federal do Amazonas, nos semestres de 2015.1 e 2016.2. Esquematiza-se a ideia da proposta deste trabalho na Figura 10.

Figura 10 - Visão geral da proposta do Framework Conceitual para Formação de Grupos



Fonte: (MONTEVERDE et al., 2018, p. 1676)

Conforme mostra a Figura 10, os dados coletados das interações dos estudantes no AVA são armazenados. Em seguida, extraem-se os atributos relevantes desse conjunto de dados de acordo com as trilhas de aprendizagem descritas por Ramos et al. (2016). Sintetizando, as trilhas são compostas de :

- **Vértice:** Corresponde a um recurso/atividade;
- **Arestas:** Pode possuir 3 valores: (i) valor padrão, definida pelo professor, e os percursos que o estudante realizou que representam as arestas de (ii) retorno e (iii) avanço;
- **Diâmetro:** Proporcional ao peso, ponderado pelo número de interações realizadas pelo estudante.

Dessa forma os vértices estão associados a uma cor e diâmetro, as trilhas consideram o percurso do estudante na realização e utilização de recursos/atividades realizadas no AVA. Nesse sentido para cada estudante utiliza-se:

- Média de acesso de vértices,
- Quantidade,
- Dispersão e variância das arestas padrão, avanço e de retorno
- Identificação do estudante.

No tocante ao processo de formação dos grupos utiliza-se a ferramenta M-Cluster, proposto por Monteverde et al. (2017) em trabalho anterior. Mediante o M-Cluster realiza-se o agrupamento de estudantes de acordo com o grau de similaridade com base nos dados das trilhas de aprendizagem, aplicando-se o algoritmo de agrupamento não supervisionado K-means e utilizando as distâncias euclidiana, cosseno ou manhattan. O professor ou tutor seleciona, manualmente, os parâmetros para formação dos grupos e analisa os clusters e grupos de estudantes gerados, ou seja, o número de clusters é pré estabelecido pelo docente.

2.5.2. Mineração de dados na identificação de grupos de estudantes com dificuldades de aprendizagem no ensino de programação

O modelo proposto por De França (2013) sugere a utilização da TB para identificação de grupos de estudantes no tocante ao nível cognitivo. De França et al. (2013) analisaram dados históricos de estudantes, utilizando-se da TB, objetivando-se a identificação de grupos de estudantes com dificuldades de aprendizagem em determinados temas abordados na disciplina. Extraíram-se dados a partir de avaliações aplicadas ao longo da disciplina. Os atributos para mapeamento do conhecimento dos estudantes estão postos conforme a listagem de atributos seguinte:

- IdAluno: Identifica um estudante (valor numérico)
- IdSessao: Identifica uma determinada avaliação (valor numérico)
- IdItem: Corresponde a uma questão formativa (valor nominal)
- Conteúdo: Corresponde ao conteúdo tratado no item avaliado (valor nominal)
- NivelCognitivo: Nível cognitivo da TB relativo ao item avaliado (valor nominal)
- ConceitoItem: Corresponde ao grau de acerto do aprendiz no item avaliado, pode compreender os valores A,B,C,D e E.
- ConceitoAvaliacao: Corresponde ao desempenho do estudante nas avaliações formativas, pode compreender os valores A,B,C,D e E.
- RotuloSomativa: efere-se à situação do estudante aprovado ou reprovado na avaliação somativa (valor nominal).

A partir de 890 registros coletados das avaliações, experimentou-se de dois a dez grupos mediante o algoritmo k-means. Para isso, foi configurada a quantidade de grupos para seis, no qual, identificaram-se grupos conforme mostra a Figura 11

Figura 11 - Formação de clusters e o nível de aprendizagem a partir de exercícios e da TB

Attribute	Full Data (890)	Cluster#		
		0 (113)	1 (144)	2 (351)
Conteudo	Arrays	Arrays	Arrays	Estruturas de Repeticao
NivelCognitivo	CRI	ENT	CRI	CRI
ConceitoAvaliacao	?	C	B	?
RotuloSomativa	REP	APR	APR	REP
		3 (122)		4 (73)
				5 (87)
Componentes Graficos do Pacote GUI + Objetos e Classes		Estruturacao de Sistemas em Camadas		Arrays
		APL		ENT
		A		E
		REP		REP

Fonte: (DE FRANÇA et al. , 2013, P.7)

No experimento descrito, permitiu-se agrupar estudantes conforme dificuldades de aprendizagem da disciplina e grupos denominados mais experientes que apresentaram bom desempenho. Identificaram-se, por exemplo, três grupos (grupo zero, grupo um e grupo cinco) que correspondem à temática *array* conforme exposto na Figura 11. O grupo zero e o grupo um foram compostos por estudantes que, respectivamente, estão situados no nível entender *arrays* (nível dois da TB) e criar *arrays* (nível seis da TB). Já os estudantes do grupo cinco apresentaram dificuldades ao criar *arrays* (nível seis da TB) e foram reprovados ao fim da disciplina.

2.5.3. Comparative Analysis of Clustering Algorithms and Moodle Plugin for Creation of Student Heterogeneous Groups in Online University Courses

O estudo de Nalli et al., (2021) visa à formação de grupos heterogêneos. Para isso, o trabalho em questão, primeiramente, realiza o agrupamento dos estudantes em clusters, considerando a similaridade do comportamento de cada estudante (instância) na interação com o AVA. Os atributos do conjunto de dados usado são:

- Número de acessos à plataforma
- Porcentagem de estudantes que concluíram a tarefa
- Frequência total de exibição
- Número médio de experimentos em vídeo
- Número médio de visualizações de vídeos,
- Média do total de Frequência de visualizações de tutoriais em vídeo
- Número médio de tutoriais em vídeo visualizados.

O trabalho de Nalli et al., (2021) experimenta seis algoritmos de clusterização, a saber: k-means, Mean-Shift Clustering, Agglomerative Clustering (WARD), Density-based spatial clustering of applications with noise (DBSCAN), Gaussian Mixture Models Clustering (GMMC), e Self-Organizing Map (SOM).

O índice de silhueta foi utilizado na avaliação do melhor algoritmo de EDM, cujo resultado apontou para o algoritmo k-means. Este foi avaliado como o algoritmo mais assertivo na alocação de um estudante a um cluster. Mediante os métodos de agrupamento, são obtidos 3 grupos (Cluster 0, Cluster 1 e Cluster 2) conforme Figura 12.

Figura 12 - Processo de formação de grupos heterogêneos

```
cluster 0:  
[4, 8, 12, 22, 25, 29, 30, 34, 43, 46]  
  
cluster 1:  
[0, 1, 2, 3, 5, 6, 7, 10, 11, 14, 15, 21, 23, 24, 26, 27, 28, 31, 33, 35, 36, 38, 39,  
40, 41, 45, 47, 51, 52]  
  
cluster 2:  
[9, 13, 16, 17, 18, 19, 20, 32, 37, 42, 44, 48, 49, 50, 53, 54]  
  
Group 0[4, 9, 0, 1, 2]  
Group 1[8, 13, 3, 5, 6]  
Group 2[12, 16, 7, 10, 11]  
Group 3[22, 17, 14, 15, 21]  
Group 4[25, 18, 23, 24, 26]  
Group 5[29, 19, 27, 28, 31]  
Group 6[30, 20, 33, 35, 36]  
Group 7[34, 32, 38, 39, 40]  
Group 8[43, 37, 41, 45, 47]  
Group 9[46, 42, 44, 51, 52]  
Group 10[48, 49, 50, 53, 54]
```

Fonte: (NALLI et al. , 2021, P.17)

Os clusters representam o grau de atividade dos estudantes, ou seja, do mais intenso (Cluster 2) ao menos ativo (Cluster 0). Por fim, os índices de estudantes presentes nos clusters

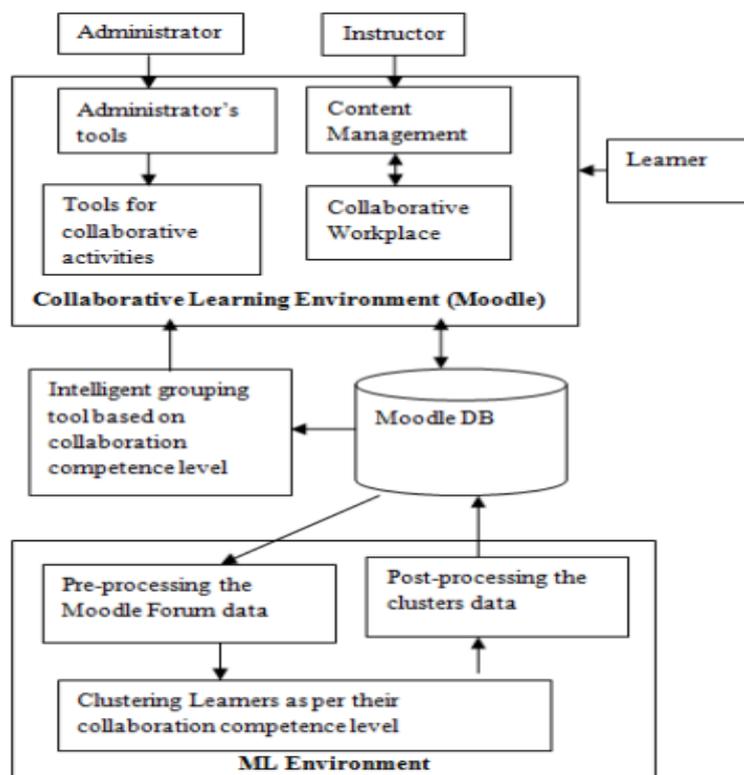
foram utilizados no algoritmo inteligente proposto resultando em grupos heterogêneos conforme exposto na Figura 12.

A partir dos 3 clusters obtidos, os estudantes foram distribuídos em 11 grupos de estudantes com perfis heterogêneos para utilização do professor conforme exemplificado na Figura 12. O trabalho implementou também um plugin para o MOODLE com a estratégia proposta.

2.5.4. Using machine learning techniques to support group formation in an online collaborative learning environment

O estudo de Maina et al. (2017) visa ao auxílio docente mediante a formação de grupos de estudantes de acordo com o nível de competência de colaboração (alto, médio e baixo). Basicamente a abordagem divide-se em duas etapas, a saber: (i) Criação de clusters (ii) Implementação de agrupamento inteligente. A visão geral da abordagem é mostrada na Figura 13.

Figura 13. - Processo realizado em duas etapas para formação de grupos heterogêneos



Fonte: (MAINA et al., 2011, P. 28)

Como mostrado na Figura 13, o trabalho utiliza dados oriundos do Moodle, mais especificamente dados relacionados a interações em fóruns de discussão. Na primeira etapa

utilizam-se os algoritmos Skmeans e Expectation Maximization (EM) para experimentação. Configurou-se os algoritmos para a geração de 3 grupos (Cluster 0, Cluster 1 e Cluster 2). Esse número é referente aos níveis de competência de colaboração (alto, médio e baixo). Os atributos presentes no conjunto de dados utilizado na clusterização são: (i) ID do usuário, (ii) Número de postagens, (iii) Número de respostas e (iv) Avaliações do fórum (i.e o tutor avalia de forma categórica a colaboração do estudante) Após agrupar, na primeira etapa, os estudantes considerando a similaridade (utilizando-se os algoritmos citados), submetem-se os resultados (cluster 0, cluster 1 e cluster 2) ao algoritmo inteligente para a formação de grupos. Neste mecanismo de formação inteligente de grupos, geram-se grupos heterogêneos, distribuindo-se estudantes originalmente pertencentes ao Cluster 0, considerado o perfil mais colaborativo até aqueles do Cluster 2, cujo perfil é o menos colaborativo. A lógica pode ser vista com mais detalhes no Anexo I algoritmo

Para o agrupamento inteligente, o algoritmo proposto por Maina (2017) define um *array* composto de valores referentes à identificação dos estudantes (*userid*) presentes nos clusters (*Cluster 0*, *Cluster 1* e *Cluster 2*) que segmentam os estudantes de acordo com o nível de interação. A partir do *array* criado, ocorre o seguinte:

1. As instâncias presentes em cada cluster (Cluster 0, Cluster 1 e Cluster 2) são embaralhadas.
2. Os estudantes são ordenados de forma decrescente por meio dos clusters.
3. A partir da ordenação, alocam-se os estudantes, de diferentes níveis de colaboração, a novos grupos.
4. O processo é realizado iterativamente até que todos os alunos sejam atribuídos a um grupo.
5. Os estudantes considerados mais interativos recebem um papel de mentor em seu grupo.

2.5.5. Identificação de perfis de interação de estudantes de educação a distância por meio de técnicas de agrupamentos

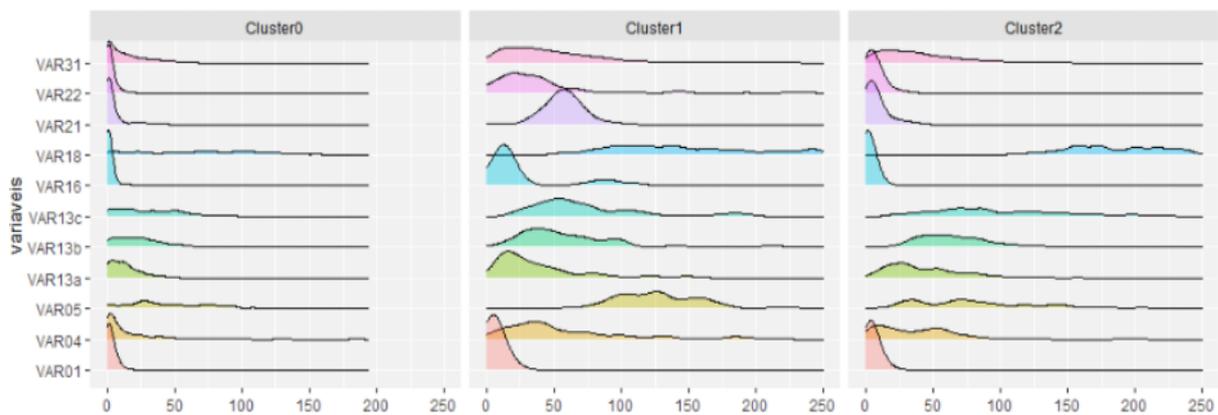
O estudo realizado por Ramos et al. (2020) utilizou agrupamento objetivando a identificação de perfis de estudantes de graduação no EAD da UNIVASF. Para isso, coletou dados oriundos das interações de seu AVA referentes aos cursos de Administração Pública e Pedagogia. Os dados foram coletados para o período de 2013 até 2018. Os atributos presentes no conjunto de dados coletado e considerados para utilização do algoritmo de agrupamento foram:

- Quantidade de postagens do aluno em fóruns por disciplina (VAR01);

- Quantidade de mensagens enviadas pelo aluno no semestre (VAR 04);
- Quantidade de mensagens recebidas pelo aluno no semestre (VAR 05);
- Quantidade de acessos por turno: manhã (VAR 13a), tarde (VAR 13b), e noite (VAR 13c), no semestre;
- Quantidade de colegas diferentes a quem o aluno enviou mensagens no semestre (VAR 16) ;
- Quantidade de acessos do aluno ao ambiente no semestre (VAR 18)
- Quantidade de mensagens recebidas de colegas no AVA no semestre (VAR 21)
- Quantidade de mensagens enviadas para outros colegas no semestre (VAR 22)
- Quantidade de acessos aos fóruns por disciplina (VAR 31).

Foi definida a necessidade de 3 clusters para identificação de perfis de interação dos grupos de estudantes conforme a seguinte discriminação: **baixa interação, média interação e alta interação**. Depois da etapa de coleta e seleção, realizou-se o pré processamento e o tratamento dos dados para aplicação dos algoritmos de clusterização hierárquico Ward (utilizado para auxiliar na escolha da quantidade de clusters, fornecendo um panorama das instâncias) e o não hierárquico K-Means que realizou o agrupamento conforme os objetivos definidos para o estudo. Visualiza-se um dos resultados do estudo em relação ao processo de clusterização na Figura 14. O eixo y representa o quantitativo de instâncias para cada valor da variável e o eixo x indica o valor da variável.

Figura 14 - Gráfico de densidade das médias dos atributos em cada cluster do curso de Pedagogia.



Fonte: (Ramos et al., 2020, p.940)

Os resultados obtidos para o curso de Pedagogia relativos ao Cluster 0 correspondem a 62% do total de estudantes e, conforme a Figura 14, indicam valores baixos e consequentemente revelam a pouca interatividade desses estudantes, o que para o autor pode refletir um baixo desempenho. Já os clusters 1 e 2 representam 38% do total e apresentam valores mais altos e intermediários de interação.

2.5.6. Síntese dos trabalhos relacionados

Quadro 5 - Síntese dos estudos relacionados à pesquisa

Estudo/ Publicação	Natureza/ Fonte de dados	Objetivo/Foco do estudo	Algoritmos de agrupamento	Tipo de grupos formados	Grupos identificados
Monteverde et al. (2018)	Dados históricos de estudantes, Moodle	Formação de grupos de estudantes para atividades de aprendizagem colaborativa em AVA	K-Means	Homogêneos	---
De França et al. (2013)	Dados históricos de estudantes, atividades aplicadas no ensino presencial	Identificação de grupos de estudantes que apresentam dificuldades de aprendizagem.	K-Means	Homogêneos	---
Nalli et al. (2021)	Dados históricos de estudantes, Moodle	Auxílio docente mediante a formação de grupos de estudantes de acordo com o nível de atividade no AVA.	k-means, Mean-Shift Clustering, WARD, DBSCAN, GMM C e SOM.	Homogêneos e Heterogêneos	Alta atividade, média atividade e baixa atividade
Maina et al. (2017)	Dados históricos de estudantes, Moodle	Auxílio docente mediante a formação de grupos de estudantes de acordo com o nível de competência de colaboração.	Skmeans e EM	Homogêneos e Heterogêneos	Alta colaboração, média colaboração e baixa colaboração
Ramos et al., (2020)	Dados históricos de estudantes, Moodle	Identificação de perfis de interação de estudantes na EAD	K-Means e WARD	Homogêneos	Alto, médio e baixo

Fonte: Elaborado pelo autor.

Conforme mostrado no Quadro 5, evidencia-se a predominância de trabalhos com utilização do AVA MOODLE de onde são coletados os dados a serem utilizados. Na maioria dos trabalhos relacionados, constata-se a utilização do algoritmo K-Means como estratégia para agrupamento dos dados. Em geral, os grupos gerados são homogêneos e

resultaram, em sua maioria, em três grupos formados, conforme critérios específicos dos trabalhos.

Os atributos comumente utilizados nos estudos referem-se hegemonicamente a aspectos interativos e colaborativos. Em sua maioria, verificam o comportamento do estudante no AVA conforme apresentado no Quadro 6.

Quadro 6 - Síntese dos atributos Considerados nos estudos

Estudo	Principais atributos Considerados
Monteverde et al. (2018)	Recurso/atividade Quantidade, Dispersão e variância das arestas padrão, avanço e de retorno
França et al. (2013)	Atividade no AVA Notas de atividades Nível da TB
Nalli et al. (2021)	Presença do estudante no AVA: número de acesso à plataforma, porcentagem de estudantes que concluíram a tarefa; frequência total de exibição experimentos em vídeo (média); número de visualizações de vídeos (média); frequência total de visualizações de tutoriais em vídeo (média); número de tutoriais em vídeo visualizados (média);
Maina et al. (2017)	Interação em Fóruns: Número de postagens; Número de respostas; Avaliações do fórum
Ramos et al., (2020)	Quantidade de: Postagens do aluno em fóruns por disciplina, Mensagens enviadas pelo estudante no semestre, Mensagens recebidas pelo aluno no semestre, Acesso por turno: manhã, tarde e noite, no semestre. Colegas diferentes que o estudante enviou mensagens no semestre, Acessos do aluno ao AVA no semestre, Mensagens recebidas de colegas no AVA no semestre,

Fonte: Elaborado pelo autor.

Comparando os trabalhos descritos com este, alguns diferenciais são apontados: (i) a abordagem ActivePlan considera dados de estudantes em turmas remotas por meio do AVA denominado Google Sala de aula (Classroom) e de sistemas administrativos educacionais como o Sistema de Processo Administrativo Eletrônico (SPAEE) denominado SUAP (Sistema Unificado da Administração Pública); (ii) A ActivePlan realiza, após a etapa de pré-processamento dos dados, uma atividade de enriquecimento dos dados. Nesta avaliação, o enriquecimento é realizado por meio da análise cognitiva dos alunos utilizando a Taxonomia de Bloom e do Coeficiente de Rendimento do Estudante (CRE); (iii) Outro diferencial do presente trabalho refere-se à busca pela formação de grupos mistos de estudantes direcionados

para PAAs que podem usufruir de grupos heterogêneos, como a Trezentos e a ABE. Os trabalhos citados, em geral, possuem mecanismos e soluções genéricas e, conforme levantamento de soluções de formação de grupos realizado no estudo de Ferreira et al., (2021), percebe-se a carência de formação de grupos específicos para PAAs. Dessa maneira, a ActivePlan estende estratégias apresentadas nos trabalhos relacionados e inclui a atividade de enriquecimento dos dados de modo direcionado ao contexto de PAAs.

3 ABORDAGEM ACTIVEPLAN

A abordagem apresentada neste trabalho permeia as cinco primeiras fases do processo CRISP-EDM, e adiciona uma fase posterior, referente à geração de grupos de estudantes com perfis heterogêneos para PAAs. Com base nessas fases, a abordagem ActivePlan pode ser compreendida a partir de duas etapas gerais, a saber:

1ª Etapa - Coleta e preparação de dados: Coleta de dados; Pré-processamento dos dados; Enriquecimento do conjunto de dados com atributos indicadores associados ao nível do estudante conforme a TB.

2ª Etapa - Formação de grupos com perfis heterogêneos de estudantes para PAAs: Aplicação de método de AM baseado em agrupamento para geração de clusters homogêneos; Avaliação da qualidade da formação dos clusters; Identificação dos perfis associados a cada cluster; Criação de grupos heterogêneos combinando diferentes estudantes de diversos grupos homogêneos de acordo com a PAA; Avaliação da qualidade da formação dos grupos heterogêneos.

As etapas que compõem a abordagem ActivePlan são descritas a seguir.

3.1 Coleta e preparação dos dados dos estudantes

A abordagem ActivePlan caracteriza-se pela utilização de dados cognitivos e interacionais de estudantes como forma de identificar seus perfis. Os perfis são utilizados para a formação de grupos de estudantes para PAAs. Assim, a primeira etapa da abordagem inclui a coleta de dados manual acerca dos estudantes de uma determinada turma. A ActivePlan recebe dados interacionais e históricos provenientes de fontes como AVAs e de sistemas administrativos de uma instituição educacional. Neste trabalho, são utilizados dados no contexto do ensino remoto oriundos do Instituto Federal da Paraíba (IFPB). Os dados são coletados a partir de interações no AVA Google Classroom e de dados oriundos de um Sistema de Processo Administrativo Eletrônico (SPAEE) denominado SUAP (Sistema Unificado da Administração Pública). Após a coleta e seleção dos dados, são realizadas atividades de pré-processamento, de modo que o conjunto de dados possa ser entrada a um método de AM não supervisionado.

Após os tratamentos necessários sobre os dados de origem, a abordagem ActivePlan inclui uma atividade adicional que é o enriquecimento do conjunto de dados. O

enriquecimento dos dados é obtido a partir da avaliação de atividades aplicadas pelo professor durante o curso ou disciplina em questão conforme a Taxonomia de Bloom (TB) ou Coeficiente de Rendimento do Estudante (CRE). A adição de indicadores associados ao nível da TB ou CRE são incluídos no conjunto de dados por meio de novos atributos.

3.1.1. Indicadores associados ao nível da TB e ao CRE

As questões referem-se a um determinado conteúdo/tema que pode ser trabalhado em atividades escolares, como, por exemplo, sobre “hardware” em uma turma de “Informática Básica”. Rotula-se manualmente as questões de acordo com o nível da TB de forma análoga ao processo realizado por França (2013). Ou seja, a partir das respostas das atividades por estudantes, os estudantes são associados ao seu respectivo nível da TB de acordo com a questão e de seus acertos ou erros. Dessa forma obtém-se um nível da TB para cada estudante referente aos conteúdos/temas associado às atividades. A Figura 15 expõe um exemplo de questão que pode ser rotulada no nível “um” da TB. Neste exemplo, a questão requer que o estudante apenas recorde ou reconheça informações, ideias ou princípios na forma aproximada em que foram aprendidos. Caso o estudante erre a questão, considera-se que o estudante não está associado ao nível da TB avaliado.

Figura 15: Exemplo de questão rotulada ao nível “um” da TB

	entrada (E)	saída (S)	entrada e saída (ES)
Assinale como periférico de entrada (E), de saída (S) ou ainda de entrada e saída (ES) quando for os dois ao mesmo tempo:			
Mouse.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pen drive.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scanner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Joystick (utilizados para ...)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Monitor touchscreen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fonte: Elaborada pelo autor.

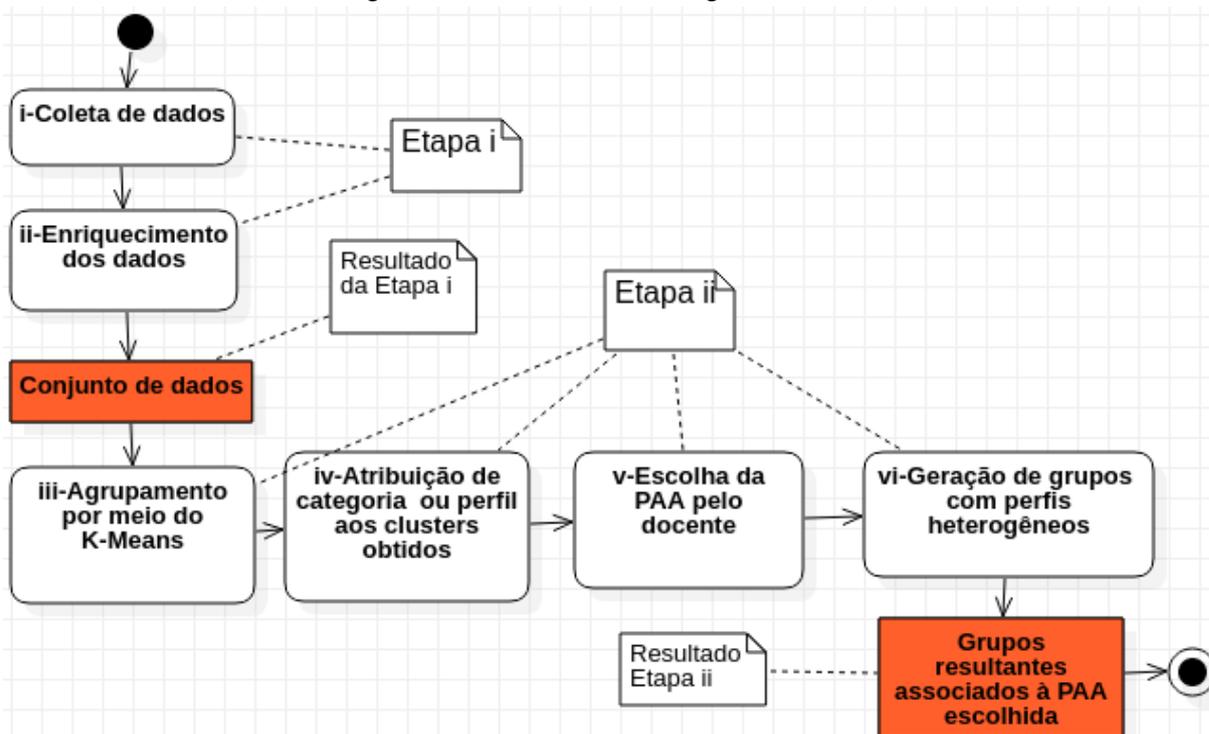
O cálculo corresponde ao coeficiente de rendimento do estudante que é a média geral de notas que engloba as disciplinas do ano anterior.

Como resultado da Etapa 1, um conjunto de dados contendo os perfis de estudantes enriquecidos com os indicadores obtidos por meio da TB ou do CRE é obtido.

3.2 Formação de grupos com perfis heterogêneos de estudantes para PAAs

A Etapa 2 da ActivePlan é descrita por meio do diagrama de atividades mostrado na Figura 16. O diagrama de atividades foi elaborado segundo as diretrizes da UML – *Unified Modeling Language* ou Linguagem de Modelagem Unificada Guedes (2018). Assim, na Etapa 2, inicialmente, mediante um método de AM, criam-se grupos (clusters) de estudantes com perfis similares (III) a partir dos dados coletados (I) e tratados e enriquecidos (II). na 1ª Etapa Os clusters resultantes são formados por estudantes que foram associados considerando a similaridade de seus perfis. Por outro ângulo, cada estudante é associado a uma classe ou categoria (perfil) conforme o cluster atribuído, tendo em vista que clusters podem ser usados como uma estratégia de classificação ou categorização (IV). A partir da identificação do perfil de cada estudante, distribuem-se os estudantes em grupos compostos de diferentes perfis identificados (V) por meio da estratégia proposta neste trabalho (Seção 3.2.4). Ao final da etapa, obtém-se uma lista de IDs dos estudantes com seus respectivos grupos/perfis e outra lista de estudantes com seu grupo heterogêneo para utilização na PAA (VII) escolhida pelo docente.

Figura 16 - Atividades da Abordagem ActivePlan



Fonte: Elaborada pelo autor.

Considerando que as atividades I e II (Etapa 1) foram descritas na seção anterior, detalham-se aquelas que constituem a Etapa 2 nas próximas seções.

3.2.1. Geração de grupos similares por meio do K-Means

A geração de grupos de estudantes a partir de perfis similares é um problema que pode ser solucionado por meio de algoritmos de clusterização, como comentado na Seção 2.3. Para a definição de qual algoritmo utilizar, dentre as várias opções existentes, pode-se fazer um estudo comparativo a respeito deles. Um estudo inicial foi realizado comparando três algoritmos, a saber: Ward, DbSCAN e K-Means. Um resumo dos resultados obtidos na comparação é mostrado no Quadro B.1 (no Apêndice B).

Para a avaliação inicial dos algoritmos, buscou-se analisar os seguintes aspectos: índice de silhueta e balanceamento dos clusters. O índice de silhueta é uma métrica que ajuda a avaliar a qualidade da atribuição de instâncias a clusters (Seção 2.3). Outro fator considerado refere-se ao balanceamento dos clusters (i.e., almeja-se, idealmente, que os clusters de estudantes de perfis similares gerados apresentem equilíbrio razoável no tocante ao montante de estudantes por cluster). Este fator é relevante pois, dependendo da estratégia do próprio método, pode-se resultar em clusters muitos dissimilares, porém, pouco balanceados (e.g., em uma turma composta de 49 estudantes, gera-se três clusters: um cluster composto de 6 estudantes, 3 estudantes e outro cluster composto de 40 estudantes).

O balanceamento dos clusters faz-se necessário considerando-se o passo seguinte que é a formação de grupos heterogêneos de estudantes. Assim, foram realizados alguns experimentos iniciais com os métodos de agrupamento citados. Conforme o Quadro B.1 pode-se analisar que o k-means obteve resultados compatíveis com os objetivos de identificação dos diferentes níveis de perfis já que atribui de forma balanceada os estudantes aos clusters satisfazendo os dois fatores avaliados.

Foi realizado também um estudo específico a respeito do K-Means que derivou em um relatório técnico (Marinho e Souza, 2020) e corroborou para sua escolha como algoritmo a ser utilizado nesta abordagem.

3.2.2. Associação dos clusters a perfis de estudantes

A associação dos clusters a perfis de estudantes é uma atividade relevante para geração de grupos heterogêneos.

A associação de perfis a um cluster pode ser vista como uma atividade de identificação do perfil de grupos de estudante, já que ao final do processo de agrupamento por

similaridade, teremos clusters compostos de estudantes com características/comportamentos similares como, por exemplo, por níveis de desempenho, interação, engajamento, colaboração etc. A identificação do estudante é um passo necessário para que em uma etapa posterior ocorra a formação dos grupos heterogêneos (i.e., ao identificar qual nível de desempenho o estudante se enquadra, torna-se possível a formação de grupos mais diversos, compostos de diferentes perfis, almejando-se ao final colaboração entre os pares).

No que tange a identificação dos perfis gerados no domínio educacional, como ilustração, clusters podem ser classificados quanto a níveis de engajamento (De Oliveira et al., 2022, Oliveira et al., 2020; Moubayed et al., 2020; Macêdo et al., 2020), ou a graus de interação (Pereira et al., 2021; Ramos et al., 2020; Pereira et al., 2017). Estes trabalhos citados usam as categorizações ou grupos para buscar quantificar características dos clusters com base nos atributos do conjunto de dados. Assim, estes trabalhos utilizam-se das médias oriundas das características/atributos do conjunto de dados para cada cluster. Outros trabalhos, além de identificar os perfis associados aos clusters, realizam a ordenação desses clusters com objetivo de delimitar uma escala de categorização (e.g., grupo com perfis mais interativos a grupo com perfis menos interativos). O trabalho de Maina et al (2017) faz essa ordenação e depois a utiliza para gerar grupos com perfis heterogêneos de estudantes. Nalli et al. (2021) também aplicam estratégia semelhante com respeito a clusters com perfis de estudantes considerados mais ativos em um AVA aos grupos com perfis menos ativos. No presente estudo define-se a categorização de clusters conforme o valor da média dos atributos (perfil dos estudantes), com base nas estratégias utilizadas nos trabalhos de Maina et al., (2017) e Nalli et al. (2021). Dessa forma, os clusters são classificados de acordo com o grau de desempenho (Desempenho Muito Elevado (DME) até Desempenho Muito Baixo (DMB)) que é calculado a partir da média de desempenho dos estudantes do cluster, tendo em vista seus valores em cada atributo.

O processo proposto para a associação de perfis de desempenho relacionado a cluster da abordagem ActivePlan é realizado da seguinte forma:

1. Para cada atributo é dado uma pontuação ao cluster (considerando as médias dos valores dos estudantes) conforme exposto na Tabela 2. Para essa atribuição da pontuação de cada cluster verifica-se as médias dos demais cluster referentes ao atributo, a partir disso o cluster recebe uma pontuação compatível com o a magnitude do valor da média. O valor da pontuação varia

conforme a quantidade de cluster (e.g., para número de cluster igual a cinco a pontuação pode variar de 0 até a quantidade de cluster). Este passo é repetido para todos os atributos, gerando-se pontuações referentes a cada atributo para determinado cluster.

2. Ao final do processo realiza-se um somatório de todas as pontuações do cluster referente a cada atributo, esse somatório é representado por *score_do_cluster*. Assim, cada cluster terá uma pontuação geral.

3. Ordena-se os clusters de forma crescente conforme a pontuação geral de cada cluster (*score_do_cluster*). Dessa forma o cluster que obtiver menos pontos ao final é considerado o cluster com o perfil de melhor desempenho e o que obtiver mais pontos é considerado o cluster de pior desempenho. Em caso de empate em relação ao valor do *score_do_cluster* (i.e., valores iguais), considera-se por padrão os atributos referentes ao aspecto cognitivos para o desempate, podendo variar de acordo com a PAA ou a escolha do professor.

4. Assim, clusters são organizados e ranqueados considerando uma escala categórica de grupos de estudantes com desempenho muito elevado até grupos com estudantes com desempenho muito baixo. Ou seja, categorias entre essas duas bordas são definidas: Desempenho Muito Elevado (DME), Desempenho Elevado (DE), Desempenho Moderado (DM), Desempenho Baixo (DB), Desempenho Muito Baixo (DMB).

Para ficar ainda mais claro o funcionamento desse processo, apresenta-se um exemplo hipotético de processo de identificação dos perfis por meio dos clusters, considere a Tabela 2. Neste exemplo, há cinco clusters. Três atributos são considerados (prova, participação e tempo_resposta) e, para cada atributo, é atribuído uma pontuação e calculado a média, referentes aos valores dos estudantes pertencentes aquele cluster.

Tabela 2: Exemplo de identificação de perfil do cluster conforme a média dos atributos

Atributos	Cluster 0		Cluster 2		Cluster 4		Cluster 3		Cluster 1	
	Média	Pontos								
Prova	0.90	0	0.71	1	0.13	4	0.64	2	0.60	3
participação	0.90	0	0.79	1	0.08	4	0.77	2	0.25	3
tempo_rep	0.02	0	0.05	3	0.86	4	0.04	2	0.04	1
<i>score_do_cluster</i>	0		-	5	-	12	-	6	-	7

Fonte: Elaborado pelo autor.

O exemplo apresentado na Tabela 2 diz respeito a um cenário fictício no qual temos três atributos, com intervalo de valores de média de 0 a 1, considerados para identificação dos perfis dos estudantes: Nota da prova (média de notas), nota de participação (média de notas), e tempo de resposta (média do tempo). Nesse exemplo, no que tange a pontuação dos clusters para o atributo “Prova” temos a verificação dos valores: 0.90 que é maior que 0.71, 0.71 que é maior que 0.64, 0.64 que é maior que 6.0 e, seguindo a logica de raciocínio, 6.0 que é maior que 0.13. De acordo com a verificação do maior valor para o menor define-se as seguintes pontuações para cada cluster: zero para cluster 0, pontuação um para o cluster 2, pontuação quatro para o cluster 4, pontuação dois para o cluster 3 e pontuação três para o cluster 1, esse processo também é realizado com os demais atributos. Para cada cluster, é obtido o valor do score geral (*score_do_cluster*) do cluster que é a soma .

De acordo com os valores apresentados na Tabela 2, tem-se:

- O *cluster* zero que pode ser visto como o perfil que detém melhor desempenho, visto que, os estudantes pertencentes ao cluster obtiveram média de 0.90 na prova, 0.90 na participação e responderam mais rapidamente as questões 0.02 (horas). Para o atributo tempo de resposta, quanto menor o tempo de resposta, isso significa que o estudante é mais ativo e engajado.
- Já os estudantes do *cluster* 4 obtiveram pior desempenho na prova com valor médio de 0.13, assim, verifica-se baixa participação no valor médio de 0.08 e eles demoraram a responder as tarefas com tempo de resposta 0.86.

Neste exemplo, considerando os pontos calculados para cada cluster apresentado no *score_do_cluster* (Tabela 2) e níveis de desempenho citados, teríamos como saída, a associação dos clusters aos perfis de níveis de desempenho da seguinte forma: Cluster 0 - DME, Cluster 4 -DE, Cluster 2 -DM, Cluster 3 - DB e Cluster 1 -DMB.

3.2.3. PAA e quantidade de membros por grupos

As PAAs possuem características diferenciadas inclusive ao uso de grupos e seu quantitativo. Também pode haver diferença em relação à quantidade de membros por grupo, conforme descrito na Seção 2.1. Na abordagem ActivePlan, inicialmente as PAA Trezentos e ABE são consideradas. Dessa maneira, com relação à quantidade de membros por grupo, a ActivePlan procede da seguinte forma:

- No caso da PAA trezentos a ActivePlan define o quantitativo de membros por grupo de maneira automática. Formam-se grupos de 5 a 6 membros em turmas com quantitativo de vinte ou mais estudantes e de 3 a 4 membros em turmas menores que vinte estudantes conforme descrito na Seção 2.1.2.2. Isso atende à regra geral definida na prática Trezentos.
- No caso da ABE, o professor deve escolher pelo quantitativo mínimo de cinco até o quantitativo máximo de sete membros por grupo.

No que tange à atribuição de papéis aos membros de um grupo, na PAA Trezentos, os dois membros que obtiverem melhor desempenho são selecionados para o papel “ajudante”. Os demais membros serão considerados “ajudados”. O ajudante com melhor desempenho é definido como “líder”. Já a PAA ABE não requer papéis diferenciados entre os estudantes, então, para esta, não há a definição de líderes ou ajudantes.

3.2.4. Geração de grupos heterogêneos para PAA

Os clusters de estudantes com perfis similares compõem uma das entradas para a estratégia de formação de grupos de estudantes com perfis diferentes proposta nessa dissertação. A estratégia concebida é adaptada do trabalho de Maina et al., (2017) e é baseada em dois algoritmos, a saber: *Algorithm_1* (Figura 17) e *Algorithm_2* (Figura 18), apresentados a seguir.

O algoritmo principal (Figura 17) tem como objetivo receber o conjunto de clusters gerado pelo k-means e retornar grupos específicos com perfis heterogêneos de estudantes para alguma PAA.

Assim, o *Algorithm_1* recebe como entrada os seguintes parâmetros: o dataset incluindo a categoria de cluster para cada instância (estudante), a PAA escolhida, e o quantitativo de membros por grupos a ser considerado para formação. O professor pode escolher esse quantitativo ou indicar “nulo” deixando a ActivePlan definir.

O *Algorithm_1* realiza um loop (linhas 01 a 10) que percorre todos os clusters obtidos. Os dados relativos a cada cluster são armazenados em listas (*all_clusters_objects* e *averages_all_clusters*) para utilização posterior nas funções *activeplan_score* e *activeplan_order* nos passos seguintes (linhas 11 e 12).

Figura 17: Algorithm_1 - Activeplan_main

```
Algorithm_1  
Input: dataset+objects_cluster_labels,paa,groups_member  
Output: groups_paa  
Begin  
//laço que percorre o conjunto de clusters de estudantes com perfis similares  
01: For each cluster ci in Cluster_List do  
02:     For each object in cluster ci do //laço que percorre os objetos do cluster ci  
03:         array_obj_cluster.add(object); //adiciona um objeto do cluster a uma lista  
04:     End For;  
05:     cluster_all_objects.add(array_obj_cluster);  
        //retorna lista com médias referentes a cada atributo dos objetos do cluster ci  
06:     mean_cluster=extract_calc_mean(array_obj_cluster)  
07:     averages_all_clusters.add(mean_cluster)  
08:     array_obj_cluster.clear() //apaga dados do array  
09:     mean_cluster.clear() //apaga dados do array  
10: End For;  
11: score_do_cluster:= activeplan_score(averages_all_clusters) //calcula pontos para os clusters  
12: lista_ids_estudantes:= activeplan_order(score_do_cluster, cluster_all_objects.index)//  
ordena ids de estudantes  
13: groups_paa := activeplan_paa(paa,lista_ids_estudantes, groups_member) //formagrupos  
14: Return groups_paa  
End;
```

Fonte: Elaborado pelo autor.

O loop interno (Linha 02 a 04) tem o objetivo de percorrer todos os objetos (estudantes) de um cluster ci, adicioná-los em uma lista bidimensional denominada array_obj_cluster que contém os índices (IDs) e os valores referentes aos atributos de cada estudante do cluster ci. A variável mean_cluster recebe o retorno da função extract_calc_mean (Linha 06), que tem o objetivo de calcular as médias dos valores de cada atributo referente a todos os estudantes de um cluster ci. A lista unidimensional de médias de um cluster ci denominada mean_cluster é adicionada à lista bidimensional averages_all_clusters (Linha 07) que visa armazenar os valores das médias para cada cluster que servirão de insumo para função activeplan_score. Em seguida, são apagados os valores mean_cluster e array_obj_cluster (Linha 08 e 09) para receber novos valores das médias de outro cluster, em uma nova iteração do loop.

A partir da média dos atributos do cluster tem-se um panorama das instâncias alocadas ao cluster conforme descrito na Seção 3.2.3 e, dessa forma, pode-se associar um perfil ao cluster. Ou seja, a variável averages_all_clusters refere-se a uma lista bidimensional contendo os atributos do dataset referente às instâncias do cluster e seus respectivos valores de média calculados. A função activeplan_score (Linha 11) recebe as médias dos atributos de cada

cluster (averages_all_clusters). A saída da função activeplan_score retorna uma lista denominada score_do_cluster com uma pontuação para cada cluster. A estratégia utilizada na activeplan_score é baseada no somatório de pontos obtidos de acordo com as médias dos atributos de cada cluster. Ao final, tem-se uma lista com a pontuação referente aos clusters.

A função activeplan_order (linha 12) ordena os clusters de acordo com a pontuação obtida de modo decrescente, ou seja, indica aqueles que contêm estudantes que apresentam alto desempenho primeiramente e, por último, aqueles clusters que possuem estudantes com baixo desempenho. Para realização da ordenação, recebe-se a lista de pontuação (score_do_cluster) (Figura 17) para cada cluster e uma lista (cluster_all_objects.index) que contém as identificações (IDs) de cada instância em cada cluster. Gera-se uma lista denominada “lista_ids_estudantes” que remete à ordenação decrescente. Dessa forma, a estratégia definida utiliza a lista_ids_estudantes para atribuir cada instância a um grupo, buscando compor o grupo com estudantes de diferentes níveis de desempenho. A lista ordenada com identificação (IDs) dos estudantes é uma das entradas da activeplan_paa (Figura 18). O parâmetro quantidade de grupos é definido porque é determinante para algumas PAAs (e.g., ABE pode conter de 5 a 7 grupos ou pode ser definido arbitrariamente pelo docente). O Algorithm_2 está especificado na Figura 18.

Figura 18: Algorithm_2 - Algoritmo de geração de grupos para PAA

```
Algorithm_0  
Input: paa, lista_ids_estudantes, groups_member  
Output: groups  
Begin  
  // Função len retorna quantidade de estudantes  
01: number_e := len(lista_ids_estudantes)  
  // Função groups_member_quant retorna o número de membros por grupo de acordo com a PAA  
  // somente para PAA que o quantitativo de membros não foi definido previamente pelo docente  
02: if groups_member is null then  
03:   groups_member = groups_member_quant(number_e, paa)  
04: End If;  
  // Cálculo do número de grupos  
05: number_groups := number_e / groups_member  
  // atribui membros aos grupos (do g0 ao gn)  
06: for g := 0 to number_groups do // laço do grupo 0 até o grupo calculado  
    // percorre a lista com ids dos estudantes  
07:   for i := g to i < number_e step number_groups do  
08:     members.add(lista_ids_estudantes [i]);  
09:   End For;  
10:   groups_other.add(members); // adiciona estudantes na lista de grupos  
11:   members := [] // limpa dados armazenados  
12: End For;  
13: Return groups;
```

```
End activeplan_paa;
```

Fonte: Elaborado pelo autor com base no algoritmo de Maina et al. (2017).

Assim, primeiramente, atribui-se a variável `number_e`, que representa a quantidade de estudantes (linha 01). Na linha 02 verifica-se o valor `groups_member`, caso esta variável apresente valor diferente de *null* significa que o professor realizou a escolha do quantitativo de estudantes. Se o valor for igual a *null*, satisfeita essa condição, a linha 03 é computada, a função `groups_member` recebe valores das variáveis `number_e` de acordo com `paa` (escolhida pelo docente). Na função `groups_member_quant` recebe a PAA e a quantidade total de estudantes, calcula-se automaticamente o número de membros por grupo de acordo com a PAA.

Calcula-se o valor do número de grupos (`number_groups`) mediante a quantidade de estudantes (`number_e`) dividida pela quantidade de membros por grupo (`groups_member`), obtendo-se o número de grupos (linha 05). Percorre-se todos os grupos (Linha 06), e aloca-se os estudantes (linha 07 até a 11) aos grupos, tendo o primeiro membro de cada grupo como o líder do grupo, o segundo membro dos grupos receberão o papel de ajudante juntamente com o líder do grupo.

3.3 ActivePlan versus trabalhos relacionados

O diferencial da abordagem ActivePlan quando comparada aos trabalhos relacionados apresentados na Seção 2.5 refere-se aos atributos considerados que, além de atributos referentes ao nível de interação, utilizam-se dados relacionados a aspectos cognitivos. Quanto à quantidade de membros por grupos e características de grupos formados, a ActivePlan busca formar grupos de modo a atender a uma PAA escolhida diferenciando-se das demais propostas apresentadas na Seção 2.5 e dos mecanismos apresentados no Anexo II.

Pode-se observar mais detalhes na comparação entre a ActivePlan e os trabalhos relacionados na Tabela A1 (Apêndice A).

4 Implementação e Experimentos

Os resultados obtidos neste trabalho incluem o desenvolvimento de uma ferramenta que assiste a abordagem ActivePlan e aqueles alcançados nas avaliações experimentais realizadas para responder às questões de pesquisa. Neste capítulo, a Seção 4.1 descreve a ferramenta implementada, a Seção 4.2 discorre sobre o dataset original coletado e o enriquecimento realizado como contribuição da ActivePlan, e a Seção 4.3 apresenta os experimentos realizados para avaliar a abordagem.

4.1 Ferramenta ActivePlan

Visando tornar todo o processo da ActivePlan acessível e transparente ao usuário (professores e/ou pesquisadores), desenvolveu-se uma ferramenta. A ferramenta foi implementada como uma aplicação WEB, desenvolvida a partir da linguagem python¹, mediante framework django². A ferramenta provê dois perfis de usuários para acesso às funcionalidades: (i) perfil do pesquisador e (ii) perfil do professor.

O módulo da ferramenta desenvolvido para o perfil pesquisador foi especificado com o intuito de auxiliar esta pesquisa na avaliação da abordagem. Dessa forma, por meio do perfil pesquisador, é possível usar a ferramenta para realizar as avaliações experimentais formuladas com base nos objetivos e questões desta pesquisa. Nesse contexto, a ferramenta possibilita acessar resultados das métricas referentes à composição de grupos similares (clusters), e de grupos heterogêneos gerados e vinculados a PAAs. O objetivo deste módulo do pesquisador é, principalmente, prover, de modo mais simples, a avaliação dos resultados dos algoritmos quanto a seu desempenho e qualidade dos grupos formados.

No caso da opção de perfil professor, a ferramenta almeja prover suporte ao usuário professor que vai, de fato, utilizar a ferramenta como apoio a suas práticas docentes. A Figura 19 apresenta um diagrama com o fluxo das atividades desenvolvidas na ferramenta

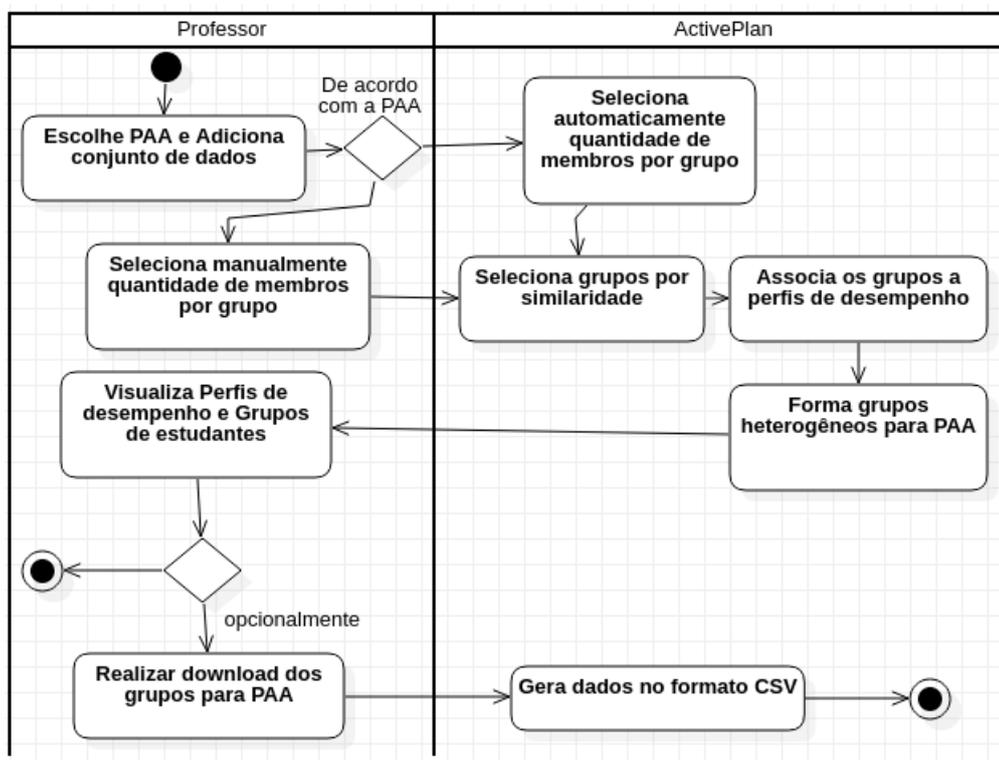
¹Disponível em:<<https://www.python.org/>>. Acessoem: 07 maio de 2022.

²Disponível em:<<https://docs.djangoproject.com/pt-br/3.2/>> . Acessoem: 07 maio de 2022.

ActivePlan no que tange às tarefas (ações) associadas ao usuário professor e as ações retornadas pela ferramenta. O diagrama de atividades segue a notação UML.

O professor é responsável por algumas configurações importantes no processo como, por exemplo, a escolha da PAA a ser considerada.

Figura 19: Diagrama de atividades da ferramenta ActivePlan para o perfil professor.



Fonte: autoria própria.

Assim, percebe-se (Figura 19) que as ações iniciam-se a partir de dados informados pelo docente. Listam-se as atividades presentes no diagrama de atividades (Figura 19) que retratam as etapas básicas do processo da ActivePlan quando o docente interage:

1. O Professor escolhe a PAA e realiza o upload de um arquivo com os dados de estudantes de uma determinada turma (CSV);
2. A ferramenta ActivePlan agrupa os estudantes (instâncias) por similaridade de perfis;
3. A ActivePlan associa cada cluster (grupo) obtido conforme um critério de categorização (e.g., nível de desempenho, no caso da PAA Trezentos e da ABE);
4. A ferramenta forma grupos heterogêneos para a PAA escolhida;
5. O professor visualiza os grupos de estudantes formados e seus perfis heterogêneos para a PAA em questão;
6. Opcionalmente o professor pode realizar o *download* dos grupos heterogêneos gerados por meio de um conjunto de dados em CSV.

Um fragmento de um conjunto de dados (no formato CSV) com dados de estudantes de uma determinada turma, um dos cenários considerados na abordagem ActivePlan, é apresentado na Figura 20.

Figura 20: Fragmento do conjunto de dados de estudantes.

Estudante	Q1	Q2	T1	T2	S1	S2	TB1	TB2	TB3	TB4	TB5
[nome]	95	100.0	36.0	24.0	5.0	5.0	1	1	1	1	1
[nome]	80	100.0	36.0	23.0	5.0	4.0	1	1	1	1	1
[nome]	50	93.33	8.0	1.0	4.0	4.0	1	1	0	1	1
[nome]	95	100.0	36.0	24.0	5.0	6.0	1	1	1	1	1
[nome]	35	50.0	39.0	17.0	99.0	99.0	1	0	0	1	0
[nome]	40	90.0	9.0	5.0	5.0	5.0	1	1	1	1	0
[nome]	80	90.0	36.0	24.0	5.0	4.0	1	1	0	1	1
[nome]	25	90.0	10.0	0.0	6.0	1.0	1	1	1	1	0
[nome]	85	70.0	6.0	7.0	6.0	4.0	0	1	0	1	1

Fonte: autoria própria.

A interface principal da ferramenta é mostrada na Figura 21a. Ela indica a possibilidade de seleção de perfil de usuário, como comentado em professor ou pesquisador. Apresentam-se na Figura 21b detalhes da configuração a ser feita pelo professor, o que inclui a informação da existência de atributos referentes à taxonomia de bloom (iii), escolha de qual PAA deve ser considerada (iv) e, dependendo da PAA, requiere-se a inserção da quantidade de membros por grupo (v). Caso o professor já tenha definido a quantidade de membros por grupo, este pode inseri-las diretamente (vii). Por fim, o professor deve indicar o conjunto de dados a ser utilizado (vii).

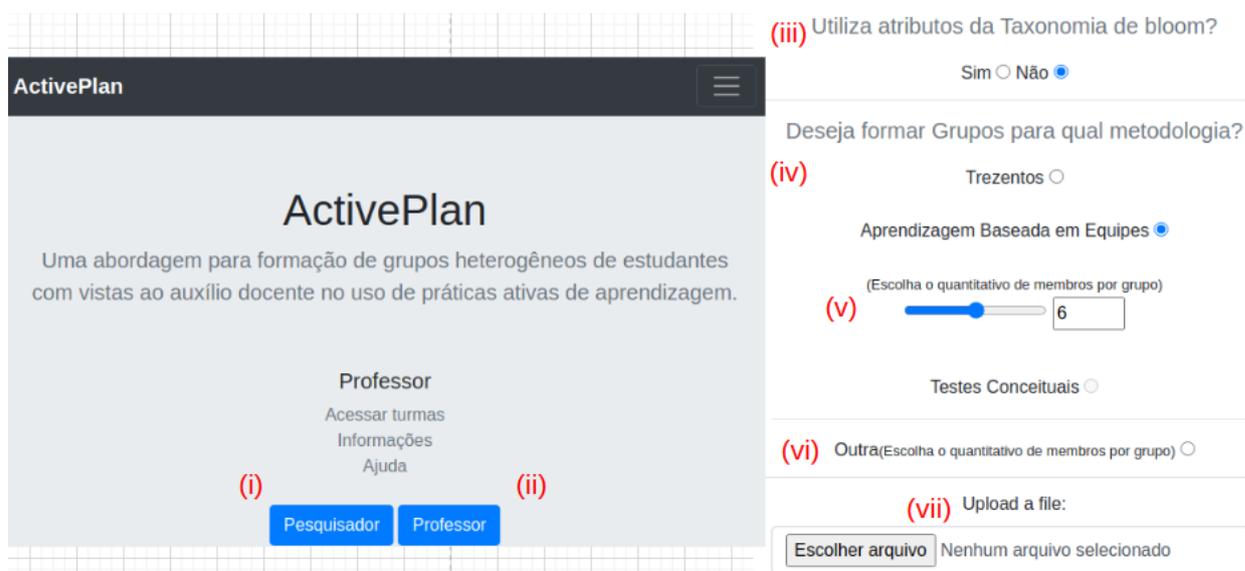


Figura 21a e 21b: a) Tela principal da interface da ActivePlan; b) Opções para o módulo do professor.

Fonte: autoria própria.

4.2 Fontes e conjuntos de dados

Para este trabalho, foram analisadas algumas possibilidades de fontes de dados como AVAs e outros meios para a coleta de dados de perfis de estudantes. O perfil esperado pela abordagem ActivePlan inclui dados associados ao domínio cognitivo (COG) e interacional (INT) do estudante.

O trabalho de Souza et al. (2020) contempla dados de interação de estudantes em turmas com execução em modalidade de ensino remoto. O trabalho inclui também alguns aspectos cognitivos dos estudantes. Tendo em vista que o trabalho citado gerou um dataset, optou-se por utilizar este dataset como base para a avaliação da ActivePlan.

O estudo de Souza et al. (2020) coletou e integrou dados a partir das seguintes fontes presentes no IFPB: SUAP (com 492 instâncias de estudantes) e Google classroom (488 instâncias de estudantes). Assim, os dados oriundos do SPAE e do AVA resultaram em dois *datasets* que contém dados pertencentes a turmas do ensino médio integrado ao curso técnico do Campus Monteiro do IFPB referentes aos cursos Técnico de Edificações, Manutenção e Suporte em Informática e Edificação. Os dados coletados correspondem às disciplinas de filosofia e informática básica, descreve-se sucintamente o montante de estudantes ao respectivo conjunto de dados do estudo de Souza et al. (2020):

- **Conjunto de dados do primeiro ano:** As turmas referentes ao primeiro ano correspondem ao montante de 259 instâncias, 125 estudantes da disciplina de informática básica e 134 da disciplina de filosofia.
- **Conjunto de dados do segundo e terceiro ano:** As turmas referentes ao segundo e terceiro ano dizem respeito apenas à disciplina de filosofia, não presencia-se registro de instâncias pertencentes à disciplina de informática básica. O conjunto inclui 128 estudantes do segundo ano e 101 estudantes do terceiro ano, totalizando o montante de 229 instâncias de estudantes.

4.2.1 Conjunto de dados 1 - Estudantes de turmas de primeiro ano pertencentes à disciplina de Informática Básica

Das 259 instâncias do conjunto de dados das turmas de primeiro ano selecionou-se um

subconjunto referente à disciplina de Informática Básica, totalizando 125 estudantes.

A Tabela 3 mostra uma descrição do primeiro *dataset* usado neste trabalho. As linhas em branco indicam os metadados oriundos do trabalho de Souza et al. (2020). As linhas em cinza são referentes aos atributos derivados no presente trabalho.

Tabela 3: Metadados estruturais do dataset original juntamente com metadados obtidos após enriquecimento

Atributos	Descrição	Domínio	Fonte	Tipo	Intervalo de valores
Q1	Nota da primeira atividade	COG	AVA	numérico	[0,0;100]
Q2	Nota da segunda atividade	COG	AVA	numérico	[0,0;100]
T1	Tempo de envio da primeira avaliação (em dias)	INT	AVA	numérico	[00;99]
T2	Tempo de envio da segunda avaliação (em dias)	INT	AVA	numérico	[00;99]
S1	Último acesso da primeira avaliação (tempo em dias)	INT	SPAE	numérico	[00;99]
S2	Último acesso da segunda avaliação (tempo em dias)	INT	SPAE	numérico	[00;99]
TB1	Ausência ou não da competência no tema 1	COG	ENR	booleano	{0,1}
TB2	Ausência ou não da competência tema 2	COG	ENR	booleano	{0,1}
TB3	Ausência ou não da competência tema 3	COG	ENR	booleano	{0,1}
TB4	Ausência ou não da competência tema 4	COG	ENR	booleano	{0,1}
TB5	Ausência ou não da competência tema 5	COG	ENR	booleano	{0,1}

Fonte: autoria própria.

As notas das atividades ao longo da disciplina (Q1 e Q2) são relevantes para a avaliação de desempenho escolar (Souza et al., 2020; Nalli et al., 2021; Oyelade et al., 2010; Chanamarn et al., 2017). Pode-se, por exemplo, identificar precocemente o desempenho insatisfatório do estudante.

Dados de acessos (S1 e S2) dos estudantes, obtidos em logs oriundos do servidor, referem-se ao número de acessos dos estudantes à plataforma na semana da realização das tarefas. São especificações de dados similares aos estudos de Ramos (2020) e Pereira (2018) que utilizam logs de acesso a sistemas visando identificar o grau de interação ou engajamento do estudante.

Para Pereira (2018), a entrega de uma atividade, quando ocorre antes do prazo, aponta para maiores chances de aprovação do estudante do que daqueles que entregam já no final do prazo. Em geral, os estudantes que fazem todas as atividades possuem média final maior que os demais. Nesta perspectiva considera-se o tempo de envio (T1 e T2) relacionado às entregas de atividades. Q1 e Q2 correspondem às notas aplicadas no ensino remoto, referindo-se ao período de envio entre o dia 24/08/2020 a 25/10/2020.

Adicionalmente aos atributos já existentes, foram especificados alguns indicadores associados por meio da Taxonomia de Bloom. Isso foi realizado manualmente por meio de questões que indicam o nível da TB para um dado estudante. O objetivo é situar o estudante

de acordo com seu desempenho em cada conteúdo. Essa definição é baseada no trabalho de Neto et al., (2016) e De França et al., (2013) que utilizaram a TB para identificação de estudantes com dificuldades de aprendizagem. Assim, os atributos TB1, TB2, TB3, TB4 e TB5 assumem valores booleanos (possui ou não possui).

Os atributos TB1 a TB5 dizem respeito aos conteúdos que recebem valores associados aos objetivos educacionais avaliados nas atividades. Neste caso, os atributos são referentes às duas semanas iniciais, para a realidade da disciplina “Informática Básica”, com os seguintes conteúdos:

- Conteúdo TB1: “Histórico”
- Conteúdo TB2: “Periféricos”
- Conteúdo TB3: “Hardware”
- Conteúdo TB4: “Memória”
- Conteúdo TB5: “CPU (Central Processing Unit)”

Durante o processo de enriquecimento dos dados, observou-se que os níveis da TB referentes às questões das atividades enquadram-se no nível “um”. Uma das possíveis explicações para isso pode estar na natureza da disciplina e no fato das atividades decorrerem nas duas primeiras semanas. Ou seja, as atividades são focadas ainda em conceitos básicos e teóricos/introdutórios.

Ressalta-se que uma das limitações relativas ao enriquecimento dos dados remete aos objetivos educacionais para cada questão rotulada referente às atividades Q1 e Q2, podendo ou não divergir do nível da TB atribuído à questão pelo docente.

4.2.2 Conjunto de dados 2 - Estudantes do segundo e terceiro ano

O outro dataset, mostrado na Tabela 4, inclui dados de estudantes do segundo ano e terceiro ano do ensino médio integrado ao técnico. Neste dataset, dados das notas de estudantes em disciplinas do ano anterior são utilizadas visando-se corroborar para identificação do perfil do estudante de acordo com suas notas nos anos anteriores.

Em revisão sistemática da literatura, Oliveira et al. (2019) indicam que a maioria dos trabalhos que buscam formar grupos para ambientes escolares são focados na formação de grupos heterogêneos. Dentre as características do estudante consideradas, o conhecimento prévio do estudante é uma das mais utilizadas entre os estudos mapeados.

Dessa forma utiliza-se a nota das disciplinas do ano anterior no conjunto de dados visando à identificação do conhecimento prévio do estudante.

Esse dataset refere-se a estudantes dos Cursos Técnico de Edificações, Manutenção e Suporte em Informática e Edificações totalizando 229 estudantes.

Tabela 4: Metadados estruturais do dataset original juntamente com médias de notas anteriores (1o ano) de turmas do segundo e terceiro ano

Atributos	Descrição	Domínio	Fonte	Tipo	Intervalo de valores
Q1	Nota da primeira atividade	COG	AVA	numérico	[00;99]
Q2	Nota da segunda atividade	COG	AVA	numérico	[00;99]
T1	Tempo de envio da primeira avaliação (em dias)	INT	AVA	numérico	[00;99]
T2	Tempo de envio da segunda avaliação (em dias)	INT	AVA	numérico	[00;99]
S1	Último acesso da primeira avaliação (tempo em dias)	INT	SPAE	numérico	[00;99]
S2	Último acesso da segunda avaliação (tempo em dias)	INT	SPAE	numérico	[00;99]
CRE	Média dos históricos das notas em disciplinas	COG	SPAE	numérico	[00;99]

Fonte: autoria própria.

O atributo CRE presente na Tabela 4 corresponde ao coeficiente de rendimento estudante que é a média geral de notas que engloba as disciplinas de: biologia, educação física, filosofia, física, geografia, história, língua portuguesa, matemática, química e sociologia, cursadas nos anos anteriores daquela turma.

4.3 Experimentos

A avaliação realizada por meio de experimentos busca analisar os resultados obtidos pela abordagem ActivePlan quanto às questões de pesquisa formuladas neste trabalho (Seção 1.2):

(Q1) Considerando perfis de estudantes, caracterizados por aspectos cognitivos e interacionais, pertencentes a turmas ofertadas em ensino remoto, como identificar perfis similares?

(Q2) A partir da identificação de perfis similares de estudantes, como associá-los a grupos, dessa vez, compostos de estudantes com perfis heterogêneos, de modo que os grupos formados possam ser usados em PAAs?

Os cenários escolhidos para os experimentos foram definidos com base nos dois conjuntos de dados descritos na seção anterior e na escolha de duas PAAs: (i) Trezentos e (ii) Aprendizagem baseada em Equipes (ABE). Para a avaliação da abordagem ActivePlan e com

o objetivo de responder as questões de pesquisa Q1 e Q2, foram planejados os seguintes experimentos, conforme os cenários definidos:

- **Experimento 1:** Objetiva verificar o nível de similaridade de perfis de estudantes agrupados em clusters, ou seja, responder à Q1.
- **Experimento 2:** Busca analisar a diversidade de grupos formados com perfis distintos de estudantes com respeito à PAA Trezentos e à PAA ABE, dessa forma, respondendo à Q2.

Os resultados do Experimento 1 são apresentados nas Seções 4.3.1.1 e 4.3.1.2, já os referentes ao Experimento 2 são abordados nas Seções 4.3.2.1 e 4.3.2.2.

Os experimentos foram realizados em máquina local, cujas configurações de hardware incluem: Processador Intel, 8ª geração, core i5-8250U, 8GB de memória RAM, 1 Terabyte de armazenamento. Para a realização de atividades de modelagem do aprendizado (por meio de algoritmo de *clusterização*) e avaliação mediante métrica silhouette, utilizou-se a biblioteca *scikit learn*³. Para a implementação de cálculos matemáticos (e.g., cálculo de distância entre instâncias) utilizou-se a biblioteca *scipy*⁴. Na exibição de alguns resultados, utilizou-se as bibliotecas *matplotlib*⁵ e *seaborn*⁶. Para auxílio no tratamento, enriquecimento e manipulação dos dados, utilizou-se as bibliotecas *numpy*⁷ e *pandas*⁸.

4.3.1 Experimento 1- Clusters de estudantes

Para a geração de clusters de estudantes, utiliza-se o algoritmo de agrupamento partitivo k-means. Para definir o número de clusters a ser obtido, utiliza-se o método *Elbow* (ou método do cotovelo), conforme introduzido em seções anteriores. Os dois cenários experimentais são descritos e discutidos a seguir.

4.3.1.1 Cenário 1.1 - Avaliação de Clusters de estudantes para turmas de primeiro ano

Este experimento tem como foco avaliar o processo de geração de grupos (clusters) com perfis de estudantes similares, utilizando-se o dataset referente às **turmas de primeiro ano**. Para este dataset o método *Elbow* indicou um quantitativo de **cinco clusters** a ser utilizado, como mostra a linha vertical pontilhada na Figura 22a.

³ Disponível em: <https://scikit-learn.org/stable/modules/clustering.html#clustering> . Acesso em: 07 maio de 2022

⁴ Disponível em: <http://scipy.org> . Acesso em: 07 maio de 2022

⁵ Disponível em: <https://matplotlib.org/> . Acesso em: 07 maio de 2022

⁶ Disponível em: <https://seaborn.pydata.org/> . Acesso em: 07 maio de 2022

⁷ Disponível em: <https://numpy.org/>. Acesso em: 07 maio de 2022

⁸ Disponível em: <https://pandas.pydata.org/> . Acesso em: 07 maio de 2022

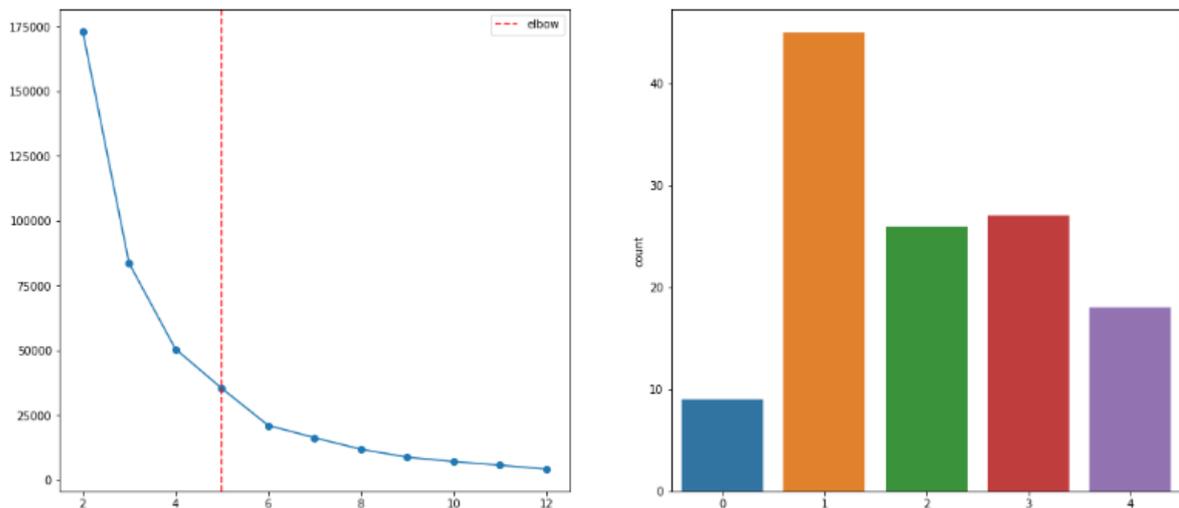


Figura 22: a) Número de clusters ideal mediante Elbow - Experimento 1.1, b) Quantitativo de estudantes por cluster - Experimento 1.1

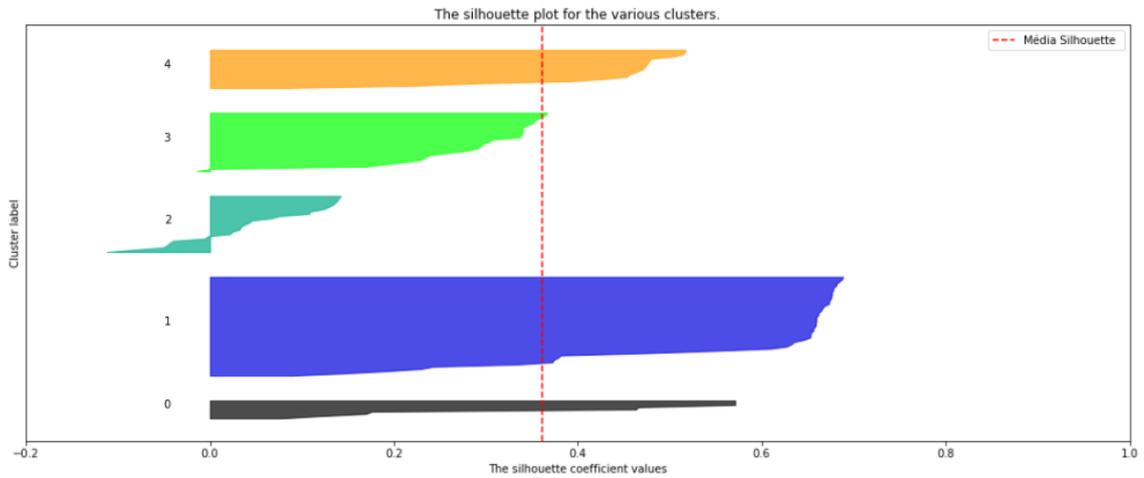
Fonte: autoria própria.

Usando então a configuração de cinco clusters, o k-means foi executado e obteve-se os resultados mostrados na Figura 22a. A Figura 22b mostra também o quantitativo de estudantes alocados em cada cluster. Nota-se uma maior concentração de estudantes no Cluster 1 (total de 45) e menor alocação de estudantes no Cluster 0 (total de 9).

A qualidade da atribuição das instâncias a um determinado cluster foi mensurada mediante o índice de silhueta (Figura23). A média da silhueta aponta para um valor próximo a quatro, o que pode indicar que o agrupamento por similaridade foi realizado de forma satisfatória para alguns clusters e insatisfatória ou inconclusiva na alocação de instâncias de outros clusters.

O Cluster 1 obteve um índice silhueta com valor próximo a 0.7. Assim, verifica-se, para este cluster, a consistência na seleção de estudantes no que tange à similaridade intra-cluster (quanto mais próximo a um, mais coesa foi a alocação da instância ao cluster). Outros agrupamentos, como o Cluster 0 e o Cluster 4 revelaram-se também consistentes, chegando a valores próximos a 0.6 (acima do limiar padrão de 0.5).

Figura 23: Métrica Silhouette - Experimento 1.1



Fonte: autoria própria.

Já o Cluster 3 apresenta valores mais próximos de zero (valores inconclusivos) e no Cluster 2 presenciaram-se valores negativos e próximos a zero, o que indica que alguns dos estudantes pertencentes a esse cluster podem estar erroneamente alocados a ele. Uma amostra dos estudantes conforme sua alocação nos clusters é mostrada na Tabela 5. A amostra indica uma variação de valores dos atributos (perfil) dos estudantes. Nesta perspectiva percebe-se que o estudante pertencente ao Cluster 0 apresenta valores nulos ou próximos a zero nos atributos Q1, Q2, TB1, TB2, TB3, TB4, TB5 (domínio cognitivos) e valores altos em T1, T2, S1 e S2 (domínio interacional), o que pode indicar que o estudante apresenta problemas no processo de ensino e aprendizagem.

Tabela 5: Amostra de resultados da atribuição de um estudante de cada cluster - Experimento 1.1

Estudantes	Q1	Q2	T1	T2	S1	S2	TB1	TB2	TB3	TB4	TB5
Estudante do Cluster 0	0.00	0.00	0.88	0.99	0.02	1.00	0	0	0	0	0
Estudante do Cluster 1	1.00	1.00	0.30	0.24	0.02	0.05	1	1	1	1	1
Estudante do Cluster 2	0.45	0.56	0.05	0.00	0.04	0.03	1	0	1	1	0
Estudante do Cluster 3	0.70	0.73	0.06	0.05	0.00	0.05	0	1	1	1	1
Estudante do Cluster 4	0.50	0.93	0.30	0.05	0.02	0.06	1	1	1	1	1

Fonte: autoria própria.

Por outro lado, o estudante do Cluster 1 apresenta altos valores nos atributos cognitivos e baixos valores nos atributos interacionais, o que pode estar associado a um bom desempenho no processo de ensino e aprendizagem. Com valores similares ao estudante do Cluster 1, o estudante do Cluster 4 demonstra um bom desempenho, com exceção do valor do

atributo Q1 que apresenta baixo valor. Em geral, o estudante do Cluster 3 revela valores intermediários, e o estudante do Cluster 2 indica valores baixos.

4.3.1.2 Cenário 1.2 - Clusters de estudantes das turmas de segundo e terceiro ano

Para a realização deste experimento, utiliza-se o dataset referente aos estudantes do **segundo e terceiro ano da disciplina de filosofia**. Visando a definição do número de clusters, foi aplicado o método Elbow, que indicou também um quantitativo de cinco clusters a ser utilizado, como mostra a linha vertical pontilhada na Figura 24a.

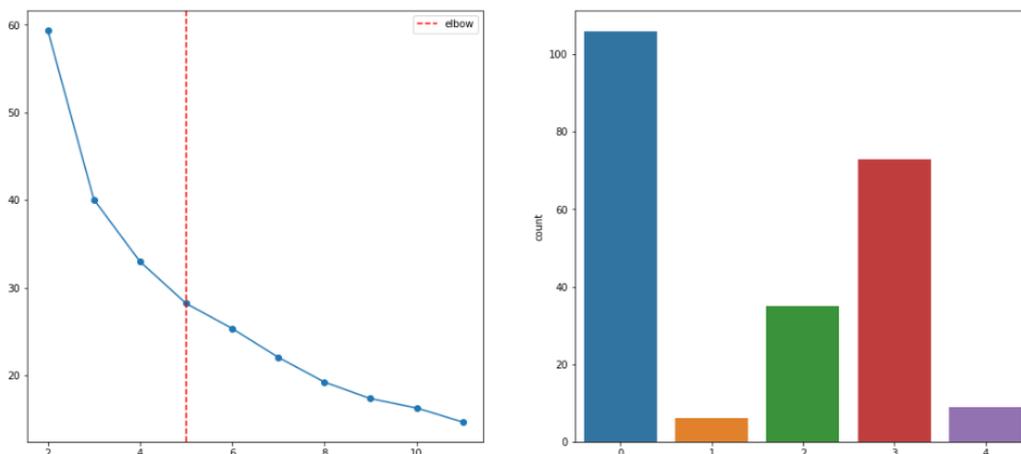


Figura 24: a) Número de clusters ideal mediante Elbow - Experimento 1.2, b) Quantitativo de estudantes por cluster - Experimento 1.2

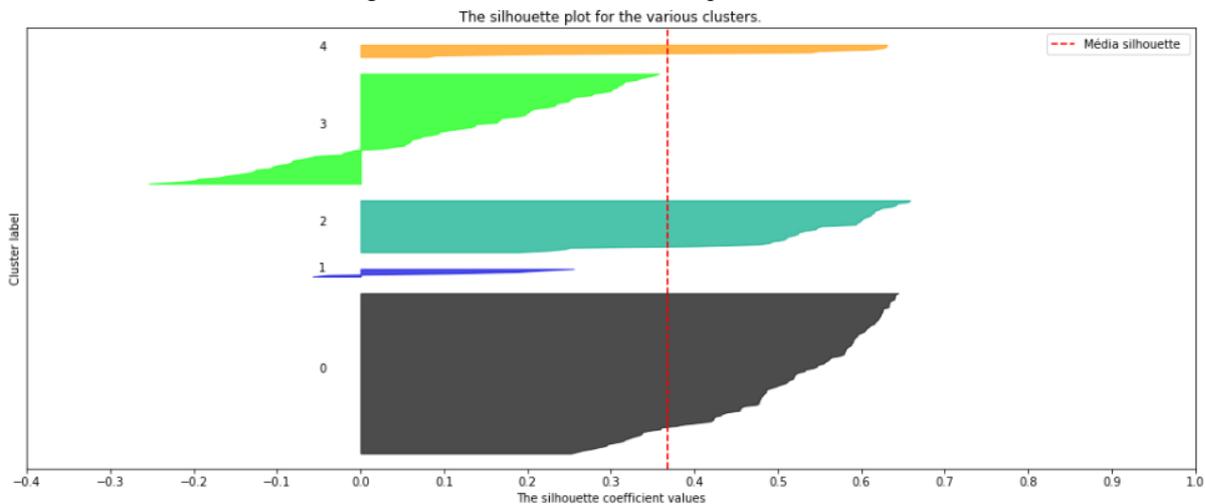
Fonte: autoria própria.

Usando então a configuração de cinco clusters, o k-means foi executado e obteve-se a distribuição de estudantes conforme mostrado na Figura 24a. A Figura 24b mostra também o quantitativo de estudantes alocados em cada cluster. Nota-se uma maior concentração de estudantes no cluster 0 (total de 106) e menor alocação de estudantes no cluster 1 (total de 6).

A qualidade da atribuição das instâncias a um determinado cluster foi mensurada mediante o índice de silhueta (Figura 25). A média da silhueta aponta para o valor 0.37, próximo a quatro, o que pode indicar que o agrupamento por similaridade foi realizado de forma satisfatória para alguns clusters e insatisfatória ou inconclusiva na alocação de instâncias de outros clusters. Dessa forma, o Cluster 4, 2 e 0 obteve um índice de silhueta próximo a 0.6. Assim, verifica-se, para estes clusters, uma consistência na seleção de estudantes no que tange à similaridade intra-cluster (quanto mais próximo a um, mais coesa

foi a alocação da instância ao cluster). Em sua maioria, a média do silhouette apresenta valores acima do limiar padrão de 0.5.

Figura 25: Métrica silhouette - Experimento 1.2



Fonte: autoria própria.

Percebe-se que parcela dos índices de silhueta apresentam valores negativos associados aos Cluster 3 e Cluster 1, o que indica que alguns estudantes pertencentes a esses clusters possivelmente foram erroneamente alocados. A Tabela 6 apresenta uma amostra de estudantes, com seus respectivos clusters e perfis definidos por meio de seus atributos e respectivos valores.

Tabela 6: Amostra de resultados da atribuição de um estudante de cada cluster - experimento 1.2

Estudantes	Q1	Q2	T1	T2	S1	S2	CRE
Estudante do Cluster 3	0.63	0.7	0.03	0.04	0.05	0.03	0.48
Estudante do Cluster 2	0.63	0.90	0.04	0.04	2.00	0.04	0.40
Estudante do Cluster 0	1.00	0.80	0.01	0.04	0.03	0.04	0.76
Estudante do Cluster 4	0.00	0.00	1.00	1.00	1.00	1.00	0.17
Estudante do Cluster 1	0.60	0.00	0.06	1.00	0.05	1.00	0.65

O estudante do Cluster 0 (Tabela 6) apresenta altos valores nos atributos cognitivos (notas mais elevadas) e baixos valores nos atributos interacionais (Tempo do último acesso ao AVA e tempo de envio de resposta). Esse estudante detêm valor da média do histórico acadêmico mais elevado que os demais. Tais características podem estar associadas a um bom desempenho no processo de ensino e aprendizagem do estudante pertencente ao Cluster 0. O estudante do Cluster 2 e o estudante do Cluster 3 apresentam valores intermediários. Por outro lado, o estudante do Cluster 3 e o estudante do Cluster 4 apresentam valores mais baixos em relação aos demais.

4.3.2 Experimento 2 - Grupos heterogêneos de estudantes para PAAs

Os experimentos desta seção são realizados a partir dos resultados obtidos na etapa implementada e avaliada na Seção 4.3.1. Nesta seção, o foco é avaliar a geração de grupos de estudantes com perfis heterogêneos de modo que esses grupos possam favorecer a aplicação de PAAs. É importante destacar a necessidade de avaliar os grupos e as características predominantes dos estudantes em cada cluster com respeito ao **nível de desempenho e, a partir dos diferentes níveis, atribuir os estudantes aos grupos com foco nas PAAs**. O critério associado ao nível de desempenho do estudante é utilizado tanto para a PAA Trezentos quanto para a ABE, práticas consideradas neste experimento.

4.3.2.1 Experimento 2.1 - Grupos heterogêneos de estudantes resultantes da turma de primeiro ano

Neste experimento, considera-se o conjunto de dados referentes ao dataset da turma de primeiro ano juntamente com os resultados da clusterização (Seção 4.3). No que tange ao reconhecimento dos perfis de cada cluster, isso pode ser mostrado por meio da Tabela 7 que aponta as médias para cada uma das características/atributos dos estudantes de acordo com o cluster identificado.

Tabela 7 : Ranking dos clusters a partir das médias de cada atributo - - experimento 2.1

Atributos	Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Média	Pontos								
Q1	0.09	4	0.91	0	0.70	1	0.58	2	0.47	3
Q2	0.23	4	0.93	0	0.66	3	0.76	2	0.90	1
T1	0.68	4	0.14	3	0.12	2	0.10	1	0.06	0
T2	0.62	4	0.14	3	0.09	2	0.07	1	0.04	0
S1	0.67	4	0.05	1	0.07	3	0.06	2	0.03	0
S2	0.89	4	0.04	2	0.11	3	0.04	1	0.03	0
TB1	0.11	3	1.00	0	0.46	2	0.00	4	1.00	0
TB2	0.11	3	1.00	0	0.0	4	1.00	0	1.00	0
TB3	0.00	4	0.93	0	0.81	1	0.67	2	0.56	3
TB4	0.33	4	0.84	1	0.69	3	0.81	2	0.89	0
TB5	0.00	3	1.00	0	0.77	2	0.92	1	0.0	3
<i>score_do_cluster*</i>		41	-	10	-	26	-	18	-	10

Fonte: Elaborados pelos autores.

**score_list*: Variável destinada ao armazenar o somatório dos valores associados aos pontos obtidos para cada cluster conforme estratégia exposta Seção 3.2.

A Tabela 7 propicia uma visão do panorama dos perfis dos estudantes conforme os grupos obtidos e indica uma pontuação para cada atributo vs cluster. A análise das médias e das

pontuações visam identificar o perfil de desempenho dos estudantes de um cluster e classificá-lo em relação aos demais conforme descrito na seção 3.2.2.

Considerando os valores apontados na Tabela 7, pode-se identificar o seguinte perfil para cada um dos clusters:

- **Cluster 1:** com maior quantitativo de estudantes (45), este cluster contém o perfil de estudantes que apresentaram maior desempenho nos questionários Q1 e Q2, que respondem rapidamente às atividades passadas e também que acessam mais ativamente o AVA. Em geral, os estudantes demonstram conhecimentos nas TB analisadas (Tabela 7). Este Cluster é classificado como **Desempenho Muito Elevado (DME)**. Isso significa que em todos os atributos analisados dos estudantes oriundos do Cluster 1, em sua maioria, os estudantes obtiveram melhores notas e não apresentaram ausência em nenhuma das competências analisadas;
- **Cluster 0:** com 9 estudantes, inclui perfis de estudantes que apresentam o pior desempenho detectado, com destaque aos valores insatisfatórios nos questionários Q1 e Q2. Em geral, os estudantes demoram mais para responder às atividades (T1 e T2) e para acessar o AVA (S1 e S2). Percebe-se adicionalmente defasagem dos estudantes quanto às TB analisadas. Portanto, devido à defasagem descrita, pode-se categorizá-lo como **Desempenho Muito Baixo (DMB)**.
- **Cluster 4:** com um montante de 18 estudantes, uma das características presentes no Cluster 4 refere-se ao alto desempenho no segundo questionário (Q2) e baixo desempenho no primeiro questionário (Q1). Tais características podem apontar um perfil de estudantes em evolução. Observa-se ausência quanto à TB5. O cluster foi classificado em **Desempenho Elevado (DE)**, ou seja, o Cluster 4 apresenta um perfil de estudante em desenvolvimento, contudo demonstra alto desempenho na Q2 e nas TBs.
- **Cluster 2:** obtendo 26 estudantes, o Cluster 2 contém perfis que correspondem a desempenhos intermediários em ambos os questionários Q1 e Q2. Quanto aos acessos e tempo de resposta das atividades (T1, T2, S1 e S2) apresentam, em geral, resultados abaixo dos encontrados em estudantes dos clusters 0, 1 e 2. Observa-se ausência quanto à TB2.
- **Cluster 3:** com um total de 27 estudantes, de modo semelhante ao Cluster 4, o Cluster 3 inclui perfis que obtiveram bom desempenho no questionário Q2 e baixo desempenho em Q1. Observa-se ausência em relação à TB1.

Com relação aos dois últimos clusters, percebe-se que os estudantes do Cluster 3 e Cluster 2 apresentaram muita semelhança, podendo-se associá-los tanto ao nível **Desempenho Baixo (DB)** ou ao **Desempenho Moderado (DM)**. Uma instância pertencente a um destes clusters pode na verdade pertencer ao outro cluster. A última linha da Tabela 7 destina-se aos somatórios das pontuações. O processo de ranqueamento dos clusters resulta no *score_list* (apresentado no Code 1) que aponta o desempenho de acordo com o somatório das médias de cada cluster. A lista de pontuações visa a indicação de perfil com respeito ao **desempenho** dos estudantes por cluster de forma automatizada. Conforme mostrado na Tabela 7, o *score_list* obtido foi, em ordem decrescente, o seguinte: Cluster 0 (41), Cluster 2 (26), Cluster 4(18), Clusters 1 e 4 (10). Ressalta-se que quanto maior a pontuação menor é o nível de desempenho. Para critério de desempate do Cluster 1 e do Cluster 4 a abordagem considera, por *default*, o aspecto cognitivo representado pelos atributos (Q1, Q2, TB1, TB2, TB3, TB4, TB5). Dessa forma, o Cluster 1 obteve maior desempenho que o Cluster 4. A ferramenta ActivePlan mostra a ordenação dos clusters, como indicado pela Figura 26. Como comentado, conforme o critério de desempenho de modo decrescente (muito elevado até muito baixo), obtém-se a lista {Cluster 1, Cluster 4, Cluster 3, Cluster 2, Cluster 0}. A Figura 26 apresenta os clusters associados aos níveis de desempenho e também indica a lista de ids (identificações) dos estudantes alocados aos clusters.

Figura 26: Resultado da clusterização de estudantes associados ao desempenho - experimento 2.1.

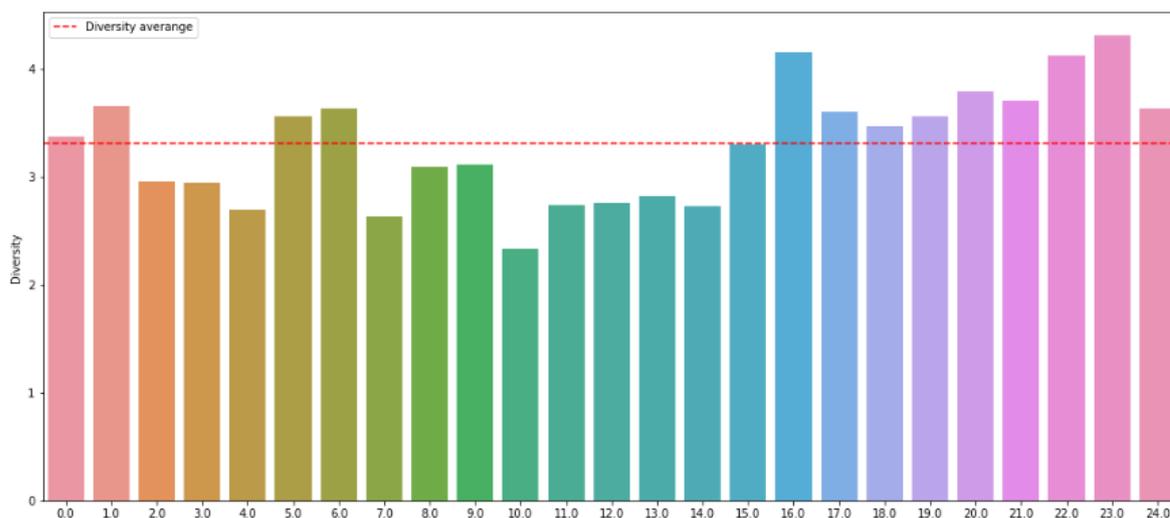
Clusters	Estudantes	Desempenhos
C0	[4, 43, 44, 75, 77, 78, 79, 103, 124]	Desempenho muito baixo
C2	[15, 17, 35, 37, 45, 47, 52, 54, 56, 57, 59, 62, 64, 66, 74, 76, 80, 82, 93, 96, 98, 100, 104, 107, 108, 114]	Desempenho baixo
C3	[8, 11, 13, 21, 31, 48, 49, 53, 58, 60, 67, 68, 69, 71, 72, 73, 81, 90, 92, 99, 109, 110, 112, 115, 117, 118, 120]	Desempenho moderado
C4	[5, 7, 14, 16, 18, 19, 22, 24, 27, 29, 32, 33, 34, 38, 40, 41, 70, 87]	Desempenho elevado
C1	[0, 1, 2, 3, 6, 9, 10, 12, 20, 23, 25, 26, 28, 30, 36, 39, 42, 46, 50, 51, 55, 61, 63, 65, 83, 84, 85, 86, 88, 89, 91, 94, 95, 97, 101, 102, 105, 106, 111, 113, 116, 119, 121, 122, 123]	Desempenho muito elevado

Fonte: autoria própria.

Após a ordenação dos clusters, distribuem-se os estudantes aos grupos novos, dessa vez, grupos heterogêneos. Isso é realizado mediante um mecanismo de atribuição do estudante ao grupo conforme exposto no Figura 26 (Seção 3.2). Em seguida, utilizando-se a métrica de diversidade, torna-se possível a avaliação da heterogeneidade de cada grupo gerado. Quanto à composição dos grupos, ao gerar grupos para PAA, espera-se que cada

grupo contenha perfis distintos de estudantes (*intra-group*). Quanto maior o valor da métrica diversidade, mais heterogêneo é o grupo, quanto menor o valor, mais semelhantes são as instâncias (estudantes) alocadas ao grupo. A Figura 27 apresenta os grupos formados para a PAA Trezentos. O eixo y representa a métrica de diversidade, e o eixo x indica os grupos gerados.

Figura 27: Resultados da métrica de diversidade para cada grupo quanto à PAA Trezentos - experimento 2.1.



Fonte: Autoria própria.

A turma considerada no conjunto de dados é grande pois possui 125 alunos. Assim, conforme a regra da PAA Trezentos, com mais de 20 estudantes, deve-se proceder à divisão da turma (total de 125 estudantes) por cinco (quantidade de membros por grupo) resultando na formação de vinte e cinco grupos (0 a 24). Por meio da Figura 27, percebe-se que nove grupos (do grupo 16 até o 24) detêm valores superiores à média de diversidade 3.3 (média referente aos vinte e quatro grupos, conforme visualiza-se na linha tracejada). Dessa maneira, considera-se a distribuição de estudantes aos grupos mais heterogêneos de acordo com a métrica de diversidade. Observa-se então a constituição obtida de membros nos grupos conforme os perfis identificados na etapa de clusterização: Cluster 1 - DME, Cluster 4 - DE, Cluster 3 - DM, Cluster 2 - DB, Cluster 0 - DMB. Pode-se constatar o seguinte em relação aos grupos heterogêneos obtidos:

- Grupos 20 ao 25, com média de diversidade de 3.9, são compostos de estudantes com os seguintes perfis: um DME (Cluster 1), um DE (Cluster 4), um DM (Cluster 3), um DB (Cluster 2) e um DMB (Cluster 0).

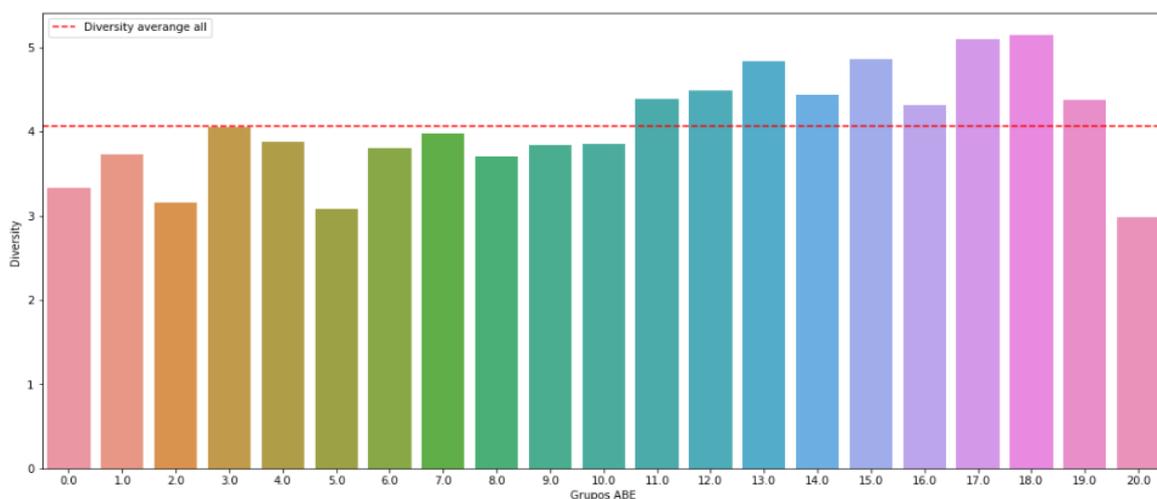
- Grupos 16 ao 19, com média de diversidade de 3.6, são formados por estudantes com os seguintes perfis: dois estudantes DME (Cluster 1), um DM (Cluster 3), um DB (Cluster 2) e um DMB (Cluster 0).

Foram verificados, adicionalmente, índices de diversidade abaixo da média, ou seja, grupos menos heterogêneos, a saber:

- Grupos 0 ao 12, com média de diversidade de 3.0, são compostos de estudantes com perfis assim: dois estudantes de perfil DME, um estudante de DE, um de DB e um estudante de perfil DM.
- Grupos 13 e 14, com média de diversidade de 2.8, são constituídos de estudantes assim: dois estudantes de perfil DME, dois estudantes de DM e um estudante de perfil DB.
- Grupo 15, com diversidade de 3.3, é constituído de estudantes assim definidos: dois estudantes de perfil DME, um estudante de DM e dois estudantes DB.

Observa-se, ainda, que todos os 24 grupos obtidos foram compostos de estudantes da categoria Desempenho Elevado (DE) e/ou da categoria Desempenho Muito Elevado (DME). Tal fato corrobora com o pré-requisito definido pela PAA Trezentos que necessita de pelo menos um dos estudantes com perfil mais especialista. Já a PAA ABE apresenta 20 grupos. Um dos grupos é composto por cinco estudantes e o restante composto por seis estudantes conforme Figura 28.

Figura 28 : Resultados da métrica de diversidade para cada grupo quanto à PAA ABE - experimento 2.1.



Fonte: Autoria própria.

A diversidade média para esse subgrupo (que engloba do grupo 11 até o 19) foi de 4.66. Outras composições demonstraram resultados de índices de diversidade mais baixos, ou seja, grupos menos heterogêneos, a saber:

- Grupos: 0, 1 e 2 (Diversidade: 3.40): apresenta três estudantes de perfil DME, dois estudantes de perfil DM e um de perfil DB.
- Grupos 3, 4 e 5 (Diversidade: 3.66): Grupo composto por dois estudantes de perfil DME, um estudante DE, dois estudantes DM e um estudante DB.
- Grupos 6, 7, 8, 9 e 10 (Diversidade: 3.83): Grupo composto por dois estudantes de perfil DME, um estudante DE, um estudante DM e dois estudantes DB.
- Grupo 20 (Diversity: 2.988): Um caso particular denota-se no Grupo 20, composto de cinco estudantes (grupo menor), conseqüentemente, apresenta quantitativo de diversidade menor. Compõe-se de dois estudantes de perfil DME e outros três estudantes com perfis DE, DM e DB.

4.3.2.2 Experimento 2.2 - Grupos heterogêneos de estudantes resultantes da turma de segundo e terceiro ano

Nesta seção objetiva-se a continuação do experimento da seção anterior, utilizando-se o conjunto de dados referentes ao dataset da turma de primeiro ano juntamente aos resultados da *clusterização* para essa turma. Dessa forma, tem-se o cluster correspondente a cada estudante. A primeira etapa para formação dos grupos heterogêneos inicia-se a partir do reconhecimento dos perfis de desempenho. Isso pode ser verificado por meio da Tabela 8 que aponta as médias para cada uma das características/atributos dos estudantes desta turma de acordo com o cluster identificado. A Tabela 8 propicia uma visão do panorama dos perfis dos estudantes conforme os grupos obtidos e indica uma pontuação para cada atributo vs cluster.

Tabela 8: Ranking dos clusters a partir das médias de cada atributo - experimento 2.2

Atributos	Cluster 0		Cluster 2		Cluster 4		Cluster 3		Cluster 1	
	Média	Pontos								
Q1	0.90	0	0.71	1	0.13	4	0.64	2	0.60	3
Q2	0.90	0	0.79	1	0.08	4	0.77	2	0.25	3
T1	0.02	0	0.05	3	0.86	4	0.04	2	0.04	1
T2	0.03	0	0.06	2	0.78	3	0.04	1	1.0	4
S1	0.03	1	1.0	3	1.0	3	0.03	0	0.52	2
S2	0.03	0	0.08	2	0.89	4	0.07	1	0.18	3
CRE	0.67	0	0.44	1	0.38	3	0.04	2	0.29	4
<i>score_do_cluster*</i>		1	-	13	-	25	-	10	-	20

Fonte: Elaborados pelos autores. **score_do_cluster*: Variável destinada ao armazenar o somatório dos valores associados aos pontos obtidos para cada cluster conforme estratégia exposta Seção 3.2.

Considerando os valores apontados na Tabela 8, pode-se identificar o seguinte perfil para cada um dos clusters:

- **Cluster 0:** com maior quantitativo de estudantes (106), este cluster contém o perfil de estudantes que apresentaram maior desempenho nos questionários Q1 e Q2, ou seja, que respondem rapidamente às atividades passadas e também que acessam mais ativamente o AVA. Em geral, os estudantes demonstram nota média mais elevada em disciplinas anteriores conforme exposto na Tabela 8. Este Cluster é classificado como **Desempenho Muito Elevado (DME)**. Isso significa que em todos os atributos analisados dos estudantes oriundos do Cluster 1, os estudantes, em sua maioria, obtiveram melhores notas.
- **Cluster 4 e Cluster 1:** com quantitativo de 9 estudantes e 6 estudantes, respectivamente, inclui perfis de estudantes que apresentam o pior desempenho detectado, com destaque aos valores insatisfatórios nos questionários Q1 e Q2. Em geral, os estudantes demoram mais para responder às atividades (T1 e T2) e para acessar o AVA (S1 e S2). Portanto, devido à defasagem descrita, pode-se categorizá-los como **Desempenho Muito Baixo (DMB) ou Desempenho Baixo (DB)**.
- **Cluster 3 e Cluster 2:** com um montante de 73 estudantes e 35 estudantes, respectivamente, uma das características presentes nos clusters refere-se ao alto desempenho no segundo questionário (Q2). Tais características podem apontar um perfil intermediário, composto por estudantes em evolução. O cluster foi classificado em **Desempenho Elevado (DE) e Desempenho Moderado (DM)**.

A última linha da Tabela 8 destina-se aos somatórios das pontuações. O processo de ranqueamento dos clusters resulta no *score_list* (definido na Seção 3.2) que aponta o desempenho de acordo com o somatório das médias de cada cluster. A lista de pontuações visa a indicação de perfil no que se refere ao desempenho dos estudantes por cluster de forma automatizada. Conforme mostrado na Tabela 8, o *score_list* obtido foi, em ordem decrescente, o seguinte: Cluster 0 (1), Cluster 2 (13), Cluster 3 (10), Clusters 1(20) e 4 (25).

Ressalta-se que quanto maior a pontuação menor é o nível de desempenho. A abordagem ActivePlan mostra a ordenação dos clusters, como indicado pela Figura 29.

Figura 29 - Resultado da clusterização de estudantes associados ao desempenho - Experimento 2.2.

Clusters	Estudantes	Desempenhos
C4	[8, 26, 41, 125, 128, 155, 168, 196, 200]	Desempenho muito baixo
C1	[0, 3, 122, 132, 161, 189]	Desempenho baixo
C2	[9, 17, 20, 21, 29, 33, 35, 39, 57, 67, 70, 71, 79, 85, 86, 88, 94, 129, 131, 135, 140, 145, 156, 167, 172, 175, 187, 188, 194, 205, 213, 218, 225, 226, 228]	Desempenho moderado
C3	[4, 5, 6, 7, 12, 16, 24, 25, 27, 28, 31, 42, 43, 44, 45, 47, 48, 50, 51, 54, 55, 62, 65, 75, 78, 87, 90, 96, 103, 106, 107, 108, 110, 111, 114, 115, 118, 123, 130, 133, 134, 136, 137, 138, 139, 141, 151, 153, 158, 159, 163, 165, 170, 171, 176, 177, 179, 181, 183, 191, 192, 193, 195, 201, 202, 207, 208, 215, 216, 219, 220, 221]	Desempenho elevado
C0	[1, 2, 10, 11, 13, 14, 15, 18, 19, 22, 23, 30, 32, 34, 36, 37, 38, 40, 46, 49, 52, 53, 56, 58, 59, 60, 61, 63, 64, 66, 68, 69, 72, 73, 74, 76, 77, 80, 81, 82, 83, 84, 89, 91, 92, 93, 95, 97, 98, 99, 100, 101, 102, 104, 105, 109, 112, 113, 116, 117, 119, 120, 121, 124, 126, 127, 142, 143, 144, 146, 147, 148, 149, 150, 152, 154, 157, 160, 162, 164, 166, 169, 173, 174, 178, 180, 182, 184, 185, 186, 190, 197, 198, 199, 203, 204, 206, 209, 210, 211, 212, 214, 217, 222, 223, 224, 227]	Desempenho muito elevado

Fonte: Elaborados pelos autores.

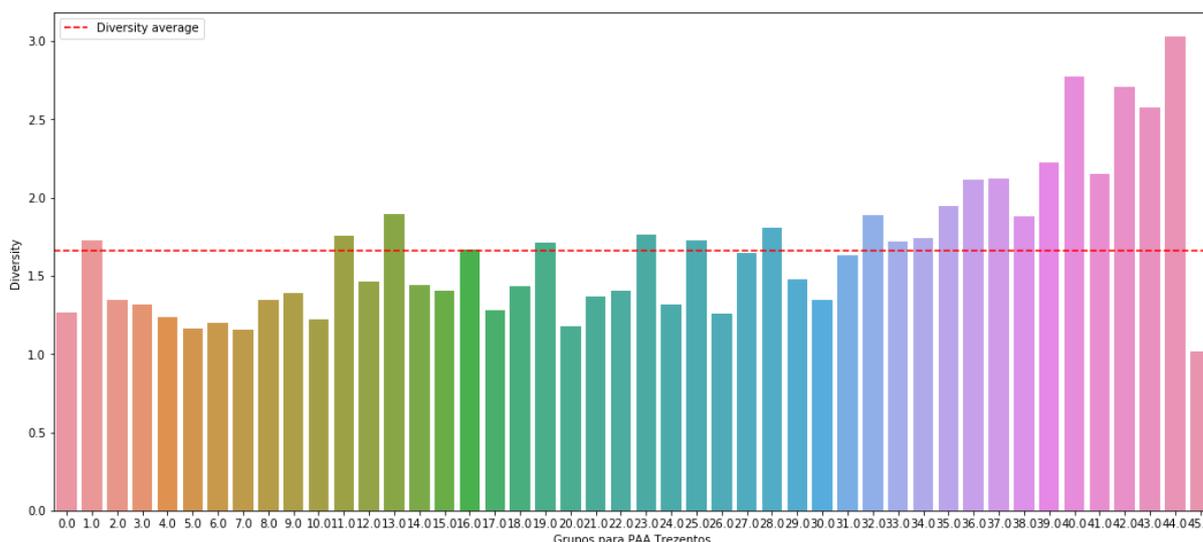
Como comentado, conforme o critério de desempenho de modo decrescente (muito elevado até muito baixo), obtém-se a seguinte lista: {Cluster 0, Cluster 3, Cluster 2, Cluster 1, Cluster 4}. A Figura 29 apresenta os clusters associados aos níveis de desempenho e também indica a lista de ids (identificações) dos estudantes alocados aos clusters.

Após a ordenação dos clusters, distribuem-se os estudantes aos grupos novos, dessa vez, grupos heterogêneos. Em seguida, utilizando-se a métrica de diversidade, busca-se a avaliação da heterogeneidade de cada grupo gerado.

A turma considerada no conjunto de dados é grande pois possui 229 alunos. Assim, conforme a regra da PAA Trezentos, para turmas com mais de 20 estudantes, deve-se proceder à divisão da turma (total de 229 estudantes) por cinco (quantidade de membros por grupo) resultando na formação de quarenta e seis grupos (0 a 45).

Quanto à composição dos grupos, ao gerar grupos para uma PAA, quanto menor o valor, mais semelhantes são as instâncias (estudantes) alocadas ao grupo. A Figura 30 apresenta os grupos formados para a PAA Trezentos. O eixo y representa a métrica de diversidade, e o eixo x indica os grupos gerados.

Figura 30: Resultados da métrica de diversidade para cada grupo quanto à PAA Trezentos - experimento 2.2.



Fonte: Elaborados pelos autores.

Por meio da Figura 30, percebe-se que 12 grupos (do grupo 32 até o 44) detêm valores superiores à média de diversidade 1,6 (média referente aos quarenta e seis grupos, conforme visualiza-se na linha tracejada). Dessa maneira, considera-se a distribuição de estudantes aos grupos mais heterogêneos de acordo com a métrica de diversidade. Observa-se então a constituição obtida de membros nos grupos conforme os perfis identificados na etapa de clusterização: Cluster 0 - DME, Cluster 3 - DE, Cluster 2 - DM, Cluster 4 - DB, Cluster 1 - DMB. Pode-se constatar o seguinte em relação aos grupos heterogêneos obtidos:

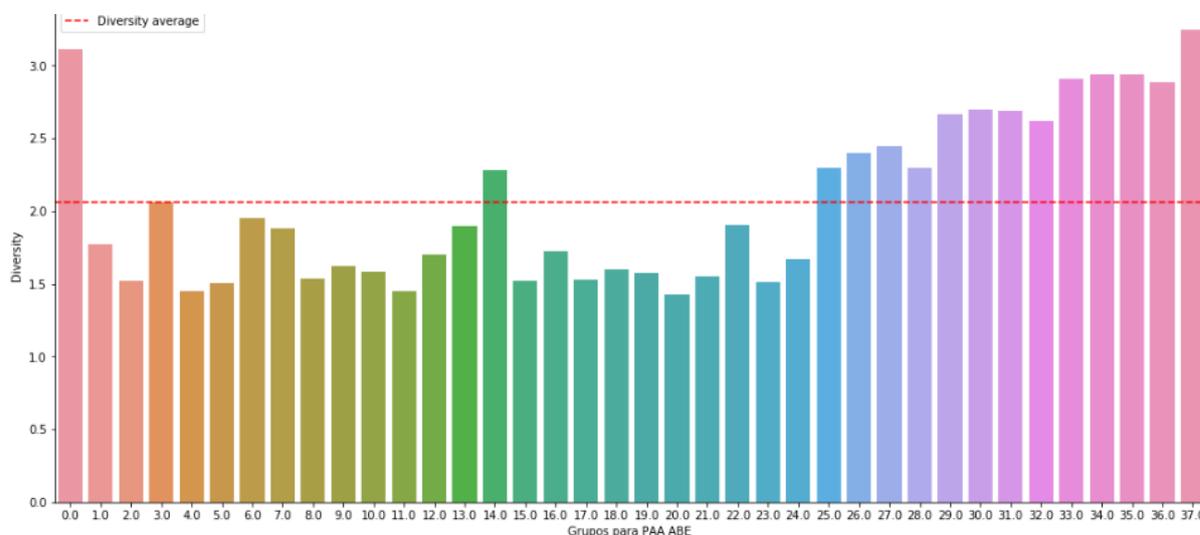
- Grupos 44 ao 36, compostos de estudantes com os seguintes perfis: Dois estudantes DME (Cluster 0), um DE (Cluster 3), um DM (Cluster 2) e um DMB (Cluster 4).
- Grupos 35 ao 30, formados por estudantes com os seguintes perfis: dois estudantes DME (Cluster 0), dois DM (Cluster 2) e um DB (Cluster 1).

Foram verificados, adicionalmente, índices de diversidade abaixo da média, ou seja, grupos menos heterogêneos, a saber:

- Grupos 29 ao 15, compostos de estudantes com perfis assim: dois estudantes de perfil DME, dois estudante DE, um DB e um estudante de perfil DM.
- Grupos 14 e 0, constituídos de estudantes assim: Três estudantes de perfil DME, um estudante DM.
- Grupo 45, constituído de quatro estudantes assim definidos: dois estudantes de perfil DME, um estudante DE e um estudante DM.

Já a PAA aprendizagem baseada em equipes apresenta 38 grupos compostos por seis estudantes conforme mostra a Figura 31.

Figura 31: Resultados da métrica de diversidade para cada grupo quanto à PAA ABE -Experimento 2.2.



Fonte: Elaborado pelo autor

Quanto à distribuição de perfis de estudantes vs grupos, percebe-se que:

- Os Grupos 37 ao 31 são compostos de estudantes com os seguintes perfis: Dois estudantes DME, dois DE, um DM e um DB .
- O Grupo 30 é formado por estudantes com os seguintes perfis: três estudantes DME, um DE, um DM e um DB.
- Os Grupos 29 ao 27 são formados por estudantes com os seguintes perfis: três estudantes DME, um DE, um DM e um DMB.
- Os Grupos 26 ao 24 são compostos de estudantes com perfis assim: três estudantes de perfil DME, dois estudantes de DE e um de DMB.
- Os Grupos 23 e 1 são constituídos de estudantes assim: Três estudantes de perfil DME, dois estudantes de DE. e um estudante de DM.
- O Grupo 0 é formado por estudantes com os seguintes perfis: três estudantes DME, dois DE e um DB.

4.3.3 Discussão geral sobre os resultados dos experimentos

O Quadro 7 e Quadro 8 apresentam amostras de estudantes pertencentes aos grupos heterogêneos formados para a turma de primeiro ano e para a turma de segundo e terceiro

ano, respectivamente, de acordo com seus perfis e níveis de desempenho (DME, DE, DM, DB e DMB).

As amostras dos Experimentos 2 e Experimentos 4 expõem a composição mista de grupos de estudantes obtidos, o que indica o potencial da abordagem ActivePlan para favorecer a aprendizagem ativa e colaborativa.

Grupos mistos propiciam um ambiente mais rico para aprendizagem por meio da interação de seus membros (Vygotsky,1984; Perrenoud, 2015; Moran 2015).

Quadro 7: Amostra de estudantes em grupos heterogêneos turmas de primeiro ano

Estudantes (ID/Papel)	Q1	Q2	T1	T2	S1	S2	T	T	T	T	T	Grupo	Nível Desempenho
							B1	B2	B3	B4	B5		
61 (Ajudante/LI)	1,0	0,9	0,09	0,09	0,03	0,06	1	1	1	0	1	23	DME
40 (Ajudante)	0,3	0,93	0,01	0,01	0,04	0,04	1	1	0	1	0		DE
81 (Ajudado)	0,75	0,86	0,04	0,06	0,04	0,04	0	1	1	0	1		DM
82(Ajudado)	1,0	0,8	0,03	0,0	0,03	0,03	1	0	1	1	1		DB
43 (Ajudado)	0,0	0,0	0,88	0,99	1,0	1,0	0	0	0	0	0		DMB
101(Ajudante/LI)	0,9	1,0	0,06	0,06	0,03	0,05	1	1	1	1	1	10	DME
23 (Ajudante)	0,8	0,86	0,05	0,06	0,04	0,05	1	1	1	1	1		DME
14(Ajudado)	0,4	0,97	0,01	0,02	0,04	0,05	1	1	1	1	0		DE
31(Ajudado)	0,55	0,86	0,01	0,01	0,05	0,05	0	1	1	1	0		DM
47(Ajudado)	1,0	0,83	0,09	0,09	0,02	0,05	1	0	1	1	1		DB
89(Ajudante/LI)	0,75	0,76	0,04	0,06	0,03	0,05	1	1	1	0	1	15	DME
6(Ajudante)	0,8	0,9	0,3	0,24	0,04	0,04	1	1	0	1	1		DME
117(Ajudado)	0,5	0,86	0,32	0,27	0,03	0,05	0	1	1	1	1		DM
114(Ajudado)	0,75	0,83	0,03	0,06	0,02	0,04	0	0	1	1	1		DB
15(Ajudado)	0,4	0,4	0,01	0,01	0,03	0,04	0	0	0	1	0		DB

Fonte: Elaborado pelos autores

*ID- Identificador, LI -Líder do grupo

Quadro 8: Amostra de estudantes em grupos heterogêneos turmas do segundo e terceiro ano

Estudantes(ID*/Papel)	Q1	Q2	T1	T2	S1	S2	MEAN	Grupo	Nível Desempenho
162(Ajudante/LI*)	0,75	0,8	0,04	0,04	0,01	0,04	0,71	28	DME
72(Ajudante)	0,8	0,9	0,02	0,01	0,02	0,04	0,57		DME
183 (Ajudado)	0,5	0,8	0	0,05	0,02	0,04	0,43		DE
43 (Ajudado)	0,22	0,7	0,31	0,23	0,05	1	0,74		DE
17 (Ajudado)	0,86	1	0,04	0	1	0,04	0,67		DE
142(Ajudante/LI*)	1	0,8	0	0,04	0,03	0,02	0,86	43	DME
52(Ajudante)	0,92	0,9	0,01	0,01	0,01	0,02	0,6		DME
151 (Ajudado)	0,5	0,9	0	0,01	0,02	0,02	0,36		DE
4 (Ajudado)	0,92	0,6	0,06	0,09	0,03	1	0,33		DE
128 (Ajudado)	0	0	1	1	1	1	0,12		DMB
206(Ajudante/LI*)	0,75	0,9	0,01	0,01	0,04	0,04	0,71	10	DME
100(Ajudante)	1	0,8	0,04	0,05	0,01	0,04	0,67		DME
13 (Ajudado)	0,88	0,8	0,01	0,04	0,01	0,04	0,55		DME
107 (Ajudado)	0,92	0,6	0,04	0,05	0,03	0,04	0,31		DE
135 (Ajudado)	0,75	0,9	0,09	0,04	1	0,03	0,36		DM

Fonte: Elaborado pelos autores

*ID- Identificador, LI -Líder do grupo

Assim, no tocante aos resultados da abordagem ActivePlan, percebe-se que cada grupo constitui-se de membros com estudantes de diferentes níveis, ou seja, de perfis heterogêneos. Dessa maneira, a abordagem conseguiu gerar grupos realmente mistos voltados à aplicação da PAA Trezentos e à aplicação da PAA ABE.

5 Considerações e Trabalhos Futuros

Este trabalho apresentou uma abordagem que visa auxiliar o docente na utilização de práticas ativas de aprendizagem para cenários de turmas com perfis heterogêneos de estudantes. Para isso, a abordagem, denominada ActivePlan, realiza o enriquecimento de perfis de estudantes com indicadores de seus desempenhos com base na Taxonomia de Bloom. A ActivePlan faz uso de método de aprendizado não supervisionado (agrupamento) e propõe uma estratégia para formação de grupos de estudantes com perfis heterogêneos.

Mediante o experimento realizado, considera-se que a formação de grupos heterogêneos para a PAA Trezentos atingiu resultados promissores. A razão para isso é o grau de heterogeneidade obtido para os grupos formados, pois todos os grupos foram constituídos de pelo menos 20% de estudantes considerados aptos (DME e DE) para exercer o papel de ajudante durante o desenvolvimento da PAA Trezentos. A métrica de diversidade avaliada possibilitou aferir o grau de heterogeneidade dos grupos formados, no qual 50% dos grupos apresentaram diversidade com valores entre 2.8 e 3.6, valores próximos à média (3.3). Isso indica a proximidade e semelhança (inter-group) entre os grupos formados e a heterogeneidade entre os membros do grupo (intra-group).

Como **trabalhos futuros**, pode-se estender a Etapa I da ActivePlan no tocante à necessidade de automatização da obtenção de indicadores conforme a Taxonomia de Bloom, evitando o trabalho manual do professor. Pode-se experimentar também outros algoritmos de clusterização em busca de possíveis melhorias no processo de agrupamento dos perfis. Além disso, é importante considerar também outras formas de avaliação dos grupos heterogêneos formados, a exemplo de uma avaliação qualitativa por meio, por exemplo, de grupo focal com especialistas na área de educação. Implantação da abordagem em sistemas acadêmicos, AVA ou SPAE.

REFERÊNCIAS BIBLIOGRÁFICAS

ABONIZIO, Gustavo. Precarização do trabalho docente: apontamentos a partir de uma análise bibliográfica. *Ensino de sociologia em debate*, v.1, n. 1, 2012. Disponível em: <http://www.uel.br/revistas/lenpes-pibid/pages/arquivos/1%20Edicao/1ordf.%20Edicao.%20Artigo%20ABONIZIO%20G.pdf> . Acesso em: 18 ago. 2021.

ARAUJO, L. C. *Configuração: uma perspectiva de Arquitetura da Informação da Escola de Brasília*. Dissertação (Mestrado) — Universidade de Brasília, Brasília, mar. 2012. Citado na página 19.

ALPAYDIN, E. . *Introduction to Machine Learning*. 2010.

ARANTES, Daniella Andrade; DA SILVA, Denise Mendes. análise do nível cognitivo do exame de suficiência contábil na perspectiva da taxonomia de bloom. *Contabilidade Vista & Revista*, v. 31, n. 2, 2020.

ARAÚJO, G. et al. Estudo da Taxonomia de Bloom e Criação de um Instrumento de Medição do Nível de Bloom no contexto de MC102. 2020 Disponível em: <<https://ic.unicamp.br/~reltech/2020/20-05.pdf>>, Acesso em: 30 de março de 2020.

ARAUJO, Ives Solano; MAZUR, Eric. Instrução pelos colegas e ensino sob medida: uma proposta para o engajamento dos alunos no processo de ensino-aprendizagem de Física. *Caderno brasileiro de ensino de física*. Florianópolis. Vol. 30, n. 2 (ago. 2013), p. 362-384, 2013.

ARAUJO, L. C. O pacote abntex2cite: Estilos bibliográficos compatíveis com a ABNT NBR 6023. [S.l.], 2015. Disponível em: <<http://www.abntex.net.br/>>. Citado na página 12.

ARAUJO, L. C. O pacote abntex2cite: tópicos específicos da ABNT NBR 10520:2002 e o estilo bibliográfico alfabético (sistema autor-data). [S.l.], 2015. Disponível em: <<http://www.abntex.net.br/>>. Citado na página 12.

AGGARWAL, Charu C. et al. *Recommender systems*. Cham: Springer International Publishing, 2016.

AGRAWAL, Rakesh; IMIELINSKI, Tomasz; SWAMI, Arun. Database mining: A performance perspective. *IEEE transactions on knowledge and data engineering*, v. 5, n. 6, p. 914-925, 1993.

AGRAWAL, Rakesh; IMIELINSKI, Tomasz; SWAMI, Arun. Database mining: A performance perspective. *IEEE transactions on knowledge and data engineering*, v. 5, n. 6, p. 914-925, 1993.

AKAY, Özlem; YÜKSEL, Güzin. Hierarchical clustering of mixed variable panel data based on new distance. *Communications in Statistics-Simulation and Computation*, p. 1-16, 2019.

ALCANTARA, Elisa FS. Inovação e renovação acadêmica: guia prático de utilização de metodologias e técnicas ativas. Volta Redonda, RJ: FERP, 2020.

ALVES, Gabriela Souza et al. Método dos trezentos: estratégia para minimizar a retenção de estudantes no curso de Química. *Brazilian Journal of Development*, v. 5, n. 12, p. 33037-33046, 2019.

ALVES, Rodrigo Santana; CHAPADEIRO, Fernando Camargo. A utilização do método 300 no ensino de arquitetura e urbanismo. 2019.

ANDERSON, Lorin W. et al. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman,, 2001.

Araujo, R. M., & Pimentel, M. (2020). #Fique em Casa, mas se mantenha ensinando-aprendendo: algumas questões educacionais em tempos de pandemia. *SBC Horizontes*. ISSN: 2175-9235. Disponível em: <http://horizontes.sbc.org.br/index.php/2020/03/30/fiqueemcasa>

BACICH, Lilian; MORAN, José. Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática. Penso Editora, 2018.

BARBOSA, Eduardo Fernandes; DE MOURA, Dácio Guimarães. Metodologias ativas de aprendizagem na educação profissional e tecnológica. *Boletim Técnico do Senac*, v. 39, n. 2, p. 48-67, 2013.

BACKES, Dirce Stein et al. Grupo focal como técnica de coleta e análise de dados em pesquisas qualitativas. *O mundo da saúde*, v. 35, n. 4, p. 438-442, 2011.

BONALS, J. O trabalho em pequenos grupos na sala de aula. Porto Alegre: Artmed, 2003.

BRASIL. Ministério da Educação. Diretrizes Curriculares Nacionais e Base Nacional Comum para a Formação Inicial e Continuada de Professores da Educação Básica. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=135951-rcp002-19&category_slug=dezembro-2019-pdf&Itemid=30192>.Citado na página 15. Acesso em: 30 de março de 2020.

BRASIL. Ministério da Educação. Diretrizes Curriculares Nacionais e Base Nacional Comum para a Formação Inicial e Continuada de Professores da Educação Básica. Disponível em: <<http://portal.mec.gov.br/docman/dezembro-2019-pdf/135951-rcp002-19/file>>.Citado na página 17. Acesso em: 30 de março de 2020.

BAKER, Ryan; ISOTANI, Seiji; CARVALHO, Adriana. Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, v. 19, n. 02, p. 03, 2011.

BARROS, arthur barretto leite de et al. desenvolvimento de um modelo de previsão do sucesso acadêmico de graduandos de engenharia visando o aumento de eficiência da instituição de ensino superior. 2020.

BARROS, C. C. A., Souza, A. da S., Dutra, F. D., Gusmão, R. S. C., & Cardoso, B. L. C.

(2021). Precarização do Trabalho Docente: reflexões em tempos de pandemia e pós pandemia. *Ensino Em Perspectivas*, 2(2), 1–23. Recuperado de <https://revistas.uece.br/index.php/ensinoemperspectivas/article/view/4975>

BLIKSTEIN, Paulo. O mito do mau aluno e porque o Brasil pode ser o líder mundial de uma revolução educacional. 2012.

BOLLELA, Valdes Roberto et al. Aprendizagem baseada em equipes: da teoria à prática. *Medicina (Ribeirão Preto)*, v. 47, n. 3, p. 293-300, 2014.

BAO, Wei. COVID-19 and online teaching in higher education: A case study of Peking University. *Human Behavior and Emerging Technologies*, v. 2, n. 2, p. 113-115, 2020.

BERBEL, Neusi Aparecida Navas. As metodologias ativas e a promoção da autonomia de estudantes. *Semina: Ciências Sociais e Humanas*, v. 32, n. 1, p. 25-40, 2011.

BLOOM, Benjamin S. et al. *Taxonomy of educational objectives. Vol. 1: Cognitive domain.* New York: McKay, 1956.

BOSCOV, CAMILA. O IMPACTO DO ENSINO CENTRADO NO ALUNO NO PROCESSO DE APRENDIZADO. *RAGC*, v. 8, n. 36, 2020.

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

BONWELL, Charles C.; EISON, James A. Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC higher education reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183, 1991.

CARNEIRO, Janalívia do Nascimento et al. Aplicação de Learning Analytics para Modelagem do Aluno de acordo com a Taxonomia de Bloom Revisada. 2018.

CASTRO JUNIOR, Amaury Antônio et al. Uma Análise Preliminar da Aplicação do Método 300 em Turmas de Algoritmos e Programação. In: *Anais do XXIX Workshop sobre Educação em Computação.* SBC, 2021. p. 171-180.

CRISPIM, Diêgo Lima; FERNANDES, Lindemberg Lima; ALBUQUERQUE, RL de O. Aplicação de técnica estatística multivariada em indicadores de sustentabilidade nos municípios do Marajó-PA. *Revista Principia-Divulgação Científica e Tecnológica do IFPB*, (46), p. 145-154, 2019.

COSTA, Evandro et al. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, v. 1, n. 1, p. 1-29, 2013

COSTA, Fellipe et al. Predição de sucesso de estudantes cotistas utilizando algoritmos de classificação. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE).* 2015. p. 997.

COELHO, Mônica. Por uma Educação sem Fronteiras e em Tempos de Pandemia-Democrática, Pública e de Qualidade. *Revista Pedagogia Social UFF*, v. 9, n. 1, 2020.

CORREIA, Creusa Fernandes; PIMENTEL, Edson Pinheiro. Mineração de dados na formação de turmas para a recuperação paralela na educação básica. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2012.

CHANAMARN, Nipaporn; TAMEE, Kreangsak. Enhancing Efficient Study Plan for Student with Machine Learning Techniques. *International Journal of Modern Education & Computer Science*, v. 9, n. 3, 2017.

CHAPMAN, Pete et al. *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS inc, v. 9, p. 13, 2000.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), p. 1-29, 2009.

CABRAL, Mirela Moraes Waldemarin. A utilização da taxonomia de Bloom no processo de ensino- aprendizagem para alunos do ensino superior. *Revista Calafiori*, v. 3, n. 1, p. 32-38, 2019.

COSTA, Evandro et al. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, v. 1, n. 1, p. 1-29, 2013.

COSTA, Fellipe et al. Predição de sucesso de estudantes cotistas utilizando algoritmos de classificação. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2015. p. 997.

DA COSTA, Newarney Torrezão; FERNANDES, Marcia Aparecida. Sequenciamento de Ações Pedagógicas por Algoritmo Genético Utilizando Taxonomia de Bloom e ASSIST. In: *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. SBC, 2020. p. 1273-1282.

CUNHA, Luísa Margarida Antunes da et al. Modelos Rasch e Escalas de Likert e Thurstone na medição de atitudes. 2007. Tese de Doutorado.

DE OLIVEIRA, Tobias Espinosa; ARAUJO, Ives Solano; VEIT, Eliane Angela. Aprendizagem Baseada em Equipes (Team-Based Learning): um método ativo para o ensino de física. *Caderno Brasileiro de Ensino de física*, v. 33, n. 3, p. 962-986, 2016.

DE ROSSO KRUGI, Rodrigo et al. O “Bê-Á-Bá” da aprendizagem Baseada em Equipe The “Bê-Á-Bá” of Team-Based Learning. *Revista Brasileira de Educação Médica*, v. 40, n. 4, p. 602-620, 2016.

DA SILVA, E. T. et al. Identifying student profiles and their impact on academic performance in a Brazilian undergraduate student sample. *European Journal of Dental Education*, v. 16, n. 1, p. e27-e32, 2012.

DOS SANTOS, Henrique et al. Recomendação de Objetos de Aprendizagem utilizando

Filtragem Colaborativa: Uma comparação entre abordagens de pré-processamento por meio de clusterização. In: Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie). 2015. p. 1127.

dos santos farias, maria arlete silva; de carvalho, antonio marcos rocha; de albuquerque junior, ailton batista. a precarização do trabalho e da formação docente: um estudo sobre a realidade do professor no brasil. a crítica da educação na sociedade do capital, 2020. p. 100.

DALE, Edgard. Audiovisual methods in teaching. 3ª Ed. New York: Holt, Reinhart & Winston, 1969.

DIESEL, Aline; BALDEZ, Alda Leila Santos; MARTINS, Silvana Neumann. Os princípios das metodologias ativas de ensino: uma abordagem teórica. Revista Thema, v. 14, n. 1, p. 268-288, 2017.

DUTT, Ashish; ISMAIL, Maizatul Akmar; HERAWAN, Tutut. A systematic review on educational data mining. Ieee Access, v. 5, p. 15991-16005, 2017.

DA SILVA, Clênio Eduardo; BORGES, Lucas Guerra. Uma Abordagem Inteligente para Classificação de Perfis de Aprendizagem de Discentes com Aplicação de Aprendizado de Máquina. Revista de Sistemas e Computação-RSC, v. 9, n. 2, 2020.

DE FRANÇA, Rozelma Soares; DO AMARAL, Haroldo José Costa. Mineração de dados na identificação de grupos de estudantes com dificuldades de aprendizagem no ensino de programação. RENOTE-Revista Novas Tecnologias na Educação, v. 11, n. 1, 2013.

FREIRE, P. (1981). Pedagogia do oprimido. Rio de Janeiro: Paz & Terra.

DE OLIVEIRA, Pamella Leticia Silva et al. Uma Análise de Algoritmos de Clusterização para Descoberta de Perfis de Engajamento. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. SBC, 2020. p. 1012-1021.

DE OLIVEIRA, Pamella Leticia Silva et al. Identificação de Pesquisas e Análise de Algoritmos de Clusterização para a Descoberta de Perfis de Engajamento. Revista Brasileira de Informática na Educação, v. 30, p. 01-19, 2022.

DE LIMA JÚNIOR, Alderí Segundo; DE OLIVEIRA, Glaydson Francisco Barros. Revisão sistemática da literatura sobre o uso do ensino híbrido em aulas de Probabilidade e Estatística no ensino básico e superior. **Revista de Ensino de Ciências e Matemática**, v. 13, n. 1, p. 1-16, 2022.

DE SOUZA, Josiane do Socorro Aguiar et al. Aprendizagem ativa em grandes turmas de engenharia: “adaptação do método dos 300”. Brazilian Journal of Development, v. 6, n. 3, p. 11892- 11903, 2020.

DE BRITO, Daniel Miranda et al. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2014. p. 882.

Ester, Martin, H. Kriegel, J. Sander & X. Xu (1996), 'A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise', Proceedings of the Second International Conference on Knowledge Discovery and Data Mining .

FARIAS, Vinicius Tasca; FACÓ, Julio Francisco Blumetti; ANDRADE, Alexandre Acácio. Mineração de Dados para Análise de Banco de Dados Empresariais. In: XVII Safety, Health and Environment World Congress. 2017. p. 159-164.

FREIRE, Paulo. Educação “bancária” e educação libertadora. **Introdução à psicologia escolar**, v. 3, p. 61-78, 1997.

Faceli, K., Lorena, A. C., Gama, J., Carvalho, A. C. P. L. F. (2011). Inteligência Artificial: Uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2, 192. 186717º WIM - Workshop de Informática Médica

FIGUEIREDO, Maria Fernanda Santos; RODRIGUES-NETO, João Felício; LEITE, Maísa Tavares Souza. Modelos aplicados às atividades de educação em saúde. Revista Brasileira de Enfermagem, v. 63, p. 117-121, 2010.

FRANÇA, R. S. de. Aplicação de Técnicas de Mineração de Dados, Baseada na Taxonomia de Bloom, para o Mapeamento do Conhecimento na Aprendizagem de Programação. Garanhuns: Universidade de Pernambuco, 2013. 45p. Monografia de Graduação.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, v. 39, n. 11, p. 27-34, 1996.

FRAGELLI, Ricardo. Método trezentos: Aprendizagem ativa e colaborativa, para além do conteúdo. Penso Editora, 2019.

FERREIRA, Gislaine Rossetti Madureira. Modelo de combinação socioafetiva: um foco na formação de grupos para um ambiente virtual de aprendizagem. 2021.

GUEDES, Gilleanes T. A.. UML 2: uma abordagem prática. 3a. Edição. São Paulo: Novatec, 2018. 496p.

GUERREIRO, Marcio Trindade et al. Análise de métodos de agrupamento de dados para detecção de anomalias na especificação e categorização de peças da indústria automotiva. 2021. Dissertação de Mestrado. Universidade Tecnológica Federal do Paraná.

GALHARDI, Antonio César; AZEVEDO, Marília Macorin de. Avaliações de aprendizagem: o uso da taxonomia de Bloom. VIII workshop de pós-graduação e pesquisa do centro Paula Souza.

GOLDSCHMIDT, R. R.; PASSOS, E. Data Mining: Um Guia Prático. Rio de Janeiro: Elsevier, 2005.

GERHARDT, Tatiana E.; SILVEIRA, Denise T. (org.). Métodos de Pesquisa. Porto Alegre, Editora UFRGS, 2009. Disponível em: <<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>>. Acesso em: 26 março 2021.

GANDHI, Jashraj et al. Comparative Study on Hierarchical and Density based Methods of Clustering using Data Analysis. Available at SSRN 3768295, 2021.

HENRY, Tyson R. Creating effective student groups: an introduction to group formation. org. In: Proceeding of the 44th ACM technical symposium on Computer science education. 2013. p. 645-650.

HARRINGTON, Peter. Machine learning in action. Manning Publications Co., 2012.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, v. 5, n. 4, p. 83-124, 2011.

HODGES, Charles et al. The difference between emergency remote teaching and online learning. Educause review, v. 27, p. 1-12, 2020.

KAPP, Karl M. The gamification of learning and instruction: game-based methods and strategies for training and education. John Wiley & Sons, 2012.

KUO, Ching-Chung; GLOVER, Fred; DHIR, Krishna S. Analyzing and modeling the maximum diversity problem by zero-one programming. Decision Sciences, v. 24, n. 6, p. 1171-1185, 1993.

KRATHWOHL, David R. A revision of Bloom's taxonomy: An overview. Theory into practice, v. 41, n. 4, p. 212-218, 2002.

LEITE, Bruno. Aprendizagem tecnológica ativa. Revista internacional de educação superior, v. 4, n. 3, p. 580-609, 2018.

LIMA, Ana Carolina Bezerra et al. Ensino híbrido na formação em saúde: uma revisão sistemática. **Revista Cuidarte**, v. 13, n. 1, p. e2051, 2022.

LIMA, Rommel Wladimir de. Mapa de Conteúdos e Mapa de Dependências: ferramentas pedagógicas para uma metodologia de planejamento baseada em objetivos educacionais e sua implementação em um ambiente virtual de aprendizagem. p. 29 2009.

LEMES, Ronaldo; DE FREITAS, Sergio Antonio Andrade. Uso do algoritmo Apriori na avaliação da motivação em uma gamificação. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2019. p. 694.

KAUFMAN, L.; ROUSSEEUW, P. J. Finding groups in data: an introduction to cluster analysis. New York: John Wiley and Sons, 1990.

MARTÍNEZ-PLUMED, Fernando et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

MADHULATHA, T. Soni. An overview on clustering methods. arXiv preprint arXiv:1205.1117, 2012.

MORAN, J. M. A educação que desejamos: Novos desafios e como chegar lá. Campinas: Papirus, 2012.

MORÁN, José. Mudando a educação com metodologias ativas. *Coleção mídias contemporâneas. Convergências midiáticas, educação e cidadania: aproximações jovens*, v. 2, n. 1, p. 15-33, 2015.

MARINHO, Johnny; SOUZA, Damires. Avaliação de técnica de clusterização para a identificação de grupos de discentes. 2020. Relatório técnico.

MACEDO, Pedro HR; SANTOS, Wylliams B.; MACIEL, Alexandre MA. Análise de Perfis de Engajamento de Estudantes de Ensino a Distância. *RENOTE*, v. 18, n. 2, p. 326-335, 2020.

MAINA, Elizaphan M.; OBOKO, Robert O.; WAIGANJO, Peter W. Using machine learning techniques to support group formation in an online collaborative learning environment. *International Journal of Intelligent Systems & Applications*, v. 9, n. 3, p. 26-33, 2017.

MARTÍ, Rafael; GALLEGO, Micael; DUARTE, Abraham. An exact method for the maximum diversity problem. *European Journal of Operational Research*, v. 200, n. 1, p. 36-44, 2010.

MONTEVERDE, Ilmara et al. M-cluster: Uma ferramenta de recomendação para formação de grupos em ambientes virtuais de aprendizagem. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2017. p. 1657.

MASCHIO, Pedro et al. Um panorama acerca da mineração de dados educacionais no Brasil. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2018. p. 1936.

MAQTARY, Naseebah; MOHSEN, Abdulqader; BECHKOUM, Kamal. Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, v. 24, n. 2, p. 169-190, 2019.

Moura, J. da S., Ribeiro, J. C. de O. A., Castro Neta, A. A. de, & Nunes, C. P. (2019). A precarização do trabalho docente e o adoecimento mental no contexto neoliberal. *Revista Profissão Docente*, 19(40), 01–17. <https://doi.org/10.31496/rpd.v19i40.1242>

MOUBAYED, Abdallah et al. Student engagement level in an e-learning environment: Clustering using k-means. *American Journal of Distance Education*, v. 34, n. 2, p. 137-156, 2020.

MICHAELSEN, Larry K.; SWEET, Michael. The essential elements of team-based learning. *New directions for teaching and learning*, v. 2008, n. 116, p. 7-27, 2008.

MITCHELL, T. M. *Machine learning* (WBC/McGraw-Hill, Boston). MA, 1997.

MORAIS, I. R. D. et al. *Ensino Remoto Emergencial: Orientações básicas para elaboração do plano de aula*. 2020. 2020.

NAVARRO, Mairlos Parra. PROPOSTA DE CLASSIFICAÇÃO DE METODOLOGIAS ATIVAS DE APRENDIZAGEM. *International Journal on Active Learning*, v. 2, n. 2, p. 8-20, 2017.

NALLI, Giacomo et al. Comparative Analysis of Clustering Algorithms and Moodle Plugin for Creation of Student Heterogeneous Groups in Online University Courses. *Applied Sciences*, v. 11, n. 13, p. 5800, 2021.

OUNNAS, Asma. *Enhancing the automation of forming groups for education with semantics*. 2010. Tese de Doutorado. University of Southampton.

ONAN, Aytuğ; TOÇOĞLU, Mansur Alp. Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts. *Computer Applications in Engineering Education*, 2020.

OLIVEIRA, Leonam; ROSA, Selma Santos; PIMENTEL, Andrey. Revisão Sistemática da Literatura: Formação de Grupos na Aprendizagem Colaborativa com Suporte Computacional. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2019. p. 1955

OLIVEIRA, Amanda Gondim de. *Contribuições ao estudo da dinâmica na teoria da informação: aplicações em clustering dinâmico*. 2018.

OLIVEIRA, Dalila Andrade. A reestruturação do trabalho docente: precarização e flexibilização. **Educação & Sociedade**, v. 25, n. 89, p. 1127-1144, 2004.

OYELADE, O. J.; OLADIPUPO, Olufunke O.; OBAGBUWA, I. C. Application of k Means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425*, 2010.

PASCHOALINO, Jussara Bueno de Queiroz; RAMALHO, Mara Lúcia; QUEIROZ, Virgínia Coeli Bueno de. *Trabalho docente: o desafio de reinventar a avaliação em tempos de pandemia*. 2020.

PAIVA, Ranilson Oscar Araujo et al. A systematic approach for providing personalized pedagogical recommendations based on educational data mining. In: *International Conference on Intelligent Tutoring Systems*. Springer, Cham, 2014. p. 362-367.

PAIVA, Ranilson et al. Mineração de dados e a gestão inteligente da aprendizagem: desafios e direcionamentos. In: *Anais do Workshop de Desafios da Computação Aplicada à Educação*. 2012. p. 158-167.

PERRENOUD, Philippe. Dez novas competências para ensinar. Artmed editora, 2015.

PERRENOUD, P. Pedagogia diferenciada: das intenções à ação. Porto Alegre: Artes Médicas, 2000.

PIMENTEL, Edson P.; DE FRANÇA, Vilma F.; OMAR, Nizam. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2003. p. 495- 504.

PIMENTEL, Mariano. Princípios do desenho didático da educação online. Revista Docência e Cibercultura, v. 2, n. 3, p. 33-53, 2018.

PEREIRA, Rodrigo. Método Ativo: Técnicas de Problematização da Realidade aplicada à Educação Básica e ao Ensino Superior. In: VI Colóquio internacional. Educação e Contemporaneidade. São Cristóvão, SE. 20 a 22 setembro de 2012.

PEREIRA, Aluisio José et al. Identificação e caracterização de níveis de interação no ensino remoto de emergência na Educação Básica. In: Anais do XXXII Simpósio Brasileiro de Informática na Educação. SBC, 2021. p. 145-156.

PEREIRA, Priscilla et al. Análise de perfis de interação de alunos no ambiente virtual Moodle via agrupamento hierárquico. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2018. p. 1413.

RAMOS, J. L.C. Uma abordagem preditiva da evasão na educação a distância a partir dos construtos da distância transacional. Tese de Doutorado (Centro de Informática –UFPE), 2016. Disponível em <http://bit.ly/TeseJorge> . Acesso em 10 jan 21

RAMOS, Jorge Luis Cavalcanti et al. A comparative study between clustering methods in educational data mining. IEEE Latin America Transactions, v. 14, n. 8, p. 3755-3761, 2016.

RAMOS, Jorge Luis C. et al. CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais. 2020.

RAMOS, Jorge Luis Cavalcanti et al. Identificação de perfis de interação de estudantes de educação a distância por meio de técnicas de agrupamentos. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. SBC, 2020. p. 932-941.

RODRIGUES, Alessandra. Ensino remoto na Educação Superior: desafios e conquistas em tempos de pandemia. SBC Horizontes, jun. 2020. ISSN 2175-9235. Disponível em: <<http://horizontes.sbc.org.br/index.php/2020/06/17/ensino-remoto-na-educacao-superior/>>. Acesso em: 03 mês. 2022.

ROMERO, Cristóbal; ROMERO, José Raúl; VENTURA, Sebastián. A survey on pre-processing educational data. In: Educational data mining. Springer, Cham, 2014. p. 29-64.

SANTANA, Roniel Venâncio; PONTES, Heráclito Lopes Jaguaribe. Aplicação da

Clusterização por K-meas.

SANTOS, Denise Cristina Augusto dos. Precarização do trabalho e suas consequências na vida do professor. 2021.

SILVA, Paulo et al. Modelos de regressão aplicados a predição do desempenho escolar de estudantes do ensino fundamental. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2019. p. 1621.

SOUZA, João Paulo Lopes; DE SOUZA FERNANDES, Damires Yluska; DUTRA, Janderson Ferreira. Predição precoce de problemas de desempenho de estudantes em modalidade de educação online: um estudo de caso no ensino médio integrado. 2020.

SCHIEHL, Edson Pedro; GASPARINI, Isabela. Modelos de ensino híbrido: Um mapeamento sistemático da literatura. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2017. p. 1.

VENTURINI, Simone Ferigolo; SILVA, Taís Oliveira. Uso e benefícios das metodologias ativas em uma disciplina de engenharia de produção. Cippus, v. 6, n. 1, p. 59-74, 2018.

VIGOTSKY, L. A. Formação social da mente. São Paulo, Martins Fontes, 1984.

WOMMER, Fernanda Gabriela Bitencourt et al. Métodos ativos de aprendizagem: uma proposta de classificação e categorização. Revista Cocar, v. 14, n. 28, p. 109-131, 2020.

WIRTH, Rüdiger; HIPPEL, Jochen. CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. London, UK: Springer-Verlag, 2000. p. 29-39.

APÊNDICE A – Tabela comparativa dos trabalhos relacionados e a abordagem proposta

Quadro A.1 - Síntese dos estudos relacionados à pesquisa

Estudo/Publicação	Natureza/Fonte de dados	Objetivo/Foco do estudo	Algoritmos de agrupamento	Principais atributos considerados	Tipo de grupos formados	Grupos identificados
Monteverde et al. (2018)	Dados históricos de estudantes, Moodle	Formação de grupos de estudantes para atividades de aprendizagem colaborativa em AVA	K-Means	<ul style="list-style-type: none"> • Recurso/atividade • Quantidade, • Dispersão e variância das arestas padrão, avanço e de retorno 	Homogêneos	-
De França et al. (2013)	Dados históricos de estudantes	Identificação de grupos de estudantes que apresentam dificuldades de aprendizagem.	K-Means	<ul style="list-style-type: none"> • Notas de atividades • Nível da TB 	Homogêneos	-
Nalli et al. (2021)	Dados históricos de estudantes, Moodle	Auxílio docente mediante a formação de grupos de estudantes de acordo com o nível de atividade no AVA.	k-means, Mean-Shift Clustering, WARD, DBSCAN, GMMC e SOM.	<ul style="list-style-type: none"> • número de acesso à plataforma, • porcentagem de estudantes que concluíram a tarefa; • frequência total de exibição experimentos em vídeo (média); • número de visualizações de vídeos (média); • frequência total de visualizações de tutoriais em vídeo (média); • número de tutoriais em vídeo visualizados (média); 	Heterogêneos	<ul style="list-style-type: none"> -Alta atividade, -Média atividade, -Baixa atividade

Estudo/ Publicação	Natureza/ Fonte de dados	Objetivo/F oco do estudo	Algoritmos de agrupame nto	Principais atributos considerados	Tipo de grupos formados	Grupos identificados
Maina et al. (2017)	Dados históricos de estudantes, Moodle	Auxílio docente mediante a formação de grupos de estudantes de acordo com o nível de competência de colaboração .	Skmeans e EM	Interação em Fóruns: <ul style="list-style-type: none"> • Número de postagens; • Número de respostas; • Avaliações do fórum. 	Heterogêneos	-Alto, -Médio -Baixo
(Ramos et al., 2020)	Dados históricos de estudantes	Identificação de perfis de interação de estudantes na EAD	K-Means e WARD	Quantidade de: <ul style="list-style-type: none"> • Postagens do aluno em fóruns por disciplina, • Mensagens enviadas pelo estudante no semestre, • Mensagens recebidas pelo aluno no semestre, • Acesso por turno: manhã, tarde e noite, no semestre. Colegas diferentes que o estudante enviou mensagens no semestre, • Acessos do aluno ao ambiente no semestre, • Mensagens recebidas de colegas no AVA no semestre, 	Homogêneos	-Alta interatividade, -Média interatividade -Baixa interatividade
ActivPlan	Dados históricos de estudantes informados pelo docente.	Identificação de perfis de estudantes e Geração de grupos para PAA.	K-Means	<ul style="list-style-type: none"> • Questionários aplicados durante o curso. • Tempo de resposta a atividade • Tempo de acesso ao AVA • Competências para desejadas ou <ul style="list-style-type: none"> • Questionários aplicados durante o curso. • Tempo de resposta a atividade • Tempo de acesso ao AVA • Média das notas referentes ao histórico acadêmico no ano anterior 	Heterogêneos	-Muito Alto -Alto, -Médio -Baixo -Muito baixo.

APÊNDICE B – Avaliação de métodos de clusterização

Quadro B1: Comparativo de resultados na avaliação de métodos de agrupamento

Método	Quantidade de instâncias por cluster	Índice de silhueta	Descrição da análise do método
dbscan	Cluster -1: 50 Cluster 0: 35 Cluster 1: 13 Cluster 2: 9 Cluster 3: 7 Cluster 4: 6 Cluster 5: 13	0,52	Um método que obteve melhor desempenho no índice de silhueta. Porém, muitos grupos desbalanceados, pode-se considerar sua aplicação em trabalhos futuros mesmo apresentando muitos grupos desbalanceados (possíveis outliers).
ward	Cluster 0:115 Cluster 1:10	0,46	Apesar de apresentar o segundo melhor resultado do índice de silhueta. Geram-se apenas dois grupos desbalanceados.
k-means (5 cluster)	Cluster 0:9 Cluster 1:45 Cluster 2:26 Cluster 3:27 Cluster 4:18	0,36	Geração de grupos mais balanceada, apesar de apresentar o índice de silhueta 0,36.b

ANEXO A – Algoritmo para formação de grupos heterogêneos (Maina et al., 2017)

A1: Algoritmo de formação de grupos heterogêneos elaborado por (Maina, 2017, p.30-31)

```
start_session := load_csv_file <- filename(mdl_cluster_temp)
  declare variable and initialize(<- inputs
int(i,j,n,a,b,no_of_cluster,userst,no_of_groups,rank)
  declare variable and initialize(<- inputs
array(random_array, new_array,test_array,group_array)

  // store cluster asignment in double dimensional
  array:(Array[i][j])
  foreach(no_of_cluster);
  userst=<get_recordset>;
  foreach(userst)
  Array[i][j]=userst;
  j++; i++;
  for(i=0;i<n;i++)
  random_array=Array[i];
  shuffle(random_array);

  // randomize the array by using shuffle function

  test_array[]=random_array;

  // assign members to groups
  for(a=0;a<no_of_groups;a++)

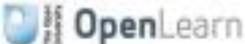
  for(b=0;b<sizeof(test_array[a]);b++)
  new_array[]=test[a][b];
  for(c=0;c<no_of_groups; c++)
  for(rank=c;rank<sizeof(new_array);rank+=no_of_groups)
  group_array[c][rank]=new_array[rank];
  add_group_data_into_Moodle:= mdl_groups_members
  <- input(group_array[c][rank])
  exit_session()
```

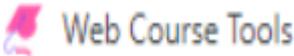
Fonte: (MAINA,2017,p.30-31)

ANEXO B – AVAs e geração de grupos de estudantes

B1: Descrição de AVA e suas aplicabilidades quanto à formação de grupos

NOME DO AVA	DESCRIÇÃO	APLICABILIDADE EM GRUPOS
	<p>Este ambiente tem funcionalidades que visam a facilitar a interação/ comunicação entre os participantes e o uso integrado de diferentes recursos.</p> <p>Disponível em: <https://ead.ufrgs.br/rooda/></p>	<p>Possibilita criar grupos de forma livre, aleatória e por combinação socioafetiva.</p>
	<p>Conhecido como o sistema de gestão da aprendizagem mais popular do mundo.</p> <p>Disponível em: <https://moodle.org/></p>	<p>Possibilita criar grupos de forma aleatória.</p>
	<p>Plataforma para cursos e treinamentos online. Disponível em:</p> <p><https://lmsstudio.com.br/></p>	<p>Possibilita criar grupos de forma aleatória.</p>
	<p>Ambiente para realização de cursos a distância através da Internet.</p> <p>Disponível em:</p> <p><http://teleduc4.multimeios.ufc.br/></p>	<p>Permite a criação de grupos para facilitar a distribuição e/ou desenvolvimento de tarefas.</p>
	<p>Ambiente para realização de cursos para ensino a distância, complementar às atividades de educação presencial e treinamento de profissionais.</p> <p>Disponível em:</p> <p>http://www.eduweb.com.br/elearning_tecnologia.asp ? ></p>	<p>Não possui a funcionalidade de grupos.</p>
	<p>Ambiente para realização de cursos.</p> <p>Disponível em: http://eproinfo.mec.gov.br/</p>	<p>Não possui a funcionalidade de grupos.</p>

NOME DO AVA	DESCRIÇÃO	APLICABILIDADE EM GRUPOS
	<p>Ambiente para criação de cursos, organização e distribuição de conteúdo.</p> <p>Disponível em: <https://blog.hotmart.com/pt-br/educacao/></p>	<p>Permite que o produtor crie diferentes turmas de acesso ao seu conteúdo, podendo vincular módulos a um grupo específico de alunos.</p>
	<p>Proporciona ensino e aprendizado conectando alunos de qualquer lugar a instrutores de todo mundo.</p> <p>Disponível em: <https://www.udemy.com/></p>	<p>A criação de grupos de usuários é um recurso opcional que permite segmentar sua base de usuários para fins de criação de relatório.</p>
	<p>Proporciona fluxos de trabalho intuitivos e integrações simplificadas, possui recursos e ferramentas para ajudar a inspirar o interesse por aprender.</p> <p>Disponível em: <https://br.blackboard.com/index.html></p>	<p>Possibilita criar grupos de forma aleatória. O professor pode remanejar e ajustar a ordem dos participantes.</p>
	<p>Plataforma de aprendizagem gratuita, fornecida pela The Open University.</p> <p>Disponível em: <https://www.open.edu/openlearn/></p>	<p>Não foram encontradas informações sobre grupos.</p>
	<p>Plataforma de cursos online.</p> <p>Disponível em: <https://es.eadbox.com/plataforma_ava/></p>	<p>Não foram encontradas informações sobre grupos.</p>
	<p>Projetada para se destacar por entregar recursos necessários para que professores e alunos possam ter as melhores experiências.</p> <p>Disponível em: <https://eadplataforma.com/></p>	<p>Possui gerenciamento de grupos. Acompanha o processo de seus alunos por meio de boletins.</p>

NOME DO AVA	DESCRIÇÃO	APLICABILIDADE EM GRUPOS
	<p>Plataforma de cursos online com recursos para empresas, consultorias e professores.</p> <p>Disponível em: <https://www.elore.com.br/#!/recursos></p>	<p>Não foram encontradas informações sobre grupos.</p>
	<p>Cursos e treinamentos online.</p> <p>Disponível em: <https://nochalks.com/></p>	<p>Não foram encontradas informações sobre grupos.</p>
	<p>Plataforma para cursos online.</p> <p>Disponível em: <https://maestrus.com/plataforma/></p>	<p>Não possui funcionalidade de grupos.</p>
	<p>Plataforma de cursos e treinamentos online.</p> <p>Disponível em: <https://doeplayer.com.br/50445042-O-uso-do-ambiente-virtual-de-aprendizagem-ava-dokeospara-formacao-de-professores-no-ensino-superior.html></p>	<p>Não foram encontradas informações sobre grupos.</p>
	<p>Sistema de gestão do aprendizado.</p> <p>Disponível em: <https://www.ufrgs.br/soft-livre-edu/software-educacional-livre-na-wikipedia/amadeus-lms/></p>	<p>Formação de grupos de estudos.</p>
	<p>Plataforma de cursos online.</p> <p>Disponível em: <https://webcoursetools.com/></p>	<p>Não foram encontradas informações sobre grupos.</p>

NOME DO AVA	DESCRIÇÃO	APLICABILIDADE EM GRUPOS
	<p>Ambiente virtual de aprendizagem que busca a construção coletiva do conhecimento.</p> <p>Disponível em: <http://eureka.pucpr.br/entrada/index.php>.</p>	<p>Possui ferramentas para criação de grupos.</p>
	<p>Ajuda alunos e professores a organizar as tarefas, aumentar a colaboração e melhorar a comunicação.</p> <p>Disponível em: <https://edu.google.com/intl/ptBR/products/classroom/></p>	<p>Possui ferramentas para criação de grupos.</p>
	<p>Oferece ferramentas digitais para professores e alunos criarem experiências de aprendizagem conectadas.</p> <p>Disponível em: <https://www.instructure.com/canvas/pt-br></p>	<p>Possui ferramentas para criação de grupos.</p>

Fonte: (FERREIRA et al., 2021)