

# Formação de Grupos Heterogêneos de Estudantes para Aplicação de Práticas Ativas de Aprendizagem

## *Formation of Heterogeneous Groups of Students for the Application of Active Learning Practices*

Johnny Yuri Solano Marinho  
Instituto Federal da Paraíba (IFPB)  
ORCID: [0000-0002-0676-0524](https://orcid.org/0000-0002-0676-0524)  
[johnnyyuri88@gmail.com](mailto:johnnyyuri88@gmail.com)

Damires Yluska de Souza Fernandes  
Instituto Federal da Paraíba (IFPB)  
ORCID: [0000-0001-9305-5323](https://orcid.org/0000-0001-9305-5323)  
[damires@ifpb.edu.br](mailto:damires@ifpb.edu.br)

### 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 dos estudantes. Isso ocorre também em cenários onde uma turma de estudantes se caracteriza por sua heterogeneidade, seja em termos de perfis cognitivos, interacionais, níveis de aprendizagem ou com relação a outros aspectos. Nessa perspectiva, este trabalho propõe uma abordagem, denominada *ActivePlan*, que objetiva identificar perfis cognitivos e interacionais de estudantes em uma turma e, com base neles, formar grupos heterogêneos. A ideia da abordagem é: (i) agrupar estudantes por similaridades cognitivas e interacionais, com vistas ao reconhecimento de perfis parecidos e, em seguida, (ii) formar grupos compostos por perfis, desta vez, heterogêneos de estudantes com vistas à aplicação de Práticas Ativas de Aprendizagem (PAA). A abordagem faz uso de um método de aprendizado de máquina não supervisionado para criação de grupos de estudantes com perfis homogêneos e propõe um algoritmo adaptado para a formação de grupos com perfis heterogêneos de estudantes para uso em PAAs. A abordagem foi avaliada com base em dados reais de estudantes em turmas do ensino técnico integrado ao médio, considerando a PAA denominada Trezentos. A PAA Trezentos faz uso de grupos de estudantes que sejam potencialmente colaborativos e, para isso, considera diferentes características ou papéis dos membros dos grupos para facilitar a colaboração. Neste trabalho, os conhecimentos comuns associados ao desempenho dos estudantes (e.g., desempenho muito alto ou desempenho baixo) foram usados como critério de formação dos grupos, tanto para identificar competências similares (etapa de agrupamento homogêneo) quanto na formação de grupos com perfis diferentes (etapa de formação heterogênea). O resultado da formação de grupos heterogêneos foi avaliado por meio da métrica diversidade, cujo valor médio obtido foi 3.3. Esse valor, juntamente com a análise realizada indicam que a abordagem traz resultados promissores no tocante à formação de grupos de estudantes para a prática Trezentos. **Palavras-Chave:** Aprendizagem ativa; Grupos heterogêneos; Aprendizado de máquina; Agrupamento; Suporte docente; Taxonomia de Bloom.

### 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, which are growing as tools for enhancing engagement, interactivity and also school performance of students, have been considered. Some active learning practices are aimed at heterogeneous student classes. In this light, this work proposes an approach, called *ActivePlan*, which proposes to identify cognitive and interactional profiles of students in a class and, based on them, forms heterogeneous groups. The idea is organized by means of two aspects: (i) to group students at first by similar cognitive and interactional profiles by recognizing common knowledge and (ii) to form heterogeneous groups of students which are associated with some active learning practices (ALP). The *ActivePlan* approach has been evaluated by means of a real dataset composed by students from a technical integrated course and by considering the ALP Three Hundred. The Three Hundred ALP is based on the formation of groups of students that are potentially collaborative and, for this, considers different characteristics or roles of the members of the groups to facilitate collaboration. In this work, therefore, common knowledge associated with student performance (e.g., very high performance or low performance) has been used as a criterion for forming groups, both to identify similar competences (homogeneous grouping stage) and also to form groups with different profiles (heterogeneous formation stage). Through these identified common competences, groups with mixed profiles of

Cite as: Marinho, J. Y. S. & Fernandes, D. Y. S. (2023). Formação de Grupos Heterogêneos de Estudantes para Aplicação de Práticas Ativas de Aprendizagem. *Revista Brasileira de Informática na Educação*, 31, 87-116. DOI: 10.5753/rbie.2023.2591

*students compatible with the practice Three hundred have been generated. The result of the formation of heterogeneous groups was evaluated using the diversity metric, whose average value was 3.3. This value, together with the analysis carried out, indicate that the approach brings promising results w.r.t. the formation of groups of students generated with regard to the practice Three hundred.*

**Keywords:** *Active learning; Heterogeneous groups; Machine learning; Clustering; Teaching support; Bloom's Taxonomy.*

## 1. Introdução

Diante do atual panorama educacional brasileiro, com baixos desempenhos escolares (OCDE, 2018) e de práticas tradicionais ainda associadas à pedagogia bancária (Freire, 1981), comumente utilizadas no processo de ensino e aprendizagem, o professor pode exercer papel fundamental para a transformação desse cenário (Moran, 2007). Para que o professor possa se amparar e ser cada vez mais um agente transformador, é importante que ele tenha suporte ferramental e diretrizes que possam ajudá-lo no planejamento de suas aulas assim como no acompanhamento aos discentes.

Um dos desafios do professor refere-se à prática docente no que tange a estratégias didáticas em turmas compostas por estudantes que possuem **perfis heterogêneos** (Perrenoud, 2015). Um cenário educacional heterogêneo se caracteriza quando o docente, em uma turma de estudantes, precisa lidar com diferentes perfis em termos de níveis cognitivos, interacionais, estilos ou ritmos de aprendizagem, entre outros fatores que podem impactar os resultados de estudos. A heterogeneidade de perfis de estudantes, um aspecto comum em cenários educacionais, por vezes não é devidamente considerada, em especial quando o docente recorre a práticas de ensino reconhecidas como mais tradicionais. A diversidade existente em perfis de estudantes requer medidas e ações docentes mais específicas (Perrenoud, 2015).

Metodologias ativas de aprendizagem vêm sendo cada vez mais consideradas como alternativas potenciais no auxílio às práticas docentes e ao processo de ensino e aprendizagem (Moran, 2015), inclusive no tocante a cenários de turmas de estudantes com perfis diferentes. Uma das características das metodologias ativas se refere à interatividade e à colaboração, consonantes com a ideia de interação social e de mediação docente como ponto central do processo educativo (Vigotsky, 1984). Para a aplicação de metodologias ativas, é comum organizar os estudantes em grupos visando à colaboração e à interação (Moran, 2015). Algumas metodologias ativas de aprendizagem como, por exemplo, Sala de Aula Invertida (Bacich et al., 2018), Instrução por Pares - IP (Mazur, 2015), Trezentos - TR (Fragelli, 2019), Aprendizagem Baseada em Equipes – ABE (Bollela, 2014), entre outras, aparecem como propostas diferenciadas para dinamizar um ambiente escolar e auxiliar o docente na condução do ensino em turmas heterogêneas. Compreender como agrupar melhor os estudantes, pode auxiliar o docente na aplicação dessas metodologias.

Metodologias ativas de aprendizagem são caracterizadas principalmente pela inserção do aluno/aprendiz como agente principal responsável pela sua aprendizagem. As metodologias ativas de aprendizagem podem ser definidas como “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). Diversas técnicas e estratégias educacionais se incluem nesse modelo metodológico. Para fins de padronização, neste artigo, utiliza-se o termo “Práticas Ativas de Aprendizagem (PAA)” na referência a métodos, técnicas, abordagens, ou estratégias de ensino centradas no estudante, com o intuito de auxiliar a condução da prática docente diante de um cenário de turma com perfis heterogêneos de estudantes. Adicionalmente, este trabalho considera, como fatores para identificação de perfis, aspectos cognitivos e interacionais de estudantes.

Além da reflexão a respeito da formação de grupos de estudantes para aplicação de práticas de aprendizagem, é importante lembrar que essas práticas podem também ser impactadas pelo tipo de modalidade empregada para o ensino (e.g., presencial, ensino a distância). Como ilustração, segundo dados do Censo da Educação Superior (INEP, 2020), o cenário educacional brasileiro encontra-se em constante transformação também em relação às modalidades de ensino existentes. Percebe-se que gradativamente o número de ingressos e de matrículas em cursos a distância (EAD) no ensino superior vem crescendo na última década. Pensar e executar práticas

de ensino na modalidade à distância ou remotamente é uma tendência principalmente em tempos recentes de pandemia.

Particularmente, o ensino remoto 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 (Morais et al., 2020). O ensino remoto se diferencia da EAD, tendo em vista que a modalidade EAD representa uma categoria de ensino com presença de atores (professor, tutores) com papéis bem definidos, desenvolvida em ambiente virtual e possui materiais específicos para o contexto da disciplina e de estudantes (Morais et al., 2020).

Conforme o panorama considerado, este trabalho busca responder a duas questões de pesquisa, a saber:

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 por estudantes com perfis heterogêneos, de modo que os grupos formados possam ser usados em PAAs?

Para responder às questões de pesquisa, este trabalho apresenta uma abordagem, denominada *ActivePlan*, que objetiva, com base nos perfis cognitivos e interacionais de estudantes em uma turma, formar grupos com perfis heterogêneos para aplicação de PAAs. A ideia é que o perfil, do ponto de vista cognitivo, pode ser obtido a partir do desempenho em atividades que lidam com conhecimentos conforme as diretrizes da Taxonomia de Bloom. O perfil interacional, por sua vez, pode ser identificado utilizando-se as interações dos estudantes obtidas em um Ambiente Virtual de Aprendizagem (AVA) online. A abordagem *ActivePlan* faz uso de um método de aprendizado de máquina não supervisionado para criação de grupos com perfis similares de estudantes e propõe um algoritmo adaptado para a formação de grupos com perfis heterogêneos, de modo que estes grupos mistos possam ser utilizados para favorecer algumas PAAs.

Neste artigo, a abordagem *ActivePlan* é instanciada e avaliada considerando a PAA denominada Trezentos (Fragelli, 2019). A PAA Trezentos se baseia na formação de grupos de estudantes que sejam potencialmente colaborativos. Para isso, considera perfis de estudantes de acordo com seu desempenho em atividades e avaliações, buscando grupos heterogêneos compostos por estudantes com bom, médio e baixo rendimentos. Para avaliar a abordagem no contexto da PAA Trezentos, foram utilizados dados obtidos em um contexto de ensino remoto, no ano de 2020, relativos a estudantes da educação profissional técnica integrada ao nível médio. Os resultados indicam que é possível identificar os diferentes estágios de desempenho dos estudantes, principalmente os estudantes de desempenho mais elevado e os de desempenho baixo. No que tange à formação de grupos para a PAA Trezentos, os resultados apontam que é possível a atribuição de estudantes de desempenho mais elevado em, pelo menos, 20% dos grupos formados. Foram obtidos grupos mistos voltados à aplicação da PAA Trezentos, o que indica que a abordagem é promissora para este tipo de formação. Com base em uma métrica que ajuda a avaliar a diversidade dos grupos, o trabalho analisa os resultados obtidos. O artigo está organizado da seguinte forma: a Seção 2 aborda conceitos e alguns trabalhos relacionados a esta temática. A Seção 3 apresenta a abordagem *ActivePlan*. Na Seção 4, são descritos os resultados obtidos por meio da implementação de uma ferramenta e da realização de uma avaliação experimental. Por fim, a Seção 5 discorre sobre conclusões e indica trabalhos futuros.

## 2. Referencial teórico e trabalhos relacionados

Nesta seção, são introduzidos alguns conceitos juntamente com a descrição de estudos relacionados a esta temática. São apresentadas também as métricas de avaliação utilizadas neste estudo.

### 2.1 Práticas ativas de aprendizagem e formação de grupos heterogêneos

Este trabalho foca em práticas ativas de aprendizagem que podem ser beneficiadas com a formação de grupos de estudantes com perfis heterogêneos. Em especial, o estudo realizado neste trabalho concentra-se na aplicação das definições organizadas por meio da abordagem *ActivePlan* à PAA denominada Trezentos. A PAA “Trezentos” estabelece a composição de grupos considerando principalmente estudantes de bom rendimento e de baixo rendimento. Normalmente, dada uma determinada turma de estudantes e, objetivando-se aplicar a prática *Trezentos*, 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, é necessário definir pelo menos um estudante com papel de líder (do tipo ajudante) no grupo, cuja consideração para esse papel ocorre a partir de um bom rendimento identificado. Dessa forma, pelo menos 20% do total de estudantes deve atingir um mínimo satisfatório de desempenho para exercer o referido papel. Quanto ao quantitativo de estudantes por grupo, considera-se o tamanho da turma, da seguinte forma (Fragelli, 2019): (i) em turmas grandes, com 20 estudantes ou mais, utiliza-se um quantitativo de cinco a seis membros por grupo; (ii) no caso de turmas com quantitativo menor que 20 estudantes, emprega-se um quantitativo de três a quatro membros por grupo. No exemplo apresentado na Tabela 1, exemplifica-se como seria a composição de grupos para a PAA trezentos.

Tabela 1 - Exemplo de formação de grupos heterogêneos para a PAA Trezentos (Fragelli, 2019).

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

A Tabela 1 mostra quatro grupos fictícios (os estudantes alocados manualmente com nomes e desempenhos fictícios para o exemplo) definidos com base no critério de desempenho dos estudantes. Cada grupo é composto por cinco membros, cujos papéis são heterogêneos conforme espera a PAA Trezentos. Assim, os grupos possuem um membro ajudante do tipo líder, um ajudante e três ajudados. Essa distribuição visa uma maior colaboração entre os participantes. O exemplo ilustra a necessidade de compor grupos com perfis heterogêneos com relação ao desempenho dos estudantes. Vale salientar a importância 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 para a aplicação da PAA Trezentos. Em particular, grupos formados com estudantes que tenham bom, médio e baixo desempenhos são almejados para que a colaboração ocorra e todos possam se beneficiar.

Existem outras metodologias que necessitam ou sugerem a formação de grupos com perfis heterogêneos de estudantes, a exemplo das metodologias Instrução por Pares - IP (Mazur, 2015) e Aprendizagem Baseada em Equipes – ABE (Bollela, 2014). A metodologia IP prevê o confronto de visões sobre um determinado tema trabalhado em sala de aula. Nesta, sugere-se a composição heterogênea de pequenos grupos de estudantes (dois a cinco membros). Sugere-se na ABE, por sua vez, a criação de grupos heterogêneos (cinco a sete) visando à complementação de perfis de estudantes distintos. Ambas as estratégias serão também aplicadas à abordagem *ActivePlan* em estudos futuros.

## 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 é estruturada em seis níveis, que estão organizados hierarquicamente, conforme apresentados na Tabela 2. Os níveis da TB se relacionam com os objetivos educacionais, conhecimentos e competências do estudante. Ressalta-se que o relacionamento hierarquizado compreende que cada nível se apresenta de forma gradativa. Ou seja, para estudantes alcançarem níveis mais abstratos e profundos, sugere-se alcançar primeiramente os níveis mais básicos.

Tabela 2: Resumo dos níveis da Taxonomia de Bloom quanto ao domínio cognitivo (LIMA, 2009) com adaptações.

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 critica 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, invente e projete.

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 Tabela 2, pode-se elaborar ou identificar quais atividades e objetivos educacionais são pertinentes a um determinado plano de aula (Costa et al., 2019; Costa et al., 2021; 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) objetiva utilizar algoritmos de Mineração de Dados (do inglês, *Data Mining*, ou DM) voltados para tratar problemas oriundos do domínio de aplicações educacionais (Baker et al., 2011). Considera-se que a EDM é uma subárea da MD e, assim, permeia áreas como educação, estatística, ciência da computação e a própria MD (Romero et al., 2013). A MDE utiliza algoritmos de aprendizado de máquina para descobrir padrões e análises a partir de dados que normalmente são gerados e explorados em Ambientes Virtuais de Aprendizagem (AVA) (Silva et al., 2021), como a exemplo do *MOODLE (Modular Object-Oriented Dynamic Learning Environment)* e do *Google Classroom* (Souza et al., 2020). Tais dados disponíveis podem originar-se de interações dos estudantes, professores e tutores, logs de acesso, ou mesmo de atividades de ensino por meio dos AVAs (Dalcin et al., 2021). Estudos de aplicação de MDE estão normalmente focados em assistir docentes e gestores educacionais na mitigação de questões como retenção de alunos, problemas de desempenho escolar, detecção de comportamentos, evasão discente, entre outros.

Visando à organização e sistematização das etapas de um processo de MD, o modelo *Cross Industry Standard Process for Data Mining* (CRISP-DM), descrito por Chapman (2000), vem sendo amplamente utilizado há mais de duas décadas (Martínez-Plumed et al., 2019). O CRISP-DM foi projetado e motivado pela necessidade de facilitar e sintetizar etapas em projetos de mineração de dados e descoberta de conhecimento na indústria. Esse modelo de processo divide-se em seis etapas que compreendem desde a escolha e seleção dos dados até a implantação de uma solução computacional associada ao modelo de aprendizado de máquina obtido. Com foco na organização das etapas para elaboração de soluções no domínio educacional surgiu o *Cross Industry Standard Process for Educational Data Mining* (CRISP-EDM) (Ramos et al., 2020). Nessa extensão, as etapas originais foram refinadas para as seguintes (Ramos et al., 2020): (i) Entendimento do domínio educacional, que visa compreender e elucidar o problema real oriundo do domínio em questão e quais os requisitos intrínsecos a serem contemplados; (ii) Entendimento dos dados educacionais a serem usados; (iii) Preparação dos dados para uso na modelagem do aprendizado de máquina e suas análises associadas; (iv) Modelagem do aprendizado de máquina, onde um método supervisionado ou não supervisionado é utilizado; (v) Avaliação do modelo de aprendizado de máquina obtido conforme métricas estabelecidas e (vi) Implantação da solução computacional no domínio educacional em questão.

O Aprendizado de Máquina (AM) forma a base técnica necessária à MD e, conseqüentemente, à EDM. Dois dos principais paradigmas do AM são o supervisionado e o não supervisionado (Harrington, 2012). No primeiro, o aprendizado ocorre a partir de exemplos rotulados. Já no segundo, o aprendizado é baseado apenas nos objetos do conjunto de dados cujos rótulos são desconhecidos. Neste trabalho, métodos de aprendizado não supervisionados baseados em agrupamento são considerados. O método de agrupamento (do inglês, *clustering*) agrupa objetos ou instâncias semelhantes observando-se similaridades entre elas (a partir de suas variáveis ou atributos) (Han et al., 2011). 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).

Existem diversas abordagens para agrupamento. Algumas são baseadas na criação de grupos seguindo uma hierarquia presente nos dados, outras consideram que grupos são definidos como regiões de alta densidade, separadas por regiões menos densas (Ramos et al., 2016; Oliveira et al., 2022). Neste trabalho um método de agrupamento baseado em partição é utilizado. Um dos principais algoritmos de agrupamento partitivo é o *K-means*, muito utilizado para soluções que utilizam agrupamento em MDE (Dutt et al., 2017). O algoritmo *K-means* faz uso de uma estratégia particional e funciona, resumida e iterativamente (etapas 2 e 3), 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 recalculando 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.

#### 2.4 Avaliação de grupos formados

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 a abordagem proposta, para isso, 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. Elas são descritas a seguir.

No tocante 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 francês *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, obtêm-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 Eq. (1)

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

A Eq. (1) apresenta o cálculo da  $S(i)$ , que representa o valor resultante da silhueta correspondente de uma instância. As variáveis  $a(i)$  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 com relação aos demais clusters. O cálculo da média dos índices *Silhouette* é mostrado na Eq. (2).

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

A Eq. (2) corresponde ao somatório dos valores do índices das instâncias dividido por  $n$ , onde  $n$  refere-se ao número de instâncias pertencentes ao conjunto de dados.

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 Eq. (3), na qual  $n$  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.

## 2.5 Trabalhos relacionados

Apresentam-se, nesta seção, alguns trabalhos relacionados a este que objetivam a formação de grupos de estudantes para suporte à prática docente (Pimentel et al. 2003; Oliveira et al. 2018; 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 de estudantes.

No trabalho de Pimentel et al., (2003), objetivou-se agrupar estudantes, considerando a similaridade entre eles, com o intuito de categorizá-los para um Sistema Tutor Inteligente (STI), conforme dados coletados em questionários e avaliações. Os dados utilizados na experimentação foram coletados em turmas de graduação de cursos de informática de duas instituições de ensino superior no formato presencial. Para o agrupamento dos estudantes, utilizou-se os algoritmos *K-means* e *Self-Organizing Map* (SOM). Um dos experimentos realizados utilizou o algoritmo *K-means*, gerando grupos de estudantes com perfis similares. O trabalho apontou resultados promissores, próximos ao esperado conforme avaliação de um docente, visando à identificação dos grupos de estudantes. A avaliação considerou que os estudantes foram atribuídos aos grupos de forma satisfatória. Apesar de apresentar uma parcela de estudantes alocados a alguns grupos erroneamente, o resultado mostrou grupos consistentes que se aproximaram do esperado pelo docente no estudo de caso. Algumas das sugestões de trabalhos futuros do estudo de Pimentel et al. (2003) indicam a formação de grupos homogêneos e heterogêneos, através do uso de pedagogias diferenciadas e, possivelmente, a criação de grupos heterogêneos a partir dos membros dos grupos homogêneos.

Oliveira et al., (2018) propõem a utilização de metodologias ativas para cursos a distância, na disciplina de programação mediante utilização do PCódigo, desenvolvido em trabalho anterior por Oliveira et al., (2015). O PCódigo é formado por uma interface web integrada ao Moodle que visa reconhecer classes de perfis de estudantes e auxiliar o docente nessa identificação. O PCódigo utiliza arquivos referentes aos códigos-fontes desenvolvidos pelos estudantes. Os códigos são tratados e pré-processados para a realização do agrupamento a partir do algoritmo Bisseting K-Means. O dataset obtido engloba registros com informações de estudantes, com atributos que representam métricas de software utilizadas para avaliação do desempenho dos estudantes. São consideradas 348 métricas de softwares, baseadas por exemplo em: verificação da compilação e execução do código, variabilidade de palavras reservadas, esforço, dificuldade de programação, uso de funções, linhas de código e média de complexidade ciclomática de funções. O experimento demonstrou a potencialidade da estratégia como auxílio docente no reconhecimento de classes de perfis de estudantes que apresentaram dificuldades de aprendizagem, assim como em aspectos como plágio.

O estudo de Monteverde et al., (2018) realiza a formação de grupos de estudantes com perfis homogêneos para atividades de aprendizagem colaborativa no AVA MOODLE. 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. No tocante ao processo de formação dos grupos, o trabalho utiliza a ferramenta M-Cluster proposta por Monteverde et al., (2017), em trabalho anterior. Através do M-Cluster, agrupam-se os estudantes de acordo com o grau de similaridade em termos de interação do estudante no AVA. Aplica-se o algoritmo de agrupamento não supervisionado *K-means*, com experimentos utilizando as distâncias Euclidiana, Cosseno e Manhattan. Dessa maneira, o professor ou tutor seleciona, manualmente, a medida de distância, a quantidade de grupos desejada para formação dos grupos e, ao final, analisa os clusters e grupos de estudantes gerados. Ou seja, o número de clusters é pré-estabelecido pelo docente ou tutor.

Nalli et al., (2021) visam 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 de estudantes incluem: número de acessos à plataforma, porcentagem de estudantes que concluíram a tarefa, frequência total de exibição, valor médio de experimentos em vídeo, número médio de visualizações de vídeos, frequência média de visualizações de tutoriais em vídeo e 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*, *Density-based spatial clustering of applications with noise (DBSCAN)*, *Gaussian Mixture Models Clustering*, e 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. No trabalho, são obtidos 3 grupos (Cluster 0, Cluster 1 e Cluster 2), onde os clusters representam o grau de atividade dos estudantes, ou seja, do mais intenso (Cluster 2) ao menos ativo (Cluster 0). A partir dos clusters obtidos, os estudantes são distribuídos em grupos de estudantes com perfis heterogêneos para utilização do professor. O trabalho implementou também um plugin para o MOODLE com a estratégia proposta.

Maina et al., (2017) também realizaram um estudo visando à geração de grupos compostos por estudantes com perfis heterogêneos. A abordagem apresentada foi dividida em duas etapas: (i) Criação de clusters com grupos homogêneos, e (ii) Implementação de agrupamento inteligente com a formação de grupos heterogêneos. A primeira etapa utiliza dados provenientes de fóruns de estudantes, da seguinte forma: o primeiro atributo é um novo post, que é uma ideia original (*number\_of\_posts*); o segundo é uma resposta a um post (*number\_of\_replies*); o terceiro é a

classificação média das postagens (*forum\_ratings*) oriundos também do ambiente MOODLE. Mediante os métodos *SK-means* e *Expectation Maximization (EM)*, obtém-se três grupos (Cluster 0, Cluster 1 e Cluster 2). Cada cluster está associado a um nível de competência de colaboração, denominados de, respectivamente, “alto”, “médio” ou “baixo”.

O algoritmo proposto por Maina et al., (2017) pode ser resumido da seguinte forma:

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.

Os três últimos trabalhos descritos (Monteverde et al., 2018; Nalli et al., 2021; Maina et al., 2017) se caracterizam pela predominância de utilização do AVA MOODLE e da utilização de dados históricos de estudantes.

De modo geral, os trabalhos relacionados utilizam o algoritmo *K-means* para agrupamento dos estudantes com perfis homogêneos.

Comparando os trabalhos descritos com este, alguns diferenciais podem ser 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; (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 em sua operacionalização, como, por exemplo, 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 proposta no presente 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 três etapas gerais, a saber:

**1ª Etapa - Coleta e preparação de dados:** Coleta de dados oriundos de AVAs e/ou SPAEs; 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 homogêneos:** Aplicação de método de AM baseado em agrupamento para geração de grupos homogêneos; Avaliação da qualidade da formação dos clusters; Identificação dos perfis associados a cada cluster;

**3ª Etapa - Formação de grupos com perfis de estudantes heterogêneos para PAAs:** Escolha da PAA; Criação de grupos heterogêneos combinando diferentes estudantes de diversos grupos homogêneos; 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* se caracteriza 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 acerca dos estudantes de uma determinada turma. A *ActivePlan* recebe dados interacionais e históricos provenientes de fontes como AVAs e/ou de sistemas administrativos. 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 existentes sobre os estudantes no sistema SUAP.

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, este trabalho 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 durante o curso conforme a Taxonomia de Bloom (TB) e da geração de novos atributos com esses indicadores.

As questões de cada atividade se referem a um determinado conteúdo/tema que pode ser trabalhado em sala de aula, 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 et al. (2013). Ou seja, a partir das respostas de suas atividades, os estudantes são associados a um 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 1 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.

Assinale como periférico de entrada (E), de saída (S) ou ainda de entrada e saída (ES) quando for os dois ao mesmo tempo:			
	entrada (E)	saída (S)	entrada e saída (ES)
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"/>

Figura 1: Exemplo de questão rotulada ao nível “um” da TB.

### 3.2 Formação de grupos com perfis homogêneos

Nesta etapa, inicialmente, são criados grupos de estudantes com perfis similares a partir dos dados coletados e tratados na Etapa I. Para isso, utiliza-se o método de agrupamento *K-means*. Os clusters resultantes são formados por estudantes que foram associados considerando a similaridade de seus perfis. Após a criação dos clusters, a qualidade da atribuição da instância a cada grupo é avaliada mediante a média dos índices de silhueta, conforme explicado na Seção 2.4.

Por outro ângulo, cada estudante é associado a uma classe ou categoria conforme o cluster atribuído, tendo em vista que clusters podem ser usados como uma estratégia de classificação ou categorização. No domínio educacional, como ilustração, clusters podem ser classificados quanto a níveis de engajamento (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., 2018).

Os trabalhos citados anteriormente usam as categorizações ou grupos para buscar quantificar características dos clusters com base nos atributos do conjunto de dados. Para isso, eles utilizam as médias oriundas dos valores das características/atributos do conjunto de dados para cada cluster. A partir do valor das médias, ranqueamentos podem ser realizados, como, por exemplo, considerando a ordenação de clusters de acordo com perfis de alunos mais engajados a clusters com estudantes menos engajados. Outro exemplo diz respeito a ordenar os clusters com respeito aos perfis de estudantes considerados mais interativos ao grupo com perfis menos interativos.

Na abordagem *ActivePlan*, define-se a categorização de clusters conforme a PAA em questão. No presente estudo, os clusters são classificados de acordo com o grau de desempenho dos estudantes. Assim, clusters são organizados e ranqueados considerando desde aquele com estudantes com **desempenho muito elevado** até aquele que possua estudantes com **desempenho muito baixo**. As categorias utilizadas neste trabalho são baseadas nos princípios para a geração de grupos da PAA Trezentos (Fragelli, 2019). Dessa forma, quanto à composição dos grupos, têm-se parcelas de estudantes considerados “ajudantes” (**desempenho mais elevado**) e outra parcela de estudantes considerados “ajudados” (**desempenho mais baixo**) conforme exemplificado na Tabela 1 da Seção 2.1. Neste trabalho, cinco categorias de grupos com perfis de estudantes são definidas e aplicadas com relação à PAA Trezentos: Desempenho Muito Elevado (DME); Desempenho Elevado (DE), Desempenho Moderado (DM), Desempenho Baixo (DB) e Desempenho Muito Baixo (DMB). A Seção 4.3.2 aborda essas categorias em mais detalhes.

Os clusters de estudantes com perfis similares compõem uma das entradas para a estratégia de formação de grupos de estudantes com perfis diferentes. A estratégia concebida é adaptada do trabalho de Maina et al. (2017) e é baseada em dois algoritmos, a saber: *Algoritmo\_1* e *Algoritmo\_2*, apresentados a seguir. Os algoritmos são apresentados por meio de um pseudocódigo não associado a nenhuma linguagem de programação específica.

O primeiro algoritmo, denominado *Algoritmo\_1: Activeplan\_main (Código 1)* tem como objetivo receber o conjunto de clusters gerado pelo *K-means* e retornar grupos específicos com perfis heterogêneos de estudantes para uma determinada PAA. Assim, o *Algoritmo\_1* recebe como entrada os seguintes parâmetros: o dataset incluindo a categoria de cluster para cada objeto (estudante), a PAA escolhida (e.g., Trezentos), e o quantitativo de membros por grupos a ser considerado para formação, no caso do professor desejar configurar esse número. Particularmente para a PAA Trezentos, o algoritmo é capaz de definir o quantitativo de grupos e de membros por grupo.

```

-----
Algoritmo_1 - Activeplan_main
-----
Input: lista_objetos_clusters,paa,quantidade_membros_grupos
Output: grupos_paa
Begin
  //laço que percorre o conjunto de clusters de estudantes com perfis similares
01: For each cluster ci in lista_objetos_clusters do
02:   For each objeto in cluster ci do //laço que percorre os objetos do cluster ci
03:     objetos_cluster.add(objeto); //adiciona um objeto do cluster a uma lista
04:   End For;
05:   lista_objetos_clusters.add(objetos_cluster);
   //retorna lista com médias referentes a cada atributo dos objetos do cluster ci
06:   lista_medias_cluster=extracao_media(objetos_cluster)
07:   listas_medias_todos_clusters.add(listas_medias_cluster)
08:   objetos_cluster.clear() //apaga dados do array
09:   lista_medias_cluster.clear() //apaga dados do array
10: End For;
   //Calcula pontos para os clusters considerando a média obtida para cada atributo
11: pontos_clusters:=activeplan_pontos(listas_medias_todos_clusters)
   //Categoriza com perfis de desempenho dos cluster conforme pontuação obtida
12: ids_estudantes:=activeplan_desemp(pontos_clusters,lista_objetos_clusters.index)
   //Forma grupos heterogeneos para PAA definida
13: grupos_paa := activeplan_paa(paa,ids_estudantes,quantidade_membros_grupos)
14: Return grupos_paa
End;

```

Código 1: Algoritmo principal.

O *Algoritmo\_1* apresenta dois *loops* que percorrem todos os clusters e os elementos de cada cluster obtidos na etapa de agrupamento. Em seguida, são coletadas as instâncias de estudantes de cada cluster e as médias referentes aos atributos do cluster. A função *activeplan\_pontos* recebe as médias dos atributos de cada cluster (*listas\_medias\_todos\_clusters*) e retorna uma lista denominada *pontos\_clusters* com uma pontuação para cada cluster. Ao final, tem-se uma lista com a pontuação referente aos clusters. Em caso de empate, o critério de desempate pode mudar de acordo com a PAA. Para este trabalho, tendo em vista que a instanciação da abordagem é realizada para a PAA Trezentos, define-se que os atributos cognitivos são decisivos devido a sua importância na formação de grupos para a PAA Trezentos.

A função *activeplan\_desemp* categoriza 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 categorização, recebe-se a lista de pontuação para cada cluster e uma lista que contém os IDs de cada instância de estudante em cada cluster. Gera-se uma lista denominada *ids\_estudantes* que remete à ordenação decrescente que possui.

Sintetizando os procedimentos realizados no *Algoritmo\_1*, os seguintes passos são realizados:

1. Extrai-se a identificação dos estudantes (IDs) de cada cluster;
2. Determinam-se as médias de cada cluster referentes aos atributos elencados;
3. Calculam-se os pontos para cada cluster a partir das médias dos atributos extraídas conforme passo 2;
4. Categorizam-se os clusters considerando a pontuação obtida, retornando as identificações de estudantes (IDs) devidamente ordenadas conforme o desempenho.
5. A lista ordenada conforme o desempenho, composta de IDs de estudantes, é submetida ao *Algoritmo\_2* (Detalhado na Seção 3.3).

### 3.3 Formação de grupos com perfis heterogêneos de estudantes para PAAs

Além da lista ordenada de IDs de estudantes, outra entrada do *Algoritmo\_2* refere-se à denominada *paa*. Neste artigo, considera-se apenas a Metodologia Trezentos. O parâmetro quantidade de grupos é definido porque pode ser importante para outras PAAs (e.g., ABE). Nesta versão do algoritmo, o parâmetro seria nulo tendo em vista que, no caso da PAA Trezentos, isso é obtido conforme a quantidade de estudantes (linha 02 do *Algoritmo\_2*). O *Algoritmo\_2* está especificado no Código 2.

```

-----
Algoritmo_2 - activeplan_paa
-----
Input: paa,ids_estudantes, quantidade_membros_grupos
Output: grupos_paa
Begin
  // Função len retorna quantidade de estudantes
01: quant_total_estudantes := len(ids_estudantes)
02: if paa == tr then
03:   if quant_estudantes < 20 then // Pequenos grupos compostos de 3 a 4 membros.
04:     quantidade_membros_grupos := 3
05:   else // Grupos grandes com 5 a 6 membros.
06:     quantidade_membros_grupos := 5
07:   End If;
08: End If;
  //Cálculo do número de grupos para PAA
09: quantidade_grupos:= quant_total_estudantes/quantidade_membros_grupos
  //Atribui membros aos grupos
10: for g:=0 to quantidade_grupos do //laço que percorre os grupos
11:   for i:=g to i<quant_total_estudantes step quantidade_grupos do
12:     membros.add(ids_estudantes[i]);
13:   End For;
14:   grupos_paa.add(membros);//adiciona um grupo de estudantes na lista de grupos
15:   membros.clear() // limpa dados armazenados
16: End For;
17: Return grupos_paa;
End activeplan_paa;

```

Código 2: Algoritmo para gerar grupos com versão atual para PAA Trezentos, adaptado do algoritmo de Maina et al. (2017).

Assim, primeiramente, atribui-se a variável *quant\_total\_estudantes* conforme a quantidade de elementos na lista de IDs. O segundo passo refere-se à verificação se é a paa Trezentos. Em caso afirmativo, atribui-se a quantidade de membros por grupos conforme especificado na Seção 2.1. Calcula-se o valor do número de grupos (*quantidade\_grupos*) que considera a quantidade total de estudantes dividida pela quantidade de membros por grupo, obtendo-se o número de grupos. Por fim, todos os grupos são percorridos e os estudantes são alocados aos grupos, sendo o primeiro membro de cada grupo definido como o líder do grupo.

## 4. Resultados e discussão

Os resultados obtidos neste trabalho dizem respeito a três aspectos principais: (i) o dataset enriquecido a partir de um trabalho que obteve dados de turmas de estudantes (Souza et al., 2020); (ii) a implementação de um protótipo para a abordagem *ActivePlan*; e (iii) a realização da avaliação da abordagem e dos resultados obtidos. Os aspectos citados juntamente com os resultados e discussões associadas são descritos a seguir.

#### 4.1 Dataset com perfis de estudantes

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 o 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 este presente trabalho.

O estudo de Souza et al., (2020) coletou e integrou dados a partir das seguintes fontes: SUAP (com 492 instâncias de estudantes) e Google classroom (488 instâncias de estudantes). Assim, os dados oriundos do SPAE e do AVA resultaram em um dataset que contém dados pertencentes a turmas do ensino médio integrado ao curso técnico do Campus Monteiro do IFPB. A Tabela 3 mostra uma descrição do dataset obtido a partir do trabalho citado. As linhas em branco indicam os metadados estruturais oriundos do trabalho de Souza et al., (2020).

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}

As notas das atividades ao longo da disciplina (Q1 e Q2) são relevantes para a avaliação do 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 dados similares aos estudos de Ramos (2020) e Pereira (2018) que utilizam logs de acesso a sistemas visando a identificar o grau de interação ou engajamento do estudante.

Para Pereira et al. (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. Os dados coletados das atividades de múltipla escolha, aplicadas no ensino remoto, correspondem ao período de envio entre o dia 24/08/2020 a 25/10/2020.

Adicionalmente aos atributos já existentes, foram especificados alguns indicadores associados à avaliação empregada 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,

cujas linhas na Tabela 3 estão em cinza, 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. O conjunto de dados produzido diz respeito a estudantes dos Cursos Técnico de Edificações, Manutenção e Suporte em Informática e Edificações totalizando 125 estudantes. Para o tratamento, enriquecimento e manipulação dos dados, foram utilizadas as bibliotecas *numpy*<sup>1</sup> e *pandas*<sup>2</sup>.

## 4.2 Ferramenta/protótipo

Visando tornar todo o processo da *ActivePlan* acessível e transparente ao usuário (professores e/ou pesquisadores), foi implementado um protótipo. A ferramenta foi implementada como uma aplicação WEB, desenvolvida a partir da linguagem python<sup>3</sup>, mediante framework *django*<sup>4</sup>. Apresenta-se a interface principal da ferramenta na Figura 2(a). A ferramenta provê dois perfis de usuários para acesso às funcionalidades: (i) perfil do pesquisador e (ii) perfil do professor.

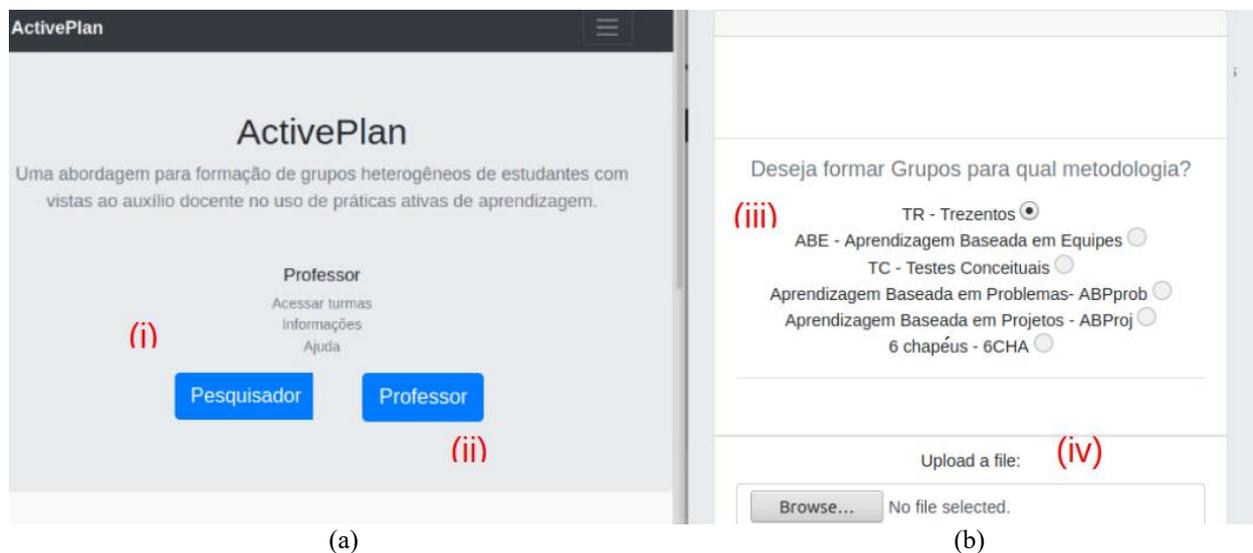


Figura 2: (a) Tela principal da interface da *ActivePlan*; (b) Opções para o módulo do professor.

<sup>1</sup> <https://numpy.org/>

<sup>2</sup> <https://pandas.pydata.org/>

<sup>3</sup> <https://www.python.org/>

<sup>4</sup> <https://docs.djangoproject.com/pt-br/3.2/>

O módulo do professor busca prover suporte ao usuário do tipo professor que vai, de fato, utilizar a ferramenta como apoio a suas práticas docentes. Apresentam-se na Figura 2(b) detalhes da configuração a ser feita pelo professor, o que inclui a inserção do conjunto de dados, (iii) e a escolha de qual PAA deve ser considerada (iv). Um fragmento do conjunto de dados, no formato CSV, utilizado para avaliação da abordagem, é apresentado na Figura 3. Cada linha representa um estudante.

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

Figura 3: Fragmento do conjunto de dados de estudantes.

O usuário professor pode observar os diferentes níveis de desempenho dos estudantes e os grupos heterogêneos formados de acordo com uma PAA escolhida. Inicialmente, esta versão implementa a opção referente à PAA Trezentos.

O módulo do pesquisador permite auxiliar na avaliação da abordagem *ActivePlan* quanto aos objetivos desta pesquisa. Nesse contexto, possibilita acessar resultados das métricas referentes à composição dos grupos similares (clusters) e de grupos heterogêneos gerados e vinculados a uma PAA. O objetivo deste módulo é, principalmente, prover, de modo mais simples, a avaliação dos resultados dos algoritmos quanto a seu desempenho e à qualidade dos grupos formados. A Figura 4 apresenta um diagrama com o fluxo das atividades desenvolvidas na abordagem *Activeplan* no que tange às tarefas (ações) do professor e as ações implementadas na ferramenta.

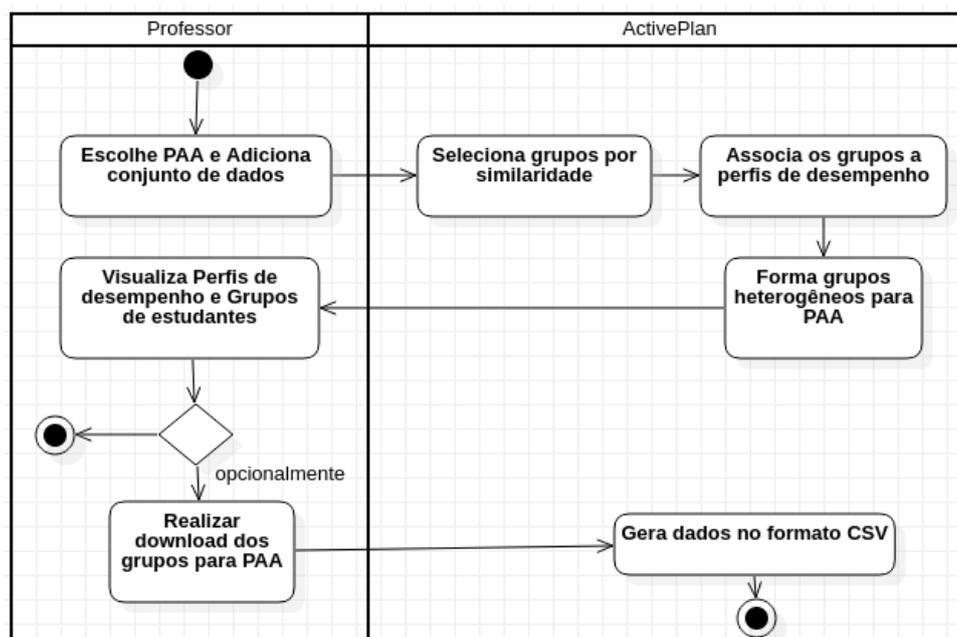


Figura 4: Diagrama de atividades da ferramenta *ActivePlan*.

A Figura 4 apresenta as ações possíveis da ferramenta que se iniciam a partir de dados

informados pelo docente. As ações ocorrem da seguinte maneira (Figura 4):

1. O Professor escolhe a PAA e realiza o upload de um arquivo com os dados (*CSV*);
2. A ferramenta *ActivePlan* agrupa os estudantes (instâncias) por similaridade;
3. A *ActivePlan* associa cada cluster (grupo) a um nível de desempenho;
4. A ferramenta forma grupos heterogêneos para a PAA escolhida;
5. O professor visualiza os perfis de desempenho e os grupos de estudantes heterogêneos para PAA;
6. Opcionalmente o professor pode realizar o *download* do conjunto de dados obtido com os grupos heterogêneos gerados.

### 4.3 Experimentos e Avaliação

A avaliação realizada por meio de experimentos busca analisar os resultados obtidos pela abordagem *ActivePlan* quanto às questões formuladas neste trabalho (Seção 1): (Q1) Considerando perfis de estudantes, caracterizados por aspectos cognitivos e interacionais, pertencentes a turmas ofertadas em ensino remoto, como identificar perfis similares? e (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?

Nesse sentido, dois experimentos foram definidos com os seguintes objetivos: (i) avaliação dos clusters formados a partir dos perfis de estudantes nos conjuntos de dados, e (iii) avaliação dos grupos heterogêneos para aplicação da PAA Trezentos. Os experimentos foram realizados em máquina local, cujas configurações são apresentadas no Quadro 1.

Quadro 1: Configurações para utilização do experimento.

<b>Hardware</b>	Processador Intel, 8ª geração, core i5-8250U, 8GB de memória RAM, 1 Terabyte de armazenamento.
<b>Software</b>	<i>scikit learn</i> , <i>matplotlib</i> , <i>scipy</i> e <i>seaborn</i> .

Conforme apresentado no Quadro 1, a realização de atividades de modelagem do aprendizado (por meio de algoritmo de *clusterização*) e a avaliação dos grupos utilizando a métrica *Silhouette* foram implementadas usando a biblioteca *scikit learn*<sup>5</sup>. Para a implementação de cálculos matemáticos (e.g., cálculo de distância entre instâncias) utilizou-se a biblioteca *scipy*<sup>6</sup>. Na exibição de alguns resultados, utilizou-se bibliotecas como *matplotlib*<sup>7</sup> e *seaborn*<sup>8</sup>.

#### 4.3.1 Clusters de estudantes

Nos métodos de agrupamento partitivos, como o *K-means*, busca-se identificar a quantidade ideal de clusters previamente ao processamento do algoritmo. Han et al., (2011) mencionam uma estratégia muito popular para essa definição que é utilizando método *Elbow* (ou método do cotovelo). O método *Elbow* é utilizado para analisar matematicamente a variação dos valores gerados a partir de cada número de clusters. A partir disso, obtém-se o melhor número de grupos esperado (Hackeling, 2017). A figura 5(i) corresponde aos valores do *elbow*.

<sup>5</sup> <https://scikit-learn.org/stable/modules/clustering.html#clustering>

<sup>6</sup> <http://scipy.org>

<sup>7</sup> <https://matplotlib.org/>

<sup>8</sup> <https://seaborn.pydata.org/>

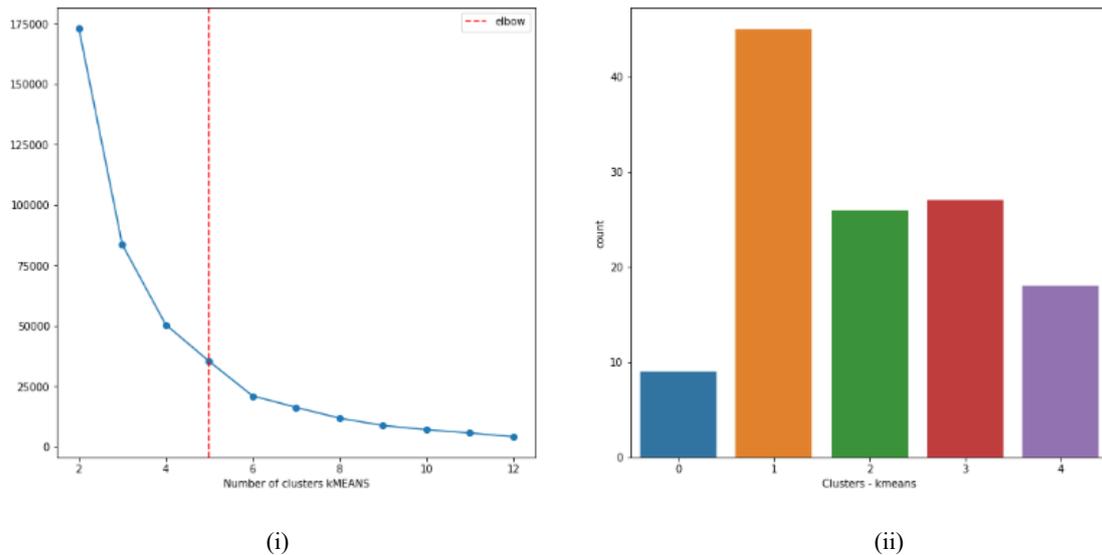


Figura 5: (i) Número de clusters ideal mediante Elbow, (ii) Quantitativo de estudantes por cluster.

A abordagem *ActivePlan* utiliza o método *Elbow* para assistir na identificação do número de clusters a ser empregado. Para o dataset utilizado neste trabalho, o método indicou um quantitativo de **cinco clusters** a ser utilizado, como mostra a linha vertical pontilhada na Figura 5(i). Usando então a configuração prévia de cinco clusters, o *K-means* foi executado, e os resultados obtidos são mostrados na Figura 5(ii). A Figura 5(ii) 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 (Figura 6). 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. Dessa forma, o Cluster 1 obteve um índice silhueta de 0.7. Assim, verifica-se, para este cluster, 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).

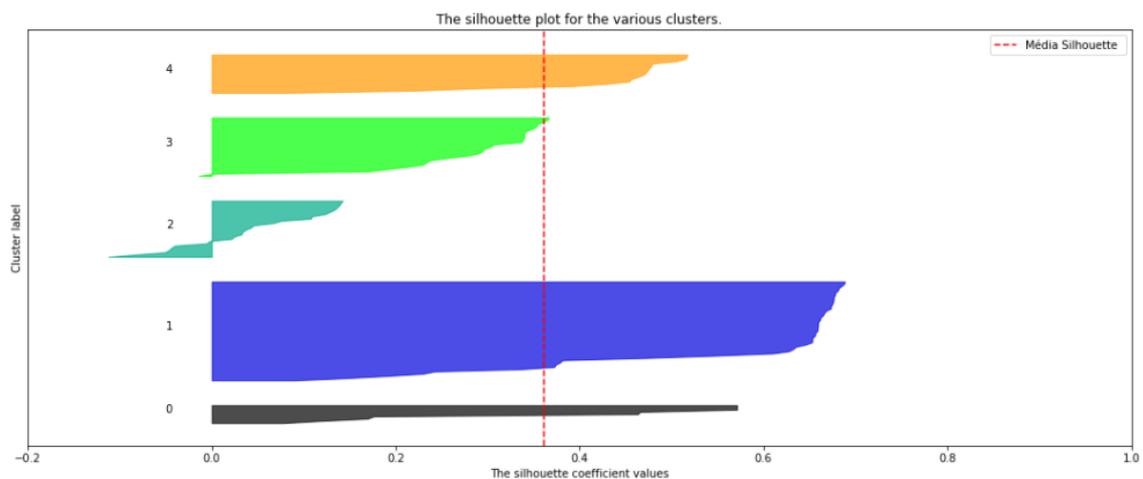


Figura 6: Métrica silhueta

Já o Cluster 3 apresenta valores mais próximos de zero (o que indica valores inconclusivos) e, no Cluster 2, são identificados valores negativos e próximos a zero, o que aponta que alguns dos estudantes pertencentes a esse cluster podem estar erroneamente alocados.

Uma amostra de 25 estudantes pertencentes ao conjunto de dados (Seção 4.1) que representa um montante de 20% do total de 125 estudantes e sua alocação aos clusters são mostrados na Tabela 4. A amostra indica a distinção dos estudantes conforme sua alocação nos clusters, considerando os diferentes valores relativos aos atributos. 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 mostrar que o estudante apresenta problemas no processo de ensino e aprendizagem. 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.

Tabela 4: Amostra de resultados da atribuição de um estudante de cada cluster

Cluster	Q1	Q2	T1	T2	S1	S2	TB1	TB2	TB3	TB4	TB5
0	0.0	0.0	0.88	0.99	1.0	1.0	0	0	0	0	0
0	0.5	0.53	0.26	0.2	1.0	1.0	0	1	0	0	0
0	0.0	0.0	0.88	0.99	0.02	1.0	0	0	0	0	0
0	0.0	0.23	0.22	0.16	1.0	0.04	0	0	0	1	0
1	0.9	1.0	0.3	0.01	0.04	0.06	1	1	1	1	1
1	0.9	1.0	0.06	0.06	0.03	0.05	1	1	1	1	1
1	0.8	0.86	0.05	0.06	0.04	0.05	1	1	1	1	1
1	1.0	1.0	0.03	0.24	0.02	0.05	1	1	1	1	1
1	1.0	1.0	0.3	0.24	0.02	0.04	1	1	1	1	1
1	1.0	0.9	0.04	0.03	0.04	0.04	1	1	1	1	1
2	0.95	0.9	0.06	0.05	0.02	0.05	0	0	1	0	1
2	0.0	0.67	0.88	0.17	0.02	0.0	0	0	1	1	1
2	0.95	0.93	0.06	0.22	0.04	0.05	0	0	1	1	0
2	0.75	0.63	0.24	0.18	0.02	0.04	0	0	1	1	1
2	0.75	0.8	0.23	0.17	1.0	1.0	0	0	1	0	1
3	0.75	0.86	0.04	0.06	0.04	0.04	0	1	1	0	1
3	0.65	0.23	0.04	0.09	0.04	0.03	0	1	1	1	1
3	0.7	1.0	0.09	0.09	0.04	0.03	0	1	1	1	1
3	0.7	0.73	0.06	0.05	0.0	0.05	0	1	1	1	1
3	0.4	0.53	0.04	0.03	1.0	0.05	0	1	0	1	1
4	0.3	0.93	0.01	0.01	0.04	0.04	1	1	0	1	0
4	0.95	1.0	0.03	0.09	0.03	0.03	1	1	1	0	0
4	0.25	0.9	0.06	0.0	0.05	0.01	1	1	1	1	0
4	0.45	0.77	0.04	0.03	0.04	0.04	1	1	0	0	0
4	0.5	0.9	0.01	0.06	0.04	0.04	1	1	1	1	0

Com valores similares ao estudante do Cluster 1, o estudante do Cluster 4 pode demonstrar 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.

#### 4.3.2 Grupos heterogêneos de estudantes

É importante destacar a necessidade de avaliar os grupos e as características predominantes dos estudantes em cada cluster com respeito ao nível de **desempenho**. Isso é mostrado por meio da Tabela 5 que aponta as médias para cada uma das características dos estudantes de acordo com o cluster identificado.

A Tabela 5 propicia uma visão do panorama dos perfis dos estudantes conforme os grupos obtidos e indica uma pontuação para cada atributo vs cluster.

Tabela 5 : Ranking dos clusters a partir das médias de cada atributo

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
<b>Pontuação (score list)</b>	-	41	-	10	-	26	-	18	-	10

Considerando os valores apontados na Tabela 5, 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 5). 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 se refere 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 5 indica o somatório das pontuações. O processo de ranqueamento dos clusters resulta no *score\_list* (apresentado no Código 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 5, o *score\_list* obtido foi, em ordem decrescente, o seguinte: Cluster 0 (41), Cluster 2 (26), Cluster 3(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 na Figura 7. 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 7 apresenta os clusters associados aos níveis de desempenho, o percentual de instâncias por cluster e a lista de IDs (identificações) dos estudantes alocados aos clusters.

Clusters	%	Estudantes	Desempenhos
C0	7,2	[4, 43, 44, 75, 77, 78, 79, 103, 124]	Desempenho muito baixo
C2	20,8	[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	21,6	[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	14,4	[5, 7, 14, 16, 18, 19, 22, 24, 27, 29, 32, 33, 34, 38, 40, 41, 70, 87]	Desempenho elevado
C1	36,0	[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

Figura 7: Resultado da clusterização de estudantes.

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 Código 2 (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 8 apresenta os grupos formados para a PAA Trezentos. O eixo y representa a métrica de diversidade, e o eixo x indica os grupos gerados ( grupo 0 ao grupo 24).

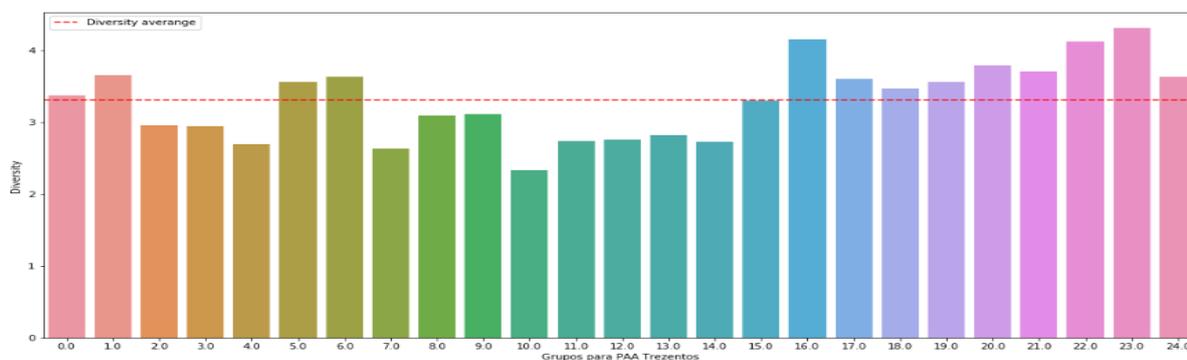


Figura 8 : Resultados da métrica de diversidade para cada grupo quanto à PAA Trezentos.

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 8, 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.

A Tabela 6 apresenta algumas amostras de estudantes pertencentes aos grupos heterogêneos formados com relação a seus diferentes níveis de desempenho (DME, DE, DM, DB e DMB).

Tabela 6: Amostra de estudantes em grupos heterogêneos

Estudantes (IDs)	Q1	Q2	T1	T2	S1	S2	T	T	T	T	Grupo	Nível Desempenho	
							B	B	B	B			B
							1	2	3	4	5		
61	1,0	0,9	0,09	0,09	0,03	0,06	1	1	1	0	1	23	DME
40	0,3	0,93	0,01	0,01	0,04	0,04	1	1	0	1	0		DE
81	0,75	0,86	0,04	0,06	0,04	0,04	0	1	1	0	1		DM
82	1,0	0,8	0,03	0,0	0,03	0,03	1	0	1	1	1		DB
43	0,0	0,0	0,88	0,99	1,0	1,0	0	0	0	0	0		DMB
101	0,9	1,0	0,06	0,06	0,03	0,05	1	1	1	1	1	10	DME
23	0,8	0,86	0,05	0,06	0,04	0,05	1	1	1	1	1		DME
14	0,4	0,97	0,01	0,02	0,04	0,05	1	1	1	1	0		DE
31	0,55	0,86	0,01	0,01	0,05	0,05	0	1	1	1	0		DM
47	1,0	0,83	0,09	0,09	0,02	0,05	1	0	1	1	1		DB
89	0,75	0,76	0,04	0,06	0,03	0,05	1	1	1	0	1	15	DME
6	0,8	0,9	0,3	0,24	0,04	0,04	1	1	0	1	1		DME
117	0,5	0,86	0,32	0,27	0,03	0,05	0	1	1	1	1		DM
114	0,75	0,83	0,03	0,06	0,02	0,04	0	0	1	1	1		DB
15	0,4	0,4	0,01	0,01	0,03	0,04	0	0	0	1	0		DB

A composição mista de grupos de estudantes tem potencial para favorecer a aprendizagem ativa e colaborativa mediante a interação de seus membros e propiciando um ambiente rico para que a aprendizagem ocorra (Vygotsky, 1984; Perrenoud, 2015; Moran, 2015). No tocante aos resultados da abordagem *ActivePlan*, percebe-se que cada grupo é composto de membros com estudantes de diferentes níveis, ou seja, de perfis heterogêneos. Assim, a abordagem conseguiu gerar grupos considerados mistos voltados à aplicação da PAA Trezentos. 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 (ajudante).

Por outro lado, é preciso ainda estender a análise para os casos de estudantes que ficam em grupos *borderline*, ou seja, grupos onde a alocação pode ter ocorrido de modo inconclusivo. Nesses casos, a instância poderia estar em mais de um grupo. Apesar disso, considerando, em especial, a PAA Trezentos onde o papel de ajudado é predominante na formação do grupo, esses casos considerados *borderline* estariam dentro dessa faixa de papel.

## 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. Neste artigo, a abordagem proposta direciona a experimentação para a PAA Trezentos, dentro de um contexto de ensino remoto.

Considerando a PAA Trezentos, a partir do método de clusterização utilizado, foi possível identificar estudantes do tipo ajudantes de forma compatível com o esperado pela prática em questão. Por meio do índice silhueta, verificou-se a qualidade da atribuição de um estudante a um cluster, obtendo-se valores próximos a 0.7 no cluster referente aos estudantes considerados de desempenho muito elevado e quase 0.6 para os clusters associados aos desempenhos elevados e desempenho muito baixo.

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*).

Em trabalhos futuros, pretende-se ampliar a instanciação da abordagem para outras PAAs, a exemplo da ABE (Aprendizagem baseada em equipes) e TC (Testes conceituais). A consideração de perspectivas de outras PAAs e a extensão de possibilidades de enriquecimento dos dados de estudantes são fundamentais para a generalização do modelo obtido.

Outro trabalho futuro refere-se à extensão da 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.

Pretende-se adicionalmente experimentar outros algoritmos de aprendizado de máquina em busca de possíveis melhorias no processo de agrupamento dos perfis. Para isso, estudos visando o refinamento dos algoritmos de agrupamentos e do processo de identificação dos perfis de estudantes podem ser realizados. Um possível trabalho futuro é gerar os grupos heterogêneos em uma só etapa de *clustering*. Em alguns algoritmos de *clustering*, é possível fazer ajustes nos parâmetros do método para gerar clusters mais homogêneos ou mais heterogêneos. Um caminho é já poder identificar as notas mais altas no mesmo cluster (considerando alta homogeneidade), assim já se pode prever os ajudantes. E nas amostras muito diferentes de todas as outras, é possível identificar os alunos com mais dificuldades, isto é, os que precisam ser ajudados. Os estudantes com perfil moderado podem se encaixar em qualquer grupo.

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.

## Agradecimentos

Os autores agradecem o conjunto de dados fornecido por Souza et al. (2020). Também agradece ao Instituto Federal da Paraíba - IFPB e ao Programa de Pós-Graduação de Tecnologia da Informação – PPGTI.

## Referências

- Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, [\[GS Search\]](#)
- Bacich, L., & Moran, J. (2018). Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática. Penso Editora. [\[GS Search\]](#)
- Baker, R., Isotani, S., & Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02), 03. doi: [10.5753/rbie.2011.19.02.03](https://doi.org/10.5753/rbie.2011.19.02.03) .[\[GS Search\]](#)
- Bollela, V. R., Senger, M. H., Tourinho, F. S. V., & Amaral, E. (2014). Aprendizagem baseada em equipes: da teoria à prática. *Medicina (Ribeirão Preto)*, 47(3), 293-300. doi: [10.11606/issn.2176-7262.v47i3p293-300](https://doi.org/10.11606/issn.2176-7262.v47i3p293-300).[\[GS Search\]](#)
- Cabral, M. M. W. (2019). A utilização da taxonomia de Bloom no processo de ensino–aprendizado para alunos do ensino superior. *Revista Calafiori*, 3(1), 32-38. [\[GS Search\]](#)
- Costa, N., Júnior, C. P., & Fernandes, M. (2019, November). Recomendação de ações pedagógicas utilizando planejamento automático e taxonomia digital de Bloom. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 30, No. 1, p. 1531). doi: [10.5753/cbie.sbie.2019.1531](https://doi.org/10.5753/cbie.sbie.2019.1531) [\[GS Search\]](#)
- Costa, N. T., & Fernandes, M. A. (2021). Sequenciamento de Ações Pedagógicas baseadas na Taxonomia de Bloom usando Planejamento Automatizado apoiado por Algoritmo Genético. *Revista Brasileira de Informática na Educação*, 29, 485-501. doi: [10.5753/rbie.2021.29.0.485](https://doi.org/10.5753/rbie.2021.29.0.485) [\[GS Search\]](#)
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS inc, 9, 13. [\[GS Search\]](#)

- Chanamarn, N., & Tamee, K. (2017). Enhancing Efficient Study Plan for Student with Machine Learning Techniques. *International Journal of Modern Education & Computer Science*, 9(3).[\[GS Search\]](#)
- Dalcin, E., Abegg, I., & Ceretta, P. S. (2021). O uso de tecnologias digitais no processo de ensino e aprendizagem no moodle: um mapeamento sistemático da literatura. *Revista Tecnologias Educacionais em Rede (ReTER)*, 2(3), 8-01. doi:[10.5753/cbie.sbie.2019.1531](https://doi.org/10.5753/cbie.sbie.2019.1531) [\[GS Search\]](#)
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005. doi:[10.1109/ACCESS.2017.2654247](https://doi.org/10.1109/ACCESS.2017.2654247) [\[GS Search\]](#)
- Fragelli, R. (2019). Método trezentos: Aprendizagem ativa e colaborativa, para além do conteúdo. Penso Editora.[\[GS Search\]](#)
- França, R. S., & do Amaral, H. J. C. (2013). Mineração de dados na identificação de grupos de estudantes com dificuldades de aprendizagem no ensino de programação. *RENOTE*, 11(1).[\[GS Search\]](#)
- Ferreira, Gislaíne Rossetti Madureira. Modelo de combinação socioafetiva: um foco na formação de grupos para um ambiente virtual de aprendizagem. 2021.[\[GS Search\]](#)
- Galhardi, A. C., & Azevedo, M. M. D. (2013, October). Avaliações de aprendizagem: o uso da taxonomia de Bloom. In *Anais do VII Workshop Pós-graduação e Pesquisa do Centro Paula Souza, São Paulo (Vol. 1, No. 1, pp. 237-247)*.[\[GS Search\]](#)
- Harrington, P. (2012). *Machine learning in action*. Simon and Schuster.[\[GS Search\]](#)
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques Third Edition* [M]. *The Morgan Kaufmann Series in Data Management Systems*, 5(4), 83-124.[\[GS Search\]](#)
- Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. Packt PublishingLtd.[\[GS Search\]](#)
- INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2020). *Sinopses Estatísticas da Educação Superior – Graduação*. Brasília: INEP, 2020. Acedido em 10 de outubro de 2020, Disponível em [\[Link\]](#).
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.[\[GS Search\]](#)
- Kuo, C. C., Glover, F., & Dhir, K. S. (1993). Analyzing and modeling the maximum diversity problem by zero-one programming. *Decision Sciences*, 24(6), 1171-1185. doi:[10.1111/j.1540-5915.1993.tb00509.x](https://doi.org/10.1111/j.1540-5915.1993.tb00509.x) [\[GS Search\]](#)
- Kelvin, T., Leandro, F., Fagundes, R., & Freitas, E. (2021, November). Políticas para Adoção de Learning Analytics: Uma Proposta Baseada nas Opiniões dos Estudantes. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação (pp. 885-896)*. SBC. doi:[10.5753/sbie.2021.218599](https://doi.org/10.5753/sbie.2021.218599) [\[GS Search\]](#)
- Lima, R. W. D. (2009). *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*. Natal: UFRN. 106p. Tese de Doutorado.[\[GS Search\]](#)
- Maina, E. M., Oboko, R. O., & Waiganjo, P. W. (2017). Using machine learning techniques to support group formation in an online collaborative learning environment. *International Journal of Intelligent Systems & Applications*, 9(3), 26-33. doi:[10.5815/ijisa](https://doi.org/10.5815/ijisa) [\[GS Search\]](#)
- Macêdo, P. H., Santos, W. B., & Maciel, A. M. (2020). Análise de perfis de engajamento de estudantes de ensino a distância. *RENOTE*, 18(2), 326-335.[\[GS Search\]](#)

- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N. J. A. H., Ramírez-Quintana, M. J., & Flach, P. A. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*.doi: [10.1109/TKDE.2019.2962680](https://doi.org/10.1109/TKDE.2019.2962680).[\[GS Search\]](#)
- Mazur, E. (2015). *Peer Instruction: a revolução da aprendizagem ativa*. Penso Editora.[\[GS Search\]](#)
- Morais, I. R. D., Garcia, T. C. M., Rêgo, M. C. F. D., Zaros, L. G., & Gomes, A. V. G. (2020). Ensino remoto emergencial: orientações básicas para elaboração do plano de aula.[\[GS Search\]](#)
- Monteverde, I., Amaral, G., Ramos, D., do Nascimento, P., Gadelha, B., & Oliveira, E. (2017, October). 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)* (Vol. 28, No. 1, p. 1657).doi: [10.5753/cbie.sbie.2017.1657](https://doi.org/10.5753/cbie.sbie.2017.1657)[\[GS Search\]](#)
- Monteverde, I., Ramos, D., Amaral, G., Gadelha, B., & Oliveira, E. (2018, October). Framework Conceitual para Formação de Grupos de Alunos utilizando Trilhas de Aprendizagem em um Ambiente Virtual de Aprendizagem. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 29, No. 1, p. 1673).DOI: [10.5753/cbie.sbie.2018.1673](https://doi.org/10.5753/cbie.sbie.2018.1673)[\[GS Search\]](#)
- Moran, J. (2015). Mudando a educação com metodologias ativas. *Coleção mídias contemporâneas. Convergências midiáticas, educação e cidadania: aproximações jovens*, 2(1), 15-33.[\[GS Search\]](#)
- Moran, J. M. (2007). *A educação que desejamos: novos desafios e como chegar lá*. Papirus Editora..[\[GS Search\]](#)
- Moubayed, A., Injadat, M., Shami, A., & Lutfiyya, H. (2020). Student engagement level in an e-learning environment: Clustering using k-means. *American Journal of Distance Education*, 34(2), 137-156. doi:[10.1080/08923647.2020.1696140](https://doi.org/10.1080/08923647.2020.1696140)[\[GS Search\]](#)
- Neto, V. S., Feitosa, R. M., Pinheiro, D. N., Lima, M. L., & Labidi, S. (2016, February). Data Mining to Identify Learning Groups with Difficulties in Programming Education. In *The International Conference on Innovations in Intelligent Systems and Computing Technologies (ICIISCT2016)* (p. 1).[\[GS Search\]](#)
- Nalli, G., Amendola, D., Perali, A., & Mostarda, L. (2021). Comparative Analysis of Clustering Algorithms and Moodle Plugin for Creation of Student Heterogeneous Groups in Online University Courses. *Applied Sciences*, 11(13), 5800. DOI: [10.3390/app11135800](https://doi.org/10.3390/app11135800)[\[GS Search\]](#)
- OCDE. Pisa 2018. Paris, 2019. Disponível em [\[Link\]](#). Acesso em: 20 out. 2020.
- Oliveira, P. L. S., Rodrigues, R. L., Ramos, J. L. C., & Silva, J. C. S. (2022). Identificação de Pesquisas e Análise de Algoritmos de Clusterização para a Descoberta de Perfis de Engajamento. *Revista Brasileira de Informática na Educação*, 30, 01-19.doi: [10.5753/rbie.2022.2508](https://doi.org/10.5753/rbie.2022.2508) [\[GS Search\]](#)
- Oliveira, P. L. S., Rodrigues, R. L., Ramos, J. L. C., & Silva, J. C. S. (2020, November). Uma Análise de Algoritmos de Clusterização para Descoberta de Perfis de Engajamento. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação* (pp. 1012-1021). SBC. [10.5753/cbie.sbie.2020.1012](https://doi.org/10.5753/cbie.sbie.2020.1012)[\[GS Search\]](#)

- Oliveira, M., Nogueira, M., & Oliveira, E. (2015, July). Sistema de apoio à prática assistida de programação por execução em massa e análise de programas. In *Anais do XXIII Workshop sobre Educação em Computação* (pp. 90-99). SBC. doi: [10.5753/wei.2015.10225](https://doi.org/10.5753/wei.2015.10225) [GS Search]
- Oliveira, M. G., Neves, A., Lopes, M. F. S., Medeiros, H. F., Andrade, M. B., & Reblin, L. L. (2018). Um curso de programação a distância com metodologias ativas e análise de aprendizagem por métricas de software. *RENOTE*, 15(1). doi: [10.22456/1679-1916.75143](https://doi.org/10.22456/1679-1916.75143) [GS Search]
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k Means Clustering algorithm for prediction of Students Academic Performance. arXiv preprint arXiv:1002.2425. [GS Search]
- Perrenoud, P. (2015). *Dez novas competências para ensinar*. Artmed editora. [GS Search]
- Pereira, P., Cunha, J. L., Sales, G., & Freire, A. (2018, October). 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)*(Vol.29,No.1,p.1413). doi: [10.5753/cbie.sbie.2018.1413](https://doi.org/10.5753/cbie.sbie.2018.1413) [GS Search]
- Pereira, A. J., Gomes, A. S., Primo, T. T., da Silva, R. M. A., Rodrigues, R. L., de Campos Filho, A. S., ... & de Melo Júnior, R. P. (2021, November). 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* (pp. 145-156).SBC. doi:[10.5753/sbie.2021.218498](https://doi.org/10.5753/sbie.2021.218498). [GS Search]
- Pimentel, E. P., de França, V. F., & Omar, N. (2003, November). 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)* (Vol. 1, No. 1, pp. 495-504). [10.5753/CBIE.SBIE.2003.495-504](https://doi.org/10.5753/CBIE.SBIE.2003.495-504) [GS Search]
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. doi:[10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). [GS Search]
- Romero, C., Romero, J. R., & Ventura, S. (2013). A survey on pre-processing educational data. In *Educational data mining* (pp. 29-64). Springer, Cham. doi: [10.1007/978-3-319-02738-8\\_2](https://doi.org/10.1007/978-3-319-02738-8_2) [GS Search]
- Ramos, J. L. C., e Silva, R. E. D., Silva, J. C. S., Rodrigues, R. L., & Gomes, A. S. (2016). A comparative study between clustering methods in educational data mining. *IEEE Latin America Transactions*, 14(8), 3755-3761. doi: [10.1109/TLA.2016.7786360](https://doi.org/10.1109/TLA.2016.7786360) [GS Search]
- Ramos, J. L. C., Santos, L. F. L., Silva, J. C. S., & Rodrigues, R. L. (2020, November). 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* (pp. 932-941). SBC. doi: [10.5753/cbie.sbie.2020.932](https://doi.org/10.5753/cbie.sbie.2020.932) [GS Search]
- Ramos, D. B., Ramos, I. M. M., Gasparini, I., & de Oliveira, E. H. T. (2021). A new learning path model for E-Learning systems. *International Journal of Distance Education Technologies (IJDET)*, 19(2), 34-54. doi: [10.4018/IJDET.20210401.0a2](https://doi.org/10.4018/IJDET.20210401.0a2) [GS Search]
- Ramos, I. M. M., Ramos, D. B., Gadelha, B., & de Oliveira, E. H. T. (2021). An Approach to Group Formation in Collaborative Learning Using Learning Paths in Learning Management Systems. *IEEE Transactions on Learning Technologies*. doi: [10.1109/TLT.2021.3117916](https://doi.org/10.1109/TLT.2021.3117916) [GS Search]

- Silva, L. M., Barbosa, J. L., & Rigo, S. J. (2021). Análise de Dados e Serviços Inteligentes Aplicados na Educação à Distância: um mapeamento sistemático. *Revista Brasileira de Informática na Educação*, 29, 331-357. doi: [10.1109/TLT.2021.3117916](https://doi.org/10.1109/TLT.2021.3117916) [[GS Search](#)]
- Souza, J. P. L., de Souza, D. Y., & Dutra, J. F. (2020). Predição precoce de problemas de desempenho de estudantes em modalidade de educação online: um estudo de caso no ensino médio integrado. doi: [10.18265/1517-0306a2021id5182](https://doi.org/10.18265/1517-0306a2021id5182) [[GS Search](#)]
- Vigotski, L. S. (1998). *A formação social da mente*. Editora Martins Fontes. [[GS Search](#)]