

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA**  
**INSTITUTO DE CIÊNCIAS EXATAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**Gilson dos Reis Dias Fonseca**

**Uma abordagem automática para extração e classificação de exercícios em  
vídeos educacionais**

Juiz de Fora

2024

**Gilson dos Reis Dias Fonseca**

**Uma abordagem automática para extração e classificação de exercícios em vídeos educacionais**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação.

Orientador: Dr. Jairo Francisco de Souza

Coorientador: Dr. Eduardo Barrére

Juiz de Fora

2024

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

D. Fonseca, Gilson dos Reis.

Uma abordagem automática para extração e classificação de exercícios  
em vídeos educacionais / Gilson dos Reis Dias Fonseca. – 2024.

64 f. : il.

Orientador: Jairo Francisco de Souza

Coorientador: Eduardo Barrére

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto  
de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computa-  
ção, 2024.

1. Videoaula. 2. Classificação automática. 3. Exercícios de videoaulas.  
4. Large Language Models. 5. IA Generativa. I. Souza, J. F., orient. II.  
Barrere, E., coorient. III. Título.

**Gilson dos Reis Dias Fonseca**

***Uma abordagem automática para extração e classificação de exercícios em vídeos educacionais***

Dissertação  
apresentada ao  
Programa de Pós-  
graduação em  
Ciência da  
Computação,  
da Universidade  
Federal de Juiz de  
Fora como requisito  
parcial à obtenção do  
título de Mestre em  
Ciência da  
Computação. Área de  
concentração:  
Ciência da  
Computação.

Aprovada em 27 de março de 2024.

**BANCA EXAMINADORA**

**Prof. Dr. Jairo Francisco de Souza** - Orientador

Universidade Federal de Juiz de Fora

**Prof. Dr. Eduardo Barrére** - Coorientador

Universidade Federal de Juiz de Fora

**Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Brugiolo Gonçalves**

Universidade Federal de Juiz de Fora

**Prof. Dr. Fabiano Azevedo Dorça**

Juiz de Fora, 02/05/2024.



Documento assinado eletronicamente por **Jairo Francisco de Souza, Professor(a)**, em 14/05/2024, às 21:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabiano, Usuário Externo**, em 15/05/2024, às 08:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gilson dos Reis Dias Fonseca, Usuário Externo**, em 20/05/2024, às 18:24, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Barrere, Professor(a)**, em 05/06/2024, às 19:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luciana Brugiolo Goncalves, Professor(a)**, em 11/06/2024, às 17:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf ([www2.ufjf.br/SEI](http://www2.ufjf.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **1795903** e o código CRC **86C45B5B**.

Dedico este trabalho à minha mãe e meu pai

## AGRADECIMENTOS

Agradeço, primeiramente aos meus pais, Gilson dos Reis Fonseca e Luiza Helena Dias Fonseca, por tudo o que fizeram por mim. Mesmo com todas as dificuldades sempre me deram apoio e condições para que pudesse focar nos estudos; sem vocês não chegaria a lugar algum.

À todos os meus irmãos e irmãs, por estarem sempre ao meu lado. Na loteria da vida tive sorte cinco vezes com vocês.

À minha namorada, Andressa Rezende Pereira, por todo o apoio, carinho e incentivo durante as dificuldades que tive para continuar com este trabalho.

Aos professores do PPGCC pela contribuição em minha formação. Em especial, aos professores Jairo Francisco de Souza e Eduardo Barrére por toda orientação, ajuda, conselhos, e, principalmente, pela paciência durante todo o período desse trabalho.

Aos meus amigos, com os quais eu compartilho minha jornada de vida e que fazem dela algo mais especial, em especial, Bernardete Aquino, Bruno Jucá, Célio, Lucas e Ludmila Larcher e João Paulo Radd, que acompanharam o processo de desenvolvimento dessa dissertação.

À banca avaliadora, professores Fabiano Azevedo Dorça, Luciana Brugiolo Gonçalves, Marcia Aparecida Fernandes e Pedro Henrique Dias Valle, pela disponibilidade e pertinentes apontamentos.

“As flores brotam e morrem, as estrelas brilham, mas um dia se apagarão. Tudo morre, a Terra o Sol, a Via Láctea e até mesmo todo esse universo não é exceção... Comparado a isso, a vida do homem é tão breve e fugidia quanto o piscar de olhos. Nesse curto instante, os homens nascem, riem, choram, lutam, sofrem, festejam, lamentam, odeiam pessoas e amam outras, tudo é transitório, e em seguida todos caem no sono eterno chamado morte”.

Shaka



## RESUMO

Os vídeos são amplamente utilizados na educação por causa de seus diversos benefícios, como a compreensão de conceitos complexos e o estímulo ao engajamento dos alunos. Sistemas de recomendação podem ser utilizados para organizar e recuperar o conteúdo adequado, a partir de metadados. Os exercícios de uma videoaula têm potencial para fornecer metadados úteis, porém, ao contrário de fontes textuais, não há marcadores claros que os identifiquem, já que o conteúdo é apresentado em fala contínua. Além disso, os exercícios, podem estar mesclados em meio a dicas e explicações fornecidas pelo professor durante a fala, o que torna a identificação ainda mais complexa. Com essa premissa, esse trabalho propõe um modelo de classificação automática de exercícios em videoaulas com base na Taxonomia de Bloom Digital, buscando proporcionar uma base sólida para sistemas de recomendação personalizados e ampliar as possibilidades de busca em repositórios educacionais. A avaliação do modelo proposto foi realizada utilizando algoritmos classificadores, onde o modelo BERT se destacou como superior aos demais.

Palavras-chave: videoaula. Metadados. Extração de metadados. Classificação automática. Classificação automática de exercícios. Identificação de exercícios. Machine learning. Inteligência artificial generativa. Large Language Models

## ABSTRACT

Videos are widely used in education due to their various benefits, such as facilitating the understanding of complex concepts and promoting student engagement. Recommendation systems can be used to organize and retrieve suitable content based on metadata. Exercises in a video lesson have the potential to provide useful metadata. However, unlike textual sources, there are no clear markers to identify them, as the content is presented in continuous speech by the instructor. Additionally, exercises may be interspersed with tips and explanations provided by the instructor during the speech, which makes identification even more complex. With this premise, this work proposes an automatic classification model for exercises in video lessons based on Digital Bloom Taxonomy, aiming to provide a solid foundation for personalized recommendation systems and expand search possibilities in educational repositories. The evaluation of the proposed model was conducted using classification algorithms, where the BERT model stood out as superior to others.

Keywords: Video lesson. Metadata. Metadata extraction. Automatic classification. Automatic exercise classification. Exercise identification. Machine learning. Generative artificial intelligence. Large Language Models

## LISTA DE ILUSTRAÇÕES

|  |    |
|--|----|
| Figura 1 – Fluxograma da solução . . . . .   | 31 |
| Figura 6 – Modelos Disponíveis . . . . .   | 32 |
| Figura 3 – Arquitetura do Workflow de Classificação Automática de Exercícios . . . | 34 |
| Figura 4 – Tela inicial do Workflow . . . . .                                      | 35 |
| Figura 5 – Tela final do Workflow . . . . .  | 35 |
| Figura 7 – Matrizes de Confusão . . . . .  | 44 |
| Figura 8 – Matriz de Confusão BEA . . . . .  | 45 |
| Figura 2 – Arquitetura do framework . . . . .                                      | 63 |

## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 1 – Matriz de termos de ação que estão em mais de um nível . . . . .   | 37 |
| Tabela 2 – Distribuição dos estilos de videoaulas por área de conhecimento, onde TH, VO, A, HD e W representam os estilos <i>Talking Head</i> , <i>Voice Over Slides</i> , <i>Animation</i> , <i>Hand Drawn</i> e <i>Whiteboard</i> , respectivamente . . . . . | 39 |
| Tabela 3 – Relacional entre o processo, base de dados e ferramenta utilizados . . .   | 41 |
| Tabela 4 – Exercícios válidos utilizados para teste . . . . .   | 42 |
| Tabela 5 – Exercícios inválidos utilizados para teste . . . . .   | 43 |
| Tabela 6 – Testes da abordagem <i>few-shots</i> . . . . .   | 43 |
| Tabela 7 – Resultado do treinamento com a Base Bloom . . . . .  | 48 |
| Tabela 8 – Tabela de resultado da classificação BEM . . . . .   | 49 |
| Tabela 9 – Tabela de resultado da classificação BEA . . . . .   | 50 |

## LISTA DE ABREVIATURAS E SIGLAS

|        |   |
|--------|---|
| ASR    | <i>Automatic Speech Recognition</i> (Reconhecimento Automático de Fala)   |
| BAVi   | Busca Avançada de Vídeos  |
| BEA    | Base de Exercícios Automática   |
| BEM    | Base de Exercícios Manual   |
| BERT   | <i>Bidirectional Encoder Representations from Transformers</i>  |
| BR     | Baseado em Regras   |
| BVA    | Base de Videoaulas  |
| DC     | <i>Dublin Core</i>  |
| DPC    | Dimensão do Processo Cognitivo  |
| ESV    | Enriquecimento Semântico de Vídeos  |
| ESOA   | Enriquecimento Semântico de Objetos de Aprendizagem   |
| FT     | <i>Fine Tuning</i>  |
| GPT    | <i>Generative Pre-trained Transformer</i> (Transformador Generativo Pré-treinado)   |
| HOTS   | <i>Higher Order Thinking Skills</i> (Habilidades de Pensamento de Ordem Superior)   |
| IEEE   | <i>Institute of Electrical and Electronics Engineers</i> (Instituto de Engenheiros Eletricistas e Eletrônicos)              |
| LAPIC  | Laboratório de Aplicações Inovadoras em Computação  |
| LLM    | <i>Large Language Model</i> (Modelo de Linguagem Grande)  |
| LMS    | <i>Learning Management System</i> (Sistema de Gerenciamento de Aprendizado)   |
| LOM    | <i>Learning Object Metadata</i> (Metadados de Objetos de Aprendizagem)  |
| LOTS   | <i>Lower Order Thinking Skills</i> (Habilidades de Pensamento de Ordem Inferior)  |
| M2P    | <i>Machine to Person</i> (Máquina para pessoa)  |
| MCC    | <i>Matthews Correlation Coefficient</i> (Coeficiente de Correlação de Matthews)   |
| MOODLE | <i>Modular Object-Oriented Dynamic Learning Environment</i> (Ambiente de Aprendizagem Dinâmico Modular Orientado a Objetos) |
| NB     | <i>Naive Bayes</i>  |
| OA     | Objeto de Aprendizagem  |
| OCR    | <i>Optical Character Recognition</i> (Reconhecimento Óptico de Caracteres)  |
| OVR    | <i>Open Video Recommendation</i> (Recomendação de Vídeos Públicos)  |
| SVM    | <i>Support Vector Machine</i> (Máquina de Vetores de Suporte)   |
| TBD    | Taxonomia de Bloom Digital  |
| TBR    | Taxonomia de Bloom Revisada   |
| TIC    | Tecnologia da Informação e Comunicação  |
| WER    | <i>Word Error Rate</i> (Taxa de Erro de Palavra)  |

## SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b> . . . . .  | <b>14</b> |
| 1.1      | HIPÓTESES . . . . .  | 16        |
| 1.2      | OBJETIVO . . . . .   | 17        |
| 1.3      | CONTRIBUIÇÕES . . . . .  | 17        |
| 1.4      | VISÃO GERAL DOS CAPÍTULOS . . . . .  | 17        |
| <b>2</b> | <b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .   | <b>19</b> |
| 2.1      | APRENDIZADO DE MÁQUINA . . . . .   | 19        |
| 2.2      | MODELOS DE CLASSIFICAÇÃO E REGRESSÃO . . . . .   | 19        |
| 2.3      | ALGORITMOS CLÁSSICOS DE CLASSIFICAÇÃO . . . . .  | 20        |
| 2.4      | PROCESSAMENTO DE LINGUAGEM NATURAL . . . . .   | 21        |
| 2.5      | APRENDIZADO PROFUNDO . . . . .   | 21        |
| 2.6      | GRANDES MODELOS DE LINGUAGEM . . . . .   | 21        |
| 2.7      | TAXONOMIA DE BLOOM REVISADA (TBR) . . . . .  | 22        |
| 2.8      | TAXONOMIA DE BLOOM DIGITAL (TBD) . . . . .   | 24        |
| 2.9      | ENRIQUECIMENTO SEMÂNTICO DE OBJETOS DE APRENDIZAGEM<br>(ESOA) . . . . .                    | 25        |
| <b>3</b> | <b>TRABALHOS RELACIONADOS</b> . . . . .  | <b>27</b> |
| 3.1      | Extração de metadados . . . . .  | 27        |
| 3.2      | Classificação automática de exercícios . . . . .   | 28        |
| 3.3      | Relevância desse estudo . . . . .  | 29        |
| <b>4</b> | <b>UMA PROPOSTA DE CLASSIFICAÇÃO AUTOMÁTICA DE<br/>EXERCÍCIOS EM VÍDEO AULAS</b> . . . . . | <b>30</b> |
| 4.1      | ARQUITETURA . . . . .  | 31        |
| 4.1.1    | Módulo de Transcrição . . . . .  | 31        |
| 4.1.2    | Módulo de Identificação dos exercícios . . . . .   | 33        |
| 4.1.3    | Módulo de Classificação dos exercícios . . . . .   | 33        |
| 4.1.4    | <i>Workflow</i> . . . . .  | 33        |
| <b>5</b> | <b>EXPERIMENTOS</b> . . . . .  | <b>36</b> |
| 5.1      | BASE DE DADOS . . . . .  | 36        |
| 5.1.1    | <i>Base Bloom</i> . . . . .  | 36        |
| 5.1.2    | <i>Base de Videoaulas (BVA)</i> . . . . .  | 38        |
| 5.1.3    | <i>Base de exercícios manual (BEM)</i> . . . . .   | 38        |
| 5.1.4    | <i>Base de exercícios automática (BEA)</i> . . . . .                                       | 40        |
| 5.2      | ANÁLISE DOS RESULTADOS . . . . .   | 40        |
| 5.2.1    | Identificação dos exercícios . . . . .   | 41        |
| 5.2.1.1  | <i>1º experimento</i> . . . . .  | 41        |
| 5.2.1.2  | <i>2º experimento</i> . . . . .  | 43        |

|       |   |           |
|-------|---|-----------|
| 5.2.2 | Classificação dos exercícios . . . . .      | 44        |
| 6     | <b>CONCLUSÃO . . . . .</b>                  | <b>51</b> |
| 6.1   | CONTRIBUIÇÕES . . . . .                     | 52        |
| 6.2   | LIMITAÇÕES . . . . .                        | 53        |
| 6.3   | TRABALHOS FUTUROS . . . . .                 | 53        |
|       | <b>REFERÊNCIAS . . . . .</b>                | <b>55</b> |
|       | <b>APÊNDICE A – Framework M2P . . . . .</b> | <b>63</b> |

## 1 INTRODUÇÃO

A aprendizagem é um processo natural e constante na vida das pessoas (40), e o ensino é um esforço realizado por pessoas em benefício de outras pessoas (18). Em (94), é dito que a motivação dos alunos é um fator importante para o sucesso do ensino, uma vez que alunos motivados têm maior probabilidade de se engajar nas atividades de aprendizagem e obter melhores resultados.

Quando a aprendizagem é intencional e ocorre em um contexto institucional, como na escola ou na universidade, é comum haver sequências estruturadas de instruções. Essas sequências são projetadas para apoiar, facilitar ou melhorar a aprendizagem e o desempenho dos alunos (40), com metas e objetivos explícitos. Essas sequências podem incluir aulas, atividades, tarefas e avaliações que são cuidadosamente planejadas e projetadas para ajudar os alunos a atingir seus objetivos de aprendizagem (2).

Conforme é dito em (71), existem diversas teorias que guiam os estudos sobre aprendizagem, as quais propõem uma abordagem sistemática para interpretar, organizar e prever como ocorre o processo de aprendizagem. Dentre as teorias que se destacam, estão o comportamentalismo, o cognitivismo e o construtivismo.

O comportamentalismo explica a aprendizagem a partir da observação (61). O processo de ensino e aprendizagem nessa abordagem ocorre por meio de reforços, recompensas e treinamento contínuo até o sucesso consistente (71). No construtivismo o aprendizado é uma construção individual e social, baseado em experiências anteriores e em interações com o meio (71). Por fim, a abordagem cognitivista surgiu como uma alternativa ao comportamentalismo na psicologia, quando este mostrou-se insuficiente para explicar a complexidade do aprendizado humano, como no caso do aprendizado de línguas (40).

Na abordagem cognitiva, as pessoas não são vistas mais como meras coleções de respostas a estímulos externos, como defendido pelos comportamentalistas, mas como processadores de informação (40). A aprendizagem é construída a partir da interação do indivíduo com informações previamente adquiridas, que são representadas em novos estágios e níveis de profundidade, adequando-se à capacidade de compreensão do indivíduo e tornando-se gradativamente mais complexas à medida que o aluno constrói novas concepções e modelos (71). Essa abordagem tem sido usada como base para alguns dos tutores inteligentes mais bem-sucedidos, nos quais os processos mentais são primeiro identificados e o conhecimento é transferido para os alunos da maneira mais eficiente e eficaz possível (94).

No século 21, o processo educacional tem sido substancialmente aprimorado pela integração de uma variedade de ferramentas de Tecnologia da Informação e Comunicação (TIC). Segundo (81), o uso dessas ferramentas tem evidenciado um aumento significativo na eficácia tanto dos educadores quanto dos alunos durante o processo de ensino e



aprendizagem. Essas ferramentas abrangem uma ampla gama de recursos, incluindo computadores, internet, sistemas de gerenciamento de aprendizado (LMS) e, especialmente, multimídia instrucional. A presença da multimídia instrucional tem se destacado como um elemento fundamental no cenário educacional contemporâneo, proporcionando uma experiência de aprendizado rica e dinâmica.

A multimídia instrucional oferece uma série de benefícios no contexto educacional. Como destacado em (1), em primeiro lugar, ela possui a capacidade de transformar conceitos abstratos em conteúdos concretos, facilitando a compreensão por parte dos alunos. Além disso, a multimídia permite a apresentação de grandes volumes de informações em um tempo limitado e com menos esforço, tornando o processo de ensino mais eficiente. Outro benefício importante é o estímulo ao interesse dos alunos pela aprendizagem, pois a variedade de formatos e recursos visuais tende a tornar as aulas mais atrativas e envolventes. Por fim, a multimídia também proporciona aos professores a capacidade de acompanhar o progresso dos alunos no aprendizado, permitindo-lhes identificar com mais precisão as necessidades individuais e adaptar suas abordagens de ensino conforme necessário.

Nos últimos anos, tem havido avanços significativos na tecnologia, facilitando o compartilhamento de conteúdos multimídia instrucionais. A multimídia instrucional abrange uma variedade de formatos, incluindo gráficos em livros didáticos, apresentações em PowerPoint com áudio, narrações em áudio ou vídeo, animações e videoaulas (80). Dentre essas áreas de multimídia, a criação de videoaulas é a que mais cresceu (43), principalmente devido à facilidade de produção e disponibilização para o público.

Os vídeos são amplamente utilizados na educação por causa de seus diversos benefícios, incluindo a ajuda na compreensão de conceitos complexos, fornecimento de exemplos práticos, estímulo da curiosidade e engajamento dos alunos e possibilidade de aprendizado em ritmo próprio. Além disso, os vídeos podem ser úteis para gerar discussões em sala de aula, oferecendo diferentes perspectivas sobre um assunto e complementando os tópicos já abordados em aula (28). No entanto, a grande quantidade de vídeos disponíveis pode tornar difícil para os alunos encontrar aqueles adequados para o seu processo de aprendizagem (17). Para solucionar esse problema, sistemas de recomendação podem ser implementados, mas é necessário identificar o tema abordado nos vídeos para recomendá-los. Essa tarefa pode ser complicada, já que as descrições textuais que acompanham os vídeos geralmente são escassas e genéricas (97) e os mecanismos de busca ainda dependem muito de informações textuais para descrever a mídia (25).

Assim, uma estratégia promissora para abordar esse desafio é o enriquecimento semântico de áudio e vídeo, buscando aprimorar a compreensão e a assimilação do conteúdo audiovisual por meio da inclusão de elementos semânticos no material original. Várias abordagens descritas na literatura exploram essa ideia, baseadas em representações semânticas explícitas (17). Essas abordagens podem incluir a adição de legendas, transcrições

e outros recursos que ofereçam informações complementares sobre o conteúdo do vídeo, como definições de termos específicos, traduções para outros idiomas ou links para recursos relacionados.

O LAPIC (Laboratório de Aplicações Inovadoras em Computação) da UFJF também tem trabalhado na geração de metadados para enriquecimento semântico. Através de abordagens construídas em módulo e baseadas em representações semânticas implícitas, como a segmentação de videoaulas utilizando a fala do professor para identificar mudanças de tópicos (88) e a forma de apresentação para identificação de tipos distintos de videoaula (7). Para gerenciamento dessas abordagens é utilizado o *framework* M2P (*Machine to Person*) que controla a fila de execução dos módulos, permitindo uma integração eficiente e coordenada das diferentes etapas do processo (Apêndice A).

Uma outra abordagem implícita que possui potencial para fornecer metadados úteis, como informações sobre tipo, dificuldade e tópicos é a identificação e classificação de exercícios. Isso pode fornecer uma base robusta para sistemas que buscam fornecer recomendações personalizadas para o processo de aprendizagem de cada aluno. Além disso, a classificação desses exercícios abre uma nova dimensão de busca em repositórios de recursos educacionais, além de melhorar a experiência do usuário e aumentar a eficiência na aprendizagem.

No entanto, identificar exercícios em videoaulas apresenta desafios distintos em comparação com mídias textuais, como documentos PDF ou livros. Enquanto em mídias textuais os exercícios podem ser facilmente destacados e localizados por meio de índices ou formatação específica, em videoaulas essa distinção é menos clara. A falta de marcadores visuais claros e a apresentação contínua do conteúdo dificultam a identificação dos exercícios, que muitas vezes estão integrados às explicações verbais do professor. Além disso, a forma como os exercícios são apresentados em videoaulas pode variar consideravelmente. Eles podem aparecer como perguntas diretas, problemas a serem resolvidos ou desafios propostos pelo instrutor. Essa variedade de formatos torna ainda mais desafiador distinguir os exercícios do restante do conteúdo.

Com base nesse contexto, as seguintes questões de pesquisa podem ser formuladas:

**Q1.** É possível a identificação dos exercícios em videoaulas, mesmo com a falta de marcadores visuais claros?

**Q2.** É possível determinar a dificuldade de exercícios em videoaulas?

## 1.1 HIPÓTESES

Com base na análise do problema apresentado, foram formuladas as seguintes hipóteses que buscam fornecer respostas às questões levantadas:

**H1.** A utilização de ferramentas de transcrição automática de áudio pode auxiliar

na identificação dos exercícios.

**H2.** O uso de modelos de Processamento de Linguagem Natural (PLN) personalizados baseados em LLM, treinados com exemplos específicos de videoaulas em diferentes disciplinas e estilos de instrução, pode permitir a identificação mais precisa dos exercícios apresentados, mesmo em situações onde os marcadores visuais são escassos.

**H3.** A aplicação de técnicas de Aprendizado de máquina supervisionado, pode permitir a construção de modelos preditivos capazes de determinar a dificuldade de novos exercícios. O uso de uma escala discreta, como a Taxonomia de Bloom, pode ser adequado para este problema.

## 1.2 OBJETIVO

Com base nas hipóteses apresentadas, o objetivo dessa dissertação de mestrado é:

Propor uma abordagem, para identificar e classificar automaticamente exercícios em videoaulas, utilizando abordagens fundamentadas em LLM, com base na Taxonomia de Bloom Digital. Atingindo esse objetivo é possível que sistemas de recuperação da informação possam ser aprimorados, facilitando o acesso e organização de conteúdos educacionais digitais, além de fornecer uma base de metadados relacionados à exercícios.

## 1.3 CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

1. A proposta de um modelo de classificação automática de exercícios em videoaulas fundamentado na Taxonomia de Bloom Digital que visa a geração de metadados relevantes para as videoaulas.
2. Modelos de identificação e classificação de exercícios baseados em abordagens fundamentadas em LLM.
3. A integração do modelo proposto com o *framework* M2P desenvolvido pelo LAPIC, contribuindo para a continuação da evolução da ferramenta.
4. Um novo *dataset* com um conjunto de vídeos do mundo real do domínio da educação de diversas áreas de conhecimento contendo exercícios.
5. Um novo *dataset* composto por exercícios do mundo real no domínio da educação de diversas áreas de conhecimento, bem como sua classificação manual e automática.

## 1.4 VISÃO GERAL DOS CAPÍTULOS

Esta dissertação está estruturada em 6 capítulos:

- **Capítulo 2:** Este capítulo tem como objetivo explorar os conceitos fundamentais que estão intimamente ligados à essência da proposta delineada neste trabalho. Além disso, serão apresentados trabalhos relacionados que contribuem para o embasamento teórico e contextual da proposta.
- **Capítulo 4:** Neste capítulo, é apresentado o modelo proposto, desenvolvido por meio de módulos construídos dentro do *framework M2P*. Serão detalhados os componentes do modelo, suas funcionalidades e como eles se inter-relacionam.
- **Capítulo 5:** Neste capítulo, serão abordados os experimentos realizados, bem como os resultados obtidos a partir da identificação e classificação dos exercícios. Serão apresentados detalhes sobre o modelo adotado, os conjuntos de dados utilizados, as métricas de avaliação empregadas e uma análise dos resultados alcançados.
- **Capítulo 6:** Por fim, neste capítulo é realizada a conclusão do trabalho, destacando as principais contribuições e limitações do estudo. Além disso, são discutidas as perspectivas para trabalhos futuros, visando a continuidade e o aprimoramento da pesquisa.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são explorados os conceitos fundamentais relacionados à proposta deste trabalho. Serão discutidas considerações sobre a Taxonomia de Bloom Revisada (TBR), as atividades digitais propostas pela Taxonomia de Bloom Digital (TBD), aprendizado de máquina, problemas de classificação e regressão, apresentação dos algoritmos clássicos de classificação, aprendizado profundo e o enriquecimento semântico de objetos de aprendizagem (ESOA). Essa análise proporciona uma base teórica sólida para fundamentar a proposta apresentada neste estudo. Além disso, incorpora uma análise dos trabalhos relacionados.

### 2.1 APRENDIZADO DE MÁQUINA

O aprendizado de máquina, ou *Machine Learning*, é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem a partir de dados, sem a necessidade de programação explícita para cada tarefa específica. Em vez disso, os algoritmos de aprendizado de máquina podem detectar padrões nos dados e fazer previsões ou tomar decisões com base nesses padrões (56).

Dentre as abordagens existentes, se encontram o aprendizado não supervisionado, no qual os algoritmos são treinados em um conjunto de dados que contém apenas as entradas, sem saídas correspondentes, visando encontrar padrões ou estruturas nos dados (16). Há também o aprendizado por reforço, no qual os agentes de aprendizado interagem com um ambiente dinâmico, aprendendo a tomar ações para maximizar uma recompensa cumulativa, comum em jogos, robótica e transportes (15).

Por último, há o aprendizado supervisionado, o qual é foco deste trabalho, que é uma técnica de aprendizado de máquina em que um modelo é treinado com um conjunto de dados rotulados, contendo pares de entrada e saída esperada. Durante o treinamento, o modelo aprende a associar as entradas às saídas correspondentes. Este tipo de aprendizado é aplicável em diversos domínios, como reconhecimento de padrões, processamento de linguagem natural e diagnóstico médico. O objetivo principal é que o modelo possa fazer previsões precisas para novos conjuntos de dados de entrada, não vistos durante o treinamento para problemas de classificação e regressão (39).

### 2.2 MODELOS DE CLASSIFICAÇÃO E REGRESSÃO

Os modelos de aprendizado supervisionado se dividem entre modelos de classificação e regressão (64). Ambos envolvem a atribuição de itens de um conjunto de dados a classes ou grupos, utilizando um modelo, com o objetivo de prever com precisão o alvo para cada

caso nos dados. Porém, um modelo preditivo com um alvo numérico utiliza um algoritmo de regressão (46, 64). Enquanto, um modelo preditivo para prever classes ou categorias é considerado um algoritmo de classificação. A classificação é subdividida em classificação binária ou multiclasse.

Na classificação binária, as instâncias de dados são divididas em duas classes distintas: verdadeiro ou falso (83). Cada instância é atribuída a uma das duas classes possíveis, cabendo ao modelo separá-las de forma eficaz. No entanto, em situações onde a classificação binária não é suficiente, como por exemplo, para determinar a probabilidade de um cliente não honrar com o empréstimo, com opções como pouco provável, provável e muito provável, é necessária a utilização da classificação multiclasse (16, 62).

### 2.3 ALGORITMOS CLÁSSICOS DE CLASSIFICAÇÃO

Os algoritmos clássicos de classificação oferecem abordagens sólidas e bem estabelecidas para resolver uma ampla gama de problemas de classificação. Como exemplo, tem-se a regressão logística, que é um método popular para problemas de classificação binária e multinomial. Baseia-se na utilização da função logística para modelar a probabilidade de uma observação pertencer a uma determinada classe. Este algoritmo é frequentemente utilizado em áreas como medicina, finanças e marketing devido à sua simplicidade e interpretabilidade (16).

As árvores de decisão são estruturas de árvore que dividem o espaço de características em regiões retangulares e atribuem uma classe a cada região. Elas são construídas de forma recursiva, dividindo o conjunto de dados com base em critérios de divisão, como ganho de informação ou índice *Gini*. As árvores de decisão são amplamente utilizadas em uma variedade de aplicações, incluindo diagnósticos médicos, detecção de fraudes e sistemas de recomendação (39).

As Máquinas de Vetores de Suporte ou *Support Vectors Machines* (SVMs) são algoritmos de aprendizado de máquina que encontram o hiperplano de separação ótimo entre classes em um espaço de características de alta dimensionalidade. Elas buscam maximizar a margem entre as classes, transformando o problema de otimização em um problema de programação quadrática. As SVMs têm uma ampla gama de aplicações, incluindo classificação de texto, reconhecimento de padrões em imagens e bioinformática (16).

Por fim, o algoritmo de *Naive Bayes*, baseado no teorema de *Bayes* e na suposição *naive* de independência condicional entre as características (34) é conhecido por sua eficiência computacional e desempenho robusto em conjuntos de dados de alta dimensionalidade (59). Ele tem sido aplicado com sucesso em diversas tarefas, como classificação de texto, filtragem de spam e diagnósticos médicos (5, 26, 29, 99).

## 2.4 PROCESSAMENTO DE LINGUAGEM NATURAL

Processamento de Linguagem Natural (PLN) busca entender e manipular a linguagem natural por meio de computadores (23). Fazendo uso de estruturas de comunicação e padrões humanos, como as análises léxica, sintática, semântica e pragmática, para melhor entendimento do contexto das informações recebidas (92), é possível realizar tarefas como transcrição e tradução de texto, detecção de *spams* em *email*, sumarização, identificação de *fake news*, sentimentos, criação de sistemas de interação com usuário, assistentes virtuais entre outros (47). Recentes avanços em aprendizado de máquinas, especialmente em aprendizado profundo, permitiram um grande avanço em PLN nos últimos anos e resoluções de tarefas em redes convolucionais e recorrentes.

## 2.5 APRENDIZADO PROFUNDO

Aprendizado Profundo ou *Deep Learning*, é uma subárea da inteligência artificial que tem ganhado destaque nos últimos anos devido à sua capacidade de aprender representações complexas dos dados. Segundo (52), o sucesso do *Deep Learning* se deve em grande parte ao uso de redes neurais profundas, que são capazes de aprender de forma automática a partir dos dados. Essas redes são compostas por múltiplas camadas de neurônios artificiais, permitindo a extração de características hierárquicas dos dados ((37)

Uma das principais aplicações do *Deep Learning* é no campo do processamento de imagens, onde algoritmos como as redes convolucionais têm alcançado resultados impressionantes em tarefas como reconhecimento de objetos e segmentação de imagens (50). Além disso, o *Deep Learning* tem sido amplamente utilizado em áreas como reconhecimento de voz, na área da saúde e processamento de linguagem natural (32).

## 2.6 GRANDES MODELOS DE LINGUAGEM

Grandes Modelos de Linguagem ou *Large Language Models* (LLMs), são modelos de aprendizado de máquina que são treinados para entender a estrutura da linguagem humana. Eles são capazes de gerar texto que é gramaticalmente correto e faz sentido em um contexto específico. O conceito de LLM foi introduzido em (45), onde foi proposto um modelo estatístico para entender a estrutura da linguagem humana.

Os LLMs podem ser usados em sistemas de recomendação para entender as preferências do usuário e fornecer recomendações personalizadas (100). Além disso, os LLMs também são aplicados em visão computacional para tarefas como reconhecimento de imagem e detecção de objetos (52).

No campo do PLN, os LLMs têm sido a força motriz por trás de muitos avanços recentes. Modelos como BERT (30), BART (53), XLNet (98), RoBERTa (55) e GPT (76) têm demonstrado bom desempenho em uma variedade de tarefas.

O BERT (30) é um modelo que utiliza o ELMo(70) (pré-treinamento como recurso adicional) e o GPT(76) (*fine-tuning*) em sua arquitetura. Tentando remover as restrições que seus predecessores possuem por serem unidirecionais, o BERT é um modelo bidirecional dividido em dois passos. No primeiro passo de pré-treinamento o modelo é treinado por diferentes tarefas usando dados sem *label*. Posteriormente, no segundo passo, é feito o *fine-tuning* onde, após ter sido inicializado com o pré-treinamento, o modelo é exposto a uma base com *labels* em tarefas supervisionadas. Cada tarefa supervisionada, embora tenha o mesmo pré-treino, possui *fine-tuning* distintos. Sua arquitetura é baseada em transformadores, onde, ao receber uma ou mais sentenças, as transforma em uma sequência de *token*. O *token* é criado utilizando *WordPiece embeddings* com um vocabulário de *tokens* de 30.000 palavras. Cada *token* inicia com [CLS] (*token* classificação especial) e termina com [SEP], que é um *token* especial para identificar quando o *token* referente a uma sentença finalizou. Isso é necessário, pois sentenças que são pares, serão postas como uma só sentença, dessa forma é possível identificar qual *token* vem de uma sentença ou outra.

A estratégia de aprendizado do BERT envolve a substituição de 15% das palavras em uma sentença. Dessas palavras, 80% são substituídas pelo *token*, 10% são substituídas por uma palavra aleatória e os outros 10% permanecem inalterados. Assim, o modelo tenta prever o valor das palavras substituídas.

O BART (53) possui uma arquitetura similar ao BERT. Como diferenças entre os modelos, é observado que em cada camada no *decoder* é feita, adicionalmente, uma *cross-attention* na última camada escondida do *encoder*. A segunda diferença entre os modelos é o fato de que, pelo fato do BART possuir 10% mais parâmetros que o BERT, ele performa uma *feed-forward network* a menos.

XLNet (98), possui uma formulação auto regressiva para o pré-treino, o que remove uma limitação vista no BERT, onde o mesmo consegue identificar a dependência entre (x = York, U = New) mas não em (x = New, U = York). Outra diferença é a maximização da vizinhança esperada sobre as permutações na ordem de fatorização, sendo assim possível o aprendizado em contextos bidimensionais.

O modelo RoBERTa (55) propõe um BERT mais robusto. Para isso seu treinamento é feito com mascaramento dinâmico, sentenças completas sem perdas de informações. Por fim, o *Generative Pretrained Transformer* (GPT) é um modelo de linguagem que utiliza a arquitetura *Transformer* para geração de textos no contexto em que é usado (76).

## 2.7 TAXONOMIA DE BLOOM REVISADA (TBR)

Uma das preocupações de educadores na década de 40 era a falta de um sistema padronizado para classificar e descrever os objetivos de aprendizado. Outra preocupação era que muitos educadores estavam se concentrando em objetivos de aprendizado de ordem



inferior, como a memorização de fatos e conceitos, em detrimento de habilidades cognitivas mais complexas, como análise, síntese e avaliação. Além disso, havia uma necessidade crescente de avaliar a aprendizagem dos alunos de maneira mais objetiva e sistemática, especialmente com o aumento do número de alunos nas escolas e a crescente demanda por prestação de contas na educação (41) (67).

Em resposta a esses desafios, em 1948, um grupo de educadores assumiu a tarefa de classificar as metas e objetivos da educação. A intenção foi desenvolver um sistema de classificação para três domínios: o cognitivo, o afetivo e o psicomotor. O trabalho no domínio cognitivo foi concluído em 1956 e é comumente referido como Taxonomia do Domínio Cognitivo de Bloom (67).

Os níveis originais da Taxonomia de Bloom foram ordenados em Conhecimento, Compreensão, Aplicação, Análise, Síntese, Avaliação. A ideia principal da taxonomia é organizar o aprendizado em uma hierarquia do menos ao mais complexo. Os níveis são entendidos como sucessivos, de modo que um nível deve ser dominado antes que o próximo nível possa ser alcançado (41) (67).

Anderson e Krathwohl (4) revisaram a taxonomia de Bloom para se adequar aos objetivos da educação moderna mais focados em resultados, incluindo mudar os nomes dos níveis de substantivos para verbos ativos e inverter a ordem dos dois níveis mais altos (49). Com as mudanças na sociedade nos últimos 50 anos a Taxonomia de Bloom Revisada, fornece uma ferramenta ainda mais poderosa que se encaixa nas necessidades dos professores de hoje, de maneira clara e concisa.

A TBR é uma estrutura hierárquica composta por seis níveis que descrevem diferentes habilidades cognitivas que os alunos podem adquirir. Cada nível é caracterizado por um verbo ativo que indica a habilidade cognitiva necessária para alcançar tal nível (2). Além disso, (4) propuseram uma segunda dimensão para seu modelo de classificação, permitindo a separação dos aspectos dos verbos e dos substantivos em dimensões distintas. Nessa nova abordagem, o substantivo serve como base para a Dimensão do Conhecimento e o verbo para a Dimensão do Processo Cognitivo (DPC). A combinação dessas modificações resultou na criação da Dimensão do Conhecimento, composta pelas categorias Factual, Conceitual, Procedural e Metacognitivo, e da DPC, composta pelas categorias Lembrar, Compreender, Aplicar, Analisar, Avaliar e Criar. As categorias da Dimensão do Conhecimento definem quatro tipos principais de conhecimento (4), que são:

- Factual: conhecimento dos elementos fundamentais necessários para compreender uma disciplina ou resolver problemas relacionados a ela;
- Conceitual: compreensão das conexões entre os elementos fundamentais dentro de uma estrutura que possibilita sua integração e funcionamento conjunto;

- Procedimental: conhecimento de como realizar algo, incluindo métodos de investigação, critérios para o uso de habilidades, algoritmos, técnicas e métodos;
- Metacognitivo: conhecimento sobre a cognição em geral, bem como consciência e conhecimento da própria cognição.

Krathwohl (49) e Ferraz e Belhot (33) apresentam as principais categorias da DPC e seus verbos associados:

- Lembrar: recuperar informações previamente aprendidas (exemplos de verbos: listar, identificar, nomear, recordar);
- Compreender: entender o significado das informações (exemplos de verbos: reescrever, resumir, explicar);
- Aplicar: usar informações de uma nova maneira (exemplos de verbos: aplicar, implementar, demonstrar);
- Analisar: quebrar informações em partes para entender as relações entre elas (exemplos de verbos: comparar, organizar, distinguir);
- Avaliar: julgar o valor ou a qualidade da informação (exemplos de verbos: julgar, avaliar, criticar);
- Criar: usar informações para criar algo (exemplos de verbos: projetar, criar, inventar).

## 2.8 TAXONOMIA DE BLOOM DIGITAL (TBD)

A TBD (24) representa uma atualização da TBR, com o objetivo de abordar os novos comportamentos, ações e oportunidades de aprendizado decorrentes do avanço tecnológico e da onipresença das TICs (Tecnologia de Informação e Comunicação). Enquanto a TBR deixou uma lacuna ao não considerar os novos objetivos, processos e ações relacionados à emergência e integração das TICs na sala de aula e seu impacto na vida dos alunos, Churches (24) desenvolveu a TBD a partir dos termos-chave de cada categoria da DPC (Dimensão do Processo Cognitivo).

Esses termos-chave foram utilizados para definir verbos digitais que orientam a seleção de atividades digitais adequadas aos objetivos educacionais relacionados a cada nível cognitivo. A TBD também inclui uma lista de atividades digitais para cada verbo digital.

Esse *framework* foi escolhido para este trabalho por permitir a personalização das recomendações com base em atividades digitais apropriadas à sequência de ações pedagógicas (2).

## 2.9 ENRIQUECIMENTO SEMÂNTICO DE OBJETOS DE APRENDIZAGEM (ESOA)

Nos últimos anos, houve esforços significativos para padronizar e desenvolver uma definição coletiva de metadados em Objetos de Aprendizagem (OAs). Esses padrões, através de diretrizes, garantem qualidade, interoperabilidade e reutilização desses recursos educacionais, facilitando seu acesso e uso eficiente na educação digital (3).

O Dublin Core (DC) e o Learning Object Metadata (LOM) são exemplos de padrões que fornecem um conjunto básico de elementos para descrever características, contexto educacional e requisitos técnicos de OAs (74). No entanto, embora esses repositórios e padrões de metadados sejam importantes para promover a interoperabilidade e a reutilização de recursos educacionais, muitas vezes é difícil encontrar materiais com metadados relevantes (91).

Segundo (86), o enriquecimento semântico emerge como uma estratégia para a geração de metadados adicionais de relevância. Ele atribui significado aos dados e facilita sua compreensão e processamento por pessoas e máquinas. Isso melhora a qualidade dos dados ao usar termos relacionados a ontologias, vocabulários e sinônimos conhecidos.

Uma prática comum de enriquecimento semântico é a indexação, onde palavras-chave são selecionadas e marcadas manualmente para identificar e descrever o conteúdo dos vídeos (42). No entanto, esse processo é demorado e condiciona a catalogação e formatação da informação às experiências pessoais (72) (93), o que reduz a eficácia dos métodos de busca existentes. Algumas soluções menos precisas incluem fazer uso da indexação de tags encontradas em wikis e blogs (90) para tratar esse problema sem muito esforço manual.

Outra técnica consiste em transformar o áudio do vídeo em texto (transcrição), permitindo que as palavras faladas sejam utilizadas como palavras-chave para indexação e busca. No entanto, essa técnica é custosa e raramente está disponível para sistemas de busca, pois exige trabalho manual de transcrição. Apesar disso, a transcrição de vídeos é uma técnica que pode ser bastante útil para melhorar a busca de conteúdo de áudio e vídeo (51) (93).

Um meio de se facilitar a transcrição é através do Reconhecimento Automático de Fala ou Automatic Speech Recognition (ASR). O ASR é uma técnica que permite o reconhecimento da fala humana por meio do processamento de sinais sonoros utilizando modelos estatísticos. Essa técnica possibilita a conversão da fala em texto, gerando transcrições que podem ser utilizadas em diversos contextos, como em sistemas de busca de vídeos, por exemplo. O ASR é capaz de processar diferentes idiomas e sotaques, embora sua eficácia possa variar em função da qualidade do áudio e das condições acústicas do ambiente. Além disso, o ASR é uma tecnologia em constante evolução, com avanços sendo realizados regularmente para melhorar sua precisão e abrangência (85).

Ao fim da transcrição textual, os dados gerados podem ser utilizados como anotações

semânticas. Em Popov et al.(73) é dito que a anotação semântica é uma técnica que permite atribuir metadados para gerar esquemas que possibilitam novos métodos de acesso à informação e aprimoram os já existentes. Já Körner et al (48) diz que o conceito de anotação semântica vai além da simples atribuição de metadados, permitindo uma categorização e descrição mais precisa dos recursos anotados. Isso resulta em uma melhora significativa na precisão dos mecanismos de recuperação da informação, já que a semântica ajuda a aumentar a relevância das informações buscadas. Dessa forma, espera-se que os resultados apresentados pelos mecanismos de busca sejam mais relevantes e adequados às necessidades do usuário.

### 3 TRABALHOS RELACIONADOS

Nessa seção, são apresentados trabalhos encontrados na literatura relacionados aos problemas apresentados nesta dissertação. Na Seção 3.1, são apresentados propostas de extração de metadados a partir de videoaulas. Já na Seção 3.2, são apresentados trabalhos que buscaram classificar automaticamente exercícios, enquanto que a Seção 3.3 apresenta a relevância deste trabalho.

#### 3.1 Extração de metadados

A extração de metadados de vídeos apresenta desafios devido à complexidade da composição contínua de imagens e áudio. Apesar das limitações, pesquisas neste campo têm progredido nas últimas décadas, com várias abordagens e tipos de extração de metadados documentados na literatura. A escolha das técnicas de extração varia de acordo com a aplicação (69).

Os estudos conduzidos por (19) e (69) apresentaram propostas de extração de metadados utilizando o padrão IEEE Learning Object Metadata (LOM). Em (19) foi desenvolvido o *Framework ABITS*, que, por meio de um conjunto de funcionalidades, possibilita a modelagem do aluno, identificando seus Objetivos de Aprendizado (OA), e elabora um currículo personalizado para alcançar tais objetivos. A eficácia dessas funcionalidades é fundamentada em um conjunto de regras para indexação de conhecimento com base em Metadados e Grafos Conceituais, aderentes ao padrão IEEE Learning Object Metadata (LOM). Por outro lado, (69) expandiu esse campo, adicionando 11 novos elementos destinados a identificar metadados educacionais, como disciplina e tópicos, a partir de videoaulas.

Outros estudos exploram abordagens diferentes, como demonstrado por (11), onde palavras-chave e segmentos baseados em tópicos foram extraídos de videoaulas por meio de transcrição. Em contraste, o trabalho apresentado por (78) propõe a geração de metadados a partir de comentários sobre trechos específicos do áudio, enriquecendo assim as informações associadas ao arquivo. Já (21, 20) desenvolveu um sistema de recomendação de Objetos de Aprendizado (OAs) de vídeos educacionais no *YouTube*, onde a categoria do vídeo é avaliada para fins de recomendação de conteúdo.

Em (54) foi desenvolvido o sistema BAVi (Busca Avançada de Vídeos) cujo principal objetivo é de facilitar a identificação de recursos didáticos pelos alunos. Ele alcança isso por meio do enriquecimento semântico desses recursos, usando a transcrição do conteúdo falado, anotações semânticas associadas a *tags* e relacionamentos entre os recursos de um repositório com base nas *tags* criadas. Em (14, 12) é proposto um sistema de recomendação de videoaulas no *Moodle* baseado no uso do BAVi para o enriquecimento semântico. O plugin desenvolvido e chamado de OVR (Open Vídeo Recommendation) tem como funções

a busca de vídeos, a recomendação por um vídeo principal e a recomendação com base no texto tópico da semana.

Buscando, também, auxiliar na recuperação de conteúdo educacional mais aderente ao resultado de aprendizagem pretendido pelo aluno, (36) apresenta uma análise comparativa entre fontes de informações presentes em videoaulas para identificar quais fontes geram melhor resultado para o processo de anotação semântica. (8, 9) utiliza características visuais dos estilos das videoaulas, como a presença de pessoas e textos, para classificar 4 diferentes estilos de videoaulas. Já (89, 88) apresenta uma proposta de segmentação de videoaulas, a partir de falas do professor que indiquem transições de tópicos. Em (13) é proposto um *framework* denominado *EasyTopic* que permite o uso de diversas técnicas para o processo de segmentação temporal de videoaulas ao construir um *pipeline* configurável a partir de módulos independentes, para validação da proposta, foi utilizada a segmentação de videoaulas apresentada em (89).

### 3.2 Classificação automática de exercícios

A automação da classificação de exercícios tem sido tema de diversos estudos, como evidenciado por uma variedade de artigos dedicados a essa questão (6, 35, 60, 77). No entanto, uma lacuna significativa na literatura é observada quando se trata da classificação automática específica dos exercícios apresentados em videoaulas. Embora os trabalhos existentes não abordem diretamente essa modalidade de classificação, alguns apresentam metodologias e abordagens que poderiam ser adaptadas para esse propósito em particular.

No trabalho de (66), é apresentada uma abordagem baseada em regras, que consiste na utilização de condicionais ou regras para realizar a classificação (57). Nesta abordagem, foram utilizados 100 exercícios extraídos de provas finais. O sistema classificava os exercícios com base nos respectivos verbos encontrados na Taxonomia de Bloom, além de utilizar regras elaboradas com base nos padrões identificados nas questões.

Por sua vez, (38) apresenta uma evolução do estudo realizado por (66). Neste trabalho, é adotada uma abordagem híbrida, na qual, para os casos em que o sistema baseado em regras não consegue identificar a classificação do exercício de forma adequada, são empregadas abordagens estatísticas, que se concentram nas características não linguísticas do texto, tais como a ocorrência de um termo em uma pergunta ou a frequência desse termo (57). A técnica utilizada para tal foi o modelo de *N-gram*, que visa calcular a probabilidade de ocorrência de uma palavra. Com essa abordagem, os resultados finais obtidos demonstraram uma melhoria significativa em relação ao estudo anterior.

Outros estudos, como os referidos na literatura (95, 84), também adotam a abordagem de aprendizado supervisionado, que guia o processo de classificação com base em características previamente extraídas ou selecionadas dos exercícios (57). Essas características são então empregadas para treinar o classificador, possibilitando a previsão da

classe da pergunta com base nos padrões identificados durante o treinamento. Em (95), o SVM é utilizado para esse propósito, onde testes de um conjunto de dados não divulgado são classificados. Já em (84), o modelo Naive Bayes foi adotado para a classificação de exercícios, sendo que, para otimizar os resultados, foi empregada a seleção de características *Chi-Square*, acompanhada pelo *Laplace Smoothing*.

### 3.3 Relevância desse estudo

Os estudos revisados na Subseção 3.1 abordam diversas técnicas de extração de metadados. Enquanto a literatura destaca a segmentação de vídeos e a utilização de palavras-chave para essa finalidade, ainda não foram exploradas amplamente as abordagens mais recentes que empregam *Large Language Models* (LLM). Os LLMs são modelos de inteligência artificial treinados em grandes volumes de texto para compreender e gerar texto em larga escala, sendo amplamente empregados em diversas aplicações de processamento de linguagem natural. Este trabalho visa preencher essa lacuna ao classificar os exercícios em videoaulas com base na Taxonomia de Bloom Digital (TBD), utilizando abordagens fundamentadas em LLM. Assim, propõe-se um modelo de classificação automática, com o intuito de aprimorar os sistemas de recuperação da informação e facilitar o acesso e organização de conteúdos educacionais digitais.

A Subseção 3.2 destaca que muitos estudos se concentram na classificação automática de exercícios provenientes de fontes textuais, como provas escritas, mas nenhum aborda especificamente a classificação automática de exercícios em videoaulas. A extração desses exercícios apresenta novos desafios significativos. Encontrar marcadores visuais que indicam exercícios em fontes textuais é fácil, mas em videoaulas, onde o conteúdo é apresentado em fala contínua, não há esses marcadores na transcrição. Isso dificulta a identificação de trechos que contenham exercícios. Adicionalmente, os exercícios podem estar inseridos em meio a dicas e explicações fornecidas pelo professor durante a apresentação do enunciado, o que torna a identificação ainda mais complexa. Esses desafios sugerem uma oportunidade para explorar e desenvolver métodos eficazes para a identificação automatizada e a categorização dos exercícios contidos em recursos de aprendizado em vídeo.

Portanto, o diferencial deste trabalho em relação às propostas existentes na literatura, reside na identificação e classificação de exercícios presentes em videoaulas de forma automática, utilizando modelos fundamentados em LLM e no uso de um conjunto real de videoaulas. Os resultados desta pesquisa podem contribuir para fornecer recomendações personalizadas de cursos para alunos.

#### 4 UMA PROPOSTA DE CLASSIFICAÇÃO AUTOMÁTICA DE EXERCÍCIOS EM VÍDEO AULAS

O objetivo deste trabalho consiste em propor um modelo de classificação automática de exercícios em videoaulas, fundamentado na Taxonomia de Bloom Digital, com o intuito de viabilizar metadados para sistemas de recomendação. Para alcançar esse propósito, é requisito identificar os exercícios a partir da transcrição do vídeo e compreender os princípios da Taxonomia de Bloom Digital, de modo a classificar de forma adequada os exercícios identificados.

Para a identificação dos exercícios, foi estabelecida uma definição do conceito de exercício. Neste trabalho, considera-se que exercícios são enunciados que abordam explicitamente conceitos, princípios ou operações educacionais, apresentando uma pergunta clara ou problema imperativo (questão expressa na forma de uma instrução) para os quais se espera uma resposta ou solução, podendo ou não ser acompanhados de um enunciado explicativo.

Exemplificando, são considerados exercícios: “A seguir, nos é fornecida uma matriz de verificação de paridade  $H$  de um código de bloco linear com parâmetros  $n$  e  $k$ , e é especificado que este código  $C$  possui tanto palavras-código de peso ímpar quanto palavras-código de peso par, ou seja, o número de 1's nas palavras-código contém tanto um número ímpar quanto um número par de 1's. Construa um novo código linear que estamos chamando de  $C1$  e a matriz de verificação de paridade do novo código  $C1$  é fornecida por esta.” e “Olhe detalhadamente este trecho das linhas 7 a 15 da fonte. Como o escritor utiliza a linguagem aqui para descrever a jornada do narrador?”. Estes exemplos são considerados exercícios, pois ambos estão alinhados a conceitos educacionais abordados em suas respectivas videoaulas apresentando. O primeiro, um problema imperativo, solicitando que se construa um novo código linear, enquanto o segundo faz uma pergunta clara a respeito a respeito do conteúdo sendo abordado. Em ambos os casos existe também, um enunciado explicativo, dando contextualização para a tarefa.

Essa definição é necessária, pois durante as videoaulas podem surgir perguntas voltadas para questões gerais, opiniões ou discussões sem foco no aprendizado ou sem exigir uma resposta específica. Também é possível que durante a resolução de um exercício, o professor faça uma pergunta com menor complexidade, semelhante ao conceito de recursividade, para assim chegar na resolução do exercício.

Por exemplo: “Vocês possuem qualquer dúvida sobre a lógica e o código de implementação para converter numeros romanos em inteiros?”. Essa pergunta constitui-se como uma pergunta de *feedback* destinada a identificar possíveis dificuldades dos alunos com o conteúdo apresentado. Outro exemplo apresenta uma pergunta feita durante a resolução de um exercício. O enunciado do exercício questiona “Qual a distância de Hamming entre  $X$  e  $Y$ ?” e durante a resolução o professor pergunta “Qual a distância

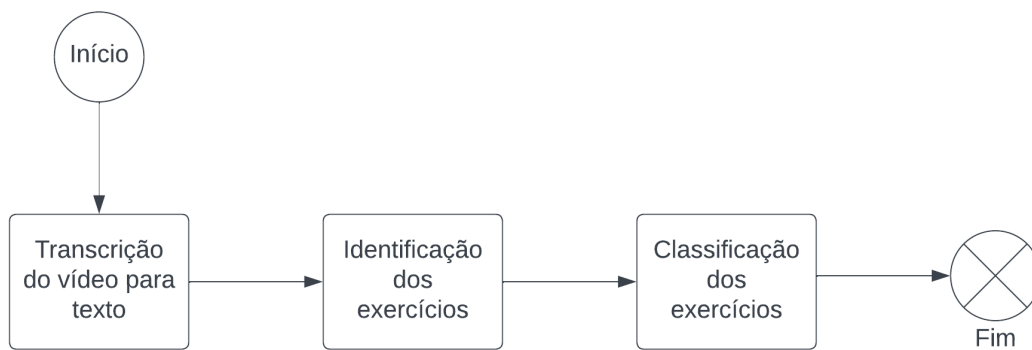


mínima entre *u plus* e *u prime*”. Neste exemplo para identificar a distância Hamming o professor, primeiramente, quis identificar a distância mínima entre o *u plus* e o *u prime*, para, a partir dessa resposta, conseguir resolver o exercício.

Essas perguntas foram consideradas como não sendo exercícios e com a definição estabelecida, é esperado identificar com maior precisão os exercícios presentes no texto.

Definido o conceito de exercício, a abordagem se inicia com a recepção da videoaula, onde seu áudio é automaticamente transcrito para texto. Em seguida, são realizadas a identificação e a classificação automáticas dos possíveis exercícios contidos na videoaula. Os exercícios são classificados de acordo com os diferentes níveis da Taxonomia de Bloom Digital. Ao final do processo, busca-se obter uma tupla contendo o exercício e sua classificação correspondente, consolidando assim o resultado da análise realizada. A Figura 1 representa o fluxograma adotado para essa abordagem.

Figura 1 - Fluxograma da solução



Fonte: Elaborada pelo autor (2024)

## 4.1 ARQUITETURA

Esta seção tem como objetivo detalhar a arquitetura da abordagem apresentada. Nela serão, detalhados os módulos de transcrição, identificação e classificação de exercícios apresentados nas subseções 4.1.1, 4.1.2 e 4.1.3. Por fim, o funcionamento do *workflow* completo utilizando o framework M2P, na subseção 4.1.4.

### 4.1.1 Módulo de Transcrição

O módulo de transcrição tem como objetivo converter o conteúdo de uma videoaula em texto. Para isso, foram considerados três modelos de transcrição. Os dois primeiros

utilizam o framework wav2vec2 criado em (10), o *wav2vec2-large-xlsr-53-english*, foi desenvolvido por Grosman<sup>1</sup> e treinado com 1.584 horas de áudio do conjunto de dados *Common\_Voice 6.1*. Já o modelo *wav2vec2-large-960h-lv60-self*, foi criado pelos mesmos autores de (10) e treinado com 960 horas utilizando o conjunto de dados *librispeech\_asr*. Finalmente, o *Whisper*, desenvolvido em (75) e treinado com 680.000 horas de áudio, compreendendo dados de vários conjuntos, incluindo *Common\_Voice* e *VoxPopuli\_En*.

O *Whisper* oferece uma variedade de modelos de transcrição, que variam desde o tamanho *tiny* até o *large*, conforme ilustrado na Figura 2. À medida que o tamanho da base aumenta, torna-se necessário um maior investimento de recursos computacionais para realizar a transcrição, resultando também em uma diminuição na velocidade relativa de execução. Verificou-se que os modelos com a extensão *.en*, especialmente otimizados para aplicações em inglês, superam em desempenho suas contrapartes generalistas de tamanho equivalente. Além disso, nota-se que a diferença de desempenho entre os modelos diminui a partir do tamanho *small*.

Nesse contexto, o modelo *small.en* foi selecionado como representante do *Whisper*, pois mantém um desempenho satisfatório em comparação com os modelos de maior tamanho, ao mesmo tempo em que requer um consumo reduzido de recursos computacionais e um tempo de execução menor.

Figura 6 - Modelos Disponíveis

| Size   | Parameters | English-only model     | Multilingual model  | Required VRAM | Relative speed |
|--------|------------|------------------------|---------------------|---------------|----------------|
| tiny   | 39 M       | <code>tiny.en</code>   | <code>tiny</code>   | ~1 GB         | ~32x           |
| base   | 74 M       | <code>base.en</code>   | <code>base</code>   | ~1 GB         | ~16x           |
| small  | 244 M      | <code>small.en</code>  | <code>small</code>  | ~2 GB         | ~6x            |
| medium | 769 M      | <code>medium.en</code> | <code>medium</code> | ~5 GB         | ~2x            |
| large  | 1550 M     | N/A                    | <code>large</code>  | ~10 GB        | 1x             |

Fonte: Figura extraída do github do *Whisper*.

Ao final, o modelo *Whisper small.en* demonstrou um desempenho superior do Word Error Rate (WER) em geral quando comparado aos demais, sugerindo uma transcrição mais eficaz das videoaulas selecionadas. Além disso, ele pode ser utilizado com formatos *.mp4*, eliminando a necessidade de uma etapa de conversão para áudio. Portanto, optou-se por prosseguir com o modelo *Whisper* para a execução da transcrição.

O resultado produzido pelo modelo é então, posteriormente, salvo em um arquivo com a extensão *.txt*. Para evitar textos muito longos, o texto é dividido em partes menores,

<sup>1</sup> <https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-english>

denominadas *chunks*. Esses pedaços menores são então salvos no mesmo arquivo, com uma quebra de linha para diferenciá-los.

#### 4.1.2 Módulo de Identificação dos exercícios

Para o módulo de identificação dos exercícios, foi optado pelo emprego do GPT-3.5 Turbo, um modelo de linguagem desenvolvido pela OpenAI. Este modelo é reconhecido por sua capacidade abrangente de processamento de linguagem natural, incluindo funções como geração de texto, tradução automática, resposta a perguntas, entre outras. Além disso, o GPT-3.5 Turbo também conta com uma quantidade significativa de parâmetros, estimada em cerca de 175 bilhões.

Este modelo que é responsável por identificar os exercícios contidos no *chunk* fornecido como entrada. Neste módulo, é feita uma requisição de conversação ao modelo escolhido, no qual são passados os parâmetros necessários para que ele possa identificar e retornar os exercícios. Após conclusão da identificação dos exercícios, as questões identificadas são salvas em um novo arquivo.

#### 4.1.3 Módulo de Classificação dos exercícios

Por fim, tem-se o módulo de classificação de exercícios o qual foi construído sob quatro diferentes perspectivas. Que se constituem de uma abordagem baseada em regras, além de estratégias mais avançadas baseadas em técnicas de aprendizado de máquina, as quais exploraram as nuances semânticas do texto.

Nas abordagens baseadas em aprendizado de máquina, foram adotados: Naive Bayes (NB), BERT e o GPT-3.5 Turbo. O Naive Bayes foi escolhido devido à sua velocidade no treinamento e na classificação, enquanto a rede neural BERT foi selecionada por suas diversas vantagens, incluindo a capacidade de pré-treinamento do modelo, o que pode economizar tempo e recursos computacionais. Já o GPT-3.5 Turbo foi escolhido por ser um representante de IA generativa.

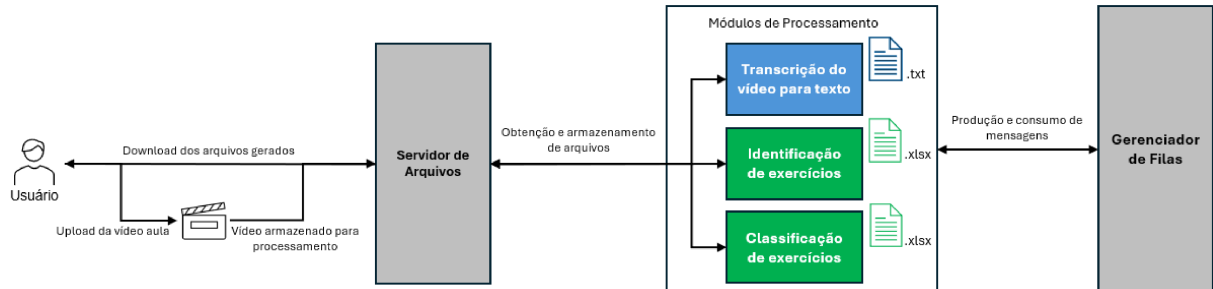
Este módulo recebe como entrada os exercícios identificados pelo módulo anterior e, em seguida, os diferentes modelos de classificação apresentados, são aplicados para classificar os exercícios em um dos diversos níveis da Taxonomia de Bloom Digital. A classificação resultante é então salva juntamente com o exercício.

#### 4.1.4 *Workflow*

A solução proposta neste trabalho foi construída utilizando o *framework* **M2P** (Apêndice A), onde para este trabalho, todos os módulos utilizados no *workflow* foram desenvolvidos. Isso se deve ao fato de que não havia módulos para executar as funcionalidades de identificação e classificação dos exercícios. Mesmo no caso do módulo de transcrição, embora existisse um módulo com uma função similar, optou-se por sua atualização para

estar em linha com o estado da arte na área. A Figura 3 apresenta a arquitetura reduzida com a utilização dos módulos.

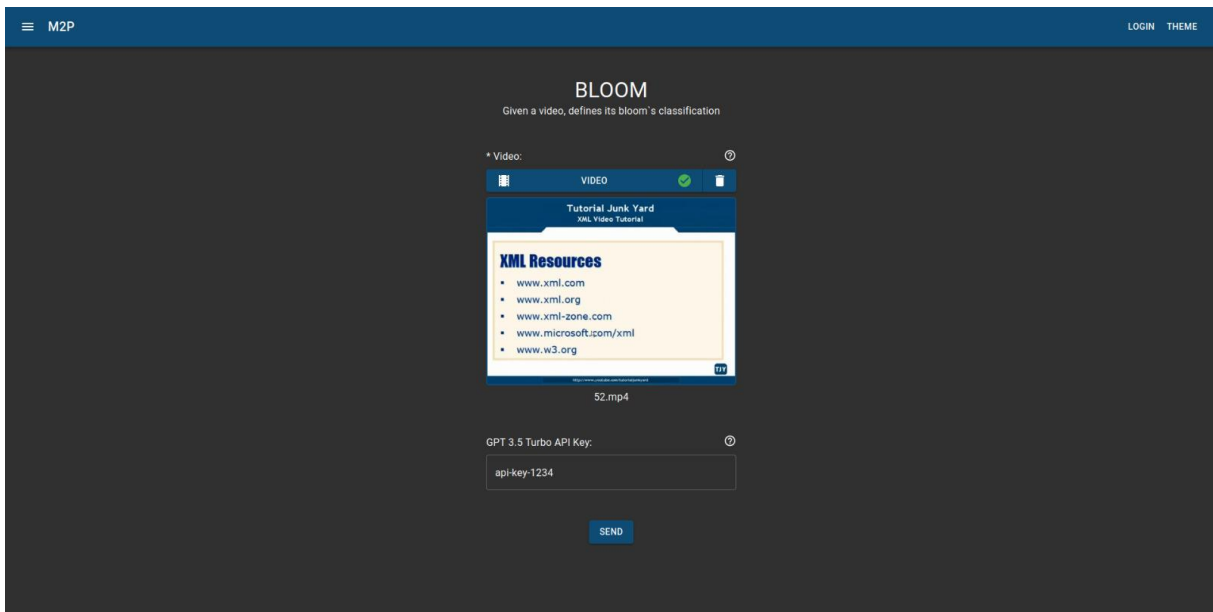
Figura 3 - Arquitetura do Workflow de Classificação Automática de Exercícios



Fonte: Elaborada pelo autor com base em (65) (2024)

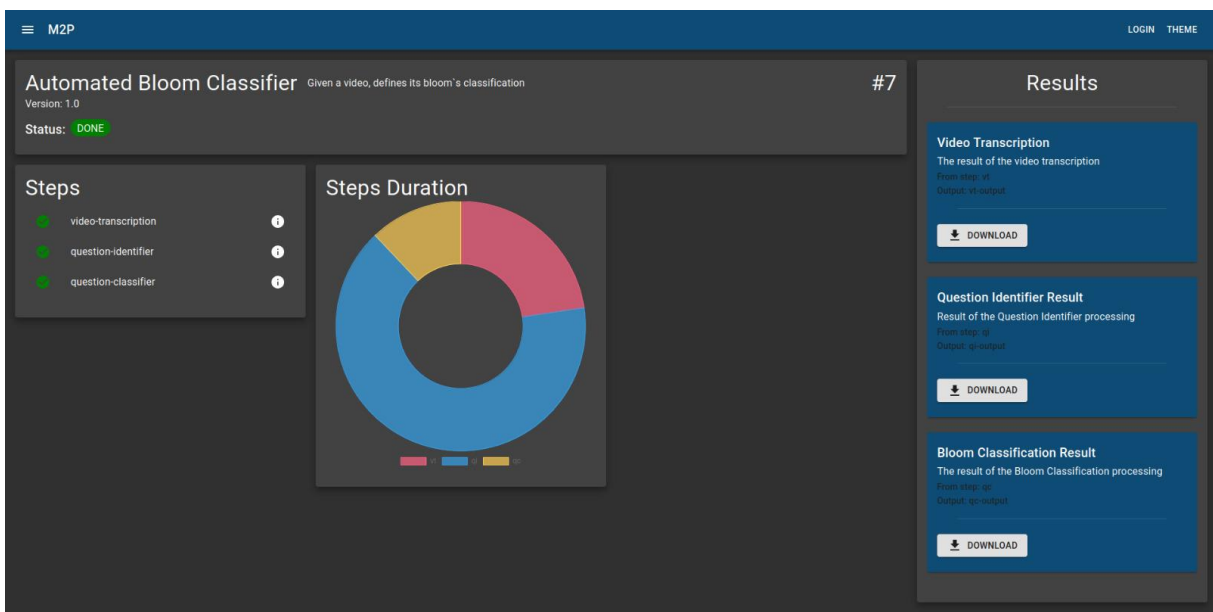
O *workflow* da solução inicia solicitando ao usuário inserir uma videoaula, conforme Figura 4. Em seguida, o sistema armazena o vídeo em um servidor de arquivos, permitindo que o módulo de transcrição acesse esse arquivo de entrada e execute as ações descritas na Subseção 4.1.1, resultando na produção do texto em *chunks*, o qual é então salvo no servidor de arquivos. Posteriormente, o módulo de identificação dos exercícios recupera o documento gerado e realiza a identificação dos exercícios conforme descrito na Subseção 4.1.2, resultando na produção de um arquivo contendo os exercícios identificados. Em seguida, o processo de classificação tem início (Subseção 4.1.3) resultando na classificação dos exercícios e possibilitando que o usuário faça o download de todos os arquivos gerados, conforme apresentado na Figura 5.

Figura 4 - Tela inicial do Workflow



Fonte: Elaborada pelo autor (2024)

Figura 5 - Tela final do Workflow



Fonte: Elaborada pelo autor (2024)

## 5 EXPERIMENTOS

No presente capítulo, será discutida a implementação da abordagem para classificação automatizada de exercícios em videoaulas, apresentada no Capítulo 4. Para avaliar a eficácia do método, foram testados diversos métodos de classificação. Além disso, uma base de dados foi desenvolvida e disponibilizada para a comunidade científica, visto que não há bases de dados públicas projetadas especificamente para a classificação de exercícios em videoaulas. Espera-se que essa classificação automática possa ser utilizada por sistemas de recomendação para sugerir videoaulas mais alinhadas às preferências dos alunos e aos objetivos de aprendizagem desejados.

A estrutura deste capítulo é organizada da seguinte forma: as bases de dados utilizadas neste trabalho são expostas em detalhes na Seção 5.1, seguidos pela análise dos resultados na Seção 5.2.

### 5.1 BASE DE DADOS

Nesta seção, serão detalhadas as bases de dados selecionadas, destacando suas características principais. Todas as bases geradas estão disponíveis no repositório público<sup>1</sup>.

#### 5.1.1 *Base Bloom*

A primeira base de dados utilizada foi a *Bloom's Taxonomy Cognitive Levels Data Set*<sup>2</sup>. Esta base é composta por uma série de exercícios coletados de várias fontes na Web, os quais possuem marcações relacionadas à Taxonomia de Bloom. O conjunto de dados empregado em estudos anteriores (96, 95), e possui duas colunas distintas. Na primeira coluna, intitulada *Question*, estão os exercícios. Estes foram previamente classificados manualmente por um especialista pedagógico em relação aos seis níveis cognitivos da Taxonomia de Bloom, sendo que cada nível possui 100 exercícios associados a ele. Tal classificação resultou na segunda coluna, denominada *Level*. A escolha desta base para o treinamento dos classificadores foi motivada pela classificação manual realizada por um profissional pedagógico, visando minimizar possíveis erros de classificação.

Para tornar a **Base Bloom** adequada à classificação automática, foi feito o remapeamento da coluna *Level* para números em uma nova coluna chamada *Label*. Essa ação facilita a classificação multinomial, em que cada número representa um nível da Taxonomia de Bloom Digital (0=Lembrar; 1=Compreender; 2=Aplicar; 3=Analisar; 4=Avaliar; 5=Criar)

<sup>1</sup> [https://drive.google.com/drive/folders/17jP72jU9R6Z1EFmXttuLUiQvBx\\_EYZas?usp=sharing](https://drive.google.com/drive/folders/17jP72jU9R6Z1EFmXttuLUiQvBx_EYZas?usp=sharing)

<sup>2</sup> [https://www.researchgate.net/publication/303608228\\_Bloom%27s\\_Taxonomy\\_Cognitive\\_Levels\\_Data\\_Set](https://www.researchgate.net/publication/303608228_Bloom%27s_Taxonomy_Cognitive_Levels_Data_Set)

Após adequação da base de dados, para identificar termos de ação comuns que aparecem em vários exercícios de diferentes níveis da TBD, foi executada uma análise manual da base. Esses termos de ação podem ser considerados palavras-chave que ajudam a identificar o nível de dificuldade do exercício base TBD. Ao fim da análise, foram identificados 26 termos que repetiam 290 vezes entre exercícios de diferentes níveis da TBD (Tabela 1). Dessa forma, os demais 310 termos repetiam apenas dentro de um dos níveis.

Tabela 1 – Matriz de termos de ação que estão em mais de um nível

| <b>Termos</b> | <b>Lembrar</b> | <b>Compreender</b> | <b>Aplicar</b> | <b>Analisar</b> | <b>Avaliar</b> | <b>Criar</b> |
|---------------|----------------|--------------------|----------------|-----------------|----------------|--------------|
| Choose        | 0              | 0                  | 2              | 0               | 2              | 0            |
| Compare       | 0              | 4                  | 0              | 16              | 0              | 0            |
| Contrast      | 0              | 1                  | 0              | 6               | 0              | 0            |
| Construct     | 0              | 0                  | 16             | 0               | 1              | 0            |
| Define        | 5              | 1                  | 0              | 0               | 1              | 0            |
| Describe      | 0              | 18                 | 0              | 0               | 1              | 0            |
| Design        | 0              | 0                  | 0              | 0               | 1              | 11           |
| Determine     | 0              | 2                  | 0              | 0               | 6              | 2            |
| Develop       | 0              | 0                  | 1              | 0               | 0              | 9            |
| Discuss       | 0              | 1                  | 0              | 0               | 7              | 0            |
| Draw          | 4              | 0                  | 3              | 0               | 0              | 1            |
| Examine       | 0              | 0                  | 0              | 7               | 1              | 0            |
| Explain       | 0              | 14                 | 0              | 2               | 3              | 14           |
| Identify      | 17             | 0                  | 0              | 0               | 0              | 1            |
| Justify       | 0              | 0                  | 0              | 0               | 2              | 1            |
| Make          | 1              | 0                  | 0              | 1               | 0              | 2            |
| Name          | 6              | 0                  | 0              | 0               | 0              | 1            |
| Predict       | 0              | 0                  | 1              | 0               | 2              | 1            |
| Pretend       | 0              | 0                  | 1              | 0               | 0              | 1            |
| Recommend     | 0              | 0                  | 0              | 0               | 1              | 1            |
| Revise        | 0              | 0                  | 0              | 0               | 1              | 3            |
| Rewrite       | 0              | 5                  | 0              | 0               | 0              | 3            |
| Select        | 0              | 1                  | 0              | 1               | 3              | 0            |
| State         | 43             | 4                  | 0              | 0               | 0              | 0            |
| Use           | 0              | 1                  | 11             | 0               | 1              | 0            |
| Write         | 0              | 0                  | 0              | 0               | 1              | 9            |

Esses termos que se repetem dentro de um dos níveis da TBD são úteis para identificar características específicas de cada nível. Por exemplo, palavras como analisar, sintetizar e avaliar podem indicar o nível de analisar da TBD, enquanto termos como lembrar, reconhecer e listar podem sugerir o nível de lembrar da TBD. Por outro lado, os demais termos, embora repetidos, tendem a se inclinar mais para um nível específico da taxonomia do que para outro, como pode ser observado na Tabela 1. Essas informações podem ser úteis para a criação de um modelo de classificação de exercícios ou mesmo para a execução da classificação manual.

### 5.1.2 Base de Videoaulas (BVA)

Para classificar os exercícios, inicialmente, foram identificadas videoaulas adequadas para a classificação sob a Taxonomia de Bloom Revisada. Foram selecionadas videoaulas de diversos temas e que estivessem licenciadas sob *Creative Commons*, assegurando sua utilização legal para fins educacionais. Ao fim da seleção, foi criada a **Base de Videoaulas (BVA)** com 42 videoaulas que incluem exercícios para reforço do aprendizado. Esses vídeos, todos em inglês, foram obtidos da plataforma Youtube<sup>3</sup>, com uma duração total de 16,28 horas, variando de 4 a 61 minutos cada, e uma média de 23 minutos de duração.

Dentre as videoaulas identificadas, 32 apresentam a explicação do conteúdo seguida por exercícios de fixação, enquanto as restantes, consistem exclusivamente de exercícios. A diversidade de abordagens apresenta um desafio adicional na identificação de exercícios, uma vez que, em videoaulas com explanação do conteúdo, é possível que o professor mescle exercícios com informações adicionais.

A base possui vídeos categorizados em 5 estilos distintos: *Talking Head*, *Voice Over Slides*, *Animation*, *Hand-drawn* e *Whiteboard*, relacionados a tipos distintos de área de conhecimento, conforme pode ser visto na Tabela 2. Esses estilos podem indicar padrões distintos relacionados aos exercícios em videoaulas diversas. No formato *Talking Head*, o foco central é o instrutor, ocupando uma parte significativa do quadro e se comunicando diretamente com a câmera na maior parte do tempo (68). Não há presença de *slides* ou outros elementos ricos em texto, embora, ocasionalmente, texto sobreposto possa ser utilizado para enfatizar ideias-chave, ou pode haver mudanças de cena para apresentar outro tipo de material (82).

No estilo *Voice Over Slides*, a apresentação é composta por slides de *PowerPoint* ou outro formato similar exibidos em tela inteira, acompanhados pela narração do instrutor (44). Por outro lado, o estilo *Animation* refere-se a imagens em movimento geradas por computador que mostram associações entre as Figuras desenhadas (58). Já o *Hand-drawn* envolve atos manuais de escrita, nos quais a mão do instrutor se torna visível enquanto ele faz anotações (22). Por fim, o *Actual Paper / Whiteboard* consiste em um monólogo feito enquanto o instrutor se move na frente do conteúdo em um quadro branco e interage com ele (27). O instrutor olha diretamente para a câmera, criando a impressão de estar falando diretamente com o espectador.

### 5.1.3 Base de exercícios manual (BEM)

A partir das videoaulas selecionadas, uma nova base de dados foi elaborada, contendo 139 exercícios identificados manualmente. Esta base é composta por 9 colunas. Na primeira coluna, denominada *Question*, estão os próprios exercícios identificados nas videoaulas correspondentes. A segunda coluna, intitulada *Source*, identifica qual videoaula

<sup>3</sup> <https://www.youtube.com/>



Tabela 2 – Distribuição dos estilos de videoaulas por área de conhecimento, onde TH, VO, A, HD e W representam os estilos *Talking Head*, *Voice Over Slides*, *Animation*, *Hand Drawn* e *Whiteboard*, respectivamente

| Área de Conhecimento | TH       | VO        | A        | HD       | W        |
|----------------------|----------|-----------|----------|----------|----------|
| Computação           | 1        | 5         | 0        | 0        | 0        |
| Biologia             | 0        | 1         | 0        | 0        | 0        |
| Probabilidade        | 0        | 0         | 0        | 0        | 1        |
| Geometria            | 0        | 0         | 0        | 0        | 1        |
| Literatura Inglesa   | 0        | 4         | 9        | 2        | 1        |
| Geografia            | 0        | 2         | 0        | 0        | 0        |
| Química              | 0        | 0         | 0        | 0        | 1        |
| Física               | 0        | 11        | 0        | 0        | 3        |
| <b>Total</b>         | <b>1</b> | <b>23</b> | <b>9</b> | <b>2</b> | <b>7</b> |

Fonte: Elaborada pelo autor (2024).

o exercício foi extraído. Por sua vez, a terceira coluna, *Subject*, identifica o tema de cada exercício e a quarta, *VideoClass Style*, o estilo da videoaula. Uma quinta coluna, *Bloom's Style*, foi desenvolvida para indicar se o exercício está formatado de acordo com o padrão usualmente utilizado pelos professores ao formularem atividades com base na Taxonomia de Bloom Digital, ou se apresenta um formato diferente. A sexta reflete a classificação realizada pelo autor, com base nos seis níveis cognitivos da Taxonomia de Bloom Digital. As demais são as classificações automáticas dos modelos testados. Além dos 139 exercícios identificados como válidos, foram também adicionados 19 exercícios considerados inválidos, contendo apenas as colunas *Question* e *Source*.

Para a classificação manual da BEM, foram utilizados como parâmetro as listas de verbos por nível disponíveis na internet. A distribuição dos exercícios entre os diferentes níveis foi a seguinte: Lembrar 54, Compreender 3, Aplicar 54, Analisar 19, Avaliar 6 e Criar 3.

Dos exercícios identificados, a maioria (93) foi encontrada em vídeos do estilo *Voice Over Slides*, seguido por 34 em *Whiteboard*, 9 em *Animation*, 2 em *Hand-drawn* e 1 em *Talking Head*. Ao analisar a adequação dos exercícios com base na Taxonomia de Bloom, observa-se que 71% dos exercícios *Voice Over Slides* estão alinhados com o padrão de exercício proposto. Para os exercícios do tipo *Hand-drawn* e *Animation*, esse número é de 100%, enquanto para *Whiteboard* reduz para 52%. No caso do estilo *Talking Head*, que possui apenas um exercício, o valor é 0%.

Os exercícios classificados como LOTS (Lower Order Thinking Skills) compreendem os três primeiros níveis da Taxonomia de Bloom: Lembrar, Compreender e Aplicar. Neste contexto, foram identificados 111 exercícios que se enquadram nesses níveis. Dentre esses exercícios, 69% estão relacionados a videoaulas no estilo *Voice Over Slides*, 28% em *Whiteboard* e 3% em *Animation*.

Os 28 exercícios restantes são considerados da ordem HOTS (Higher Order Thinking Skills) que compreendem Analisar, Avaliar e Criar. A distribuição desses exercícios por estilo de videoaula é a seguinte: 57% em *Voice Over Slides*, 11% em *Whiteboard*, 21% em *Animation*, 7% em *Hand Drawing* e 4% em *Talking Head*.

Com base no número amostral de exercícios de videoaulas dos estilos *Voice Over Slides* e *Whiteboard*, é possível inferir que esses estilos estão normalmente associados a exercícios LOTS. Por outro lado, embora os estilos *Animation*, *Hand Drawing* e *Talking Head* tendam a representar exercícios HOTS, a amostragem disponível não é suficiente para fazer uma afirmação definitiva sobre essa associação.

Identificou-se, também, que muitas vezes os instrutores não explicavam os exercícios conforme o enunciado. Esse fenômeno ocorreu com maior frequência nos estilos *Voice Over Slides*, *Whiteboard* e *Talking Head*, onde os instrutores utilizavam uma abordagem de explicação mais semelhante à encontrada em sala de aula, o que pode ter colaborado para a identificação de exercícios mais discrepantes em relação a Taxonomia de Bloom. Em contrapartida, para *Hand-drawn* e *Animation*, os instrutores explicavam os exercícios como um passo a passo para a resolução.

#### 5.1.4 Base de exercícios automática (BEA)

A **Base de Exercícios Automática (BEA)**, foi gerada a partir do resultado do processo de identificação automática de exercícios. Esta base contém um total de 107 exercícios identificados automaticamente, dos quais 102 foram considerados válidos, pois estão de acordo com a definição de exercícios apresentada no trabalho, enquanto 5 exercícios foram classificados como inválidos, pois ou são perguntas relacionadas a *feedback* ou perguntas feitas durante a resolução de um exercício.

Assim como a BEM (Subseção 5.1.3), esta base também é composta por 9 colunas, contendo as mesmas informações e também utilizada para classificação automática. A distribuição manual dos exercícios entre os diferentes níveis da TBD, foi a seguinte: Lembrar 42, Compreender 2, Aplicar 37, Analisar 14, Avaliar 5 e Criar 2. Os demais 5 exercícios inválidos, não possuem classificação.

## 5.2 ANÁLISE DOS RESULTADOS

Nesta seção, serão abordados os processos de identificação e classificação dos exercícios presentes nas videoaulas. O objetivo principal desta seção é analisar os resultados obtidos com os métodos e técnicas utilizados para identificar os exercícios nos textos transcritos das videoaulas e, em seguida, classificá-las de maneira automatizada. A Tabela 3 apresenta a relação entre as bases de dados, as ferramentas utilizadas e os processos apresentados.

Tabela 3 – Relacional entre o processo, base de dados e ferramenta utilizados

| <b>Processo</b> | <b>Base de Dados</b>  | <b>Ferramenta</b>                |
|-----------------|-----------------------|----------------------------------|
| Transcrição     | BVA                   | <i>Whisper</i>                   |
| Identificação   | BEM e BEA             | <i>GPT-3.5 few-shot</i>          |
| Classificação   | Base Bloom, BEM e BEA | BR, NB, BERT e <i>GPT-3.5 FT</i> |

Fonte: Elaborada pelo autor (2024).

### 5.2.1 Identificação dos exercícios

Foram realizados dois experimentos distintos para identificar os exercícios. No primeiro experimento, o objetivo era determinar exercícios como válidos e inválidos. No segundo experimento, foram identificados e retornados os exercícios válidos a partir da transcrição.

#### 5.2.1.1 1º experimento

A identificação automática dos exercícios foi realizada utilizando a abordagem de *few-shot learning* no GPT-3.5 Turbo. Para isso, uma amostra composta por cinco exemplos de exercícios considerados válidos e seis exemplos de exercícios considerados inválidos, conforme definido neste trabalho, foi empregada. Esses exemplos foram selecionados da base apresentada na Subseção 5.1.3. Além disso, a configuração utilizada incluiu uma temperatura de 0, o que indica uma preferência por respostas mais determinísticas e coerentes.

Uma amostra de 29 exercícios, extraídos da base da Subseção 5.1.3, foi utilizada para realizar testes. Destes, 16 exercícios (Tabela 4) eram considerados válidos e 13 inválidos (Tabela 5). Os resultados obtidos são apresentados na Tabela 6, demonstrando que a abordagem utilizando *few-shot* foi eficaz, já que os resultados alcançados são significativos para a classificação. Foram alcançados valores de precisão, revocação e F1-score de 93,8%, quanto a acurácia, o valor obtido foi de 93,1%.

A Figura 6 exibe a matriz de confusão resultante do teste realizado. Foi identificado um erro, no qual o GPT classificou erroneamente, como válida, a pergunta: “Então, vamos analisar o trecho e nos perguntar. Qual é o efeito específico que o escritor está tentando alcançar?”. Esta pergunta, embora contextualizada no âmbito educacional de literatura inglesa, tratava-se de uma interrogação retórica destinada a instigar a reflexão sobre um determinado tema. Da mesma forma, ocorreu o oposto, onde o exercício: “Uma muito semelhante. A única diferença é que agora, em vez de se mover a uma velocidade constante, diz que está sendo baixado a uma velocidade decrescente. Qual é verdadeiro?” foi erroneamente classificado como inválido. Neste caso, foi possível observar que o exercício não seguia a estrutura típica dos demais, o que, provavelmente, levou o GPT a encontrar dificuldades em classificá-lo corretamente, porém ele condiz com a definição de exercício,

Tabela 4 – Exercícios válidos utilizados para teste

| <b>Exercícios</b>   |
|---|
| Qual é o valor de latitude do Trópico de Câncer?  |
| O peso molecular do ácido benzóico no benzeno, conforme determinado pelo método de depressão do ponto de congelamento, corresponderá a quanto?  |
| Olhe detalhadamente este trecho das linhas 7 a 15 da fonte. Como o escritor usa a linguagem aqui para descrever a jornada do narrador?  |
| Um site de viagens está realizando uma competição de escrita criativa. Ou descreva uma jornada sugerida por esta imagem, ou escreva uma história com o título As Férias.  |
| Como Priestley explora a importância do tempo em Uma Visita de Inspetor?  |
| Compare como os poetas apresentam a realidade da guerra em Carga da Brigada Ligeira e em outro poema.   |
| O que é o movimento harmônico simples e quais são as condições necessárias para que um corpo execute o movimento harmônico simples?   |
| Desenhe uma onda de transferência com uma amplitude de 2 centímetros e um comprimento de onda de 4 centímetros, rotulando uma crista e um vale na onda.   |
| São 120 metros de um portão para outro em um aeroporto. Use uma correia inclinada a 45 graus para puxar sua mala pelo aeroporto. A tensão na correia é de 20 newtons, então quanto trabalho você faz?   |
| Uma formiga zigzagueia para frente e para trás em uma mesa de piquenique como mostrado. A distância percorrida e o deslocamento da formiga são o quê?   |
| Uma bola vermelha pesada é liberada do repouso a 2 metros acima de uma superfície horizontal plana. Ao mesmo tempo, uma bola amarela com a mesma massa é disparada horizontalmente a 3 metros por segundo. Qual bola atinge o solo primeiro?  |
| Um muito semelhante. A única diferença é que, agora, em vez de se mover a uma velocidade constante, diz que está sendo abaixado a uma velocidade decrescente. O que é verdade.  |
| Uma bola foi lançada verticalmente para cima. Qual dos seguintes é o diagrama de corpo livre correto logo após a bola deixar a mão? Como sempre, ignore a resistência do ar.  |
| Uma bola com peso de 50 Newtons é puxada para trás e mantida no lugar por uma corda a um ângulo de 20 graus em relação à vertical. Qual é a tensão na corda de tração?  |
| Os dois gráficos mostrados mostram dois gráficos de força versus tempo suave para colisões. Qual dos dois entrega o maior impulso?  |
| Você acorda à noite para descobrir que sua sala de estar está em chamas. Esperançosamente, isso nunca acontece. Eu não escrevi este apenas para ser claro. Sua única chance de se salvar é lançar algo que atingirá a parte de trás da porta do seu quarto e fechá-la, dando-lhe tempo para escapar pela janela. Você acontece de ter tanto uma bola de argila pegajosa quanto uma Superbola super saltitante ao lado da sua cama, ambas do mesmo tamanho e mesma massa. Você só tem tempo para jogar uma. Qual você escolherá? |

Fonte: Elaborada pelo autor (2024).

Tabela 5 – Exercícios inválidos utilizados para teste

| <b>Exercícios</b>  |
|--|
| Como não temos realmente ninguém na plateia, você tem alguma pergunta sobre o primeiro?  |
| Então, com isso, teremos a resposta como ABC com um comprimento de três. Então, por que ainda não terminamos com isso?                             |
| Agora, se adicionarmos esse código X a todos os elementos no conjunto S, o que obtemos?  |
| Na verdade, a maioria deles está bem, mas, em geral, alguém pode apenas olhar e me dizer uma ordem que funcione para eles?                         |
| E então, e quanto aos negativos?   |
| Então, vamos olhar o trecho e nos perguntar, qual é o efeito específico que o escritor está tentando alcançar?                                     |
| O que isso significa?  |
| Ele também vai revisar nossa pergunta. Há alguma pergunta?   |
| E então, como estamos obtendo esta condição?   |
| Agora, por que convulsões artificiais?   |
| Alguma pergunta sobre o que estaremos fazendo hoje?  |
| E então, e quanto aos outros dois?   |
| Se você colocasse 2 aqui ou não tivesse essas chaves curvas ao redor daqui, isso não seria correto e você teria pontos marcados. Isso faz sentido? |

Fonte: Elaborada pelo autor (2024).

Tabela 6 – Testes da abordagem *few-shots*

| <b>Precisão</b> | <b>Revocação</b> | <b>F1</b> | <b>Acurácia</b> |
|-----------------|------------------|-----------|-----------------|
| 0,938           | 0,938            | 0,938     | 0,931           |

Fonte: Elaborada pelo autor (2024).

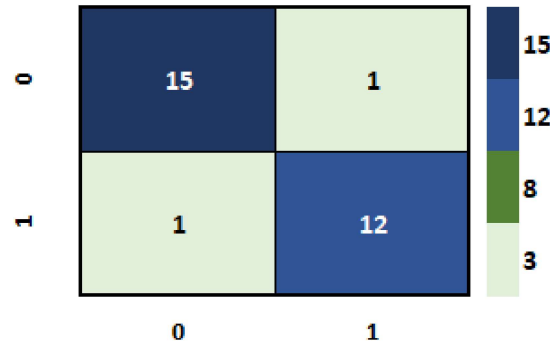
pois possui explícito conceito educacional e uma pergunta explícita, sobre o decaimento da velocidade.

### 5.2.1.2 2° experimento

O segundo experimento consistiu em uma reformulação do prompt, no qual foi solicitado que, utilizando o conhecimento sobre exercícios válidos e inválidos, o modelo identificasse e retornasse os exercícios válidos contidos no texto transcrito. Foram removidos dessa etapa, os 21 exercícios válidos utilizados no primeiro experimento. Foi observado que o modelo GPT, criou exercícios com base nas informações apresentadas nas aulas, extrapolando os resultados fornecidos. Esses exercícios foram classificados como alucinações, sendo consequentemente excluídos dos experimentos.

Entre os 118 exercícios restantes, o GPT demonstrou uma precisão de 95,3% na identificação de exercícios, com um revocação de 86,4%, um F1-score de 91,0% e uma

Figura 7 - Matrizes de Confusão

**Matriz BEM 29**

Fonte: Figura elaborada pelo autor (2024).

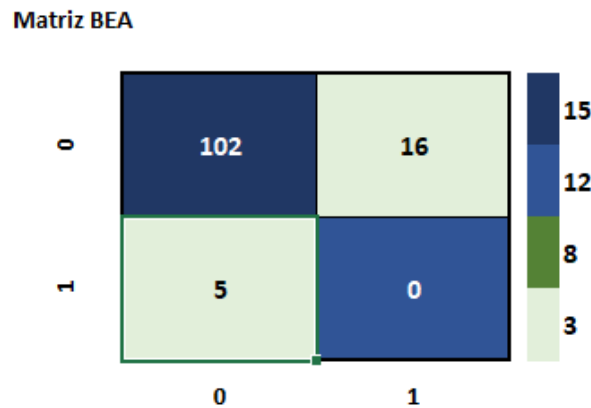
acurácia de 83,0%. A Matriz de Confusão (Figura 7) revela que 102 exercícios foram identificados como válidos, enquanto 16 foram considerados inválidos. Além disso, foram identificados outros 5 que estavam presentes na transcrição textual, mas não correspondiam a exercícios válidos. Como resultado, foi criada a Base de Exercícios Automática (Subseção 5.1.4) 107 exercícios. Dos 16 exercícios não reconhecidos pelo GPT, todos estavam de acordo com a definição de exercício, porém apenas 7 estavam estruturados conforme a Taxonomia Bloom. Em relação aos 5 exercícios inválidos classificados erroneamente como válidos e posteriormente recuperados, um deles era uma pergunta solicitando *feedback* sobre a solução de um exercício (“Vocês têm alguma dúvida sobre a lógica e a implementação do código para converter números romanos em inteiros?”), enquanto os demais eram perguntas geradas durante a resolução de um exercício. Essa análise dos erros de identificação possibilita possíveis melhorias na geração do *prompt* enviado para o GPT.

### 5.2.2 Classificação dos exercícios

Para todos os classificadores propostos foi-se utilizada a Base Bloom apresentada na Subseção 5.1.1, contendo 600 exercícios.

A abordagem baseada em regras (BR) faz uso direto de um conjunto suplementar de dados que inclui verbos associados a diversos níveis da Taxonomia de Bloom. Em suma, o exercício é comparado com esta lista de verbos até que se encontre uma correspondência, permitindo assim a determinação do nível cognitivo associado ao exercício em análise. A fim de aprimorar a precisão dessa técnica, os exercícios foram convertidos em vetores de palavras, e palavras contendo menos de 3 caracteres, bem como sinais de pontuação, foram removidos. Essa medida visou simplificar o processo de correspondência, uma vez que o segundo conjunto de dados não incluía verbos com menos de 3 caracteres. Como resultado,

Figura 8 - Matrizes de Confusão BEA



foi possível reduzir a complexidade da busca nos vetores de palavras. Posteriormente, procedeu-se com uma comparação entre as palavras presentes no exercício e os verbos contidos na base de verbos. Se houvesse correspondência, isso indicava que o exercício se referia a esse nível específico da Taxonomia de Bloom.

Esta abordagem possui implementação e interpretação descomplicadas, contudo, sua eficácia pode ser prejudicada pela precisão dos verbos escolhidos. Por exemplo, alguns exercícios podem não se alinhar claramente com nenhum dos verbos disponíveis no segundo conjunto de dados, o que pode ocasionar classificações imprecisas ou incorretas. Além disso, as relações semânticas entre os verbos e os níveis cognitivos da Taxonomia de Bloom também pode se apresentar como um desafio para esta abordagem.

Para a classificação com os métodos NB e BERT, inicialmente, a base foi dividida em duas partes: 498 exercícios foram destinados a validação, enquanto os outros 102 exercícios foram reservados para teste. A base de validação foi subdividida em conjuntos de treinamento e validação usando a técnica de validação cruzada 10-fold. Nesse método, o conjunto de dados é dividido em 10 partes iguais, sendo que em cada iteração do treinamento, 9 partes são usadas para treinar o modelo e 1 parte é reservada para validar seu desempenho. Esse processo é repetido 10 vezes para garantir que todas as partes do conjunto de dados sejam usadas para treinamento e validação, permitindo a comparação dos resultados.

No treinamento do Naive Bayes (NB), após a etapa de separação das bases de treinamento e validação, as amostras passam por uma etapa de vetorização para prepará-las para a execução do treinamento. Devido à natureza do problema, que envolve múltiplas respostas, foi utilizada a versão Multinomial do NB. Essa escolha permite identificar não apenas se um exercício pertence ou não a um determinado nível cognitivo em uma

classificação binária, mas sim determinar o nível cognitivo específico ao qual o exercício pertence.

Ao treinar o modelo BERT, é possível utilizar um modelo que tenha sido pré-treinado em uma grande quantidade de dados para modelagem de linguagem. Para a classificação dos exercícios, utilizou-se o modelo pré-treinado *bert-base-uncased*<sup>4</sup>, introduzido em 2018 por (31). Esse modelo foi treinado com dados da *Wikipedia* em inglês e do *Toronto Book Corpus*, que contém mais de 11 mil livros não publicados. Os textos são escritos em letras minúsculas, tokenizados usando a técnica *WordPiece*, e um vocabulário de tamanho 30 mil foi usado. O modelo utiliza uma combinação de objetivos de modelagem de linguagem, incluindo a aplicação de máscaras e a predição da próxima sentença do texto.

Para o modelo BERT, as *features* das instâncias foram *tokenizadas*, transformando as palavras em identificadores de *tokens*. Além disso, foi realizado um tratamento para limitar o tamanho das sentenças a 256 *tokens*, devido às limitações do *hardware*. Foram criadas máscaras de atenção para o modelo a partir dos *tokens*. As máscaras de atenção indicam quais *tokens* são relevantes para a classificação do exercício e quais são *paddings*, ou seja, *tokens* adicionais que são adicionados para que todas as sentenças tenham o mesmo tamanho. Dessa forma, os *paddings* foram criados de modo que, se um *token* estiver presente, a máscara de atenção receberá o valor 1; caso contrário, receberá o valor 0. O processo de treinamento fez uso do otimizador *AdamW* com *learning rate* de 2E-5 e  $\epsilon$  de 1e-8, e por fim o treinamento foi realizado em oito épocas.

A classificação com o GPT empregou a técnica de *fine-tuning*. Para este modelo, foi necessário adequar a base para o formato necessário para a execução do *fine-tuning*, que requer um arquivo *jsonl* como entrada. Após a conversão, a base foi separada em validação, contendo 498 exercícios, e na base de testes, com 102 exercícios para teste. A base de validação foi subdividida em treinamento e validação e, então, os arquivos foram carregados na plataforma da *OpenAI*, que realiza o *fine-tuning* para criar o novo modelo.

Finalizada a configuração dos classificadores, os resultados obtidos foram avaliados. O método NB obteve uma média de acurácia de 67% durante o treinamento e validação dos dados, com um desvio padrão excepcionalmente baixo de apenas 6%. Esta consistência sugere uma precisão considerável nas classificações realizadas pelo NB. Por outro lado, ao empregar o método BERT, alcançou-se uma média de 74% de acurácia, porém, acompanhada por um desvio padrão ligeiramente superior de 12%, indicando uma maior variabilidade nos resultados. Por fim, a implementação utilizando o GPT 3.5 com Fine-tuning destacou-se com uma média de 91% de acurácia, embora tenha apresentado o maior desvio padrão de 14%, sugerindo uma variação mais ampla nos resultados.

Após o treinamento, a base de teste foi usada para avaliar a classificação utilizando os três métodos: NB, BERT e GPT. Os resultados demonstraram que esses métodos

<sup>4</sup> <https://github.com/google-research/bert>



apresentaram desempenho mais satisfatório em comparação com a abordagem baseada em regras. Os modelos treinados foram salvos para uso posterior na etapa de classificação da base de exercícios, a qual foi construída a partir das videoaulas.

A Tabela 7 apresenta os resultados, além de trazer informações sobre o modelo apresentado em (96). Para medir a qualidade das predições, foi utilizado o Coeficiente de Correlação Mathews (*Matthews correlation coefficient* ou MCC), uma vez que há um desbalanceamento entre as classes. Este coeficiente é usado no aprendizado de máquina como uma medida da qualidade das classificações binárias e seu valor varia de -1 a +1. Um coeficiente de +1 representa uma predição perfeita, 0 uma predição aleatória média e -1 uma predição inversa. Posterior a isso, foram medidos também os valores de acurácia, precisão, revocação e o F1.

Para o valor do MCC foi obtido 23% para o BR 36% para o GPT e 65% para o NB enquanto que para o BERT esse valor foi de 84%. Os valores de acurácia macro para o BR, GPT e NB, foram respectivamente de 27%, 49% e 73%, enquanto que para o BERT foi de 88%. Verificando a acurácia micro os valores não se alteram muito para o GPT, NB e BERT, sendo 47%, 71% e 86%, porém para o BR a diferença foi maior, tendo obtido 34%.

A abordagem baseada em regras teve desempenho inferior comparado aos demais classificadores, tal resultado pode ser atribuído à presença de termos implícitos ou com múltiplos significados, dificultando a classificação precisa com essa abordagem. Essa conclusão é corroborada pela matriz apresentada na Tabela 1, que evidencia termos presentes em mais de um nível da TBD e que, embora tendam a se inclinar para um determinado nível, podem gerar confusão na classificação.

Foi possível identificar também que a abordagem GPT-3.5 Turbo *Fine-Tuning* teve um desempenho ruim se comparado ao NB e o BERT. Tal diferença pode ter ocorrido, porque o processo de *fine-tuning* foi feito utilizando a plataforma web da *OpenAI*, não sendo possível configurar certos hiperparâmetros, como por exemplo, o número de *epochs* a serem executados, tendo o processo sido feito com 3 *epochs*. Outro ponto a se considerar é o fato do GPT-3.5 contar com uma lista de 175 bilhões de parâmetros, o que pode acarretar um baixo desempenho do *fine-tuning* em uma base pública e conhecida, como é a Base Bloom.

Já método BERT apresentou um desempenho superior em relação aos demais métodos apresentados, na tarefa de classificação. O valor mais elevado do MCC para o BERT sugere uma correlação mais forte entre as classificações previstas e as observações reais em comparação com o NB.

Comparando o resultado do BERT com o modelo SVM apresentado por (96), há uma maior precisão de um conjunto de classes para cada abordagem, o que reflete em uma acurácia micro ligeiramente similar (87% para o SVM e 86% para a nossa abordagem com BERT). Embora a abordagem de (96) seja, para algumas classes, mais confiável predizer

Tabela 7 – Resultado do treinamento com a Base Bloom

| Método           | MCC         | Nível cognitivo da TBD | Precisão    | Revocação   | Valor F1    |
|------------------|-------------|------------------------|-------------|-------------|-------------|
| BR               | 0,22        | Lembrar                | 0,43        | 0,83        | 0,57        |
|                  |             | Compreender            | 0,37        | 0,48        | 0,42        |
|                  |             | Aplicar                | 0,57        | 0,67        | 0,61        |
|                  |             | Analisar               | 0,09        | 0,02        | 0,04        |
|                  |             | Avaliar                | 0,07        | 0,03        | 0,04        |
|                  |             | Criar                  | 0,08        | 0,10        | 0,09        |
|                  |             | Acurácia               | -           | -           | 0,34        |
|                  |             | Média macro            | 0,27        | 0,36        | 0,29        |
| NB               | 0,65        | Lembrar                | 0,83        | 0,59        | 0,69        |
|                  |             | Compreender            | 0,67        | 0,71        | 0,69        |
|                  |             | Aplicar                | 0,60        | 0,71        | 0,65        |
|                  |             | Analisar               | 0,93        | 0,76        | 0,84        |
|                  |             | Avaliar                | 0,72        | 0,76        | 0,74        |
|                  |             | Criar                  | 0,60        | 0,71        | 0,65        |
|                  |             | Acurácia               | -           | -           | 0,71        |
|                  |             | Média macro            | 0,73        | 0,71        | 0,71        |
| BERT             | <b>0,84</b> | Lembrar                | 1,00        | 1,00        | <b>1,00</b> |
|                  |             | Compreender            | 0,89        | 0,94        | <b>0,91</b> |
|                  |             | Aplicar                | 0,83        | 0,88        | <b>0,86</b> |
|                  |             | Analisar               | 0,93        | 0,82        | <b>0,87</b> |
|                  |             | Avaliar                | 0,68        | 0,88        | <b>0,77</b> |
|                  |             | Criar                  | 0,92        | 0,65        | <b>0,76</b> |
|                  |             | Acurácia               | -           | -           | 0,86        |
|                  |             | Média macro            | <b>0,88</b> | <b>0,86</b> | <b>0,86</b> |
| GPT-3.5 Turbo FT | 0,36        | Lembrar                | 0,62        | 0,94        | 0,74        |
|                  |             | Compreender            | 0,83        | 0,59        | 0,69        |
|                  |             | Aplicar                | 0,53        | 0,47        | 0,50        |
|                  |             | Analisar               | 0,79        | 0,65        | 0,71        |
|                  |             | Avaliar                | 0,09        | 0,12        | 0,10        |
|                  |             | Criar                  | 0,08        | 0,06        | 0,07        |
|                  |             | Acurácia               | -           | -           | 0,47        |
|                  |             | Média macro            | 0,49        | 0,47        | 0,47        |
| (96)             | N/D         | Lembrar                | 1,00        | 0,09        | 0,17        |
|                  |             | Compreender            | 0,50        | 0,25        | 0,33        |
|                  |             | Aplicar                | 1,00        | 0,08        | 0,14        |
|                  |             | Analisar               | 0,75        | 0,19        | 0,30        |
|                  |             | Avaliar                | 1,00        | 0,50        | 0,67        |
|                  |             | Criar                  | 0,90        | 0,64        | 0,75        |
|                  |             | Acurácia               | -           | -           | <b>0,87</b> |
|                  |             | Média macro            | 0,86        | 0,29        | 0,39        |

uma instância para aquela classe, a abordagem não consegue alcançar um equilíbrio aceitável entre precisão e revocação. De fato, em (96) os autores argumentam que os baixos valores de revocação e F1 se dão por conta do baixo volume de instâncias da base (apenas 600 registros). Por outro lado, percebe-se que a atual proposta consegue contornar esse problema, alcançando um F1 muito superior à abordagem anterior para todas as classes, o que resulta em uma maior acurácia macro (86% vs 39%).

Após treinamento e teste dos modelos, foram classificadas as bases BEM e BEA, utilizando os classificadores que obtiveram um desempenho satisfatório na etapa anterior, as Tabelas 8 e 9 apresentam os resultados obtidos.

Tabela 8 – Tabela de resultado da classificação BEM

| Método           | MCC         | Nível cognitivo da TBD | Precisão    | Revocação   | Valor F1    |
|------------------|-------------|------------------------|-------------|-------------|-------------|
| NB               | 0,23        | Lembrar                | 0,74        | 0,31        | 0,44        |
|                  |             | Compreender            | 0,06        | 0,33        | 0,10        |
|                  |             | Aplicar                | 0,52        | 0,59        | 0,55        |
|                  |             | Analisar               | 0,47        | 0,37        | 0,41        |
|                  |             | Avaliar                | 0,11        | 0,17        | 0,13        |
|                  |             | Criar                  | 0,08        | 0,33        | 0,12        |
|                  |             | Acurácia               | -           | -           | 0,42        |
|                  |             | Média macro            | 0,33        | 0,35        | 0,29        |
| BERT             | <b>0,64</b> | Lembrar                | 0,71        | 0,94        | <b>0,81</b> |
|                  |             | Compreender            | 0,60        | 1,00        | <b>0,75</b> |
|                  |             | Aplicar                | 0,82        | 0,70        | <b>0,76</b> |
|                  |             | Analisar               | 1,00        | 0,50        | <b>0,67</b> |
|                  |             | Avaliar                | 0,60        | 0,50        | <b>0,55</b> |
|                  |             | Criar                  | 0,67        | 0,67        | <b>0,67</b> |
|                  |             | Acurácia               | -           | -           | <b>0,76</b> |
|                  |             | Média macro            | <b>0,73</b> | <b>0,71</b> | <b>0,69</b> |
| GPT-3.5 Turbo FT | 0,48        | Lembrar                | 0,68        | 0,83        | 0,75        |
|                  |             | Compreender            | 0,50        | 0,33        | 0,40        |
|                  |             | Aplicar                | 0,73        | 0,59        | 0,65        |
|                  |             | Analisar               | 0,64        | 0,47        | 0,55        |
|                  |             | Avaliar                | 0,18        | 0,33        | 0,24        |
|                  |             | Criar                  | 0,50        | 0,33        | 0,40        |
|                  |             | Acurácia               | -           | -           | 0,65        |
|                  |             | Média macro            | 0,54        | 0,48        | 0,50        |

Os resultados para a base de exercícios manualmente identificados foram superiores aos obtidos automaticamente, sugerindo que a forma como os exercícios foram retornados pode não ser a mais adequada para a etapa posterior de classificação. O método Naive Bayes (NB) apresentou valores semelhantes em ambas as classificações, com MCC de 23% e 20% para as bases BEM e BEA, respectivamente. Quanto a acurácia houve uma divergência de apenas 3% entre elas, o que demonstra que os resultados, embora não se comparem aos obtidos com os outros 2 métodos, são consistentes.

Os resultados obtidos com os demais métodos, foram superiores aos de NB, porém é possível notar que, em ambos, não existe uma consistência similar ao NB entre os resultados obtidos da BEM e da BEA. Para o GPT-3.5 Turbo *Fine-Tuned*, o MCC reduziu de 48% para 35%, a acurácia de 65% para 53%. Já para o BERT, o MCC foi de 64% para 48% e a acurácia saiu de 76% para 65%. Por outro lado, os resultados dos demais métodos, GPT e BERT, foram superiores aos do NB, porém não exibiram a mesma consistência entre as classificações da BEM e da BEA. Para o GPT, o MCC diminuiu de

Tabela 9 – Tabela de resultado da classificação BEA

| Método           | MCC         | Nível cognitivo da TBD | Precisão    | Revocação   | Valor F1    |
|------------------|-------------|------------------------|-------------|-------------|-------------|
| NB               | 0,20        | Lembrar                | 0,68        | 0,36        | 0,47        |
|                  |             | Compreender            | 0,11        | 0,50        | 0,18        |
|                  |             | Aplicar                | 0,44        | 0,43        | 0,44        |
|                  |             | Analisar               | 0,33        | 0,36        | 0,34        |
|                  |             | Avaliar                | 0,10        | 0,20        | 0,13        |
|                  |             | Criar                  | 0,20        | 1,00        | 0,33        |
|                  |             | Acurácia               | -           | -           | 0,39        |
|                  |             | Média macro            | 0,31        | 0,47        | 0,32        |
| BERT             | <b>0,48</b> | Lembrar                | 0,68        | 0,81        | <b>0,74</b> |
|                  |             | Compreender            | 0,67        | 1,00        | <b>0,80</b> |
|                  |             | Aplicar                | 0,63        | 0,51        | <b>0,57</b> |
|                  |             | Analisar               | 0,67        | 0,43        | <b>0,52</b> |
|                  |             | Avaliar                | 0,50        | 0,60        | <b>0,55</b> |
|                  |             | Criar                  | 0,50        | 1,00        | <b>0,67</b> |
|                  |             | Acurácia               | -           | -           | <b>0,65</b> |
|                  |             | Média macro            | <b>0,61</b> | <b>0,73</b> | <b>0,64</b> |
| GPT-3.5 Turbo FT | 0,35        | Lembrar                | 0,65        | 0,36        | 0,46        |
|                  |             | Compreender            | 0,00        | 0,00        | 0,00        |
|                  |             | Aplicar                | 0,57        | 0,70        | 0,63        |
|                  |             | Analisar               | 0,40        | 0,71        | 0,51        |
|                  |             | Avaliar                | 0,50        | 0,40        | 0,44        |
|                  |             | Criar                  | 0,50        | 0,50        | 0,50        |
|                  |             | Acurácia               | -           | -           | 0,53        |
|                  |             | Média macro            | 0,44        | 0,45        | 0,42        |

48% para 35%, enquanto a acurácia caiu de 65% para 53%. No caso do BERT, o MCC foi de 64% para 48%, e a acurácia diminuiu de 76% para 65%.

Pode-se observar que o método BERT manteve um desempenho superior em comparação aos demais. Além disso, ao analisar ambas as bases, observa-se que os valores de precisão, revocação e F1-score do BERT são consistentes entre os diferentes níveis da TBD, demonstrando um certo equilíbrio. Este padrão é evidenciado pelo fato de que o método obteve uma acurácia superior a 65% em ambas as bases.

Em contraste, tanto o NB quanto o GPT enfrentaram dificuldades na classificação dos níveis com poucos exercícios identificados, como Compreender, Avaliar e Criar. Por exemplo, o método NB teve uma precisão máxima de 20% para o nível Criar da base BEA, enquanto o GPT obteve 0% para o nível Compreender da mesma base. O que destaca desafios específicos enfrentados por esses métodos ao lidar com dados escassos ou desbalanceados, evidenciando a necessidade de aprimoramento e ajustes para melhorar sua eficácia nessas situações.

## 6 CONCLUSÃO

Na literatura, embora se encontrem diversos estudos sobre a classificação de exercícios em mídias textuais, a mesma atenção não foi observada no contexto das videoaulas. O uso do vídeo como recurso educacional, tem crescido cada vez mais, e durante a pandemia de *COVID-19* houve uma rápida transição das universidades para o ensino online (63). Isso ressalta a importância de desenvolver interfaces e serviços avançados para maximizar o potencial das videoaulas, especialmente dada sua ampla aplicação em várias áreas do conhecimento (7). No entanto, a ausência de metadados para definir e classificar videoaulas apresenta um desafio significativo, dificultando a localização de conteúdo relevante durante sessões de aprendizagem (20). Portanto, surge a necessidade de abordagens que tratem essa lacuna, como a geração de novos metadados, incluindo a classificação automática de exercícios. Essa abordagem pode não apenas facilitar a organização e a recuperação de conteúdo, mas também permitir que sistemas de recomendação direcionem os alunos para materiais alinhados com seus objetivos educacionais.

Este trabalho apresentou um modelo para a classificação automatizada de exercícios, com base na Taxonomia Digital de Bloom. O processo começa com a transcrição de videoaulas para texto, o que, de acordo com a hipótese **H1**, auxilia na identificação dos exercícios. Os resultados obtidos utilizando o GPT foram satisfatórios, apesar do conjunto limitado de dados para treinamento, confirmando assim as hipóteses **H1** e **H2**. Esse modelo oferece uma estrutura organizada para a geração de metadados e a classificação do conteúdo das videoaulas de acordo com os princípios da Taxonomia Digital de Bloom.

A avaliação do modelo proposto foi realizada utilizando diversos algoritmos classificadores, onde o modelo BERT se destacou como superior aos demais. Os resultados obtidos demonstram que a classificação automatizada, utilizando aprendizado de máquina, tem a capacidade de distinguir efetivamente os diferentes níveis da Taxonomia de Bloom quando se é observado a semântica, confirmando a hipótese **H3**. Portanto, classificadores que se baseiam em semântica tendem a alcançar resultados superiores. Ao considerar as nuances semânticas dos exercícios, os classificadores podem capturar com mais precisão as características que diferenciam os diferentes níveis cognitivos na TBD, resultando em uma classificação mais acurada e confiável. Essa constatação ressalta a importância de explorar a semântica dos exercícios como um recurso valioso para aprimorar a classificação automática e, por consequência, o desempenho dos sistemas de recomendação educacional.

Além disso, essa classificação possibilita a incorporação de informações semânticas adicionais em videoaulas, permitindo que sistemas de recomendação educacionais ofereçam conteúdos mais personalizados e precisos de acordo com as necessidades de aprendizado do aluno. Isso abre caminho para uma maior eficácia no fornecimento de materiais educacionais alinhados aos objetivos de aprendizado do aluno.

Foram, também exploradas possíveis associações entre os estilos de videoaulas e os tipos de exercícios propostos, destacando-se uma clara afinidade entre os estilos *Voice Over Slides* e *Whiteboard* com exercícios LOTS. No entanto, não foi possível determinar se os demais estilos identificados (*Animation*, *Hand Drawing* e *Talking Head*) representam exercícios HOTS, devido à limitada amostragem disponível para análise.

Por fim, esta proposta de classificação automática demonstra uma boa capacidade de generalização, embora tenha sido limitada ao uso de videoaulas em língua inglesa neste estudo. Para sua aplicação em outros idiomas, como português ou espanhol, não basta apenas adquirir novas videoaulas nesses idiomas, também serão necessárias algumas adaptações durante a escolha da base pré-treinada do modelo BERT. Além disso, a implementação do modelo foi realizada utilizando o **Sistema de Gerenciamento de Workflows M2P**. Para sua implantação em ambiente de produção, ainda é necessário um esforço adicional na construção de códigos mais eficientes.

## 6.1 CONTRIBUIÇÕES

As principais contribuições deste trabalho incluem:

- Uma abordagem de classificação automática de exercícios, fundamentada na Taxonomia Digital de Bloom, que oferece novas possibilidades para aplicações em Informática na Educação. Isso permite a exploração do uso desses exercícios e a entrega de videoaulas mais pertinentes ao momento de aprendizagem do aluno.
- A integração do modelo proposto com o framework M2P.
- Um novo *dataset* com um conjunto de vídeos educativos com exercícios que permite replicar e avançar em pesquisas na área de extração de metadados de videoaulas. Até onde é de conhecimento do autor, este é o primeiro *dataset* anotado com exercícios de videoaulas.
- Um novo *dataset* com um conjunto de exercícios educativos contendo a classificação dos mesmos para os pesquisadores que permite replicar e avançar em pesquisas na área de classificação de exercícios baseado na Taxonomia de Bloom. Ressalta-se que foram encontrados poucos *datasets* públicos para utilização.
- Durante os estudos para esta dissertação, foram publicados dois artigos que contribuíram para o desenvolvimento deste trabalho, embora não estejam diretamente relacionados ao tema discutido. O primeiro artigo analisa o comportamento de usuários no contexto de *Search as Learning* (79), enquanto o segundo artigo apresenta um estudo de caso para analisar o processo de recuperação de informação e a busca como um processo de aprendizagem em um ambiente de busca sobre *COVID-19* (87).

## 6.2 LIMITAÇÕES

As limitações desse trabalho podem ser divididas em quatro partes: (i) Execução da transcrição, (ii) Identificação de exercícios, (iii) Classificação dos exercícios, e (iv) Relação entre exercícios e o estilo da videoaula.

**Execução da transcrição:** A escolha da ferramenta utilizada para transcrever o texto a partir do vídeo pode impactar o resultado final da abordagem e deve ser selecionada com base no idioma do vídeo. No entanto, mesmo com essa consideração, é possível que o resultado contenha elementos que possam dificultar a identificação dos exercícios.

**Identificação de exercícios:** A identificação dos exercícios foi baseada em um técnica de engenharia de prompt. No entanto, as escolhas feitas pelo autor podem não ser as mais adequadas para garantir os melhores resultados. Além disso, este trabalho depende do modelo GPT-3.5 Turbo, que possui 175 bilhões de parâmetros, uma quantidade considerada limitada em comparação com o modelo 4.0, que possui 1 trilhão de parâmetros.

**Classificação dos exercícios:** O modelo não foi submetido à avaliação por especialistas em áreas educacionais. Possíveis deficiências na definição da classificação manual proposta podem ser abordadas por meio de uma revisão do trabalho pela comunidade científica ou por meio de validação manual realizada por um especialista na área.

**Relação entre exercícios e o estilo da videoaula:** O escasso número de exercícios por estilo de videoaula pode ter levado a conclusões potencialmente equivocadas. É possível que estudos com conjuntos de dados mais abrangentes revelem resultados diferentes dos encontrados neste trabalho.

## 6.3 TRABALHOS FUTUROS

Este estudo proporciona um vasto campo para investigações subsequentes, pois até onde o autor tem conhecimento, esta é a primeira abordagem que trata da classificação de exercícios a partir de videoaulas, possibilitando a criação de metadados para esse tipo de conteúdo.

Foram apresentadas duas abordagens distintas para a classificação dos exercícios: uma baseada em regras e as demais baseadas em *machine learning*, que utilizam apenas o canal de áudio. No entanto, é relevante notar que os vídeos possuem múltiplos canais de comunicação. Abordagens híbridas, como a proposta em (38), mostram-se promissoras para alcançar resultados superiores. Por exemplo, a utilização da segmentação dos vídeos (89) permite a extração de características adicionais, que podem ser integradas a abordagens baseadas em regras e posteriormente combinadas com as abordagens de classificação automática apresentadas.

Determinar estilos relacionados às videoaulas, conforme abordado por (9), e identificar a correlação entre esses estilos e os exercícios, também pode oferecer ganhos para a

classificação. Através de abordagens híbridas, é possível determinar, por exemplo, que determinados tipos de videoaulas estão associados a uma maior incidência de exercícios de níveis cognitivos LOTS ou HOTS. Essa análise pode permitir ajustar o peso atribuído a cada nível cognitivo no classificador, melhorando assim sua capacidade de discriminação e precisão.

Outras abordagens, como o reconhecimento óptico de caracteres (OCR), podem agregar valor à classificação, especialmente ao identificar enunciados de exercícios que podem não ter sido verbalizados, mas apenas apresentados visualmente no vídeo.

Além disso, observar a classificação de exercícios em uma videoaula pode fornecer entendimentos sobre a complexidade da aula como um todo. Isso ocorre porque cada nível da Taxonomia de Bloom tem um objetivo distinto de aprendizagem, e os níveis são dependentes um do outro, ou seja, para alcançar o nível Criar, é necessário que o nível anterior tenha sido compreendido. A partir dessa informação, talvez seja possível classificar a videoaula com base no nível taxonômico dos exercícios apresentados, o que gera mais metadados ao vídeo.

Por fim, experimentos com conjuntos de dados mais amplos e diversificados em termos de domínio, como duração e conteúdo abordado, podem revelar que outros modelos de classificação venham a apresentar uma precisão superior a gerada pelo BERT.



## REFERÊNCIAS

- 1 ABDULRAHAMAN, M.; FARUK, N.; OLOYEDE, A.; SURAJUDEEN-BAKINDE, N.; OLAWOYIN, L.; MEJABI, O.; IMAM-FULANI, Y.; FAHM, A.; AZEEZ, A. Multimedia tools in the teaching and learning processes: A systematic review. **Heliyon**, [S.l.], v.6, n.11, 2020.
- 2 ALMEIDA, D. J. **Sequenciamento e recomendação de ações pedagógicas baseados na Taxonomia de Bloom Revisada utilizando PSO multiobjetivo**. 2023. 111p. Dissertação (Mestrado em Ciência da Computação) — Faculdade de Computação, Universidade Federal de Uberlândia, Uberlândia, Minas Gerais. 2023.
- 3 AMANE, M.; AISSAOUI, K.; BERRADA, M. Multi-Label Classification of Learning Objects Using Clustering Algorithms Based on Feature Selection. **International Journal of Emerging Technologies in Learning (iJET)**, [S.l.], v.17, n.20, p.pp. 248–260, Oct. 2022.
- 4 ANDERSON, L.; KRATHWOHL, D.; AIRASIAN, P.; CRUIKSHANK, K.; RICHARD, M.; PINTRICH, P.; J., R.; WITTROCK, M. **A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives**. [S.l.]: Addison Wesley Longman, Inc., 2001. 336p.
- 5 ANDROUTSOPOULOS, I.; KOUTSIAS, J.; CHANDRINOS, K. V.; PALIOURAS, G.; SPYROPOULOS, C. D. An Evaluation of Naive Bayes’ Classifier for Email Filtering. In: WORKSHOP ON MACHINE LEARNING IN THE NEW INFORMATION AGE, 2000. **Proceedings...** [S.l.: s.n.], 2000. p.26–33.
- 6 ANINDITYA, A.; HASIBUAN, M. A.; SUTOYO, E. Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom’s Taxonomy. In: IEEE INTERNATIONAL CONFERENCE ON INTERNET OF THINGS AND INTELLIGENCE SYSTEM (IOTAIS), 2019., 2019, Bali, Índia. **Anais...** IEEE, 2019. p.112–117.
- 7 AQUINO, B. **Classificação automática de estilos de videoaulas**. 2023. 54p. Dissertação (Mestrado em Ciência da Computação) — Faculdade de Computação, Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais. 2023.
- 8 AQUINO, B.; BARRÉRE, E.; SOUZA, J. F. de. Classificação Automática de Estilos de Videoaulas. In: XXXIII SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2022, Porto Alegre, RS, Brasil. **Anais...** SBC, 2022. p.956–967.
- 9 AQUINO, B.; SOUZA, J. F. de; BARRÉRE, E. Automatic Classification of Learning Material Styles. **Revista Brasileira de Informática na Educação**, [S.l.], v.31, p.906–924, nov. 2023.
- 10 BAEVSKI, A.; ZHOU, Y.; MOHAMED, A.; AULI, M. Wav2vec 2.0: A framework for self-supervised learning of speech representations. **Advances in neural information processing systems**, [S.l.], v.33, p.12449–12460, 2020.
- 11 BALASUBRAMANIAN, V.; DORAISAMY, S. G.; KANAKARAJAN, N. K. A multimodal approach for extracting content descriptive metadata from lecture videos. **Journal of Intelligent Information Systems**, [S.l.], v.46, p.121–145, 2016.

- 12 BARRÉRE, E.; ALMEIDA, M. Alvim de; APARECIDA VITOR, M.; SOUZA, J. Francisco de. Utilização de Enriquecimento Semântico para a Recomendação Automática de Videoaulas no Moodle. **Revista Brasileira de Informática na Educação**, [S.l.], v.28, n.1, 2020.
- 13 BARRÉRE, E.; SOUZA, J.; SOARES, E. R. Framework para segmentação temporal de vídeos educacionais. In: XXXI SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2020. **Anais...** [S.l.: s.n.], 2020. p.972–981.
- 14 BARRÉRE, E.; SOUZA, J.; VITOR, M. A.; ALMEIDA, M. A. de. Recomendação automática de videoaulas no Moodle. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE), 2018. **Anais...** [S.l.: s.n.], 2018. v.29, n.1, p.1613.
- 15 BAZZAN, A. L. Contribuições de aprendizado por reforço em escolha de rota e controle semafórico. **Estudos Avançados**, [S.l.], v.35, p.95–110, 2021.
- 16 BISHOP, C. M. **Pattern Recognition and Machine Learning**. [S.l.]: Springer, 2006.
- 17 BORGES, M. V. M.; REIS, J. C. dos. Semantic-enhanced recommendation of video lectures. In: 2019 IEEE 19TH INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES (ICALT), 2019. **Anais...** [S.l.: s.n.], 2019. v.2161, p.42–46.
- 18 BROWN, M.; MCCORMACK, M.; REEVES, J.; BROOK, D. C.; GRAJEK, S.; ALEXANDER, B.; BALI, M.; BULGER, S.; DARK, S.; ENGELBERT, N. et al. **2020 educause horizon report teaching and learning edition**. [S.l.]: Educause, 2020.
- 19 CAPUANO, N.; MARSELLA, M.; SALERNO, S. ABITS: An agent based Intelligent Tutoring System for distance learning. In: INTERNATIONAL WORKSHOP ON ADAPTIVE AND INTELLIGENT WEB-BASED EDUCATION SYSTEMS, ITS, 2000. **Proceedings...** [S.l.: s.n.], 2000.
- 20 CARVALHO, H. C. F. B.; DORÇA, F. A.; PITANGUI, C. G.; ASSIS, L. P. de; ANDRADE, A. V.; TRINDADE, E. A. C. Classificação automática de vídeos educacionais por meio de comentários apoiada por técnicas de Aprendizado de Máquina: uma análise experimental utilizando o Youtube. **Revista Brasileira de Informática na Educação**, [S.l.], v.30, p.419–448, 2022.
- 21 CARVALHO, H.; PITANGUI, C.; ASSIS, L.; VIVAS, A. Educavideos: Um sistema de recomendação de objetos de aprendizagem de vídeos educacionais do youtube. In: CONGRESSO BRASILEIRO DE ENSINO SUPERIOR A DIST ^ANCIA, XVII, 2020. **Anais...** [S.l.: s.n.], 2020.
- 22 CHEN, H.-T. M.; THOMAS, M. Effects of lecture video styles on engagement and learning. **Educational Technology Research and Development**, [S.l.], v.68, p.2147–2164, 2020.
- 23 CHOWDHURY, G. G. Natural language processing. **Annual review of information science and technology**, [S.l.], v.37, n.1, p.51–89, 2003.

- 24 CHURCHES, A. **Bloom's digital taxonomy**. [S.l.]: Australian School Library Association NSW Incorporated, 2010.
- 25 COELHO, S. A.; SOUZA, J. F. de. Anotação semântica de transcritos para indexação e busca de vídeos. In: CONFERÊNCIA IBERO AMERICANA (WWW/INTERNET).[GS SEARCH], 2014. **Anais...** [S.l.: s.n.], 2014.
- 26 COOPER, G. F.; HERSKOVITS, E. A Bayesian Method for the Induction of Probabilistic Networks from Data. **Machine Learning**, [S.l.], v.9, n.4, p.309–347, 1992.
- 27 CROOK, C.; SCHOFIELD, L. The video lecture. **The Internet and Higher Education**, [S.l.], v.34, p.56–64, 2017.
- 28 DE OLIVEIRA, F. K.; SANTANA, J. R.; OLIVEIRA PONTES, M. G. de. O vídeo como ferramenta educacional a partir de múltiplas plataformas. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE), 2010. **Anais...** [S.l.: s.n.], 2010. v.1, n.1.
- 29 DENG, X.; LI, Y.; WENG, J.; ZHANG, J. Feature selection for text classification: A review. **Multimedia Tools and Applications**, [S.l.], v.78, n.3, p.3797–3816, 2019.
- 30 DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, [S.l.], 2018.
- 31 DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. **BERT**: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- 32 ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, [S.l.], v.542, n.7639, p.115–118, 2017.
- 33 FERRAZ, A. P. d. C. M.; BELHOT, R. V. Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Gestão & produção**, [S.l.], v.17, p.421–431, 2010.
- 34 FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The Elements of Statistical Learning**. [S.l.]: Springer, 2001.
- 35 GANI, M. O.; AYYASAMY, R. K.; SANGODIAH, A.; FUI, Y. T. Bloom's Taxonomy-based exam question classification: The outcome of CNN and optimal pre-trained word embedding technique. **Education and Information Technologies**, [S.l.], p.1–22, 2023.
- 36 GOMES JR, J.; SOUZA, J.; BARRÉRE, E. Comparativo entre fontes de dados para anotação automática de videoaulas. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION (SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO-SBIE), 2017. **Anais...** [S.l.: s.n.], 2017. v.28, n.1, p.1127.
- 37 GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.

- 38 HARIS, S. S.; OMAR, N. BLOOM'S TAXONOMY QUESTION CATEGORIZATION USING RULES AND N-GRAM APPROACH. **Journal of Theoretical & Applied Information Technology**, [S.l.], v.76, n.3, 2015.
- 39 HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. [S.l.]: Springer Science & Business Media, 2009.
- 40 HUANG, R. **Educational technology a primer for the 21st century**. [S.l.]: Springer, 2019.
- 41 HUITT, W. Bloom et al.'s taxonomy of the cognitive domain. **Educational psychology interactive**, [S.l.], v.22, p.1–4, 2011.
- 42 HUNTER, J. **Multimedia Content and the Semantic Web: Standards, Methods, and Tools**. [S.l.]: John Wiley & Sons, 2005.
- 43 IBRAHIM, M. Implications of designing instructional video using cognitive theory of multimedia learning. **Critical questions in education**, [S.l.], v.3, n.2, 2012.
- 44 INMAN, J.; MYERS, S. Now Streaming: Strategies that improve video lectures. **IDEA Paper**, [S.l.], v.68, 2018.
- 45 JELINEK, F.; MERCER, R. L. Interpolated estimation of Markov source parameters from sparse data. **Proceedings of the Workshop on Pattern Recognition in Practice**, [S.l.], v.1, p.381–397, 1980.
- 46 KESAVARAJ, G.; SUKUMARAN, S. A study on classification techniques in data mining. In: ICCCNT), 2013., 2013. **Anais. . .** [S.l.: s.n.], 2013. p.1–7.
- 47 KHURANA, D.; KOLI, A.; KHATTER, K.; SINGH, S. Natural language processing: State of the art, current trends and challenges. **arXiv preprint arXiv:1708.05148**, [S.l.], 2017.
- 48 KÖRNER, C.; BENZ, D.; HOTH, A.; STROHMAIER, M.; STUMME, G. Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In: WORLD WIDE WEB, 19., 2010. **Proceedings. . .** [S.l.: s.n.], 2010. p.521–530.
- 49 KRATHWOHL, D. R. A revision of Bloom's taxonomy: An overview. **Theory into practice**, [S.l.], v.41, n.4, p.212–218, 2002.
- 50 KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Advances in neural information processing systems**, [S.l.], v.25, p.1097–1105, 2012.
- 51 LAWRENCE, R. Fundamentals of speech recognition. **AT&T**, [S.l.], 1993.
- 52 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, [S.l.], v.521, n.7553, p.436–444, 2015.
- 53 LEWIS, M.; LIU, Y.; GOYAL, N.; GHAZVININEJAD, M.; MOHAMED, A.; LEVY, O.; STOYANOV, V.; ZETTLEMOYER, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. **arXiv preprint arXiv:1910.13461**, [S.l.], 2019.

- 54 LIMA DIAS, L.; BARRÉRE, E.; SIQUEIRA BARBOSA, J.; SOUZA, J. F. de. Uma Abordagem para Identificação de Similaridade Entre Recursos Educacionais Utilizando Bases de Conhecimento Externas. **Revista Brasileira de Informática na Educação**, [S.l.], v.25, n.2, 2017.
- 55 LIU, Y.; OTT, M.; GOYAL, N.; DU, J.; JOSHI, M.; CHEN, D.; LEVY, O.; LEWIS, M.; ZETTLEMOYER, L.; STOYANOV, V. Roberta: A robustly optimized bert pretraining approach. **arXiv preprint arXiv:1907.11692**, [S.l.], 2019.
- 56 MAHESH, B. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**. [Internet], [S.l.], v.9, n.1, p.381–386, 2020.
- 57 MAKHLOUF, K.; AMOURI, L.; CHAABANE, N.; NAHLA, E.-H. Exam Questions Classification Based on Bloom’s Taxonomy: Approaches and Techniques. In: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION SCIENCES (ICCIS), 2020., 2020. **Anais. . .** [S.l.: s.n.], 2020. p.1–6.
- 58 MAYER, R. E.; MORENO, R. Animation as an aid to multimedia learning. **Educational psychology review**, [S.l.], v.14, p.87–99, 2002.
- 59 MCCALLUM, A.; NIGAM, K. A comparison of event models for Naive Bayes text classification. In: AAAI-98 WORKSHOP ON LEARNING FOR TEXT CATEGORIZATION, 1998. **Anais. . .** [S.l.: s.n.], 1998. v.752, p.41–48.
- 60 MOHAMMED, M.; OMAR, N. Question classification based on Bloom’s taxonomy cognitive domain using modified TF-IDF and word2vec. **PloS one**, [S.l.], v.15, n.3, p.e0230442, 2020.
- 61 MOREL, G. M.; SPECTOR, J. M. **Foundations of educational technology: Integrative approaches and interdisciplinary perspectives**. [S.l.]: Taylor & Francis, 2022.
- 62 MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**. [S.l.]: MIT Press, 2012.
- 63 NOETEL, M.; GRIFFITH, S.; DELANEY, O.; SANDERS, T.; PARKER, P.; POZO CRUZ, B. del; LONSDALE, C. Video improves learning in higher education: A systematic review. **Review of educational research**, [S.l.], v.91, n.2, p.204–236, 2021.
- 64 OLIVEIRA, A. L. **Perceptron Dilatação-Erosão Linear com Treinamento Baseado em Otimização DC: Aplicações em Problemas de Regressão e Classificação**. 2023. Tese (Doutorado em Ciência da Computação) — [sn].
- 65 OLIVEIRA JUCá, B. de. **Sistema de gerenciamento de workflows M2P**. 2023. 47p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) — , Juiz de Fora, Minas Gerais. 2023.
- 66 OMAR, N.; HARIS, S. S.; HASSAN, R.; ARSHAD, H.; RAHMAT, M.; ZAINAL, N. F. A.; ZULKIFLI, R. Automated analysis of exam questions according to Bloom’s taxonomy. **Procedia-Social and Behavioral Sciences**, [S.l.], v.59, p.297–303, 2012.
- 67 OREY, M. **Emerging perspectives on learning, teaching and technology**. [S.l.]: CreateSpace North Charleston, 2010.

- 68 OZAN, O.; OZARSLAN, Y. Video lecture watching behaviors of learners in online courses. **Educational Media International**, [S.l.], v.53, n.1, p.27–41, 2016.
- 69 PAL, S.; PRAMANIK, P. K. D.; MAJUMDAR, T.; CHOUDHURY, P. A semi-automatic metadata extraction model and method for video-based e-learning contents. **Education and Information Technologies**, [S.l.], v.24, p.3243–3268, 2019.
- 70 PETERS, M. E.; NEUMANN, M.; IYYER, M.; GARDNER, M.; CLARK, C.; LEE, K.; ZETTLEMOYER, L. Deep contextualized word representations. **arXiv preprint arXiv:1802.05365**, [S.l.], 2018.
- 71 PINTO, L. F. G. TEORIAS DE APRENDIZAGEM APLICADAS AO E-LEARNING: UMA ABORDAGEM DA TEORIA COGNITIVA DE APRENDIZAGEM MULTIMÍDIA. **Anais do CIET:EnPED:2020 - (Congresso Internacional de Educação e Tecnologias | Encontro de Pesquisadores em Educação a Distância)**, [S.l.], 2020.
- 72 POLYVYANYYY, A. Evaluation of a novel information retrieval model: eTVSM. **Master's thesis, Hasso Plattner Institut**, [S.l.], 2007.
- 73 POPOV, B.; KIRYAKOV, A.; MANOV, D.; KIRILOV, A.; OGNYANOFF, D.; GORANOV, M. Towards semantic web information extraction. In: HUMAN LANGUAGE TECHNOLOGIES WORKSHOP (ISWC'03), 2003. **Proceedings...** [S.l.: s.n.], 2003. p.1–22.
- 74 PÖTTKER, L. M. V.; FERNEDA, E.; MOREIRO-GONZÁLEZ, J. A. Mapeamento relacional entre padrões de metadados educacionais. **Perspectivas em Ciência da Informação**, [S.l.], v.23, p.25–38, 2018.
- 75 RADFORD, A.; KIM, J. W.; XU, T.; BROCKMAN, G.; MCLEAVEY, C.; SUTSKEVER, I. Robust speech recognition via large-scale weak supervision. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2023. **Anais...** [S.l.: s.n.], 2023. p.28492–28518.
- 76 RADFORD, A.; NARASIMHAN, K.; SALIMANS, T.; SUTSKEVER, I. et al. Improving language understanding by generative pre-training. , [S.l.], 2018.
- 77 RAZZAGHNOORI, M.; SAJEDI, H.; JAZANI, I. K. Question classification in Persian using word vectors and frequencies. **Cognitive Systems Research**, [S.l.], v.47, p.16–27, 2018.
- 78 REIS DIAS FONSECA, G. dos. **Audioteca Gerenciador de acervo de áudios**. 2019. 44p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) — , Juiz de Fora, Minas Gerais. 2019.
- 79 REIS DIAS FONSECA, G. dos; SOUZA, J. F. de; BARRERE, E. Predicting the User Intention in Web Search. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 2021. **Proceedings...** [S.l.: s.n.], 2021. p.133–136.
- 80 RUDOLPH, M. Cognitive theory of multimedia learning. **Journal of Online Higher Education**, [S.l.], v.1, p.1–10, 2017.

- 81 SAIRA, A. F., & Hafeez, M.(2020). Assessment of Student’s Academic Achievement by Flipped Classroom Model and Traditional Lecture Method. **Global Educational Studies Review**, [S.l.], v.4, p.10–19, 2020.
- 82 SANTOS-ESPINO, J. M.; AFONSO-SUÁREZ, M. D.; GUERRA-ARTAL, C. Speakers and boards: A survey of instructional video styles in MOOCs. **Technical Communication**, [S.l.], v.63, n.2, p.101–115, 2016.
- 83 SANTOSO, L. W.; SINGH, B.; RAJEST, S. S.; REGIN, R.; KADHIM, K. H. A genetic programming approach to binary classification problem. **EAI Endorsed Transactions on Energy Web**, [S.l.], v.8, n.31, p.e11–e11, 2021.
- 84 SETYANINGSIH, E. R.; LISTIOWARNI, I. Categorization of exam questions based on bloom taxonomy using naïve bayes and laplace smoothing. In: EAST INDONESIA CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY (EICONCIT), 2021., 2021. **Anais...** [S.l.: s.n.], 2021. p.330–333.
- 85 SEYMORE, K.; ROSENFELD, R.; CHEN, S.; ESKENAZI, M.; GOUVEA, E.; REDDY, R.; RAVISHANKAR, M.; SIEGLER, M.; STERN, R.; THAYER, E. The 1997 CMU Sphinx-3 English broadcast news transcription system. **Carnegie Mellon University**, [S.l.], 1998.
- 86 SILVA, G. S.; LUZ, L. P. d. et al. Aplicativo que gera enriquecimento semântico de dados abertos: Estudo de caso do metrô de São Paulo. **Revista Eletrônica eF@tec**, [S.l.], 2020.
- 87 SILVA, J. P. R.; FONSECA, G. R.; BARRERE, E.; MENDONÇA, F. M.; SOUZA, J. F. Recuperação da Informação e Pesquisa como Processo de Aprendizagem: um estudo de caso no contexto da Covid-19. **Frontiers of Knowledge Representation**, [S.l.], v.1, n.1, p.126–144, 2021.
- 88 SOARES, E. R. **Temporal Segmentation of Video Lectures**: a speech-based optimization framework. 2020. 79p. Dissertação (Mestrado em Ciência da Computação) — Faculdade de Computação, Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais. 2020.
- 89 SOARES, E. R.; BARRÉRE, E. An optimization model for temporal video lecture segmentation using word2vec and acoustic features. In: BRAZILLIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 25., 2019. **Proceedings...** [S.l.: s.n.], 2019. p.513–520.
- 90 SPECIA, L.; MOTTA, E. Integrating folksonomies with the semantic web. In: EUROPEAN SEMANTIC WEB CONFERENCE, 2007. **Anais...** [S.l.: s.n.], 2007. p.624–639.
- 91 TAVAKOLI, M.; ELIAS, M.; KISMIHÓK, G.; AUER, S. Metadata analysis of open educational resources. In: LAK21: 11TH INTERNATIONAL LEARNING ANALYTICS AND KNOWLEDGE CONFERENCE, 2021. **Anais...** [S.l.: s.n.], 2021. p.626–631.
- 92 TORFI, A.; SHIRVANI, R. A.; KENESHLOO, Y.; TAVVAF, N.; FOX, E. A. Natural language processing advancements by deep learning: A survey. **arXiv preprint arXiv:2003.01200**, [S.l.], 2020.

- 93 TURNBULL, D.; BARRINGTON, L.; TORRES, D.; LANCKRIET, G. Semantic annotation and retrieval of music and sound effects. **IEEE Transactions on Audio, Speech, and Language Processing**, [S.l.], v.16, n.2, p.467–476, 2008.
- 94 WOOLF, B. P. **Building intelligent interactive tutors**: Student-centered strategies for revolutionizing e-learning. [S.l.]: Morgan Kaufmann, 2010.
- 95 YAHYA, A. A.; TOUKAL, Z.; OSMAN, A. Bloom’s Taxonomy–Based Classification for Item Bank Questions Using Support Vector Machines. In: MODERN ADVANCES IN INTELLIGENT SYSTEMS AND TOOLS, 2012, Berlin, Heidelberg. **Anais...** Springer Berlin Heidelberg, 2012. p.135–140.
- 96 YAHYA, A. Automatic Classification of Questions into Bloom’s Cognitive Levels using Support Vector Machines. In: THE INTERNATIONAL ARAB CONFERENCE ON INFORMATION TECHNOLOGY, 2011. **Anais...** [S.l.: s.n.], 2011.
- 97 YANG, H.; MEINEL, C. Content based lecture video retrieval using speech and video text information. **IEEE transactions on learning technologies**, [S.l.], v.7, n.2, p.142–154, 2014.
- 98 YANG, Z.; DAI, Z.; YANG, Y.; CARBONELL, J.; SALAKHUTDINOV, R.; LE, Q. V. Xlnet: Generalized autoregressive pretraining for language understanding. **arXiv preprint arXiv:1906.08237**, [S.l.], 2019.
- 99 ZHANG, H.; LEE, W. S. A Simple and Effective Algorithm for Unsupervised Learning of Verb Taxonomies. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2003., 2003. **Proceedings...** [S.l.: s.n.], 2003. p.241–248.
- 100 ZHANG, S.; YAO, L.; SUN, A.; TAY, Y. Deep learning based recommender system: A survey and new perspectives. **ACM Computing Surveys (CSUR)**, [S.l.], v.52, n.1, p.1–38, 2019.

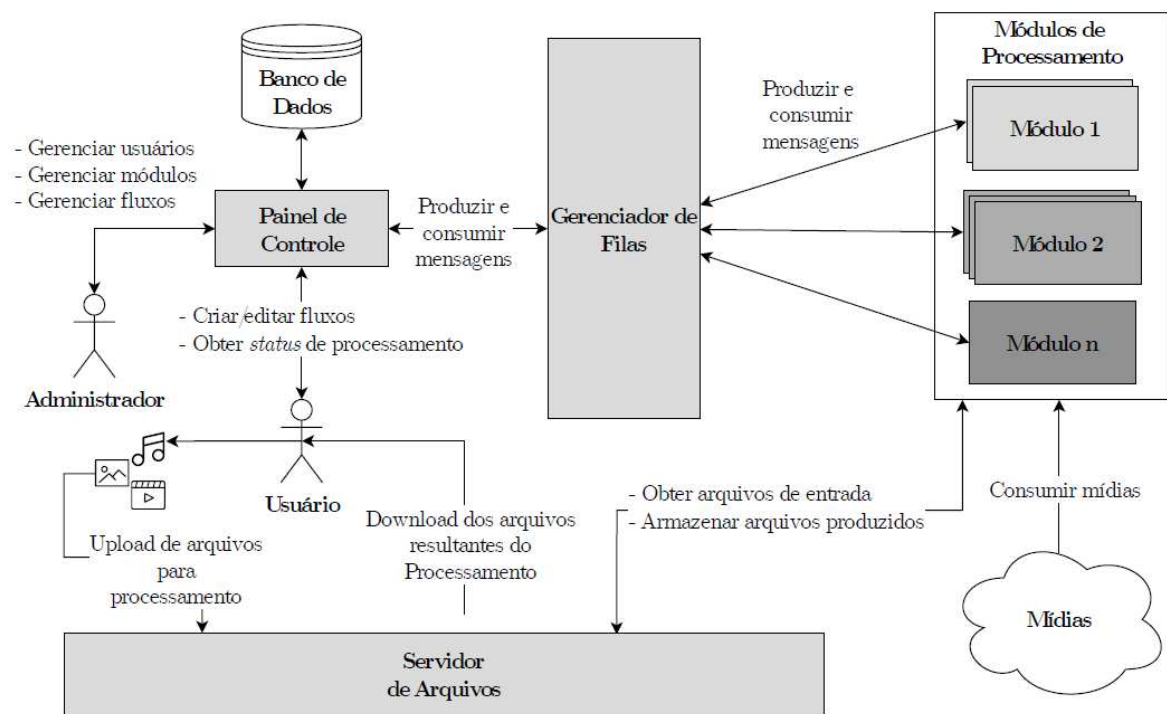


## APÊNDICE A – Framework M2P

O *framework* denominado **M2P**, foi desenvolvido por (65) no laboratório LAPIC, para controle de módulos dentro de um fluxo de trabalho (*workflow*). Esse sistema tem como finalidade administrar um *workflow*, que consiste em um conjunto de módulos, gerenciando as entradas e saídas de cada um deles e culminando na geração dos resultados desejados desse mesmo *workflow*. O *framework* permite a construção de *workflows* com base nos módulos nele disponíveis e também a criação de novos módulos conforme necessário.

A Figura 8 apresenta a arquitetura do *framework*.

Figura 2 - Arquitetura do framework



Fonte: Elaborada por (65)

O Framework tem sido utilizado para adaptar soluções de atribuição automática de metadados em videoaulas e, além da proposta apresentada neste trabalho, ele possui outros três *workflows* distintos. O primeiro é o OCR, um *workflow* simples composto por apenas um módulo, o de Reconhecimento Óptico de Caracteres (OCR). Este módulo recebe uma imagem como entrada e gera como saída os textos presentes na imagem.

O segundo *workflow*, denominado *Easytopic Workflow*, foi proposto por (13) em sua pesquisa. Ele é composto por 5 módulos diferentes. Ao receber um vídeo como entrada, o primeiro módulo realiza a extração do áudio. Em seguida, o segundo módulo detecta

atividades de voz no áudio e envia a saída para dois módulos subsequentes: um para transcrição do áudio e outro para extração de características de baixo nível. Por fim, ambos os resultados obtidos são utilizados como entrada para o último módulo, responsável pela segmentação em tópicos do vídeo.

Já o terceiro, apresentado em (9, 7), possui 4 módulos. O *workflow* inicia com a videoaula como entrada sendo recebida pelos 2 primeiros módulos de transcrição e detecção de pessoa no vídeo. Finalizadas essas etapas os resultados são enviados para o terceiro módulo de criação de *features* e por fim o resultado é enviado para o módulo de treinamento do modelo, que possui como saída o estilo da videoaula.