**Classificação automática de questões baseada em
competências: ENEM - Estudo de caso**

Valtemir de Alencar e Silva

Tese de Doutorado do Programa de Pós-Graduação em Ciências de
Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Valtemir de Alencar e Silva

Classificação automática de questões baseada em competências: ENEM - Estudo de caso

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. José Carlos Maldonado

Coorientador: Prof. Dr. Ig Ibert Bittencourt

USP – São Carlos
Maio de 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S586c Silva, Valtemir de Alencar e
Classificação automática de questões baseada em
competências: ENEM - Estudo de caso / Valtemir de
Alencar e Silva; orientador José Carlos Maldonado;
coorientador Ig Ibert Bittencourt. -- São Carlos,
2020.
158 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2020.

1. Classificação de questões. 2. Aprendizado
de Máquina. 3. Modelo espaço-vetorial. 4.
Representação de textos em rede. 5. ENEM. I. Carlos
Maldonado, José, orient. II. Ibert Bittencourt, Ig,
coorient. III. Título.

Valtemir de Alencar e Silva

**Competency-based automatic question classification: ENEM
- Case study**

Thesis submitted to the Institute of Mathematics and
Computer Sciences – ICMC-USP, in accordance with
the requirements of the Computer and Mathematical
Sciences Graduate Program for the degree of
Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. José Carlos Maldonado

Co-advisor: Prof. Dr. Ig Ibert Bittencourt

**USP – São Carlos
May 2020**

*Às pessoas que dedicam sua vida à promoção dos conceitos,
e extinção dos preconceitos.*

AGRADECIMENTOS

Agradeço inicialmente a minha família (pai, mãe, filho e irmãos) pelo apoio e amor incondicional de todas as horas.

No âmbito profissional, meu orientador José C. Maldonado e Ig Ibert foram essenciais em apontar os caminhos corretos a seguir e exemplos marcantes da excelência profissional. Também agradeço o IFSP (Instituto federal de São Paulo), meu local de trabalho e que me proporcionou um período de afastamento remunerado para elaboração deste trabalho. Da mesma forma, meus colegas/amigos de trabalho foram importantes no suporte tanto no período de saída para o afastamento, quanto no período de retorno.

Um agradecimento especial ao Rafael Rossi pelo excelente suporte científico e ao pessoal do NEES da UFAL em Alagoas pela ajuda essencial na troca de informações e conhecimento para a realização deste trabalho.

Por fim, agradeço a minha namorada/companheira/amante/amiga Irani Almeida pelo incentivo e carinho.

*“Consciência crítica,
pensamento transformador.”
(Maria Nilde Mascellani)*

RESUMO

SILVA, V. A. **Classificação automática de questões baseada em competências: ENEM - Estudo de caso.** 2020. 155 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Introdução: A grande quantidade de informação digital em formato textual existente na internet transforma em atividades essenciais a organização, análise e extração de conhecimento, tanto no universo acadêmico quanto no mercado de trabalho e, com isso, a classificação automática de textos vêm ganhando cada vez mais destaque. A classificação de questões é um subgrupo da classificação de textos e consiste, basicamente, da associação de um ou mais rótulos a cada questão, de acordo com um critério pré-determinado, mas com menos texto disponível que os documentos em geral. As principais aplicações dos sistemas de classificação automática de questões são: QA (*Question/Answering*), IR (*Information Retrieval*), Ambiente educacional e tratamento de linguagens específicas. Os sistemas de QA e RI têm como ponto de partida uma pergunta escrita em linguagem natural e, a partir daí, buscam uma coleção de documentos na *Web* que sejam compatíveis com o assunto descrito. Considerando especificamente o ambiente educacional, a geração automática de testes de avaliação tem aplicação prática imediata nos sistemas de *e-learning* por viabilizar a personalização do ensino através da busca de questões adequadas a um determinado perfil de aprendizado, os chamados sistemas de ensino adaptativo. Para viabilizar a personalização, é essencial classificar as questões dentro de uma variedade representativa de competências e habilidades. As avaliações em larga escala (ENEM, SAEB, Prova Brasil) podem ser uma fonte de informação para essa geração, pois utilizam Matrizes de Referência de Avaliação (MRA) para classificar as questões de acordo com as áreas do conhecimento, disciplinas, competências e habilidades esperadas dos estudantes. Uma forma de realizar esta classificação é através de algoritmos de Aprendizado de Máquina, capazes de extrair padrões ou generalizar classes através de geração de modelos matemáticos a partir dos dados disponíveis. São exemplos de algoritmos de Aprendizado de Máquina: redes neurais, árvores de decisão, máquinas de vetor de suporte *SVM*, naive bayes, *KNN*, entre outros. As diferentes formas de representação de texto para algoritmos de Aprendizado de Máquina possuem vasta pesquisa realizada quando se trata da classificação de documentos com grandes quantidades de texto; quando se trata de pequenos trechos (caso das questões), essa tarefa se torna mais complexa pois a quantidade de texto disponível para análise é reduzida em comparação aos outros tipos de documentos textuais. Além disso, a imensa maioria das pesquisas atuais se voltam para o problema de QA ou IR e há pouca pesquisa disponível considerando o ambiente educacional.

Objetivos: (i) Identificar a arquitetura de um classificador ou conjunto de classificadores de forma a maximizar o desempenho do processo de classificação de questões no contexto educacio-

nal; (ii) realizar uma avaliação empírica para comparar o desempenho das diferentes combinações utilizadas; (iii) disponibilizar as representações, algoritmos, códigos-fonte e ferramentas desenvolvidos para a comunidade científica avaliar e replicar resultados e (iv) disponibilizar ferramentas de integração e aplicação dos conteúdos desenvolvidos para utilização por outras plataformas e instituições (escolas, empresas) interessadas em utilizar as competências do ENEM para realizar diagnósticos e acompanhamento de aprendizagem.

Método: Estudo de caso com base em cerca de 25 mil questões pré-processadas retiradas de avaliações do ENEM até 2017 e classificadas por especialistas em educação dentro das disciplinas, competências e habilidades contidas na matriz de avaliação. A partir disso, foram realizados experimentos com classificadores gerados pelas diferentes combinações de representações de texto, cálculos de peso dos termos e algoritmos de Aprendizado de Máquina supervisionados produzindo, ao final, um comparativo com os indicadores de desempenho.

Resultados: (i) modelo geral para construção de classificadores automáticos de questões voltadas a avaliações em larga escala; (ii) identificação do estado da arte na classificação de questões para contexto educacional e avaliações em larga escala; (iii) estado da arte nos resultados de classificação de questões em redes no contexto educacional; (iv) efetividade do método $IQF * QF * ICF$ para o cálculo de pesos/frequências na classificação de questões; (v) banco de dados constantemente atualizável composto de questões classificadas e representativas do ENEM e suas competências; (vi) API para integração com outras plataformas das funcionalidades e (vii) análise de desempenho com comparativo dos indicadores gerados pelos classificadores de questões.

Conclusões: Com base nos resultados obtidos, a combinação da representação de textos no modelo espaço-vetorial, técnica $IQF * QF * ICF$ e o algoritmo KNN , se mostrou a melhor arquitetura para o classificador de questões do ENEM com desempenho geral médio em torno de 80% de acurácia. Destacam-se como principais contribuições: i) apoio às plataformas de ensino adaptativo; ii) ferramentas de classificação para as editoras de material didático; iii) apoio a professores e equipes pedagógicas na elaboração de avaliações diagnósticas contínuas sobre as facilidades e dificuldades de seus alunos. Relativamente aos trabalhos futuros, pretende-se gerar classificadores específicos para os exames Prova Brasil e SAEB, integrar a solução com um LMS de código aberto (Moodle, por exemplo) e desenvolver um algoritmo para conversão das questões do ENEM em imagens para classificação via rede neural convolucional.

Palavras-chave: Classificação de questões, Aprendizado de máquina, Modelo espaço-vetorial, Representação de textos em rede, ENEM.

ABSTRACT

SILVA, V. A. **Competency-based automatic question classification: ENEM - Case study.** 2020. 155 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Introduction: The large amount of digital textual information available on the Internet makes the organization, analysis and extraction of knowledge essential both in the academic world and in the job market, making automatic text classification increasingly important. Question classification is a subgroup of text classification and basically consists of associating one or more labels with each question, according to a predetermined criterion, but with less text available than the general documents. The main applications of automatic question classification systems are: QA (Question/Answering), IR (Information Retrieval), educational environment, and specific languages processing. The QA and IR systems have as their starting point a question written in natural language and, from there, search a collection of documents in the web that are compatible with the subject described. Considering specifically the educational environment, the automatic generation of assessment tests has immediate practical application in e-learning systems by enabling the personalization of teaching through the search for questions that are appropriate to a particular learning profile, the so-called adaptive learning systems. To enable personalization, it is essential to classify questions within a representative range of appropriate competencies and skills. Large-scale evaluations (ENEM, SAEB, Prova Brasil) could be a source of information for this generation, as they use evaluation reference matrices to classify questions according to the areas of knowledge, disciplines, competencies and expected skills of students. One way to perform this classification is through Machine Learning algorithms that are able to extract patterns or generalize classes by generating mathematical models from the available data. Examples of Machine Learning algorithms are: neural networks, decision trees, support vector machines (SVM), naive bayes, among others. The different forms of text representation and Machine Learning algorithms have extensive research done when it comes to classifying documents with large amounts of text; when it comes to short excerpts (such as questions), this task becomes more complex because the amount of text available for analysis is reduced when compared to other types of textual documents. In addition, the majority of current research addresses the problem of QA or IR, and there is not a lot of research available considering the educational environment.

Objectives: (i) Identify the architecture of a classifier or set of classifiers in order to maximize the performance of the question classification process in the educational context; (ii) perform an empirical evaluation to compare the performance of the different combinations used; (iii) make available representations, algorithms, source codes and tools developed for the scientific community to evaluate and replicate results; and (iv) make available tools for integration and

application of content developed for use by other platforms and institutions (schools, companies) interested in use ENEM skills to perform diagnostics and learning monitoring.

Method: Case study based on about 25,000 preprocessed questions taken from ENEM assessments until 2017 and rated by education experts within the disciplines, competencies and skills contained in the assessment matrix. From this, experiments were performed with classifiers generated by the different combinations of text representations, term weighting calculations, and supervised Machine Learning algorithms producing, in the end, a comparison with the performance indicators.

Results: (i) general model for building automatic classifiers of large-scale assessments questions; (ii) state of the art identification in the question classification of educational context and large scale assessments; (iii) state of the art in the results of question classifying based in network text representation in the educational context; (iv) effectiveness of the $IQF * QF * ICF$ method for term-weighting in the question classification tasks; (v) constantly updated database composed of classified and representative questions of ENEM and its competencies; (vi) API for integration with other functionality platforms; and (vii) performance analysis comparing the indicators generated by the question classifiers.

Conclusions: Based on the results obtained, the combination of the text representation in the space-vector model, technique $IQF * QF * ICF$ and the algorithm KNN , proved to be the best architecture for the ENEM queton classifier with average overall performance. around 80 % accuracy. The main contributions include: (i) support for adaptive education platforms; ii) rating tools for educational material publishers; iii) support to teachers and pedagogical teams in the elaboration of continuous diagnostic evaluations about the facilities and difficulties of their students. For future work, we intend to generate specific classifiers for the Prova Brasil and SAEB exams, integrate the solution with an open source LMS (Moodle, for example) and develop an algorithm for converting ENEM questions into images and perform classification using a convolutional neural network.

Keywords: Question classification, Machine learning, Vector space model, Network text representation, ENEM.

LISTA DE ILUSTRAÇÕES

Figura 1 – Competências e habilidades	30
Figura 2 – Categorias definidas por Li e Roth	38
Figura 3 – Estágios do processo de classificação	39
Figura 4 – Representação de um espaço vetorial	41
Figura 5 – Redes homogêneas e heterogêneas	43
Figura 6 – Rede documento-termo	43
Figura 7 – Lista de adjacências para uma rede homogênea	44
Figura 8 – Tipos de termos	45
Figura 9 – <i>Support Vector Machine</i>	49
Figura 10 – Redes Neurais: (a) Perceptron e (b) MLP	50
Figura 11 – Fluxo de aplicação dos critérios de inclusão e exclusão	57
Figura 12 – Referências mais citadas nos estudos selecionados	58
Figura 13 – Quantidade de publicações através dos anos	64
Figura 14 – Distribuição de estudos por país	65
Figura 15 – Distribuição de estudos por base de questões	66
Figura 16 – Distribuição de estudos por critérios/taxonomia	69
Figura 17 – Algoritmo para aplicação das regras de classificação	83
Figura 18 – Exemplo de matriz de referência de avaliação (Prova Brasil)	86
Figura 19 – Competências da matriz de referência do ENEM	88
Figura 20 – Esquema ilustrativo da classificação automática de textos por meio de aprendizado supervisionado	89
Figura 21 – Foco da pesquisa	90
Figura 22 – Modelo de estruturação e aplicação de classificadores baseado em competências para avaliações em larga escala	92
Figura 23 – Exemplo de funcionamento do algoritmo <i>IMBHN</i>	98
Figura 24 – Aplicação do modelo geral para o caso das questões do ENEM	102

LISTA DE QUADROS

Quadro 1 – Taxonomia de Bloom	51
Quadro 2 – Questões de pesquisa	54
Quadro 3 – Seleção de estudos para grupo de controle	55
Quadro 4 – String de busca	55
Quadro 5 – Critérios de inclusão/exclusão	56
Quadro 6 – Formulário para extração de dados	63
Quadro 7 – Seis categorias do domínio cognitivo baseadas na taxonomia de Bloom	81

LISTA DE ALGORITMOS

Algoritmo 1 – IMBHN (Adaptado)	97
--	----

LISTA DE TABELAS

Tabela 1 – Matriz documento-termo	41
Tabela 2 – Matriz de adjacências	44
Tabela 3 – Tabela de contingência entre f e c	48
Tabela 4 – Matriz de contingência para o processo de classificação (Adaptado)	52
Tabela 5 – Questões para avaliação de qualidade	58
Tabela 6 – Estudos selecionados pela avaliação de qualidade	59
Tabela 7 – Distribuição dos algoritmos unitários	67
Tabela 8 – Distribuição dos algoritmos nos estudos com combinações	68
Tabela 9 – Identificação dos estudos em cada taxonomia	69
Tabela 10 – Estudos por tipo de extração de termos	70
Tabela 11 – Estudos por tipo de seleção de termos	72
Tabela 12 – Métricas utilizadas nos estudos	74
Tabela 13 – Distribuição da base de questões por Área e Disciplina	103
Tabela 14 – Estado da arte	106
Tabela 15 – Classificação geral por disciplina	106
Tabela 16 – Classificação geral por competência	107
Tabela 17 – Resultado final a partir da combinação dos classificadores por disciplina e suas competências	109
Tabela 18 – Classificação geral por disciplina das versões do algoritmo <i>IMBHN</i>	109
Tabela 19 – Classificação geral por competência das versões do algoritmo <i>IMBHN</i> in- cluindo melhores resultado da <i>baseline</i>	110

LISTA DE ABREVIATURAS E SIGLAS

X^2	<i>Chi-Square</i>
AM	Aprendizado de Máquina
ANN	<i>Artificial Neural Network</i>
API	<i>Application Program Interface</i>
BNB	<i>Bernoulli Naive Bayes</i>
BOW	<i>Bag-of-Words</i>
CNB	<i>Complement Naive Bayes</i>
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
DF	<i>Document Frequency</i>
ENEM	Exame Nacional do Ensino Médio
FUVEST	Fundação Universitária para o Vestibular
GNB	<i>Gaussian Naive Bayes</i>
IG	<i>Information Gain</i>
IMBHN	Inductive Model Bipartite Heterogeneous Network
IR	<i>Information Retrieval</i>
KNN	<i>K-Nearest Neighbor</i>
LMS	Learning Management System
ME	Máxima Entropia
MEV	Modelo Espaço-Vetorial
MI	<i>Mutual Information</i>
MLP	<i>MultiLayer Perceptron</i>
MNB	<i>Multinomial Naive Bayes</i>
OR	<i>Odds Ratio</i>
POS	<i>Part-of-Speech</i>
QA	<i>Question Answering</i>
RS	Revisão Sistemática
SAEB	Sistema de Avaliação de Educação Básica
SVM	<i>Support Vector Machine</i>
TF	<i>Term Frequency</i>
TF*ICF	<i>Inverse Class Frequency</i>
TF*IDF	<i>Inverse Document Frequency</i>

TREC	<i>Text REtrieval Conference</i>
TRI	Teoria de Resposta ao Item

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Contexto	29
1.2	Motivação	32
1.3	Hipóteses	34
1.4	Objetivos	34
1.5	Resultados	35
1.6	Organização do Trabalho	36
2	REVISÃO DA LITERATURA	37
2.1	Considerações Iniciais	37
2.2	Classificação de Questões	37
2.2.1	<i>Representação de Textos</i>	39
2.2.1.1	<i>Modelo Espaço-Vetorial</i>	40
2.2.1.2	<i>Redes</i>	42
2.2.2	<i>Pré-Processamento</i>	44
2.2.3	<i>Extração de termos</i>	45
2.2.4	<i>Term-Weighting</i>	46
2.2.5	<i>Seleção de termos</i>	47
2.2.6	<i>Algoritmos</i>	48
2.2.6.1	<i>KNN</i>	48
2.2.6.2	<i>Árvores de Decisão</i>	48
2.2.6.3	<i>Naïve Bayes</i>	49
2.2.6.4	<i>SVM</i>	49
2.2.6.5	<i>Redes Neurais</i>	50
2.2.7	<i>Taxonomias</i>	50
2.2.8	<i>Indicadores de Desempenho</i>	51
2.3	Revisão Sistemática	52
2.3.1	<i>Questões de Pesquisa</i>	53
2.3.2	<i>Seleção de Fontes e Estudos</i>	54
2.3.3	<i>Avaliação de Qualidade</i>	58
2.3.4	<i>Extração de Dados</i>	63
2.3.5	<i>Análise dos Resultados</i>	64
2.3.5.1	<i>Ano de Publicação</i>	64

2.3.5.2	<i>País</i>	65
2.3.5.3	<i>Base de Questões e Meios de Publicação</i>	65
2.3.5.4	<i>QP1: Métodos Computacionais</i>	66
2.3.5.5	<i>QP2: Taxonomias para Classificação</i>	68
2.3.5.6	<i>QP3: Extração e Seleção dos termos</i>	70
2.3.5.7	<i>QP4: Métricas de Resultados</i>	74
2.4	Considerações finais	76
3	CLASSIFICAÇÃO AUTOMÁTICA DE QUESTÕES COM BASE NAS COMPETÊNCIAS DO ENEM	79
3.1	Considerações Iniciais	79
3.2	Trabalhos Relacionados	80
3.2.1	<i>Trabalhos Relacionados Baseados na Taxonomia de Bloom</i>	80
3.2.2	<i>Outros Trabalhos</i>	83
3.3	Matriz de Referência de Avaliação do ENEM	85
3.4	Propostas para Classificação de Questões Baseada em Competências do ENEM	89
3.4.1	<i>Modelo Geral</i>	90
3.4.2	<i>Método</i>	93
3.4.3	<i>Geração da Baseline</i>	94
3.4.4	<i>Algoritmo IMBHN^{COS}</i>	96
3.4.4.1	<i>Algoritmo IMBHN</i>	96
3.4.4.2	<i>Exemplo de Funcionamento</i>	97
3.4.4.3	<i>Otimização da Classificação através da Medida do Cosseno</i>	99
3.5	Considerações Finais	100
4	AVALIAÇÃO EXPERIMENTAL	101
4.1	Considerações Iniciais	101
4.2	Configuração	102
4.3	Resultados	104
4.3.1	<i>Resultados para o algoritmo IMBHN^{COS}</i>	109
4.4	Considerações Finais	110
5	CONCLUSÕES	113
5.1	Resultados e Contribuições	113
5.1.1	<i>Resultados</i>	114
5.1.2	<i>Contribuições</i>	115
5.2	Limitações e Trabalhos Futuros	115
	REFERÊNCIAS	117

ANEXO A	MATRIZ DE REFERÊNCIA DO ENEM	131
----------------	---	------------

INTRODUÇÃO

1.1 Contexto

A grande quantidade de informação digital existente na internet, especialmente em formato textual, transforma em atividades essenciais a organização, análise e extração de conhecimento, tanto no universo acadêmico quanto no mercado de trabalho. Uma parte considerável dos dados está no formato textual, como *e-mails*, relatórios, boletins, artigos, registros de pacientes e conteúdo de páginas *web* e utilizar profissionais treinados para classificá-los manualmente implica alto gasto de recursos (humano, tempo, por exemplo). Essa dificuldade acaba impedindo a aplicação do processo em outros locais e domínios e, como consequência, fomenta a busca de alternativas mais viáveis como classificação automática de textos (WEISS; INDURKHYA; ZHANG, 2015; AGGARWAL; ZHAI, 2012).

Dentro do universo da classificação de documentos textuais, há um subgrupo mais restrito chamado classificação de questões que corresponde, basicamente, à associação das questões a um rótulo (*label*) de acordo com um critério pré-determinado. O problema se torna complexo, pois a quantidade de informação disponível numa questão é bem menor quando comparada à dos textos presentes nos documentos em geral (LI; ROTH, 2002). As principais aplicações da classificação automática de questões são (SANGODIAH; MUNIANDY; HENG, 2015):

1. **Question Answering (QA):** aplicação primária da classificação de questões, a partir de questões formuladas em linguagem natural, visa a recuperar um conjunto de documentos associados e encontrar as respostas mais compatíveis. Exemplo: *Yahoo Answers*.
2. **Information Retrieval (IR):** semelhante à QA, mas foca apenas na recuperação dos documentos. Exemplo: motor de busca do Google.
3. **Ambiente educacional (e-learning):** recuperação de questões (normalmente de múltipla escolha) em categorias de competências ou níveis de dificuldade para testes de avaliação

cognitiva. Exemplo: classificadores via taxonomia de Bloom.

4. **Linguagens específicas:** recuperação e conversão de documentos envolvendo características específicas de um idioma. Exemplo: classificadores chineses e persas.

Os sistemas de QA e IR têm como ponto de partida uma pergunta escrita em linguagem natural e, a partir daí, buscam uma coleção de documentos na *Web* que sejam compatíveis com o assunto descrito. A análise dessa compatibilidade é feita através do processamento do texto contido na pergunta e da execução de um algoritmo de classificação que utiliza como base o critério de organização dos documentos armazenados. Dessa forma, é possível retornar as respostas esperadas (QA) ou o conjunto de documentos relacionados à pergunta proposta (IR) (SANGODIAH; MUNIANDY; HENG, 2015).

Considerando especificamente o ambiente educacional, a geração automática de testes de avaliação tem aplicação prática imediata nos sistemas de *e-learning* por viabilizar a personalização do ensino através da busca que questões adequadas a um determinado perfil de aprendizado (ROMERO; VENTURA, 2013). Esses sistemas de ensino adaptativo fazem uso de bancos de questões em formatos diversos e os utiliza para gerar testes de avaliação adaptados ao desempenho de cada estudante (SONWALKAR, 2007). Para viabilizar essa personalização, é essencial classificar as questões dentro de uma variedade representativa de competências, habilidades e níveis de dificuldade adequados (SONWALKAR, 2007; MAHATME; BHOYAR,).

Habilidades e competências possuem definição semelhante e estão associadas à capacidade e destreza na realização de uma determinada tarefa. O presente trabalho utiliza o conceito de competência como um conjunto de habilidades; já a habilidade corresponde a um saber fazer de algo mais simples, como mostrado na Figura 1 (BLASIS *et al.*, 2014):

Figura 1 – Competências e habilidades

Fonte: Blasis *et al.* (2014).

Considerar competências e habilidades do estudante como critérios de classificação torna complexa a criação de um banco categorizado de questões mas, conforme já descrito, a alta disponibilidade de questões não classificadas que poderiam ser utilizadas nessa tarefa

presentes em materiais didáticos digitais representa uma oportunidade de pesquisa na otimização do processo para torná-lo mais viável.

Para realizar a classificação automática de questões, podem ser utilizados sistemas especialistas ou algoritmos de Aprendizado de Máquina. No primeiro caso, um ou mais especialistas de domínio criam um conjunto de regras (testes condicionais baseados na presença ou frequência de um conjunto de palavras) que definem a classe de um documento. Porém, regras geradas manualmente são difíceis de serem atualizadas ou mesmo de serem geradas para diferentes locais ou domínios de aplicação, além de serem dependentes da presença e esforço de especialistas de domínio ou engenheiros de conhecimento (MANNING; RAGHAVAN; SCHÜTZE, 2010; SEBASTIANI, 2002; DUMAIS *et al.*, 1998).

Uma alternativa aos sistemas especialistas é o uso de algoritmos de Aprendizado de Máquina para realizar a classificação automática de textos. O objetivo dos algoritmos de Aprendizado de Máquina é aprender, generalizar, ou ainda extrair padrões ou características das classes das coleções com base nos documentos textuais e rótulos (identificadores de classe) dos documentos informados por um usuário ou especialista de domínio. Apesar de ainda haver intervenção humana, a atividade de rotulação dos documentos é mais simples do que a definição de regras por especialistas ou engenheiros de conhecimento.

O uso de algoritmos de Aprendizado de Máquina para a classificação de textos demonstra ter uma *performance* de classificação comparável às alcançadas por sistemas especialistas. Além disso, algoritmos de Aprendizado de Máquina exigem um menor esforço humano; os modelos de classificação são mais rápidos de serem construídos e, portanto, mais fáceis de serem utilizados em diferentes domínios e aplicações (BENBRAHIM; BRAMER, 2009). Tais algoritmos facilitam a personalização da classificação automática de acordo com as necessidades dos usuários. Por exemplo, em uma coleção de questões de vestibular, o usuário poderia querer dividi-las em tópicos gerais como as áreas do conhecimento (**Linguagens, Matemática, Ciências da Natureza, Ciências Humanas**) ou tópicos mais detalhados como **gramática, literatura** para Linguagens ou **trigonometria e conjuntos** para Matemática.

Neste trabalho, são considerados apenas os algoritmos de Aprendizado de Máquina supervisionados, ou seja, aqueles que contém uma base de questões rotuladas definidas por especialistas de forma a induzir um modelo de classificação. Um modelo de classificação consiste em algum mecanismo capaz separar um novo conjunto de dados em classes ou categorias pré-estabelecidas. Esse modelo pode ser um conjunto de regras, hiperplanos de separação, documentos ou protótipos representantes de classes, probabilidades ou valores de relevância de termos para as classes.

Independentemente da categoria, os documentos devem ser representados em um formato apropriado para a manipulação por parte dos algoritmos de Aprendizado de Máquina. A maioria dos trabalhos envolvendo classificação de textos considera que a coleção de documentos é representada no Modelo Espaço-Vetorial (MEV), no qual cada documento (questão) é representado por

um vetor e cada posição desse vetor corresponde a uma dimensão (atributo ou termo/*feature*) da coleção (SALTON, 1989). Costuma-se usar a palavra "termo" para expressar palavra(s) do texto que representam dimensões; atributos para dimensões não baseadas em palavras do texto como autor e localização e *features* para os termos considerados mais relevantes. Por simplicidade, este trabalho considera termo como a porção do texto selecionada para a geração das dimensões. O modelo espaço-vetorial tem como principais características a alta dimensionalidade (centenas de questões podem gerar milhares de dimensões) e a alta esparsidade (cada documento contém poucos termos frente a todos os termos da coleção, e portanto, a maioria dos valores das dimensões é 0). Essas características fazem com que vários algoritmos de Aprendizado de Máquina baseados no modelo espaço-vetorial consumam alta quantidade de memória e apresentem queda de desempenho.

Além do modelo espaço-vetorial, coleções textuais também podem ser representadas como uma rede de termos. O modelo de representação de textos em rede corresponde, basicamente, a um grafo que armazena os objetos pertencentes a um documento (questão) e as relações entre estes objetos. Diversos tipos de objetos e relações podem ser utilizados para representar textos individuais ou coleções de textos. Pode-se considerar como objetos da rede, por exemplo: sílabas, termos, sentenças e documentos. Como relações, pode-se considerar a precedência de sílabas, a coocorrência de termos, ou a similaridade de sentenças ou documentos, entre outros. Alguns domínios específicos contendo informações textuais proporcionam mais possibilidades de tipos de objetos e relações. Por exemplo, nas redes de informações bibliográficas, pode-se ter objetos como autor, conferência, artigo e termo e relações como autor-escreve-artigo e artigo-publicado-em-conferência. No caso deste trabalho, os objetos são representados pelas questões e seus termos e somente a relação questão-termo se aplica. As representações em redes permitem a extração de padrões de classes que dificilmente são capturados por algoritmos baseados no modelo espaço-vetorial, além de melhorarem a performance de classificação dos algoritmos de Aprendizado de Máquina (BREVE *et al.*, 2011).

Outras formas de representação de texto para classificação em algoritmos de Aprendizado de Máquina são as sequências de caracteres (LODHI *et al.*, 2002), árvores de sufixo (ZAMIR; ETZIONI, 1998) e tensores (LIU *et al.*, 2005). Entretanto, esses tipos de representações não foram abordados nesta tese devido à sua baixa aplicação nas tarefas de classificação automática de documentos textuais.

1.2 Motivação

Para se avaliar periodicamente a qualidade do ensino no Brasil, avaliações externas em larga escala como o Exame Nacional do Ensino Médio (ENEM), a Prova Brasil e o exame do Sistema de Avaliação de Educação Básica (SAEB) fazem uso de bancos de itens classificados segundo descritores de competências e habilidades organizados em matrizes de referência.

Estudos recentes mostram que os resultados dessas avaliações externas têm sido utilizados por professores e diretores para, por exemplo, planejar atividades pedagógicas que contemplem as necessidades de aprendizagem identificadas por esses exames (SILVA; GIMENES, 2013; PIMENTA; ISHII; SANTOS, 2014).

Um importante aspecto do diagnóstico fornecido por esses exames é a capacidade dos instrumentos de comparação de seus resultados entre diferentes populações mesmo ao se utilizar diferentes exames, o que só é possível devido à utilização da Teoria de Resposta ao Item (TRI) (ANDRADE; TAVARES; VALLE, 2000; BORTOLOTTI *et al.*, 2013). Por meio da TRI, é possível determinar estatisticamente a probabilidade de acerto de um determinado item de avaliação em termos de parâmetros relacionados ao estudante e ao próprio item e, assim, possibilitar a criação de um banco calibrado de itens, sendo capaz de avaliar o atual nível de habilidades de um estudante e sugerir itens para a continuidade da trilha de aprendizado (WAUTERS; DESMET; NOORTGATE, 2010).

No entanto, apesar da excelente qualidade técnica, os resultados dessas avaliações externas não têm como objetivo apoiar um processo de ensino adaptativo centrado no aluno. O primeiro aspecto deficiente para esse fim é a frequência e o prazo da disponibilização desses resultados: o acesso só acontece anual ou semestralmente ao final de cada ciclo de aprendizagem e, muitas vezes, com defasagem de meses entre a realização dos exames e a liberação dos resultados. Dessa forma, não existe tempo hábil para implementar adaptações no ensino para os mesmos alunos que passaram pelo processo de avaliação. Outro aspecto crítico é a granularidade em que os resultados das avaliações são enviados às escolas. Professores e diretores têm acesso, considerando o nível mais detalhado de informação, apenas ao desempenho geral de cada turma, o que torna inviável sua utilização para identificar necessidades individuais de cada aluno. Em contrapartida, o interesse em individualizar a trilha de aprendizado só tem aumentado (WAUTERS; DESMET; NOORTGATE, 2010). De acordo com Fei *et al.* (2003), a utilização de itens classificados automaticamente para modificar a trilha de aprendizado pode melhorar a qualidade e eficiência da educação, além de reduzir a possibilidade de erro humano. Além da utilização em sistemas computacionais adaptativos, o fato de disponibilizar itens classificados para utilização pelos professores para a realização de avaliações diagnósticas de maior qualidade já representa um importante passo para o apoio de um ensino de qualidade.

Após uma breve revisão bibliográfica, foram encontrados trabalhos relacionados à classificação de questões associados ao ambiente educacional utilizando a taxonomia de Bloom (HARRISON; DIKKEN; PEER, 2017; ABDULJABBAR; OMAR, 2015a; DIAB; SARTAWI, 2017; DUBEY; GOYAL, 2016; OSADI *et al.*, 2017; JAYAKODI *et al.*, 2016; FACHRURROZI; SILABAN; YUSLIANI, 2015), nível de dificuldade (FEI *et al.*, 2003), mapas conceituais (SUN *et al.*, 2018) e, a grande maioria, baseada na taxonomia de Li e Roth (LI; ROTH, 2002), conforme levantamento realizado na revisão sistemática descrita no Capítulo 2. Não foi observada na literatura nenhuma pesquisa relacionada à classificação de questões no ambiente educacional

com base, especificamente, nas competências e habilidades, elementos básicos de avaliações em larga escala como o ENEM.

1.3 Hipóteses

Apesar da crescente produção de conteúdos na internet, especialmente em formato textual, e a necessidade de organização, análise e extração de conhecimento ser um diferencial competitivo fazendo com que a classificação automática de textos venha ganhando cada vez mais destaque (WEISS; INDURKHYA; ZHANG, 2010; AGGARWAL; ZHAI, 2012), há um problema com a classificação de questões. Mesmo funcionando de forma similar à classificação de textos, existe a dificuldade em conseguir índices de precisão razoáveis devido à pequena quantidade de informação presente nas questões, se comparadas aos documentos comuns de texto (RAHMAN, 2015). Com isso, fica limitada a capacidade de processamento dos algoritmos, o que representa um amplo desafio de pesquisa devido às diferentes aplicações da classificação de questões e pouca pesquisa realizada nessa direção.

O ambiente educacional é um contexto de aplicação pertencente ao desafio de pesquisa. O ENEM se enquadra como um caso específico nesse contexto mas, por outro lado, é baseado em competências e habilidades direcionam o processo de aprendizagem de milhões de estudantes em todo o Brasil. Para contribuir nesse processo, foram elaboradas as seguintes hipóteses de pesquisa para esta tese de doutorado:

1. **(Principal)** Existe(m) algoritmo(s) de Aprendizado de Máquina ou combinação(ões) de algoritmos capaz(es) de classificar questões baseadas nas competências definidas para avaliações em larga escala como o ENEM com desempenho similar ou superior ao estado da arte na classificação de textos variadas.
2. **(Suporte)** Existe um formato de representação texto independente de idioma mais eficiente para a classificação de questões em uma ou mais competências que embasam o ENEM.
3. **(Suporte)** Existe um método para cálculo da relevância dos termos mais eficiente para a classificação de questões em uma ou mais competências que embasam o ENEM.

1.4 Objetivos

Para desenvolvimento deste trabalho de pesquisa é necessário traçar objetivos para demonstrar a veracidade das hipóteses levantadas na seção anterior. Todos os objetivos descritos abaixo consideram as competências direcionadoras do ENEM como a base de extração das categorias para a classificação de questões:

- **(Principal)** Identificar a arquitetura de um classificador ou conjunto de classificadores composto pela melhor combinação possível entre a tripla <algoritmo de Aprendizado de Máquina supervisionado, representação de texto, cálculo de pesos dos termos> de forma a maximizar o desempenho do processo de classificação de questões no contexto educacional.
- **(Secundário)** Realizar uma avaliação empírica para comparar o desempenho das diferentes combinações utilizadas.
- **(Secundário)** Disponibilizar as representações, algoritmos, códigos-fonte e ferramentas desenvolvidos para a comunidade científica avaliar e replicar resultados.
- **(Secundário)** Disponibilizar ferramentas de integração e aplicação dos conteúdos desenvolvidos para utilização por outras plataformas e instituições (escolas, empresas) interessadas em utilizar as competências do ENEM para realizar diagnósticos e acompanhamento de aprendizagem.

1.5 Resultados

Com base nos objetivos propostos para demonstrar a veracidade das hipóteses apresentadas, foram obtidos os seguintes resultados:

1. **Identificação do estado da arte na classificação de questões para contexto educacional/avaliações em larga escala:** apoiado no modelo espaço-vetorial e no método $IQF * QF * ICF$, o algoritmo KNN obteve o melhor desempenho tanto na classificação no nível mais agregado (disciplinas) quanto no nível mais detalhado (competências).
2. **Algoritmo $IMBHN^{COS}$:** estado da arte nos resultados de classificação de questões para algoritmos de Aprendizado de Máquina baseado em redes.
3. **Método $IQF * QF * ICF$:** efetividade do cálculo de pesos/frequências em comparação aos métodos tradicionais TF e $TF * IDF$.
4. **Base de questões:** banco de dados constantemente atualizável composto questões classificadas e representativas do ENEM e suas competências.
5. **Application Program Interface (API):** interface para integração com outras plataformas das funcionalidades de pré-processamento de arquivos com questões, geração das representações de texto, cálculo de pesos, indução do modelo de classificação e avaliação de resultados a partir dos indicadores medidos.
6. **Análise de desempenho:** comparativo dos indicadores gerados pelos classificadores de questões com diferentes combinações de componentes e níveis de critérios de classificação (disciplinas e competências).

1.6 Organização do Trabalho

Os próximos capítulos deste trabalho estão organizados da seguinte forma:

- **Capítulo 2:** descreve os resultados de uma revisão sistemática sobre a classificação automática de questões. São analisadas as principais formas de representação do texto, processos para extração e seleção de termos, algoritmos mais utilizados, juntamente com critérios de classificação e indicadores/métricas de desempenho.
- **Capítulo 3:** detalha, inicialmente, os trabalhos relacionados a classificação de questões do ambiente educacional. Em seguida, apresenta um modelo geral para a construção de classificadores automáticos voltados a questões elaboradas para avaliações em larga escala; a partir desse modelo são definidas propostas específicas para a classificação de questões do ENEM usando o modelo espaço-vetorial e o modelo em redes de forma a viabilizar o uso das competências da respectiva MRA como critério de classificação.
- **??:** apresenta a configuração definida para o experimento, incluindo a base de dados, preparação para processamento dos classificadores e algoritmos de classificação. Após a definição da configuração utilizada, é descrito um resultado comparativo dos algoritmos das propostas apresentadas.
- **Capítulo 5:** relata as contribuições científicas do trabalho, lista as limitações e aponta os possíveis trabalhos futuros.

REVISÃO DA LITERATURA

2.1 Considerações Iniciais

Este capítulo apresenta inicialmente uma revisão dos conceitos relacionados à classificação automática de questões e, em seguida, descreve a execução de uma revisão sistemática desenvolvida por [Silva, Bittencourt e Maldonado \(2018\)](#) realizada em 2017 e publicada em 2018. Para este projeto de doutorado, os dados da revisão foram atualizados considerando trabalhos de pesquisa publicados até o final de agosto de 2019. O objetivo dessa revisão sistemática foi apresentar o estado da arte nas tecnologias utilizadas na classificação automática de questões de forma a gerar um ponto de partida para o alcance dos objetivos de pesquisa deste trabalho, conforme descrito no [Capítulo 1](#).

A [Seção 2.2](#) descreve os conceitos envolvendo a representação de textos, etapas básicas para o processo de classificação de questões, principais algoritmos classificadores, taxonomias e indicadores de desempenho. Já a [Seção 2.3](#) trata do método adotado para executar o protocolo da revisão: questões de pesquisa, seleção de trabalhos, avaliação de qualidade, extração de dados e análise de resultados.

2.2 Classificação de Questões

A classificação de questões, conforme já descrito no [Capítulo 1](#), faz parte do algoritmos de Aprendizado de Máquina supervisionados e funciona de forma similar à classificação de textos, com o agravante da dificuldade em conseguir índices de precisão razoáveis ([LI; ROTH, 2002](#)). Isso se deve pela pequena quantidade de informação presente nas questões se comparados aos documentos de texto. A ideia central é associar rótulos às questões de acordo com o critério pretendido de forma efetiva. O critério pode ser, entre outros, a atribuição de um nível de dificuldade (fácil, difícil, por exemplo), tipos de respostas esperadas (política, esportes, saúde etc.)

ou ainda um tópico escolar específico (Geometria/Matemática ou Análise Sintática/Português, por exemplo).

Um dos primeiros trabalhos de destaque desenvolvidos para a classificação de questões, e que é referenciado até o hoje, é o estudo descrito em [Li e Roth \(2002\)](#), desenvolvido na Universidade de Illinois em Urbana-Champaign (UIUC). Trata-se de um classificador hierárquico SNoW com uma base de 6000 questões de inglês para classificação em 6 categorias principais e 50 específicas ([Figura 2](#)). O classificador teve uma acurácia de 84,2% e seus resultados e base de dados continuam a ser uma referência na área, inclusive em grande parte dos estudos selecionados nesta Revisão Sistemática (RS).

Figura 2 – Categorias definidas por Li e Roth

Class	Class	Class	Class
ABBREV	letter	individual	NUMERIC
exp	other	title	code
abb	plant	description	count
ENTITY	product	LOCATION	date
animal	religion	city	distance
body	sport	country	money
color	substance	mountain	order
creative	symbol	other	other
currency	technique	state	period
dismed	term	DESCRIPTION	percent
event	vehicle	definition	speed
food	word	description	temp
instrument	HUMAN	manner	size
lang	group	reason	weight

Fonte: [Sangodiah, Muniandy e Heng \(2015\)](#).

Outro trabalho que é referência na área é o apresentado em [Huang, Thint e Qin \(2008\)](#), utilizando um conjunto compacto de termos através de *head words* e *hypernyms* (estes termos serão explicados mais à frente) com acurácia de 89.2%, utilizando o algoritmo *Support Vector Machine* (SVM) Linear e 89%, com o de (Máxima Entropia (ME)).

[Alahmadi \(2016\)](#) descreve diversas formas de dividir os tipos de classificadores. Considerando o número de categorias de classificação, pode-se dividir em binário (2 classes, uma única associação à questão), *single-label* (diversas classes, uma única associação à questão) e *multi-label* (diversas classes, diversas associações). Por outro lado, considerando o relacionamento entre as categorias, pode-se dividir os classificadores nos tipos *flat* (sem relacionamento) e hierárquico (categorias e sub-categorias). Já com base na decisão de classificação pode-se agrupar nos tipos *hard* (verdadeiro/falso para definir se a questão pertence a uma categoria) e *soft* (indicador numérico que mede o grau de confiabilidade da classificação da questão numa determinada categoria).

Considerando [Taqi e Ali \(2016\)](#), [Jayalakshmi e Sheshasaayee \(2015\)](#), [Hardy e Cheah \(2013\)](#), a principal forma de categorizar os classificadores de questões é com relação à sua estratégia de associação de classes e consiste em 3 tipos:

1. **Regras (*Rule-based*)**: abordagem não estatística por meio de fórmulas com base na estrutura das questões para previsão das classes com alto grau de acurácia, mas que decai com um número grande de questões;
2. **Aprendizado de Máquina (AM) (*Machine Learning*)**: busca de padrões através de análise estatística da estrutura de um conjunto de questões. Seu grau de precisão tende a melhorar com o aumento da quantidade de dados. Pode ser do tipo supervisionado (modelo gerado a partir de um conjunto de dados rotulados, utilizados como treinamento), não supervisionado (sem conjunto de dados rotulados para treinamento) e semi-supervisionado (modelo gerado a partir de um conjunto inicial de dados rotulados e não rotulados);
3. **Híbrido**: combinação de regras e algoritmo de AM.

Baseado em [Patra e Singh \(2013\)](#), na [Figura 3](#) apresentam-se os diferentes estágios durante o processo de classificação de questões. A ideia é transformar o texto inicial da questão em um conjunto de atributos relevantes (termos) que influenciam diretamente no desempenho do classificador.

Figura 3 – Estágios do processo de classificação

Fonte: Adaptada de [Patra e Singh \(2013\)](#).

2.2.1 Representação de Textos

Normalmente, os algoritmos de Aprendizado de Máquina utilizam valores tabelados para a busca de algum padrão que reflita o comportamento destes dados. Nos algoritmos supervisionados, estas tabelas relacionam atributos/termos, valores (peso) e um rótulo de classificação. No

caso da classificação de textos, não existem tabelas com essa característica, apenas um fluxo de texto normalmente de tamanho pequeno em comparação a um documento textual comum. Para resolver este problema, existe a tarefa de representação de textos responsável por elencar quais são os termos mais relevantes e qual o impacto dos mesmos no processamento da classificação.

A principal forma de representação de textos para processamento de textos é o modelo espaço-vetorial, onde os documentos (questões) são representados por vetores no espaço compostos pela combinação dos termos que fazem o papel das dimensões (AGGARWAL; ZHAI, 2012; AGGARWAL, 2014). Outra forma de representação de textos que vem ganhando destaque e se tornando competitiva em relação ao modelo espaço-vetorial são as representações em rede. Tais representações se caracterizam por enfatizar as relações entre os diferentes tipos de componentes de um texto (documento/questão, termo, autor etc.) (SUN; HAN, 2012; SILVA; ZHAO, 2012; ROSSI *et al.*, 2014). Existem ainda representações de texto alternativas como sequências de caracteres (LODHI *et al.*, 2002) e tensores (LIU *et al.*, 2005), mas que não serão tratadas nesta tese de doutorado.

2.2.1.1 Modelo Espaço-Vetorial

Proposto por Salton (1989), o modelo espaço-vetorial é uma associação de vetores numéricos a documentos (questões) em conjunto de documentos (*corpus*). Cada vetor é composto por valores associados aos termos do respectivo documento e descreve a posição do mesmo em um espaço multidimensional, onde cada termo corresponde a uma dimensão. Um termo pode ser uma palavra simples, sequência de palavras, conjuntos de palavras ou mesmo conceitos associados a um conjunto. Os valores associados aos termos são chamados de pesos e indicam a relevância de cada termo dentro do documento. Para calcular esses valores, os documentos passam por um pré-processamento para eliminação de palavras comuns (*stopwords*) como artigos, pronomes e números, agrupamento de variações de uma mesma palavra em um único termo (*stemming*), contagem da repetição dos termos e cálculo do peso final do termo a partir desta contagem em cada documento e/ou também da sua distribuição no conjunto de documentos.

Através da representação vetorial é possível calcular a distância entre os vetores e assim determinar quais vetores estão mais próximos gerando um grau de **similaridade** entre eles. A ideia central é que documentos similares tendam a produzir vetores similares e a estar dentro de uma mesma classe ou categoria de assunto, como acontece na separação de livros dentro de uma biblioteca. A Figura 4 ilustra a representação de três documentos $DOC_A(0.2, 0.35, 0.1)$, $DOC_B(0.3, 0.0, 0.5)$ e $DOC_C(0.5, 0.4, 0.3)$ em um espaço vetorial tridimensional formado pelos termos t_1 , t_2 e t_3 .

Seja $D = \{d_1, d_2, \dots, d_n\}$ o conjunto de N documentos e $T = \{t_1, t_2, \dots, t_m\}$ o conjunto de M termos distintos que compõem uma coleção de textos. Cada documento d_i é representado por um vetor de termos $\vec{d}_i = \{w_{i1}, w_{i2}, \dots, w_{im}\}$ com M dimensões, no qual cada peso w_{ij} quantifica a relevância do termo $t_j \in T$ para o documento $d_i \in D$. Caso um determinado termo t_j não esteja

Figura 4 – Representação de um espaço vetorial

Fonte: Adaptada de Ferneda (2012).

presente em documento d_i , o valor do peso é definido $w_{ij} = 0$; caso contrário, normalmente o cálculo representa a frequência do termo ou uma fórmula derivada dele (LIU, 2007). A união dos vetores relativos a todos os documentos da coleção gera uma matriz W de ordem $N \times M$ chamada de matriz **documento-termo** (Tabela 1). No caso desta tese, que trata da tarefa de classificação, há uma última coluna adicional para representar a classe à qual o documento foi rotulado (y_i).

Tabela 1 – Matriz documento-termo

	t_1	t_2	...	t_m	Y
d_1	w_{11}	w_{12}	...	w_{1m}	y_1
d_2	w_{21}	w_{22}	...	w_{2m}	y_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
d_n	w_{n1}	w_{n2}	...	w_{nm}	y_n

A eficiência dos algoritmos de classificação de questões dependem diretamente de quais termos são extraídos, qual a relevância de cada termo e se é necessário filtrar ou acrescentar novos termos.

No caso da extração de termos, a abordagem mais utilizada é a **Bag-of-Words (BOW)** onde os termos são as diferentes palavras do texto. Isso acaba gerando uma alta dimensionalidade e esparsidade na matriz documento-termo pois gera uma grande quantidade final de termos, muitos dos quais sem ocorrência em grande parte das questões. Para geração dos termos podem ser usadas técnicas de pré-processamento conforme descrito na Subseção 2.2.2, além do uso de *thesaurus* e dicionários de domínio para usar sinônimos e generalizações para substituição e redução do número de termos.

O cálculo do peso/relevância de cada termo, mais conhecido como *term-weighting*, denota a importância daquele termo perante ao documento/questão em que se encontra e também

a toda a coleção de questões. No modelo espaço-vetorial, os pesos correspondem a valores reais positivos e partem, normalmente, do cálculo da frequência do termo em cada questão. A abordagem mais utilizada é a $TF * IDF$ que usa a frequência do termo em cada documento multiplicada pelo inverso da sua frequência em toda a coleção de questões, ou seja, seu valor aumenta cada vez que o termo aparece mais vezes em menos questões.

As técnicas para seleção de termos utilizam um ou mais critérios de avaliação para eliminar termos que não contribuem para a qualidade dos resultados e velocidade de processamento dos algoritmos de classificação.

As técnicas de extração, cálculo de pesos e seleção de termos são melhor detalhados na [Subseção 2.2.3](#), [Subseção 2.2.4](#) e [Subseção 2.2.5](#).

2.2.1.2 Redes

Há muitas definições para o termo redes e que podem variar dependendo do problema ou área de aplicação. Dentro do contexto desta tese, foi adotado o conceito formulado por [Newman \(2010\)](#): “Uma rede, em sua forma mais simples, é uma coleção de pontos, nos quais pares de pontos são conectados por uma linha”. No caso da computação, as redes também são chamadas de grafos e as linhas e pontos são chamados de arestas e nós, respectivamente. Sistemas que possuam elementos com algum tipo de relacionamento entre si podem ser modelados na representação de rede. Cita-se como exemplos redes físicas de computadores, rede elétrica, redes sociais e própria *World Wide Web*.

Numa definição mais formal, pode-se definir uma rede como uma tripla $N = \{E, R, W\}$, onde E representa os elementos da rede, R corresponde às relações entre os elementos e W os pesos destas relações. Da mesma forma que descrito no modelo espaço-vetorial, estes pesos serão valores reais positivos.

Considerando a capacidade de representação de textos de questões e a aplicação de algoritmos de Aprendizado de Máquina é essencial definir quais diferentes categorias de elementos podem pertencer à rede como, por exemplo, identificador, enunciado do problema, instituição de aplicação do questionário, ano de aplicação e assim por diante. Quando todos os elementos são da mesma categoria tem-se uma rede **homogênea**; caso contrário, a rede é chamada **heterogênea** ([Figura 5](#)). Um exemplo clássico de rede homogênea são as redes de páginas *web*, que se relacionam através de hiperlinks; uma rede bibliográfica contendo elementos como autor, livro, editora, local é um bom exemplo para rede heterogênea.

A classificação de questões possui dois elementos distintos: o identificador da questão (documento) e as palavras que compõem o enunciado da questão (termos). Diferentes combinações desses elementos geram diferentes tipos de rede, por exemplo, documento-documento, documento-termo, documento-documento-termo e assim, por diante. Considerando os objetivos desta tese, a pesquisa foi direcionada sobre as redes documento-termo ([Figura 6](#)), a qual

Figura 5 – Redes homogêneas e heterogêneas

Fonte: Rossi (2016).

corresponde a uma rede heterogênea do tipo bi-partida, na qual há apenas dois tipos de elementos e um tipo de objeto se conecta apenas com o outro tipo e nunca com o mesmo tipo. Conforme apresentado em Rossi (2016), as redes bipartidas são muito utilizadas para representação de documentos de texto e apresentam ótima eficiência na tarefa de classificação supervisionada de textos executada por algoritmos de Aprendizado de Máquina.

Figura 6 – Rede documento-termo

Fonte: Rossi (2016).

Similarmente ao modelo espaço-vetorial, as redes também possuem uma representação na forma de matriz. Chamada de matriz de adjacências, essa matriz armazena em cada uma de suas células o valor do peso atribuído à relação entre cada um dos elementos. No caso de redes homogêneas, a ordem da matriz é $|E| \times |E|$, conforme mostra a Tabela 2; já para as redes heterogêneas bipartidas (caso desta tese), a ideia é semelhante, pois só existe um tipo de relação entre os dois tipos possíveis de elementos com ordem $|M| \times |N|$, considerando M e N as respectivas quantidades para cada tipo de elemento. No caso geral para redes heterogêneas, pode-se criar uma matriz de adjacências para cada combinação de tipos de objetos, 2 a 2, ou

Tabela 2 – Matriz de adjacências

	E_1	E_2	...	$E_{ E }$
E_1	$w_{E_1E_1}$	$w_{E_1E_2}$	...	$w_{E_1E_{ E }}$
E_2	$w_{E_2E_1}$	$w_{E_2E_2}$	...	$w_{E_2E_{ E }}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$E_{ E }$	$w_{E_{ E }E_1}$	$w_{E_{ E }E_2}$	...	$w_{E_{ E }E_{ E }}$

reúne-se essas submatrizes em uma única matriz, ordenando os objetos por tipo.

A matriz de adjacências tem a característica de ser esparsa, pois há muito mais elementos do que relações acarretando desperdício de memória. Uma alternativa muito utilizada para representação de redes são as listas de adjacências (NEWMAN, 2010), uma lista para cada elemento. Tanto para redes homogêneas quanto para as redes bipartidas, cada elemento da lista é formado por um índice e um valor de peso, ambos associados ao elemento com o qual existe uma relação [Figura 7](#). Cada lista é formada apenas por elementos com peso diferente de zero, que reduz muito a esparsidade.

Figura 7 – Lista de adjacências para uma rede homogênea

Fonte: [Rossi \(2016\)](#).

2.2.2 Pré-Processamento

Conforme [Alahmadi \(2016\)](#), a fase de pré-processamento consiste em converter o texto inicial da questão em um conjunto bem definido de termos. Isso inclui remover todas as palavras irrelevantes como pronomes, preposições, pontuação, além da unificação de palavras semelhantes. Entre as principais operações possíveis de pré-processamento, destacam-se ([POTA et al., 2015](#); [ALAHMADI, 2016](#)):

1. **Tokenização:** divisão do texto em fragmentos elementares chamados *tokens* e que são separados por caractere específico. No caso das questões, os *tokens* são as palavras e o caractere específico é o espaço;
2. **Remoção de stop Words:** retirada de palavras comuns como pronomes, preposições e artigos (todos chamados de *stop words*) e que não têm relevância para uma determinada

classificação. Também são eliminados outros caracteres irrelevantes como os numéricos e de pontuação.

3. **Tagging**: associação de classes morfológicas-léxicas a cada *token*, por exemplo, artigo, verbo, advérbio;
4. **Stemming**: substituição de cada *token* pela sua palavra de origem, por exemplo, “escritor”, “escrita” e “escreveram” por “escrever”.
5. **Parsing**: geração de uma estrutura de *tokens* no formato de árvore a partir de um conjunto de regras gramaticais, de forma a representar a estrutura sintática da questão.

Na [Figura 8](#) ilustram-se alguns exemplos de aplicação dos diferentes tipos de *feature*:

Figura 8 – Tipos de termos

Feature Space	Features
Unigram	{{Who, 1} (was, 1) (elected, 1) (president, 1) (of, 1) (South, 1) (Africa, 1) (in, 1) (1994, 1) (?, 1)}
Bigram	{{(Who-was, 1), (was-elected, 1), (elected-president, 1), (president-of, 1), (of-South, 1), (South-Africa, 1), (Africa-in, 1), (in-1994, 1), (1994-?, 1)}
Trigram	{{ Who-was-elected, 1), (was-elected-president, 1), ..., (in-1994-?, 1)}
Wh-Word	{{Who, 1}}
Word-Shapes	{{(lowercase, 5) (mix, 3) (digit, 1) (other, 1)}

Question	Category
What <u>city</u> has the zip code of 35824 ?	LOC:city
Who developed the <u>vaccination</u> against polio ?	HUM:ind
Who invented the slinky ?	HUM:ind
George Bush purchased a small interest in which baseball <u>team</u> ?	HUM:gr
When did Idaho become a state ?	NUM:date
What <u>river</u> flows between Fargo, North Dakota and Moorhead, Minnesota ?	LOC:other
What is the oldest <u>city</u> in Spain ?	LOC:city

“What river flows between Fargo, North Dakota and Moorhead. Minnesota?”.

Feature Space	Features
Hypernyms	{{(river, 1) (stream, 1) (body-of-water, 1) (thing, 1) (physical-entity, 1) (entity, 1)}
Related Words	{{(rel:What, 1) (rel:list.tar, 2) (rel:loca, 2)}
Question Category	{{other, 1}}
Query Expansion	{{(river, 1) (stream, 0.6) (body-of-water, 0.36) (thing, 0.22) (physical-entity, 0.13) (entity, 0.08)}

Fonte: Adaptada de [Van-Tu e Anh-Cuong \(2016\)](#).

2.2.3 Extração de termos

Como resultado da fase de pré-processamento, diversos tipos de termos podem ser extraídos e, com isso, influenciar diretamente na performance do processo de classificação, de

acordo com a estratégia adotada. Normalmente, são divididos em léxicos, sintáticos e semânticos (POTA *et al.*, 2015; JAYALAKSHMI; SHESHASAYEE, 2015):

1. **Léxicos:** basicamente, são as palavras no contexto de uma questão. Alguns exemplos são:
 - a) *Unigram ou Bag-of-Words*: correspondem a cada par (t, f) onde t representa uma palavra da questão e f a quantidade de vezes que essa palavra t aparece;
 - b) *N-grams*: generalização da *BOW* para N termos consecutivos que aparecem f vezes em uma questão. *Unigrams* é o caso para N=1, *Bigrams*, N=2 e assim por diante;
 - c) *Wh-Words*: palavras que implicam o sentido de pergunta (*what, who, where* etc.).
2. **Sintáticos:** termos derivados da estrutura sintática da questão. Exemplos são:
 - a) *Part-of-Speech (POS)-tags*: marcação de palavras de acordo com sua classe gramatical ou de palavras como no caso dos adjetivos, substantivos, advérbios etc.;
 - b) *Head words*: palavras-chaves nas questões obtidas com a operação de *parsing* e que contém a informação principal para o processo classificação (Figura 8).
3. **Semânticos:** termos associados a uma determinada classificação da questão. Os principais exemplos são *Hypernyms/Hiponyms*, ou seja, palavras que representam o conceito semântico de generalização/especialização. Por exemplo: (Cor)/(Azul, Vermelho, Verde) ou (Flor)/(Rosa, Jasmim e Orquídea).

2.2.4 Term-Weighting

Term-weighting é a tarefa de calcular pesos para os termos extraídos. Os pesos são usados para criar um vetor com estes termos para cada questão (Modelo Espaço-Vetorial) e, a partir dele, aplicar técnicas para selecionar quais são mais relevantes. As técnicas mais utilizadas são:

1. Binária (EMMANUEL; KHATRI; BABU, 2013): técnica simples que consiste em atribuir valor 1 se o *feature* aparece em uma determinada questão e 0 se não houver ocorrência;
2. *Term Frequency (TF)* (EMMANUEL; KHATRI; BABU, 2013): corresponde ao total de ocorrências de um *feature* em uma mesma questão;
3. *Inverse Document Frequency (TF*IDF)* (EMMANUEL; KHATRI; BABU, 2013): a ideia dessa técnica de seleção é ajustar o valor de TF pelo fato de que termos comuns em muitas questões normalmente não contribuem para a classificação e possuem alta frequência. A fórmula para cálculo do TF*IDF para cada *feature* é:

$$TF * IDF = TF * \log \left(\frac{N}{n} \right)$$

Onde:

N = Total de questões

n = número de questões nas quais o *feature* está presente (DF).

4. *Inverse Class Frequency* ($TF*ICF$) (EMMANUEL; KHATRI; BABU, 2013): método com o mesmo raciocínio do $TF*IDF$, mas considerando a incidência de um *feature* em muitas categorias no lugar das questões. Funciona como um fator de ajuste ao próprio $TF * IDF$ original:

$$TF * ICF = TF * \left[1 + \log \left(\frac{N}{n} \right) \right] * \left[1 + \log \left(\frac{C}{c} \right) \right]$$

Onde:

N = Total de questões

n = número de questões nas quais o *feature* está presente (DF)

C = Total de categorias

c = número de categorias nas quais o *feature* está presente.

2.2.5 Seleção de termos

A seleção de termos consiste na tarefa de descobrir quais termos são mais relevantes que outros para o problema de classificação. Diminuir a quantidade de termos tende a tornar o algoritmo de classificação mais rápido e eficiente através da melhora nos indicadores de desempenho como a acurácia. A ideia principal é partir do cálculo de pesos e/ou cálculo de índices para diferenciar os termos mais relevantes através de algum tipo de critério de avaliação. Entre as técnicas mais comuns para seleção de termos destacam-se:

1. *Document Frequency* (DF) (SHAH; PATEL, 2016): número de documentos (questões) no qual um determinado *feature* ocorre. É estabelecido um valor limite e todos os termos com valores menores de DF são descartados;
2. *Chi-Square* (X^2) (ALAHMADI, 2016): medida do grau de dependência entre um *feature* f e uma categoria c . Quando maior o valor dessa estatística, maior a associação entre f e c . Para o cálculo é necessário tomar como base a tabela de contingência entre f e c Tabela 3, onde A é o número de vezes que f aparece em c , B é o número de vezes que f não aparece em c , E é o número de vezes que c aparece sem f , K é o número de vezes que f e c não aparecem e N é o número total de documentos:

$$CHI^2(f, c) = \frac{N(AK - EB)^2}{(A + E)(B + K)(A + B)(E + K)}$$

Tabela 3 – Tabela de contingência entre f e c

	c	NOT c	Total
f		B	A + B
NOT f	E	K	E + K
Total	A + E	B + K	N

Outras técnicas também muito utilizadas são o *Information Gain (IG)*, *Odds Ratio (OR)* e *Mutual Information (MI)* e estão detalhadas em [Taqi e Ali \(2016\)](#).

2.2.6 Algoritmos

Conforme descrito no início do capítulo, os classificadores de questões dividem-se praticamente em algoritmos a base de regras, Aprendizado de Máquina e combinação entre eles. As regras, normalmente, são instruções *hard-coded*, ou seja, são fórmulas definidas de acordo com o problema e não tem algoritmos pré-definidos. Já os algoritmos de Aprendizado de Máquina possuem uma grande variedade e as seções seguintes destacam os mais utilizados na classificação de questões.

2.2.6.1 KNN

De acordo com [Patra e Singh \(2013\)](#), *K-Nearest Neighbor (KNN)* é um algoritmo onde os objetos são classificados por votação de vários exemplos de treinamento rotulados com suas menores distâncias possíveis para cada objeto. O algoritmo é conhecido por sua capacidade de reconhecimento de padrões e sua maior desvantagem é a necessidade de utilizar todos os termos na computação das distâncias, o que gera um alto custo computacional ([PATRA; SINGH, 2013](#); [GHAREHCHOPOGH; LOTFI, 2013](#)).

2.2.6.2 Árvores de Decisão

Uma árvore de decisão reconstrói os dados de treinamento no formato de árvore utilizando condições bem definidas no formato verdadeiro-falso ([GHAREHCHOPOGH; LOTFI, 2013](#)). A ideia é que os nós-folha representem um conjunto de termos e assim um ou mais grupos ou categorias. Com a árvore gerada, novos dados podem ser confrontados às condições de cada nó, resultando na previsão de uma classificação para os mesmos. A principal vantagem é a facilidade de compreender o modelo através de sua estrutura e a principal desvantagem é que os dados de treinamento possuem muitos termos resultando em problemas de *performance* e *overfitting* ([SHAH; PATEL, 2016](#)). Há uma variedade de algoritmos para árvores de decisão tais como ID3, ID4-hat, ID5, C4.5 e CART.

2.2.6.3 Naïve Bayes

O algoritmo *Naive Bayes* também é um algoritmo conhecido na área de classificação de textos e consiste, basicamente, da análise de probabilidades a partir de uma base de dados de treinamento a aplicação destas probabilidades para classificação de questões através dos termos extraídos (PATRA; SINGH, 2013). O *Naive Bayes* assume independência entre os termos e o desempenho de aprendizado é rápido e simples. O problema é a perda de *performance* no caso de algum tipo de relação entre os termos (atributos). De acordo com a distribuição dos dados possui implementações diferentes como, por exemplo, o *Multinomial Naive Bayes* (MNB), *Gaussian Naive Bayes* (GNB), *Complement Naive Bayes* (CNB) e *Bernoulli Naive Bayes* (BNB).

2.2.6.4 SVM

O SVM (Máquina de Vetor de Suporte) é um método de classificação para dados lineares e não lineares. Usa mapeamento não linear para transformar os dados de treinamento em dimensões superiores e então encontra um hiperplano ótimo linear para separação das categorias (PATRA; SINGH, 2013). O classificador linear é definido pela função $WTx + B = 0$ Figura 9, onde W é a direção do hiperplano e B sua exata posição (SHAH; PATEL, 2016).

Figura 9 – Support Vector Machine

Fonte: Elaborada pelo autor.

Os itens fora dos hiperplanos representam 2 categorias separadas e as coordenadas pertencentes aos hiperplanos são conhecidas como vetores de suporte. Para Shah e Patel (2016), o SVM é robusto e tem valores ótimos de acurácia, mas possui alta complexidade e requer uso extensivo de memória para tarefas de larga escala.

2.2.6.5 Redes Neurais

A Rede Neural Artificial (*Artificial Neural Network* (ANN)) é um sistema composto de unidades de processamento chamadas neurônios que recebem entradas (*inputs*) e geram cada uma saída (*output*). Conforme Gharehchopogh e Lotfi (2013), o objetivo é simular algumas das funcionalidades do neurônio biológico através da interconexão entre os neurônios. Sua estrutura pode variar desde uma simples camada de neurônios (Perceptron) a múltiplas camadas como, por exemplo, o *MultiLayer Perceptron* (MLP), conforme mostra a Figura 10:

Figura 10 – Redes Neurais: (a) Perceptron e (b) MLP

Fonte: Adaptada de Wikipédia (2019).

Ainda segundo Gharehchopogh e Lotfi (2013), uma das suas vantagens principais é a boa performance com uma grande quantidade de termos, o que acarreta como desvantagem alto custo de processamento e memória, além da dificuldade inerente no entendimento do funcionamento das redes neurais por parte dos usuários.

2.2.7 Taxonomias

O processo de classificação de questões necessita de critérios para separação das categorias. Esses critérios podem ser expressos através de taxonomias, as quais detalham os requisitos em comum para agrupamento. Conforme apresentado na Figura 2, a taxonomia proposta por Li e Roth (2002) é uma das mais utilizadas devido às classes bem definidas e base de dados disponível para utilização. Outra taxonomia muito referenciada é a de Bloom que consiste em

um sistema baseado em objetivos educacionais para classificação em níveis de aprendizado e entendimento (ABDULJABBAR; OMAR, 2015a). A taxonomia possui 3 domínios: cognitivo, afetivo e psicomotor, mas os trabalhos de classificação de itens de avaliação costumam considerar apenas o domínio cognitivo. Este domínio é similar às avaliações escolares e consiste em seis níveis hierárquicos, iniciando do nível mais simples, o conhecimento, até o mais complexo, a avaliação. O Quadro 1 descreve cada nível do domínio cognitivo, ilustrando exemplos simples:

Quadro 1 – Taxonomia de Bloom

Categoria	Exemplo de Questões
Conhecimento	Defina o conceito de herança
Compreensão	Explique a estrutura de um método em um programa
Aplicação	Demonstre a relação entre pacotes, classes e métodos
Análise	Liste as vantagens de utilizar uma classe do tipo <i>container</i>
Síntese	Crie um programa em Java mostrando o conceito de sobrecarga
Avaliação	Justifique o conceito de herança e escreva um código-fonte exemplo

2.2.8 Indicadores de Desempenho

O grande objetivo dos classificadores automáticos é chegar o mais perto possível da exatidão na atribuição dos rótulos, diminuindo assim a intervenção humana. Além de uma possível aplicação inadequada das questões que estiverem classificadas incorretamente, há a necessidade de gastos de tempo, dinheiro e recursos humanos para o trabalho de correção. Para isso, é necessário avaliar o andamento do processo e isso é realizado por meio de indicadores de desempenho. Os indicadores são extraídos por meio do levantamento de questões corretamente categorizadas e também dos erros de classificação, os quais são obtidos através da matriz de contingência gerada. A matriz de contingência é a base para a maioria dos indicadores (Tabela 4), mas existem outros, por exemplo, o tempo de processamento.

Os indicadores mais conhecidos são (ALAHMADI, 2016; TAQI; ALI, 2016):

1. Acurácia: proporção das questões que são classificadas corretamente.

$$Acuracy = (TP + TN) / TOTAL$$

Onde:

TP: *True Positives* (questões corretamente classificadas na categoria)

TN: *True Negatives* (questões corretamente não classificadas na categoria)

TOTAL = Total de questões

2. Precisão: proporção de acerto entre as questões previstas para uma determinada categoria.

$$Precision = TP / (TP + FP)$$

Onde:

TP: *True Positives* (questões corretamente classificadas na categoria)

Tabela 4 – Matriz de contingência para o processo de classificação (Adaptado)

			Classificação atual		
		Classificação Positiva	Classificação Negativa		
Classificação Prevista	Previsão Positiva	TP (True Positives)	FP (False Positives)	Precisão (P) = $TP / (TP + FP)$	
	Previsão Negativa	FN (<i>False Negatives</i>)	TN (<i>True Negatives</i>)		
		$Recall (R) = TP / (TP + FN)$		Acurácia (A) = $(TP + TN) / Total$	$F1-score (F1) = 2 * P * R / (P + R)$

Fonte: [Wikipédia \(2018\)](#).

FP: *False Positives* (questões incorretamente classificadas na categoria)

TP + FP = Total de questões classificadas em uma determinada categoria

3. *Recall*: proporção de acerto para as questões pertencentes a uma determinada categoria.

$$Recall = TP / (TP + FN)$$

Onde:

TP: *True Positives* (questões corretamente classificadas na categoria)

FN: *True Negatives* (questões incorretamente não classificadas na categoria)

TP + TN = Total de questões pertencentes a uma determinada categoria

4. *F1-score (F1-measure)*: Média harmônica entre Precisão e *Recall*.

$$F1 = (2 * Precision * Recall) / (Precision + Recall)$$

2.3 Revisão Sistemática

Utilizar a revisão sistemática como método de pesquisa significa identificar, avaliar e interpretar uma busca de informações associada a uma questão de pesquisa, área ou fenômeno para gerar evidências que possam embasar possíveis conclusões ([KITCHENHAM; CHARTERS, 2007](#)). O contexto a ser realizada a RS é a classificação automática de questões, também chamadas de itens de avaliação, conforme introduziu a [Seção 2.2](#). As diretrizes e o *template* do protocolo da revisão foram baseados nos trabalhos de [Nakagawa et al. \(2017\)](#), [Kitchenham \(2004\)](#), [Kitchenham e Charters \(2007\)](#), [Biolchini et al. \(2005\)](#) e incluem as atividades de planejamento, execução e análise de resultados.

O planejamento da revisão sistemática é responsável por identificar se há a necessidade de realizar a revisão sistemática sobre o tema escolhido e, em caso positivo, qual a estratégia a ser

utilizada para a busca de estudos primários através da definição de um protocolo (BIOLCHINI *et al.*, 2005). Tal protocolo inclui os métodos para a definição da questão de pesquisa, termos de e fontes para busca, critérios de seleção e exclusão, além da extração e síntese de dados. O protocolo especifica os métodos que serão utilizados para empreender uma específica revisão sistemática. Kitchenham (2004) define os seguintes componentes para um protocolo:

1. Justificativa, ou seja, a razão da pesquisa;
2. Questões de pesquisa que a revisão pretende responder;
3. Estratégia que será utilizada para procurar os estudos primários, incluindo os termos de busca e as fontes para extração dos dados. Isso inclui desde a formulação de palavras-chaves e criação de *strings* de busca através da combinação entre elas até a seleção dos recursos a serem pesquisados, que podem incluir, por exemplo, bancos de dados indexados, periódicos específicos, e anais de eventos;
4. Critérios de inclusão e exclusão de estudos de acordo com os objetivos definidos como, por exemplo, o idioma original do estudo e sua área de aplicação;
5. *Checklists* para avaliação da qualidade dos estudos selecionados;
6. Estratégia para extração de dados. Utilizada para definir a forma como as informações exigidas a partir de cada estudo preliminar serão obtidas, ou seja, se será necessária a validação dos dados através de alguma inferência ou manipulação sobre os mesmos;
7. Síntese e análise dos dados extraídos;
8. Calendário do projeto para definir a duração de cada etapa da revisão.

Deve-se ressaltar que esses componentes/etapas não são sequenciais e, durante o trâmite da revisão, podem ser executados mais de uma vez e sofrer alterações devido aos processos de aprovação tanto na fase de planejamento quanto de execução. O protocolo da revisão sistemática representa o planejamento propriamente dito no sentido de descrever todas as estratégias, métodos e considerações a serem aplicados durante a sua condução. Para apoiar a definição do protocolo e condução da RS, foi utilizado o software StArt (*State of Art through Systematic Reviews*), que vem mostrando resultados positivos nessa atividade (HERNANDES *et al.*, 2012).

2.3.1 Questões de Pesquisa

O objetivo desta revisão sistemática é identificar o estado da arte na aplicação de algoritmos de Aprendizado de Máquina para a classificação automática de questões. Há uma quantidade razoável de algoritmos para classificação de textos envolvendo a tecnologia do Aprendizado de Máquina. No caso de questões, há a falta de levantamentos sobre o estado da arte dos classificadores de questões, suas possíveis aplicações e forma de medição dos resultados obtidos. Dessa

forma, a meta é responder a seguinte questão primária de pesquisa:

Quais os principais algoritmos para classificação automática para questões?

Com base na questão primária, surgem questões específicas de interesse, conforme apresentado no [Quadro 2](#).

Quadro 2 – Questões de pesquisa

Questão de pesquisa	Motivação
QP1: Quais métodos computacionais utilizados na implementação dos classificadores?	Esta questão identifica a metodologia adotada na construção dos principais algoritmos de Aprendizado de Máquina para classificação de questões.
QP2: Quais taxonomias adotadas para classificação?	A resposta a esta questão permite identificar os critérios de classificação aplicados.
QP3: Quais as principais técnicas utilizadas para extração e seleção dos termos?	A extração e seleção dos atributos (termos) relevantes impactam diretamente na performance dos classificadores.
QP4: Quais os principais instrumentos de medição de resultados de classificação utilizados?	Verificar as diferentes formas de analisar a eficiência do processo de classificação.

2.3.2 Seleção de Fontes e Estudos

O critério de seleção das fontes de arquivos considerou a disponibilidade de consulta dos estudos na web, publicações atualizadas regularmente, disponibilidade dos textos, qualidade dos resultados, possibilidade de exportação das referências bibliográficas e mecanismo de busca por título, resumo e palavras-chave, todos escritos no idioma inglês. Também foram consideradas revisões formais e informais da literatura realizadas por especialistas e a opinião de pesquisadores, ambos envolvendo de alguma forma a área de classificação automática de textos e questões, conforme orientações descritas em [Nakagawa et al. \(2017\)](#). Com base nos requisitos supracitados, foram selecionadas as seguintes bases de dados eletrônicas para pesquisa: *ACM Digital Library* ¹, *Scopus* ², *Elsevier Science Direct* ³, *IEEEExplore Digital Library* ⁴, *Web of Science* ⁵ e *Google Acadêmico* ⁶. Antes da criação da *string* de busca, foi criada uma lista de estudos primários considerados como referências de controle. Essa lista foi extraída a partir do trabalho de pesquisa dos autores e também de referências secundárias sobre o assunto analisadas anteriormente, como o trabalho de [Taqi e Ali \(2016\)](#), os quais desenvolveram uma RS

¹ <<http://dl.acm.org>>

² <<http://www.scopus.com>>

³ <<http://www.sciencedirect.com>>

⁴ <<http://ieeexplore.ieee.org>>

⁵ <<https://webofknowledge.com>>

⁶ <<http://scholar.google.com>>

focada na classificação de questões em exames de informática e a revisão informal publicada por [Sangodiah, Muniandy e Heng \(2015\)](#) que considerou a classificação de questões em ambientes de *Question Answering* (QA), *Information Retrieval* (IR), educação e conversão de idiomas. A lista de controle é apresentada no [Quadro 3](#).

Quadro 3 – Seleção de estudos para grupo de controle

Referência	Título
Abduljabbar e Omar (2015a)	Exam Questions Classification Based on Bloom's Taxonomy Cognitive Level Using Classifiers Combination
Haris e Omar (2012a)	A rule-based approach in Bloom's Taxonomy question classification through natural language processing
Haris e Omar (2015)	Bloom's taxonomy question categorization using rules and n-gram approach
Jayakodi, Bandara e Meedeniya (2016)	An automatic classifier for exam questions in Engineering: A process for Bloom's taxonomy
Omar et al. (2012a)	Automated analysis of exam questions according to Bloom's taxonomy
Yahya et al. (2013)	Analyzing the cognitive level of classroom questions using machine learning techniques
Yahya e Osman (2014)	Classification of high dimensional Educational Data using Particle Swarm Classification

Com base nas questões de pesquisa e diretrizes da RS, foi definida uma *string* de busca considerando as áreas e algoritmos envolvidos, ou seja, Aprendizado de Máquina, Mineração de Dados (*Data Mining*) e Classificação de Questões . Considerando os possíveis sinônimos, obteve-se o seguinte resultado ([Quadro 4](#)):

Quadro 4 – String de busca

```
((("MACHINE LEARNING" OR
"DATA MINING" OR
"TEXT MINING" OR
"DEEP LEARNING")
AND
("QUESTION CLASSIFICATION" OR
"QUESTION CLASSIFIER" OR
"QUESTION ANALYSIS" OR
"QUESTION ANALYZE" OR
"QUESTION CATEGORIZATION")))
```

Entre os tipos de estudos possíveis destacam-se os estudos primários publicados em periódicos, conferências e capítulos de livro com preferência para os estudos mais recentes, no caso de trabalhos semelhantes. Com base nas diretrizes propostas por [Kitchenham \(2004\)](#), no [Quadro 5](#) apresentam-se os critérios de inclusão/exclusão definidos para essa RS.

Quadro 5 – Critérios de inclusão/exclusão

Critérios de inclusão
CI1: Estudos primários
CI2: Estudos que apresentam algoritmos para classificação de questões
CI3: Estudos publicados entre janeiro de 2012 e agosto de 2019
CI4: Avaliação de qualidade com pontuação maior ou igual a 50% do possível
Critérios de exclusão
CE1: Estudos secundários
CE2: Estudos incompletos ou com menos de 5 páginas
CE3: Estudos escritos em língua diferente do Inglês
CE4: Estudos com texto completo indisponível
CE5: Estudos com foco nas áreas de QA ou IR
CE6: Estudos com foco na classificação em outros idiomas (Ex: chinês, persa, hindi, árabe)
CE7: Estudos não abordam algoritmos para a classificação de questões

Na primeira etapa do processo, chamada de Seleção Preliminar, a *string* de busca foi adaptada e executada para cada base de dados eletrônica da lista de fontes. A busca inicial resultou na seleção de 5491 estudos primários compatíveis. A [Figura 11](#) mostra as etapas de aplicação dos critérios de inclusão/exclusão, detalhando ao lado direito quais os tipos de estudo descartados.

Para limitar o número de estudos a serem carregados, foi utilizado o mecanismo de filtro existentes nas próprias bases de dados para selecionar apenas os estudos publicados posteriormente ao início de janeiro de 2012 (CI3). Este primeiro filtro resultou em 2485 estudos, os quais foram identificados e carregados na ferramenta StArt para organização e depuração das informações e eliminação das entradas duplicadas (202), resultando em 2283 registros.

O próximo passo foi a aplicação dos critérios de inclusão. Num primeiro momento, foi realizada a leitura do título, palavras-chave e área de conhecimento abordada pelo estudo para excluir os estudos que não atendiam os critérios de inclusão, não estando assim relacionados às questões de pesquisa. Nesse passo foram descartados 1820 registros, restando 463 estudos primários para análise. Os estudos foram submetidos à leitura de seu resumo (*abstract*) e, novamente, à aplicação dos critérios de inclusão/exclusão resultando na exclusão de mais 252 entradas (detalhes da exclusão são apresentados no segundo quadro à direita na [Figura 11](#)).

Os 211 estudos selecionados tiveram seu texto completo recuperado e a leitura das seções de introdução e conclusão realizada. Dessa etapa, mais 84 estudos foram descartados pelos critérios de inclusão/exclusão, a grande maioria por estarem voltados à tradução de questões para outro idioma (24 estudos, 28,5%) ou não tratarem especificamente de classificação de questões (40 estudos, 47,6%). Considerando as referências bibliográficas destes estudos selecionados e os estudos secundários (revisões) descartados nas etapas anteriores, 4 foram adicionados à seleção ([Figura 12](#)) por serem representativos na área de classificação de questões. O autor também um artigo de 2018 por considerá-lo o estado da arte na área. Com isso, resultaram 132 estudos para

aplicação do CI4 referente à pontuação mínima nos critérios de qualidade.

Figura 11 – Fluxo de aplicação dos critérios de inclusão e exclusão

Fonte: Elaborada pelo autor.

Nessa etapa final, 35 estudos foram eliminados restando 97 estudos para a condução da revisão sistemática. Os critérios com menor pontuação foram CQ4 (descrição das limitações, média de 0,09), CQ7 (disponibilização dos dados, média 0,29), CQ5 (análise estatística, média de 0,59) e CQ6 (experimento para validação, média de 0,59).

Figura 12 – Referências mais citadas nos estudos selecionados

Fonte: Elaborada pelo autor.

2.3.3 Avaliação de Qualidade

A avaliação de qualidade permite selecionar apenas os estudos mais relevantes, com resultados concretos dentro do tema de pesquisa desejado. Cinco questões foram obtidas a partir da literatura e outras duas propostas de acordo com o escopo e questões de pesquisa.

Tabela 5 – Questões para avaliação de qualidade

Questão	Respostas permitidas e pontuação
CQ1: Existe uma justificativa para o motivo pelo qual o estudo foi realizado? (MAHDAVI-HEZAVEHI; GALSTER; AVGERIOU, 2013)	S = 1, P = 0,5 e N = 0
CQ2: Existe uma declaração clara dos objetivos da pesquisa? (DYBA; DINGSOYR, 2008; SALLEH; MENDES; GRUNDY, 2011)	S = 1, P = 0,5 e N = 0
CQ3: A técnica proposta está claramente descrita? (ACHIMUGU <i>et al.</i> , 2014))	S = 1, P = 0,5 e N = 0
CQ4: As limitações deste estudo são explicitamente discutidas? (DING <i>et al.</i> , 2014)	S = 1, P = 0,5 e N = 0
CQ5: Os resultados do experimento foram obtidos de forma confiável como, por exemplo, através de análise estatística? (SALLEH; MENDES; GRUNDY, 2011)	S = 1, P = 0,5 e N = 0
CQ6: O estudo descreve um experimento para validação do algoritmo?	S = 1, P = 0,5 e N = 0
CQ7: O estudo disponibiliza os dados para replicação do experimento?	S = 1, P = 0,5 e N = 0

A escala de pontuação foi baseada na dicotomia [Sim (S)]/[Não (N)], com pontuação 1 para respostas afirmativas e 0 para as negativas, com a possibilidade do atendimento parcial da questão (P) e respectiva pontuação 0,5. Para utilização como critério mínimo de exclusão,

considerou-se a pontuação menor que 3,5 por representar 50% de aproveitamento dos 7 pontos possíveis (Tabela 5).

A Tabela 6 apresenta os estudos selecionados a partir do resultado da avaliação de qualidade e aplicação do critério de inclusão CI4. Analisando os resultados da avaliação, a média de pontuação dos estudos ficou em 4,48 (64%), o estudo E05 obteve a maior pontuação (6,5 ou 92,9%), enquanto 13 estudos conseguiram a pontuação limite (50%) para a extração de dados e análise dos resultados (E4, E10, E13, E18, E21, E39, E41, E42, E45, E61, E63, E71 e E75). Considerando individualmente os critérios de qualidade, percebe-se que os critérios Q1 e Q2 obtiveram maior pontuação, o que indica justificativas bem fundamentadas e objetivos claramente definidos; por outro lado, o critério Q4 obteve uma média muito baixa (0,12) implicando que a maior parte dos estudos não se preocupou em descrever as possíveis limitações da sua pesquisa ou detalhá-las em seções específicas. Além disso, tais estudos abrem a perspectiva de generalização de seus resultados e aplicação em contextos diferentes, algo que pode não ser factível.

Tabela 6 – Estudos selecionados pela avaliação de qualidade

ID	Autor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Tt	%
E01	Abduljabbar e Omar (2015a)	1,0	1,0	0,5	0,0	1,0	0,5	0,5	4,5	64,3
E02	Alahmadi (2016)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E03	Bandyopadhyay (2012)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E04	Braum, Rigo e Barbosa ()	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E05	Chan <i>et al.</i> (2013)	1,0	1,0	1,0	0,5	1,0	1,0	1,0	6,5	92,9
E06	Chen, Han e Liu (2014)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E07	Chen, Zhang e Levene (2012)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E08	Chernov, Petukhova e Klakow (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E09	Dalavi e Cheke (2014)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E10	Diab e Sartawi (2017)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E11	Dubey e Goyal (2016)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E12	Emmanuel, Khatri e Babu (2013)	1,0	0,5	0,5	0,0	1,0	0,5	0,5	4,0	57,1
E13	Filice, Croce e Basili (2015)	1,0	0,5	0,5	0,0	0,5	0,5	0,5	3,5	50,0
E14	Foley e Allan (2016)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E15	Hao, Xie e Xu (2015)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3

Continua na próxima página

Tabela 6 – continuação

ID	Autor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Tt	%
E16	Hardy e Cheah (2013)	1,0	1,0	1,0	0,5	1,0	1,0	0,5	6,0	85,7
E17	Haris e Omar (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E18	Hoque, Goncalves e Quaresma (2013)	0,5	1,0	0,5	0,5	0,5	0,5	0,0	3,5	50,0
E19	Huang, Thint e Qin (2008)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E20	Huo, Zhang e Zhao (2016)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E21	Hutzler <i>et al.</i> (2014)	1,0	0,5	1,0	0,0	0,5	0,5	0,0	3,5	50,0
E22	Jayakodi, Bandara e Medeniya (2016)	1,0	1,0	1,0	0,5	0,5	0,5	0,5	5,0	71,4
E23	Jayakodi, Bandara e Perera (2015)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E24	Jayakodi <i>et al.</i> (2016)	0,5	1,0	0,5	0,5	0,5	0,5	0,5	4,0	57,1
E25	Blunsom, Grefenstette e Kalchbrenner (2014)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E26	Karyawati <i>et al.</i> (2015)	1,0	1,0	1,0	0,5	1,0	0,5	0,0	5,0	71,4
E27	Kim (2014)	0,5	0,5	1,0	0,0	0,5	1,0	0,5	4,0	57,1
E28	Komninos e Manandhar (2016)	1,0	1,0	0,5	0,0	0,5	0,5	1,0	4,5	64,3
E29	La <i>et al.</i> (2012)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E30	Le e Zuidema (2015)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E31	Lee e Dernoncourt (2016)	0,5	0,5	0,5	0,0	1,0	0,5	1,0	4,0	57,1
E32	Le-Hong, Phan e Nguyen (2015)	1,0	1,0	1,0	0,5	0,5	0,5	0,0	4,5	64,3
E33	Li e Roth (2002)	1,0	1,0	0,5	0,0	0,5	0,5	1,0	4,5	64,3
E34	Li, Tripathi e Srinivasan (2016)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E35	Lin, Wang e McClean (2016)	0,5	1,0	1,0	0,0	0,5	0,5	0,5	4,0	57,1
E36	Lu <i>et al.</i> (2016)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E37	Luo <i>et al.</i> (2014)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E38	Huang e Zhou (2015)	0,5	1,0	1,0	0,0	0,5	0,5	1,0	4,5	64,3
E39	Madabushi e Lee ()	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0

Continua na próxima página

Tabela 6 – continuação

ID	Autor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Tt	%
E40	Mahatme e Bhoyar ()	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E41	Marincic, Kompara e Gams (2012)	0,5	0,5	0,5	0,0	0,5	0,5	1,0	3,5	50,0
E42	Mishra, Kumar e Saha (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E43	Mou <i>et al.</i> (2015)	0,5	1,0	1,0	0,0	0,5	0,5	1,0	4,5	64,3
E44	Obasa, Salim e Khan (2016)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E45	Osadi <i>et al.</i> (2017)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E46	OSMAN e YAHYA (2016b)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E47	OSMAN e YAHYA (2016a)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E48	Patrick e Li (2012)	1,0	1,0	1,0	1,0	1,0	1,0	0,0	6,0	85,7
E49	Pillai e Narayanan (2012)	0,5	0,5	1,0	0,0	1,0	1,0	0,0	4,0	57,1
E50	PING <i>et al.</i> (2012)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E51	Pota <i>et al.</i> (2015)	0,5	1,0	1,0	0,0	0,5	0,5	0,5	4,0	57,1
E52	Pota, Esposito e Pietro (2016)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E53	Poyraz, Kilimci e Ganiz (2014)	1,0	1,0	1,0	0,5	1,0	0,5	0,5	5,5	78,6
E54	Qi <i>et al.</i> (2014)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E55	Qu <i>et al.</i> (2012)	1,0	1,0	0,5	0,5	1,0	1,0	0,5	5,5	78,6
E56	Rahman e Roy ()	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E57	Roberts <i>et al.</i> (2014a)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E58	Roberts <i>et al.</i> (2014b)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E59	SANGODIAH <i>et al.</i> (2017)	1,0	0,5	1,0	0,0	0,5	0,5	0,5	4,0	57,1
E60	Sarrouti, Lachkar e Ou- atik (2015)	1,0	1,0	1,0	0,5	0,5	0,5	0,5	5,0	71,4
E61	Shanthi e Krishna- murthi (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E62	Silva <i>et al.</i> (2011)	1,0	1,0	1,0	0,5	0,5	0,5	1,0	5,5	78,6
E63	Singh e Dwivedi (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0

Continua na próxima página

Tabela 6 – continuação

ID	Autor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Tt	%
E64	Tomas e Vicedo (2013)	1,0	1,0	1,0	0,5	1,0	1,0	0,5	6,0	85,7
E65	Van-Tu e Anh-Cuong (2016)	1,0	1,0	1,0	0,0	0,5	0,5	0,5	4,5	64,3
E66	Wan, Huang e Wei (2013)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E67	Wang e Zhang (2013)	1,0	1,0	1,0	0,5	1,0	1,0	0,5	6,0	85,7
E68	Wang et al. (2016a)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E69	Wang et al. (2016b)	1,0	1,0	1,0	0,0	1,0	1,0	0,5	5,5	78,6
E70	Wang et al. (2015)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E71	Yadav e Mishra ()	0,5	0,5	1,0	0,0	0,5	0,5	0,5	3,5	50,0
E72	Yahya, Osman e El-Bashir (2016)	1,0	1,0	0,5	0,0	1,0	1,0	0,5	5,0	71,4
E73	Yahya e Osman (2014)	1,0	1,0	0,5	0,0	1,0	0,5	0,0	4,0	57,1
E74	Yoshikawa, Iwata e Sawada (2014)	1,0	1,0	0,5	0,0	1,0	1,0	1,0	5,5	78,6
E75	Zhang et al. (2014)	1,0	1,0	0,5	0,0	0,5	0,5	0,0	3,5	50,0
E76	Zhang e Lee (2003)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E77	Zhang, Lee e Radev (2016)	1,0	1,0	1,0	0,0	1,0	0,5	0,5	5,0	71,4
E78	Zhang, Roller e Wallace (2016)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E79	Zhang et al. (2013)	1,0	1,0	0,5	0,0	0,5	0,5	0,5	4,0	57,1
E80	Zhou et al. (2015)	1,0	0,5	1,0	0,0	0,5	0,5	0,5	4,0	57,1
E81	Croce, Filice e Basili (2019)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E82	Dernoncourt (2017)	1,0	1,0	1,0	0,0	1,0	1,0	0,0	5,0	71,4
E83	Fentaw e Kim (2019)	1,0	0,5	1,0	0,0	1,0	1,0	0,0	4,5	64,3
E84	Harrison, Dikken e Peer (2017)	1,0	1,0	1,0	1,0	0,5	0,5	0,0	5,0	71,4
E85	Li, Zhang e Li (2018)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E86	Liu e Guo (2019)	1,0	1,0	1,0	0,0	1,0	1,0	0,0	5,0	71,4
E87	Qiu (2018)	1,0	1,0	1,0	0,5	1,0	1,0	0,0	5,5	78,6
E88	Shin et al. (2018)	0,5	1,0	1,0	0,5	0,5	0,5	0,0	4,0	57,1
E89	Shuang et al. (2020)	1,0	1,0	1,0	0,0	0,5	1,0	0,5	5,0	71,4
E90	Sun et al. (2018)	1,0	0,1	1,0	0,5	1,0	1,0	0,0	5,5	78,6

Continua na próxima página

Tabela 6 – continuação

ID	Autor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Tt	%
E91	Supraja et al. (2017)	1,0	0,1	1,0	0,0	0,5	0,5	0,0	4,0	57,1
E92	Tan (2018)	1,0	1,0	1,0	0,5	1,0	1,0	0,0	5,5	78,6
E93	Wang, Cui e Wang (2017)	0,5	0,5	1,0	0,5	1,0	1,0	0,0	4,5	64,3
E94	Xia et al. (2018)	0,5	1,0	1,0	0,5	1,0	1,0	1,0	6,0	85,7
E95	Yahya (2017)	1,0	0,5	1,0	0,0	1,0	1,0	0,0	4,5	64,3
E96	Zhang et al. (2018)	1,0	1,0	1,0	0,0	1,0	1,0	0,0	5,0	71,4
E97	Zhou e Du (2017)	1,0	1,0	1,0	0,0	0,5	0,5	0,0	4,0	57,1
	Média	0,93	0,93	0,81	0,12	0,68	0,65	0,37	4,48	64

Fonte: Dados da pesquisa.

2.3.4 Extração de Dados

Com base nas diretrizes de [Kitchenham e Charters \(2007\)](#) e na leitura completa dos 97 estudos selecionados, foi elaborado um formulário ([Quadro 6](#)) para extração das informações relevantes. Basicamente, tais informações descrevem uma visão geral de cada estudo e detalham as respostas para as questões de pesquisa propostas.

Quadro 6 – Formulário para extração de dados

Informação para Extração	Descrição	Relevância
ID	Identificador único do estudo	Geral
Ano	Ano de publicação	Geral
País	País de origem	Geral
Base de Questões	Fonte de origem das questões/documentos	Geral
Meio de Publicação	<i>Journal</i> , conferência, <i>workshop</i> , capítulo de livro e arquivo eletrônico	Geral
Tamanho da Amostra	Quantidade de questões utilizadas para análise dos classificadores	Geral
Algoritmos	Métodos computacionais utilizados como Redes Neurais Artificiais, SVM, <i>Naive Bayes</i> etc.	QP1
Crítérios de Classificação	Taxonomias utilizadas para Classificação (ex: Bloom)	QP2
Seleção de Atributos	Extração dos atributos com impacto direto no processo de classificação (TF*IDF, Chi quadrado etc.)	QP3
Métricas de Resultados	Indicadores para análise como acurácia, precisão, <i>recall</i> e <i>F1 Score</i>	QP4

As informações gerais selecionadas foram: ano, país, base de dados das questões ou documentos, meio de publicação do estudo e tamanho da amostra. Considerando as questões de pesquisa, os dados tabulados foram: algoritmos (classificadores), critérios de classificação (taxonomias), técnicas de pré-processamento, seleção de atributos (termos) e indicadores utilizados (métricas).

2.3.5 Análise dos Resultados

A partir dos dados extraídos dos 97 estudos selecionados, este capítulo apresenta os resultados obtidos considerando suas informações gerais e as questões de pesquisa propostas.

2.3.5.1 Ano de Publicação

O intervalo original de observação dos estudos foi entre 2012 e 2017, mas acabou se expandindo para 2019 pela atualização das informações para 2019. Além disso, os 4 estudos mais citados nas referências bibliográficas incluídos no final possuem ano de publicação entre 2002 e 2011. 2016 é o ano com a maioria das publicações com 23,71% dos estudos selecionados, seguido por 2015 (18,56%), 2014 (14,43%), 2012 e 2017 (9,28%), 2013 (8,25%), 2018 (7,22%) e 2019 (4,12%). 2002, 2003, 2008, 2011 e 2020 correspondem às inclusões pontuais com 1,03% cada (Figura 13).

Figura 13 – Quantidade de publicações através dos anos

Fonte: Elaborada pelo autor.

2.3.5.2 País

A China (24%), Estados Unidos (14%) e Índia (10%) concentraram quase metade (48%) dos estudos primários selecionados. Considerando os continentes, a Ásia obteve predominância na publicação dos trabalhos com 60 estudos selecionados (62%); a África teve menos de 1% de representação com um único trabalho publicado (Marrocos); completando a lista, temos Europa com 17,5% das publicações, América com 16,5% e Oceania com 3%. A [Figura 14](#) mostra a distribuição dos estudos entre os países com maior número de publicações:

Figura 14 – Distribuição de estudos por país

Fonte: Elaborada pelo autor.

2.3.5.3 Base de Questões e Meios de Publicação

Todos os 97 estudos primários selecionados tiveram sua origem no ambiente acadêmico o que determinou, por muitas vezes a confecção manual da base de questões utilizando as avaliações produzidas na própria instituição (20%) ou a utilização de bases de dados utilizadas como referência na área de classificação de questões, por exemplo, a base de dados gerada pelo trabalho de [Li e Roth \(2002\)](#) (17,5%) realizado na Universidade de Illinois-Champaign (UIUC) nos Estados Unidos e disponibilizado pela *Text REtrieval Conference* (TREC), que é uma das principais conferências mundiais sobre o tema. Se considerarmos a utilização não exclusiva da base da UIUC/TREC, o índice de uso sobe para 36% dos estudos. Além da área acadêmica, há também diversas fontes de questões disponíveis na web, como os sistemas de QA *Yahoo! Answers* (4%) e *Stack Overflow* (2%); além disso, há bases com armazenamento de textos curtos (similar às questões) como, por exemplo, Reuters (1%), 20NewsGroups (1%) e Baidu (1%), as quais possuem também utilização em conjunto diversas vezes. Um gráfico com as principais bases de questões levantadas é exibido na [Figura 15](#).

Conforme já descrito no [Quadro 6](#), os tipos de meio de publicação adotados neste trabalho foram *journal*, conferência, *workshop*, capítulo de livro e arquivo eletrônico. Os capítulos de livro incluem trabalhos de mestrado e doutorado, enquanto categoria de arquivo eletrônico foi criada para rotular os estudos armazenados na biblioteca eletrônica arXiv.org da Universidade de Cornell e que não foram publicados em outro meio. A maioria de estudos foram publicados em conferências (48,5%), seguidos por *journals* (37%), eletrônicos (7%), capítulos de livro (6%) e *workshops* (1%). .

Figura 15 – Distribuição de estudos por base de questões

Fonte: Elaborada pelo autor.

2.3.5.4 QP1: Métodos Computacionais

A finalidade desta questão de pesquisa foi identificar as diferentes técnicas utilizadas na implementação dos classificadores. Foram observadas duas situações: a primeira se refere à utilização de um simples algoritmo ou combinação/comparação de algoritmos para realizar a classificação; a segunda trata da contribuição principal do estudo, no caso, o foco na lógica de classificação (adaptação de algoritmos-base como SVM, Rede Neural ou criação de novos algoritmos) ou na identificação, extração e organização dos termos como base para o mecanismo classificador.

Analisando os métodos computacionais dos 97 estudos selecionados, 68 (70%) optaram por utilizar um único algoritmo como base para o classificador. A [Tabela 7](#) mostra a distribuição da escolha dos algoritmos e a consequente predominância das Redes Neurais, SVM e Regras.

Tabela 7 – Distribuição dos algoritmos unitários

Algoritmo	Estudos	Qtde.	% (Geral)
Redes Neurais	E16, E25, E27, E30, E31, E38, E43, E69, E71, E77, E78, E80, E82, E83, E85, E86, E87, E88, E89, E90, E94, E96, E97	23	23,7%
SVM	E05, E08, E09, E12, E15, E36, E42, E48, E49, E50, E51, E52, E57, E59, E65, E92, E93	17	17,5%
Regras	E10, E17, E22, E23, E24, E26, E39, E60	8	8,24%
Clustering	E11, E34, E40, E54, E75, E79	6	6,1%
kNN	E06, E66	2	2%
QACS	E35	1	1%
Registro Linear	E61	1	1%
QPT	E03	1	1%
SMM	E74	1	1%
DC2	E64	1	1%
FSM	E18	1	1%
SDM	E14	1	1%
LMQC	E68	1	1%
SNoW	E33	1	1%
<i>Naive Bayes</i>	E53	1	1%
Árvore de Decisão	E41	1	1%
Particle Swarm	E73	1	1%

Com relação aos 29 estudos com combinação de algoritmos, verificou-se que os SVM's também foram a escolha preferida, seguido por *Naive Bayes*, Regras, Árvore de Decisão e Rede Neural (Tabela 8).

Por fim, incorporando os 29 estudos com algoritmos combinados à análise inicial dos 68 estudos com algoritmos unitários, o SVM (39,17%), Rede Neural (29,9%) e Regras (14,4%) se mantiveram nas três primeiras posições, enquanto o *Naive Bayes* (12,37%) acabou entrando na quarta colocação. Considerando a contribuição principal dos estudos na construção de classificadores de questões, os estudos com foco na lógica de classificação prevaleceram com 64 trabalhos (66%, com predominância das Redes Neurais - 19 estudos), enquanto aqueles voltados à estruturação dos termos somaram 33 (34%, com predominância do SVM - 15 estudos).

Os resultados demonstraram que Redes Neurais e SVM foram os dois tipos de classificadores mais utilizados no aprimoramento de algoritmos unitários para classificação. Por outro lado, o SVM foi a opção principal quando a escolha foi realizar a combinação de algoritmos ou propôr uma nova lógica para o tratamento dos termos. Considerando especificamente os 38 estudos com SVM, 19 (50%) eram da área de QA, 7 (18,4%) do e-Learning e 3 (7 %) de RI.

A predominância no uso de Rede Neural e SVM demonstra que o custo de processamento

Tabela 8 – Distribuição dos algoritmos nos estudos com combinações

Algoritmo	Estudos	Qtde.	%(Parcial)	% (Geral)
SVM	E01, E02, E04, E07, E19, E21, E28, E44, E45, E46, E47, E55, E56, E58, E62, E63, E67, E76, E81, E84, E91	21	72,41%	21,65%
<i>Naive Bayes</i>	E01, E02, E20, E21, E32, E44, E45, E46, E47, E55, E76	11	37,93%	11,34%
Regras	E37, E44, E45, E58, E62, E63	6	20,68%	6,18%
Árvore de Decisão	E02, E21, E44, E46, E47, E76	6	20,68%	6,18%
Rede Neural Artificial	E20, E21, E28, E44, E56, E70	6	20,68%	6,18%
KNN	E01, E29, E45, E67, E76	5	17,24%	5,15%
Regressão Logística	E04, E37, E46, E47	4	13,80%	4,12%
Máxima Entropia	E19, E32, E55	3	10,34%	3,09%
SMO	E20, E44	2	6,89%	2,06%
SNoW	E76	1	3,44%	1,00%
Rochio	E72	1	3,44%	1,00%
RHC	E55	1	3,44%	1,00%
Clustering	E70	1	3,44%	1,00%
MPH	E55	1	3,44%	1,00%
S-EM	E07	1	3,44%	1,00%
Probability Estmation	E07	1	3,44%	1,00%
SPH	E55	1	3,44%	1,00%
PSO	E72	1	3,44%	1,00%
Adaboost	E29	1	3,44%	1,00%
Random Forest	E02	1	3,44%	1,00%

vem sendo uma barreira cada vez menor para a viabilização dos experimentos de pesquisa, além da busca por classificadores com os melhores indicadores de desempenho possíveis na tarefa de classificação de questões. Vale ressaltar que essa revisão sistemática não teve objetivo de comparar os resultados dos algoritmos ou suas combinações, pois isso demandaria uma minuciosa análise estatística comparativa.

2.3.5.5 QP2: Taxonomias para Classificação

Esta questão de pesquisa teve como objetivo identificar os critérios mais utilizados para classificação de questões e, por conseguinte, o uso de taxonomias já definidas na literatura. Observou-se que a utilização de critérios próprios (autorais) de cada pesquisa foi predominante. Com a incidência de 41 estudos (42%), englobou desde categorias pertencentes a áreas geográficas e requisitos médicos a sistemas de QA como o *Yahoo! Answers*. Também foi observada

a utilização da classificação binária (Sim/Não, Positivo/Negativo, Relevante/Irrelevante) em 4 estudos (4%). Dentre estes estudos, o SVM foi o algoritmo mais utilizado com 11 estudos (29,7% dos 37), seguido pela combinação de algoritmos com 9 (24% dos 37).

Entre as taxonomias observadas na [Figura 16](#), a taxonomia de Li e Roth [Li e Roth \(2002\)](#) foi a mais adotada com 33 estudos (34%), seguida pela taxonomia de Bloom [Bloom, Krathwohl e Masia \(1984\)](#) com 17 estudos (17,5%) e taxonomias de Costa [Shanthi e Krishnamurthi \(2015\)](#) e Sekine [Tomas e Vicedo \(2013\)](#), ambas com 1 estudo cada (1%).

Figura 16 – Distribuição de estudos por critérios/taxonomia

Fonte: Elaborada pelo autor.

A [Tabela 9](#) mostra a identificação dos estudos em cada uma das taxonomias/critérios encontrados:

Tabela 9 – Identificação dos estudos em cada taxonomia

Taxonomia	Estudos	Qtde.	% (Geral)
Autoral	Demais estudos	39	40,21%
Binária	E44, E54, E56, E75	4	4,12%
Bloom	E01, E10, E11, E17, E22, E23, E24, E45, E46, E47, E59, E72, E73, E84, E91, E92, E95	17	17,53%
Costa	E61	1	1%
Li e Roth	E04, E08, E13, E16, E19, E25, E28, E30, E32, E33, E35, E38, E39, E43, E51, E52, E62, E63, E65, E69, E70, E71, E76, E77, E78, E80, E81, E84, E85, E86, E87, E88, E89, E94, E96, E97	33	36%
Sekine	E64	1	1%

Considerando-se o relacionamento das taxonomias de Li e Roth e de Bloom e os algoritmos unitários implementados (sem combinações, nem taxonomias autorais), verificou-se que as Regras possuem maior utilização com a taxonomia de Bloom (5 dos 17 estudos, 29,4%), enquanto as Redes Neurais com a taxonomia de Li e Roth (18 dos 35 estudos, 51,4%).

Esse resultado reflete a compatibilidade entre a utilização de regras para mapear os níveis de aprendizado propostos por Bloom e também a maior facilidade de adaptação das Redes Neurais a classificações hierárquicas como propuseram Lee e Roth.

2.3.5.6 QP3: Extração e Seleção dos termos

Esta questão de pesquisa procurou identificar quais formas foram utilizadas para a extração dos termos após o pré-processamento inicial das questões e também quais técnicas responsáveis por atribuir pesos ou índices aos termos para seleção dos mais relevantes a uma determinada questão. Após análise dos 97 estudos selecionados, verificou-se a predominância pela escolha extração/seleção, ou seja, apenas 34% dos estudos analisados (33 estudos) descreveram tanto a extração quanto a seleção dos termos utilizados para classificação de questões. Individualmente, a extração esteve presente em 68 estudos (70%) e a seleção em 61 (62%), enquanto apenas 9 estudos não apresentaram técnica para extração/seleção.

Considerando as diferentes formas de extração de termos (Tabela 10), a técnica *Bag-of-Words*, foi a mais utilizada contando com 30 estudos (31%), seguida pelos *N-Grams* com 20 estudos (20,6%) e *Headwords* e *POS-tags*, com 12 estudos cada (12,37%); por outro lado, a seleção de termos esteve presente em 51 estudos ou 52,57%. De forma geral, percebeu-se o grande predomínio dos *N-Grams* e suas variações (*Bag-of-Words*, *Bigrams*, *Trigrams*) com 55,6% dos estudos, ou seja, a utilização de palavras (*tokens*) em sequência extraídos da questão prevaleceu sobre as outras formas envolvendo conceitos/significados, palavras-chaves ou relacionamentos entre os termos da questão.

Tabela 10 – Estudos por tipo de extração de termos

Tipo de Extração	Estudos	Qtde.	% (Geral)
<i>Bag-of-Words</i>	E02, E04, E08, E13, E26, E32, E41, E42, E44, E46, E47, E48, E51, E52, E55, E57, E58, E59, E62, E63, E64, E65, E69, E74, E81, E84, E87, E89, E91, E92	30	31%

Continua na próxima página

Tabela 10 – continuação

Tipo de Extração	Estudos	Qtde.	% (Geral)
<i>N-grams</i>	E04, E05, E08, E17, E19, E25, E34, E37, E38, E42, E46, E47, E55, E65, E70, E76, E80, E84, E87, E91	20	20,60%
<i>Head words</i>	E16, E18, E19, E39, E51, E52, E59, E62, E63, E65, E71, E87	12	12,37%
<i>POS-tags</i>	E05, E08, E17, E33, E45, E46, E47, E58, E59, E69, E71, E87	12	12,37%
<i>Wh-words</i>	E16, E19, E32, E39, E42, E44, E51, E52, E63, E87	10	10,30%
<i>Word Vector</i>	E31, E45, E82, E83, E88, E89, E90, E94, E96, E97	10	10,30%
Entidades	E26, E33, E60, E68, E69, E71, E85	7	7,20%
<i>Bigrams</i>	E41, E48, E51, E58, E63, E64	6	6,18%
<i>Question words</i>	E41, E42, E48, E61, E71, E87	6	6,18%
<i>Hypernyms/Hiponyms</i>	E15, E16, E19, E62, E65, E87	6	6,18%
Verbos	E39, E51, E52, E63	4	4,12%
<i>Skipgrams</i>	E28, E69	4	4,12%
Substantivos	E51, E52, E63	3	3,10%
<i>Synsets</i>	E56, E62	2	2,50%
<i>Keywords</i>	E05, E18	2	2,00%
<i>Multi-Head words</i>	E51, E52	2	2,00%
<i>Trigrams</i>	E63, E64	2	2,00%
<i>Related words</i>	E63, E65	2	2,00%
Conceitos	E68	1	1,00%
<i>Taxonomy-based</i>	E59	1	1,00%
CBOW	E27	1	1,00%
<i>Chunks</i>	E33	1	1,00%
<i>Tree Kernel</i>	E76	1	1,00%
<i>Word Segmentation</i>	E54	1	1,00%
<i>What-WDT</i>	E58	1	1,00%
<i>After-How</i>	E52	1	1,00%
KCRF	E29	1	1,00%
<i>Question Mark</i>	E44	1	1,00%

Continua na próxima página

Tabela 10 – continuação

Tipo de Extração	Estudos	Qtde.	% (Geral)
TOPSIS	E41	1	1,00%
<i>Question Patterns</i>	E65	1	1,00%
Tree Tags	E58	1	1,00%
<i>Focus-words</i>	E08	1	1,00%
Mapas Sintáticos	E39	1	1,00%
<i>Focus-words lemmas</i>	E08	1	1,00%
<i>What-WP</i>	E58	1	1,00%
Árvores Sintáticas	E05	1	1,00%
<i>Bag-of-concepts</i>	E02	1	1,00%
<i>Clue Words</i>	E34	1	1,00%
<i>Question Categories</i>	E65	1	1,00%
<i>Question Expansion</i>	E65	1	1,00%

Fonte: Dados da pesquisa.

No caso das formas de seleção de termos, a técnica TF*IDF foi a mais utilizada com 15 estudos (15,5%), seguida por *Cosine Similarity* e *Word2Vector* com 6 estudos (6,18%), TF com 5 estudos (5,15%), *Binary Vector* com 4 estudos (4,12%) e *Chi-Square*, GLoVe e LDA com 3 estudos ou 3,1% cada. Esse resultado demonstra uma “pulverização” na utilização das técnicas de seleção com um leve domínio do TF*IDF.

Tabela 11 – Estudos por tipo de seleção de termos

Tipo de Seleção	Estudos	Qtde.	% (Geral)
TF*IDF	E04, E05, E17, E44, E46, E47, E49, E54, E56, E79, E84, E87, E91, E93, E95	15	15,50%
<i>Cosine Similarity</i>	E05, E22, E23, E24, E28, E68	6	6,18%
<i>Word2vec</i>	E27, E69, E83, E86, E89, E96	6	6,18%
TF	E44, E53, E72, E73, E95	5	5,15%
<i>Binary Vector</i>	E20, E44, E59, E92	4	4,12%
<i>Chi-Square</i>	E01, E36, E44	3	3,10%
GLoVe	E69, E88, E97	3	3,10%
LDA	E11, E66, E91	3	3,10%
<i>Word Embedding</i>	E28, E89	2	2,00%

Continua na próxima página

Tabela 11 – continuação

Tipo de Seleção	Estudos	Qtde.	% (Geral)
<i>Action Verbs Similarity</i>	E10	1	1,00%
SNOMED	E48	1	1,00%
QF*ICF	E49	1	1,00%
<i>Feature Graph</i>	E25	1	1,00%
CCE	E50	1	1,00%
<i>Gain Ratio</i>	E44	1	1,00%
<i>PIF - Positive Impact Factor</i>	E12	1	1,00%
<i>Semantically Related Words</i>	E33	1	1,00%
ICF	E67	1	1,00%
<i>Symmetrical Uncertainly</i>	E44	1	1,00%
<i>Information Gain</i>	E44	1	1,00%
<i>Yahoo! Place-Maker</i>	E07	1	1,00%
IQF*QF*ICF	E49	1	1,00%
PAS	E48	1	1,00%
KCRF	E29	1	1,00%
<i>Probability</i>	E68	1	1,00%
KFM	E18	1	1,00%
<i>Relatedness</i>	E68	1	1,00%
<i>Term-Weighting Scheme</i>	E09	1	1,00%
<i>SMO wrapper</i>	E44	1	1,00%
BM25	E75	1	1,00%
SVDRD	E50	1	1,00%
<i>Typed Dependencies</i>	E32	1	1,00%
DC2	E64	1	1,00%
Apriori	E06	1	1,00%
TF*ICF	E67	1	1,00%
<i>Levenshtein Distance</i>	E58	1	1,00%
VRF	E49	1	1,00%
MDS	E35	1	1,00%
<i>Word Ranking</i>	E57	1	1,00%
<i>Mutual Information</i>	E01	1	1,00%
WSD	E16	1	1,00%
N2WET	E50	1	1,00%
NWNET	E50	1	1,00%
<i>Odds Ratio</i>	E01	1	1,00%

Fonte: Dados da pesquisa.

2.3.5.7 QP4: Métricas de Resultados

Acurácia, Precisão, *Recall* e F1 (*F-measure*) são indicadores de resultados muito utilizados em estudos acadêmicos e ratificaram essa ideia também nesta RS. A Acurácia foi a métrica mais utilizada, estando presente em 69 dos 97 estudos selecionados (71,13%); em seguida apareceram a Precisão (37 estudos/38,14%), *Recall* (34 estudos, 35,05%) e F1 (30 estudos, 30,92%).

A [Tabela 12](#) mostra o levantamento completo:

Tabela 12 – Métricas utilizadas nos estudos

Métrica	Estudos	Qtde.	% (Geral)
Acurácia	E02, E06, E09, E10, E12, E13, E15, E16, E18, E19, E20, E22, E23, E24, E25, E27, E28, E30, E31, E32, E33, E34, E35, E36, E38, E39, E41, E42, E43, E45, E48, E49, E50, E51, E52, E53, E54, E57, E59, E60, E61, E62, E63, E64, E65, E68, E69, E70, E71, E74, E75, E76, E77, E78, E80, E81, E82, E83, E84, E85, E86, E88, E89, E92, E93, E94, E96, E97	69	71,13%
Precisão	E01, E02, E03, E04, E06, E07, E08, E10, E11, E16, E17, E18, E19, E26, E29, E44, E45, E46, E47, E48, E49, E56, E57, E58, E64, E66, E68, E72, E79, E82, E85, E86, E90, E91, E93, E94, E95	37	38,14%
<i>Recall</i>	E01, E02, E03, E04, E06, E07, E10, E11, E16, E17, E18, E19, E26, E29, E44, E45, E46, E47, E48, E49, E56, E57, E58, E66, E68, E72, E79, E82, E85, E86, E90, E91, E93, E95	34	35,05%

Continua na próxima página

Tabela 12 – continuação

Métrica	Estudos	Qtde.	% (Geral)
F1	E01, E03, E04, E06, E07, E10, E13, E17, E29, E44, E45, E46, E47, E48, E49, E53, E55, E57, E58, E66, E68, E72, E75, E79, E82, E85, E86, E90, E91, E93, E95	30	30,92%
Macro F1	E02, E05, E07, E37, E50, E55, E58, E67, E73	9	9,27%
Micro F1	E02, E05, E07, E37, E50, E55, E58, E67	8	8,24%
Tempo Processamento	E09, E12, E35, E61	4	4,12%
Desvio Padrão	E16, E21, E53, E92	4	4,12%
Taxa de Erro	E10, E21, E34	3	3,09%
<i>Test T</i>	E64, E69	2	2,06%
<i>Total Hits</i>	E46, E47	2	2,06%
Correlação	E46, E47	2	2,06%
Kappa	E46, E47	2	2,06%
<i>M-Average</i>	E10, E26	2	2,06%
<i>Target Hits</i>	E46, E47	2	2,06%
Macro Acurácia	E64	1	1,03%
P@1	E14	1	1,00%
NDCg	E14	1	1,00%
<i>Macro Recall</i>	E37	1	1,00%
Ganho	E50	1	1,00%
RR	E14	1	1,00%
Micro Acurácia	E64	1	1,00%
<i>Overgeneration</i>	E26	1	1,00%
Eficiência	E61	1	1,00%
P@10	E26	1	1,00%
Perplexidade	E36	1	1,00%
Kernel gamma	E49	1	1,00%
<i>Average Sysset</i>	E23	1	1,00%
Distância Média	E21	1	1,00%
Macro Precisão	E37	1	1,00%
<i>Boolean Success Rate</i>	E21	1	1,00%

Continua na próxima página

Tabela 12 – continuação

Métrica	Estudos	Qtde.	% (Geral)
Teste de Significância	E76	1	1,00%
Micro Precisão	E37	1	1,00%
<i>Undergeneration</i>	E26	1	1,00%
<i>Micro Recall</i>	E37	1	1,00%
MRR	E26	1	1,00%
Não informado	E40	1	1,00%

Fonte: Dados da pesquisa.

2.4 Considerações finais

O foco desta revisão sistemática foi buscar o estado-da-arte nas técnicas, critérios e indicadores associados à classificação de questões. Este artigo fez um levantamento sobre estudos envolvidos direta e especificamente com a classificação de questões e desconsiderou trabalhos com escopo mais amplo como a classificação de textos (que engloba as questões) ou escopo associado e foco diferente (sistemas de QA, IR e conversão de idiomas) como, por exemplo, tratamento de documentos, arquitetura do sistema ou expressões de um determinado idioma.

Noventa e sete estudos foram selecionados através dos critérios de inclusão, exclusão e de qualidade nos quais dois tipos de métodos computacionais principais foram identificados, 6 taxonomias/critérios, 2 formas de controle dos termos e os 4 indicadores mais utilizados na aferição dos classificadores de questões.

Entre os dois tipos de métodos computacionais utilizados, a grande maioria dos estudos optou por contribuições na lógica de classificação. As Redes Neurais, o SVM e a combinação de algoritmos se destacaram como principais métodos computacionais utilizados.

O tratamento de termos ocupou destaque tanto na identificação dos algoritmos quanto nas técnicas de seleção e extração, pois cerca de um terço dos trabalhos (32,98%) implementaram novos algoritmos para esse tratamento e quase 90% (91,75%) possuíam algum mecanismo de extração/seleção.

Os critérios/taxonomias e os indicadores não apresentaram novidades em relação ao descrito na literatura ([Subseção 2.2.7](#) e [Subseção 2.2.8](#)): a utilização de categorias definidas pelos próprios autores se sobressaiu da mesma forma que o grupo de indicadores Acurácia/Precisão/*Recall*/*F1-score* se mostraram as métricas mais adotadas para aferição do desempenho dos classificadores.

Como ameaças à validade e limitações encontradas destacam-se a utilização de apenas um pesquisador no levantamento e a possível adoção de decisões subjetivas aplicadas aos critérios de inclusão/exclusão/qualidade aplicados aos estudos selecionados. Estes fatores podem ter influenciado diretamente na ausência de alguma fonte de busca que acarretou na não inclusão de algum estudo importante e/ou exclusão de estudos relevantes aos objetivos de pesquisa.

CLASSIFICAÇÃO AUTOMÁTICA DE QUESTÕES COM BASE NAS COMPETÊNCIAS DO ENEM

3.1 Considerações Iniciais

Conforme já explicitado anteriormente neste trabalho, o grande volume de dados textuais disponível em formato digital culminou com a necessidade de organização da informação disponível para extração do conhecimento. Inicialmente de forma manual, depois através de sistemas especialistas e, finalmente, via algoritmos de Aprendizado de Máquina, a automatização dessa organização ganhou força, principalmente em tarefas de classificação ou agrupamento de dados.

Nesse contexto, a área educacional sempre foi grande geradora de informações digitais, disponibilizando na *web* uma infinidade de recursos como livros digitais, avaliações *online*, listas de exercícios e vídeo-aulas. Os sistemas de *e-learning* são um dos principais geradores de dados textuais e, entre eles, há os sistemas de aprendizagem adaptativa capazes de recomendar atividades específicas de acordo com o desempenho individual de cada aluno.

Para atingir esse objetivo, é necessário organizar os conteúdos e recursos educacionais dentro de algum tipo de critério, de forma a viabilizar o acompanhamento das diferentes trilhas de aprendizagem que cada estudante percorre. Como trilha de aprendizagem entende-se, neste trabalho, o caminho de atividades realizado pelo estudante no processo de aprendizagem de um determinado conhecimento. Por exemplo, se o estudante tem dificuldades nas disciplinas de Português e Biologia, faz todo sentido acompanhar quais assuntos dentro destas disciplinas necessitam ter algum tipo de reforço ou mudança na estratégia de ensino.

O Exame Nacional do Ensino Médio (ENEM) é uma avaliação em larga escala realizada

no Brasil. Tem abrangência nacional e é utilizado como processo de entrada para estudantes nas principais universidades públicas do país. Seus conteúdos são organizados em um conjunto de competências e habilidades denominado Matriz de Referência de Avaliação (MRA). Apesar do desempenho de cada aluno avaliado por competência/habilidade ser devidamente registrado, neste momento, só são disponibilizados resultados de forma agregada (escola, cidade, região) e isso impede o acompanhamento individualizado dos estudantes pelos seus professores.

No contexto supracitado, esta tese de doutorado realizou um estudo de caso baseado em milhares de questões do ENEM com o objetivo principal de buscar a melhor arquitetura possível para um classificador automático baseado em competências. As competências definidas na matriz de referência constituíram as categorias usadas para classificação, enquanto a arquitetura do classificador foi estruturada em 3 componentes principais: (i) modelos de representação do texto das questões, (ii) fórmulas para cálculo da relevância (peso) dos termos e (iii) algoritmos de Aprendizado de Máquina supervisionados para indução do modelo de classificação.

Este capítulo está organizado da seguinte forma: a [Seção 3.2](#) faz uma revisão da literatura sobre classificadores automáticos de questões considerando o ambiente educacional; a matriz de referência de avaliação do ENEM é apresentada na [Seção 3.3](#), enquanto a [Seção 3.4](#) descreve um modelo geral para a construção de um classificador automático a partir de seus componentes principais e níveis hierárquicos de competências, além das propostas nele baseadas.

3.2 Trabalhos Relacionados

Relacionados ao escopo desta pesquisa, em alguns trabalhos foram propostas abordagens para o diagnóstico educacional através da classificação de questões. Nessa área, uma possibilidade é classificar os itens de forma que eles possam representar um nível de avaliação cognitiva dos estudantes associada a uma determinada competência ([RAHMAN, 2015](#); [ABDULJABBAR; OMAR, 2015b](#)). O levantamento realizado na literatura encontrou uma diversidade de algoritmos computacionais mas utilizando, em sua maioria, a taxonomia de Bloom ([BLOOM; KRATHWOHL; MASIA, 1974](#)) como critério para classificação e o Aprendizado de Máquina como paradigma computacional. Esses trabalhos são apresentados nas seções seguintes.

3.2.1 Trabalhos Relacionados Baseados na Taxonomia de Bloom

A taxonomia de Bloom consiste em um sistema baseado em objetivos educacionais para classificação do grau de aprendizado e entendimento a partir da proposição de níveis hierárquicos ([ABDULJABBAR; OMAR, 2015b](#)). A taxonomia possui 3 domínios (cognitivo, afetivo e psicomotor), mas os trabalhos de classificação de questões costumam considerar apenas o domínio cognitivo pela maior disponibilidade de avaliações para o diagnóstico de competências na área educacional. Este domínio é similar às avaliações escolares e é composto por seis níveis hierárquicos iniciando do nível mais simples, o conhecimento, até o mais complexo,

Quadro 7 – Seis categorias do domínio cognitivo baseadas na taxonomia de Bloom

Categoria	Exemplos de questões
Conhecimento	Defina o conceito de herança
Compreensão	Explique a estrutura de um método em um programa
Aplicação	Demonstre a relação entre pacotes, classes e métodos
Análise	Liste as vantagens de utilizar uma classe do tipo <i>container</i>
Síntese	Crie um programa em Java mostrando o conceito de sobrecarga
Avaliação	Justifique o conceito de herança e escreva um código-fonte exemplo

a avaliação. O [Quadro 7](#) lista cada nível do domínio cognitivo, ilustrando exemplos simples ([ABDULJABBAR; OMAR, 2015b](#)):

[SANGODIAH et al. \(2017\)](#) propuseram um novo tipo de termo chamado de *taxonomy feature*. Basicamente, após a extração da *Bag-of-Words* e uso do *POS-Tagger* para identificação dos verbos utilizados, há um tratamento para identificar a forma como eles foram aplicados na questão e o seu relacionamento com os níveis da taxonomia. Feito isso, a partir destes verbos, os *taxonomy features* são gerados e incorporados aos termos extraídos da questão. O experimento utilizou classificador *SVM*, cálculo binário para os pesos e bases de treinamento e teste geradas através da técnica *10-fold cross validation* resultando numa acurácia final de 72,9 %.

O trabalho de [Osadi et al. \(2017\)](#) propôs um *ensemble* com 4 classificadores: regras, *SVM*, *Naive Baye* e *KNN*. A partir de módulos de pré-processamento e dois tipos de termos são gerados: *POS-tag* e Word Vectors. O primeiro baseado em análise morfológica e usado para o classificador baseado em regras e o segundo seguindo a *Bag-of-Words* e usado para os demais classificadores. Por fim, há um componente que combina o resultado dos classificadores e a similaridade com termos e conceitos da base de termos léxicos da *WordNet* gerando uma acurácia de 82%.

Também usando a base de termos léxicos da *WordNet*, [Diab e Sartawi \(2017\)](#) desenvolveram um algoritmo para comparar verbos de ação extraídos das questões e verbos associados à taxonomia com base nos termos da *WordNet* para análise de similaridade. A acurácia obtida foi de 83% e um valor de 90% para o *F1-score*.

[Yusof e Hui \(2010\)](#) utilizaram redes neurais artificiais para propôr um modelo de classificação de questões de exames baseado na taxonomia de Bloom. Tal taxonomia é largamente utilizada no sentido de mensurar e avaliar objetivos educacionais, mas a identificação do nível de cognição é realizada manualmente. Outros problemas apontados são a pouca quantidade de acadêmicos com qualificação para realizar a identificação, muitos erros de classificação e o alto consumo de tempo no processo. Apesar das redes neurais serem muito utilizadas na categorização de objetos, [Yusof e Hui \(2010\)](#) mostraram que se fazia necessário um estudo com redes que pudessem classificar contextos textuais diferenciados, tais como os níveis cognitivos de questões. O modelo proposto implementa a abordagem da rede artificial neural, a qual é treinada utilizando o algoritmo de aprendizagem do gradiente conjugado escalonado. São utilizadas

diversas técnicas de pré-processamento de texto como extração de palavras, remoção por *stop word* e *stemming* (redução de palavras derivadas à sua palavra raiz) resultando em um conjunto inicial de termos. A partir deste conjunto, é criado um vetor de características para cada item de avaliação que será processado pela rede neural artificial para determinar a qual nível de cognição ele está associado. Os autores reportaram 65,9% de precisão para a melhor abordagem avaliada.

Haris e Omar (2012b), Omar *et al.* (2012b) utilizaram uma abordagem baseada em regras para analisar e classificar questões através do processamento da linguagem natural utilizando o domínio cognitivo de Bloom como referência. Este domínio foi escolhido por ser largamente reconhecido e familiar na academia, genérico e aplicável a uma gama diversa de assuntos. O domínio também é facilmente aplicado a diferentes tipos de questões devido à sua estrutura simplificada. A metodologia aplicada foi dividida em duas fases: pré-processamento e desenvolvimento de regras de classificação. A fase de pré-processamento consistiu em preparar as questões para aplicação das regras. Entre as operações de pré-processamento estão a conversão do texto do item para formato minúsculo, remoção via *stopwords*, *stemming*, geração de *tokens* (quebra do texto em um conjunto de palavras únicas e significativas, símbolos, para análise) e marcação dos *tokens* através da utilização de um *parser* (analisador sintático). Para a definição das regras de classificação (Figura 17) de cada item de avaliação, algumas condições foram consideradas. Entre elas estão o significado do verbo utilizado na questão e se ele está relacionado com ações associadas a um nível cognitivo, palavras específicas de questionamento (*what, why, when, which e how*), além das ambiguidades detectadas pelo *parser* e resolvidas por expressões regulares. Com isso, são identificadas estruturas sintáticas e então classificadas em um determinado nível cognitivo. O classificador construído com este método obteve 77% de precisão.

O trabalho desenvolvido por Yahya e Osama (2011) utilizou um classificador baseado no SVM para aproveitar seu ótimo desempenho na classificação de questões utilizando os níveis cognitivos de Bloom como critério. O SVM é um algoritmo de aprendizado de máquina supervisionado cuja habilidade é a separação de elementos entre dois conjuntos e que é muito eficaz na classificação de textos. O método aplicado utilizou 3 etapas principais: a representação do texto (responsável pelo pré-processamento, mapeamento do texto para um formato compactado) e a construção e avaliação dos classificadores SVM. Como resultado, obteve-se uma precisão de 87,4% com os classificadores.

Chang e Chung (2009) projetaram um sistema de avaliação *online* baseado em palavras-chaves também associadas aos níveis cognitivos de Bloom. O sistema conta com uma base de questões, os quais são analisados a partir, especificamente, dos verbos contidos em cada item. Os verbos possuem um relacionamento com as palavras-chaves e, assim, o nível de cognição é determinado. O experimento com o sistema piloto mostrou eficiência apenas para o nível "Conhecimento" com um índice de 75%.

Abduljabbar e Omar (2015b) também propuseram um novo método de classificação automática de acordo com os níveis cognitivos de Bloom. A estratégia é baseada em um algo-

Figura 17 – Algoritmo para aplicação das regras de classificação

```
Input: question, Qi
category = ""
preprocessing
read POS/Phrases stored in dictionary
return POS/Phrases

if Qi.keyword == 'define' then proceed to check next
occurring word
    if regex = "struct", then
        if POS/Phrases == "IN CD NN" from the question
            and followed by
            specific term match to "of type"
            assign sub-category = "Define"
            Apply rule1 for "Define-Application"
            Apply rule2 for "Define-Synthesis"
            Make a selection
            search in the array of subcategory to get category
            return category that contains "Define"
        end if
```

Fonte: Haris e Omar (2012b).

ritmo de votação que combina três classificadores de aprendizado de máquina. Os classificadores utilizados são o *SVM*, *Naive Bayes* e o *KNN* que são integrados em um algoritmo para aproveitamento de suas melhores vantagens. A classificação encontra o melhor resultado aplicando a estratégia de combinação através do modelo chamado *Mutual Information* (Informação Mútua), o qual provou ser promissor e comparável a modelos similares. Os experimentos demonstraram eficiência na integração de diferentes métodos e algoritmos de classificação. Os resultados foram comparados à pesquisa de Haris e Omar (2012b) e provou ser mais eficiente com 92% de acerto.

O grande problema dos classificadores baseados na taxonomia de Bloom é a sua dependência do idioma. Todos os estudos citados fazem algum tipo de processamento do texto para extrair termos que tenham algum tipo de compatibilidade com os níveis de cognitivos da taxonomia. Isso implica em criar uma tabela ou regra para associação desses termos, o que gera a necessidade direta de uma análise morfológica do texto em cada idioma para identificar componentes (verbos e substantivos, por exemplo) para estabelecer a correlação da questão com a taxonomia.

3.2.2 Outros Trabalhos

Sun *et al.* (2018) propuseram um algoritmo de rede neural profunda (*deep neural network*) capaz de associar *tags* a questões de múltipla-escolha. A partir dos termos extraídos da

resposta de cada questão e um mapa de conhecimento baseado no currículo-base de conteúdos, o algoritmo associa *tags* com base na posição (se for uma questão com espaços em branco para preenchimento) ou cálculo de peso. Apesar da característica multi-rótulo, este estudo foi destacado pela associação de um mapa conceitual como referência para atribuição das *tags* e o uso de questões de multi-escolha (caso do ENEM). O estudo relatou o valor do F1-score em 82%.

Com base na taxonomia de Li e Roth, [Huang e Zhou \(2015\)](#) desenvolveram uma rede convolucional usando árvores de termos gerados com *n-grams* de diferentes tamanhos organizados em uma árvore hierárquica de termos. Apesar do contexto não ser exatamente educacional, este trabalho foi referenciado por estar entre as maiores acurácias encontradas para problema de classificação de questões: 95,6% para o nível mais agregado da taxonomia e 89% para o nível mais detalhado.

[Rahman \(2015\)](#) realizou uma revisão bibliográfica sobre a classificação de questões utilizando uma abordagem estatística. A pesquisa relata o problema da manutenção e da especificidade associadas às regras de classificação, além do seu caráter artesanal. O foco do trabalho é realizar um levantamento das pesquisas em classificação de questões que utilizam técnicas de aprendizado de máquina como *SVM* e redes neurais onde a base ferramental de suporte é estatística. As áreas abordadas envolveram sistemas de resposta, recuperação de informação, ambientes educacionais e classificação de questões em idiomas diferentes do inglês. Os resultados da revisão apontaram que os trabalhos recentes estão utilizando análises sintáticas e semânticas profundas para enfrentar o problema da pouca quantidade de texto presente nas questões. Também se constatou que a eficiência dos classificadores se mostrou mais aparente nas áreas correspondentes a sistemas de resposta e recuperação de informação. Também foi verificada uma predominância dos classificadores *SVM* com boa performance em dados textuais não estruturados.

[Fei et al. \(2003\)](#) implementaram técnicas de redes neurais artificiais para classificação automática de questões de múltipla escolha utilizadas em sistemas de *e-learning*. O critério de classificação foi representado pela utilização de três diferentes níveis de dificuldade (difícil, médio e fácil). No trabalho foi utilizado um vetor de características de cinco dimensões como entrada para uma rede neural do tipo *backpropagation*. As dimensões do vetor consistem em relevância do texto, frequência do termo significativo, comprimento da pergunta e resposta, distribuição da frequência do termo e distribuição da pergunta e resposta no texto. Os autores reportaram precisão de cerca de 73% para a realização dessa tarefa.

O modelo de classificação proposto por [Cetintas et al. \(2014\)](#) aborda o universo dos problemas matemáticos no sentido de identificar dois tipos de sentenças nele embutidos: "relevantes" ou "irrelevantes". Esta classificação representa um importante passo para o cálculo dos níveis de dificuldade de tais problemas. Trata-se de um modelo de classificação probabilística conjunta que utiliza a correlação entre sentenças matemáticas utilizando dois classificadores *SVM*. Um

extensivo conjunto de experimentos determinou a efetividade do modelo probabilístico com 71% de acurácia, mas reconheceu que esse desempenho necessita ser melhorado com a continuidade da pesquisa.

Nuntiyagul *et al.* (2008) apresentaram uma ferramenta de apoio à aprendizagem adaptável para gerenciamento de bancos de questões. Além da classificação automática de itens em classes pré-definidas, também foi capaz de recuperar itens pela especificação de classes ou níveis de dificuldade. O sistema adaptou o modelo de aprendizagem de categorização para aprimorar o desempenho da classificação utilizando questões de entrada. Os resultados mostraram que a acurácia foi aceitável e satisfaz a necessidade dos usuários.

Ma *et al.* (2015) utilizaram redes neurais convolucionais com *n-grams* baseados em árvores para descobrir relações entre palavras distantes dentro de uma mesma sentença. A partir de experimentos com classificação de questões oriundas da base de dados TREC, os pesquisadores conseguiram atingir o **estado da arte**, obtendo acurácia de 95,6% para as classes mais gerais e 89% para as classes mais detalhadas, de acordo com a taxonomia de Li e Roth.

3.3 Matriz de Referência de Avaliação do ENEM

Sistemas de ensino-aprendizagem, caso dos sistemas de *e-learning* e, por conseguinte, dos sistemas de aprendizagem adaptativa, têm como objetivo final aferir a proficiência dos estudantes em uma determinada área do conhecimento. Para isso, é necessário definir quais os requisitos mínimos de conhecimento necessários de modo que possam ser elaboradas questões para composição dos testes diagnósticos. Uma forma de definição e organização desses requisitos é a **Matriz de Referência de Avaliação (MRA)**. Considerando especificamente o contexto de avaliações em larga escala, a matriz de referência estabelece o que deve ser avaliado em cada etapa da escolarização, orienta a elaboração de testes e níveis de proficiência para definir o grau de conhecimento efetivamente adquirido pelo estudante.

Na prática, uma matriz de referência de avaliação corresponde a um conjunto de habilidades e competências essenciais organizadas em agrupadores hierárquicos chamados descritores (BLASIS *et al.*, 2014; PESTANA, 1998). Os descritores são cruzamentos ou associações entre dois pontos básicos para avaliação: o conteúdo programático a ser avaliado em cada período de escolarização e o nível de operação mental necessário para a habilidade avaliada (FERREIRA, 2014). Exemplos de avaliações baseadas em matrizes de referência aplicadas no Brasil são: ENEM, SAEB, Prova Brasil e Fundação Universitária para o Vestibular (FUVEST). Os descritores para a disciplina Língua Portuguesa no caso da Prova Brasil podem ser observados na Figura 18.

No caso do ENEM, foco deste trabalho, a organização da matriz de referência é feita de forma hierárquica com as áreas de conhecimento (Ensino Médio) e suas disciplinas no topo da árvore; no próximo nível há a divisão por competências e essas, por sua vez, detalhadas nas

Figura 18 – Exemplo de matriz de referência de avaliação (Prova Brasil)

Língua Portuguesa – 5º ano do Ensino Fundamental	Língua Portuguesa – 9º ano do Ensino Fundamental
Descritores do Tópico I. Procedimentos de Leitura	Descritores do Tópico I. Procedimentos de Leitura
D1 – Localizar informações explícitas em um texto. D3 – Inferir o sentido de uma palavra ou expressão. D4 – Inferir uma informação implícita em um texto. D6 – Identificar o tema de um texto. D11 – Distinguir um fato da opinião relativa a esse fato.	D1 – Localizar informações explícitas em um texto. D3 – Inferir o sentido de uma palavra ou expressão. D4 – Inferir uma informação implícita em um texto. D6 – Identificar o tema de um texto. D11 – Distinguir um fato da opinião relativa a esse fato.
Descritores do Tópico II. Implicações do Suporte, do Gênero e /ou do Enunciador na Compreensão do Texto	Descritores do Tópico II. Implicações do Suporte, do Gênero e /ou do Enunciador na Compreensão do Texto
D5 – Interpretar texto com auxílio de material gráfico diverso (propagandas, quadrinhos, foto etc.). D9 – Identificar a finalidade de textos de diferentes gêneros.	D5 – Interpretar texto com auxílio de material gráfico diverso (propagandas, quadrinhos, foto etc.). D12 – Identificar a finalidade de textos de diferentes gêneros.
Descritores do Tópico III. Relação entre Textos	Descritores do Tópico III. Relação entre Textos
D15 – Reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema, em função das condições em que ele foi produzido e daquelas em que será recebido.	D20 – Reconhecer diferentes formas de tratar uma informação na comparação de textos que tratam do mesmo tema, em função das condições em que ele foi produzido e daquelas em que será recebido. D21 – Reconhecer posições distintas entre duas ou mais opiniões relativas ao mesmo fato ou ao mesmo tema.
Descritores do Tópico IV. Coerência e Coesão no Processamento do Texto	Descritores do Tópico IV. Coerência e Coesão no Processamento do Texto
D2 – Estabelecer relações entre partes de um texto, identificando repetições ou substituições que contribuem para a continuidade de um texto. D7 – Identificar o conflito gerador do enredo e os elementos que constroem a narrativa. D8 – Estabelecer relação causa /consequência entre partes e elementos do texto. D12 – Estabelecer relações lógico-discursivas presentes no texto, marcadas por conjunções, advérbios etc.	D2 – Estabelecer relações entre partes de um texto, identificando repetições ou substituições que contribuem para a continuidade de um texto. D7 – Identificar a tese de um texto. D8 – Estabelecer relação entre a tese e os argumentos oferecidos para sustentá-la. D9 – Diferenciar as partes principais das secundárias em um texto. D10 – Identificar o conflito gerador do enredo e os elementos que constroem a narrativa. D11 – Estabelecer relação causa/consequência entre partes e elementos do texto. D15 – Estabelecer relações lógico-discursivas presentes no texto, marcadas por conjunções, advérbios etc.
Descritores do Tópico V. Relações entre Recursos Expressivos e Efeitos de Sentido	Descritores do Tópico V. Relações entre Recursos Expressivos e Efeitos de Sentido
D13 – Identificar efeitos de ironia ou humor em textos variados. D14 – Identificar o efeito de sentido decorrente do uso da pontuação e de outras notações.	D16 – Identificar efeitos de ironia ou humor em textos variados. D17 – Reconhecer o efeito de sentido decorrente do uso da pontuação e de outras notações. D18 – Reconhecer o efeito de sentido decorrente da escolha de uma determinada palavra ou expressão. D19 – Reconhecer o efeito de sentido decorrente da exploração de recursos ortográficos e/ou morfosintáticos.
Descritores do Tópico VI. Variação Linguística	Descritores do Tópico VI. Variação Linguística
D10 – Identificar as marcas linguísticas que evidenciam o locutor e o interlocutor de um texto	D13 – Identificar as marcas linguísticas que evidenciam o locutor e o interlocutor de um texto

Fonte: Blasis *et al.* (2014).

habilidades necessárias. As áreas e disciplinas são organizadas da seguinte forma:

- Linguagens, códigos e suas tecnologias, que abrange o conteúdo de Língua Portuguesa, Literatura, Língua Estrangeira (Inglês ou Espanhol), Artes, Educação Física, e Tecnologias da Informação e Comunicação.
- Matemática e suas tecnologias.
- Ciências da Natureza e suas tecnologias, que abrange os conteúdos de Química, Física e Biologia.
- Ciências Humanas e suas tecnologias, que abrange os conteúdos de Geografia, História, Filosofia e Sociologia.

As competências são relacionadas a cada área do conhecimento. Dessa forma, disciplinas de uma mesma área tem competências e habilidades comuns para avaliação, mudando apenas a forma de aplicação determinada pelos conteúdos próprios de cada disciplina. A [Figura 19](#) apresenta as 30 competências abordadas pela matriz de referência do ENEM, cuja última atualização foi no ano de 2012. A matriz completa, incluindo as 120 habilidades (divididas por área e competência) é apresentada no [Apêndice A](#).

Figura 19 – Competências da matriz de referência do ENEM

Fonte: Adaptada de Miranda, Ferreira e Dias (2019).

3.4 Propostas para Classificação de Questões Baseada em Competências do ENEM

O objetivo central deste trabalho concentrou-se na obtenção do melhor classificador automático possível para questões, com base nas competências definidas na matriz de referência do ENEM. A ideia inicial do trabalho era realizar a classificação usando diretamente as habilidades como critério, mas a grande quantidade de classes possíveis (120), a distribuição muito irregular de questões entre as habilidades dentro da mesma competência (poucas questões geram poucos termos) impossibilitou, nesse momento da pesquisa, a obtenção desse nível de detalhe com a configuração de dados disponível. Dessa forma, decidiu-se por utilizar as 30 competências definidas para as 4 áreas do conhecimento como critério final de classificação [Seção 3.3](#).

Com as competências do ENEM definidas como critério de classificação e disponibilidade de uma base com questões assim classificadas, a pesquisa foi direcionada para utilização de classificadores supervisionados, ou seja, com modelos gerados a partir de uma base de questões com rótulos definidos (classes ou categorias) manualmente. A [Figura 20](#) ilustra um esquema geral para esse tipo de classificação de textos.

Figura 20 – Esquema ilustrativo da classificação automática de textos por meio de aprendizado supervisionado

Fonte: [Rossi \(2016\)](#).

Considerando o esquema da [Figura 20](#), as propostas apresentadas nesta tese impactam diretamente as etapas de indução do modelo de classificação e a aplicação do modelo propriamente dito. Na etapa de indução, as propostas pretenderam contribuir na representação do texto das questões e no cálculo de pesos dos termos relevantes; já na aplicação do modelo, o objetivo foi contribuir na elaboração/aplicação dos algoritmos. Associando esse esquema ilustrativo com

Figura 21 – Foco da pesquisa

os estágios do processo de classificação apresentado no [Capítulo 2](#), gerou-se um esquema com o foco de atuação desta trabalho de pesquisa, conforme mostra a [Figura 21](#).

Para a construção das propostas, decidiu-se por definir um modelo geral com um sequenciamento dos processos essenciais para a construção da melhor arquitetura para o classificador automático. Esse modelo é apresentado na próxima seção.

3.4.1 Modelo Geral

A construção de um classificador envolve uma série de requisitos que englobam os dados que serão classificados, critérios de classificação, representação computacional destes dados, seleção das informações mais relevantes, aplicação de algoritmos classificadores, entre outros. Considerando os objetivos deste trabalho e seu contexto educacional, um modelo geral para a construção do classificador automático de questões foi desenvolvido de forma a estruturar a sequência essencial de processos. Esse modelo é representado na [Figura 22](#), se baseou na técnica *Cross Industry Standard Process for Data Mining* (CRISP-DM) e seguiu as seguintes diretrizes para sua aplicação:

- As questões devem pertencer a avaliações em larga escala (ENEM, SARESP, Prova Brasil etc.) e compostas por questões objetivas.

- A elaboração as questões deve ter sido baseada em um conjunto de competências organizadas de forma hierárquica do nível mais agregado para o mais detalhado.
- As avaliações em larga escala devem ter seu formato digital disponível para utilização, preferencialmente, pela internet em formatos de fácil importação como texto ou PDF.
- Deve haver uma base de dados de questões pré-classificadas por especialistas de acordo com as competências associadas.
- Deve haver um ambiente para desenvolvimento, execução e coleta de resultados (programadores, algoritmos, linguagens de programação como Java e Phyton ou sistemas específicos como Weka e R). O ambiente necessita de capacidade de processamento compatível com o tamanho da base de dados desejada e os requisitos computacionais dos algoritmos.
- Nenhum processo deve obrigar a implementação de regras específicas para um idioma.

Descrição e sequenciamento dos processos:

1. **Compreensão do Contexto:** Avaliar a disponibilidade das avaliações em larga escala de forma a poder extrair seu conteúdo em formato digital de forma rápida e segura. Ao mesmo tempo, avaliar se as questões estão organizadas em níveis hierárquicos de competências.
2. **Compreensão dos Dados:** A partir das questões extraídas em formato digital, é necessário criar um estrutura de armazenamento computacional para gravação e consulta dessa informação (sistema de arquivos/diretórios ou tabelas de banco de dados). Para armazenamento, são criados códigos numéricos associados às questões, competências e seus relacionamentos hierárquicos. Cada nível de competência possui sua base de questões classificadas. Essa etapa pode retornar à etapa anterior caso haja algum problema com as competências definidas ou com a classificação das questões nessas competências.
3. **Preparação dos Dados:** O próximo passo é realizar as operações de pré-processamento para excluir termos não relevantes e unificar variações de termos em um único termo. Em seguida, são avaliadas as formas computacionais para representação do texto das questões e aplicadas técnicas para cálculo da relevância (peso) dos termos em relação ao conjunto de questões. Com isso, são extraídos e selecionados os termos mais importantes e que serão utilizados como entradas para o algoritmo de classificação.
4. **Estimar Modelos:** São selecionados os algoritmos com melhor desempenho prévio publicado na literatura para classificação de textos curtos e questões. Um ambiente computacional é preparado para realização de experimentos controlados que envolvem a parametrização dos algoritmos, controle de repetições e do processo gerenciamento dos conjuntos de questões para treinamento e teste. Todo o processo é executado para cada nível de competência definido com sua respectiva base de questões.

Figura 22 – Modelo de estruturação e aplicação de classificadores baseado em competências para avaliações em larga escala

5. **Avaliar e Interpretar:** Realizados os experimentos, são gerados indicadores para medição do desempenho. A melhor combinação entre as possíveis representações de texto, cálculos de pesos e algoritmos de classificação obtida é registrada para cada nível de competência. Esse registro é feito através de indicadores de desempenho padrão obtidos com a matriz de contingência gerada pelo processo anterior. Entre os indicadores padrão podem ser utilizados a acurácia, precisão, *recall* e *F1-score* (micro e macro). Caso os indicadores não sejam satisfatórios, todo o processo volta para a etapa inicial e é realizada uma reinterpretação do contexto.

6. **Implementar e Divulgar:** Com os resultados em mão, resta montar a arquitetura final com as combinações obtidas para cada nível e disponibilizar o modelo computacional obtido para a classificação de novas questões em ambiente de produção.

O modelo apresentado possui algumas limitações: (i) dependendo da quantidade de termos extraídos das questões e algoritmos escolhidos, pode ser que, mesmo utilizando uma capacidade computacional elevada, os tempos de resposta para geração de resultados não sejam factíveis para utilização do classificador obtido em ambientes de produção; (ii) uma distribuição muito irregular das questões entre as competências definidas nas questões pré-classificadas podem inviabilizar o balanceamento e prejudicar a confiabilidade dos resultados; (iii) avaliações em larga escala sem estrutura hierárquica de competências não possuem aplicação para o modelo. Uma evolução possível para esse modelo seria embutir os níveis de competências em algoritmos de classificação hierárquica como o classificador SNoW de desenvolvido por [Li e Roth \(2002\)](#).

3.4.2 Método

Como estratégia para obter o melhor classificador de questões possível, foram definidas duas propostas: (i) identificar qual a melhor combinação de componentes de um classificador possível para gerar uma versão chamada aqui de *baseline* e, a partir desta versão, (ii) apresentar um algoritmo para aprimoramento dos resultados. Foram considerados como componentes do classificador, o modelo adotado para representação do texto da questão, o cálculo de importância dos termos e o algoritmo-base para o classificador. Há de se considerar que o modelo de representação de texto impacta diretamente na escolha do algoritmo, pois estes componentes envolvem a tarefa de pré-processamento do texto para definir como serão definidos os termos e a forma que os mesmos se relacionam com cada documento/questão, enquanto o cálculo de pesos é uma etapa independente de modelo de representação e algoritmo. As etapas para apresentação das propostas foram organizadas da seguinte forma:

- **Geração da *baseline*:**

1. **Disciplinas e competências:** comparação dos principais classificadores/representações de texto, com as mesmas fórmulas de cálculo de peso dos termos utilizando dois critérios de classificação: áreas/disciplinas e competências do ENEM. Como as competências são definidas por áreas e compartilhadas pelas disciplinas da área, decidiu-se realizar o processo de classificação em 2 etapas: em primeiro lugar, o nível mais agregado para obter a disciplina relacionada e, em seguida, a competência final no nível mais detalhado.
2. **Cálculo de pesos (*term-weighting*):** em conjunto com a comparação dos classificadores e representações de texto, foram analisadas diferentes técnicas para o cálculo de relevância dos termos. Foram utilizados os métodos tradicionais TF e $TF * IDF$, conforme mostrou o [Capítulo 2](#), além do método $IQF * QF * ICF$, considerado o melhor quando se trata de tarefas de classificação de questões, de acordo com o estudo de [Quan, Wenyin e Qiu \(2010\)](#) que não englobou o ambiente educacional, mas apenas os ambientes de *Question Answering*.

3. **Conjunto de classificadores:** a partir do melhor resultado individual da avaliação dos classificadores no nível de área/disciplina, foi realizado o cálculo acumulado dos indicadores em reunião com os melhores resultados da classificação por competências para o processamento do resultado final.
- **Algoritmo $IMBHN^{COS}$:** considerando os resultados obtidos na geração da *baseline* e o algoritmo *textitIndutive Model Bipartite Heterogeneous Network* (*IMBHN*) desenvolvido por [Rossi et al. \(2014\)](#) (baseado no modelo de redes) foi proposta a medida de similaridade do cosseno para otimizar os resultados da classificação de novas questões de acordo com o modelo gerado pelo algoritmo *IMBHN*, estado da arte na classificação de documentos textuais ([ROSSI, 2016](#)).

3.4.3 Geração da Baseline

A busca da melhor forma de classificar questões envolve construir um ponto de partida (*baseline*) a partir do estado-da-arte existente nos componentes necessários para a montagem de um classificador. Dessa forma, as etapas associadas à comparação de classificadores para definição da *baseline* já tiveram seus principais possíveis componentes descritos no [Capítulo 2](#). Relativamente a essas etapas, vale destacar que a combinação de componentes foi definida para gerar um classificador com lógica hierárquica, ou seja, composto por um conjunto de classificadores com melhor desempenho no nível mais agregado (disciplina) e no nível mais detalhado (competências).

O [Capítulo 2](#) mostrou também que o método $TF * IDF$ foi a principal técnica utilizada para cálculo de pesos entre os estudos avaliados na revisão sistemática e representa uma referência quando o assunto é classificação de textos. [Quan, Wenying e Qiu \(2010\)](#) desenvolveram um estudo sobre cálculo de pesos voltado especificamente para a classificação de questões e propuseram três novos métodos supervisionados: $QF * ICF$, $IQF * QF * ICF$ e VRF . São chamados de métodos supervisionados porque, como os classificadores, também dependem de questões rotuladas (classificadas) nas respectivas categorias para que os cálculos sejam realizados. Para poder apresentar as formulações relacionadas aos métodos e considerando T como o termo sendo analisado, é necessário definir alguns conceitos ([QUAN; WENYIN; QIU, 2010](#)):

- **Categoria positiva:** classe da questão à qual T pertence no momento da análise.
- **Categoria negativa:** classes diferentes da classe positiva.
- N : total de questões da coleção.
- C : total de categorias da coleção.
- tp : número de questões que contém T na categoria positiva.
- fp : número de questões que não contém T na categoria positiva.

- **tn**: número de questões que contém T na categoria negativa.
- **fn**: número de questões que não contém T na categoria negativa.
- **qf**: frequência de questões (*question frequency*) que contém T na categoria positiva, dada pela fórmula: $\log(tp + 1)$.
- **cf**: número de categorias na qual não contém T ocorre (*category frequency*).
- **icf**: força discriminativa de um termo sobre todas as categorias (*inverse category frequency*), dada pela fórmula: $\log\left(\frac{|C|}{cf} + 1\right)$.

A partir destes conceitos [Quan, Wenyin e Qiu \(2010\)](#) propuseram 3 métodos para cálculo de pesos:

- **$QF * ICF$** : valoriza os termos com alto valor de qf na categoria positiva mas que aparecem muito pouco nas outras categorias:

$$QF * ICF = \log(tp + 1) * \log\left(\frac{|C|}{cf} + 1\right)$$

- **$IQF * QF * ICF$** : baseado no $QF * ICF$, também considera o número de questões na categoria negativa (fn) através do termo IQF , gerando a seguinte fórmula:

$$IQF * QF * ICF = \log\left(\frac{N}{tp + fn}\right) * \log(tp + 1) * \log\left(\frac{|C|}{cf} + 1\right)$$

- **VRF** : também relaciona tp e fn medindo a proporção entre eles de forma a indicar que quanto maior a proporção da categoria positiva sobre a negativa em relação ao termo analisado, maior a influência do termo na classificação final.

$$VRF = \frac{\log(tp + 1)}{\log(fn + 1)}$$

Os estudos desenvolvidos por [Quan, Wenyin e Qiu \(2010\)](#) utilizaram dados do ambiente de *Question Answering* e que, apesar de não estarem no foco desta tese dessa forma, representam um ótimo indicador pois todas as técnicas comparadas compartilharam os mesmos dados. Entre as técnicas utilizadas no comparativo estão TF , $TF * IDF$, $TF * IG$, $TF * \chi^2$, $TF * RF$ e $TF * OR$, além de diferentes de coleções de texto e os algoritmos *KNN* e *SVM* como classificadores. O método $IQF * QF * ICF$ obteve ampla vantagem e, por isso, juntamente com os métodos tradicionais TF e $TF * IDF$, foi selecionado como etapa da geração da *baseline*, cujos resultados da análise comparativa foram descritos no [Capítulo 4](#).

3.4.4 Algoritmo $IMBHN^{COS}$

O trabalho de Rossi (2016) demonstrou a superioridade do algoritmo $IMNH$ na classificação de textos. No caso das questões, pela menor quantidade de texto e, consequentemente, menor número de relações entre questões e termos, o mesmo resultado não foi observado. Conforme apresentado mais a frente no Capítulo 4, a versão para redes bipartidas (somente relações de documentos e textos) $IMBHN$ obteve desempenho inferior aos algoritmos KNN e MNB na maioria dos resultados. Dessa forma, decidiu-se por aprimorar o processo de classificação desse algoritmo.

O algoritmo $IMBHN$, semelhantemente aos algoritmos SVM , MNB e as redes neurais, se baseia na indução de peso dos termos para geração do modelo de classificação. As seções seguintes descrevem os detalhes do algoritmo, exemplo de funcionamento e a otimização realizada.

3.4.4.1 Algoritmo $IMBHN$

O objetivo do algoritmo é induzir informações de classes para os termos ($F(T)$) como modelo de classificação usando a relação documento-termo (questão-termo) (ROSSI, 2016). O algoritmo $IMBHN$ é a versão para redes bipartidas do algoritmo $IMHN$ e essa opção foi feita devido à característica das coleções de questões, onde somente a relação documento-termo é representada. O algoritmo possui duas versões para indução das informações de classe: (i) $IMBHN^C$, baseado na produção de um erro de classificação e (ii) $IMBHN^R$, que realiza uma regressão. As funções de custo e atualização de pesos para as duas versões do algoritmo $IMBHN$ não são foco desta tese e podem ser encontradas em Rossi *et al.* (2014) e Rossi (2016).

O pseudocódigo para algoritmo $IMBHN$ é apresentado no algoritmo 1. A função $ClassificarDocumento(w_{d_i}, F(T))$ retorna um vetor com a informação de classes para o documento corrente: na versão $IMBHN^C$, o vetor possui valores binários com 1 para a classe encontrada e 0 para as demais; no caso da versão $IMBHN^R$, o vetor traz pesos para todas as classes. A chamada da função é repetida até que um dos valores pré-definidos para limiar de erro ou número de iterações seja atingido.

A informação de classe de um documento d_i para uma classe c_j é definida por uma função linear pesada apresentada na Equação 3.1. A função f_{t_k, c_j} usada para ponderar o peso associados a cada termo do documento representa exatamente a consulta aos pesos do modelo induzido.

$$f_{d_i, c_j} = \sum_{t_k \in T} w_{d_i, t_k} * f_{t_k, c_j} \quad (3.1)$$

Para classificar um novo documento, em ambas as versões do algoritmo $IMBHN$, as informações devem ser inseridas na Equação 3.1 que produz um vetor com informações para todas as classes. O maior valor do vetor (arg-max) é usado então para definir o rótulo de classe para o novo documento.

Algoritmo 1 – IMBHN (Adaptado)**Entrada:**

D = Coleção de documentos

T = Coleção de termos

 ε = Conjunto de relações da rede

W = Conjunto de pesos

Y(D) = Conjunto das classes rotuladas, uma para cada documento

 η = Taxa de correção de erro**Saída:**

F(T) = Modelo induzido com a informação de classes

```

1  início
2  repita
3      para cada  $d_i \in D^L$  faça
4          InfClasses = ClassificarDocumento( $w_{d_i}, F(T)$ )
5          para cada  $c_j \in C$  faça
6              erro =  $y_{d_i, c_j} - \text{InfClasses}[c_j]$ 
7              para cada  $t_k \in T | \exists e_{d_i, t_k} \in \varepsilon$  faça
8                   $f_{t_k, c_j} = f_{t_k, c_j} + \eta * w_{d_i, t_k} * \text{erro}$ 
9              fim
10         fim
11     fim

```

Fonte: Rossi (2016).

3.4.4.2 Exemplo de Funcionamento

Considere uma rede bipartida com 2 classes possíveis, 7 termos e 6 documentos com todos os pesos das relações iguais e igual a 1. A lista de adjacências com os termos de cada documento são representadas pelas ligações com seus termos (sem a exibição do peso de cada termo por documento, apenas o peso final da classe por termo). A Figura 23 apresenta a sequência de passos do algoritmo para classificação de 2 novos documentos, utilizando como base o processamento de informações de classes da versão *IMBHN*^R. Sobre cada documento e cada termo há um vetor com o valor de peso para cada uma das classes.

Inicialmente, como mostra a Figura 23a, os vetores sobre os termos estão zerados e os 7 documentos rotulados tem o seu valor correspondente no vetor marcado com o valor 1 e demais valores do vetor com 0. Na sequência, a Figura 23b são apresentados nos vetores sobre os termos as informações de classe dos objetos da rede obtidos pelo algoritmo *IMBHN*^R com $\eta=1.0$ e $\varepsilon=0.01$. Pelos pesos calculados, percebe-se que os termos que ocorrem em vários documentos rotulados em uma classe possuem valores para a classe na posição correspondente de seu vetor pesos. Menores ocorrências em documentos de uma respectiva classe geram valores menores

(a) Rede bipartida de exemplo com as informações de classes iniciais dos objetos da rede

(b) Definição das informações de classes dos termos utilizando o algoritmo $IMBHN^R$ com $\eta = 1.0$ e $\epsilon = 0.01$

(c) Definição das informações de classe de novos documentos por meio da propagação das informações de classes dos termos relacionados

(d) Definição dos rótulos dos documentos considerando o $\arg\max$ das informações de classe

(e) Legenda.

(e) Legenda

Figura 23 – Exemplo de funcionamento do algoritmo $IMBHN$

que podem, inclusive, ser negativos.

A partir da [Figura 23c](#) é ilustrado o processo de classificação de um novo documento. Os termos dos novos documentos são analisados e, caso já tenham pesos calculados anteriormente, tem a informação de classes atualizadas pela [Equação 3.1](#). Verificando, por exemplo, o documento Doc 7, percebe-se que o mesmo está associado a termos que possuem um valor de informação de classe no vetor de pesos maior para a Classe 1, enquanto o Doc 8 possui maior valor no vetor de pesos para a Classe 2. O resultado final da indução de informações de classe e atribuição do rótulo de classe para os 2 novos documentos é mostrado na [Figura 23d](#).

3.4.4.3 Otimização da Classificação através da Medida do Cosseno

Conforme visto nas seções anteriores, o algoritmo *IMBHN* utiliza uma função para indução de pesos e geração da informação de classes e, após a indução, usa a mesma função para obter qual a classe associada aos termos do documento com maior valor agregado de pesos (arg-max). Esta estratégia tem ótimos resultados em coleções de textos comuns, com capacidade de gerar uma grande quantidade de termos e relações documento-termo. O problema é que questões (foco desta tese) possuem uma pequena quantidade de texto, na maioria das vezes, e isso prejudica o desempenho do algoritmo *IMBHN*, como pode ser visto no [Capítulo 4](#).

Para otimizar o funcionamento do algoritmo *IMBHN*, optou-se por dar uma visão do modelo espaço-vetorial para os vetores de pesos das classes gerados pelo modelo induzido. O modelo em rede utilizado nesta tese para rede bipartidas utilizou uma lista de adjacências relacionada a cada questão para armazenar os pesos dos respectivos termos. A partir do processamento das questões e de suas listas pelo algoritmo *IMBHN* e utilização da função de indução apresentada na seção anterior, são gerados vetores de pesos para as classes contendo a ponderação final com relevância de cada termo em relação a uma determinada classe.

A ideia elaborada para a otimização do algoritmo visou utilizar a abordagem do modelo espaço-vetorial para analisar os vetores de classes induzidos pelo modelo gerado. Dessa forma, cada classe pode ter seu vetor gerado pelo modelo com N termos representado em um espaço N -dimensional. Dessa forma, pode-se imaginar que, ao invés de utilizar o arg-max do valor de peso entre os vetores de classes para a classificação de um novo documento via função de indução do modelo, pode-se usar uma função para cálculo da similaridade que compare o vetor gerado pelos termos de um novo documento com o vetor de termos gerado pelo modelo para cada classe. A função escolhida para este cálculo foi a medida do cosseno ([SALTON, 1989](#)), cujo valor final definido pela [Equação 3.2](#) reflete a similaridade do vetor de pesos de um documento d_i com o vetor de pesos induzidos dos termos para uma classe c_j ; termos não presentes no documento d_i ficam com valor zero no respectivo vetor. Esta função mede o ângulo entre os vetores gerados, ou seja, quanto mais próximo o resultado for de 1, mais próximos estão os

vetores, e assim, mais similares eles são.

$$\text{Cos}(d_i, c_j) = \frac{\sum_{t_k \in T} w_{d_i, t_k} * f_{t_k, c_j}}{\sqrt{\sum_{t_k \in T} w_{d_i, t_k}^2} * \sqrt{\sum_{t_k \in T} f_{t_k, c_j}^2}} \quad (3.2)$$

Através da similaridade cosseno, tornou-se possível analisar a distribuição de pesos dos termos induzidos para os vetores das classes comparada aos pesos nos vetores de cada documento, ao invés de usar um simples valor máximo como fazia o *arg-max*. Essa versão do algoritmo *IMBHN* foi nomeada como *IMBHN^{COS}* e teve como ponto de partida o algoritmo *IMBHN^R* pela sua característica de induzir os pesos das classes via regressão, por manter o processamento do erro de classificação sem zerar os pesos das classes não selecionadas no processo de indução. Os resultados de desempenho na classificação de questões, incluindo a avaliação da *baseline* e a comparação das duas versões do algoritmo *IMBHN* são detalhados no [Capítulo 4](#).

3.5 Considerações Finais

Este capítulo apresentou, inicialmente, os trabalhos relacionados à classificação de questões dentro do ambiente educacional ([Seção 3.2](#)) e a estrutura definida para matriz de referência de avaliação do ENEM ([Seção 3.3](#)). Já na [Seção 3.4](#), um modelo geral para a construção de classificadores automáticos foi proposto, considerando as avaliações em larga escala. A partir desse modelo, foram apresentadas as possibilidades existentes para construção do classificador automático com base em três componentes: representação do textos da questão, cálculo do peso/relevância dos termos da questão e algoritmo de Aprendizado de Máquina Supervisionado. Nesse contexto, foi proposta a construção de uma combinação destes componentes para a geração de um classificador final denominado de *baseline*. Juntamente com essa *baseline*, também foi proposta a otimização do algoritmo *IMBHN* baseado no modelo em redes (*IMBHN^{COS}*).

AVALIAÇÃO EXPERIMENTAL

4.1 Considerações Iniciais

Este capítulo apresenta os resultados da análise comparativa entre diferentes configurações de classificadores de questões baseados em algoritmos de Aprendizado de Máquina.

A MRA do ENEM foi utilizada como base para o critério de classificação, onde a cada competência foi atribuído um código de classe único para associação às questões que cumpram os requisitos da respectiva competência.

A partir de uma base de dados com questões do ENEM oriundas de diversos anos, foram comparados algoritmos com representações de texto baseadas no modelo espaço-vetorial e em redes, métodos diferentes para o cálculo de peso dos termos, além do nível de detalhe das competências.

O objetivo final foi encontrar a configuração que pudesse gerar o melhor desempenho possível e assim satisfazer as hipóteses e objetivos de pesquisa. Com essa finalidade, a partir do modelo geral definido no [Capítulo 3](#), foi gerada uma adaptação dos processos voltada especificamente para o caso das questões do ENEM conforme mostra a [Figura 24](#).

Este capítulo está organizado da seguinte forma: a [Seção 4.2](#) descreve os componentes existentes na configuração experimental, incluindo base de dados, tarefas de pré-processamento e algoritmos de classificação, enquanto a [Seção 4.3](#) apresenta uma avaliação empírica das propostas apresentadas através da comparação de seus indicadores de desempenho.

Figura 24 – Aplicação do modelo geral para o caso das questões do ENEM

4.2 Configuração

Foi gerada uma base de dados com 25321 mil questões classificadas por especialistas em educação oriundas de exames do ENEM aplicados até 2017. As questões estão distribuídas nas 4 áreas do conhecimento, 7 disciplinas principais, 30 competências e 120 habilidades, conforme mostra a [Tabela 13](#). O detalhamento da relação entre áreas/disciplinas e suas respectivas competências e habilidades pode ser conferido no [Apêndice A](#).

O número de termos extraídos foi relativamente pequeno quando comparado a uma base de documentos textuais convencional. Com exceção da disciplina Português, o número de termos não chegou ao dobro do número de questões para cada disciplina. Isso dificulta o processo de descoberta de padrões e indução de modelos para classificação. Para se chegar a essa quantidade de termos extraídos, foram realizadas as seguintes tarefas de pré-processamento:

Tabela 13 – Distribuição da base de questões por Área e Disciplina

Disciplina	Nº Questões	Nº Termos	Área	Detalhamento
Português	3901	10507	Linguagens, Códigos e suas Tecnologias	9 competências e 30 habilidades
Matemática	4438	3306	Matemática e suas Tecnologias	7 competências e 30 habilidades
Biologia	3599	5406	Ciências da Natureza e suas Tecnologias	8 competências e 30 habilidades
Física	4578	4179		
Química	3851	4507		
História	2505	5343	Ciências Humanas e suas Tecnologias	6 competências e 30 habilidades
Geografia	2449	4589		

- **BOW**: uso da *Bag-of-Words*, onde os termos são compostos por palavras simples.
- **Padronização**: remoção de números, símbolos e *stopwords*.
- **Stemming**: redução de variações da mesma palavra (gênero, número, grau, por ex.) ao seu radical (*stem*) mais simples. Foi utilizado o algoritmo de Porter aplicado à língua portuguesa (PORTER, 2006).
- **Redução da dimensionalidade**: não foi aplicada pelo número relativamente pequeno já discutido.

Para não permitir distorções na análise dos resultados, foi realizado um balanceamento dos dados, ou seja, para cada classe (disciplina ou competência) foi atribuída, de forma aleatória, o mesmo número de questões para classificação. O processo de execução dos classificadores foi realizado através da técnica *10-fold cross validation*, ou seja, cada conjunto de questões foi dividido, aleatoriamente, em 10 partes iguais com 9 usadas para treinamento e 1 para teste do classificador; esse processo é repetido 10 vezes trocando apenas a parte utilizada para testes de forma que, ao final do processo, todas as 10 divisões tenham estado tanto no treinamento quanto no teste. Como resultado final, são gerados indicadores médios (TAN; STEINBACH; KUMAR, 2016) que foram selecionados de acordo com os resultados da revisão sistemática apresentada no Capítulo 2: acurácia, precisão (micro e macro), *recall* (micro e macro) e F1 (micro e macro). Para melhorar o entendimento, decidiu-se basear a avaliação nos indicadores acurácia e macro-F1 (micro-F1 é equivalente à acurácia quando a tarefa de classificação atribui apenas um rótulo/classe): o primeiro, por mostrar diretamente quantidade de acertos; já o segundo, por ser uma média harmônica dos indicadores precisão e *recall*, demonstrar a qualidade deste acertos na coleção de questões.

Com relação aos algoritmos de classificação, buscou-se na literatura aqueles com melhor desempenho na classificação de textos, considerando tanto os baseados no modelo espaço-vetorial (AGGARWAL; ZHAI, 2012; WEISS; INDURKHYA; ZHANG, 2015; BENBRAHIM;

BRAMER, 2009; DUMAIS *et al.*, 1998), quanto os algoritmos baseados em rede (ROSSI, 2016; ROSSI *et al.*, 2014). São eles:

1. **KNN**: versões com e sem votos ponderados pela distância com a medida cosseno, com $K \in \{1; 3; 5; 7; 9; 11; 13; 15; 17; 19; 21; 25; 29; 35; 41; 49; 57; 73; 89\}$.
2. **MNB**: sem parâmetros.
3. **SVM**: utilizados os tipos mais comuns (*linear*, *polynomial* e *radial basis*), com parâmetro $C \in 10^{-2}; 10^{-1}; 10^0; 10^1; 10^2$.
4. **IMBHN**: algoritmo baseado em rede com melhor desempenho em classificação de textos desenvolvido por Rossi *et al.* (2014). foram usadas tanto as versões com inferência de classe via classificação quanto regressão (IMBHN^C e IMBHN^R, respectivamente). Como parâmetros, foram considerados $\varepsilon = 0,01$ (erro mínimo), $\eta \in \{0,01; 0,05; 0,1; 0,5\}$ (taxa de correção) e número de interações fixado em 1000.

Para comparação dos algoritmos, em cada análise foram utilizadas três técnicas para cálculo dos pesos: TF , $TF * IDF$ e $IQF * QF * ICF$. Além disso, pela grande variedade de parâmetros, apenas o melhor resultado para cada algoritmo/técnica foi considerado.

O gerenciamento da base de dados, tarefas de pré-processamento, processamento de diferentes classificadores parametrizados e geração de resultados foram reunidos em uma **API** para viabilizar a integração destas funcionalidades em diferentes plataformas voltadas à área educacional. Entre as principais funcionalidades da **API** destacam-se:

- **Base de Dados**: leitura e gravação de arquivos texto e importação de questões e níveis de competências.
- **Pré-processamento**: tokenização em *BOW*, *Bigrams* e *Trigrams*, *stemming* (inglês e português).
- **Classificadores**: parametrização e implementação de diversos algoritmos classificadores, geração de matrizes documento-termo (Espaço-Vetorial) e lista de adjacências (redes), cálculo e gravação de pesos (TF , $TF * IDF$ e $IQF * QF * ICF$).
- **Experimentos**: configuração de repetições e conjuntos treinamento/teste (k-fold-validation), geração da matriz de contingências e formatação/exportação dos resultados para planilhas de cálculo.

4.3 Resultados

A primeira etapa da avaliação consistiu em comparar o resultado dos classificadores listados na Seção 4.2 de 2 maneiras: primeiro, para efeitos comparativos, considerando toda a base

de questões classificadas diretamente nas competências da MRA para avaliar o desempenho dos algoritmos considerados o estado da arte (Tabela 14); depois, usando as disciplinas presentes nas áreas de conhecimento (Português, Matemática, Física, Química, Biologia, História e Geografia) para realizar um primeiro filtro antes da classificação por competência (Tabela 15).

Considerando estas disciplinas como critério de classificação, toda a coleção de questões foi submetida aos classificadores, cujos resultados são apresentados na . Conforme já mencionado, tanto os algoritmos baseados no modelo espaço-vetorial, quanto o algoritmo baseado em rede tiveram seus termos gerados a partir da *bag-of-words* para a representação do texto das questões. Os melhores resultados foram destacados e assinalados em negrito.

O algoritmo *KNN* obteve o melhor desempenho geral, seguido por *MNB*, *SVM* e *IMBHN^C*, todos com indicadores próximos ao valor de 95%. Vale ressaltar o papel essencial do método $IQF * QF * ICF$ que comprovou gerar melhores resultados que os métodos tradicionais para todos os algoritmos, sem exceção. O melhor exemplo disso é o próprio *KNN* que teve a pior *performance* entre todos os algoritmos usando $TF * IDF$ e a segunda pior com TF , ou seja, a ponderação via $IQF * QF * ICF$ impactou diretamente no cálculo da distância entre os vetores gerados pelos termos. Considerando apenas os métodos TF e $TF * IDF$, os melhores resultados ficaram com *SVM* e *MNB*, respectivamente.

Com relação, especificamente, ao desempenho das versões do algoritmo *IMBHN* (baseado no modelo em redes), todas obtiveram desempenho inferior aos algoritmos baseados no modelo espaço-vetorial, diferentemente do resultado obtido por Rossi (2016) com coleções de textos variadas. O algoritmo *IMBHN^C* ficou em quarto lugar e *IMBHN^R* em quinto, respectivamente (ambos com $IQF * QF * ICF$). Mesmo assim, os valores de acurácia e macro-F1 foram relevantes, tanto para o algoritmo *IMBHN^C* (94,36%, 0,9497) quanto para o *IMBHN^R* (91,90%, 0,9140).

Os resultados apresentados na Tabela 15 mostram que há uma grande probabilidade do classificador acertar a disciplina correta em uma nova coleção de questões. Com isso, o próximo passo foi identificar quais dos classificadores utilizados produzem os melhores resultados com base nas competências como critério de classificação. Nesse contexto, a Tabela 16 apresenta os indicadores de desempenho obtidos, todos usando agora apenas a técnica $IQF * QF * ICF$, devido à sua superioridade em relação às demais.

Analisando os maiores valores para Acurácia e macro-F1 nota-se que os algoritmos *KNN* e *MNB* levam ampla vantagem: O algoritmo *KNN* teve melhor desempenho em 4 disciplinas (Matemática, Física, Química e História), enquanto o *MNB* em 3 (Português, Biologia e Geografia). Além disso, nos casos em que não tiveram os melhores resultados, estes algoritmos estiveram perto disso: *KNN* ficou em segundo lugar em 2 disciplinas (Biologia e Geografia) e terceiro em uma (Português), enquanto o *MNB* ficou em segundo em 3 disciplinas (Física, Química e História) e terceiro em uma (Português).

Tabela 14 – Estado da arte

Algoritmo / Cálculo de Pesos	Acurácia (%)	Erro (%)	Desvio Padrão (%)	Macro-F1
KNN / TF	40,7304	59,2695	0,8347	0,1586
KNN / TF*IDF	43,1970	56,8029	0,8353	0,1765
KNN / IQF*QF*ICF	54,1824	45,8175	0,6393	0,2735
MNB / TF	43,0722	56,9277	0,8495	0,1336
MNB / TF*IDF	44,5613	55,4386	0,5625	0,1889
MNB / IQF*QF*ICF	56,0209	43,9790	1,0642	0,2493
SVM / TF	41,0506	58,9493	0,7399	0,1816
SVM / TF*IDF	41,8048	58,1951	1,6880	0,5166
SVM / IQF*QF*ICF	61,1871	38,8128	0,9785	0,3800
IMHN ^C / TF	35,3895	64,6104	0,7129	0,1669
IMBHN ^C / TF*IDF	35,2231	64,7768	0,7436	0,1645
IMBHN ^C / IQF*QF*ICF	56,4952	43,5047	0,8678	0,3450
IMBHN ^R / TF	31,8206	68,1793	1,0880	0,0719
IMBHN ^R / TF*IDF	37,9351	62,0648	0,9037	0,0988
IMBHN ^R / IQF*QF*ICF	44,9481	55,0518	0,8286	0,1487

Tabela 15 – Classificação geral por disciplina

Algoritmo / Cálculo de Pesos	Acurácia (%)	Erro (%)	Desvio Padrão (%)	Macro-F1
KNN / TF	83,6256	16,3744	0,5018	0,8267
KNN / TF*IDF	86,1730	13,8270	0,8552	0,8587
KNN / IQF*QF*ICF	95,1540	4,8460	0,2810	0,9497
MNB / TF	88,9376	11,0624	0,5431	0,8845
MNB / TF*IDF	89,4392	10,5608	0,6029	0,8901
MNB / IQF*QF*ICF	94,7354	5,2646	0,2667	0,9447
SVM / TF	88,9731	11,0269	0,4939	0,8832
SVM / TF*IDF	88,9692	11,0308	0,4869	0,8832
SVM / IQF*QF*ICF	94,6011	5,3989	0,9410	0,9415
IMHN ^C / TF	88,8349	11,1651	0,4860	0,8808
IMBHN ^C / TF*IDF	89,2378	10,7622	0,5089	0,8857
IMBHN ^C / IQF*QF*ICF	94,3681	5,6319	0,9395	0,9395
IMBHN ^R / TF	83,4518	16,5482	0,8924	0,8256
IMBHN ^R / TF*IDF	87,7607	12,2393	0,6562	0,8708
IMBHN ^R / IQF*QF*ICF	91,9076	8,0924	0,9117	0,9140

Tabela 16 – Classificação geral por competência

Disciplina /Algoritmo	Acurácia (%)	Erro (%)	Desvio Padrão (%)	Macro-F1
Português				
KNN	74,6193	25,3807	4,30915	0,64613
MNB	77,03675	22,96325	2,38459	0,68101
SVM	66,79484	33,20516	3,94876	0,57546
IMBHN ^C	76,57966	23,42034	3,25829	0,65912
IMBHN ^R	73,90587	26,09413	2,39737	0,63619
Matemática				
KNN	92,47967	7,52032	2,28205	0,92629
MNB	88,59466	11,40534	6,44778	0,88187
SVM	84,47735	15,52265	4,47358	0,86499
IMBHN ^C	87,38095	12,61905	4,32518	0,86643
IMBHN ^R	88,60627	11,39373	5,52403	0,88468
Biologia				
KNN	89,73005	10,26995	2,88904	0,89913
MNB	91,47157	8,52842	4,61951	0,91729
SVM	79,78858	20,21142	5,48732	0,80421
IMBHN ^C	84,70139	15,29861	4,68700	0,85091
IMBHN ^R	85,9042	14,0958	4,03323	0,86120
Física				
KNN	88,64268	11,35732	3,121491	0,887376
MNB	87,9545	12,0455	4,42148	0,87862
SVM	84,52979	15,47021	3,94223	0,84953
IMBHN ^C	86,10318	13,89682	4,03765	0,86252
IMBHN ^R	85,80525	14,19475	4,22660	0,85830
Química				
KNN	89,64286	10,35714	5,59412	0,89689
MNB	87,60629	12,39371	4,702321	0,876283
SVM	83,84354	16,15646	6,273759	0,841145
IMBHN ^C	84,2602	15,7398	5,00126	0,84441
IMBHN ^R	80,34864	19,65136	5,14925	0,80537
História				
KNN	84,32323	15,67677	5,40508	0,84308
MNB	83,88889	16,11111	5,34786	0,84192
SVM	67,80303	32,19697	6,37097	0,70393
IMBHN ^C	74,70202	25,29798	7,40826	0,75445
IMBHN ^R	81,43939	18,56061	5,95870	0,81813
Geografia				
KNN	79,47529	20,52470	4,06325	0,80503
MNB	83,10166	16,89833	2,94877	0,83154
SVM	73,01861	26,98138	4,41035	0,73629
IMBHN ^C	77,23156	22,76843	3,32898	0,77335
IMBHN ^R	77,59018	22,40981	3,46808	0,77780

Com relação aos piores desempenhos, fica a nota negativa para o algoritmo *SVM*, considerado na literatura um dos melhores algoritmos de Aprendizado de Máquina supervisionados: obteve o pior desempenho em 6 das 7 disciplinas (Português, Matemática, Biologia, Física, História e Geografia) e foi penúltimo somente em (Química). Conclui-se com isso que os tipos de *kernel* e demais parâmetros selecionados não foram os ideais para a coleção de questões apresentada. Uma alternativa possível seria desenvolver um estudo de álgebra linear para identificar um espaço gerado pelos vetores para definição de um *kernel* mais adequado para execução do algoritmo para o perfil reduzido de texto utilizado utilizado nas questões do ENEM.

Comparando o desempenho por competência, os maiores valores para os indicadores (acurácia e macro-F1) foram encontrados nas disciplinas de Matemática (*KNN*, 92,47%, 0,9262) e Biologia (*MNB*, 91,47%, 0,9172); já os menores valores foram encontrados para Português (*SVM*, 66,79%, 0,6810) e História (*SVM*, 92,47%, 0,9262).

Por fim, é necessário reunir os resultados obtidos na classificação por disciplinas e competências para geração de um resultado final de classificação para *baseline*. Para isso, os melhores resultados apresentados nas Tabelas 15 e 16 foram cruzados para descobrir as melhores combinações, todos com a técnica $IQF * QF * ICF$. Os algoritmos *KNN* e *MNB* demonstraram ser os melhores classificadores tanto no nível de disciplinas quanto no nível de competências, como mostra a Tabela 17.

O melhor resultado, novamente, foi apresentado pelo algoritmo *KNN* na disciplina de Matemática com 87,99% de acurácia, seguido pelo algoritmo *MNB*, na disciplina Biologia com 86,65%. Já o pior resultado final foi obtido com o algoritmo *MNB* na disciplina de Português (72,98%). Vale destacar que todas as combinações obtiveram acurácia maior que 70% sendo que e 9 das 14 combinações possíveis obtiveram acurácia entre 80 e 90%.

Considerando apenas o melhor resultado final, somente na disciplina Português a classificação obteve um desempenho mais baixo (73,30 %); os demais resultados foram de, no mínimo, 79% de acurácia.

Tabela 17 – Resultado final a partir da combinação dos classificadores por disciplina e suas competências

Disciplina	Algoritmo - p/ Competência	Acurácia (%)	Algoritmo - p/ Disciplina	Acurácia (%)	Acurácia Final (%)
Português	MNB	77,03675	KNN	95,1540	73,30355
			MNB	94,7354	72,98107
Matemática	KNN	92,47967	KNN	95,1540	87,99811
			MNB	94,7354	87,61099
Biologia	MNB	91,47157	KNN	95,1540	87,03886
			MNB	94,7354	86,65596
Física	KNN	88,64268	KNN	95,1540	84,34706
			MNB	94,7354	83,97600
Química	KNN	89,64286	KNN	95,1540	85,29877
			MNB	94,7354	84,92352
História	KNN	84,32323	KNN	95,1540	80,23693
			MNB	94,7354	79,88395
Geografia	MNB	83,10166	KNN	95,1540	79,07455
			MNB	94,7354	78,72669

4.3.1 Resultados para o algoritmo $IMBHN^{COS}$

Nesta seção são os apresentados resultados de classificação do algoritmo $IMBHN^{COS}$ em comparação com as duas versões do algoritmo $IMBHN$. O objetivo foi avaliar em quais tipos de questão houve evolução no comportamento do algoritmo. Juntamente com esta análise, também foi realizada a comparação com os resultados finais obtidos na geração da *baseline*.

A [Tabela 18](#) apresenta o resultado geral de classificação para as versões do algoritmo $IMBHN$ usando toda a coleção de questões e tendo as disciplinas como critério de classificação. Os indicadores de acurácia e macro-F1 melhoraram com relação à sua versão-base $IMBHN^R$, mas ainda ficaram atrás da versão $IMBHN^C$. De forma geral, o algoritmo $IMBHN^{COS}$ manteve uma acurácia acima de 90% com a macro-F1 próxima de 1.

Tabela 18 – Classificação geral por disciplina das versões do algoritmo $IMBHN$

Algoritmo	Acurácia (%)	Erro (%)	Desvio Padrão (%)	Macro-F1
$IMBHN^C$	94,3681	5,6319	0,9395	0,9395
$IMBHN^R$	91,9076	8,0924	0,9117	0,9140
$IMBHN^{COS}$	92,9936	7,0063	0,3050	0,9275

Relativamente à classificação das questões no nível de competências dentro de cada disciplina, a [Tabela 19](#) apresenta o comparativo entre as versões do $IMBHN$. Além disso, também inclui os melhores resultados gerados na *baseline*. O principal resultado observado foi a superação do algoritmo $IMBHN^C$ em relação às outras versões, em **todos** os casos. Além disso, o desempenho também melhorou em comparação aos outros classificadores: o $IMBHN^{COS}$ obteve o melhor resultados para a classificação no nível de competências dentro da disciplina de Português e foi o segundo colocado nas outras 6 disciplinas. Com isso, reunindo os resultados obtidos com a *baseline*, o resultado final apresentado na [Tabela 17](#) sofre uma alteração para a disciplina Português: a acurácia final muda de 73,30% para 74,05%.

Tabela 19 – Classificação geral por competência das versões do algoritmo *IMBHN* incluindo melhores resultado da *baseline*

Disciplina /Algoritmo	Acurácia (%)	Erro (%)	Desvio Padrão (%)	Macro-F1
Português				
MNB	77,03675	22,96325	2,38459	0,68101
IMBHN ^C	76,57966	23,42034	3,25829	0,65912
IMBHN ^R	73,90587	26,09413	2,39737	0,63619
IMBHN ^{COS}	77,81767	22,18232	2,65572	0,778176
Matemática				
KNN	92,47967	7,52032	2,28205	0,92629
IMBHN ^C	87,38095	12,61905	4,32518	0,86643
IMBHN ^R	88,60627	11,39373	5,52403	0,88468
IMBHN ^{COS}	91,74796	8,25203	4,89960	0,90930
Biologia				
MNB	91,47157	8,52842	4,61951	0,91729
IMBHN ^C	84,70139	15,29861	4,68700	0,85091
IMBHN ^R	85,9042	14,0958	4,03323	0,86120
IMBHN ^{COS}	88,74581	11,25418	3,94224	0,88890
Física				
KNN	88,64268	11,35732	3,121491	0,887376
IMBHN ^C	86,10318	13,89682	4,03765	0,86252
IMBHN ^R	85,80525	14,19475	4,22660	0,85830
IMBHN ^{COS}	87,56520	12,43479	3,98732	0,87600
Química				
KNN	89,64286	10,35714	5,59412	0,89689
IMBHN ^C	84,2602	15,7398	5,00126	0,84441
IMBHN ^R	80,34864	19,65136	5,14925	0,80537
IMBHN ^{COS}	84,68537	15,31462	5,63194	0,84681
História				
KNN	84,32323	15,67677	5,40508	0,84308
IMBHN ^C	74,70202	25,29798	7,40826	0,75445
IMBHN ^R	81,43939	18,56061	5,95870	0,81813
IMBHN ^{COS}	83,44444	16,55555	4,58694	0,83717
Geografia				
MNB	83,10166	16,89833	2,94877	0,83154
IMBHN ^C	77,23156	22,76843	3,32898	0,77335
IMBHN ^R	77,59018	22,40981	3,46808	0,77780
IMBHN ^{COS}	81,14461	18,85538	2,64125	0,81284

4.4 Considerações Finais

Este capítulo apresentou a configuração experimental adotada com relação à base de questões e algoritmos/parametrizações escolhidos. Durante a geração da *baseline*, os algoritmos *KNN* e *MNB* demonstraram obter os melhores resultados tanto no nível de disciplinas quanto no nível de competências.

Considerando especificamente os algoritmos baseados em redes, a aplicação em coleção de questões demonstrou um resultado inferior ao obtido com coleções de documentos textuais

com maior quantidade de informação (ROSSI, 2016). Isso indica que a pouca quantidade de texto presente nas questões pode acabar não gerando uma rede heterogênea bipartida com termos e relações suficientes para um desempenho superior. Isso corroborou para o desenvolvimento do $IMBHN^{COS}$, uma versão do algoritmo $IMBHN$ com otimização na análise de informação de classes para novos documentos. OS resultados demonstraram desempenho superior em relação às outras duas versões do $IMBHN$, além dos indicadores ficarem entre os dois melhores resultados para a combinação da classificação no nível de competências, independentemente da nível de disciplina escolhido. Uma outra conclusão é que a forma de avaliar a informação de classes gerada pelo algoritmo $IMBHN$ impacta diretamente no desempenho do classificador. Com isso, há o potencial de que novas formas de análise das classes possam tornar o algoritmo $IMBHN$ vencedor nos comparativos de classificadores de questões, da mesma forma que já ocorreu na classificação de coleções de texto.

Como resultado final, concluiu-se que o modelo espaço-vetorial, juntamente com o algoritmo KNN e a técnica $IQF * QF * ICF$ demonstraram ser a melhor combinação de componentes para utilização na classificação de questões. A menor acurácia final ficou com a combinação final para a disciplina de Português, com 74,05%; as demais combinações finais obtiveram desempenho entre 79 a 87%, estando próxima dos melhores resultados obtidos nos estudos referenciados como estado da arte, casos dos trabalhos de Rossi (2016), Rossi *et al.* (2014) e Huang e Zhou (2015), comprovando as hipóteses (principal e de suporte) apresentadas no Capítulo 1.

CONCLUSÕES

Este capítulo apresenta as contribuições e resultados obtidos durante o desenvolvimento desta pesquisa com base nas hipóteses levantadas e objetivos traçados, juntamente com os artigos publicados e em preparação. Para finalizar, são descritas as limitações encontradas no processo e quais caminhos projetados para a continuação da pesquisa.

5.1 Resultados e Contribuições

Considerando a pequena quantidade de texto nas questões quando comparada às coleções de texto existentes na *web*, há a dificuldade de extrair uma quantidade de termos relevantes para a construção de um classificador eficiente. Desta forma, foi necessário analisar diferentes combinações de componentes de um classificador e também resultados de nos diferentes níveis de classificação (agregado e detalhado). Os componentes analisados foram as representações de texto, cálculo de pesos dos termos e algoritmos classificadores.

Os resultados de acurácia demonstraram que as hipóteses levantadas no [Capítulo 1](#) foram satisfeitas: (i) o conjunto de classificadores obteve acurácia geral média em torno dos 80% e se baseou (ii) na representação do texto usando o modelo espaço-vetorial tanto para a classificação mais geral (disciplinas) quanto para a mais detalhada (competências), sem tratamento específico (regras) e adaptável a qualquer idioma; e (iii) o método $IQF * QF * ICF$ se mostrou superior e com impacto direto no desempenho final dos classificadores. Considerando especificamente os algoritmos baseados em redes, a aplicação do algoritmo *IMBHN* em coleção de questões demonstrou um resultado inferior ao obtido com coleções de documentos textuais com maior quantidade de informação.

5.1.1 Resultados

Com base nos objetivos propostos para demonstrar a veracidade das hipóteses apresentadas, foram obtidos os seguintes resultados:

- **Modelo geral para construção de classificadores automáticos de questões voltadas a avaliações em larga escala:** com seu conjunto sequenciado de processos, define uma forma de classificar questões elaboradas para avaliações em larga escala. A ideia foi aproveitar a organização hierárquica de competências associadas a esse tipo de avaliação para e assim viabilizar o reconhecimento de padrões na base de questões associada ao nível de competência. O modelo geral foi usado como base para a elaboração das propostas de pesquisa e obtenção dos demais resultados.
- **Identificação do estado da arte na classificação de questões para contexto educacional/avaliações em larga escala:** o modelo espaço-vetorial, juntamente com o algoritmo *KNN* e a técnica $IQF * QF * ICF$ demonstraram ser a melhor combinação de componentes para ser utilizada na classificação de questões. A acurácia final variou de 74% (79, sem Português) a 87% comprovando as hipóteses principal e de suporte desta tese apresentadas no [Capítulo 1](#).
- **Algoritmo $IMBHN^{COS}$:** o algoritmo apresentou melhoria nos resultados que as versões $IMBHN^C$ e $IMBHN^R$ e também, em muitos casos, aos algoritmos *MNB* e *SVM*. Mesmo o algoritmo *KNN*, que obteve os melhores resultados, foi superado na classificação por competências dentro da disciplina Português. Dessa forma, essa otimização se enquadra no estado da arte nos resultados de classificação de questões para algoritmos de Aprendizado de Máquina baseado em redes.
- **Estado-da-arte no cálculo de pesos para classificação de questões:** técnica $IQF * QF * ICF$ demonstrou superioridade às técnicas *TF* e $TF * IDF$ na classificação de questões para todos os algoritmos, tanto para algoritmos baseados no modelo espaço-vetorial quanto para baseados em redes.
- **Banco de dados de questões representativas e constantemente atualizável:** gerado a partir da classificação de novas questões do ENEM, constituindo assim um repositório renovado periodicamente através da importação de dados e organizado por áreas do conhecimento, disciplinas e competências.
- **API (Application Program Interface):** interface para integrar as funcionalidades de geração e utilização do modelo gerado pelo classificador para outras plataformas. Possui funcionalidades para importação e pré-processamento de questões; cálculo de pesos para os termos extraídos, geração da matriz documento-termo (modelo espaço-vetorial) e da lista de adjacências (modelo em redes); processamento de algoritmos de classificação e

geração de experimentos (*k-fold cross validation*, matriz de contingências, controle de repetições e exportação de resultados).

- **Análise comparativa:** desempenho dos algoritmos tabulados e avaliados através dos indicadores acurácia, precisão (micro e macro), *recall* (micro e macro) e F1 ((micro e macro), divida em uma visão geral (por disciplinas) e uma outra mais detalhada (competências).

5.1.2 Contribuições

A partir dos resultados obtidos neste trabalho, pode-se destacar as seguintes contribuições:

- apoio às plataformas de ensino adaptativo para integrar funcionalidades de importação e classificação de questões, além da utilização da base de questões classificadas, o que viabiliza a recomendação de atividades a partir da avaliação de competências.
- ferramentas de classificação para as editoras de material didático, facilitando a produção de conteúdos organizados de acordo com as competências definidas na matriz de referência do ENEM.
- apoio a professores e equipes pedagógicas na elaboração de avaliações diagnósticas contínuas sobre as facilidades e dificuldades de seus alunos dentro de cada competência do ENEM, habilitando assim o acompanhamento da trilha do aprendizado e a realização de processos de recuperação paralela mais customizados às necessidades individuais do aluno.

Para a elaboração deste trabalho, foi fundamental o desenvolvimento de uma revisão sistemática sobre classificadores automáticos de questões considerando o contexto educacional. Esta revisão foi atualizada e incluída nesta tese de doutorado através do [Capítulo 2](#). A qualidade da pesquisa realizada foi reconhecida e disponibilizada à comunidade científica através da seguinte publicação:

SILVA, V. A.; BITTENCOURT, I. I.; MALDONADO, J. C. *Automatic question classifiers: a systematic review*. IEEE Transactions on Learning Technologies, IEEE, 2018. (**Qualis A2**)

5.2 Limitações e Trabalhos Futuros

Dentro do universo da classificação de textos, esta tese abrangeu um contexto muito específico: questões para avaliações em larga escala, no caso, questões do ENEM. Este contexto se relaciona diretamente ao ambiente educacional, pode ser usado como referência, mas não pode ser generalizado para outras aplicações como *Question Answering* e *Information Retrieval*, por exemplo.

A avaliação comparativa se baseou nos principais classificadores de texto encontrados na literatura. Classificadores que consomem maior tempo de processamento como árvores de decisão e redes neurais foram descartados, mas podem oferecer resultados comparáveis na classificação de textos e questões.

O algoritmo $IMBHN^{COS}$ utilizou a similaridade do cosseno para a classificação final de questões, a partir das informações de classe induzidas no modelo de classificação. A escolha dessa medida foi pela simplicidade de aplicação, rapidez no tempo de processamento e indicadores de desempenho obtidos nos resultados de classificação. Mesmo assim, considerando a abordagem vetorial dada ao vetor de peso gerado pelo modelo em redes, outras medidas de similaridade podem ser utilizadas, como a distância euclidiana e distância Manhattan.

Pretende-se, como próximo passo para este trabalho de pesquisa, avaliar o desempenho dos classificadores de texto para outras avaliações em larga escala com competências definidas em uma matriz de referência, entre elas a Prova Brasil e o SAEB.

Ainda considerando questões do ENEM, será utilizado um *Learning Management System* (LMS) (por exemplo, Moodle), para criar um módulo de importação e classificação de questões e outro módulo de geração de testes e recomendação de conteúdos a partir do diagnóstico obtido com as competências avaliadas.

Por fim, será dada continuidade a um trabalho de pesquisa referente à transformação das questões em imagens usando características dos textos das questões (por exemplo, classes de palavras e termos associados às competências). A partir da coleção de imagens, as disciplinas e competências do ENEM serão utilizadas como critério de classificação para processamento através de uma rede neural convolucional.

REFERÊNCIAS

- ABDULJABBAR, D.; OMAR, N. Exam questions classification based on bloom's taxonomy cognitive level using classifiers combination. **Journal of Theoretical and Applied Information Technology**, v. 78, n. 3, p. 447–455, 2015. Cited By 3. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84940667147&partnerID=40&md5=0b7f8c78cb54652c54bc18e8b9ae2e01>>. Citado nas páginas 33, 51, 55 e 59.
- ABDULJABBAR, D. A.; OMAR, N. Exam questions classification based on bloom's taxonomy cognitive level using classifiers combination. **Journal of Theoretical and Applied Information Technology**, v. 78, n. 3, 2015. Citado nas páginas 80, 81 e 82.
- ACHIMUGU, P.; SELAMAT, A.; IBRAHIM, R.; MAHRIN, M. N. A systematic literature review of software requirements prioritization research. **Information and software technology**, Elsevier, v. 56, n. 6, p. 568–585, 2014. Citado na página 58.
- AGGARWAL, C. C. **Data classification: algorithms and applications**. [S.l.]: CRC press, 2014. Citado na página 40.
- AGGARWAL, C. C.; ZHAI, C. **Mining text data**. [S.l.]: Springer Science & Business Media, 2012. Citado nas páginas 29, 34, 40, 103 e 104.
- ALAHMADI, A. Automatic text classification using bag of words and bag of concepts based representations. University of Limerick, 2016. Citado nas páginas 38, 44, 47, 51 e 59.
- ANDRADE, D. F. de; TAVARES, H. R.; VALLE, R. da C. Teoria da resposta ao item: conceitos e aplicações. **ABE, Sao Paulo**, 2000. Citado na página 33.
- BANDYOPADHYAY, S. B. S. Question classification and answering from procedural text in english. In: **24th International Conference on Computational Linguistics**. [S.l.: s.n.], 2012. p. 11. Citado na página 59.
- BENBRAHIM, H.; BRAMER, M. Text and hypertext categorization. In: **Artificial Intelligence An International Perspective**. [S.l.]: Springer, 2009. p. 11–38. Citado nas páginas 31, 103 e 104.
- BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. Systematic review in software engineering. **System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES**, v. 679, n. 05, p. 45, 2005. Citado nas páginas 52 e 53.
- BLASIS, E.; TILGER, M.; LONGATO, S.; GUEDES, P. **Avaliação educacional: os desafios da sala de aula e a promoção da aprendizagem**. [S.l.]: Cenpec: Fundação Itaú Social, 2014. Citado nas páginas 30, 85 e 86.
- BLOOM, B. S.; KRATHWOHL, D. R.; MASIA, B. B. Taxonomia dos objetivos educacionais, vols 1 e 2. **Editora Globo**, 1974. Citado na página 80.
- _____. Bloom taxonomy of educational objectives. In: **Allyn and Bacon**. [S.l.]: Pearson Education, 1984. Citado na página 69.

- BLUNSOM, P.; GREFFENSTETTE, E.; KALCHBRENNER, N. A convolutional neural network for modelling sentences. In: PROCEEDINGS OF THE 52ND ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics**. [S.l.], 2014. Citado na página 60.
- BORTOLOTTI, S. L. V.; TEZZA, R.; ANDRADE, D. F. de; BORNIA, A. C.; JÚNIOR, A. F. de S. Relevance and advantages of using the item response theory. **Quality & Quantity**, Springer, v. 47, n. 4, p. 2341–2360, 2013. Citado na página 33.
- BRAUM, H. M.; RIGO, S. J.; BARBOSA, J. L. Modelo de classificação automática de questões na língua portuguesa. **RENOTE**, v. 12, n. 2. Citado na página 59.
- BREVE, F.; ZHAO, L.; QUILES, M.; PEDRYCZ, W.; LIU, J. Particle competition and cooperation in networks for semi-supervised learning. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 24, n. 9, p. 1686–1698, 2011. Citado na página 32.
- CETINTAS, S.; SI, L.; XIN, Y. P.; ZHANG, D.; PARK, J. Y.; TZUR, R. A joint probabilistic classification model of relevant and irrelevant sentences in mathematical word problems. **arXiv preprint arXiv:1411.5732**, 2014. Citado na página 84.
- CHAN, W.; YANG, W.; TANG, J.; DU, J.; ZHOU, X.; WANG, W. Community question topic categorization via hierarchical kernelized classification. In: ACM. **Proceedings of the 22nd ACM international conference on Conference on information & knowledge management**. [S.l.], 2013. p. 959–968. Citado na página 59.
- CHANG, W.-C.; CHUNG, M.-S. Automatic applying bloom’s taxonomy to classify and analysis the cognition level of english question items. In: **2009 Joint Conferences on Pervasive Computing (JCPC)**. [S.l.: s.n.], 2009. p. 727–734. Citado na página 82.
- CHEN, C.; HAN, H.; LIU, Z. Knn question classification method based on apriori algorithm. 2014. Citado na página 59.
- CHEN, L.; ZHANG, D.; LEVENE, M. Identifying local questions in community question answering. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 7675 LNCS, p. 306–315, 2012. Cited By 0. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84871595876&doi=10.1007%2f978-3-642-35341-3_26&partnerID=40&md5=55ed944049f6bb72d57fed8dda1c55da>. Citado na página 59.
- CHERNOV, A.; PETUKHOVA, V.; KLAKEW, D. Linguistically motivated question classification. In: **Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)**. [S.l.: s.n.], 2015. p. 51–59. Citado na página 59.
- CROCE, D.; FILICE, S.; BASILI, R. Making sense of kernel spaces in neural learning. **Computer Speech & Language**, Elsevier, v. 58, p. 51–75, 2019. Citado na página 62.
- DALAVI, M.; CHEKE, S. Hadoop mapreduce implementation of a novel scheme for term weighting in text categorization. In: IEEE. **Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on**. [S.l.], 2014. p. 994–999. Citado na página 59.

DERNONCOURT, F. **Sequential short-text classification with neural networks**. Tese (Doutorado) — Massachusetts Institute of Technology, 2017. Citado na página 62.

DIAB, S.; SARTAWI, B. Classification of questions and learning outcome statements (los) into blooms taxonomy (bt) by similarity measurements towards extracting of learning outcome from learning material. **arXiv preprint arXiv:1706.03191**, 2017. Citado nas páginas 33, 59 e 81.

DING, W.; LIANG, P.; TANG, A.; VLIET, H. V. Knowledge-based approaches in software documentation: A systematic literature review. **Information and Software Technology**, Elsevier, v. 56, n. 6, p. 545–567, 2014. Citado na página 58.

DUBEY, M.; GOYAL, V. **Classifying stack overflow questions based on bloom’s taxonomy**. Tese (Doutorado), 2016. Citado nas páginas 33 e 59.

DUMAIS, S.; PLATT, J.; HECKERMAN, D.; SAHAMI, M. Inductive learning algorithms and representations for text categorization. 1998. Citado nas páginas 31, 103 e 104.

DYBA, T.; DINGSOYR, T. Empirical studies of agile software development: A systematic review. **Information and software technology**, Elsevier, v. 50, n. 9-10, p. 833–859, 2008. Citado na página 58.

EMMANUEL, M.; KHATRI, S. M.; BABU, D. R. A novel scheme for term weighting in text categorization: Positive impact factor. In: IEEE. **Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on**. [S.l.], 2013. p. 2292–2297. Citado nas páginas 46, 47 e 59.

FACHRURROZI, M.; SILABAN, L. I.; YUSLIANI, N. Questions classification software based on bloom’s cognitive levels using naive bayes classifier method. **PROCEEDING IC-ITECHS 2014**, v. 1, n. 01, 2015. Citado na página 33.

FEI, T.; HENG, W. J.; TOH, K. C.; QI, T. Question classification for e-learning by artificial neural network. In: IEEE. **Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on**. [S.l.], 2003. v. 3, p. 1757–1761. Citado nas páginas 33 e 84.

FENTAW, H. W.; KIM, T.-H. Design and investigation of capsule networks for sentence classification. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 9, n. 11, p. 2200, 2019. Citado na página 62.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação**. [S.l.]: Ciência moderna, 2012. Citado na página 41.

FERREIRA, J. H. C. Enetec—um modelo de referência para avaliação em larga escala da educação técnica profissional na modalidade ead. **SIED: EnPED-Simpósio Internacional de Educação a Distância e Encontro de Pesquisadores em Educação a Distância**, 2014. Citado na página 85.

FILICE, S.; CROCE, D.; BASILI, R. A stratified strategy for efficient kernel-based learning. In: **AAAI**. [S.l.: s.n.], 2015. p. 2239–2245. Citado na página 59.

FOLEY, J.; ALLAN, J. Retrieving hierarchical syllabus items for exam question analysis. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9626, p. 575–586, 2016. Cited By 0. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84962622134&partnerID=40&md5=fe7c7b5479810cfd8f2cf40118d3d802>>. Citado na página 59.

GHAREHCHOPOGH, F. S.; LOTFI, Y. Machine learning based question classification methods in the question answering systems. **Int J Innovat Appl Stud**, v. 4, n. 2, 2013. Citado nas páginas 48 e 50.

HAO, T.; XIE, W.; XU, F. A wordnet expansion-based approach for question targets identification and classification. In: **Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data**. [S.l.]: Springer, 2015. p. 333–344. Citado na página 59.

HARDY; CHEAH, Y.-N. Question classification using extreme learning machine on semantic features. **Journal of ICT Research and Applications**, v. 7, n. 1, p. 36–58, 2013. Cited By 5. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84888405312&doi=10.5614%2ftbj.ict.res.appl.2013.7.1.3&partnerID=40&md5=b8fd4334d74abac3f19b29fab95e8a0>>. Citado nas páginas 39 e 60.

HARIS, S. S.; OMAR, N. A rule-based approach in bloom's taxonomy question classification through natural language processing. In: IEEE. **Computing and Convergence Technology (ICCCT), 2012 7th International Conference on**. [S.l.], 2012. p. 410–414. Citado na página 55.

_____. A rule-based approach in bloom's taxonomy question classification through natural language processing. In: IEEE. **Computing and Convergence Technology (ICCCT), 2012 7th International Conference on**. [S.l.], 2012. p. 410–414. Citado nas páginas 82 e 83.

_____. Bloom's taxonomy question categorization using rules and n-gram approach. **Journal of Theoretical & Applied Information Technology**, v. 76, n. 3, 2015. Citado nas páginas 55 e 60.

HARRISON, J.; DIKKEN, O.; PEER, D. Question classification according to bloom's revised taxonomy. **Bachelor End Project Report**, 2017. Citado nas páginas 33 e 62.

HERNANDES, E.; ZAMBONI, A.; FABBRI, S.; THOMMAZO, A. D. Using gqm and tam to evaluate start-a tool that supports systematic review. **CLEI Electronic Journal**, Centro Latinoamericano de Estudios en Informática, v. 15, n. 1, p. 3–3, 2012. Citado na página 53.

HOQUE, M. M.; GONCALVES, T.; QUARESMA, P. Classifying questions in question answering system using finite state machines with a simple learning approach. **Sponsors: National Science Council, Executive Yuan, ROC Institute of Linguistics, Academia Sinica NCCU Office of Research and Development**, p. 409, 2013. Citado na página 60.

HUANG, M. M. L.; ZHOU, B. X. B. Dependency-based convolutional neural networks for sentence embedding. 2015. Citado nas páginas 60, 84 e 111.

HUANG, Z.; THINT, M.; QIN, Z. Question classification using head words and their hypernyms. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the Conference on Empirical Methods in Natural Language Processing**. [S.l.], 2008. p. 927–936. Citado nas páginas 38 e 60.

- HUO, L.-F.; ZHANG, L.-M.; ZHAO, X.-Q. Question recognition based on subject. **Lecture Notes in Electrical Engineering**, v. 375, p. 353–361, 2016. Cited By 0. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85007530389&doi=10.1007%2f978-981-10-0539-8_35&partnerID=40&md5=939cc33c9e958342534c21b74b381ff1>. Citado na página 60.
- HUTZLER, D.; DAVID, E.; AVIGAL, M.; AZOULAY, R. Learning methods for rating the difficulty of reading comprehension questions. In: IEEE. **Software Science, Technology and Engineering (SWSTE), 2014 IEEE International Conference on**. [S.l.], 2014. p. 54–62. Citado na página 60.
- JAYAKODI, K.; BANDARA, M.; MEEDENIYA, D. An automatic classifier for exam questions with wordnet and cosine similarity. In: IEEE. **Moratuwa Engineering Research Conference (MERCon), 2016**. [S.l.], 2016. p. 12–17. Citado nas páginas 55 e 60.
- JAYAKODI, K.; BANDARA, M.; PERERA, I. An automatic classifier for exam questions in engineering: A process for bloom’s taxonomy. In: IEEE. **Teaching, Assessment, and Learning for Engineering (TALE), 2015 IEEE International Conference on**. [S.l.], 2015. p. 195–202. Citado na página 60.
- JAYAKODI, K.; BANDARA, M.; PERERA, I.; MEEDENIYA, D. Wordnet and cosine similarity based classifier of exam questions using bloom’s taxonomy. **International Journal of Emerging Technologies in Learning**, v. 11, n. 4, 2016. Citado nas páginas 33 e 60.
- JAYALAKSHMI, S.; SHESHASAYEE, A. A review on question classification using machine learning based on semantic features. **International Journal of Engineering Sciences & Research Technology**, Citeseer, v. 4, p. 561–564, 2015. Citado nas páginas 39 e 46.
- KARYAWATI, A. E.; WINARKO, E.; AZHARI, A.; HARJOKO, A. Ontology-based why-question analysis using lexico-syntactic patterns. **International Journal of Electrical and Computer Engineering**, IAES Institute of Advanced Engineering and Science, v. 5, n. 2, p. 318, 2015. Citado na página 60.
- KIM, Y. Convolutional neural networks for sentence classification. **arXiv preprint arXiv:1408.5882**, 2014. Citado na página 60.
- KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004. Citado nas páginas 52, 53 e 55.
- KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. ebse technical report. keele university & department of computer science university of durham. 2007. Citado nas páginas 52 e 63.
- KOMNINOS, A.; MANANDHAR, S. Dependency based embeddings for sentence classification tasks. In: **Proceedings of NAACL-HLT**. [S.l.: s.n.], 2016. p. 1490–1500. Citado na página 60.
- LA, L.; GUO, Q.; YANG, D.; CAO, Q. Multiclass boosting with adaptive group-based knn and its application in text categorization. **Mathematical Problems in Engineering**, Hindawi Publishing Corporation, v. 2012, 2012. Citado na página 60.
- LE-HONG, P.; PHAN, X.-H.; NGUYEN, T.-D. Using dependency analysis to improve question classification. In: **Knowledge and Systems Engineering**. [S.l.]: Springer, 2015. p. 653–665. Citado na página 60.

- LE, P.; ZUIDEMA, W. The forest convolutional network: Compositional distributional semantics with a neural chart and without binarization. In: **EMNLP**. [S.l.: s.n.], 2015. p. 1155–1164. Citado na página 60.
- LEE, J. Y.; DERNONCOURT, F. Sequential short-text classification with recurrent and convolutional neural networks. **arXiv preprint arXiv:1603.03827**, 2016. Citado na página 60.
- LI, D.; ZHANG, J.; LI, P. Representation learning for question classification via topic sparse autoencoder and entity embedding. In: IEEE. **2018 IEEE International Conference on Big Data (Big Data)**. [S.l.], 2018. p. 126–133. Citado na página 62.
- LI, X.; ROTH, D. Learning question classifiers. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 19th international conference on Computational linguistics-Volume 1**. [S.l.], 2002. p. 1–7. Citado nas páginas 29, 33, 37, 38, 50, 60, 65, 69 e 93.
- LI, Y.; TRIPATHI, A.; SRINIVASAN, A. Challenges in short text classification: The case of online auction disclosure. 2016. Citado na página 60.
- LIN, Z.; WANG, H.; MCCLEAN, S. Tree similarity measurement for classifying questions by syntactic structures. In: SPRINGER. **International Conference on Intelligent Computing**. [S.l.], 2016. p. 379–390. Citado na página 60.
- LIU, B. **Web data mining: exploring hyperlinks, contents, and usage data**. [S.l.]: Springer Science & Business Media, 2007. Citado na página 41.
- LIU, G.; GUO, J. Bidirectional lstm with attention mechanism and convolutional layer for text classification. **Neurocomputing**, Elsevier, v. 337, p. 325–338, 2019. Citado na página 62.
- LIU, N.; ZHANG, B.; YAN, J.; CHEN, Z.; LIU, W.; BAI, F.; CHIEN, L. Text representation: From vector to tensor. In: IEEE. **Fifth IEEE International Conference on Data Mining (ICDM'05)**. [S.l.], 2005. p. 4–pp. Citado nas páginas 32 e 40.
- LODHI, H.; SAUNDERS, C.; SHAW-TAYLOR, J.; CRISTIANINI, N.; WATKINS, C. Text classification using string kernels. **Journal of Machine Learning Research**, v. 2, n. Feb, p. 419–444, 2002. Citado nas páginas 32 e 40.
- LU, Z.; LIN, Y.-R.; ZHANG, Q.; CHEN, M. Classifying questions into fine-grained categories using topic enriching. In: IEEE. **Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on**. [S.l.], 2016. p. 166–174. Citado na página 60.
- LUO, Y.; BOUCHER, T.; ORAL, T.; OSOFSKY, D.; WEBER, S. A Study on Expert Sourcing Enterprise Question Collection and Classification. In: Calzolari, N and Choukri, K and Declerck, T and Loftsson, H and Maegaard, B and Mariani, J and Moreno, A and Odijk, J and Piperidis, S (Ed.). **LREC 2014 - NINTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION**. [S.l.], 2014. ISBN 978-2-9517408-8-4. 9th International Conference on Language Resources and Evaluation (LREC), Reykjavik, ICELAND, MAY 26-31, 2014. Citado na página 60.
- MA, M.; HUANG, L.; XIANG, B.; ZHOU, B. Dependency-based convolutional neural networks for sentence embedding. **arXiv preprint arXiv:1507.01839**, 2015. Citado na página 85.
- MADABUSHI, H. T.; LEE, M. High accuracy rule-based question classification using question syntax and semantics. Citado na página 60.

MAHATME, V. P.; BHOYAR, K. Questions categorization in e-learning environment using data mining technique. **World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering**, v. 10, n. 1, p. 93–97. Citado nas páginas 30 e 61.

MAHDAVI-HEZAVEHI, S.; GALSTER, M.; AVGERIOU, P. Variability in quality attributes of service-based software systems: A systematic literature review. **Information and Software Technology**, Elsevier, v. 55, n. 2, p. 320–343, 2013. Citado na página 58.

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. Introduction to information retrieval. **Natural Language Engineering**, Cambridge university press, v. 16, n. 1, p. 100–103, 2010. Citado na página 31.

MARINCIC, D.; KOMPARA, T.; GAMS, M. Question Classification with Active Learning. In: Sojka, P and Horak, A and Kopecek, I and Pala, K (Ed.). **TEXT, SPEECH AND DIALOGUE, TSD 2012**. [S.l.], 2012. (Lecture Notes in Computer Science, 7499), p. 673–680. ISBN 978-3-642-32789-6; 978-3-642-32790-2. ISSN 0302-9743. 15th International Conference on Text, Speech and Dialogue (TSD), Brno, CZECH REPUBLIC, SEP 03-07, 2012. Citado na página 61.

MIRANDA, L. A. S. d.; FERREIRA, A. C. F.; DIAS, G. R. M. Análise de conteúdo das questões de fisiologia humana da prova de ciências da natureza e suas tecnologias do exame nacional do ensino médio (1998-2016). **Ciência & Educação (Bauru)**, SciELO Brasil, v. 25, n. 2, p. 375–393, 2019. Citado na página 88.

MISHRA, S. K.; KUMAR, P.; SAHA, S. K. A support vector machine based system for technical question classification. In: SPRINGER. **International Conference on Mining Intelligence and Knowledge Exploration**. [S.l.], 2015. p. 640–649. Citado na página 61.

MOU, L.; PENG, H.; LI, G.; XU, Y.; ZHANG, L.; JIN, Z. Tree-based convolution: A new neural architecture for sentence modeling. In: **Proceedings of Conference on Empirical Methods in Natural Language Processing (to appear)**. [S.l.: s.n.], 2015. Citado na página 61.

NAKAGAWA, E. Y.; SCANNAVINO, K. R. F.; FABBRI, S. C. P. F.; FERRARI, F. C. **Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática**. [S.l.]: Elsevier Brasil, 2017. Citado nas páginas 52 e 54.

NEWMAN, M. Networks: an introduction: Oxford university press, inc. **Nielsen, M., Elmes, G. and Kipyatkov**, v. 1999, 2010. Citado nas páginas 42 e 44.

NUNTIYAGUL, A.; NARUEDOMKUL, K.; CERCONE, N.; WONGSAWANG, D. Adaptable learning assistant for item bank management. **Computers & Education**, Elsevier, v. 50, n. 1, p. 357–370, 2008. Citado na página 85.

OBASA, A. I.; SALIM, N.; KHAN, A. Hybridization of bag-of-words and forum metadata for web forum question post detection. **Indian Journal of Science and Technology**, v. 8, n. 32, 2016. Citado na página 61.

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**, Elsevier, v. 59, p. 297–303, 2012. Citado na página 55.

- _____. Automated analysis of exam questions according to bloom's taxonomy. **Procedia-Social and Behavioral Sciences**, Elsevier, v. 59, p. 297–303, 2012. Citado na página 82.
- OSADI, K.; FERNANDO, S.; WELGAMA, W. *et al.* Ensemble classifier based approach for classification of examination questions into bloom's taxonomy cognitive levels. **International Journal of Computer Applications**, Foundation of Computer Science, v. 975, p. 8887, 2017. Citado nas páginas 33, 61 e 81.
- OSMAN, A.; YAHYA, A. A. Classifications of exam questions using linguistically-motivated features: A case study based on bloom's taxonomy. 2016. Citado na página 61.
- _____. Classifications of exam questions using natural language syntatic features: A case study based on bloom's taxonomy. 2016. Citado na página 61.
- PATRA, A.; SINGH, D. A survey report on text classification with different term weighing methods and comparison between classification algorithms. **International Journal of Computer Applications**, Foundation of Computer Science, v. 75, n. 7, 2013. Citado nas páginas 39, 48 e 49.
- PATRICK, J.; LI, M. An ontology for clinical questions about the contents of patient notes. **J. of Biomedical Informatics**, Elsevier Science, San Diego, USA, v. 45, n. 2, p. 292–306, abr. 2012. ISSN 1532-0464. Disponível em: <<http://dx.doi.org/10.1016/j.jbi.2011.11.008>>. Citado na página 61.
- PESTANA, M. I. O sistema de avaliação brasileiro. **Revista Brasileira de Estudos Pedagógicos**, v. 79, n. 191, p. 65–73, 1998. Citado na página 85.
- PILLAI, P. G.; NARAYANAN, J. Question categorization using svm based on different term weighting methods. **International Journal on Computer Science and Engineering**, Engg Journals Publications, v. 4, n. 5, p. 938, 2012. Citado na página 61.
- PIMENTA, C. O.; ISHII, I.; SANTOS, M. P. dos. Usos de resultados de avaliações externas por coordenadores pedagógicos e professores: implicações para o currículo? **VII Reunião da ABAVE–Avaliação e currículo: um diálogo necessário**, n. 1, p. 443–456, 2014. Citado na página 33.
- PING, Y.; ZHOU, Y. jian; XUE, C.; YANG, Y. xian. Efficient representation of text with multiple perspectives. **The Journal of China Universities of Posts and Telecommunications**, v. 19, n. 1, p. 101 – 111, 2012. ISSN 1005-8885. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1005888511602343>>. Citado na página 61.
- PORTER, M. F. An algorithm for suffix stripping. **Program**, Emerald Group Publishing Limited, 2006. Citado na página 103.
- POTA, M.; ESPOSITO, M.; PIETRO, G. D. A Forward-Selection Algorithm for SVM-Based Question Classification in Cognitive Systems. In: DePietro, G and Gallo, L and Howlett, RJ and Jain, LC (Ed.). **INTELLIGENT INTERACTIVE MULTIMEDIA SYSTEMS AND SERVICES 2016**. [S.l.], 2016. (Smart Innovation Systems and Technologies, 55), p. 587–598. ISBN 978-3-319-39345-2; 978-3-319-39344-5. ISSN 2190-3018. 9th KES International Conference on Intelligent Interactive Multimedia Systems and Services (IIMSS), Puerto de la Cruz, SPAIN, JUN 15-17, 2016. Citado na página 61.

POTA, M.; FUGGI, A.; ESPOSITO, M.; PIETRO, G. D. Extracting Compact Sets of Features for Question Classification in Cognitive Systems: A Comparative Study. In: Xhafa, F and Barolli, L and Messina, F and Ogiela, MR (Ed.). **2015 10TH INTERNATIONAL CONFERENCE ON P2P, PARALLEL, GRID, CLOUD AND INTERNET COMPUTING (3PGCIC)**. [S.l.: s.n.], 2015. p. 551–556. ISBN 978-1-4673-9473-4. 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing 3PGCIC, Krakow, POLAND, NOV 04-06, 2015. Citado nas páginas 44, 46 e 61.

POYRAZ, M.; KILIMCI, Z. H.; GANIZ, M. C. Higher-order smoothing: a novel semantic smoothing method for text classification. **Journal of Computer Science and Technology**, Springer, v. 29, n. 3, p. 376–391, 2014. Citado na página 61.

QI, X.; SU, L.; YANG, B.; CHEN, J.; LI, Y.; LIU, J. Question classification based on hadoop platform. In: SPRINGER. **International Conference on Cloud Computing**. [S.l.], 2014. p. 107–115. Citado na página 61.

QIU, Y. **Questionnaire integration system based on question classification and short text semantic textual similarity**, A. Tese (Doutorado) — Colorado State University. Libraries, 2018. Citado na página 62.

QU, B.; CONG, G.; LI, C.; SUN, A.; CHEN, H. An evaluation of classification models for question topic categorization. **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 63, n. 5, p. 889–903, 2012. Citado na página 61.

QUAN, X.; WENYIN, L.; QIU, B. Term weighting schemes for question categorization. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 33, n. 5, p. 1009–1021, 2010. Citado nas páginas 93, 94 e 95.

RAHMAN, A. K. M. M. M.; ROY, C. K. Embedded emotion-based classification of stack overflow questions towards the question quality prediction. Citado na página 61.

RAHMAN, T. A. Question classification using statistical approach: A complete review. **Journal of Theoretical and Applied Information Technology**, v. 71, n. 3, 2015. Citado nas páginas 34, 80 e 84.

ROBERTS, K.; KILICOGU, H.; FISZMAN, M.; DEMNER-FUSHMAN, D. Automatically classifying question types for consumer health questions. **AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium**, v. 2014, p. 1018–1027, 2014. Cited By 2. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84964315969&partnerID=40&md5=6e0fef25b320de90395e0cedf0aa18fa>>. Citado na página 61.

_____. Decomposing consumer health questions. In: **BioNLP Workshop**. [S.l.: s.n.], 2014. p. 29–37. Citado na página 61.

ROMERO, C.; VENTURA, S. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 3, n. 1, p. 12–27, 2013. Citado na página 30.

ROSSI, R. G. **Classificação automática de textos por meio de aprendizado de máquina baseado em redes**. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado nas páginas 43, 44, 89, 94, 96, 97, 98, 104, 105 e 111.

- ROSSI, R. G.; LOPES, A. de A.; FALEIROS, T. de P.; REZENDE, S. O. Inductive model generation for text classification using a bipartite heterogeneous network. **Journal of Computer Science and Technology**, Springer, v. 29, n. 3, p. 361–375, 2014. Citado nas páginas 40, 94, 96, 104 e 111.
- SALLEH, N.; MENDES, E.; GRUNDY, J. Empirical studies of pair programming for cs/se teaching in higher education: A systematic literature review. **IEEE Transactions on Software Engineering**, IEEE, v. 37, n. 4, p. 509–525, 2011. Citado na página 58.
- SALTON, G. Automatic text processing: The transformation, analysis, and retrieval of. **Reading: Addison-Wesley**, v. 169, 1989. Citado nas páginas 32, 40 e 99.
- SANGODIAH, A.; AHMAD, R.; AHMAD, W.; FATIMAH, W. Taxonomy based features in question classification using support vector machine. **Journal of Theoretical & Applied Information Technology**, v. 95, n. 12, 2017. Citado nas páginas 61 e 81.
- SANGODIAH, A.; MUNIANDY, M.; HENG, L. Question classification using statistical approach: A complete review. **Journal of Theoretical and Applied Information Technology**, v. 71, n. 3, p. 386–395, 2015. Cited By 0. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84921988552&partnerID=40&md5=cecb56daef19411106e45c53a006050c>>. Citado nas páginas 29, 30, 38 e 55.
- SARROUTI, M.; LACHKAR, A.; OUATIK, S. E. A. Biomedical question types classification using syntactic and rule based approach. In: IEEE. **Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 2015 7th International Joint Conference on**. [S.l.], 2015. v. 1, p. 265–272. Citado na página 61.
- SEBASTIANI, F. Machine learning in automated text categorization. **ACM computing surveys (CSUR)**, ACM, v. 34, n. 1, p. 1–47, 2002. Citado na página 31.
- SHAH, F. P.; PATEL, V. A review on feature selection and feature extraction for text classification. In: IEEE. **Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on**. [S.l.], 2016. p. 2264–2268. Citado nas páginas 47, 48 e 49.
- SHANTHI, P.; KRISHNAMURTHI, I. A semantic approach for question classification using register linear based model. **Middle-East Journal of Scientific Research**, v. 23, n. 4, p. 685–694, 2015. Citado nas páginas 61 e 69.
- SHIN, J.; KIM, Y.; YOON, S.; JUNG, K. Contextual-cnn: A novel architecture capturing unified meaning for sentence classification. In: IEEE. **2018 IEEE International Conference on Big Data and Smart Computing (BigComp)**. [S.l.], 2018. p. 491–494. Citado na página 62.
- SHUANG, K.; ZHANG, Z.; LOO, J.; SU, S. Convolution–deconvolution word embedding: An end-to-end multi-prototype fusion embedding method for natural language processing. **Information Fusion**, Elsevier, v. 53, p. 112–122, 2020. Citado na página 62.
- SILVA, J.; COHEUR, L.; MENDES, A. C.; WICHERT, A. From symbolic to sub-symbolic information in question classification. **Artificial Intelligence Review**, Springer, v. 35, n. 2, p. 137–154, 2011. Citado na página 61.
- SILVA, T. C.; ZHAO, L. Stochastic competitive learning in complex networks. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 23, n. 3, p. 385–398, 2012. Citado na página 40.

SILVA, V. A.; BITTENCOURT, I. I.; MALDONADO, J. C. Automatic question classifiers: a systematic review. **IEEE Transactions on Learning Technologies**, IEEE, 2018. Citado na página 37.

SILVA, V. G.; GIMENES, N. Uso da avaliação externa por equipes gestoras e profissionais docentes: um estudo em quatro redes de ensino público. **São Paulo: FCC/SEP**, 2013. Citado na página 33.

SINGH, V.; DWIVEDI, S. K. An Integrated Pattern Matching and Machine Learning Approach for Question Classification. In: DEFENCE R&D ORG, MINIST DEFENCE; DEPT SCI & TECH, GOVT INDIA; SCI & TECH COUNCIL, UTTARAKHAND; ONGC; IEEE UP SEC, INDIA. **2015 1st International Conference on Next Generation Computing Technologies (NGCT)**. [S.l.], 2015. p. 762–767. ISBN 978-1-4673-6809-4. 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, INDIA, SEP 04-05, 2015. Citado na página 61.

SONWALKAR, N. Adaptive individualization: The next generation of online education. In: ASSOCIATION FOR THE ADVANCEMENT OF COMPUTING IN EDUCATION (AACE). **EdMedia+ Innovate Learning**. [S.l.], 2007. p. 3056–3063. Citado na página 30.

SUN, B.; ZHU, Y.; XIAO, Y.; XIAO, R.; WEI, Y. Automatic question tagging with deep neural networks. **IEEE Transactions on Learning Technologies**, IEEE, v. 12, n. 1, p. 29–43, 2018. Citado nas páginas 33, 62 e 83.

SUN, Y.; HAN, J. Mining heterogeneous information networks: principles and methodologies. **Synthesis Lectures on Data Mining and Knowledge Discovery**, Morgan & Claypool Publishers, v. 3, n. 2, p. 1–159, 2012. Citado na página 40.

SUPRAJA, S.; HARTMAN, K.; TATINATI, S.; KHONG, A. W.; HU, X.; BARNES, T.; HERSHKOVITZ, A.; PAQUETTE, L. Toward the automatic labeling of course questions for ensuring their alignment with learning outcomes. In: **EDM**. [S.l.: s.n.], 2017. Citado na página 63.

TAN, H. K. **Standalone application of question classification in higher education institutions**. Tese (Doutorado) — UTAR, 2018. Citado na página 63.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. [S.l.]: Pearson Education India, 2016. Citado na página 103.

TAQI, M. K.; ALI, R. Automatic question classification models for computer programming examination: A systematic literature review. **Journal of Theoretical and Applied Information Technology**, v. 93, n. 2, 2016. Citado nas páginas 39, 48, 51 e 54.

TOMAS, D.; VICEDO, J. L. Minimally supervised question classification on fine-grained taxonomies. **KNOWLEDGE AND INFORMATION SYSTEMS**, 36, n. 2, p. 303–334, AUG 2013. ISSN 0219-1377. Citado nas páginas 62 e 69.

VAN-TU, N.; ANH-CUONG, L. Improving question classification by feature extraction and selection. **Indian Journal of Science and Technology**, v. 9, n. 17, 2016. Citado nas páginas 45 e 62.

WAN, Q.; HUANG, S.; WEI, M. Research on pretreatment of questions based on large-scale real questions set. **Journal of Networks**, Academy Publisher, v. 8, n. 8, p. 1810–1817, 2013. Citado na página 62.

- WANG, D.; ZHANG, H. Inverse-category-frequency based supervised term weighting schemes for text categorization. **Journal of Information Science and Engineering**, Institute of Information Science Academia Sinica, v. 29, n. 2, p. 209–225, 2013. Citado na página 62.
- WANG, F.; YANG, Z.; LI, Z.; ZHOU, J. Query classification by leveraging explicit concept information. In: SPRINGER. **Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings 12**. [S.l.], 2016. p. 636–650. Citado na página 62.
- WANG, P.; XU, B.; XU, J.; TIAN, G.; LIU, C.-L.; HAO, H. Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification. **Neurocomputing**, Elsevier, v. 174, p. 806–814, 2016. Citado na página 62.
- WANG, P.; XU, J.; XU, B.; LIU, C.-L.; ZHANG, H.; WANG, F.; HAO, H. Semantic clustering and convolutional neural network for short text categorization. In: **ACL (2)**. [S.l.: s.n.], 2015. p. 352–357. Citado na página 62.
- WANG, R.; CUI, X.; WANG, C. Research of test questions classification based on hybrid frame mixing semantic comprehension and machine learning. In: ACM. **Proceedings of the 2017 International Conference on Robotics and Artificial Intelligence**. [S.l.], 2017. p. 1–5. Citado na página 63.
- WAUTERS, K.; DESMET, P.; NOORTGATE, W. V. D. Adaptive item-based learning environments based on the item response theory: possibilities and challenges. **Journal of Computer Assisted Learning**, Wiley Online Library, v. 26, n. 6, p. 549–562, 2010. Citado na página 33.
- WEISS, S. M.; INDURKHYA, N.; ZHANG, T. **Fundamentals of predictive text mining**. [S.l.]: Springer Science & Business Media, 2010. Citado na página 34.
- _____. **Fundamentals of predictive text mining**. [S.l.]: Springer, 2015. Citado nas páginas 29, 103 e 104.
- WIKIPÉDIA. 2018. Disponível em: <https://en.wikipedia.org/wiki/F1_score>. Acesso em: 01/12/2018. Citado na página 52.
- _____. **Perceptron multicamadas — Wikipédia, a enciclopédia livre**. 2019. [Online; accessed 9-junho-2019]. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Perceptron_multicamadas&oldid=55443187>. Citado na página 50.
- XIA, W.; ZHU, W.; LIAO, B.; CHEN, M.; CAI, L.; HUANG, L. Novel architecture for long short-term memory used in question classification. **Neurocomputing**, Elsevier, v. 299, p. 20–31, 2018. Citado na página 63.
- YADAV, K.; MISHRA, M. Question classification: Using support vector machine and lexical, semantic and syntactic features. Citeseer. Citado na página 62.
- YAHYA, A. A. Swarm intelligence-based approach for educational data classification. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, 2017. Citado na página 63.
- YAHYA, A. A.; OSAMA, A. Automatic classification of questions into bloom's cognitive levels using support vector machines. Naif Arab University for Security Sciences, 2011. Citado na página 82.

- YAHYA, A. A.; OSMAN, A. Classification of high dimensional educational data using particle swarm classification. In: IEEE. **Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on**. [S.l.], 2014. p. 34–41. Citado nas páginas 55 e 62.
- YAHYA, A. A.; OSMAN, A.; EL-BASHIR, M. S. Rocchio algorithm-based particle initialization mechanism for effective {PSO} classification of high dimensional data. **Swarm and Evolutionary Computation**, p. –, 2016. ISSN 2210-6502. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2210650216304515>>. Citado na página 62.
- YAHYA, A. A.; OSMAN, A.; TALEB, A.; ALATTAB, A. A. Analyzing the cognitive level of classroom questions using machine learning techniques. **Procedia - Social and Behavioral Sciences**, v. 97, p. 587 – 595, 2013. ISSN 1877-0428. The 9th International Conference on Cognitive Science. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877042813037221>>. Citado na página 55.
- YOSHIKAWA, Y.; IWATA, T.; SAWADA, H. Latent support measure machines for bag-of-words data classification. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2014. p. 1961–1969. Citado na página 62.
- YUSOF, N.; HUI, C. J. Determination of bloom's cognitive level of question items using artificial neural network. In: IEEE. **Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on**. [S.l.], 2010. p. 866–870. Citado na página 81.
- ZAMIR, O.; ETZIONI, O. Web document clustering: A feasibility demonstration. In: CITESEER. **SIGIR**. [S.l.], 1998. v. 98, p. 46–54. Citado na página 32.
- ZHANG, D.; LEE, W. S. Question classification using support vector machines. In: ACM. **Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval**. [S.l.], 2003. p. 26–32. Citado na página 62.
- ZHANG, D.; TIAN, L.; HONG, M.; HAN, F.; REN, Y.; CHEN, Y. Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification. **IEEE Access**, IEEE, v. 6, p. 73750–73759, 2018. Citado na página 63.
- ZHANG, L.; LI, Y.; XU, Y.; TJONDRONEGORO, D.; SUN, C. Centroid training to achieve effective text classification. In: **2014 International Conference on Data Science and Advanced Analytics (DSAA)**. [S.l.: s.n.], 2014. p. 406–412. Citado na página 62.
- ZHANG, R.; LEE, H.; RADEV, D. Dependency sensitive convolutional neural networks for modeling sentences and documents. **arXiv preprint arXiv:1611.02361**, 2016. Citado na página 62.
- ZHANG, Y.; ROLLER, S.; WALLACE, B. Mgnc-cnn: A simple approach to exploiting multiple word embeddings for sentence classification. **arXiv preprint arXiv:1603.00968**, 2016. Citado na página 62.
- ZHANG, Z.; LIN, H.; LI, P.; WANG, H.; LU, D. Improving semi-supervised text classification by using wikipedia knowledge. In: SPRINGER. **International Conference on Web-Age Information Management**. [S.l.], 2013. p. 25–36. Citado na página 62.
- ZHOU, C.; SUN, C.; LIU, Z.; LAU, F. A c-lstm neural network for text classification. **arXiv preprint arXiv:1511.08630**, 2015. Citado na página 62.

ZHOU, R.; DU, X. A hybrid neural network for sentence classification. In: SISSA MEDIALAB. **The 7th International Conference on Computer Engineering and Networks**. [S.l.], 2017. v. 299, p. 057. Citado na página [63](#).

MATRIZ DE REFERÊNCIA DO ENEM

MINISTÉRIO DA EDUCAÇÃO
INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

MATRIZ DE REFERÊNCIA ENEM

EIXOS COGNITIVOS (comuns a todas as áreas de conhecimento)

- I. **Dominar linguagens (DL):** dominar a norma culta da Língua Portuguesa e fazer uso das linguagens matemática, artística e científica e das línguas espanhola e inglesa.
- II. **Compreender fenômenos (CF):** construir e aplicar conceitos das várias áreas do conhecimento para a compreensão de fenômenos naturais, de processos histórico-geográficos, da produção tecnológica e das manifestações artísticas.
- III. **Enfrentar situações-problema (SP):** selecionar, organizar, relacionar, interpretar dados e informações representados de diferentes formas, para tomar decisões e enfrentar situações-problema.
- IV. **Construir argumentação (CA):** relacionar informações, representadas em diferentes formas, e conhecimentos disponíveis em situações concretas, para construir argumentação consistente.
- V. **Elaborar propostas (EP):** recorrer aos conhecimentos desenvolvidos na escola para elaboração de propostas de intervenção solidária na realidade, respeitando os valores humanos e considerando a diversidade sociocultural.

Matriz de Referência de Linguagens, Códigos e suas Tecnologias

Competência de área 1 - Aplicar as tecnologias da comunicação e da informação na escola, no trabalho e em outros contextos relevantes para sua vida.

H1 - Identificar as diferentes linguagens e seus recursos expressivos como elementos de caracterização dos sistemas de comunicação.

H2 - Recorrer aos conhecimentos sobre as linguagens dos sistemas de comunicação e informação para resolver problemas sociais.

H3 - Relacionar informações geradas nos sistemas de comunicação e informação, considerando a função social desses sistemas.

H4 - Reconhecer posições críticas aos usos sociais que são feitos das linguagens e dos sistemas de comunicação e informação.

Competência de área 2 - Conhecer e usar língua(s) estrangeira(s) moderna(s) como instrumento de acesso a informações e a outras culturas e grupos sociais*.

H5 – Associar vocábulos e expressões de um texto em LEM ao seu tema.

H6 - Utilizar os conhecimentos da LEM e de seus mecanismos como meio de ampliar as possibilidades de acesso a informações, tecnologias e culturas.

H7 – Relacionar um texto em LEM, as estruturas linguísticas, sua função e seu uso social.

H8 - Reconhecer a importância da produção cultural em LEM como representação da diversidade cultural e linguística.

Competência de área 3 - Compreender e usar a linguagem corporal como relevante para a própria vida, integradora social e formadora da identidade.

H9 - Reconhecer as manifestações corporais de movimento como originárias de necessidades cotidianas de um grupo social.

H10 - Reconhecer a necessidade de transformação de hábitos corporais em função das necessidades cinestésicas.

H11 - Reconhecer a linguagem corporal como meio de interação social, considerando os limites de desempenho e as alternativas de adaptação para diferentes indivíduos.

Competência de área 4 - Compreender a arte como saber cultural e estético gerador de significação e integrador da organização do mundo e da própria identidade.

H12 - Reconhecer diferentes funções da arte, do trabalho da produção dos artistas em seus meios culturais.

H13 - Analisar as diversas produções artísticas como meio de explicar diferentes culturas, padrões de beleza e preconceitos.

H14 - Reconhecer o valor da diversidade artística e das inter-relações de elementos que se apresentam nas manifestações de vários grupos sociais e étnicos.

Competência de área 5 - Analisar, interpretar e aplicar recursos expressivos das linguagens, relacionando textos com seus contextos, mediante a natureza, função, organização, estrutura das manifestações, de acordo com as condições de produção e recepção.

H15 - Estabelecer relações entre o texto literário e o momento de sua produção, situando aspectos do contexto histórico, social e político.

H16 - Relacionar informações sobre concepções artísticas e procedimentos de construção do texto literário.

H17 - Reconhecer a presença de valores sociais e humanos atualizáveis e permanentes no patrimônio literário nacional.

Competência de área 6 - Compreender e usar os sistemas simbólicos das diferentes linguagens como meios de organização cognitiva da realidade pela constituição de significados, expressão, comunicação e informação.

H18 - Identificar os elementos que concorrem para a progressão temática e para a organização e estruturação de textos de diferentes gêneros e tipos.

H19 - Analisar a função da linguagem predominante nos textos em situações específicas de interlocução.

H20 - Reconhecer a importância do patrimônio linguístico para a preservação da memória e da identidade nacional.

Competência de área 7 - Confrontar opiniões e pontos de vista sobre as diferentes linguagens e suas manifestações específicas.

H21 - Reconhecer em textos de diferentes gêneros, recursos verbais e não-verbais utilizados com a finalidade de criar e mudar comportamentos e hábitos.

H22 - Relacionar, em diferentes textos, opiniões, temas, assuntos e recursos linguísticos.

H23 - Inferir em um texto quais são os objetivos de seu produtor e quem é seu público alvo, pela análise dos procedimentos argumentativos utilizados.

H24 - Reconhecer no texto estratégias argumentativas empregadas para o convencimento do público, tais como a intimidação, sedução, comoção, chantagem, entre outras.

Competência de área 8 - Compreender e usar a língua portuguesa como língua materna, geradora de significação e integradora da organização do mundo e da própria identidade.

H25 - Identificar, em textos de diferentes gêneros, as marcas linguísticas que singularizam as variedades linguísticas sociais, regionais e de registro.

H26 - Relacionar as variedades linguísticas a situações específicas de uso social.

H27 - Reconhecer os usos da norma padrão da língua portuguesa nas diferentes situações de comunicação.

Competência de área 9 - Entender os princípios, a natureza, a função e o impacto das tecnologias da comunicação e da informação na sua vida pessoal e social, no desenvolvimento do conhecimento, associando-o aos conhecimentos científicos, às linguagens que lhes dão suporte, às demais tecnologias, aos processos de produção e aos problemas que se propõem solucionar.

H28 - Reconhecer a função e o impacto social das diferentes tecnologias da comunicação e informação.

H29 - Identificar pela análise de suas linguagens, as tecnologias da comunicação e informação.

H30 - Relacionar as tecnologias de comunicação e informação ao desenvolvimento das sociedades e ao conhecimento que elas produzem.

Matriz de Referência de Matemática e suas Tecnologias

Competência de área 1 - Construir significados para os números naturais, inteiros, racionais e reais.

H1 - Reconhecer, no contexto social, diferentes significados e representações dos números e operações - naturais, inteiros, racionais ou reais.

H2 - Identificar padrões numéricos ou princípios de contagem.

H3 - Resolver situação-problema envolvendo conhecimentos numéricos.

H4 - Avaliar a razoabilidade de um resultado numérico na construção de argumentos sobre afirmações quantitativas.

H5 - Avaliar propostas de intervenção na realidade utilizando conhecimentos numéricos.

Competência de área 2 - Utilizar o conhecimento geométrico para realizar a leitura e a representação da realidade e agir sobre ela.

H6 - Interpretar a localização e a movimentação de pessoas/objetos no espaço tridimensional e sua representação no espaço bidimensional.

H7 - Identificar características de figuras planas ou espaciais.

H8 - Resolver situação-problema que envolva conhecimentos geométricos de espaço e forma.

H9 - Utilizar conhecimentos geométricos de espaço e forma na seleção de argumentos propostos como solução de problemas do cotidiano.

Competência de área 3 - Construir noções de grandezas e medidas para a compreensão da realidade e a solução de problemas do cotidiano.

H10 - Identificar relações entre grandezas e unidades de medida.

H11 - Utilizar a noção de escalas na leitura de representação de situação do cotidiano.

H12 - Resolver situação-problema que envolva medidas de grandezas.

H13 - Avaliar o resultado de uma medição na construção de um argumento consistente.

H14 - Avaliar proposta de intervenção na realidade utilizando conhecimentos geométricos relacionados a grandezas e medidas.

Competência de área 4 - Construir noções de variação de grandezas para a compreensão da realidade e a solução de problemas do cotidiano.

H15 - Identificar a relação de dependência entre grandezas.

H16 - Resolver situação-problema envolvendo a variação de grandezas, direta ou inversamente proporcionais.

H17 - Analisar informações envolvendo a variação de grandezas como recurso para a construção de argumentação.

H18 - Avaliar propostas de intervenção na realidade envolvendo variação de grandezas.

Competência de área 5 - Modelar e resolver problemas que envolvem variáveis socioeconômicas ou técnico-científicas, usando representações algébricas.

H19 - Identificar representações algébricas que expressem a relação entre grandezas.

H20 - Interpretar gráfico cartesiano que represente relações entre grandezas.

H21 - Resolver situação-problema cuja modelagem envolva conhecimentos algébricos.

H22 - Utilizar conhecimentos algébricos/geométricos como recurso para a construção de argumentação.

H23 - Avaliar propostas de intervenção na realidade utilizando conhecimentos algébricos.

Competência de área 6 - Interpretar informações de natureza científica e social obtidas da leitura de gráficos e tabelas, realizando previsão de tendência, extrapolação, interpolação e interpretação.

H24 - Utilizar informações expressas em gráficos ou tabelas para fazer inferências.

H25 - Resolver problema com dados apresentados em tabelas ou gráficos.

H26 - Analisar informações expressas em gráficos ou tabelas como recurso para a construção de argumentos.

Competência de área 7 - Compreender o caráter aleatório e não-determinístico dos fenômenos naturais e sociais e utilizar instrumentos adequados para medidas, determinação de amostras e cálculos de probabilidade para interpretar informações de variáveis apresentadas em uma distribuição estatística.

H27 - Calcular medidas de tendência central ou de dispersão de um conjunto de dados expressos em uma tabela de frequências de dados agrupados (não em classes) ou em gráficos.

H28 - Resolver situação-problema que envolva conhecimentos de estatística e probabilidade.

H29 - Utilizar conhecimentos de estatística e probabilidade como recurso para a construção de argumentação.

H30 - Avaliar propostas de intervenção na realidade utilizando conhecimentos de estatística e probabilidade.

Matriz de Referência de Ciências da Natureza e suas Tecnologias

Competência de área 1 – Compreender as ciências naturais e as tecnologias a elas associadas como construções humanas, percebendo seus papéis nos processos de produção e no desenvolvimento econômico e social da humanidade.

H1 – Reconhecer características ou propriedades de fenômenos ondulatórios ou oscilatórios, relacionando-os a seus usos em diferentes contextos.

H2 – Associar a solução de problemas de comunicação, transporte, saúde ou outro, com o correspondente desenvolvimento científico e tecnológico.

H3 – Confrontar interpretações científicas com interpretações baseadas no senso comum, ao longo do tempo ou em diferentes culturas.

H4 – Avaliar propostas de intervenção no ambiente, considerando a qualidade da vida humana ou medidas de conservação, recuperação ou utilização sustentável da biodiversidade.

Competência de área 2 – Identificar a presença e aplicar as tecnologias associadas às ciências naturais em diferentes contextos.

H5 – Dimensionar circuitos ou dispositivos elétricos de uso cotidiano.

H6 – Relacionar informações para compreender manuais de instalação ou utilização de aparelhos, ou sistemas tecnológicos de uso comum.

H7 – Selecionar testes de controle, parâmetros ou critérios para a comparação de materiais e produtos, tendo em vista a defesa do consumidor, a saúde do trabalhador ou a qualidade de vida.

Competência de área 3 – Associar intervenções que resultam em degradação ou conservação ambiental a processos produtivos e sociais e a instrumentos ou ações científico-tecnológicos.

H8 – Identificar etapas em processos de obtenção, transformação, utilização ou reciclagem de recursos naturais, energéticos ou matérias-primas, considerando processos biológicos, químicos ou físicos neles envolvidos.

H9 – Compreender a importância dos ciclos biogeoquímicos ou do fluxo energia para a vida, ou da ação de agentes ou fenômenos que podem causar alterações nesses processos.

H10 – Analisar perturbações ambientais, identificando fontes, transporte e(ou) destino dos poluentes ou prevendo efeitos em sistemas naturais, produtivos ou sociais.

H11 – Reconhecer benefícios, limitações e aspectos éticos da biotecnologia, considerando estruturas e processos biológicos envolvidos em produtos biotecnológicos.

H12 – Avaliar impactos em ambientes naturais decorrentes de atividades sociais ou econômicas, considerando interesses contraditórios.

Competência de área 4 – Compreender interações entre organismos e ambiente, em particular aquelas relacionadas à saúde humana, relacionando conhecimentos científicos, aspectos culturais e características individuais.

H13 – Reconhecer mecanismos de transmissão da vida, prevendo ou explicando a manifestação de características dos seres vivos.

H14 – Identificar padrões em fenômenos e processos vitais dos organismos, como manutenção do equilíbrio interno, defesa, relações com o ambiente, sexualidade, entre outros.

H15 – Interpretar modelos e experimentos para explicar fenômenos ou processos biológicos em qualquer nível de organização dos sistemas biológicos.

H16 – Compreender o papel da evolução na produção de padrões, processos biológicos ou na organização taxonômica dos seres vivos.

Competência de área 5 – Entender métodos e procedimentos próprios das ciências naturais e aplicá-los em diferentes contextos.

H17 – Relacionar informações apresentadas em diferentes formas de linguagem e representação usadas nas ciências físicas, químicas ou biológicas, como texto discursivo, gráficos, tabelas, relações matemáticas ou linguagem simbólica.

H18 – Relacionar propriedades físicas, químicas ou biológicas de produtos, sistemas ou procedimentos tecnológicos às finalidades a que se destinam.

H19 – Avaliar métodos, processos ou procedimentos das ciências naturais que contribuam para diagnosticar ou solucionar problemas de ordem social, econômica ou ambiental.

Competência de área 6 – Apropriar-se de conhecimentos da física para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

H20 – Caracterizar causas ou efeitos dos movimentos de partículas, substâncias, objetos ou corpos celestes.

H21 – Utilizar leis físicas e (ou) químicas para interpretar processos naturais ou tecnológicos inseridos no contexto da termodinâmica e(ou) do eletromagnetismo.

H22 – Compreender fenômenos decorrentes da interação entre a radiação e a matéria em suas manifestações em processos naturais ou tecnológicos, ou em suas implicações biológicas, sociais, econômicas ou ambientais.

H23 – Avaliar possibilidades de geração, uso ou transformação de energia em ambientes específicos, considerando implicações éticas, ambientais, sociais e/ou econômicas.

Competência de área 7 – Apropriar-se de conhecimentos da química para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

H24 – Utilizar códigos e nomenclatura da química para caracterizar materiais, substâncias ou transformações químicas.

H25 – Caracterizar materiais ou substâncias, identificando etapas, rendimentos ou implicações biológicas, sociais, econômicas ou ambientais de sua obtenção ou produção.

H26 – Avaliar implicações sociais, ambientais e/ou econômicas na produção ou no consumo de recursos energéticos ou minerais, identificando transformações químicas ou de energia envolvidas nesses processos.

H27 – Avaliar propostas de intervenção no meio ambiente aplicando conhecimentos químicos, observando riscos ou benefícios.

Competência de área 8 – Apropriar-se de conhecimentos da biologia para, em situações problema, interpretar, avaliar ou planejar intervenções científico-tecnológicas.

H28 – Associar características adaptativas dos organismos com seu modo de vida ou com seus limites de distribuição em diferentes ambientes, em especial em ambientes brasileiros.

H29 – Interpretar experimentos ou técnicas que utilizam seres vivos, analisando implicações para o ambiente, a saúde, a produção de alimentos, matérias primas ou produtos industriais.

H30 – Avaliar propostas de alcance individual ou coletivo, identificando aquelas que visam à preservação e a implementação da saúde individual, coletiva ou do ambiente.

Matriz de Referência de Ciências Humanas e suas Tecnologias

Competência de área 1 - Compreender os elementos culturais que constituem as identidades

H1 - Interpretar historicamente e/ou geograficamente fontes documentais acerca de aspectos da cultura.

H2 - Analisar a produção da memória pelas sociedades humanas.

H3 - Associar as manifestações culturais do presente aos seus processos históricos.

H4 - Comparar pontos de vista expressos em diferentes fontes sobre determinado aspecto da cultura.

H5 - Identificar as manifestações ou representações da diversidade do patrimônio cultural e artístico em diferentes sociedades.

Competência de área 2 - Compreender as transformações dos espaços geográficos como produto das relações socioeconômicas e culturais de poder.

H6 - Interpretar diferentes representações gráficas e cartográficas dos espaços geográficos.

H7 - Identificar os significados histórico-geográficos das relações de poder entre as nações

H8 - Analisar a ação dos estados nacionais no que se refere à dinâmica dos fluxos populacionais e no enfrentamento de problemas de ordem econômico-social.

H9 - Comparar o significado histórico-geográfico das organizações políticas e socioeconômicas em escala local, regional ou mundial.

H10 - Reconhecer a dinâmica da organização dos movimentos sociais e a importância da participação da coletividade na transformação da realidade histórico-geográfica.

Competência de área 3 - Compreender a produção e o papel histórico das instituições sociais, políticas e econômicas, associando-as aos diferentes grupos, conflitos e movimentos sociais.

H11 - Identificar registros de práticas de grupos sociais no tempo e no espaço.

H12 - Analisar o papel da justiça como instituição na organização das sociedades.

H13 - Analisar a atuação dos movimentos sociais que contribuíram para mudanças ou rupturas em processos de disputa pelo poder.

H14 - Comparar diferentes pontos de vista, presentes em textos analíticos e interpretativos, sobre situação ou fatos de natureza histórico-geográfica acerca das instituições sociais, políticas e econômicas.

H15 - Avaliar criticamente conflitos culturais, sociais, políticos, econômicos ou ambientais ao longo da história.

Competência de área 4 - Entender as transformações técnicas e tecnológicas e seu impacto nos processos de produção, no desenvolvimento do conhecimento e na vida social.

H16 - Identificar registros sobre o papel das técnicas e tecnologias na organização do trabalho e/ou da vida social.

H17 - Analisar fatores que explicam o impacto das novas tecnologias no processo de territorialização da produção.

H18 - Analisar diferentes processos de produção ou circulação de riquezas e suas implicações sócio-espaciais.

H19 - Reconhecer as transformações técnicas e tecnológicas que determinam as várias formas de uso e apropriação dos espaços rural e urbano.

H20 - Selecionar argumentos favoráveis ou contrários às modificações impostas pelas novas tecnologias à vida social e ao mundo do trabalho.

Competência de área 5 - Utilizar os conhecimentos históricos para compreender e valorizar os fundamentos da cidadania e da democracia, favorecendo uma atuação consciente do indivíduo na sociedade.

H21 - Identificar o papel dos meios de comunicação na construção da vida social.

H22 - Analisar as lutas sociais e conquistas obtidas no que se refere às mudanças nas legislações ou nas políticas públicas.

H23 - Analisar a importância dos valores éticos na estruturação política das sociedades.

H24 - Relacionar cidadania e democracia na organização das sociedades.

H25 – Identificar estratégias que promovam formas de inclusão social.

Competência de área 6 - Compreender a sociedade e a natureza, reconhecendo suas interações no espaço em diferentes contextos históricos e geográficos.

H26 - Identificar em fontes diversas o processo de ocupação dos meios físicos e as relações da vida humana com a paisagem.

H27 - Analisar de maneira crítica as interações da sociedade com o meio físico, levando em consideração aspectos históricos e(ou) geográficos.

H28 - Relacionar o uso das tecnologias com os impactos sócio-ambientais em diferentes contextos histórico-geográficos.

H29 - Reconhecer a função dos recursos naturais na produção do espaço geográfico, relacionando-os com as mudanças provocadas pelas ações humanas.

H30 - Avaliar as relações entre preservação e degradação da vida no planeta nas diferentes escalas.

ANEXO

Objetos de conhecimento associados às Matrizes de Referência

1. Linguagem, Códigos e suas Tecnologias

- **Estudo do texto: as sequências discursivas e os gêneros textuais no sistema de comunicação e informação** - modos de organização da composição textual; atividades de produção escrita e de leitura de textos gerados nas diferentes esferas sociais - públicas e privadas.
- **Estudo das práticas corporais: a linguagem corporal como integradora social e formadora de identidade** - *performance* corporal e identidades juvenis; possibilidades de vivência crítica e emancipada do lazer; mitos e verdades sobre os corpos masculino e feminino na sociedade atual; exercício físico e saúde; o corpo e a expressão artística e cultural; o corpo no mundo dos símbolos e como produção da cultura; práticas corporais e autonomia; condicionamentos e esforços físicos; o esporte; a dança; as lutas; os jogos; as brincadeiras.
- **Produção e recepção de textos artísticos: interpretação e representação do mundo para o fortalecimento dos processos de identidade e cidadania** – Artes Visuais: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade. Teatro: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Música: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Dança: estrutura morfológica, sintática, o contexto da obra artística, o contexto da comunidade, as fontes de criação. Conteúdos estruturantes das linguagens artísticas (Artes Visuais, Dança, Música, Teatro), elaborados a partir de suas estruturas morfológicas e sintáticas; inclusão, diversidade e multiculturalidade: a valorização da pluralidade expressada nas produções estéticas e artísticas das minorias sociais e dos portadores de necessidades especiais educacionais.
- **Estudo do texto literário: relações entre produção literária e processo social, concepções artísticas, procedimentos de construção e recepção de textos** - produção literária e processo social; processos de formação literária e de formação nacional; produção de textos literários, sua recepção e a constituição do patrimônio literário nacional; relações entre a dialética cosmopolitismo/localismo e a produção literária nacional;

elementos de continuidade e ruptura entre os diversos momentos da literatura brasileira; associações entre concepções artísticas e procedimentos de construção do texto literário em seus gêneros (épico/narrativo, lírico e dramático) e formas diversas.; articulações entre os recursos expressivos e estruturais do texto literário e o processo social relacionado ao momento de sua produção; representação literária: natureza, função, organização e estrutura do texto literário; relações entre literatura, outras artes e outros saberes.

- **Estudo dos aspectos linguísticos em diferentes textos: recursos expressivos da língua, procedimentos de construção e recepção de textos** - organização da macroestrutura semântica e a articulação entre idéias e proposições (relações lógico-semânticas).

- **Estudo do texto argumentativo, seus gêneros e recursos linguísticos: argumentação: tipo, gêneros e usos em língua portuguesa** - formas de apresentação de diferentes pontos de vista; organização e progressão textual; papéis sociais e comunicativos dos interlocutores, relação entre usos e propósitos comunicativos, função sociocomunicativa do gênero, aspectos da dimensão espaçotemporal em que se produz o texto.

- **Estudo dos aspectos linguísticos da língua portuguesa: usos da língua: norma culta e variação linguística** - uso dos recursos linguísticos em relação ao contexto em que o texto é constituído: elementos de referência pessoal, temporal, espacial, registro linguístico, grau de formalidade, seleção lexical, tempos e modos verbais; uso dos recursos linguísticos em processo de coesão textual: elementos de articulação das sequências dos textos ou à construção da micro estrutura do texto.

- **Estudo dos gêneros digitais: tecnologia da comunicação e informação: impacto e função social** - o texto literário típico da cultura de massa: o suporte textual em gêneros digitais; a caracterização dos interlocutores na comunicação tecnológica; os recursos linguísticos e os gêneros digitais; a função social das novas tecnologias.

2. Matemática e suas Tecnologias

- **Conhecimentos numéricos:** operações em conjuntos numéricos (naturais, inteiros, racionais e reais), desigualdades, divisibilidade, fatoração, razões e proporções, porcentagem e juros, relações de dependência entre grandezas, sequências e progressões, princípios de contagem.
- **Conhecimentos geométricos:** características das figuras geométricas planas e espaciais; grandezas, unidades de medida e escalas; comprimentos, áreas e volumes; ângulos; posições de retas; simetrias de figuras planas ou espaciais; congruência e semelhança de triângulos; teorema de Tales; relações métricas nos triângulos; circunferências; trigonometria do ângulo agudo.
- **Conhecimentos de estatística e probabilidade:** representação e análise de dados; medidas de tendência central (médias, moda e mediana); desvios e variância; noções de probabilidade.
- **Conhecimentos algébricos:** gráficos e funções; funções algébricas do 1.º e do 2.º grau, polinomiais, racionais, exponenciais e logarítmicas; equações e inequações; relações no ciclo trigonométrico e funções trigonométricas.
- **Conhecimentos algébricos/geométricos:** plano cartesiano; retas; circunferências; paralelismo e perpendicularidade, sistemas de equações.

3. Ciências da Natureza e suas Tecnologias

3.1 Física

- **Conhecimentos básicos e fundamentais** - Noções de ordem de grandeza. Notação Científica. Sistema Internacional de Unidades. Metodologia de investigação: a procura de regularidades e de sinais na interpretação física do mundo. Observações e mensurações: representação de grandezas físicas como grandezas mensuráveis. Ferramentas básicas: gráficos e vetores. Conceituação de grandezas vetoriais e escalares. Operações básicas com vetores.

- **O movimento, o equilíbrio e a descoberta de leis físicas** – Grandezas fundamentais da mecânica: tempo, espaço, velocidade e aceleração. Relação histórica entre força e movimento. Descrições do movimento e sua interpretação: quantificação do movimento e sua descrição matemática e gráfica. Casos especiais de movimentos e suas regularidades observáveis. Conceito de inércia. Noção de sistemas de referência inerciais e não inerciais. Noção dinâmica de massa e quantidade de movimento (momento linear). Força e variação da quantidade de movimento. Leis de Newton. Centro de massa e a idéia de ponto material. Conceito de forças externas e internas. Lei da conservação da quantidade de movimento (momento linear) e teorema do impulso. Momento de uma força (torque). Condições de equilíbrio estático de ponto material e de corpos rígidos. Força de atrito, força peso, força normal de contato e tração. Diagramas de forças. Identificação das forças que atuam nos movimentos circulares. Noção de força centrípeta e sua quantificação. A hidrostática: aspectos históricos e variáveis relevantes. Empuxo. Princípios de Pascal, Arquimedes e Stevin: condições de flutuação, relação entre diferença de nível e pressão hidrostática.

- **Energia, trabalho e potência** - Conceituação de trabalho, energia e potência. Conceito de energia potencial e de energia cinética. Conservação de energia mecânica e dissipação de energia. Trabalho da força gravitacional e energia potencial gravitacional. Forças conservativas e dissipativas.

- **A Mecânica e o funcionamento do Universo** - Força peso. Aceleração gravitacional. Lei da Gravitação Universal. Leis de Kepler. Movimentos de corpos celestes. Influência na Terra: marés e variações climáticas. Concepções históricas sobre a origem do universo e sua evolução.

- **Fenômenos Elétricos e Magnéticos** - Carga elétrica e corrente elétrica. Lei de Coulomb. Campo elétrico e potencial elétrico. Linhas de campo. Superfícies equipotenciais. Poder das pontas. Blindagem. Capacitores. Efeito Joule. Lei de Ohm. Resistência elétrica e resistividade. Relações entre grandezas elétricas: tensão, corrente, potência e energia. Circuitos elétricos simples. Correntes contínua e alternada. Medidores elétricos.

Representação gráfica de circuitos. Símbolos convencionais. Potência e consumo de energia em dispositivos elétricos. Campo magnético. Ímãs permanentes. Linhas de campo magnético. Campo magnético terrestre.

- **Oscilações, ondas, óptica e radiação** - Feixes e frentes de ondas. Reflexão e refração. Óptica geométrica: lentes e espelhos. Formação de imagens. Instrumentos ópticos simples. Fenômenos ondulatórios. Pulsos e ondas. Período, frequência, ciclo. Propagação: relação entre velocidade, frequência e comprimento de onda. Ondas em diferentes meios de propagação.

- **O calor e os fenômenos térmicos** - Conceitos de calor e de temperatura. Escalas termométricas. Transferência de calor e equilíbrio térmico. Capacidade calorífica e calor específico. Condução do calor. Dilatação térmica. Mudanças de estado físico e calor latente de transformação. Comportamento de Gases ideais. Máquinas térmicas. Ciclo de Carnot. Leis da Termodinâmica. Aplicações e fenômenos térmicos de uso cotidiano. Compreensão de fenômenos climáticos relacionados ao ciclo da água.

3.2 Química

- **Transformações Químicas** - Evidências de transformações químicas. Interpretando transformações químicas. Sistemas Gasosos: Lei dos gases. Equação geral dos gases ideais, Princípio de Avogadro, conceito de molécula; massa molar, volume molar dos gases. Teoria cinética dos gases. Misturas gasosas. Modelo corpuscular da matéria. Modelo atômico de Dalton. Natureza elétrica da matéria: Modelo Atômico de Thomson, Rutherford, Rutherford-Bohr. Átomos e sua estrutura. Número atômico, número de massa, isótopos, massa atômica. Elementos químicos e Tabela Periódica. Reações químicas.

- **Representação das transformações químicas** - Fórmulas químicas. Balanceamento de equações químicas. Aspectos quantitativos das transformações químicas. Leis ponderais das reações químicas. Determinação de fórmulas químicas. Grandezas Químicas: massa, volume, mol, massa molar, constante de Avogadro. Cálculos estequiométricos.

- **Materiais, suas propriedades e usos** - Propriedades de materiais. Estados físicos de materiais. Mudanças de estado. Misturas: tipos e métodos de separação. Substâncias químicas: classificação e características gerais. Metais e Ligas metálicas. Ferro, cobre e alumínio. Ligações metálicas. Substâncias iônicas: características e propriedades. Substâncias iônicas do grupo: cloreto, carbonato, nitrato e sulfato. Ligação iônica. Substâncias moleculares: características e propriedades. Substâncias moleculares: H_2 , O_2 , N_2 , Cl_2 , NH_3 , H_2O , HCl , CH_4 . Ligação Covalente. Polaridade de moléculas. Forças intermoleculares. Relação entre estruturas, propriedade e aplicação das substâncias.

- **Água** - Ocorrência e importância na vida animal e vegetal. Ligação, estrutura e propriedades. Sistemas em Solução Aquosa: Soluções verdadeiras, soluções coloidais e suspensões. Solubilidade. Concentração das soluções. Aspectos qualitativos das propriedades coligativas das soluções. Ácidos, Bases, Sais e Óxidos: definição, classificação, propriedades, formulação e nomenclatura. Conceitos de ácidos e base. Principais propriedades dos ácidos e bases: indicadores, condutibilidade elétrica, reação com metais, reação de neutralização.

- **Transformações Químicas e Energia** - Transformações químicas e energia calorífica. Calor de reação. Entalpia. Equações termoquímicas. Lei de Hess. Transformações químicas e energia elétrica. Reação de oxirredução. Potenciais padrão de redução. Pilha. Eletrólise. Leis de Faraday. Transformações nucleares. Conceitos fundamentais da radioatividade. Reações de fissão e fusão nuclear. Desintegração radioativa e radioisótopos.

- **Dinâmica das Transformações Químicas** - Transformações Químicas e velocidade. Velocidade de reação. Energia de ativação. Fatores que alteram a velocidade de reação: concentração, pressão, temperatura e catalisador.

- **Transformação Química e Equilíbrio** - Caracterização do sistema em equilíbrio. Constante de equilíbrio. Produto iônico da água, equilíbrio ácido-base e pH. Solubilidade dos sais e hidrólise. Fatores que alteram o sistema em equilíbrio. Aplicação da velocidade e do equilíbrio químico no cotidiano.

- **Compostos de Carbono** - Características gerais dos compostos orgânicos. Principais funções orgânicas. Estrutura e propriedades de Hidrocarbonetos. Estrutura e propriedades de compostos orgânicos oxigenados. Fermentação. Estrutura e propriedades de compostos orgânicos nitrogenados. Macromoléculas naturais e sintéticas. Noções básicas sobre polímeros. Amido, glicogênio e celulose. Borracha natural e sintética. Polietileno, poliestireno, PVC, Teflon, náilon. Óleos e gorduras, sabões e detergentes sintéticos. Proteínas e enzimas.

- **Relações da Química com as Tecnologias, a Sociedade e o Meio Ambiente** - Química no cotidiano. Química na agricultura e na saúde. Química nos alimentos. Química e ambiente. Aspectos científico-tecnológicos, socioeconômicos e ambientais associados à obtenção ou produção de substâncias químicas. Indústria Química: obtenção e utilização do cloro, hidróxido de sódio, ácido sulfúrico, amônia e ácido nítrico. Mineração e Metalurgia. Poluição e tratamento de água. Poluição atmosférica. Contaminação e proteção do ambiente.

- **Energias Químicas no Cotidiano** - Petróleo, gás natural e carvão. Madeira e hulha. Biomassa. Biocombustíveis. Impactos ambientais de combustíveis fósseis. Energia nuclear. Lixo atômico. Vantagens e desvantagens do uso de energia nuclear.

3.3 Biologia

- **Moléculas, células e tecidos** - Estrutura e fisiologia celular: membrana, citoplasma e núcleo. Divisão celular. Aspectos bioquímicos das estruturas celulares. Aspectos gerais do metabolismo celular. Metabolismo energético: fotossíntese e respiração. Codificação da informação genética. Síntese protéica. Diferenciação celular. Principais tecidos animais e vegetais. Origem e evolução das células. Noções sobre células-tronco, clonagem e tecnologia do DNA recombinante. Aplicações de biotecnologia na produção de alimentos, fármacos e componentes biológicos. Aplicações de tecnologias relacionadas ao DNA a investigações científicas, determinação da paternidade, investigação criminal e identificação de indivíduos. Aspectos éticos relacionados ao desenvolvimento biotecnológico. Biotecnologia e sustentabilidade.

- **Hereditariedade e diversidade da vida** - Princípios básicos que regem a transmissão de características hereditárias. Concepções pré-mendelianas sobre a hereditariedade. Aspectos genéticos do funcionamento do corpo humano. Antígenos e anticorpos. Grupos sanguíneos, transplantes e doenças auto-imunes. Neoplasias e a influência de fatores ambientais. Mutações gênicas e cromossômicas. Aconselhamento genético. Fundamentos genéticos da evolução. Aspectos genéticos da formação e manutenção da diversidade biológica.

- **Identidade dos seres vivos** - Níveis de organização dos seres vivos. Vírus, procariontes e eucariontes. Autótrofos e heterótrofos. Seres unicelulares e pluricelulares. Sistemática e as grandes linhas da evolução dos seres vivos. Tipos de ciclo de vida. Evolução e padrões anatômicos e fisiológicos observados nos seres vivos. Funções vitais dos seres vivos e sua relação com a adaptação desses organismos a diferentes ambientes. Embriologia, anatomia e fisiologia humana. Evolução humana. Biotecnologia e sistemática.

- **Ecologia e ciências ambientais** - Ecossistemas. Fatores bióticos e abióticos. Habitat e nicho ecológico. A comunidade biológica: teia alimentar, sucessão e comunidade clímax. Dinâmica de populações. Interações entre os seres vivos. Ciclos biogeoquímicos. Fluxo de energia no ecossistema. Biogeografia. Biomas brasileiros. Exploração e uso de recursos naturais. Problemas ambientais: mudanças climáticas, efeito estufa; desmatamento; erosão; poluição da água, do solo e do ar. Conservação e recuperação de ecossistemas. Conservação da biodiversidade. Tecnologias ambientais. Noções de saneamento básico. Noções de legislação ambiental: água, florestas, unidades de conservação; biodiversidade.

- **Origem e evolução da vida** - A biologia como ciência: história, métodos, técnicas e experimentação. Hipóteses sobre a origem do Universo, da Terra e dos seres vivos. Teorias de evolução. Explicações pré-darwinistas para a modificação das espécies. A teoria evolutiva de Charles Darwin. Teoria sintética da evolução. Seleção artificial e seu impacto sobre ambientes naturais e sobre populações humanas.

- **Qualidade de vida das populações humanas** - Aspectos biológicos da pobreza e do desenvolvimento humano. Indicadores sociais, ambientais e econômicos. Índice de desenvolvimento humano. Principais doenças que afetam a população brasileira:

caracterização, prevenção e profilaxia. Noções de primeiros socorros. Doenças sexualmente transmissíveis. Aspectos sociais da biologia: uso indevido de drogas; gravidez na adolescência; obesidade. Violência e segurança pública. Exercícios físicos e vida saudável. Aspectos biológicos do desenvolvimento sustentável. Legislação e cidadania.

4. Ciências Humanas e suas Tecnologias

• Diversidade cultural, conflitos e vida em sociedade

- Cultura Material e imaterial; patrimônio e diversidade cultural no Brasil.
- A Conquista da América. Conflitos entre europeus e indígenas na América colonial. A escravidão e formas de resistência indígena e africana na América.
- História cultural dos povos africanos. A luta dos negros no Brasil e o negro na formação da sociedade brasileira.
- História dos povos indígenas e a formação sócio-cultural brasileira.
- Movimentos culturais no mundo ocidental e seus impactos na vida política e social.

• Formas de organização social, movimentos sociais, pensamento político e ação do Estado

- Cidadania e democracia na Antiguidade; Estado e direitos do cidadão a partir da Idade Moderna; democracia direta, indireta e representativa.
- Revoluções sociais e políticas na Europa Moderna.
- Formação territorial brasileira; as regiões brasileiras; políticas de reordenamento territorial.
- As lutas pela conquista da independência política das colônias da América.
- Grupos sociais em conflito no Brasil imperial e a construção da nação.
- O desenvolvimento do pensamento liberal na sociedade capitalista e seus críticos nos séculos XIX e XX.
- Políticas de colonização, migração, imigração e emigração no Brasil nos séculos XIX e XX.
- A atuação dos grupos sociais e os grandes processos revolucionários do século XX: Revolução Bolchevique, Revolução Chinesa, Revolução Cubana.
- Geopolítica e conflitos entre os séculos XIX e XX: Imperialismo, a ocupação da Ásia e da África, as Guerras Mundiais e a Guerra Fria.

- Os sistemas totalitários na Europa do século XX: nazi-fascista, franquismo, salazarismo e stalinismo. Ditaduras políticas na América Latina: Estado Novo no Brasil e ditaduras na América.
- Conflitos político-culturais pós-Guerra Fria, reorganização política internacional e os organismos multilaterais nos séculos XX e XXI.
- A luta pela conquista de direitos pelos cidadãos: direitos civis, humanos, políticos e sociais. Direitos sociais nas constituições brasileiras. Políticas afirmativas.
- Vida urbana: redes e hierarquia nas cidades, pobreza e segregação espacial.

• **Características e transformações das estruturas produtivas**

- Diferentes formas de organização da produção: escravismo antigo, feudalismo, capitalismo, socialismo e suas diferentes experiências.
- Economia agro-exportadora brasileira: complexo açucareiro; a mineração no período colonial; a economia cafeeira; a borracha na Amazônia.
- Revolução Industrial: criação do sistema de fábrica na Europa e transformações no processo de produção. Formação do espaço urbano-industrial. Transformações na estrutura produtiva no século XX: o fordismo, o toyotismo, as novas técnicas de produção e seus impactos.
- A industrialização brasileira, a urbanização e as transformações sociais e trabalhistas.
- A globalização e as novas tecnologias de telecomunicação e suas consequências econômicas, políticas e sociais.
- Produção e transformação dos espaços agrários. Modernização da agricultura e estruturas agrárias tradicionais. O agronegócio, a agricultura familiar, os assalariados do campo e as lutas sociais no campo. A relação campo-cidade.

• **Os domínios naturais e a relação do ser humano com o ambiente**

- Relação homem-natureza, a apropriação dos recursos naturais pelas sociedades ao longo do tempo. Impacto ambiental das atividades econômicas no Brasil. Recursos minerais e energéticos: exploração e impactos. Recursos hídricos; bacias hidrográficas e seus aproveitamentos.
- As questões ambientais contemporâneas: mudança climática, ilhas de calor, efeito estufa, chuva ácida, a destruição da camada de ozônio. A nova ordem ambiental

internacional; políticas territoriais ambientais; uso e conservação dos recursos naturais, unidades de conservação, corredores ecológicos, zoneamento ecológico e econômico.

- Origem e evolução do conceito de sustentabilidade.
- Estrutura interna da terra. Estruturas do solo e do relevo; agentes internos e externos modeladores do relevo.
- Situação geral da atmosfera e classificação climática. As características climáticas do território brasileiro.
- Os grandes domínios da vegetação no Brasil e no mundo.

• **Representação espacial**

- Projeções cartográficas; leitura de mapas temáticos, físicos e políticos; tecnologias modernas aplicadas à cartografia.