



**UNIVERSIDADE DE ITAÚNA
PRÓ-REITORIA DE ENSINO
COORDENAÇÃO DE CIÊNCIA DA COMPUTACAO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

Eugênio Cunha

**APRENDIZADO DE MÁQUINA APLICADO À VALORAÇÃO DE
REDAÇÕES**

**ITAÚNA
2017**

EUGÊNIO CUNHA

APRENDIZADO DE MÁQUINA APLICADO À VALORAÇÃO DE REDAÇÕES

Projeto submetido à Coordenadoria do Curso de Bacharelado em Ciência da Computação da Universidade de Itaúna - Campus Verde, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Área de pesquisa: Aprendizagem de Máquina

Orientador: Prof. Dr. Marco Túlio Alves N Rodrigues

Itaúna
2017



UNIVERSIDADE DE ITAÚNA
COORDENAÇÃO DE CIÊNCIA DA COMPUTAÇÃO

EUGÊNIO CUNHA

Este projeto foi julgada adequada para a obtenção do Grau de Bacharel em Ciência da Computação, sendo aprovada pela coordenação de ciência da computação do curso de Bacharelado em Ciência da Computação do Campus verde da Universidade de Itaúna e pela banca examinadora:

Orientador: Prof. Dr. Marco Túlio Alves N
Rodrigues
Universidade de Itaúna- UIT

Avaliador: Prof. Me. Felipe Domingos da
Cunha
Universidade de Itaúna- UIT

Itaúna, 19 de Junho de 2017

Dedico este trabalho a Davi, meu filho, sempre preocupado em proporcionar um minuto de pausa para brincadeiras durante minhas horas de estudo, “meus melhores minutos”!

“Os computadores são incrivelmente rápidos, precisos e burros; os homens são incrivelmente lentos, imprecisos e brilhantes; juntos, seus poderes ultrapassam os limites da imaginação”.

Albert Einstein

Resumo

Este trabalho baseou-se na avaliação das competências exigidas em um texto de redação do tipo dissertativo-argumentativo com temas diversificados de ordem social, científica, cultural ou política. Fundamentou-se no estudo das técnicas de aprendizado de máquina supervisionado que provê uma gama diversificada de algoritmos poderosos para classificações de textos.

O objetivo deste trabalho é classificar as competências exigidas em um texto de redação a partir do treinamento de um modelo de aprendizado de máquina com base em um *corpus* de redações avaliadas seguindo as competências exigidas em uma redação do tipo dissertativa-argumentativa.

A compilação de um *corpus* de redações para treinamento e teste de um modelo de aprendizado de máquina exigiu a prática de extração de informações que compreende técnicas e algoritmos que realizam duas tarefas importantes: a identificação de informações desejadas a partir de documentos estruturados e não-estruturados, e o armazenamento dessas informações em um formato apropriado para uso futuro.

Afim de se avaliar a eficácia dos classificadores, vários experimentos foram executados usando-se um *corpus* extraído do banco de redações da UOL (UOL, 2017), um serviço que estimula o estudante a treinar produção de textos, em especial do gênero dissertativo-argumentativo.

O resultado de classificação das competências exigidas em um texto de redação obtidas experimentalmente mostraram que o resultado geral do sistema proposto é comparável à avaliação manual por avaliadores capacitados.

Palavras-chaves: Aprendizado de máquina, Banco de Redações, Classificação, Redação

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

Lista de Abreviações

1	Introdução	12
1.1	Definição do Problema de Pesquisa	13
1.2	Motivação	13
1.3	Objetivos Gerais e Específicos	14
1.3.1	Objetivos Especificos	14
1.4	Contribuições	14
1.5	Organização do trabalho	15
2	Trabalhos Relacionados	16
2.1	Competências requeridas pela avaliação de redação do enem	16
2.2	Modelo de representação de texto mais adequado à classificação	19
2.3	Aprendizado de máquina	20
2.4	Mineração de dados <i>Orange Data Mining</i>	22
3	Fundamentação Teórica	23
4	Método Proposto	24

5	Desenvolvimento	27
6	Resultados Experimentais	28
7	Conclusão e Trabalhos Futuros	29
	Referências Bibliográficas	30
	Apêndice A – Título do Apêndice	32
	Apêndice B – Exemplo do pacote Algorithm	33

Lista de Figuras

1	A hierarquia do aprendizado indutivo	21
2	Processo de Classificação	22
3	Um <i>Web Crawler</i> , navega entre as páginas HTML do banco de redações UOL de forma metódica e automatizada indexando textos de redações que posteriormente serão filtrados e coletados.	24
4	Os textos são submetidos aos algoritmos de normalização e posteriormente estruturados e armazenados no padrão JSON.	25
5	O <i>corpus</i> será utilizado em um fluxo de trabalho da ferramenta <i>Orange</i> para treinar os modelos classificadores.	25
6	O modelos ajustados e treinados serão submetidos a testes, e os resultados comparados graficamente.	26

Lista de Tabelas

1	Matriz de referência elaborada pelo INEP.	16
2	Modelo <i>Bag of Words</i>	19

Lista de Símbolos

<i>TAG</i>	Rótulo
<i>BOW</i>	Representação da frequência acumulada de palavras em documentos diferentes a partir de um dicionário pré-criado

Lista de Abreviações

BOW	<i>Bag of Words</i>
CEBRASPE	Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos
CSF	Ciência sem fronteiras
ENEM	Exame Nacional do Ensino Médio
HTML	<i>HyperText Markup Language</i>
IA	Inteligência artificial
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais
JSON	<i>JavaScript Object Notation</i>
SISU	Sistema de Seleção Unificada
UNB	Universidade de Brasília

INTRODUÇÃO

O desenvolvimento de uma redação é uma atividade prática presente na cultura civilizada desde a invenção da escrita. Já faz pelo menos uma década que um bom desempenho na redação do Exame Nacional de Ensino Médio - ENEM virou sinônimo de chances maiores para ser aprovado no processo seletivo de acesso a inúmeras universidades públicas (SISU, 2017) e a importantes programas de governo como Ciência sem fronteiras (CSF, 2017).

Em todo processo seletivo é comum o uso de marcações em gabaritos afim de automatizar o processo de correção, uma alternativa rápida e segura, até mesmo aplicações de provas eletrônicas são cada vez mais comum. Um exemplo seria o processo seletivo para avaliador das redações do ENEM que durante a “FASE II” respondem uma prova eletrônica eliminatória (CEBRASPE, 2016a). É notável que todo o processo evoluiu com objetivo de agilidade, confiança e segurança do resultado, entretanto a avaliação das competências de uma redação ainda depende exclusivamente da supervisão de duas ou mais pessoas envolvidas (INEP, 2016).

A redação é aplicada no ENEM desde a primeira edição 1998, hoje o maior exame do Brasil, que na edição de 2016 teve 8.627.195 escritos confirmados, e a participação direta de 11.360 profissionais externos na correção de 5.825.134 redações, entre eles, 378 supervisores e 10.982 avaliadores de acordo com a CEBRASPE (CEBRASPE, 2016b).

Segundo o edital do ENEM 2016 (INEP, 2016) cada redação foi avaliada por, pelo menos, dois avaliadores, de forma independente, contabilizando um número mínimo de 11.650.268 avaliações manuais, das competências exigidas em um texto de redação pelo ENEM.

Machine Learning ou aprendizado de máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Os

algoritmos de aprendizado de máquina procuram padrões dentro de um conjunto de dados (MITCHELL, 1997). Esses algoritmos existem há bastante tempo, uma ciência que não é nova, mas que está ganhando um novo impulso enquanto o processamento computacional cresce e fica mais barato.

A hipótese desta monografia é que um ou mais modelos de aprendizado de máquina na valoração das competências de uma redação pode ser tão eficiente quanto o processo de avaliação manual.

1.1 Definição do Problema de Pesquisa

Dado um corpus de redações avaliar as competências exigidas em um texto de redação do tipo dissertativo-argumentativo substituindo a etapa de avaliação manual.

1.2 Motivação

Com crescente volume e variedade de dados disponíveis, o processamento computacional que está mais barato e mais poderoso, e o armazenamento de dados de forma acessível, o aprendizado de máquina está no centro de muitos avanços tecnológicos atingindo áreas antes exclusivas de seres humanos. Os carros autônomos do Google são o exemplo de uma atividade antes exclusiva de um humano e hoje exercida e aperfeiçoada por algoritmos de aprendizado de máquina (WAYMO, 2017).

Aplicações de aprendizado de máquina estão presentes na nossa vida cotidiana como, resultados de pesquisa web, análise de sentimento baseado em texto e na detecção de fraudes em operações com cartões de crédito (BATISTA *et al.*, 1999).

As competências exigidas em uma redação podem ser avaliadas por aprendizado de máquina, diferente de um ser humano um algoritmo de aprendizado de máquina está livre de ansiedade, fadiga, *stress* entre outros fatores emocionais que afetam uma avaliação imparcial a opinião do autor.

1.3 Objetivos Gerais e Específicos

Este trabalho tem como objetivo geral aplicar aprendizado de máquina na avaliação das competências exigidas em um texto de redação do tipo dissertativo-argumentativo.

1.3.1 Objetivos Específicos

O método de construção do conhecimento deste trabalho terá como fundamentos processos de pesquisas relacionadas às áreas descritas. O mesmo será dividido em etapas dentro do escopo geral de forma detalhada e refinada para alcançar o objetivo geral acima, são particularizadas como os seguintes objetivos específicos:

- Percorrer o banco de redações UOL (UOL, 2017) em páginas HTML, filtrar e coletar redações avaliadas;
- Normalizar os textos coletados, separar o tema, título, texto e competências avaliadas em uma estrutura no formato JSON;
- Montar um fluxo de trabalho utilizando a ferramenta para mineração de dados *Orange* (DEMŠAR *et al.*, 2013) com modelos classificadores de múltiplas classes;
- Ajustar e treinar os modelos classificadores com o corpus de redações;
- Realizar testes de acurácia, *overfitting* e *noise* sobre os classificadores;
- Representar e comparar graficamente os resultados obtidos.

1.4 Contribuições

O presente estudo contribuirá na área do aprendizado de máquina e diretamente no processo de avaliação de um texto em prosa do tipo dissertativo argumentativo.

1.5 Organização do trabalho

Capítulo 2: Trabalhos Relacionados cita alguns dos trabalhos lidos para embasamento teórico que serviram de base para solucionar o problema proposto.

Capítulo 4: Método proposto apresenta as etapas passo a passo para desenvolver e resolver o problema proposto deste trabalho.

Capítulo 5: Desenvolvimento descreve cada procedimento metodológico que será utilizado para a realização da pesquisa.

Capítulo 6: Resultados Experimentais apresenta os resultados obtidos do trabalho desta pesquisa.

TRABALHOS RELACIONADOS

2.1 Competências requeridas pela avaliação de redação do enem

De acordo com (SILVA; CARVALHO, 2017) a prova de Redação do ENEM é avaliada levando em conta uma matriz de referência. Essa matriz, desenvolvida pelo (INEP, 2016), com a colaboração de especialistas, foi elaborada com o objetivo de operacionalizar o exame. A matriz apresenta cinco competências, para cada competência expressa para redação existem níveis de conhecimento associados de 0 a 5. Isso pode ser visto na Tabela 1.

Tabela 1: Matriz de referência elaborada pelo INEP.

Demonstrar domínio da norma padrão da língua escrita.	
I	0 Demonstra desconhecimento da modalidade escrita formal da língua portuguesa.
	1 Demonstra domínio precário da modalidade escrita formal da língua portuguesa, de forma sistemática, com diversificados e frequentes desvios gramaticais, de escolha de registro e de convenções da escrita.
	2 Demonstra domínio insuficiente da modalidade escrita formal da língua portuguesa, com muitos desvios gramaticais, de escolha de registro e de convenções da escrita.
	3 Demonstra domínio mediano da modalidade escrita formal da língua portuguesa e de escolha de registro, com alguns desvios gramaticais e de convenções da escrita.
	4 Demonstra bom domínio da modalidade escrita formal da língua portuguesa e de escolha de registro, com poucos desvios gramaticais e de convenções da escrita.

Continua na próxima página

Tabela 1 (continuação)

	5	Demonstra excelente domínio da modalidade escrita formal da língua portuguesa e de escolha de registro. Desvios gramaticais ou de convenções da escrita serão aceitos somente como excepcionalidade equando não caracterizem reincidência.
II	Compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa.	
	0	“Fuga ao tema/não atendimento à estrutura dissertativo-argumentativa”.
	1	Apresenta o assunto, tangenciando o tema ou demonstra domínio precário do texto dissertativo-argumentativo, com traços constantes de outros tipos textuais
	2	Desenvolve o tema recorrendo à cópia de trechos dos textos motivadores ou apresenta domínio insuficiente do texto dissertativo-argumentativo, não atendendo à estrutura com proposição, argumentação e conclusão.
	3	Desenvolve o tema por meio de argumentação previsível e apresenta domínio mediano do texto dissertativo-argumentativo, com proposição, argumentação e conclusão.
	4	Desenvolve o tema por meio de argumentação consistente e apresenta bom domínio do texto dissertativo-argumentativo, com proposição, argumentação e conclusão.
III	5	Desenvolve o tema por meio de argumentação consistente, a partir de um repertório sociocultural produtivo e apresenta excelente domínio do texto dissertativo-argumentativo.
	Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista.	
	0	Apresenta informações, fatos e opiniões não relacionados ao tema e sem defesa de um ponto de vista.
III	1	Apresenta informações, fatos e opiniões pouco relacionados ao tema ou incoerentes e sem defesa de um ponto de vista.

Continua na próxima página

Tabela 1 (continuação)

	2	Apresenta informações, fatos e opiniões relacionados ao tema, mas desorganizados ou contraditórios e limitados aos argumentos dos textos motivadores, em defesa de um ponto de vista.
	3	Apresenta informações, fatos e opiniões relacionados ao tema, limitados aos argumentos dos textos motivadores e pouco organizados, em defesa de um ponto de vista.
	4	Apresenta informações, fatos e opiniões relacionados ao tema, de forma organizada, com indícios de autoria, em defesa de um ponto de vista.
	5	Apresenta informações, fatos e opiniões relacionados ao tema proposto, de forma consistente e organizada, configurando autoria, em defesa de um ponto de vista.
IV	Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação.	
	0	Não articula as informações.
	1	Articula as partes do texto de forma precária.
	2	Articula as partes do texto, de forma insuficiente, com muitas inadequações e apresenta repertório limitado de recursos coesivos.
	3	Articula as partes do texto, de forma mediana, com inadequações, e apresenta repertório pouco diversificado de recursos coesivos.
	4	Articula as partes do texto com poucas inadequações e apresenta repertório diversificado de recursos coesivos.
	5	Articula bem as partes do texto e apresenta repertório diversificado de recursos coesivos.
V	Elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.	
	0	Não apresenta proposta de intervenção ou apresenta proposta não relacionada ao tema ou ao assunto.
	1	Apresenta proposta de intervenção vaga, precária ou relacionada apenas ao assunto.
	2	Elabora, de forma insuficiente, proposta de intervenção relacionada ao tema, ou não articulada com a discussão desenvolvida no texto.

Continua na próxima página

Tabela 1 (continuação)

3	Elabora, de forma mediana, proposta de intervenção relacionada ao tema e articulada à discussão desenvolvida no texto.
4	Elabora bem proposta de intervenção relacionada ao tema e articulada à discussão desenvolvida no texto.
5	Elabora muito bem proposta de intervenção, detalhada, relacionada ao tema e articulada à discussão desenvolvida no texto.

2.2 Modelo de representação de texto mais adequado à classificação

Segundo a autora (ALVES, 2010) o BOW (*Bag of Words*) é o modelo mais utilizado em aplicações de classificação de texto. Com baixo custo em termos de processamento este modelo transforma a cadeia de caracteres de um documento num conjunto de palavras, registrando além da presença de uma palavra, a sua frequência.

Entretanto propriedades básicas do texto, como a ordem em que as palavras ocorrem e a pontuação, são ignoradas, além da incapacidade em capturar a semântica do texto, isto é, há palavras com significados distintos que apesar de serem exactamente iguais têm significados diferentes, dependendo do contexto em que são utilizadas.

Obviamente, termos que aparecem em todos os documentos não serão analisados, geralmente são os pronomes, artigos e as preposições. Estes termos são chamados de *stop-words*, não são úteis dado que têm uma semântica fraca e somente desempenham um papel funcional no texto. Para melhorar os métodos de processamento normalmente são removidas, em vários casos a remoção das *stop-words* não traz consequências graves.

Na Tabela 2, w_i representa uma palavra, d_j representa um documento e p_{ij} o peso atribuído a cada palavra no documento.

Tabela 2: Modelo *Bag of Words*

	w_1	w_2	w_3	$w_{...}$	w_n
d_1	p_{11}	p_{12}	p_{13}	$p_{...}$	p_{1n}
d_2	p_{21}	p_{22}	p_{23}	$p_{...}$	p_{2n}

Continua na próxima página

Tabela 2 (continuação)

d₃	p ₃₁	p ₃₂	p ₃₃	p...	p _{3n}
d₄	p ₄₁	p ₄₂	p ₄₃	p...	p _{4n}
d_n	p _{n1}	p _{n2}	p _{n3}	p...	p _{nn}

Ainda segunda a autora existem várias medidas para calcular os valores dos pesos de p_{ij} . Essas medidas podem ser classificadas em dois tipos distintos: baseadas em frequências e binárias. Os pesos baseados em frequência visam contabilizar o número de ocorrências de um dado termo num determinado documento, servindo como base para diversas medidas estatísticas e os pesos binários indicam a ocorrência ou não de um dado termo num determinado documento.

2.3 Aprendizado de máquina

Segundo (MONARD; BARANAUSKAS, 2003) “A indução é a forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos.” Na indução, um conceito é aprendido efetuando-se inferência indutiva sobre os exemplos apresentados. O aprendizado indutivo pode ser dividido em supervisionado e não-supervisionado como ilustrado na Figura 1.

No aprendizado não-supervisionado, o algoritmo de aprendizado analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando clusters ou agrupamentos. Já no aprendizado supervisionado é fornecido ao algoritmo de aprendizado um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido.

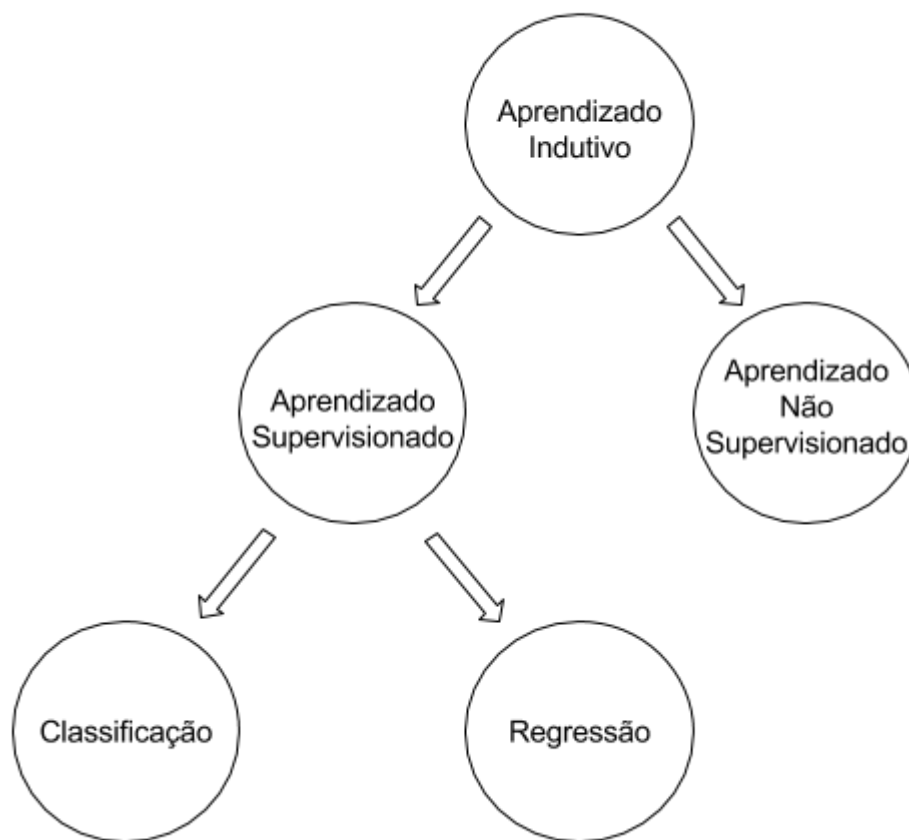


Figura 1: A hierarquia do aprendizado indutivo

De acordo com (MOTTA, 2016) classificadores são utilizados para a predição de classes de objetos e pode ser dita como o processo de generalização dos dados a partir de diferentes instâncias. Existe uma tendência de se referir a problemas com uma resposta quantitativas como problemas de regressão e aqueles com uma resposta qualitativa como problemas de classificação.

Dado um conjunto de exemplos como ilustrado na Figura 2, os classificadores devem encontrar uma função geral capaz de prever adequadamente as saídas para novos exemplos, após o treinamento, o classificador é avaliado e se necessário o processo de classificação pode ser ajustado usando o conhecimento sobre o domínio do problema para escolher os dados de entrada ao algoritmo de aprendizado.

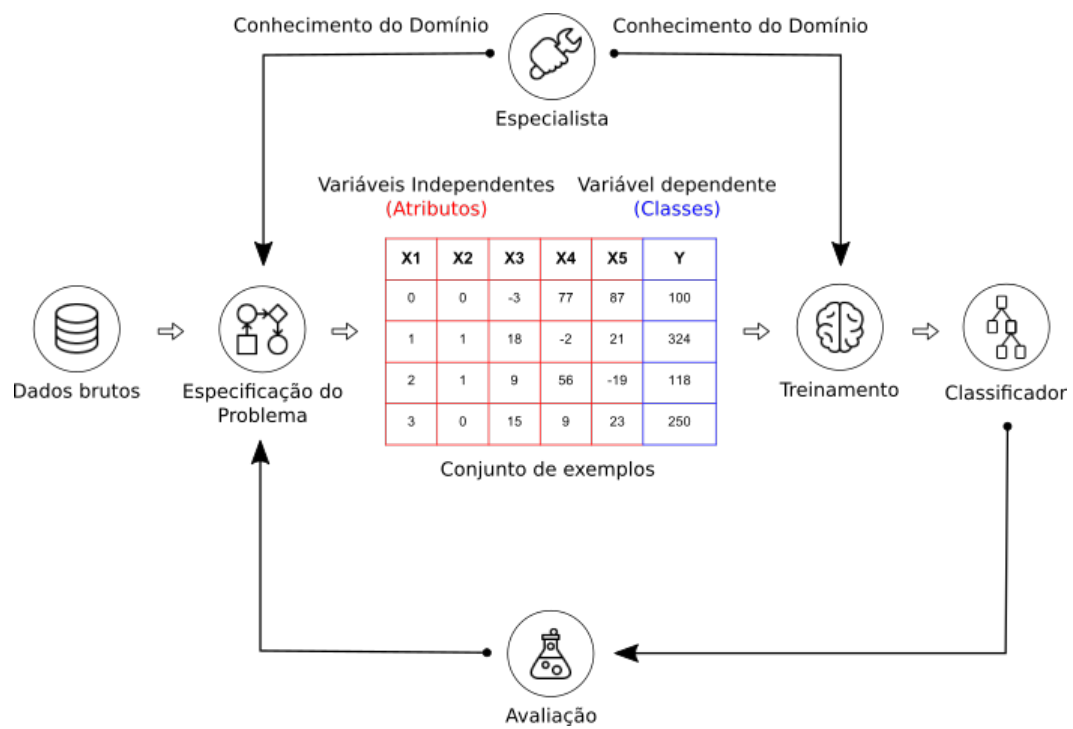


Figura 2: Processo de Classificação

2.4 Mineração de dados *Orange Data Mining*

(DEMŠAR *et al.*, 2013)

CAPÍTULO 3

FUNDAMENTAÇÃO TEÓRICA

MÉTODO PROPOSTO

Para concluir com êxito o desenvolvimento deste trabalho e consequentemente os objetivos propostos, o método utilizado para solução do problema é composto das seguintes etapas sequenciais:

Como já foi dito o banco de redações UOL foi desenvolvido e armazenado em páginas HTML, o que permite o uso de um *Web Crawler*, um algoritmo que explora a estrutura de grafo da *Web* para navegar de uma página para outra. A Figura 3 ilustra a etapa que o *Web Crawler* recupera as páginas, filtra as redações avaliadas e coleta cada uma para um repositório local.



Figura 3: Um *Web Crawler*, navega entre as páginas HTML do banco de redações UOL de forma metódica e automatizada indexando textos de redações que posteriormente serão filtrados e coletados.

Na etapa subsequente a Figura 4 ilustra a normalização dos textos, que consiste em uma técnica de remoção de caracteres não alfa-numéricos presentes no HTML e espaços desnecessários, tal que o valor textual ainda seja o mesmo que o original. Após a normalização será organizado as as diversas partes que compõem a redação (tema, título, texto e nota) em uma estrutura JSON para armazenamento e uso futuro.



Figura 4: Os textos são submetidos aos algoritmos de normalização e posteriormente estruturados e armazenados no padrão JSON.

Na terceira etapa ilustrada pela Figura 5 será utilizada a ferramenta de mineração de dados *Orange* (DEMŠAR *et al.*, 2013). Será necessário realizar estudo e análise para obter o conhecimento necessário para desenvolvimento de um fluxo de trabalho, seleção e treinamento dos modelos classificadores, concluindo todos os objetivos propostos nesta etapa.



Figura 5: O *corpus* será utilizado em um fluxo de trabalho da ferramenta *Orange* para treinar os modelos classificadores.

A quarta e última etapa é ilustrada pela Figura 6, onde os modelos classificadores previamente ajustados e treinados serão submetidos aos testes de Acurácia (taxa de predições corretas ou incorretas realizada pelo modelo para um determinado conjunto de dados), *Overfitting* (super-ajustamento que ocorre quando o modelo se especializa nos dados utilizados no seu treinamento) e *Noise* (*noise* ou ruído é classificação errada do conjunto de dados de entrada), os resultados serão representados e comparados graficamente.

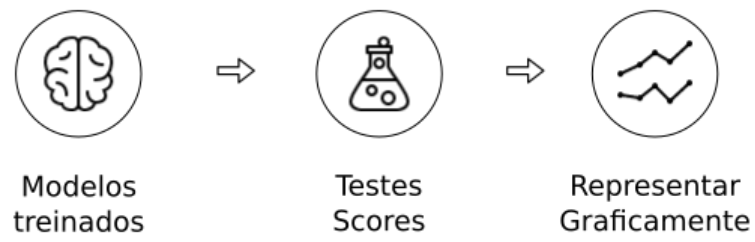


Figura 6: O modelos ajustados e treinados serão submetidos a testes, e os resultados comparados graficamente.

CAPÍTULO 5

DESENVOLVIMENTO

CAPÍTULO 6

RESULTADOS EXPERIMENTAIS

CAPÍTULO 7

CONCLUSÃO E TRABALHOS FUTUROS

Referências Bibliográficas

ALVES, Alexandra Isabel Magalhães. **Modelo de representação de texto mais adequado à classificação**. Dissertação (Mestrado) — Instituto Superior de Engenharia do Porto, 11 2010. Online; acessado 06 Junho 2017.

BATISTA, GEAPA; CARVALHO, ACPLF; MONARD, Maria C; BRASIL, Silicon Graphics. Aplicando seleção unilateral em conjuntos de exemplos desbalanceados: Resultados iniciais. In: **XIX CONGRESSO NACIONAL DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO “EDUCAÇÃO E APRENDIZAGEM NA SOCIEDADE DA INFORMAÇÃO**. [S.l.: s.n.], 1999. v. 20, p. 327–340.

CEBRASPE. **CENTRO BRASILEIRO DE PESQUISA EM AVALIAÇÃO E SELEÇÃO E DE PROMOÇÃO DE EVENTOS (CEBRASPE) PROGRAMA DE ATUALIZAÇÃO, QUALIFICAÇÃO E SELEÇÃO DE AVALIADORES DAS REDAÇÕES DO ENEM 2016**. 2016. Online; acessado 07 Abril 2017. Disponível em: <http://www.cespe.unb.br/colaboradores/ENEM_16_AVALIADOR_REDACAO/arquivos/ENEM_REDA___ES_2016___AVALIADOR___REGULAMENTO.PDF>.

CEBRASPE, CESPE UNB. **Relatório de Gestão CEBRASPE**. 2016. 1–20 p. Online; acessado 07 Abril 2017. Disponível em: <http://www.cespe.unb.br/cebraspe/arquivos/Relatorio_de_Gestao_2016.pdf>.

CSF, Ciência sem Fronteiras. **Estudante de Graduação**. 2017. Online; acessado 07 Abril 2017. Disponível em: <<http://www.cienciasemfronteiras.gov.br/web/csf/estudante>>.

DEMŠAR, Janez; CURK, Tomaž; ERJAVEC, Aleš; GORUP, Črt; HOČEVAR, Tomaž; MILUTINOVIČ, Mitar; MOŽINA, Martin; POLAJNAR, Matija; TOPLAK, Marko; STARIČ, Anže; ŠTAJDOHAR, Miha; UMEK, Lan; ŽAGAR, Lan; ŽBONTAR, Jure; ŽITNIK, Marinka; ZUPAN, Blaž. Orange: Data mining toolbox in python. **Journal of Machine Learning Research**, v. 14, p. 2349–2353, 2013. Disponível em: <<http://jmlr.org/papers/v14/demsar13a.html>>.

INEP. Edital anual do exame nacional do ensino médio, **EDITAL No 10, DE 14 DE ABRIL DE 2016**. 2016. Online; acessado 05 Junho 2017. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/edital/2016/edital_enem_2016.pdf>.

MITCHELL, Tom M. Mitchell. **Machine Learning**. [S.l.]: Book News, Inc., 1997. ISBN 0070428077.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas Inteligentes-Fundamentos e Aplicações**, v. 1, n. 1, 2003.

MOTTA, Porthos Ribeiro de Albuquerque. **Estudo Exploratório do Uso de Classificadores para a Predição de Desempenho e Abandono em Universidades**. Dissertação (Mestrado) — Universidade Federal de Goiás Instituto de Informática, 11 2016. Online; acessado 06 Junho 2017.

SILVA, Sílvia Ribeiro da; CARVALHO, Taynan Lima. Produção de texto escrito no ensino médio: Competências requeridas pela avaliação de redação do enem em (des)uso no livro didático de português. **Caminhos em linguística aplicada**, 1o sem 2017, v. 16, n. 1, p. 1–25, 2017. Disponível em: <<http://periodicos.unitau.br/ojs-2.2/index.php/caminhoslinguistica>>.

SISU, Sistema de seleção unificada. **O que é o Sisu**. 2017. Online; acessado 07 Abril 2017. Disponível em: <<http://sisu.mec.gov.br/>>.

UOL. **Banco de redações**. 2017. Online; acessado 01 Junho 2017. Disponível em: <<https://educacao.uol.com.br/bancoderedacoes/>>.

WAYMO. **We're building a safer driver for everyone**. 2017. Online; acessado 07 Abril 2017. Disponível em: <<https://waymo.com/>>.

APÊNDICE A – Título do Apêndice

APÊNDICE B – Exemplo do pacote Algorithm

Algoritmo 1 Estimador ML otimizado.

- 1: Inicializar o contador: $j \leftarrow 1$;
 - 2: Fixar o limiar de variação das estimativas: $e_{\text{out}} \leftarrow 10^{-4}$;
 - 3: Fixar o número máximo de iterações: $N \leftarrow 1000$;
 - 4: Computar o ponto inicial: $\hat{\gamma}(0)$;
 - 5: Determinar o limiar inicial: $e_1 \leftarrow 1000$;
 - 6: Estabelecer o valor inicial de α : $\hat{\alpha}(0) \leftarrow -10^{-6}$;
 - 7: **enquanto** $e_j \geq e_{\text{out}}$ e $j \leq M$ **fazer**
 - 8: Solucionar $\hat{\alpha}_j \leftarrow \arg \max_{\alpha} l_1(\alpha; \gamma_{j-1}, \mathbf{z}, n)$;
 - 9: Solucionar $\hat{\gamma}_j \leftarrow \arg \max_{\gamma} l_2(\gamma; \alpha_j, \mathbf{z}, n)$;
 - 10: $j \leftarrow j + 1$
 - 11: Computar o critério de convergência: e_j ;
 - 12: **fim enquanto**
-