

Recuperação de Padrões na Valoração Textual de Redações

Eugênio Cunha¹, Dr. Marco Túlio Alves Nolasco Rodrigues¹

¹Universidade de Itaúna (UIT)

Caixa Postal 100 – 35.680-142 – Itaúna – MG – Brasil

{genio.py, tulio.rodrigues}@gmail.com

Abstract. Nowadays, there is an intense amount of essays being produced and evaluated in entrance exams, competitions and exams. Unlike the manuals, which process and evaluate the essays manually, this work addresses an automatic way, through machine learning, capable of generalizing, learning and extracting standards from writing based on the labeled content. The method needs little human intervention and allows the valuation of large amounts of texts. This job is based on the problem of manual assessment of competences required in a essay-argumentative writing text with diverse themes social, scientific, cultural or political. Given a corpus of the main objective is to induce a model to classify the required by writing an evaluation note on the text. Embasado in the main metrics of analysis of the classifiers cited in the literature of machine learning, the solution proposed in this work was shown to be useful and useful for use in issues involving text valuation manual by trained professionals.

Resumo. Nos dias atuais, há uma quantidade intensa de redações sendo produzida e avaliada em vestibulares, concursos e exames. Diferentemente dos métodos existentes, que processam e avaliam as redações de maneira manual, este trabalho aborda uma forma automática, por meio de aprendizagem de máquina, capaz de generalizar, aprender e extrair padrões das classes de redações com base no conteúdo rotulado. O método precisa de pouca intervenção humana e permite a valoração de grandes quantidades de textos. Este trabalho fundamenta-se no problema de avaliação manual das competências exigidas em um texto de redação do tipo dissertativo-argumentativo com temas diversificados de ordem social, científica, cultural ou política. Dado um “corpus” de redações o objetivo principal é induzir um modelo a classificar as competências exigidas compondo uma nota avaliativa sobre o texto. Embasado nas principais métricas de análise dos classificadores citados na literatura de aprendizado de máquina, a solução proposta neste trabalho demonstrou ser útil e propícia a ser utilizada em problemas que envolva a valoração de texto manual por profissionais capacitados.

1. Introdução

O desenvolvimento de uma redação é uma atividade prática presente na cultura civilizada desde a invenção da escrita. Na década de 70 iniciou-se processo de redemocratização que consequentemente restituiu a palavra ao estudante. O decreto 79.298, de 24 de Fevereiro de 1977 definiu a volta da redação à escola pela “inclusão obrigatória da prova ou questão de redação em língua portuguesa” nos concursos e vestibulares (Art. 1º, alínea d) (LARA, 1994).

A redação é aplicada no ENEM desde a primeira edição 1998, hoje o maior exame do Brasil, que na edição de 2016 contou com 8.627.195 escritos confirmados, e a participação direta de 11.360 profissionais externos na correção de 5.825.134 redações. Com o advento do ENEM ser um requisito para o processo seletivo de acesso a inúmeras universidades públicas (SISU, 2017) e a importantes programas de governo (CSF, 2017), este número tem aumentado incessantemente. Segundo o edital do ENEM 2016, cada redação foi avaliada por, pelo menos, dois avaliadores, de forma independente, uma estimativa mínima de 11.650.268 avaliações manuais das competências exigidas em um texto de redação pelo ENEM (INEP, 2016a). Devido à grande quantidade de redações produzidas, torna-se humanamente difícil e caro organizar e avaliar as competências de uma redação manualmente.

Com o processamento computacional mais barato e poderoso, a crescente variedade e volume de dados disponíveis, e o armazenamento de dados de forma acessível, o Aprendizado de Máquina está no centro de muitos avanços tecnológicos, alcançando áreas antes exclusivas de seres humanos. Os carros autônomos do Google são o exemplo de uma atividade antes exclusivamente humana e hoje exercida e aperfeiçoada por algoritmos de Aprendizado de Máquina (WAYMO, 2017). O Aprendizado de Máquina está presente 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 entre outras aplicações (BATISTA et al., 1999).

A avaliação de redações automática pode ser realizada utilizando sistemas especialistas ou algoritmos de Aprendizado de Máquina. A primeira hipótese dependente essencialmente da presença de especialistas que detem o conhecimento sobre o domínio do problema para desenvolver um conjunto de regras. O sistema especialista deve ser capaz de tomar suas decisões, ou seja, as regras são disparadas para atingir determinada decisão (NEGNEVITSKY, 2005). Entretanto regras desenvolvidas manualmente tem um processo de manutenção e atualização complexo, o que torna mais difícil a sua utilização em diferentes domínios do problema proposto.

O uso de algoritmos de Aprendizado de Máquina para valoração de redações é uma alternativa ao sistema especialista, exige um menor esforço humano com a abstração simples de extrair padrões ou características, aprender e generalizar. Dados os benefícios, a hipótese deste artigo é que um algoritmo de Aprendizado de Máquina pode ser útil e propício a ser utilizada em problemas que envolva a valoração de texto manual por profissionais capacitados.

Além disso, para avaliar e validar a hipótese o método de construção do conhecimento deste trabalho terá como fundamento o problema de recuperação de padrões na valoração textual. Dado um “corpus” de redações o objetivo principal é induzir um modelo a classificar as competências exigidas compondo uma nota avaliativa sobre a redação.

O presente estudo com base na proposta do problema descrito contribuirá na área do Aprendizado de Máquina e diretamente no processo de valoração de um texto em prosa do tipo dissertativo-argumentativo.

2. Trabalhos Relacionados

Este tópico é destinado a listar uma sequência de trabalhos científicos nas áreas abordadas por este estudo com o objetivo de adquirir conhecimento para a elaboração do mesmo.

De acordo com (SILVA; CARVALHO, 2017), à prova de redação do ENEM é avaliada levando em conta uma matriz de referência listada na Tabela 1. Essa matriz, desenvolvida pelo (INEP, 2016a), 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.

Tabela 1: Competência I de V da matriz de referência elaborada pelo (INEP, 2016b).

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.
	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 e quando não caracterizem reincidência.

De acordo com (BRAGA, 2015), no texto de redação, o candidato defenderá uma opinião a respeito do tema proposto, de forma coerente e coesa, embasado em argumentos consistentes. O texto será redigido respeitando a escrita formal da Língua Portuguesa. Ao fim, o candidato elabora uma proposta de intervenção social para o problema apresentado no desenvolvimento do texto que respeite os direitos humanos.

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 ilustrada a 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.

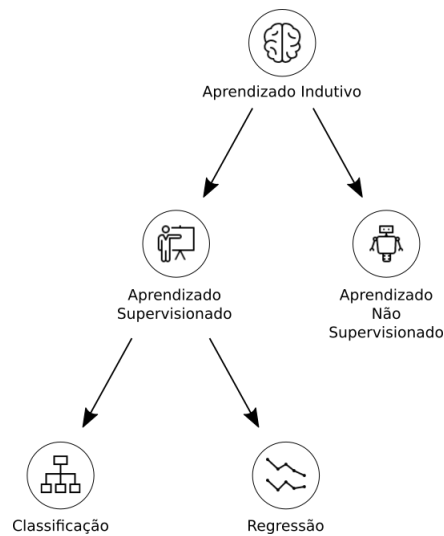


Figura 1. Árvore hierárquica do aprendizado indutivo, a qual é dividida em algoritmos supervisionado e não-supervisionado.

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.

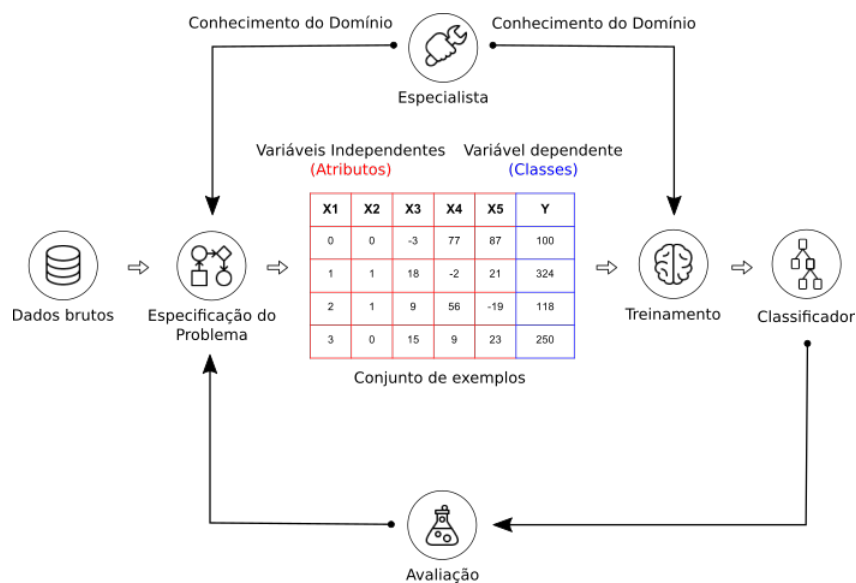


Figura 2. Fluxo do processo de classificação, o modelo encontra uma função geral capaz de prever as saídas, a especificação do problema pode ser reajustada com o conhecimento do domínio para obter um melhor resultado.

Diversas ferramentas disponíveis para exploração de dados dispõem de soluções para o processamento e a análise das informações de forma ágil e simples. Em uma análise comparativa (BOSCARIOLI; VITERBO; TEIXEIRA, 2014) demonstra que não existe uma única ferramenta com características melhores para todas as aplicações em mineração de dados. Em um estudo que comparou quatro ferramentas (KMINE, *Orange*, Tanagra, Weka), todas de código aberto, gratuitas e muito utilizadas na pesquisa e na academia, (WAHBEH et al., 2011) concluiu que a ferramenta Weka apresentou o melhor desempenho, seguido pelo *Orange*, e, depois, pelo KMINE e Tanagra. De acordo com (DEMŠAR et al., 2013), a ferramenta *Orange* na atual versão 3.5 desenvolvida pelo laboratório de Inteligência Artificial da Faculdade de Computação e Ciência da Informação da Universidade de *Ljubljana* na *Eslovênia* sob a licença GPL, possui uma interface gráfica denominada *Orange Canvas*. Por meio de sua interface ilustrada na Figura 3 é possível conectar e interligar os objetos montando um fluxo de trabalho para o desenvolvimento de modelos de classificação, incluindo *Adaboost*, *Naive Bayes*, Regras de Decisão, Árvore de Decisão, etc..

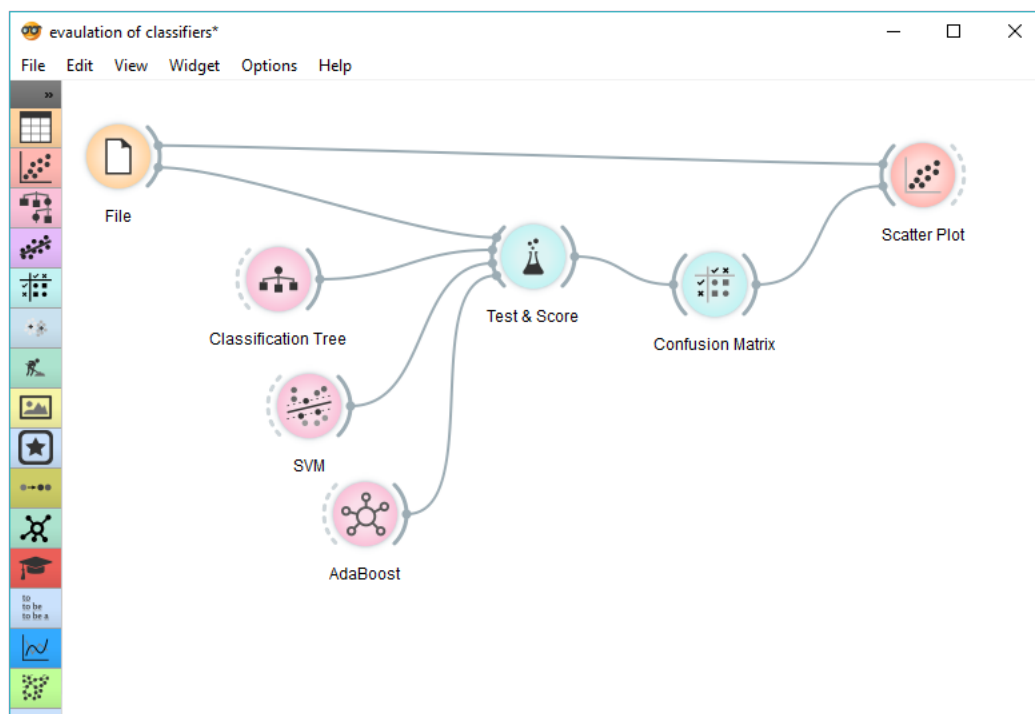


Figura 3. Ferramenta de mineração de dados *Orange Canvas* executando teste de desempenho dos classificadores *AdaBoost*, *SVM* e *Classification Tree* na matriz de confusão.

No processo de mineração de dados, segundo (MATSUBARA; MARTINS; MONARD, 2003), na etapa de pré-processamento de textos, um dos métodos geralmente adotado é a representação usando a abordagem “bag-of-words”, uma das representações estruturadas mais simples, que utiliza técnicas de redução do termo ao seu radical e remoção de termos irrelevantes. Cada documento é representado como um vetor de palavras que ocorrem no documento, especificamente uma tabela atributo-valor.

Segundo o estudo de (BRITO, 2017), o classificador *Naive Bayes* é um progenitor

probabilístico para textos, está entre um dos mais utilizados no Aprendizado de Máquina, devido a seu comportamento simplista que traz bons resultados em muitos casos. Baseado no Teorema de Bayes, criado por Thomas Bayes no século XVIII, este classificador é eleito o mais eficiente na precisão e rotulação de novas amostras (CHAKRABARTI, 2002). Objetivo da classificação Naive Bayes é encontrar a melhor classe, uma característica atraente desse classificador.

Um dos algoritmos mais famoso e utilizado no Aprendizado de Máquina é *AdaBoost* ou *Adaptive Boosting* derivado do *Boosting*, uma técnica de Aprendizado de Máquina que usa diversos classificadores fracos com a finalidade de aumentar a acurácia geral. Segundo (REIS, 2015), o seu sucesso deve-se ao merito de conseguir adaptar-se aos classificadores base. Neste algoritmo, classificadores são gerados de forma a favorecer os exemplos erroneamente classificados pelos classificadores anteriores.

3. Metodologia

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 4 ilustra a etapa que o *Web Crawler* recupera as páginas, filtra as redações avaliadas e coleta cada uma para um repositório local.



Figura 4. 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 5 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 5. Os textos são submetidos aos algoritmos de normalização e posteriormente estruturados e armazenados no padrão JSON.

Na terceira etapa ilustrada pela Figura 6 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 6. 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 7, onde o classificador previamente ajustado e treinado será submetido aos testes de Acurácia, Sensitividade, Precisão, Especificidade, Curva ROC e Matriz de Confusão. Os resultados serão representados graficamente e comparados para analisar o desempenho do classificador.

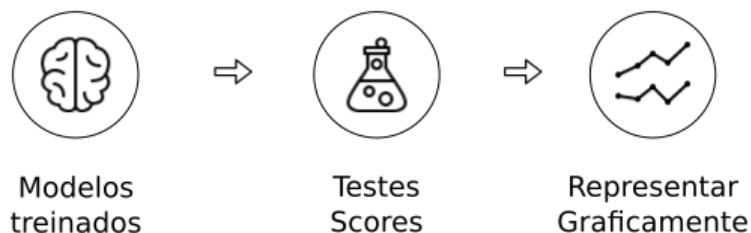


Figura 7. O classificador ajustado e treinado será submetido a testes, e os resultados comparados graficamente com o objetivo de analisar o desempenho do classificador induzido.

4. Resultados

Este capítulo é dedicado a apresentar os resultados preliminares e adversidades obtidas na indução do classificador *AdaBoost*, um algoritmo que utiliza *Boosting* como método de aprendizagem, flexível para se combinar com os vários algoritmos base de aprendizagem disponíveis.

4.1. Dados Desbalanceados

Para a indução do classificador AdaBoost sobre a primeira competência exigida em um texto de redação, isto é, “Demonstrar domínio da norma padrão da língua escrita.”, a ferramenta *Orange* selecionou de forma aleatória uma amostra de 30% do corpus de redações, ou seja, aproximadamente 131 redações de 436.

A qualidade dos dados no *dataset* é uma fator fundamental neste processo de indução, e trabalhar com dados desbalanceados tende à produzir regras de classificação que beneficiam as classes majoritárias, isto é, com maior probabilidade de ocorrência, resultando em uma baixa taxa de predição para o grupo minoritário.

Como demonstrado no Gráfico 8, a amostra de 30% das redações selecionadas no *dataset* apresentou dados desbalanceados, ou seja, 70% da amostra selecionada tende para as classes 0.50 e 1.00, os demais 30% restante, para as classes 0.00, 1.50 e 2.00.

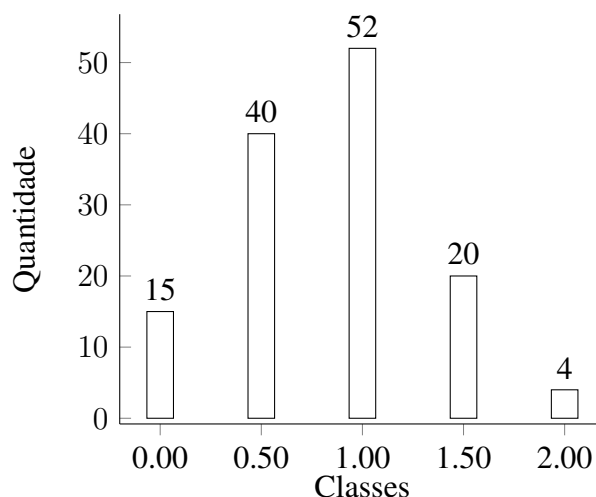


Figura 8. Distribuição das classes em uma amostra de 131 redações selecionadas aleatoriamente no *dataset*.

O desbalanceamento de dados presente na amostra era uma condição esperada, no processo de valoração, a competência é avaliada por, pelo menos, dois avaliadores, de forma independente, se ocorrer diferenças nas notas da competência inferior a 20% entre elas, é calculado o valor médio, fator que colabora efetivamente para tendência de uma ou mais classes presentes na amostra.

4.2. Métricas de Desempenho

Nos resultados preliminares, este estudo utilizou as principais métricas da literatura para análise do desempenho de classificadores, tendo como foco as métricas: Curva ROC, Acurácia, *F-Score*, *Precision* e *Recall*.

A Tabela 2 exibe os resultados das principais métricas de desempenho de classificadores sobre cada classe induzida e também a média geral de cada métrica.

Classes	Resultado da avaliação				
	ROC	Acurácia	<i>F-Score</i>	<i>Precision</i>	<i>Recall</i>
0.00	0.498	0.828	0.096	0.845	0.828
0.50	0.552	0.640	0.349	0.653	0.640
1.00	0.499	0.509	0.422	0.506	0.509
1.50	0.549	0.579	0.222	0.755	0.759
2.00	0.541	0.915	0.140	0.899	0.915
Média	0.529	0.694	0.246	0.737	0.730

Tabela 2. Resultado das métricas de desempenho do classificador AdaBoost.

A adversidade de classes desbalanceadas, pode produzir um modelo com elevadas taxas de acurácia global para determinadas classes, como o ocorrido nas classes 0.00 e 2.00 de 0.828 e 0.915 respectivamente, entretanto frequentemente tende a prejudicar a identificação de exemplos pertencentes a grupos minoritários.

Ilustrada na Figura 9, a representação gráfica da Curva ROC de cada classe induzida pelo classificador.

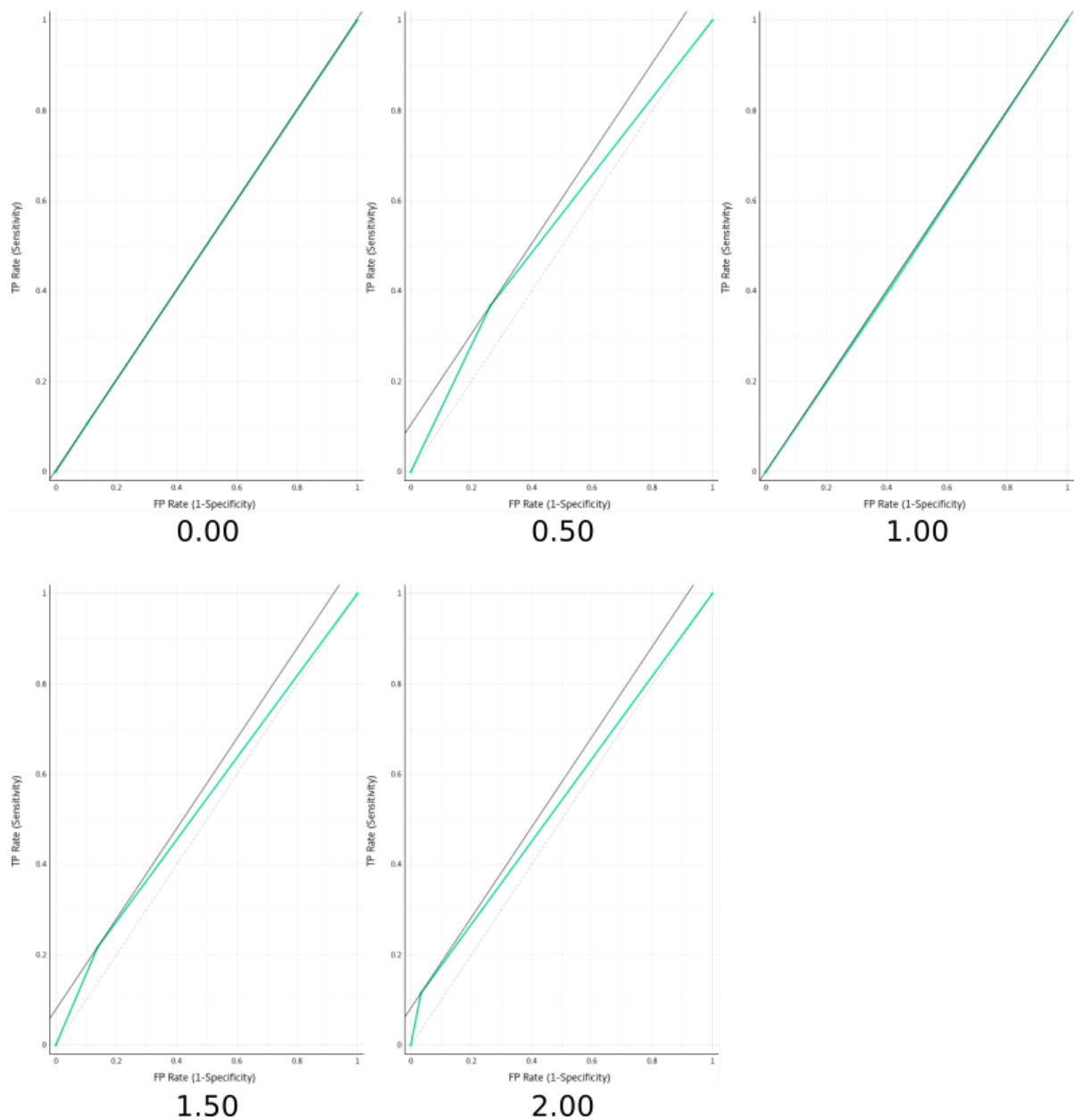


Figura 9. Representação gráfica da Curva ROC para cada classe (0.00, 0.50, 1.00, 1.50 e 2.00) induzida no modelo AdaBoost.

O comportamento esperado para a curva, é que a mesma, se aproxime o máximo possível de 1 em cada classe. Entretanto a métrica se apresenta como uma reta nas classes 0.00 e 1.00, nas demais classes tende sutilmente a 1. Conclui-se que a predição das classes 0.00 e 1.00 nos testes estão ocorrendo de forma aleatório pelo classificador.

Por fim a tabela de contingência ou matriz de confusão, que através da discriminação dos erros ou acertos preditos para cada classe demonstra o desempenho do classificador, uma das métricas mais eficiente de se analisar um classificador.

A Tabela ?? exibe ao longo da diagonal em tons de cinza as decisões corretas: número de verdadeiros positivos TP e verdadeiros negativos TN; já os elementos fora dessa

diagonal representam os erros cometidos: número de falsos positivos FP e falsos negativos FN. É notável que o valor ideal fora da diagonal seja sempre igual a 0.

4.3. Considerações Finais

É notável que adversidade de classes desbalanceadas influenciou consideravelmente nos resultados preliminares. A próxima etapa deste estudo merece destaque em uma seção exclusiva para discussão do tema e a análise dos principais métodos na literatura para balanceamento de classes.

As dificuldades observadas no estudo do problema proposto motivam melhorias e o surgimento de novas estratégias para a continuidade do trabalho.

5. Discussão

Texto texto texto ...

6. Conclusão

Texto texto texto ...

7. Trabalhos Futuros

Texto texto texto ...

8. Considerações finais

Texto texto texto ... ([UOL, 2017](#))

Referências

BATISTA, G. et al. 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. Citado na página 2.

BOSCARIOLI, C.; VITERBO, J.; TEIXEIRA, M. F. Avaliação de aspectos de usabilidade em ferramentas para mineração de dados. *Anais da I Escola Regional de Sistemas de Informação do Rio de Janeiro*, v. 1, n. 1, p. 107–114, 2014. Citado na página 5.

BRAGA, B. M. d. A. *Teoria da resposta ao item: o uso do modelo de Samejima como proposta de correção para itens discursivos*. Dissertação (Mestrado) — Universidade de Brasília Instituto de Ciências Exatas Departamento de matemática, 7 2015. Online; acessado 06 Junho 2017. Citado na página 3.

BRITO, E. M. N. D. Mineração de textos: detecção automática de sentimentos em comentários nas mídias sociais. *Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento*, v. 6, n. 1, 2017. Citado na página 5.

CHAKRABARTI, S. *Mining the Web: Discovering knowledge from hypertext data*. [S.l.]: Elsevier, 2002. Citado na página 6.

CSF, C. s. F. *Estudante de Graduação*. 2017. Online; acessado 07 Abril 2017. Disponível em: (<http://www.cienciasemfronteiras.gov.br/web/csf/estudante>). Citado na página 2.

DEMŠAR, J. et al. 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>. Citado 2 vezes nas páginas 5 e 7.

DUARTE, J. *O Algoritmo Boosting at Start e suas aplicações*. Tese (Doutorado) — PUC-Rio, 2009. Nenhuma citação no texto.

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. Citado 2 vezes nas páginas 2 e 3.

INEP. Edital anual do exame nacional do ensino médio, *Matriz De Referência Para Redação*. 2016. 89–90 p. Online; acessado 02 Setembro 2017. Disponível em: http://download.inep.gov.br/educacao_basica/enem/edital/2016/edital_enem_2016.pdf. Citado na página 3.

LARA, G. M. P. A redação como tema de pesquisa. In: *Leitura: Teoria e Prática*. [S.l.]: 1994, 1994. v. 13, n. 24, p. 62–82. Citado na página 1.

MATSUBARA, E. T.; MARTINS, C. A.; MONARD, M. C. Pretext: Uma ferramenta para pré-processamento de textos utilizando a abordagem bag-of-words. *Technical Report*, v. 209, 2003. Citado na página 5.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, v. 1, n. 1, 2003. Citado na página 3.

MOTTA, P. R. d. A. *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. Citado na página 4.

NEGNEVITSKY, M. *Artificial intelligence: a guide to intelligent systems*. [S.l.]: Pearson Education, 2005. Citado na página 2.

REIS, W. A. D. dos. Detecção de sinais de trânsito através do método de classificação adaboost. *UNOPAR Científica Ciências Exatas e Tecnológicas*, v. 12, n. 1, p. 28, 2015. Citado na página 6.

SILVA, S. R. da; CARVALHO, T. L. 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>. Citado na página 3.

SISU, S. d. s. u. *O que é o Sisú*. 2017. Online; acessado 07 Abril 2017. Disponível em: <http://sisu.mec.gov.br/>. Citado na página 2.

UOL. *Banco de redações*. 2017. Online; acessado 01 Junho 2017. Disponível em: <https://educacao.uol.com.br/bancoderedacoes/>. Citado na página 11.

WAHBEH, A. H. et al. A comparison study between data mining tools over some classification methods. *International Journal of Advanced Computer Science and Applications*, v. 8, n. 2, p. 18–26, 2011. Citado na página 5.

WAYMO. *We're building a safer driver for everyone*. 2017. Online; acessado 07 Abril 2017. Disponível em: <https://waymo.com/>. Citado na página 2.