



Ernane Rosa Martins
(Organizador)

Princípios e Aplicações da Computação no Brasil 3

Atena
Editora
Ano 2019

Ernane Rosa Martins
(Organizador)

Princípios e Aplicações da Computação no Brasil 3

Atena Editora
2019

2019 by Atena Editora
Copyright © Atena Editora
Copyright do Texto © 2019 Os Autores
Copyright da Edição © 2019 Atena Editora
Editora Chefe: Profª Drª Antonella Carvalho de Oliveira
Diagramação: Lorena Prestes
Edição de Arte: Lorena Prestes
Revisão: Os Autores



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição 4.0 Internacional (CC BY 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Conselho Editorial

Ciências Humanas e Sociais Aplicadas

Profª Drª Adriana Demite Stephani – Universidade Federal do Tocantins
Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia
Prof. Dr. Edvaldo Antunes de Faria – Universidade Estácio de Sá
Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima
Prof. Dr. Fabiano Tadeu Grazioli – Universidade Regional Integrada do Alto Uruguai e das Missões
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Ivone Goulart Lopes – Istituto Internazionele delle Figlie di Maria Ausiliatrice
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Profª Drª Keyla Christina Almeida Portela – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Marcelo Pereira da Silva – Universidade Federal do Maranhão
Profª Drª Miranilde Oliveira Neves – Instituto de Educação, Ciência e Tecnologia do Pará
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Profª Drª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa
Profª Drª Sandra Regina Gardacho Pietrobon – Universidade Estadual do Centro-Oeste
Profª Drª Sheila Marta Carregosa Rocha – Universidade do Estado da Bahia
Prof. Dr. Rui Maia Diamantino – Universidade Salvador
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Ciências Agrárias e Multidisciplinar

Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista
Profª Drª Diocléa Almeida Seabra Silva – Universidade Federal Rural da Amazônia
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Júlio César Ribeiro – Universidade Federal Rural do Rio de Janeiro
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

Ciências Biológicas e da Saúde

Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás
Prof. Dr. Edson da Silva – Universidade Federal dos Vales do Jequitinhonha e Mucuri
Profª Drª Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará
Profª Drª Magnólia de Araújo Campos – Universidade Federal de Campina Grande
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Ciências Exatas e da Terra e Engenharias

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto
Prof. Dr. Alexandre Leite dos Santos Silva – Universidade Federal do Piauí
Profª Drª Carmen Lúcia Voigt – Universidade Norte do Paraná
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande
Profª Drª Neiva Maria de Almeida – Universidade Federal da Paraíba
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)	
P957	Princípios e aplicações da computação no Brasil 3 [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa, PR: Atena Editora, 2019. – (Princípios e aplicações da computação no Brasil; v. 3) Formato: PDF Requisito de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web Inclui bibliografia ISBN 978-85-7247-702-4 DOI 10.22533/at.ed.024191510 1. Computação. 2. Informática. 3. Programação de computador. I. Martins, Ernane Rosa. II. Série. CDD 004
Elaborado por Maurício Amormino Júnior – CRB6/2422	

Atena Editora
Ponta Grossa – Paraná - Brasil
www.atenaeditora.com.br
contato@atenaeditora.com.br

APRESENTAÇÃO

O volume 3 da referida obra “Princípios e Aplicações da Computação no Brasil” apresenta 15 novos capítulos, que abordam assuntos importantes sobre o panorama atual da computação no Brasil, tais como: desenvolvimento de aplicativos móveis, VoIP, modelagem computacional, simulação, recurso educacional aberto, projeto de banco de dados, mobile learning, robótica, avaliação automática de código-fonte e agrupamento difuso multivariado.

Deste modo, esta obra reúne debates e análises acerca de questões relevantes, tais como: desenvolvimento de uma aplicação móvel para realização de recargas, consultas e simulação de saldo na carteira de transporte público, usando o método Design Constructal para melhorar a configuração geométrica de um novo layout para TCSA com quatro dutos; análise dos benefícios da tecnologia VoIP como alternativa de baixo custo a um serviço de telefonia institucional gerenciado por minicomputadores Raspberry e adaptadores ATA; análise do comportamento mecânico de placas finas de materiais compósitos laminados reforçados por fibras submetidas a esforços de flexão, comparando-as com o de uma placa de aço, usando modelagem computacional utilizando o software Abaqus 2017 Student Version®; apresentação de uma arquitetura de sistemas de simulação para área de convergência; desenvolvimento de uma abordagem para auxiliar o docente na criação de atividades com REA's que quando executados pelos alunos gerem o registro de suas experiências; análise da gestão de recursos humanos na administração pública por meio de um projeto de banco de dados; análise dos recursos disponíveis no WhatsApp Messenger, que oferecem possibilidades de uso no ambiente escolar como recurso pedagógico que pode auxiliar o processo de ensino e aprendizagem; simulação de um agente robótico capaz de realizar busca e exploração visual num ambiente virtual; e o desenvolvimento de uma ferramenta de análise e avaliação automática de código-fonte como alternativa para a otimização do processo de ensino-aprendizagem de programação.

Nesse sentido, esta obra apresenta enorme potencial para contribuir com análises e discussões aprofundadas sobre assuntos relevantes da área da computação, podendo servir de referência para novas pesquisas e estudos nesta área. Agradecemos em especial aos autores dos capítulos apresentados, e desejamos aos leitores, inúmeras reflexões proveitosas sobre as temáticas abordadas nesta obra.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 1	1
WALLET BUS - TECNOLOGIA E TRANSPORTE PÚBLICO	
Carlos Adriano Carneiro Pestana Juan Castro Ribeiro Leonardo Carlos Balbino	
DOI 10.22533/at.ed.0241915101	
CAPÍTULO 2	16
UM PROJETO CONSTRUCTAL DE TROCADORES DE CALOR SOLO-AR COMPOSTOS POR QUATRO DUTOS	
Ruth da Silva Brum Liércio André Isoldi Jairo Valões de Alencar Ramalho Luiz Alberto Oliveira Rocha Michel Kepes Rodrigues Elizaldo Domingues dos Santos	
DOI 10.22533/at.ed.0241915102	
CAPÍTULO 3	30
TELEFONIA DIGITAL DE BAIXO CUSTO COM VOIP E RASPBERRY: UM ESTUDO DE CASO NO IFPI CAMPUS ANGICAL DO PIAUÍ	
Francisco Alysson da Silva Sousa Matheus Lopes Pereira	
DOI 10.22533/at.ed.0241915103	
CAPÍTULO 4	40
SIMULAÇÃO NUMÉRICA PARA ANÁLISE DE DEFLEXÕES E TENSÕES EM PLACAS FINAS DE AÇO E DE MATERIAI COMPÓSITOS SUBMETIDAS A ESFORÇOS DE FLEXÃO	
João Vítor de Azambuja Carvalho Daniel Helbig	
DOI 10.22533/at.ed.0241915104	
CAPÍTULO 5	49
SIMULAÇÃO DE ÁREA DE CONVERGÊNCIA NUVEM-HPC	
Alessandro Kraemer	
DOI 10.22533/at.ed.0241915105	
CAPÍTULO 6	63
RECREATE: RECOMENDAÇÃO DE REA'S POR MEIO DA DESCOBERTA DO ESTILO DE APRENDIZAGEM PELA EXPERIÊNCIA DO ESTUDANTE E EXPECTATIVA DO PROFESSOR	
Marco Antonio Colombo Da Silva Marilde Terezinha Prado Santos	
DOI 10.22533/at.ed.0241915106	
CAPÍTULO 7	76
PROJETO DE BANCO DE DADOS COM FOCO NA ALOCAÇÃO DEPARTAMENTAL DE SERVIDORES PÚBLICOS E SUAS ATRIBUIÇÕES FUNCIONAIS	
José Nazareno Alves Rodrigues	
DOI 10.22533/at.ed.0241915107	

CAPÍTULO 8	87
M-LEARNING ATRAVÉS DO WHATSAPP: UMA ANÁLISE DE RECURSOS	
Nilson Alves dos Santos Mayara Kaynne Fragoso Cabral Beatriz Francisco Coelho	
DOI 10.22533/at.ed.0241915108	
CAPÍTULO 9	98
EXPLORAÇÃO E BUSCA VISUAL ROBÓTICA EM AMBIENTE SIMULADO	
Felipe Duque Belfort de Oliveira Hansenclever de França Bassani Aluizio Fausto Ribeiro Araujo	
DOI 10.22533/at.ed.0241915109	
CAPÍTULO 10	112
ESTUDO SOBRE A ESPECTROSCOPIA NIR PARA DETECÇÃO DO NÍVEL ALCOÓLICO NO SANGUE	
Luis Gustavo Longo da Silva Cláudio Luis D'Elia Machado Fabiano Sandrini Moraes Rafael Galli	
DOI 10.22533/at.ed.02419151010	
CAPÍTULO 11	125
ESTUDO COMPUTACIONAL DO COEFICIENTE DE IMPACTO VERTICAL EM LONGARINAS DE PONTES RODOVIÁRIAS BRASILEIRAS	
Vinícius Heidtmann Avila Mauro de Vasconcellos Real Márcio Wrague Moura	
DOI 10.22533/at.ed.02419151011	
CAPÍTULO 12	140
DESENVOLVIMENTO DE UMA ABORDAGEM DE CLASSIFICAÇÃO DE TRAÇOS DE PERSONALIDADE DO BIG-FIVE VIA ANÁLISE DE TEXTOS EM PORTUGUÊS DO BRASIL	
Angelo Travizan Neto Taís Borges Ferreira Márcia Aparecida Fernandes	
DOI 10.22533/at.ed.02419151012	
CAPÍTULO 13	148
CODETEACHER: UMA FERRAMENTA PARA CORREÇÃO AUTOMÁTICA DE TRABALHOS ACADÊMICOS DE PROGRAMAÇÃO EM JAVA	
Francisco Alan de Oliveira Santos Plácido Soares das Chagas Segundo Mardoqueu Sousa Telvina	
DOI 10.22533/at.ed.02419151013	
CAPÍTULO 14	158
CLASSIFICAÇÃO DE TENDÊNCIAS POLÍTICAS EM NOTÍCIAS VIA MINERAÇÃO DE TEXTO E REDES NEURAIS SEM PESO	
Rafael Dutra Cavalcanti	
DOI 10.22533/at.ed.02419151014	

CAPÍTULO 15	169
ANÁLISE DE DADOS DE EXPRESSÃO GÊNICA POR MEIO DE MÉTODOS DE AGRUPAMENTO DIFUSO MULTIVARIADO	
Bruno Almeida Pimentel	
Marcilio Carlos Pereira de Souto	
DOI 10.22533/at.ed.02419151015	
SOBRE O ORGANIZADOR	182
ÍNDICE REMISSIVO	183

CLASSIFICAÇÃO DE TENDÊNCIAS POLÍTICAS EM NOTÍCIAS VIA MINERAÇÃO DE TEXTO E REDES NEURASIS SEM PESO

Rafael Dutra Cavalcanti

Universidade Federal do Rio de Janeiro
Rio de Janeiro - RJ

RESUMO: Uma notícia tendenciosa é, às vezes, bem suave para o interlocutor, e alcança seu objetivo de influenciar a opinião do leitor no mesmo sentido. Nos dias atuais, devido à quantidade de informações existentes, muitas pessoas sentem dificuldades em avaliar a ideia principal do conteúdo de uma notícia ou se existe alguma tendência, no caso deste trabalho, política.

Neste artigo, buscamos a identificação de polaridade em notícias políticas em português através do processo de mineração de dados textuais com a utilização da Rede Neural sem Peso WiSARD e de uma derivação, a ClusWiSARD. O WiSARD funciona através de uma estrutura de discriminadores, onde cada discriminador é responsável por identificar uma classe. Realizamos avaliações relacionadas ao corpo da notícia e à manchete da notícia e realizamos uma avaliação de um veículo de mídia amplamente conhecido. Obtivemos acurácia de cerca de 90% ao utilizar o corpo da notícia completo e acurácia de cerca de 75% ao considerar apenas manchetes. Além disso, também fazemos uma análise temporal sobre a dinâmica política das tendências.

PALAVRAS-CHAVE: Mineração de Textos, Redes Neurais sem Peso, Descoberta de Conhecimento em Dados não Estruturados.

CLASSIFICATION OF POLITICAL BIAS IN NEWS THROUGH TEXT MINING AND WEIGHTLESS NEURAL NETWORKS

ABSTRACT: Biased news can influence the reader's opinion in subtle ways. Nowadays, due to the unprecedented amount of information created and made available through social media, the identification of biases is increasingly challenging. In the domain of politics, addressing the challenge is particularly relevant.

In this paper, we seek the identification of polarity in Portuguese political news through the process of textual data mining using the WiSARD Weightless Neural Network, and one of its extensions, the ClusWiSARD. The WiSARD classifier works through a structure of discriminators, where each discriminator is responsible for identifying a class. We assessed polarity using the body and the headline of news published in widely known media vehicles. The obtained results are encouraging, indicating the feasibility of automatic and efficient bias detection. We obtained accuracy of about 90% when using full body news and accuracy of the 75% when considering only headlines. In addition, we also perform a temporal analysis

on the political dynamics of bias.

KEYWORDS: Text Mining, Weightless Neural Networks, Discovery of Knowledge in Non-Structured Data.

1 | INTRODUÇÃO

A mineração de textos é um campo relacionado com diversas disciplinas como por exemplo aprendizado de máquina, recuperação da informação, processamento de linguagem natural e estatística. Com a mineração de texto, é possível obter informações e chegar a conclusões sobre um conjunto de documentos, através da análise dos dados.

A rede neural sem peso WiSARD é um classificador utilizado para o reconhecimento de padrões e utiliza valores binários em suas entradas. A WiSARD pode realizar a classificação de acordo com os padrões apresentados previamente durante a fase de treinamento. Buscamos classificar automaticamente as notícias políticas recentes do Brasil com base em sua polaridade política.

As fontes de dados selecionadas foram os feeds de notícias dos sítios dos partidos: Partido do Movimento Democrático Brasileiro (PMDB), Partido da Social Democracia Brasileira (PSDB), Partido dos Trabalhadores (PT) e o Partido Socialismo e Liberdade (PSOL).

Levamos em consideração neste trabalho que as fontes de notícias utilizadas gerem feeds de notícias com sua respectiva opinião implícita. Nossos resultados de classificação explicam o estilo de escrita dos conjuntos de autores e dos grupos por posição ideológica, juntamente com as diferenças no vocabulário que normalmente usam. Portanto, enquadrámos o problema de identificação de polarização como um problema de reconhecimento de origem e polaridade do texto.

Abordamos as vantagens e desvantagens do classificador WiSARD [3] em relação à precisão e eficiência/desempenho, comparada aos classificadores SVM [20], Regressão Logística [10], Naive Bayes [12] e Gradient Tree Boosting [13]. A terceira é a utilização da ClusWiSARD [5] na avaliação de um noticiário, por meio de aprendizado semi-supervisionado.

Identificar a viabilidade de classificar as fontes e as polaridades apenas com base no conteúdo do texto, com acurácia de cerca de 90%, é a nossa maior contribuição.

1.1 Considerações Sobre as Posições Partidárias

Identificar a posição ideológica de um partido político é ainda não possui um consenso e proporciona muitos debates por cientistas políticos, sociólogos e pesquisadores [21]. O trabalho de [18] apresenta essa dificuldade ao relacionar coligações partidárias e fatos noticiados. Para exemplificar uma mudança de

posicionamento partidário, podemos citar um fato que ocorreu na Dinamarca. Um partido denominado Esquerda Radical (Radikale Venstre) depois de variações de posicionamentos durante anos, hoje assume uma postura como partido de centro [19, 25].

Consideramos que o PMDB e o PSDB possuem afinidades, de acordo com [4, 7, 26] e consideramos que os partidos PT e PSOL possuem uma afinidade, conforme pode ser observado em [23, 22, 17]. Para fins de classificação por polaridade política, consideramos que as notícias podem pertencer a uma de duas classes, PMDB/PSDB e PT/PSOL.

2 | REDE NEURAL SEM PESO WISARD E A CLUSWISARD

A WiSARD é uma rede neural sem peso composta por um grupo de discriminadores. Cada discriminador é responsável pelo reconhecimento de uma única classe. Ela utiliza vetores binários como entrada e o armazenamento da informação é realizado pelos neurônios RAM, que não são capazes de generalizar de forma isolada, porém, conseguem generalizar quando compõe o discriminador [14].

A ClusWiSARD é uma variação da WiSARD, que busca melhorar o desempenho da WiSARD ao utilizar discriminadores como clusters, com a finalidade prevenir a classificação incorreta da WiSARD [5].

3 | METODOLOGIA

3.1 Coleta de Documentos

Foi implementado um crawler para obter textos das páginas dos respectivos partidos políticos. As notícias foram separadas por partido e em arquivo no formato TXT. Foram coletadas 1147 notícias políticas, dos anos de 2016 e 2017.

3.2 Pré-Processamento

Utilizamos o modelo *bag-of-words* para a representação dos documentos. Inicialmente, são removidos os caracteres especiais, dígitos, acentos e pontuações, pois não serão úteis ao processamento. Em seguida, todos os caracteres são convertidos para caixa baixa. Foi utilizado um dicionário para a remoção das stopwords. Consideramos a palavra 'não' como uma stopword e, em seguida, realizamos o *stemming*, onde as palavras são reduzidas ao seu radical.

3.3 Transformação dos Dados

Considerando n documentos indexados e m termos do conjunto, podemos

representar a matriz termo-documento A de ordem $m \times n$. Foi escolhido o modelo TF-IDF para representar a relevância de cada termo.

Foi escolhida Decomposição em Valores Singulares (SVD) para a redução da dimensionalidade. Baseado na escolha dos maiores valores singulares, podemos reduzir o “ruído” da matriz original, isto é descartar informações insignificantes.

Usualmente formaliza-se o espaço conceito como $A_k = U_k \Sigma_k V_k^T$ onde A_k é a representação da matriz termo-documento reduzida, U_k representa a relação entre os termos e os conceitos, V_k representa a relação entre os documentos e os conceitos e Σ_k representa a intensidade de cada conceito. Para a escolha do valor de k que melhor representa a aproximação de A , foi utilizada uma forma empírica para encontrar o valor. Na Figura 3.1, podemos observar e concluir sobre os valores de k que melhor representam a aproximação em relação à matriz original. Cada documento foi representado por $d^{\wedge} = d^T U_k$ onde d^T representa o vetor do documento e d^{\wedge} é a representação reduzida.

Por fim, cada documento foi representado por um vetor binário, similar a um ‘termômetro’, de acordo a relevância de cada termo.

3.4 Análise

Os padrões descobertos podem ser avaliados pelo usuário final, especialista do domínio e analista de dados, com o intuito de validar o conhecimento obtido [8].

Foi utilizado o método K-fold com 90% da massa de dados como treinamento e 10% para testes para avaliação.

4 | EXPERIMENTOS E RESULTADOS

4.1 Objetivo dos Experimentos

Foi avaliada a notícia completa e as manchetes das notícias, devido a muitas pessoas compartilharem reportagens em redes sociais sem ler seu conteúdo [9].

Como apoio à implementação, as principais bibliotecas para Python foram utilizadas: Scikit-Learn [15], PyWANN [16], Numpy [24] e Matplotlib [11].

4.3 Comparação Entre Classificadores

Este experimento foi demonstrado em [6] e realiza uma avaliação através do processo utilizando o corpo da notícia como base de dados. Para a redução de dimensionalidade, estratégia adotada foi aplicar a técnica de atribuição de pesos TF-IDF aos termos. Um limite estabelecido empiricamente Θ foi utilizado para selecionar somente os termos mais significativos. A tabela 4.3 retrata uma comparação entre o desempenho de variações das arquiteturas do WiSARD e alguns classificadores.

Modelo	Unigramas		Bigramas		Trigramas	
	Ac	DP	Ac	DP	Ac	DP
Regressão Logística	88%	4%	95%	2%	94%	3%
WiSARD 4 bits com <i>bleaching</i>	88%	5%	93%	3%	95%	2%
WiSARD 8 bits com <i>bleaching</i>	88%	5%	93%	3%	94%	3%
SVM (kernel linear)	87%	4%	93%	4%	92%	4%
WiSARD 16 bits com <i>bleaching</i>	89%	6%	93%	6%	93%	4%
WiSARD 16 bits	89%	4%	92%	5%	93%	3%
WiSARD 8 bits	90%	6%	91%	3%	94%	3%
WiSARD 4 bits	91%	4%	91%	4%	94%	3%
GB (estimators=150, learning rate=0.01, depth=5)	83%	8%	86%	4%	85%	5%
Bernoulli Naive Bayes	80%	5%	85%	5%	86%	5%
WiSARD 32 bits com <i>bleaching</i>	83%	6%	84%	4%	93%	3%
WiSARD 32 bits	82%	6%	82%	6%	93%	4%

Tabela 4.1: Comparativo entre classificadores [6].

4.4 Classificação e Identificação de Polaridade

4.4.1 Classificação Baseada no Corpo da Notícia

O corpo da notícia é onde existe todo conteúdo da informação, através dos termos relevantes, podemos identificar a essência do mensagem.

Após a binarização das dimensões reduzidas, foi atribuído um rótulo para cada notícia de acordo com sua fonte partidária. Para avaliação de polaridade, foram atribuídos os rótulos da acordo com o grupo de sua fonte partidária.

Observamos que nas Figuras 4.1 e 4.2 que a acurácia decai ao se reduzir o tamanho dos bits de endereçamento de memória. Este fato explica o aumento do potencial de generalização do modelo relacionado ao tamanho da memória.

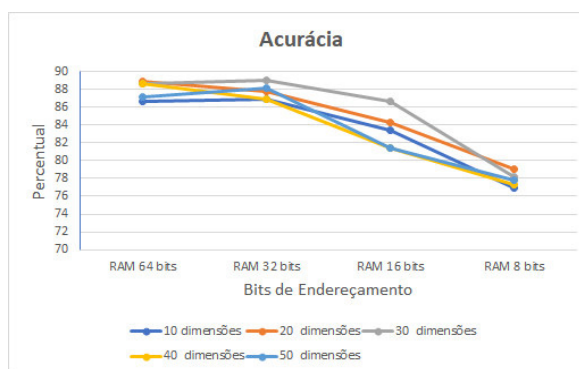


Figura 4.1: Acurácia relativa à classificação por polaridade baseada no corpo da notícia

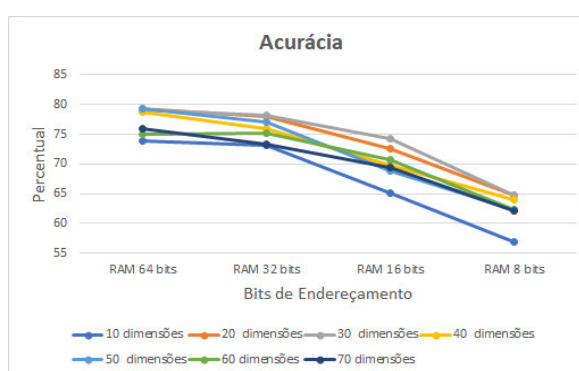


Figura 4.2: Acurácia relativa à classificação por partido baseada no corpo da notícia

4.4.2 Classificação Baseada na Manchete

A manchete da notícia é a parte de destaque da mensagem, geralmente são utilizados termos de destaque ou até mesmo apelativos.

Foram realizadas avaliações utilizando a redução de dimensionalidade através da SVD. Foram rotuladas as notícias de acordo com suas fontes, através de suas siglas.

Observamos que nas Figuras 4.3 e 4.4, que a acurácia decai ao se reduzir a quantidade dos bits de endereçamento de memória. Este fato explica o aumento do potencial de generalização do modelo relacionado ao tamanho da memória.

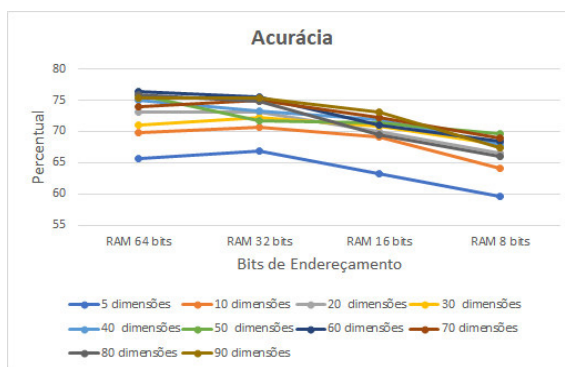


Figura 4.3: Acurácia relativa à classificação por polaridade baseada na manchete da notícia

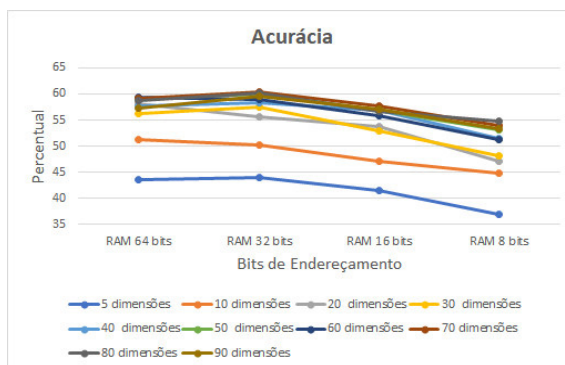


Figura 4.4: Acurácia relativa à classificação por partido baseada na manchete da notícia

4.5 Avaliação da Evolução Temporal da Polaridade

O objetivo deste experimento é avaliar, dentro do conjunto de notícias coletados de um noticiário amplamente conhecido, a evolução temporal das notícias de acordo com o resultado do classificador. Foram consideradas que as notícias possuem um viés político em sua elaboração. Como base de dados do noticiário a ser avaliado, foram coletadas notícias dos anos de 2016 e 2017 da página do Jornal Nacional (<http://g1.globo.com/jornalnacional/>). Seleccionamos notícias sobre alguns personagens políticos deste período. Para seleccionar os personagens, filtramos as notícias de acordo com as palavras-chave ‘Lula’, ‘Dilma’, ‘Temer’ e ‘Moro’, utilizadas na manchete.

Foram utilizados os corpos das notícias coletadas. Para a classificação, foi utilizado o algoritmo ClusWiSARD [5].

Os dados coletados não receberam nenhum rótulo e foram apresentados em ordem temporal. Como parâmetros da ClusWiSARD, utilizamos clusters de tamanho 40, similaridade mínima de 50% e tamanho de endereçamento de memória de 32 bits.

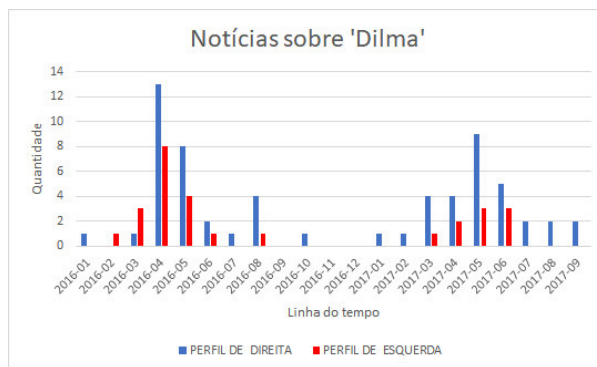


Figura 4.5: Distribuição temporal de notícias com a palavra-chave 'Dilma' por polaridade

Após a classificação das notícias, é possível extrair algumas conclusões de acordo com fatos políticos ocorridos nos anos de 2016 e 2017. Na Figura 4.5, podemos observar o grande volume de notícias entre abril e agosto de 2016, período em que houve a votação da Câmara dos Deputados sobre o impeachment da ex-presidente Dilma. O segundo período de grande volume de notícias é relacionado ao julgamento da chapa Dilma-Temer e delações que envolveram a ex-presidente.

Na Figura 4.6, que é referente às notícias relativas ao termo 'Temer', podemos observar um aumento na quantidade de notícias em maio de 2017, mês em que houve, no dia 18, a divulgação um áudio pela imprensa. Nesse período também estava ocorrendo o julgamento da chapa Dilma-Temer. Este alto volume de notícias se deu até o arquivamento do processo contra o presidente Michel Temer por corrupção passiva pela Câmara dos Deputados. Nos meses de julho e agosto, é possível observar uma transição no perfil das notícias referentes ao termo.

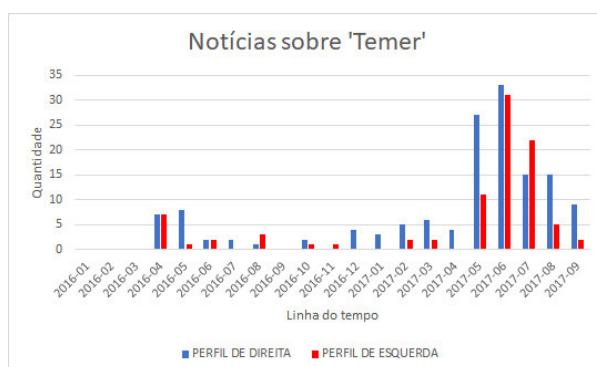


Figura 4.6: Distribuição temporal de notícias com a palavra-chave 'Temer' por polaridade

Com relação às notícias referentes ao ex-presidente Lula, observamos que o volume de notícias um pouco maior antes da votação pela Câmara dos Deputados pelo impeachment da ex-presidente Dilma. Este volume é relacionado à nomeação do ex-presidente Lula como ministro da Casa Civil e o próximo grande volume de notícias coincide com o volume da Figura 4.8 e é referente aos depoimentos feitos pelo ex-presidente Lula ao Juiz Sérgio Moro e notícias relacionadas às delações.

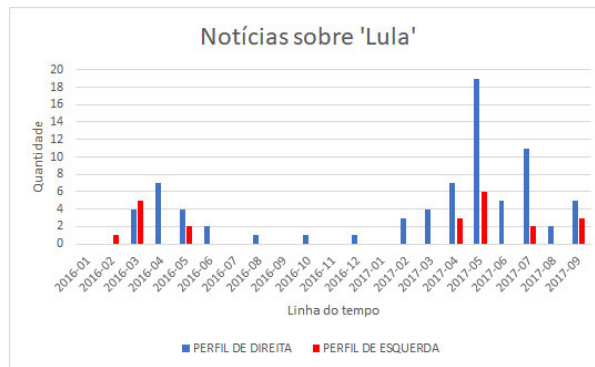


Figura 4.7: Distribuição temporal de notícias com a palavra-chave 'Lula' por polaridade



Figura 4.8: Distribuição temporal de notícias com a palavra-chave 'Moro' por polaridade

5 | CONCLUSÃO

A capacidade de aprendizado on-line do WiSARD e da ClusWiSARD, permitindo a inclusão de novos elementos em seu conjunto de treinamento sem a necessidade de se processar todos os dados, não é característica usual dos métodos de classificação e aglomeração. Isso as torna boas opções para utilização em identificação de polaridade e em classificação de notícias.

Os resultados demonstraram que a utilização do corpo da notícia no processo de mineração de textos obtiveram melhores resultados em acurácia. A utilização das manchetes das notícias se mostrou como mais uma alternativa para identificar a polaridade em notícias.

5.1 Trabalhos Futuros

Durante o desenvolvimento, observaram-se algumas possibilidades de melhorias na metodologia adotada. Como sugestão, podem ser abordadas as fake news, notícias imparciais e a exploração da avaliação temporal.

REFERÊNCIAS

- [1] ALEKSANDER, I. **From WISARD to MAGNUS: a family of weightless virtual neural machines. RAM-Based Neural Networks**, Canterbury, v.9, 1998.

- [2] ALEKSANDER, I.; DE GREGORIO, M.; FRANÇA, F. M. G.; LIMA, P. M. V.; MORTON, H. **A brief introduction to Weightless Neural Systems**. In: ESANN, Bruges, Bélgica. Anais. . . UCL/ELEN, 2009. p.299-305.
- [3] ALEKSANDER, I.; THOMAS, W.; BOWDEN, P. **WISARD: a radical step forward in image recognition**. Sensor review, Bingley, Reino Unido, v.4, n.3, p.120-124, 1984.
- [4] BRASIL, B. **De criação na ditadura até o rompimento com governo: o PMDB em 10 capítulos**. Acesso em: 2017-10-05, Disponível em:<<http://www.bbc.com/portuguese/noticias/2016/03/160330-pmdb-historia-ms-ss>>.
- [5] CARDOSO, D. d. O.; CARVALHO, D. S.; ALVES, D. S.; SOUZA, D. F. P. de; CARNEIRO, H. C.; PEDREIRA, C. E.; LIMA, P. M.; FRANÇA, F. M. **Credit analysis with a clustering RAM-based neural classifier**. In: ESANN, Bruges, Bélgica. Anais. . . UCL/ELEN, 2014.
- [6] CAVALCANTI, R. D.; LIMA, P. M. V.; GREGORIO, M. D.; MENASCHE, D. S. **Evaluating weightless neural networks for bias identification on news**. In: IEEE 14TH INTERNATIONAL CONFERENCE ON NETWORKING, SENSING AND CONTROL (ICNSC), 2017., Calabria, Italy. Anais. . . IEEE, 2017. p.257-262.
- [7] DIAP, A. **Eleições 2014: direita política se populariza no brasil**. Acesso em: 2017-10-05, Disponível em:<<https://tinyurl.com/y9q772hb>>.
- [8] EBECKEN, N. F.; LOPES, M. C. S.; COSTA, M. C. et al. **Mineração de textos. Sistemas inteligentes: fundamentos e aplicações**, São Carlos, p.337-370, 2003.
- [9] GRANJA, B. **Geração só Manchete**. Acesso em: 2017-09-20, Disponível em:<<https://tinyurl.com/yd4wwdwh>>.
- [10] HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. New York: John Wiley & Sons, 2013. v.398.
- [11] HUNTER, J. D. **Matplotlib: a 2d graphics environment**. Computing In Science & Engineering, IEEE Computing in Science Engineering, v.9, n.3, p.90- 95, 2007.
- [12] LEWIS, D. D.; RINGUETTE, M. **A comparison of two learning algorithms for text categorization**. In: THIRD ANNUAL SYMPOSIUM ON DOCUMENT ANALYSIS AND INFORMATION RETRIEVAL, Las Vegas, Nevada. Anais. . . University of Nevada, 1994. v.33, p.81-93.
- [13] MAYRINK, V. T. d. M. **Avaliação do algoritmo Gradient Boosting em aplicações de previsão de carga elétrica a curto prazo**. 2015. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Juiz de Fora.
- [14] PÁDUA BRAGA, A. de; LEON FERREIRA, A. C. P. de; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro, Brasil: LTC Editora, 2014.
- [15] PEDREGOSA, F. et al. **Scikit-learn: machine learning in python**. Journal of Machine Learning Research, USA, v.12, n.Oct, p.2825-2830, 2011.
- [16] RANGEL, F.; FIRMINO, F. **Python Weightless Artificial Neural Network**. Acesso em: 2016-01-20, Disponível em:<<https://github.com/firmino/PyWANN>>.
- [17] ROCHA, V. da. **Saiba como surgiu e o que defende cada partido**. Acesso em: 2017-10-05, Disponível em:<<https://tinyurl.com/yaa3wy2l>>.
- [18] SAMUELS, D.; ZUCCO, C. **The power of partisanship in Brazil: evidence from survey**

experiments. American Journal of Political Science, Michigan, v.58, n.1, p.212-225, 2014.

[19] STAFF, R. **Danish parties agree on tougher border controls.** Acesso em: 2017-10-05, Disponível em:<<https://tinyurl.com/y6wh329a>>.

[20] SUYKENS, J. A.; VANDEWALLE, J. **Least squares support vector machine classifiers.** Neural processing letters, Kluwer Academic Publishers, v.9, n.3, p.293-300, 1999.

[21] TAROUÇO, G. d. S.; MADEIRA, R. M. **Partidos, programas e o debate sobre esquerda e direita no Brasil.** Revista de Sociologia e Política, Curitiba, v.21, p.149 - 165, 03 2013.

[22] THIAGO GUIMARÃES, R. S. e. **Como as eleições municipais desidratam os partidos de esquerda.** Acesso em: 2017-10-05, Disponível em:<<http://www.bbc.com/portuguese/brasil-37710397>>.

[23] VASCONCELLOS, F. **Maioria dos partidos se posiciona como de centro. Veja quem sobra no campo da direita e da esquerda.** Acesso em: 2017-10-05, Disponível em:<<https://tinyurl.com/y72rvf2r>>.

[24] WALT, S. v. d.; COLBERT, S. C.; VAROQUAUX, G. **The NumPy array: a structure for efficient numerical computation.** Computing in Science & Engineering, arXiv:1102.1523v1, v.13, n.2, p.22-30, 2011.

[25] WIKIPEDIA. **Radikale Venstre.** Acesso em: 2017-10-05, Disponível em:<https://da.wikipedia.org/wiki/Radikale_Venstre>.

[26] WINTER, B. **Exclusive: brazil opposition leader will seek economic reforms.** Acesso em: 2017-10-05, Disponível em:<<https://tinyurl.com/y9w9n7gf>>.

ÍNDICE REMISSIVO

A

Abaqus 40, 41, 43, 46, 48
Agrupamento 81, 140, 169, 170, 175, 176
Aplicação 5, 6, 28, 31, 33, 34, 37, 49, 50, 51, 52, 53, 58, 59, 60, 61, 68, 70, 80, 83, 89, 129, 135, 148, 150, 152, 157, 169, 174
Aplicativo móvel 1, 7, 88
Aprendizado online 98, 100
Aprendizagem 63, 64, 65, 66, 67, 69, 70, 71, 72, 73, 74, 75, 87, 88, 89, 90, 92, 93, 94, 95, 96, 97, 100, 101, 140, 147, 148, 149, 150, 154, 156
Aprendizagem colaborativa 89, 95, 140
Área de convergência 49, 50, 51, 52, 58, 60, 61, 62
Atenção visual 98, 101, 103, 106, 107, 110
Avaliação automática 148, 149

B

Banco de dados 8, 59, 66, 76, 77, 79, 81, 82, 84, 85, 86, 170, 182
Busca visual 98, 99, 100, 101, 102, 107, 110

C

Carteira estudantil 1
Celular 87, 88, 90, 91, 96, 179
Centrais de dados 49, 50, 54, 62
Coeficiente de impacto 125, 130, 138
Condicionamento de Ar 16
Custos 30, 31, 32, 36, 37

D

Design constructal 16, 18, 19, 28
Desvio funcional 76
Difuso 169

E

Efeitos dinâmicos 125, 126, 127, 130, 136, 137
Energias renováveis 16
Ensino 2, 34, 38, 63, 64, 74, 87, 88, 90, 93, 94, 95, 96, 97, 148, 149, 156
Espectroscopia 112, 114, 115, 116, 118, 121, 122
Etilômetro 112, 113, 114, 116
Exercícios de programação 148, 156
Expressão gênica 169, 171, 174

G

Gestão pública 76, 85

I

Infravermelho 112, 114, 115, 116, 117, 118, 121

J

Java 53, 148, 149, 150, 151, 156

L

LIWC 140, 143, 144, 146

M

Materiais compósitos 40, 41, 42, 43, 44, 46, 47

Mineração de textos 158, 166, 167

Mobile learning 87, 88, 89, 91, 93, 95, 96, 97

Modelagem computacional 16, 40, 42, 125, 138

Modelo de estilo de aprendizagem 63, 65, 69

Modelos de cargas móveis 125

Multivariado 169, 172, 176, 178

P

Pontes rodoviárias 125, 126, 138, 139

Processamento de linguagem natural 140, 159

Processo ensino-aprendizagem 63

R

Recomendação 63, 64, 67, 70, 71, 72, 73

Recurso educacional aberto 63, 64

Recursos humanos 76, 79, 86

Redes 7, 30, 31, 32, 33, 35, 37, 38, 39, 81, 97, 158, 161, 167

Redes neurais sem peso 158

Robótica 98, 99, 100, 101, 110

S

Simulação 5, 13, 14, 40, 41, 43, 49, 50, 52, 54, 57, 58, 60, 61, 62, 68, 102

Simulação numérica 40, 41, 43

Sistemas de informação 76, 77, 79, 80, 86, 182

T

Tecnologia 1, 30, 32, 33, 34, 38, 39, 63, 90, 97, 112, 176, 182

Telefonia 30, 31, 32, 33, 34, 35, 37, 38, 39

Transporte público 1, 2, 3, 5, 6, 7, 15

Trocadores de Calor Solo-Ar (TCSA) 16, 17

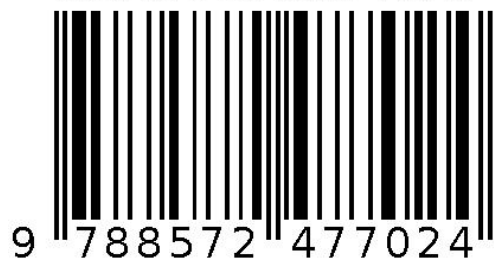
V

VoIP 30, 32, 33, 34, 38, 39

W

Whatsapp 87, 91, 96, 97

Agência Brasileira do ISBN
ISBN 978-85-7247-702-4



9 788572 477024