Opinião no Twitter Sobre o Atual Presidente da República: Uma Aplicação de Análise de Sentimento

Denilson Grupp Fernandes, Felipe Ciolacov, Felipe Julio da Costa, Ricardo Araujo dos Santos, Ricardo Barboza

Orientador: Prof. Me Fernando Sequeira Sousa
Faculdade Impacta de Tecnologia
São Paulo, SP, Brasil
29 de maio de 2019

Resumo. Este trabalho objetiva realizar a análise de sentimento de tweets através de técnicas de aprendizado de máquina sobre o candidato eleito à Presidência da República em 2018. As mensagens de Twitter foram capturadas entre o primeiro turno das eleições para a presidência da república de 2018 até o segundo mês de mandato em 2019. Os modelos preditivos aplicados mostraram que 62,91% das publicações foram negativas em relação ao candidato eleito.
Palavras-chaves: análise de sentimento, tweets, eleições, presidência, *machine learning*.

Abstract. This project aims to perform an analysis of tweets using machine learning techniques on the candidate for the Presidency of the Republic in 2018. Tweets were captured between the first round of the elections for the presidency of the republic from 2018 until the February 2019. The predictive models applied showed 62.91% of the publications were negative in relation to the elected candidate.
Keywords: sentiment analysis, tweets, elections, presidency, machine learning.

1. Introdução

Com o surgimento da internet e diversas redes sociais a comunicação entre as pessoas mudou radicalmente. Segundo a Academia Americana de Pediatria, grande parte do desenvolvimento social, emocional e cultural da chamada geração do milênio é proveniente da internet (Liu B. 2012).

As informações que são publicadas na web através do Twitter e Facebook estão armazenadas em textos que podem ser visualizadas por qualquer pessoa conectada à rede. Parte desses textos contém argumentos de opiniões e expressões de sentimentos de um determinando assunto. As opiniões são importantes, pois são fontes de conteúdo para organizações e dão a possibilidade de indivíduos se afirmarem na sociedade (Liu B. 2012).

Porém, a forma mais contundente de afirmação e expressão de opinião de um indivíduo é o evento democrático da eleição, na qual todos os cidadãos escolhem os seus representantes políticos. Sendo esses, pessoas que irão em um futuro próximo, determinar quais os próximos passos que o país irá tomar tanto economicamente quanto ideologicamente. Dentro dessa máxima, a figura do presidente da república é a representação magna desse movimento democrático (Nicolau J 2012).

1.1. Problema

As eleições de 2018 para Presidência foram definitivamente o ponto principal de mudanças e divisões de opiniões sobre o perfil do eleitorado, os fatores determinantes que mais foram pautados estão em: Renovação, alternância de poder, combate a corrução, impunidade e o melhor para o Brasil. As estatísticas da Justiça Eleitoral mostram que houve um aumento de 3,14% de eleitores em relação a 2014, no exterior revela um aumento de 41,37% resultando na acessibilidade em poder votar no seu país de origem. (Tribunal Superior Eleitoral 2018). Diante desse cenário desafiador para qualquer candidato que precise de o máximo de votos possíveis, é imprescindível compreender a opinião do eleitorado de forma rápida e assertiva (El Pais 2018).

1.2. Objetivos

O objetivo desse trabalho é avaliar as opiniões dos eleitores através de tweets relacionados ao candidato eleito à presidência da República do Brasil nas eleições de 2018, por meio de ferramentas de mineração de dados avaliando seu desempenho no geral, observando se houve melhora ou piora na classificação dos sentimentos durante as eleições e os dois primeiros meses do mandato.

1.3. Justificativa

Diante de um histórico saturado de crises de corrupção envolvendo diversos partidos políticos no Brasil, o cenário atual político se mostra fragmentando por um sistema que favoreceu o desvio de muitos recursos em instituições brasileiras para outros fins ou bens próprios desde a ditadura militar. A população encontrou nas redes sociais uma maneira de se manifestar, quase que diariamente suas opiniões e sentimentos sobre a realidade política e organizacional do país e cada vez mais os candidatos necessitam saber a opinião da população em relação a sua campanha eleitoral (BBC Brasil 2016).

2. Fundamentação teórica

O desenvolvimento do projeto foi fundamentado principalmente em 4 pilares:

- Análise de sentimento
- Mineração de dados
- Processamento de Linguagem Natural
- Machine Learning

2.1. Análise de Sentimento

Quando uma decisão é tomada em muitos casos envolvem argumentos e opiniões já estabelecidas. Esse conceito está sobre estudo da análise de sentimento.

A análise de sentimento recebeu grande atenção desde o início de 2000, as opiniões são centrais em quase todas atividades humanas e determinantes para expressar o posicionamento sobre um determinado assunto, são influenciadoras referente a nossa crença, escolha, avaliação

e comportamento. Com o aumento e propagação de dados opinativos via web através de vlogs, blogs, fórum de notícias e rede sociais o estudo e análise de sentimento cresceram, a quantidade de informações geradas vem aumentando de forma expressiva. Por isso a análise manual de sentimentos presente nesses dados é inviável (Liu B. 2012).

2.2. Mineração de Dados

O termo Mineração de Dados não representa a complexidade desse processo. Mesmo a referência original que seria o processo de mineração de ouro deveria ser contemplada de maneira diferente já que a mineração é feita nas rochas e em outros minerais para encontrar as pepitas. O termo mais correto seria "Mineração de conhecimento dos dados". É o processo de conhecimento através de padrões interessantes extraídos em grandes quantidades de dados incluindo diversas fontes como banco de dados, arquivos, internet, etc. (Jiawei H. 2006).

2.3. Processamento de Linguagem Natural (PLN)

Um dos maiores paradigmas da tecnologia atualmente é diminuir a barreira e aumentar de forma contundente a interação entre homem e máquina através da comunicação. Diversos meios foram criados para que o homem possa se comunicar com a máquina através da fala e da escrita, mas a grande dificuldade é fazer com que a máquina possa responder e interagir como se fosse outro ser humano, sendo esse o objetivo do campo de pesquisa da programação de linguagem natural (Muller 2003).

2.4. Machine Learning

Diferentemente do desenvolvimento de algoritmos e a programação propriamente dita para a execução de tarefas definidas e engessadas, o conceito de *machine learning* ou "Aprendizado de Máquina" consiste em que os computadores, amparados por um conjunto de procedimento e regras, criem a capacidade de tomar decisões e agir dentro de tarefas específicas através dos dados. É muito usado em aplicações onde não se é possível utilizar algoritmos explícitos, como exemplo, aplicações de reconhecimento de fala, motores de busca, processamento de linguagem natural, diagnósticos médicos, entre outros (Supawan L 2014).

Novos dados são utilizados como fonte para modelos de *machine learning* no intuito de aprender e juntamente com seus erros, aprender e melhorar.

Atuando como uma vertente da Inteligência Artificial, que seria uma ciência capaz de reproduzir artificialmente as capacidades humanas, o *machine learning* utiliza e compartilha, em teoria e metodologia, da estatística e da matemática para proporcionar o aprendizado (Mitchell T. 1997).

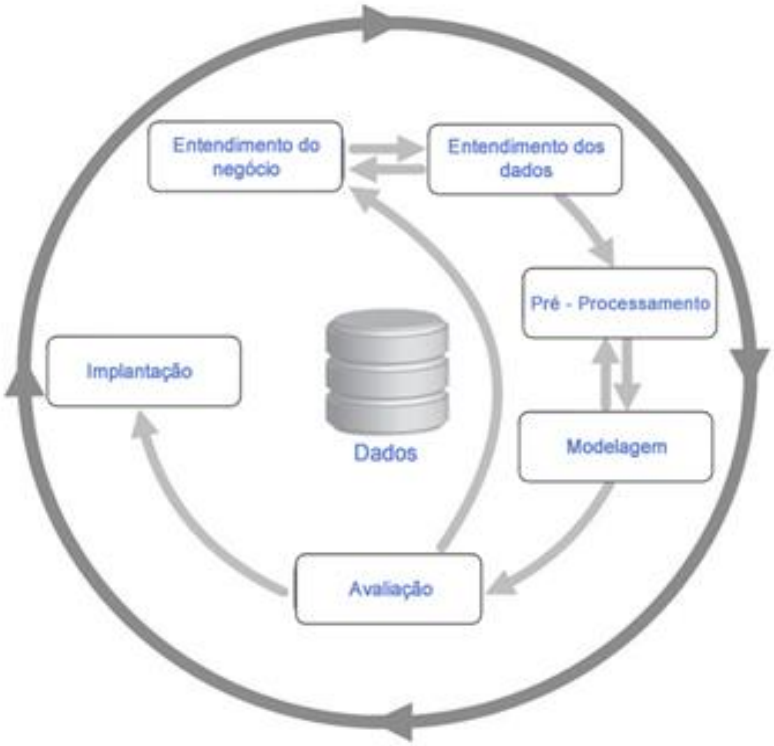
Na área de análise de dados ou análise preditiva, o *machine learning* proporciona aos profissionais da área a produção de modelos capazes de auxiliar a tomada de decisões através de resultados confiáveis, além da obtenção de informações "escondidas" na descoberta de tendências históricas e no aprendizado das relações e repetições (Supawan L 2014).

O uso de técnicas de processamento de linguagem natural na mineração de texto tem o objetivo de identificar e mensurar a real importância de um termo em um determinado contexto possibilitando maximizar os resultados obtidos na mineração de texto (Aranha 2007).

3. Metodologia

Para o desenvolvimento do projeto foi adotada a metodologia CRISP-DM (IBM 2011). Como modelo de processo o CRISP-DM fornece uma visão geral do ciclo de vida de um projeto de mineração de dados. O processo CRISP-DM consiste de seis fases organizadas de maneira cíclica e interligadas. Entretanto, as fases não precisam necessariamente seguir um fluxo contínuo, uma determinada fase pode ser repetida diversas vezes ou ser pouca explorada conforme a necessidade do projeto, ou até mesmo sofrer alterações em caso de erros nos resultados ou mudanças de objetivos (Olson, D. L., & Delen, D. 2008).

Figura 1 – Diagrama CRISP-DM



Fonte: IBM 2011

3.1. Entendimento do Negócio

Nos últimos anos uma análise do cenário político brasileiro mostra que a população está cada vez mais interessada em política, isso ficou mais evidente nas últimas eleições presidenciais com as intensas discussões políticas que povoaram as reuniões familiares e as redes sociais (Folha de São Paulo 2018).

Dessa forma, será utilizado o Twitter para realizar uma análise de sentimentos em relação ao candidato eleito presidente do Brasil em 2018. A análise consiste em classificar os comentários entre positivo e negativo, sempre sob a ótica do candidato, utilizando uma aplicação criada para esse fim. Os comentários classificados viabilizam a realização de uma avaliação evolutiva do candidato; antes, durante e depois das eleições a respeito do sentimento dos usuários do Twitter em relação ao candidato eleito e que influências teve no resultado e em alguns dos pontos marcantes de sua campanha política.

3.2. Entendimento dos Dados

As principais tarefas dessa fase são (Chapman, P., Clinton, J. & Kerber, R. 2000):

- Coleta inicial dos dados.
- Visualização dos dados
- Descrição e relatórios dos dados coletados
- Exploração dos dados
- Verificação da qualidade dos dados.

Para coleta dos tweets, foi desenvolvido um algoritmo na linguagem de programação *Python*, utilizando a biblioteca *tweepy* para conexão e acesso a API do Twitter. Nesse algoritmo foi utilizado o conceito de *listener*, onde a aplicação em execução "escuta" tweets postados aleatoriamente conforme o filtro de pesquisa, que nesse caso foram o nome, apelido e jargões no intuito de obter somente textos sobre o então candidato (Python docs. 2019).

Foram coletados tweets nos seguintes períodos:

- Antes do 1º turno (Setembro de 2018);
- Entre o 1º e 2º turnos (Outubro 2018);
- Após eleição (Outubro e Novembro 2018);
- Após a posse (Janeiro e Fevereiro 2019).

Os tweets coletados foram salvos em um formato de texto estruturado. Foi analisado que esse texto contém não só a mensagem do tweet digitado, mas também diversas informações adicionais, com código identificador único, a data e hora da criação do tweet, o usuário de criação, a localização do usuário (se o mesmo preencheu essa informação), entre outros.

Constatou-se então que esses arquivos carregam informações importantes para serem usadas adicionalmente na análise de sentimento, possibilitando análises no tempo, por regiões etc.

3.3. Pré-Processamentos dos Dados

Com um volume adequado de tweets coletados e após o entendimento e armazenamento dos mesmos, foram iniciados os estudos das formas para classificação de uma base de treinamento. Esse fator é essencial para a aplicação de modelos supervisionados de mineração de textos e *machine learning*, onde essa base será utilizada como treino para os modelos (Liu B 2012).

A classificação da base de treinamento consiste em separar um percentual dos dados coletados, aleatoriamente, para então classificá-los manualmente, conforme o conceito que será aplicado na análise do sentimento em relação ao alvo, no caso, o atual presidente da República. Cada texto foi classificado em positivo e negativo, tendo como foco a perspectiva do sentimento em relação ao presidente (Liu B 2012).

Para facilitar a classificação, foi desenvolvido um aplicativo web responsivo, podendo ser acessado por PCs e aparelhos *mobile*. Esse aplicativo foi desenvolvido na linguagem JAVA, utilizando a ferramenta case Genexus que é capaz de gerar o código fonte da aplicação abstraindo isso do desenvolvedor, podendo o mesmo focar nos conceitos de negócio da aplicação. Para viabilizar o desenvolvimento, foram desenvolvidos programas para conversão dos arquivos dos tweets para uma base de dados relacional. O banco de dados utilizado foi o MS SQL Server.

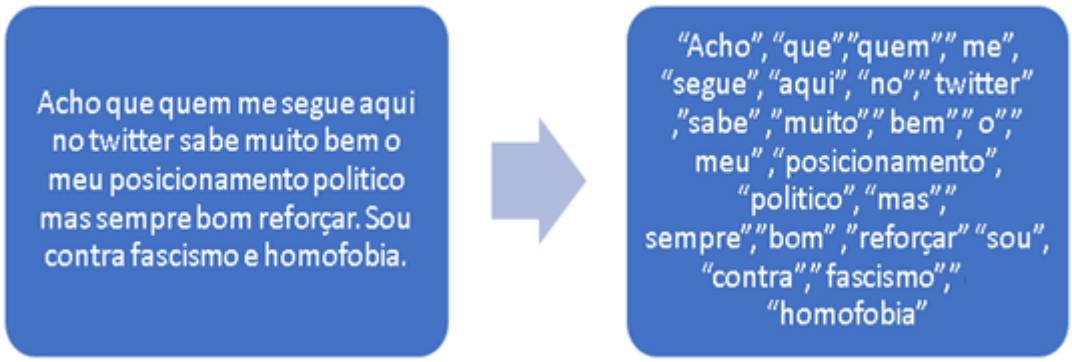
Foram classificados manualmente através da API desenvolvida nesse projeto aproximadamente 3.000 tweets, sendo que 64,5% eram negativos e o restante positivo. Na sequência, foram aplicadas as seguintes técnicas de transformação de dados:

- O arquivo da amostra foi codificado para UTF-8, dessa forma o *Python* reconhece caracteres especiais e normaliza sua leitura.
- Caracteres especiais, números e datas foram excluídos da amostra, a fim de utilizar apenas as palavras do texto.
- Foram excluídos os nomes dos usuários e as palavras que continham *Hashtags*.
- As classificações dos tweets são transformados em 1 positivo e 0 negativo.

Após o tratamento da amostra, as técnicas de pré-processamento de texto são aplicadas através de métodos da biblioteca de *machine learning* (mlib) do Spark.

Todas as palavras viram tokens, como na Figura 2:

Figura 2 – Tokenização

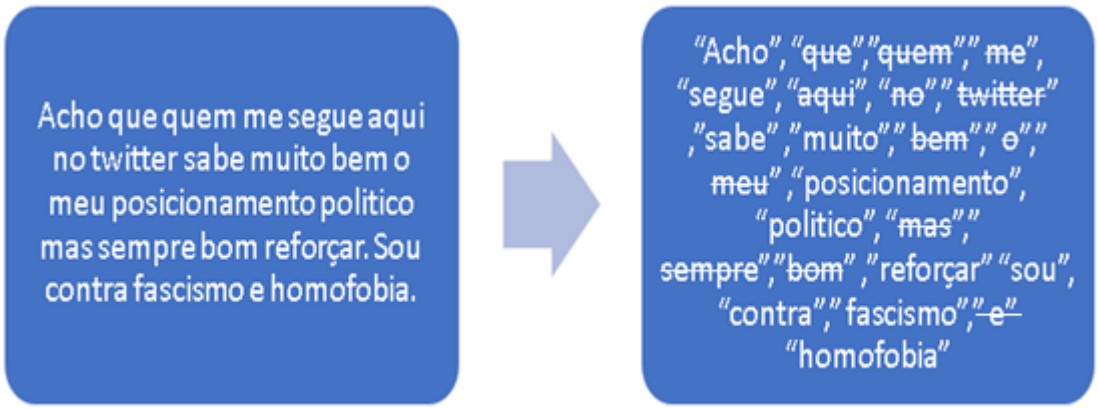


Fonte: Elaborada pelo autor

Com o objetivo de limitar a quantidade de termos principais e com o intuito de se manter apenas os termos chaves de cada documento são eliminadas palavras que não possuem relevância significativa no texto. Toma-se por exemplo: pronomes, preposições, artigos e interjeições. Essas palavras tem a função de conectar os termos, sendo assim, não há a necessidade de adicioná-las na estrutura de índices. Essa etapa é conhecida como *stopwords* (Muller, 2003).

Na amostra utilizada não foi necessário incluir novas condições, resultando na Figura 3:

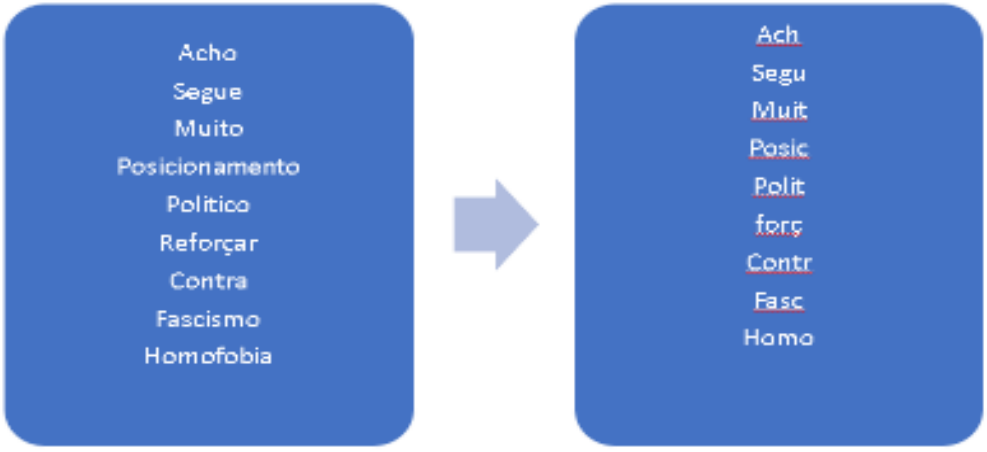
Figura 3 – Stop Words



Fonte: Elaborada pelo autor

Após a retirada de *stopwords*, todas as palavras são radicalizadas, ou seja, toda palavra é transformada para sua forma radical para diminuir as variações. Essa é a fase de *stemming* (Feldman & Sanger, 2007) e pode ser visualizada na figura abaixo:

Figura 4 – Stemming



Fonte: Elaborada pelo autor

Por fim, é aplicado a técnica *Bag of Words* para criar um vetor numérico das frequências de palavras que aparecem nos textos das publicações do *twitter*. Depois de vetorizado o texto está pronto para a aplicação do modelo preditivo (Muller, 2003).

3.4. Modelagem

Na fase de modelagem os modelos matemáticos são aplicados para extraírem conhecimento e encontrarem padrões nas amostras de dados (McDonald D. 2012). Foram

criadas duas tabelas com os dados recém-tratados na fase de tratamento de dados. Sendo uma tabela para teste e outra para treinamento. A base de treinamento é composta de 70% da base original e a de teste 30% dívida aleatoriamente através da biblioteca do Spark com o método randomSplit (McDonald D. 2012).

Foram aplicados os seguintes modelos de aprendizado supervisionado, pelo fato da base do projeto já estar classificada (McDonald D. 2012):

- **Regressão Logística** - Modelo estatístico destinado às análises de dados em que as variáveis são categóricas, ou seja, restringem a uma opção positiva ou negativa e pode ser representada de forma binária, por 0 e 1 (Hosmer, D. W., & Lemeshow, S. 2000).
- **Naive Bayes** - Naive Bayes é um algoritmo de classificação que utiliza dados históricos para prever um evento futuro calculando sua probabilidade. Ele foi amplamente usado para algoritmos que filtram spam de e-mails, mas pode ser utilizado em diversas áreas (Konchady, M. 2006).
- **Support Vector Machine** - É um algoritmo que busca encontrar um hiperplano entre N dimensões para tentar classificar uma amostra de dados. Exemplificando, o algoritmo cria uma linha, ou um plano, dividindo os elementos classificados da amostra de dados. Quando um novo elemento é inserido o algoritmo busca classificar em qual plano ele está inserido (Olson, D. L., & Delen, D. 2008).
- **Redes Neurais** - As Redes Neurais podem ter centenas ou milhares unidades de processamento, simulando o cérebro de grandes mamíferos e apresentando um modelo inspirado na sinapse dos neurônios, onde a experiencia adquirida é passada através de uma rede para ser dividida em todo o sistema neural. Essa rede possui uma camada de entrada onde as informações são consumidas, diversas camadas intermediárias que processam e trocam informações entre si e uma camada de saída com o resultado apresentado. As redes neurais buscam encontrar padrões nas informações inseridas e buscar pesos e medidas determinísticas (Gurney, K. 1997).

Para aplicar os modelos de *machine learning*, foi utilizada a técnica de *crossvalidation* da biblioteca de *machine learning tunning* do *Spark.* O *crossvalidation* divide os dados aleatoriamente em um número pré-determinado de partições, realiza treino e teste dos dados separadamente buscando o melhor resultado (Apache Org 2019).

3.5. Avaliação

Depois de aplicado os modelos de *machine learning* na base previamente classificada, o modelo retorna uma coluna com resultado do algoritmo aplicado, com o valor de 1 para positivo ou 0 para negativo. Foram aplicadas técnicas de classificação da biblioteca de *machine learning (mlib)* do *Spark* para obter os resultados das métricas dos modelos aplicados (Apache Org 2019). Para o projeto foram escolhidas as técnicas e determinadas métricas com o objetivo de indicar (Lu Peter & Garisson Eric 2018):

- **Acurácia** - Número de acertos divididos pelo número total da base.
- **Precisão** - Número de vezes que a uma classe foi predita corretamente dividia pela total base de testes.
- **Revocação** - Número de vezes que uma classe foi predita corretamente dividia pelo total da base.
- **F1 Score** - Métrica de avaliação que equilibra precisão e revocação.

3.6. Implantação

Com o modelo escolhido, toda a base de dados com aproximadamente 22.000 tweets foi inserida no ambiente do Spark. Novamente foram aplicadas as técnicas de padronização e pré-processamento de dados e aplicado o modelo escolhido, que no projeto foi o *Support Vector Machine (SVM*). Depois de preditos os dados foram disponibilizados em um arquivo excel para que sejam consumidos por uma aplicação de *Business Intelligence* e possam ser desenvolvidos relatórios analíticos.

4. Resultados e Discussão

O primeiro resultado obtido no projeto foi o dos modelos de machine learning aplicados. Abaixo na figura 5 são ilustrados os valores de cada métrica desejada, obtida por modelo aplicado em aproximadamente 3.000 tweets.

Figura 5 – Resultados dos Modelos

Naive Bayes	
Acurácia	80%
Precisão	80,10%
Revocação	80,11%
F1 Score	80,10%

Redes Neurais	
Acurácia	44%
Precisão	19%
Revocação	44%
F1 Score	27%

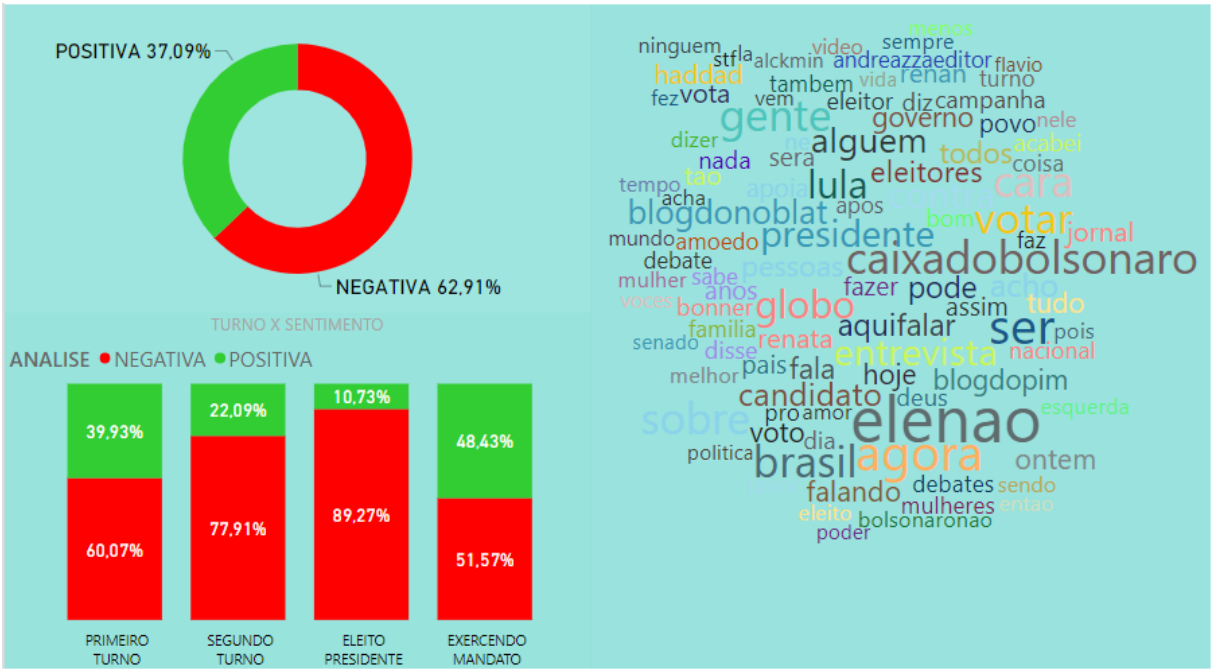
LinearSVC	
Acurácia	81%
Precisão	81%
Revocação	81%
F1 Score	81%

Regressão logística	
Acurácia	78,52%
Precisão	80,73%
Revocação	78,52%
F1 Score	78,69%

Fonte: Elaborado pelo autor

Analisando o todo, fica claro que pelos números apresentados que o modelo SVM é o que possui os maiores indicadores em todos os atributos e foi o escolhido para ser aplicado nas bases de tweets não classificadas. Os modelos SVM e Naive Bayes são indicados para analise de texto por sua facilidade no treinamento, eficiência e rapidez para avaliação (Susan D. 1998). Com o modelo SVM implementado na base total, foi desenvolvido um dashboard com os resultados da predição, conforme a figura 6:

Figura 6 – Dashboard Resultado



Fonte: Elaborado pelo autor

O dashboard evidencia a cronologia vivida pelo candidato durante os períodos escolhidos:

- PRIMEIRO TURNO – 01/09/2018 A 07/10/2018
- SEGUNDO TURNO – 08/10/2018 A 28/10/2018
- ELEITO PRESIDENTE – 29/10/2018 A 31/12/2018
- EXERCENDO MANDATO – 01/01/2019 A 28/02/2019

Com base nos resultados do relatório desenvolvido é possível supor que:

- No primeiro turno a quantidade de candidatos era maior, mesmo assim percebe-se uma possível mobilização negativa contra o presidente eleito. Também pode ser atribuído o número considerável de analises positivas devido ao possível apoio dos eleitores ao episódio do atentado a vida do candidato no dia 06/09/2018.
- No segundo turno, com apenas dois candidatos, os eleitores contrários ao atual presidente podem ter tentado uma mobilização para impedir uma possível vitória, por isso o número maior de análises negativas e também as críticas pela não participação do candidato nos debates políticos.
- No período em que o atual presidente já estava eleito temos dois fatores que talvez explicam os números majoritariamente negativos. Primeiramente os apoiadores do atual presidente já tinham conquistado a vitória, portanto as suas manifestações nas redes sociais diminuíram consideravelmente. Já os eleitores contrários ao atual presidente com a certeza da derrota aproveitaram para realizar ainda mais manifestações negativas movidos pela frustração e aumentaram ainda mais o discurso de oposição.
- Já no período de início do mandato percebe-se um equilíbrio maior nas avaliações, possivelmente pelo entusiasmo dos apoiadores com a posse e a perspectiva de um novo mandato e maior resignação dos eleitores que votaram contra o atual presidente.

5. Considerações Finais

Entendendo o poder das redes que rompem barreiras, da voz a todos os usuários, como foi impactante e utilizada incansavelmente na corrida eleitoral, pode-se observar que a análise de sentimentos é muito importante para tentar entender a opinião dos internautas. Neste projeto foi utilizada a rede social Twitter buscando publicações do candidato eleito na Eleição Presidencial de 2018 para analisar o sentimento dos seus usuários a respeito do atual presidente da república. Aplicando a metodologia CRISP-DM e algoritmos de *machine learning* conclui-se que:

- Durante o primeiro turno das eleições as publicações negativas foram de 60,07% e as positivas 39,93%.
- Durante o segundo turno das eleições as publicações negativas foram de 77,91% e as positivas 22,09%.
- Depois de eleito, período de novembro a dezembro de 2018, as publicações negativas foram de 89,27% e as positivas 10,73%.
- Contando apenas publicações no atual mandato do presidente da república em 2019 as publicações negativas foram de 51,57% e as positivas de 48,43%.

Os resultados apontam que a maioria das publicações dos usuários do Twitter é de um tom negativo em relação ao presidente eleito em 2018. Nota-se um avanço no negativismo depois da eleição, provavelmente pelo tom de cobrança da população. Nas próximas etapas o projeto visa correlacionar os índices de análise de sentimento com fatos políticos em destaque.

6. Referências

Apache org. (2019 Cross) Validator Forum https://spark.apache.org/docs/2.2.0/api/scala/index.html#org.apache.spark.ml.tuning.CrossValidator.

Aranha, C. N. (2007). Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional. Tese de Doutorado, Departamento de Engenharia Elétrica, PUC- Rio.

BBC Brasil (2016) Pagamento de propinas por empreiteiras se consolidou durante ditadura, diz historiador. https://www.bbc.com/portuguese/brasil-38337544

Chapman, P., Clinton, J., & Kerber, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

El Pais (2018) Por que devemos nos preocupar com a influência das redes sociais nas eleições? https://brasil.elpais.com/brasil/2018/09/21/opinion/1537557693_143615.html

Feldman & Sanger, J. (2007). The Text Mining Handbook. Cambridge: Cambridge University Express.

Folha de São Paulo (2018). Aumento de participação popular nas eleições amplia democracia no Brasil em 2018, afirma Economist. https://www1.folha.uol.com.br/poder/2019/01/aumento-de-participacao-popular-nas-eleicoes-amplia-democracia-no-brasil-em-2018-afirma-economist.shtml

Gurney, K. (1997) Na introduction to Neural Network (1. edition) London U.K, UCL Press
Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression (Vol. 2). New York: Jhon Wiley & Sons.

IBM (2011). IBM SPSS Modeler CRISP-DM Guide (14 ed.).

Jiawei, H. (2006). Data Mining: Concepts and Techniques (Second Edition ed.). San Francisco: Elsevier.

Konchady, M. (2006). Text Mining Application Programming. Michigan: Charles River Media.

Liu B (2012) Sentiment Analysis and Opinion Mining. Morgan & Claypool

Lu Peter & Garisson Eric (2018) Evalute Model. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/evaluate-model

Mitchell, T. (1997) Machine Learning. New York: McGraw Hill.

Muller, D. N. (2003). Processamento de linguagem Natural. Universidade Federal do Rio Grande do Sul, Porto Alegre.

McDonald D. U. K. (2012). Value and benefits of text mining. Digital Infrastructure.

Nicolau J. (2012) Eleições no Brasil: Do Império aos dias atuais. Zahar: São Paulo

Olson, D. L., & Delen, D. (2008). Advanced Data Mining Techniques. Springer.

Python docs. (2019) Python Org. https://python-can.readthedocs.io

Supawan, P., Chen, Y., Ping, Y., & Chen, P. (2005). Machine Learning in Bioinformatics. Melbourne: Springer.

Susan D. (1998) 7th International Conference on Information and Knowledge Management https://www.microsoft.com/en-us/research/publication/inductive-learning-algorithms-and-representations-for-text-categorization/

Tribunal Superior Eleitoral (2018). Perfil do Eleitorado Brasileiro
https://politica.estadao.com.br/blogs/fausto-macedo/wp-content/uploads/sites/41/2018/08/Slide-1.pdf

Glossário

Genexus – Ferramenta de Desenvolvimento
Java – Linguagem de programação
MS SQL SERVER – Ferramenta de banco de dados
Python – Linguagem de programação

Agradecimentos

Gostaríamos de agradecer a todos os familiares, aos amigos, o nosso orientador Prof. Me Fernando Sequeira Sousa e a faculdade Impacta de Tecnologia,