

Análise de Sentimento de Insegurança Através do *Twitter*

André S. Nascimento, Marcelo I. Reis

Graduação Tecnológica em Análise e Desenvolvimento de Sistemas – Universidade
Católica do Salvador (UCSal)
41740-090 – Salvador – BA – Brasil

andresoteropolitano10@gmail.com, marcelo.reis@pro.ucsal.br

Abstract. *The sense of insecurity has a direct influence on the citizen's quality of life. This work used machine learning algorithms to identify the feeling of insecurity of the Brazilian citizen, expressed in posts on the social network Twitter. For this, a balanced database was generated with 400 tweets, which were previously classified between feelings of insecurity or others. Prediction models were created to classify the tweets posted in the network with the Random Forest, SVM and Logistic Regression algorithms. The algorithms reached indexes of up to 0.34 for the Matthews correlation coefficient and 0.69 for the F1 measure. A web application was developed for online ranking of the new tweets published using a better strategy of three, demonstrating the viability to use this approach to accompany the sensation of insecurity among the users of this network.*

Resumo. *A sensação de insegurança tem influência direta na qualidade de vida do cidadão. Este trabalho utilizou algoritmos de aprendizado de máquina para identificar a sensação de insegurança do cidadão brasileiro, expressa em postagens na rede social Twitter. Para isso foi gerada uma base de dados balanceada com 400 tweets, os quais foram classificados previamente entre sentimento de insegurança ou outros. Foram criados modelos de previsão para classificar os tweets postados na rede com os algoritmos Floresta Aleatória, SVM e Regressão Logística. Os algoritmos alcançaram índices de até 0,34 para o coeficiente de correlação de Matthews e 0,69 para a medida F1, sendo desenvolvida uma aplicação web para classificação on-line dos novos tweets publicados utilizando uma estratégia de melhor de três, demonstrando a viabilidade de se utilizar tal abordagem para acompanhar a sensação de insegurança entre os usuários dessa rede.*

1. Introdução

Desde a década de 1990 e ao longo do início do século XXI que a população brasileira vem questionando a eficiência das políticas de segurança no combate à criminalidade [Reishoffer e Bicalho, 2009]. A falta de eficácia dessas políticas tem aumentado o sentimento de insegurança da população, que teme sofrer danos materiais e físicos.

Sendo assim, as autoridades vêm buscando estratégias no combate a violência e insegurança. Esta conceitualmente é o receio de ser vítima de um criminoso e, conseqüentemente, ter a sensação de impotência para inibir o delito [Frattari, 2009].

As redes sociais *Twitter*, *Facebook*, e *Instagram* são espaços virtuais bastante utilizados pelos brasileiros. Os usuários dessas redes comentam fatos, citam suas experiências e observações sobre situações cotidianas. Tais espaços podem vir a ser utilizados pelas autoridades, através de monitoramentos, para obter informações sobre práticas criminosas e a conseqüente sensação do cidadão, auxiliando nas decisões estratégicas para o combate a violência.

Diante de tal situação, o problema abordado neste trabalho foi a identificação do sentimento de insegurança descrito em postagens realizadas no *Twitter*. Como solução, foi utilizada a inteligência artificial, através de técnicas de aprendizado de máquina para classificar as postagens realizadas nessa rede.

Foram utilizados os algoritmos Floresta Aleatória, SVM e Regressão Logística para a geração dos modelos preditivos, e utilizado as matrizes de confusão, o coeficiente de correlação de Matthews, medidas de precisão, revocação, medida F1, acurácia e taxa de erro para avaliar esses modelos. Ainda, foi construída uma aplicação web com um painel de indicadores para realizar previsões em tempo real sobre os novos *tweets* registrados.

2. Fundamentação Teórica

2.1 A Violência e o Sentimento de Insegurança

A violência pode ser definida como uma ação que ultrapassa os limites humanos, seja em suas condições física e psíquica ou em seus aspectos sociais, ou seja, é quando os indivíduos são observados como objetos [Sena et al., 2013]. A violência age contra a natureza do ser humano através do constrangimento, tortura, violação psicológica, privação de liberdade, etc [Silva, 2007].

Para Frattari (2009) a violência modifica a forma de interação social. Dessa maneira as pessoas buscam ambientes mais fechados ou distantes de espaços pouco movimentados, além de, às vezes, repreender com "suas próprias mãos" as atividades delituosas.

A insegurança é a falta da garantia na preservação da ordem pública e da confiança no poder de polícia [Anjos, 2011]. A insegurança está associada à probabilidade de ser vítima da violência, ao aumento nas taxas de criminalidade nas cidades e às experiências vivenciadas pela população [Frattari, 2009].

2.2 Análise de Sentimento e Mineração de Textos

A análise de sentimento é um processo de avaliação de emoção, sentimento, estimativas, etc sobre organizações, produtos, serviços e eventos. Bo Pang e Lillian Lee (2008) destacam o crescimento do uso e importância de técnicas computacionais que obtêm essas informações em textos disponibilizados amplamente pela Web.

A mineração de textos é o processo de busca de padrões e informações em bases textuais, ou seja, dados não estruturados a fim de obter informações em relação a sentimentos, emoções e opiniões [Madeira, 2015]. Através de técnicas de aprendizado de máquina como classificação, associação ou agrupamento pode-se obter conhecimento de forma automatizada dessas bases [Mitchell, 1997][Bishop, 2006].

A mineração de textos permite três abordagens técnicas de classificação para o conhecimento de opinião ou análise de sentimento. São elas: a aprendizagem de máquina que ensina um algoritmo na predição de uma característica; método léxico que atribui uma polaridade para cada palavra do documento e sendo que várias técnicas podem ser aplicadas para rotular cada instância do modelo; e o método baseado em análise linguística que considera a função sintática das palavras, ou seja, a classificação gramatical e a posição de cada palavra no documento [Santos, 2013].

A aprendizagem de máquina requer um modelo de dados classificado por um conhecedor do assunto buscado. A técnica de rotular o modelo para a aprendizagem de máquina é denominada de aprendizagem supervisionada, porque o algoritmo de classificação precisa do modelo rotulado anteriormente para ser treinado para as predições. Quando cada instância do modelo não é rotulada, a técnica aplicada para polarização é chamada de aprendizagem não supervisionada.

2.3 Trabalhos Correlatos

Caragea et. al, (2014) realiza a classificação do sentimento de posts de usuários no *Twitter* durante o furacão Sandy e visualiza esses sentimentos em um mapa geográfico centrado em torno do furacão. Terpstra et. al, (2012) investiga as possibilidades de análise automatizada em tempo real do *Twitter* durante as crises. Ele analisou quase 97.000 tweets que foram publicados pouco antes, durante e depois que uma tempestade atingiu o Pukkelpop 2011 festival na Bélgica. Já Zielinski, (2013) utiliza técnicas combinadas de Análise de Confiança (TA), Classificação Multilíngue de *Tweeting* (MTC) e Geoprocessamento (GEO) na mineração de texto em mídias sociais para apoio à decisão em gestão de crises naturais

O uso de técnicas de mineração de texto e análise de sentimentos tem se tornado comum no domínio da gestão de desastres. Trabalhos como os relatados acima tentam capturar informações em tempo real dos envolvidos em algum evento catastrófico, como forma de fornecer às autoridades que atuam na emergência uma maneira alternativa de colher informações de campo, informações essas que no passado eram geradas apenas pelas equipes de emergência no local do sinistro. Assim é propiciada a produção de consciência situacional durante esses eventos.

Já no domínio da segurança pública essa ainda não é uma abordagem comum. Apesar disso, considerando que possa existir uma relação entre esse e a gestão de desastres, é possível que as autoridades se beneficiem com esse tipo de tecnologia para colher a percepção da sociedade em relação à violência, ocorrência de crimes e, como

neste trabalho, a sensação de insegurança. Esta última tem grande influência na qualidade de vida da população, e também por isso pode ser monitorada através do uso de técnicas de mineração de textos, análise de sentimento e aprendizagem supervisionada em redes sociais, comuns entre este trabalho e os demais citados.

3. A Metodologia

3.1 Preparação da Base de Tweets

A linguagem de programação *Python* V 3.6.2 e o *Twitter* foram utilizados na coleta de dados. Bibliotecas *Python* foram utilizadas no pré-processamento, mineração e pós-processamento do modelo. O *Twitter* disponibiliza uma API na arquitetura REST (*Representation State Transfer*) para o consumo de seus web services.

Na fase de pré-processamento foi utilizada a biblioteca *nlTK* V 3.2.5 para as tarefas de remoção de *stop words* e aplicação de *stemming*, e a biblioteca *pandas* V 0.20.1 que foi utilizada para remover os *tweets* duplicados. Nesta mesma fase, funções *Python* foram desenvolvidas para remoção de URL's e mencionamentos de *tweets*, acentuação gráfica, *hashtags*, números e pontuações dos textos. Também as letras maiúsculas foram passadas para minúsculas. Todas essas tarefas padronizam palavras escritas de várias maneiras, corrigem dados inconsistentes e retiram trechos irrelevantes para a classificação [Castanheira, 2008].

Finalmente, nesta etapa, foram criados vetores com as frequências das palavras de cada *tweet* utilizando a biblioteca *nlTK* V 3.2.5. Para isso foi utilizada a técnica de *bag of words*.

Com a coleta de dados, foram obtidos 400 *tweets*. A rotulação dos mesmos foi baseada na análise pessoal e subjetiva do autor, que se embasou nos conceitos de violência e insegurança bem como no contexto expresso nas mensagens. A validação dessa análise foi procedida pelo coautor que possui experiência na área de segurança pública. Podemos observar nos exemplos da tabela 1 que palavras como violência, medo, assédio, etc, contextualizam o sentimento de insegurança por parte de quem publicou os *tweets*.

Tabela 1. Exemplos de tweets rotulados como sentimento de insegurança

“Minha mãe tá muito preocupada com esses lance de violência que tem acontecido com os LGBT aqui em Chape”
“da até medo de sair na rua à noite de tanto que tá a violência”
“Ser mulher é sofrer assédio as 8h da manhã no trem. Estou quase chorando de tão nervosa”

Os 400 *tweets* geraram um dataset com 1821 palavras, utilizadas para treinar os algoritmos. Termos como violência, medo, assédio, etc figuraram entre aqueles com as maiores frequências. Especificamente a palavra violência teve 64 ocorrências. Podemos observar na distribuição de frequência da figura 1 esses exemplos.

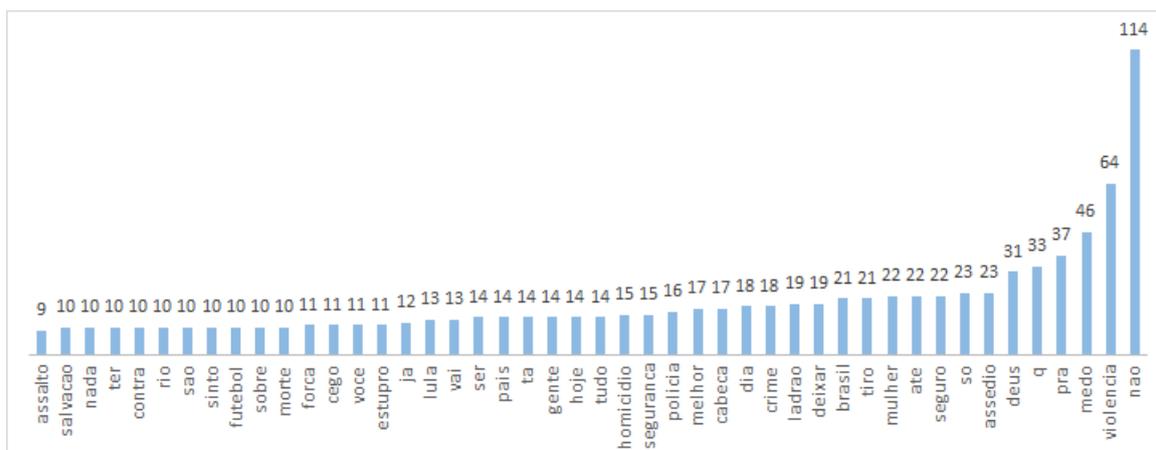


Figura 1. Distribuição de frequência com palavras usadas para treinar os algoritmos

É interessante ressaltar, contudo que, metade dos *tweets* foram classificados como sentimento de insegurança, o que ajudou a gerar essa alta frequência de termos negativos. Mesmo assim, isso ajuda a compreender como todas essas palavras citadas tem relação com o sentimento de insegurança.

4. Os Algoritmos de Classificação e Seus Resultados

4.1 Classificação

A classificação atribui um rótulo a um registro através de um algoritmo, de acordo com o conjunto de dados previamente classificado como parâmetro de entrada para o treinamento do modelo, através da aprendizagem supervisionada. Assim, o algoritmo associa as características de cada registro do modelo de treino com a sua classe, gerando um aprendizado de máquina para prever a classe de novos registros [Mitchell, 1997].

Para treinar o algoritmo de classificação, foi usada a biblioteca *scikit-learn* 0.18.1, passando a frequência das palavras e seus rótulos para os algoritmos aplicados neste trabalho: Floresta Aleatória, SVM e Regressão Logística. A escolha desses algoritmos e mais os de Árvore de Decisão e *Naive Bayes* que foram posteriormente descartados, se deu pelos bons resultados apresentados por Joachims, (1998), Porotsky, (2011) e Maia, (2013) quando da realização de tarefas de text mining pelos dois primeiros e classificação de documentos pelo último.

Após o treinamento do modelo, é necessário testar o resultado da aprendizagem com alguma abordagem de validação. A técnica de *cross validation* foi utilizada no teste dos algoritmos, sendo k a divisão em k partes do conjunto de *tweets* de teste, que neste trabalho teve o valor arbitrado para 10. Os algoritmos realizam um treinamento com cada 9 partes e a outra divisão serve para teste. Assim, é estabelecido um ciclo de treino e teste do conjunto total dos dados que é executado 10 vezes [Mitchell, 1997].

Existem alguns métodos para medir a eficiência dos algoritmos, entre eles a matriz de confusão e o coeficiente de correlação de Matthews que é como uma medida da qualidade para classificações binárias (duas classes). É levado em conta os verdadeiros e falsos positivos e negativos, sendo considerado uma medida balanceada que pode ser usada mesmo se as classes forem de tamanhos muito diferentes [Matthews, 1975]. O seu valor varia de -1 a +1. Quanto mais próximo de +1 melhor o resultado. Na figura 2 podemos observar a sua fórmula.

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

Figura 2. Fórmula do coeficiente de correlação de Matthews

Já a matriz de confusão monitora os acertos e erros da classificação. Para isso é gerada uma matriz que contém os números das classificações com verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN).

4.2 Os Resultados

Para a geração dos modelos preditivos foi gerada uma base balanceada com 400 *tweets* previamente classificados, sendo a metade como inseguros e a outra metade como outros. Foram também testados os algoritmos de Árvore de Decisão e *Naive Bayes*, mas devido a sua baixa performance na classificação os mesmos foram descartados. Na figura 3 observamos a matriz de confusão com os resultados do algoritmo Floresta Aleatória.

```

INSEGURO - OUTRO
INSEGURO [[138 62]
OUTRO [ 77 123]]

```

Figura 3. Classificação com algoritmo Floresta Aleatória

Na figura 4 observamos os resultados do algoritmo SVM, e na figura 5 os resultados do algoritmo Regressão Logística. Os valores do coeficiente de Mathews podem ser observados na tabela 2.

```

INSEGURO - OUTRO
INSEGURO [[149 51]
OUTRO [ 82 118]]

```

Figura 4. Classificação com algoritmo SVM

```

INSEGURO - OUTRO
INSEGURO [[151 49]
OUTRO [ 84 116]]

```

Figura 5. Classificação com algoritmo Regressão Logística

Os resultados obtidos com a execução dos três algoritmos são praticamente homogêneos na criação dos modelos. As matrizes de confusão indicam uma tendência nos modelos de produção de falsos positivos de 20% e falsos negativos de 13,5% entre os três algoritmos, o que dá uma taxa de erro de 33,5% e de acerto (acurácia) de 66,5% no total. Também foram calculadas as métricas de precisão, revocação e medida F1, como podemos ver na tabela 2 os valores, e na figura 6 um gráfico radar representativo.

Tabela 2. Métricas de avaliação

	Precisão	Revocação	Medida F1	Correlação de Matthews
Floresta Aleatória	0,64	0,69	0,66	0,30
SVM	0,64	0,74	0,68	0,33
Regressão Logística	0,64	0,75	0,69	0,34

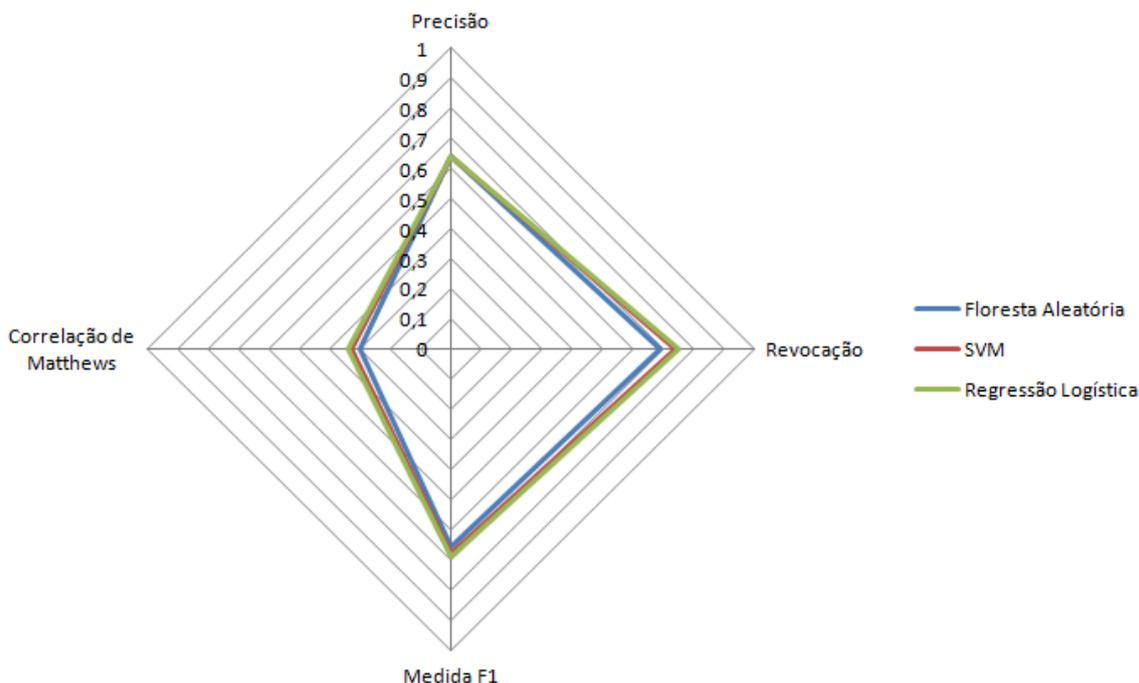


Figura 6. Gráfico radar de métricas avaliativas

Muitos fatores podem influenciar essa taxa de acerto não tão performática como, por exemplo, o contexto da frase, ambiguidades, sarcasmos ou mesmo regionalismos. Neste último caso devemos considerar que os *tweets* analisados foram gerados em todo o território brasileiro.

A precisão, figura 7, é a taxa que indica dentre os classificados como positivos, quais realmente o são [Powers, 2007]. Observamos na tabela 2 que todos os três algoritmos obtiveram desempenhos similares.

$$prec(\hat{f}) = \frac{VP}{VP + FP}$$

Figura 7. Fórmula da precisão

Já a revocação, figura 8, indica quando realmente é positivo, o quanto é classificado como positivo [Powers, 2007]. Na tabela 2 observamos que o algoritmo da regressão logística obteve o melhor desempenho. O objetivo a ser perseguido deve ser o equilíbrio da precisão e revocação com o mesmo valor, que tem no valor +1 a expressão do máximo de desempenho. Esse é um objetivo difícil de ser atingido.

$$rev(\hat{f}) = \frac{VP}{VP + FN}$$

Figura 8. Fórmula da revocação

Uma medida que avalia ao mesmo tempo precisão e revocação é a medida F1, figura 9 [Powers, 2007]. Precisão e revocação devem ser observadas em conjunto, e não de forma isolada. Observamos então na tabela 2 que o algoritmo de regressão logística apresentou o melhor desempenho para essa métrica, que somado ao coeficiente de

correlação de Matthews indica que esse foi de uma maneira geral o algoritmo com melhor performance.

$$F_1(\hat{f}) = \frac{2 \times rev(\hat{f}) \times prec(\hat{f})}{rev(\hat{f}) + prec(\hat{f})}$$

Figura 9. Medida F1

Uma aplicação web foi desenvolvida utilizando a linguagem *Python* e o *Google Charts*, que através da utilização do modelo de predição criado monitora novos *tweets* adicionados à rede. Na figura 10 pode-se ver um painel de indicadores existente na aplicação demonstrando a classificação dos *tweets* em tempo real.

Para tornar a classificação on-line mais efetiva foi utilizada a estratégia de Aquino, (2017) de se classificar cada novo *tweet* com todos os modelos disponíveis, sendo escolhida como classe aquela que obtiver a maioria, no caso, pelo menos duas classificações iguais. Isso ajuda a compensar a taxa de acerto não tão satisfatória dos modelos conforme visto nas matrizes de confusão.

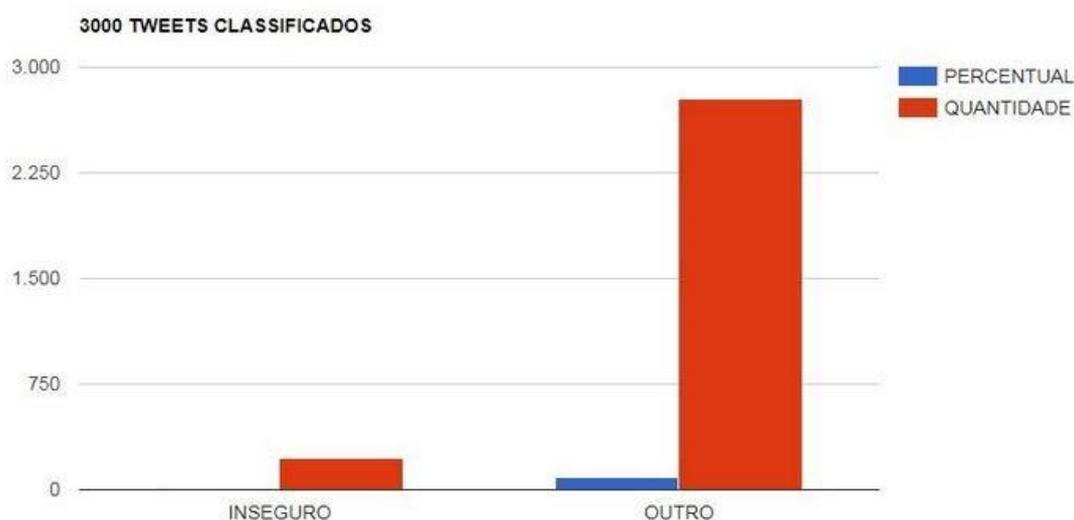


Figura 10. Painel de indicadores da aplicação web para classificação de tweets

5. Conclusão

Nos últimos anos o Brasil vem sofrendo com um aumento substancial da violência, causando reflexos na qualidade de vida dos seus cidadãos que já não conseguem mais se sentir seguros. Dessa maneira, as autoridades vêm buscando novas formas de combate a esse mal.

Considerando o fato das pessoas utilizarem as redes sociais para expressar os seus sentimentos e emoções, buscamos neste trabalho uma maneira automatizada de capturar esse conhecimento. Através de técnicas de aprendizado de máquina foram analisados os sentimentos expressos em postagens do *Twitter*.

Uma base de dados balanceada de 400 *tweets* foi gerada e cada *tweet* foi classificado como em sensação de insegurança ou outros. Após isso foram utilizados três algoritmos para a geração de modelos preditivos, os quais foram validados através da técnica de *cross validation*.

Para melhorar a confiabilidade dos modelos criados, foi implementada na aplicação web desenvolvida uma classificação com os resultados melhores de três. Assim foi realizada a predição com cada algoritmo e eleita a classe indicada pela maioria deles.

Os resultados alcançados indicaram que os modelos produzidos tiveram em média uma taxa de acerto de 66,5% sendo apresentadas as matrizes de confusão que explicitam esses resultados. Também foi utilizado o coeficiente de correlação de Matthews, obtendo-se índices de até 0,34, indicando a viabilidade dessa técnica ser utilizada pelas autoridades para avaliar a sensação de segurança da população. Além disso foram calculadas as métricas de precisão, revocação e medida F1 que indicaram um melhor desempenho do algoritmo de regressão logística.

Como trabalhos futuros sugerimos a criação de uma ferramenta de mapas de calor que poderia georreferenciar os termos de interesse relacionados à sensação de insegurança e violência. Assim seria possível realizar análises do sentimento de insegurança da população, de forma a se poder aplicar medidas mitigadoras direcionadas, como mais policiamento, blitz etc.

Referências

- Anjos, N. N. Funcionalidade do sistema de segurança pública no Brasil e a violência social. Departamento de Estudos da Escola Superior de Guerra. Rio de Janeiro, 2011. Disponível em: <<http://www.esg.br/images/monografias/2011/anjos.pdf>>. Acesso em: 11 out. 2017.
- Aquino, P., & Batista, V. (2017). Sistema de minería de opiniones para el análisis de sentimiento en Twitter. *Avances en Informática y Automática*. Décimo Workshop, pp.38-54. ISBN 978-84-617-9707-3. Retrieved from <http://hdl.handle.net/10366/133039>
- Bicalho, J. C. R. e Gastalho, P. P. Insegurança e produção de subjetividade no Brasil contemporâneo. Universidade Federal do Rio de Janeiro. *Revista de Psicologia*, v. 21 – n. 2, p. 425-444. Instituto de Psicologia, 2009. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=s1984-02922009000200015&lng=pt&nrm=iso>. Acesso em: 30 set. 2017.
- Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer. 2006
- Bo Pang and Lillian Lee (2008), *Opinion Mining and Sentiment Analysis*, *Foundations and Trends® in Information Retrieval*: Vol. 2: No. 1–2, pp 1-135. <http://dx.doi.org/10.1561/1500000011>
- Caragea, C. et al Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy. In: 11th International Conference on Information Systems for Crisis Response and Management. 2014. Conference Proceedings.
- Frattari, N. F. Insegurança: as práticas e discursos do medo na cidade de Goiânia. Universidade Federal De Goiás. Goiânia, 2009. Disponível em: <https://pos-sociologia.cienciassociais.ufg.br/up/109/o/najla_frango_frattari.pdf>. Acesso em: 29 set. 2017.
- Joachims, T. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proceedings of ECML*, Springer, 1998. Disponível em:

- <https://www.cs.cornell.edu/people/tj/publications/joachims_98a.pdf>. Acesso em: 29 set. 2017.
- Madeira, R. O. C. Aplicação de técnicas de mineração de texto na detecção de discrepâncias em documentos fiscais do Rio de Janeiro. Escola de Matemática Aplicada da Fundação Getúlio Vargas. Rio de Janeiro, 2015. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/14593/texto%20disser%20ta%20c3%87%20c3%83o%20vfinal1.pdf>>. Acesso em: 04 out. 2017.
- Maia, P. et al. Application of text mining techniques for classification of documents: a study of automation of complaints screening in a Brazilian Federal Agency. Universidade de Brasília. Departamento de Ciência da Computação, 2013. Disponível em: http://bdm.unb.br/bitstream/10483/7711/1/2013_fernandoleandro dossantos.pdf. Acesso em: 25 out. 2017.
- Matthews, B. W.. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*. 1975.
- Mitchell, T. *Machine Learning*. McGraw Hill. 1997
- Porotsky, S. Bluvband, Z. Advanced Text Mining Algorithms for Aerospace Anomaly Identification. Proceedings of ESREL Conference. 2011. DOI: 10.1201/b11433-54. Acesso em: 08 out. 2017.
- Powers, D. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Technical Report SIE-07-001. 2007. Disponível em: <<https://csem.flinders.edu.au/research/techreps/SIE07001.pdf>>
- Santos, F.. Mineração de opinião em textos opinativos utilizando algoritmos de classificação. Universidade de Brasília. Departamento de Ciência da Computação, 2013. Disponível em: <http://bdm.unb.br/bitstream/10483/7711/1/2013_fernandoleandro dossantos.pdf>. Acesso em: 08 out. 2017.
- Sena, J. A., Lopes, K. T. e Pereira, A. O transtorno de estresse pós-traumático e a violência urbana. *Psicologia ciências biológicas e da saúde*. Maceió, 2013. Disponível em: <<https://periodicos.set.edu.br/index.php/fitbiosauade/article/view/575>>. Acesso em: 25 set. 2017.
- Silva, H. C. Políticas públicas para o idoso: marcos referenciais no trato da violência no município de Florianópolis. Universidade Federal de Santa Catarina, 2007. Disponível em: <<http://tcc.bu.ufsc.br/ssocial285753.pdf>>. Acesso em: 10 out. 2017.
- Terpstra, T. et al. Towards a realtime Twitter analysis during crises for operational crisis management. In: 9 th International Conference on Information Systems for Crisis Response and Management. 2012. Conference Proceedings.
- Zielinski, A. et al. Social Media Text Mining and Network Analysis for Decision Support in Natural Crisis Management. In: 10th International Conference on Information Systems for Crisis Response and Management. 2012. Conference Proceedings.