

UNIVERSIDADE ESTADUAL DE MONTES CLAROS
Centro de Ciências Exatas e Tecnológicas
Curso de Engenharia de Sistemas

Lucas Franklin Silva
Mateus Felliipe Alves Lopes

PROCESSAMENTO DE LINGUAGEM NATURAL
ANÁLISE DE SENTIMENTOS DO TWITTER

Montes Claros - MG
Novembro / 2018

ANÁLISE DE SENTIMENTOS DO TWITTER

1. INTRODUÇÃO

Processamento de Linguagem Natural (PLN ou *NLP do inglês*) consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural. O objetivo do PLN é fornecer aos computadores a capacidade de entender e compor textos. “Entender” um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos, analisar sentimentos e até aprender conceitos com os textos processados.

A análise de sentimentos é uma técnica que consiste em extrair informações de textos em linguagem natural. O objetivo dessa técnica é obter, de forma automática, a polaridade de um texto ou sentença. O interesse da indústria e da academia neste campo de estudo é, em parte, devido ao seu potencial de aplicações, tais como: marketing, relações públicas e campanhas políticas (FREITAS; VIEIRA, 2015).

Freitas e Vieira (2015) afirmam que pesquisas na área de análise de sentimentos frequentemente são aplicadas a textos em inglês e que outras línguas são menos exploradas. As bibliotecas utilizadas para o desenvolvimento do projeto não suportam nativamente a análise de sentimentos em português, assim duas opções são possíveis, treinar modelos de machine learning com dicionários em português já classificados, ou usar ferramentas de tradução de texto, a primeira opção foi a adotada para o desenvolvimento do projeto.

2. OBJETIVOS

O objeto do projeto foi desenvolver uma ferramenta para análise de sentimentos de pessoas que utilizam a rede social Twitter a cerca das eleições presidenciais de 2018 no Brasil, analisando tweets em português com as bibliotecas de processamento de linguagem natural *TextBlob* e *NLTK*, disponível para Python.

3. DESENVOLVIMENTO

O projeto foi desenvolvido utilizando a linguagem Python, versão 3.6.7, e como IDE o Jupyter Notebooks, disponíveis na distribuição Anaconda, que já é preparada com pacotes de *data science* e *machine learning*.

Para coleta dos tweets foi utilizado a biblioteca *Tweepy*, uma biblioteca que facilita o acesso a API do Twitter através do Python. Foi utilizado a função de busca da API do Twitter para coleta das informações, foram solicitadas 1500 tweets para cada busca realizada. A API do Twitter limita essa busca ao período máximo de 1 semana.

Para a primeira análise de sentimentos foi utilizada o pacote de bibliotecas *TextBlob*. *TextBlob* é um pacote de bibliotecas para processamento de dados textuais disponíveis para Python 2 e 3. Ele fornece uma API simples para realizar tarefas comuns de processamento de linguagem natural, como análise de sentimento, classificação, tradução e muito mais. Apesar de ser poderosa, ela não dá suporte a língua portuguesa para análise de sentimentos. Inicialmente a ferramenta de tradução do *Textblob* foi utilizada para tradução dos tweets, mas como ele usa a API do Google Translator para isso, o número de requisição é bastante limitada, e pelo fato do número de tweets analisados serem grande, se tornou uma solução inviável. Para contornar esse problema, utilizamos um dicionário de palavras em português previamente classificadas como positivas ou negativas, assim foi possível treinar um modelo para a língua portuguesa.

Para criação do modelo de treinamento, foi utilizado o *NaiveBayesClassifier*, disponibilizado pelo próprio *textblob*. Como base para o treinamento, foi utilizado o *ReLi-Lex*, *ReLi* (REsenha de Livros) que foi criado no âmbito do projeto Anotadores Semânticos baseados em Aprendizado Ativo, do LEARN, coordenado por Ruy Milidiú (Departamento de Informática - PUC-Rio). Esse dicionário contém 1600 resenhas, de 14 livros diferentes, totalizando 12.470 frases e 259.978 palavras.

Outro pacote de bibliotecas utilizado no projeto foi o *NLTK* (*Natural Language Toolkit*), que é especializada em processamento de linguagem natural e simbólica. Além disso ela dá suporte a lingüística empírica, ciência cognitiva, inteligência artificial, recuperação de informação e machine learning. Apesar de dar suporte a língua portuguesa em algumas de suas funções, ela não está presente nas bibliotecas de análise de sentimentos, sendo assim um modelo de machine

learning foi treinado para ter suporte a língua portuguesa, utilizando-se também do dicionário *ReLi-Lex*. O modelo foi treinado usando o algoritmo de *MultinomialNB*.

Para melhor entendimento dos dados e resultados obtidos utilizamos alguns recursos de visualização disponíveis em bibliotecas de data science do python, os métodos escolhidos foram o de gráfico de pizza e de barras, nuvem de palavras e mapa de calor de pontos geográficos dos tweets coletados.

Gráfico de pizza e de barras - Através do pacote *Matplotlib* do Python foi possível criar um gráfico de pizza que representa a porcentagem de tweets positivos e negativos dos tweets, como ele representa apenas dois possíveis valores ele possui apenas duas cores distintas. Deste mesmo pacote ainda foi gerado um gráfico de barras que demonstra a quantidade de cada fonte dos tweets coletados.

Nuvem de palavras - Através do pacote *Wordcloud* do Python foi criado uma nuvem das principais palavras que apareceram nos nossos bancos de dados, utilizando de cores e tamanho das fontes para enfatizar as principais palavras. Como o trabalho foi feito sobre eleições no Brasil, decidimos utilizar o mapa do Brasil para servir de molde para a nossa nuvem, para isso importamos uma imagem png do mapa brasileiro com fundo branco. Também limpamos os tweets antes da criação da nuvem, retirando “*stopwords*” da língua portuguesa das frases.

Mapa de calor - Para criação deste mapa de calor dos tweets foi necessário a utilização de alguns pacotes do Python, visto que quase nenhum tweet vinha com sua geolocalização disponível (a maioria dos usuários não disponibiliza esses dados no momento de publicação do tweet). A alternativa encontrada foi coletar os dados em um campo do perfil do usuário, campo em que o usuário pode digitar da maneira que quiser sua localização, tornando os dados não muito confiáveis. Após a obtenção desses dados utilizamos a função *Nominatim* do pacote *Geopy.geocoders*, que busca o texto enviado como endereço na API do Google Maps, retornando, caso encontrado, a latitude e longitude do endereço. Após a obter o ponto no mapa utilizamos o pacote *Folium* do Python para marcação do ponto como mapa de calor.

4. RESULTADOS

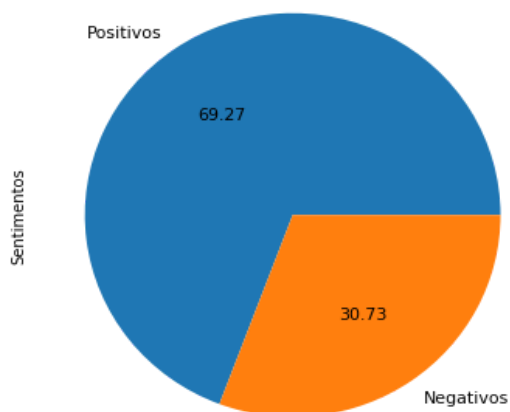
A seguir serão apresentados os resultados obtidos com a análise de tweets, baseados nas palavras chaves definidas para o projeto.

Palavras Chave: Lula Povo

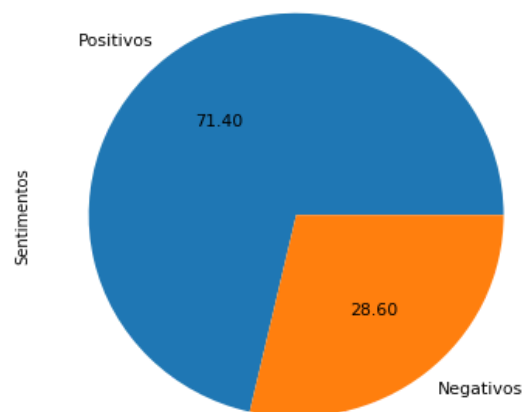
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Estão matando não só Lula, mas matando a todos nós brasileiros. Mito estão hipnotizados e não conseguem acordar para a realidade assassina que paira sobre o país. Não sejamos omissos. Sejam firmes e vamos insistir na liberdade de Lula e do povo brasileiro.	Lula tá preso , impeachment da Dilma agora me fala uma coisa : tu viu algum militar pagar pelas atrocidades que fizeram com o povo Brasileiro ? Entre fascismo e democracia me parece óbvio ! Votaria até no Temer contra o Fascistoide!
Número de curtidas: 304	Número de retweets: 11484

Análise de Sentimentos

TextBlob

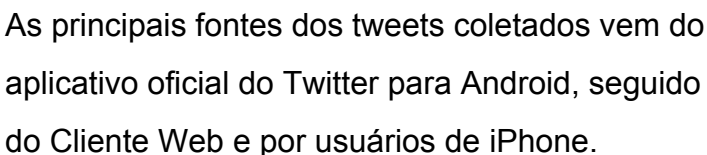


NLTK



A análise de sentimentos mostra positividade em relação ao termo “Lula Povo”. A maioria das pessoas vê Lula e seu governo como positivo, e tendem a falar coisas

Principais Fontes dos Tweets



Word cloud visualization of tweets related to the 2018 Brazilian election. The most prominent words are 'Lula', 'povo', 'representa', 'dignidade', 'povo brasileiro', 'primeira vez', 'Lula livre', 'conheceu', 'matem', 'fome', 'Brasil', and 'natural'. The words are arranged in a circular pattern, with 'Lula' and 'povo' being the largest. The colors range from light blue to dark blue.

The map illustrates the geographical distribution of the dengue virus across South America. The color scale indicates the density of the virus, with red representing the highest density and blue representing the lowest. The most significant hotspots are located in Brazil, especially in the southeastern and northeastern regions, and in Colombia. Other countries like Peru, Bolivia, Paraguay, Argentina, and Uruguay show much lower densities, indicated by blue and green colors. Major cities such as Lima, Bogotá, and Buenos Aires are marked with dots.

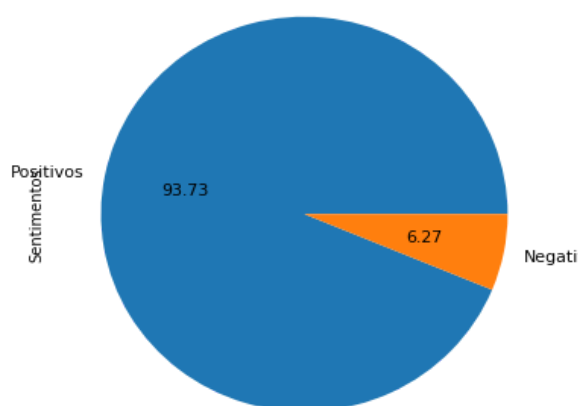
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, sendo a maioria delas concentrados no sul, sudeste e no litoral e norte do nordeste.

Palavras Chave: STF Lula

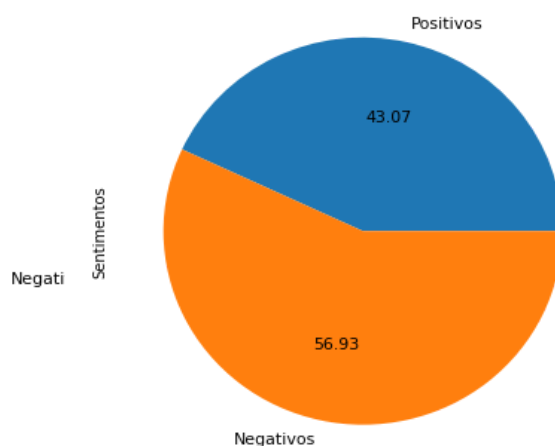
O Tweet Mais Curtido	O Tweet Mais Compartilhado
General Villas Bôas confessou pressionar STF para manter a prisão política do LULA e q o exército ameaçou intervenção no país caso ele fosse solto. DITADURA MILITAR É O NOME DISSO! começou assim, disfarçada de “justiça” e pode terminar com vc preso só por compartilhar isso aqui.	O Adv do MBL, o mesmo que tirou os direitos políticos do Lula @RubinhoNunesMBL já entrou com ação popular contra o aumento do salário do STF. #VetaTemer
Número de curtidas: 6114	Número de retweets: 3491

Análise de Sentimentos

TextBlob



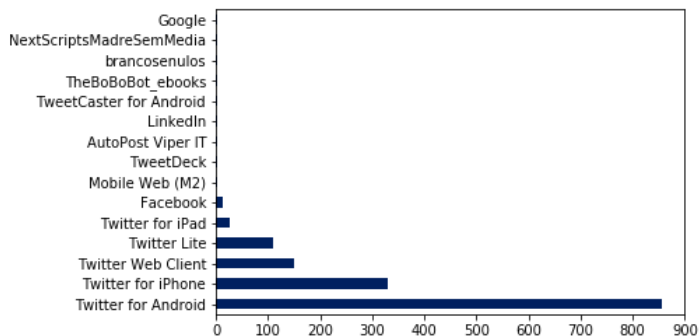
NLTK



A análise de sentimentos entre as duas ferramentas divergiram completamente para a busca com as palavras chave “Lula STF”. Enquanto o *Textblob* classificou a grande maioria dos tweets como positivo, o *NLTK* classificou um pouco mais que a

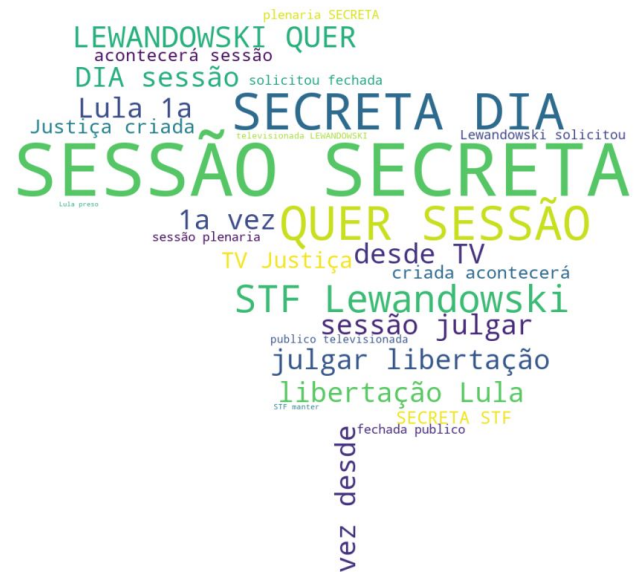
metade como negativos. Analisando os tweets coletados podemos notar que o *NLTK* fez as melhores estimativas de sentimento para o assunto buscado.

Principais Fontes dos Tweets

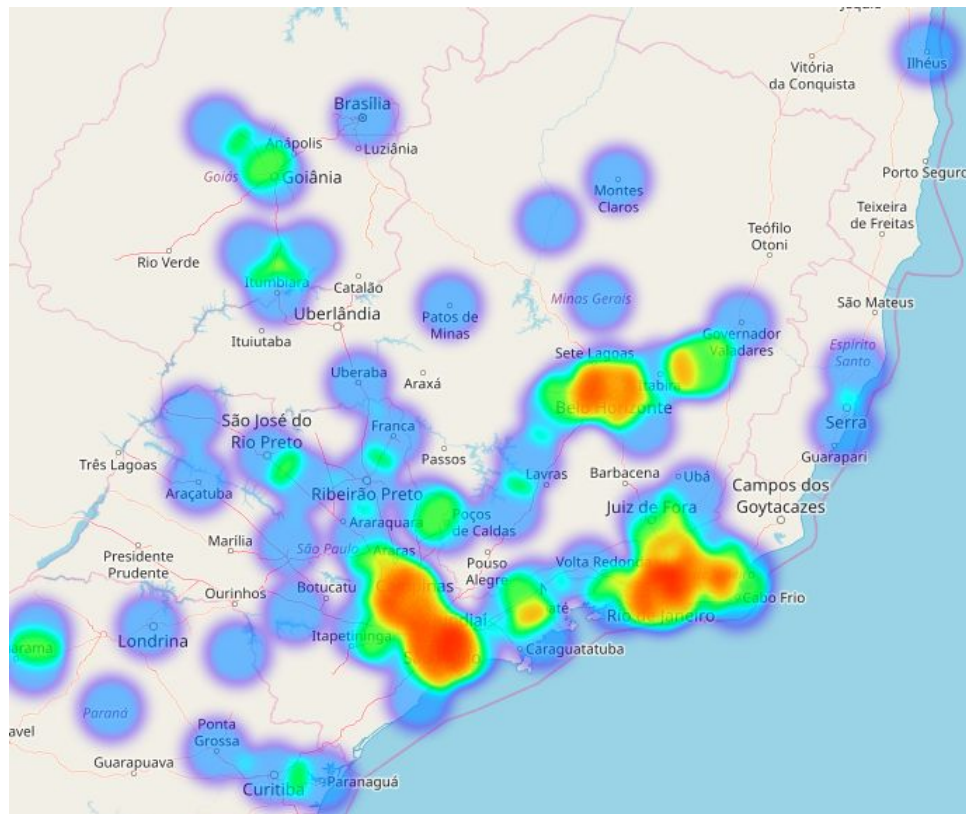


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



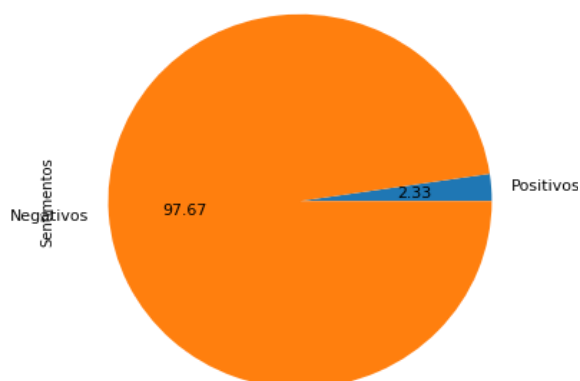
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, um zoom na região sudeste do Brasil nos mostra a concentração de pessoas nessa região que tuitaram sobre o assunto buscado, é possível ver a concentração de pessoas nas capitais e regiões metropolitanas dessa área.

Palavras Chave: Lula Haddad

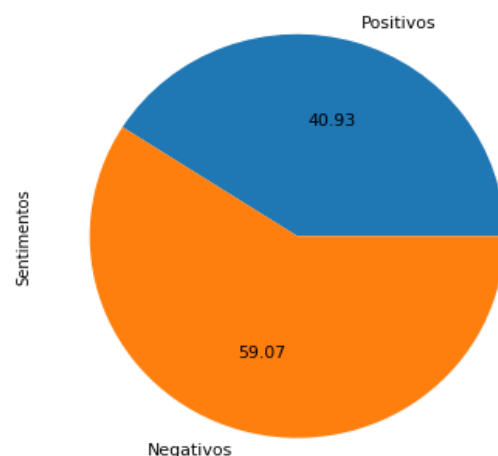
O Tweet Mais Curtido	O Tweet Mais Compartilhado
<p>“Estive hoje pela manhã com o presidente Lula e ele está tranquilo e sereno, com a certeza de que será absolvido se tiver um julgamento justo” - @Haddad_Fernando, na chegada à Justiça Federal para acompanhar depoimento de Lula. #LulaLivre https://t.co/yCaPHszrtm</p>	<p>JOSÉ DIRCEU diz que o PT "cometeu erros". LULA diz que o PT "cometeu erros". HADDAD diz que o PT "cometeu erros". A ESQUERDA diz que o PT "cometeu erros". Não, o PT não "cometeu erros". O PT cometeu CRIMES! Não podemos deixar que esse discurso mole de "erros" prevaleça.</p>
Número de curtidas: 1302	Número de retweets: 2154

Análise de Sentimentos

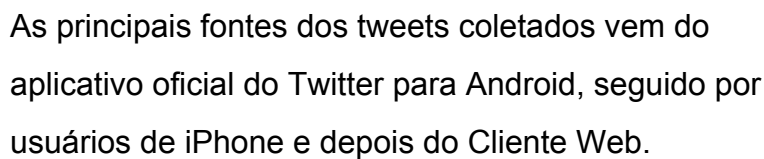
TextBlob



NLTK



Principais Fontes dos Tweets



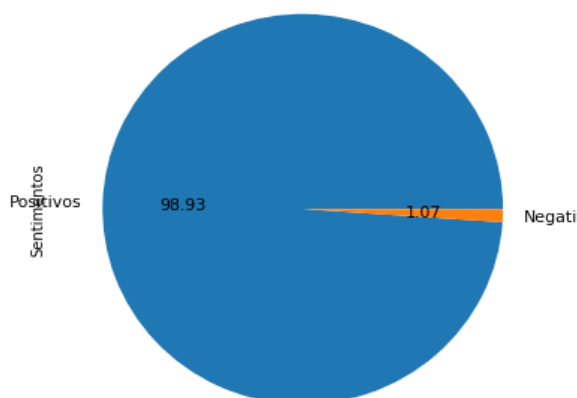
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, um zoom na região nordeste, reduto eleitoral do PT mostra as cidades com maior concentração de usuários do twitter que citaram Lula e Haddad. É possível verificar a concentração de tweets nas grandes cidades próximo ao litoral, enquanto a parte central da região tem poucos ou nenhum usuário declarado.

Palavras Chave: General Bolsonaro

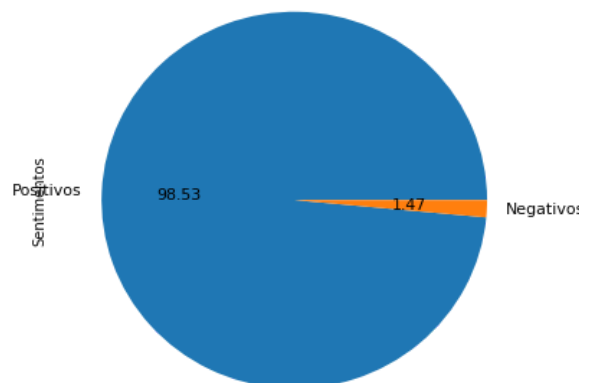
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Até onde chega a estupidez humana que não há limites. General Mourão, o vice: “os livros contarão a história do homem que libertou 8 mil cubanos da escravidão. Jair Bolsonaro”	Ministros do Governo Bolsonaro: 1 Casa Civil: Onyx Lorenzoni 2 Justiça: Sergio Moro 3 Defesa: General Fernando Azevedo 4 Relações Exteriores: Ernesto Araújo 5 Economia: Paulo Guedes 6 GSI: General Heleno 7 Ciência e Tecnologia: Marcos Pontes 8 Agricultura: Teresa Cristina
Número de curtidas: 569	Número de retweets: 2073

Análise de Sentimentos

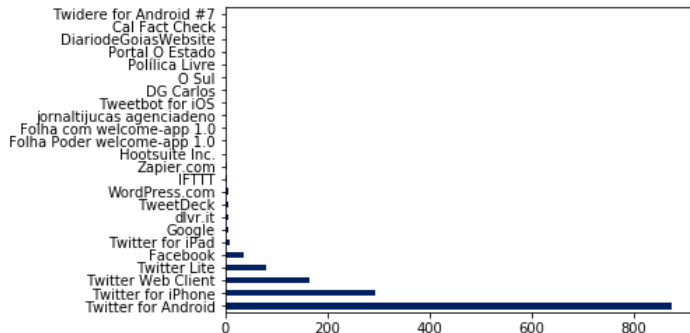
TextBlob



NLTK



Principais Fontes dos Tweets

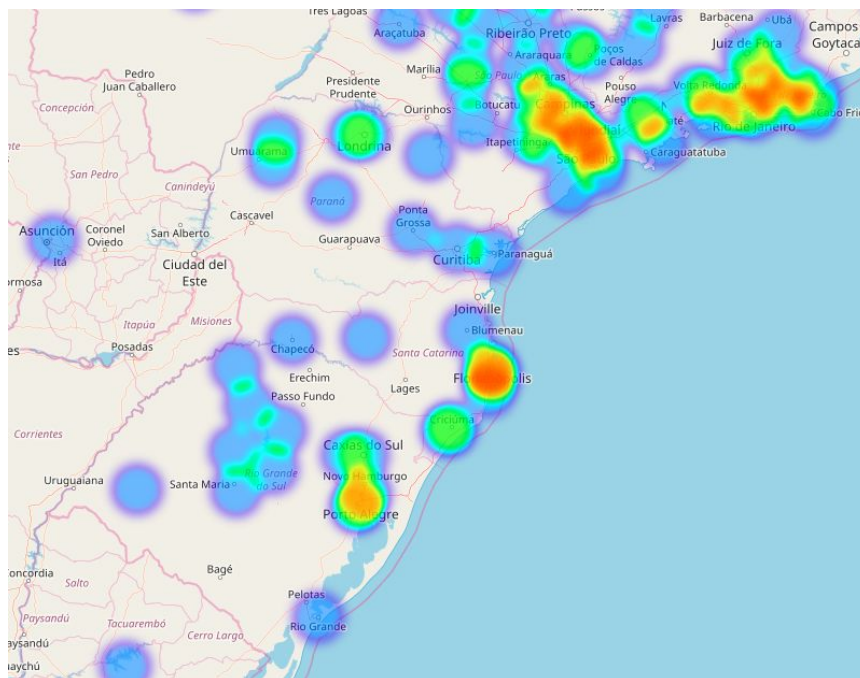


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



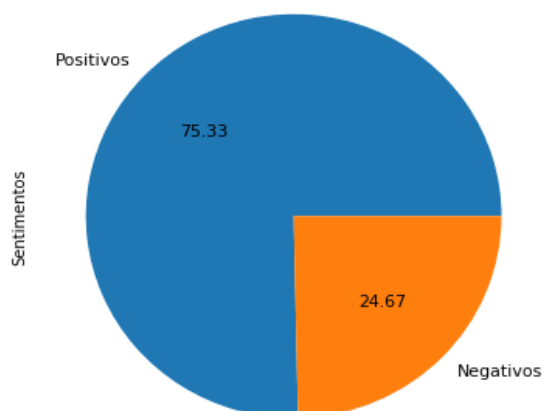
Distribuição Geográfica dos Usuários



O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, um zoom na região sul do brasil mostra a grande concentração de usuários nas grandes cidades próximo ao litoral, e também usuários espalhados pelos estados.

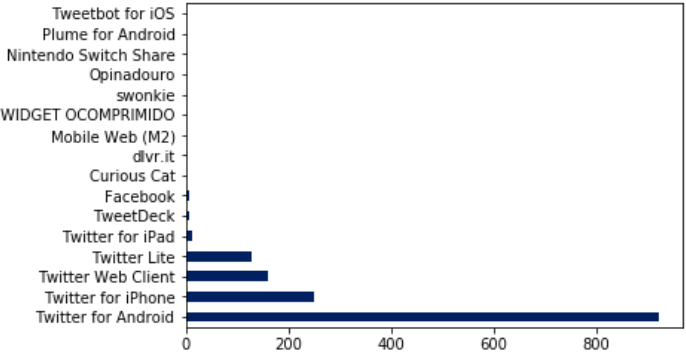
O Tweet Mais Curtido	O Tweet Mais Compartilhado
<p>@lsentoes engraçado que ele foi o primeiro a espalhar a mentira da Folha dizendo que bolsominion espalhava fake news no zap. A casa caiu bonito kkkkkkkkkkkkkkkkkkkkkkkkkkkkkkk</p>	<p>O TSE anda está analisando a questão dos 15 empresários que doaram R\$ 12,5 milhões, cada um, para pagar 40 mil grupos no WhatsApp para espalhar fake news contra Haddad. Se o TSE condenar, a chapa toda é cassada. https://t.co/uLwTxTwhjs</p>
Número de curtidas: 312	Número de retweets: 338

TextBlob



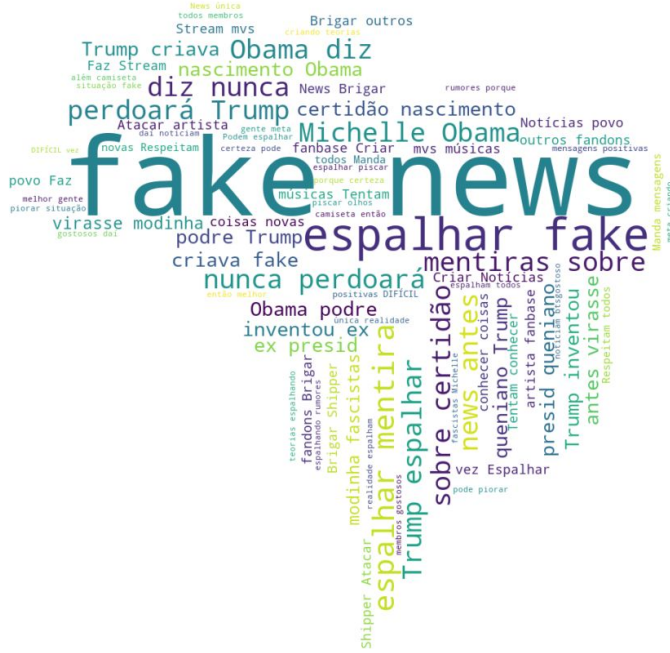
As duas ferramentas de análise de sentimento retornaram valores próximos para a busca “Espalhar Fake News”. A maioria dos sentimentos são positivos, de pessoas

Principais Fontes dos Tweets

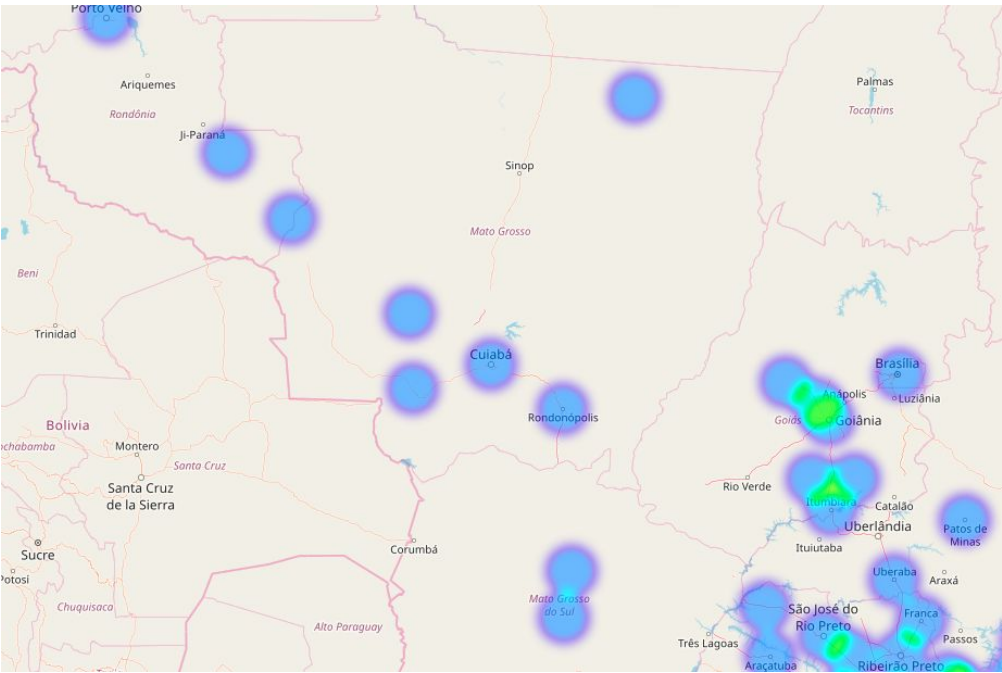


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, um zoom na região centro-oeste do país mostra pouca concentração de usuários, mesmo em grandes cidades para todas as buscas realizadas. As regiões de

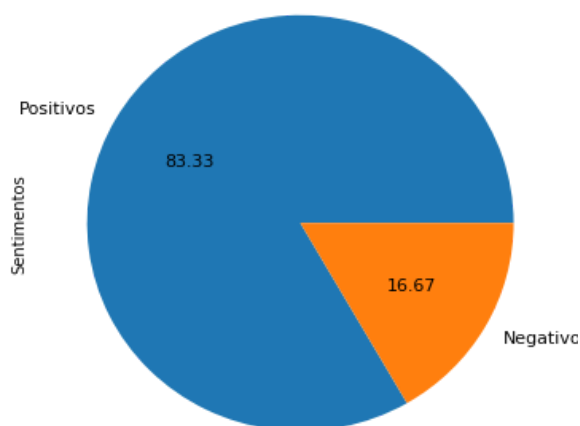
Goiânia, Capital de Goiás e a região de Itumbiara concentram o maior número de usuários.

Palavras Chave: Socialismo Haddad

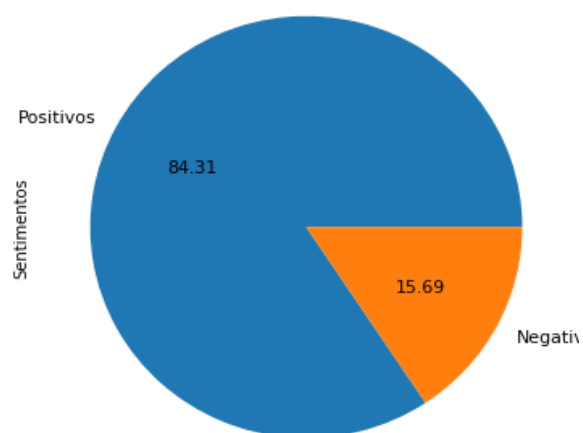
O Tweet Mais Curtido	O Tweet Mais Compartilhado
@NARSC77 @dudunaweb Boa tarde, o Haddad perdeu por causa do rancor e das Fake News. As eleições brasileiras foram marcadas pelo medo. Medo do socialismo, medo da permanência do pt (pt estava no poder há 13 anos)Bolsonaro se apresentou como o "novo" como o "salvador da Pátria" aliado a isso crise.	- Lula é inocente. - Não existe provas contra Lula. - Haddad não criou o Kit Gay. - Haddad e Manuela continuam indo a Missa. - PT é o partido mais honesto do mundo. - Comunismo e Socialismo funcionam. - Venezuela do Maduro é uma Democracia. #FaçaSuaFakeNews
Número de curtidas: 165	Número de retweets: 4

Análise de Sentimentos

TextBlob



NLTK



As duas ferramentas de análise de sentimento retornaram valores próximos para a busca "Socialismo Haddad". O Twitter retornou para essa busca apenas 101 tweets, a maioria dos sentimentos desses tweets são positivos, apesar do conteúdo da busca. Isso

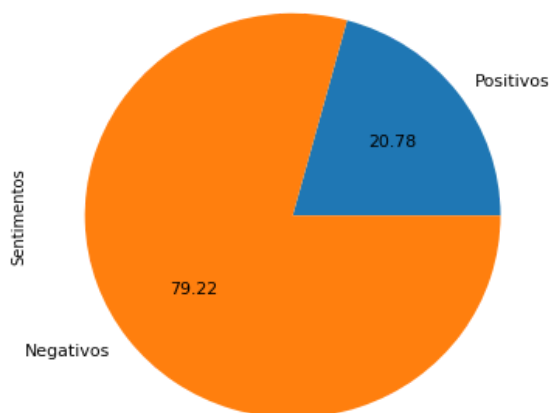
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, nele é possível perceber que a maior concentração de usuários para essa busca se encontram na região sudeste do país.

Palavras Chave: Ditadura Haddad

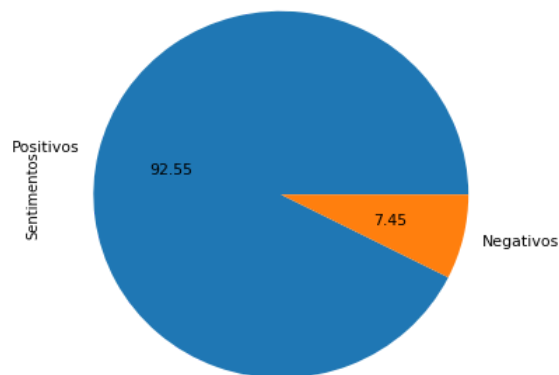
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Coisas que já caíram desde que Bolsonaro foi eleito: - o dólar - a frequência da Manu nas missas - a máscara os esquerdistas que se diziam contra ditadura - a audiência do Amor & Sexo Coisas que ainda não caíram: - os depósitos da vaquinha do Haddad	Haddad diz que sou responsável pela campanha mais baixa da história. Logo ele, que é orientado por um presidiário, esconde as cores do partido, finge ser religioso, joga bíblia no lixo, esconde apoio à ditadura venezuelana e espalha um monte de porcaria mentirosa ao meu respeito.
Número de curtidas: 3225	Número de retweets: 24672

Análise de Sentimentos

TextBlob

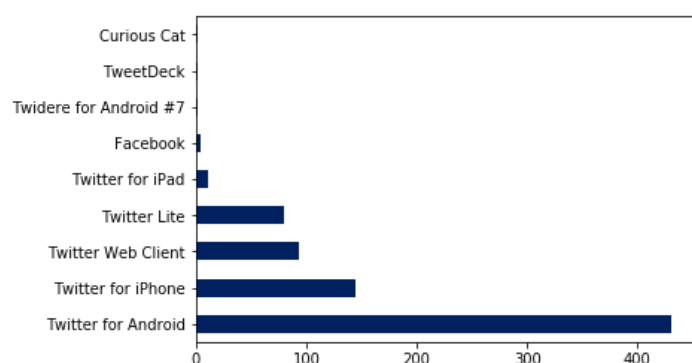


NLTK



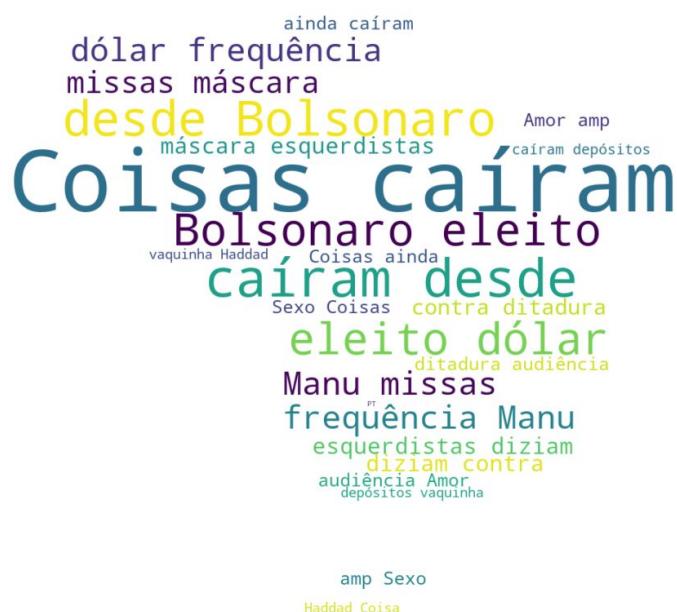
As duas ferramentas de análise de sentimento retornaram valores divergentes para a busca “Ditadura Haddad”. O Twitter retornou para essa busca apenas 750 tweets. Um determinado tweet aparece repetido muitas vezes nessa consulta, pois foi retuitado por vários usuários diferentes, umas das ferramentas avaliou o texto como negativo, enquanto o outro avaliou como positivo, o que levou a essa divergência. Avaliando os tweets podemos dizer que o *NLTK* se saiu melhor na classificação, pois a maioria das pessoas falam da “vitória sobre a ditadura”, e o “fim dela”.

Principais Fontes dos Tweets

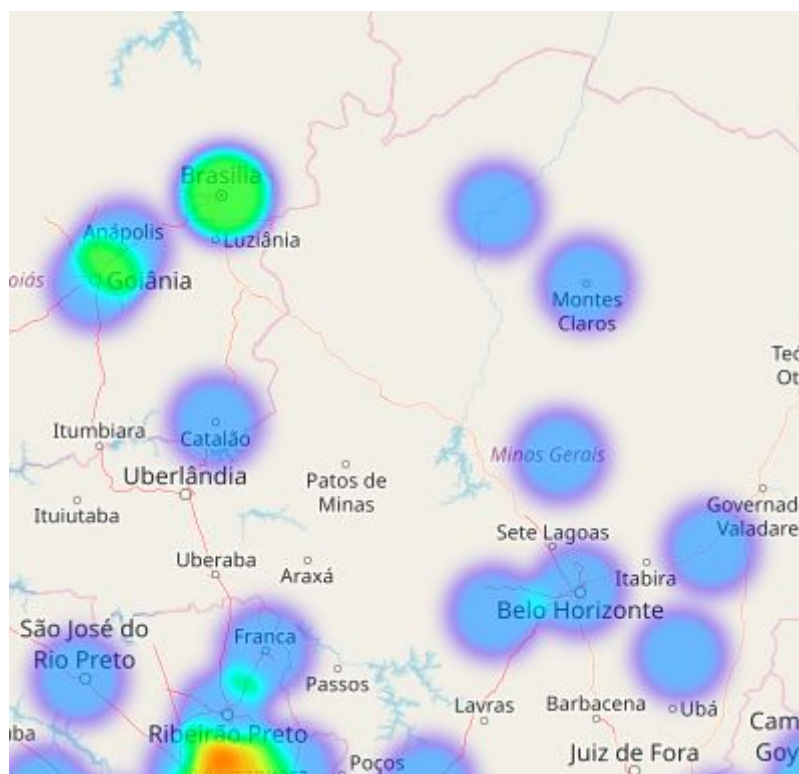


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido pelos usuários de iPhone e depois pelo Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



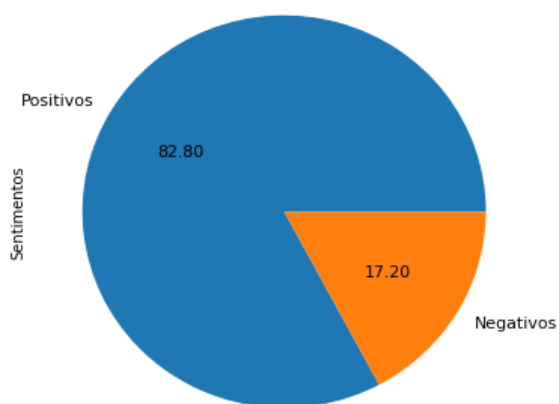
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, com um zoom no estado de .minas gerais, podemos ver a maior concentração de pessoas se localizam em cidades grandes, como a capital Belo Horizonte, Governador Valadares e Montes Claros.

Palavras Chave: Lula Livre

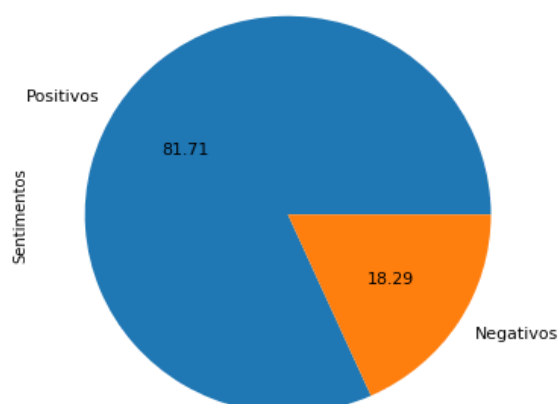
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Lula Livre acaba por falta de gente https://t.co/YczuTXhLQ5	Se o ENEM permanecer com o atual nível de exigência, em breve, o português será um dialeto. As necessidades dos alunos se resumirão em grunhir, e dominar os preceitos marxistas sobre disforia sexual, homossexualismo, males do capitalismo, racismo, fascismo, Lula Livre et cetera.
Número de curtidas: 2880	Número de retweets: 3362

Análise de Sentimentos

TextBlob

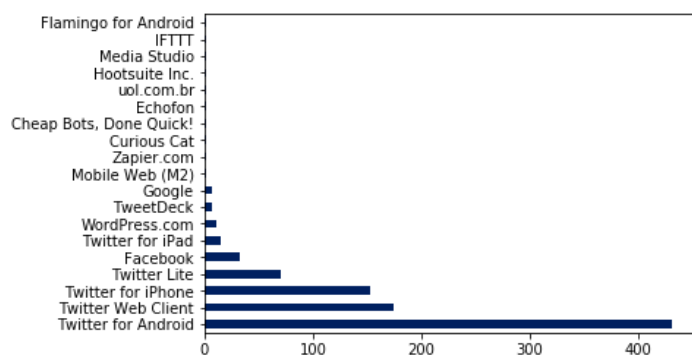


NLTK



As duas ferramentas de análise de sentimento retornaram valores próximos para a busca “Lula Livre”. A maioria dos sentimentos são positivos, de pessoas que apoiam o ex-presidente e acham ele inocente. Como é possível de ver na nuvem de palavras, Lula Livre é a frase mais comum entre os tweets.

Principais Fontes dos Tweets

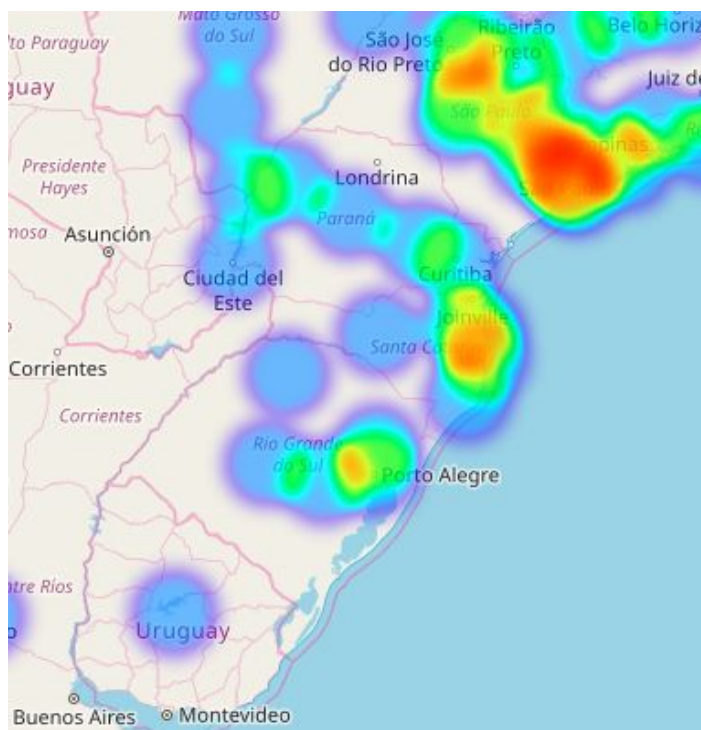


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



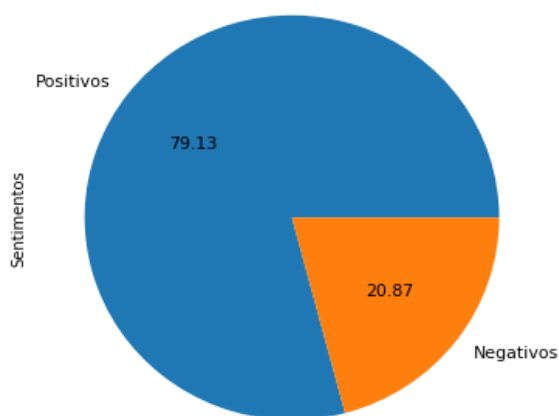
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, além da concentração de tweets em São Paulo, podemos ver também uma grande concentração próximo a Curitiba, onde o ex-presidente Lula se encontra preso.

Palavras Chave: Lula Bolsonaro

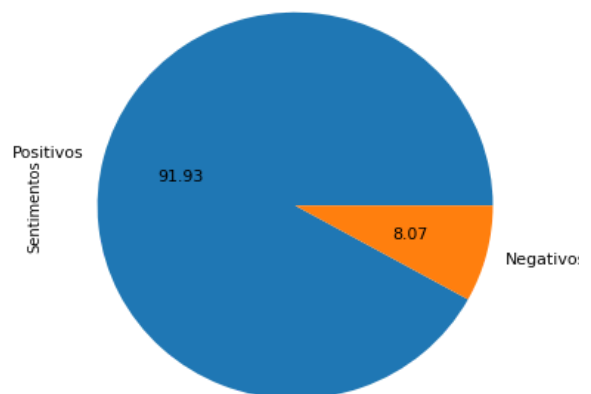
O Tweet Mais Curtido	O Tweet Mais Compartilhado
O ministro Luís Roberto Barroso disse, durante um evento no exterior, que Lula não é um perseguido político e que Bolsonaro não representa ameaça à democracia. Todo mundo sabe isso no Brasil. E cada vez menos pessoas acreditam nesta lorota contada pelos esquerdistas.	Frota, ator pornô fala em Moral. Ônix, Deputado corrupto, fala em honestidade. Janaína, ajudou a dar um Golpe, fala em Constituição. Bolsonaro, elogia o torturador Ustra, fala em Democracia. Moro, q condenou Lula, fala em Imparcialidade. Vcs estão delirando e acreditando nisso?
Número de curtidas: 3414	Número de retweets: 15408

Análise de Sentimentos

TextBlob

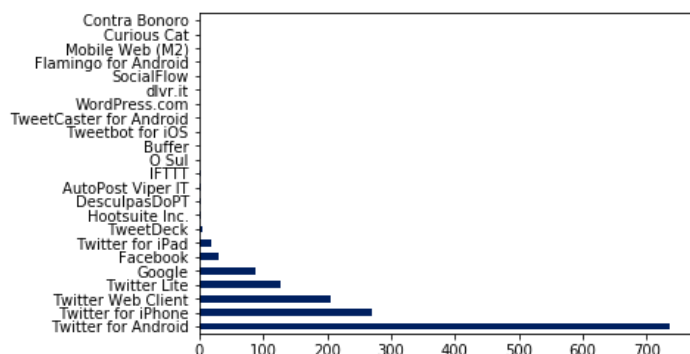


NLTK



As duas ferramentas de análise de sentimento divergiram pouco para a busca “Lula Bolsonaro”. A maioria dos sentimentos são positivos, os tweets são relacionados ao ex-presidente lula ser inocente ou não, e relação do governo de Bolsonaro com Sergio Moro.

Principais Fontes dos Tweets

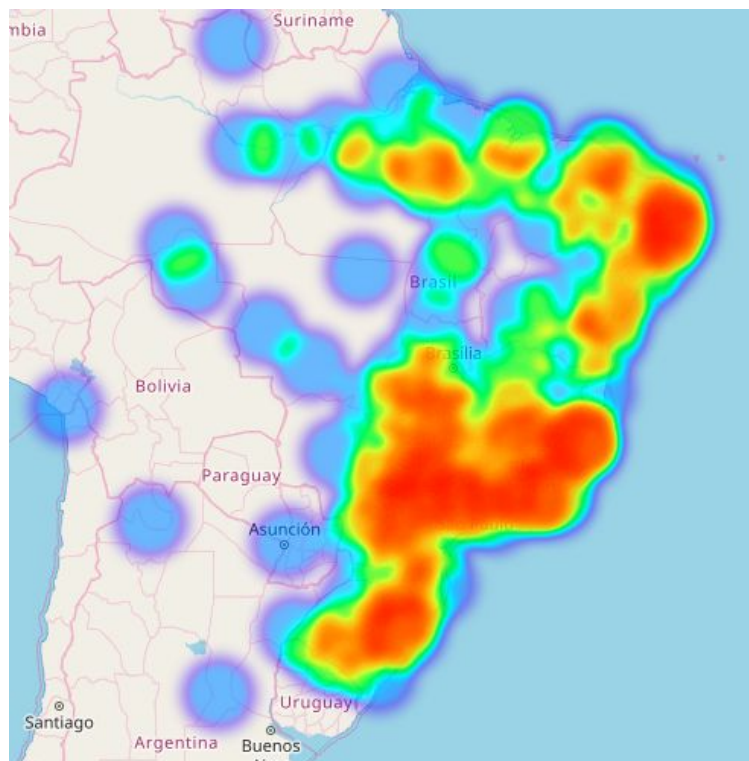


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



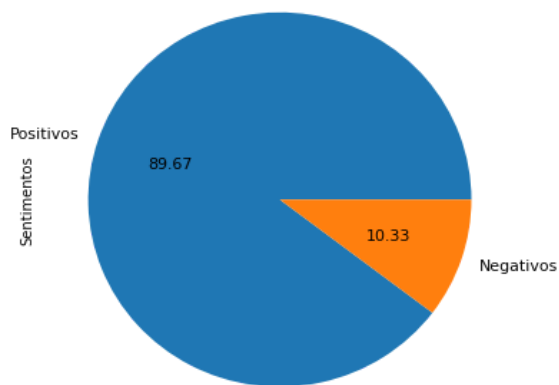
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, ele mostra a concentração de usuários espalhados pelo Brasil que tuitaram sobre o termo buscado, com grande concentração de pessoas do sul, sudeste e nordeste do Brasil.

Palavras Chave: Lula Ladrão

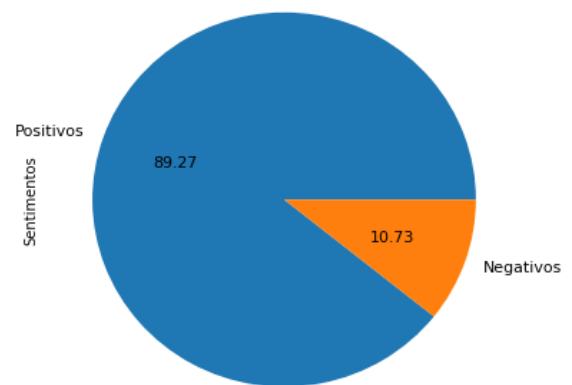
O Tweet Mais Curtido	O Tweet Mais Compartilhado
"Nunca foi tão fácil ser ladrão nesse país." (Lula, 14/11/2018)	"Nunca foi tão fácil ser ladrão nesse país." (Lula, 14/11/2018)
Número de curtidas: 15017	Número de retweets: 3662

Análise de Sentimentos

TextBlob

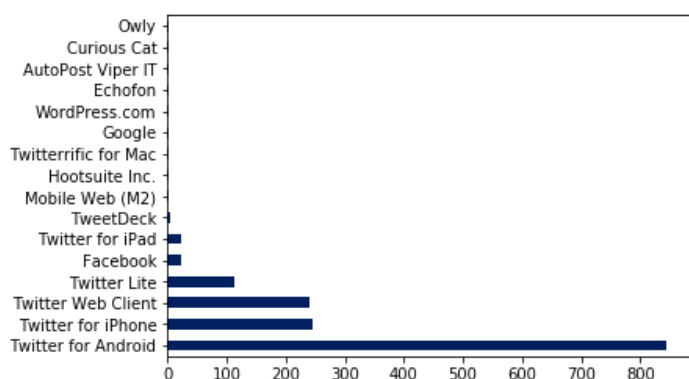


NLTK



As duas ferramentas de análise de sentimento convergem quanto aos resultados para a busca “Lula Ladrão”. A maioria dos sentimentos são positivos, baseados no sarcasmo das pessoas e pela frase que dominou essa busca nessa semana: “Nunca foi tão fácil ser ladrão no Brasil”. Apesar do sarcasmo, os algoritmos entenderam isso como um sentimento positivo.

Principais Fontes dos Tweets

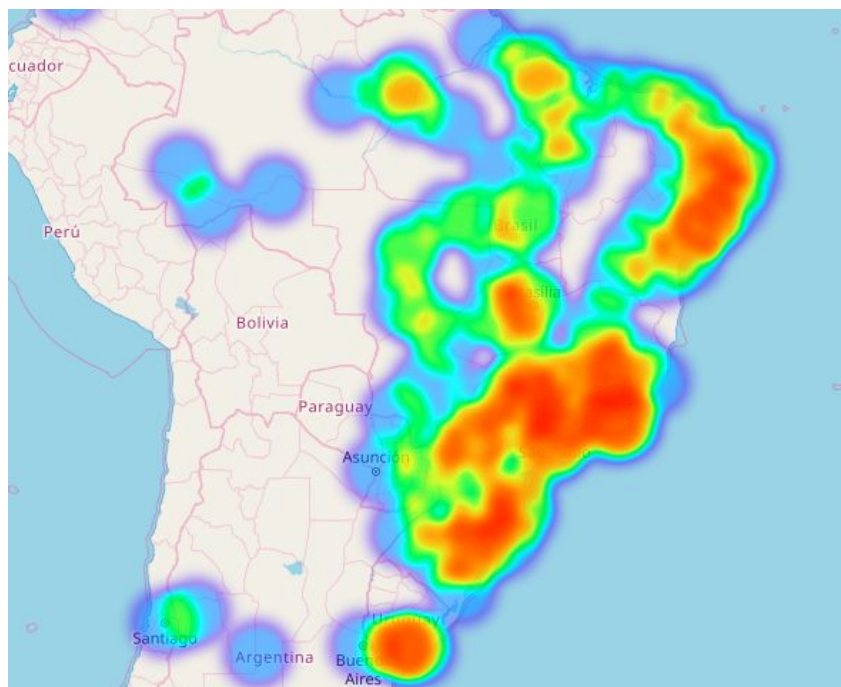


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Cliente Web.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



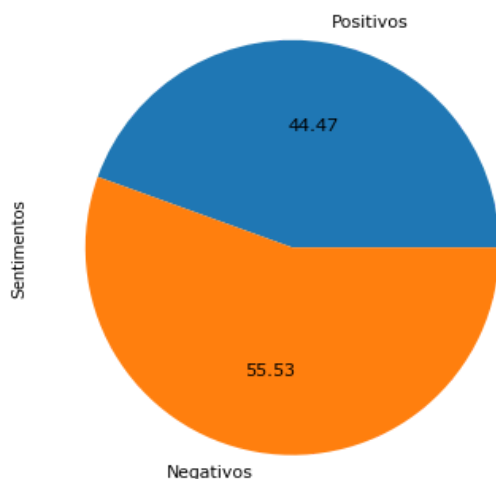
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, dessa vez ele mostra uma grande interação das pessoas em todas as regiões do Brasil sobre o assunto.

Palavras Chave: Ditadura Bolsonaro

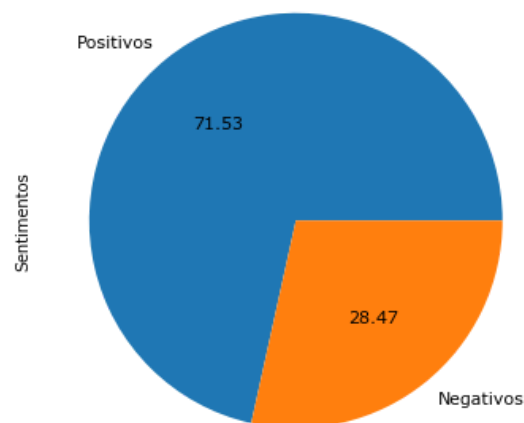
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Não foi Bolsonaro que abandonou o Mais Médicos. Foi Cuba, indisposta a (1) tratar seus profissionais como trabalhadores em vez de escravos a serviço da ditadura; e (2) a esperar que o novo governo se estabeleça para retirá-los. Cuba tenta pegar país desprevenido para gerar crise.	o dia é 1 de janeiro de 2019 bolsonaro assume que na verdade é gay todo o discurso de ódio foi proposital pra ter os votos dos preconceituosos e mostrar como o ódio se espalha rápido seu vice na verdade é a pablo vittar a ditadura gay é implantada no brasil
Número de curtidas: 29874	Número de retweets: 13859

Análise de Sentimentos

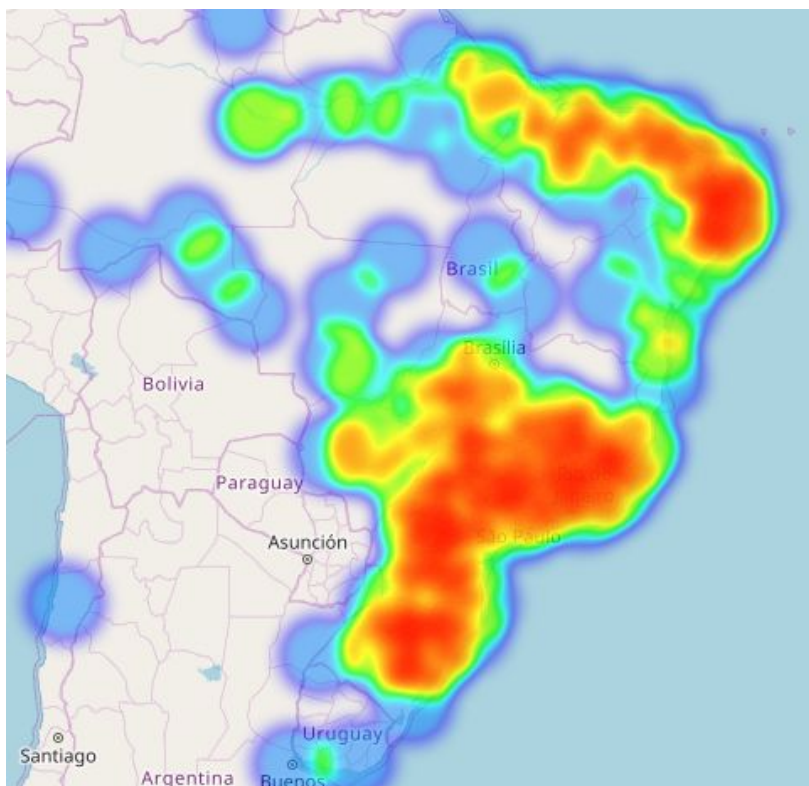
TextBlob



NLTK



As duas ferramentas de análise de sentimento divergem quanto aos resultados para a busca “Ditadura Bolsonaro”. Enquanto o *TextBlob* demonstrou que a maioria dos tweets é negativo o *NLTK* afirmou que a grande maioria é positiva. Acreditamos que o resultados divergiram pois os métodos não distinguem para quem é positivo ou negativo. Como o assunto mais comentado durante essa semana foi da retirada dos médicos cubanos do Brasil, como se pode visualizar no mapa de palavras e no tweet mais curtido,



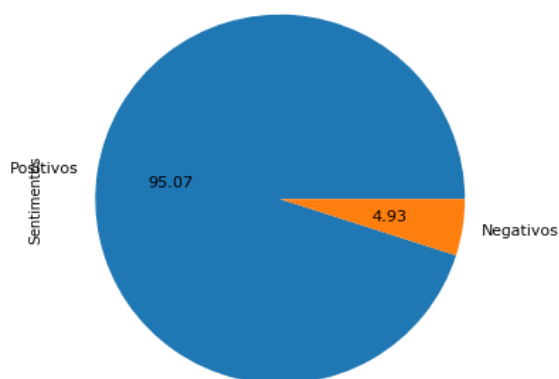
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, ele mostra uma grande interação das pessoas nas regiões sul, sudeste e parte do centro-oeste do Brasil sobre o assunto.

Palavras Chave: Fascismo Bolsonaro

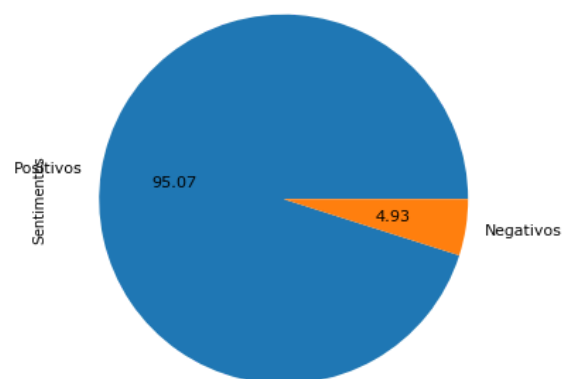
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Spike Lee: “O fascismo usa o medo. Não é só Trump, Bolsonaro é igualmente ruim” Cultura EL PAÍS Brasil https://t.co/gS8S9zLgju	Ver gente boicotando o Teleton por causa de política é asqueroso. A eleição já passou, podem tirar essa máscara de fingimento e parar de usar esses termos de doente mental, como "fascismo", "tortura" e etc. Bolsonaro será presidente do país por 4 anos (no mínimo). CRESÇAM!
Número de curtidas: 671	Número de retweets: 2320

Análise de Sentimentos

TextBlob



NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Fascismo Bolsonaro” obtendo dois resultados iguais. Acreditamos que este resultado ocorreu devido ao grande número de retuïtes do tweet mais compartilhado, em

O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, com um zoom na região sudeste, podemos ver a maior concentração de

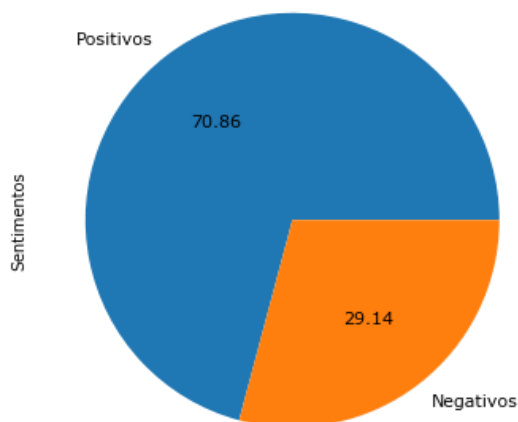
pessoas se localizam em cidades grandes, como as capitais Belo Horizonte, São Paulo e Rio de Janeiro.

Palavras Chave: Medo Bolsonaro

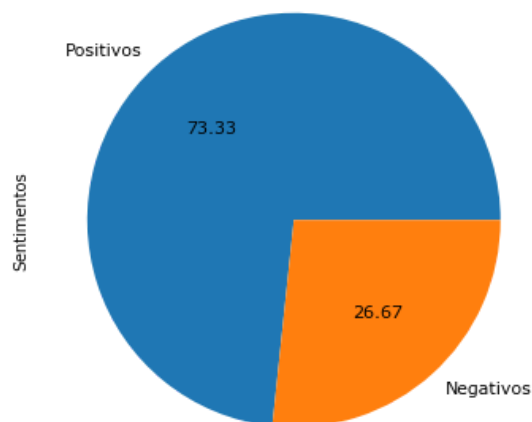
O Tweet Mais Curtido	O Tweet Mais Compartilhado
(Des) governo Bolsonaro terá Ministro das Relações ANTERIORES: o indicado ainda acha que a China é "maoísta", que a globalização é marxista e que todo o planeta tem que ser "ocidental e cristão" (à moda dele). O cara que será a cara do Brasil no mundo tem medo do mundo!	O SBT está usando o lema da ditadura militar: "Brasil, ame-o ou deixe-o". A mesma imprensa que foi censurada na ditadura revela um saudosismo masoquista na tentativa de agradar Bolsonaro. O medo de perder verba publicitária é maior que a dignidade.
Número de curtidas: 1626	Número de retweets: 10311

Análise de Sentimentos

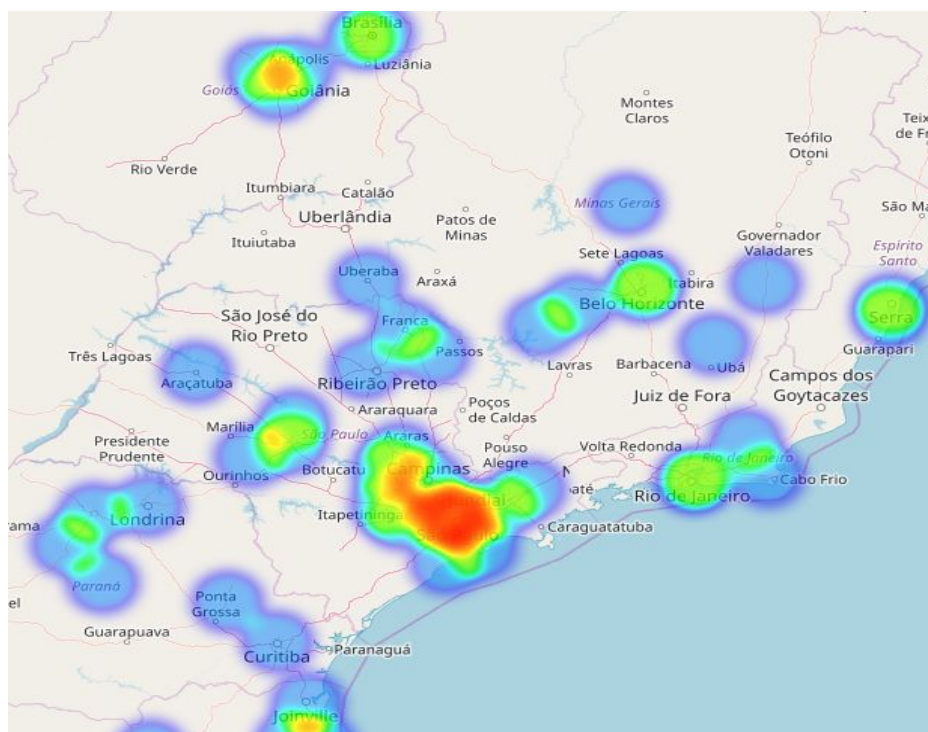
TextBlob



NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Medo Bolsonaro” obtendo dois resultados bem próximos, para essa busca menos tweets foram encontrados, cerca de 530 tweets. Os resultados obtidos



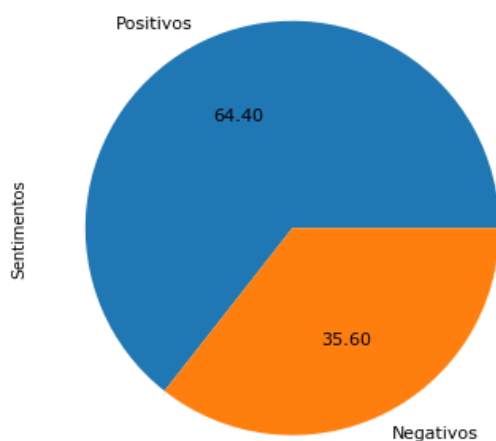
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, com um zoom na região sudeste, podemos notar uma maior concentração na região de São Paulo em comparação com outros mapas gerados anteriormente, isso pode ter acontecido devido a São Paulo possuir um dos maiores índices de casamentos homoafetivo do Brasil.

Palavra Chave: Eleições

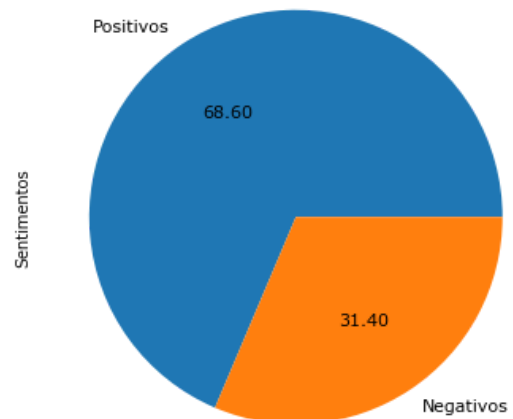
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Bom dia. Passadas as eleições o candidato fujão eleito por fake news até churrasco já anda fazendo. Pela agenda dele fica clara a manobra para nao participar de debates. Mentiou ao povo.	Após as eleições, grandes empresas já anunciaram milhões em investimentos no Brasil nos próximos anos. É só o começo! Comércio com o mundo todo sem viés ideológico + Redução de impostos + Desburocratização = Mais confiança, mais investimentos e mais empregos.
Número de curtidas: 774	Número de retweets: 17973

Análise de Sentimentos

TextBlob

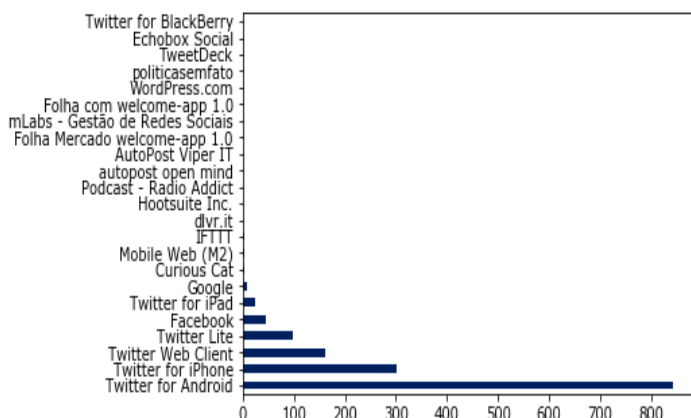


NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Eleições” obtendo dois resultados bem próximos. Os resultados obtidos condiz com os resultados esperados pelos tweets mais compartilhados, pois podemos notar uma otimismo para economia e geração de empregos das pessoas após o fim das eleições, assim os métodos avaliaram a maioria dos tweets como positivos.

Principais Fontes dos Tweets

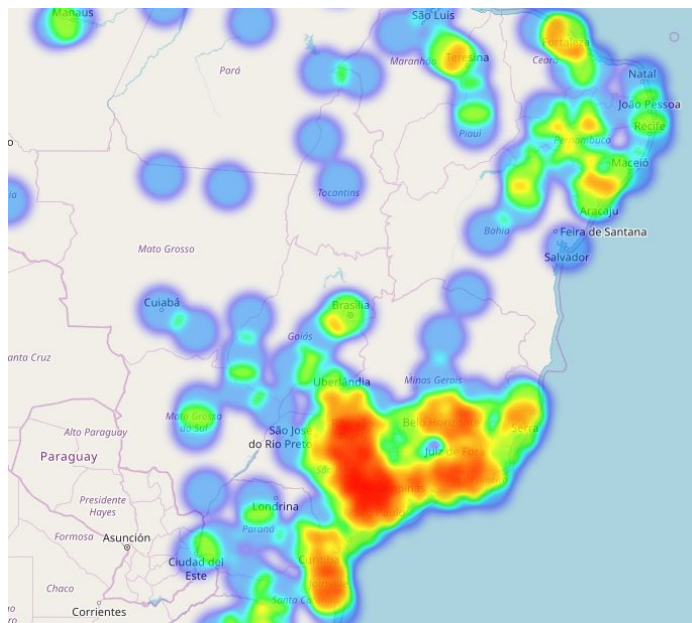


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Twitter Web Client.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



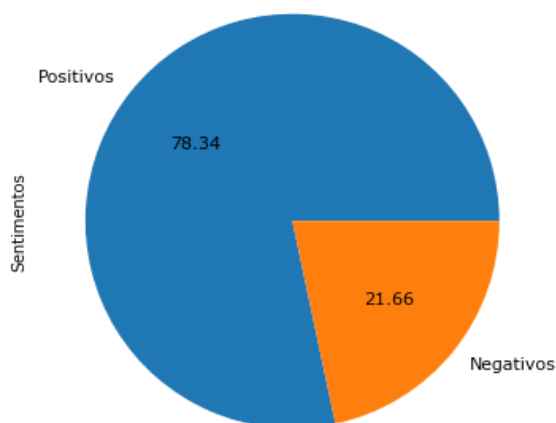
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, podemos notar uma maior concentração nas principais cidades do Brasil, principalmente na região Sudeste.

Palavra Chave: Voto Bolsonaro

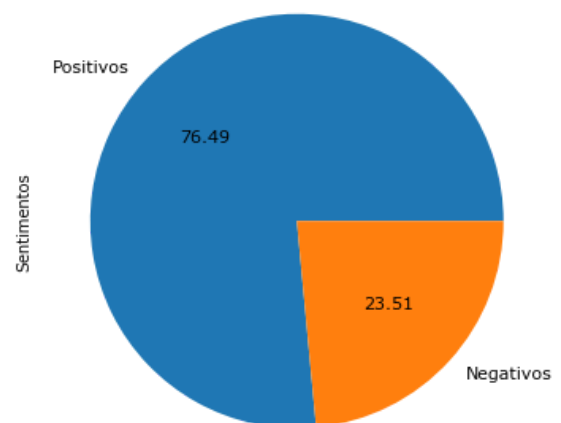
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Amoedo acaba de dizer em entrevista que votou em Bolsonaro. O Brasil não tem um partido com destaque no cenário e que seja de fato liberal. O Novo é economicamente neoliberal, fica em cima do muro nos costumes e sua principal liderança declara voto no ultra-conservador do PSL.	Em suas redes sociais, Joaquim barbosa divulga voto em Haddad, mas já está na história que ele mesmo disse que só Bolsonaro não foi comprado pelo PT no esquema de corrupção conhecido como Mensalão, que feria gravamente a democracia do nosso país anulando o Poder Legislativo. https://t.co/79dRHsvGwH
Número de curtidas: 3023	Número de retweets: 16526

Análise de Sentimentos

TextBlob



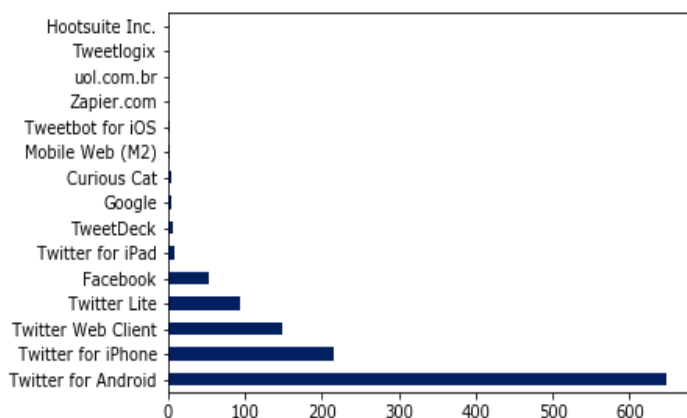
NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Voto Bolsonaro” obtendo dois resultados bem próximos, sendo que esta busca gerou 1191 tweets. Os resultados obtidos refletem um cenário que pode ser observado na nuvem de palavras, em que os usuários fizeram tweets demonstrando orgulho de Bolsonaro ter apoiado a campanha de doações da Telecon. Sendo assim, a maioria dos tweets foram considerados positivos pelos dois métodos.

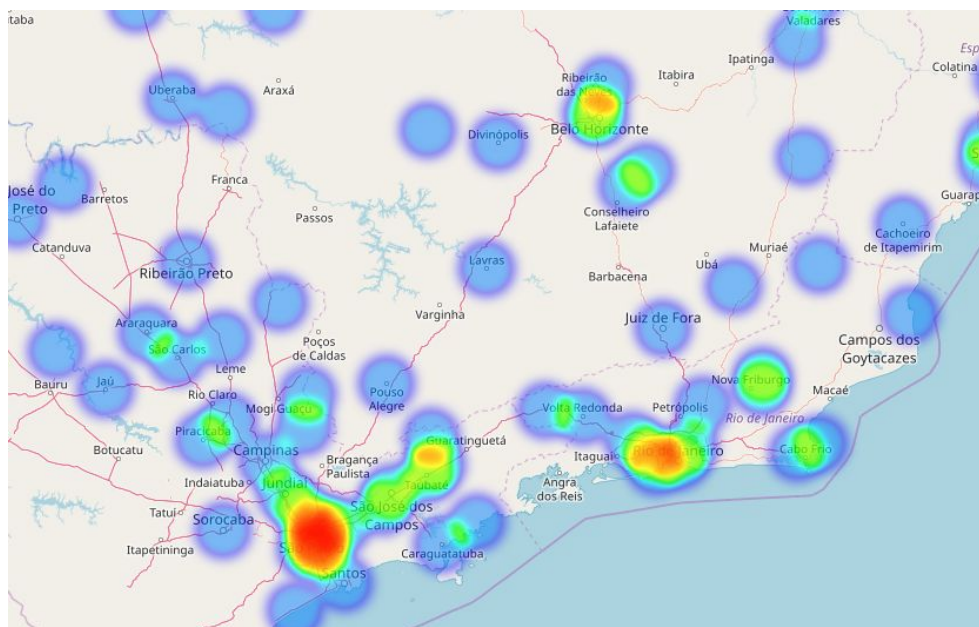
Principais Fontes dos Tweets

Nuvem de Palavras



As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois do Twitter Web Client.

Distribuição Geográfica dos Usuários



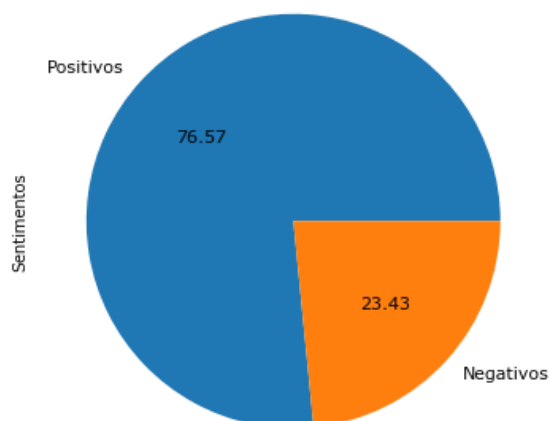
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, com um zoom na região sudeste podemos notar uma maior concentração nas principais cidades do Brasil e seus vizinhos locais.

Palavra Chave: Voto Haddad

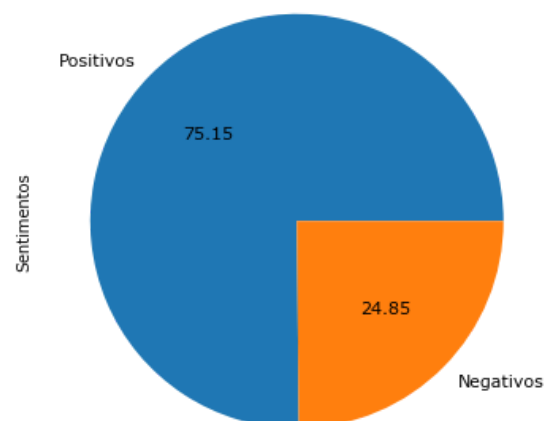
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Tipo declarar voto no haddad com discurso requeentado das minorias etc https://t.co/iNHyzIGrt7	Votar é fazer uma escolha racional. Eu, por exemplo, sopesei os aspectos positivos e os negativos dos dois candidatos que restam na disputa. Pela primeira vez em 32 anos de exercício do direito de voto, um candidato me inspira medo. Por isso, votarei em Fernando Haddad.
Número de curtidas: 391	Número de retweets: 106872

Análise de Sentimentos

TextBlob



NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Voto Haddad” obtendo dois resultados bem próximos, sendo que esta



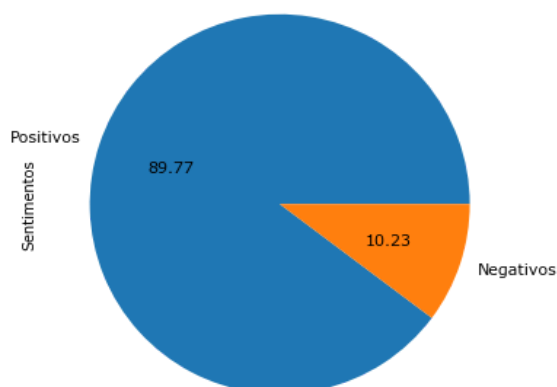
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, podemos observar que mesmo os tweets sendo positivos para o candidato Haddad, a maior concentração ainda ocorreu no sudeste (um dos locais em que houve maior rejeição do candidato). Este fato pode refletir a quantidade de usuários que utilizam o twitter, que deve ser mais fraco na região nordeste, ou até mesmo o índice de pessoas que possuem acesso a internet na região nordeste.

Palavra Chave: Fake News Whatsapp

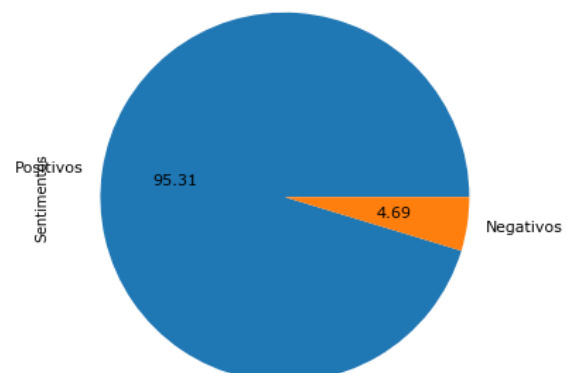
O Tweet Mais Curtido	O Tweet Mais Compartilhado
Depois de Facebook e Twitter, agora foi a vez do WhatsApp afirmar que a campanha de Bolsonaro não contratou impulsionamento. Aquela fake news lançada na época da campanha vai ficando cada vez mais... fake News.	Fake News podem destruir vidas. Cheque informações de saúde via WhatsApp, no seguinte número: (61) 99289-4640. https://t.co/8CN9nx1dO5 https://t.co/EXilwjXBsp
Número de curtidas: 11528	Número de retweets: 14347

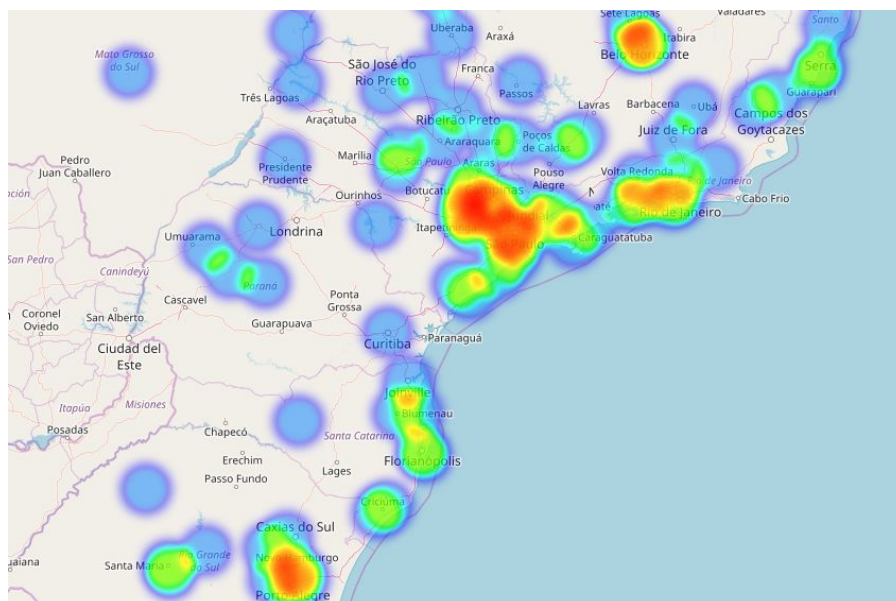
Análise de Sentimentos

TextBlob



NLTK





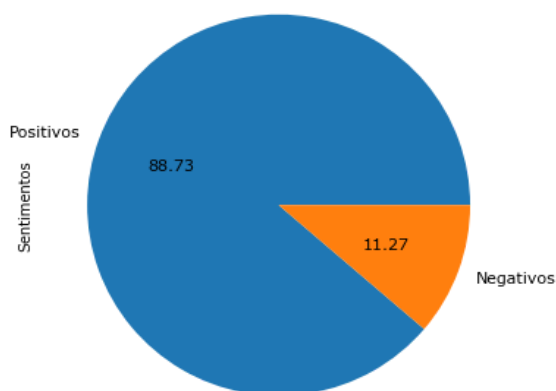
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, podemos observar as localizações do Brasil onde houve a maior concentração de tweets deste assunto.

Palavra Chave: Mourão Bolsonaro

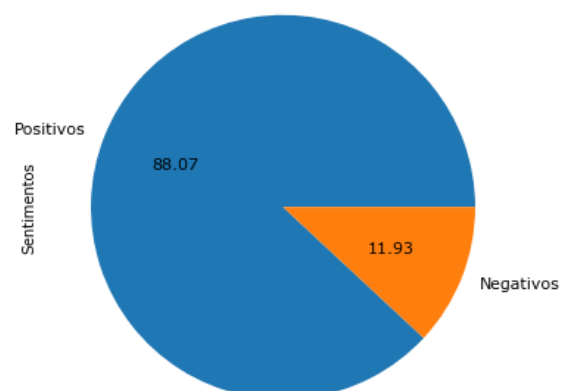
O Tweet Mais Curtido	O Tweet Mais Compartilhado
O presidente eleito Jair Bolsonaro deve dividir as atribuições da Casa Civil entre o deputado Onyx Lorenzoni e o vice-presidente, Hamilton Mourão. https://t.co/QDOnMKoy3b	Bolsonaro defende o torturador Brilhante Ustra. Seu vice, Mourão, também defende torturadores. Agora, seu filho, faz apologia à tortura. Não é por acaso. É absurdo que as autoridades não tomem providências. Nós temos que encerrar essa cultura de ódio. #TorturaNuncaMais
Número de curtidas: 626	Número de retweets: 20227

Análise de Sentimentos

TextBlob



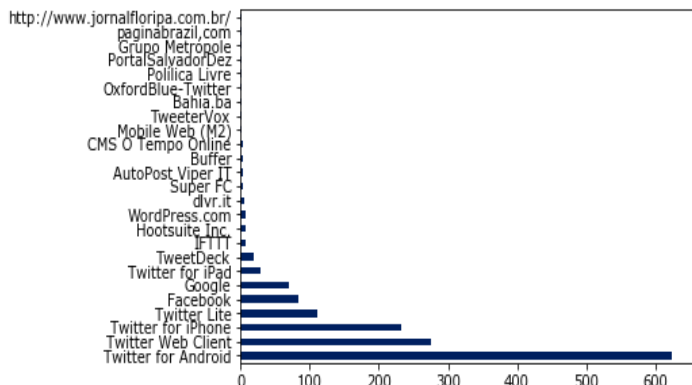
NLTK



As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Mourão Bolsonaro” obtendo dois resultados bem próximos. Acreditamos que os tweets em sua maioria foram considerados positivos devido a alta do assunto dos

médicos cubanos no Brasil, em que os usuários dizem que os médicos não devem ser tratados como escravos e apoiam a decisão de Bolsonaro.

Principais Fontes dos Tweets

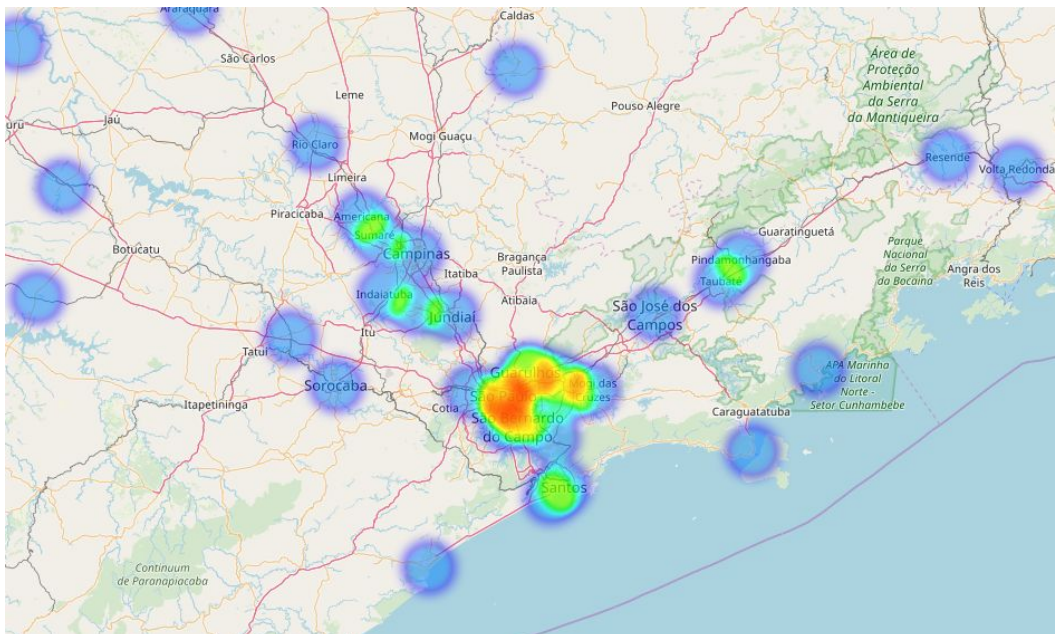


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por Twitter Web Client e depois por usuários de iPhone.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



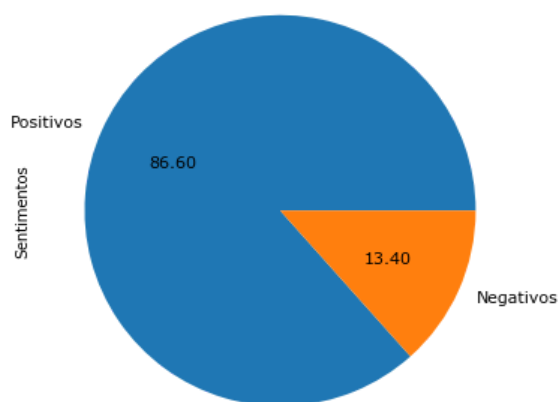
O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, dando um zoom na região do estado de São Paulo podemos notar que os tweets se concentram nas principais cidades do estado.

Palavra Chave: Manuela Haddad

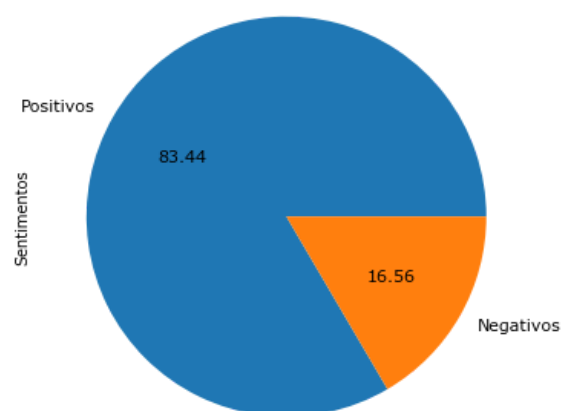
O Tweet Mais Curtido	O Tweet Mais Compartilhado
O Lula foi calado na maior injustiça que já vi mas temos aqui um novo Cara à altura. Estarei sempre com Lula, Haddad, Manuela e Boulos. O facista e os bolsominions que faça o timinho deles.	gente pq vcs estão tão ofendidos dizendo que a manuela e o haddad foram na igreja fingir que são pessoas boas? maioria de vcs fazem isso todo domingo
Número de curtidas: 551	Número de retweets: 24331

Análise de Sentimentos

TextBlob

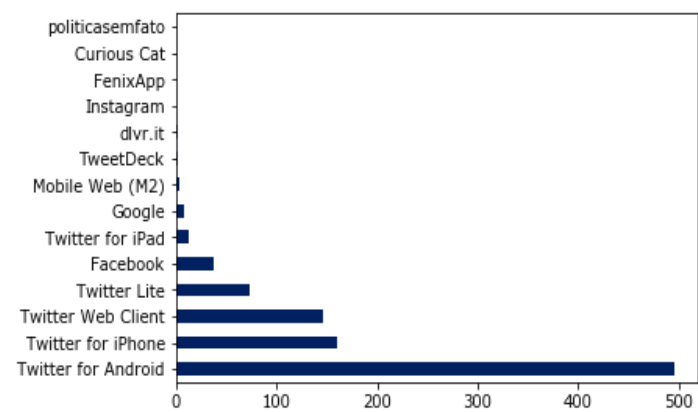


NLTK



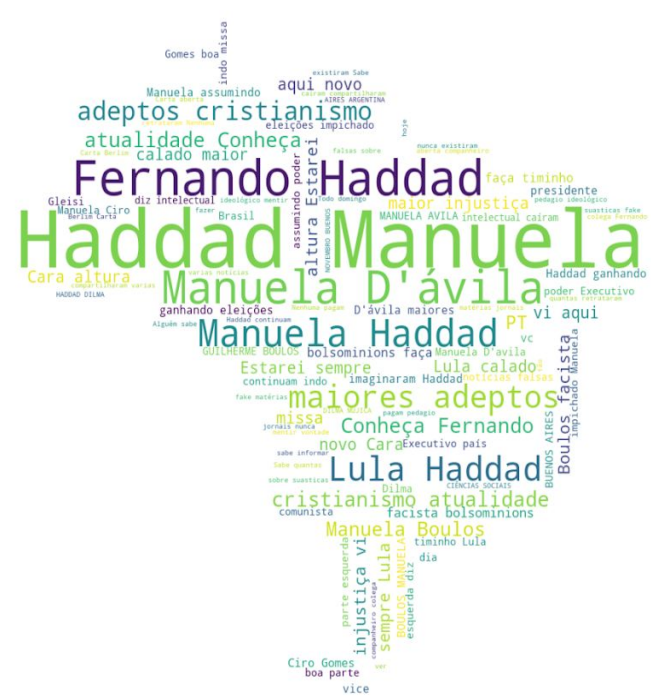
As duas ferramentas de análise de sentimento convergiram quanto aos resultados para a busca “Mourão Bolsonaro” obtendo dois resultados bem próximos, esta busca gerou menos tweets, foram 948. A maioria dos tweets tem como assunto a relação de Manuela e Haddad com a Igreja, como pode ser visto no mapa de palavras e no tweet mais retuitado. Como em sua maioria os usuários utilizaram de certa ironia do assunto e os dois métodos não conseguem distinguir tal comportamento, então a classificação da maioria dos tweets foi dada como positiva.

Principais Fontes dos Tweets

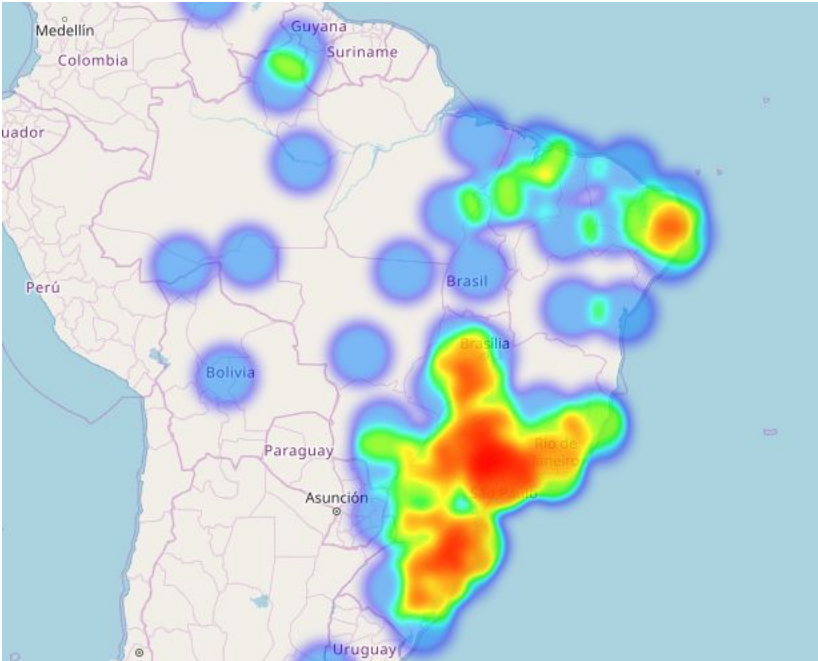


As principais fontes dos tweets coletados vem do aplicativo oficial do Twitter para Android, seguido por usuários de iPhone e depois por Twitter Web Client.

Nuvem de Palavras



Distribuição Geográfica dos Usuários



O mapa de calor apresenta a localização que os usuários declararam no seu perfil do twitter, podemos observar a concentração de tweets em todo Brasil da última busca.

5. CONCLUSÃO

Por fim, após a análise de todos os resultados podemos notar alguns padrões existentes nos dados dos tweets, tais como:

- A grande maioria dos usuários dos tweets utilizados no trabalho foram provenientes do Android, em seguida pelo iPhone e em terceiro o Web Client do Twitter.
- A grande maioria dos tweets foi dada das mesmas regiões do Brasil, como as maiores cidades, principalmente na região Sudeste.
- Em média, a maioria dos tweets foram classificados como positivos pelo *Textblob* e *NLTK*.

Outro ponto a ser observado é sobre o aprendizado do *Textblob* e *NLTK* com palavras em português, notamos que os métodos são bem simples e não conseguem distinguir ironias e sarcasmos nas frases, avaliando apenas as palavras em si. Também observamos a necessidade de treinamento com mais palavras em português, pois os métodos não conseguiram identificar algumas palavras que seriam negativas, e as classificava como positivas. Este comportamento provavelmente diminuiu a precisão dos resultados obtidos.

Em geral, o trabalho foi bastante interessante e motivador, trazendo uma boa experiência prática sobre *data science* e *machine learning* em Python. Experiência que certamente será muito útil para nossas vidas acadêmicas e profissionais.

6. REFERÊNCIAS

7.

NATURAL Language Toolkit. Disponível em:

<https://en.wikipedia.org/wiki/Natural_Language_Toolkit>. Acesso em: 22 nov. 2018.

CORPUS ReLi. Disponível em: <<https://www.linguateca.pt/Repositorio/ReLi/>>. Acesso em: 22 nov. 2018.

TEXTBLOB: Simplified Text Processing. Disponível em:

<<https://textblob.readthedocs.io/en/dev/>>. Acesso em: 22 nov. 2018.

RODRIGUES, Jéssica. O que é o Processamento de Linguagem Natural?. Disponível em:

<<https://medium.com/botsbrasil/o-que-%C3%A9-o-processamento-de-linguagem-natural-49ece9371cff>>. Acesso em: 22 nov. 2018.

8. ANEXO 1: CÓDIGO FONTE

Twitter Setiments Analysis

Imports

Basic Imports

```
import tweepy
import pandas as pd
import numpy as np
import time
import os
import re
```

Plotting and Visualization

```
from IPython.display import display
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud, ImageColorGenerator, STOPWORDS
%matplotlib inline
```

TextBlob Imports

```
from textblob import TextBlob
from textblob import TextBlob
from textblob.classifiers import NaiveBayesClassifier
```

NLTK Imports

```
import nltk
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
```

#MAP

```
from geopy.geocoders import Nominatim
import folium
from folium import plugins
from geopy.geocoders import Nominatim
```

Twitter Authentication Keys

#Keys for authentication

```
consumer_key='PI4AzCjWqxtu49moOM6fbY6w'
consumer_secret='NkE1jTLOlux5M9SmfZ1I0IvliATOQALnhyb1waKyecWk2xtJGR'

access_token='1036744604576636929-L6sPRS1vR705PKUpqV0gDrvfAybP7M'
access_token_secret='hKfheORbmZ4NOnSbtSsrll6aTeoclvUOlqRamT85hDgNa'
```

Twitter Authentication

#Authentication Methods

```
auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_token_secret)
api = tweepy.API(auth)
```

Searching for Tweets

#Searchin Twitter Timelines

```
tweets = []
```

```
info = []
```

```
for tweet in tweepy.Cursor(api.search,
                            q="lula ladrão",
                            tweet_mode='extended',
                            rpp=100,
                            result_type="mixed",
                            include_entities=True,
                            lang="pt").items(1500):
    if 'retweeted_status' in dir(tweet):
        aux=tweet.retweeted_status.full_text
    else:
        aux=tweet.full_text

    newtweet = aux.replace("\n", " ")

    tweets.append(newtweet)
    info.append(tweet)

    file = open("lula-ladrão.txt", "a", -1, "utf-8")
    file.write(newtweet+"\n")
    file.close()

    time.sleep(0.5)
```

Creating the data frame

#Construction of the data frame

```
tweets_df = pd.DataFrame(tweets, columns=['Tweets'])

tweets_df['len'] = np.array([len(tweet) for tweet in tweets])
tweets_df['ID'] = np.array([tweet.id for tweet in info])
tweets_df['Date'] = np.array([tweet.created_at for tweet in info])
tweets_df['Source'] = np.array([tweet.source for tweet in info])
tweets_df['Likes'] = np.array([tweet.favorite_count for tweet in info])
tweets_df['RTs'] = np.array([tweet.retweet_count for tweet in info])
```

```

tweets_df['User Location'] = np.array([tweet.user.location for tweet in info])
tweets_df['Geo'] = np.array([tweet.geo for tweet in info])
tweets_df['Coordinates'] = np.array([tweet.coordinates for tweet in info])

tweets_df.to_csv("lula-ladrão.csv")
tweets_df.head()

```

	Tweets	len	ID	Date	Source	Likes	RTs	User Location	Geo	Coordinates
0	"Nunca foi tão fácil ser ladrão nesse país." (...)	63	1062825499670573056	2018-11-14 21:52:00	Twitter Web Client	15017	3662	Rio de Janeiro	None	None
1	LULA""NUNCA FOI TÃO FÁCIL SER LADRÃO"" HUMMMMP...	105	1062895238384103425	2018-11-15 02:29:07	TweetDeck	680	117	São Paulo	None	None
2	'Nunca foi tão fácil ser ladrão neste País', d...	91	1062847668165189632	2018-11-14 23:20:05	Hootsuite Inc.	172	73	São Paulo, SP, Brasil	None	None
3	@Ronaldo @Corinthians @realvalladolid @realmad...	176	1064304119027023872	2018-11-18 23:47:30	Twitter for Android	0	0		None	None

Top Tweets

Top Searched Tweets

```
fav_max = np.max(tweets_df['Likes'])
```

```
rt_max = np.max(tweets_df['RTs'])
```

```
fav = tweets_df[tweets_df.Likes == fav_max].index[0]
```

```
rt = tweets_df[tweets_df.RTs == rt_max].index[0]
```

Max FAVs:

```
print("O tweet com mais curtidas é: \n{}".format(tweets_df['Tweets'][fav]))
```

```
print("Número de curtidas: {}".format(fav_max))
```

```
print("\n")
```

Max RTs:

```
print("O tweet com mais retweet é: \n{}".format(tweets_df['Tweets'][rt]))
```

```
print("Número de retweets: {}".format(rt_max))
```

O tweet com mais curtidas é:

"Nunca foi tão fácil ser ladrão nesse país." (Lula, 14/11/2018)

Número de curtidas: 15017

O tweet com mais retweet é:

"Nunca foi tão fácil ser ladrão nesse país." (Lula, 14/11/2018)

Número de retweets: 3662

Source of Tweets

#Source of the Tweets

```
sources = []
```

```
for source in tweets_df['Source']:
```

```
    if source not in sources:
```

```
        sources.append(source)
```

```
percent = np.zeros(len(sources))
```

```
for source in tweets_df['Source']:
```

```
    for index in range(len(sources)):
```

```
        if source == sources[index]:
```

```
            percent[index] += 1
```

```
            pass
```

```
newDF = pd.DataFrame({
```

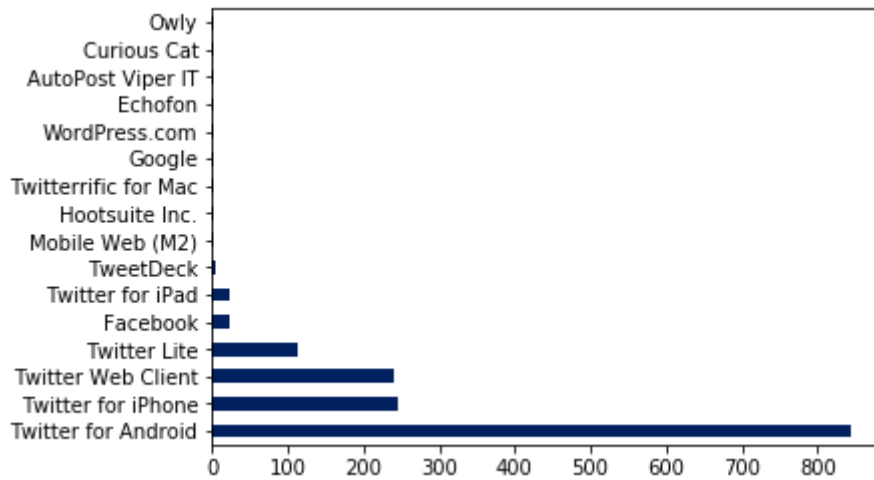
```
    'source':percent,
```

```
}, index=sources)
```

```
sources_sorted = newDF.sort_values('source',ascending=False)
```

```
ax = sources_sorted.source.plot(kind='barh',color='#002060')
```

```
ax.get_xaxis().set_major_formatter(plt.FuncFormatter(lambda x, loc: "{:,}".format(int(x))))
```



Training Model

#Sentiment Analysis in Português and English

```
base_path = 'c:/ReLi-Lex'
```

```
train = []
```

```
wordsPT = []
```

```
wordsPT_sentiments = []
```

```
files = [os.path.join(base_path, f) for f in os.listdir(base_path)]
```

```
for file in files:
```

```
    t = 1 if '_Positivos' in file else -1
```

```
    with open(file, 'r') as content_file:
```

```
        content = content_file.read()
```

```
        all = re.findall('[. *? \]', content)
```

```
        for w in all:
```

```
            wordsPT.append((w[1:-1]))
```

```
            wordsPT_sentiments.append(t)
```

```
            train.append((w[1:-1], t))
```

```
cl = NaiveBayesClassifier(train)
```

```
def clean_tweet(tweet):
```

```
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|(^0-9A-Za-z \t)|(\w+:\w\S+)", "", tweet).split())
```

Sentiment Analysis with TextBlob

```
def sentiment(tweet):

    polarity = 0

    #blob = TextBlob(tweet, classifier=cl)
    blob = TextBlob(clean_tweet(tweet), classifier=cl) #Clean Tweets

    for s in blob.sentences:
        polarity = s.classify() + polarity

    if polarity > 0:
        return 1
    elif polarity < 0:
        return -1
    else:
        return 0

def analyze_sentimentEN(tweet):

    analysis = TextBlob(tweet)

    if analysis.detect_language() != 'en':
        analysis = TextBlob(str(analysis.translate(to='en')))
        time.sleep(0.5)

    if analysis.sentiment.polarity > 0:
        return 1
    elif analysis.sentiment.polarity == 0:
        return 0
    else:
        return -1

#tweets_df['SA'] = np.array([ analyze_sentimentEN(tweet) for tweet in tweets_df['Tweet'] ])
#English Dictionary

tweets_df['SA TextBlob'] = np.array([ sentiment(tweet) for tweet in tweets_df['Tweets'] ]) #Portuguese Dictionary
```

Results of Sentiment Analysis

#Sentiment Analysis Results

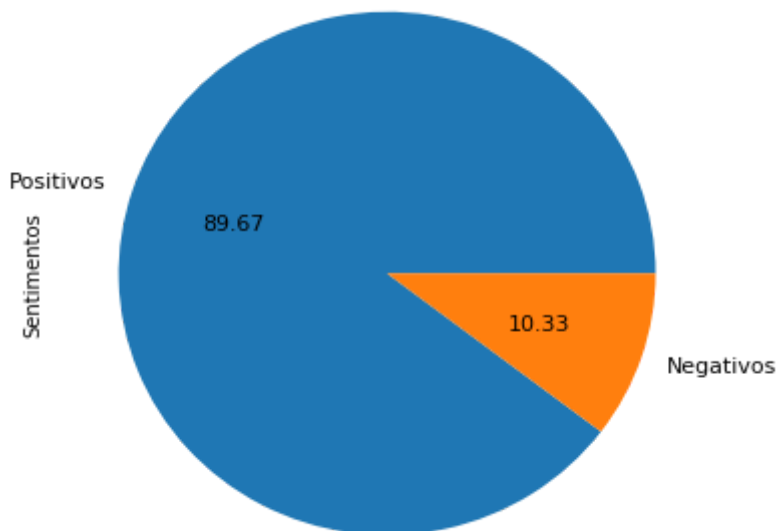
```
pos_tweets = [ tweet for index, tweet in enumerate(tweets_df['Tweets']) if tweets_df['SA TextBlob'][index] > 0]
neg_tweets = [ tweet for index, tweet in enumerate(tweets_df['Tweets']) if tweets_df['SA TextBlob'][index] < 0]
```

```
print("Porcentagem de Tweets Positivos: {}".format(len(pos_tweets)*100/len(tweets_df['Tweets'])))
print("Porcentagem de Tweets Negativos: {}".format(len(neg_tweets)*100/len(tweets_df['Tweets'])))
```

Porcentagem de Tweets Positivos: 89.66666666666667%
Porcentagem de Tweets Negativos: 10.333333333333334%

```
sentiments = ['Positivos', 'Negativos']
percents = [len(pos_tweets), len(neg_tweets)]
```

```
pie_chart = pd.Series(percents, index=sentiments, name='Sentimentos')
pie_chart.plot.pie(fontsize=11, autopct='% .2f', figsize=(6, 6));
```



Word Cloud

```
stopwords= set(STOPWORDS)

new_words = []
with open("stopwords_portuguese.txt", 'r') as f:
    [new_words.append(word) for line in f for word in line.split()]

new_stopwords = stopwords.union(new_words)

words = ''.join(tweets_df['Tweets'])

words_clean = "".join([word for word in words.split()
                        if 'https' not in word
                        and not word.startswith('@')
                        and word != 'RT' ])
```

```

from imageio import imread
import warnings

warnings.simplefilter('ignore')

twitter_mask = imread('brasil_mask.png')

wc = WordCloud(min_font_size=10,
               max_font_size=300,
               background_color='white',
               mode="RGB",
               stopwords=new_stopwords,
               width=2000,
               height=1000,
               mask=twitter_mask,
               normalize_plurals= True).generate(words_clean)

plt.imshow(wc, interpolation="bilinear")
plt.axis("off")
plt.savefig('lula-bolsonaro.png', dpi=300)
plt.show()

```



Time Series

#Tweets per Day

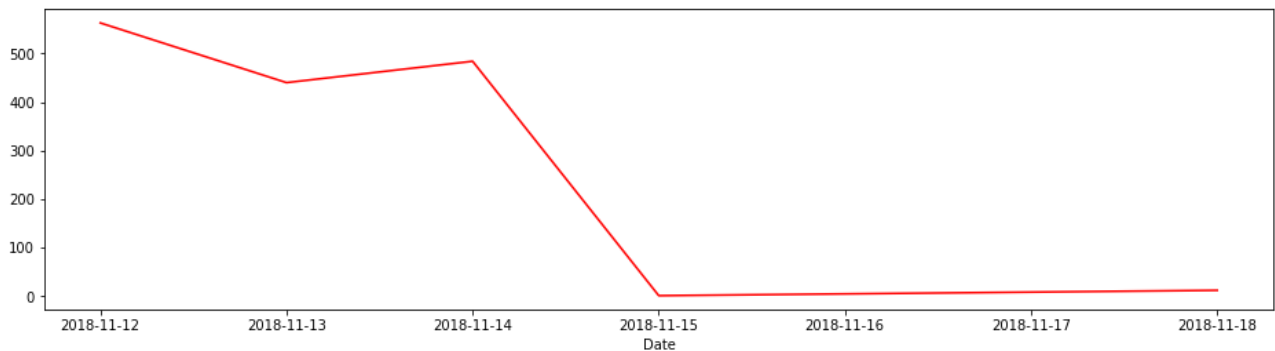
```

data = tweets_df

data['Date'] = pd.to_datetime(data['Date']).apply(lambda x: x.date())

tlen = pd.Series(data['Date'].value_counts(), index=data['Date'])
tlen.plot(figsize=(16,4), color='r');

```



Twitter Setiments Analisys with NLTK

Training Model

```
vectorizer = CountVectorizer(analyzer="word")
freq_tweets = vectorizer.fit_transform(wordsPT)
modelo = MultinomialNB()
modelo.fit(freq_tweets, wordsPT_sentiments);
```

Sentiment Analysis

```
tweetsarray = []
```

```
for tw in tweets_df['Tweets']:
    text = clean_tweet(tw)
    tweetsarray.append(text)
```

```
predictionData = vectorizer.transform(tweets_df['Tweets'])
tweets_df['SA NLTK'] = modelo.predict(predictionData)
```

Results of Sentiment Analysis

#Sentiment Analysis Results

```
pos_tweets = [ tweet for index, tweet in enumerate(tweets_df['Tweets']) if tweets_df['SA NLTK'][index] > 0]
neg_tweets = [ tweet for index, tweet in enumerate(tweets_df['Tweets']) if tweets_df['SA NLTK'][index] < 0]
```

```
print("Porcentagem de Tweets Positivos: {}".format(len(pos_tweets)*100/len(tweets_df['Tweets'])))
print("Porcentagem de Tweets Negativos: {}".format(len(neg_tweets)*100/len(tweets_df['Tweets'])))
```

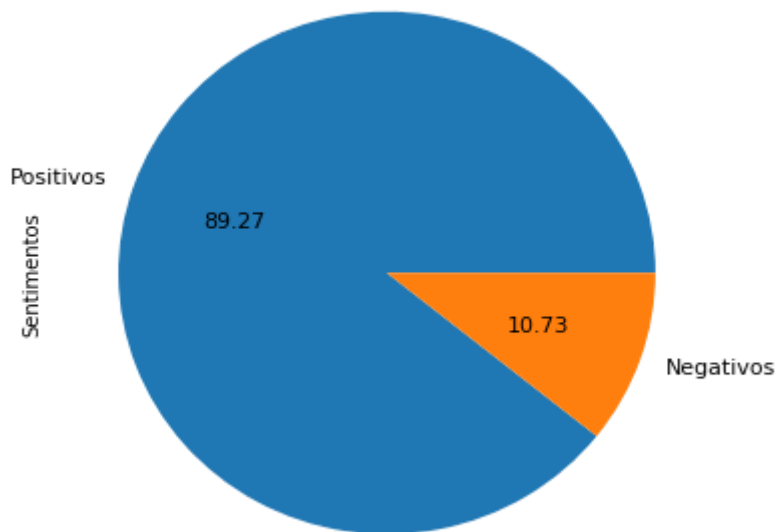
Porcentagem de Tweets Positivos: 89.26666666666667%
 Porcentagem de Tweets Negativos: 10.733333333333333%


```

sentiments = ['Positivos', 'Negativos']
percents = [len(pos_tweets), len(neg_tweets)]

pie_chart = pd.Series(percents, index=sentiments, name='Sentimentos')
pie_chart.plot.pie(fontsize=11, autopct='%0.2f', figsize=(6, 6));

```



Map of Tweets

```

geolocator = Nominatim(user_agent="TweeterSentiments")

latitude = []
longitude = []

for user_location in tweets_df['User Location']:
    try:
        location = geolocator.geocode(user_location)
        latitude.append(location.latitude)
        longitude.append(location.longitude)
    except:
        continue

coordenadas = np.column_stack((latitude, longitude))

mapa = folium.Map(location=[-15.788497,-47.879873],zoom_start=4.)

mapa.add_child(plugins.HeatMap(coordenadas))
mapa.save('lula-ladrão.html')
mapa

```