



**INSTITUTO
FEDERAL**

Paraíba

Campus
Campina Grande

Disciplina: Estatística Aplicada a Computação

Curso: Engenharia de Computação

Professor: Paulo Ribeiro

Alunos:

- ❖ Lucas Matheus Torres Costa
- ❖ Marcelo Alves da Silva Filho
- ❖ Carlos Vinícius

Turno: Tarde

Turma: 2017.2

RELATÓRIO

Análise de sentimentos do Twitter

O principal objetivo da Análise de sentimentos no Twitter é avaliar o que aquele tweet está tentando expressar, raiva, rejeição, aprovação, entre outros sentimentos. Muito utilizada por empresas, para saber se o produto que foi lançado no mercado está sendo bem aceito ou não pelos compradores, essa técnica consiste em extrair informações do texto em linguagem natural.

Optamos por usar a linguagem de programação Python, pois ela nos fornece um maior suporte na análise, treinamento do algoritmo, entre outras vantagens. Porém dessa vez não iremos analisar nenhum produto, mas sim a aceitação do filme “Deadpool 2” e como a população está lidando com a recente greve dos caminhoneiros.

Antes de começarmos as análises precisamos coletar os dados, para isso fizemos o pré-processamento dos mesmos através da biblioteca tweepy, removendo alguns sinais de pontuação, links, tweets retweetados, tweets irrelevantes e etc. Depois de pré-processados já podemos criar o nosso arquivo csv.

Através dos dados coletados anteriormente, construímos nossa base de dados, que possui mais de 2000 linhas e 8 colunas, tanto para Deadpool 2 quanto para Greve dos caminhoneiros, com algumas informações como: nome do usuário, texto do tweet, data e hora de postagem, aparelho em que foi feita a postagem, quantidade de retweets no tweet, quantidade de curtidas no tweet, localização do usuário e verificação da conta do usuário.

Porém, para que o algoritmo consiga classificar os tweets precisamos treiná-lo. Para isso vamos usar o Naive Bayes que é um classificador probabilístico baseado no teorema de Bayes. Este utiliza os dados de treino para formar um modelo probabilístico baseado na evidência de características das frases. Esse algoritmo supõe que as palavras dentro da frase não possuem relações com as outras. Sendo assim as sentenças pode ter polaridades positiva caso a maioria das palavras presentes na sentença possuam probabilidade maior de ser positiva.

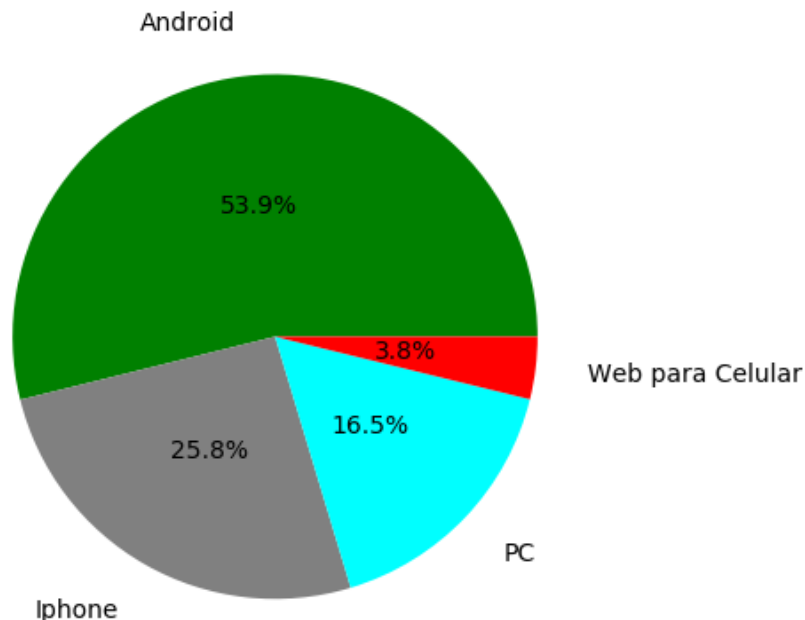
Depois de treinado, já podemos utilizar nossa base de dados para que o algoritmo classifique as frases como positiva negativa ou neutra. Optamos por criar uma nova base de dados, porém essa possui a classificação das frases e as informações anteriores. Agora sim começamos a analisar os dados, criar gráficos, fazer perguntas e consequentemente respondê-las.

Os dados foram extraídos no dia 25/05/18 do site twitter.com.

Análise sobre o filme Deadpool 2

Aparelhos

- Quais os dispositivos mais utilizados para twittar sobre o deadpool 2?

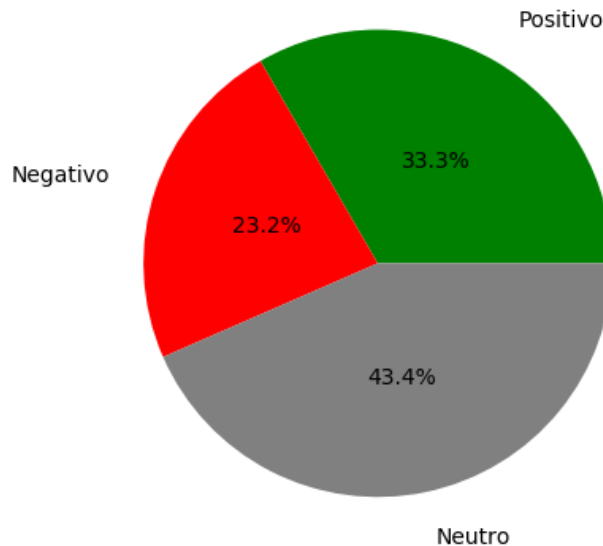


No gráfico em verde, os usuários que utilizaram android, em cinza os usuários que utilizaram Iphone, em ciano, os usuários que utilizaram computador desktop e em vermelho os usuários que utilizaram web para celular.

De acordo com o gráfico coletado pode-se observar que os dispositivos Android detém boa parte do acesso com 53,9%, logo em seguida o acesso pelo Iphone com 25,8%, depois o acesso pela PC com 16,5% e por ultimo através da Web para Celular (ou Twitter Lite) com 3,8%. Ao observar os dados tratados pode-se ver que com o passar do tempo os usuários fizeram uma migração da utilização da web para smartphones, isso decorre também do desenvolvimento dos dispositivos móveis que trouxe mais praticidade para os usuários. Através desses dados também para os desenvolvedores surge um leque de possibilidades no desenvolvimento de aplicativos móveis que ajudam na usabilidade do twitter. Também aplicativos que analisam o comportamento dos usuários para lançamento de produtos e serviços.

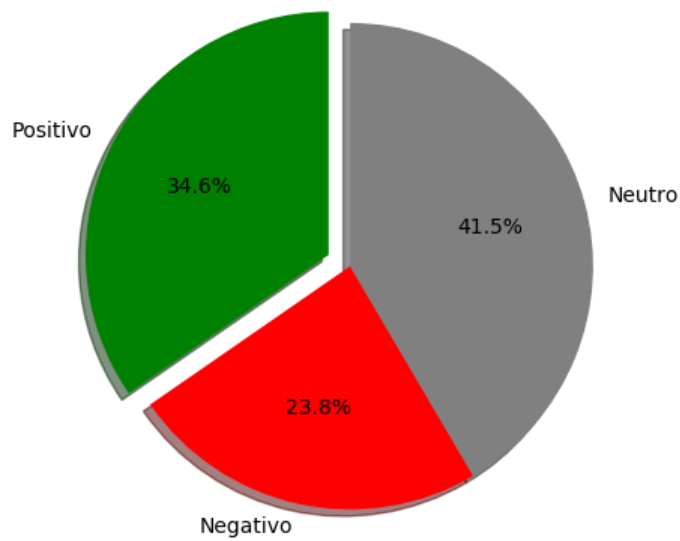
Avaliação

- Qual a opinião dos usuários sobre o filme Deadpool 2 que estreou no dia 17/05/18?



No gráfico, a cor verde representa os usuários que gostaram do filme, em vermelho, os usuários que não gostaram e em cinza os usuários não apresentaram decisão.

Ao analisarmos os twitters coletados pode-se observar que a maioria dos comentários são positivos mostrando um bom desempenho do filme com os seus telespectadores, houve algumas mudanças em relação ao filme Deadpool 1 o que gerou preocupação se o filme Deadpool 2 teria o mesmo bom desempenho, no entanto essa preocupação pode ser desfeita ao analisar o gráfico. É importante ressaltar que desconsideramos os comentários neutros, pois maioria desses são frases como: "eu vou assistir deadpool 2 hoje", "Alguém pra assistir deadpool comigo?". Dessa forma eles não acrescentam na avaliação positiva ou negativa do filme.

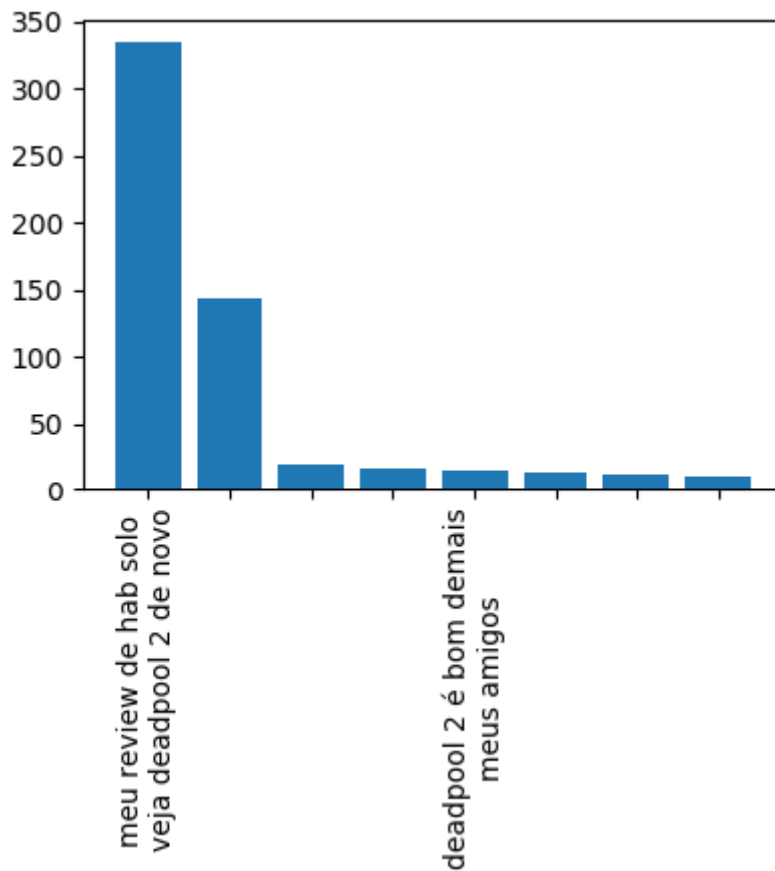


No gráfico, a cor verde representa os usuários que gostaram do filme, em vermelho, os usuários que não gostaram e em cinza os usuários não apresentaram decisão.

Porém, os gráficos gerados no dia 05/06, que possui tweets até 02/06, mostra que a aceitação do filme ainda é semelhante aos gerados anteriormente (quando o filme ainda estava em alta).

Retweets

- Qual Twitter tem mais retweets e o que estão falando sobre o filme?

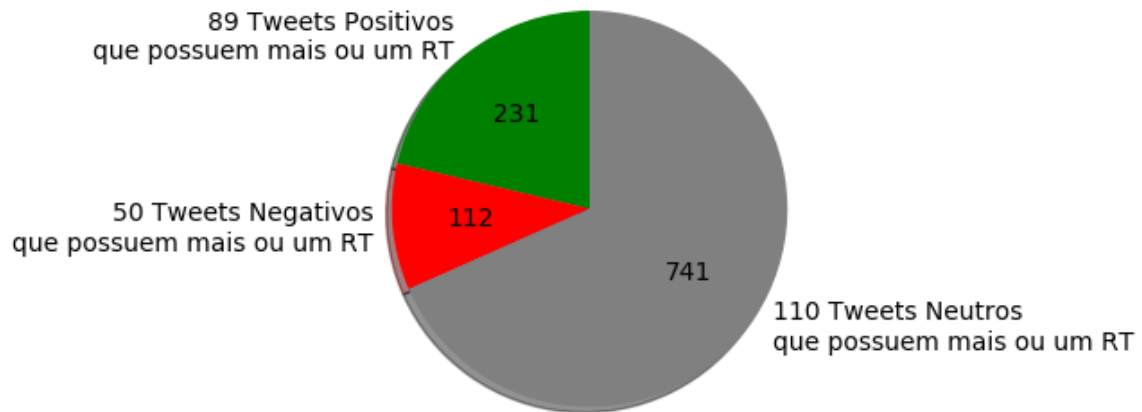


No gráfico, estão representados do maior para o menor, na ordem da esquerda para a direita os tweets mais retweetados.

De acordo com o gráfico o twitter que mais possui RT(retweets) não trás informações relevantes sobre o filme, da mesma forma o segundo(143) e o terceiro, com uma quantidade muito inferior aos anteriores(19). Mas logo atrás com 15 RTs o usuário avalia o filme como muito bom, tornando-se uma informação mais relevante que as anteriores, porém com um número menor de retweets. Concluimos que os tweets com mais RTs nem sempre são aqueles que expressão uma opinião. Em alguns casos possuem essa quantia, pois os usuários que postam, são mais conhecidos que os demais alguns possuindo canais no YouTube e outros conhecidos no Twitter. Não foi possível colocar os outros twitters devido o seu tamanho e se alguma informação do twitter for

omitida pode perder o sentido da frase e gerar uma tomada de decisão errônea.

Os comentários retweetados foram mais positivos ou negativos?

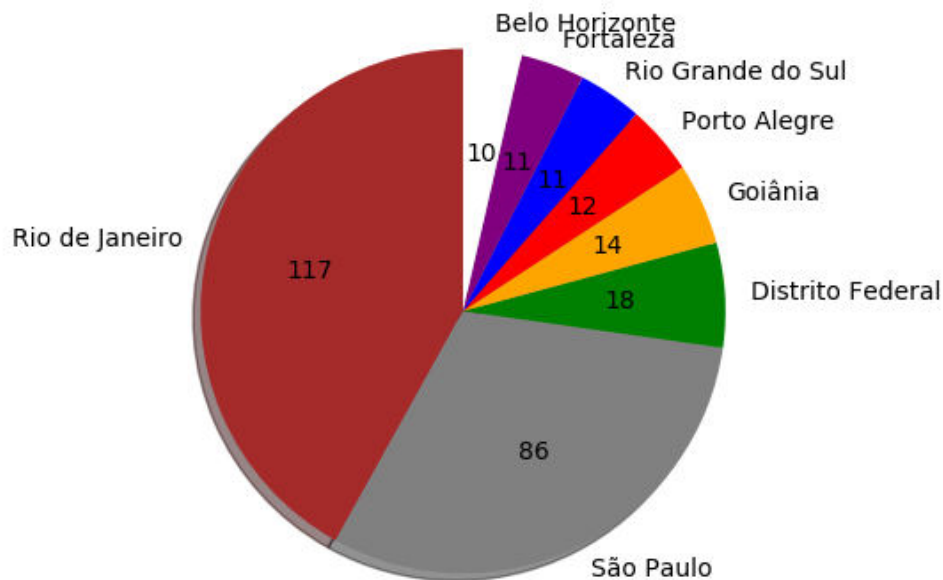


No gráfico, em verde os tweets positivos com mais ou um retweet, em vermelho, os tweets negativos com mais ou um retweet e em cinza os tweets neutros com mais ou um retweet.

Como já esperado, os tweets classificados como neutro estão em grande maioria. Mas desconsiderando-os podemos concluir que além de ter um número maior de tweets que possuem RTs classificados como positivos, a soma dos retweets desses 89 tweets é maior do que as do negativo.

Localização

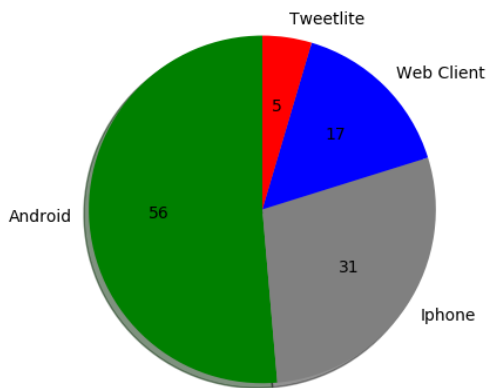
Qual estado teve a maior quantidade de tweets?



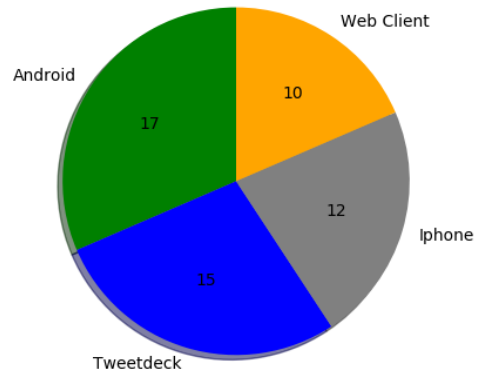
No gráfico, em vinho Rio de Janeiro, em branco Belo Horizonte, em roxo Fortaleza, em azul Rio Grande do Sul, em vermelho Porto Alegre, capital do Rio Grande do Sul, em laranja Goiânia, em verde Distrito Federal, em cinza São Paulo.

Separamos algumas localizações da base de dados, não foi possível analisar todas, pois alguns usuários optaram por não colocar suas localizações e outros colocaram localizações fictícias. Mas, através da análise feita sobre as informações que obtemos do estado do Rio de Janeiro, o estado foi o que mais postou sobre o assunto, seguido por São Paulo e Distrito Federal, curiosamente as maiores metrópoles nacionais.

Quais os tipos de aparelhos que mais twittaram por estado?



Rio de Janeiro



São Paulo

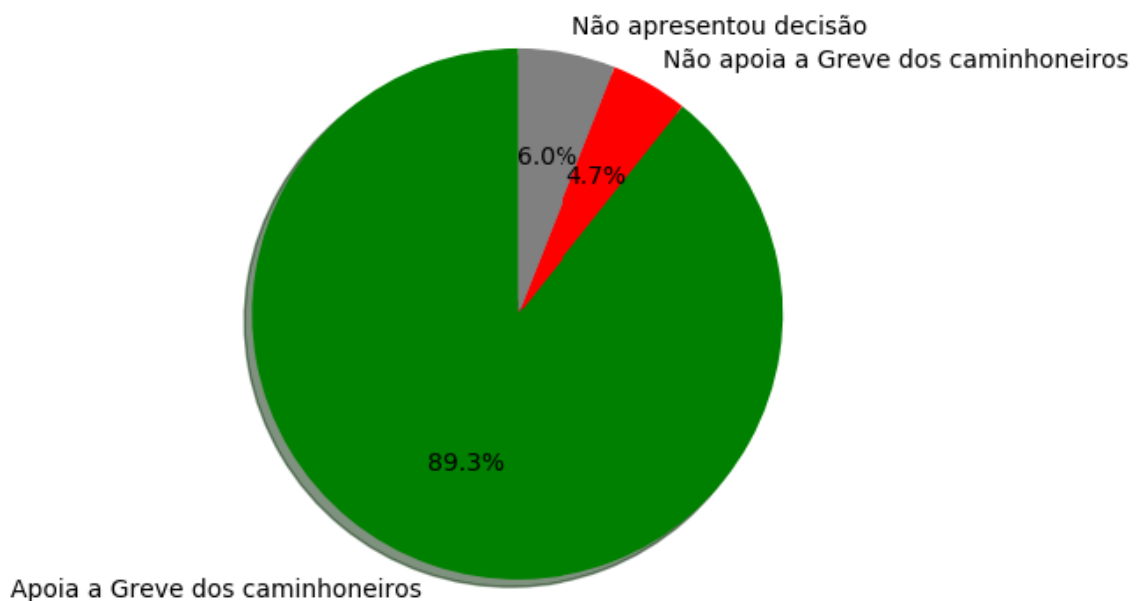
No primeiro gráfico, em verde android, em vermelho Tweetlite, em azul Web Client, em cinza Iphone. Já no segundo gráfico, em verde android, em laranja Web Client, em cinza Iphone, em azul Tweetback.

Separamos apenas os dados de dois estados, o primeiro e segundo onde houve maior incidência de usuários que twittaram sobre deadpool 2, e também apenas os tipos de aparelhos mais usados no estado em questão. Como já esperávamos, devido a gráficos anteriores onde o Android é o tipo de aparelho mais usado para twittar sobre deadpool, os gráficos acima seguem o mesmo modelo, onde em ambos os casos o Android está como o mais utilizado. Mas o que chama atenção é a presença do Tweetdeck (Aplicativo de mídias socais, que integra mensagens do Twitter e do Facebook.) como o segundo mais usado no estado de São Paulo. Já no Rio de Janeiro o Iphone continua com a segunda colocação.

Análise sobre o tema Greve dos Caminhoneiros

Avaliação

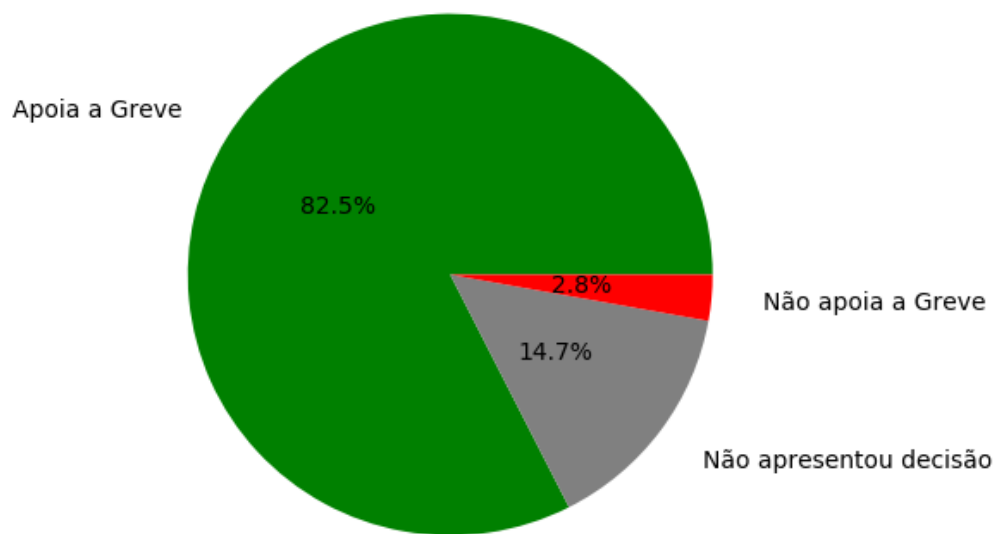
Em relação à Greve dos caminhoneiros os usuários estão a favor, contra ou neutros?



No gráfico, a cor verde representa os usuários que são a favor da Greve dos Caminhoneiros, em vermelho, os usuários que são contra e em cinza os usuários não apresentaram decisão.

Análise feita dia 25/05/18 – Ápice do movimento grevista:

Sobre o tema Greve dos caminhoneiros ao analisar este gráfico pode-se observar que existe no país um apoio quase que unânime a greve, mostrando a força que o movimento tomou em todo país, estes dados foram coletados apenas cinco dias após o início do movimento.



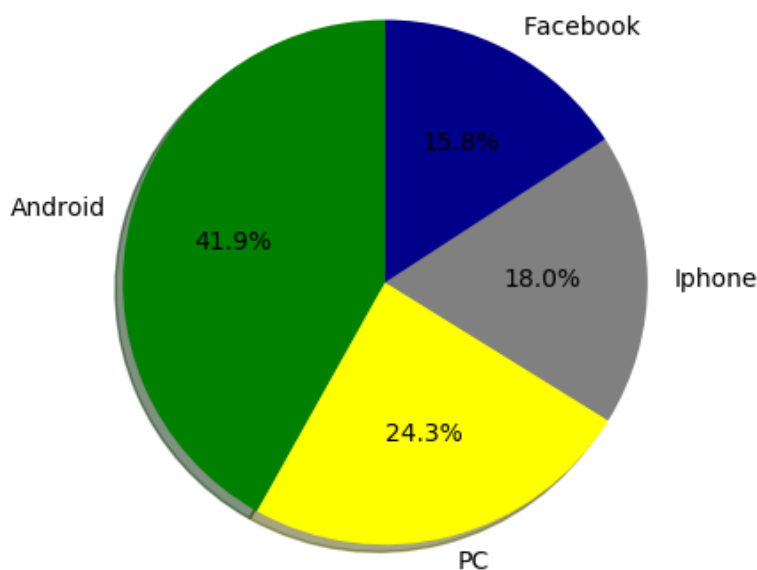
No gráfico, a cor verde representa os usuários que são a favor da Greve dos Caminhoneiros, em vermelho, os usuários que são contra e em cinza os usuários não apresentaram decisão.

Análise feita dia 05/06/18:

Alguns dias depois do ápice do movimento já nota-se que houve uma baixa nas avaliações a favor e contra o movimento, o que mostra que na sociedade o movimento perdeu adeptos e as pessoas optaram por ficar neutras em relação ao movimento.

Plataforma

Quais as plataformas que foram mais utilizadas para falar sobre a greve dos caminhoneiros?

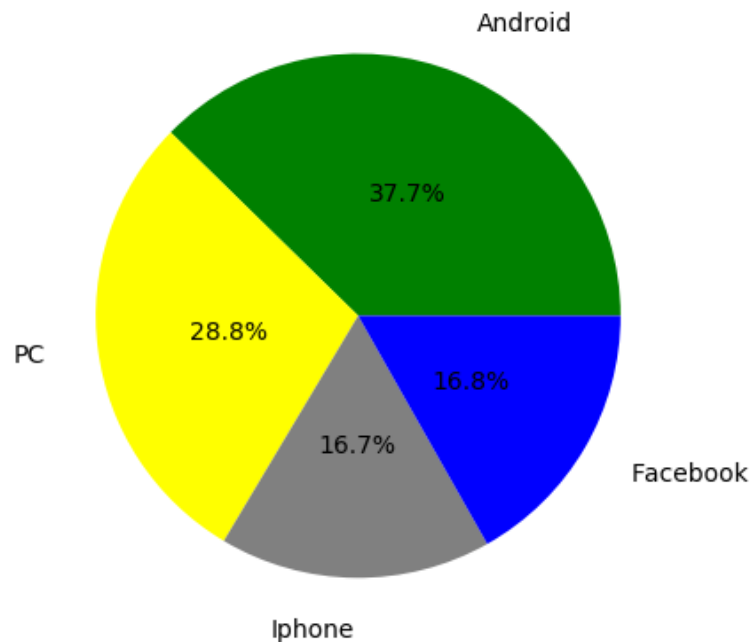


No gráfico, em verde, o acesso ao twitter pelo Android, em amarelo, o acesso ao twitter pelo Computador, "Desktop", em cinza, o acesso ao twitter pelo Iphone e em azul, o acesso ao twitter pelo Facebook.

Análise feita dia 25/05/18 – Ápice do movimento grevista:

Neste gráfico, nós mapeamos os aparelhos que foram utilizados para postar no twitter.com; em relação ao filme Deadpool 2 que foi objeto de análise também, os usuários postaram em segundo lugar mais pela web e também houve uma troca de quarto lugar onde o facebook aparece tirando o lugar do twitter lite que foi o quarto lugar da análise de Deadpool 2. Mas o que isso no mostra? Ao analisar os dados de Deadpool 2 e de Greve dos caminhoneiros sobre os aparelhos utilizados temos a conclusão que existem dois diferentes

públicos e que cada público tem sua característica, com isso os dados coletados são importantes para a tomada de decisão.



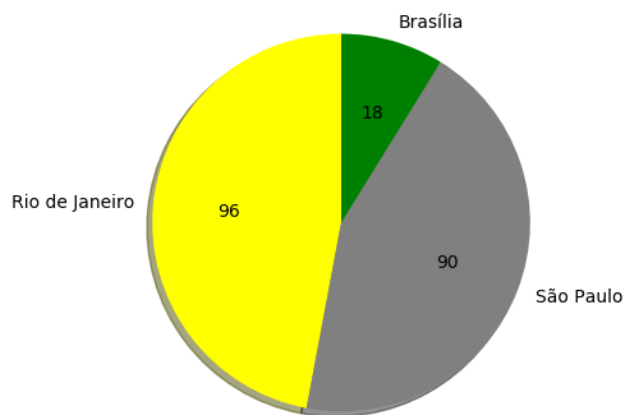
No gráfico, em verde, o acesso ao twitter pelo Android, em amarelo, o acesso ao twitter pelo Computador, "Desktop", em cinza, o acesso ao twitter pelo Iphone e em azul, o acesso ao twitter pelo Facebook.

Análise feita dia 05/06/18:

Ao analisar o segundo gráfico, nota-se que não houveram grandes mudanças do uso das plataformas pelo gráfico, porém o facebook teve um crescimento, mostrando o seu papel em meio ao tema, ultrapassando pouco, mas ultrapassando o acesso pelo Iphone.

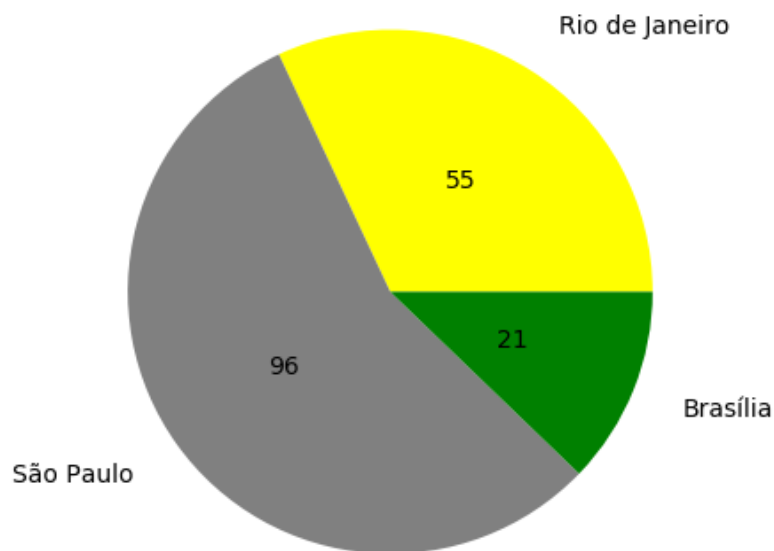
Localização

Quais as cidades que tiveram mais twitters sobre a Greve dos caminhoneiros?



No gráfico, em verde, a quantidade de comentários em Brasília, em amarelo, a quantidade de comentários no Rio de Janeiro e em Cinza, a quantidade de comentários em São Paulo.

No gráfico de comparação entre as cidades, as cidades de Rio de Janeiro, São Paulo e Brasília se destacam em número de comentários o que mostra a preocupação do povo destes locais com o destino do seu país e também o gráfico mostra um dado interessante que a princípio numa comparação entre São Paulo e Rio de Janeiro no número de habitantes, São Paulo ganha por ter maior numero de habitantes, mas de acordo com o gráfico as cidades estão equivalentes tendo o Rio de Janeiro uma leve superioridade. Vale ressaltar que na análise foi desconsiderado os dados que o twitter apresentou como “brasil”, o foco da análise foi nos estados brasileiros.



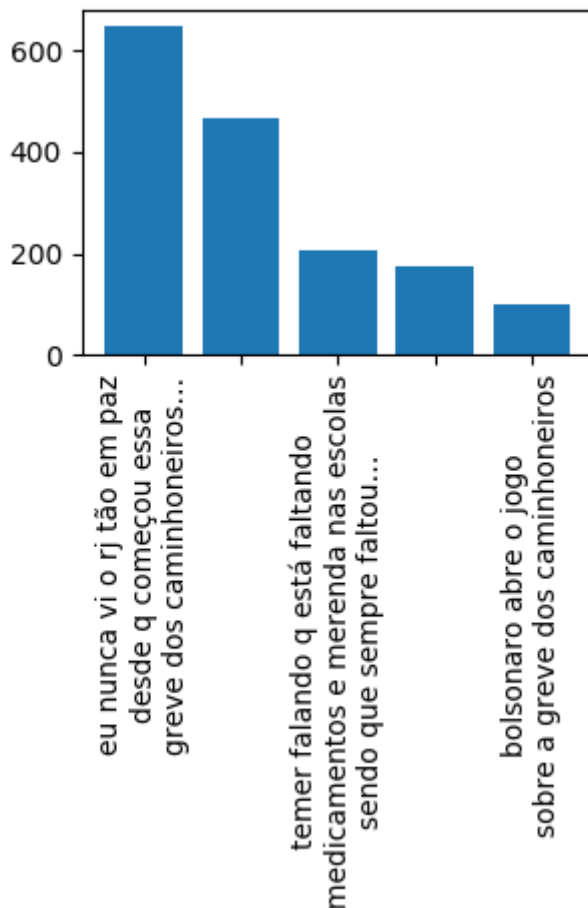
No gráfico, em verde, a quantidade de comentários em Brasília, em amarelo, a quantidade de comentários no Rio de Janeiro e em Cinza, a quantidade de comentários em São Paulo.

Análise feita dia 05/06/18:

Neste gráfico, nota-se que houve uma mudança de posições entre São Paulo e Rio de Janeiro em relação ao outro gráfico onde São Paulo aparecia atrás do Rio de Janeiro. Isso mostra que houve uma caída na procura no Rio de Janeiro, porém em São Paulo o índice permaneceu alto, sobre o tema.

Retweets

Quais os twitters que foram mais retweetados sobre a Greve dos caminhoneiros?

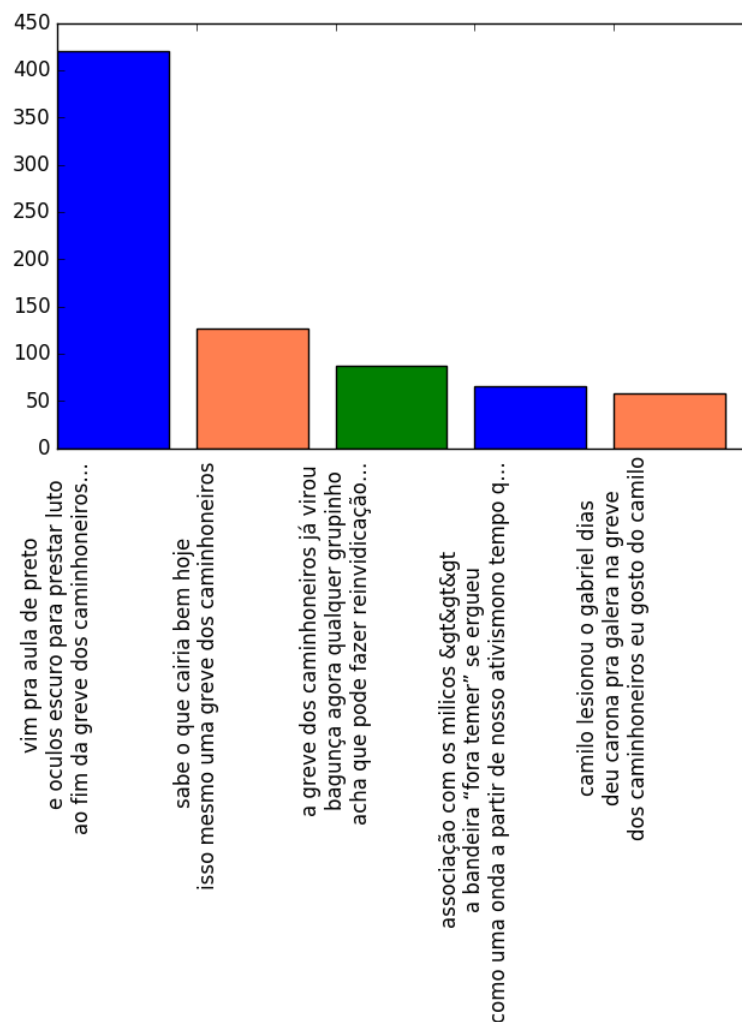


No gráfico, estão representados do maior para o menor, na ordem da esquerda para a direita os tweets mais retweetados.

Análise feita dia 25/05/18 – Ápice do movimento grevista:

Esse gráfico serve para mostrar quais tweets foram mais compartilhados, obtiveram um alcance maior dos usuários, através da análise destes tweets pode-se ver qual o assunto dentro do tema está sendo mais divulgado isso além de formar opinião, também ajuda na tomada de decisão em âmbito local, nacional. Não foi possível colocar os outros tweets devido o

seu tamanho e se alguma informação do tweet for omitida pode perder o sentido da frase e gerar uma tomada de decisão errônea.



No gráfico, estão representados do maior para o menor, na ordem da esquerda para a direita os tweets mais retweetados.

Análise feita dia 05/06/18:

Nesta análise não houve tanta diferença em relação a outra análise feita anteriormente, no entanto pode-se observar que os comentários mais retweetados foram num sentido diferente dos comentários do outro gráfico, mostrando a modificação de pensamento dos usuários ao longo do tempo.

Conclusão

Com isso, concluiu-se que o projeto análise de sentimentos do twitter atendeu as nossas expectativas e tudo que foi planejado ocorreu como planejado; houve também um bom aprendizado sobre a captação e análise de um banco de dados o que contribuiu de maneira significativa para o nosso crescimento profissional.