

CONSULTORIA E APOIO AO INVESTIDOR

Bruno F. Martins Fialho, Caio F. de Lima, Fernando R. Landuche, Gabriel L. Souza, Vitor H. C. da Silva Oliveira, Yago H. L. Viana, André Vogelsanger

Faculdade Impacta de Tecnologia
São Paulo, SP, Brasil
Dezembro de 2019

Resumo:

Este artigo apresenta como foi gerado um dashboard para apoio de decisão de compra de ação usando análise de sentimento de base de comentários extraídos do Twitter usando a dollartag("\$"). Foi feito a conexão através da biblioteca TWEETPY(Python), que realiza conexão com o Twitter via API e com isso consegue realizar a extração das informações necessárias,

Os dados foram utilizados para o "Entendimento do negócio" , "avaliação do modelo" e assim poder definir um Padrão das informações. Para tratamento dos dados e o dataframe foi criado a partir da biblioteca PANDAS com auxílio da biblioteca NUMPY, apresentando no final um DashBoard integrado que o PowerBi irá fornecer para uma melhor visibilidade para o cliente agilizando o uso.

Palavras-chaves: Integração, bibliotecas, Metodologias

Abstract:

This article presents how a dashboard for stock purchase decision support was generated using sentiment analysis based on comments extracted from Twitter using the dollartag (""). The connection was made through the library TWEETPY (Python), which connects to Twitter via API and thus can extract the necessary information,

The data were used for "Business Understanding", "Model Evaluation" and thus to be able to define an information pattern. For data processing and dataframe was created from the PANDAS library with the aid of the NUMPY library, presenting at the end a Integrated DashBoard that PowerBi will provide for better customer visibility for faster use.

Keywords: integration, libraries, methodologies

¹ Os autores podem ser contatados respectivamente pelos seus correios eletrônicos:

bruno.fialho@aluno.faculdadeimpacta.com.br, caio.lima@aluno.faculdadeimpacta.com.br,
fernando.landuche@aluno.faculdadeimpacta.com.br, vitor.oliveira@aluno.faculdadeimpacta.com.br,
andre.vogelsanger@aluno.faculdadeimpacta.com.br, gabriel.souza@aluno.faculdadeimpacta.com.br,
yago.viana@aluno.faculdadeimpacta.com.br,

Agradecimentos:

Agradecemos a Deus primeiramente, aos familiares que nos apoiam, e a todos os professores que forneceram sua sabedoria e dedicação para nossa evolução. Nossos sinceros agradecimentos a todos envolvidos.

1. Introdução

As redes sociais têm sido utilizadas como fontes de extração de informação direta e muitas vezes real e atualizada sobre uma empresa específica, entretanto nem todas as informações são relevantes a todos usuários, criando assim grupos diferentes de utilização da informação obtida. A utilização depende de cada propósito e usuário, tendo critério único de cada utilizador. (infomoney2018)

Através dos tweets extraídos da empresa foi feito um resumo e definição de positivo | negativo | neutro para os resultados . Os posts que foram utilizados são os posts que contenham o símbolo “\$” e o nome específico da empresa analisada, e a partir desta extração foi feito o agrupamento dos dados e criado os padrões de identificações; para o uso inicial deste modelo foi feita a extração dos dados e analisada manualmente pelos integrantes do grupo em uma primeira etapa, ao qual foi usada como base de treinamento para a aplicação

Identificando assim um padrão de evolução da empresa, onde foi utilizado como consultoria e base de apoio para aqueles que iriam investir em algum capital. Em cada um dos treinamentos da aplicação os resultados foram refinados e apresentados aos usuários , o qual ficava de critério de cada um dos usuários em como a informação seria utilizada, servindo assim como base de apoio para uma tomada de decisões.

1.1 Apresentação do Problema

Este projeto tem como finalidade auxiliar investidores que enfrentam dificuldades em tomar uma decisão segura de compra/venda de ações de uma empresa baseado em informações extraídas da rede social Twitter.

Baseado nos reflexos da Rede Social, foi apresentado outra perspectiva de análise capaz de orientar o investidor, atenuando os riscos do investidor..

Com os resultados obtidos o investidor pode definir se o dinheiro será investido ou não baseado na própria análise.

1.2. Objetivos

Os objetivos deste trabalho são:

- Fornecer a análise da atual situação e uma possível previsão futura de opiniões e sentimentos sobre a carteira de ações analisadas.
- Criar uma ferramenta que auxilie no investimento de dinheiro no mercado de ações

- Gerar dashboards que contenham os resultados da análise de sentimentos e que possam ser usados para identificar se o investimento terá maior ou menor risco.
- Servir como base de apoio na tomada de decisões financeiras

2. Estudo de Viabilidade

O projeto mostrou-se viável, porquê serviu para ajudar os investidores a minimizar os riscos de acordo com a opinião pública de uma determinada empresa através em uma rede social.

2.1 Soluções de Mercado e OPE

Este projeto foi inicialmente baseado em um projeto criado por (lilianweng - 2018), onde foi utilizado RNN para previsão do movimento de ações baseado em regressão logística. Outras referências de pesquisa se encontram na seção 5, em Referências.

2.2 Justificativa

Após análise das ferramentas encontradas, não foi encontrado uma ferramenta completa e de fácil utilização. Através deste projeto, o investidor teve disponível de forma clara todos os dados públicos referentes às ações da empresa quanto os posts relacionados em um dashboard integrado.

O software criado realizava a integração utilizando uma base de dados do twitter e foi usado como consultoria para entender a atual situação na época para cada empresa, a extração e análise foi feita baseada em períodos separados e empresas únicas de grande mercado.

3. Arquitetura da Solução

Após solicitado a análise da ação da empresa, como por exemplo Apple, a biblioteca tweepy faz a conexão com o Twitter através de uma API, e extrai os posts da determinada empresa. Em cima dos dados obtidos foi realizado a limpeza usando a biblioteca RE (Regex - Regular - Expressions), e a análise de dados usando a biblioteca Textblob. Todo Dataframe foi criado em cima da biblioteca Pandas e Numpy.

Os dados extraídos são exportados para um arquivo CSV para tratamento e disponibilização do dashboard pelo PowerBI, em conjunto com base financeira extraída do Yahoo Finanças.

O software será utilizado para fornecer uma melhor experiência para o investidor, onde será gerado dashboards que apresentam os dados atuais, gerando assim o resultado final, conforme figura 1, o que mostra o crescimento da opinião pública:

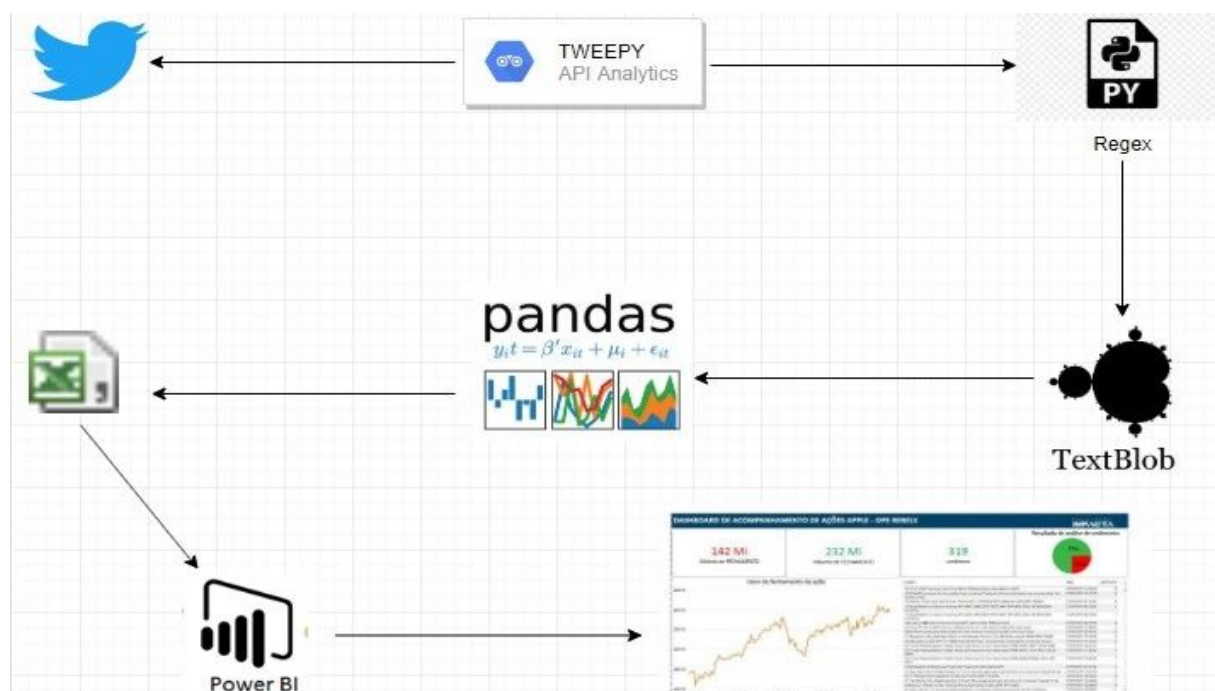


Figura 1- Ilustração da arquitetura de solução.

3.1 Diagrama de Componentes

Temos na figura 2 o modelo do diagrama de componentes feito:

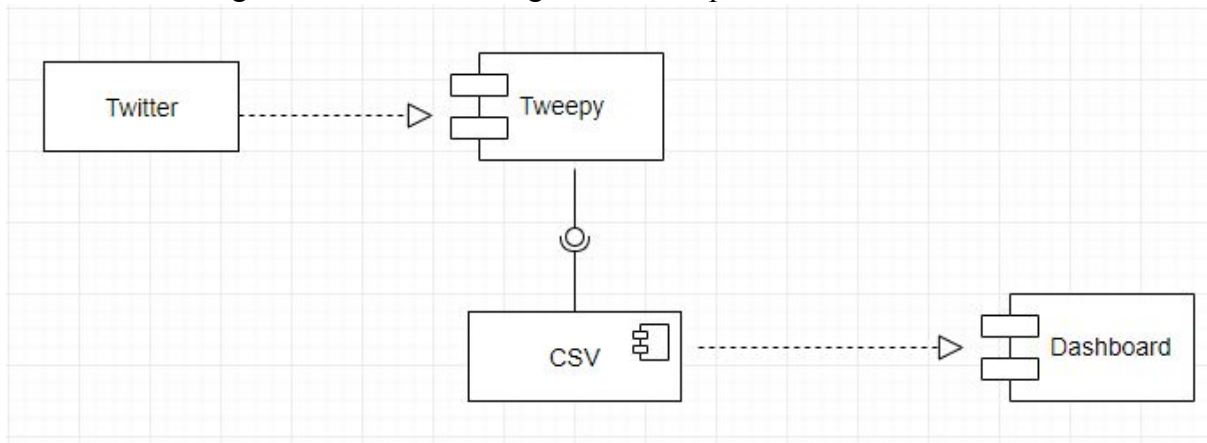


Figura 2- Ilustração o diagrama de componentes.

3.2 Infraestrutura

Solicitamos os hardwares e softwares abaixo para a integração do software :

- Notebook ou desktop core i3-600 6U 4GB 500GB ou AWS com configuração compatível
- Sistema Operacional Windows 7 ou 10 (qualquer versão)
- Acesso a internet com a velocidade de 5MB
- Softwares necessários (listados abaixo)

3.3 Tecnologias Utilizadas

Na tabela temos listado as tecnologias utilizadas na criação deste projeto :

Tabela 1 - Tecnologias utilizadas na criação deste projeto.

Tecnologia	Camada/SubSistema	Justificativa
Python	Desenvolvimento	Linguagem com ampla biblioteca
PowerBI	Apresentação dos Dados	Software utilizado com licença gratuita para a criação de Dashboards e de fácil utilização
Excel	Armazenamento dos dados	É um aplicativo que permite a criação e o gerenciamento de planilhas eletrônicas.

4. Resultados Obtidos

Analisando os resultados obtidos da extração e análise de sentimentos, foi observado (figura 3) que a empresa Apple analisada, tinha postagens com mais de 70% de conteúdo positivo, reduzindo assim os riscos do investimento baseado na visibilidade da Rede Social.

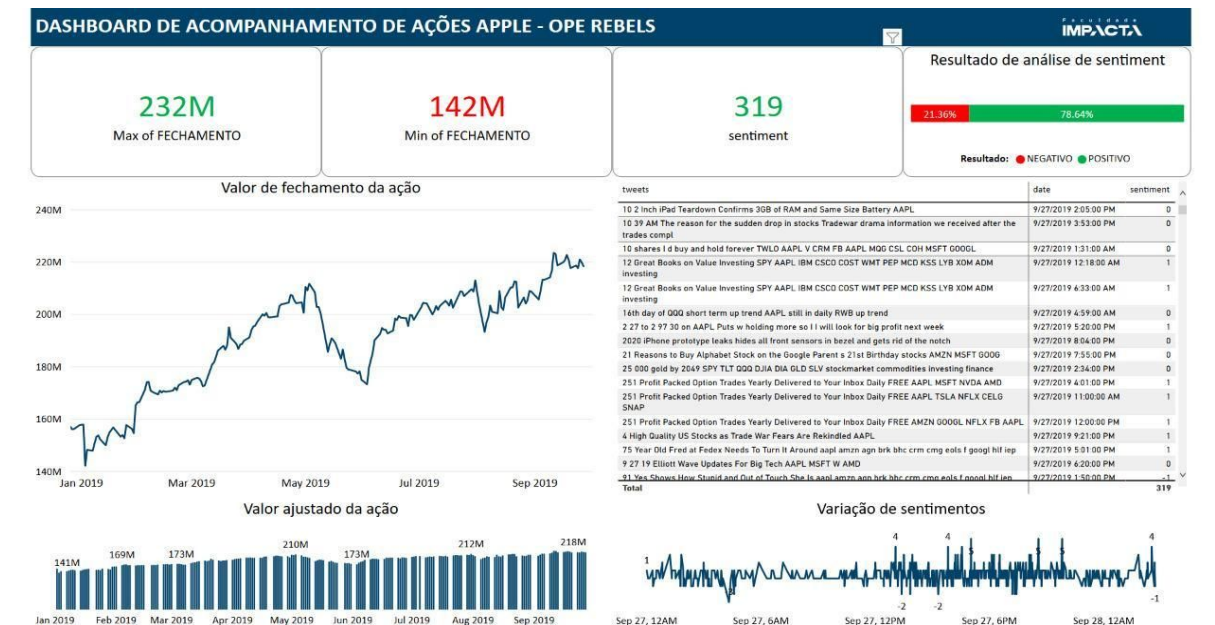


Figura 3 - Dashboard com análise de sentimentos.

4.1. Comparativo com Soluções

A solução apresentada foi baseada em posts do twitter, o que garante uma melhor análise em real time das informações. Outras soluções do mercado apresentam a informação baseada em bancos de dados que já armazenam esta informação em um período de tempo maior.

Comparado com os trabalhos estudados, o fator confiança está melhorado e assim temos um dashboard que foi mais rápido, prático e eficaz para o investidor.

4.2 Prototipação

Todo o código do python foi criado com base no tutorial visto no youtube do usuário (FreeCodeCamp.org - link nas referências)

4.2.1 Módulo de conexão com o Twitter

Na criação do módulo de conexão, foi utilizado chamada externa para que as informações confidenciais não fossem reveladas na disponibilização do código. Todo o processo de conexão é feito através da biblioteca Tweepy como mostrado abaixo

```
21 #####
22 # Módulos de autenticação do Twitter - código base usado: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
23 #####
24
25 auth = tweepy.OAuthHandler(tc.CONSUMER_KEY, tc.CONSUMER_SECRET)
26 auth.set_access_token(tc.ACCESS_TOKEN, tc.ACCESS_TOKEN_SECRET)
27 api = tweepy.API(auth, wait_on_rate_limit=True)
28
29
```

Figura 4 - Módulo de conexão com o Twitter.

4.2.2 Tratamento do texto

Para limpar os tweets extraídos para que a análise de sentimento seja feita de forma mais precisa, eliminando caracteres inválidos ou informações “suja”, foi utilizado a biblioteca RE conforme mostrado abaixo

```
14 #####
15 # Tratamento do tweet - código base usado: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
16 #####
17
18 def clean_tweet(tweet):    ##([A-Za-z0-9+)]|(^0-9A-Za-z \t)|(\w+:\|\/\|S+) - MOD ([^0-9A-Za-z\u0000-\uffff \t])|(\w+:\|\/\|S+)
19     return ' '.join(re.sub("([^\0-9A-Za-z\u0000-\uffff \t])|(\w+:\|\/\|S+)", " ", tweet).split())
20
```

Figura 5 - Limpeza dos tweets extraídos.

4.2.3 Criação do Dataframe

Temos na figura 6 o código da criação do dataframe, utilizando a biblioteca Pandas em conjunto com a biblioteca Numpy, com a criação da coluna de tweets passando por tratamento no módulo de limpeza de tweets.

```
39
40 #####
41 # Criação do dataframe para análise de sentimento - código base usado: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
42 #####
43
44 df = pd.DataFrame(data=[clean_tweet(tweet.text) for tweet in tweets], columns=['Tweets']) #limpeza removida - clean_tv
45 df['date'] = np.array([tweet.created_at for tweet in tweets])
46 df['source'] = np.array([tweet.source for tweet in tweets])
47 df['id'] = np.array([tweet.id for tweet in tweets])
48 df['len'] = np.array([len(tweet.text) for tweet in tweets])
49 df['date'] = np.array([tweet.created_at for tweet in tweets])
50 df['source'] = np.array([tweet.source for tweet in tweets])
51 df['likes'] = np.array([tweet.favorite_count for tweet in tweets])
52 df['retweets'] = np.array([tweet.retweet_count for tweet in tweets])
53
```

figura 6 - exemplo da criação do dataframe

4.2.4 Análise de sentimento

Na figura 7 vemos como a análise de sentimento é realizada, utilizando a biblioteca Textblob e passando como parâmetro a coluna tratada de tweets

```
54 #####
55 # Análise de sentimento dos dados extraídos - código base usado: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
56 #####
57
58 def analyze_sentiment(tweet):
59     analysis = TextBlob(clean_tweet(tweet))
60
61     if analysis.sentiment.polarity > 0:
62         return 1
63     elif analysis.sentiment.polarity == 0:
64         return 0
65     else:
66         return -1
67
68
69
70 df['sentiment'] = np.array([analyze_sentiment(tweet) for tweet in df['Tweets']])
71 df['polarity'] = np.array([(TextBlob(clean_tweet(tweet)).polarity) for tweet in df['Tweets']])
72
```

figura 7 - Analisando o sentimento do tweet

4.2.5 Salvando o Dataframe

Utilizando a biblioteca Pandas, vimos na figura 8 como exportamos o arquivo para formato .CSV para utilização no PowerBI.

```
72
73 #####
74 # Salvando o arquivo - código base usado: https://www.youtube.com/watch?v=1gQ6uG5Ujiw
75 #####
76 df.to_csv('base_aapl20-11.csv', index=True)#, encoding='utf-8')
77
```

Figura 8 - Salvando o resultado da análise

4.2.6 Exemplo do dataframe antes de ser importado pelo PowerBI

Temos na figura 9 um demonstrativo de como ficou o dataframe completo e tratado, antes do envio para criação do Dashboard no PowerBI

df - DataFrame

Index	Tweets	date	source	id	len	likes	retweets	sentiment	polarity
0	Apple adds more than 20 new magazines to Apple News+ \$AAPL	2019-11-22 14:52:32	divr.it	1197890612936...	82	0	0	1	0.318182
1	#Retailers are Embracing #MobilePayments (@Marketer) #digita...	2019-11-22 14:50:04	Hootsuite Inc.	1197889990518...	131	0	0	0	0
2	\$JMIA 300k volume in first 18 minutes \$SDC \$UBER \$LYFT \$I \$AAPL \$WORK NEW ...	2019-11-22 14:49:09	Twitter for iPhone	1197889763287...	97	0	0	1	0.193182
3	\$spcy \$gspc \$aapl \$nflx \$goog \$tsla Stocks edge higher at open as Dow, S...	2019-11-22 14:49:08	Zapier.com	1197889755779...	141	0	0	1	0.15
4	RT @trickmitch1: Man oh man if people only knew what happens here at #trick...	2019-11-22 14:47:26	Twitter for Android	1197889330301...	140	0	1	-1	-0.1
5	Man oh man if people only knew what happens here at #tricktrades everyda...	2019-11-22 14:45:40	TweetDeck	1197888886078...	139	2	1	-1	-0.1
6	RT @philstockworld: Friday Follies - Market Drifts Into the Weekend Sligh...	2019-11-22 14:44:22	Twitter Web App	1197888559279...	139	0	3	-1	-0.166667
7	RT @philstockworld: Friday Follies - Market Drifts Into the Weekend Sligh...	2019-11-22 14:44:07	Twitter Web App	1197888495085...	139	0	3	-1	-0.166667
8	@TheDeepDive_ca \$AAPL is a great short though. Been propped up the pa...	2019-11-22 14:43:56	Twitter Web App	1197888447400...	80	1	0	1	0.183333
9	5 Trades to Make 500% in 5 Months Updated and Short-Term Portfolio Rev...	2019-11-22 14:41:30	Twitter Web App	1197887835874...	139	0	0	0	0
10	\$AAPL poor signal if the market leader cannot get a bid.	2019-11-22 14:41:19	Twitter Web App	1197887790421...	56	2	0	-1	-0.4
11	\$SPY testing the pivot to the upside here, watching \$AAPL along side, so ...	2019-11-22 14:38:11	Twitter Web App	1197887000717...	87	1	0	1	0.35
12	BOT \$AAPL Nov 22 2019 W 262.50 Put @.83	2019-11-22 14:38:08	Twitter Web App	1197886988113...	39	0	0	0	0
13	#Spotify Sinks on a Double Whammy of Bad News @themotleyfool #stocks \$SPQ...	2019-11-22 14:37:59	HubSpot	1197886951233...	128	0	0	-1	-0.266667

Format Resize Background color Column min/max Save and Close Close

figura 9 - Dataframe antes de ser enviado ao PowerBI.

4.3. Considerações Finais

Apesar de não ser análise mais significativa no processo decisório com certeza este é mais um dado que pode auxiliar o investidor a tomar uma decisão com menor risco.

Portanto o projeto prometido cumpre as necessidades do investidor e como resultado temos um dashboard simples e intuitivo para auxiliar em sua tomada de decisão.

Este projeto serviu de apoio a investidores independentes e de grandes empresas. E com o bom resultado destes dashboards auxiliamos na análise mais precisa do investidor, através de informações de tweets.

4.3.1. Conclusão sobre o Projeto

Este projeto ainda está em andamento e todas as idéias têm sido alimentadas diariamente, novos updates serão fornecidos nas próximas atualizações deste projeto.

4.3.2. Sugestões de continuidade

Este projeto abre portas para novas análises para que possa melhorar a confiabilidade neste tipo de análise de sentimento.

5. Referências

(lilianweng 2018 - <https://github.com/lilianweng/stock-rnn> - acessado em 21/08/2018)
(FreeCodeCamp.org - Tweet Visualization and Sentiment Analysis in Python - Full Tutorial - <https://www.youtube.com/watch?v=1gQ6uG5Ujiw> acessado em 20/11/2018)
(infomoney2018 - <https://www.infomoney.com.br/patrocinados/noticias-corporativas/redes-sociais-de-investimentos-ganham-cada-vez-mais-atencao-do-mercado/> Acessado em 15/10/2019)
<https://www.quora.com/What-is-the-advantage-of-learning-Python-R-over-SSBI-tools-Tableau-and-PowerBI> (Acessado em 20/08/2019)
<https://www.edureka.co/blog/PowerBI-tutorial/> (Acessado em 17/11/2019)
<https://www.mssqltips.com/sqlservertip/5628/linear-regression-with-python-in-sql-server-2017/> (Acessado em 20/08/2019)
<http://investidordesucesso.com.br/bolsa-de-valores/8-sofware-para-investir-na-bolsa-de-valores/> (Acessado em 15/08/2019)
<https://docs.microsoft.com/pt-br/azure/storage/files/storage-python-how-to-use-file-storage>
<https://azure-storage.readthedocs.io/> (Acessado em 07/10/2019)
<https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (Acessado em 20/08/2019)
<https://medium.com/machina-sapiens/raspagem-de-dados-com-python-e-beautifulsoup-1b1b7019774c> (Acessado em 05/11/2019)
<https://www.quora.com/What-is-the-advantage-of-learning-Python-R-over-SSBI-tools-Tableau-and-Power-BI> (Acessado em 20/10/2019)
<https://www.edialog.com.br/novas-redes/redes-sociais-para-investidores/> (Acessado em 10/08/2019)