

Stock Market Movement By Analysing Tweets



Priya Sinha, Pawan Bisht
Supervised by Mr.Suresh Kumar Choudhary
Department of Data Science and Analytics, Central University Of Rajasthan, Ajmer, Rajasthan.
curaj.ac.in

1. Project Overview

Stock market moment is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock’s future moment will maximize investor’s gains. Stock prices rise and fall every second due to variations in supply and demand. Various Data mining techniques are frequently involved to solve this problem. But technique using machine learning will give more accurate, precise and simple way to solve such issues related to stock and market moment. Under this project we are analysing the moment of stock market by analysing tweets or by analysing sentiments of peoples during stock change.

2. Dataset

Twitter dataset(TCS)

| | Text | Sentiment | Sentiment_score |
|---|---|-----------|-----------------|
| 0 | RT In You fourth final instalment Jon Goyens ... | 0 | 0.000000 |
| 1 | RT Weve got As Soon As Possible exciting news... | 1 | 0.219048 |
| 2 | The outcome PRIOR CHESTNUTS FUNDRAISER March U... | 0 | 0.000000 |
| 3 | New india old india UFEEUFF | 1 | 0.118182 |
| 4 | TCS Developing Five Blockchain Platforms | 0 | 0.000000 |
| 5 | Sir TCS ruined Department Post SAP Ask empl... | 0 | 0.000000 |
| 6 | Why Flipkart Amazon OYO Paytm TCS Reliance Ind... | 0 | 0.000000 |

Stock market dataset(TCS)

| | Date | Close | collective sentiment | relative change |
|---|-----------|-------------|----------------------|-----------------|
| 0 | 4/10/2018 | 1468.800049 | 2.142911 | |
| 1 | 4/11/2018 | 1507.069946 | 3.604040 | 0.0253936 |
| 2 | 4/12/2018 | 1569.619995 | 3.370833 | 0.0398504 |
| 3 | 4/13/2018 | 1576.650024 | 0.369048 | 0.00445884 |
| 4 | 4/16/2018 | 1593.819946 | 3.777237 | 0.0107728 |
| 5 | 4/17/2018 | 1583.300049 | 2.736364 | -0.00664429 |
| 6 | 4/18/2018 | 1579.550049 | 3.511071 | -0.00237409 |

4. Twitter Development

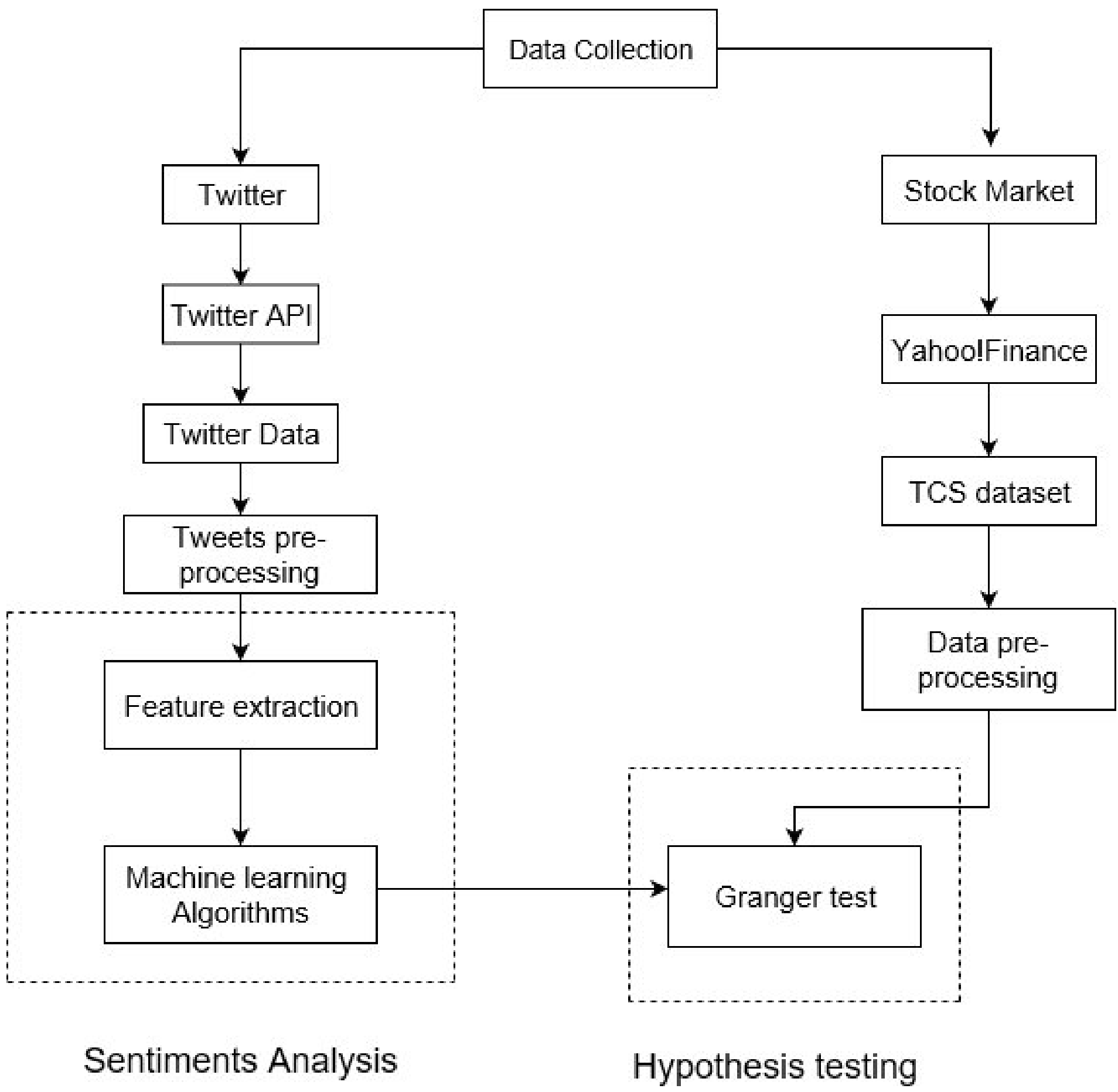


Accessing the Twitter APIs requires a set of credentials that you must pass with each request. These credentials can come in different forms depending on the type of authentication that is required by the specific endpoint that you are using, i.e., user context requires an API key and secret and a set of access tokens that are specific to the user that you are making the request on behalf of the owning user.

The keys and tokens necessary to make Twitter API requests

- 1.API key
- 2.API secret key
- 3.Access token
- 4.Access-token-secret

3. Flow chart



5. Results

Granger Causality Test

The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another.

In general case:

We create two models which predict y, one with only past values of y, and the other with past values of y and x.

Model1: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \text{error}$

Model2: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \dots + \beta_k Y_{t-k} + \alpha_1 X_{t-1} + \dots + \alpha_k X_{t-k} + \text{error}$

The F-test, t-test are calculated to test the following null and alternate hypotheses:

$H_0: \alpha_i = 0$ for each i of the element [1,k]
 $H_1: \alpha_i \neq 0$ for at least 1 I's of the element [1,k]

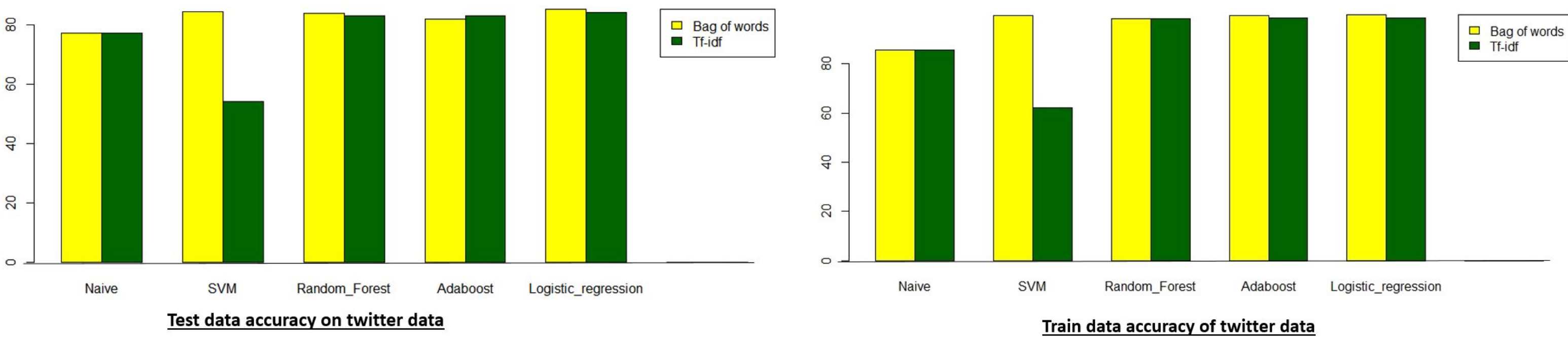
Let X= Collective Sentiment and Y= Relative change, then AR(1) is :

Our Model: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \alpha X_{t-1}$

$H_0: \alpha = 0$
 $H_1: \alpha \neq 0$

For X->Y(X Granger Causes Y)
 $\text{Pr}(>F) = 0.024 < 0.05$ (Reject H_0) (dependent)

For Y->X(Y Granger Causes Y)
 $\text{Pr}(>F) = 0.052 > 0.05$ (Not reject H_0) (independent)



7. References

<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> Intelligent Networks and Systems Society (INASS) <http://www.inass.org/> www.ijsr.net
<https://www.kaggle.com/general/12595>
<https://www.researchgate.net/publication/311843931>
<https://www.leadingindia.ai/downloads/projects/SMA/sma7.pdf>
<https://developer.twitter.com/en/apply>
<https://marcobonzanini.com/2015/03/09/mining-twitter-data-with-python-part-2/>
<https://pythonprogramming.net/preprocessing-for-machine-learning-python-programming-for-finance/>
<http://dcs/courses/grangertest.html>

6. Conclusion

In this project, we are providing a comparative study of existing techniques for opinion mining including machine learning algorithms, together with cross domain and cross-lingual methods. In the conventional stock market analysis, stock level indicators such as Moving average, Moving average convergence/ Divergence stock level indicators are used for stock market prediction. Project result shows that machine learning algorithm, such as svm, random forest and logistic regression have the highest accuracy and can be regarded as the baseline learning algorithm, we have also studied the effect of various features on classifier using ‘bag of words’ and ‘tf-idf’. We can conclude that more the clean data, more accurate result will be obtained. We have focused on the study of combining machine learning method into hypothesis testing of time series data in order to improve the accuracy of sentiment classification and adaptive capacity of variety of domains.