Sentiment analysis of tweets and their effects on the Stock Market

Hrishikesh Mahajan*
Department of Computer Science
Binghamton University - State University of New York
hmahaja1@binghamton.edu

Jainil Parikh†
Department of Computer Science
Binghamton University - State University of New York
jparikh1@binghamton.edu

ACM Reference Format:
Hrishikesh Mahajan and Jainil Parikh. 2020. Sentiment analysis of tweets and their effects on the Stock Market. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 DESCRIPTION OF DATA SOURCE

The Data source is generated by scraping the data from a website named 'investing.com'. We will scrape the data for the top 30 companies and select 8 attributes for analysis. We plan to run the script every 10 minutes. Following are the attributes we plan to perform analysis on:- Name, Last, High, Low, Change, ChangePercent, volume, and time.

2 HOW TO COLLECT DATA

The data is scraped from a website named 'investing.com'. We use http requests to hit the website and use beautiful soup to parse the html for us. Using this parsed html we access the data attributes we want and dump this into a table. We then perform ETL on the data and store the data of a Stock into its own table for ease of analysis.

3 MEASUREMENT AND ANALYSIS IDEAS

We plan to perform sentiment analysis on tweets made on twitter and correlate them to the changes in stock market prices. A future scope of the project would be to predict the change percent of the stocks based on current parameters. We will have to develop a Machine learning algorithm to train a model on this data for the given attributes and then offer insights for the Stock Market

4 ESTIMATED AMOUNT OF DATA TO BE COLLECTED

We are scraping 8 attributes for the top 30 companies, which means we will have 30 rows and 8 columns every time we hit the website. The size of the data for 1 cycle is about 32 KB. We plan to run this every 10 minutes, that is we will be generating approximately 4608 KB (4.5 MB) of data per day, totaling it to 32256 KB (31.5 MB) per week.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn