

Social Media based Stock Market Analysis using Big-Data Infrastructure

Shashanka Venkatesh
Venkataraman Nagarajan
Vishakan Subramanian

SSN College of Engineering, Chennai

October 26, 2021

Outline

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- 1 Motivation
- 2 Problem Statement
- 3 Justification
- 4 Feasibility Study
- 5 Literature Survey
- 6 Proposed System
- 7 Module Split-Up
- 8 Timeline
- 9 References

Motivation

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- In recent years, news events are influencing stock values to a great extent.
- A greater influence can be felt from social media trends.
- Unexpected stock market fluctuations are becoming more frequent.
- Immediate reactions to the stock market can be seen when incidents are reported/propagated using extreme sentiments.

Problem Statement

Shashanka,
Venkataraman
& Vishakan

Motivation

**Problem
Statement**

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- In a nutshell, we wish to analyse the correlation between social media trends & news articles relating to specific sectors and its effect on their current stock valuations.
- To develop a big-data architecture that is capable of handling real-time streaming information from sources like Google News, Bing, Yahoo Finance, Twitter etc.
- To assess and quantify the effect of news and social media on the economics of different sectors.

Justification

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

Observation

A lot of people invest in the stock market as a means to make money quickly. Shareholders react adversely and immediately towards their owned stock of relevant companies based on news trends, making their valuations more volatile.

Conclusion

Presently, with the majority of the world being online, people tend to be more engaged on social media & use it as a source of news. Finding the correlation between such events & stock valuations is hence justified.

Feasibility Study

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

Existing Work

The feasibility of the project is based on the research done by Lee et al. [1], where they collect Tweets and news from specific sectors to find a correlation between the expressed Tweet sentiments and the stock value.

Literature Survey

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- Lee et al.'s^[1] research delves into analyzing Twitter data, by classifying it into categories and performing sentiment analysis to predict the trend of stock prices, and comparing it to reality, to find the correlation between the two.
- Sources of improvement
 - Statistically insignificant volume of data, which reduces the confidence in their conclusion.
 - An increase of nearly 70 million users on Twitter between 2017 and 2021, giving reason to believe that the results may be different.
 - Usage of better machine learning techniques to improve classification & sentiment analysis.

Proposed System

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

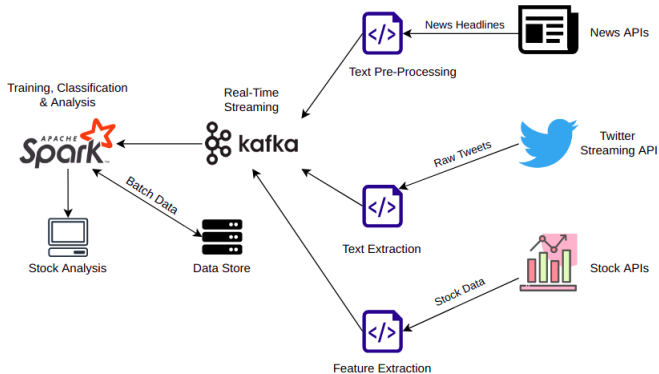


Figure: Proposed Architecture

Proposed System (cont.)

- Twitter Data Classification
 - Classifying Tweets with respect to their relevant sectors.
 - Analysing expressed Tweet sentiments in each sector.

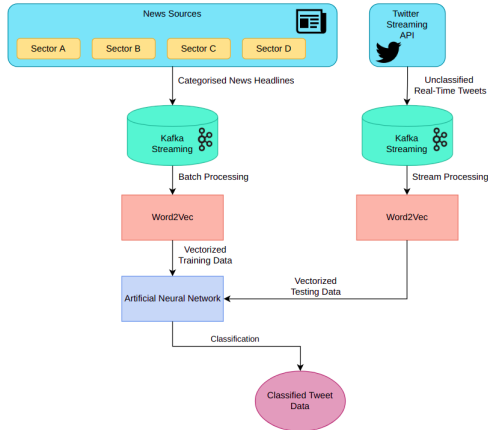


Figure: ML Pipeline

Proposed System (cont.)

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- Stock Data Analysis
 - Raw stock data is collected and pre-processed in order to find the current valuation trend.
 - Stock trend is then compared & correlated with the trends observed in Tweets of the relevant sector.
- Correlation Analysis
 - Correlation analysis quantifies the effect of social media, news & current events on the valuation of a stock in a particular sector.
 - Some sectors may be influenced more by social media than other sectors - for example, cryptocurrency is significantly influenced by Twitter hype around it.

Module Split-Up

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study


Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

- News Classification
 - News 
 - Tokenization
 - Feature selection
 - Training
- Twitter 
 - Tokenization
 - Feature Vector
 - Classification
- Stock Analysis 
 - Real-Time Data Collection.
 - Efficient Storage of data
 - Trend Analysis
- Correlation Analysis

Timeline

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

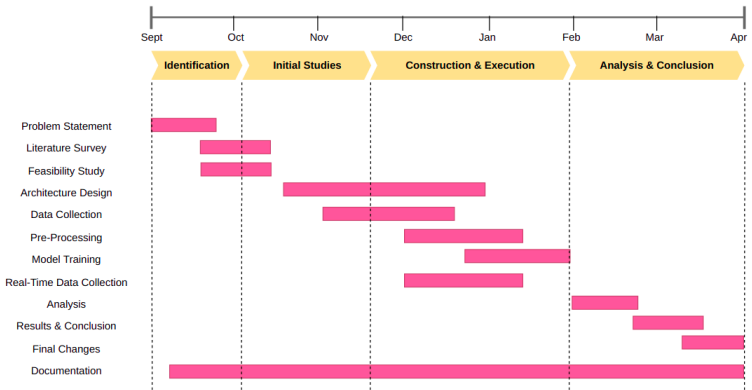


Figure: Timeline Chart

References

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References



[1] Chungho Lee, Incheon Paik, *Stock Market Analysis from Twitter and News Based on Streaming Big Data Infrastructure*



[2] Vaanchitha Kalyanaraman, Sarah Kazi, Rohan Tondulkar, Sangeeta Oswal, *Sentiment Analysis on News Articles for Stocks*



[3] Altti Ilari Maarala, Mika Rautiainen, Miikka Salmi, Susanna Pirttikangas, Jukka Riekk, *Low latency analytics for streaming traffic data with Apache Spark*



[4] Zhihao Peng, *Stocks Analysis and Prediction Using Big Data Analytics*, 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS).



[5] Gurcan, Fatih & Berigel, Muhammet, *Real-Time Processing of Big Data Streams: Lifecycle, Tools, Tasks, and Challenges*, 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT).



[6] Aparna Nayak, M. M. Manohara Pai and Radhika M. Pa, *Prediction Models for Indian Stock Market*



[7] Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra, *Indian stock market prediction using artificial neural networks on tick data*



[8] Kanavos, Andreas & Vonitsanos, Gerasimos & Mohasseb, Alaa & Mylonas, Phivos. (2020). *An Entropy-based Evaluation for Sentiment Analysis of Stock Market Prices using Twitter Data*



[9] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, *Efficient Estimation of Word Representations in Vector Space*



[10] Marsland, Stephen, *Machine Learning: An Algorithmic Perspective*, 2014

Thank You

Shashanka,
Venkataraman
& Vishakan

Motivation

Problem
Statement

Justification

Feasibility
Study

Literature
Survey

Proposed
System

Module
Split-Up

Timeline

References

Thank You
Q & A