Kammavari Sangham (R) - 1952
K.S.INSTITUTE OF TECHNOLOGY
(Affiliated to VTU, Belgaum. Approved by AICTE, New Delhi)
No.14, Raghuvanahalli, Kanakapura Main Road, Bengaluru - 560109
Phone : 080-28435722 / 24, Fax : 080-28435723, Email : principal.ksit@gmail.com, Website : www.ksit.ac.in

Stock Market Prediction using news articles

Batch 21

Guided By:
Dr Rekha B Venkatapur
Professor, Dept of CSE,KSIT

Presented By:
Sampath Kumar
Shravan Bhat
Siddhanth M

Questions

1) How do you retrieve stock prices in real time?
2) How to calculate the correlation of stock prices?
3) How many keywords are present in the bag of words?
4) How do you measure the performance analysis of the model?
5) How do you verify the authenticity of the news?

Contents

Introduction

- Stock market is a place where shares of stocks of publicly listed companies are traded.
- A stock exchange facilitates stock brokers to trade company stocks and other securities.
- The major indian stock exchanges are Bombay Stock Index(Sensex) and National Stock Exchange(Nifty).
- Some companies on the stock market include Reliance, L&T, Bharti Airtel, Wipro, Vodafone and many more.

Introduction

- Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange.
- There are many factors involved in the prediction – physical factors, physiological factors, rational and irrational behaviour, etc.
- All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

Literature Survey

1) “**A Data mining algorithm to analyse stock market data using lagged correlation”**. Cicil Fonseka, School of Computing and Mathematics University of Western Sydney, Campbelltown, Australia

- The ability of one stock to predict the future usually short term future trends of a closely correlated another stock.
- Algorithm for predicting the market direction more accurately when two stocks are strongly correlated to each other with a lag of K number of trading days.

2) “**Predicting the Effects of News Sentiments on the Stock Market.”** Dev Shah School of Computing Queens University, Haruna Isah School of Computing Queens University, Farhana Zulkernine School of Computing Queens University.

- To classify the sentiment of the news article/headline and categorizing accordingly to its respective category for further understanding.
- Worked on the development of a sentiment analysis dictionary for the financial sector, the development of a dictionary-based sentiment analysis model, and the evaluation of the model for gauging the effects of news sentiments on stocks.

Literature Survey

3) “**Stock Price Prediction Using Financial News Articles”.** M. ø. Yasef Kaya, M. Elif KarslÕgil, Department of Computer Engineering YÕldÕz Technical University Istanbul, Turkey

- A prediction model, finding and analyzing correlation between contents of news articles and stock prices and then making predictions for future prices
- Accomplished the stock prediction system using financial news articles

- Analyzing textual data, use word couples consisting of a noun and a verb as features instead of using single words.
- Strong relationship between financial news and stock price movements.

Problem Statement

“To predict stock market movement using daily financial news articles”

Input: Stock market data along with daily financial news article to predict the stock market movement.

Output: Expected movement(Prediction) of the specific shares of stock

Scope

The scope of the project is that it focuses on the financial market as of now which can easily be expanded onto the other platforms in the future.

Goals

- The goal of the project is to give the investors an informed decision so that they can invest at the right time.
- Application should be scalable so that it can be expanded to other platforms and markets.
- Ability to simulate mock news and see how it might affect the price of a company.

Objectives

- Our project aims to make gathering and parsing of news easier for stock market prediction.
- We also use the sentiment for defining the effect of individual news by assigning a polarity score to each news article.
- For checking the movement between stocks whose news is unknown, we can use the covariance between the stocks using

$$\text{Covariance} = \frac{\sum (Return_{ABC} - Average_{ABC}) * (Return_{XYZ} - Average_{XYZ})}{(\text{Sample Size}) - 1}$$

- Using these polarities and the stock returns, we will assign a prediction.

Methodology

1. **Gathering of information**

- First step is to download data from various news sources and their respective api's.
- We also download the data of the stock market exchange for which we are trying to predict the outcome. This will involve all the market data like high,low volume traded etc.
- This scraping of information will be done with help of BeautifulSoup4 **[6]** - A library in python for extracting data.

2. **Cleaning the dataset**

- We will now parse the given information which has been downloaded, to process and remove any unnecessary information.
- From the news articles, only the financial news will be loaded and any extra tags or information will be discarded.
- The relevant fields from the stock market data will also be parsed in similar manner.

Methodology

3. **Classify the cleaned dataset**

- Now with the cleaned data of the news articles, we will now apply sentiment analysis and try to determine if the news article is a positive, neutral or negative sentiment**[4]**. **For eg**: "**TCS introducing a new platform about AI**" or "**TCS share prices up by 7%**" is considered to be a **positive news** whereas "**TCS share drops 5%**" or "**TCS posts 2000 crore loss"** is considered to be negative news. News such as "**Earthquake near TCS headquarters, 10 people injured**" or "**TCS opens up anew tech park to accommodate more employees**" is considered to be **neutral news**.

4. **Correlation of data**

- With all the classified information we will now perform correlation between news articles of the day and the companies involved in the news which affected the corresponding day's prices. We will implement a regression model which will calculate how much the positive news affected the stock and by how much percent.
- Percentage changes involved in the companies sector will also be checked in case there is no news about the company on that particular day.
- This can be used to perform mock news articles and see how it will affect a company's stock based on previous data.

Methodology

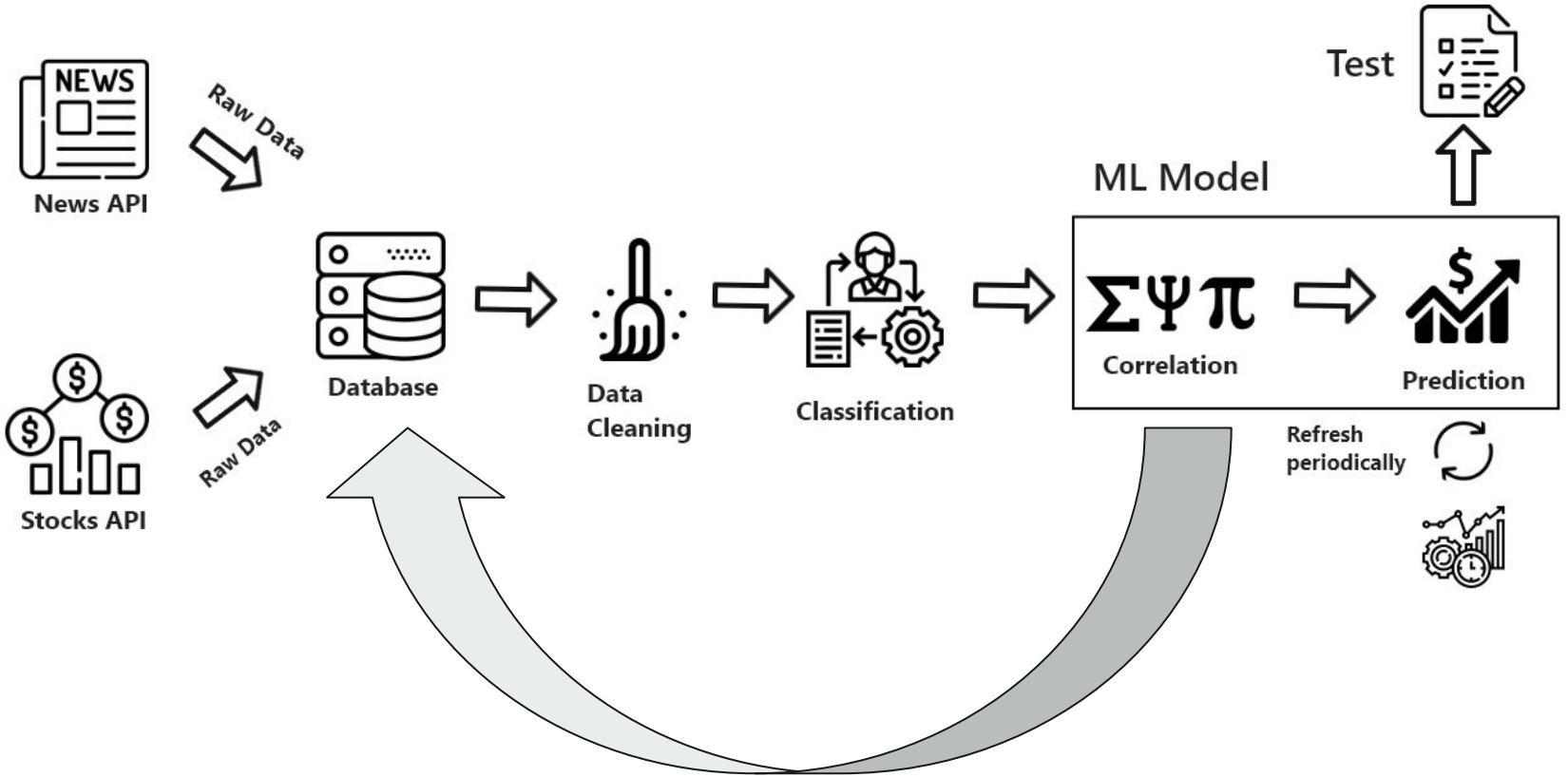
5. Prediction

- We will be monitoring all news outlets in real time along with the share market prices.
- Any news changes will be displayed to the user with the percentage change of how much the news article could affect it based on the dataset we have trained[5].

6. Performance Analysis

- We will be comparing the dataset with test cases to check and thus perform an analysis of the prediction.
- We will be comparing it to our model to check its precision, accuracy.
- A **confusion matrix** is formed from the four outcomes produced as a result of binary classification and we will use it to find accuracy of the model.
- These have 4 outcomes of positives and negatives which will be used to determine accuracy

FLOW DIAGRAM



Expected Output

Our application will display the positive or negative news along with the prediction of market movement of a company or stock in real time.

Contribution to society

- Our project aims to help layman have sufficient knowledge before investing.
- It also helps seasoned investors know in real time before investing.
- Makes a single platform for gathering of various news sources.

Requirements Specification

Hardware Requirements:

- Memory: 3GB
- CPU - Intel i3 onwards

Software Requirements:

- Python 3
- MySQL Database
- Django (Web Platform)
- BeautifulSoup4, NLTK and SpaCy for language processing
- Sklearn for machine learning

*Internet connection

References

[1] A Data mining algorithm to analyse stock market data using lagged correlation - Cicil Fonseka, Liwan Liyanage, University of Western Sydney Campbelltown, Australia l.liyanage@UWS.edu.au
[2] Predicting the Effects of News Sentiments on the Stock Market - Dev Shah , Haruna Isah, Farhana Zulkernine - Queen's University , Kingston, ON, Canada
[3] Stock Price Prediction Using Financial News Articles - M. Yasef Kaya, M. Elif Karslgil - Yildiz Technical University, Istanbul, Turkey , miykaya@yahoo.com, elif@ce.yildiz.edu.tr
[4] NLTK 3.4.5 documentation - https://www.nltk.org/
[5] Scikit Documentation - https://scikit-learn.org/stable/
[6] BeautifulSoup4 - https://www.crummy.com/software/BheautifulSoup/bs4/doc/

THANK YOU!