

TOPIC – Project Report

Stock Sentiment Analysis Using Machine Learning Techniques

The project aims to develop a sentiment analysis model to predict the movement of stock prices based on textual data from news articles, social media posts, and other sources of financial news and opinions. By analyzing the sentiment expressed in these texts, the model will seek to uncover insights into investor sentiment and market sentiment, which can be valuable indicators for making informed trading decisions.

Table Of Contents

Following is the list of contents in this report of mine

1. Overview
2. Work flow
3. Comprehensive information on how data was collected
4. Elaborating sentiment analysis
5. Constraints
6. References used
7. Author

Overview

Stock sentiment analysis using machine learning techniques involves analyzing text data (such as news articles, social media posts, or financial reports) to determine the sentiment (positive, negative, or neutral) towards a particular stock or the market. It is a powerful approach to extract insights from textual data and understand market sentiment. It combines techniques from natural language processing (NLP) with traditional machine learning and can provide valuable inputs for investment strategies and risk management in financial markets.

Work Flow

1. Data Collection:

- Obtaining textual data from various sources such as financial news websites (Yahoo Finance, Business Insider), social media platforms and financial blogs.
- Historical stock price data can also be collected.

2. Text Preprocessing:

- Clean and preprocess the text data

3. Sentiment Labeling:

- Label the textual data with sentiment labels (positive, negative, neutral). This can be done manually or using pre-labeled datasets.

4. Machine Learning Model Selection:

- Do sentiment analysis of your data using machine learning.

5. Training and Evaluation:

- Split the labeled data into training and testing sets.
- Train the chosen machine learning model on the training set and evaluate its performance on the testing set using metrics such as accuracy, precision, recall, and F1-score.

Comprehensive Information On How Data Was Collected

An extensive analysis was done by me on how to scrape financial and news data from websites.

- *Steps for web scraping data:-*

1. **Identify Target Websites**

I used Business Insider (<https://www.businessinsider.in/>) to scrape news data and external API of Yahoo Finance to scrape stock price data.

2. **Understand Website Structure**

3. **Choose a Web Scraping Tool or Library:** Using libraries like BeautifulSoup to scrape data.

4. **Write the Scraping Code**

5. **Handle Pagination and Dynamic Content**

6. **Data Cleaning and Preprocessing**

7. **Store Data:** I stored data in form of csv file.

Elaborating Sentiment Analysis

I have merged the stocks and news data on the 'date' column. Then I created a function to get subjectivity and polarity. Then in the next step I worked on a function to get Sentiment scores and stored the sentiment scores in merged datasets. Then I made featured data set and target data set and split the data into 80% training and 20% testing datasets. I performed model evaluation using appropriate metrics, such as accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curve analysis, to assess the performance of the sentiment analysis model. I also used following metrics to judge the performance - Sharpe ratio, Maximum drawdowns, number of trades executed and win ratio.

Constraints

The effectiveness and consistency of sentiment analysis can be significantly influenced by several factors, including the nuances of language, the presence of sarcasm, and the context-specific nature of interpretations. These complexities underscore the necessity for employing robust natural language processing (NLP) techniques and rigorous validation procedures. Moreover, the accessibility and quality of textual data can vary widely across different stocks and timeframes, underscoring the importance of meticulous data selection and preprocessing to ensure coherence and relevance in analysis.

References Used

1. Algo Trading using Python - Refer to free code camp.
2. Basic Blog for Sentiment Analysis overview - <https://blog.quantinsti.com/sentiment-analysis-trading/>
3. Example of a model for analysis - [Stock Market Sentiment Analysis Using Python & Machine Learning](#)

Author

Devanshi, IITR Roorkee- if you have any suggestions please let me know!