

Project Report: Sentiment Analysis and Stock Price Prediction

Overview

The objective of this project is to investigate the relationship between public sentiment and stock prices. By leveraging sentiment analysis on news headlines and financial data, we aim to develop a predictive model that could potentially inform trading decisions. The project is structured into several key steps, including data collection, sentiment analysis, stock data retrieval, sentiment-adjusted moving averages, and finally, signal generation and performance evaluation for trading strategies.

Project Flow

1. Data Collection:

- Web Scraping of News Headlines: Using Selenium, we scraped news headlines related to specific stocks from Yahoo Finance. The process involved dynamically loading pages and extracting relevant data within a 12-month timeframe.
- API Fetching from New York Times: To complement our dataset, we fetched additional news articles from the New York Times using their archive API, filtering articles related to Apple Inc.

2. Data Preprocessing:

- Date Conversion: Scraped news dates, often in relative formats like "2 hours ago" or "yesterday," were converted to actual dates.
- Merging Datasets: We merged the Yahoo Finance and New York Times datasets, ensuring a comprehensive collection of news headlines for our target stocks.

3. Sentiment Analysis:

- VADER Sentiment Scoring: Each headline was analyzed using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool to assign a polarity score, classifying the sentiment as positive, negative, or neutral.

4. Stock Data Retrieval:

- Yahoo Finance API: Stock data for the selected companies was retrieved using the yFinance API, spanning from July 1, 2018, to June 14, 2024. The data included open, high, low, close, adjusted close prices, and volume.

5. Data Integration:

- **Merging Stock and Sentiment Data:** We integrated the sentiment scores with the stock data based on the date and stock ticker.

6. Feature Engineering:

- **Sentiment-Adjusted Moving Averages:** We developed a sentiment-adjusted moving average (SMA) to account for sentiment scores alongside traditional price data.

- **Preprocessing Headlines:** The headlines were preprocessed using tokenization, lemmatization, and TF-IDF vectorization to prepare for model training.

7. Model Training and Evaluation:

- **Random Forest Classifier:** A Random Forest classifier was trained to predict sentiment labels. The model was evaluated using cross-validation and test sets.

- **Signal Generation:** Based on sentiment scores and moving averages, buy/sell signals were generated for trading decisions.

- **Performance Metrics:** We evaluated the trading strategy using performance metrics like the Sharpe Ratio, win ratio, and overall portfolio value.

8. Visualization:

- **Plotting Results:** We visualized the stock price, sentiment scores, buy/sell signals, and portfolio value over time to understand the strategy's effectiveness.

Detailed Project Flow

Data Collection

Web Scraping with Selenium: We utilized Selenium WebDriver to scrape news headlines from Yahoo Finance. By simulating user interactions, we collected news headlines related to specific stock tickers over a 12-month period.

Fetching Data from New York Times API: Using the New York Times archive API, we retrieved additional news articles that mentioned Apple Inc. The articles were filtered based on keywords related to Apple stock.

Sentiment Analysis

Using VADER for Sentiment Analysis: Each news headline was analyzed using VADER to determine its sentiment. VADER provided a polarity score for each headline, which we used to classify the sentiment as positive, negative, or neutral.

Stock Data Retrieval and Integration

Fetching Stock Data: Stock price data for the selected companies was retrieved using the yFinance API. The data included various metrics such as open, high, low, close, adjusted close prices, and trading volume.

Integrating Sentiment and Stock Data: We merged the sentiment scores with the stock data based on the date and stock ticker. This integration allowed us to analyze the impact of sentiment on stock prices.

Feature Engineering and Model Training

Sentiment-Adjusted Moving Averages: We created sentiment-adjusted moving averages (SMAs) by incorporating sentiment scores into the calculation of traditional moving averages. This adjustment aimed to better capture the influence of public sentiment on stock prices.

Random Forest Classifier for Sentiment Prediction: We trained a Random Forest classifier to predict the sentiment of news headlines. The classifier was trained on a dataset of headlines labeled with sentiment scores and evaluated using cross-validation techniques.

Signal Generation and Performance Evaluation

Trading Signals Based on Sentiment-Adjusted Moving Averages: Trading signals were generated based on sentiment-adjusted moving averages. The signals indicated potential buy or sell actions, aiming to optimize trading performance by incorporating sentiment analysis.

Performance Metrics: The trading strategy's performance was evaluated using various metrics such as the Sharpe Ratio, win ratio, and overall portfolio value. These metrics helped assess the effectiveness of the sentiment-adjusted strategy.

Conclusion

This project demonstrates the potential of integrating sentiment analysis with stock price prediction. By analyzing the sentiment of news headlines and incorporating it into traditional moving averages, we developed a sentiment-adjusted trading strategy. While the performance evaluation of the strategy will reveal its effectiveness, the approach provides a promising direction for enhancing trading algorithms with sentiment analysis.

Future Work

- Model Optimization: Further tuning and optimization of the Random Forest model and other machine learning algorithms.
- Real-time Sentiment Analysis: Implementing a real-time system to analyze live news feeds and adjust trading strategies accordingly.
- Exploration of Other Sentiment Analysis Tools: Experimenting with advanced NLP models like BERT for more accurate sentiment predictions.
- Expansion to Other Stocks and Markets: Extending the analysis to a broader range of stocks and exploring other financial markets.

By combining sentiment analysis with financial data, this project lays the groundwork for developing more sophisticated and responsive trading strategies that leverage the power of natural language processing and machine learning.