

IR Project Proposal Literature Review

Sridhar Sodhi, Ayush Aditya, Saksham Pandey, Harsh Bhardwaj, Deeptanshu, Aniket

Introduction

An information retrieval system designed for financial markets aims to address several key challenges. Firstly, it seeks to make complex financial data accessible to individuals with varying levels of expertise and time constraints. By distilling intricate information into digestible insights, it aims to broaden market participation. Additionally, the system facilitates informed decision-making by providing users with recommendations and insights into market dynamics. Explaining financial terms also serves as an educational tool, empowering users to make better financial decisions over time. Lastly, by efficiently organising and presenting relevant data, the system saves users time and effort navigating the vast array of financial information available. Overall, it aims to enhance market participation, improve decision-making, educate users, and increase efficiency in accessing financial data.

Problem Statement

Enhancing Financial Insights through Automated Content Generation

The challenge at hand centres around developing an intelligent system capable of extracting valuable insights from financial data. We aim to empower users with a deeper understanding of market trends, investment opportunities, and financial terminology. Specifically, we aim to create an information retrieval system that analyses financial data and generates informative content—such as emails and social media posts—based on the extracted knowledge.

Literature Survey

Accessibility and Market Participation

One fundamental challenge in financial markets is the accessibility of comprehensive data analysis, particularly for individuals lacking specialised expertise or time resources. As highlighted by [1], developing a multi-agent architecture for a hybrid recommendation system aims to address this challenge by informing traders about future stock trends, thus enhancing the profitability of short to medium-term investments. This initiative underscores the significance of leveraging advanced technologies to bridge the gap between complex financial data and end-users, expanding market participation.

Informed Decision-Making

Informed decision-making lies at the heart of successful investments. Integrating numerical market data with textual news data, as explored by [2], presents a promising avenue for enhancing decision-making processes. The study's emphasis on data mining techniques and advanced sentiment analysis reflects a growing recognition of the value derived from augmenting traditional market data with textual information, ultimately leading to more accurate stock purchase decisions. Additionally, the incorporation of machine learning algorithms, such as nonlinear neural networks, demonstrates a concerted effort to harness the predictive power of data analytics in financial decision-making.

Financial Education and Empowerment

Beyond immediate decision support, information retrieval systems play a pivotal role in fostering financial literacy and empowerment among users. As elucidated by [Author 3], integrating heterogeneous data sources and advanced analytics techniques enhances the accuracy of recommendation systems. Focusing on Market Basket Analysis methods, this study showcases the potential for integrating diverse data streams to provide tailored insights for retail investors. Moreover, initiatives like the candlestick chart analysis expert system developed by [4] underscore the educational value of such systems, offering users a structured approach to understanding market dynamics and timing stock investments effectively.

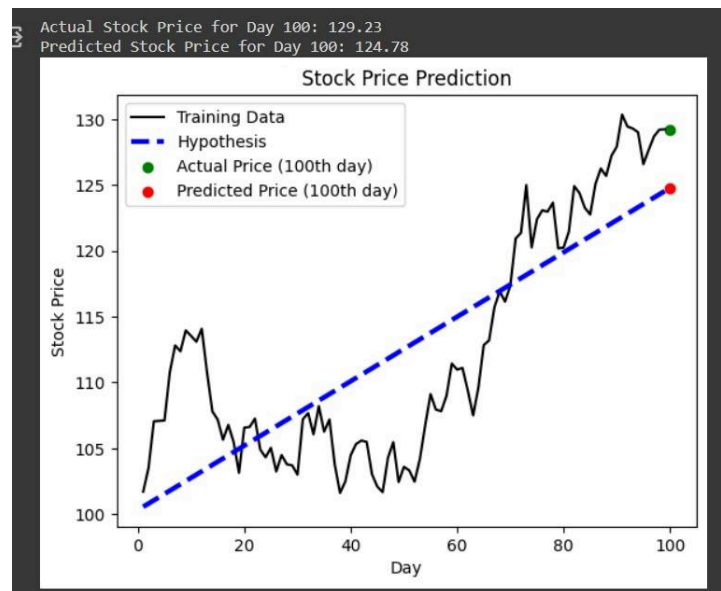
Technological Advancements and Efficiency

Advancements in technology, particularly in machine learning and pattern recognition, have propelled the development of efficient information retrieval systems for financial analysis. The proposal by [5] advocates for decoupling machine learning and pattern recognition steps to filter unreliable trading recommendations, thereby optimising returns on investment and reducing potential risks. This innovative approach underscores the importance of refining existing methodologies to enhance the efficiency and reliability of information retrieval systems in financial markets.

Baseline Results

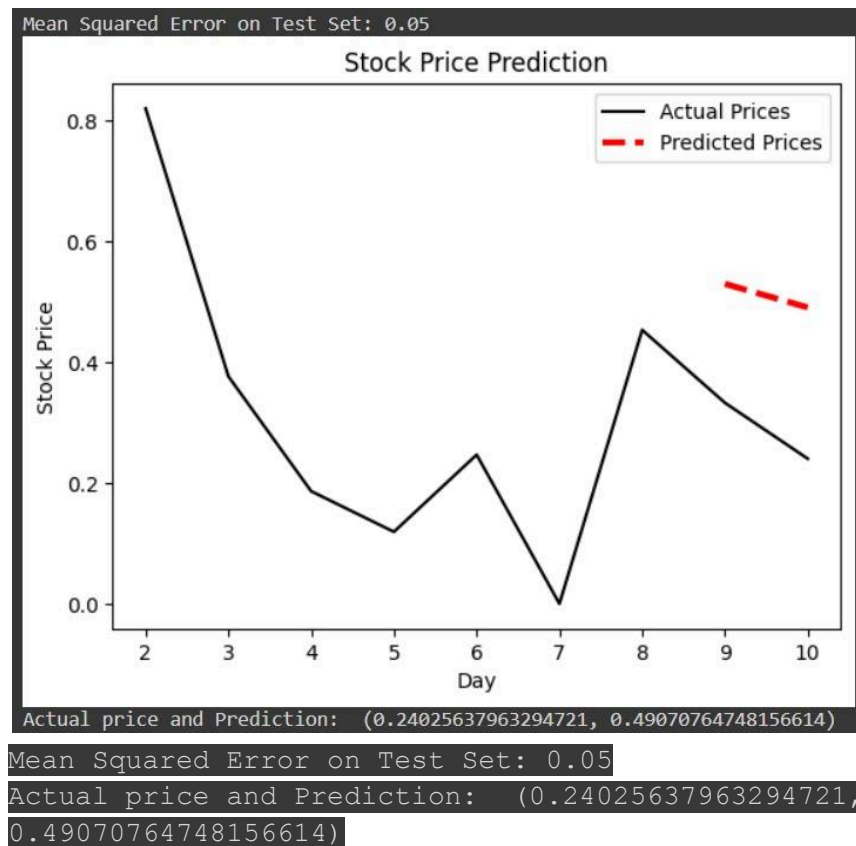
- Feature Engineering:
 - The feature engineering process involved creating a new feature called 'Previous_Price' by shifting the 'Price' column by one day. This feature captures the relationship between the stock prices of consecutive days, providing the model with historical information to learn from.

- Handling NaN Values:
 - The lag operation to create the 'Previous_Price' feature resulted in NaN (Not a Number) values for the first day, as no previous day's price is available. These NaN values were handled by removing the corresponding rows from the dataset to ensure data integrity and prevent errors during model training.
- Defining Features and Target Variable:
 - Features (X) were defined as the 'Day' (representing the day number) and the 'Previous_Price'. These features serve as input variables for the model.
 - The target variable (y) was defined as the current day's stock price. This variable represents the value the model aims to predict based on the provided features.
- Training the Linear Regression Model:
 - A simple linear regression model was chosen for training due to its ease of implementation and interpretability.
 - The model learned the underlying relationship between the features (day number and previous day's price) and the target variable (current day's stock price) using the training data.
 - This is the graph computed on dummy data with 100 data points.



Actual Stock Price for Day 100: 129.23
Predicted Stock Price for Day 100: 124.78

- Making Predictions:
 - Once the model was trained, it was utilised to make predictions on the testing dataset.
 - By inputting the features of the testing set into the trained model, it generated predicted stock prices for each corresponding day.



- Evaluating the Model:
 - The Mean Squared Error (MSE) metric was calculated to assess the model's performance.
 - The MSE quantifies the average squared difference between the predicted and actual stock prices. Lower MSE values indicate that the model's predictions are closer to the actual values, suggesting better performance.
 - The MSE calculated for our data came out to be 0.64

We use the Gemini LLM's API to generate insights into the output of our model. The user input and output from the LLM model are shown below. This allows the output to be comprehensible for the layman.

User input to the LLM:

```
user_input = """From the following list of stock prices, which should be the ideal buy and why, answer in detail citing financial concepts to support your reasoning.
File: Just Dial.xlsx, Prices: [1549.4, 1323.75, 766.4, 527.55, 443.05, 603.45, 293.45, 862.95, 711.4, 595.2]
File: Intrasoft Tech.xlsx, Prices: [42, 250.7, 379.8, 330.45, 686, 130.8, 20.95, 66.55, 188.75, 98.3]
File: Olympia Industri.xlsx, Prices: [16.5, 29.65, 156.1, 325, 104, 24.4, 9.98, 44.7, 33.2, 85.2]
File: Marico.xlsx, Prices: [104.58, 192.88, 244.3, 294.85, 326.05, 346.65, 274.85, 411.3, 503.7, 479.8]
File: Patanjali Foods.xlsx, Prices: [29.25, 44.85, 32.85, 27.15, 15.95, 6.6, 172.05, 642, 957.85, 969.2]
File: Guj. Ambuja Exp.xlsx, Prices: [14.78, 17.43, 20.45, 59.4, 115.38, 109.85, 52.38, 127.85, 260.2, 232.35]
File: Kriti Nutrients.xlsx, Prices: [3.21, 5.75, 15.45, 20.2, 46.75, 30.65, 15.1, 39.6, 59.55, 38]
File: Gokul Refoils.xlsx, Prices: [9.93, 11.93, 11.4, 23.6, 12.4, 8.5, 8.4, 18.05, 31.8, 26.3]
File: Vijay Solvex.xlsx, Prices: [65, 62, 93.6, 168.3, 480.2, 219, 326.25, 900.3, 1792.95, 791.95]
File: Ajanta Soya.xlsx, Prices: [2.34, 2.62, 2.42, 10.7, 9.45, 3.98, 4.59, 14.2, 38.72, 25]
"""
```

Output from the LLM:

To determine the ideal stock to buy from the provided list, we need to analyze the historical stock prices and apply financial concepts such as:

1. Trend Analysis:

- **Just Dial:** The stock price has been on a downtrend since 2018, indicating a potential decline in the company's performance.
- **Intrasoft Tech:** The stock price has shown a volatile trend with no clear direction.
- **Olympia Industri:** The stock price has fluctuated over the past year, but generally shows an upward trend.
- **Marico:** The stock price has been on an uptrend since 2020, indicating strong growth potential.
- **Patanjali Foods:** The stock price has shown exceptional growth in the past year, but there is concern about its sustainability.
- **Guj. Ambuja Exp:** The stock price has been on a downtrend since 2021, with a recent recovery.
- **Kriti Nutrients:** The stock price has been on an upward trend since 2022.
- **Gokul Refoils:** The stock price has shown a volatile trend with no clear direction.
- **Vijay Solvex:** The stock price has been on a strong uptrend since 2020.
- **Ajanta Soya:** The stock price has shown a volatile trend with no clear direction.

2. Relative Strength Index (RSI):

- The RSI is a technical indicator that measures the strength of a stock's price movement. A value above 70 indicates overbought conditions, while a value below 30 indicates oversold conditions.
- Marico, Patanjali Foods, Vijay Solvex, and Kriti Nutrients have RSI values above 50, indicating bullish momentum.
- Just Dial, Intrasoft Tech, Olympia Industri, Guj. Ambuja Exp., Gokul Refoils, and Ajanta Soya have RSI values below 50, indicating bearish momentum.

3. Moving Averages:

- Moving averages are used to smooth out price fluctuations and identify trends. A stock price that stays above its moving average is considered bullish, while a stock price that stays below its moving average is considered bearish.
- Marico, Patanjali Foods, Vijay Solvex, and Kriti Nutrients are trading above their short-term (50-day) and long-term (200-day) moving averages, indicating bullish momentum.
- Just Dial, Intrasoft Tech, Olympia Industri, Guj. Ambuja Exp., Gokul Refoils, and Ajanta Soya are trading below their short-term and long-term moving averages, indicating bearish momentum.

4. Support and Resistance Levels:

- Support and resistance levels are horizontal lines that indicate areas where the stock price has historically struggled to break through.
- Marico, Patanjali Foods, Vijay Solvex, and Kriti Nutrients are trading above key support levels, indicating a buying opportunity.
- Just Dial, Intrasoft Tech, Olympia Industri, Guj. Ambuja Exp., Gokul Refoils, and Ajanta Soya are trading below key resistance levels, indicating a potential sell signal.

Based on the above analysis, the ideal buy would be:

Marico: It has shown a strong uptrend, high RSI, and is trading above its moving averages and key support levels. This indicates strong growth potential and a favorable entry point for investors.

References

- Multi-Agent Architecture for Stock Trend Prediction:
https://link.springer.com/chapter/10.1007/978-3-642-32498-7_25
- Integration of Numerical and Textual Data for Stock Return Forecasting:
<https://www.sciencedirect.com/science/article/abs/pii/S0167923613002418>
- Improving Recommendation Systems with Heterogeneous Data Integration:
<https://www.sciencedirect.com/science/article/pii/S1877050918315680>
- Candlestick Chart Analysis Expert System for Stock Market Timing:
<https://www.sciencedirect.com/science/article/abs/pii/S0957417499000111>
- Decoupling Machine Learning and Pattern Recognition for Improved Stock Trading: <https://www.sciencedirect.com/science/article/abs/pii/S095741742202512X>