**Investment Strategy Suggestion System Using NLP and LSTM**

**Project Proposal**

**Team Members:**

* Hozefa Patel [N01686385]
* Jenil Pancholi [N01665133]
* Dev Ariwala [N01664568]

**Course – Natural Language Processing**

**Course Code – 5501**

**Section Code – ONA**

**Submitted to – Prof. Zeeshan Ahmad1. Project Objectives**

The primary objective of this project is to develop a system that provides buy, hold, or sell recommendations for stocks based on an analysis of financial news articles, company financial reports, and historical stock price data. The system will leverage Natural Language Processing (NLP) for sentiment analysis and Long Short-Term Memory (LSTM) networks for time series prediction.

**2. Project Scope**

This project will focus on:

1. Collecting and preprocessing financial news articles, company financial reports, and historical stock price data.
2. Performing sentiment analysis on news articles to gauge market sentiment.
3. Integrating sentiment scores with financial ratios and stock price data.
4. Using LSTM networks to predict stock price movements.
5. Generating buy, hold, or sell recommendations based on the predictions.
6. Providing an interface for users to input stock names and receive real-time recommendations.

**3. Background and Motivation**

Investing in the stock market requires analyzing vast amounts of data, including financial news, company reports, and stock prices. Traditional methods of analysis can be time-consuming and prone to human error. By leveraging modern machine learning techniques such as NLP and LSTM, this project aims to automate and enhance the accuracy of investment strategy recommendations. This system will help investors make informed decisions quickly and efficiently.

**4. Methodology**

**Step 1: Data Collection**

* **APIs for News Articles**: Utilize NewsAPI, Alpha Vantage, or Finnhub to fetch financial news articles.
* **Stock Data**: Use Yahoo Finance, Alpha Vantage, or IEX Cloud to download historical and real-time stock price data.
* **Financial Reports**: Scrape the SEC EDGAR database or company websites for financial reports.

**Step 2: Text Preprocessing**

* Convert text to lowercase, remove punctuation, and tokenize the text.
* Remove stop words and perform stemming/lemmatization.

**Step 3: Sentiment Analysis**

* Use TextBlob, VADER, or Hugging Face Transformers for sentiment analysis of news articles.

**Step 4: Feature Extraction**

* Transform text data into numerical features using TF-IDF or word embeddings.
* Extract financial ratios from company reports.

**Step 5: Data Integration**

* Merge sentiment scores with financial ratios and historical stock prices.

**Step 6: LSTM for Time Series Prediction**

* Scale the data and create sequences for LSTM input.
* Build and train the LSTM model to predict stock price movements.

**Step 7: Recommendation Generation**

* Define logic to generate buy, hold, or sell recommendations based on model predictions.

**Step 8: User Interface**

* Develop a command-line or web-based interface for users to input stock names and receive recommendations.

**5. Tools and Technologies**

* **Programming Languages**: Python
* **Libraries**: yfinance, requests, BeautifulSoup, pandas, scikit-learn, TextBlob, tensorflow, keras
* **APIs**: NewsAPI, Alpha Vantage, Yahoo Finance, Finnhub
* **Platforms**: SEC EDGAR, company websites
* **Development Tools**: Jupyter Notebook

**6. Expected Outcomes**

* A functional system that provides real-time buy, hold, or sell recommendations for stocks.
* A user-friendly interface for inputting stock names and receiving recommendations.
* A detailed analysis of the system’s performance and potential improvements.

**7. Challenges and Mitigation**

* **Data Quality**: Ensuring the quality and relevance of data collected. Mitigation: Use reliable data sources and perform thorough preprocessing.
* **Model Accuracy**: Achieving high accuracy in predictions. Mitigation: Experiment with different model architectures and hyperparameters.
* **Integration of Multiple Data Sources**: Effectively integrating sentiment, financial ratios, and stock price data. Mitigation: Use robust data merging techniques and feature engineering.

**8. Conclusion**

This project aims to develop a robust investment strategy suggestion system that leverages NLP and LSTM for accurate and real-time stock market recommendations. By automating the analysis process and providing actionable insights, the system will assist investors in making informed decisions, ultimately enhancing their investment outcomes.