ECGR 5106 – Real Time Machine Learning

Final Project Proposal

Project Title: Real Time Stock Prediction

Team Members: Nahush Tambe, Naseeruddin Lodge, & Patrick Flynn

Project Manager: Nahush Tambe

Summary and Broader Impact: The Real-Time Stock Price Prediction project is an ambitious effort to develop a machine learning model that can accurately predict the stock price movements of 20 major publicly traded American companies across a range of sectors, including Finance, Automobile, Technology, Communications, Consumer Goods, and Energy. By leveraging historical stock price data and technical indicators, the project aims to build a powerful predictive model using cutting-edge machine learning techniques such as LSTM and GRU neural networks.

The potential impact of this project is vast. With real-time insights into stock price movements, investors and traders can make more informed decisions, potentially resulting in better returns on investment. Moreover, by increasing financial literacy and making complex financial concepts more accessible to the general public, the project can help people make more informed financial decisions, leading to greater financial stability and improved overall economic outcomes.

The project can also support innovation and growth in the technology sector and beyond, as investors and companies alike use the insights generated to inform their investment and business decisions. This, in turn, can drive economic growth, job creation, and improved living standards.

While there are risks associated with any predictive model, such as over-reliance on past data or unforeseen market fluctuations, the Real-Time Stock Price Prediction project has the potential to provide significant benefits to investors, traders, companies, and society as a whole. By improving market efficiency and increasing financial literacy, this project can help build a more prosperous and sustainable future for all.

Project Details: The Real-Time Stock Price Prediction project will begin with the critical task of collecting and preprocessing data. The Huge Stock Market Dataset, available on Kaggle, will serve as the primary dataset for the project, containing historical daily prices and volumes of all U.S stocks and ETFs. However, the focus will be narrowed down to the data required for the selected 20 companies.

Once the data is collected, the next step will involve concatenating the separate files into a big DataFrame with over 150,000 data points. This step will require careful data preprocessing, including data cleaning, normalization, and feature engineering. After preprocessing, an appropriate model will be selected to predict stock prices, such as LSTM, GRU, or a combination of both. The model will then be trained with the available data while optimizing the hyperparameters to achieve the best possible results.

Once the training is complete, the model's performance will be evaluated using a test dataset, comparing the actual and predicted stock prices using metrics such as mean squared error (MSE) or root mean squared error (RMSE).

In addition to the primary objectives of the project, there will be further analysis to identify the factors that most influence stock price movements. This will be achieved by conducting a feature importance analysis to determine the most significant factors driving stock price changes.

Overall, the Real-Time Stock Price Prediction project is a complex and challenging task, requiring a comprehensive data collection and preprocessing phase, model selection, training, and testing. However, the potential benefits of the project are enormous, including enabling investors and traders to make informed decisions and improving overall financial literacy, supporting innovation and growth in the technology sector, and driving economic growth and development.

Individual Responsibility:

Nahush: I will be responsible for managing the project, ensuring that it progresses smoothly and is completed on time. This includes coordinating with team members, setting deadlines, and ensuring that everyone is working towards the project's objectives. I will also be responsible for gathering and preprocessing the data needed for the project, such as historical stock prices, and combining it into a single file for the model to use.

Patrick: Patrick will be responsible for developing the LSTM and GRU models and training them using the preprocessed data. This includes selecting the appropriate hyperparameters and adjusting them to optimize the model's performance. Patrick will also be responsible for troubleshooting any issues that arise during the training process and ensuring that the models are performing optimally.

Naseeruddin: Naseeruddin will be responsible for evaluating the trained models using the testing dataset. This includes using metrics such as mean squared error (MSE) or root mean squared error (RMSE) to compare the actual and predicted stock prices. Naseeruddin will also be responsible for identifying any areas where the models can be improved and providing recommendations to the team for further optimization.

As far as the project report is concerned, it will be a combined team effort. All team members will contribute to the report, documenting their contributions, discussing the results, and providing recommendations for future work. Nahush will be responsible for ensuring that the report is complete, well-organized, and submitted on time.

Additional Goals (If Possible): If we have enough time and resources, it is possible to incorporate a chatbot into this project. The chatbot could be designed to provide real-time insights into stock prices, answer common questions related to investing, and offer personalized investment advice based on the user's investment profile. This could enhance the user experience by providing a more interactive and personalized experience, making complex financial concepts more accessible and understandable to the general public. Additionally, a chatbot could allow

investors to make more informed decisions about their investments and stay up to date with the latest market trends.

However, building a chatbot would require additional resources, including more work with natural language processing and conversational AI, and would need to be carefully integrated into the existing project workflow.