MANIKANTA NAIDU DOREPALLI

B00116737

Question 1: Elevator Pitch: Our project aims to predict Google stock prices using a Recurrent Neural Network (RNN). Leveraging historical stock market data, the RNN will capture temporal patterns and trends, providing reliable predictions for informed decision-making. This project will integrate advanced deep learning techniques and financial data analysis to offer a robust stock prediction model.

Question 2: Dataset Details

1) Collector(s):

Yahoo Finance or Alpha Vantage (depending on the selected source for stock market data).

2) Year:

The dataset will include data spanning the past 10 years (e.g., 2014–2024).

3) Title of Dataset:

"Google Stock Historical Data"

4) Version Number (if any):

No specific version number; the data is updated dynamically by the provider.

5) Publisher:

Yahoo Finance or Alpha Vantage.

6) DOI or URL:

- **Yahoo Finance:** https://finance.yahoo.com/
- **Alpha Vantage:** https://www.alphavantage.co/

7) Study/Paper/Reason:

The data was collected for financial analysis, stock market research, and algorithmic trading purposes.

Question 3: Language and Libraries

Language:

- Python 3.13.1

Libraries:

- **Data Collection & Preprocessing:** Pandas, NumPy
- **Visualization:** Matplotlib, Seaborn
- **Model Building:** TensorFlow/Keras, PyTorch
- **Evaluation:** Scikit-learn
- **Data Access:** Alpha Vantage or Yahoo Finance API

Question 4: Code will write our own

- **Data Preprocessing:** Cleaning and normalizing stock data, handling missing values, and creating time-series datasets.
- **Feature Engineering:** Selecting features such as closing price, volume, and moving averages.

- **Model Architecture:** Defining and training the RNN model (e.g., LSTM or GRU).
- **Hyperparameter Tuning:** Writing code to experiment with learning rates, batch sizes, and sequence lengths.
- **Evaluation Metrics:** Implementing functions to calculate metrics like RMSE and MAE.
- **Visualization:** Plotting actual vs. predicted stock prices.

Question 5: Best Choice of Model

Model Choice:

Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU)

- **Why:** These models are designed to capture long-term dependencies in time-series data, making them ideal for stock market predictions, where historical trends significantly influence future prices.

Question 6: Hyperparameters and Optimization

Key Hyperparameters:

1. Learning Rate
2. Batch Size
3. Sequence Length (number of past days considered)
4. Number of LSTM/GRU units
5. Dropout Rate

Optimization Strategy:

- Use a grid search or random search for initial hyperparameter tuning.
- Refine using Bayesian optimization or Hyperband for efficiency.
- Cross-validate results to ensure robustness.

Question 7: Performance Evaluation

Metrics:

1. **Root Mean Squared Error (RMSE):** To measure prediction accuracy.
2. **Mean Absolute Error (MAE):** To evaluate the average deviation.
3. **R-squared (R^2):** To assess how well the model fits the data.

Techniques:

- Compare predicted vs. actual stock prices on unseen test data.
- Visualize predictions and trends to assess alignment with real data.