Name: **Shahid Ali Khowaja**

Student ID: **23034289**

Project Overview:

Project Title: **Predicting NVIDIA Stock Prices Using Machine Learning Models**

Summary of the project topic: Stock price prediction is one of the most essential activities for investors and analysts in financial markets. With the recent uprising of machine learning, modern forecasting techniques have outperformed traditional models regarding accuracy and adaptability. NVIDIA is a leading technology company that shows high volatility in stock prices, which is influenced by factors such as AI advancements and semiconductor demand. This project compares LSTM and ARIMA models in predicting NVIDIA's stock prices by analysing historical trends and market fluctuations. It will determine which model, either LSTM or ARIMA, is better regarding predictive accuracy, stability, and adaptability in volatile conditions.

This refined question allows us to analyse the following:

1. Short-term vs. Long-term Forecasting:
    - Does LSTM perform better for short-term (daily/weekly) predictions?
    - Does ARIMA provide better results for long-term (monthly/yearly) trends?
2. Predictive Accuracy:
    - Which model has lower error rates (MAE, RMSE, R^2)?
    - Which machine learning algorithms (e.g., Linear Regression, Random Forest, LSTM, ARIMA) yield the most accurate predictions for stock prices?
    - Does LSTM outperform ARIMA in volatile periods?
3. Market Volatility Adaptability:
    - How do the models handle sudden stock price fluctuations?
    - Are the predictions reliable in highly volatile periods?

Project Objectives

The primary objective of this project is to develop and evaluate machine learning models for predicting NVIDIA's stock prices using historical market data. Specifically, the project aims to:

1. Develop Predictive Models:
    - Implement LSTM (Long Short-Term Memory) for deep learning-based forecasting.
    - Apply ARIMA (AutoRegressive Integrated Moving Average) for statistical time-series forecasting.
2. Compare Short-Term vs. Long-Term Forecasting:
    - Analyze the performance of models in short-term predictions and long-term predictions.
3. Evaluate Model Accuracy:
    - Assess performance using error metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.
4. Analyse Market Volatility Impact:

- Determine how well LSTM and ARIMA models adapt to sudden fluctuations or external events affecting NVIDIA stock prices.

5. Provide Insights for Investors & Traders:

   - Identify which model offers the most reliable predictions for different market conditions.
   - Evaluate whether machine learning (LSTM) outperforms traditional statistical methods (ARIMA) in stock price forecasting.

Reference List:

1. https://www.researchgate.net/publication/347611169_Comparison_of_ARIMA_ANN_and_LSTM_for_Stock_Price_Prediction
2. https://royalsocietypublishing.org/doi/10.1098/rsos.240699
3. https://www.researchgate.net/publication/370573980_A_Comparison_of_Linear_Regression_LSTM_model_and_ARIMA_model_in_Predicting_Stock_Price_A_Case_Study_HSBC's_Stock_Price

Project Plan:

Task List & Timeline:

Below is a structured project plan outlining key tasks, deadlines, and assessment preparation activities. The project is divided into four phases: Research and Data Collection, Model Development, Evaluation and optimization, and Report Writing.

Task List with Dates

No	Task Name	Description	Start Week	End Week
1	Define Research Question & Objectives	Finalize research question, objectives, and scope of the study.	6	10
2	Literature Review	Gather and review at least 3 relevant academic papers on LSTM, ARIMA, and stock price forecasting.		
3	Data Collection	Download NVIDIA stock data from Kaggel		
4	Data Preprocessing	Clean and preprocess stock data (handle missing values, normalize data, create new features).		
Phase 2: Model Development				
5	Implement ARIMA Model	Train and test the ARIMA model for time-series forecasting.	11	12
6	Implement LSTM Model	Train and test the LSTM deep learning model for stock price prediction.		
7	Compare Models	Evaluate model performance using MAE, RMSE, and R2 metrics. Compare LSTM vs. ARIMA.		
Phase 3: Evaluation & Optimization				
8	Hyperparameter Tuning	Optimize LSTM and ARIMA models using grid search and hyperparameter tuning.	12	13
9	Market Volatility Analysis	Assess how models perform under market fluctuations and economic events.		
10	Trading Strategy Development	Develop a simple trading strategy based on model predictions.		

Phase 4: Report Writing & Finalization				
11	Draft Report	Write the initial draft of the report (Introduction, Methodology, Results, Discussion).	14	17
12	Review & Proofreading	Revise the report, improve clarity, check citations, and proofread.		
13	Final Submission	Submit the completed project report.		

Data Management Plan:

1. **Description of the Data:** The dataset contains 1,699 entries and 7 columns related to NVIDIA's stock price movements. The data represents daily trading activity, including Open, High, Low, Close, Adjusted Close Prices, and Trading Volume. However, the dataset has issues, such as non-numeric values in the first two rows, missing values in some columns, and object-type numerical data that needs cleaning.

2. **Columns in the Dataset:**

Name	Description
Date	Likely represents date information.
Adj Close	Adjusted closing price after corporate actions.
Close	Official closing price of NVIDIA stock for the day.
High	Highest price recorded during the trading session.
Low	Lowest price recorded during the trading session.
Open	Opening price of NVIDIA stock on that trading day.
Volume	Total number of shares traded during the day.

3. **Origin of the Data:**
    - Source: The dataset is most likely collected from financial market data providers
        - Kaggle (https://www.kaggle.com/datasets/muhammaddawood42/nvidia-stock-data)
        - The data will be downloaded as a CSV file from Kaggle.
    - Original Purpose:
        - Financial institutions, stock traders, and AI researchers collect stock data to analyse price trends, predict future stock movements, and develop trading strategies.
        - The dataset is typically used for technical analysis, algorithmic trading, and financial forecasting.

4. Document control
    - Version Control: https://github.com/smkhowaja08/data-science-project