KINGSTON
ENGINEERING COLLEGE

COLLEGE CODE: 5113

APPLIED DATA SCIENCE

Project No.6 - STOCK PRICE PREDICTION

Batch Members:

1. ABHINANDAN A - (au511321104001) - abhiarjun0610@gmail.com
2. ABISHEK V P - (au511321104002) – abishekreddy7112003@gmail.com
3. MAHEDHAR V – (au511321104051) - vmahedhar7@gmail.com
4. BHAKTHULA CHANDU – (au511321104009)- chanchandu353@gmail.com

PHASE 2: INNOVATION

INTRODUCTION:

The problem at hand is of significant importance in the world of finance and investment. Developing a predictive model for forecasting stock prices is a formidable challenge, but the potential benefits are equally substantial. The primary aim of this project is to equip investors with a powerful tool that enables them to navigate the complex and often unpredictable world of financial markets with confidence.

By creating a reliable predictive model, we seek to provide investors with the foresight they need to make well-informed decisions. In financial markets, where volatility and uncertainty are constants, such a tool can be invaluable. It allows investors to identify trends, spot opportunities, and mitigate risks effectively. The path to achieving this goal is a structured and multifaceted one. It starts with the collection of historical stock market data, including a wealth of information that encapsulates the intricacies of market behaviour over time. This data, in its raw form, is often messy and unstructured, necessitating thorough preprocessing.

Feature engineering is a key component, as it empowers the model to capture intricate patterns within the data. Selecting the right model is pivotal, and this decision must be informed by the dataset's characteristics and the specific nature of the problem. Training and evaluation, the final steps, require meticulous attention to detail. The chosen model must be honed to offer accurate predictions, and its performance must be rigorously assessed using appropriate metrics. In essence, this project seeks to transform data into knowledge and, ultimately, into an innovative solution that enhances the investment decisions of individuals and institutions alike.

ABOUT DATASET:

Where did we get the dataset?

Kaggle:

The dataset provided on Kaggle, titled "Microsoft Lifetime Stocks Dataset" (accessible at https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset), offers a valuable resource for our project aimed at forecasting stock prices. This dataset primarily focuses on Microsoft's stock market performance over a substantial period, making it an excellent choice for our predictive modelling task.

Dataset Details:

This dataset comprises a comprehensive set of attributes that are essential for analysing and forecasting stock prices:

1. **Date:** A crucial component of time series data, the date allows us to track stock price changes over time.
2. **Open Price:** This represents the opening price of Microsoft's stock on a given trading day. It is one of the primary indicators of daily market dynamics.

3. **High Price:** The highest price reached during the trading day, providing insights into intraday fluctuations.
4. **Low Price:** The lowest price recorded on the same trading day, indicating the day's lowest level of market activity.
5. **Close Price:** The closing price of Microsoft's stock, which holds significance as it often reflects investors' sentiment and can influence trading decisions.
6. **Volume:** This attribute records the trading volume for the stock on a specific date, helping to identify days of high market activity.

This dataset spans a significant time frame, which is vital for training and testing our predictive model. The availability of additional features, such as high and low prices, further enriches the dataset, enabling us to capture a wide range of market dynamics. Overall, the "Microsoft Lifetime Stocks Dataset" on Kaggle is an ideal resource for our stock price forecasting project.

COLUMNS THAT WE USED*:*

In our stock price forecasting project using the "Microsoft Lifetime Stocks Dataset," we carefully select specific columns for analysis and model training. These columns have been chosen based on their critical relevance to understanding and predicting stock price movements.

1. **Date:** The date column is the fundamental aspect of time series data, allowing us to track stock price changes over

time. It serves as the temporal reference point for our predictions.

2. **Open Price:** The opening price of a stock on a given trading day is of paramount importance. It sets the initial tone for the market, and analysing its historical patterns can provide insights into market sentiment.
3. **Close Price:** The closing price, reflecting the end-of-day value, often plays a central role in trading decisions. Its historical trends are invaluable for predicting future prices.
4. **Volume:** Trading volume indicates market activity, with high volumes often accompanying significant price movements. Analysing this attribute helps us identify days of heightened market interest.
5. **High Price and Low Price:** High and low prices provide a range of prices within a trading day, allowing us to assess intraday fluctuations. This information is vital for understanding daily volatility and price ranges.

These selected columns offer a comprehensive view of Microsoft's stock performance, incorporating temporal and pricing aspects along with trading activity. By focusing on these attributes, we aim to capture the essential features that influence stock prices. This meticulous column selection is integral to building a predictive model that can provide investors with valuable insights for making informed decisions and optimizing their investment strategies.

DETAILS OF LIBRARIES USED AND WAY TO DOWNLOAD:

In our stock price forecasting project, we leverage a range of powerful libraries to facilitate data manipulation, model development, and evaluation. These libraries play a crucial role in transforming our design into an innovative solution. Here are the key libraries and a detailed explanation of how to download and install them:

1. Pandas:

- **Download:** Pandas can be installed using Python's package manager, pip. Open a command prompt or terminal and run **pip install pandas**.
- **Installation:** After downloading, Pandas can be imported in your Python script using **import pandas as pd**.

2. NumPy:

- **Download:** NumPy can also be installed using pip. Run **pip install numpy** in your command prompt or terminal.
- **Installation:** Import NumPy in your Python script using **import numpy as np**.

3. Matplotlib and Seaborn:

- **Download:** These visualization libraries can be installed with pip. Use **pip install matplotlib seaborn** in your command prompt or terminal.

- **Installation:** Import Matplotlib and Seaborn in your Python script using **import matplotlib.pyplot as plt** and **import seaborn as sns**, respectively.

4. Scikit-Learn:

- **Download:** Scikit-Learn can be installed with pip. Run **pip install scikit-learn** in your command prompt or terminal.
- **Installation:** Import Scikit-Learn in your Python script using **import sklearn**.

5. Keras with TensorFlow Backend:

- **Download:** Install Keras with TensorFlow using **pip install tensorflow keras** in your command prompt or terminal.
- **Installation:** Import Keras in your Python script using **import keras**.

6. Jupyter Notebooks:

- **Download:** You can install Jupyter Notebooks by running **pip install jupyter** in your command prompt or terminal.
- **Installation:** After installation, start a Jupyter Notebook server by running **jupyter notebook** in your command prompt or terminal. This will open a web-based interface for creating and running notebooks.

Once these libraries are downloaded and installed, you can import them into your Python scripts to access their functionality. These libraries provide a robust ecosystem for data analysis, machine learning, and data visualization, essential for transforming your design into an innovative solution for stock price forecasting.

HOW TO TRAIN AND TEST:

Training and testing a predictive model for stock price forecasting involves a series of innovative steps aimed at achieving accurate and reliable results. The process is integral to transforming our design into a practical and impactful solution. Here's a detailed explanation of how to train and test our model:

1. Data Preprocessing:

- Begin by preparing the dataset. Clean the data, address missing values, and convert categorical features into numerical representations. Ensuring data consistency is crucial for accurate modelling.

2. Feature Engineering:

- Enhance the dataset by creating additional features. This could include generating moving averages, technical indicators, and lagged variables. These engineered features provide valuable insights into stock price trends.

3. Model Selection:

- Choose an appropriate model for stock price forecasting. Given the nature of the data and the complexities of financial markets, Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), can be an innovative choice. LSTMs excel at capturing sequential patterns in time series data.

4. Splitting the Dataset:

- Divide the pre-processed data into training and testing sets. A common approach is an 80-20 or 70-30 split, where the larger portion is used for training and the rest for testing.

5. Hyperparameter Tuning:

- Fine-tune the model's hyperparameters, such as the number of LSTM units, batch size, and learning rate, to optimize its performance. This is an iterative process involving experimentation to find the best settings.

6. Training the Model:

- Use the training dataset to train the LSTM model. The model learns from historical patterns and relationships between features to make predictions. Training can be done for multiple epochs, ensuring that the model converges to an optimal state.

7. Model Validation:

- After training, use the testing dataset to validate the model's performance. The model makes predictions on this data, and the results are compared to the actual stock prices.

DESIGN INTO INNOVATION:

In the transition from design to innovation, our aim is to take the conceptual framework for stock price forecasting and transform it into a practical and impactful solution. This phase marks the journey from theory to implementation, with a strong emphasis on turning our design into a real-world application.

1. Data Integration and Preprocessing: The first step involves integrating the "Microsoft Lifetime Stocks Dataset" and applying rigorous data preprocessing. This stage ensures that the data is cleaned, any missing values are addressed, and feature engineering is executed to extract valuable insights.
2. Model Development: Here, we put theory into action by selecting and implementing the chosen predictive model, such as the LSTM neural network. This involves configuring the model architecture and fine-tuning hyperparameters to optimize its predictive capabilities.
3. Training and Testing: We feed historical data into the model, allowing it to learn from past trends and patterns. The training process involves iterative adjustments and optimization for maximum accuracy. Subsequently, the model is rigorously tested on unseen data to evaluate its forecasting performance.

4. Continuous Improvement: Innovation doesn't stop at the initial implementation. We iterate and improve by continually monitoring the model's performance. Ongoing assessment helps us adapt to changing market conditions and enhance accuracy.
5. Interpretability and Visualization: Transforming the model's output into insights is vital. Visualizations and interpretability tools help investors understand the rationale behind predictions, making it easier to act on them.
6. Deployment: To drive innovation, our model is deployed to a real-world environment where it can be used to support investment decisions. It becomes a practical tool for investors to leverage.
7. Feedback Loop: To ensure sustained innovation, we establish a feedback loop. We collect feedback from users, monitor the model's performance, and continuously update and improve its accuracy and usability.

This phase represents the transition from theoretical design to a tangible innovation—a predictive tool that empowers investors to navigate financial markets with confidence. Through data-driven insights and machine learning, this innovation enhances decision-making, optimizes investment strategies, and ultimately reshapes the way we approach stock price forecasting.

METRICS USED FOR ACCURACY CHECK:

In our stock price forecasting project, accurate evaluation is essential to assess the performance of our predictive model. The choice of appropriate metrics is crucial, as it directly impacts our ability to measure the model's accuracy and effectiveness. We employ the following key metrics for this purpose:

1. **Mean Absolute Error (MAE):** MAE calculates the average of the absolute differences between the predicted and actual stock prices. It provides a straightforward and interpretable measure of the model's predictive accuracy. A lower MAE indicates a better fit of the model to the data.
2. **Root Mean Squared Error (RMSE):** RMSE measures the square root of the average of the squared errors between predicted and actual prices. RMSE gives more weight to larger errors, making it sensitive to outliers. It is a valuable metric when we want to penalize larger prediction errors more significantly.
3. **Mean Absolute Percentage Error (MAPE):** MAPE calculates the percentage difference between predicted and actual prices. This metric is particularly useful for understanding the relative accuracy of predictions. It quantifies the error as a percentage of the actual value, making it easy to interpret.
4. **R-squared (R2):** R-squared measures the proportion of the variance in the stock price that the model can explain. It indicates how well the model fits the data, with a higher R2

value suggesting a better fit. However, it should be used in conjunction with other metrics, as it doesn't account for prediction errors.

5. **Back testing:** Back testing is a practical metric for assessing the model's performance in a trading scenario. It involves implementing the model's predictions in a simulated trading environment and measuring the strategy's profitability and risk.

These metrics provide a comprehensive view of the model's accuracy, its ability to capture trends and patterns, and its overall performance in forecasting stock prices. By employing a combination of these metrics, we ensure a robust evaluation of the predictive model, enabling investors to make well-informed decisions and optimize their investment strategies with confidence.

CONCLUSION:

In conclusion, our stock price forecasting project is built on a foundation of rigorous evaluation. The chosen metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R-squared (R2), and practical back testing, serve as reliable measures of the model's accuracy and performance. By carefully assessing the model's predictions, we ensure that it provides investors with trustworthy insights into stock price movements. These metrics are instrumental in our mission to empower investors to make well-informed decisions, optimize their investment strategies, and navigate the complex world of financial markets with confidence.