KINGSTON
ENGINEERING COLLEGE

COLLEGE CODE: 5113

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

APPLIED DATA SCIENCE

Project No.6 - STOCK PRICE PREDICTION

PREDICTIVE MODEL FOR STOCK PRICE FORECASTING: PROBLEM UNDERSTANDING AND DESIGN APPROACH

Abstract:

Forecasting stock prices is a critical task that can greatly impact investors' decision-making and investment strategies. In this document, we will comprehensively address the problem statement of building a predictive model for stock price forecasting. We will outline the steps involved in understanding the problem, data collection, preprocessing, feature engineering, model selection, training, and evaluation. This document serves as a guide for approaching this complex task with a structured and effective design thinking approach.

1. Problem Definition:

The problem at hand involves developing a predictive model that forecasts stock prices based on historical market data. The goal is to create a tool that assists investors in making well-informed decisions and optimizing their investment strategies. This project entails multiple stages, each of which plays a pivotal role in the model's accuracy and effectiveness.

1.1 problem understanding:

Certainly, the problem at hand is to create a predictive model for forecasting stock prices using historical market data. The primary objective is to provide investors with a tool that helps them make well-informed decisions and optimize their investment strategies. This involves collecting and structuring data, enhancing the model's predictive capabilities through feature engineering, selecting an appropriate forecasting model, training it, and evaluating its performance. The overall goal is to develop a reliable and accurate stock price prediction tool for investors.

2. Design Thinking:

2.1 Data Collection:

The first step in solving this problem is gathering historical stock market data. The dataset should contain a variety of features, including the date, open price, close price, volume, and other relevant indicators. This comprehensive data is essential for training a model that can capture the nuances of stock price movements.

2.2 Data Preprocessing:

Clean and structured data is a prerequisite for any meaningful analysis. Data preprocessing involves:

- Handling missing values: Identify and address missing data points to avoid errors during analysis.
- Converting categorical features: Convert categorical attributes into numerical representations suitable for model training.
- Ensuring data consistency: Ensure the data is consistent and formatted correctly for analysis.

2.3 Feature Engineering:

Feature engineering plays a crucial role in enhancing the predictive power of the model. In this step, we will:

- Create additional features: Generate new features that might provide valuable insights into stock price movements, such as moving averages, technical indicators, and lagged variables.
- Normalize and scale data: Ensure that all features are on the same scale to prevent certain attributes from dominating the modelling process.

2.4 Model Selection:

Choosing the right algorithm is pivotal to the success of the predictive model. Potential models include:

- ARIMA (Auto Regressive Integrated Moving Average): A time series forecasting model widely used for its simplicity and effectiveness.

- LSTM (Long Short-Term Memory): A type of recurrent neural network (RNN) designed for sequence prediction, suitable for capturing complex patterns in time series data.

The model selection should consider the characteristics of the dataset, the nature of the problem, and the computational resources available.

2.5 Model Training:

Once the model is selected, it needs to be trained with the pre-processed data. This step involves:

- Splitting the data: Divide the dataset into training and testing sets to assess the model's performance accurately.
- Hyperparameter tuning: Optimize the model's hyperparameters to achieve the best forecasting results.
- Handling overfitting and underfitting: Implement techniques to prevent the model from being overly complex or too simplistic.

2.6 Evaluation:

Evaluating the model's performance is a critical aspect of this project. The performance can be assessed using various time series forecasting metrics, including but not limited to:

- Mean Absolute Error (MAE): Provides an average of the absolute errors between predicted and actual values.
- Root Mean Squared Error (RMSE): Measures the square root of the average of squared errors, giving higher weight to larger errors.
- Mean Absolute Percentage Error (MAPE): Calculates the percentage difference between predicted and actual values.

The chosen metrics should align with the project's objectives and the specific needs of investors.

Conclusion:

In conclusion, building a predictive model for stock price forecasting is a multifaceted task that involves several critical stages. By understanding the problem statement and following a structured design thinking approach, we can develop a powerful tool that empowers investors with accurate forecasts. Proper data collection, preprocessing, feature engineering, model selection, training, and evaluation are all integral

components of this process. It is essential to choose the right techniques and metrics to ensure the model's success in this challenging endeavor.