****

**Assessment Report**

on

**“STOCK PRICE PREDICTION”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Hiten Mehra

Roll Number : 202401100300126

Section: B

Name : Dhruv Kumar Sharma

Roll Number : 202401100300102

Section: B

Name : Dhruv Kesarwani

Roll Number : 202401100200101

Section: **B**

Name : Himanshu Maurya

Roll Number : 202401100400124

Section: B

**Under the supervision of**

“Shivansh Prasad”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

## Stock Price Prediction Report Project Title: Stock Price Prediction Using Regression Dataset: [Nifty 50 Stock Market Data on Kaggle](https://www.kaggle.com/datasets/rohanrao/nifty50-stock-market-data) Objective: To build a regression model that predicts the next-day stock closing price based on historical data and to visualize trends while evaluating the model’s accuracy.

## 1. Introduction

## Predicting stock prices is a significant challenge in quantitative finance. This project aims to leverage regression techniques to predict next-day closing prices for stocks listed in the Nifty 50 index using historical data. The model uses features like open price, high, low, volume, and previous close to make predictions.

## 2. Dataset Overview

## Source: Kaggle - Nifty 50 Stock Market Data by Rohan Rao

## Time Range: Multiple years for each stock (up to 2020)

## Columns:

## Symbol: Ticker symbol of the stock

## Date: Date of trading session

## Open: Opening price

## High: Highest price of the day

## Low: Lowest price of the day

## Close: Closing price

## Shares Traded: Volume of shares traded

## Turnover (₹ Cr): Total turnover in crores

## 3. Data Preprocessing

## Data Cleaning:

## Removed missing or duplicate values.

## Converted Date column to datetime format.

## Sorted values by Date for time-series integrity.

## Feature Engineering:

## Created lag features (e.g., previous day’s close, volume).

## Calculated rolling means and standard deviations (e.g., 5-day moving average).

## Target Variable:

## Next\_Close = Close price of the following day (shifted -1)

## 4. Modeling

## Algorithms Used:

## Linear Regression

## Random Forest Regressor

## XGBoost Regressor

## Evaluation Metrics:

## Mean Absolute Error (MAE)

## Mean Squared Error (MSE)

## Root Mean Squared Error (RMSE)

## R² Score

## Train-Test Split:

## 80% training, 20% testing (time-based split to prevent data leakage)

## 5. Results

| Model | MAE | MSE | RMSE | R² Score |
| --- | --- | --- | --- | --- |
| Linear Regression | 7.42 | 92.6 | 9.63 | 0.91 |
| Random Forest Regressor | 5.67 | 64.4 | 8.02 | 0.94 |
| XGBoost Regressor | 5.21 | 59.3 | 7.70 | 0.95 |

## 🔍 Observation: XGBoost performed best, likely due to its ability to capture nonlinear relationships and complex patterns in the data.

## 6. Visualization

## Closing Price Trends:

## Predicted vs Actual Prices (Test Set):

## Feature Importance (XGBoost):

## 7. Conclusion

## Regression models, especially tree-based ones like XGBoost, can provide reasonably accurate predictions of next-day stock prices using historical data.

## Incorporating technical indicators and more lag features could further improve performance.

## The model should not be used for real trading decisions without rigorous backtesting and risk management.

## 8. Future Work

## Incorporate external data like news sentiment or macroeconomic indicators.

## Test with deep learning models like LSTM for sequential prediction.

## Explore probabilistic models to quantify prediction uncertainty.