PRICE PREDICTION MODEL

MUHAMMAD ARSIM ANSARI (FA21-BSCS-0100)
MUHAMMAD FURQAN (FA21-BSCS-0020)

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF COMPUTING

MOHAMMAD ALI JINNAH UNIVERSITY

SPRING 2024

Acknowledgements

We would like to express our gratitude to our instructors, Dr. Asim Imdad Wagan & Ms. Fozia Khan, for their continuous support and guidance throughout this project. We also thank our peers and family for their encouragement and understanding.

Abstract

This project explores the application of AI to forecast stock prices using Linear Regression and Random Forest Regressor models. The main objective is to accurately predict the future prices of a particular stock based on historical data and analyzed trends. We have utilized these predictive models due to their interpretability and efficiency. The model training involved handling various features like historical prices, volume, and daily price movements. The results, evaluated using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-Squared (R^2), indicate the models' potential in providing reliable price forecasts. This project demonstrates the viability of AI in enhancing predictive analytics for the financial sector, facilitating informed decision-making.

Keywords: AI-based price prediction, Linear regression, Random Forest, Predictive modeling, Model accuracy, Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2), Predictive analytics.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Overview

This project is focused on developing a price prediction model for stocks using AI techniques. The model aims to forecast future stock prices based on historical data, assisting stakeholders in making informed investment decisions.

1.2 Motivation

Accurate stock price predictions are invaluable in the financial industry for strategic planning and investment. Leveraging machine learning models can significantly enhance the accuracy of these predictions, providing a competitive edge in market analysis.

CHAPTER 2

LITERATURE REVIEW

2.1 Existing Models

Various models have been proposed for stock price prediction, ranging from traditional statistical methods to advanced machine learning techniques. Linear Regression is frequently used due to its simplicity and interpretability, while more complex models like Random Forest offer enhanced predictive power through ensemble learning. [1]

2.2 Research Gap

While many studies focus on single-factor models, our project integrates multiple features, including historical prices and volume, to improve prediction accuracy. The comparative analysis of different models also provides insights into their respective strengths and weaknesses.[2]

CHAPTER 3

DEFINITION

This section provides an overview of key concepts related to stock price prediction, including the definition of features like 'Open', 'Close', 'High', 'Low', and 'Volume', and metrics like MAE, MSE, and R^2, which are used to evaluate the models' performance.

CHAPTER 4

DATASET

4.1 Overview

The dataset used in this project is a historical record of stock prices from the Pakistan Stock Exchange (PSX). It includes daily observations of the stock's 'Open', 'Close', 'High', 'Low', and 'Volume' from May, 2014 to May, 2024

4.2 Dataset Features

- **Date**: The trading date.
- **Open**: The opening price of the stock.
- **Close**: The closing price of the stock.
- **High**: The highest price of the stock during the trading day.
- **Low**: The lowest price of the stock during the trading day.
- **Volume**: The number of shares traded.

CHAPTER 5

METHODOLOGY

5.1 Data Preprocessing

Data preprocessing involves converting the 'Date' column to datetime format and dropping the 'Adj Close' column, which is redundant for this analysis. The price difference between consecutive days is calculated to determine the daily price movement. [3]

5.2 Feature Selection

The features selected for the model are 'Open', 'High', 'Low', 'Volume', and a binary variable 'Price_Up' that indicates whether the price increased from the previous day.

5.3 Model Training

Two models were trained:

- **Linear Regression**: A simple yet effective model for predictive analysis.
- **Random Forest Regressor**: An ensemble model that provides robust predictions by averaging multiple decision trees. [4]

5.4 Evaluation Metrics

The models were evaluated using:

- **R-Squared (R^2)**: Measures the proportion of variance explained by the model.
- **Mean Absolute Error (MAE)**: Measures the average magnitude of errors.
- **Mean Squared Error (MSE)**: Measures the average squared magnitude of errors.
- **Root Mean Squared Error (RMSE)**: Provides the square root of MSE for easier interpretation.

CHAPTER 6

RESULTS

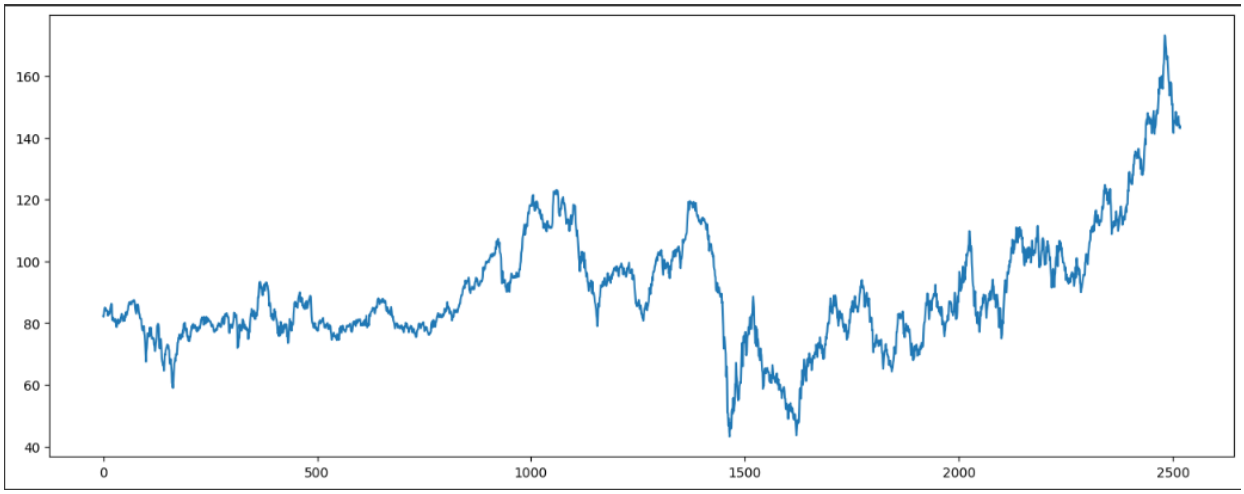
6.1 Model Performance

The Linear Regression model achieved an R^2 score of 0.999 indicating good accuracy. The Random Forest Regressor performed better with an R^2 score of 0.998, suggesting a more accurate model for this dataset.

Model	R^2 Score	MAE	MSE	RMSE
Linear Regression	0.999	0.458	0.364	0.603
Random Forest Regressor	0.998	0.543	0.606	0.778

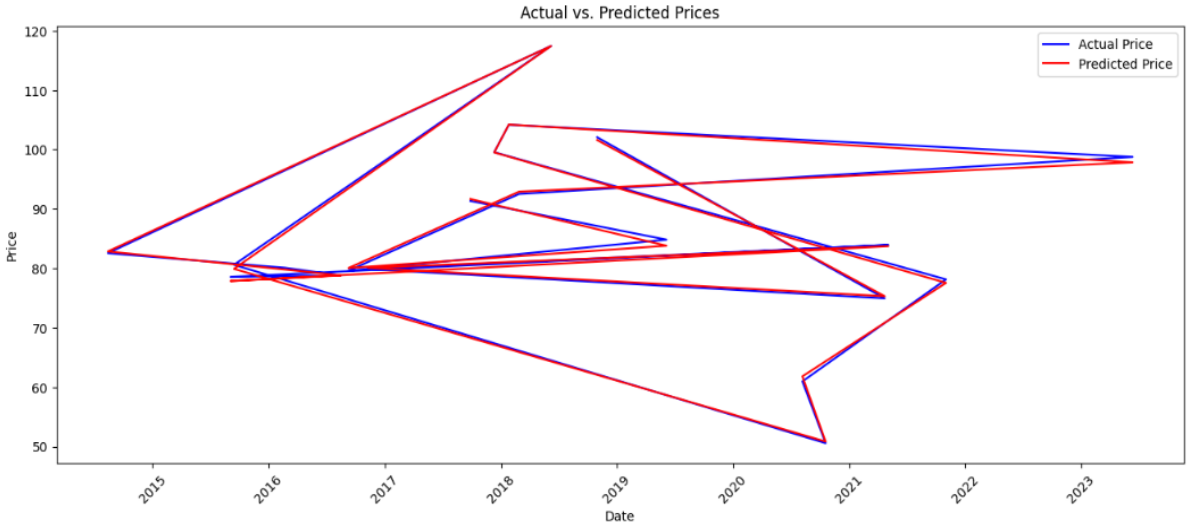
6.2 Visualization

6.2.1 Plot of 'Open' Prices

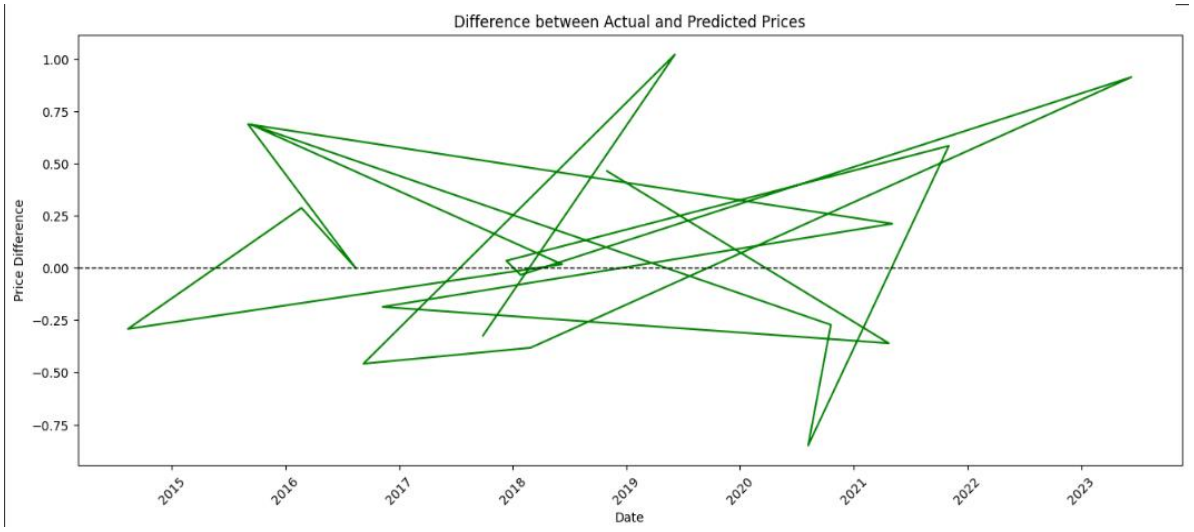


6.2.2 Actual vs. Predicted Price

- **Linear Regression:**



- **Random Forest Regressor:**



6.2.3 Price Movement Analysis

CHAPTER 7

CONCLUSION

The project successfully demonstrates the application of Linear Regression and Random Forest models for stock price prediction. The Random Forest model outperformed Linear Regression, highlighting the benefits of using ensemble methods for such tasks. Future work could explore incorporating more features and testing other advanced machine learning algorithms.

References

[~~1~~1] Mitchell, M., McKinney, M., Laramey, J., Neill, C., & Varian, H. (2019). Machine Learning. Ziff Davis Publishing Company http://www.cs.cmu.edu/~tom/mlbook.html

[2] Huang, W., Nakamori, Y., & Wang, S. (2014). Forecasting stock prices using a seasonal Holt-Winters exponential smoothing model with rough sets and least squares support vector machines. Journal of Applied Mathematics, 2014, 614342 https://scholar.archive.org/work/ci2gmtezyjcmpeher7lil4rpxy/access/wayback/htt

[3] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer Publishing https://link.springer.com/book/10.1007/978-1-0716-1418-1

[4] sklearn: Machine Learning in Python https://scikit-learn.org/