

# PROJECT REPORT

# Stock Price Prediction Using Machine Learning

Submitted to: Mrs Rita Chhikara

Submitted by: Geetu R. Sanjay

Roll No: 22CSU340

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S.No** |  | **Page No.** |
| **1.** | **Project Description** | **PG 3** |
| **2.** | **Problem Statement** | **PG 3** |
| **3.** | **Analysis**  **3.1 Hardware Requirements**  **3.2 Software Requirements** | **PG 3** |
| **4.** | **Design**  **4.1 Data/Input Output Description:**  **4.2 Algorithmic Approach / Algorithm / DFD / ER diagram/Program Steps** | **PG 3** |
| **5.** | **Implementation and Testing (stage/module wise)** | **PG 4** |
| **6.** | **Output (Screenshots)** | **PG 5-9** |
| **7.** | **Conclusion and Future Scope** | **PG 9** |

## Introduction

**Background:**  
Stock price prediction plays a vital role in financial decision-making and risk management. Bajaj Finance's stock price movements are influenced by numerous factors, including technical indicators and market trends. Accurately forecasting these movements is crucial for optimizing trading strategies and minimizing risks.  
  
1. **Project Description**

To build a robust machine learning model that predicts stock price trends for Bajaj Finance using features like moving averages, stochastic oscillators, and volatility indicators. This project focuses on developing machine learning models to predict stock price trends for Bajaj Finance using technical indicators. The objective is to create an efficient and accurate predictive system for stock price movements, which can aid in informed financial decision-making.

**2. Problem Statement**

The problem involves predicting stock price trends for Bajaj Finance based on historical data and technical indicators. By analyzing these features, the project aims to enhance decision-making for investors and stakeholders.

### ****3. Analysis****

#### ****3.1 Hardware Requirements****

* **Processor:** Intel i5 or higher
* **RAM:** 8GB or higher
* **Storage:** Minimum 20GB of free space
* **Operating System:** Windows 10 or Linux

#### ****3.2 Software Requirements****

* **Tools and Libraries:**
  + Python 3.x
  + Jupyter Notebook
  + Libraries: pandas, numpy, matplotlib, seaborn, sklearn

### ****4 Design****

#### ****4.1 Data/Input Output Description****

* **Input:** Bajaj Finance dataset with multiple technical indicators like moving averages, RSI, and price changes.
* **Output:** Predicted stock price or stock price trend using trained machine learning models.

#### ****4.2 Algorithmic Approach****

1. **Data Preprocessing:**
   * Handle missing values.
   * Perform feature scaling.
   * Handle outliers using statistical techniques.
2. **Feature Engineering:**
   * Select important indicators based on correlation.
   * Create lag-based features for trends.
3. **Model Selection and Training:**
   * Linear Regression, Multiple Linear Regression, Polynomial Regression.
4. **Evaluation:**
   * Use metrics like RMSE, R², and adjusted R².

### ****5. Implementation and Testing****

* **Implementation Steps:**
  + Preprocessing the dataset: Scaling and handling missing values.
  + Splitting data into training and testing sets.
  + Training models: Linear Regression, Multiple Linear Regression, and Polynomial Regression.
  + Validating models using cross-validation techniques.
* **Testing Metrics:**
  + Root Mean Square Error (RMSE) for performance evaluation.
  + R² score to assess model accuracy.

## Dataset Overview

- Size: 58646 rows, 60 columns.  
- Data Types: Integer, Float, Boolean, Date, Char.  
- Purpose: Predictive analysis for financial regression and classification tasks.  
- Key Features:  
 - Price Indicators: Open, High, Low, Close.  
 - Volume: Trading volume data.  
 - Technical Indicators: SMA (Simple Moving Average), ATR (Average True Range), BETA (risk coefficient), FastD, FastK (Stochastic Oscillator).

## Data Preprocessing

1. Missing Values: No significant missing data identified.  
2. Duplicates: Checked and removed.  
3. Outlier Treatment:  
 - Methods: IQR and Z-Score.  
 - Outcome: Removed 24 outlier rows to improve model reliability.

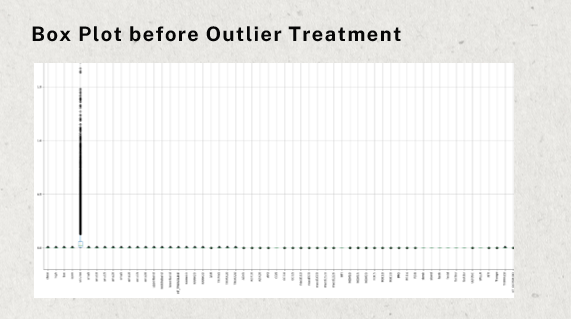
## Machine Learning Models

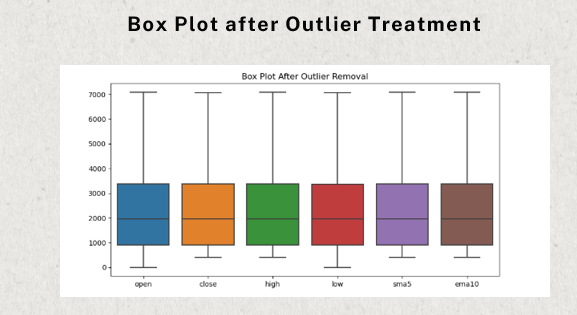
**1. Linear Regression:**  
- Purpose: Establish a relationship between stock prices (Open) and closing prices.  
- Findings: Showed initial predictive capability but was sensitive to outliers and limited by its linearity assumption.  
  
**2. Multiple Linear Regression (MLR):**  
- Purpose: Leverage multiple predictors (Open, High, Low, SMA5, EMA10).  
- Findings: Captured more variance but required additional tuning for accuracy.  
  
**3. Robust Regression:**  
- Techniques Explored:  
 - Huber Regression: Balanced sensitivity to outliers.  
 - Theil-Sen Regression: Non-parametric and robust.  
 - RANSAC Regression: Best suited for datasets with a high proportion of outliers.  
- Outcome: RANSAC Regression emerged as the most reliable method in this category.  
  
**4. Polynomial Regression:**  
- Purpose: Address non-linear relationships.  
- Insights: Improved prediction accuracy but risked overfitting the data.  
  
**5. Ridge Regression:**  
- Purpose: Regularization to control coefficient magnitudes and balance bias-variance trade-off.  
- Performance: Achieved low error and high accuracy with optimal α=0.001.  
  
**6. Lasso Regression:**  
- Purpose: Perform feature selection by reducing some coefficients to zero.  
- Findings: Underperformed compared to Ridge due to over-regularization.

## Model Evaluation

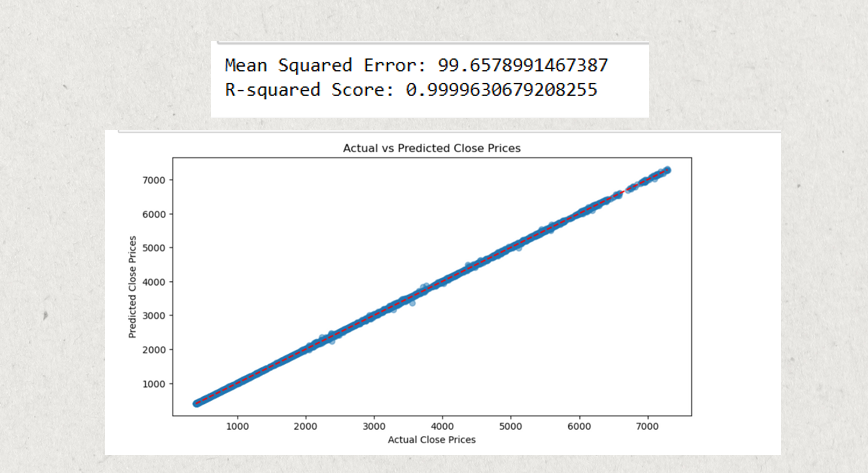
Final Model: Ridge Regression  
- Training MSE: 15.623  
- Testing MSE: 19.202  
- R² Score: 0.99999

**Output (Screenshots)**

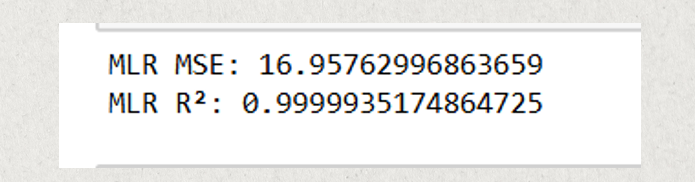
****

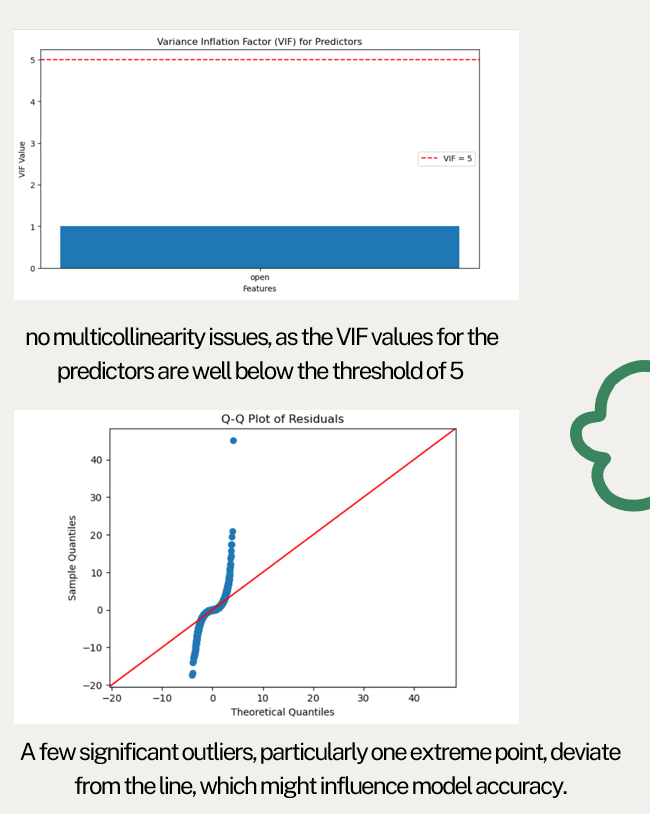
****

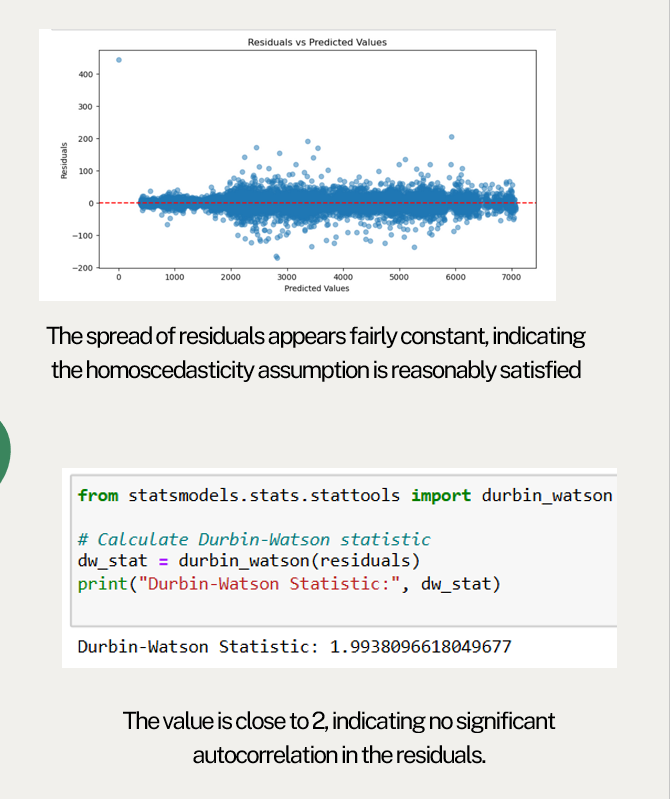
**LINEAR REGRESSION**

****

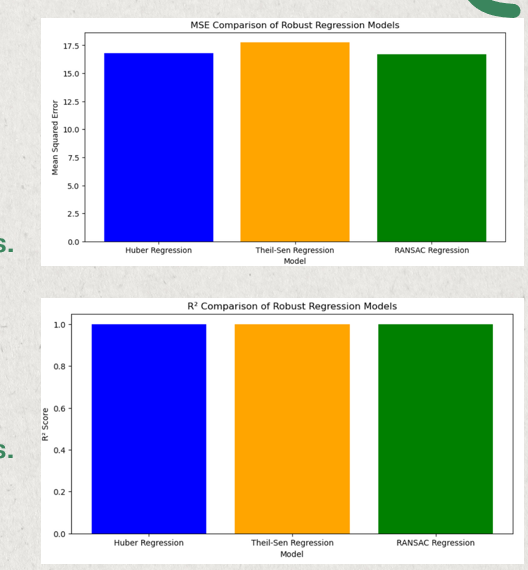
**MLR**

****

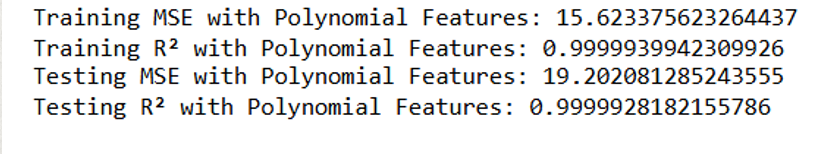
****

****

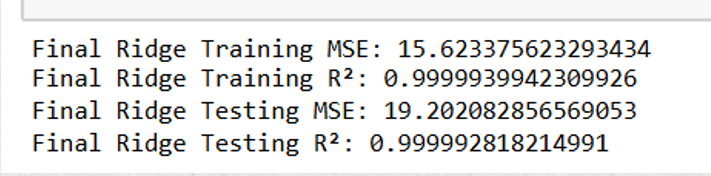
**ROBUST REGRESSION**

****

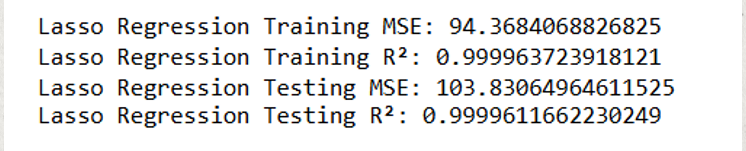
**POLYNOMIAL REGRESSION**

****

**RIDGE REGRESSION**

****

**LASSO REGRESSION**

****

## Conclusion and Future Scope

* **Conclusion:**
  + The developed models successfully predict stock price movements. Among the tested models, Polynomial Regression provided better performance due to its ability to handle non-linearity in data.
* **Future Scope:**
  + Integrating time-series analysis techniques like ARIMA or LSTM for better predictions.
  + Incorporating additional external factors like market news and sentiments for a more robust model.