GROUP M15

Statistical Inference and Multivariate Analysis

<u>Group Members</u>

Ayushman Verma	210123013
Inaganti Chakradhar	210123027
Rahul Kumar	210123048
Shivam Singh	210123058
Yatharth Yash	210123071

STATISTICAL INFERENCE IN SIGNAL PREDICTION

INTRODUCTION

Statistical inference in signal prediction uses data analysis to forecast future outcomes based on observed patterns. It involves modelling relationships between signals, making predictions, and validating their accuracy.

1. Data Source: Reliance NSE stock prices for 100 days were collected as the dataset for analysis.

2. Prediction Task: The objective was to forecast the stock prices for the next 100 days based on the observed patterns in the historical data.

3. Models Used

- Linear Regression: A simple regression model that establishes a linear relationship between the input features (historical stock prices) and the target variable (future stock prices).

- Ridge Regression: A regression technique that adds a penalty term to the linear regression model to prevent overfitting and improve generalization.

- Lasso Regression: Similar to Ridge Regression, but with a different penalty term which tends to produce sparse coefficient vectors by forcing some coefficients to be exactly zero.

4. Bootstrap Techniques:

- Local Block Bootstrapping: This method involves resampling blocks of data from the original dataset, preserving the local dependencies and patterns present in the time series data.

- Moving Block Bootstrapping: In this approach, overlapping blocks of data are randomly sampled from the original dataset, allowing for the preservation of temporal dependencies in the data.

5. Implementation:

- The dataset was split into input features (X) and target variable (Y) based on a sliding window approach.

- Each model (Linear Regression, Ridge Regression, and Lasso Regression) was trained using the bootstrap resampled data.

- The trained models were then used to predict the stock prices for the next 100 days.

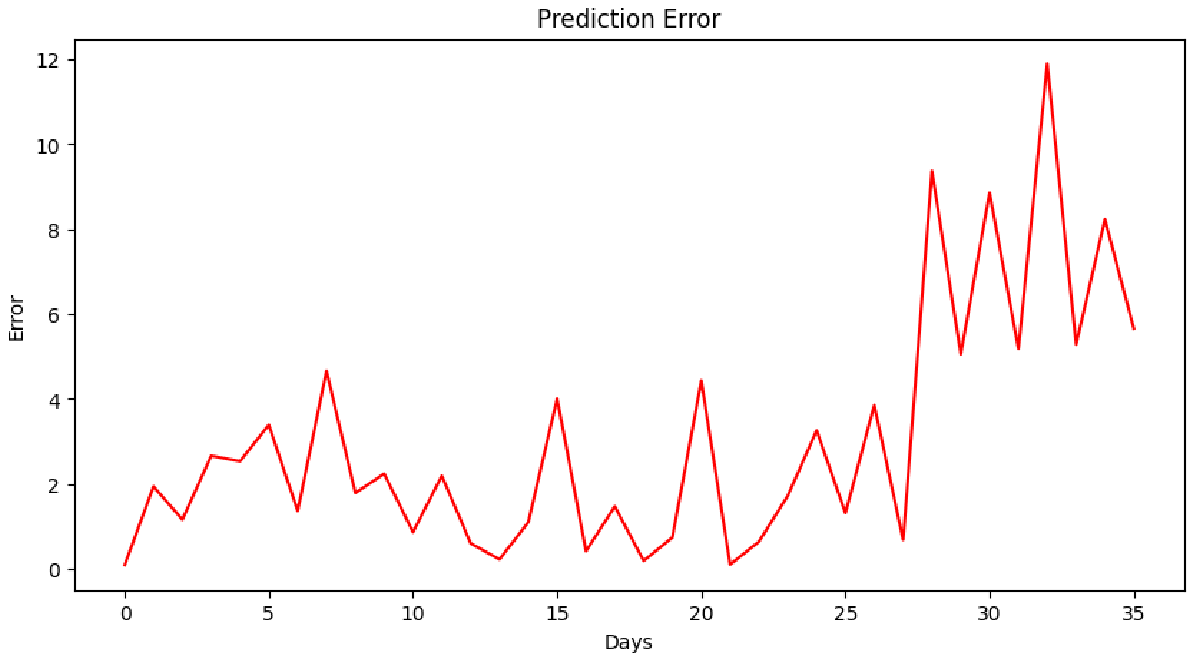
6. Evaluation:

- The accuracy of the predictions was evaluated using Mean Relative Percentage Error.

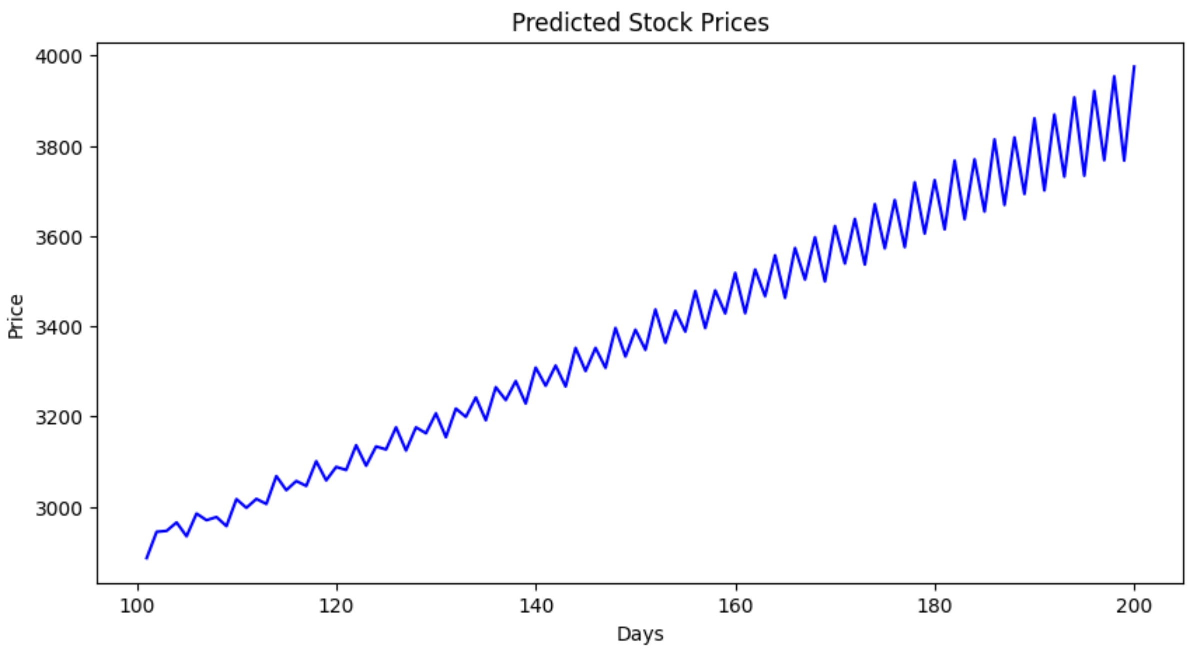
- The performance of each model and bootstrap technique was compared to determine the most effective approach for predicting stock prices.

Overall, this approach combined different regression models with bootstrap resampling techniques to make reliable predictions about future stock prices based on historical data.

Linear Regression using Moving Block Bootstrap

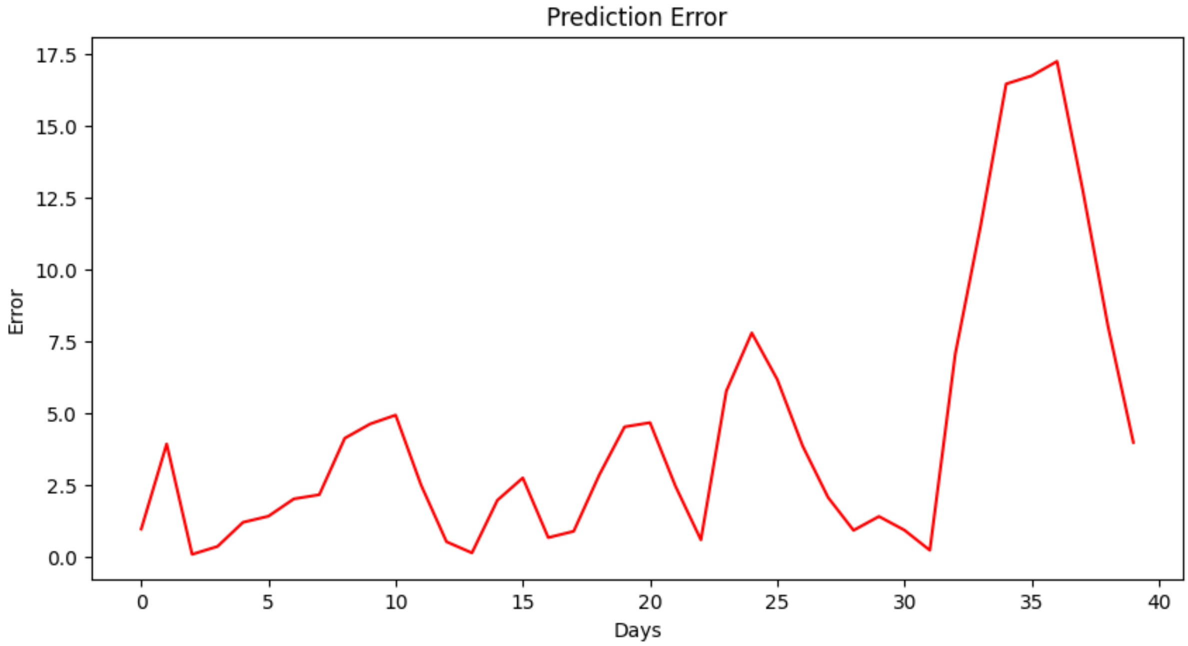


Above graph depicts error for the predicted data and testing data for 40 Days.

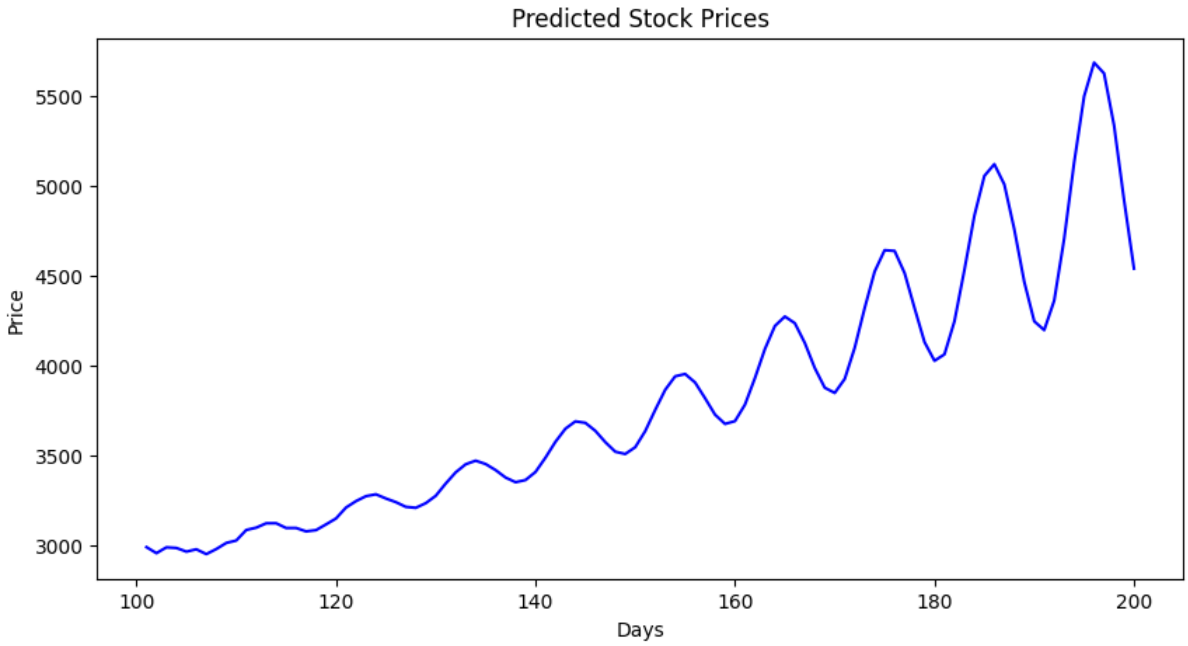


Above graph depicts error for the predicted stock price for next 100 days.

Linear Regression using Local Block Bootstrap

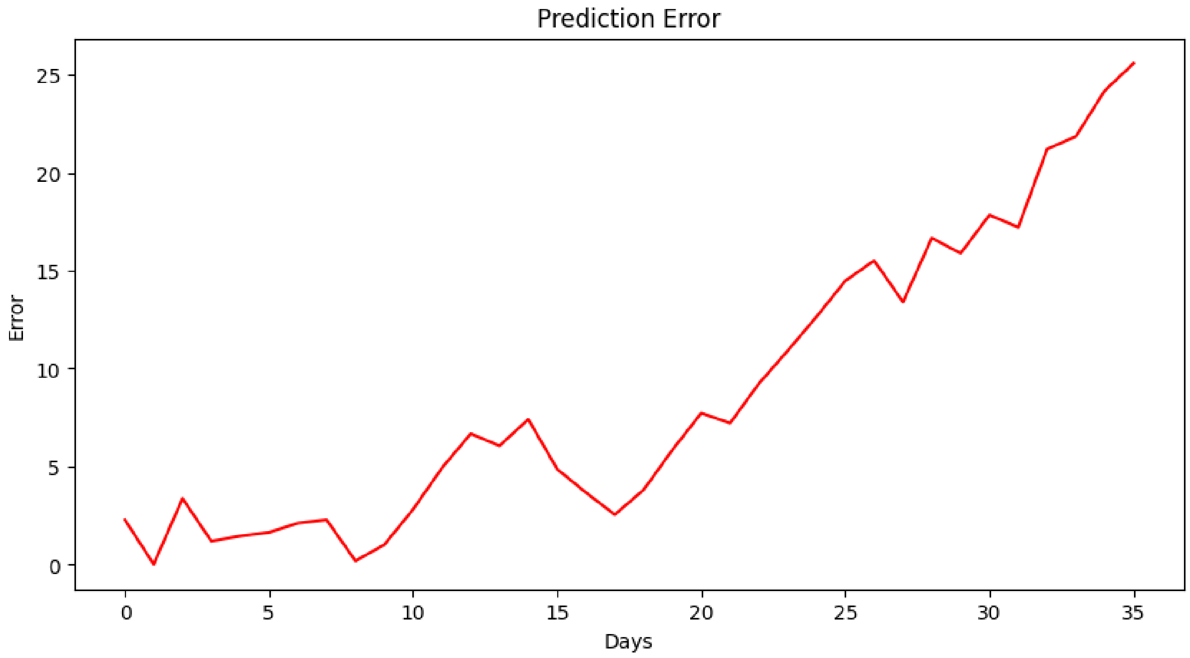


Above graph depicts error for the predicted data and testing data for 40 Days.

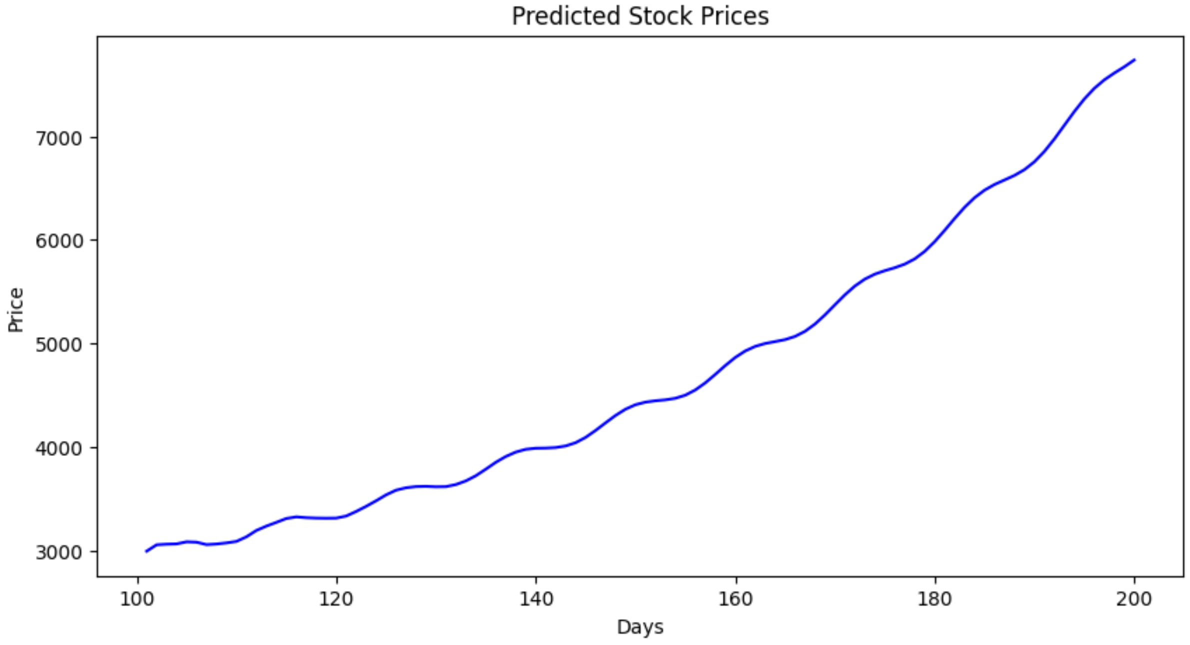


Above graph depicts error for the predicted stock price for next 100 days

Ridge Regression using Moving Block Bootstrap

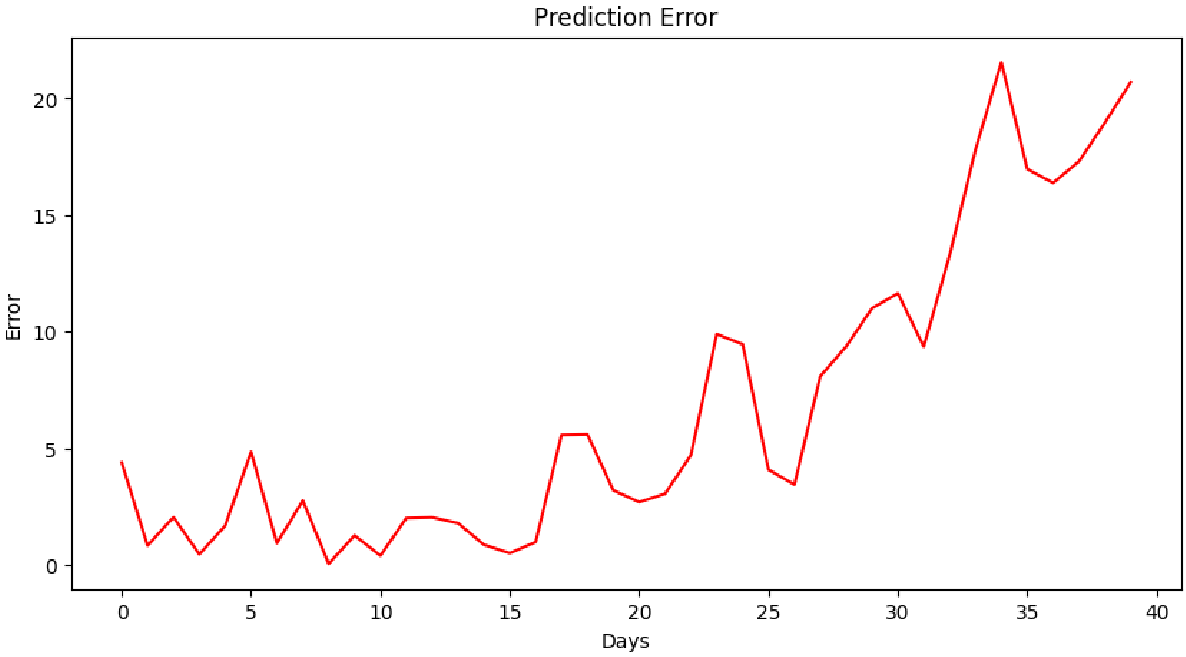


Above graph depicts error for the predicted data and testing data for 40 Days.

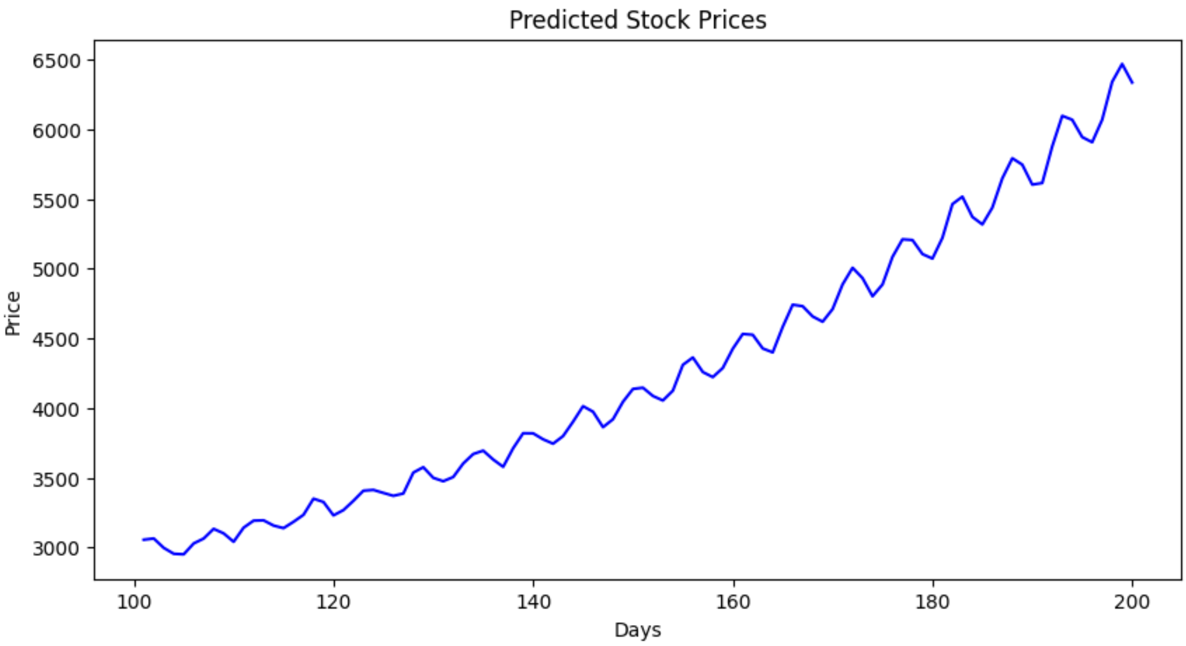


Above graph depicts error for the predicted stock price for next 100 days

Ridge Regression using Local Block Bootstrap

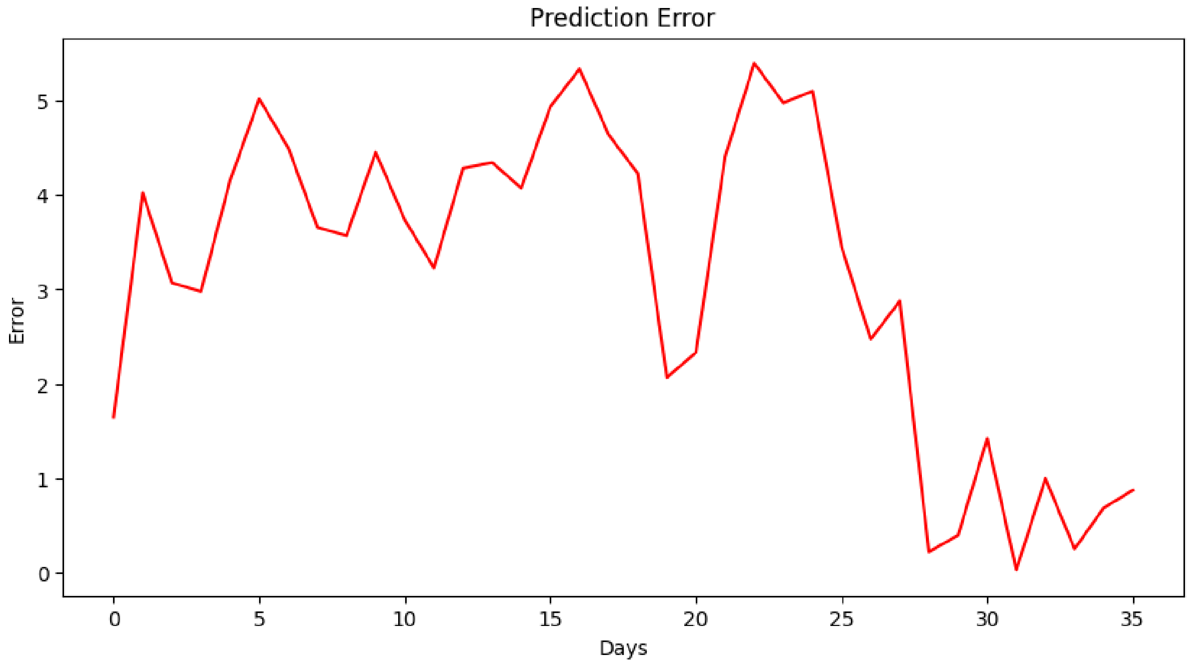


Above graph depicts error for the predicted data and testing data for 40 Days.

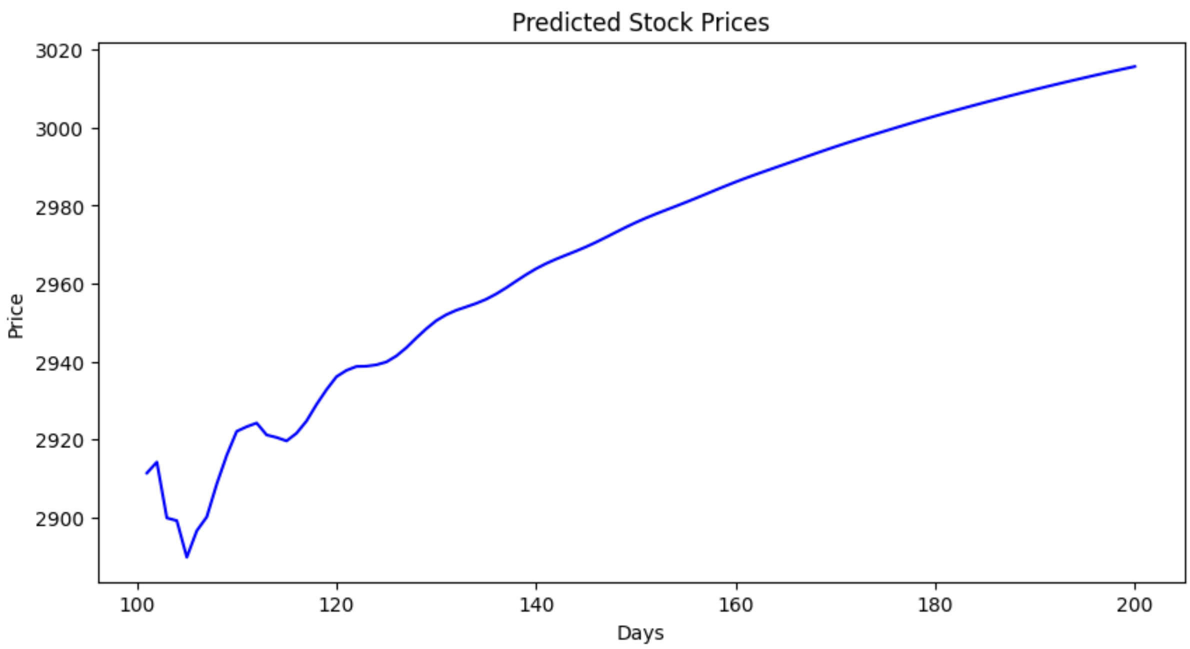


Above graph depicts error for the predicted stock price for next 100 days

Lasso Regression using Moving Block Bootstrap

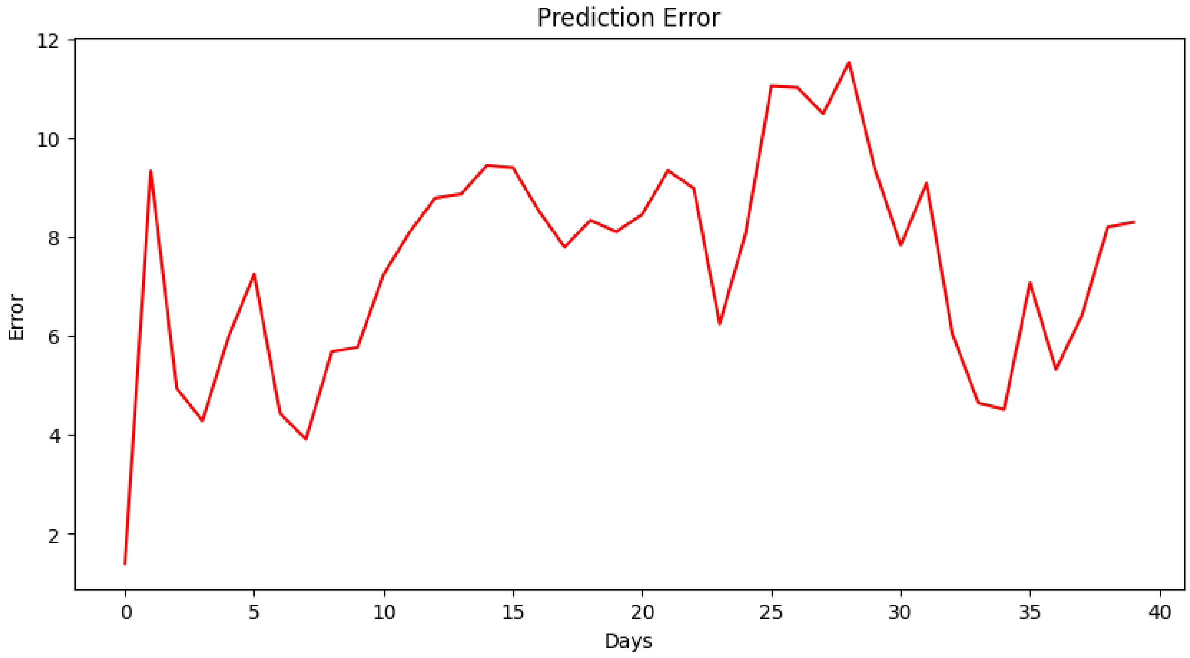


Above graph depicts error for the predicted data and testing data for 40 Days.

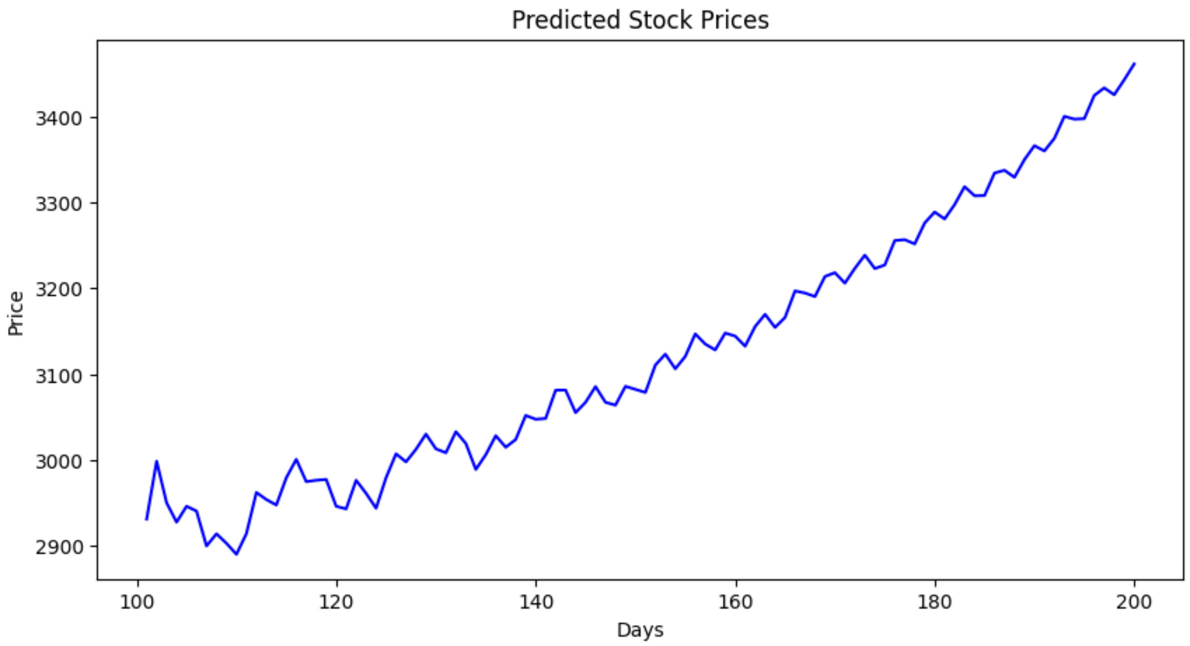


Above graph depicts error for the predicted stock price for next 100 days

Lasso Regression using Local Block Bootstrap



Above graph depicts error for the predicted data and testing data for 40 Days.



Above graph depicts error for the predicted stock price for next 100 days

CONCLUSION :

The minimum relative error observed in the Lasso Regression model suggests its superior performance compared to Linear Regression and Ridge Regression. Lasso Regression's feature selection ability, regularization, and robustness to multicollinearity contribute to its effectiveness in minimizing prediction errors. Additionally, the optimal alpha selection via GridSearchCV ensures the model's fine-tuning for optimal performance on the dataset.