Machine Learning for Option Pricing: A Comparative Analysis

Project Overview:

This project explores the application of machine learning models to the problem of option pricing. Options are complex financial derivatives, and accurately pricing them is crucial for traders and financial institutions. Traditional methods like the Black-Scholes model have limitations, particularly in capturing the complexities of the market. This project investigates the use of various machine learning techniques to enhance the accuracy of option pricing.

1. Introduction to Option Pricing:

Option pricing involves determining the fair value of an option contract. Traditional models, such as the Black-Scholes model, rely on assumptions that may not hold in real market conditions, such as constant volatility and log-normal distribution of returns. Machine learning offers a data-driven approach that can potentially improve the accuracy of pricing by learning from historical market data.

2. Machine Learning Models Implemented:

The project implements several machine learning models to predict option prices, including:

- Linear Regression: A basic model that attempts to predict the option price by fitting a linear relationship between the input features and the target variable.
- Support Vector Regression (SVR): A type of Support Vector Machine used for regression tasks, aiming to find the best fit within a threshold defined by a margin of tolerance.
- Random Forest: An ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees.
- Neural Network: A complex model capable of capturing non-linear relationships by passing inputs through multiple layers of neurons.

3. Model Evaluation Metrics:

The models were evaluated based on the following metrics:

- Root Mean Square Error (RMSE): Measures the average magnitude of the errors between predicted and actual values. Lower values indicate better fit.
- Mean Absolute Error (MAE): The average of the absolute errors, providing a sense of the model's prediction accuracy.
- R-Squared (R^2): Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. Values closer to 1 suggest a better fit.

4. Results:

The performance of the machine learning models on the option pricing problem was evaluated, and the following results were obtained:

- Linear Regression: RMSE of 19.09, MAE of 13.70, and R^2 of 0.93.
- Support Vector Regression (SVR): RMSE of 47.09, MAE of 24.33, and R^2 of 0.60.
- Random Forest: RMSE of 14.34, MAE of 7.35, and R^2 of 0.96.
- Neural Network: RMSE of 39.75, MAE of 31.16, and R^2 of 0.71.

5. Key Findings:

- Random Forest emerged as the best-performing model, achieving the lowest RMSE and MAE, and the highest R^2 value, indicating it captured the underlying patterns in the data most effectively.
- Linear Regression performed well, with a reasonably low RMSE and high R^2, making it a strong baseline model.
- Neural Network and SVR did not perform as well as expected, suggesting that these models may require further tuning or different feature engineering techniques to improve their accuracy in this context.

6. Conclusion:

This project demonstrates the potential of machine learning models in improving the accuracy of option pricing. The Random Forest model, in particular, showed strong predictive power, suggesting that ensemble methods may be particularly well-suited for this type of financial prediction task. The results highlight the importance of model selection and tuning in achieving accurate option pricing and suggest avenues for further research and model improvement.