Machine Learning and Option Pricing: A Proposal for Predicting Option Price for the Ticker ‘SPY’

This is a proposal for option pricing strategies by incorporating advanced machine learning techniques. Given the complexity and dynamic nature of financial markets, especially in the realm of options trading, traditional models like the Black-Scholes, while foundational, often fall short of capturing real-world intricacies. By leveraging machine learning, we can potentially achieve more accurate and robust option pricing; we study the ticker symbol ‘SPY’, an S&P 500 ETF.

Problem Statement

The primary problem we aim to solve is improving the accuracy of our option pricing models. Traditional models, such as Black-Scholes, assume constant volatility and interest rates, which are rarely reflective of market realities. Our goal is to predict the price of put options for the ‘SPY’ ticker with higher precision by utilizing machine learning models that can learn from historical data and adapt to market conditions dynamically.

Constraints and Challenges

While machine learning offers promising avenues for enhancing option pricing, there are several constraints and challenges we must consider:

1. **Data Quality and Availability**: The accuracy of our predictions heavily relies on the quality and completeness of the historical data we feed into our models. Inaccurate or missing data can significantly skew our results.
2. **Market Anomalies**: Sudden market events or anomalies, such as geopolitical tensions or economic shifts, can create scenarios that our models may not be adequately trained to handle.
3. **Overfitting**: There is a risk that our models might overfit to historical data, capturing noise instead of underlying patterns, which would result in poor predictive performance on new, unseen data.
4. **Computational Resources**: Training complex machine learning models can be resource-intensive, requiring significant computational power and time.

Data Source

For this project, we will use historical data of ‘SPY’ put options downloaded from Yahoo Finance. This platform provides a comprehensive and accessible repository of financial data, ensuring we have a reliable source for our analyses.

Scope

The scope of this project encompasses the following key areas:

1. Gathering historical data of SPY put options, cleaning it, and preparing it for analysis.
2. Developing relevant features such as historical prices, implied volatility to feed into our machine learning models.
3. Building and training machine learning models to predict option prices, evaluating various algorithms to identify the most effective approach.

Criteria for Success

The success of this project will be measured by the following criteria:

1. **Accuracy**: Our models should significantly outperform traditional pricing models in terms of prediction accuracy, measured by metrics such as Mean Squared Error (MSE) or Mean Absolute Error (MAE).
2. **Robustness**: The models must demonstrate resilience to market anomalies and maintain predictive performance across different market conditions.
3. **Efficiency**: The models should be computationally efficient, providing timely predictions that can be effectively integrated into our trading strategies.

The following is the information of the put option data that we use:

```
RangeIndex: 62 entries, 0 to 61
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   contractSymbol     62 non-null     object
 1   lastTradeDate      62 non-null     datetime64[ns, UTC]
 2   strike             62 non-null     float64
 3   lastPrice          62 non-null     float64
 4   bid                62 non-null     float64
 5   ask                62 non-null     float64
 6   change             62 non-null     float64
 7   percentChange      62 non-null     float64
 8   volume             61 non-null     float64
 9   openInterest       62 non-null     int64
 10  impliedVolatility  62 non-null     float64
 11  inTheMoney         62 non-null     bool
 12  contractSize       62 non-null     object
 13  currency           62 non-null     object
dtypes: bool(1), datetime64[ns, UTC](1), float64(8), int64(1), object(3)
```