Kriti 2024

TRADING STRATEGY

CHALLENGE WITH OPTIONS DATA

KAMENG

Data Acquisition

We have used the data provided data for the **NIFTY BANK** index, spanning over a period of 6 months from **July 2023 to December 2023** along with using the '**yfinance**' python library for data acquisition of the underlying.

Data Preprocessing

This process enhances the quality of data, leading to more accurate models and meaningful insights.

HANDLING MISSING VALUES

In order to handle the missing values, we have used the imputation method, which involved:

- Filling the empty values in **'No. of contracts', 'Open Int' and 'Change in OI' with zero**.
- Filling the empty values in **'Open', 'High' and 'Low' with the 'Close'** of the corresponding day.

We have imputed the missing values in both train and test dataset in order to run the strategy.

Data Preprocessing

OUTLIER ANALYSIS & FEATURE ENGINEERING

It helps us to identify and handle data points that significantly deviate from the majority of the dataset. It improves model performance, and aids in identifying anomalies.

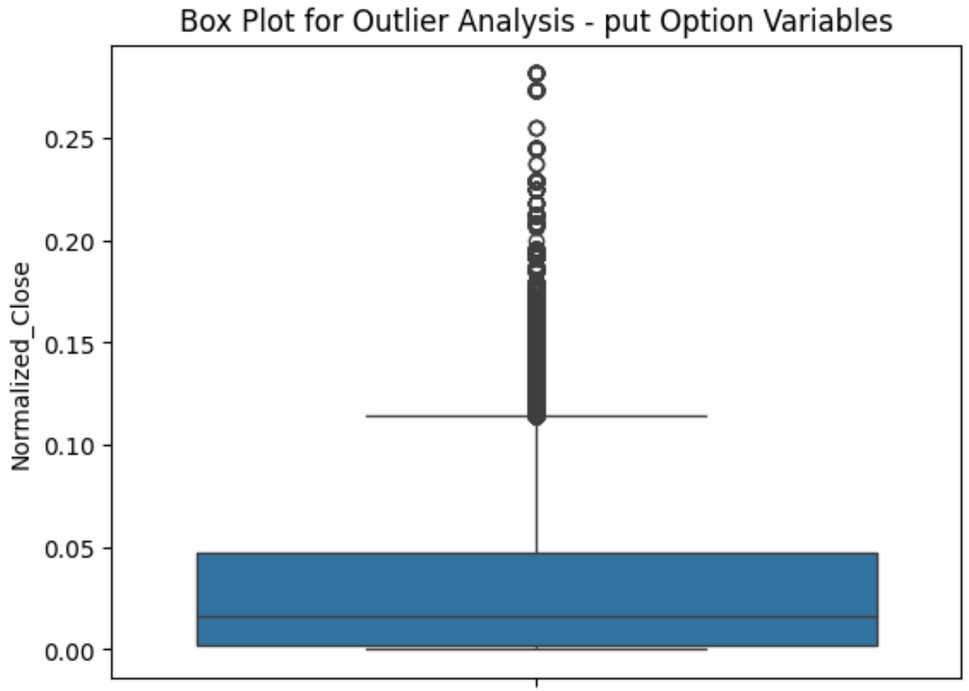
In order to define outliers, we came up with a new feature - **'Normalized Close'.**

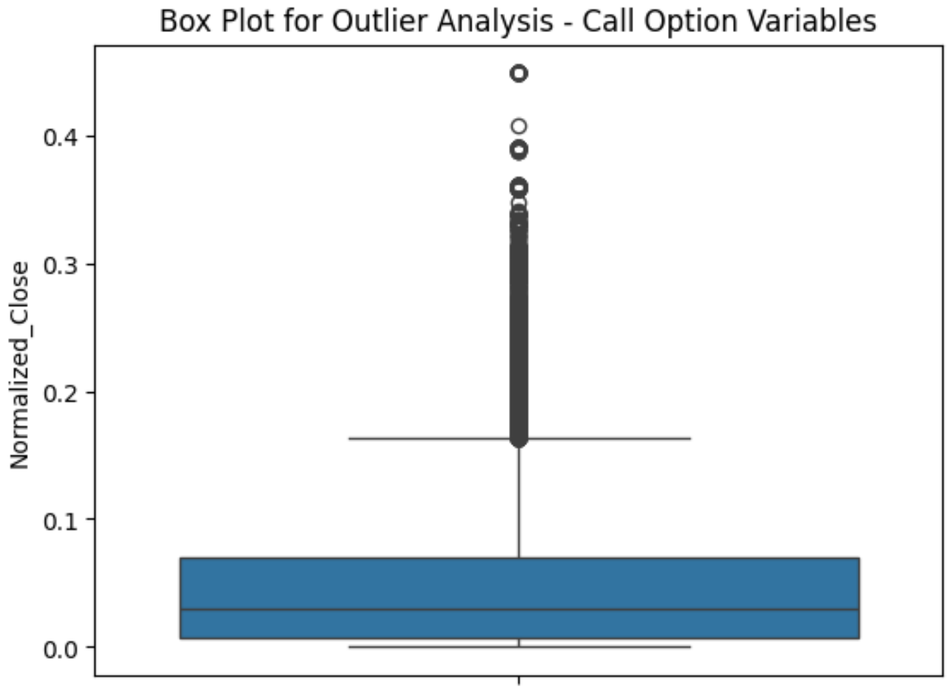
Normalized Close = Close Price/Strike Price

A normalized measurement of the close price in relation to the strike price is provided by this feature. It is beneficial to our outiler analysis because it helps us in:

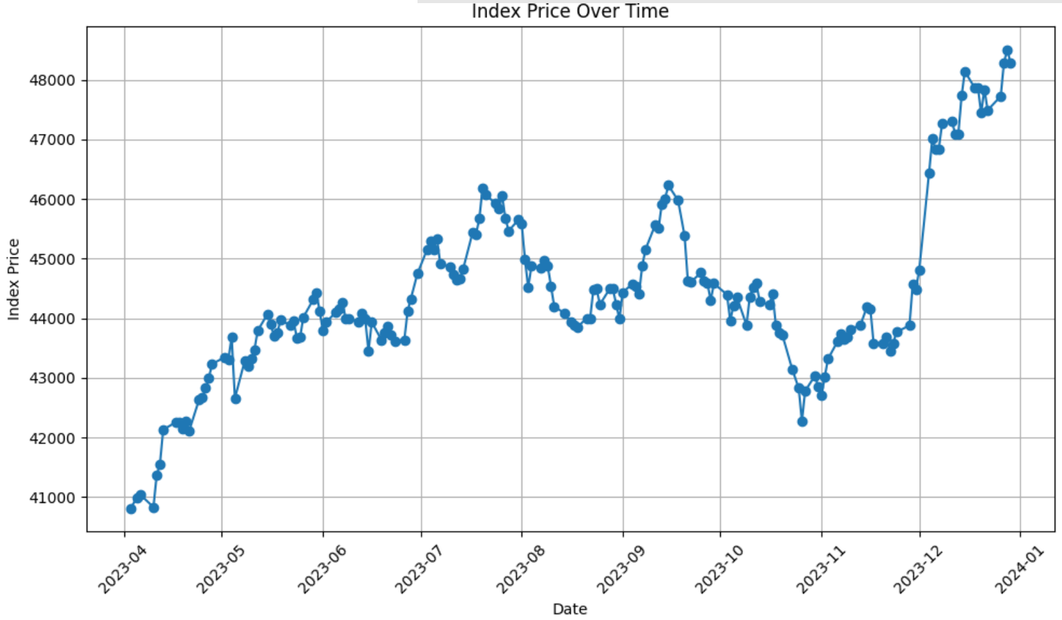
- **Recognizing Abnormal Price Changes**
- **Finding Unusual Option Contracts**

Using the Inter-Quartile Range, we have removed 1,416 data points out of 37,904 train data points provided. The box-plots for the same is shown on the right side.

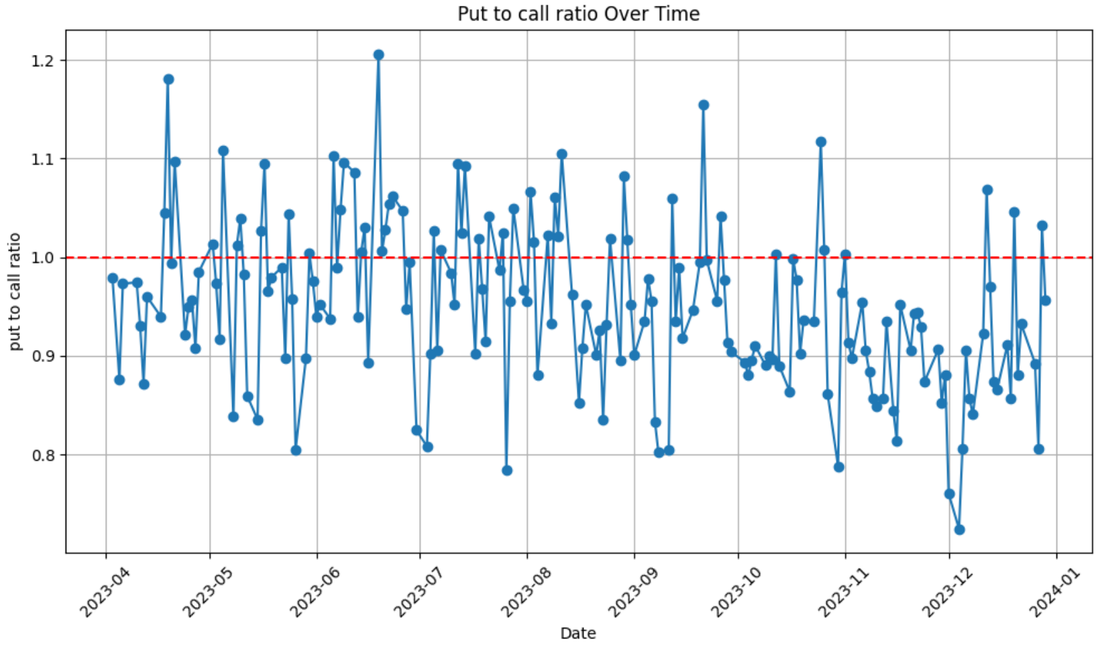




Data Visualization



The graphs of Index price vs Date displays the movement of an index's price over time. Price Trends shows whether the index's price is rising, falling, or remaining stable. It indicates the overall market direction (bullish, bearish) during the period. This helps identify patterns and trends for future market predictions.

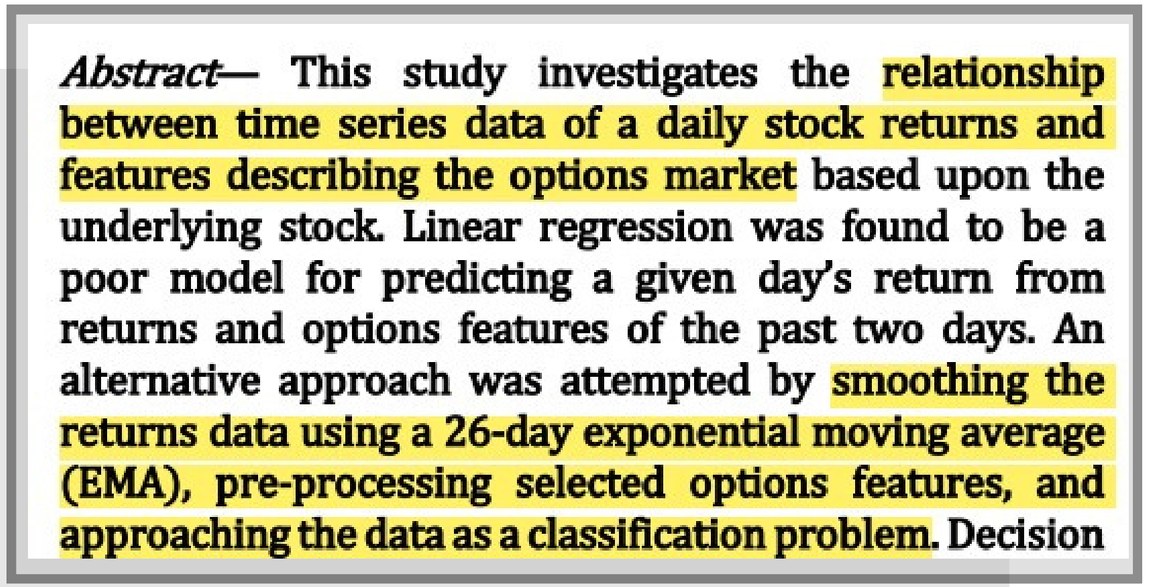


The Graph is showing put-to-call ratio changes over time, with dates on the x-axis and the ratio on the y-axis. It reflects market sentiment—high ratio suggests bearishness, low ratio indicates bullishness. Changes in ratio may indicate shifts in volatility expectations.

Strategy Development

We have taken inspiration from a research paper published by a professor from the Stanford University, in which he studied the relation between the daily stock returns and the option chain data.

The paper included Regression and Classification ML models and the features which were used as inputs to the model.



Abstract— This study investigates the relationship between time series data of a daily stock returns and features describing the options market based upon the underlying stock. Linear regression was found to be a poor model for predicting a given day's return from returns and options features of the past two days. An alternative approach was attempted by smoothing the returns data using a 26-day exponential moving average (EMA), pre-processing selected options features, and approaching the data as a classification problem. Decision

Enhancements to the paper

- Adding a new feature to the input list - **'OS Volume'.**
- Using a different classification algorithm - **'CatBoost'**. The paper originally used **'AdaBoost'**, which showed inferior performance in comparision to the catboost model.
- Added another Machine Learning model - **'Regressor Trees'** which uses the same features and predicts the magnitude of the returns.
- Used buy threshold value as a part of the trading strategy which helps us in risk management.

KEY FEATURES

EMA RETURNS

EMA Returns indicate the profit or loss generated by following buy/sell signals derived from Exponential Moving Averages (EMA).

EMA PUT-TO-CALL RATIO

It's a version of the put-to-call ratio using Exponential Moving Averages (EMA) to smoothen fluctuations, aiding in options market sentiment trend identification.

IMPLIED VOLATILITY CALL AND PUT

IV measures the expected future volatility of an underlying asset's price, as indicated by the options market.

IMPLIED VOLATILITY SPREAD

Implied Volatility Spread refers to the difference in implied volatility levels among options with the same underlying asset and expiration date.

OPTION INDEX VOLUME RATIO(OS VOLUME)

OS Volume is a financial metric that represents the ratio of total option volume to index volume. It measures the level of options trading activity relative to the trading volume of the underlying index.

EMA RETURNS

- EMA is a type of Moving Average (MA) that emphasises recent data points.
- EMA = (Closing Price × Multiplier) + (EMA Previous Day × (1 - Multiplier)).
- EMA Returns indicate the profit or loss generated by following buy/sell signals derived from Exponential Moving Averages (EMA).
- Relies on EMA crossovers for trade signals.
- Aims to capture trending market movements.

EMA PUT TO CALL RATIO

- It's a version of the put-to-call ratio using Exponential Moving Averages (EMA) to smoothen fluctuations, aiding in options market sentiment trend identification.
- To calculate the EMA put to call ratio find put-to-call ratio (put options divided by call options), then apply EMA over a chosen period.
- EMA Put-to-Call Ratio aids traders in understanding options market sentiment trends, offering insights for strategic decision-making alongside other analysis tools.

IMPLIED VOLATILITY CALL

- **Definition**: Implied Volatility (IV) for calls is a measure of the expected future volatility of the underlying asset's price, as implied by the market price of call options.
- **Calculation**: Implied Volatility for calls is derived from the market prices of call options using mathematical models like the Black-Scholes model.

IMPLIED VOLATILITY PUT

- **Definition**: Implied Volatility (IV) for put options is a measure of the expected future volatility of an underlying asset's price, as inferred from the market prices of put options.
- **Calculation**: Similar to call options, IV for put options is computed using options pricing models like the Black-Scholes model. It takes into account factors such as the current market price, strike price, time to expiration, interest rates, and dividends.

IMPLIED VOLATILITY SPREAD

- **Definition**: Implied Volatility Spread refers to the difference in implied volatility levels among options with the same underlying asset and expiration date.
- **Purpose**: It helps traders identify potential pricing discrepancies or biases in market sentiment.
- **Significance**: A wide spread may indicate heightened market uncertainty or differing expectations among traders.

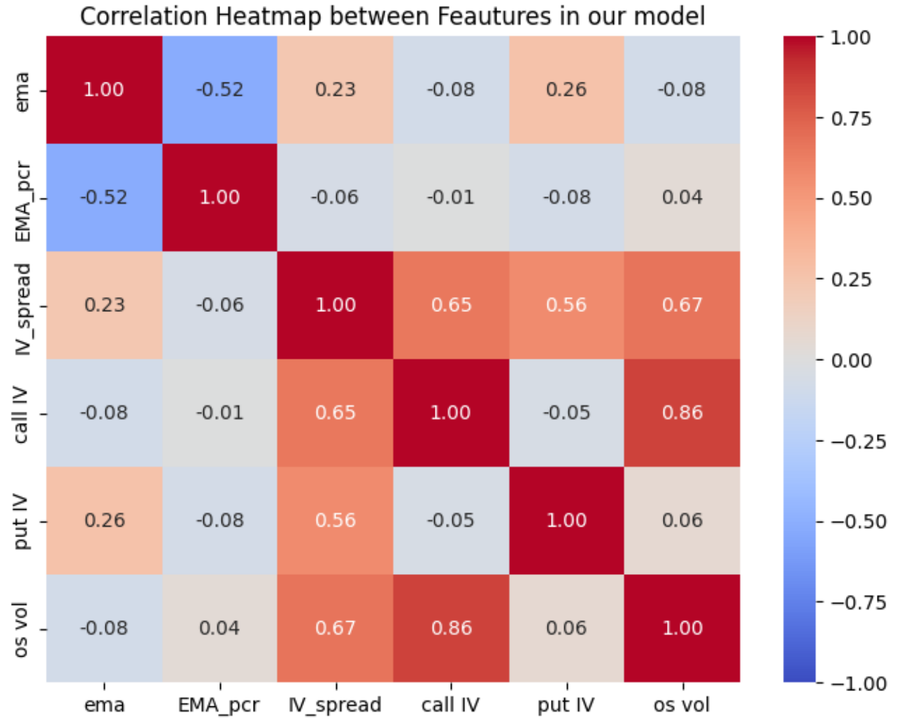
OPTION-INDEX VOLUME RATIO (OS VOLUME)

- **Definition**: OS Volume is a financial metric that represents the ratio of total option volume to index volume. It measures the level of options trading activity relative to the trading volume of the underlying index.
- **Calculation**: OS Volume is calculated by dividing the total volume of options traded by the volume of the underlying index.

Feature Selection

The small correlations imply a weaker linear relationship between features. In predictive modeling, weak correlations suggest each feature contributes unique information not captured by others, enhancing the model's ability to learn without redundancy.

This independence allows features to potentially influence the target variable individually, broadening the predictive scope.
Here, as we can see from the heatmap, none of the features are highly correlated to each other.
Therefore, this set of features is the most optimal set and there is no need of redundant feature present.



HEAT MAP

Strategy Development

We have used outputs from two ML models - **“CatBoost”** and **“Regressor Trees”**.

- Catboost is a classification algorithm which given 1/0 as outputs indicating the upward/downward movement of price on the next day.
- Regressor Trees give us the magnitude of the predicted return on the next day.

Along with these we have used a buy threshold value. This threshold is a hyperparameters which are tuned by the team for maximum profit.

Buy Threshold = **0.001**

Generating Trading Signals

BUY CONDITION:

- The algorithm gives a buy signal if the output from the CatBoost model is 1, which means that the return for the coming day will be positive.
- Along with it, the output from the Regressor Trees model must be greater than the buy threshold to generate a buy signal.

SELL CONDITION:

- The algorithm gives a sell signal if the output from the CatBoost model is 0, which means that the return for the coming day will be negative.
- Other than this, the position is squared off if the stop loss is hit.

Backtesting

We have used the given datasets to test the our trading strategy.
The results are as follows:

Profit	**7.83%**	**Maximum Drawdown**	**8.75%**
Sharpe Ratio	**1.49**	**Total Trades**	**27**

Risk Management

We have applied two risk management measures to minimize our losses:

- **Stop Loss** - The trade is squared off as soon as we hit a stop loss of **2%**.
- **Buy Threshold** - A new trade is not opened until the magnitude predicted by the regressor trees model is greater than the buy threshold value.

THANK YOU

FOR YOUR ATTENTION