International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-5, January 2021

Stock Market (NIFTY) Forecasting using Machine Learning Analysis on Option Chain

Saurabh Gupta, Vaishali, Raghuvansh Tahlan, Navya Sanjna Joshi, Ritvik Agarwal

Abstract*: Stock market prediction is a long-time intriguing topic to researchers from different fields. Stock market data is extremely volatile and hence laborious to model. In particular, innumerable studies have been conducted to predict the movement of stock market using Machine Learning algorithms such as Regression Techniques, Time Series Forecasting, Indices Modelling, Natural Language Processing and more, but there is still room for improvement. Also, Option chain and Options have been the subjects that not many have ventured into, leading us to this subject. Mainly, NIFTY and BANKNIFTY Options account for 70% of total derivatives traded and much more turnover than all stocks combined. This research paper attempts to figure out the utility of Option Chain in predicting the direction of movement in NIFTY. We have tried how different features from Option chain can be extracted, and the resulting problem can be solved using Machine Learning techniques and Deep Learning techniques.

Keywords: NIFTY Forecasting, Options, Option Chain, Stock Market Forecasting.

I. INTRODUCTION

Prediction of stock trend has long been an intriguing topic and is extensively studied by researchers from different fields. However, only a handful of studies on the relationship between the Stock Market and Option Chain/Options for data and analysis exist today. In the financial realm, Options are financial instruments that are derived from an underlying asset such as stocks or index[1]. An Options contract is an opportunity for the investor to buy or sell the underlying asset depending on the type of contract he holds. A call option gives the trader the right but not the obligation to buy a fixed quantity of the underlying asset at a given price.

Manuscript received on August 24, 2020.
Revised Manuscript received on January 28, 2021.
Manuscript published on January 30, 2021.

Dr Saurabh Gupta, Computer Science and Engineering Dr AkhleshDas Gupta Institute of Technology and Management, New Delhi, India, saurabh@adgitmdelhi.ac.in

Vaishali, Computer Science and Engineering Dr AkhleshDas Gupta Institute of Technology and Management, New Delhi, India, vaishalig1012@gmail.com

Raghuvansh Tahlan, Computer Science and Engineering Dr AkhleshDas Gupta Institute of Technology and Management, New Delhi, India,raghuvanshtahlan@gmail.com

Navya Sanjna Joshi, Computer Science and Engineering Dr AkhleshDas Gupta Institute of Technology and Management, New Delhi, India, joshinavya3@gmail.com

Ritvik Agarwal, Computer Science and Engineering Dr AkhleshDas Gupta Institute of Technology and Management, New Delhi, India,ritvikagarwal2@gmail.com

A put option, on the other hand, gives the trader the right but not the compulsion to sell the given quantity of an asset at a given price.An option chain lists all available option contracts, both puts, and calls, for given security. It shows all puts, calls, strike prices, and pricing information for a single underlying asset within a given maturity period[1].

II. OPTIONS AND OPTION CHAIN

A strike price, also known as the exercise price, is a predefined price at which a contract can be bought or sold when exercised[1]. For the case of call options, the strike price is the price at which the holder can buy the security; for put options, the strike price is the price at which the security can be sold. Also, depending upon the difference between the strike prices and underlying asset, the strike prices can be categorized into three categories. 'At the Money'(ATM) is a situation where the price of an option contract is equal to the price of the underlying asset. 'In the Money'(ITM) is when an option contract's price is less than the asset in case of Call option and more than the asset in case of Put option. 'Out of the Money' commonly referred to as OTM, is when the option contract's price is more than the asset for the call option and less than the asset for the put option. Open Interest (OI) refers to the total number of option contracts that are active/traded but not yet liquidated.On the other hand, Volume is the total number of contracts traded. Implied Volatility (IV) is the expected volatility of the underlying asset over the option's life. The expiry date for Options refers to the last date till which Options contracts are valid, and Option holders can exercise their rights. For instance,in the case of NIFTY and BANKNIFTY weekly options, they expire every Thursday at 3:30 pm.

Option Chain data is available on various online websites and portals. However, the one provided by the National Stock Exchange(NSE)[2] is the most accurate, reliable, and acts as a reference for all. The data is volatile and keeps changing every few minutes, so we collected the data every 5 minutes for six months, which became the basis of our research.

The data collected consists of Strike Price, Last Traded Price(LTP), Implied Volatility(IV), Open Interest(OI) and Traded Volume. Data from Option Contracts corresponding to these 7 Strike Prices are consolidated into a single row to have one row of data per Timestamp.Some of the features are derived from Statistics like Sum of Open Interest(OI) and Sum of Volume of all Strike Price for both Put and Call Option contracts.However, Trading Tutorials have also inspired us to extract some of the features, for example, Put Call Ratio commonly called as PCR for Open Interest and Volume[4].

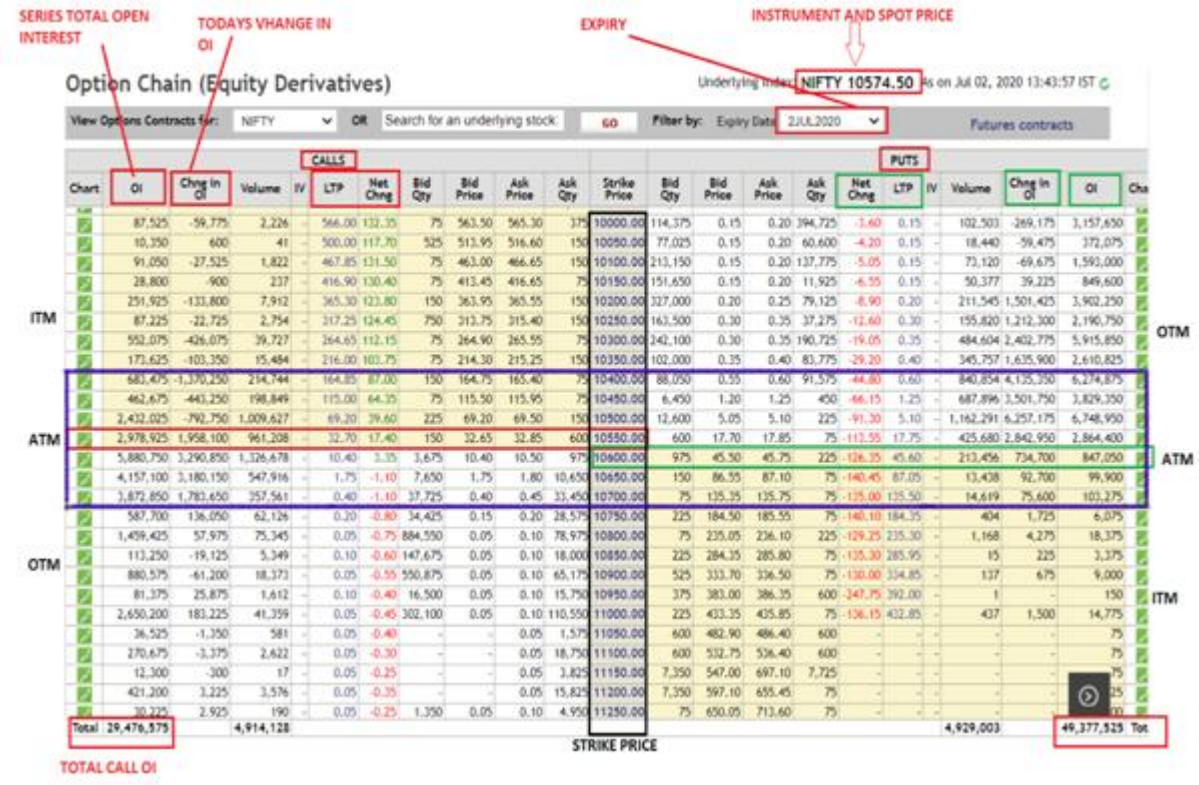

Figure 1: Key Option Chain concepts[3]

PCR is a Contrarian indicator in the trading world and denotes the ratio of the total number of Put option contracts and Call Option Contracts. Changes in the valueof Open Interest(OI), Implied Volatility(IV), Last Traded Price(LTP) and Volume in respectof preceding Timestamp are also used as features. The Strike Price where Open Interest is Maxiumum for Call Option Contracts is considered as 'Resistance' for the underlying asset. At the same time, 'Support' is the strike Price where Open Interest is Maximum for Put Option Contracts. Depending upon the signs of change in Open Interest and Last Traded Price we can also classify the Strike Price into one of the four Categories – Long BuildUp, Long Liquidation, Short BuildUp and Short Covering.[5]By observing the traded Volume at different times of the day, we have found that significant trades.happen during the first and last two market hours, so we have created a feature keeping that in mind.

Table 1: Interpretation of Change in OI and LTP in Call Contract.

CHANGE IN OPEN INTEREST CALL	CHANGE IN LAST TRADED PRICE CALL	INTERPRETATION	CALL SIGNAL
RISE	RISE	LONG BUILDUP	BULLISH
RISE	FALL	SHORT BUILDUP	BEARISH
FALL	RISE	SHORT COVERING	BULLISH
FALL	FALL	LONG LIQUIDATION	BEARISH

Table 2: Interpretation of Change in OI and LTP in Put Contract.

CHANGE IN OPEN INTEREST(OI) PUT	CHANGE IN LAST TRADED PRICE(LTP) PUT	INTERPRETATION	PUT SIGNAL
RISE	RISE	LONG BUILDUP	BEARISH
RISE	FALL	SHORT BUILDUP	BULLISH
FALL	RISE	SHORT COVERING	BEARISH
FALL	FALL	LONG LIQUIDATION	BULLISH

A typical row in the dataset consists of Timestamp, Expiry Date, PCR values, Interpretation of Change in Open Interest and Change in Last Traded Price(LTP), the value of the underlying asset(NIFTY), time of the Trading day, Trading Day(Monday, Tuesday, etc.) and data from ATM, Support and Resistance Strike Prices.Since we are in the initial stages of the research, we are not trying to predict the actual values of the NIFTY, but we are attempting to at least predict the trend of the NIFTY different time periods.The research will help usfigure out if the data extracted from Option Chain does have the power to predict and is worth researching further.Over some time we have observed that changes in the values of Option Chain are not readily reflected in the prices of NIFTY, so we conducted a series of experimentsto see how the accuracy of the model changes with different time periods.

Over some time we have observed that changes in the values of Option Chain are not readily reflected in the pricesWe formed a binary Target variable where it values 0 if the future value of NIFTY is less than the current value and 1 otherwise. Since our target variable is binary and binary and our data is numerical, we can use Classification Machine Learning as well as Deep Learning Models. For this research, we have used CatBoost Classifier[6], which is fast, open-sourceand provides good results with relatively fewer data. We have divided our data into two parts- 70% for Training and 30% for Testing and a particular random state to get steady results. We have trained our model for 300 iterations with 100 early stopping rounds and used scikit-learn's accuracy as well as ROC_AUC score as metrics for our research.

III. RESULTS AND OBSERVATIONS

This research is based mostly on predicting trends (upward/downward) in NIFTY. We did not have pre hand knowledge about the time it takes for the changes in Option chain to get reflected in the values of NIFTY. So we have tried multiple experiments in which we calculated the target variable using future values of NIFTY ranging from the immediate row to next 6^{th} row. As the timeframe or the time difference between each row is 5 minutes, this implies that we have tried predicting the trend of NIFTY using durations ranging from 5 minutes to 30 minutes. Training scores were always superior to testing scores, but both scores follow a similar trend, where the scores dropped for the 'Next 2^{nd} row' (10 minutes in future) and then rising to their maximum in 'Next 6^{th} row' (30 minutes in future) and saturating at 'Next 5^{th} row.' We also believed that by using preceding rows of data along with the current row might produce some improvements in results. So we conducted experiments where up to 3 previous rows were used in addition to the current row to predict the target variable. However, this approach did not lead to any significant improvement in the result.

While understanding the 'feature importance' of the most effective model, we encountered something interesting. Trading Tutorials and videos have always emphasized on the use of Put Call Ratio(PCR) of Open Interest,and we do agree with this because PCR of Volume was our most important feature.

However, surprisingly PCR of Open Interest does not feature even in the top 15 'features'. Some other notable features include Open Interest values of the Put contract for Resistance row and Call Contract's for Support row. Implied Volatility(IV) does not feature anywhere in the top 15.

IV. CONCLUSION

In the project, we researched the use of the Option chain data to help predict the future trend of the underlying asset/NIFTY. Numerical results suggest high efficiency with 60% accuracy after 25 minutes and 70% accuracy after 30 minutes for the prediction of the NIFTY trend. PCR of Volume was a prominent feature, but surprisingly PCR of Open Interest and Implied Volatility(IV) proved to be of little utility. Some noteworthy mentions include Open Interest values of the Put contract for Resistance and Call Contract's for Support. However, it is still far from what could be used single-handedly in trading because knowing the trend after 30 minutes might not give the confidence to invest. Option Chain alone data may not be powerful enough. However, when coupled with the data of Equities, Futures and Sentiments and appropriate Deep Learning Technologies, it may become a game-changer.

REFERENCES

1. https://www.investopedia.com/terms/o/option.asp
2. https://www.nseindia.com/option-chain
3. https://dotnettutorials.net/lesson/option-chain-analysis/
4. https://www.investopedia.com/trading/forecasting-market-direction-with-put-call-ratios/
5. https://fyers.in/school-of-stocks/chapter/options/reading-option-chain-using-volume-and-ol.html
6. https://catboost.ai/docs/concepts/python-reference_catboostclassifier.html
7. Chia, C., 2020. *Prediction Of Stock Price Movements Using Options Data*., http://cs229.stanford.edu/proj2016spr/report/052.pdf

Table 3: Results of various Experiments

EXPERIMENT		TRAINING ACCURACY	TESTING ACCURACY	TRAINING ROC_AUC	TESTING ROC_AUC
TARGET	PREVIOUS ROWS				
NEXT ROW 1	NA	0.67	0.57	0.67	0.57
NEXT ROW 2	NA	0.71	0.55	0.7	0.54
NEXT ROW 3	NA	0.91	0.61	0.91	0.6
NEXT ROW 4	NA	0.91	0.65	0.91	0.65
NEXT ROW 5	NA	0.93	0.69	0.93	0.68
NEXT ROW 6	NA	0.93	0.7	0.93	0.69
NEXT ROW 1	PREV1	0.71	0.57	0.71	0.57

NEXT ROW 2	**PREV1**	0.69	0.56	0.68	0.55
NEXT ROW 3	**PREV1**	0.9	0.61	0.9	0.6
NEXT ROW 4	**PREV1**	0.94	0.63	0.94	0.63
NEXT ROW 5	**PREV1**	0.94	0.66	0.94	0.65
NEXT ROW 6	**PREV1**	0.94	0.7	0.94	0.69
NEXT ROW 1	**PREV1 + PREV2**	0.65	0.56	0.64	0.56
NEXT ROW 2	**PREV1 + PREV2**	0.81	0.56	0.8	0.55
NEXT ROW 3	**PREV1 + PREV2**	0.87	0.58	0.87	0.57
NEXT ROW 4	**PREV1 + PREV2**	0.93	0.64	0.93	0.63
NEXT ROW 5	**PREV1 + PREV2**	0.94	0.66	0.94	0.65
NEXT ROW 6	**PREV1 + PREV2**	0.94	0.69	0.94	0.69
NEXT ROW 1	**PREV1 + PREV2 + PREV3**	0.64	0.56	0.64	0.56
NEXT ROW 2	**PREV1 + PREV2 + PREV3**	0.85	0.56	0.85	0.55
NEXT ROW 3	**PREV1 + PREV2 + PREV3**	0.95	0.6	0.95	0.59
NEXT ROW 4	**PREV1 + PREV2 + PREV3**	0.96	0.61	0.96	0.61
NEXT ROW 5	**PREV1 + PREV2 + PREV3**	0.96	0.65	0.95	0.64
NEXT ROW 6	**PREV1 + PREV2 + PREV3**	0.95	0.69	0.95	0.68

AUTHORS PROFILE

Dr Saurabh Gupta works asProfessor in Computer Science and Engineering Department in Dr. Akhilesh Das Institute of Technology & Management affiliated to GGSIPU, Delhi.

Vaishali works as Assistant Professor in Computer Science and Engineering Department in Dr. Akhilesh Das Institute of Technology & Management affiliated to GGSIPU, New Delhi. She has contributed in publishing papers in the area which includes web attacks, Internet of things, vulnerability analysis, brief introduction to machine learning and data mining.

Raghuvansh Tahlanis a final year student in Computer Science and Engineering Department in Dr. Akhilesh Das Institute of Technology & Management affiliated to GGSIPU, New Delhi. He is interested and enthusiastic in researching the application of Data Science in the field of Sports Analytics and Algorthmic Trading.

Navya Sanjna Joshiis a final year student in Computer Science and Engineering Department in Dr. Akhilesh Das Institute of Technology & Management affiliated to GGSIPU, New Delhi. She is interested and enthusiastic in researching about design methodologies and formal approaches in fields of computer science like, social computing, data analysis and more.

Ritvik is a final year student in Computer Science and Engineering Department in Dr. Akhilesh Das Institute of Technology & Management affiliated to GGSIPU, New Delhi. He is interested in puzzles of all kinds and cracking difficult logical scenarios and also am an enthusiast in the world of machine learning.