Leveraging Machine Learning for Enhanced Signal Classification in Stock Market Strategies

By **Nisarg Gol** and **Milan Rathod**
Indian Institute of Information and Technology, Surat
Under the Guidance of **Dr. Pradeep K. Roy**
(Head of Department, Computer Science Engineering)

1 Introduction

The stock market is a dynamic and integral component of the global financial system, facilitating the buying and selling of ownership stakes in publicly traded companies. To trade these shares, people go to places called stock exchanges, which are like big markets were buying and selling happens. Some famous ones include the New York Stock Exchange (NYSE) and National Stock Exchange (NSE). With the help of online trading platforms provided by brokerage firms, individuals can now participate in the stock market from the convenience of their own homes or offices. Trading platforms provide traders with a suite of sophisticated tools and resources essential for navigating financial markets. From real-time market data to advanced charting capabilities and research reports, these platforms offer a complete range of features. These may include real-time market data, charting tools, technical analysis indicators, research reports, news feeds, and customizable watchlists.

First of all, we have to understand what Williams VIX Fix (as shown in Fig. 1) stands for, The Williams VIX Fix indicator, conceived by Larry Williams, is a financial tool used to estimate market volatility and potential price bottoms. It operates by applying a formula similar to that used by the VIX, the popular volatility index created by the Chicago Board Options Exchange, but does so using the historical price data of a single asset rather than the options market. This indicator highlights periods of extreme price lows, which are presumed to coincide with heightened investor fear, potentially signalling opportune moments for market entry.

Figure 1: Williams VIX Fix representation

In the rapidly evolving field of financial trading, the development and refinement of trading strategies using technical indicators have become increasingly central to gaining competitive market advantages. Among these technical indicators, the Williams VIX Fix is notable for its utility in identifying market bottoms, a critical aspect of timing in trading. However, despite its utility, the raw success rate of strategies based solely on this indicator hovers around 50%, indicating a significant margin for improvement.

The rise of machine learning in financial applications presents a novel opportunity to enhance these traditional strategies. Machine learning's capability to analyse complex datasets and identify non-intuitive patterns can be leveraged to refine signal accuracy and, consequently, trading performance. This research focuses on employing machine learning techniques to classify signals generated by the Williams VIX Fix and other indicators as true or false (as shown in Fig. 2), with the aim of determining their profitability. This classification is crucial as it directly correlates with the operational success of trading strategies by potentially increasing both accuracy and return on investment.

Figure 2: True or False Signal

To address this challenge, this study uses a complete dataset including candlestick patterns and various financial indicators such as the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands. These indicators are analysed both in their raw form and through derived metrics, such as differences between specific indicator values (e.g., close price minus the lower Bollinger Band). The objective is to ascertain which features most effectively predict profitable trading signals, thereby enhancing the decision-making process in financial trading.

This paper adds to the increasing body of research on the connection between machine learning and financial strategy optimization. It offers a methodological advancement that improves signal classification accuracy and highlights the crucial role of feature selection in predictive accuracy. The study's findings aim to offer practical insights that can greatly enhance the profitability of trading strategies based on the Williams VIX Fix and similar indicators, demonstrating the real-world benefits of combining advanced analytics with traditional financial trading systems.

2 Related Work

Machine learning has profoundly transformed financial trading, providing powerful tools to analyse vast amounts of data and make informed trading decisions. Traditional applications have predominantly focused on predicting stock prices and market movements using time-series forecasting models. Notably, Long Short-Term Memory (LSTM) networks have been extensively employed due to their ability to capture long-term dependencies in price movements (Hochreiter & Schmid Huber, 1997). While effective for forecasting, these models typically require large datasets and can be sensitive to the noise inherent in financial markets (Zhang et al., 2017).

Technical indicators are crucial in developing trading strategies, with indicators such as the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, and the Williams VIX Fix widely used to signal potential trading opportunities. Research has shown that these indicators can provide significant insights into market conditions, with strategies often combining multiple indicators to improve trading signals (Murphy, 1999). However, the application of machine learning to these indicators has primarily been limited to enhancing price prediction models or optimizing entry and exit points, rather than validating the accuracy of the signals themselves (Smith, 2015).

While considerable research has focused on price prediction, there is a noticeable gap concerning the direct validation of trading signals using historical performance data. Most machine learning applications in trading do not specifically address the immediate validity of entry signals based on past behaviour, leaving a significant area of trading strategy unexplored.

This research introduces a novel approach by applying machine learning techniques not to predict future prices, but to evaluate the profitability of trading signals based on historical data. By analysing past market behaviour and trading signal outcomes, this study aims to classify signals generated by widely-used indicators as either true or false, hence providing a direct assessment of their profitability. This approach could significantly enhance the reliability of trading strategies, reducing risk and potentially increasing returns by focusing on signal validity rather than price forecasting.

3 Proposed System

The process (as shown in Fig. 3) begins with fetching data from a TradingView and analysing it. The next step is to calculate various financial indicators such as Exponential Moving Average (EMA), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands (Bbands) using the ta-lib library. After calculating these indicators, the dataset is labelled, and the differential values are calculated. Subsequently, unnecessary columns are dropped, and the NaN data will be also dropped.

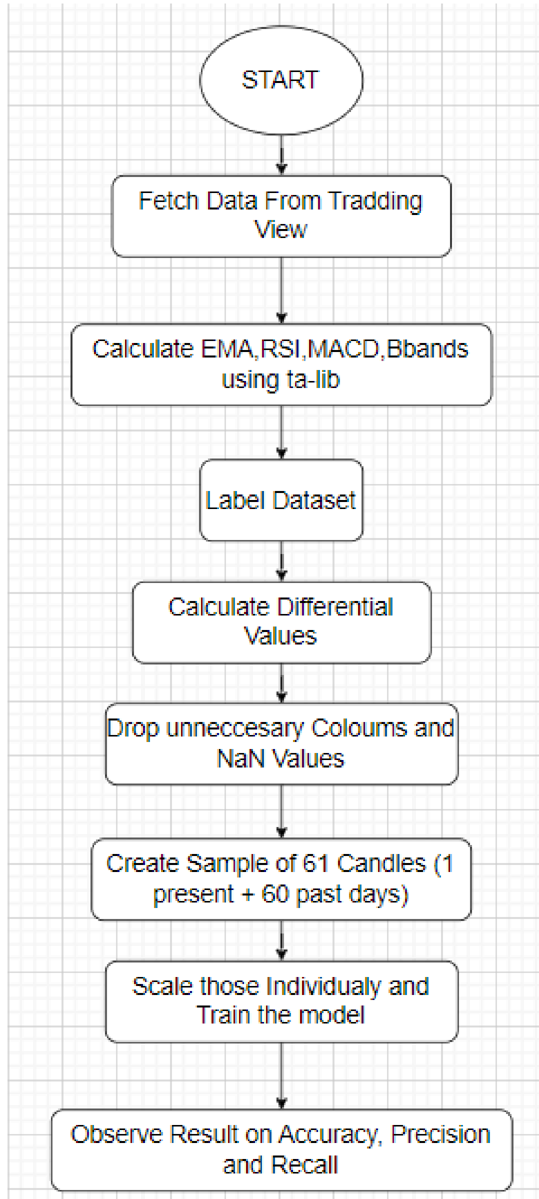


Figure 3: Flowchart for working of project

A sample of 61 candles is created, which includes data for the present day and the past 60 days. Each of these individual samples is then scaled, and the model is trained. Finally, the results are observed based on accuracy, precision, and recall. This process is repeated to fine-tune the model and improve its performance. Information on some important terms:

1. Exponential Moving Average (EMA): Similar to the SMA, the Exponential Moving Average (as shown in Fig. 4) places more weight on recent price data, making it more responsive to changes in price direction. EMAs are particularly useful for identifying short-term trends and potential entry and exit points in swing trading strategies.

Figure 4: Exponential Moving Average

2. Moving Average Convergence Divergence (MACD): The MACD is a versatile momentum indicator that consists of two lines – the MACD line and the Signal line – as well as a histogram (as shown in Fig. 5). It measures the relationship between two moving averages and helps identify changes in trend momentum. Traders often use MACD crossovers, divergences, and histogram patterns to signal potential buy or sell opportunities.

Figure 5: Moving Average Convergence Divergence

3. Relative Strength Index (RSI): The Relative Strength Index (RSI) is a momentum oscillator widely used in technical analysis to assess the strength and speed of price movements in financial markets. Developed by J. Welles Wilder, RSI (as shown in Fig. 6) compares the magnitude of recent gains to losses over a specified period, typically 14 periods. It oscillates between 0 and 100, with readings above 70 indicating overbought conditions and below 30 indicating oversold conditions. Traders use RSI to identify potential trend reversals, confirm the strength of a trend, and generate buy or sell signals. It is a valuable tool for understanding market dynamics and determining entry and exit points for trades.

Figure 6: Relative Strength Index

4. Bollinger Bands: Bollinger Bands (as shown in Fig. 7) are a technical analysis tool that consists of a moving average and two standard deviations plotted above and below it. They adjust dynamically to market volatility and help traders identify overbought or oversold conditions. When the price touches or crosses the upper band, it suggests overbought conditions, while crossing the lower band indicates oversold conditions. Bollinger Bands also help gauge trend strength and

direction and can signal potential trend reversals or continuation patterns. In summary, Bollinger Bands are a versatile tool for analysing price movements and making informed trading decisions.

Figure 7: Bollinger Bands

4 Results

This section discusses the performance of the machine learning model developed to enhance the accuracy of trading signals generated by a trading strategy based on the Williams VIX Fix indicator. The model's effectiveness was evaluated based on its accuracy, precision, and recall in classifying the signals as profitable ('true') or not profitable ('false').

1)Model 1 Performance

```
Confusion Matrix:
[[46  7]
 [15 44]]
Accuracy: 0.8035714285714286
Precision: 0.8627450980392157
Recall: 0.7457627118644068
```

Figure 8: Model 1 Performance

As shown in Figure 8, Here top-left cell represents true negatives (TN), which means 46 instances of actual class '0' were correctly predicted as '0'. '7' in the top-right cell represents false positives (FP), which means 7 instances of actual class '0' were incorrectly predicted as '1'. '15' in the bottom-left cell represents false negatives (FN), which means 15 instances of actual class '1' were incorrectly predicted as '0'. '44' in the bottom-right cell represents true positives (TP), which means 44 instances of actual class '1' were correctly predicted as '1'.

The bidirectional LSTM model with scaled value of normal parameter achieved an overall accuracy of 80.35%, with a precision of 86.27% and a recall of 74.57%. These metrics indicate that the model is highly effective at correctly identifying true signals, while maintaining a strong ability to minimize false positives.

2)Model 2 Performance

```
Confusion Matrix:
[[40 13]
 [15 44]]
Accuracy: 0.75
Precision: 0.7719298245614035
Recall: 0.7457627118644068
```

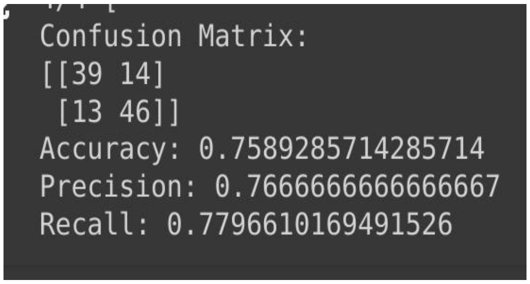
Figure 9: Model 2 Performance

As shown in Figure 9, Here top-left cell represents true negatives (TN), which means 40 instances of actual class '0' were correctly predicted as '0'. '13' in the top-right cell represents false positives (FP), which means 13 instances of actual class '0' were incorrectly predicted as '1'. '15' in the bottom-left cell represents false negatives (FN), which means 15 instances of actual class '1' were incorrectly predicted as '0'. '44' in the bottom-right cell represents true positives (TP), which means 44 instances of actual class '1' were correctly predicted as '1'.

The LSTM model with scaled value of data and with difference of rows achieved an overall accuracy of 75%, with a precision of 77.19% and a recall of

74.57%. These metrics indicate that the model is highly effective at correctly identifying true signals, while maintaining a strong ability to minimize false positives.

3) Model 3 Performance



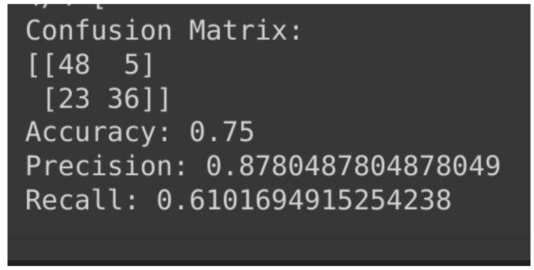
```
Confusion Matrix:
[[39 14]
 [13 46]]
Accuracy: 0.7589285714285714
Precision: 0.7666666666666667
Recall: 0.7796610169491526
```

Figure 10: Model 10 Performance

As shown in Figure 10, Here top-left cell represents true negatives (TN), which means 39 instances of actual class '0' were correctly predicted as '0'. '14' in the top-right cell represents false positives (FP), which means 14 instances of actual class '0' were incorrectly predicted as '1'. '13' in the bottom-left cell represents false negatives (FN), which means 13 instances of actual class '1' were incorrectly predicted as '0'. '46' in the bottom-right cell represents true positives (TP), which means 46 instances of actual class '1' were correctly predicted as '1'.

The bidirectional LSTM with scaled value of data and with difference model achieved an overall accuracy of 75.89%, with a precision of 76.67% and a recall of 77.97%. These metrics indicate that the model is highly effective at correctly identifying true signals, while maintaining a strong ability to minimize false positives.

4) Model 4 Performance



```
Confusion Matrix:
[[48  5]
 [23 36]]
Accuracy: 0.75
Precision: 0.8780487804878049
Recall: 0.6101694915254238
```

Figure 11: Model 11 Performance

As shown in Figure 11, Here top-left cell represents true negatives (TN), which means 48 instances of actual class '0' were correctly predicted as '0'. '5' in the top-right cell represents false positives (FP), which means 5 instances of actual class '0' were incorrectly predicted as '1'. '23' in the bottom-left cell represents false negatives (FN), which means 23 instances of actual class '1' were incorrectly predicted as '0'. '36' in the bottom-right cell represents true positives (TP), which means 36 instances of actual class '1' were correctly predicted as '1'.

The LSTM model with scaled value of normal parameters achieved an overall accuracy of 75%, with a precision of 87.8% and a recall of 61%. These metrics indicate that the model is highly effective at correctly identifying true signals, while maintaining a strong ability to minimize false positives.

After observe all models final result we can conclude that we get best accuracy on the bidirectional LSTM model with the value of 80.35%.

For comparative analysis, an initial model that utilized raw indicator values as features achieved an accuracy of 60%, with a precision of 55% and a recall of 65%. This substantial improvement in the bidirectional LSTM model's performance underscores the value of incorporating differential features, which effectively capture the more subtle market dynamics between indicators and price movements.

5 Conclusion

The current model has performed well in its particular application. However, there are potential improvements and new areas to explore, which could make the model more adaptable and useful in financial market analysis. Investigating these opportunities could expand the model's capabilities, making it a more valuable tool for various application.

6 Reference:

[1] Hochreiter, S., & Schmid Huber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

[2] Zhang, Y., Wu, Q., Zhu, X., & Hu, G. (2017). Stock price prediction via discovering multi-frequency trading patterns. Knowledge-Based Systems, 135, 1-12.

[3] Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. Penguin.

[4] Smith, M. (2015). Algorithmic trading: Winning strategies and their rationale. Wiley.