SANGAM International Journal of Multidisciplinary Research

Vol. 01, Issue 02, March (2024) **E-ISSN: 2584-086X**

# DATA ANALYSIS AND MATHEMATICAL APPROACH FOR TRADING IN NIFTY FUTURES FOR PROFITABILITY

**Vinod Kumar Joshi**

Deputy Superintendent of Police, National Crime Records Bureau Ministry of Home Affairs, New Delhi
Technical Resident, Office of Digital Transformation, Indian Institute of Technology, Jodhpur, Rajasthan
Research Scholar, Sangam University, Bhilwara, Rajasthan

**Available online at: www.sijmr.org**

***Abstract —*** The Indian Futures and Options (F&O) stock market is highly volatile, and conventional trading methods are challenging. Our research focuses on reducing the financial losses of retail intraday traders. In this study, we have analysed the open, high, low, and close (OHLC) prices of stock data from the previous five years and applied a novel mathematical approach to buy and sell the stock for intraday trading in the NIFTY Futures. This Probabilistic Profitable Model (PPM) framework suggests a trading method based on mathematically proven results. We have focused on intraday trading methods in which buying and selling are frequent. We aim to buy low and sell high to become profitable. Our data analysis method provides a trading accuracy of 90% for the NIFTY futures.

***Keywords****— Probabilistic Profitable Model, Probability, Algorithm, Set Theory, OHLC Data, Trading Method, Intraday, NIFT, Futures and Options*

## I. INTRODUCTION

The Securities and Exchange Board of India (SEBI) has released a comprehensive report detailing the financial performance of individual traders engaged in the Equity Futures and Options (F&O) Segment. According to the findings of the SEBI report, a significant majority of individual traders, specifically 89% of them, experienced financial losses within the equity F&O segment, signifying that 9 out of every 10 individual traders incurred losses [1]. The stock market consists of mainly three trader categories. The first category is characterized by The Uninformed Investor, who adopts a gambling-like approach, relying predominantly on luck rather than thorough analysis. In the second category, we find the independent systematic trader, who keeps trade journals to assess and enhance decision-making for future strategies. These traders also leverage technology by utilizing APIs to facilitate efficient trading. The third category is represented by the algorithmic trader, who not only employs computers for trade execution but also relies on them to formulate trading strategies. Our focus is to minimise the losses by leveraging algorithmic technology.

The stock market analysis is complex because stock market volatility and fluctuations are very high. The volatility of the stock market is unpredictable and difficult [2]. Retail investors and traders often lose money because they do not have the expertise in statistics to predict market trends. A depth of statistics and programming knowledge is necessary to anticipate probabilistic forecasting [3]. The Probabilistic Profitable Model (PPM) will use mathematics, statistics, programming, and quantitative analysis for algorithmic investment strategies that can be more profitable. The historical data from the stock exchanges, commodity exchanges, and currency data can be analysed to check the historical movements and patterns. Our algorithm is based on the probability of occurrence of the same event multiple times.

It has been observed that automated strategies based on strict mathematical and statistical models offer better returns. Instead of predicting the prices for the upcoming days, strict mathematical and statistical models are better because mathematical and statistical models are not affected by factors like news and public sentiment [4].

Retail traders and investors commonly employ both fundamental and technical analyses in their decision-making processes. Fundamental analysis serves as a foundation for a long-term investment strategy, aiming to progressively accumulate wealth through vehicles like mutual funds or actively managing a diversified portfolio of stocks, bonds, or stock baskets. On the other hand, technical analysis is instrumental in navigating the short-term and volatile nature of the market, guiding frequent transactions based on the discernment of trends within the stock market [5].

Some experts do not favour the predictability of stock markets. Economic events and announcements make it hard to predict the stock market trends. At the same time, other researchers have found that the stock market often follows similar trends and patterns based on past experiences.

Recently, it has been observed that intraday trading methods are very popular, especially after COVID-19 [6]. Intraday trading, also known as day trading, means the traders buy and sell stocks within the same day. The purpose is to earn profit based on the movement of the stock market. Indian brokers like Groww, AngelOne, Zerodha, Kotak, and many others provide the leveraging means brokers lend money to their customers to amplify investment returns.

Utilizing leverage in intraday trading presents an increased array of trading opportunities, albeit accompanied by a heightened susceptibility to losses. Overleveraging stands out as the predominant factor contributing to substantial setbacks in the stock market. A considerable number of inexperienced traders find themselves at risk of depleting their entire capital when engaging in trading activities with an extensive margin.

Stock market price manipulation is also done by the operators to trap the retailers [7]. Our research developed a probabilistic profitable model that can be applied to any stock and futures, making our model an efficient dynamic trading strategy based on the current pattern of stock prices. However, initially, our research has been conducted on the Indian futures contract market, especially the NIFTY futures contract. A futures contract constitutes a legally binding arrangement to purchase or sell a standardized asset on a specified date or within a particular month. These contracts are characterized by their standardization, ensuring interchangeability, and they explicitly outline various contract specifications.

## I. RELATED WORK

The Bombay Stock Exchange (BSE) holds the distinction of being India's premier and oldest securities market, with its establishment dating back to 1875. In the realm of stock market analysis, various methodologies for predicting market patterns have been explored by data analysts. These predictive approaches encompass indicators, machine learning, and empirical analysis.

In our research, we undertake an empirical analysis specifically focused on technical trading rules. This form of empirical analysis involves monitoring key indicators such as price moving averages, momentum, and trading volume [8]. Employing a statistical analysis framework, our methodology adheres to rule-based trading principles, relying on the observed movements within stock market trends rather than making decisions based on forecasts or external information. This approach dictates when to initiate buy or sell actions for a given stock [9].

Machine learning prediction collects historical data and applies various machine learning models to predict the prices. Many researchers presented a detailed review of stock market predicting methodologies, like the Bayesian model, Support Vector Machine (SVM) classifier, Fuzzy classifier, Artificial Neural Networks (ANN), Neural Networks (NN), and similar Machine Learning Methods. [10].

One of the standard empirical analysis methods of trading based on opening, high, low, and closing (OHLC) prices of stocks has a hidden pattern that has yet to be extensively studied in the various literature. Many researchers study candlestick patterns (sequence of opening, high, low, and closing (OHLC) price data structures) to predict futures price trends in financial markets. A candle or candlestick is a visual representation used to show four essential price indicators of a stock over time. Each candlestick represents a specific price change within a single day, including the opening, closing, highest, and lowest prices [11].

Diverging from conventional methodologies, our approach diverges from exclusive reliance on sentiment analysis of social media posts, examination of candlestick chart patterns, or predictions derived from machine learning algorithms. Instead, this study introduces a rigorously mathematical strategy, substantiated by proof, for forecasting stock movements. [12].

Recently, Artificial Intelligence (AI) has been employed to create techniques for categorizing and forecasting stock market trends. Analysts utilize Machine Learning methods to leverage the predictive capabilities of Opening, High, Low, and Closing (OHLC) data. Implementing these technologies in real-life scenarios demands substantial computational power and in-depth knowledge. For less experienced traders, the options are investing in AI technology to predict trends or relying on market sentiments for trading. We've introduced a Probabilistic Profitable Model that novice traders can use for any stock.

## II. METHODOLOGY

Within this section, we elucidate the fundamental definitions of candlestick patterns, encompassing all pertinent terms and functions integral to their specifications. It is noteworthy that the terms "candle" and "candlestick" are used interchangeably throughout this paper. A candlestick is comprised of four essential indicators: the opening price, high price, low price, and closing price, commonly denoted as open, high, low, and close. Figure 1 provides visual representations of two examples, one featuring a green candlestick and the other exhibiting a red candlestick. [11].

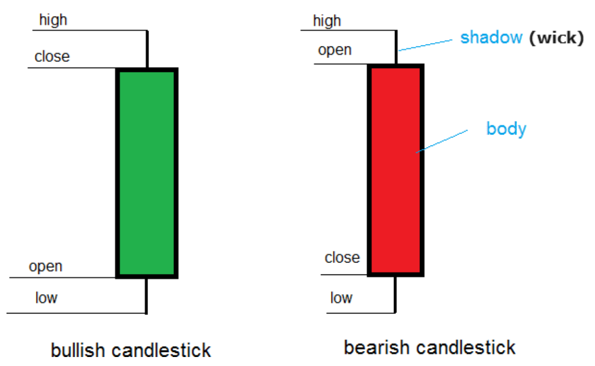


*Figure 1: Pictorial representation of Japanese candlestick chart for OHLC data.*

3.1. Definitions of candlesticks and patterns

A candlestick, originating from Japan, is a specialized type of price chart that visually represents the high, low, open, and closing prices of a security within a specific time frame. The candlestick takes on a rectangular shape, and its extremities are referred to as wicks.

Each individual candlestick conveys crucial price data for a stock, encapsulating four key elements: the opening price, closing price, high price, and low price. The coloration of the central rectangle within the candlestick serves as an indicator of whether the opening or closing price held greater prominence.

A red-hued candlestick signifies that the closing price for the designated period was lower than the opening price. This red candlestick serves as a bearish signal, indicative of selling pressure in the market. Conversely, a green-hued candlestick denotes that the closing price exceeded the opening price, serving as a bullish signal and reflecting buying pressure.

Historical data

It has been noticed that looking at how a stock has performed in the past can give us clues about what might happen in the future. There are two types of analysts – technical and fundamental. Technical analysts study past prices to find levels where the stock was supported or faced resistance. On the other hand, fundamental analysts use historical prices as one of the factors to figure out how much a company is worth and its potential for growth [13].

Historical data can be obtained from multiple sources, including brokers such as Yahoo Finance, Zerodha, Angel One, Upstox, 5paisa, Dhan, and Fyers [14]. We have acquired data for the past five years for analysis purposes. The logic and pseudo code are depicted in Figure 2.

```
#This script to fetch historical data from various broker
#The broker may have syntax differences.

SET exchange = name of exchange NSE, NFO, BSE
SET symboltoken = each stock must a have token
SET interval = candle interval can of one minute, five minute etc
SET fromdate = start date of historical data
SET todate = end  date of historical data
SET array_of_date as five years calender dates

FOR each item in array_of_date
    SET historical_array = Historical_data_from_broker
    FOR each item in historical_array
        IF item not empty
            INSERT item to database
        ENDIF
    ENDFOR
ENDFOR
```

*Figure 2: Pseudo script to fetch historical data.*

Concerning the pseudo-code written in Figure 2, historical data has been stored in MongoDB to analyse the data. Historical data includes information over time, with a focus on opening, high, low, and closing prices and volume data. These details help recognize patterns for predicting stock prices. While relational databases have conventionally stored this data, the increasing volume, and the emergence of non-relational databases challenge whether relational databases are still the best choice. The findings indicate that switching to a MongoDB database would be more cost-effective and space-efficient and offer better throughput [15].


```
{
  "status": true,
  "message": "SUCCESS",
  "errorcode": "",
  "data": [
    [
      "2023-12-01T09:15:00+05:30", /*Time Stamp*/
      20194.1, /*open*/
      20263.55, /*high*/
      20188.3,/*low*/
      20258.65,  /*close*/
      0
    ],
```


*Figure 3: JSON format for database.*

 

The MongoDB uses JSON format to store the database. Most brokers provide the historical data in JSON format. JSON, or JavaScript Object Notation, offers numerous advantages. Its human-readable format simplifies both reading and writing, while its lightweight nature ensures efficiency in data transmission and storage. Being language-agnostic, JSON facilitates seamless integration across various programming environments. Parsing is straightforward, with built-in support in many languages. Its support for nested structures makes it versatile for representing complex data hierarchies. JSON's widespread adoption, schema-less nature, and compatibility with web browsers contribute to its popularity. Additionally, the format supports various data types, providing flexibility in data representation. Overall, JSON stands out as a versatile, widely accepted, and efficient data interchange format [16].

## Constraints

The candlestick is characterized by four float-type parameters: open, high, low, and close. A float, denoting a floating-point number, implies a numerical value with a decimal place. Illustrated in Figure 3, the candlestick displays values of 20194.1 for the open, 20263.55 for the high, 20188.3 for the low, and 20258.65 for the close prices.

In the specific context of the Indian stock market, the open price is the value at 9:15 hours when the market commences trading. Conversely, the close price is the value at 15:30 hours when the market concludes trading for the day. Between 9:15 and 15:30 hours, the highest recorded price is termed the high, while the lowest is termed the low.

The values of open, high, low, and close vary as time progresses—for example, the candlestick type changes based on the time interval. When the market opens at 9:15 hours, the corresponding candlestick represents the open price. After one hour, from 9:15 to 10:14 hours, the candlestick becomes a one-hour candle, encapsulating its specific open, high, low, and close values within that timeframe.

Let's assume the open, high, low, and close prices of a specific stock as $N_O$, $N_H$, $N_L$, and $N_C$, respectively. The prediction tasks should adhere to the following constraints:

(i). The prices must be positive; hence, they should not be negative or zero:

$$N_O > 0, \tag{1}$$

$$N_O > 0, \tag{2}$$

$$N_H > 0, \tag{3}$$

$$N_L > 0, \text{ and} \tag{4}$$

$$N_C > 0, \tag{5}$$

(ii). The high price must surpass the low price:

$$N_H > N_L \tag{6}$$

(iii). The open and close prices should fall within the range defined by the high and low prices:

$$N_O \in [N_L, N_H], \text{ and} \tag{7}$$

$$N_C \in [N_L, N_H] \tag{8}$$

The candles depict the sentiments of the market; therefore, it is essential to record accurate data for processing [17].

## Proposed Method

The proposed approach relies on algorithmic probability, mainly when the seller exhibits aggression in selling futures contracts or the buyer is assertive in acquiring futures contracts following the initial balance. Put plainly, the term "initial balance" (IB) pertains to the price data established in the initial hour of a trading session, which spans from 09:15 am to 10:14 am in the context of the Indian market. Traders' activity during this period shapes the Initial Balance, a concept introduced by Peter Steidlmayer when presenting the market profile [18].

Post-initial balance, the opening and closing of candles serve as support and resistance. Our strategy emphasizes selling futures contracts near the closing price during a green candle (close greater than opening) and, conversely, supporting buying near the close price during a red candle (close lower than opening).

Let's assume the opening and closing prices after the Initial Balance (IB) hours are represented as No and Nc, respectively. In the case of Nc > No (green candle), we focus on selling futures contracts, while for Nc < No (red candle), we emphasize buying futures contracts. Utilizing the Probabilistic Profitable Model (PPM), we aim to determine the probability of achieving a 5-point gain through either selling or buying the futures contracts.

 


*Figure 4: Initial Balance (IB) & Green Candle*

As depicted in Figure 4, the initial equilibrium was set between 9:15 am and 10:14 am. Within this period, given that the closing price surpasses the opening price (indicated by a green candle, where Nc > No), our attention is directed towards selling futures contracts in proximity to the closing price.


*Figure 5: Initial Balance (IB) & Red Candle*

As depicted in Figure 5, the initial balance was set between 9:15 am and 10:14 am. Within this period, owing to a lower closing price compared to the opening (as denoted by a red candle, where Nc < No), our primary focus is on buying futures contracts near the closing price.

According to the Probabilistic Profitable Model (PPM) theory, opting to buy and sell in proximity to the closing price increases the likelihood of securing favourable returns. This is grounded in the theory's assertion that, within the last hour, either sellers or buyers are typically depleted capitals, thereby significantly raising the chances of a reversal post the initial balance.

## III. Results and Discussion

Our algorithm produces distinct outcomes tailored to various target groups, each categorized based on specific criteria. The first category pertains to instances where the target exceeds zero, indicating a profitable outcome for that day. The magnitude of profit is contingent upon the quantity of lots transacted during the intraday session. In Futures and Options, a lot represents the minimum number of shares defined by regulatory bodies such as SEBI, encompassing all tradable stocks and indices. For instance, the Nifty 50 lot size is fixed at 50 shares, allowing options trading in multiples of 50. Consequently, if a user buys or sells one lot (50 shares) and the NIFTY moves one point up or down, the user stands to gain or lose 50 Indian Rupees, under the Probabilistic Profitable Model (PPM) theory [19].

Our experimental analysis involves the examination of the last five years of NIFTY futures data obtained from the Angel One broker. According to the Profitable Model (PPM) algorithm, when the target is greater than zero, the anticipated accuracy of the algorithm stands at an impressive 98.50%. This signifies that users employing the algorithm are likely to achieve profitable results in the majority of cases. Figure 6 illustrates the PPM algorithm's success in hitting 723 out of 734 targets over the past five years.

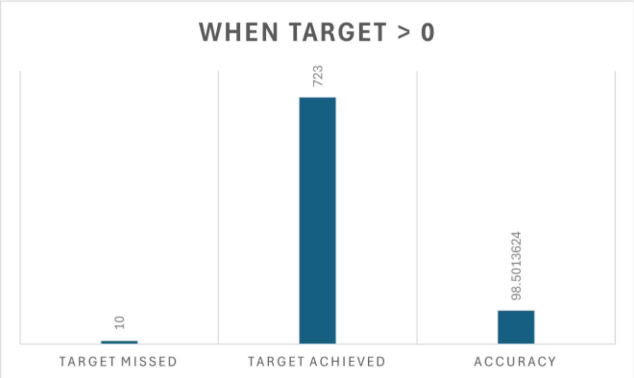


*Figure 6: Target greater than zero with accuracy*

Moving on to the second target group, where the expectation is to gain a profit of more than five points in the trade, Figure 7 showcases the PPM algorithm's performance, achieving 629 out of 734 targets with an accuracy of 85.69%.

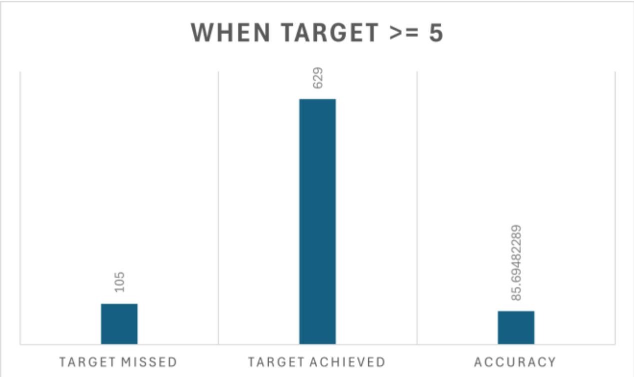


*Figure 7: Target greater than five with accuracy*

In the third target group, where the aim is to secure a profit of more than ten points in the trade, Figure 8 demonstrates that the PPM algorithm has successfully reached 543 out of 734 targets, achieving an accuracy rate of 73.97%.

 

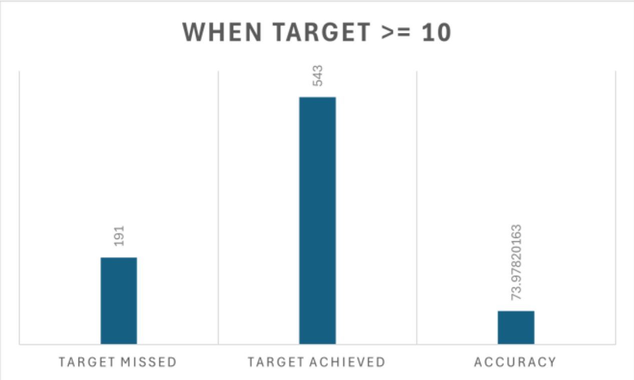


*Figure 8: Target greater than ten with accuracy*

## IV. CONCLUSION and Future Scope

The accuracy of the Probabilistic Profitable Model (PPM) hypothesis relies on assessing the probabilities associated with event occurrences. Probability, in this context, serves as a metric to gauge the likelihood or chance of a specific event transpiring. Our strategic approach focuses exclusively on current events and their potential outcomes. To enhance the probability of success, we may integrate volume profile, volume weighted average price (VWAP), and machine learning algorithms into the PPM hypothesis. Notably, the Long Short-Term Memory (LSTM) machine learning algorithm is instrumental in leveraging past events to predict historical occurrences. It is pertinent to note that our current analysis does not factor in the wick and body size of candlesticks.

The LSTM algorithm, however, possesses the capability to effectively incorporate the wick and body size, enhancing its predictive capacity for upcoming Open, High, Low, and Close (OHLC) candlestick patterns [20]. The judicious integration of machine learning techniques with the Probabilistic Profitable Model (PPM) constitutes a strategic direction that we plan to pursue in the future.

## REFERENCES

[1]. SEBI, "Study - Analysis of Profit and Loss of Individual Traders dealing in Equity F&O Segment," 1 2024. [Online].
[2]. K. Adam, A. Marcet and J. P. Nicolini, "Stock market volatility and learning," The Journal of finance, vol. 71, no. 1, pp. 33-82, 2016.
[3]. C.-A. S. S. Von Holstein, "Probabilistic forecasting: An experiment related to the stock market," Organizational Behavior and Human Performance, vol. 8, no. 1, pp. 139-158, 1972.
[4]. X. Li, P. Wu and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," Information Processing & Management, vol. 57, no. 5, p. 102212, 2020.
[5]. K. Nti, A. F. Adekoya and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artificial Intelligence Review, vol. 53, no. 4, pp. 3007-3057, 2020.
[6]. J. Nie, X. Qiao and S. Yan, "COVID-19 Effects on Intraday Stock Market Behavior," World Scientific, 2022, pp. 229-252.
[7]. S.-L. Chung, C.-C. Chang and P.-F. Hsieh, "Price Manipulation on Expiration: Evidence of Taiex Futures and Options," Available at SSRN 4670330, 2023.
[8]. K.-Y. Kwon and R. J. Kish, "Technical trading strategies and return predictability: NYSE," Applied Financial Economics, vol. 12, no. 9, pp. 639-653, 2002.
[9]. S. Fong, J. Tai and Y. W. Si, "Trend Following Algorithms for Technical Trading in Stock Market.," Journal of Emerging Technologies in Web Intelligence, vol. 3, no. 2, 2011.
[10]. D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniques," Computer Science Review, vol. 34, p. 100190, 2019.
[11]. W. Hu, Y.-W. Si, S. Fong and R. Y. K. Lau, "A formal approach to candlestick pattern classification in financial time series," Applied Soft Computing, vol. 84, p. 105700, 2019.
[12]. T.-T. Ho and Y. Huang, "Stock price movement prediction using sentiment analysis and CandleStick chart representation," Sensors, vol. 21, no. 23, p. 7957, 2021.
[13]. J. M.-T. Wu, Z. Li, N. Herencsar, B. Vo and J. C.-W. Lin, "A graph-based CNN-LSTM stock price prediction algorithm with leading indicators," Multimedia Systems, vol. 29, no. 3, pp. 1751-1770, 2023.
[14]. Chittorgarh, "Top 20 Share Brokers in India 2024 Chittorgarh," 1 2024. [Online]. Available: https://www.chittorgarh.com/report/top_20_share_brokers_in_india_by_clients_at_nse/1/.
[15]. B. Singh, R. Martyr, T. Medland, J. Astin, G. Hunter and J.-C. Nebel, "Cloud based evaluation of databases for stock market data," Journal of Cloud Computing, vol. 11, no. 1, p. 53, 2022.
[16]. H. K. Dhalla, "A Performance Analysis of Native JSON Parsers in Java, Python, MS. NET Core, JavaScript, and PHP," 2020.
[17]. D. Chakraborty, S. Ghosh and A. Ghosh, "Hybrid Autoencoder-Multi-Task Lstm Model for Indian Stock Market Ohlc Prediction and Profitable Recommendations," Available at SSRN 4565910.
[18]. J. Gopalakrishnan, "Market Profile Basics," TECHNICAL ANALYSIS OF STOCKS AND COMMODITIES-MAGAZINE EDITION-, vol. 17, pp. 17-22, 1999.
[19]. Groww, "Lot Size in Options Trading," 1 2024. [Online]. Available: https://groww.in/blog/lot-size-in-options-trading.
[20]. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," Procedia Computer Science, vol. 170, pp. 1168-1173, 2020.

 

**AUTHORS PROFILE**

Mr. Vinod Kumar Joshi completed his Master of Technology (MTech) degree at Birla Institute of Technology and Science, Pilani (BITS Pilani). Presently, he serves as the Deputy Superintendent of Police at the National Crime Records Bureau (NCRB) under the Ministry of Home Affairs (MHA) in New Delhi. Additionally, he holds a technical resident position on deputation at the Office of Digital Transformation, Indian Institute of Technology, Jodhpur, Rajasthan. He is a research scholar at Sangam University, Bhilwara, Rajasthan. With extensive experience spanning over a decade, Mr. Joshi has demonstrated expertise in handling various cybercrimes and a profound understanding of criminal psychology. His professional focus revolves around data analysis and machine learning.