International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Predicting Intraday Trend Reversals in Index Derivatives Using Supervised Machine Learning Algorithms

Payas Deshpande, Sridhar Subramanian, Shivali Amit Wagle, Preksha Pareek

Submitted: 24/12/2023 **Revised**: 30/01/2024 **Accepted**: 06/02/2024

Abstract: This research paper delves into the realm of financial market forecasting, specifically focusing on predicting intraday trend reversals in index derivatives using supervised machine learning algorithms. The study encompasses a comprehensive examination of various machine learning techniques, including Support Vector Machines, Random Forests, XGBoost, and LSTM, to develop models capable of navigating the complexities inherent in the financial markets.The primary objective of the research is to enhance the predictive accuracy of stock market movements by incorporating a range of factors such as market conditions, liquidity, and external influences. This multifaceted approach aims to capture the dynamic and often unpredictable nature of financial markets, offering a more nuanced and effective prediction model.Through meticulous analysis and evaluation, the paper demonstrates the significant potential of machine learning technologies in the field of computational finance. It explores the strengths and limitations of each algorithm, providing an in-depth understanding of their applicability in real-world market scenarios.Furthermore, the research identifies key areas for future exploration, emphasizing the need for a more detailed examination of macroeconomic and sociopolitical factors, as well as the utilization of high-frequency data, particularly in emerging markets. These insights pave the way for ongoing advancements in the application of machine learning for financial market analysis.Overall, this paper makes a notable contribution to the field of computational finance, offering valuable perspectives and tools for academics and practitioners alike. It lays the groundwork for further research that aims to refine and expand the use of machine learning in stock market prediction, ultimately leading to more robust and versatile forecasting models.

Keywords: *Support Vector Machine (SVM), Random Forest (RF), XGBoost, Exploratory Data Analysis, LSTM, Stock market, S&P Index (SPX), Feature Scaling*

Introduction

The application of machine learning algorithms in finance has increased dramatically in recent years. One area of interest is the prediction of intraday trend reversals in index derivatives.Index derivatives are financial contracts that draw their value from the performance of an underlying index, such as the S&P 500 or the Nikkei 225. Intraday trend reversals are abrupt shifts in price movement direction throughout a single trading day.Due to the complexity and volatility of financial markets, predicting intraday trend reversals in index derivatives can be difficult. The potential advantages, however, are substantial because precise predictions can aid traders in making wise selections and enhancing their earnings.Machine learning models have long been used in the financial sector, particularly in the stock market. With varying levels of accuracy, multiple methods such as Linear Regression, Support Vector Machines (SVM), and Auto-Regressive Integrated Moving Average (ARIMA), among others have been employed (Md et al., 2023). The market has three main regimes UP(Bullish), Down(Bearish), and Sideways. (Ma et al., 2023). We intend to start with models such as XGBoost, Random Forest, or K-means to classify the market regime and then use a time series model such as LSTM and GCN to predict the stock price. (Dudukcu et al., 2023; Ren et al., 2023; Zhao et al., 2023)

There have been countless attempts to predict stock prices but a workable approach that gives good accuracy and is adequate for all types of people has not been accomplished. When a generalized user base is taken the number of available features to be considered explode and tank accuracy. (Deng et al., 2023; Toochaei and Moeini, 2023)In this study, we investigate using supervised machine learning techniques combined with Temporal CNN or RNN techniques to create a process flow that can reduce the number of features considered and is also usable for a more generalized user base. The Use of the Adam Optimizer may be required to increase performance score. (Md et al., 2023)Many external factors affect the stock market, such as Liquidity, Market condition, and exogenous events such as Climate. The Consideration of these factors is crucial for the accuracy of the model. Our method for this involves using relational models to include the influence of these features on the market. (Ren et al., 2023)We start by

Symbiosis Institute of Technology, (Pune campus), symbiosis International Deemed University, Pune 411215, India
payas.deshpande.mtech2023@sitpune.edu.in,sridhar.subramanian.mtech2023@sitpune.edu.in, Shivali.wagle@sitpune.edu.in, Preksha.pareek@sitpune.edu.in
Corresponding author: Shivali Amit Wagle (shivali.wagle@sitpune.edu.in)

reviewing the related papers on this topic and finding the methodologies that they employ and their Gaps. Stock market prediction appears to be a simple task, but a market is a chaotic system in which the price of a stock is influenced not only by external events but also by the price of other stocks.

Related work

The field of stock market forecasting has witnessed substantial progress through the integration of diverse computational strategies and analytic models. In an innovative approach to discern the interconnections among stocks, Zhao and colleagues (2023) employed a novel Time Series Relationship model augmented by a trio of sophisticated algorithms: K-Means, Long Short-Term Memory (LSTM), and Graph Convolutional Networks (GCN). This methodology demonstrated superior performance compared to extant models, as evidenced by its application to the Shanghai and Shenzhen 300 indices. In another study, Ren and associates (2023) delved into the dynamic intraday relationships using high-resolution, one-minute data, applying a combination of TOP and Vector Error Correction Model with Integrated Structure (VECM-IS). Their findings indicate the potential for refining pair trading strategies through additional parameter optimization.

Further adding to the body of knowledge, Deng et al. (2023) implemented an Explainable Xtreme Gradient Boosting algorithm for binary classification tasks using daily price data from the Shanghai Index, which outperformed traditional Ordinary Least Squares (OLS) forecasting. However, they noted that the model's reliance on constrained sentiment features may limit its applicability to a broader investor base.

A study by (Toochaei et al. (2023)) embraced a multiclass classification framework to ascertain the favorability of stocks listed on the Tehran Stock Exchange. This research utilized a robust ensemble of machine-learning algorithms, including Random Forest, LightGBM, and XGBoost, among others. The ensemble's predictive prowess was commendable, although the large feature set warrants further refinement to simplify the model without sacrificing accuracy.

(Rajendiran and collaborators (2023)) conducted an inquiry into the complex relationship between stock prices and market sentiment by leveraging a Q-learning framework. Their model exhibited a comparatively lower error rate than its predecessors, albeit the reduction in prediction time remains an area for future enhancement. Gupta and co-researchers (2023) proposed a non-parametric investment portfolio construction for the NSE 100, utilizing the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Despite its technical focus, the methodology did not fully incorporate fundamental financial parameters, suggesting an avenue for comprehensive analysis.

In a similar vein,(Segnon et al. (2023)) applied a Markov-switching GARCH model with mixed-data-sampling (AR-MSGARCH-MIDAS) to monthly observations of the Dow Jones Industrial Average, incorporating macroeconomic indicators such as the US consumer price index and interest rates. Concurrently, Dudukcu et al. (2023) addressed the challenge of stock market prediction through the lens of chaotic time series data, employing Temporal Convolutional Neural Networks (CNN) and LSTM to achieve remarkable predictive accuracy.

Continuing this trajectory,(A.Q. Kapoor and associates (2023)) documented a commendable 95% accuracy rate in stock price prediction on data sourced from Yahoo Finance, utilizing a meticulously structured Multi-Layer Sequential LSTM model. On a different front, Zhang et al. (2019) advocated for the adoption of a Finite Automata optimizer in conjunction with Support Vector Regression to enhance forecasting precision.

(Kuber et al.) presented a dualistic LSTM-based predictive model that accounted for closing prices and technical indicators, while Han and colleagues (2023) set a precedent with their application of a comprehensive algorithmic suite that included XGBoost, SVM, and Bi-LSTM. Their research extended the frontier of predictive accuracy. Adding to the discourse, Argotty-Erazo et al. (2023) prioritized the identification and validation of discriminant features, while Singh et al. (2022) refined time-series forecasting models for the NASDAQ and NSE.

Chandrika and team (2023) offered an extensive review of machine learning approaches applied to the Bombay Stock Exchange, albeit with an acknowledgment of the need for a more granular analysis of macroeconomic and sociopolitical influences. In an exploratory study, Hani'ah and associates (2023) synthesized stock price data with Google Trends insights to forecast market trends with noteworthy precision. Arismendi-Zambrano and co-researchers (n.d.) evaluated various machine learning methodologies in the context of the Brazilian stock market, recognizing a research gap in the utilization of high-frequency data for emerging markets.

Mandal et al. (2023) embarked on a quest to refine data analysis and forecasting accuracy, employing SVR and Multiple Linear Regression on a diverse array of Indian stock data. In the realm of academia, Sonkavde et al. (2023) critically assessed a spectrum of computational models, seeking to elevate the accuracy of stock predictions. Lastly, Uma and Srinath Naidu (2021) scrutinized trend reversals through a multifaceted analytical approach, though their iterative methodology requires further elucidation.

This comprehensive review encapsulates a series of empirical studies, delineating the continuous evolution of analytic methodologies in the ambit of financial market prediction. Each contribution underscores not only the strides made in computational finance but also the persistent quest for models that marry precision with practical utility. The related work on stock price classification is shown in Table 1.

Table 1: Related work on stock price classification

Sno	Methodology	Algorithm used	Dataset used	Gaps	Remarks	Reference
1	Relationship between various stocks is calculated using a Time Series Relationship model.	K-Means, LSTM, and GCN.	Shanghai and Shenzhen 300 indices (Jan,19 to Sept,20)	Did not use external data to supplement analysis	The model performed better than the already-used model	*Zhao, C et.al (2023).*
2	investigates fluctuations in intraday relationships using 1-minute data.	TOP & VECM-IS	SSE 50 from the Wind database	Inability to consider influencing factors such as liquidity market conditions and exogenous events.	The pair trading strategy can be expanded by parameter tuning and further exploring the relationships.	*Ren, F. et.al (2023).*
3	The model utilized Sentiment characteristics for a binary classification model.	Explainabl e Xtreme Gradient Boosting	Daily prices of the Shanghai Index	The sentiment features are too constrained making the model work well for a specialized group of investors	The forecasting ability of XGBoost- S Superior to OLS and other algorithms	*Deng, S. et.al (2023).*
4	Estimates the Favorability of stock using Multiclass classification.	Random Forest, LightGBM , XGBoost, Extra-Trees, AdaBoost, CatBoost	Stock market data for the Tehran Stock Exchange	The list of features selected is too huge, increasing the complexity of the model	The use of Ensembles is an excellent idea and provides far better results. XGBoos and ADAboost provide the Best and Worst prediction	*Toochaei, M. R. et.al (2023).*
5	investigates the nonlinear link	Q-learning	Stock Market Sentiment Data	The Q-learning	error rate of Q-learning is	*Rajendiran, P. et.al (2023).*

	between stock price and sentiment.			model fails to reduce prediction time	comparatively lesser than the improved stock market classification model	
6	The project aimed to create a non-parametric investment portfolio for the NSE 100.	TOPSIS (Technique for order of performance by similarity to ideal solution)	NSE 100	Inability to consider information priors. It is not a multi-stage framework. Only concentrated on the technical parameters of the stocks and exclude the fundamental parameters like Return on Equity, Return on Capital Employed,		*Gupta, S. et.al (2023).*
7		Markov-switching GARCH mixed-data-sampling (AR-MSGARCH-MIDAS)	Dow Jones Industrial Average (DJIA) index values are reported, as monthly observations of the US consumer price index (CPI), interest rate, and adverse events.			*Segnon, M. et.al (2023).*
8	The stock market prediction is based on chaotic time series data.	Temporal CNN, LSTM	Lorenz, Rössler, and a Lorenz-like chaotic equation sets, and twenty-one electrocardiogram (ECG) recordings of patients with arrhythmias.		The proposed model achieved the best prediction performance with an average root-mean-square error (RMSE) value of 0.0022 for the chaotic dataset and 0.0082 for the ECG arrhythmia dataset	Dudukcu, H. V. et.al (2023).
9	Multi-Layer Sequential LSTM with	Multi-layered Sequential	Yahoo Finance		The Sequential LSTM consists of 3 Vanilla and 1 Dense layer.	A. Q., Kapooret.al (2023).

	Adam Optimized is employed on normalized time series data.	LSTM			The model achieves 95% accuracy.	
10	As an optimizer, the algorithm employs a Finite Automata known as an MFA.	Support Vector Regression			The model performs better than most of the models.	Zhang, J. et.al (2019).

A salient and innovative aspect of this research lies in its two-tiered approach to predicting intraday trend reversals in index derivatives. Unlike conventional methodologies that solely focus on either classification of market trends or regression analysis for price prediction, this study ingeniously amalgamates both techniques into a cohesive analytical framework [28-33]. Initially, the research implements a classification model to categorize market trends as bullish, bearish, or sideways, thereby establishing a foundational understanding of market direction. Subsequently, it ingeniously incorporates the predicted trend category as an additional feature into a regression model tasked with forecasting the specific opening or closing prices of the index derivatives. This hybrid approach not only leverages the strengths of both classification and regression techniques but also provides a more nuanced and comprehensive analysis of market dynamics. The integration of trend classification into the regression model represents a novel contribution to the field, potentially enhancing the accuracy and robustness of predictive models in financial market analysis. It addresses a critical gap in traditional forecasting methodologies, offering a more holistic and contextually informed perspective that is pivotal for effective decision-making in financial markets' complex and volatile domain.

Proposed Methodology

Our proposed workflow aims to leverage long-range relationships between different stocks over time where changes in one or more of the stocks can influence the trend of another stock. To achieve a birdseyeview of these relationships we intend to implement a 2-step prediction process. The workflow of the model is shown in Figure 1. First, the data will be classified by trends. This will be a multiclass classification. The classes are as follows:

a. Up Trend or Bullish
b. Down Trend or Bearish
c. Unchanged Trend of Sideways.

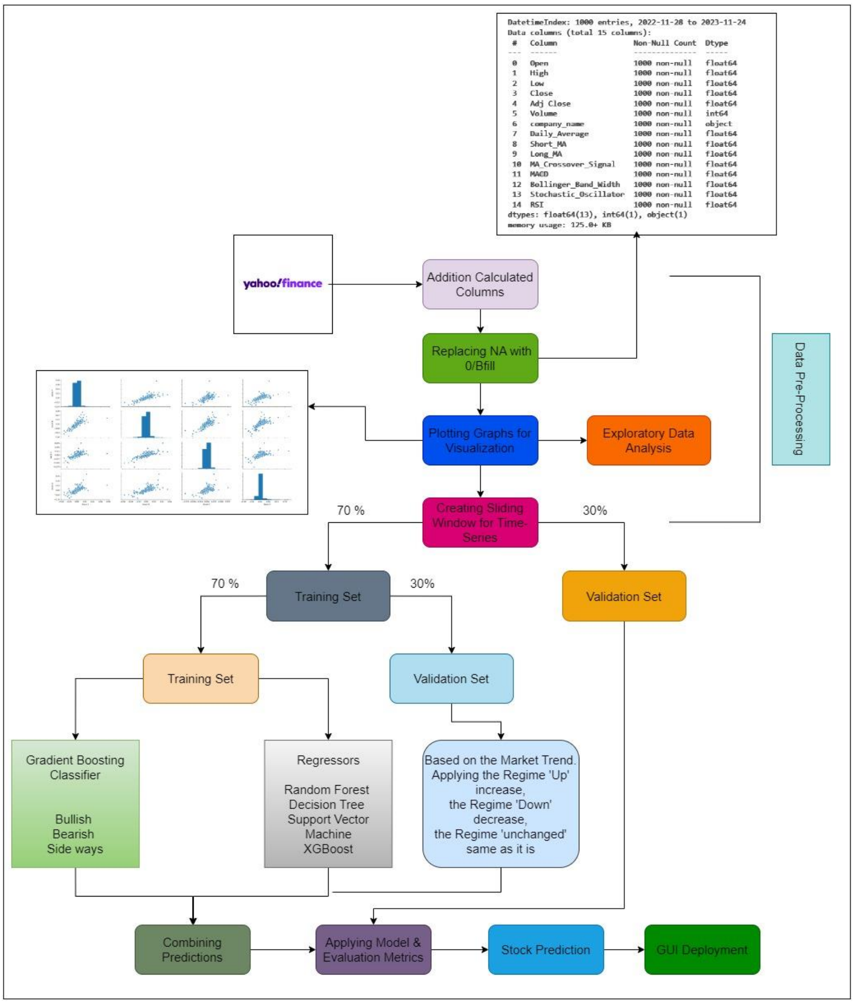


Fig 1: Workflow of Model

These trends will provide an additional dimension to the data aiding the second step of the process where the Opening or Closing price is predicted.

Dataset

In this paper, we have implemented a dynamic data model that allows for flexibility in predicting stock market trends. The model is adaptable to various stocks, time spans, and prediction targets, such as the opening or closing prices of stocks. Real-time functionality is a key feature of this model, enabled by the use of live data acquired through the Yahoo Finance API. The data, comprising opening, closing, and volume information of stocks, is stored in multi-level pandasdata frame. This data frame is structured such that Level 0 columns represent indices like Open, Close, High, Low, etc., while Level 1 columns denote the stock names, for instance, AAPL, MSFT, INFY, and HDB.

To enhance the visibility of market trends, the data is processed to calculate eight additional features. These include the Daily Average, which indicates day-to-day percentage changes in the closing price, and two forms of Moving Averages: a 50-day Short Moving Average and a 200-day Long Moving Average. The Moving Average Crossover Signal, which calculates the disparity between short-term and long-term moving averages, provides insights into potential bullish or bearish regimes. Additionally, we incorporate the Moving Average Convergence Divergence, a widely-used momentum indicator, and the Bollinger Band Width, which signifies the standard deviation of the price and can hint at potential market turning points. The Stochastic Oscillator, employed to identify overbought and oversold conditions, along with the Relative Strength Index (RSI), used to assess the velocity and magnitude of price movements, further enrich our analysis.

The inclusion of these calculated features expands the multi-level data frame, resulting in 15 Level 0 columns and 4 Level 1 columns for each stock. This comprehensive and dynamic approach to data representation and feature calculation underpins the robustness and adaptability of our predictive model, providing a nuanced view of market trends and movements. Distribution of the data used for the analysis is shown in Figure 2.

				Adj Close				Close				High			
	AAPL	HDB	INFY	MSFT	AAPL	HDB	INFY	MSFT	AAPL	HDB	INFY	MSFT	AAPL	HDB	INFY
count	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000	251.000000
mean	153.726056	61.053651	19.402625	265.729238	154.835060	61.946016	20.202231	268.917091	156.907809	62.611155	20.391275	272.496095	152.691315	61.203426	20.004263
std	12.790990	5.128210	2.211153	24.843657	13.056081	5.233816	2.414609	25.761774	12.937389	5.222035	2.446577	25.973312	13.108058	5.232885	2.384173
min	125.504539	49.605835	16.003296	212.199982	126.040001	50.700001	16.610001	214.250000	129.949997	51.790001	16.709999	220.410004	125.870003	50.610001	16.389999
25%	143.846024	57.046482	17.765683	244.753220	144.645004	57.645000	18.430000	247.180000	146.709999	58.555000	18.554999	249.430000	142.324997	57.115000	18.240000
50%	152.968796	60.802078	18.778099	262.797974	154.089996	61.849998	19.430000	265.899994	155.830002	62.509998	19.549999	268.100006	151.940002	61.049999	19.260000
75%	164.569420	65.837799	21.386150	285.799683	165.915001	66.949997	22.445000	289.744995	167.989998	67.735001	22.725000	292.954987	164.044998	65.970001	22.160000
max	180.190979	70.181992	24.964008	329.394867	182.009995	71.730003	26.200001	334.750000	182.940002	72.199997	26.389999	338.000000	179.119995	71.010002	25.580000

Fig 2: Snapshot of the Distribution of the Data

Data Preprocessing-

The distribution of NA values in the dataset can be seen in the Figure 3.

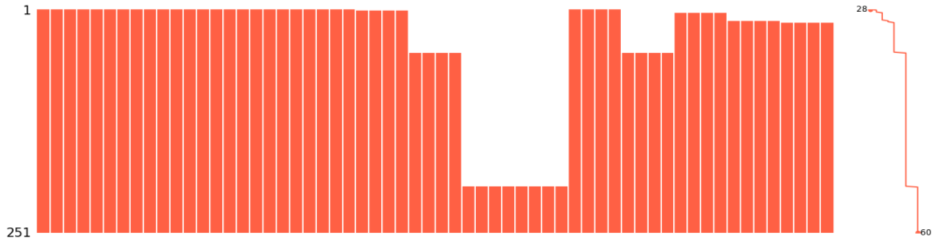


Fig 3: Distribution of Columns Before Replacing NA values

As it is displayed in the above visualization there are a lot of NA values in the data. However, removing these values would shield us from seeing a huge market trend. These NA values signify the Unchanged Market Regime. Most of the NA values are in the calculated columns. These need to be replaced with 0 to obtain a more complete picture of the market. The distribution of data after replacement in Figure 4.

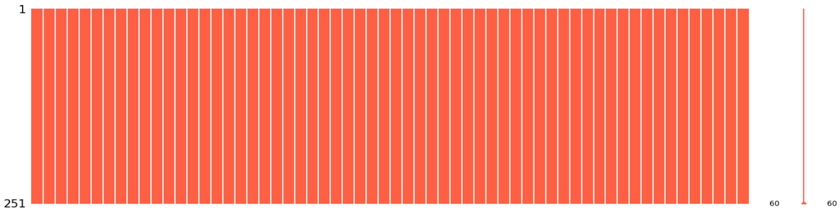


Fig 4: Distribution of Columns After Replacing NA values

The Dataframe acquired from Yahoo Finance is in the form of a multi-level dataframe. To make it usable for machine learning it needs to be flattened. This is done by merging the 2 levels of the data frameand concatenating the column names as per Figure 5.

	Adj Close_AAPL	Adj Close_HDB	Adj Close_INFY	Adj Close_MSFT	Close_AAPL	Close_HDB	Close_INFY	Close_MSFT	High_AAPL	High_HDB	...	Relative Strength Index_MSFT	Relative Strength Index_INFY	Relative Strength Index_HDB
Date														
2010-01-04	6.478999	12.164503	5.129956	23.474922	7.643214	13.346	7.09500	30.950001	7.660714	13.348	...	0.0	0.0	0.0
2010-01-05	6.490199	12.216457	5.149840	23.482498	7.656429	13.403	7.12250	30.959999	7.699643	13.403	...	0.0	0.0	0.0
2010-01-06	6.386963	12.271147	5.071210	23.338392	7.534643	13.463	7.01375	30.770000	7.686786	13.510	...	0.0	0.0	0.0
2010-01-07	6.375157	12.467111	4.916660	23.095678	7.520714	13.678	6.80000	30.450001	7.571429	13.689	...	0.0	0.0	0.0
2010-01-08	6.417542	12.504480	4.929314	23.254959	7.570714	13.719	6.81750	30.660000	7.571429	13.750	...	0.0	0.0	0.0

Fig 5: Flattened DataFrame

The values in the data need to be standardized. This is done using the Standard Scaler. It scales the data to fit a Standard Normal Distribution, with a mean of 0 and a standard deviation of 1. Following the standardization, the data is split into training and test sets for model training.

Data Visualization

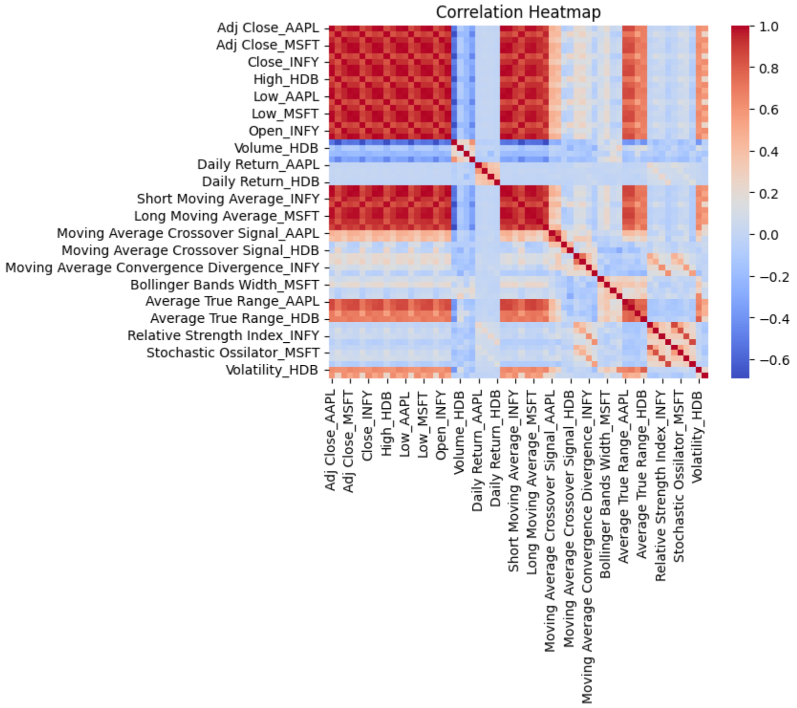


Fig 6: Correlation Heatmap

This heatmap shows the correlation between the features of different stocks in Figure 6. In the above visualization,the features of 4 stocks are shown. From the above chart,it can be seen that the short-moving Average and long-moving average of the four stocks are closely related showcasing the long-term relationships.

Data Labeling

In our proposed approach, supervised machine learning is used for classification, the crucial step of labeling data for model training is accomplished through K-Means Clustering. We divided the data into three distinct clusters, with the trends each cluster represents being inferred from the values of its centroid. Key indicators like the Short Moving Average (SMA), Long Moving Average (LMA), Moving Average Crossover, and Bollinger Bandwidth values were instrumental in signifying market trends. The labeling system was as follows: Label 0, representing a 'Sideways' trend, was assigned when the SMA and LMA were not significantly divergent, and the Moving Average Crossover hovered around zero. Label 1, indicative of a 'Bullish' trend, was used when the SMA was considerably higher than the LMA, and the Moving Average Crossover showed a large positive value, often referred to as the 'golden signal.' Lastly, Label 2, denoting a 'Bearish' trend, was applied when the SMA was lower than the LMA, coupled with a negative Moving Average Crossover. This methodical labeling process, leveraging the insights provided by these financial indicators, laid the foundation for the effective training and functioning of our classification models.

Model Selection

Classification

In our research, we implemented and evaluated ten different classification algorithms to analyze their efficacy in predicting market trends. The algorithms we explored were Logistic Regression, K-Nearest Neighbor, Linear SVM, SVM with Radial Basis Function (RBF), Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Gaussian Naïve Bayes, and Multi-Layer Perceptron. We applied these models to a dataset with a 70:30 split between training and testing data. The performance of each model was meticulously assessed using four key metrics: Accuracy, Precision, Recall, and F1 Score.

To enhance prediction accuracy, we employed an ensemble method that combined the four best-performing models: Gradient Boosting, Decision Tree, AdaBoost, and Linear SVM. This ensemble was executed as a stacking classifier, where Logistic Regression was utilized to integrate the decisions made by each model. Despite achieving high performance, this ensemble method could not surpass the effectiveness of the single best-performing model.

Among all the classifiers, the Gradient Boost Classifier stood out with the highest performance metrics. Consequently, we applied this model to our data for predicting market trends. The predicted trends were then appended to our data frame, creating an enriched dataset that further facilitates accurate and insightful market trend predictions. This approach underscores the potential of advanced machine learning techniques in enhancing the precision of financial market analysis.

Regression

In our study, we employed a suite of eleven different regression algorithms to predict the opening and closing prices of index derivatives. These included Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Bayesian Ridge, Support Vector Regression (SVR), K-Nearest Neighbors (KNN) Regression, Decision Tree, Random Forest Regressor, Gradient Boosting Regressor, and AdaBoost. These regression techniques were applied after the use of a Gradient Boosting Classifier. After a comprehensive comparison of the performance of these models, it was determined that Linear Regression and Bayesian Ridge Regression emerged as the top performers. However, between these two, Bayesian Ridge Regression demonstrated a lower Mean Squares Error, thereby establishing it as the superior model in our analysis. This selection was based on the model's effectiveness in accurately predicting the market's opening and closing values, a critical factor in the realm of financial market forecasting.

Model Evaluation

This paper implements both Classification and Regression techniques so the evaluation metrics for each differ.

For Classification, the metrics used are shown in Equation 1-4.

1. Accuracy: -

Accuracy is calculated as the ratio of correctly predicted instances to the total number of instances in the dataset. Mathematically, it is represented as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \quad (1)$$

2. Precision: -

It is a metric that evaluates the model's accuracy in predicting positive instances. It is particularly crucial in scenarios where the cost of false positives (incorrectly predicting a positive instance) is high.

Precision is defined as the ratio of true positive predictions to the total positive predictions made by the model. Mathematically, it is expressed as:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP)+False\ Positives\ (FP)} \quad (2)$$

3. Recall: -

Recall is a critical metric used to evaluate the performance of a classification algorithm. Recall, also known as sensitivity or the true positive rate, measures the proportion of actual positive cases the model correctly identifies.

Formally, recall is defined as:

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP)+False\ Negatives\ (FN)} \quad (3)$$

4. F1-Score: -

The F1 score is a crucial metric used in evaluating the performance of models, particularly in classification problems. It represents a balance between precision and recall, two critical aspects of a model's accuracy.

The F1 score harmonizes these two measures by calculating their harmonic mean. The formula is:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision+Recall} \quad (4)$$

For Regression, the metrics used are: -

1. Mean Squared Error: -

Mean Squared Error (MSE) is a widely used metric for evaluating the performance of regression models. The MSE quantify the difference between the predicted valuesby a model andthe actual value:$MSE = \Sigma_i^n (Y_i - Y_{i(mean)})^2$) (5)

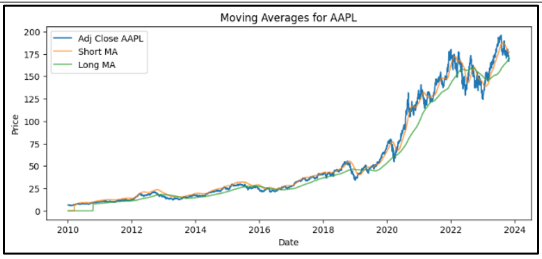
2. R Squared Value: -

It is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. It is a key indicator of the model's accuracy and goodness of fit. In simpler terms, it measures how well the model's predictions approximate the real data points.

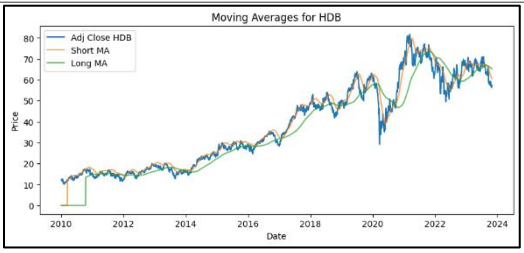
Results:

In this section we present the results for predicting performance of the model for stock market analysis. The short and moving average are shown in Figure 7 wherein Figure 7(a) presents results for Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Apple stock between 2010 and 2023. Figure 7(b) presents results for Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for HDFC stock between 2010 and 2023. Figure 7(c) presents results for Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Infosys stock between 2010 and

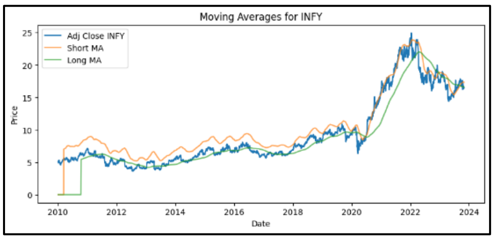
2023. Figure 7(d) presents results for Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Microsoft stock between 2010 and 2023.



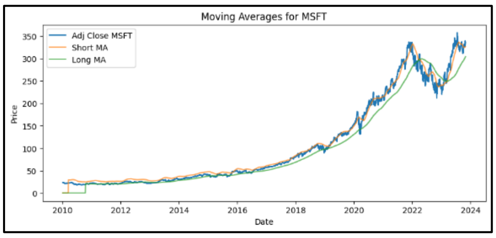
(a) Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Apple stock between 2010 and 2023.



(b) Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for HDFC stock between 2010 and 2023.



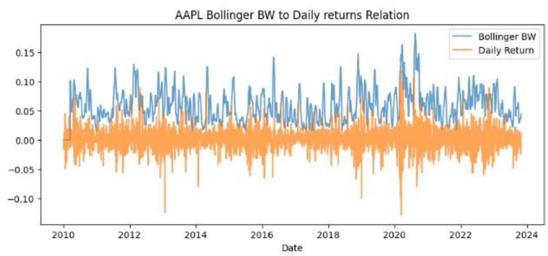
(c) Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Infosys stock between 2010 and 2023.



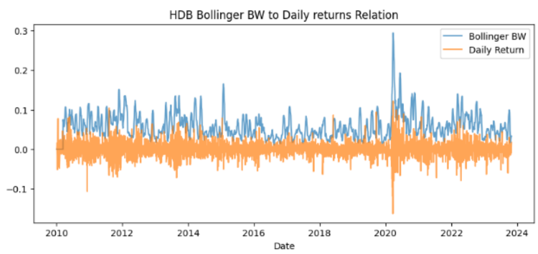
(d) Short (50 Day) and Long (200 Day) Moving Average and Adjustive Close for Microsoft stock between 2010 and 2023

Fig 7: Short and Long Moving Average

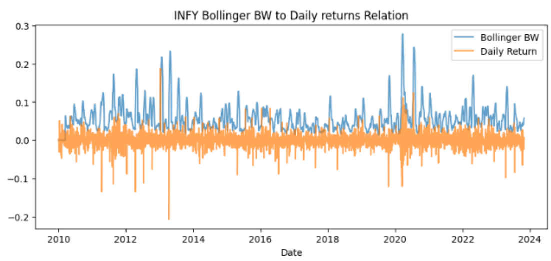
The 50-day Moving Average provides a more accurate representation of the stock, and the 200-day moving average provides a more generalized picture of the stock's performance on a long-term basis.



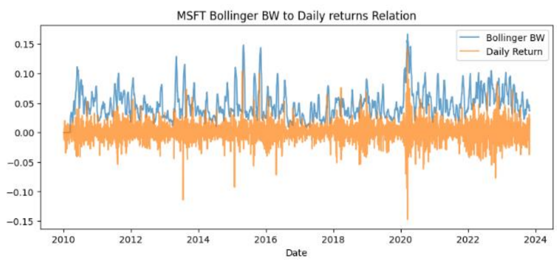
(a) Establishing a Relationship between the Bollinger Bandwidth and Daily returns for Apple Stock



(b) Establishing a Relationship between the Bollinger Bandwidth and Daily returns for HDFC Stock



(c) Establishing a Relationship between the Bollinger Bandwidth and Daily returns for Infosys Stock



(d) Establishing a Relationship between the Bollinger Bandwidth and Daily returns for Microsoft Stock

Fig 8: Bollinger Bandwidth Trend

In Figure 8, Bollinger Bandwidth Trend is presented for Apple, HDFC, Infosys, and Microsoft respectively.Bollinger Bandwidth signifies the volatility of the stock. This diagram establishes a relationship between the Bollinger Bandwidth and Daily Average Returns.

The results of the two steps for Classification and Regression, are shown in Table 2 and 3 respectively. Gradient Boosting Classifier, Decision Tree, AdaBoost, Stacking Classifier, SVM Linear, K-Nearest Neighbor, SVM-RBF, Random Forest, MLP Classifier, Logistic Regression, Gaussian Naïve Bayes model results are used to analyze the classification task. Gradient Boost classifier outperforms the other classification model for prediction of stock prices.

Table 2: Performance parameter with the classification models

Classifier	Accuracy	Precision	Recall	F1 Score
Gradient Boosting Classifier	99.617	99.6	99.617	99.61
Decision Tree	99.524	99.520	99.57	99.52
AdaBoost	99.5	99.528	99.521	99.53
Stacking Classifier	99.5	99.50	99.52	99.520
SVM Linear	99.42	99.425	99.4	99.42
K-Nearest Neighbor	99.32	99.34	99.33	99.32
SVM-RBF	99.32	99.33	99.32	99.32
Random Forest	99.32	99.32	99.329	99.327
MLP Classifier	99.329	99.327	99.325	99.328
Logistic Regression	99.232	99.233	99.234	99.231
Gaussian Naïve Bayes	98.754	98.768	98.754	98.748

Bayesian Ridge, Linear Regression, Ridge Regression, Random Forest, Gradient Boosting Regressor, Decision Tree, Lasso Regression, KNN Regression, Adaboost, Elastic Net, SVR regression models are used for the prediction of stock prices. Bayesian Ridge model performs best amongst the other regression model used.

Table 3: Performance parameter with the regression models

Regressor	MSE	R^2
Bayesian Ridge	0.1752	99.993
Linear Regression	0.1762	99.98
Ridge Regression	0.3266	99.98
Random Forest	0.4419	99.984
Gradient Boosting Regressor	0.5221	99.981
Decision Tree	0.8797	99.96
Lasso Regression	2.2275	99.92
KNN Regression	5.8808	99.79
Adaboost	6.5288	99.771
Elastic Net	20.683	99.275
SVR	114.97	95.97

Given the impressive performance of the Bayesian Ridge Regressor. The model seems well suited for capturing the complex relationshipsin each regime.The Classification performed serves as a feature engineering, adding a layer of information that helps the model understand the context of the data better. In the classification stage, the best-performing model is the Gradient Boosting Classifier.The Gradient Boosting Classifier can model the complexities in the data. By Classifying the Market regimes, the model sets a strong foundation for the subsequent regression analysis. It also allows the Bayesian Ridge Regressor to model the complex relations in the data better.This two-step approach allows for a more nuanced understanding of the data.

Conclusion

This paper embarked on a comprehensive exploration of the application of supervised machine learning techniques in predicting intraday trend reversals in index derivatives. The research synthesized a wide array of algorithms, such as Support Vector Machines, Random Forests, XGBoost, LSTM, and others, to forge a refined approach that balances feature complexity and user accessibility.The study notably addressed the challenges inherent in financial market forecasting, acknowledging the multifaceted nature of stock movements influenced by a variety of factors including market conditions, liquidity, and external events. By integrating these elements into its models, the research provided a nuanced view of the market dynamics. This holistic approach was essential in enhancing the accuracy of the predictive models.

The paper's methodical analysis and rigorous evaluation of diverse machine learning models demonstrated the potential of these technologies in financial market analysis. The results showed promising avenues for improving the accuracy and efficiency of stock market predictions, offering valuable insights for both academics and practitioners in the field.However, the paper also highlighted the limitations and potential areas for further research. It identified the need for a more granular analysis of macroeconomic and sociopolitical influences, as well as the exploration of high-frequency data in emerging markets. These aspects underscore the ever-evolving nature of the field and the continuous need for innovation and adaptation in machine learning applications for financial market forecasting.In conclusion, this paper contributes significantly to the existing body of knowledge in computational finance. It not only underscores the advancements made in the use of machine learning for stock market prediction but also paves the way for future research to further refine these models, making them more robust, accurate, and applicable to a broader range of scenarios in the financial domain.

Future Scope

The research presented in this paper offers a promising foundation for several future research avenues in the dynamic fields of financial markets and machine learning. A critical area for expansion involves integrating macroeconomic and sociopolitical data into predictive models, enhancing their accuracy by considering global economic events and policy changes. Exploring high-frequency trading data, particularly in emerging markets, represents another significant direction, requiring the development of algorithms capable of real-time processing and learning. Innovations in deep learning and neural networks, such as reinforcement learning and convolutional neural networks, present exciting prospects for advanced predictive capabilities.The implementation of cutting-edge AI methods, such as Graph Neural Networks (GNN), could lead to an even better framework. Providing a way to model relationships between various stocks and their effect on each other.

Future work could also focus on creating algorithms that adapt in real-time to changing market conditions, maintaining relevance and accuracy over time. Cross-market analysis and the development of global predictive models could provide insights into the interconnected nature of financial markets. Moreover, as machine learning applications in finance expand, addressing regulatory compliance and ethical considerations becomes paramount. Customizing models for individual and institutional investors, with user-friendly interfaces and decision-support systems, can cater to various expertise levels and investment strategies. Lastly, applying machine learning to enhance risk management and detect market anomalies could significantly impact the field. This research paves the way for innovative exploration in computational finance, with the evolving landscape of machine learning and financial market dynamics promising rich opportunities for future study and application.

Reference

[1] Deng, S., Huang, X., Zhu, Y., Su, Z., Fu, Z., Shimada, T., 2023. Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments. North American Journal of Economics and Finance 64. https://doi.org/10.1016/j.najef.2022.101848

[2] Dudukcu, H.V., Taskiran, M., Taskiran, Z.G.C., Yildirim, T., 2023. Temporal Convolutional Networks with RNN approach for chaotic time series prediction. Appl Soft Comput 133. https://doi.org/10.1016/j.asoc.2022.109945

[3] Ma, Y., Mao, R., Lin, Q., Wu, P., Cambria, E., 2023. Multi-source aggregated classification for stock price movement prediction. Information Fusion 91. https://doi.org/10.1016/j.inffus.2022.10.025

[4] Md, A.Q., Kapoor, S., Chris, C.J., Sivaraman, A.K., Tee, K.F., Sabireen, H., Janakiraman, N., 2023. A novel optimization approach for stock price forecasting using multi-layered sequential LSTM. Appl Soft Comput 134. https://doi.org/10.1016/j.asoc.2022.109830

[5] Ren, F., Cai, M.L., Li, S.P., Xiong, X., Chen, Z.H.J., 2023. A Multi-market Comparison of the Intraday Lead–Lag Relations Among Stock Index-Based Spot, Futures and Options. Comput Econ 62, 1–28. https://doi.org/10.1007/s10614-022-10268-0

[6] Toochaei, M.R., Moeini, F., 2023. Evaluating the performance of ensemble classifiers in stock returns prediction using effective features. Expert Syst Appl 213. https://doi.org/10.1016/j.eswa.2022.119186

[7] Zhao, C., Hu, P., Liu, X., Lan, X., Zhang, H., 2023. Stock Market Analysis Using Time Series Relational Models for Stock Price Prediction. Mathematics 11. https://doi.org/10.3390/math11051130

[8] Kuber, V., Yadav, D., & Yadav, A. K. (2022). Univariate and Multivariate LSTM model for Short-Term Stock Market Prediction. arXiv (Cornell University).https://doi.org/10.48550/arxiv.2205.06673.(Kuber et al., n.d.)

[9] Han, Y., Kim, J., & Enke, D. (2023). A machine learning trading system for the stock market based on N-period Min-Max labeling using XGBoost. Expert Systems With Applications, 211, 118581.https://doi.org/10.1016/j.eswa.2022.118581 (Han et al., 2023)

[10] Argotty-Erazo, M., Blázquez-Zaballos, A., Argoty-Eraso, C. A., Lorente-Leyva, L. L., Sánchez-Pozo, N. N., & Peluffo-Ordóñez, D. H. (2023). A Novel Linear-Model-Based Methodology for Predicting the Directional Movement of the Euro-Dollar Exchange Rate. IEEE Access, 11, 67249–67284.https://doi.org/10.1109/access.2023.3285082(Argotty-Erazo et al., 2023)

[11] Singh, T., Kalra, R., Mishra, S., Satakshi, & Kumar, M. (2022). An efficient real-time stock prediction exploiting incremental learning and deep learning. Evolving Systems.https://doi.org/10.1007/s12530-022-09481-x(Singh et al., 2022)

[12] G. Naga Chandrika, Sai, Ashish Kalleru, Vijitha Kambhampati, & Pragnika Kandaggatlta. (2023). Comparative Analysis of Machine Learning Algorithms to Forecast Indian Stock Market. ITM Web of Conferences, 56, 05009–05009.https://doi.org/10.1051/itmconf/20235605009 (Chandrika et al., 2023)

[13] MamluatulHani'ah, Moch Zawaruddin Abdullah, Wilda Imama Sabilla, Akbar, S. A., &Dikky Rahmad Shafara. (2023). Google Trends and Technical Indicator based Machine Learning for Stock Market Prediction. Matrik: JurnalManajemen, Teknik Informatika, Dan RekayasaKomputer, 22(2), 271–284.https://doi.org/10.30812/matrik.v22i2.2287 (Hani'ah et al., 2023)

[14] Juan Arismendi Zambrano, Alan De Genaro, & Henrique Leone Alexandre. (2023). Intraday Returns Forecasting Using Machine Learning: Evidence from the Brazilian Stock Market. Social Science Research Network.https://doi.org/10.2139/ssrn.4557985 (Arismendi-zambrano et al., n.d.)

[15] Mandal, U., Chakarborty, A., Mahato, P., & Das, G. (2023). LinVec: A Stacked Ensemble Machine Learning Architecture for Analysis and Forecasting of Time-Series Data. Indian Journal of Science and Technology, 16(8), 570–582.https://doi.org/10.17485/ijst/v16i8.2197 (Mandal et al., 2023)

[16] Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. International Journal of Financial Studies, 11(3), 94.https://doi.org/10.3390/ijfs11030094(Sonkavde et al., 2023)

[17] Uma, K. S., & Naidu, S. R. (2020). Prediction of Intraday Trend Reversal in Stock Market Index Through Machine Learning Algorithms. Advances in Intelligent Systems and Computing.https://doi.org/10.1007/978-3-030-51859-2_30(Uma & Srinath Naidu, 2021)

[18] Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Ijaz, M. F. (2021). A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. Mathematics, 9(21), 2646.(Padhi et al., 2021)

[19] Sadorsky, P. (2022). Forecasting solar stock prices using tree-based machine learning classification: How important are silver prices? The North American Journal of Economics and Finance, 61, 101705.(Sadorsky, 2022)

[20] Sadorsky, P. (2022). Using machine learning to predict clean energy stock prices: how important are market volatility and economic policy uncertainty? Journal of Climate Finance, 1, 100002.(Sadorsky, 2022)

[21] Jiang, M., Liu, J., Zhang, L., & Liu, C. (2020). An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Physica A: Statistical Mechanics and its Applications, 541, 122272.(Jiang et al., 2020)

[22] Chatigny, P., Patenaude, J. M., & Wang, S. (2021). Spatiotemporal adaptive neural network for long-term forecasting of financial time series. International Journal of Approximate Reasoning, 132, 70-85.(Chatigny et al., 2021)

[23] Lim, B., &Zohren, S. (2021). Time-series forecasting with deep learning: a survey.

Philosophical Transactions of the Royal Society A, 379(2194), 20200209.(Lim &Zohren, 2021)

[24] Lara-Benítez, Pedro, Manuel Carranza-García, and José C. Riquelme. "An experimental review on deep learning architectures for time series forecasting." International journal of neural systems 31.03 (2021): 2130001.(Lara-Benítez et al., 2021)

[25] Deng, S., Zhu, Y., Huang, X., Duan, S., & Fu, Z. (2022). High-frequency direction forecasting of the futures market using a machine-learning-based method. Future Internet, 14(6), 180.(Deng et al., 2022)

[26] Deng, S., Huang, X., Wang, J., Qin, Z., Fu, Z., Wang, A., & Yang, T. (2020). A decision support system for trading in the Apple futures market using predictions fusion. IEEE Access, 9, 1271-1285.(Deng et al., 2020)

[27] Pham, L., & Do, H. X. (2022). Green bonds and implied volatilities: Dynamic causality, spillovers, and implications for portfolio management. Energy Economics, 112, 106106.(Pham & Do, 2022)

[28] Khetani, V., Gandhi, Y., Bhattacharya, S., Ajani, S. N., & Limkar, S. (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262.

[29] Sable, N. P., Shende, P., Wankhede, V. A., Wagh, K. S., Ramesh, J. V. N., & Chaudhary, S. (2023). DQSCTC: design of an efficient deep dyna-Q network for spinal cord tumour classification to identify cervical diseases. Soft Computing, 1-26.

[30] Borkar, P., Wankhede, V. A., Mane, D. T., Limkar, S., Ramesh, J. V. N., & Ajani, S. N. (2023). Deep learning and image processing-based early detection of Alzheimer disease in cognitively normal individuals. Soft Computing, 1-23.

[31] Kaulage, A. ., Mane, D. ., Upadhye, G. ., Rajput, S. D. ., Kale, S. ., & Zope , B. . (2023). Exercise Movement Detection Using Spearman Correlation-based Sliding Window Technique. International Journal of Intelligent Systems and Applications in Engineering, 12(2s), 48–54.

[32] Kumar, A., & Sharma, S. K. (2022). Information cryptography using cellular automata and digital image processing. Journal of Discrete Mathematical Sciences and Cryptography, 25(4), 1105-1111.

[33] Sairise, Raju M., Limkar, Suresh, Deokate, Sarika T., Shirkande, Shrinivas T. , Mahajan, Rupali Atul & Kumar, Anil(2023) Secure group key agreement protocol with elliptic curve secret sharing for authentication in distributed environments, Journal of Discrete Mathematical Sciences and Cryptography, 26:5, 1569–1583, DOI: 10.47974/JDMSC-1825