Implication of Varying Estimation Period of Technical Indicators on Trading Performance using Traditional ML Models

Subhajit Chattopadhyay
Research Scholar
Lovely Professional University, Punjab , India, chatterjee.subhajit01@gmail.com,
https://orcid.org/0000-0003-2206-2694

Abstract. Recent studies of Security Exchange Board of India (SEBI) underlines the precarious condition of the retail traders of India. Keeping the retail traders in mind, the present study explored how retail traders' performance can be improved with minimal difficulty in terms complexity of computation techniques and requirement of computation power. The study clearly underlines that for short term trading, reduction of evaluation period of technical indicators considerably improves trading performance. For 1-day and 3-day ahead forecasting of direction of all indices can be best performed with 3-day evaluation period. For 5-day ahead prediction, 5-day based evaluation suits the best for Quality30 and Alpha50. For Nifty50, it is 7 days of evaluation period suits the best. The maximum return of Buy & Hold (B&H), strategy from Nifty 50 stands at 205.5%. But with customised evaluation period the maximum return generated by Nifty 50, Nifty Alpha 50 and Nifty 200 Quality 30 are 289.3%, 647% and 349% respectively. Such performance also outperforms the return from B & H strategy of respective indices. The study also identified that William's R, Relative Strength Index (RSI), FastK are important predictors of periodic return of all indices under Ridge and LASSO. Similarly, volatility indicators such as VIX close, normalized average true range (NATR), T3 moving average are important predictors of periodic return of Nifty50 under spline. Mostly importantly, the study shows if trades are executed in 5 days, based on trading signal generated with 7days of data yield highest return per trade for all indices.

Keywords: strategy indices, regularised regression, multivariate adaptive regression spline, prediction of market movement, prediction of periodic return

1 Introduction

In the recent past, SEBI has published two reports, Handbook of Statistics (2022) [1] and Analysis of Profit and Loss of Individual Traders dealing in Equity F&O Segment (2022) [2]. These reports underline following patterns and characteristics of Indian stock market and trading behaviour of the retail traders, Following patterns have been observed for NSE (National Stock Exchange) & BSE (Bombay Stock Exchange) -

- at NSE, use of non-algo trade has decreased by 57.48% over the period 2018 - 2023 (in derivatives trading). At BSE, the same has been by reduced by 99.98% over the same period. Over the same period, mobile trading increased by 2.78 times (approx. 3 times).
- at NSE, use of facilities like Co-location has increased by 44% (approx). At BSE, use of Co-location increased by 1.57 times.

Following are the salient observations about trading performance of individual traders –

- active traders (who traded through top 10 brokers) increased 5 times between 2019 and 2022. Top 1% and top 5% of the active traders account for 51% and 75% of the total net profit respectively.
- between 2019 and 2022, number of individual traders trading in index options and stock options went up by 8 and 5 times respectively. The preference to trade in options has increased compared to that of future. 89% of the individual traders who traded in index options incurred an average loss of Rs. 77,000. 89% of the individual traders incurred losses averaging Rs. 110,000 in 2022. For active traders, the average loss was over 15 times of the average profit.

Central Issue: Skewness of profit distribution, low profitability underlines abysmal performance of retail traders which may be the result of inability to use technology and /or sophisticated algorithms, given the cost associated with such options [3]. Poor performance of retail traders can also be attributed to their inability to predict the direction of movement of the stock and market along with lack of expertise in predicting future periodic return. Direction and periodic return helps to ascertain entry and exit points while placing trading orders.

Approach: Therefore, the present study aims to explore how choice of strategy indices, customized technical indicators and traditional machine learning techniques can help retail traders. Traders pursue various trading objectives, such as following trend, seeking alpha or to chase momentum etc. In line with the trading objectives various strategy indices have been created. In this study, two strategy indices such as Nifty Alpha50 (henceforth Alpha50) and Nifty 200 Quality30 (henceforth Quality30) have been selected. To understand future movement of market logistic regression has been

applied. And for prediction of periodic return regularised and adaptive regressions have been used.
The prediction of direction for strategy indices has been done in two ways. In the first approach, prediction has been done solely based upon technical indicators and independent of broad market direction. In the second approach, apart from the technical indicators, direction of broad market movement (i.e. movement of Nifty50) has also been considered to predict the direction of strategy indices.

Objectives: In this context the objectives of the study are as follows -

- to examine the performance of logistic regression with respect to prediction of direction under customised evaluation period of technical indicators (i.e. input variables) for short term prediction (i.e. 1, 3, 5 days) for the indices under study (i.e. Nifty 50, Alpha 50, Quality 30).
- to identify the important technical indicators that govern short term periodic return of indices under study using Spline and Multiple Adaptive regression models.
- to explore trading performance (cumulative & return per trade) under customized evaluation period for indices under study and to compare the same under standard evaluation period.

2 Literature Review

Ayyildiz & Iskenderoglu examined the performance of 7 traditional Ml algorithms (decision tree(DT), random forest (RF), K-nearest neighbour (KNN), Naïve Bayes (NB), logistic regression (LR), support vector machines (SVM), artificial neural network (ANN)) to predict direction of 6 indices. There is no one algorithm which outperformed others for all indices. It is observed that ANN has achieved highest accuracy at 93.48%., but for 3 indices LR outperformed ANN [4]. Chandar & Punjabi proposed Cat Swarm Optimization (CSO) to train multi-layer perceptron to predict the closing price of the stocks on the same day. The model with 9 technical indicators was tested 10 stocks and the performance of the model is superior compared to Genetic Algorithm (GA) trained LSTM (Long short term memory based network), Functional link neural network (FLNN) optimized using Differential Evolution (DE), GA-optimized Radial Basis Function (RBF), CSO optimized ARMA, ELMAN neural network optimized using particle swarm optimization (PSO), PSO optimized MLP, MLP optimized using biogeography based optimization [5]. Chen et. al proposed a model where turning points are predicted using PLR and trained the model with WSVM, backward propagation neural network , gradient boosted decision trees, random forest and long short term memory. It was observed that PLR with weighted support vector machine performed the best [6]. Joginpally proposed Spiking Quantum Neural Network (SQNN) for prediction of S&P100 , ASX200 and SSE50 indices. It is observed that SQNN outperformed support vector regression (SVR), long-short term memory (LSTM), particle swarm optimization (PSO) with LSTM amd Improved PSO with LSTM. The accuracy of SQNN for US , Australia and China markets are 95.2% , 94% and 93.56 respectively. As look-up period increased errors increased for LSTM and PSO-LSTM and the same decreased for SVR & IPSO- LSTM. RMSE decreased and MAPE increased for SQNN as look back period increased [7]. Kilimci & Duvar did a comparative studies where 3 deep learning algorithms (convolutional neural network , recurrent neural network, LSTM) along with 4 types (Word-to-vec , Glove, BERT, FasText) of word embedding for 9 most traded bank stocks to predict the direction BIST100 index. It observed that a combination with word embedding and a deep learning algorithms outperforms individual performance of word embedding and deep learning techniques. Performance vary with source of texts and the accuracies varied from 77.82% to 86.39% [8]. Lv & Yuan et al. compared the performance of six machine learning (ML) algorithms (i.e. support vector machine , logistic regression, random forest, naive bayes, XG boost, classification and regression trees) with six deep neural network models (DNN) (i.e. MLP, DBN, LSTM, SAE, GRU, RNN). It is observed that traditional ML algorithms at par with DNN models with respect to prediction of direction. But trading strategies based on DNN result into better profit after consideration of transaction cost [9]. Ohaegbulem & Kalu used logistic regression to predict the direction NSX ALL Share index using other sectoral index such as banking, insurance, oil & gas, consumer goods. In the process they could achieve an accuracy of 70.17% [10]. Ren & Wu et. al. predicted next day market direction using support vector machine and also factored in sentiment analysis and day of the week effect. Apart from 8 sentiment index they also used open, high, low, close, volume, financial volume, and changes in volume. The study shows that without sentiment component accuracy for both SVM and logistic regression was around 71% . After factoring in sentiment indices, the accuracy went up to 89.93% and 86.59% for SVM and logistic regression respectively [11]. Tang & Dong et al. used weighted support vector machine (WSVM) coupled with piece-wise linear representation (PLR) to predict change in direction. Along with WSVM and PLR, the proposed model also added another RSI (relative strength index) based trading rule along with over or under sampling correction for turning point prediction. The proposed method outperformed models where only PLR & WSVM, PLR , WSVM & RSI, PLR & ANN are used [12]. Zhou & Li et.al. predicted the direction of daily return using 60 indicators for deep learning models. They has used three variants first all the raw indicators (with out transformation), secondly all transformed indicators and subsequently 31 principal components. The structure of model varied from 10 to 1000 layers. Dataset with 31 components has marginally done better compared dataset with 60 indicators (raw). The highest accuracy achieved was 59.6% [13]. The current study aims to explore the implication of change of estimation horizon on the performance of the proposed models. The proposed study also aims to examine the role or return generated by strategy indices based on recommendations generated by the logistic regression. Further, more than prediction accuracy emphasis has been placed to identify the nature of technical indicators that govern periodic return for adaptive and regularised regression techniques.

3 Materials & Methods

Given the ease of computing and the ability to capture various market characteristics, the technical indicators (TIs) are increasingly being used for various predictions (direction of movement, return, value of index or stocks ([14], [15]). TIs are also easily interpretable. Keeping retail investors in mind, computationally inexpensive methods are chosen - logistic regression for prediction of direction and regularized regression and splines for estimation of periodic return. Splines not only help in addressing non-linearity issues, but also looks into interaction among technical indicators.

3.1 Methods & Target variable

Proposed Work: There are numerous studies with respect to application of machine learning algorithms for prediction of direction and values ([12], [16], [17]). Since the study is on short term trading, the forecasting horizon has been limited to 1 day, 3 days, 5 days.

Following is the scope of work of the present study –

- **to predict the direction of movement (i.e. up or down) using logistic regression under standard and customised evaluation period for all the indices**
- **to identify the important factors that governs the periodic returns using Ridge, Spline and LASSO**
- **compare the trading performance (based on prediction of direction) under customized and standard evaluation periods of technical indicators.**

Target Variable: The target or the dependent variable has been defined as the logarithm of periodic return for 1 , 3 and 5 days i.e. $Y = log(Y_{t+i} / Y_t)$, (where i = 1,3,5). To capture the direction of return, the values of Y has been discretized such as values greater than 0 has been considered as 1 else 0.

Trading Rule: If the fitted value of the logistic regression is greater than or equal to 0.5, it is considered upward movement and a buy signal is triggered. But, short selling has not been considered

Input Variables: In line with the study [15], technical indicators, such as, AD (Accumulation Distribution), ADX (Average Directional Index), ATR (Average Trading Range), Bollinger Band (upper, BB_UP), NATR (Normalized Average True Range), ADXR (Average Directional Movement Rating), MFI (Money Flow Index), OBV (On Balance Volume), RSI (Relative Strength Index), T3 – Triple moving average, WR (William'R) , Stochastics – FastD, FastK

The standard formulae of technical indicators imply that the evaluation period generally ranges in between 14 to 26 days [18]. For the purpose of the study, the standard evaluation period stands at 14 days. Further, to examine the impact to change in evaluation period, standard evaluation periods of TIs have been changed with 3, 5 and 7 days.

Data Processing: All the input variable have been standardized prior using for the aforementioned models. The split criteria for cross- validation of the aforementioned models stands at 70:30 and the number of folds of cross validation has been set at 10. Further, for the adaptive regression (Spline) model, a hyper grid has deployed which can take care of linear to cubic relations among the various technical indicators and its interaction variables. Number of such terms has been limited to 100 and only top most important 10 terms will be displayed and first 3 variables has been captured for comparison purpose.

3.2 Data

This study involved one broad market index CNX Nifty50 and two other strategy indices, Nifty200 Quality30 and Nifty Alpha 50. To understand the volatility of the market index VIX has been used. The time frame of the study is 01/03/2019 to 28/03/2024. All the information has been taken from www.niftyindices.com, www.nseindia.com and www.finance.yahoo.com .

4. Results & Discussion

4.1 Direction prediction with standard formulae

In the context of prediction of direction of movement, Table 1 summarizes the performance of logistic regression on Alpha50 and Quality30 with standard formulae of the technical indicators. It is also evident that as prediction horizon increased accuracy of the model decreased except for Nifty50 which stayed almost constant for all forecasting period.

Table 1. Performance of logistic regression with standard formulae

Akaike Information Criteria (AIC)	Strategy / Broad Market Index	Prediction Horizon		
		1-day ahead	3-day ahead	5-day ahead
AIC	Alpha 50	1187.9	1494.9	1510.1
AIC	Quality30	807.61	1427.5	1505.6
AIC	Nifty50	1675.1	1639.7	1675.1

The initial observations are as follows (significance levels of each indicated within brackets) -

- Nifty 50: Under standard formulae of TIs, variables like AD (90%), NATR (99%), BB_UP (99%), MFI (95%), OBV (99.99%), WR (95%), VIX_close (95%) found to be significant for 1 day ahead prediction. For 3 day ahead prediction, only 3 variables such as AD (95%), NATR (99%), VIX_close (99.9%) are significant. And ADXR (99.9%), VIX_close (95%) are relevant for 5 day ahead prediction of direction.
- Alpha50: For 1 day ahead prediction, ADX(95%), NATR (99%), BB_UP (99.99%), T3 (99.99%), WR (99.99%), FastD (99.99%), SlowD (99%), RSI (99.99%) are significant. TIs such as AD (95%), AXR (95%), NATR (99%), WR (99%), FastD (99.99%), RSI (99.99%), VIX_close (95%) are significant for 3 day ahead prediction. In case of 5 day ahead prediction, AD (99.9%), NATR (95%), BB_UP (99%), T3 (95%), WR (90%), FastD (95%), RSI (99.9%), VIX_close (90%) found to be relevant at various levels.
- Quality30: In case of 1 day ahead prediction, T3 (99%), FastD (99.9%), SlowD (99.9%), WR (99.9%) are relevance. TIs such as ADX (95%), WR (99.99%), ADXR (95%), FastD (99.9%) , SlowD (99%) are significant for 3 day ahead prediction. For 5 day ahead prediction, AD (99%), ADX (99.9%), NATR (99%), T3 (99%), WR (99.9%), FastD (95%), RSI (99.9%), VIX_close (99.9%) are important.
- AIC values for Nifty50 are higher than that of Alpha50 and Quality30 indices. The AIC value for Nifty 50 remained constant for all the forecasting horizons. For strategy indices, AIC value is least for 1 day ahead forecasting and it gradually increased with the length of the forecasting horizon. Across all prediction horizons, AIC values are least for Quality30, followed by Alpha50 and Nifty50. The number of significant indicators of Nifty50 are lesser than the number of significant variables for Alpha50 and Quality30 indices.

4.2 Direction prediction with customized evaluation periods

To explore the possibilities to enhance the performance of logistic regression, the evaluation periods are tweaked to 3 days, 5-days and 7 days. Table 2 clearly shows that change in the evaluation period of the technical indicators improved the performance of the models as average AIC values of all prediction horizon decreased. It shows, 1 day and 3-day ahead movement of Nifty50 is best predicted with 3-day evaluation period. For 5- day ahead movement prediction, evaluation period of 7-days works the best. Tables 3 & Table 4 summarizes the performance logistic regression for Quality30 and Alpha50. In case of Quality30, Table. 3 shows that for 1-day and 3-day ahead prediction, an evaluation period of 3 days works best and for 5 day ahead prediction, 5-day based evaluation outperforms other. Similar trend can also be observed for Alpha50 (Table 4). Further, the impact of change in evaluation period is least for 5-day ahead prediction for both Quality30 and Alpha50.

Table 2. Logistic regression - Performance - Nifty 50

Akaike Information Criteria (AIC)	1-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	732.08	1028.3	1167
	3-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1508	1555.2	1552.7
	5-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1536.5	1537.3	1508.7

Table 3. Logistic regression - Performance – Quality30

Akaike Information Criteria (AIC)	1-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	762.07	1522	1170
	3-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1463.1	1522	1497.6
	5-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1534.8	1522	1524

Table 4. Logistic regression - Performance - Nifty Alpha50

Akaike Information Criteria (AIC)	1-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	703.51	1010.2	1197
	3-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1452.9	1506.8	1518.9
	5-day ahead prediction		
	3-day data	5-day data	7-day data
AIC	1565.7	1557.3	1561.7

4.3 Performance of cross validated models

Analysis of Nifty50: From Table no. 5, it can be observed that for 1-day and 3-day ahead prediction of direction, 3-day evaluation period based model provides best result. The same for 5-day ahead prediction is 5-day based evaluation period.

Analysis of strategy indices: Table no. 5 also shows that for 1-day and 3-day ahead prediction, 3-day evaluation period of technical indicators leads to highest accuracy for test data for all the prediction horizon. For 5-day ahead prediction, there is no unequivocal choice. For Alpha50, the accuracy remains same with 5-day or 7-day based evaluation. For Quality30, 3-day or 5-day based evaluation does not show any substantial difference.

Table 5: Accuracy (in %) of cross validated logistic regression models (test data)

Evaluation Period / Forecast Horizon	**1-day ahead prediction**			**3-day ahead prediction**			**5-day ahead prediction**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
	Nifty50			Alpha50			Quality30		
#1-D.AH	87	76	76	76	74	72	86	77	78
3-D.AH	67	66	63	69	64	68	69	64	62
5-D.AH	66	67	65	62	65	65	65	65	64

(*D-EP – day / day of evaluation period; #D.AH – day / days ahead)

4.3 Prediction of periodic return

4.3.1 Use of Regularized Regression

Table 6 to Table 8 captures the performance of ridge regression and LASSO for all the indices for all prediction horizon. Under regularized regression Ridge regression and Least Absolute Shrinkage and Selection Operator (LASSO) have been applied to predict periodic return of the index. Root Mean Square Error (RMSE) of both train and test dataset indicated and RMSE of test data has been used for comparison purpose.

Analysis of Nifty50: The error rate (RMSE) is marginally different between ridge regression and LASSO for all indices. Error rates are found to increases with length of prediction period. Both under ridge and LASSO , it is observed that for 1-day , 3-day and 5-day ahead prediction, error rate is least when evaluation period of the technical parameters is 3-day. Across all prediction horizons, Willam's R (WR), RSI & FastK are found to be the common significant indicators governing periodic return.

Table 6. Performance of Ridge and LASSO - Nifty50

Train , Test	1-day ahead prediction			3-day ahead prediction			5-day ahead prediction		
	3D-EP*	5D-EP	7D-EP	3D-EP*	5D-EP	7D-EP	3D-EP*	5D-EP	7D-EP
RMSE	0.008, 0.013	0.001, 0.011	0.012, 0.009	0.012, 0.19	0.020, 0.020	0.021, 0.017	0.009, 0.013	0.026, 0.025	0.026, 0.023
** Var. Imp. (3)	WR, RSI, FastK	WR, RSI,, FastK	WR, RSI, FastK	RSI, FastK, ATR	WR, RSI	WR, RSI, FastK	WR, RSI, FastK	VIX close, RSI, WR	VIX close, RSI, FastK
LASSO - Nifty 50									
RMSE	0.008, 0.013	0.10, 0.011	0.012, 0.009	0.020, 0.019	0.02 , 0.02	0.021 , 0.017	0.011 , 0.011	0.026 , 0.25	0.025, 0.023
** Var. Imp. (3)	RSI, FastD, FastK	WR, RSI, FastK	FastD, RSI, WR	RSI, FastD, NATR	WR, RSI	WR, VIX close, RSI	RSI, WR, FastD	VIX close, WR, RSI	RSI, VIX close, NATR

(*D-EP – day / days of evaluation period; ** 3 most important technical indicators)

Analysis of strategy indices: The RMSE of Ridge and LASSO for strategy indices are marginally different (Table 7 & 8). It is observed that for 1 day and 3-day ahead prediction, models with 3-day evaluation period performs better. For 5-day ahead prediction, 5-day based evaluation period shows superior performance. For Quality 30 index, William's R, RSI and Fast K are top three predictors of periodic return using ridge regression. With respect to LASSO, in most cases the number of influential predictors is limited only one i.e. William's R.

Table 7. Performance of Ridge and LASSO - Quality30

Train, Test	1-day ahead prediction			3 day ahead prediction			5 day ahead prediction		
	3D-	5D-	7D-	3D-	5D-	7D-	3D-	5D-	7D-
RMSE	0.008, 0.008	0.008, 0.010	0.10, 0.10	0.017, 0.017	0.017, 0.017	0.016, 0.017	0.023 , 0.022	0.024, 0.019	0.022 , 0.024
**Var. Imp. (3)	WR, RSI FastK	WR, RSI FastK	WR, RSI FastK	WR, RSI FastK	WR, RSI FastK	WR, RSI FastD	WR, RSI Fast K	WR, RSI FastK	WR, RSI FastK

LASSO									
RMSE	0.008 0.008	0.008, 0.009	0.010, 0.009	0.017, 0.017	0.017, 0.017	0.017, 0.017	0.023 , 0.021	0.023, 0.021	0.021 , 0.023
****Var. Imp. (3)**	WR	WR, FastD	WR	WR	WR	WR	WR	WR, RSI, NATR	RSI, FastD WR

(*D-EP – day / days of evaluation period; ** 3 most important technical indicators)

The periodic return of Alpha50 is majorly governed by William's R, RSI , Fast K in all cases. Considering RMSE of test data, it can be seen that 1 day ahead return is not sensitive to evaluation period. 3 day ahead return can be better estimated when indicators are estimated on the basis of 5 day. An evaluation period of 7 days better predict 5 day ahead return.

Table 8. Performance of Ridge and LASSO – Alpha50

Train, Test	**1-day ahead forecast**			**3-day ahead forecast**			**5-day ahead forecast**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
RMSE	0.14, 0.013	0.013 0.012	0.013 0.012	0.010 0.015	0.013 0.012	0.011, 0.015	0.031, 0.034	0.032, 0.029	0.034, 0.030
***Var. Imp. (3)**	NATR , MFI VIX close	WR, RSI FastK	WR, RSI FastK	RSI, FastK Slow D	WR, RSI FastK	WR, RSI FastD	VIX close, RSI, FastK	VIX close, RSI, WR	VIX close, WR, RSI
LASSO									
RMSE	0.014 0.013	0.13 0.013	0.013 0.012	0.009, 0.014	0.012, 0.012	0.011, ,0.015	0.031, 0.034	0.032 , 0.09	0.031, 0.030
***Var. Imp. (3)**	Slow D, RSI, NATR	WR	WR, RSI	FastD FastK RSI	WR, RSI, Slow D	WR, RSI, FastD	RSI VIX close	VIX close, NATR , RSI	VIX close, NATR , WR

(*D-EP – day / days of evaluation period; ** 3 most important technical indicators)

4.3.2 Use of Multiple Adaptive Regression Spline

The models have been compared in term of RMSE of test data which has been indicated after RMSE of training dataset. For Nifty50, 1 day and 3 day ahead returns can be better predicted when LASSO is used with the parameters evaluated over 3 days. 5 day ahead return can be better predicted using 5 day evaluation period. The important factors governing periodic return are very different that of ridge regression. It can be observed that VIX close, T3 and NATR are the important predictors across all prediction horizon. Moreover, linear models with 2 to 12 terms can be used to predict 1 day ahead return. The best pruned models to predict 3 day and 5 day ahead return are mostly quadratic or cubic in nature with 23 to 32 terms. For Alpha50, being a momentum based strategy, 1 day, 3 day and 5 day ahead return can best predicted using 5 days evaluation period. RSI , VIX close , ATR or NATR governs the periodic return for Alpha50. But, no significant predictor of 1 day ahead return can be identified. Most best pruned models have 23 terms and approximately 44% models are linear. The error rate is least for Quality 30 compared other two indices. 1 day and 3 day ahead returns can be better predicted using 3 day evaluation period. And 5 day ahead return can be better predicted using 7 day evaluation period. William's R, NATR / ATR, T3 are the most common predictors across all prediction horizon. Most of the best performing models have 23 terms and 67% of the models are linear.

Table 9. Performance of Spline - Nifty50

Train, Test	**1 day ahead prediction**			**3 day ahead prediction**			**5 day ahead prediction**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
RMSE	0.012 0.012	0.012, 0.012	0.011, 0.015	0.017, 0.022	0.017, 0.022	0.017, 0.023	0.018, 0.035	0.019, 0.034	0.019, 0.034
****Var. Imp.** (3)	NATR, T3, VIX close	NATR, T3, VIX close	VIX close, T3, NATR	VIX close, BB_UP, AD	VIX close, T3, NATR	VIX close, T3, NATR	VIX close, T3, OBV	VIX close, ADXR, AD	VIX close, BB_UP , NATR

(*D-EP – day / days of evaluation period; ** 3 most important technical indicators)

Table 10. Performance of Spline – Alpha50

Train, Test	**1 day ahead prediction**			**3 day ahead prediction**			**5 day ahead prediction**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
RMSE	0.012, 0.013	0.008, 0.011	0.009, 0.011	0.02,0.022	0.02, 0.022	0.021, 0.021	0.026, 0.028	0.029, 0.026	0.024, 0.031
****Var. Imp.** (3)		RSI, ATR FastD	WR RSI, NATR	RSI, VIX close, T3	RSI, WR, BB_UP	RSI, NATR BB_UP	VIX close, BB_UP, RSI	VIX close, T3, RSI	VIX close, ADX, RSI

(*D-EP – day / days of evaluation period; ** 3 most important technical indicators)

Table11. Performance of Spline - Quality30

Train, Test	**1 day ahead prediction**			**3 day ahead prediction**			**5 day ahead prediction**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
RMSE	0.005, 0.005	0.004, 0.006	0.009 0.008	0.014, 0.014	0.014, 0.016	0.014, 0.016	0.018, 0.020	0.018, 0.20	0.017, 0.019
****Var. Imp. (3)**		WR, ATR, MFI		T3, NATR FastK,	FastD, T3 NATR	WR, RSI, BB_UP	T3, NATR VIX close	RSI, NATR MFI	BB_UP, AD, WR

(*D-EP – day / days of evaluation period; ** 3 most important 3 technical indicators)

4.4 Comparison among Ridge, LASSO & Spline

There is no significant difference in the performance of ridge and LASSO with respect to prediction of periodic return. RMSE of test data of Nifty 50 is least for ridge regression , followed by LASSO and Spline across all prediction period. Momentum indicators have been found to be influential predictors of periodic return in the case of regularised regression (i.e. ridge and LASSO). In case Spline, its volatility indicators like average true range (ATR) / normalised true range (NATR), VIX close and Bollinger band (upper) and T3 (a smoothening variable) found to be influential. For Alpha50, Spline outperforms Ridge and LASSO to estimate return over 1 day and 5 days. But, Ridge and LASSO outperforms to predict return of 3 days. For Quality 30, spline outperforms ridge and LASSO for all prediction horizon. Further, momentum based indicators such as William's R, RSI, FastK, FastD dominate the performance of ridge and LASSO for all indices. In case of Spline, periodic return of Nifty 50 is governed by volatility based indicators as such VIX close, NATR, BB_UP and a smoothening variable T3. But for the strategy indices its a mix of momentum and volatility based indicators which governs the periodic return of the indices.

4.5 Estimation of return from trades

- **Standard evaluation period (14 days):**
  - **Nifty 50:** Trades executed within a day generated 64.4% return. Whereas the same for trade executed in 3 days and 5 days stands at 71.1% and 94% respectively.
  - **Alpha 50:** Trades executed in 1-day yielded 81.7%. Return from trades executed in 3 days and 5 days are 287.1% AND 260.3% respectively.
  - **Quality 30:** The returns from trades executed in 1-day, 3-days and 5-days stands at 54.3% , 134.2%, 270.3%.
- **Customised evaluation period**:

- **Nifty 50:** The cumulative return of Buy & Hold (henceforth B&H) strategy is 200%. Nifty50 generated a maximum return of 297% when traded in 5 days based on 5-day evaluation period. Further, the maximum return generated for trades executed in 1 day is 71.1% based on 5-day data. The same for trades executed in 3rd day is 172.1% based on 3-day evaluation period.
- **Alpha 50:** The cumulative return under B& H for Alpha50 is 300%. It generates a maximum cumulative return of 647% when traded for 5-day period based 5-day evaluation period. Trades executed next day can generate a maximum return of 138.5% based on 5-day evaluation period. Trades executed in 3-days can lead to a maximum return of 405% based on 5-day evaluation period.
- **Quality 30:** The cumulative return on B& H strategy over the period for Quality30 is 95%. And it generates a maximum cumulative return of 349% when traded for a 5-day period based on 5-day evaluation period. Trades executed in 1-day can lead to a maximum return of 77.7% based on 3-day data and the same for trades executed in 3 days is 209.5%. based on 5-day evaluation period.

- **Incorporating trend of Nifty50: F**or strategy indices, trades have been executed based on both technical indicators and also factoring in movement of Nifty50. When buy signals are generated after incorporating nifty return, it is found that the highest cumulative return stands at 349% for Quality30 and the same for Alpha50 stands at 525% (approx.). Both the are less than cumulative returns generated by the stand alone models.

- **Trade-wise profitability (Table 12.):**
  - **Standard evaluation period (14days):** Highest return per trade for Quality 30 & Alpha 50 stands at 0.21% and 0.617% respectively. The same for Nifty 50 stands at 0.132%. For Quality 30 & Alpha50 profit per trade is highest for trades executed in 5 days. For Nifty50 the same is highest for trades executed in 3 days.
  - **Customised evaluation period:** For all indices, one can note that as evaluation period increased, profit per trade also increases. Further, for all the indices, highest return is generated when trade is executed in 5 days with 7 days of evaluation period of technical indicators. For Nifty50, 3-day and 5-day based trades, an evaluation period of 3 days generated substantially greater return. The same for Alpha50 and Quality30 is 7days of evaluation period.

Forecast Horizon	**Quality 30**			**Alpha 50**			**Nifty 50**		
	3D-EP*	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**	**3D-EP***	**5D-EP**	**7D-EP**
#1-D.AH	0.39	0.42	0.42	0.084	0.172	0.129	0.032	0.056	- 0.06
3-D.AH	0.77	0.81	0.82	0.336	0.420	0.312	0.137	0.118	0.11

5-D.AH	1.09	1.17	1.2	0.630	0.647	0.574	0.194	0.236	0.229

Table 12. Return per trade (in %)
(*D-EP – day / days of evaluation period, #D.AH – day / days ahead)

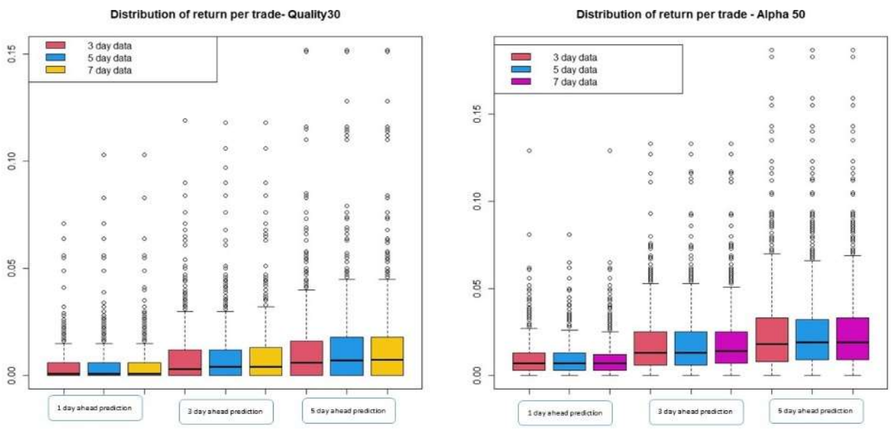

Fig. 1. Distribution of return per trade – Quality30 & Alpha50

4.6 Comparison with previous literatures

Ayyildiz, & Iskenderoglu used linear model, artificial neural network, random forest, support vector machines ad decision tree. The range of accuracies for the aforementioned models are 80.83 to 86.45%, 81.01 to 93.48%, 50.59 to 64.09%, 72.22 to 84.66% and 53.37 to 59.57% respectively[4]. Joginpally also deployed logistic regression, random forest, support vector machine and artificial neural network with accuracy ranges 85.51 to 90.60%, 84.45 to 91.27%, 82.68 to 883.59% and 84.45 to 89.93% respectively [7]. Lv & Yaun used LASSO and Elastic Net where error rate was 0.659 to 0.859 and 0.621 to 0.787 respectively [9]. Ohaegbulem & Kalu used deep learning, random forest and gradient boosting machines with 78%, 82% and 81% respectively. The error rates for algorithm were 0.17, 0.14 and 0.14 respectively [10]. Gandhamal & Kumar used logistic regression with accuracy level of 59.17% [17]. Milunovich deployed logistic regression with an accuracy of 72.7% [19]. Mokhtari & Yen et. al used logistic regression with an accuracy of 71.6 to 72.8% [20].The performance shows that for all indices, accuracy is highest while predicting one day ahead movement, but it decreases as the forecasting horizon increases. Further, for 1 day ahead prediction accuracy of LR matched with complex ML algorithm, but for 3-day and 5-day ahead prediction its accuracy is less than 70%.

5 Conclusion

In terms of accuracy of prediction of direction, 3-day evaluation period works best for 1 & 3 day ahead forecasting for indices. Same holds for prediction of periodic return of all indices. For Quality 30 & Alpha50, 5-day ahead market direction can be predicted best with 5 days of estimation period and the same for Nifty50 is 7 days. For Nifty 50 and Alpha 50, 7 days of evaluation period accurately predicts periodic return of 5 days and the same for Quality 30 is 5-day evaluation. It is also evident that, Spline outperformed the rest for Nifty 50 and Alpha 50. Moreover, VIX close, T3 important predictor of periodic return of Nifty 50. Momentum based indicators such as William's R, RSI, FastK are important predictors of periodic return of all indices for Ridge and LASSO. For Spline, volatility based indicators such VIX close, NATR and T3 are important predictors of periodic return of Nifty50. Most importantly, trades executed in 5 days based on 7-day data yield maximum return, both in terms of cumulative as well as return per trade. This holds for all indices. Thus, the study shows that simplistic models like logistic regression and prudent choice of evaluation period can enable a trader to outperform market & B&H strategy. In future, there is a scope to perform research on optimization of evaluation period for various market conditions (such as upward or downward trend, high or low volatility etc.)

References

[1] Security Exchange Board of India (SEBI), Handbook of Statistics. https://www.sebi.gov.in/reports-and-statistics/publications/may-2023/handbook-of-statistics-2022_74606.html (2022).

[2] Security Exchange Board of India (SEBI), Analysis of Profit and Loss of Individual Traders dealing in Equity F&O Segment, https://www.sebi.gov.in/reports-and-statistics/research/jan-2023/study-analysis-of-profit-and-loss-of-individual-traders-dealing-in-equity-fando-segment_67525.html (2023).

[3] National Stock Exchange (NSE), Real Time Tariff For Domestic Vendors, https://nsearchives.nseindia.com/web/sites/default/files/inline-files/Download_Real_Time_Tariff_Domestic_01042023.pdf (2023).

[4] Ayyildiz, N., Iskenderoglu , O., How effective is machine learning in stock market predictions? Heliyon 10(2), (2024).

[5] Chandar, K. S., Punjabi, H., Cat swarm optimization algorithm tuned multilayer perceptron for stock price prediction. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 17(7):1–15, (2021).

[6] Chen, X., Hirota , K., Dai, Y., Jia, Z., A model fusion method based on multi-source heterogeneous data for stock trading signal prediction. Soft Computing, 27(10):6587–6611, (2023).

[7] Joginpally,B., An intelligent soft computing framework for forecasting stock movement direction in international stock markets. Journal of Theoretical and Applied Information Technology, 101(9), (2023).

[8] Kilimci, Z. H., Duvar, R., An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: A case of istanbul stock exchange (bist 100). *IEEE Access*, 8:188186–188198, (2020).

[9] Lv, D., Yuan, S., Li, M., Xiang, Y., An empirical study of machine learning algorithms for stock daily trading strategy. Mathematical Problems in Engineering, (2019).

[10] Ohaegbulem, E. U., Kalu, O. C., Using logistic regression to predict the direction of Nigeria stock market returns (based on some selected sector indexes, 2015-2022). European Journal of Theoretical and Applied Sciences, 1(4):1180–1200, (2023).

[11] Ren, R., Wu, D. D., Liu ,T., Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Systems Journal, 13(1):760–770, (2018).
[12] Tang , H., Dong, P., Shi, Y., A new approach of integrating piecewise linear representation and weighted support vector machine for forecasting stock turning points. Applied Soft Computing, 78:685–696, 2019.
[13] Zhou, T., Li, X., Wang, P., Statistics and practice on the trend's reversal and turning points of chinese stock indices based on gann's time theory and solar terms effect. Mathematics, 9(15):1713, (2021).
[14] Alsubaie, Y., K. Hindi, El., Alsalman, H., Cost-sensitive Prediction of Stock Price Direction: Selection of Technical Indicators. IEEE Access, 7:146876–146892, (2019).
[15] Naza´rio, R. T. F., e Silva, J. L. , Sobreiro,V. A., Kimura , H., A literature review of technical analysis on stock markets. The Quarterly Review of Economics and Finance, 66:115–126, (2017).
[16] Faghihi Nezhad, M., Minaei Bidgoli ,B., Development of an ensemble learning-based intelligent model for stock market forecasting. Scientia Iranica, 28(1):395–411, (2021).
[17] Gandhmal, D. P., Kumar, K., Systematic analysis and review of stock market prediction techniques. Computer Science Review, 34:100190, (2019).
[18] Colby, R. W., The Encyclopedia of Technical Market Indicators. McGraw-Hill, (2003)
[19] Milunovich, G., Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39(7):1098–1118, (2020).
[20] Mokhtari, S., Yen , K. K., Liu , J.. Effectiveness of artificial intelligence in stock market prediction based on machine learning. *arXiv preprint arXiv:2107.01031*, (2021).
[21] Khan, A. H., Shah , A., Ali, A., Shahid , R., Zahid, Z. U., Sharif , M. U., Jan, T., Zafar, M. H., A Performance comparison of machine learning models for stock market prediction with novel investment strategy. PLOS One, 18(9):e0286362, (2023).
[22] Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A., Predicting Stock Market Trends Using Machine Learning And Deep Learning Algorithms Via Continuous And Binary Data; A Comparative Analysis, IEEE Access, 8:150199–150212, (2020).