International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.9, September 2021 E-ISSN: 2347-2693

Process Improvement in the Criteria of Investment on Stock Market Using Data Mining Techniques

S.K. Sharma

Department of Computer Science & IT, Kalinga University, Nawa Raipur(C.G.) India

*Corresponding Author: shishirsharma51@gmail.com, Tel.: +919827460305

DOI: https://doi.org/10.26438/ijcse/v9i9.3138 | Available online at: www.ijcseonline.org

Received: 13/Sept/2021, Accepted: 20/Sept/2021, Published: 30/Sept/2021

Abstract - Exact expectation of stock trade returns could be a difficult undertaking in view of unpredictable and non-direct nature of the monetary securities exchanges. The financial exchange information, as S&P500 Index is gigantic, perplexing, non-straight and noised. Foreseeing stock costs is a difficult undertaking as it relies upon different elements including however not restricted to worldwide economy, political conditions, organization's monetary reports and execution and so on The speculation models utilizing this data have been a test. Along these lines, to augment the benefit and limit the misfortunes, procedures to anticipate estimations of the stock in advance by examining the pattern over the past couple of years, could end up being exceptionally valuable for making securities exchange developments [42,43]. This investigation proposes the accompanying momentary bit by bit technique: to consolidate two data sources that the financial backers can break down to settle on a choice. In the first place, the file information comprises the contribution for Profound Learning Neural Organization preparing, for addressing and estimating following day stock worth. Second, this exploration distinguishes the principal delegate endeavors, remembered for File, which address the List social inclination, utilizing Highlight Determination Investigation. At long last, the yields are supplemented and verified; the technique shows promising outcomes to upgrade the financial backer's choice. Especially, for stock trade investigation, the data size is enormous and furthermore non-direct. To influence such an information proficient model is required which will recognize the secret examples and muddled relations during this huge informational index. AI strategies during this region have demonstrated to improve efficiencies by 60-86 percent when contrasted with the past techniques.

Keywords - Stock Exchange, Machine Learning, Predict, Feature Selection and Forecasting.

1. INTRODUCTION

Securities exchange is described as powerful, eccentric and non-direct in nature. The consistency of the stock trade has been long an examination theme. Estimating market hazard requires quantitative procedures to explore individual monetary instruments and an arrangement of resources. This quantitative measure or model catches patterns and practices in information which are then wont to find future qualities [1]. Like other comparable exploration addresses stock expectation is additionally an endeavor to encourage the dynamic assignment for speculation. Nevertheless, it holds its own novel attributes, as far as an alternate explicit undertaking and thus various procedures.

As per Fang [2], the stock records are for the most part simpler to figure with than singular stocks. Zheng and Chen [3] and Olden [4] recommend that albeit the specialists can't concede to if the financial exchanges are unsurprising, contemplating whether one can foresee them is an energizing topic.

Stock predictions are frequently made by utilizing a factual methodology, which regards the stock information as a measurement and uses no other data. Models incorporate Remarkable Smoothing Models (ESM), ARIMA models, Curve and GARCH models, among others [1]. The monetary models are upheld measurable decencies and suspicions inside the fundamental information. In certain cases, ridiculous presumptions are made to improve on the issue or to permit the numerical inference of the model. Given the mind boggling conduct of monetary business sectors, these models can possibly neglect to address basic highlights of fundamental information [1].

Then again, highlight based AI approaches exploit financial information, just as verifiable stock information. Models like Support Vector Machines, Genetic Algorithms, and Artificial Neural Network (ANN). ANN has been one among the premier fruitful applications [5,6,7,8].

As of late there has been an extraordinary interest in profound organization structures. Deep Neural Network (DNN) has different effective reports in machine learning [2]. There are alluring highlights of neural network, which make them a fitting instrument for market hazard displaying. As per Mostafa et al. [1] and Qian [5], the outcomes accomplished show the pervasiveness of neural networks over factual models. Additionally, they're normally reasonable to demonstrate nonlinearities in information.

The depicted situation encourages investigating choices to improve the rules to securities exchange ventures, utilizing the most forecast methods of AI, as is presented in this study, and considering the S&P 500 Record information.

2. RELATED WORK AND BACKGROUND

A. S&P 500 Index

The standard & poor's 500 is an American financial exchange list dependent available capitalizations of 500 huge organizations having regular stock recorded on the New York Stock Exchange (NYSE) or Nasdaq stock exchange (NASDAQ). Every endeavor name is addressed utilizing an abbreviation; like, INTC for Intel.

S&P Dow Jones Files decide the S&P 500 record segments and their weightings; it is viewed as probably the best portrayal of the U.S. stock exchange.

The referenced information can be obtained from URL: http://www.financeyahoo.com. Table I shows part of a concentrate of the every day information accessible for the S&P500 File for every Endeavor. Open, High, Low and Close allude to stock worth; Volume alludes to the quantity of portions of the stock exchange.

B. Stock Forecasting Using Neural Networks for Deep Learning

Numerous artificial intelligence strategies are utilized to foresee stock trade costs [6,7]. ANN stays as a popular decision for this errand and is broadly considered and are appeared to display incredible execution [8], and cutting-edge writing recommends that specialists are endeavoring to utilize deep learning for stock forecast utilizing DNN [8,9].

DNN is an AI worldview for demonstrating complex nonlinear mappings among info and yield, in which the inward boundaries are refreshed iteratively to make the given data sources fit with target yields [10,11]. Deep Learning approaches comprise in adding various repeatable layers to a neural network. Examining the matter, most deep learning systems depend in any event on the accompanying five sorts of models [12,13,14,15]: (I) Convolutional Neural Network, (ii) Intermittent and Recursive Neural Network; (iii) Multi-facet Perceptron; (iv) Back Proliferation Neural Network (BPNN); and, (v) Standard DNNs, which are a blend of layers of various kinds with no specific request.

TABLE I: S&P500 INDEX DATA (EXTRACT)

Date	Open	High	Low	Close	Volume
13/05/2019	2840.19	2840.19	2801.43	2811.87	3894030000
14/05/2019	2820.12	2852.54	2820.12	2834.41	3322720000
15/05/2019	2820.38	2858.68	2815.08	2850.96	3125950000
16/05/2019	2855.80	2892.15	2855.80	2876.32	3338060000
17/05/2019	2858.60	2885.48	2854.23	2859.53	3257950000
20/05/2019	2841.94	2853.86	2831.29	2840.23	3288870000
21/05/2019	2854.02	2868.88	2854.02	2864.36	3218700000
22/05/2019	2856.06	2865.47	2851.11	2856.27	3192510000
23/05/2019	2836.70	2836.70	2805.49	2822.24	3891980000

Standard DNN gives a widespread structure to displaying mind boggling and high-dimensional information. A particularly fascination of DNN approach is that the natural capacity of covering all phases of information driven displaying (highlights determination, information change, and grouping/relapse) inside one structure, i.e., preferably, the specialist can begin with crude information in the space of intrigue and prepare to-utilize arrangement; in addition, this profound component pecking order empowers DNNs to accomplish great execution in numerous errands [16,17].

There are some exploration respect to the suggested utilization of Deep Learning Structures. Nevertheless, the old style engineering for Deep Learning (Fig. 1) has the layers in each completely associated stack with less hubs than the previous [9,18,19,20]. The yield from a definitive stack delivers an expectation of the objective variable.

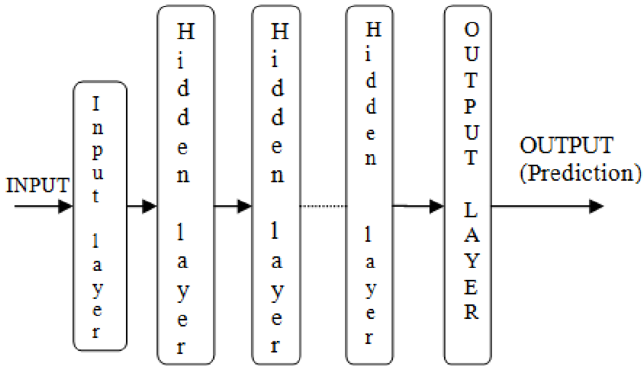

Fig 1 : Back propagation neural network for deep learning architecture

C. Feature Selection Techniques

A feature is an individual quantifiable property of an interaction being noticed, addressed by a variable. Feature Selection (variable elimination) helps in getting information, decreasing calculation prerequisite, diminishing the impact of dimensionality and improving the indicator execution. Thusly, the focal point of highlight determination is to choose a subset of information factors that can depict information, decreasing impacts from commotion, or superfluous factors and still give better prescient outcomes [21,22,23].

As to of mark data, include determination procedure can be generally arranged into three families: administered techniques, semi-directed strategies, and solo strategies [24,25,26]. The accessibility of name data permits administered include choice calculations to productively choose discriminative and pertinent highlights, to separate examples from various classes.

In view of various systems of looking, include determination likewise can be grouped into three techniques, i.e., channel, covering and inserted strategies [22,23,24]. Covering strategies include a learning calculation as a recorder and join utilizing its forecast execution to survey the general value of subsets of factors. At the end of the day, the element determination calculation utilizes a learning strategy (Classifier) as a subroutine with the computational weight that comes from calling a learning calculation to assess every subset of highlights [27,28] (See Fig. 2).

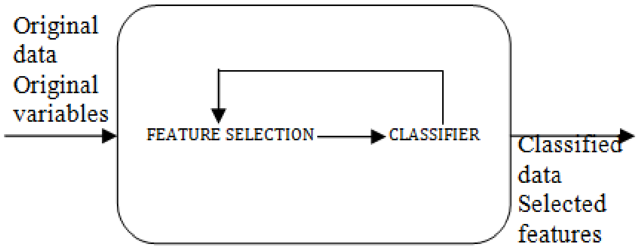

Fig. 2 : Wrapper method configuration [26]

Choices to Highlight determination are Covering Subset Evaluator, Correlation-based Feature Subset Selection (CFS), Head Segments Examination, among others [29,30,31]. On the contrary hand, alternatives for Search Techniques (Classifiers) are Avaricious Stepwise, Developmental Pursuit or Best First [22,32,33].

In a primer test round, utilizing the Legitimacy, or Pearson's coefficient of connection [34,35] as an exhibition measure, the best outcomes were appeared by the CFS as a classifier and Eager Stepwise forward binding for include determination.

CFS might be a basic multivariate channel calculation that positions include subsets reliable with to a connection based heuristic assessment work. The action utilized is:

$$Qzc = (m\ Qzi) / (\ m + m(m-1)\ Qii)^{½}$$

where Qzc is that the correlation between the individual features and also the output class, m is that the number of features, Qzi is that the measure of correlation between each feature and also the output variable, and Qii is the average intercorrelation among features. So, the measure used assigns high values to the subsets that are highly correlated with the output while being weakly correlated with each other. Irrelevant features should be ignored because they'll have a low correlation with the category. Redundant features should be screened out as they're going to be highly correlated with one or more of the remaining features. The acceptance of a feature will depend upon the extent to which it predicts classes in areas of the instance space not already predicted by other features [28].

On the opposite, Greedy Stepwise performs a greedy forward or backward search through the space of attribute subsets. The process stops when the addition/deletion of any remaining attributes leads to an evaluation decrease [33]. The process solves the following model:

$$\max R^2 (G, S)$$
$$S \subset P$$
$$\text{s. t. } |S| = k$$

where:

k number of data sources to choose
P data sources
G target data
α_i are the regression coefficients from fitting G using de
P_i's

$$R^2 (G,S) = \frac{Var(G) - Var(G - \sum_{i \epsilon s} \alpha\ i\ pi)}{Var\ (G)}$$

Var corresponds to the Variance.

3.METHODS

The financial exchange information, as S&P500, are monstrous, unpredictable, non-straight and noised. Subsequently, the venture rules have been a test. This examination proposes the accompanying technique: produce consolidated data that the financial backers can dissect. To begin with, file esteem gauging can be made utilizing a directed learning approach. Furthermore, the most influent ventures with volume esteems that address the List conduct, are distinguished. In this way, the strategies to be utilized are the accompanying:

1) A forecasting model based on deep learning neural networks is defined and trained.

The anticipating interaction requires a model to address the marvel. In this way, the model should extrapolate utilizing the new information, because of the express or implied presence of time as a variable.

The ANNs can display unusual conduct in extrapolation cases, as some old style examines propose [36,37] just as more as of late examinations [38,39]. In this study, an underlying explorative work shows social peculiarities in extrapolation estimating. For addressing the matter, Barnard and Wessels [36] recommend mimicking the ANN with values round the prepared information

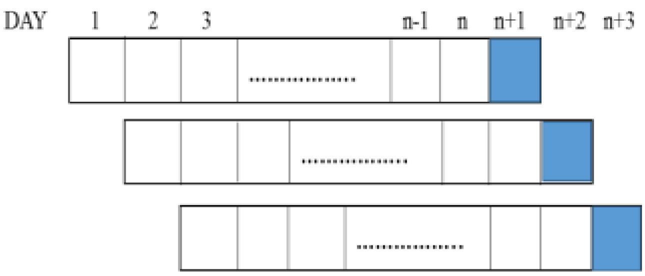

Fig 3 : Sliding window technique

The anticipating model should address information sufficiently, produce quantitative future list esteems and recommend the subjective attribute of the estimated esteem: UP for a value increment, DOWN for a value diminishing, and STABLE for an invariant worth. As per the above contemplations, utilizing information of the past advance, the following worth in time is determined. The sliding window method is appeared in Fig. 3.

Data of n days is used to forecast the value of the day (n+1). At subsequent step, the important data of the day (n+1) is added, and therefore the data on the primary day is deleted. Then, the value of the day (n+2) is forecasted and so on. Besides, as is showed later, new variables must be defined to obtain an implicit representation time and transform the extrapolation into an interpolation problem.

Feature selection: A data mining technique, is used to diminish the complexity and noise of data. So, the investors can identify the most influents enterprises in monitoring the index behaviour. In this case, the forecasting model must represent the data adequately and produce the future quantitative values of variable Open and suggest the qualitative characteristic of it (UP, DOWN, STABLE). The enterprise's data are used to corroborate and complement the forecasting model results.

Thus, the investment decision has two complementary sources. The proposed approach is shown in Fig. 4.

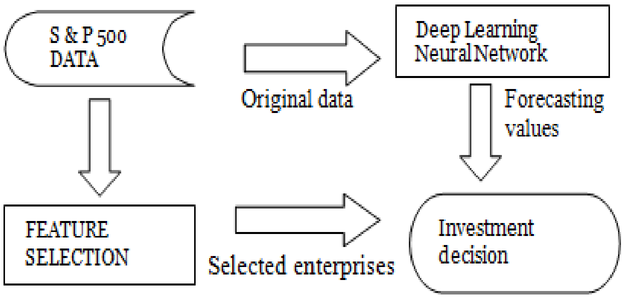

Fig 4 : Investment decision process

4.RESULTS

A. Data

The S&P 500 Index data correspond to market activity days, from May 25, 2014, to May 24, 2019. For initial data processing, enterprises constitute 486 variables and 1259 examples of Open values for each one. The missing values were replaced with mean values.

Fig. 5 shows an extract of the most interesting descriptive statistical results: nonlinearly distributed variables (In the diagonal of the matrix), and the proportionality relationships between them. The relations between variables have high dispersion producing clouds images, as shown in the figure. Alike, Table III shows the Correlation Matrix of data for the first eight enterprises (Vi). There is a high correlation between some variables, which suggest the data duplication presence.

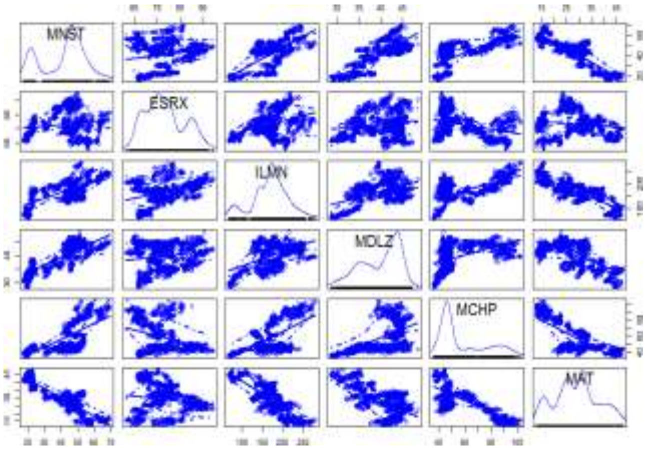

Fig 5 : Scatter plot matrix for some variables

TABLE III : Correlation Matrix Some Variables

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
V_1	1.0	0.1	0.5	0.2	0.2	0.6	-0.4	0.0
V_2	0.1	1.0	0.7	0.3	0.3	0.6	-0.7	0.3
V_3	0.5	0.7	1.0	0.3	0.5	0.9	-0.8	0.5
V_4	0.2	0.3	0.3	1.0	-0.1	0.5	-0.3	-0.4
V_5	0.2	0.3	0.5	-0.1	1.0	0.3	-0.2	-0.6
V_6	0.6	0.6	0.9	0.5	0.3	1.0	-0.8	0.1
V_7	-0.4	-0.7	-0.8	-0.3	-0.2	-0.8	1.0	-0.2
V_8	0.0	0.3	0.5	-0.4	0.6	0.1	-0.2	1.0

Besides, the Covariance matrix (Table IV) shows the different proportionality between variables, which corroborate the scatter plot matrix results.

TABLE IV : Covariance Matrix for Some Variables

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
V_1	10.6	1.7	12.2	1.4	3.5	18.7	-7.0	0.4
V_2	1.7	38.9	30.8	3.3	7.3	37.8	-21.5	22.9
V_3	12.2	30.8	47.8	3.6	14.6	56.7	-25.4	38.0
V_4	1.4	3.3	3.6	4.0	-0.8	8.8	-3.1	-9.2
V_5	3.5	7.3	14.6	-0.8	19.6	13.5	-4.9	28.5
V_6	18.7	37.8	56.7	8.8	13.5	88.2	-38.4	10.8
V_7	-7.0	-21.5	-25.4	-3.1	-4.9	-38.4	23.4	-8.9
V_8	0.4	22.9	38.0	-9.2	28.5	10.8	-8.9	126.6

The significance of the foregoing facts can be summarized as:

i) The frequency of high values on linear correlations suggest that some variables can introduce duplications in data, a way of noise, as a recognized characteristic of stock market behavior.

ii) The distribution of variables are nonlinear; therefore, possible models for treating data must support nonlinear data.

B. Stock Forecasting Using DNN

For constructing, training, and testing the model, MATLAB software [40] is used. The performance of DNN is evaluated using the correlation value, R, of the modeled output. R is determined as follows:

$$R^2 = 1 - \left(\sum(y_{exp} - y_{pred})^2 / \sum(y_{exp} - \bar{Y})^2 \right)$$

Where $yexp$ is the experimental (real) value, $ypred$ is the predicted value, and $\bar{Y}$ is the mean value.

The following eleven variables are derived and used in the forecasting process. These variables definition convert the initial extrapolation forecasting into an interpolation forecasting. Note that the data of year is eliminated. As a notation, S_i and S_{i+1} are successive days.

Input Variable:

- Month: it refers to the month to which a given record belongs.

Table V: Data for OpenPerc variable forecasting

Case	Month	Month day	Week day	Low Diff	High Diff	Close Diff	Vol Diff	Range Diff	Open Close	Open Perc
1	6	10	2	13.99	4.29	-0.57	-391260000	-9.70	1.29	1.19
2	6	11	3	-16.34	-8.56	-16.68	456980000	7.78	-4.17	-0.37
3	6	12	4	-12.00	-2.42	-13.61	-233160000	9.58	3.81	-0.53
......	....	....	....	......	....	......	....	......	....	....
1258	6	6	4	8.95	19.78	23.55	133850000	10.83	4.45	0.17
1259	6	7	5	8.95	19.78	23.55	133850000	10.83	4.45	0.24

- MonthDay: day of the month to which a given record belongs.
- WeekDay: refers to the day of the week corresponding to a given stock record.
- LowDiff: For two consecutive slots $S1$ and $S2$. If $L1$ and $L2$ refer to the Low values for $S1$ and $S2$ respectively, then LowDiff for S2 is computed as ($L2$ - $L1$).
- HighDiff: the difference between the High values of two successive slots. The computation is identical to LowDiff.
- CloseDiff: If two successive slots $S1$ and $S2$ have close values $C1$ and $C2$ respectively, then CloseDiff for $S2$ is calculated as ($C2$ - $C1$).
- VolDiff: For two consecutive slots $S1$ and $S2$, if the mean values of Volume for both the slots are $V1$ and $V2$ respectively, the VolDiff for $S2$ is ($V2$ - $V1$).
- RangeDiff: For two consecutive slots $S1$ and $S2$, suppose the High and Low values are $H1$, $H2$, $L1$ and $L2$ respectively. Hence, the Range value for $S1$ is $R1 = (H1 - L1)$ and for $S2$ is $R2 = (H2 - L2)$. The RangeDiff for the slot $S2$ is ($R2$ - $R1$).
- OpenClose: Suppose two consecutive slots: $S1$ and $S2$. Let the Open price of $S2$ be $X2$, and for the Close price of $S1$ be $X1$. The OpenClose for the slot $S2$ is ($X2$ - $X1$).

Output Variable:

Two possible Output variables are defined:

- OpenPerc: Suppose two consecutive slots: $S1$ and $S2$. Let the Open price of the stock for the record of $S1$ be $X1$, and that for $S2$ be $X2$, the OpenPerc for the slot S2 is computed as $(X2 - X1)/X1*100$.
- AveragePerc: Let Average = (Open + Close)/2. Suppose two consecutive slots: $S1$ and $S2$. Let the Average price of the stock for the record of $S1$ be $X1$, and that for $S2$ be $X2$, AveragePerc for slot $S2$ is computed as $(X2 - X1)/X1 \times 100$.

OpenPerc gives early daily information about stock price and can be used for an "instantaneous" early investment decision. AveragePerc is a good option for investment, that takes into account a daily period. One, or both of them, can be utilized depending on the investment strategy.

According to the current theory and practice, Fig. 6 shows the used architecture: four feedforward backpropagation fully connected layers, with sixty-six neurons.

The first hidden layer attempt to produce weights from the activations, to be used for regression. The second and third hidden layer act as pooling layers that perform down sampling along the spatial dimensionality of the given input, further reducing the number of parameters of the phenomena. The linear transfer function on the output layer acts as a regression layer.

Adjust to the training, validation, testing, and complete data are presented in Fig. 7, including R values. The training process uses 70% of the data, the validation uses 15% of the data, and 15 % of the data is taken into account on test calculations. The results show an excellent capability of DNN to represent the phenomena, using the OpenPerc Output variable.

Fig. 6 shows a graphical representation of experimental data, using the values generated by the DNN for variable Open, against real Open values.

Table V shows the data used to predict the variable OpenPerc for the next slot, based on the historical behaviour of stock prices. In other words, if the current time slot is $S1$, the technique will attempt to predict OpenPerc for the next slot $S2$.

Fig 6 : Real and computed open values.

A positive predicted OpenPerc, indicates that there is an expected rise in stock price of $S2$ (UP option), while a negative OpenPerc indicates a fall in stock price for the next slot (DOWN option). An equal value, for both slots, constitutes a STABLE option.

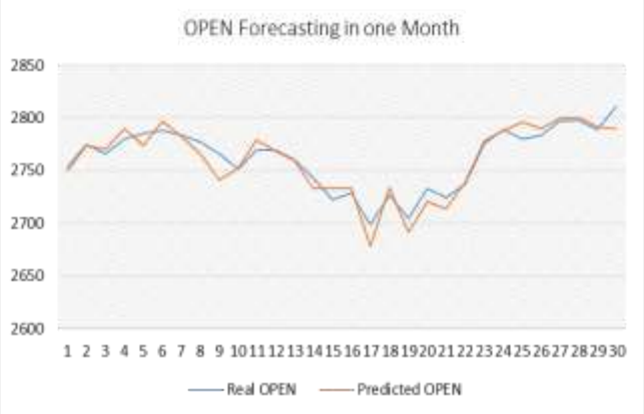

Fig 7 : Adjust of the index open forecasted values.

At the last line of Table V (Case 1259) only the Weekday and therefore the Month day values are modified (Time variables). Other values are identical as the previous day. Then, the value of OpenPerc is obtained simulating the trained DNN. In the showed case the predicted value of OpenPerc is 0.24, with a "UP" behaviour.

Thirty steps (from June 7, 2019, to July 19, 2019) are forecasted Using the previous procedure, as a proof of the adjustment quality. Fig. 7 shows a graph of forecasted and real values of the Open variable, with a correlation of *R*= 0.9322, that is acceptable.

TABLE VI: ESTIMATED AND REAL BEHAVIOUR OF STOCK OPEN VALUES TO THE NEXT DAY

Enterprises	06/06/2019 open real value	06/07/2019 open real value	Open forecasting value	Estimated behavior	Real behaviour
MAR	137,97	141,66	136,31	DOWN	UP
INTC	56,53	56,92	55,52	DOWN	UP
GD	201,83	201,98	201,86	UP	UP
IPG	22,84	23,00	22,81	DOWN	UP
KEY	19,99	20.55	20,38	UP	UP
KO	43,06	43,22	42,77	DOWN	UP
APD	165,16	167,51	166,42	UP	UP
APH	90,00	90,5	92,99	UP	UP
BK	55,8	57,72	56,19	UP	UP
BRK-A	287680	291900	287678,55	UP	UP
UTF	22,68	22,7	22,76	UP	UP
CHK	4,33	4,49	4,36	UP	UP
FDX	253,93	257,3	256,31	UP	UP
F	11,87	11,97	11,93	UP	UP
ECL	144,98	146,53	145,12	UP	UP
PKI	76,73	78,41	77,39	UP	UP
ORCL	47,42	47,93	47,62	UP	UP
SPGI	203,85	206,03	205,78	UP	UP
SHW	390,51	395,37	389,59	UP	DOWN
SCHW	56,63	58,11	56,55	UP	DOWN
RSG	67,92	68,25	67,59	UP	DOWN
MMC	81,3	81,85	82,25	UP	UP
WY	38,16	38,06	38,13	DOWN	DOWN
USB	51,39	51,94	51,95	UP	UP
TXT	68,33	68,88	69,11	UP	UP
ABT	63,11	63,56	61,58	DOWN	UP
PBCT	18,93	19,14	19,03	UP	UP
CSCO	43,76	44,24	44,20	UP	UP
CMCSA	24,8	24,98	24,83	UP	UP
ADP	134,43	136,32	134,12	DOWN	UP

C. Selecting Features of S&P500 Index

The daily values of variable Open from 486 enterprises are processed, to identify the essential features.
With the Index value (Open variable) as label value, the relevant variables (features) are selected, using CFS, and Greedy Stepwise, which implementations and functionality can be revisited in [34,35].
Thirty Enterprises (Attributes) are selected as relevant for the index data tendency: MAR, INTC, GD, IPG, KEY, KO, APD, APH, BK, BRK-A, UTF, CHK, FDX, F, ECL, PKI, ORCL, SPGI, SHW, SCHW, RSG, MMC, WY, USB, TXT,
ABT, PBCT, CSCO, CMCSA, ADP. The Merit, or Pearson's correlation coefficient value, is 0.998.
The computed values of enterprises were obtained training the data of each representative enterprise, with the same model architecture used for the SP&500 Index forecasting. See the important and calculate values of variable Open of thirty representative enterprises for the Case 1259 (Table 6). The value of the Open variable of the day 06/07/2018 is forecasted using the values of variables for day 06/06/2019.

D. The Investment Decision

The investment decision depends on the consideration of the forecasting of the Index value behaviour for the values of variable Open, and the behavior of the group of enterprises identified by feature selection. So, in Case 1259 the Index value has a positive variation (UP with OpenPerc value of 0,24% or 2759,86 for Open variable). Alike, the two-thirds of selected enterprises information show, as the Index, a current similar representative tendency, UP (Table VI).
Then, according to the criteria shown in Table II, to maintain the stocks of the Index, it is a good option as an investment decision, for the next day.

5.DISCUSSION AND CONCLUSIONS

The way toward processing one anticipating esteem is laborious (for this situation, the next day Open value). To start with, it is important to address information and afterward forecast it. As per Prastyo et al. [41], the extrapolated anticipated qualities are progressively less definite, as demonstrated in their research. This reality certifies that the Neural Nets can display strange conduct in extrapolation cases. Here, a heuristic is utilized to control the potential inconsistencies: to find out about a point it is important to know it; at that point, in the learning cycle, just past related information is utilized, and a sliding window strategy is received. This strategy refreshes the information for the estimating cycle, in a bit by bit style (step by step). Also, the information factors of the wonder secure to change the extrapolation into an introduction issue.

The qualities produced by the DNN model are critical, as is demonstrated with the rule of change. As more point by point in time are information, more accurate the anticipating will be, on the grounds that more expectation factors can be characterized; additionally, the information combination empowers other forecast periods (minute, hour, every day, week by week, and so on) Hence, the models utilized in this work are substantial for transient dynamic, that is, step by step.

Moreover, the consequences of this research show, whenever contrasted and unique factors, scarcely any significant undertakings as illustrative of the S&P500 Record. That is, it is feasible to decrease the intricacy and commotion of information and encourage the information examination and choice interaction. The data of undertakings supplements the choice models. In any case, it is important more exact outcome examination to improve the model design and the mix with undertakings information.

Respect to the investigated period; it is crucial for show that it is molding the outcomes legitimacy; for this situation, to the day by day time of the S&P500 List. Further, it is feasible to break down every endeavor or undertaking gathering, utilizing a few securities exchange records.

The model outcomes (portrayal and figure) can be considered as cycle improvement, perhaps to be utilized as a part of a Decision Support System (DSS) for corporate securities. This is a point that requires more itemized hypothetical and experimental examinations.

At long last, this research portrays a hypothetical and a useful instrument for scholastics and professionals. The scholastics can return to another experience of utilizing elective information mining and learning measures. To the specialists, it adds to the development of the current information, concerning the standards to help the stock exchange ventures.

REFERENCES

[1] F. Mostafa, T. Dillon, and E. Chang, "Computational intelligence applications to option pricing, volatility forecasting and value at risk, Studies in Computational Intelligence," *Springer International Publishing*, **vol. 697, 2017.**

[2] Y. Fang, "Feature selection, deep neural network, and trend prediction," *Journal of Shanghai Jiaotong University (Science)*, **vol. 23, no. 2, pp. 297–307, 2018.**

[3] X. Zheng and B. Chen, "Stock market modeling and forecasting," *LNCIS*, Vol. 442, pp. 1–11, London: Springer-Verlag, **2013.**

[4] M. Olden, "Predicting stocks with machine learning," Master's Thesis, **2016.**

[5] X.-Y. Qian and S. Gao, "Financial series prediction: Comparison between precision of time series models and machine learning methods," *Mathematics, Computer Science, Economics*, **2017.**

[6] S. Banik and A. Khan, "Forecasting US NASDAQ stock index values using hybrid forecasting systems," in *Porc. 18th International Conference on Computer and Information Technology (ICCIT)*, **2015 [6].**

[7] Z. Cao, L. Wang, and G. Melo, "Multiple-weight recurrent neural networks," in *Proc. the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, **2017.**

[8] R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools Application*, **vol. 76, pp. 18569–18584, 2017.**

[9] B. Yong, M. Rahim, and A. Abdullah, "A stock market trading system using deep neural network," *AsiaSim 2017*, Part I, Singapore, **2017.**

[10] B. Yang, Z. Gong, and W. Yang, "Stock market index prediction using deep neural network ensemble," in *Proc. the 36th Chinese Control Conference*, Dalian, **2017.**

[11] J. Ma, M. Yu, S. Fong, K. Ono, E. Sage, B. Demchak, R. Sharan, and T. Ideker, "Using deep learning to model the hierarchical structure and function of a cell," *Nature Methods*, **pp. 1-12, 2018.**

[12] P. Addo, D. Guegan and B. Hassani, "Credit risk analysis using machine and deep learning models," *Documents de Travail du Centre d'Economie de la Sorbonne*, **2018.**

[13] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks," *FAU Discussion Papers in Economics*, **2017.**

[14] H. Abdou, "Prediction of financial strength ratings using machine learning and conventional techniques," *Investment Management and Financial Innovation*, **vol. 14, no. 4, pp. 194-211, 2017.**

[15] S. Edet., Recurrent Neural Networks in Forecasting S&P 500 Index, **2017.**

[16] V. Gavrishchaka, Z. Yang, R. Miao, and O. Senyukova, "Advantages of hybrid deep learning frameworks in applications with limited Data," *International Journal of Machine Learning and Computing*, **vol. 8, no. 6, pp. 549-558, 2018.**

[17] V. Sze, Y. Chen, T. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proceedings of the IEEE*, **vol. 105, no. 12, pp. 2296-2329, 2017.**

[18] T. Cook and A. Hall, *Macroeconomic Indicator Forecasting with Deep Neural Networks*, **2017.**

[19] M. Abe and H. Nakayama, *Deep Learning for Forecasting Stock Returns in the Cross-Section*, **2017.**

[20] A. Moghaddama, M. Moghaddamb, and M. Esfandyaric, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science*, **vol. 21, pp. 89–93, 2016.**

[21] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, **vol. 3, pp. 1157-1182, 2003.**

[22] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers and Electrical Engineering*, **vol. 40, pp. 16–28, 2014.**

[23] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics Review*, **vol. 23, no. 19, pp. 2507–2517, 2007.**

[24] J. Miao and L. Niub, "A survey on feature selection," *Information Technology and Quantitative Managemen*, **2016.**

[25] J. Dy and C. Brodley, "Feature selection for unsupervised learning," *Journal of Machine Learning Research*, **vol. 5, pp. 845–889, 2004.**

[26] M. Law, M. Figueiredo, and A. Jain, "Simultaneous feature selection and clustering using mixture models.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **vol. 26, no. 9, pp. 1154-1166, 2004.**

[27] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artificial Intelligence*, **vol. 97, pp. 273-324, 1997.**

[28] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, *Feature Selection for High-Dimensional Data*, Switzerland: Springer International Publishing, **2015.**

[29] H. Abusamra, "A comparative study of feature selection and classification methods for gene expression data of glioma," *Procedia Computer Science*, **vol. 23, pp. 5-14, 2013.**

[30] R. Kaur, M. Sachdeva, and G. Kumar, "Study and comparison of feature selection approaches for intrusion detection," *International Journal of Computer Applications*, **2016.**

[31] E. Mahsereci, S. Ayşe, and T. İbrikçib, "A comparative study on the effect of feature selection on classification accuracy," *Procedia Technology*, **vol. 1, pp. 323-327, 2012.**

[32] W. Puch, E. Goodman, M. Pei, L. Chia-Shun, P. Hovland and R. Enbody, "Further research on feature selection and classification using genetic algorithm," in *Porc. International Conference On Genetic Algorithm*, **1993.**

[33] B. Arguello, "A survey of feature selection methods: Algorithms and software," *Austin*, **2015.**

[34] The University of Waikato, *WEKA Manual for Version 3-7-8*, Hamilton: New Zealand, **2013.**

[35] I. Witten, F. Eibe and M. Hall, "Data Mining: Practical machine learning tools and techniques," *The Morgan Kaufmann Series in Data Management Systems*, **2011.**

[36] E. Barnard and L.Wessels, "Extrapolation and interpolation in neural network classifiers," *IEEE Control Systems Magazine*, **vol. 12, no. 5, pp. 50-53, 1992.**

[37] P. Haley and D. Soloway, "Extrapolation limitations of multilayer feedforward neural networks," in *Proc. International Joint Conference on Neural Networks*, Baltimore, **1992.**

[38] A. Pektas and H. Cigizoglu, "Investigating the extrapolation performance of neural network models in suspended sediment data," *Hydrological Sciences Journal*, **vol. 62, no. 10, pp.1694-1703, 2017.**

[39] P. Hettiarachchi, M. Hall and A. Minns, "The extrapolation of artificial neural networks for the modeling of rainfall-runoff relationships," *Journal of Hydroinformatics*, **vol. 07, no. 4, pp. 291-296, 2005.**

[40] MathWorks, *Neural Network Toolbox™. User's Guide. R2014a*, The MathWorks, Inc., **2014.**

[41] A. Prastyo, D. Junaedi, and M. Sulistiyo, "Stock price forecasting using artificial neural network," in *Proc. Fifth International Conference on Information and Communication Technology (ICoICT)*, **2017.**

[42] MH. Najeb, Masoud, "The impact of stock market performance upon economic growth." *International Journal of Economics and Financial Issues* **3 (4) : 788–798, 2017.**

[43] Murkute, Amod, and Tanuja Sarode, "Forecasting market price of stock using artificial neural network." *International Journal of Computer Applications* **124 (12) : 11-15, 2015.**

[44] C. Montenegro and M. Malina, "Improving the Criteria of the Investment on Stock Market Using Data Mining Techniques: The Case of S&P500 Index," International Journal of Machine Learning and Computing, **Vol. 10, No. 2, 2020.**