

Recommending System for Penny Stock Trading

S Shabana Begum, N. Kasiviswanath,



Abstract: Penny stocks at times makes the investors wealthy by turning to be a multi-bagger stocks or erode the wealth of the investors with poor performance in volatile conditions. While there are many machine learning-based prediction models that are used for stock price evaluation, very few studies have focused on the dynamics to be considered in penny stock conditions. Though the pattern might remain the same for normal stocks and the penny stock classification, still some of the parameters to be evaluated in the process needs changes. The model discussed in this report is a comprehensive solution discussed as scope for evaluation of the penny stock pick, using trading and reporting financial metrics. Experimental study of the test data indicates that the model is potential and if can be used effectively with reinforcement learning pattern, it can turn to be sustainable solution.

Key words: penny stock prediction, ML based penny stock prediction, Stock analysis using machine learning.

I. INTRODUCTION

Stock market holds critical importance in the economic development of the respective countries and for the global economic conditions. It is imperative factor from the history that, the up's and down's in one stock exchange market conditions have direct impact on the other markets bullish or bearish levels of the other global stock exchanges too. The fact in the stock market remains that the prediction of the stock pricing movements both in intraday and equity conditions has short term and long term bearing over the conditions that prevail for the developments in the stock market and sequential impact on the profits or losses that are envisaged by the investors the market. In the stock market domain, one of the key areas wherein the studies and researches were taking place consistently is about the penny stocks related developments. Profoundly small-scale investors and some of the tactical investments from the stock trading agencies, long term investors focus on the penny stock markets. While it is evident that in the case of certain penny stocks the impact has been a kind of multi-bagger opportunity, in many of the penny stock solutions, the investor's wealth was eroded in the market fluctuations. Such volatile conditions in the stock market environments lead to the fact that the stocks are to be more focused and emphatically addressed and it is paramount importance that the penny stock movement must be tactically reviewed and considered for its up/down movement for investments [1].

In the other dimension, the role of technologies like the predictive analytics in the stock markets prediction is gaining momentum and there are numerous research studies that have focused in developing predictive models that can support in improving the accuracy of penny stocks prediction. Though majority of the models have focused on the predictive analytics solutions for all categories of stock movements, some of the studies were being more specific about focusing on the penny stock related predictive modeling solutions.

The volatile is of the penny stocks leads to conditions wherein majority of the investors stay away from the facets of penny stocks, but at times the penny stocks can become multi-baggers. For instance, the stocks of Ford Motor Company, American Airlines, and Titan etc. have started off as a penny stocks but over period they have given substantial returns with more than 100X returns, which has created significant wealth for the investors in such stocks.

The trend factors of investments in the penny stocks has both success and the failure stories in proportion, which indicates the fact about if right kind of stocks are picked and managed for sustainable period, there are potential chances to ensure right kind of outcome from the process [2].

1.1 Scope of Penny Stocks

Penny stocks if are chosen rightly can be a profit bagger for the investors. However, there are various factors that are integral to handling the process of predicting the penny stocks. Some of the machine learning models that are discussed in the literature turns promising solutions towards specific kind of penny stock related analysis.

However, in general, some of the key factors to be taken in to account in the process of handling the penny stock requirement could be attributed to some of the following factors.

The scope of bagging profits in the case of the penny stocks is relatively about choosing right kind of penny stocks while there are tens of such stocks that are available in the respective stock exchange [3]. The other dimension about understands the time frame where in the stock must be hold by an investor, to reap maximum benefit from the respective stock. For instance, some of the stocks might be having a steady progress for longer periods of time and in some stocks, there is a significant volatile condition, wherein the investors pitching-in and heading out of a specific stock must be more timed. Taking such factors in to account, the classification for prediction requirement could be stated as

- Selection of a stock as a right pick
- Timing factor detection as to whether the specific penny stock must be short timed or long-term hold
- Right pricing at which a stock could give maximum returns.

Manuscript published on 30 December 2019.

* Correspondence Author (s)

S Shabana Begum, Research Scholar, Rayalaseema University, Kurnool, Andhra Pradesh.

Email: shabanasphd@gmail.com.

Dr.N. Kasiviswanath, Professor & Head, Dept. of Computer Science and Engineering,

G. Pulla Reddy Engineering College, Kurnool, Andhra Pradesh.

Email: csehod@gpec.ac.in.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The afore mentioned three factors are the key areas wherein the role of predictive analytics and machine learning models can be more significant in the penny stock models empowering conditions.

However, taking in to account the conditions that are limited to the scope of research, in this manuscript, the first objective of choosing a penny stock as right pick in terms of “yes or no” is tested for a machine learning model.

In the further sections of this report, the emphasis is on some of the earlier studies that have discussed the usage of predictive learning models and its relative impact is reviewed. Followed by, in section three, the proposed model, selection of datasets, training model classifiers used for the machine learning solution, the proposed algorithm, experimental results of the model are discussed.

II. RELATED WORK

In [4], a contemporary model of deep learning solution was introduced as RSR (Relational Stock Ranking) in the case of the stock prediction. The two key aspects that are integral to the plans are about customizing the deep learning models for enhancing the stock ranking and towards capturing the stock relations across the time-sensitive conditions.

Novelty of the model is about a niche component of neural network-based modeling wherein the Temporal Graph Convolution is discussed as a joint patter, wherein the assessment of a stock is about temporal evaluation and the relational impact with the similar kind of stocks. Experimental analysis of the model was carried out using the back-testing of historical data across two stock markets NYSE and NSADAQ, with results of 98% on the NYSE and 71% average return ratio for NASDAQ solutions.

In a distinct kind of study, author of [5], discusses about the predicament with the AI based stock market analysis conditions. The key argument presented in the study is that any kind of wrong guidance from the system towards a pricing time of a stock or entry/exit for a stock could be more treacherous and could have adverse impact conditions. The other major argument presented in the study is about the certainty that not all the algorithms can be 100% accurate and even in terms of 90% accuracy, still the 10% inaccuracy results from bouts of sudden changes impacting the trading sessions. The imperative statement from the author is about how the ML models has the limitation in terms of same sum being applied for every set of input variable (variables being price high, lows and closing price etc.) wherein none of them reflects sudden change till the real-instance and only small amounts of movement estimations can be predicted based on earlier fluctuations.

The authors of the study [5] conclude that while some of the complex machine learning models might be effective in sensing the emerging changes of interest rates depending on the analysis of news, prices etc. or any other industry related insights, getting to more accurate predictions could be a challenging affair, and in the instances of such predictions going awry even by 10% margin could be resulting in significant loss to the investors.

In [6] the authors discuss the scope of deep learning usage in the trading algorithms for stock price movement prediction accuracy. There is considerable gap in

the real-time deployment of the breakthrough. For instance, in the case of the professional traders across long-term careers, there are sizeable accumulated practices in terms of human interpretation of the conditions. Whereas, in the case of the deep learning models, the emphasis is highly interpretable. The model has presented the scope of a text-based deep learning models, wherein the end users have the opportunity for interpretation of the stock price movement, classification of predictable and unpredictable metrics, the scope of integrated visualization systems being integral to the prediction model of a stock price depending on the news emerging from the different platforms like the social media.

In [7], the researchers have developed trading strategies using the application of machine learning models for technical indicators based on the conditions of high frequency stock data conditions. The results in the case of the system is about usage of automated trading system solution application for a stock generating significant returns which are higher than the market returns without considering significant rise in the volatile conditions. The authors of the study claim that the proposed model can be more effective for algorithmic trading which can be used by the individuals or even the retail traders having the elementary level of technical skills.

In [8], the study has focused on using the dispersed information which is processed in instantaneous manner and is profoundly used for making price adjustment. The proposed solution takes fundamentally the embedded information in textual news as to be inferred by the investors. In the recent times, the researchers started to automatically develop news sentiment for managing the stock price movements. The model proposed is about usage of the supervised and reinforcement-based learning algorithms, wherein the news-based data is used as feed for the decision-making process of the investment system.

In [9], the model of using the predictive machine learning for analysis of the financial news articles in terms of distinct set of textual representations are evaluated for its relative impact the movement of the stock prices for the information discussed in the financial news articles. Based on the insights from the experimental analysis using the set of financial articles and the related stock movement, the inferences in the study are made by the model.

The other contribution “Temporal Relational Ranking for Stock Prediction (TRRSP)” [10] is a contemporary contribution of the recent past, which has intended to resolve the constraints observed in other contemporary models. The method TRRSP is a machine learning strategy that considering the temporal validity of the features in given training corpus. However, the method has considering the ranking system to select the optimal features. These features are further used to perform deep learning. However, the method TRRSP is evincing the process overhead due to the process of using deep learning strategy. Due to the sparseness of the dimensionality in features used in learning phase, the scope of deep learning is least significant [11]. Moreover, the temporal validity is not significant in penny stock trading [12].

In summarization of the inputs garnered from the studies, it is evident that the scope of using the machine learning and predictive models are being tried in the domain of stock trading and some of the studies in the domain are also focusing in-depth on the conditions of such solutions in the penny stocks movement too.

III. PENNY STOCK PREDICTION SYSTEM

1.2 Gap Analysis

The inputs garnered from the model reflects positive factors in the form of successful application of the predictive modeling and machine learning solutions for the stock market price movement. In the other dimension, the gap that remains as emphasis of issues are about if majority of the machine learning solutions think alike for prediction, then the problem exists in the form of unique pattern in the model of analysis which shall be used by multiple trading firms. But in a realistic term, the possibilities of distinct areas of expertise from the people involving in the process might have varying results of trade analysis. This is to be essentially integrated to the system.

Secondarily, the scope of relying on thousands of traders making distinct kind of decisions will vary with higher frequency range of decisions. But if the AI models are deployed, there is a specific fixed pattern in which the model might approach the decision-making process, thus resulting in higher percentage of decision going to one specific condition, upon using the machine learning models [13].

Traditional models of stock prediction using the time-series analysis models as proposed in Kalman filters auto regressive models and its extensions as discussed in reflect that when a specific kind of models reflects on the stochastic process in the stock prediction, there are three key blocks that impact the outcome. Firstly, high reliance on the indicators which are manual inputs, secondarily the issue of hypothesized process which might not fit the volatile stock predictions in the real-time and thirdly the issue of considering very few indicators for analysis. Some of the earlier models has addressed such complexities and has provided scope of improved solutions.

Considering the afore mentioned problem conditions, in the further section, the emphasis is about developing a comprehensive system for practice, which can lead to potential ways in which the penny stock conditions are evaluated [14].

1.3 Proposed Solution Framework

Selecting the penny stocks has impact of positive or negative outcome depending on the various aspects that are integral to the process of choosing the right kind of stock for the conditions. As discussed in the section-1, while there are three key dimensions that are important to decide on the scope with a particular set of penny stock model, still in this manuscript, the framework solution is about choosing a penny stock for a pick as a recommended or non-recommended solution.

For the proposed solution, the scope is about a simple framework as discussed below.

When a penny stock must be evaluated, there are distinct parameters in which the stock must be evaluated. Some of the most prevalently used computing models for analysis of a stock are closing price, price range, historical performance, historical movement pattern, market capitalization, futuristic inputs like the financial news related to the company etc.

Using the distinct set of predictive models, in many of the earlier solutions, using several of the above-mentioned metrics, some of the commonly used models of prediction are moving average conditions, regression analysis, ARIMA, PROPHET, and KNN.

In the study [13], the author has discussed the scope, implementation algorithm and the predicted outcome for each of the above-mentioned model. Taking the inputs discussed in the models as the fundamental for application in to the system, the system proposed is to focus on some of the critical metrics that are integral to the process of evaluating the viability of the proposed penny stock model using the time-series analysis.

Considering such system insights in to account, for the proposed solution, the emphasis is on developing a system which takes certain metrics in to account and accordingly finalize on some of the key aspects that are considered integral to the recommendation or non-recommendation of a penny stock.

1.4 Metrics Considered for ML Model

Some of the key components that are considered as variables or metrics for the model assessment are some of the key financial ratios and other metrics considered integral to the system development.

The financial metrics chosen as one group of parameters are in Table 1.

Table 1: The financial metrics group of parameters

Ratio Segment	Metrics of Ratios	Metric Label
Liquidity Ratio	Current Ratio	R1
	Quick Ratio	R2
	Cash Ratio	R3
Leverage Ratio	Operating Cash Flow Ratio	R4
	Debt Ratio	R5
Performance Ratio	Interest Coverage Ratio	R6
	Price to Earnings Ratio	R7
	Price to sales ratio	R8
	Price to Cashflow Ratio	R9

In the other dimension, the factors of trade parameters of the following are chosen as the other set of group metrics in Table 2.

Table 2: the factors of trade parameters

Variable	Metric Label
Market Capitalization	T1
Trade Volume Rating on the analysis day	T2
Trade volume rating average in last 6 months	T3
Trade price range volume on the analysis time	T4
Weighted moving average	T5
Market Sentiment rating	T6
Consistency score	T7

In the final evaluation of the scoring pattern, there is equal weightage for both the conditions and the pattern scores higher in terms of outcome based on the time-series data, the recommendation towards recommended or non-recommended pick for a product shall not be categorized.

IV. CLASSIFIER USAGE

The proposed machine learning model is trained for the datasets using the NB (Naïve Bayes) classifier, which is seen as generative learning model solution. It is a kind of classification technique that relies on the Bayes Theorem wherein it is perceived that there is independence among the predictors. In the other way, it can be stated that the NB classifier presumes the presence of a specific kind of feature in the class as unrelated for the presence of any other kind of features. Despite that the features could have some dependency on the each other, still in terms of existence; all the properties have independent contribution towards the probable outcome. NB classifiers are model which can be effective for large datasets and is known to outperform many of the complex kind of classification methods too. The algorithm used for the proposed model has explored in following description

Let S be the penny stock Indicator // S is to be reviewed as eligible penny stock based for further level recommendation analysis

The notation C_p denotes the current price of a stock

The notation F_p denotes the fixed slab price to consider a stock as fixed price

If $C_p < F_p$, then the stock S is considered as the penny stock and further evaluation is triggered

Let Q be the selection score used for rating and classification of a stock as recommended or non-recommended using the two subcategories for evaluation as

Rn – Features of Ratio Parameters (R1-R9)

Tn – Features of Trading Parameters (T1-T7)

{
{
 P is Correlation Co-efficiency P-value of all the parameters together

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x^2)][n\sum y^2 - (\sum y^2)]}}$$

The correlation values are analyzed to assess the significance and relativity among the patterns considered in the labeled parameters.

Whereas, individually all the sub-score (Rn_{1-5}) are significant for the system and the value range is

$$Rn = \geq 1 \leq 90 // \text{All the five parameters given scoring range of 1-10}$$

If the $Rn \geq 60$, the case is considered for “Highly Recommended” rated as HR

If the $Rn \geq 40 \leq 59$, the case is considered for “Moderately Recommended” rated as MR

If the $Rn \geq 19 \leq 39$, the case is considered for “Low Ranked” rated as LR

If the $Rn \geq 1 \leq 18$, the case is considered for “Not Recommended” denoted as NR }

{
The Trade score ranking category is considered as Tn defined with seven parameters considered for the stream of scoring range 1-10

$$Tn = \geq 1 \leq 70$$

$$Tn = (\sum (Tn1, \dots, Tn5))$$

If the $Tn \geq 50$, the case is considered for “Highly Recommended” rated as W

If the $Tn \geq 30 \leq 49$, the case is considered for “Moderately Recommended” rated as X

If the $Tn \geq 19 \leq 29$, the case is considered for “Low Ranked” rated as Y

If the $Tn \geq 1 \leq 18$, the case is considered for “Not Recommended” denoted as Z }

$$Tn = \text{either } W \text{ or } X \text{ or } Y \text{ or } Z$$

}



When the input “Q” is assessed for R and T score value allocation, the process followed is about

{

Eligibility score

If the Rn has the label of HR then only TW must be used for analysis

If the Rn has the label of MR ...then Tx or Tw or Ty can be considered for allocation accuracy prediction

If the Rn has a score of LR , only Tx or Ty shall be review

If the Rn has a score of NR , Tx or Ty shall be chosen for review

}

{

If the Q score validation is effective, then the scoring output for assessing the stock as a valued for recommended or non-recommended is estimated.

$$Q = 2 * ((Rn * T) / Rn + T)$$

}

If the $Q \geq 27$, the Task stock is recommended as highly recommended

If the $Q \geq 20 \leq 26$, the Task is rated as Moderate Recommended

If the $Q \geq 15 \leq 19$, the task is rated as Low Recommendation

If the $Q \geq 1 \leq 18$, the task is rated as Non recommendation

}

Some of the integral factors for the proposed algorithm are

The T1-T7 and R1-R7 rating must be initially human intervention during the training sets, wherein the corresponding inputs like the range for classification as scoring pattern is to be provided to the system.

As a loop, the information is not discussed in this algorithm, as the discretion of the trading or analyzing agencies could have different rating patterns. The pattern is adapted for keeping the loop available for defining the scoring pattern is considering the gaps discussed in the section of gap analysis, as to the training patterns alone can define the accuracy rate and in the absence for such feasibility all the models shall act similar.

1.5 Data Sets

For the experimental study of the proposed study, the stocks categorized as penny stocks in the list of economic time’s financial news articles are chosen from the period of last one year, based on the stocks listed in both NSE and BSE. In a cumulative approach, total of 60 stocks are chosen at different rates to be recommended or non-recommended category. In the recommended category, the classification is of three sections, whereas in the non-recommended category the classification is only one segment in Table 3.

Table 3: the classification among the various parameters in terms of classification

Total number of records	100000
Total Positives	56000
Total negatives	44000
positives for training	42000
negatives for training	33000
positives for testing	14000
negatives for testing	11000

1.6 Test Results

The test results that are integral to the process are evaluated based on some of the key factors considered integral to the system. Upon the testing of the data chosen for test, the following are some of the significant levels of results that are integral from the process.

Table 4: The detection from test records, Metrics and Proportion

Metrics	Detection from Test Records	Proportion
True positive	226	94.17
False Positive	14	5.83
False Negative	3	5.00
True Negative	37	61.67

The Table 3 indicates the classification among the various parameters in terms of classification which is envisaged as the process integral to the test outcome envisaged from the proposed machine learning model. In the following Table 4, Table 5, the scope of key metrics integral to the process and the related dynamics are discussed in significant manner.

Table 5: The performance and metrics

Metrics	Performance
Precision	94.17
Recall	98.69
Accuracy	93.93
F1 Score	96.38
Specificity	72.55

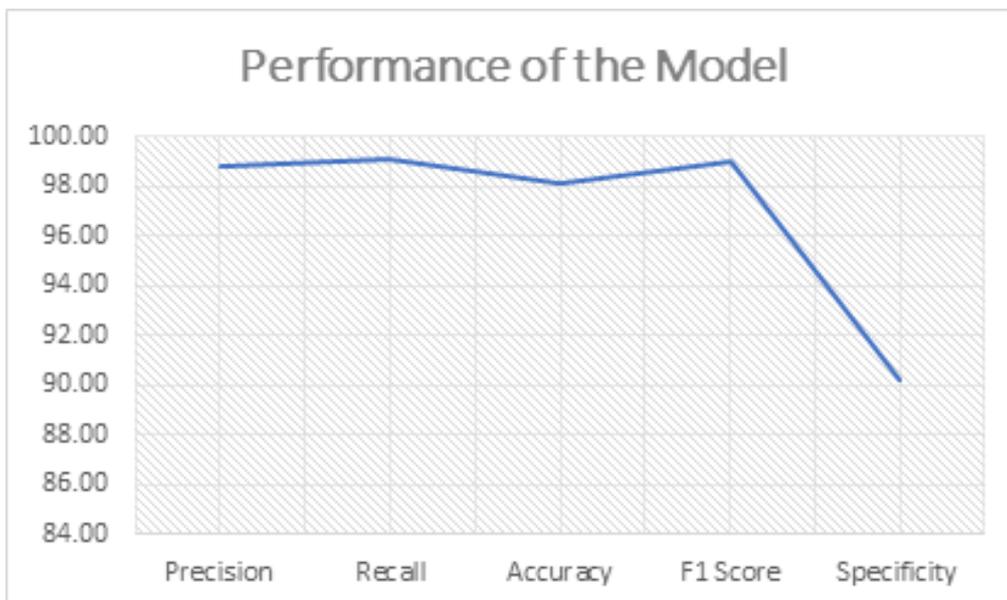


Figure 1: Performance of the Model

Tabulated information reflects that the proposed model has considerable performance and still there is more effective ways in which the accuracy of the proposed application must be improved. However, considering the quantum of gaps in the system and the need for more effective kind of analysis patterns over the real-time, it can be stated that if there is combinational approach followed in the process, it can lead to more effective outcome.

V. EXPERIMENTAL STUDY

In the experimental study, database is segregated into small clusters in 4-fold approach. The effectiveness & efficiency of this mechanism by comparing with other projected mechanisms is defined by divergent classification standards that comprise precision, sensitivity, specificity, accuracy, F-measure, Matthews Correlation Coefficient metrics. The comparative outcomes of this experimental stage are represented.

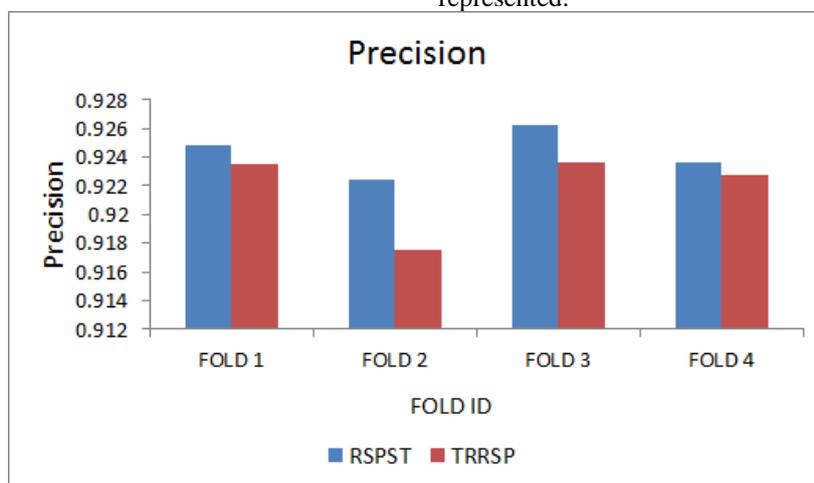


Figure 1: The average precision for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The metric precision is defined as relevant instances fraction among the instances that are retrieved. In Figure 1, the graph is plotted among precision and four folds for the proposed RSPST method and contemporary TRRSP method. The average standard deviation (ASD) of precision for the

RSPST model is (0.9243 ± 0.0015) and for contemporary TRRSP model [10] is (0.92185 ± 0.0026) . From the statistics, it is evinced that the proposed model RSPST model performs better when compared to other contemporary TRRSP model.

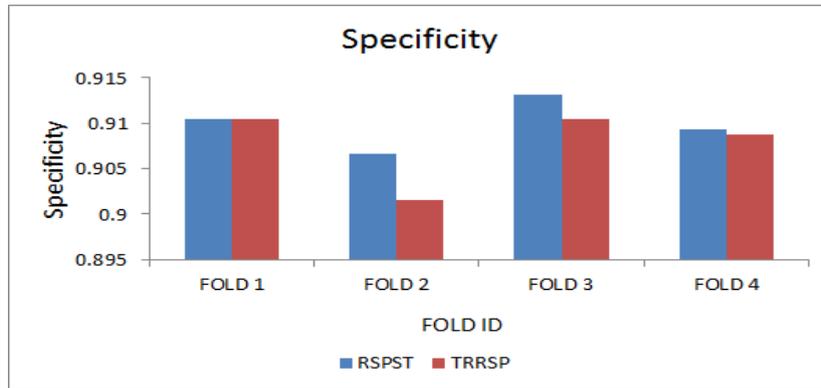


Figure 2: The average specificity for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The metric specificity is also called true negative rate (TNR), which measures ratio of the actual negatives that identified correctly. In Figure 2, the graph is plotted among Specificity and four folds for the proposed RSPST method and contemporary TRRSP method. The ASD of specificity

for the RSPST model is (0.909875 ± 0.0024) and for contemporary TRRSP model is (0.9078 ± 0.0037) . It is exhibited from the statistics that, the proposed model RSPST model performs superior when compared to other contemporary TRRSP model.

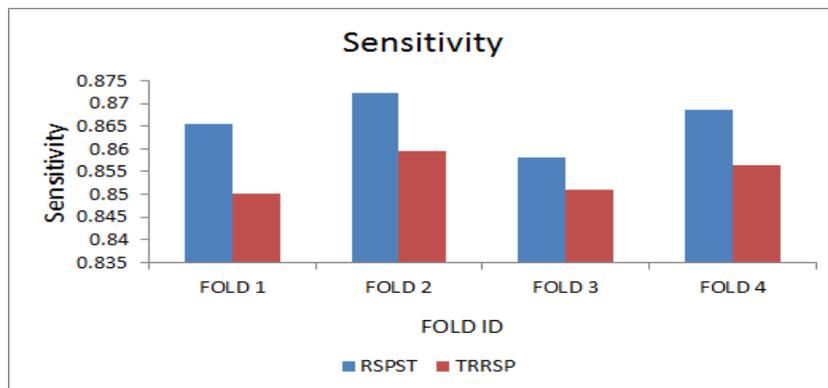


Figure 3: The average sensitivity for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The term recall is also known as sensitivity that measures ratio of the actual positive instances, which is estimated to be positive. The graph is plotted among Sensitivity and four folds for the proposed RSPST method and contemporary TRRSP method in Figure 3. The ASD of sensitivity for

RSPST model is (0.86615 ± 0.0053) and for contemporary TRRSP model is (0.854275 ± 0.0038) . It is exhibited from the statistics that, the proposed model RSPST model is more advantage when compared to other contemporary TRRSP model.

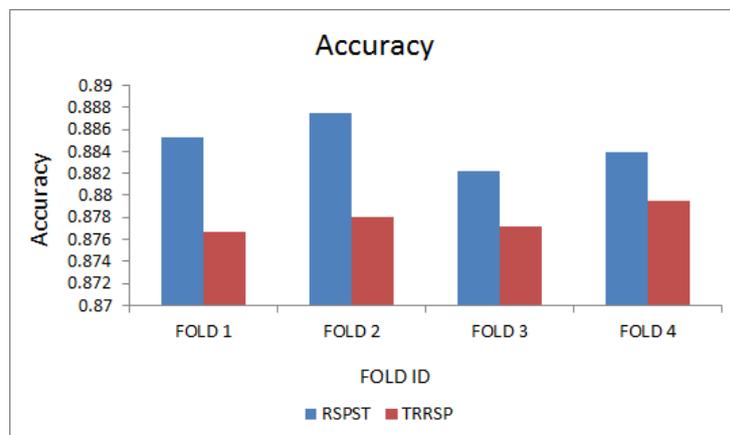


Figure 4: The average accuracy for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The metric accuracy is one of the evaluating classification methods. It is defined as ratio of number of correct estimations to the total amount of estimations. In Figure 4, the graph is plotted among Accuracy and four folds for the proposed RSPST method and contemporary TRRSP method. The average standard deviation (ASD) of accuracy

for the RSPST model is (0.884725 ± 0.00194) and for contemporary TRRSP model is (0.87785 ± 0.00105) . From the statistics, it is evinced that the proposed model RSPST model performs better when compared to other contemporary TRRSP model.

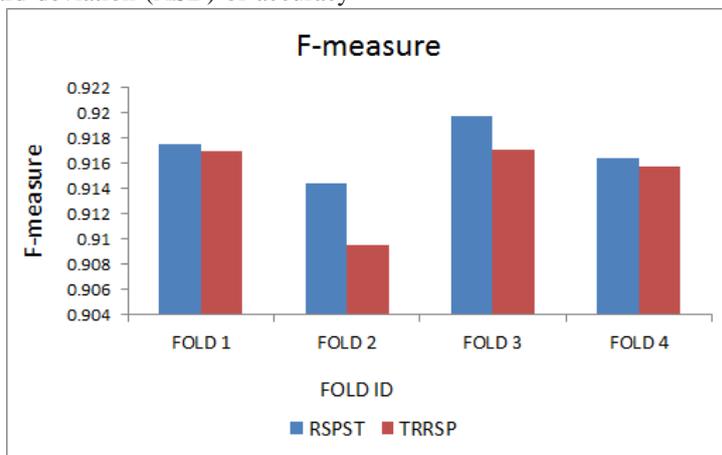


Figure 5: The average F-measure for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The metric F-measure is defined as evaluation of test's accuracy and could also be defined as weighted-harmonic-mean of recall & precision of test. The graph is plotted among F-measure and four folds for the proposed RSPST method and contemporary TRRSP method in Figure 5. The

ASD of F-measure for RSPST model is (0.917 ± 0.00191) and for contemporary TRRSP model is (0.914775 ± 0.00308) . It is exhibited from the statistics that, the proposed model RSPST model is more advantage when compared to other contemporary TRRSP model.

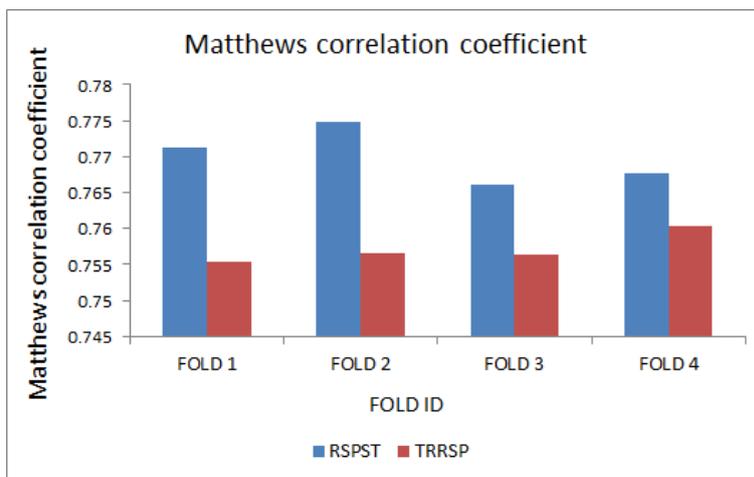


Figure 6: The average Matthews correlation coefficient for the proposed RSPST model and contemporary TRRSP model are observed from 4-fold classification

The metric Matthews correlation coefficient is used. The graph is plotted among Matthews's correlation coefficient and four folds for the proposed RSPST method and contemporary TRRSP method in Figure 6. The ASD of Matthews correlation coefficient for RSPST model is (0.76992 ± 0.00336) and for contemporary TRRSP model is (0.75712 ± 0.00187) . It is exhibited from the statistics that, the proposed model RSPST model is more advantage when compared to other contemporary TRRSP model.

towards understanding how the proposed model can be significant in the selection process. The experimental study of the model using the datasets of penny stocks indicate that the model has potential scope of accuracy in the prediction and can be reliable solution with more enforced learning metrics. However, to keep the accuracy to be more suitable for the pragmatic conditions of the market, in defining the scoring patterns, the values shall be decided by the analysts to rate as good or moderate scores. This will enable flexibility in ranking the stocks and towards estimation for enhanced accuracy.

VI. CONCLUSION

The model discussed in this manuscript is about predicting the selection metric of a penny stock in the market, and



As a scope of future model, if the solutions can be interpreted using the combination of three dynamics like the selection of a penny stock, timing and the pricing conditions, it can result as a more comprehensive solution.

REFERENCES

1. Feuerriegel, Stefan, and Helmut Prendinger. "News-based trading strategies." *Decision Support Systems* 90 (2016): 65-74.
2. Shen, Shunrong, Haomiao Jiang, and Tongda Zhang. "Stock market forecasting using machine learning algorithms." *Department of Electrical Engineering, Stanford University, Stanford, CA* (2012): 1-5.
3. Zheng, Alice, and Jack Jin. *Using AI to Make Predictions on Stock Market*. Stanford University, Tech. Rep, 2017.
4. Feng, Fuli, et al. "Temporal relational ranking for stock prediction." *ACM Transactions on Information Systems (TOIS)* 37.2 (2019): 27.
5. <https://blog.usejournal.com/why-stock-predicting-ai-will-never-take-over-the-world-b1b411decc21>
6. Shi, Lei, et al. "DeepClue: visual interpretation of text-based deep stock prediction." *IEEE Transactions on Knowledge and Data Engineering* 31.6 (2018): 1094-1108.
7. Obi Bertrand Obi. "Application of Machine Learning in High Frequency Trading of Stocks." *International Journal of Scientific & Engineering Research* Volume 10, Issue 5, May-2019, ISSN 2229-5518,1592-1603.
8. Feuerriegel, Stefan, and Dirk Neumann. "Evaluation of news-based trading strategies." *International Workshop on Enterprise Applications and Services in the Finance Industry*. Springer, Cham, 2014.
9. Schumaker, Robert P., and Hsinchun Chen. "Textual analysis of stock market prediction using breaking financial news: The AZFin text system." *ACM Transactions on Information Systems (TOIS)* 27.2 (2009): 12.
10. Feng, Fuli, et al. "Temporal relational ranking for stock prediction." *ACM Transactions on Information Systems (TOIS)* 37.2 (2019): 27.
11. Yang, Yaodong, et al. "Can Deep Learning Predict Risky Retail Investors? A Case Study in Financial Risk Behavior Forecasting." *arXiv preprint arXiv:1812.06175* (2018).
12. Tsuruta, Tokimaru, and Tomoya Suzuki. "Technical Trading Strategy Using Reactions to Stock Price Jumps." *International Symposium on Nonlinear Theory and its Applications*. Vol. 11. 2016.
13. Haliassos, Michael, and Carol C. Bertaut. "Why do so few hold stocks?." *the economic Journal* 105.432 (1995): 1110-1129.
14. Kim, Raehyun, et al. "HATS: A Hierarchical Graph Attention Network for Stock Movement Prediction." *arXiv preprint arXiv:1908.07999* (2019).