



## An Extensive Work on Stock Price Prediction Using Ant Colony Optimization Algorithm (ACO-SPP)

**Saravanan Ramalingam<sup>1\*</sup>      Pothula Sujatha<sup>1</sup>**

<sup>1</sup> *Department of Computer Science, Pondicherry Univeristy, Puducherry, India*

\* Corresponding author's Email: r.saravanan26@gmail.com

**Abstract:** Stock Price Prediction (SPP) is a potential fiscally rewarding arena as it involves several dependent factors and complex structure. As the stock price fluctuates during every business day, it is nontrivial to forecast the stock prices by the buyers and sellers. Numerous methods have been developed for SPP based on Machine Learning (ML) techniques. In this work, an Ant Colony Optimization (ACO) based SPP model is introduced to forecast the stock prices accurately. The efficiency of the proposed ACO-SPP model is analyzed by comparing the performance with a set of traditional ML based classifiers such as Naive Bayes (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Radial Basis Function (RBF), J48, RF, Classification and Regression Tree (CART) and Olex-GA. To emphasize the benefits of ACO-SPP model, it is applied to a benchmarked dataset named Dow Jones Index (DJI) dataset and three datasets from Yahoo finance on daily, weekly and monthly basis. The experimental results infer that the ACO-SPP accurately predicts stock prices than the compared methods in terms of accuracy, F-score, AUC, Discriminant Power and Youden's Index.

**Keywords:** Ant colony optimization, Ant-miner, Dow Jones Index dataset, Machine learning, Stock price prediction.

### 1. Introduction

Recent trends in research contend that one needs to look at financial market data on stock indices from an evolutionary point of view due to their divergent nature. Stock is a crucial investment in financial sectors and presently it seeks much attention among the research community. Investors and research people are considering multidimensional factors to suggest a methodology that will choose the stock precisely with the guaranteed profit in return. Due to the diverging nature of SPP, it is diligent to predict stock prices each and every business day. An accurate SPP can lead the investors to gain higher profit while exchanging stocks. Hence unambiguously, SPP has become more popular and highly helpful for investors [1]. A study [2] revealed that a proper SPP can lead the trading to a more profitable way and it also recommended that researchers should concentrate on precisely identifying the direction of

movement and to reduce the variations from the real observed values.

The precise SPP will be a huge benefit to the investors [3]. But, the performance of stock exchange is based on several qualitative factors like politics, economics, weather, etc. The stock markets rapidly changing and they show more variation. This nonlinear and complex nature of stock market makes SPP more difficult [4-5]. The prediction of future stock prices is defined by factors like data intensity, noise, non-stationary, random nature, uncertainties and invisible interlinks [6-8]. Generally, there are two classifications of SPP which includes the SPP movement and predicting the value of the stock price. The first type is considered as a classification problem and the other is treated as a regression problem. But, both depend on auto regression and multi-variable regression models. The auto regression model treats the problem based on time sequences and the sequences are partitioned to various segments. The partitioned segments are employed as raw data for SPP. In multi-variable

regression model, the technical variables are chosen as raw data for SPP.

In recent years, many methods have been proposed to predict the trends in stock price variations. Prediction of upcoming stock price using the existing financial information is highly valuable for investors. Investors like to predict the increasing or decreasing trends in stock price over a period of time. Prior to making an investment decision in a firm, a prediction method was used to foresee stock prices using the existing and present financial data about the firm. Financial balance sheets and several ratios represent the status of company and it is the fundamental for technical investigation. The experts applied various mathematical models using the previous data to validate company's intrinsic value like Graham number. Graham number and Graham's criterion is assumed as a popular method of SPP [9]. Because of increasing instability in the present market scenario, it is very hard to identify a company which satisfies Graham's principles on current situation. Due to these modifications, the requirement for adjusted models started. In addition, the stock market trends change over time [10-11].

Latest investment models and techniques are mushrooming which eliminates the existing models outdated. The application of algorithms in trading surely modifies the perspective of stock exchange which made prediction easier and react faster to some events on stock market. ML techniques make experts to design models for SPP using previous information. In the last two decades, several methodologies like Genetic Algorithms (GAs), Neural Networks (NN) and other artificial intelligence techniques [12-14] are used for predicting stock price. In previous works, GA based SPP models are developed on stock market index. GA is used for rule classification and also GA hybrid approaches are used for prediction [15]. Several researches have been carried out by the use of evolutionary algorithms to do black-box investing.

Though GA based approaches are employed in SPP previously, it fails to achieve efficient results and also rate of convergence is less. ACO is also a popular technique used for classification problem. ACO is introduced in [16], which is inspired from the foraging nature of ants. Ant miner [17] is the ACO based algorithm used for rule discovery. This paper proposes an Ant Colony Optimization (ACO) for classification purposes in SPP. ACO-SPP technique is introduced to effectively forecast the stock prices. In the context of rule discovery, ACO algorithm has the ability to perform a flexible robust search for a good combination of terms (logical conditions) involving values of the predictor

attributes. Hence, the ACO algorithm can be used for rule discovery in SPP.

In the field of data mining, a classification algorithm named ant miner algorithm is used, based on the principle of ACO. The contribution of the paper is as follows:

- Implementation of novel hybrid ACO-SPP model to make SPP accurate and the rate of convergence of results faster
- The above said algorithm is applied on a benchmarked dataset Dow Jones and datasets collected from Yahoo finance on the basis of daily, weekly and monthly
- The performance of the ACO-SPP algorithm is compared with Naive Bayes (NB), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Radial Basis Function (RBF), J48, RF, Classification and Regression Tree (CART), Olex-GA and recently proposed XGBoost classifiers. A detailed analysis is performed to prove the efficacy of the proposed algorithm. The Performance measures used are Accuracy, F-score, Area Under the Curve (AUC), Discriminant Power and Youden's Index

The remaining section of the paper is arranged as follows. Section 2 briefly outlines the earlier works on ACO and SPP. Section 3 elucidates the proposed ACO-SPP method in detail. Section 4 presents the performance analysis of the ACO-SPP model and the Section 5 depicts the conclusion.

## 2. Related works

In this section, the applications of ACO in different domains are discussed especially in classification problems. In addition, the existing SPP models are also explained. The previous works on SPP are discussed with the aim, methodology, performance measures, benefits and some of the demerits. Several SPP techniques have been proposed based on mathematical frameworks, ML methods and bio-inspired algorithms. Here, the existing works related to the proposed method are discussed in brief.

Alrasheedi et al. [18] employ various classifiers such as LR, MLP, LDA, SVM, LS-SVM for SPP in Saudi stock exchange from the period of 2006-2013. Dow Jones dataset was used with five-fold cross validation. On comparing the hit rates of LDA, LR and SVM, SVM give similar results for the applied dataset. Though it achieves better results than MLP and LS-SVM, it performs well when same economic situation is encountered and it lacks reliability.

Nikola Milosevic et al. [19] presented a ML based method to analyze the equity's future price for

a longer duration. This model accurately identifies the rise of company value by 10% in a year, the main aim of using ML algorithms is to train the previous data, it is likely to foresee the stock price and also the finds the ratio of movement for a particular time period. They compared different classifiers for SPP such as C4.5, SVM, LR, NB and RF. By comparing with other classifiers, RF achieves better results interms of with a maximum F-score of only 76.5. In addition, the algorithm still makes mistakes and labels some stocks that will perform well as the ones that will perform badly and vice versa. Leung et al. [20] propose structural SVM (SSVM) to forecast SSP. It performs classification on complex inputs like the nodes of a graph. The proposed method allows SSVM to learn a prediction model for a complex graph input with many edges for every node. The resultant model is applied for SPP in which the positive and negative class labels represent the increase and decrease in stock prices. In addition, it uses 3-fold cross validation to identify the actual value and SSVM parameter C is set to 1000. It achieves an accuracy of 78%, recommended that the model has learnt with no over-fitting. The performance analysis shows that the ML approaches are better to predict stock prices. Though the SSVM achieves better results, the usage of 3-fold cross validation limits the classification performance.

Qiu et al. [21] introduced Artificial Neural Network (ANN) for SPP in Japanese stock exchange. The main goal of this work is to identify the next day stock prices. To increase the classification prediction, ANN is integrated to GA algorithm and presented GA-ANN method for efficient SPP. GA is chosen to improvise the precision level of ANN and also to eliminate the convergence issue of the back-propagation method. The performance analysis reported that hybrid GA-ANN achieves a hit rate of 81.27% which is higher than the existing works. But, the prediction performance of this study may be enhanced by combining two types of input indicators, or test a subset of these variables. Zhiqiang Guo et al. [22] implemented a hybrid model which integrates 2-dimensional principal component analysis (2D) (PCA), RBF and NN for SPP in Shanghai stock market. They chose 36 stock market variables as input features, and a sliding window is employed to get input data. Then, 2D-PCA is used to minimize the dimension of data and filter its intrinsic features. At the end, RBFNN uses the data processed by 2D-PCA to foresee the tomorrow's stock price. The simulation results imply that the proposed method outperforms the MLP. While the model obtains high accuracy forecasting at low computational cost, the input dimension of

the RBFNN is still high. Despite the fact that the RBFNN training has fast convergence, this high dimensionality and associated training complexity may not be suitable for some real-time forecasting contexts where models must be rapidly built, "on-the-fly".

Khalid Alkhatib et al. [23] uses k-Nearest Neighbor (KNN) algorithm and non-linear regression method for SPP for a set of six major companies listed on the Jordanian stock exchange to help sellers and buyers to take appropriate decisions. Based on the results, kNN algorithm seems to be robust and achieves less error and the prediction results are almost closer to the actual stock prices. Due to the lack of knowledge of financial econometrics in Jordan, there is still long way to utilize advanced predicting models to help the financial markets and brokerage houses and to move forward and be part of the developed international financial market.

OguzAkbulgic et al. [24] introduced a Hybrid RBR-NN which combines regression trees, NN and RBF for SPP in Istanbul stock exchange. HRBF-NN is highly useful to model complex non-linear relationships and dependencies between the stock indices. It produces better results even for non-linear data. A drawback of this approach is the assumption made that random noise follows a Gaussian distribution, which should be relaxed by considering a more general distributional assumption on the random noise such as the Power Exponential (PE) distribution. Zhiqiang Guo et al. [25] presented a SPP model using PCA, canonical correlation analysis and SVM. Initially, two features are filtered from the previous closing prices and 39 technical variables area obtained from independent component analysis. Next, a canonical correlation analysis technique is used to integrate two types of features and filter intrinsic features to enhance the prediction accuracy. At last, SVM is used to foresee the next day closing price. The experiment is carried out using Dow Jones index and results indicate proposed method is better than other methods. The existing methods fail to attain better classification performance for SPP. Apart from several advantages of the proposed work, the forecasting accuracy of the model is not particularly high. Another weakness is that the optimal feature dimensionality of the ICA-CCA-SVR model may sometimes be higher than that of the other models, due to the serial features fusion method which needs to be resolved by parallel fusion method.

A predictive system is designed for SPP using sentimental analysis [26]. It integrates both sensex points and Really Simple Syndication (RSS) feeds

for efficient prediction. It is claimed that the sentiment analysis of RSS news feeds has an impact on stock market values. Hence, RSS news feed data are gathered along with the stock market investment data for a certain time period. Using this algorithm for sentiment analysis, the correlation between the stock market values and sentiments in RSS news feeds are established. Here, the proposed work can further improve the accuracy by combining one or more stock level indicators with RSS feed stock news.

Though several works have been done on the classification problem of SPP, none of the work uses ACO algorithm to address the same. ACO provides accurate results within a minimum time span. Hence, an ACO based SPP classification model is proposed to attain enhanced results.

### 3. ACO based SPP

#### 3.1 Overview

An ACO algorithm is essentially a system based on agents that simulate the natural behaviour of ants, including mechanisms of cooperation and adaptation. The use of this system as a new metaheuristic was proposed in order to solve combinatorial optimization problems [27]. This new metaheuristic has been shown to be both robust and versatile in the sense that it has been applied successfully to a range of different combinatorial optimization problems [28]. In the context of rule discovery, ACO algorithm has the ability to perform a flexible robust search for a good combination of terms (logical conditions) involving values of the predictor attributes. Hence, the ACO algorithm can be used for rule discovery in SPP. In the field of data mining, a classification algorithm named ant miner algorithm is used, based on the principle of ACO. In this work, an enhanced Ant miner algorithm ACO-SPP is implemented to classify the SPP. The work flow and steps involved in ant-miner are discussed below. In the classical ACO algorithm, every individual ant generates an optimal solution for the applied problem. Every classification rule is represented as IF<term1 && term2 ...>THEN<class>.

Every term is a triplet <attribute, operator, value>, where value belongs to the domain of attribute. The operator in the triplet is a relational operator. The present version of Ant-Miner depends on categorical and the operator element is considered as '='. Ant-Miner uses a sequential covering method to discover a list of classification rules which cover all training cases. Initially, the list

of discovered rules is empty and the training set holds all training cases. Once the ant-miner discovers a rule, it is appended to the discovered rule list and deleted from the training set. This step is iterated until the number of uncovered training cases crosses a specified threshold value. When the process is done, the best rule from the discovered rule list is appended to the list of discovered rules, and the system reinitiates the new iteration with equal quantity of pheromone. The ACO-SPP algorithm is given in Algorithm 1.

#### 3.2 Rule pruning

It is a widely used technique in data mining [29]. This process intends to eliminate irrelevant terms which are improperly included in the rule. This process significantly enhances the predictive power of the rule, helps to reduce over fitting to the training data. The other aim of rule pruning is to enhance the rule simplification where a short rule can be easily understandable when compared to a long rule. Once the ant finishes the process of rule construction, then the rule pruning procedure is invoked.

#### 3.3 Pheromone updating

Pheromone updating represents the volatile nature of ant pheromone in the real world. Artificial ants employ the pheromone updating procedure to discover simple rules. Because of positive feedback mechanism, the errors in the heuristic measure are rectified and attained enhanced classification accuracy.

---

#### Algorithm 1: ACO-SPP

---

```

TrainingSet = {all training cases};
DiscoveredRuleList = [ ]; /* rule list is initialized with an
empty list */
WHILE (TrainingSet > Max_uncovered_cases)
  t = 1; /* ant index */
  j = 1; /* convergence test index */
  Initialize all trails with the same amount of pheromone;
  REPEAT
    Antt starts with an empty rule and
    constructs a classification rule Rt
    Prune rule Rt;
    Update the pheromone;
    IF (Rt is equal to Rt-1) /* update convergence
test */
      THEN j = j + 1;
      ELSE j = 1;
    END IF
  
```

---

---

```

t = t + 1;
UNTIL (i ≥ No_of_ants) OR (j ≥ No_rules_converg)
Choose the best rule Rbest among all rules Rc constructed
by all the ants;
Add rule Rbest to DiscoveredRuleList;
TrainingSet = TrainingSet - {set of cases correctly
covered by Rbest};
END WHILE
    
```

---

The quality of a rule (Q) is calculated as [30],

$$Q = sensitivity \times specificity \quad (1)$$

$$Q = \frac{TP}{(TP + FN)} \times \frac{TN}{(FP + TN)} \quad (2)$$

where, TP-True Positive, TN- True Negative, FP-False Positive and FN-False Negative

(1) *Pheromone initialisation*

The operation is to “Initialize all trails with the same amount of pheromone”. It is defined in Eq. (3), where a is the number of attributes; b<sub>i</sub> is the number of values of Attribute<sub>i</sub>; t is the sequence number of iteration.

$$\tau_{ij}(t = 0) = \frac{1}{\sum_{i=1}^a b_i} \quad (3)$$

(2) *Pheromone updating for explored nodes*

The amount of pheromone on the nodes which have been used by the current rule will be updated because the artificial ant deposits pheromone during path exploration. Meanwhile, the pheromone evaporation also needs to be simulated. Therefore, the integrative operation is performed according to Eq. (4).

$$\tau_{ij}(t) = (1 - \rho) \tau_{ij}(t-1) + (1 - \rho / (1 + Q)) \tau_{ij}(t-1) \quad (4)$$

where ρ is the pheromone evaporation rate which controls how fast the pheromone evaporates from the trails; Q is the quality of the rule which is calculated from Eq. (3); t is the sequence number of iteration. This equation adopted from the pheromone updating function of ACO algorithm has higher classification accuracy because Ant-Miner’s function does not consider pheromone evaporation for explored nodes.

(3) *Pheromone updating for unexplored nodes*

The nodes which have not been used by the current rule will only have pheromone evaporation. The evaporation is performed according to Eq. (5):

$$\tau_{ij}(t) = \frac{\tau_{ij}(t-1)}{\sum_{i=1}^a \sum_{j=1}^{b_i} b_i = 1 \tau_{ij}(t-1)} \quad (5)$$

where a is the number of attributes; b<sub>i</sub> is the number of values of Attribute<sub>i</sub>; t is the sequence number of iteration. The equation means that the amount of pheromone of unexplored nodes will be decreased as time goes by.

### 3.4 Usage of discovered rules

For classifying a new test case, the discovered rules in the training process is used to discover the unobserved cases in the way training cases are discovered. Initially the rule that covers the new case is used, it is also likely that none of the rules covers the new case. In this scenario, new case is classified by a standard rule which identifies the majority class in the set of uncovered training cases.

## 4. Performance evaluation

The performance of the ACO-SPP model is validated by comparing the results with existing ML classifiers and also a GA based classifier using a set of datasets. The classifiers used for comparison are NB [31], SVM [32], MLP [33], RBF [34], J48 [35], RF [36], CART [37], Olex-GA [38] and XGBoost classifier [39]. The parameters used, datasets employed, performance metrics and the results are discussed in the following subsections.

### 4.1 Parameter setting

The parameters used in this experiment are tabulated in Table 1. The parameter values were determined by experimentation prior to running the eight trials. Min\_cases\_per\_rule = 5, Max\_uncovered\_cases=10 and Number of converged rules = 10.

Table 1. Parameter setting

Parameter	Variable	Value
Number of ants	Number_of_ants	2000
Minimum number of cases	Min_cases_per_rule	5
Maximum number of uncovered cases	Max_uncovered_cases	10
Number of converged rules	No_of_rules_converge	10

Table 2. Performance measures

Measure	Mathematical formula
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$
F-score	$2TP / (2TP + FP + TN)$
AUC	$1/2 (Sensitivity + Specificity)$
Youden's index	$Sensitivity - (1 - Specificity)$
Discriminant Power	$\sqrt{3} / \pi (\log (Sensitivity (1 - Specificity)) + \log (Specificity (1 - Sensitivity)))$

### 4.2 Dataset description

The proposed method is validated by applying Dow Jones Index dataset [40]. In addition, we have collected three datasets from Yahoo finance on daily, weekly and monthly basis. Four datasets are applied to validate the performance of the ACO-SPP model. Dow Jones index dataset comprises of 750 instances and 16 attributes. The other three datasets consist of 1257,727 and 253 instances with 7 attributes namely date, opening price, high price, lower price, closing price, adjusted close price and stock volume respectively.

### 4.3 Performance measures

The performance measures used in this experiment for comparison purposes are Accuracy, F-score, AUC, Discriminant Power and Youden's Index. The measures are tabulated in Table 2 along with their mathematical formulas.

### 4.4 Results and discussion

From Table 3 and Fig. 1, it is clear that ACO-SPP achieves better results than existing classifiers and attains an accuracy of 96%. It is noted that NB classifier produces poor results than other classifiers. RF achieves 91.46% accuracy which is better one among traditional classifiers but lesser than GA based classifier named Olex-GA. Though Olex-GA produce accuracy of 92.89% which is higher than other classifiers, but it is not efficient than ACO-SPP model. In terms of F-score, SVM attains poor results and Olex-GA is better than traditional classifiers. But, Olex-GA achieves an F-score of 91 whereas ACO-SPP model attains a value of 96.77 respectively. Among all the other methods, XGBoost outperforms all the classifiers except ACO-SPP algorithm.

Moreover, the proposed method showed maximum classification performance with accuracy of 96, F-score of 96.77, AUC of 96.77, Youden's

Table 3. Comparison of Dow Jones Index dataset on various classifiers with proposed model

Method	Accu.	F-score	AUC	Youden's Index	DP
ACO-SPP	96	96.77	96.77	93.55	37.5
XGBoost	94.2	92.35	94.17	88.33	29.5
Olex-GA	92.89	91	92	81	27
CART	90.66	89.9	77	54	15
RF	91.46	90	90	80	26
J48	90	89.8	89	79	24
MLP	88.4	88.37	88.63	77.3	22.4
NB	83.33	82.35	84.17	68.33	27.9
RBF	77.2	80	76.84	53.68	14.8
SVM	50.13	50.98	50.49	98	3

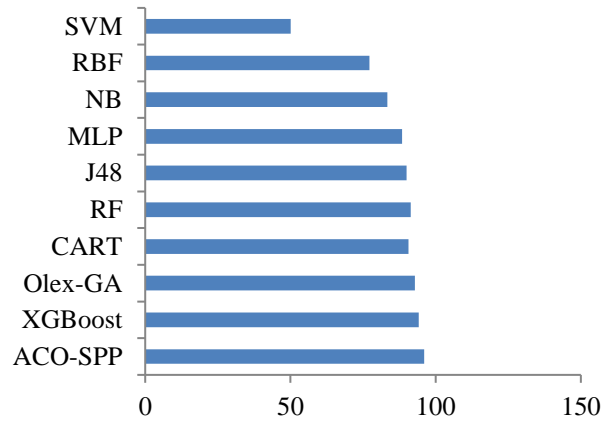


Figure. 1 Comparison of Dow Jones dataset of various classifiers interms of accuracy

index of 93.55 and DP of 37.5. Contrastingly, SVM showed worst classification performance with accuracy of 50.13, F-score of 50.98, AUC of 50.49, Youden's index of 98 and DP of 3. In overall, the proposed method is the better choice for SPP on Dow Jones dataset.

From Table 4 and Fig. 2, it is clear that ACO-SPP achieves better results than existing classifiers for daily SPP dataset from Yahoo finance. ACO-SPP attains an accuracy of 56.16%. It is noted that RF classifier produces poor results than other classifiers. RF achieves 52.31% accuracy which is better one among traditional classifiers but lesser than FA based classifier named Olex-GA. Though Olex-GA produce accuracy of 54.89% which is higher than other classifiers, but it is not efficient than ACO-SPP model. At the same time, XGBoost outperforms all the other classifiers except ACO-SPP. In terms of F-score, RF attains poor results and Olex-GA is better than traditional classifiers. But,



Table 4. Comparison of Daily stock prediction dataset on various classifiers with proposed model

Method	Accu.	F-score	AUC	Youden's Index	DP
ACO-SPP	56.16	56.89	56.51	13.04	29
XGBoost	55.15	55.80	55.01	13.01	28
Olex-GA	54.89	55	55	12.9	27
CART	53.76	51.99	53.9	73.7	14
RF	52.31	50.3	52.12	54.8	8.5
J48	53.94	52	54.1	73.89	13
MLP	52.67	50.98	52.76	55.3	8.8
NB	52.52	53.33	52.52	50.57	13
RBF	53.63	51.98	53.6	73.4	11
SVM	53.94	52.82	53.69	73.7	14

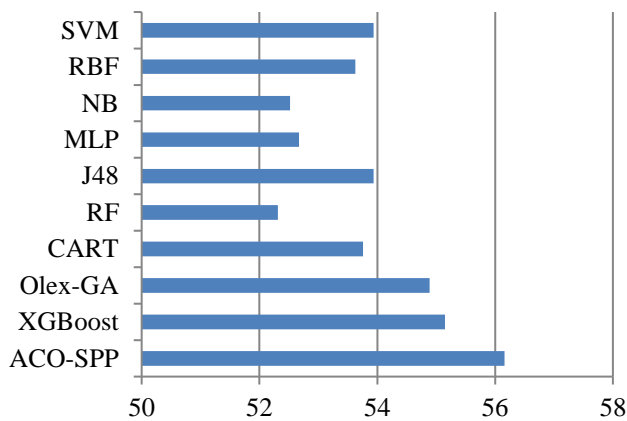


Figure.2 Comparison of Daily stock prediction dataset on various classifiers with proposed model

Table 5. Comparison of Weekly stock prediction dataset on various classifiers with proposed model

Method	Accu.	F-score	AUC	Youden's Index	DP
ACO-SPP	61.41	61.54	61.54	23.08	52
XGBoost	60.59	59.64	60.56	24.9	52
Olex-GA	60.15	59.14	60.23	24.7	52
CART	56.82	57.63	56.09	12.17	31
RF	49.3	49.12	50	10	12
J48	56.82	57.63	56.09	12.17	31
MLP	55.59	50.98	54.99	9.99	12
NB	51.23	52.03	51.26	2.51	10
RBF	55.77	50.99	55.2	10	24
SVM	56.82	57.63	56.09	12.17	31

Olex-GA achieves an F-score of 55 whereas ACO-SPP model attains a value of 56.89 respectively.

Furthermore, the proposed ACO-SPP algorithm depicted closest classification performance with accuracy of 56.16, F-score of 56.89, AUC of 56.51, Youden's index of 13.04 and DP of 29. Contrastingly, RF showed worst classification performance with accuracy of 52.31, F-score of 50.3, AUC of 52.12, Youden's index of 54.8 and DP of 8.5. In overall, the proposed method is the better choice for SPP on Yahoo daily dataset.

From Table 5 and Fig. 3, it is clear that ACO-SPP achieves better results than existing classifiers for weekly SPP dataset from Yahoo finance. ACO-SPP attains an accuracy of 61.41%. It is noted that RF classifier produces poor results than other classifiers. RF achieves 49.3% accuracy which is better one among traditional classifiers but lesser than FA based classifier named Olex-GA. Though Olex-GA produce accuracy of 60.15% which is higher than other classifiers, but it is not efficient than ACO-SPP and XGBoost model. At the same time, XGBoost is found to be efficient than the compared algorithms except the ACO-SPP model. In terms of F-score, RF attains poor results and Olex-GA is better than traditional classifiers. But, Olex-GA achieves an F-score of 59.14 whereas ACO-SPP model attains a value of 61.54 respectively. Additionally, the proposed ACO-SPP algorithm reported closest classification performance with accuracy of 61.41, F-score of 61.54, AUC of 61.54, Youden's index of 23.08 and DP of 52. Contrastingly, RF showed worst classification performance with accuracy of 49.3, F-score of 49.12, AUC of 50, Youden's index of 10 and DP of 12. In overall, the proposed method is the better choice for SPP on yahoo weekly dataset.

From Table 6 and Fig. 4, it is clear that ACO-SPP achieves better results than existing classifiers for monthly SPP dataset from Yahoo finance. ACO-SPP attains an accuracy of 84.61%. It is noted that NB classifier produces poor results than other classifiers. RF achieves 50.38% accuracy which is better one among traditional classifiers but lesser than FA based classifier named Olex-GA. Though Olex-GA produce accuracy of 79.49% which is higher than other classifiers, but it is not efficient than ACO-SPP model. In terms of F-score, NB classifier attains poor results and Olex-GA is better than traditional classifiers. But, Olex-GA achieves an F-score of 76 whereas ACO-SPP model attains a value of 80 respectively. Moreover, the XGBoost classifier is found to be efficient than all the compared methods except ACO-SPP model.

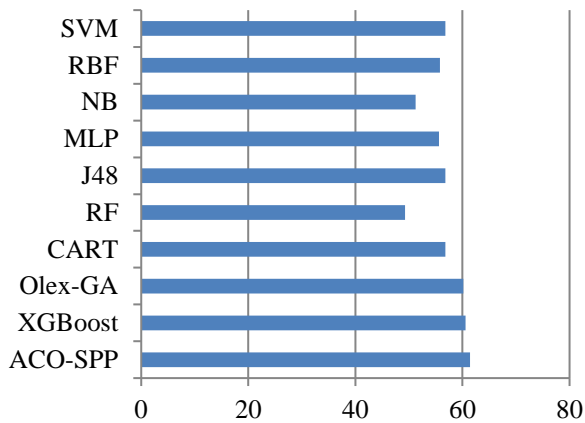


Figure.3 Comparison of Weekly stock prediction dataset on various classifiers with proposed model

Table 6. Comparison of Monthly stock prediction dataset on various classifiers with proposed model

Method	Accu.	F-score	AUC	Youden's Index	DP
ACO-SPP	84.61	80	83.9	67.8	14.8
XGBoost	82.41	81.54	82.91	6582	13.2
Olex-GA	79.49	76	79	63	12
CART	57.89	57	56	13	31
RF	59.4	58	59	24	51
J48	63.91	62	63	25	53
MLP	62.41	61.54	62.91	25.82	54
NB	50.38	49.12	50	0	0
RBF	60.15	59.14	60.23	24.7	52
SVM	63.91	62.69	63.49	26.97	59

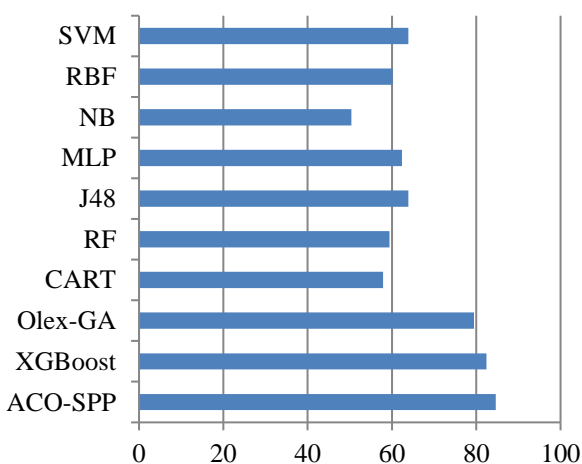


Figure.4 Comparison of Monthly stock prediction dataset on various classifiers with proposed model

Additionally, the proposed ACO-SPP algorithm resulted in maximum classification performance with accuracy OF 84.61, F-score of 80, AUC of 83.9, Youden's index of 67.8 and DP of 14.8. Contrastingly, NB showed worst classification performance with accuracy of 50.38, F-score of 49.12, AUC of 50, Youden's index of 0 and DP of 0. In overall, the proposed method is the better choice for SPP on yahoo weekly dataset. From the above discussion, it is clear that ACO-SPP is the better model for SPP than other techniques. The proposed method produces better results than weekly stock prediction dataset. From the above discussion, it is clear that ACO-SPP is the better model for SPP than other techniques. The effectiveness of the proposed method is due to the following reasons: inherent parallelism nature and also the positive feedback mechanism for rapid discovery of good solutions.

### 5. Conclusion

SPP is a trivial task since the stock prices are tremendously volatile and diverge extensively both within a single stock and comparatively amongst other stocks. Several methods have been proposed for SPP based on ML techniques. In this paper, an Ant Colony Optimization (ACO) based SPP technique is introduced to effectively forecast the stock prices. The effectiveness of the proposed ACO-SPP model is validated by comparing its performance with a set of traditional ML based classifiers such as NB, SVM, MLP, RBF, J48, RF, CART and Olex-GA. To highlight the efficiency of ACO-SPP model, it is applied to a benchmarked dataset named Dow Jones dataset. We have collected three datasets from Yahoo finance on daily, weekly and monthly basis. The experimental results imply that ACO-SPP attains an accuracy of 96%, 56.16%, 61.41% and 84.61% for the applied datasets. It is also observed that better accuracy is achieved for Dow Jones dataset when compared to other datasets. In future, this work can be extended by the use of different bio-inspired algorithms to solve the classification problem of SPP.

### References

[1] D. Gholamiangonabadi, S.D. Taheri, A. Mohammadi, and M.B. Menhaj, "Investigating the performance of technical indicators in electrical industry in Tehran's Stock Exchange using hybrid methods of SRA, PCA and Neural Networks", In: *Proc. of the 5th Conference on Thermal Power Plants*, pp. 75-82, 2014.

[2] M.T. Leung, H. Daouk, and A.S. Chen, "Forecasting stock indices: a comparison of



- classification and level estimation models”, *International Journal of Forecasting*, Vol.16, No.2, pp.173-190, 2000.
- [3] M. M. Mostafa, “Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait”, *Expert Systems with Applications*, Vol. 37, No.9, pp. 6302-6309, 2010.
- [4] E. Guresen, G. Kayakutlu, and T.U. Daim, “Using artificial neural network models in stock market index prediction”, *Expert Systems with Applications*, Vol. 38, No. 8, pp. 10389-10397, 2011.
- [5] T.S. Lee and C.C. Chiu, “Neural network forecasting of an opening cash price index”, *International Journal of Systems Science*, Vol. 33, No. 3, pp. 229-237, 2002.
- [6] M.A.I. Khan, “Financial volatility forecasting by nonlinear support vector machine heterogeneous autoregressive model: evidence from Nikkei 225 stock index”, *International Journal of Economics and Finance*, Vol.3, No.4, pp.138, 2011.
- [7] F.E. Tay and L. Cao, “Application of support vector machines in financial time series forecasting”, *Omega*, Vol. 29, No. 4, pp. 309-317, 2001.
- [8] G. Deboeck, “Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets”, *John Wiley & Sons*, Vol. 39, 1994.
- [9] P. Holloway, “A filosofia value investing na gestão de fundos de investimentos brasileiros”, *Ph. D. Diss.*, 2012.
- [10] R. B. Barsky and J. B. DeLong, “Why does the stock market fluctuate?”, *The Quarterly Journal of Economics*, Vol. 108, No. 2, pp. 291-311, 1992.
- [11] T. Hendershott and P.C. Moulton, “Automation, speed, and stock market quality: The NYSE's hybrid”, *Journal of Financial Markets*, Vol.14, No.4, pp.568-604, 2011.
- [12] S. Mahfoud and G. Mani, “Financial forecasting using genetic algorithms”, *Applied Artificial Intelligence*, Vol.10, No.6, pp.543-566, 1996.
- [13] E. Tsang, S. Markose, and E. Hakan, “Chance discovery in stock index option and futures arbitrage”, *New Mathematics and Natural Computation*, Vol.1, No.03, pp.435-447, 2005.
- [14] L. Wagman, “Stock portfolio evaluation: An application of genetic-programming-based technical analysis”, *Genetic Algorithms and Genetic Programming at Stanford*, pp.213-220, 2003.
- [15] J.J. Wang, J.Z. Wang, Z.G. Zhang, and S.P. Guo, “Stock index forecasting based on a hybrid model”, *Omega*, Vol.40, No.6, pp.758-766, 2012.
- [16] M. Dorigo and G. D. Caro, “Ant colony optimization: a new meta-heuristic”, *In: Proc. of the 1999 Congress on Evolutionary Computation*, Vol. 2, 1999.
- [17] R.S. Parpinelli, S. Lopes, and A.A. Freitas, “Data mining with an ant colony optimization algorithm”, *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 4, pp. 321-332, 2002.
- [18] M. Alrasheedi and A. Alghamdi, “Comparison of Classification Methods for Predicting the Movement Direction of Saudi Stock Exchange Index”, *Journal of Applied Sciences*, Vol.14, No.16, pp.1883-1888, 2014.
- [19] N. Milosevic, “Equity forecast: Predicting long term stock price movement using machine learning”, *arXiv preprint arXiv:*, pp.1603.00751, 2016.
- [20] C.K.S. Leung, R.K. MacKinnon, and Y.Wang, “A machine learning approach for stock price prediction”, *In: Proc. of the 18th International Database Engineering & Applications Symposium*, 2014.
- [21] M. Qiu and Y. Song, “Predicting the direction of stock market index movement using an optimized artificial neural network model”, *PloS one*, Vol.11, No.5, 2016.
- [22] Z.Guo, et al. “A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network”, *PloS one*, Vol.10, No.4, pp. 1-19, 2015.
- [23] K. Alkhatib, H. Najadat, I. Hmeidi, and M.K.A. Shatnawi, “Stock price prediction using k-nearest neighbor (knn) algorithm”, *International Journal of Business, Humanities and Technology*, Vol. 3, No. 3, pp. 32-44, 2013.
- [24] O. Akbilgic, H. Bozdogan, and M. E. Balaban, “A novel Hybrid RBF Neural Networks model as a forecaster”, *Statistics and Computing*, Vol.24, No.3, pp.365-375, 2014.
- [25] Z. Guo, et al. “A feature fusion based forecasting model for financial time series”, *PloS one*, Vol.9, No.6, pp. 1-13, 2014.
- [26] S. Bharathi and A. Geetha, “Sentiment Analysis for Effective Stock Market Prediction”, *International Journal of Intelligent Engineering and Systems*, Vol.10, No.3, pp. 146-154, 2017.

- [27] M. Dorigo and G. Di Caro, “The ant colony optimization meta-heuristic”, *New Ideas in Optimization*, pp. 11–32, 1999.
- [28] M. Dorigo, G. D. Caro, and L. M. Gambardella, “Ant algorithms for discrete optimization”, *Artif. Life*, Vol. 5, No. 2, pp.137–172, 1999.
- [29] L.A. Breslow and D.W. Aha, “Simplifying decision trees: A survey”, *The Knowledge Engineering Review*, Vol.12, No.1, pp.1-40, 1997.
- [30] H.S. Lopes, M.S. Coutinho, and W. C. de Lima, “An evolutionary approach to simulate cognitive feedback learning in medical domain”, *Genetic Algorithms and Fuzzy Logic Systems: Soft Computing Perspectives*, pp.193-207, 1997.
- [31] I. Rish, “An empirical study of the naive Bayes classifier”, In: *Proc. of IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. 3, No. 22, pp. 41-46, 2001.
- [32] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines”, *IEEE Intelligent Systems and their applications*, Vol. 13, No. 4, pp. 18-28, 1998.
- [33] S.K. Pal and S. Mitra, “Multilayer perceptron, fuzzy sets, and classification”, *IEEE Transactions on neural networks*, Vol. 3, No. 5, pp. 683-697, 1992.
- [34] J. Park and I.W. Sandberg, “Universal approximation using radial-basis-function networks”, *Neural Computation*, Vol. 3, No. 2, pp. 246-257, 1991.
- [35] T.R. Patil and S.S. Sherekar, “Performance analysis of Naive Bayes and J48 classification algorithm for data classification”, *International Journal of Computer Science and Applications*, Vol. 6, No. 2, pp. 256-261, 2013.
- [36] M. Pal, “RF classifier for remote sensing classification”, *International Journal of Remote Sensing*, Vol. 26, No. 1, pp. 217-222, 2005.
- [37] W.Y. Loh, “Classification and regression trees”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp. 14-23, 2011.
- [38] A. Pietramala, V.L. Policicchio, P. Rullo, and I. Sidhu, “A genetic algorithm for text classification rule induction”, In: *Proc. of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp.188-203, 2008.
- [39] T. Chen, T. He, and M. Benesty, “Xgboost: extreme gradient boosting”, *R package version 0.4-2*, pp.1-4, 2015.
- [40] A. Robles-Kelly, Antonio, and R. Wei, “A spectral clustering approach for online and streaming applications”, In: *Proc. of the International Joint Conference on Neural Networks*, pp. 3904-3911, 2017.