



Social Spider Optimization Algorithm with Gradient Boosting Tree Model for Decision Making in Telemarketing Sector

Abd Al-Aziz Hosni El-Bagoury¹, Sundus Naji AL-Aziz^{2*}, S.S.ASKAR³

¹Higher Institute of Engineering and Technology, El-Mahala El-Kobra, Egypt

²Department of Mathematics, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, KSA.

³Faculty of Science, King Saud University, Riyadh, Saudi Arabia

Emails: azizhel2013@yahoo.com; snalaziz@pnu.edu.sa ; saskar@ksu.edu.sa

*Corresponding Author: Sundus Naji AL-Aziz. Email: Sundus.alaziz3@gmail.com

Abstract

Telemarketing becomes a major tool in enhancing the services of different business sectors. On banking industry, telemarketing is applied to sell products or services. Banking advertisements as well as marketing are majorly based on the detailed information of neutral data related to marketing market and original needs of user for the banks. Decision making becomes an essential part in the telemarketing field that computes a particular class of automated fact in assisting the companies for making decision. Artificial intelligence (AI) is applied for decision making in the telemarketing sector. In this aspect, this paper introduces a social spider optimization (SSOA) with gradient boosting tree (GBT) model for decision making in the telemarketing sector. The main aim of the SSOA-GBT method is to make proper decisions in the telemarketing sectors. To accomplish this, the SSOA-GBT model initially exploits the GBT model for data classification purposes. Next, for improving the performance of the GBT classifier, the SSOA is applied. The performance validation of the SSOA-GBT model is performed using benchmark dataset and the outcomes are investigated in several aspects. The simulation outcomes indicated the better outcomes of the SSOA-GBT approach over the recent approaches.

Keywords: Telemarketing; Banking sector; Decision making; Social spider optimization; Gradient boosting tree

1. Introduction

Telemarketing is a system or technique of direct marketing, where a businessperson interacts with client to encourage them to avail or buying the product and services, whether by in-person interaction or connected through telephone [1]. Recently, with the enormous acceptance of mobile phone telemarketing has received considerable attention as an effective manner of marketing. During the banking field, it is major support scheme for exchanging services and products. Banking services and products promotion to rise the enterprise needs an extensive knowledge of present market data and the real user expectation. The major way bank reaches the customer is through telemarketing service [2, 3]. As a marketing technique, this cover the product promotion by making call out of a curated customer list that utilize low cost and high yield when compared to other models while a client is targeted precisely. In another word, performing telemarketing without a plan might leads to redundant cost expenditure.

Many customers face uneasiness during telemarketing campaign since they aren't attracted in product or think telemarketing is an attack of the confidentiality [4, 5]. Therefore, some critics, telemarketing complaint, and customer opinion are found easily on social networking platform. For regulating telemarketing better and reducing negative perception, state government and regulator have launched laws and regulation for telemarketing in the jurisdiction to guarantee the authority of telemarketing campaign and defend customer confidentiality [6]. Furthermore, businesses make use of self-regulation in professional training, telemarketing, and strategy to control the behavior of telemarketer and telemarketing practice [7]. There has been continuing progress in machine learning method which offer opportunity to sales needs and strategize marketing [8]. Thus, it is necessary to implement telemarketing sales prediction study with ML method for guiding marketer in designing telemarketing strategy. ML is a technique of data mining, and a predecessor to artificial intelligence offers a technique for scaling complicated challenges by constructing prediction methods according to the past dataset and utilize it for making prediction in the future [9, 10].

In [11], common techniques of SVM, DT, RF, and ANN classifier are executed. In order to decrease the dimensional, FS optimum subsets LR, LASSO, and RF approaches were utilized. A purpose is for verifying the prediction performance and accuracy of this technique then FS. Sun et al. [12] define a mining statistics model for extracting helpful data in a present Portuguese bank telemarketing function in the UCIKDD ML-Dataset. In order to validate prediction performance, the projected approach was connected for general classification method which contains NB, SVM, and DT. Che et al. [13] purpose at the difficult maximum dimensional nonlinear feature of the issue affecting the attainment rate of telemarketing, a t-SNE FE method, and later proceeds the extracting minimum dimensional feature as input, employ nonlinear SVM to trained and prediction. The outcome illustrates that the bank phones dependent on t-SNESVM projected under these studies. The marketing prediction techniques are generalized and learning ability which is offer particular decision making references to bank and another business to attain accuracy marketing.

Kim et al. [14] presented the DCNN infrastructure that forecasts if a given user has been suitable for banks telemarketing/not. Numerous layers, rate of learning, initial value of node, and another variable that is set to generate DCNN are presented and analyzed. Ghatasheh et al. [15] designed at improving the performance of forecast the drive of bank client to execute for term deposits from extremely imbalanced dataset. It presents increased ANN approaches (i.e., cost sensitive) to mitigate the drastic effect of extremely imbalanced data, without distortion of new data instances. The created method was validation, evaluation, and afterward related to various ML approaches. Real-time telemarketing data sets in Portuguese bank is utilized from all experiments. In [16], DL techniques (SimpleRNN, LSTM, and GRU) are utilized to forecast the chance of contributes for deposits then the customer is named in the possibility of bank's telemarketing function. The implemented techniques are verified utilizing the datasets and experiment outcomes are taken as well as related. In order to enhance the obtained efficiency level various techniques were utilized to data set. Due to the unbalanced infrastructure of utilized datasets, SMOTE system was utilized for attaining maximum precise outcomes.

This paper introduces a social spider optimization (SSOA) with gradient boosting tree (GBT) model for decision making in the telemarketing sector. The main aim of the SSOA-GBT method is to make proper decisions in the telemarketing sectors. To accomplish this, the SSOA-GBT model initially exploits the GBT model for data classification purposes. Next, for improving the performance of the GBT classifier, the SSOA is applied. The performance validation of the SSOA-GBT model is performed using benchmark dataset and the outcomes are investigated in several aspects. The simulation outcomes indicated the better outcomes of the SSOA-GBT model over the recent approaches.

2. Process involved in SSOA-GBT Model

This paper has developed a new SSOA-GBT method is to make proper decisions in the telemarketing sectors. The SSOA-GBT model initially exploits the GBT model for data classification purposes. Next, for improving the performance of the GBT classifier, the SSOA is applied.

2.1 Level I: GBT Based Classification

In order to classify the data related to telemarketing in the banking sector, the GBT model is used. The basic model of the GBT is the combination of a range of weak base classifications with robust element. Generally, boosting process are positive and negative samples, GBT realizes global convergence utilizing negative gradient [17].

Let $\{x_i, y_i\}_{i=1}^n$ signifies the data set and softmax is termed as loss function. The gradient descent (GD) approach was executed for assuring the convergence of GBT. Figure 1 illustrates the structure of GBT. The basic learner is $h(x)$, whereas $\chi_i = (x_{1i}, x_{2i}, \dots, x_{pi})$. p signifies the amount of forecasting variables. y_i stands for the forecasting label. The steps contained in GBT are presented as follows:

- An initial constant value of model β is given as:

$$F_0(x) = \arg \min \sum_{i=1}^N L(y_i, \beta) \tag{1}$$

- In order to the iteration count $m = 1: M$ (M determines the iteration), the gradient direction of residual are estimated under.

$$y_i^* = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x) = F_{m-1}(x)}, i = \{1, 2, \dots, N\} \tag{2}$$

- The basic classifications are employed to set sample data and achieve the initial technique. Along with the least square, parameter a_m has been obtained and executed $h(x_i; a_m)$.

$$a_m = \arg \min \sum_{i=1}^N [y_i^* - \beta h(x_i; a)]^2 \tag{3}$$

- At this point, the loss function was decreased. As stated by Eq. (4), a novel step size of model as the present weight has been measured.

$$\beta_m = \arg \min \sum_{i=1}^N L(y_i, F_{m-1}(x) + \beta h(x_i; a)) \tag{4}$$

- Eventually, the model upgrade was implemented as:

$$F_m(x) = F_{m-1}(x) + \beta_m h(x_i; a) \tag{5}$$

However, the size of sample data are decreased, Information Gain (IG) of feature is a vital and calculated to several iterations if the arbitrary data was produced as input for GBT and keep the search procedure [18]. It concludes in iteration number and decreases the convergence speed.

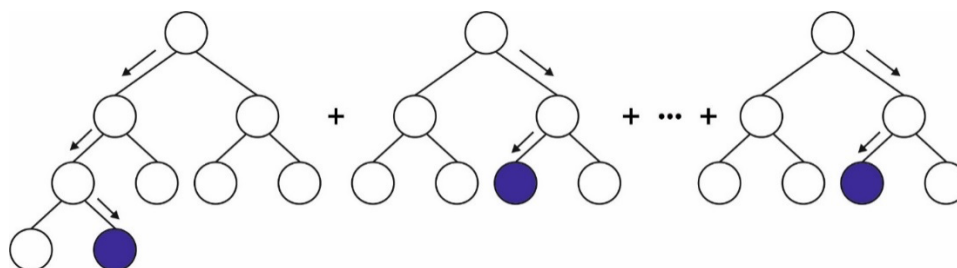


Figure 1. Structure of GBT

2.2 Level II: SSOA Based Parameter Optimization

To adjust the parameters of the GBT model, the SSOA is applied to it. The SSOA technique considering the entire searching region as a common network, whereby every social spider interacted with each other [19]. The solution in the searching space designates the location of the spider in the common netting. All the spiders would receive a weight as per the fitness values viz. determined by social spider. The SSOA method determines 2 kinds of searching agents such as male and female spiders. The SSOA approach begins by the population initialization of N spider position [20]. The entire population comprises male m_i and female f_i spiders. The amount of f_i (N_f) is arbitrarily designated and is defined as follows and also the amount of m_i (N_m) is defined as follows:

$$N_f = \text{floor}[(0.9 - \text{rnd}(0,1)) \cdot 0.25 \cdot N] \quad (6)$$

$$N_m = N - N_f \quad (7)$$

The location of f_u is arbitrarily created among the lower and upper bounds p_v^{hg} and p_v^{lw} that is given as follows

$$f_{u,v}^0 = p_v^{\text{lw}} + \text{rand}(0,1) \cdot (p_v^{\text{hg}} - p_v^{\text{lw}}) \quad (8)$$

Here $u = 1, 2, \dots, N_f$; $v = 1, 2, \dots, n$. also, The location of m_u is generated arbitrarily in the following

$$m_{u,v}^0 = p_v^{\text{lw}} + \text{rand}(0,1) \cdot (p_v^{\text{hg}} - p_v^{\text{lw}}) \quad (9)$$

When $u = 1, 2, \dots, N_m$; $v = 1, 2, \dots, n$.

The size of spider is the major features i.e., utilized for evaluating spider ability to achieve the assigned procedure. In the suggested technique, spider receive a weight w_u that designates the solution quality of spider u of the population S . it can be defined in the following

$$\frac{J(s_u) - \text{wrst}_s}{\text{bst}_s - \text{wrst}_s} \quad (10)$$

Whereas $J(s_u)$ denotes the fitness value, wrst_s and bst_s indicates the maximal and minimal values. The wrst_s and bst_s value are show as follows

$$\text{bst}_s = \max_{k \in \{1,2,\dots,N\}} (J(s_k)) \quad \text{and} \quad \text{wrst}_s = \min_{k \in \{1,2,\dots,N\}} (J(s_k)) \quad (11)$$

data exchanges among the u and v colony members, the vibration is arithmetically determined by the following equation:

$$\text{Vib}_{u,v} = w_v e^{-d_{u,v}^2} \quad (12)$$

Now $d_{u,v}$ indicates the Euclidian distance among u and v colony members. With this vibration, the attribute values of the instance would be transmitted through the member u to member v . It comprises three types of vibration occurs among u and v and denoted as Vibc_u , Vibb_u and Vibf_u .

Vibration Vibf_u can be received through the individual u (s_u) because of the data transmitted with the member c (s_c) i.e., nearer to u also with high weight than u ($w_c > w_u$).

$$\text{Vibc}_u = w_c e^{-d_{u,c}^2} \quad (13)$$

The vibration Vibb_u received through the individual u through the data communicated with the member b (s_b) that has optimal weight of the full population S ,

$$\text{Vibb}_u = w_b e^{-d_{u,b}^2} \quad (14)$$

In conclusion, Vibf_u determines the data communicated from the member u to the adjacent female individual $f(s_f)$ as follows

$$Vibf_u = w_f e^{-d_{u,f}^2} \tag{15}$$

The f_u represent a repulsion or attraction to another spider irrespective of gender. The motion of repulsion or attraction based on distinct conditions. An arbitrary value r_m is created within $[0, 1]$. Once r_m is smaller when compared to a predefined threshold PF as follows

$$f_u^{t+1} = \begin{cases} f_u^t + \alpha \cdot Vibc_u \cdot (s_c - f_u^t) + \beta \cdot Vibb_u \cdot (s_b - f_u^t) + \delta \cdot (rnd - 0.5) \text{ with probability PF} \\ f_u^t - \alpha \cdot Vibc_u \cdot (s_c - f_u^t) - \beta \cdot Vibb_u \cdot (s_b - f_u^t) + \delta \cdot (rnd - 0.5) \text{ with probability } 1 - PF \end{cases} \tag{16}$$

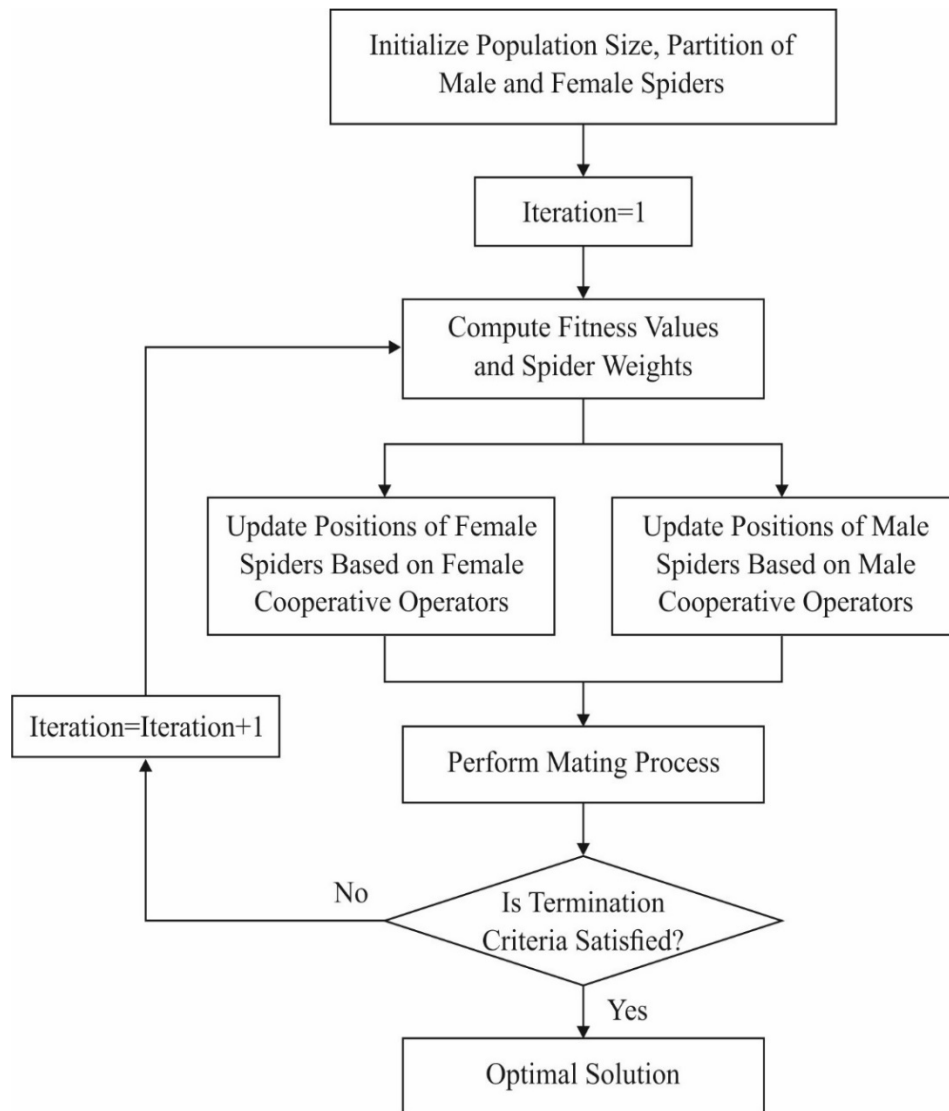


Figure 2. Flowchart of SSO technique

Male cooperative operator

The m_u is a weighted value over the median of N_m is named as dominant D and the residual m_i are named as non-dominant ND. The median weighted has been represented as N_{f+m} . The position of m_u is expressed by

$$m_u^{t+1} = \begin{cases} m_u^t + \alpha \cdot \text{Vib}f_u \cdot (s_f - m_u^t) + \delta \cdot (\text{rnd} - 0.5) \text{if}(w_{N_{f+u}} > w_{N_{f+m}}) \\ m_u^t + \alpha \cdot \left(\frac{\sum_{h=1}^{N_m} - h - N_f + n}{\sum_{h=1}^{N_m} \cdot w_{N_f} + h} - m_u^t \right) \end{cases} \quad (17)$$

Mating operator

Generally, mate can be done amongst D and f_u while f_u is found by D in a certain extent, it can be expressed as follows

$$R = \frac{\sum_{v=1}^n (p_v^{hg} - p_v^{lw})}{2 \cdot n} \quad (18)$$

The spider with extra weights have high possibility of creating the off-spring. Figure 2 demonstrates the flowchart of SSO technique.

3. Performance Validation

In this section, a comprehensive experimental validation of the proposed model is carried out using benchmark dataset [21, 22]. The dataset holds 11162 samples with 16 features and 2 classes.

Table 1 Result analysis of SSOA-GBT technique with distinct runs

No. of Runs	$Prec_n$	$Reca_l$	Acc_y	F_{Score}	Kappa
Run-1	93.76	92.26	92.19	92.96	91.44
Run-2	92.39	92.31	92.87	90.79	89.33
Run-3	93.60	93.38	93.38	92.20	89.51
Run-4	93.68	93.82	92.41	92.70	90.74
Run-5	93.46	92.47	92.34	92.31	89.67
Average	93.38	92.85	92.64	92.19	90.14

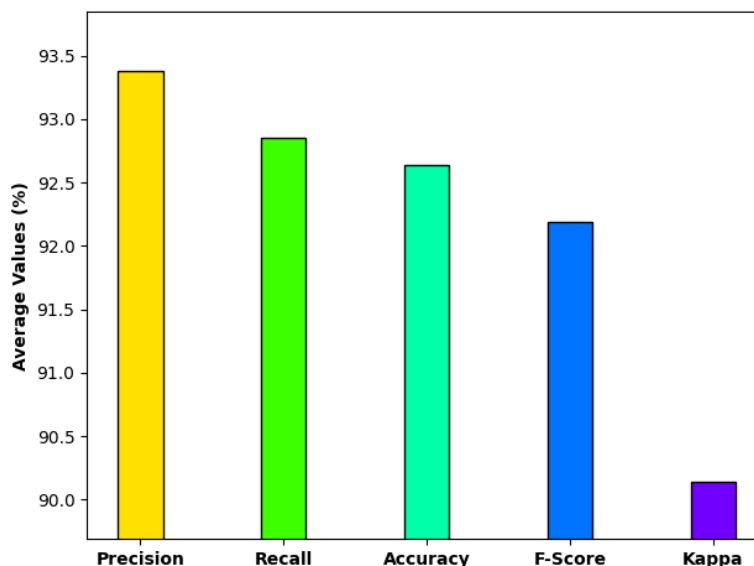


Figure 3. Average analysis of SSOA-GBT algorithm with various measures

Table 1 and Figure 3 demonstrates an overall classifier outcomes of the SSOA-GBT model. The experimental results indicated that the SSOA-GBT model has accomplished effectual outcome under all runs. For sample, with run-1, the SSOA-GBT model has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score} , and

kappa of 93.76%, 92.26%, 92.19%, 92.96%, and 91.44% correspondingly. Similarly, with run-2, the SSOA-GBT model has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score} , and kappa of 92.39%, 92.31%, 92.87%, 90.79%, and 89.33% respectively. Followed by, with run-3, the SSOA-GBT model has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score} , and kappa of 93.60%, 93.38%, 93.38%, 92.20%, and 89.51% respectively. Also, with run-4, the SSOA-GBT model has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score} , and kappa of 93.68%, 93.82%, 92.41%, 92.70%, and 90.74% respectively. Besides, with run-5, the SSOA-GBT model has resulted to $prec_n$, $reca_l$, $accu_y$, F_{score} , and kappa of 93.46%, 92.47%, 92.34%, 92.31%, and 89.67% respectively.

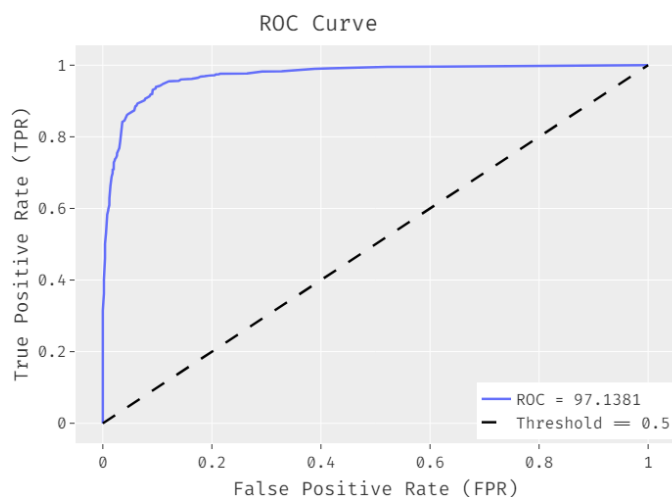


Figure 4. ROC analysis of SSOA-GBT technique

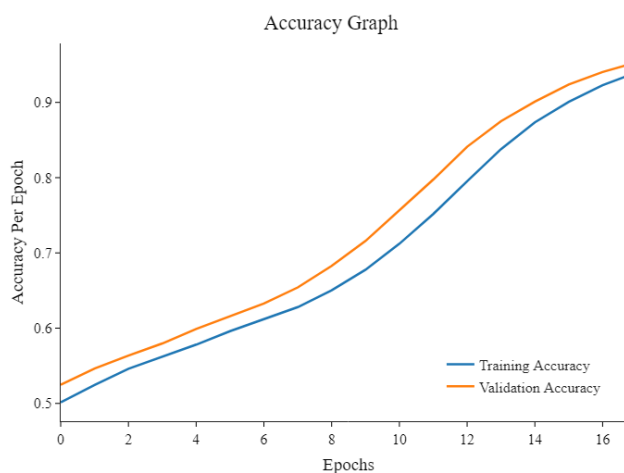


Figure 5. Accuracy graph analysis of SSOA-GBT algorithm

The ROC analysis of the SSOA-GBT system on the test dataset is depicted in Figure 4. The figure revealed that the SSOA-GBT approach has the ability to accomplish improved classification performance with the maximum ROC of 97.1381.

Figure 5 validates the accuracy assessment of the SSOA-GBT system on test dataset. The outcomes described that the SSOA-GBT methodology has the aptitude of gaining improved values of training and validation accuracy. It can be visible that the validation accuracy values are somewhat superior to training accuracy.

A brief training and validation loss offered by the SSOA-GBT system are reported in Figure 6 on the test dataset. The outcomes revealed that the SSOA-GBT approach has been able minimum values of training and validation losses on test dataset.

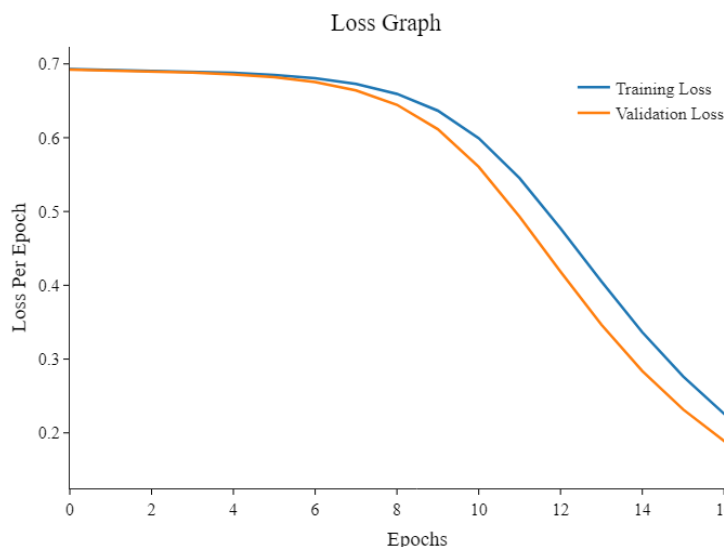


Figure 6. Loss graph analysis of SSOA-GBT algorithm

Table 2 and Figure 7 offers a comparative examination of the SSOA-GBT model with other methods under diverse measures [23]. The experimental results implied that the SSOA-GBT model has showcased effectual outcome over the other models under all measures. For instance, with respect to $prec_n$, the SSOA-GBT model has offered higher $prec_n$ of 93.38% whereas the PTML-DSS, RF, NB-Tree, and DT models have reached lower $prec_n$ of 89.70%, 85.62%, 85.16%, and 85.12%.

Table 2 Comparative analysis of SSOA-GBT technique with recent algorithms

Methods	$Prec_n$	$Recal_l$	Acc_y	F_{Score}	Kappa
SSOA-GBT	93.38	92.85	92.64	92.19	90.14
PTML-DSS Model	89.70	91.48	90.13	86.99	80.91
RF Model	85.62	85.78	85.58	86.37	71.94
NB-Tree Model	85.16	85.56	86.31	85.70	72.28
DT Model	85.12	85.32	84.79	84.67	71.17

At the same time, with respect to $recal_l$, the SSOA-GBT model has offered higher $recal_l$ of 92.85% whereas the PTML-DSS, RF, NB-Tree, and DT models have reached lower $prec_n$ of 91.48%, 85.78%, 85.56%, and 85.32%. Moreover, with respect to acc_y , the SSOA-GBT model has offered higher acc_y of 92.64% whereas the PTML-DSS, RF, NB-Tree, and DT models have reached lower acc_y of 90.13%, 85.58%, 86.31%, and 84.79%. Along with that, with respect to F_{score} , the SSOA-GBT model has offered higher F_{score} of 92.19% whereas the PTML-DSS, RF, NB-Tree, and DT models have reached lower F_{score} of 86.99%, 86.37%, 85.70%, and 84.67%. Finally, with respect to $kappa$, the SSOA-GBT model has offered higher $kappa$ of 90.14% whereas the PTML-DSS, RF, NB-Tree, and DT models have reached lower $kappa$ of 80.91%, 71.94%, 72.28%, and 71.17%.

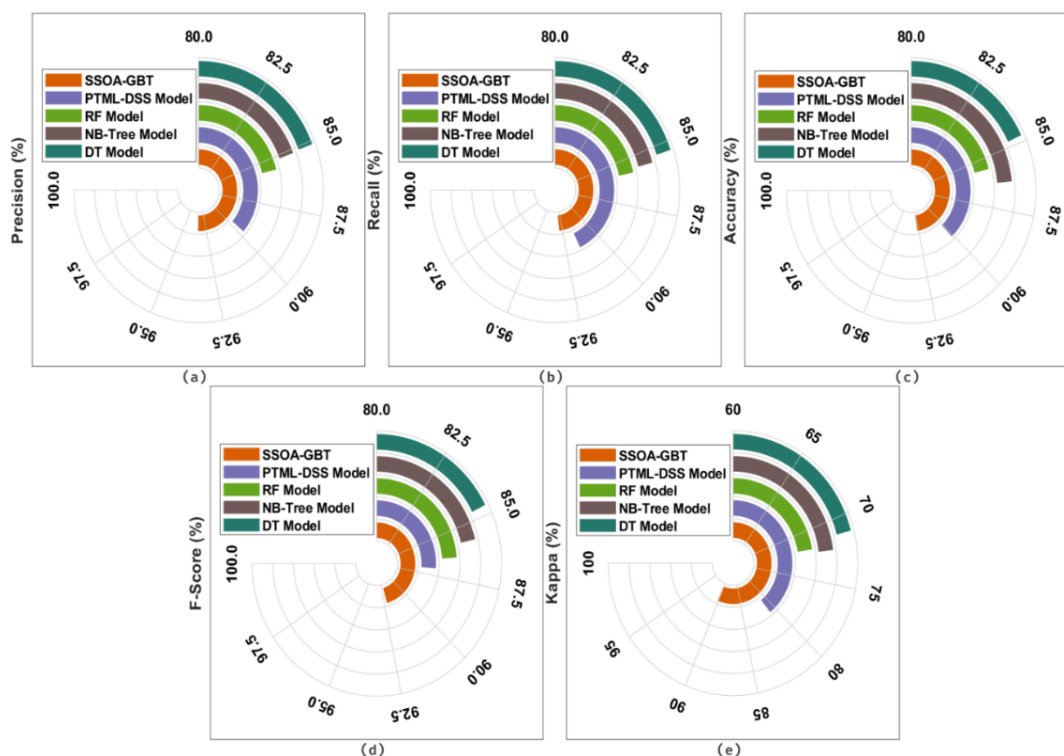


Figure 7. Comparative analysis of SSOA-GBT technique with recent algorithms

4. Conclusion

This paper has established a new SSOA-GBT technique is to make proper decisions in the telemarketing sectors. The SSOA-GBT model initially exploits the GBT model for data classification purposes. Next, for improving the performance of the GBT classifier, the SSOA is applied. The performance validation of the SSOA-GBT model is performed using benchmark dataset and the outcomes are investigated in several aspects. The simulation outcomes indicated the better outcomes of the SSOA-GBT technique over the recent approaches. Thus, the SSOA-GBT model has resulted to effective outcome over the other approaches. In future, feature selection and deep learning models can be designed to enhance the classification outcomes of the presented model.

References

- [1] Rahayu, M., Rasid, F. and Tannady, H., 2018. Effects of self efficacy, job satisfaction, and work culture toward performance of telemarketing staff in banking sector. *South East Asia J. Contemp. Business, Econ. Law*, 16(5), pp.47-52.
- [2] Farooqi, R. and Iqbal, N., 2019. Performance evaluation for competency of bank telemarketing prediction using data mining techniques. *International Journal of Recent Technology and Engineering*, 8(2), pp.5666-5674.
- [3] Prompreing, K. and Prompreing, T., 2021. A Telemarketing Guidance in Selling Banking Services: A Data Mining Approach. *Indonesian Journal of Business Analytics*, 1(1), pp.1-16.
- [4] Prompreing, K. and Prompreing, T., 2021. A Telemarketing Guidance in Selling Banking Services: A Data Mining Approach. *Indonesian Journal of Business Analytics*, 1(1), pp.1-16.
- [5] Desai, R., 2021. Performance Enhancement of Hybrid Algorithm for Bank Telemarketing. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(8), pp.2703-2711.
- [6] Borugadda, P., Nandru, P. and Madhavaiah, C., 2021. Predicting the Success of Bank Telemarketing for Selling Long-term Deposits: An Application of Machine Learning Algorithms. *St. Theresa Journal of Humanities and Social Sciences*, 7(1), pp.91-108.
- [7] Karakuş, M.Ö., 2021. A Multi-Layer Neural Network Approach to Predict The Success of Bank Telemarketing. *Artificial Intelligence Theory and Applications*, 1(1).

- [8] Puteri, A.N., Arizal, A. and Achmad, A.D., 2021. Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), pp.335-342.
- [9] Tékouabou Koumético, S.C. and Touluni, H., 2021. Improving KNN Model for Direct Marketing Prediction in Smart Cities. In *Machine Intelligence and Data Analytics for Sustainable Future Smart Cities* (pp. 107-118). Springer, Cham.
- [10] Ilham, A., Khikmah, L. and Iswara, I.B.A.I., 2019, March. Long-term deposits prediction: a comparative framework of classification model for predict the success of bank telemarketing. In *Journal of Physics: Conference Series* (Vol. 1175, No. 1, p. 012035). IOP Publishing.
- [11] Asif, M., 2018. Predicting the Success of Bank Telemarketing using various Classification Algorithms.
- [12] Sun, Y., Cheng, X., Li, H., Lin, Y., Sung, J. and Wang, A., 2019. Prediction of Bank Customer Telemarketing Success Using Data Mining Models. *ADRRRI Journal of Engineering and Technology*, 4(5), pp.18-32.
- [13] Che, J., Zhao, S., Li, Y. and Li, K., 2020. Bank Telemarketing Forecasting Model Based on t-SNE-SVM. *Journal of Service Science and Management*, 13(3), pp.435-448.
- [14] Kim, K.H., Lee, C.S., Jo, S.M. and Cho, S.B., 2015, November. Predicting the success of bank telemarketing using deep convolutional neural network. In *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (pp. 314-317). IEEE.
- [15] Ghatasheh, N., Faris, H., AlTaharwa, I., Harb, Y. and Harb, A., 2020. Business analytics in telemarketing: cost-sensitive analysis of bank campaigns using artificial neural networks. *Applied Sciences*, 10(7), p.2581.
- [16] Turkmen, E., 2021, February. Deep Learning Based Methods for Processing Data in Telemarketing-Success Prediction. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 1161-1166). IEEE.
- [17] Zhang, Y. and Haghani, A., 2015. A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58, pp.308-324.
- [18] Rao, H., Shi, X., Rodrigue, A.K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X. and Gu, L., 2019. Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing*, 74, pp.634-642.
- [19] Cuevas, E., Cienfuegos, M., Zaldívar, D. and Pérez-Cisneros, M., 2013. A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Systems with Applications*, 40(16), pp.6374-6384.
- [20] Luque-Chang, A., Cuevas, E., Fausto, F., Zaldivar, D. and Pérez, M., 2018. Social spider optimization algorithm: modifications, applications, and perspectives. *Mathematical Problems in Engineering*, 2018.
- [21] <https://www.kaggle.com/janiobachmann/bank-marketing-dataset>
- [22] <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
- [23] Noura Metawa, Amany Ahmed Elshimy, "Parameter Tuned Machine Learning based Decision Support System for Bank Telemarketing", *American Journal of Business and Operations Research (AJBOR)*, Vol. 4, No. 1, PP. 28-38, 2021