Int. J. Adv. Eng. Pure Sci. 2024, 36(4): <354-366>
DOI: 10.7240/jeps.1519469

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

The Parameter Optimization of Support Vector Machine with Genetic Algorithm in Risk Early Warning Models

Muhammed IŞIK [1]

[1]*Marmara University, Institute of Pure and Applied Sciences, Department of Industrial Engineering, 34722, Istanbul, Turkey*

Abstract
Machine Learning algorithms are widely used by lenders in risk early warning models. With Machine Learning, the risk levels of individual and corporate customers are determined at the account and customer level. Lenders want to manage risk by evaluating the payment performance of customer or account with the help of Machine Learning algorithms. Banks, which have an important place among lenders, develop risk early warning models with the help of learning algorithms using customer information. In the development process of risk early warning models, while banks generally use customer information and credit bureau information for the individual segment, they use financial, non-financial and behaviour-based information for the corporate segment. In this study, it is planned to develop a risk early model for customers in corporate service segment. For the customers of corporate service segment, Balance Sheet and Income Statement items were used and the financial ratios were calculated for risk early warning models. In the development of risk early warning models, Mutual Information method was used as a novel feature selection approach and Support Vector Machine method (linear function, radial basis function and sigmoid function) was used as a supervised learning approach. By changing the neighbourhood metric (k), important patterns were discovered with the Mutual Information method in feature selection process. The optimal C and gamma parameters for Support Vector Machine models have been tried to be determined with the Genetic Algorithm, which is among the Meta-Heuristic algorithms. In order to find the optimal metrics in this study, the metric values for all parameters of the SVM model (function specific) have been kept quite wide. In this dataset of corporate service customers, the small neighbourhood metric has been found to have a significant impact on model learning and performance.
Keywords: Risk Early Warning Models, Mutual Information, Support Vector Machine, Parameter Optimization, Genetic Algorithm

I. INTRODUCTION

In measurement and management of credit risk, Machine Learning (ML) models are frequently used by many institutions, especially banking, leasing, factoring. ML models are quite successful compared to the classic models such as regression models in credit risk [1,2]. Although they are difficult to implement in organizational processes compared to classical models, ML models are generally preferred by managers or decision makers due to their high performance. It is very important to know the use cases, assumptions and how to interpret the results obtained depending on the modelling process of ML models [3]. In the world of credit, ML models are frequently used to predict financial crisis by institutions [4]. ML models are used by financial institutions in many modelling studies such as early warning, disruption and anomaly detection [5]. In different risk modelling studies, financial information obtained from financial statements are used such as Balance Sheet (BALSH), Income Statement (INCSTAT) and Cash Flow (CASHFL). The information obtained from financial statements show how the customer will perform in the future. It is very important that the financial statements received from customers are not made up. Inaccurate financial information complicates the requirements of data quality such as continuity, accuracy and completeness. The validity of the models developed using unreliable information is questioned and the performance of these models is quite low [6].

In the modelling process, financial ratios are obtained by using the information requested from financial institutions. In the model development process, financial ratios are used as explanatory variables for features. In the univariate and multivariate analysis process of the modelling, the valuable financial ratios and the components obtained from the financial ratios are tried to be determined. Various feature selection methods are used in the

Corresponding Author: MUHAMMED IŞIK, **Tel: :** 0(543) 353 58 75, **E-mail:** muhammedisik@marun.edu.tr
Submitted: 20.07.2024, **Revised:** 13.10.2024, **Accepted:** 20.10.2024

determination of important financial ratio variables and components. In feature selection, analysis studies can be carried out in linear space such as dimension reduction, as well as analysis studies in non-linear space such as manifold learning. The main goal of feature selection is to reveal important patterns with various techniques. Significant patterns obtained with feature selection are used as inputs in development of the model pattern. Supervised, unsupervised, semi-supervised and hybrid models are widely used in the development of the ML model pattern. The results of model trials are compared with each other and tested. There are many performance metrics used to compare models in the modelling world. In comparison of models, accuracy, recall, precision, gmean and f1-score metrics are frequently used as well as receiver operating characteristic curve, confusion matrix and lift table. The use of performance metrics may vary taking into structure and distribution of the data set [7]. In models using balanced data sets, accuracy can be used as a performance measure. However, it may be wrong to use only accuracy metric in models using imbalanced data sets. It would be more accurate to use recall, precision and f1-score in models where imbalanced datasets are used. It is known that banks, which are constantly faced with crisis, resort to analytical solutions in order to define and manage credit risk with ML models. In the world of banking, analytical and expert opinion-based scorecards are developed to define and manage credit risk. With the developed scorecards, the performances of customers or accounts are measured, and performance outputs are produced from the scorecard [8]. In scorecards, assigning the customer to the correct risk class is very important for risk management [9]. Early payment of loans and additional collateral may be requested from low-performing customers.

On the other hand, customers with high performance can be offered options by banks such as limit increase and cross-selling. In the credit world, risk early warning scorecards are developed, and analytical decision processes are created within the bank. It is known that risk early warning models make significant contributions to bank credit management. Credit allocation processes will be simplified in terms of time and work process by analysing customer behaviour with risk early warning models. In addition, by using the outputs of these models in the bank's internal systems, the workload of the disruption and monitoring units will be reduced.

II. LITERATURE REVIEW

2.1.Financial ratios

Financial statement analysis, financial intelligence and scoring are studies to measure the financial status of Small and Medium-sized Enterprises (SMEs), companies and institutions. Financial statements provide information about the financial situation of SMEs, companies and institutions at certain periods of each year. In order to show the financial situation, the BALSH is generally shared in the last period of the fiscal year. The BALSH consisting of assets, liabilities and shareholders' equity and has a static structure on the contrary the INCSTAT [10]. The INCSTAT is the statement that shows the net profit or loss status by using the sales and expense items of the company. The INCSTAT is the statement showing the company's performance in the financial year and provides convenience to decision makers in many ways [11]. In financial analysis, BALSH and INCSTAT information is frequently used in determining expert opinions and improving analytical processes. These financial statements not only show the current situation of the companies but also provide information about their future performance.

Financial ratios created using financial statements have an importance on sector basis [12]. By looking at financial ratios, it is possible to comment on the future of the financial institution in the sector and turnover basis. It is very important to make the necessary analyses of the financial statements and to interpret the results correctly. The information obtained from the financial statements affects both the financial institution requesting the loan and the lending institution. The scoring model is developed by evaluating the financial, non-financial and behaviour information of the financial institution with analysis methods. Scoring studies are based on modelling information from institution and non- institution with appropriate model patterns. By using the results of the scoring model, risk warning information are obtained at the customer or account level for institution. Scorecards are frequently used to measure and determine financial performance in risk early warning models. With risk early warning models, closer timely risk measurements can be made compared to traditional credit risk parameters (such as Probability of Default). In particular, the running of the scorecards developed for the early warning model on a daily and weekly basis will be very useful in understanding and managing risks.

2.2. Risk early warning decision systems

Data Science and Big Data have profoundly affected many areas from the individual segment to the corporate segment in financial systems. In risk early warning models, there are significant differences between segments regarding datasets, predictive purposes, and processes. For example, while credit bureau information is important for individual segment customers in risk early warning models, cash flow information may be very important for corporate segment customers. Risk early warning models are developed in order to predict the payment performance and financial status of borrowers [13]. Risk early warning models make predictions about the future performance of the borrower obtaining patterns from historical information. In early warning models,

developments are made by considering that the risk may be disrupted for 30 days or more within 3 or 6 months. Early warning models differ from traditional credit risk modelling techniques where definitions are known. Also, this approach used in creating the disruption status (target variable) may differ among modelling teams. In this respect, risk early warning models are included in business models that progress with the definition of model developers. By using the historical information of the customers, the disruption status is examined within a certain date range in risk early warning models.

Financial, non-financial and behavioural datasets are used in commercial risk early warning models developed by banks. With the help of patterns obtained from financial, non-financial and behavioural datasets, scorecards are developed as a commercial risk early warning model. In commercial risk early warning models, disruption information is estimated before the customer legal follows up [14]. BALSH, INCSTAT and CASHFL statements are included in the financial module data set [15]. In the non-financial module datasets, the partnership status and information of the partners are included. In behaviour module datasets, the customer's performance is examined by obtaining credit payment information at the product or customer level. Risk situations of customers are analysed with risk early warning models and different actions are taken for customers predicted to be risky [16]. Additional collateral and limit reductions can be made for customers who are predicted to be risky in early warning models. In risk warning models, after the model development process, the model template is embedded in bank processes and run at certain periods. By using the patterns obtained from the historical information, the financial changes of the customers are analysed, and decision systems can be developed for the management of the results. It is very important that the decision system established in the risk early warning model is easily implemented into the bank process [17]. In the decision system of the early warning model, not only the predictive power should be considered, but also the intelligibility and ease of use of the model.

III. METHODOLOGY

3.1. Support vector machine

Support Vector Machine (SVM) has gained more attention and adopted in classification and regression problems so as to find a good solution space. SVM is a ML algorithm that wants to create a high-performance model pattern without overfitting problem while developing a model [18]. SVM algorithm aims to create the model pattern by moving the data set from the input space to the feature space with kernel transform functions [19]. In kernel transformation process, SVM algorithm tries to control the margin between the positive hyperplane ($\vec{w}.\vec{x} + b = 1$) and the negative hyperplane ($\vec{w}.\vec{x} + b =$ -1) with support vectors by minimizing the loss function [20,21].

In SVM algorithm, when linear structure is used as kernel transformer for training dataset, linear vectors are used as parser. In this linear approach, which is called the linear SVM model, there is only the C parameter for regularization tuning. If misclassification is acceptable at the end of the model development process, soft margin is selected in regularization tuning. If misclassification is not accepted by the decision maker or modeler, hard margin is selected in model development. In nonlinear approach of SVM, the model tries to draw curvilinear boundaries that can best separate the training datasets. On the other hand, in non-linear SVM algorithms, besides the C regularization parameter, there is also the gamma free parameter [22]. In non-linear SVM algorithms, kernel functions with C and gamma parameters varied according to the training datasets.

3.2. Genetic algorithm

Genetic Algorithm (GA) is a adaptive Meta-Heuristic algorithm based on developed over natural selection inspired by evolution process of genetics [23]. GA is among evolutionary algorithms and GA is used in real life problems such as Traveling Salesman Problem, Network Design Problem, Scheduling Problem, Feature Selection, Data Clustering and Parameter Optimization. In GA, optimal solutions are tried to be determined by applying selection, crossover, mutation and elitism stages on the population. While completing the basic stages of GA, the fitness function is tried to be optimized for each cycle [24]. Especially, while the fitness function is evaluated in selection phase, the most suitable members are directed to the next generation in elitism phase. In each generation of GA, changes are made on the candidate chromosomes through crossover and mutation stages. Fitness function is controlled by using the offsprings created in GA and decisions are made by looking at the criteria of the Meta-Heuristic algorithm. GA is very easy to code in several programming languages and GA is used as an auxiliary model in many different fields.

3.3. Parameter optimization

Parameters have a huge impact on the efficiency and effectiveness in search [25]. In the Parameter Optimization (PO) of ML algorithms, parameter tuning can provide greater flexibility and robustness but requires a good initialization in the tuning process. While developing models, it is very important to determine the initial parameters of the training dataset. In process called hyper parameter tuning, the parameters of the developed models are assigned. PO is a time-consuming process and as the number of parameters increases, it becomes more difficult to determine the optimal values. In PO, iteratively progresses over all values of the parameters used in modelling. In PO, the used program and environment

capacity have also a very important place. In the parameter assignment process, the model success (score) is maximized, and the process is terminated for model patterns. In the PO of ML models, GA is widely used in practical problems that focus on searching for optimal model parameters [26]. First of all, for PO of models, model population (generations) is created with some predefined hyperparameters on training data set [27]. In models, performance metric values are calculated for each model population (generation) such as accuracy ratio, recall, precision, f-1 score, etc. By comparing the values of model performance metrics, the most successful model is tried to be determined with GA.

IV. EMPIRICAL ANALYSIS

4.1. Data description

In corporate segment, risk models are developed by using the ratios obtained from BALSH and INCSTAT items [28]. In this study, service customers were taken as reference to financial information between January 2022 and September 2023 time intervals in order to develop risk early warning model [29]. Financial ratios were calculated by taking financial information of service customers from BALSH and INCSTAT items. In the dataset of this risk study, financial ratios of service customers were used as exploration variables in analysis and classification stages. After financial ratios were obtained for service customers, customers (324 uniq rows) were observed for 1 year and target flags were determined as 'Default' (‘flag1’-128 uniq rows) and 'Live' (‘flag0’-196 uniq rows) in this study. In the credit monitoring process, if bank considers the service customer's performance bad during fiscal period, the reference customer status is assigned as 1 for time interval. In credit life cycle, if service customer performance is not bad, the reference customer status is assigned as 0 for fiscal period. The data set of financial information is shown in Table 1.

Table 1. Exploration variables

Variables	Resource
Current rate	BALSH
Acid-Test ratio	BALSH
Cash rate	BALSH
Stocks to total assets	BALSH
Financial leverage ratio	BALSH
Short-term receivables to total assets	BALSH
Long-term liabilities to total resources	BALSH
Short-term liabilities to total resources	BALSH
Long-term foreign resources to continuous capital	BALSH
Stock values turnover speed	BALSH
Current assets to total assets	BALSH
Receivable turnover speed	INCSTAT
Stock turnover speed	INCSTAT
Fixed assets to equity	INCSTAT
Rotating asset turnover rate	INCSTAT
Fixed asset turnover rate	INCSTAT
Equity turnover	INCSTAT
Total asset turnover rate	INCSTAT
Profitability ratio of equity	INCSTAT
Profit before interest and tax to total resources	INCSTAT
Profitability ratio of total assets	INCSTAT
Financing expenses to net sales	INCSTAT
Gross profit margin	INCSTAT
Net profit margin	INCSTAT
Operating profit margin	INCSTAT

4.2. Feature selection process

In ML world, models developed with a noise-free training data set are easier and more effective to interpret and adapt. In ML studies, feature selection approaches are used to eliminate noise on the training data set [30]. With feature selection approach, it is tried to obtain high quality variables by reducing the noise in training data set. In feature selection, Mutual Information (MI) is an effective method for interdependence degree among variables which is not restricted linear and curvilinear relationships [31]. In the feature selection process where MI is used, the most important features select and ranks them starting with the most relevant [32,33]. In feature selection with MI, Kullback–Leibler (KL) divergence is used to examine whether there is a distributional relationship between the variables X (independent) and y (dependent). The calculation of MI coefficient is shown in Eq.1.

$$MI(X,y) = \sum_y y_i \sum_X X_i \left[P(X,y) \log\left(\frac{P(X,y)}{P(X)P(y)}\right)\right] \quad \textbf{(1)}$$

In the modelling process, all data set is divided into training data set (75%) and testing data set (25%) using stratified sampling approach. MI method was applied to training dataset for feature selection. It has been tried to determine appropriate variables for modelling by changing neighbours parameter (k). In feature selection process, among the 25 variables, the most relevant 10 features were tried to be determined. In feature selection stage, python libraries were widely used such as numpy, pandas and scikit-learn. When important features are examined, it is seen that variables 'Proportion of profitability of equity', 'Profit before interest and tax to total resources' and 'Operating profit margin' are in a strong relationship with target variable. There are strong relationships between equity & profitability variables and target variable in the training data set. At the end of the feature selection process, the most important features according to the k metric are shown in Table 2.

Table 2. The most relevant 10 features

k = 2	k = 3	k = 4
Important features	**Important features**	**Important features**
Profitability ratio of equity	Profitability ratio of equity	Profitability ratio of equity
Operating profit margin	Profit before interest and tax to total resources	Profit before interest and tax to total resources
Profit before interest and tax to total resources	Operating profit margin	Long-term liabilities to total resources
Acid-Test ratio	Financial leverage ratio	Operating profit margin
Fixed assets to equity	Long-term liabilities to total resources	Receivable turnover speed
Receivable turnover speed	Financing expenses to net sales	Fixed assets to equity
Long-term liabilities to total resources	Fixed assets to equity	Current assets to total assets
Net profit margin	Acid-Test ratio	Financial leverage ratio
Long-term foreign resources to continuous capital	Receivable turnover speed	Financing expenses to net sales
Current assets to total assets	Long-term foreign resources to continuous capital	Fixed asset turnover rate

In this modelling process, the SVM as classification algorithm is trained on the training dataset with important features and forms a pattern with the help of the patterns they have learned. In the next step, the SVM model predictions on the test data with this model pattern. Within the scope of this study, SVM model tries were carried out with important features determined according to neighbours k coefficient.

4.3. Experimental setup

In banking, SVM models are especially used to predict various cases, including risk early warning studies. In risk early warning studies, the efficiency of SVM model relies on the correct setting of hyperparameters such as C, gamma and tolerance value. In this paper, different (linear function, radial basis function and sigmoid function) topological approaches of SVM models were applied to training dataset. In model tries with different SVM approaches, C is in ranges [1, 1000], gamma is in ranges [1e-5, 100] and tolerance is ranges in [1e-5, 1e-1].

In this study, in order not to miss the optimum metrics, the metric values for all parameters of the SVM approaches were kept quite wide. In this paper, for each one kernel function and neighbours k coefficient, GA algorithm was used to determine optimal C, gamma and tolerance parameters. For each model trial, tpot tools of python programming language were used in order to determine optimal parameters. Google Colab (GPU)

platform was used in all analysis and modelling studies. The parameters of the GA used for local search are shown in Table 3.

Table 3. Parameters of the GA

Parameter	Value
Number of generations	5
Population size	25
Offspring size	25
Crossover rate	0.1
Mutation rate	0.9
Crossover type	Two-point crossover
Selection method	Elite selection

V. RESULTS AND DISCUSSIONS

In paper, model tries were carried out on the basis of different kernel functions {'linear function', ' radial basis function' and 'sigmoid function'} and neighbours k {2,3,4} coefficients. In each model trial, local search studies were performed on the training dataset via GA. In model trials, optimal parameters (C, gamma and tolerance values) for SVM models were determined by GA local search method. In order to analyse model trials, it is necessary to examine the model in terms of model performance metrics. The performance of classification models such as SVM is generally controlled by model performance metrics derived from the confusion matrix. In classification models, model performance metrics such as True Positive Rate (TPR) and False Positive Rate (FPR) are calculated by using the confusion matrix. Not only the Receiver Operating Characteristic (ROC) curve but also the optimal threshold values (default threshold: 0.5) can be calculated from TPR and FPR metrics.

In this article, optimum threshold values were calculated on the training dataset for each model trial. Recall, precision, f-1 score and accuracy ratio performance metrics were calculated using reference optimal threshold values. Detailed tables containing all performance metrics are in the appendices section. Among the performance metrics of classification models, accuracy ratio is generally used for balanced datasets. In this paper, since the training and testing datasets were balanced in terms of target variable, accuracy ratio referenced as the main model evaluation metric. When the performance metrics of the model tries is analysed in terms of accuracy ratio values of the training and testing datasets, it is seen that the best model is radial basis function (RBF) Kernel (C=654, gamma= 0.01, tolerance value=0.001) SVM model with neighbours k (2) parameter. In fact, when model trials are examined, it is concluded that RBF Kernel SVM model with the neighbours k (2) is the best model in terms of other performance (recall, precision and f1 score) metrics as well. Fig.1. shows the ROC curve of the most successful model among the model trials. For RBF Kernel (C=654, gamma= 0.01, tolerance value=0.001) SVM model with neighbours k (2) parameter, while Fig.2(a) and Fig.2(b) depict the probability distribution on the basis of target flag in the training dataset, Fig.3(a) and Fig.3(b) depict the probability distribution on the basis of target flag in the testing data set.

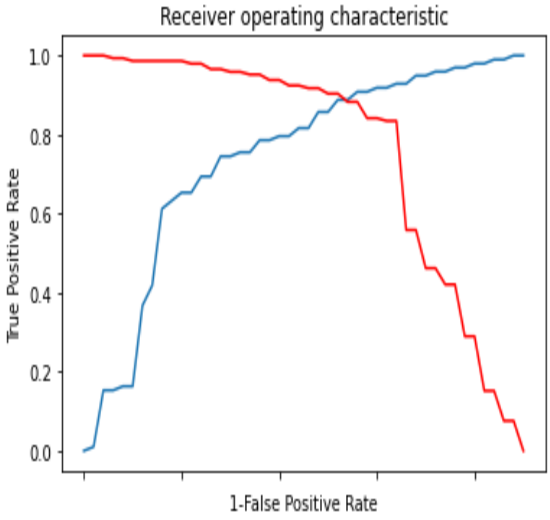

Figure. 1. The ROC curve of RBF kernel SVM model with neighbours k (2) parameter

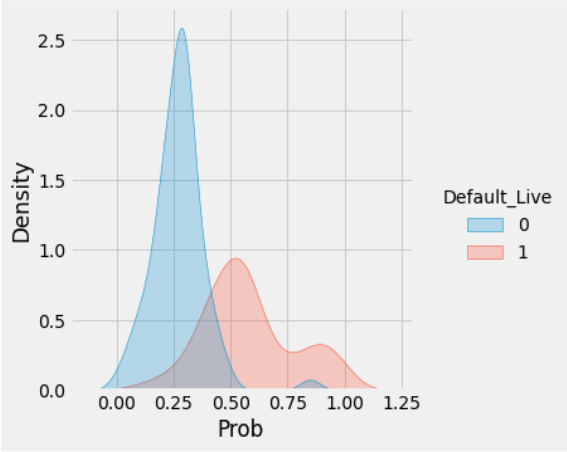

Figure. 2(a). The probability distribution graph in training data set

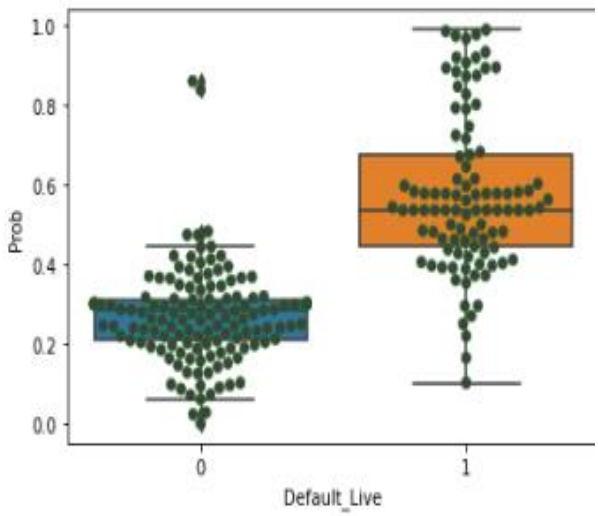

Figure. 2(b). The box and whisher diagram in training data set

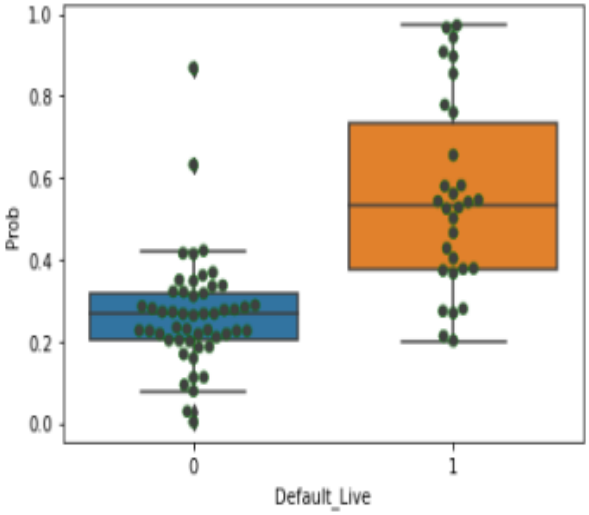

Figure. 3(b). The box and whisher diagram in testing data set

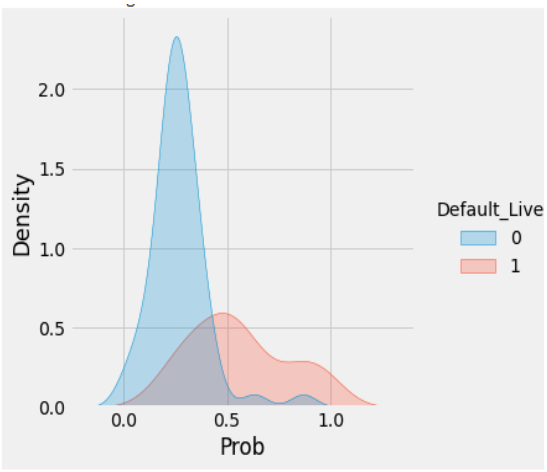

Figure. 3(a). The probability distribution graph in testing data set

For benchmarking, successful models have been identified based on the k parameter and kernel approach. In the paper, not only the GA technique was used, but also search algorithms such as grid search and randomized search were used. Table 4 shows the results of optimization approaches based on the k parameter and kernel approach. When Table 4 is examined, in PO studies, it is concluded that the GA technique is more successful than search algorithms such as grid search and randomized search.

Table 4. Benchmarking studies with Grid Search and Randomized Search

n	Approach	Optimization Techniques	C	Gamma	Tolerance	Accuracy Ratio
k = 2	**Linear**	Genetic Algorithm	13	---	0,01	86,24%
		Grid Search	9	---	0,00001	84,72%
		Randomized Search	8	---	0,001	84,66%
	RBF	Genetic Algorithm	654	0,01	0,001	87,45%
		Grid Search	56	0,01	0,00001	86,35%
		Randomized Search	12	0,001	0,01	86,12%
	Sigmoid	Genetic Algorithm	73	0,001	0,00001	62,46%
		Grid Search	15	0,001	0,00001	61,90%
		Randomized Search	12	0,01	0,0001	61,57%

Table 4. Benchmarking studies with Grid Search and Randomized Search (cont.)

		Genetic Algorithm	12	---	0,0001	85,76%
	Linear	Grid Search	8	---	0,00001	82,75%
		Randomized Search	8	---	0,01	82,46%
		Genetic Algorithm	763	0,001	0,001	86,17%
k = 3	**RBF**	Grid Search	79	0,001	0,0001	85,33%
		Randomized Search	21	0,001	0,01	85,12%
		Genetic Algorithm	76	0,001	0,001	62,12%
	Sigmoid	Grid Search	20	0,001	0,0001	61,21%
		Randomized Search	14	0,01	0,01	61,02%
		Genetic Algorithm	8	---	0,00001	84,68%
	Linear	Grid Search	5	---	0,00001	82,23%
		Randomized Search	3	---	0,1	82,12%
		Genetic Algorithm	752	0,01	0,001	85,90%
k =4	**RBF**	Grid Search	88	0,001	0,00001	85,44%
		Randomized Search	26	0,01	0,001	85,05%
		Genetic Algorithm	64	0,001	0,001	61,06%
	Sigmoid	Grid Search	50	0,001	0,00001	60,89%
		Randomized Search	50	50	0,00001	60,65%

VI. CONCLUSIONS

PO has vital importance to the model development processes of supervised and unsupervised learning. Especially, in supervised learning algorithms, model performance varies depending on the dataset and algorithm parameters. High-performance models can be developed by determining optimal parameters on training datasets. In this study, optimal parameters for SVM models were determined by using the corporate service customer dataset. Local search tools are frequently used such as grid search and random search in parameter determination. Apart from local search algorithms, Meta-Heuristic algorithms have started to be used such as GA for PO. With Meta-Heuristic algorithms, optimal coordinates are determined, and appropriate parameters are detected in the solution space. In this study, training datasets were created by grouping the features determined by variable selection. In the feature selection of risk early warning model, MI method as a novel method was used. In the next stage, the parameters of different kernel functions on the SVM algorithm were tried to be determined using GA. Instead of the 0.5 threshold (default) value for SVM-based candidate models, the optimal threshold value as a novel approach was calculated with TPR and FPR. This study has some limitations. The desire of corporate service companies not to share their data prevents the increase in the number of rows. The conservative approach of corporate service companies negatively affects the knowledge discovery process in models. In the study, BALSH and INCSTAT were used as financial dataset. In future studies, the model process can be developed by using tables such as cash flow in addition to BALSH and INCSTAT. In addition, the modelling process can be enriched by using non-financial and behavioural data as well as financial data in the modelling process.

REFERENCES

[1] Farooq, U., Jibran Qamar, M. A., & Haque, A. (2018). A three-stage dynamic model of financial distress. Managerial Finance, 44(9), 1101–1116. https://doi.org/10.1108/MF-07-2017-0244

[2] Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 197, 116659. https://doi.org/10.1016/j.eswa.2022.116659

[3] Geršl, A., & Jašová, M. (2018). Credit-based early warning indicators of banking crises in emerging markets. Economic Systems, 42(1), 18–31. https://doi.org/10.1016/j.ecosys.2017.05.004

[4] Shen, C., Lee, Y., & Fang, H. (2020). Predicting banking crises based on credit, housing and capital booms. International Finance, 23(3), 472–505. https://doi.org/10.1111/infi.12367

[5] Zhang, C., Wang, Z., & Lv, J. (2022). Research on early warning of agricultural credit and guarantee risk based on deep learning. Neural Computing and Applications, 34(9), 6673–6682. https://doi.org/10.1007/s00521-021-06114-3

[6] Feng, Q., Chen, H., & Jiang, R. (2021). Analysis of early warning of corporate financial risk via deep learning artificial neural network. Microprocessors and Microsystems, 87, 104387. https://doi.org/10.1016/j.micpro.2021.104387

[7] Kou, G., Xu, Y., Peng, Y., Shen, F., Chen, Y., Chang, K., & Kou, S. (2021). Bankruptcy prediction for SMEs using transactional data and two-stage multiobjective feature selection. Decision Support Systems, 140, 113429. https://doi.org/10.1016/j.dss.2020.113429

[8] Du, G., Liu, Z., & Lu, H. (2021). Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment. Journal of Computational and Applied Mathematics, 386, 113260. https://doi.org/10.1016/j.cam.2020.113260

[9] Wen, C., Yang, J., Gan, L., & Pan, Y. (2021). Big data driven Internet of Things for credit evaluation and early warning in finance. Future Generation Computer Systems, 124, 295–307. https://doi.org/10.1016/j.future.2021.06.003

[10] Rosa, N. L. (2020). Analysing Financial Performance: Using Integrated Ratio Analysis (1st ed.).Routledge. https://doi.org/10.4324/9781003092575

[11] Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of Financial Statement Fraud and Feature Selection using Data Mining Techniques. Decision Support Systems, 50(2), 491–500. https://doi.org/10.1016/j.dss.2010.11.006

[12] Zhang, J. (2020). Investment risk model based on intelligent fuzzy neural network and VaR. Journal of Computational and Applied Mathematics, 371, 112707. https://doi.org/10.1016/j.cam.2019.112707

[13] Lin, M. (2022). Innovative Risk Early Warning Model under Data Mining Approach in Risk Assessment of Internet Credit Finance. Computational Economics, 59(4), 1443–1464. https://doi.org/10.1007/s10614-021-10180-z

[14] Lahmiri, S., Bekiros, S., Giakoumelou, A., & Bezzina, F. (2020). Performance assessment of ensemble learning systems in financial data classification. Intelligent Systems in Accounting, Finance and Management, 27(1), 3–9. https://doi.org/10.1002/isaf.1460

[15] Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine Learning Techniques for Credit Risk Evaluation: A Systematic Literature Review. Journal of Banking and Financial Technology, 4(1), 111–138. https://doi.org/10.1007/s42786-020-00020-3

[16] Zhang, W., He, H., & Zhang, S. (2019). A Novel Multi-Stage Hybrid Model with Enhanced Multi-Population Niche Genetic Algorithm: An Application in Credit Scoring. Expert Systems with Applications, 121, 221–232. https://doi.org/10.1016/j.eswa.2018.12.020

[17] Catullo, E., Gallegati, M., & Palestrini, A. (2015). Towards a credit network based early warning indicator for crises. Journal of Economic Dynamics and Control, 50, 78–97. https://doi.org/10.1016/j.jedc.2014.08.011

[18] Luo, C., Wu, D., & Wu, D. (2017). A deep learning approach for credit scoring using credit default swaps. Engineering Applications of Artificial Intelligence, 65, 465–470. https://doi.org/10.1016/j.engappai.2016.12.002

[19] Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. Expert Systems with Applications, 193, 116429. https://doi.org/10.1016/j.eswa.2021.116429

[20] Bequé, A., & Lessmann, S. (2017). Extreme Learning Machines for Credit Scoring: An Empirical Evaluation. Expert Systems with Applications, 86, 42–53. https://doi.org/10.1016/j.eswa.2017.05.050

[21] Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services, 27, 11–23. https://doi.org/10.1016/j.jretconser.2015.07.003

[22] Nguyen, M. H., & de la Torre, F. (2010). Optimal feature selection for support vector machines. Pattern Recognition, 43(3), 584–591. https://doi.org/10.1016/j.patcog.2009.09.003

[23] Jadhav, S., He, H., & Jenkins, K. (2018). Information Gain Directed Genetic Algorithm Wrapper Feature Selection for Credit Rating. Applied Soft Computing, 69, 541–553. https://doi.org/10.1016/j.asoc.2018.04.033

[24] Vijayanand, R., Devaraj, D., & Kannapiran, B. (2018). Intrusion Detection System for Wireless Mesh Network using Multiple Support Vector Machine Classifiers with Genetic-Algorithm-Based Feature Selection. Computers & Security, 77,304–314. https://doi.org/10.1016/j.cose.2018.04.010

[25] Talbi, E.-G. (2009). Metaheuristics: From Design To Implementation. John Wiley & Sons.

[26] Manurung, J., Mawengkang, H., & Zamzami, E. (2017). Optimizing Support Vector Machine Parameters with Genetic Algorithm for Credit Risk Assessment. Journal of Physics: Conference Series, 930, 012026. https://doi.org/10.1088/1742-6596/930/1/012026

[27] İlhan, İ., & Tezel, G. (2013). A genetic algorithm–support vector machine method with parameter optimization for selecting the tag SNPs. Journal of Biomedical Informatics, 46(2), 328–340. https://doi.org/10.1016/j.jbi.2012.12.002

[28] Onay, C., & Öztürk, E. (2018). A review of credit scoring research in the age of Big Data. Journal of Financial Regulation and Compliance, 26(3), 382–405. https://doi.org/10.1108/JFRC-06-2017-0054

[29] Işık, M. (2023). Dataset. Işık, Muhammed (2023), "Early Warning Model Dataset for Corporate Segment", Mendeley Data, V5, doi: 10.17632/pp599dy9c8.5

[30] Bishop, C.M., 2006. Pattern recognition and machine learning, Information science and statistics. Springer, New York

[31] Su, X., Li, L., Shi, F., & Qian, H. (2018). Research on the Fusion of Dependent Evidence Based on Mutual Information. IEEE Access, 6, 71839–71845. https://doi.org/10.1109/Access.2018.2882545

[32] Barraza, N., Moro, S., Ferreyra, M., & de la Peña, A. (2019). Mutual Information and Sensitivity Analysis for Feature Selection in Customer Targeting: A Comparative Study. Journal of Information Science, 45(1), 53–67. https://doi.org/10.1177/0165551518770967

[33] Yan, C., Kang, X., Li, M., & Wang, J. (2021). A Novel Feature Selection Method on Mutual Information and Improved Gravitational Search Algorithm for High Dimensional Biomedical Data. 2021 13th International Conference on Computer and Automation Engineering (ICCAE), 24–30. https://doi.org/10.1109/ICCAE51876.2021.9426130

APPENDICES

	Kernel Approach	FPR	TPR	1-FPR	TF	Optimal Threshold	The Best Parameter (C)	The Best Parameter (gamma)	Tolerance Value
neighbours (2)	SVM Linear Model	0.1379	0.8571	0.8621	-0.005	0.3856	13	---	0.01
	SVM RBF Model	0.1172	0.8878	0.8828	-0.005	0.3759	654	0.01	0.001
	SVM Sigmoid Model	0.3655	0.6227	0.6345	-0.002	0.4244	73	0.001	0.00001

	SVM Linear Model	Target Flag	Precision	Recall	F1-Score	Accuracy
neighbours (2)	Training Dataset	Live	0.90	0.86	0.88	0.86
		Default	0.81	0.86	0.83	
	Testing Dataset	Live	0.87	0.90	0.88	0.85
		Default	0.82	0.77	0.79	

	SVM RBF Model	Target Flag	Precision	Recall	F1-Score	Accuracy
neighbours (2)	Training Dataset	Live	0.92	0.88	0.90	0.88
		Default	0.84	0.89	0.86	
	Testing Dataset	Live	0.87	0.90	0.88	0.85
		Default	0.82	0.77	0.79	

	SVM Sigmoid Model	Target Flag	Precision	Recall	F1-Score	Accuracy
neighbours (2)	Training Dataset	Live	0.72	0.63	0.67	0.63
		Default	0.54	0.63	0.58	
	Testing Dataset	Live	0.68	0.71	0.69	0.60
		Default	0.46	0.43	0.45	

	Kernel Approach	FPR	TPR	1-FPR	TF	Optimal Threshold	The Best Parameter (C)	The Best Parameter (gamma)	Tolerance Value
neighbours (3)	SVM Linear Model	0.1586	0.8469	0.8613	0.005	0.4022	460	---	0.01
	SVM RBF Model	0.1241	0.8775	0.8758	0.001	0.3853	327	0.01	0.00001
	SVM Sigmoid Model	0.4482	0.5612	0.5517	0.009	0.4098	113	0.1	0.1

neighbours (3)	SVM Linear Model	Target Flag	Precision	Recall	F1-Score	Accuracy
	Training Dataset	Live	0.89	0.84	0.87	0.84
		Default	0.78	0.85	0.81	
	Testing Dataset	Live	0.86	0.86	0.86	0.82
		Default	0.77	0.77	0.77	

neighbours (3)	SVM RBF Model	Target Flag	Precision	Recall	F1-Score	Accuracy
	Training Dataset	Live	0.91	0.88	0.89	0.87
		Default	0.83	0.88	0.85	
	Testing Dataset	Live	0.87	0.88	0.87	0.84
		Default	0.79	0.77	0.78	

neighbours (3)	SVM Sigmoid Model	Target Flag	Precision	Recall	F1-Score	Accuracy
	Training Dataset	Live	0.65	0.55	0.59	0.55
		Default	0.45	0.55	0.50	
	Testing Dataset	Live	0.68	0.59	0.63	0.56
		Default	0.43	0.53	0.48	

neighbours (4)	Kernel Approach	FPR	TPR	1-FPR	TF	Optimal Threshold	The Best Parameter (C)	The Best Parameter (gamma)	Tolerance Value
	SVM Linear Model	0.1517	0.8469	0.8482	-0.001	0.3894	274	---	0.1
	SVM RBF Model	0.1241	0.8673	0.8758	-0.008	0.3640	441	0.001	0.01
	SVM Sigmoid Model	0.4965	0.5000	0.5034	-0.003	0.4064	586	0.0001	0.00001

neighbours (4)	SVM Linear Model	Target Flag	Precision	Recall	F1-Score	Accuracy
	Training Dataset	Live	0.88	0.85	0.87	0.84
		Default	0.79	0.84	0.81	
	Testing Dataset	Live	0.88	0.90	0.89	0.86
		Default	0.83	0.80	0.81	

neighbours (4)	SVM RBF Model	Target Flag	Precision	Recall	F1-Score	Accuracy
	Training Dataset	Live	0.90	0.88	0.89	0.87
		Default	0.82	0.86	0.84	
	Testing Dataset	Live	0.86	0.84	0.85	0.81
		Default	0.74	0.77	0.75	

	SVM Sigmoid Model	Target Flag	Precision	Recall	F1-Score	Accuracy
neighbours (4)	Training Dataset	Live	0.60	0.50	0.55	0.50
		Default	0.40	0.50	0.45	
	Testing Dataset	Live	0.67	0.43	0.52	0.50
		Default	0.40	0.63	0.49	