International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7402
***Article Received**: 03 May 2023 **Revised:** 29 June 2023 **Accepted:** 23 July 2023*

A Framework for Credit Risk Prediction Using the Optimized-FKSVR Machine Learning Classifier

Usha Devi[1], Dr. Neera Batra[2]
[1]Maharishi Markandeshwar (Deemed To Be University) Mullana
Ambala Haryana, India.
e-mail: usha16feb@gmail.com
[2]Maharishi Markandeshwar (Deemed To Be University) Mullana
Ambala Haryana, India
e-mail: neera.batra@mmumullana.org

Abstract— Transparency is influenced by several crucial factors, such as credit risk (CR) predictions, model reliability, efficient loan processing, etc. The emergence of machine learning (ML) techniques provides a promising solution to address these challenges. However, it is the responsibility of banking or nonbanking organizations to control their approach to incorporate this innovative methodology to mitigate human preferences in loan decision-making. The research article presents the Optimized-Feature based Kernel Support Vector Regression (O-FKSVR) model which is an ML-based CR analysis model in the digital banking. This proposal aims to compare several ML methods to identify a precise model for CR assessment using real credit database information. The goal is to introduce a classification model that uses a hybrid of Stochastic Gradient Descent (SGD) and firefly optimization (FFO) methods with Support Vector Regression (SVR) to predict credit risks in the form of probability, loss given, and exposure at defaults. The proposed O-FKSVR model extracts features and predicts outcomes based on data gathered from online credit analysis. The proposed O-FKSVR model has increased the accuracy rate and resolved the existing problems. The experimental study is conducted in Python, and the results demonstrate improvements in accuracy, precision, and reduced error rates compared to previous ML methods. The proposed O-FKSVR model has achieved a maximum accuracy rate value of 0.955%, precision value of 0.96%, and recall value of 0.952%, error rate value of 4.4 when compared with the existing models such as SVR, DT, RF, and AdaBoost.

Keywords- Credit Risk Analysis (CRA), Machine Learning (ML), Optimized-Feature based Kernel Support Vector Regression (O-FKSVR), Firefly Optimization (FFO), Stochastic Gradient Descent (SGD).

I. INTRODUCTION

The goals of credit risk supervision in banks are essential to achieve the credit risk characteristic of the whole group and the hazard of distinct credits or communications. Banks also have to consider the associations between CR and other risks. The active supervision of CR is a compulsory component of an inclusive risk management method and is crucial to the long-term achievement of any finance association [1]. The term "risk" must be defined as everything that might present a threat or restrict the ability of e-business. It is an unexpected or projected occurrence that is dangerous or can limit an organization's capacity. Specific danger methods are physical, such as building destruction, fires, financial loss, and theft [2]. This analysis in financial organizations has become increasingly significant due to the prevalence of commercial transactions. However, it is also crucial for managing the risk of default in both booming and struggling economies. Credit risk analysis is a fundamental concept that governs commercial performance, sustainable development, and reliable operations. Financial institutions' survival and ability to provide services to clients for generating revenue, gaining a competitive advantage, and meeting investors' expectations depend on these facilities. Loan services are provided to clients at an agreed-upon interest rate and under specific terms and conditions. In the past, financial institutions and banks relied on traditional methods to assess the creditworthiness of clients [3]. However, with the advancement of technology and data analytics, credit risk analysis has evolved significantly. By using advanced algorithms and statistical models, financial institutions can now analyze large amounts of data to identify potential risks and assess the creditworthiness of clients more accurately. It has enabled financial institutions to make informed lending decisions and manage credit risk more effectively.

The major possibilities now faced by profitable banks, as represented in Figure 1, include credit, legal, market, interest rate, and operational risks. The significance of these risks denotes the risk that borrowers may default on their contractual responsibilities and fail to repay bank loans or debts in full and on time. The key drivers of credit risk are external macroeconomic issues and the precise internal operating environment of the business. The company's financial position in these two areas is reflected in its financial statements [4]. With the growth of the customer base and technological advancements, traditional mathematical approaches for handling large amounts of data have transitioned to machine learning (ML) procedures. Adopting cost-effective and efficient techniques has allowed banks to manage credit risk more effectively. Effectively managing credit risk is vital for the existence and development of financial organizations. It requires a robust credit risk management system incorporating advanced data analytics and machine learning techniques. By identifying potential risks and assessing the creditworthiness of clients accurately, banks can make informed lending decisions and manage CR effectively. It is an important risk faced by commercial banks. Effective CR management is vital for financial organizations' survival and growth. By adopting advanced data analytics and machine learning techniques, banks can manage credit risk more effectively and make informed lending decisions.

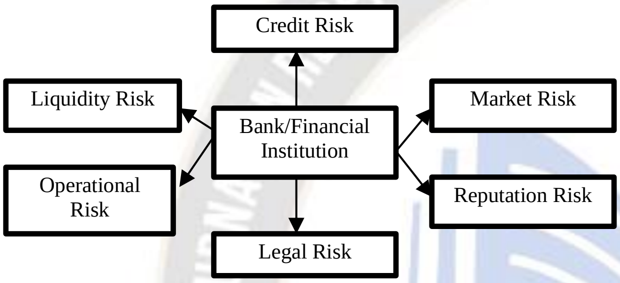

Fig. 1 Several Risks Exist in Banking Sector [5]

Various risks occur in the banking sector, but credit risk is the most significant due to the higher risks resulting from some of the characteristics of clients and the business environments they operate in. Banks must handle CR well because it is essential to the loan application process. To protect the bank from the negative consequences of CR, it maximizes its risk-adjusted return by minimizing the CR experience. The connection between CR and the presentation of profitable banks has been the focus of several revisions, indicating that CR is one of the main features affecting the effectiveness and production of good banks. In Ethiopia, credit risk is the main challenge facing bank performance [5]. Banks rely on various credit risk estimation methods to manage credit risk and categorize credit applicants into decent and corrupt classes. Applicants in proper courses are more likely to repay the bank, while those in evil classes have a lower chance of repayment and are considered high-risk debtors. Accurate credit risk analysis is essential in banking, and a significant challenge banks face. The growth in the debtor's ratio in the credit risk dataset further motivates the development and use of reliable credit risk estimation methods to avoid financial loss. The assistance of consistent credit risk analysis reduces the cost of credit scoring and enables banks to make informed lending decisions, minimizing credit risk exposure and avoiding the adverse effects of credit risk.

Numerous researchers have dedicated themselves to CR analysis, calculation, and management. **Jiang et al. [6]** used random forest (RF) and persistence analysis to create a forecast-driven crossbreed model. They empirically tested the typical peer-to-peer (P2P) lending data and achieved a high accuracy rate. **Liu et al. [7]** introduced an improved GB-DT method for credit scoring. The technique is built on two trees, which enhance the variety of a single base classifier and retain the interpretability of GB-DT methods. **Zhang et al. [8]** used a voting method for outlier detection, a bagging procedure for sampling, and method for collective modeling to estimate CR. The model's performance was evaluated on five UC Irvine ML repository databases, and the investigational outcomes indicate a higher version of the planned method.

The main involvement of the research work is:

- A comparative analysis of various existing credit risk prediction models such as RF, DT, GB, etc.
- To optimize the existing approaches for credit risk prediction with hybrid Firefly-SGD algorithms. The Hybrid firefly-SGD method was used to reduce the selected feature to pass on to the classification model to detect the risks in credit.
- To develop and implement a framework for credit risk prediction using the optimized-FKSVR (a hybrid of PCA and kernel SVR classification) machine learning classifier. The proposed O-FKSVR model combines different layers: feature extraction, feature selection, and classification. The initial layer extracts the feature sets as a kernel PCA method. After that, the next layer works as an optimizer to reduce the error probability and enhance the detection rate of accuracy. This optimizer layer works as a firefly algorithm to select the feature sets, and the fitness function has calculated the best score value to get the final chosen features. The last layer works as a classification model using the SVR method. The proposed O-FKSVR model has been used to classify credit risks in probability: a loss is given, and exposure at defaults. The proposed model has attained a high accuracy rate and reduced existing problems, such as errors.
- Verify and validate the proposed framework using several performance evaluation parameters such as accuracy, precision, etc.

The following sections are described as trails: Section 2 offers an overview of several existing ML and DL models and different feature extraction methods. Section 3 presents the proposed work, including dataset description, pre-processing techniques, feature extraction, selection, and classification methods. The simulation setup and tools used are discussed in Section 4. Lastly, Section 5 accomplishes the research work and highlights areas for further improvement.

II. RELATED WORK

This section describes the various existing investigations on credit risk analysis or prediction models, tools, and methods that have been widely considered. The significance of economic, operational, improvement, and undesirable action evidence has been established in forecasting. The literature has primarily focused on statistical and ML methods [9][10][11][12]. **Apostolos Ampountolas et al. (2021) [13]** proposed a comparative study using machine learning (ML) methods on actual micro-lending data to evaluate their effectiveness in categorizing customers into credit categories. The random forest (RF) classifier was the most effective multi-class classifier. RF performed well using only customer attributes such as age, profession, and location. This approach provides a low-cost and reliable way for microlending institutions worldwide to assess creditworthiness without relying on credit history or complete credit databases. **Trilok Nath Pandey et al. (2017) [14]** described the banking trade as the main movement of loaning money to those who need money. The collection represents the principal borrowers' interest to pay back the amount borrowed from the depositor. CR analysis is a significant area of financial risk management. Several CR analysis methods were used to evaluate the CR of the customer database. Assessing the credit risk database information to decide whether to approve or reject a customer's loan application is a challenging task that involves a deep investigation of the user's credit database or the statistics provided by the user. The proposed method surveys several methods for the CR analysis utilized to estimate the CR datasets. **Yong Hu et al. (2022) [15]** proposed a credit risk analysis investigation based on fourteen commercial indicators to create a credit risk assessment structure constructed on the outcomes of previous pieces of training. They also combined cluster and factor analysis to control the tangible recognition score of the model statistics. This investigation offered data classification labels and integrated recognized and associated traditional binary CR forecast models. Additionally, they evaluated three commonly utilized ANN models and finally compared their prediction performance. **Henry Ivan Condori-Alejo et al. (2020) [16]** described the micro-credit mechanism and its significant components, which played an essential role in advancing the Peruvian rural economy. Micro-credit institutes evaluated the rural population, reducing the high-risk directory conventionally organized through rural commercial consultants. The consultants assessed and verified the customers' demand for these microcredits. The authors recommended a model that presented the finest level of insistence for the micro-credit valuation procedure, constructed based on the study of rural variables gathered from the literature. This model assisted rural business advisors in making decisions to lower the CR of the rural microfinance organization.

The authors utilized the pre-processing process and the micro-finance division estimate. The proposed model achieved an accuracy rate of 93.72% compared to other methods. **Emanuele Dri et al. (2023) [17]** developed a risk system for each ability in a selection, allowing it to reflect several general risk issues. This results in a more accurate and complex model for each strength's default prospect. They improved the loss-specified avoidance input by eliminating the control of using only whole number standards, which permits using actual statistics from the commercial sector to establish impartial benchmarking properties. The projected irregularity of the credit-risk analysis quantum method was reported as an important limitation of existing methods. It highlighted an increased budget in terms of circuit complexity and measurement. The proposed model achieved better performance compared to other methods. **Guina Sotomayor Alzamora et al. (2022) [18]** proposed a Peruvian system based on professional evaluations. The projected method uses numerous ML models to capture the utmost importance of the credit permitting procedure and the resulting credit risk fall.

At last, the proposed model reached 96.20% accuracy using the Light GBM model. **Jingyuan Li et al. (2022) [19]** described the enhancement of the economic market; the credit risk matter regarding recorded companies had developed gradually. So, forecasting the credit danger of registered companies was a crucial alarm for banks, supervisors, and depositors. The generally utilized models were the Z-score, logistic regression (LR), kernel-based virtual machine (KVM), and NN methods. But, the outcomes were more acceptable. The authors implemented a CR forecast model for enumerated concerns, constructed on a CNN-LSTM and a care apparatus. The projected model was built with the advantages of the LSTM model for more long-period series forecasts composed through the CNN model. The authors presented a consideration tool to allocate weights autonomously and enhance the model to diminish difficulties. The investigation of the CR forecast of the citation method had an important significance. **Liukai Wang et al. (2022) [20]** constructed new prediction models using a disparity sampling approach based on machine learning (ML) methods. They used these cutting-edge models to forecast SMEs' credit risk in China using predictors such as economic and task-related data, performance metrics, and negative events. The proposed models' outcomes indicate that economic data-based models were the most effective in forecasting the CR of SMEs in supply chain finance. The multiple-source evidence fusion was considered significant in improving forecasting credit risk. In accumulation, the ideal CSL-RF model is constructed, which covers cost-sensitive skills using an RF method. Table I analyzes various existing credit risk prediction methodologies that describe research tools and performance metrics.

TABLE I COMPARISON ANALYSIS OF DIFFERENT MODELS

Author Name	**Comparison Methods**	**Tools**	**Parameter With Values**

Apostolos et al. (2021) [13]	• MLP
	• XGBoost
	• Adaboost
	• Random Forest
		Python	Acc = 71.5 %
			Prec = 80%
			Rec = 68%
			F1-Score=74%

Trilok et al. (2017) [14]	• ML methods • Ensemble methods	MATLAB	Accuracy = 96.35%
Yong Hu et al. (2021) [15]	• BP neural network (NN) • Radial basis function	Python	Accuracy = 98.8%
Henry Ivan et al. (2020) [16]	• LR • ANN • SVM • KNN	Python	Accuracy = 93.72%
Guina et al. (2022) [18]	• ML, • SMOTE, and • K-fold methods	Python	Accuracy = 96.20%
Jingyuan Li et al. (2022) [19]	• Logistic • KVM • SVM method	*	Accuracy = 98.43%

Liukai et al. (2022) [20]	• SVM
	• DT
	• RF
	• GB
	• NN
	• Bagging.
		*	Recall =95.29%
			AUC = 64.90%

Acronyms: NN(neural network), LR(logistic regression),DT(decision tree),GB(gradient boost), ML(machine learning), DL (deep learning), ANN(artificial neural network), KNN(k-nearest neighbour), SVM(support vector machine), RF(random forest), MLP(multilayer perceptron layer)Acc (accuracy), Pre (precision), Rec(recall).

III. PROPOSED CREDIT RISK PREDICTION USING AN OPTIMIZED MACHINE LEARNING SYSTEM

In this section, we represent the CR forecast dataset description, mathematical formulas, proposed models, and proposed methodology of the several ML (machine learning) methods defined in this article. This article's ML model prediction parameters are based on the test set.

A. Dataset: Credit Risk Analysis

We aim to analyze the leading records from 2007–2010 [25] and develop a model that can acutely classify and predict the loan repayment behavior of borrowers. The research study identified a '*.csv" file with missing values, which has since been updated to remove those values.

The columns include definitions for credit policy, purpose, installment, interest rate, etc.

- Credit_policy → 1 (if the user meets the credit financing scenario of LendingClub.com), and 0 (Otherwise).
- Determination →The main motive of the loan (proceeds values " Major Purchase", "Debt. Consolidation", "Credit Card"," Educational"," Small Bussiness", and "all others").
- Interest Rate → A loan interest rate, as a proportion (of 11 percent would be saved as 0.11). Debtors refereed by LeadingClub.com to be more dangerous are allocated maximum Int_rates.
- Installment → Monthly installments payable by the debtor if the loan is funded.
- Log Annual Income → Natural log of the self-reported annual income of the debtor.
- DTi → Debit-to-income ratio of the debtor.
- Fico → fico credit_score of the debtor.
- Day with CR line → no. of days the debtor has a credit line.
- Revol balance → debtor revolving balance.
- Revol.util → debtor revolving line utilization rate.
- Inq Last 6 Months → no. of inquiries (debtor) by creditors in the last six months.
- Delinq 2 years → no. of time the debtor had been 30+ days past due on a payment in the previous 2 years.
- Pub Rec → no. of derogatory (debtor) public records like taxlines, judgments, etc.

B. Performance Metrics

This research section presents the formulas used as performance parameters for the credit-risk prediction model, including precision, recall, and accuracy, F1-score, and MSE rate.

- **Precision:** It procedures true positives (TP) to all positives documented through the model to amount to the models' reliability while perceiving the TPs. It is well-defined as the ratio of TN and the sum of TN and FN.
  It is also well-defined as eq (1);

$$precision = \frac{TP}{TP + FP} \tag{1}$$

- **Recall:** It is described as the ratio of the sum of appropriately classified positive data to the sum of positive data. It must be as extraordinary as potential. It is also well-defined as the ratio of total positive divided by the sum of TP and false negative (FN).
  It is also well-defined as eq (2);

Article Received*: 03 May 2023* ***Revised:*** *29 June 2023* ***Accepted:*** *23 July 2023*

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \qquad (2)$$

- **Accuracy Rate:** It is the ratio of the total sum of TP and true negative (TN) and the total addition of all true positive negatives and all FN and false positives (FP).

It is also defined in eq (3)

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \qquad (3)$$

- **F1-Score:** Relating binary models using high precision (P) and high recall (R) simultaneously can be challenging. Therefore, to balance their performance and to better evaluate the models, we use the F1-score. It permits the dimension of the Recall (R) and Precision (P) morals at a similar time. It practices the harmonic mean instead of the arithmetic mean, penalizing the extreme standards added.

It is also defined in eq (4).

$$F1-Score = \frac{2*P*R}{P+R} \qquad (4)$$

- **MSE Rate (Error Rate):** It processes the sum of errors in arithmetical models. It evaluates the average squared difference between the experimental values (yk) and the predicted values (y_k). If a model has no error, then MSE is equivalent to zero. As system error rises, its rate rises. The MSE is also identified as the mean squared deviation (MSD). Total experimental values denote N.

It is also defined in eq (5).

$$MSE = \frac{\sum(yk+y\_k)^2}{N} \qquad (5)$$

C. Proposed Methodology

This section describes a novel framework for predicting and assessing credit risk using machine learning methods. Figure 2 defines the proposed flowchart of the research work. The proposed model will explain the different steps, such as (i) Credit risk-based data gathering, (ii) Credit risk pre-processed dataset values, (iii) Feature Engineering, and (iv) Prediction Model.

The research model has collected the dataset from an online site for further pre-processing. This step deals with data preparation for use in modeling. It has removed the missing values and normalized the data within a specific range. After data pre-processing, feature engineering comes next. This feature engineering process has been divided into feature extraction and selection. This proposed model uses the PCA (principal component analysis) method for feature extraction. This extraction process has converted, normalized, and evaluated the covariance matrix. After determining the covariance matrix, the Eigen matrix is calculated, which includes the Eigenvalues (E) and Eigenvectors (V). The Eigen matrix is then sorted in decreasing order. In the next step, the transformation matrix is calculated to develop a feature selection using a hybrid of Stochastic Gradient Descent and Firefly optimization algorithms. This feature selection has reduced the feature sets with the help of the fitness function. This fitness function has calculated the best score of the feature sets. This process has improved the reliability of the data by cleaning the dataset and selecting a subset of data features. A credit risk prediction system has collected the relevant feature sets to improve the organization's accuracy and minimize the computation costs connected with machine learning models.

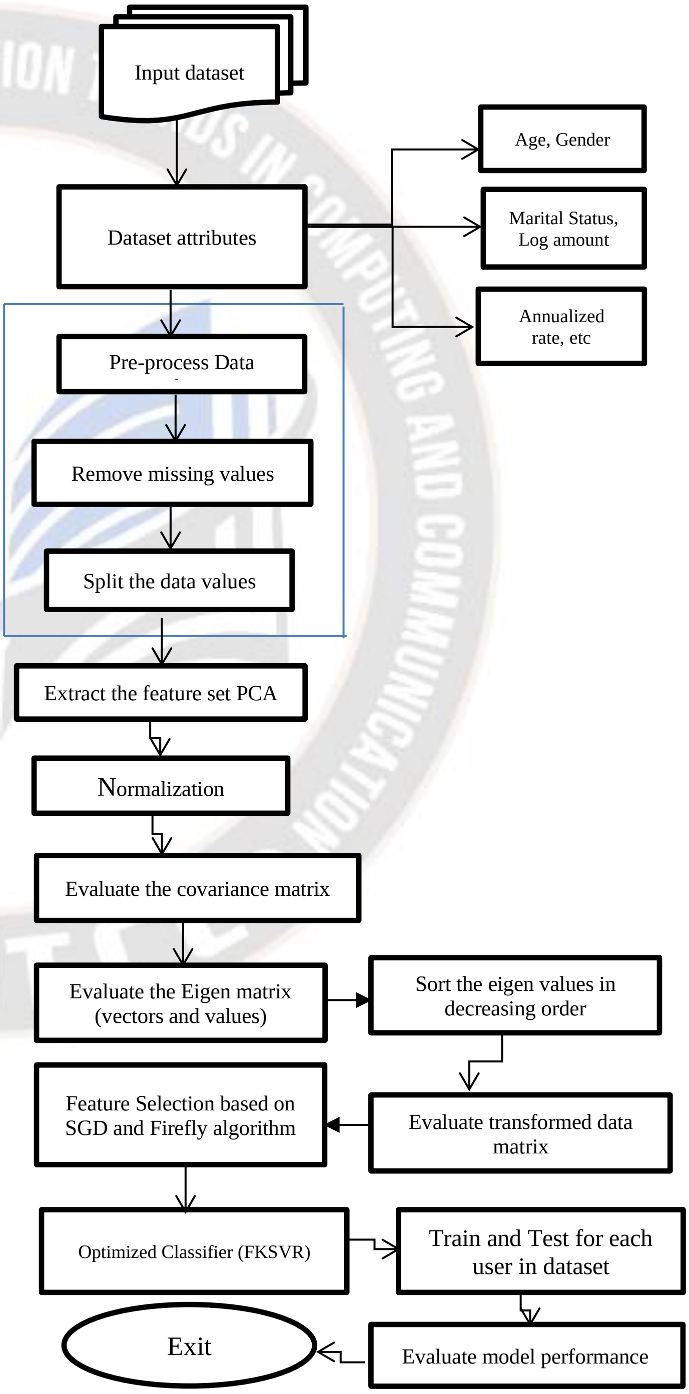
Input dataset
Dataset attributes
Age, Gender
Marital Status, Log amount
Annualized rate, etc
Pre-process Data
Remove missing values
Split the data values
Extract the feature set PCA
Normalization
Evaluate the covariance matrix
Evaluate the Eigen matrix (vectors and values)
Sort the eigen values in decreasing order
Evaluate transformed data matrix
Feature Selection based on SGD and Firefly algorithm
Optimized Classifier (FKSVR)
Train and Test for each user in dataset
Evaluate model performance
Exit

Figure 2 Proposed Flow Chart: Optimized-FSVR Model

***Article Received**: 03 May 2023 **Revised:** 29 June 2023 **Accepted:** 23 July 2023*

It reduces the risk of overfitting by feature selection since it makes the data less dimensional. Once the feature selection process is complete, an optimized feature-based Kernel-SVR classification model is introduced. The network has been trained using the ML method. The method diminishes or optimizes the kernel weights that reduce the difference between actual and desired outcomes. The outcome of this model has created binary values, which can be used as a classifier that has helped banks verify whether the borrower has been the payer. Generally, the credit risk dataset has been trained to test the research model's system prediction performance on the test dataset. It has evaluated performance metrics like accuracy, precision, etc., and compared them with the existing model.

D. Proposed Methods

(1) Feature Extraction Using PCA Method

PCA stands for "Principal Component Analysis". This method is used for feature extraction. The PCA algorithm in this article requires a multiple-variation data analysis approach to reduce the size of multi-dimensional statistics while maintaining data density. This method [21] offers several benefits, including simple functions, no parameter restrictions, and a wide range of applications across domains such as face detection, image reduction, and feature extraction. This paper utilizes the concept of dimension reduction in model building by extracting the principal components that have the greatest impact on enterprise capacity estimation. A measurable calculation is then established to evaluate the enterprise's total volume value.

The creation procedure of the proposed model is as follows:

- PCA is used to eliminate dimensional effects by utilizing the min-max standard technique for normalizing data processing, followed by the use of the following metric for linear variation shown in eq (6),

$$xj \frac{yj-min\,\{yj\}}{max\{yj\}-min\,\{yj\}} \tag{6}$$

Here, eq (6) xj data ∈ [0, 1] and dimensionless.

- Create the data matrix and compute the model correlation coefficient matrix for the filtered statistics as expressed in eq (7).

$$R_{ij} = \frac{1}{n}\sum_{i=1}^{n}(x_{i_j} * x_{i_j})\ \ i,j = 1,2,3 \tag{7}$$

- Evaluate the E (eigenvalues) and V (eigenvectors) of the correlation coefficient matrix ‘r’ using the Jacobian technique.
- Evaluate the PC (principal component) contribution rate and CCR (cumulative contribution rate):

$$C1 = \frac{l}{\sum_{i=1}^{7}(l\_i\,)} \tag{8}$$

- Compute the PC weight mentioned in eq (9).

$$r_{jk} = P1\big(z_{j\,,}y_k\big) = \sqrt{\lambda_j}\, e_{jk}\,, (i,j = 1,2, \ldots..9) \tag{9}$$

- Analyze the PC co-efficient as eq(10).

$$P1 = \sqrt{\frac{x}{\lambda}} \tag{10}$$

Eq (10), P1 is the PC defines the co-efficient, x defines the value in the PC co-efficient, λ denotes the value of the characteristics.

(2) Hybrid Feature Selection Using SGD and FFO (firefly Optimization) Methods

This section describes the hybrid feature selection method that combines the Stochastic Gradient Descent (SGD) and Firefly Optimization (FFO) algorithms.

a) SGD Optimizer: Stochastic Gradient Decent refers to a method that produces a different set of values each time it is run due to its underlying random (stochastic) nature. Distinct output values occur because SGD [22] is not achieved on the complete volume of input data. But, stochastically chooses only a pre-defined number of records for each iteration. A set of specific documents is known as a batch.

This method involves expressions like the following:

- Arbitrarily shuffle the database.
- Choose the k needed samples [pure SD k =1, in minimum batch k>1].
- Evaluation of novel values of variables is defined in eq (11).

$$w = w - \gamma * g\big(f(w)\big) = w - \frac{\gamma}{k}\sum_{j=1}^{k} g\big(fj\,(w)\big), \tag{11}$$

Here, j < 0; γ > 0-a properly minimum value, the step of learning.

- Repeat steps 1-3 until g(f(wj)) > ε , where ε > 0 – a proper minimum constant.

b) FFO (Firefly Optimizer) Method: FFO (Firefly Optimization) is categorized as an intelligent swarming [23] method through the more active representation that has gradually been utilized in resolving optimization issues constructed on rare occurrences of fireflies. This method was inspired by the capability to develop irregular light by fireflies. Their primary procedures are communicated briefly in two steps. They selected breeding associates for a prominent predator by repeating the Firefly optimization method and reviewing the results over several periods. The method was then familiarized through three steps.

Initialization: All fireflies' speedy indications are allocated, permitting the two fireflies' gap. The coefficient of distinctive preoccupation produces an eq (12).

$$L(r1) = L0e^{(-x_r)^2} \tag{12}$$

In eq(12) L= power of the light source, r = gap between two fireflies, L0 = power of light source during r = 0.

Attractiveness: It is described as;

$$B\_r(r1) = B\_r\_0e^{(-x_r)^2} \tag{13}$$

***Article Received**: 03 May 2023 **Revised:** 29 June 2023 **Accepted:** 23 July 2023*

In eq(13), B_0 = attractiveness of the Firefly at r =0.

Moving: In every group of fireflies for the complete residents, the smaller amount suitable Firefly is successful to the cost-efficient ones, with eq (13).

$$Y_j^{T+1} = Y_J^T + B_{r_{0e}}{}^{(-x_r)^2}\left(Y_i^T - Y_J^T\right) + a_1 * r1, n \qquad (14)$$

In eq (14), $a_1 =$ mutation coefficient.

c) Classification Using Kernel SVR Model: In the SVR method considered for classification by Frank [24], a decision tree structure is used. Instead of having final class values at the leaves, linear regression functions are employed. To predict the class, we choose the one whose model tree creates the highest estimated possibility values. However, local learning restrictions and the risk of over-learning may limit the effectiveness of linear regression. To address these issues, we propose using Support Vector Regression (SVR) which works well with high-dimensional feature spaces and transforms the optimization issue into dual convex quadratic programs. Similar to **E. Frank's** Model Trees, we will use the SVR for classification and approximate function values to identify the most probable class. The SVM algorithm is not limited to classification tasks and can also be used for regression problems. However, it shares the key characteristics of the maximum margin approach. By transforming the data into a high-dimensional feature space, the ML model obtains a nonlinear function. (FS) using a kernel function. SVM regression involves mapping the input instance, x, into an m-dimensional FS using secure mapping. Once the mapping is done, a model is constructed within this FS.

The calculated notation for the SVR model in the feature space F(x,ω) is given by:

$$F(x,\omega) = \sum_{i=1}^{M} \omega_i \, G_i \,(x) + b \qquad (15)$$

Here (15), where, G_i (x), I = 1,….M defines a set of non-linear revolutions, and b is the *bias term.*

d) Implemention Using O-FKSVR Model: The proposed model is an optimized feature-based kernel SVR (Support vector regression) model used to predict credit risks. The O-FKSVR model is a machine learning model that combines the strengths of SVR with a combination of SGD and Firefly Optimizer to enhance the credit risk prediction accuracy rate. The O-FKSVR algorithm involves the following phases:

- **Data pre-processing:** The credit records are pre-processed to extract features such as the PCA algorithm.
- **Hybrid Features with Optimization:** SGD and Firefly optimization methods are used to enhance the set of features used by SVR, improving accuracy and precision by selecting the optimal features that maximize the performance of the classification model.
- **Training and Testing:** The SVR model can then be trained using the training set and optimized using grid search or cross-validation techniques. Once the model is introduced, it can be tested using the testing set, and its performance can be evaluated using metrics such as mean squared error (MSE).

Credit risk prediction is an important application of support vector regression (SVR) in finance. An optimized feature-based kernel SVR model can enhance the accuracy and reliability of credit risk classification. To implement this model, a dataset can be collected and pre-processed to eliminate noise and standardize the data. Feature selection methods such as recursive feature elimination or correlation analysis can be applied to identify the most important features that affect credit risk. Once the important parts are identified, a suitable kernel function can be selected, and the hyperparameters of the SVR model can be optimized using techniques such as grid search or cross-validation. The optimized feature-based kernel SVR model can then be trained and tested on the dataset. The model's performance can be evaluated during testing using metrics such as mean squared error (MSE). If the model's performance is not satisfactory, additional feature selection and hyperparameter optimization techniques can be applied to improve the model's accuracy further.

IV. SIMULATION RESULT AND ANALYSIS

The Optimized Feature-based KSVR model (optimized featured-based kernel Support vector regression) is used as a credit risk prediction model in this proposed work and is executed using computer applications. Feature sets are extracted, and training and testing steps are designed using the Python tool, which runs on the Windows operating system. The research method is designed with the help of the ML toolbox in Python. When the SGD and FFO methods create a huge architecture that exceeds the memory limitation, a fitness of 0 is allocated to the candidate solution (CS). The metrics discussed above are intended and described to calculate the performance of the credit risk calculation system using PCA, SGD, FFO, and SVR. The credit risk prediction system is an FS-based system. Various feature vectors from the sample set are recognized and saved for comparing and corresponding at the previous stage. Input credit feature sets are matched with train sets. Various features are used for credit risk prediction. The o-FKSVR technique is used in this system for better performance. This technique subdivides the feature sets into groups and predicts the match based on the subdivided feature sets.

This analysis evaluated the performance of the implemented method on the credit risk dataset and compared it with different metrics, which are discussed below. Figure 3 shows that the output accuracy rate is one of the most critical calculation measures. The number of training data in the assessment is changeable, and the production accuracy rate has been compared based on the training data. Figure 3 demonstrates that the implemented approach is more precise than DT, RF, and AdaBoost and accurately increases the number of data points included in the outcomes. Table II shows the accuracy rate's performance as compared to other methods. The proposed model has improved accuracy compared to the existing models.

TABLE II PERFORMANCE METRICS

Methods	SVR	DT	RF	AdaBoost	O-FKSVR
Accuracy	0.863	0.817	0.908	0.933	0.955
Precision	0.944	0.828	0.919	0.944	0.964
Recall	0.871	0.929	0.928	0.871	0.952
F1-Score	0.906	0.876	0.923	0.906	0.958
Error Rate	13.6	18.0	9.1	6.6	4.4

As represented in Figure 5, the error rate of the implemented method has been compared to existing models such as DT, RF, AdaBoost, etc. The figure shows that the research method has the lowest error rate associated with the existing models. Table 2 provides a comparison of the error rates. The proposed approach was compared with DT, RF, SVR, and AdaBoost methods using well-defined Python software and other ML methods. The Python software allows the use of various ML methods and tools for data processing. In Table II, the accuracy of the research approach is shown to be better than other approaches on this specific dataset. The accuracy of each method, including the implemented strategy, is evident. In this research, the error of the research approach was minimal compared to the other techniques, resulting in a higher accuracy rate. Table II compares the performance of these methods and other techniques.

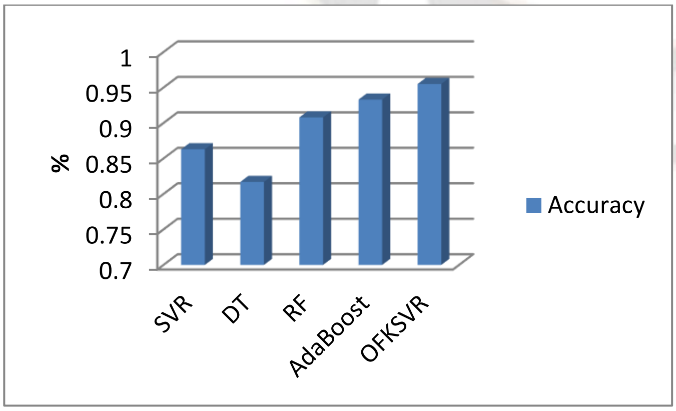

Figure 3 Comparisons –Accuracy %

Figure 3 shows that the suggested approach has the maximum accuracy rate. In this research, the defined O-FKSVR method achieved the highest precision of 96.4% on the credit risk analysis dataset, which is better than other methods shown in Figure 4. The proposed O-FKSVR model also achieved a high recall value is 95.2 % and an F1-Score of 95.8% on the credit risk analysis dataset, which is better than other models.

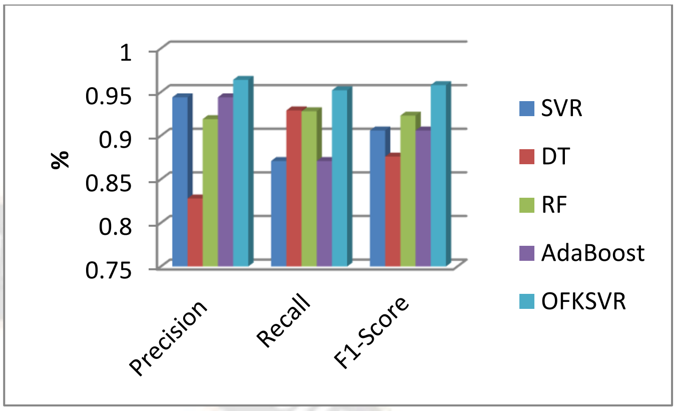

Figure 4 Comparison Analysis with different parameters: precision, recall, f1-score

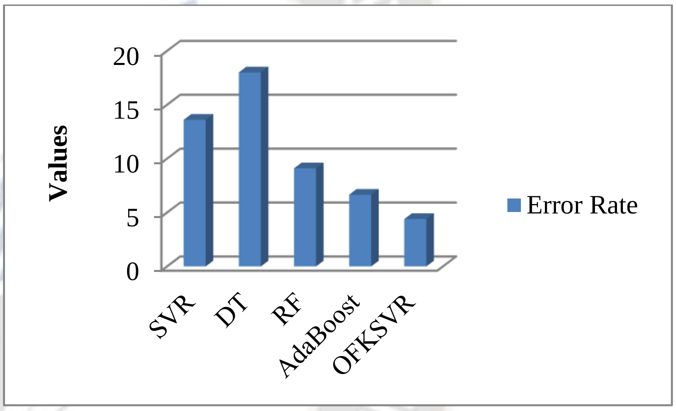

Figure 5 Comparison Analysis: Error Rate

As per Table II above, the proposed O-FKSVR method has the lowest error rate among the compared techniques and performed better than the DT, RF, AdaBoost method, etc. Table II provides a comprehensive mathematical comparison between the proposed approach and the previously defined approaches, indicating the maximum accuracy achieved by the proposed approach.

V. CONCLUSION AND FUTURE SCOPE

This article presents an optimized feature-based Kernel SVR (O-FKSVR) model for CR analysis in digital banking. The primary objective is to analyze various machine learning techniques and develop a precise model for CR assessment based on the credit risk analysis database. The O-FKSVR model is a machine learning model that combines the strengths of SVR with a combination of SGD and Firefly Optimizer to enhance the credit risk prediction accuracy rate. The research focuses on CR assessment and investigates the effect of different research methods in classifying business errors,

DOI: https://doi.org/10.17762/ijritcc.v11i9s.7402
***Article Received**: 03 May 2023 **Revised:** 29 June 2023 **Accepted:** 23 July 2023*

probability defaults, loss given defaults, etc. The prediction-based method design uses the Firefly optimizer with a kernel SVR model to predict credit risk. The proposed O-FKSVR model has demonstrated high accuracy, precision, recall, f1-score, and a reduced error rate compared to other methods. Compared with existing methods, the proposed O-FKSVR model has achieved a minimum error rate of 4.4 and an accuracy value of 95.5%.

Future work may introduce hybrid deep-learning models to detect credit risks and digital banking defaults. The evaluation may also consider time and speed to enhance performance.

REFERENCES

[1] Basle Committee on Banking Supervision, and Bank for International Settlements. "Principles for the management of credit risk", Bank for International Settlements, 2000.

[2] A. H. Mohammad, S. Ghwanmeh, and A. Al-Ibrahim, "Establishing Effective Guidelines to avoid Failure and Reducing Risk in E-Business", International Journal of Current Engineering and Technology, vol 4, no. 1, pp. 28-31, 2014.

[3] O. Awodele, S. Alimi, O. Ogunyolu, O. Solanke, S. Iyawe, F. Adegbie, F."Cascade of Deep Neural Network And Support Vector Machine for Credit Risk Prediction", In 2022 5th Information Technology for Education and Development (ITED), pp. 1-8, IEEE, 2022.

[4] Y. Hu, and J. Su, "Research on credit risk evaluation of commercial banks based on an artificial neural network model", Procedia Computer Science, vol 199, pp.1168-1176, 2022.

[5] T. G. L. Anwen, and M. S. Bari, "Credit Risk Management and Its Impact on Performance of Commercial Banks: In of Case Ethiopia", Research Journal of Finance and Accounting, vol 6, no. 24, pp. 53-64, 2015.

[6] C. Jiang, Z. Wang, and H. Zhao, "A prediction-driven mixture cure model and its application in credit scoring", European Journal of Operational Research, vol 277, no. 1, pp. 20-31, 2019.

[7] W. Liu, H. Fan, and M. Xia, "Credit scoring based on tree-enhanced gradient boosting decision trees", Expert Systems with Applications, vol 189, pp. 116034, 2022.

[8] W. Zhang, D. Yang, and S. Zhang, "A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring", Expert Systems with Applications, vol 174, pp. 114744, 2021.

[9] S. Goyal, N. Batra, K. Chhabra, "Diabetes Disease Diagnosis Using Machine Learning Approach", Lecture Notes in Networks and Systems, vol 473, pp 229–237, 2023.

[10] Poonam, N. Batra, "Evaluation of Various Machine Learning Based Existing Stress Prediction Support Systems (SPSSs) for COVID-19 Pandemic", Communications in Computer and Information Science, vol 1798 CCIS, pp. 408–422, 2023.

[11] N. Batra, S. Goyal, "Real-Time Smart Traffic Analysis Employing a Dual Approach Based on AI", Lecture Notes in Networks and Systems, vol 600, pp. 713–723, 2023.

[12] S. Goyal, N. Batra, K. Chhabra, "Lung Disease Detection Using Machine Learning Approach", "Lecture Notes in Networks and Systems, vol 473, pp. 251–260, 2023.

[13] Ipseeta Nanda, Monika SIngh, Lizina Khatua. (2023). Automated Irrigation System Using IoT Cloud Computing. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 360–365. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2728

[14] A. Ampountolas, "A machine learning approach for Micro-Credit scoring, MDPI. Multidisciplinary Digital Publishing Institute", Available at: https://www.mdpi.com/2227-9091/9/3/50 (Accessed: April 4, 2023), 2021.

[15] Ms. Elena Rosemaro. (2014). An Experimental Analysis Of Dependency On Automation And Management Skills. International Journal of New Practices in Management and Engineering, 3(01), 01 - 06. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/25

[16] T. N. Pandey, A. K. Jagadev, S. K. Mohapatra, and S. Dehuri," Credit risk analysis using machine learning classifiers", In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1850-1854). IEEE,2017.

[17] Y. Hu, and I. Su, "Research on credit risk evaluation of commercial banks based on artificial neural network model", Procedia Computer Science, vol 199, pp. 1168-1176, 2022.

[18] Condori-Alejo, H. I., Aceituno-Rojo, M. R., and G. S. Alzamora, "Rural micro credit assessment using machine learning in a peruvian microfinance institution", Procedia Computer Science, vol 187, pp. 408-413, 2021.

[19] E. Dri, A. Aita, E. Giusto, D. Ricossa, D. Corbelletto, B. Montrucchio, and R. Ugoccioni, "A More General Quantum Credit Risk Analysis Framework", Entropy, vol 25, no. 4, pp. 593, 2023.

[20] G. S. Alzamora, M. R. Aceituno-Rojo, and H. I. Condori-Alejo, "An Assertive Machine Learning Model for Rural Micro Credit Assessment in Peru", Procedia Computer Science, vol 202, pp.301-306, 2022.

[21] J. Li, C. Xu, B. Feng, and H. Zhao, "Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism", Electronics, vol 12, no. 7, 1643, 2023.

[22] L. Wang, F. Jia, L. Chen, and Q. Xu, "Forecasting SMEs' credit risk in supply chain finance with a sampling strategy based on machine learning techniques", Annals of Operations Research, pp. 1-33,2022.

[23] Davis, W., Wilson, D., López, A., Gonzalez, L., & González, F. Automated Assessment and Feedback Systems in Engineering Education: A Machine Learning Approach. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/102

[24] Z. Huang, Z. Yan, Q. Zhao, and K. Ma, "Credit risk assessment in bank based on SMEs using PCA", In Journal of Physics: Conference Series (Vol. 1848, No. 1, p. 012071). IOP Publishing, 2021.

[25] Y. Tian, Y. Zhang, and H. Zhang, "Recent Advances in Stochastic Gradient Descent in Deep Learning", Mathematics, vol 11, no. 3, pp. 682, 2023.

[26] Z. Hassani, M. Alambardar Meybodi, and V. Hajihashemi, "Credit risk assessment using learning algorithms for feature selection", Fuzzy Information and Engineering, vol 12, no. 4,pp. 529-544, 2020.

[27] B. Huang, Z. Cai, Q. Gu, and C. Chen, "Using Support Vector Regression for Classification", In Advanced Data Mining and Applications: 4th International Conference, ADMA 2008, Chengdu, China Proceedings 4, pp. 581-588. Springer Berlin Heidelberg, 2008.

[28] Karthickaravindan, "Decision trees and Random Forest", Kaggle. Kaggle. Available at: https://www.kaggle.com/code/karthickaravindan/decision-trees-and-random-forest/notebook (Accessed: April 26, 2023), 2018.