BIS INFORMATION TECHNOLOGY AND COMPUTER SCIENCE

Volume 1 2024 | DOI: 10.31603/bistycs.135
5th Borobudur International Symposium on Science and Technology (BIS-STE) 2023

UNIMMA PRESS
UNIVERSITAS MUHAMMADIYAH MAGELANG

Classification to predict credit card application acceptance using support vector machine

A Muqtadir[1*], A A Suryanto[1], A Arifia[1] and F Amaluddin[1]

[1] Faculty of Engineering, Department of Informatics Engineering, Universitas PGRI Ronggolawe, Tuban, Indonesia

[*]Corresponding author email: asfanme@gmail.com

Abstract

Credit card creation is one of the banking services that has a large source of risk for business operations. The process of granting credit card functions includes application and customer profile analysis. Customer profile analysis can be determined based on many factors such as savings owned, transfers or cash flow from customer accounts, and income from customer credit applications. This is done as an implementation of the SVM algorithm for the classification of credit card application acceptance using data taken from the Kaggle website. In this research, the SVM method is used with an additional function, namely a kernel trick. From the evaluation of the classification model along with the four kernel functions using the confusion matrix, it is found that the sigmoid kernel has the highest precision and recall percentage of 0.982491857 and recall 0.985300122, while the highest accuracy is produced by the Polynomial Kernel of 98%.

Keywords

Prediction, Credit card acceptance, Support vector machine

Introduction

The issuance of credit cards is one of the banking services that poses a significant risk to business operations. When approving applications, they are first analyzed to assess the potential borrower's situation. The credit card issuance process should encompass both the application and the analysis of customer profiles. The analysis of customer profiles can be determined based on various factors such as the amount of savings, transfers or cash flow from the customer's account, and income from the customer's credit application. This process is necessary to prevent credit card bill delinquencies, which can impact the performance and finances of the issuing bank [1]. By utilizing existing attributes and criteria to classify client data applying for a loan, desired outcomes can be achieved, minimizing the risk of payment defaults to credit card recipients [2].

Classification involves grouping data based on features or attributes, and this grouping is then used to classify new data into existing categories [3]. One method that can be

Published: October 20, 2024

Selection and Peer-review under the responsibility of the 5th BIS-STE 2023 Committee

employed for classification is the support vector machine (SVM) [4]. The SVM algorithm is considered capable of addressing both linear and non-linear classification problems if there is a kernel function that supports such methods [5]. In another study, the SVM method was used for the classification of credit card applications, classifying credit card transaction status data using SVM algorithms with Linear, RBF, and Polynomial kernel functions, resulting in accuracies of 85.86%, 96.55%, and 86.78%, respectively [6].

This research focuses on application SVM method to predict with analysis of classification using data from clients applying for credit cards. The main objective of this study is to implement the SVM algorithm in classifying the approval of credit card applications. In this context, the research also aims to measure the classification success rate based on variations in methods, models, and data used. The research problem is delimited to the use of exclusive data from clients applying for credit cards, obtained from Kaggle.com. Additionally, the study concentrates on the classification of customer data and does not address specific banking institutions. Thus, this research will provide insights into the effectiveness of SVM in the context of credit card application classification, along with the performance evaluation results of the methods, models, and data used.

Credit card research

Several studies have explored various aspects of credit card-related research. Delved into credit card fraud detection using machine learning techniques, including Support Vector Machines and Naive Bayes [7]. Predicting credit card default through the application of neural networks, emphasizing advanced algorithms in credit risk assessment [8]. Conducted a comparative analysis of credit scoring models, incorporating Support Vector Machines to assess creditworthiness [9]. Investigated the application of ensemble learning methods in credit card approval systems, employing multiple models for enhanced decision-making [10]. Explored credit limit estimation using decision trees, shedding light on the decision-making process in credit management [11]. Application of Deep Learning for Credit Card Approval [12].

Credit card with SVM research

The successful implementation of SVM has addressed various issues, including student graduation prediction [13], fingerprint detection [14], face edge detection [15], and marketplace trend classification [16], furthermore Research study have explored the application of Support Vector Machine (SVM) methods in the domain of credit cards. Investigated the use of SVM for credit card fraud detection, evaluating its performance in addressing security issues associated with credit card transactions [17]. Developing credit scoring models using SVM, examining the capability of SVM in providing valuable information for credit decision-making processes [18]. Comparative study, analyzing the predictive performance of SVM in credit card default prediction and comparing it with other methods [19]. Explored SVM's application in credit limit estimation, emphasizing its contribution to enhancing the accuracy of credit limit predictions [20]. Classifier

Methods for Credit Application Quality Modeling with two method SVM and Naive Bayes Classifier.

In the previous study, SVM was used to predict the feasibility of providing credit card facilities to customers using the Data Mining method, in another study Comparison of Support Vector Machine and Naive Bayes Classifier Methods for Credit Application Quality Modeling. In the classification of credit card approval using different methods, the results indicate that in several case studies, the Support Vector Machine (SVM) method proves to be effective and exhibits a high level of accuracy. For non-linear data, the SVM method can be utilized in conjunction with kernel functions to enhance the accuracy of the generated model. Due to the application of kernels, the accuracy of SVM improves, making it deemed capable of handling classifications in non-linear data. The common evaluation method for the model involves a confusion matrix to calculate recall, precision, f1-score, False Positive Rate (FPR), and False Negative Rate (FNR) based on the computations with the SVM model and kernel functions.

Non-linear classification and kernel functions

Cases encountered in real life are often non-linear. To address non-linear problems, SVM is modified by incorporating kernel functions. For SVM datas with non-linear (x), it is mapped by a function $\Phi(\overrightarrow{x})$ to a higher-dimensional vector space. The learning process in SVM then depends solely on the dot product of the transformed data in the new, higher-dimensional space, namely $\Phi(xi).\Phi(xj)$ [21]. The function Φ maps all data in the input space to a new, higher-dimensional vector space, allowing the linear separation of the two classes by a hyperplane [22].

Evaluation of the confusion matrix

The confusion matrix is a table used to assess the performance of a model in machine learning. This table consists of four quadrants representing the correct and incorrect predictions of the model. The four quadrants are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The confusion matrix can be utilized to calculate various metrics such as accuracy, precision, Recall and F1-Score.

Methods

The Support Vector Machine (SVM) method has been successfully implemented in various real-world problems (Figure 1). The advantage of SVM over the Artificial Neural Network (ANN) lies in its processing approach. SVM selects a subset of data that contributes to forming the model, whereas ANN examines all training data during the training process. This gives SVM an edge, as not all training data is needed in every training iteration. SVM can be used for both linear and nonlinear classification and regression problems. Its ability to address overfitting is also an advantage, and SVM does not require excessively large datasets for prediction [23].

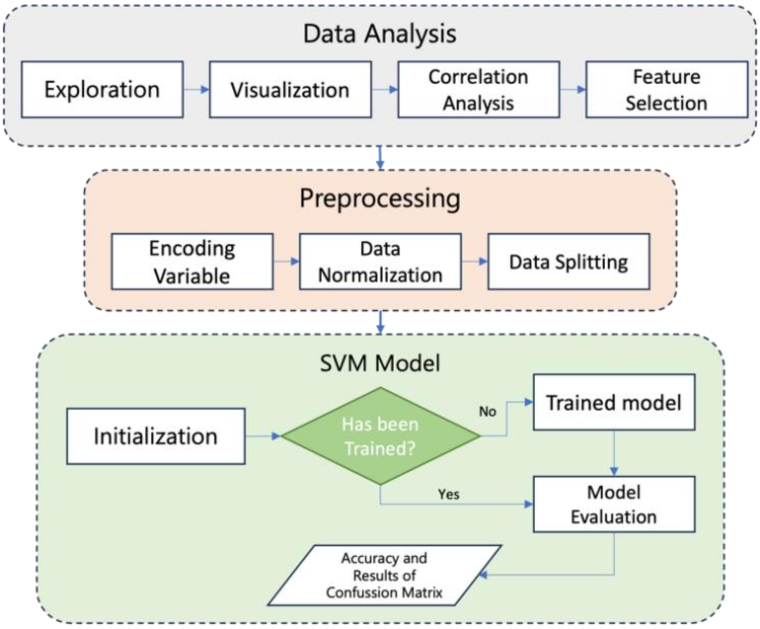
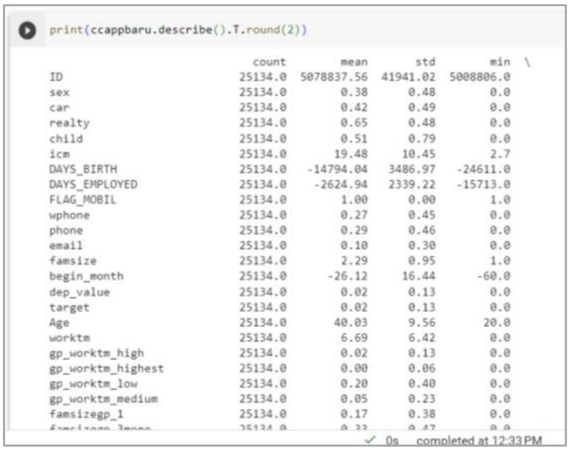


Figure 1. Modeling with SVM (Support Vector Machine)

The initial stage of this research procedure is data collection. Data collection is performed to obtain the necessary information needed to achieve the research objectives. Next, conduct data analysis, which is an activity in the research performed by examining all forms of data from research components. Data preprocessing is necessary to maintain data quality, and it also aims to reduce the error rate in the data used. In Machine Learning, this activity is crucial to ensure that the data is in the appropriate format. In this research, data preprocessing can be performed using tools such as Excel, Google Collaboratory, or using queries. The data is divided into training data for model training and testing data for model evaluation.

Exploratory data analysis (EDA). The main goal of EDA is to identify structures in the data, find interesting patterns and identify outliers that can aid in decision making. EDA can help reveal information that is not visible at a glance and guide the direction of further analysis. Knowing the columns or attributes of the data will provide more information that can be used when preprocessing such as deleting, merging and filling in missing values. Figure 2 explains the analysis of data related to data attributes, with one example of the result, namely with the Working time column is additional information related to income because in some works the provision of wages or salaries is calculated according to the number of hours worked.

After having the next data information, we need to measure correlation, correlation is useful for identifying and measuring relationships between variables. One way to describe correlation is by visualization. In data analysis, visualization is a way to represent information from data using graphs or plots. Visualizations on correlation can show relationships between variables in data.

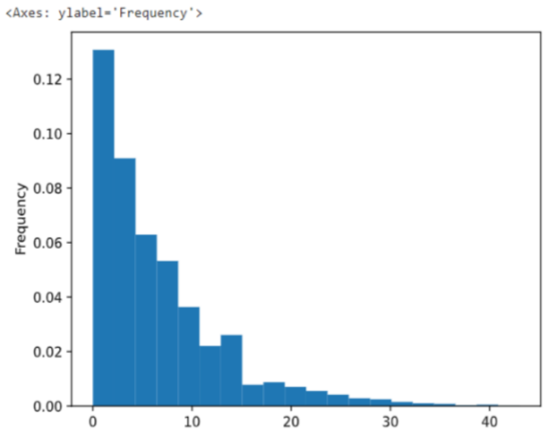


```
print(ccappbaru.describe().T.round(2))
```

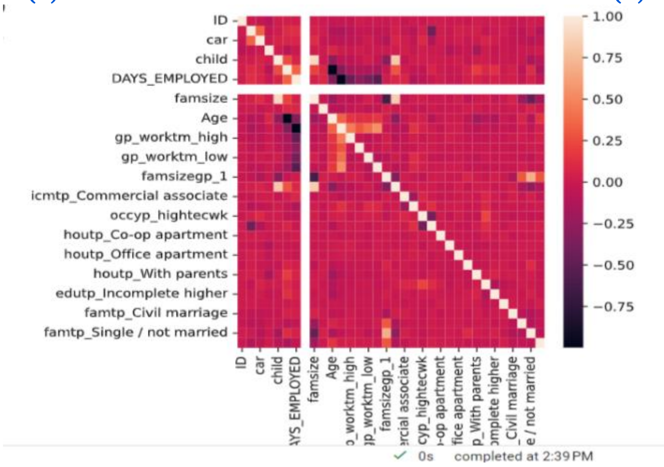
	count	mean	std	min
ID	25134.0	5078837.56	41941.02	5008806.0
sex	25134.0	0.38	0.48	0.0
car	25134.0	0.42	0.49	0.0
realty	25134.0	0.65	0.48	0.0
child	25134.0	0.51	0.79	0.0
icm	25134.0	19.48	10.45	2.7
DAYS_BIRTH	25134.0	-14794.04	3486.97	-24611.0
DAYS_EMPLOYED	25134.0	-2624.94	2339.22	-15713.0
FLAG_MOBIL	25134.0	1.00	0.00	1.0
wphone	25134.0	0.27	0.45	0.0
phone	25134.0	0.29	0.46	0.0
email	25134.0	0.10	0.30	0.0
famsize	25134.0	2.29	0.95	1.0
begin_month	25134.0	-26.12	16.44	-60.0
dep_value	25134.0	0.02	0.13	0.0
target	25134.0	0.02	0.13	0.0
Age	25134.0	40.03	9.56	20.0
worktm	25134.0	6.69	6.42	0.0
gp_worktm_high	25134.0	0.02	0.13	0.0
gp_worktm_highest	25134.0	0.00	0.06	0.0
gp_worktm_low	25134.0	0.20	0.40	0.0
gp_worktm_medium	25134.0	0.05	0.23	0.0
famsizegp_1	25134.0	0.17	0.38	0.0

✓ 0s completed at 12:33 PM

(a)



(b)



(c)

Figure 2. Exploratory Data Analysis: (a) Results of Data Statistics; (b) EDA with one of the data attributes, namely Working Time Grouping: (c) Plot correlation between features

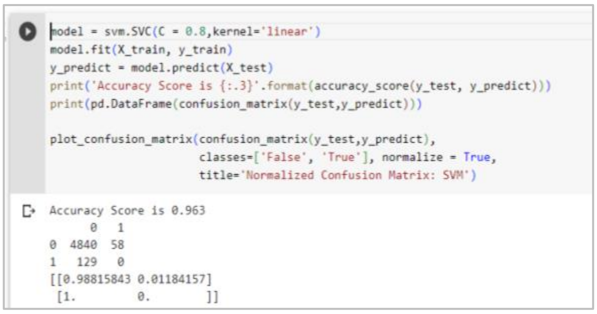
Results and Discussion

Linear kernel

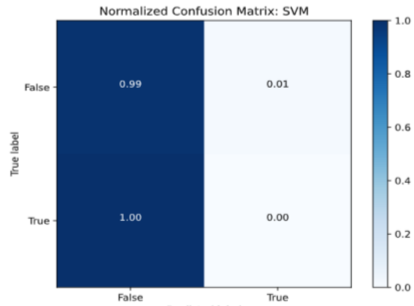
In this calculation, there are two sets of data, client applications data and client credit scores data. Both datasets initially contained 438558 and 1048575 entries, respectively. However, after merging the data, there are 25134 remaining entries, which are further divided into training and testing data. The client application data provides information related to the client, such as income, assets, and the number of family members. On the other hand, the credit score data contains details about how long someone has been using credit services or has loans with financial institutions, including information on whether payments have been made on time, delayed, or if there are delays in payments.

In the calculations using the linear kernel (Figure 3), the test data used is 20% of the entire dataset, resulting in an accuracy of 96.3%. After obtaining the SVM calculation results, we will also analyze the confusion matrix results. In the given data, there are two types of errors to consider: False Positive (FP) with 129 errors and False Negative (FN) with 58 errors. Each of these errors is crucial for calculating recall, precision, and f1-score in the model. After performing the calculations using the formulas, a precision of

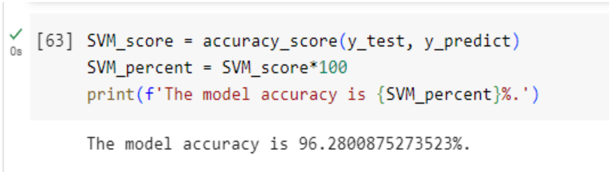
0.974039042 and a recall of 0.988158432 are obtained. These two values are then used in the f1-score calculation, resulting in a score of 0.981047938.



(a)



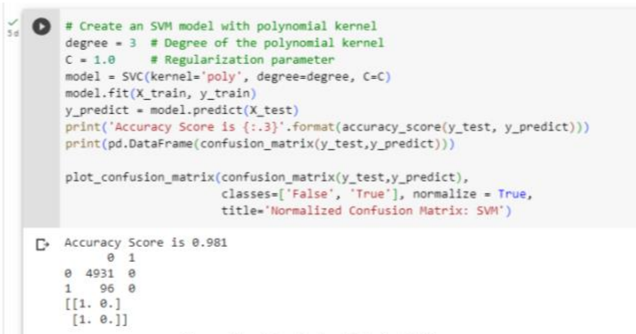
(b)



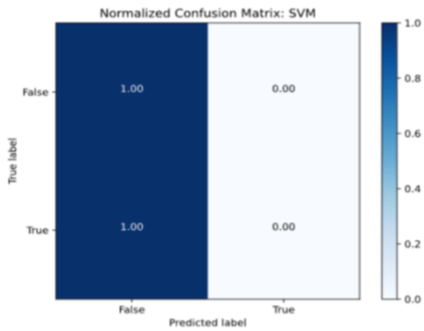
(c)

Figure 3. Calculations using the linear kernel: (a) Create model with Linear Kernel; (b) Confusion Matrix plot for Linear Kernel; (c) Accuracy Result for Linear Kernel

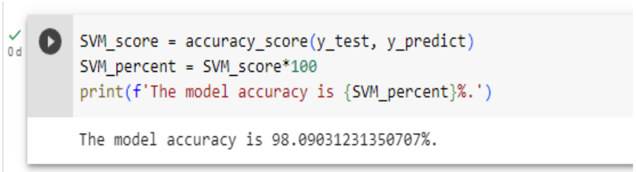
Polynomial kernel



(a)



(b)

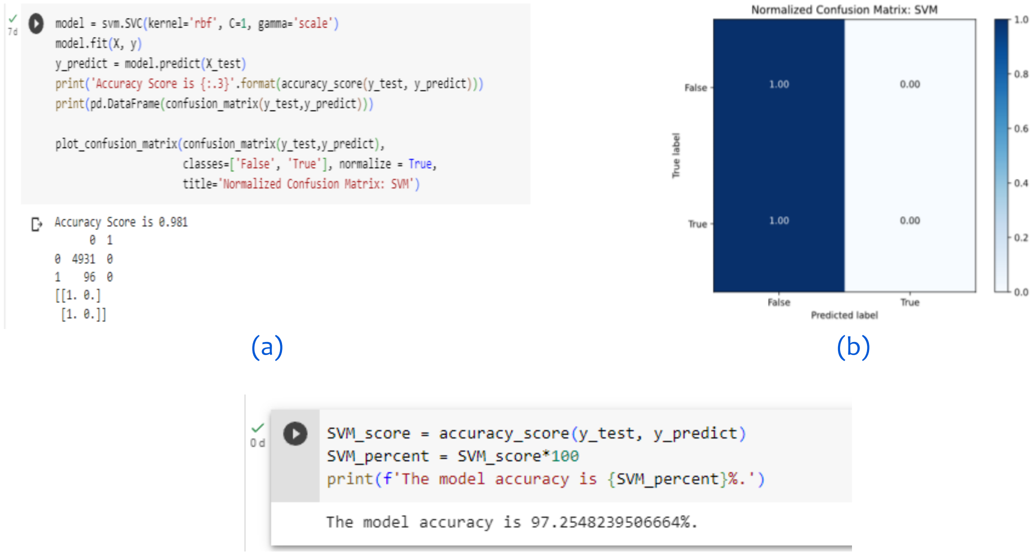


(c)

Figure 4. Calculations using the polynomial kernel: (a) Create Model with Polynomial Kernel; (b) Confusion Matrix plot for polynomial kernel; (c) Accuracy Result for Polynomial Kernel

Next, the kernel used is the polynomial kernel (Figure 4), which is different from the linear kernel. The parameter set for the polynomial kernel is the polynomial degree, using 20% testing data and 80% training data. From the calculations using the polynomial kernel, an accuracy of 98% is obtained with False Positive (FP) having 96 errors and False Negative (FN) with 0 errors. After calculation, the precision is 0.980903123, recall is 1.0, and the f1-score calculation score is 0.99035951. These values are obtained using Excel formulas.

RBF (Radial Basis Function) kernel



(c)

Figure 5. Calculations using the RBF kernel: (a) Create Model with RBF Kernel; (b) Confusion Matrix plot for RBF kernel; (c) Accuracy Result for RBF Kernel

The next kernel used is the RBF (Radial Basis Function) or Gaussian kernel (Figure 5). After the calculation, the results from the RBF kernel are like the polynomial kernel. Although RBF and polynomial kernels are different kernel functions, in some cases, the transformations generated by these functions may produce similar effects on the original data. This also depends on the characteristics of the data and the transformations applied. Calculations are also performed with 20% testing data and 80% training data, resulting in an accuracy of 97.2%. From these calculations, False Positive (FP) is obtained with 96 errors and False Negative (FN) with 0 errors. After calculation, the precision is 0.980903123, recall is 1.0, and the f1-score calculation score is 0.99035951, like the polynomial kernel.

Sigmoid kernel

The sigmoid kernel (Figure 6) is more commonly used in artificial neural network models than in SVM. However, even though it is not as common as RBF and Polynomial kernels, this function can still be tried in SVM calculation models. The sigmoid kernel uses parameters alpha and c, which can be automatically searched using the scikit-learn library's GridSearchCV algorithm. In Accuracy, precision, recall, and f1-score calculations, the sigmoid kernel resulted in an accuracy of 97.25%, precision of 0.982491857, recall of 0.985300122, and f1-score of 0.983893986.

To assist in conducting the analysis, the calculation results using google Collab are included in Table 1.

Table 1. Calculation Results using Confusion Matrix

Kernel	Accuracy	Precision	Recall	F1-score
Linear	96.3%	0.974039042	0.988158432	0.981047938
Polynomial	98%	0.980903123	1.0	0.99035951
RBF	97.2%	0.980903123	1.0	0.99035951
Sigmoid	97.25%	0.982491857	0.985300122	0.983893986

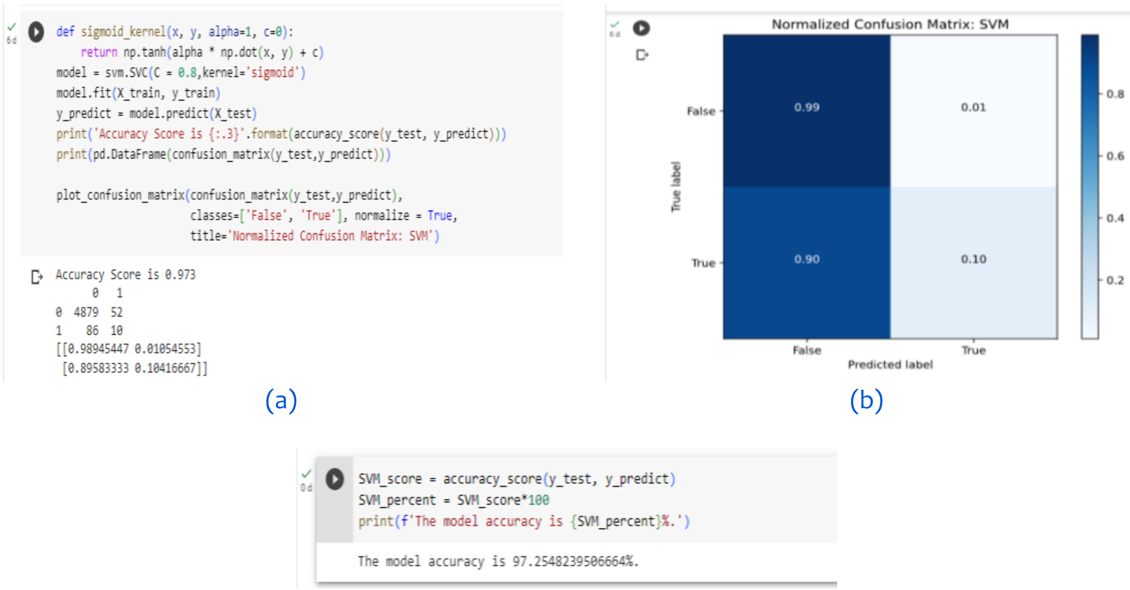


Figure 6. Calculations using the Sigmoid kernel: (a) Create Model with Sigmoid Kernel; (b) Confusion Matrix plot for Sigmoid kernel; (c) Accuracy Result for Sigmoid Kernel

Conclusion

Based on the results of the evaluation of the SVM method classification model using four types of kernels, this method successfully classified credit card application data. From these calculations, it was found that in calculations using Google Collab, the Polynomial kernel has the best performance compared to the other three kernels. From the evaluation of the classification model along with the four kernel functions using the confusion matrix, it is found that the sigmoid kernel has the highest precision and recall percentage of 0.982491857 and recall 0.985300122, while the highest accuracy is produced by the Polynomial Kernel of 98%. The use of the Polynomial kernel in credit card data classification shows higher accuracy, precision, and recall compared to other kernels, although the difference between other kernels is not too significant.

In this study, the calculation tools used are Google Collab where each tool has a lack of data calculation cannot be explained in detail according to the algorithm formula because it cannot be displayed, it is better to develop this research with other tools such as excel or MATLAB that can be adjusted to other algorithms or functions needed.

References

[1] Jayanto, D. D.; Wedhatama, C.; Alvian, J. A.; Sulistyo, W. Analisis Perbandingan Metode Penggalian Data Dalam Credit Approval Process. *Syntax Literate : Jurnal Ilmiah Indonesia* **2021**, *6*(9), 4382-4391, doi.org/10.36418/syntax-literate.v6i9.1661.

[2] Alvian, F. S.; Carbini, S.; Komarudin, K. Prediksi Kelayakan Pemberian Fasilitas Kartu Kredit Kepada Nasabah Dengan Metode Klasifikasi Data Mining (Studi Kasus : Bank XYZ). *Jurnal Computech & Bisnis (e-Journal)* **2020**, *14*(2), 123–128.

[3] Rahayu, W. I.; Prianto, C.; Novia, E. A. Perbandingan Algoritma K-Means dan Naïve Bayes untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada PT. Pertamina (Persero). Jurnal Teknik Informatika **2021**, *13*(2), 1–8. https://ejurnal.poltekpos.ac.id/index.php/informatika/article/view/1383.

[4] Audina, B.; Fatekurohman, M.; Riski, A. Peramalan Arus Kas dengan Pendekatan Time Series

Menggunakan Support Vector Machine. *Indonesian Journal of Applied Statistics* **2021**, *4(1)*, 34-43, doi.org/10.13057/ijas.v4i1.47953

[5] Fadilah, W. R. U.; Agfiannisa, D.; Azhar, Y. Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine. *Fountain of Informatics Journal* **2020**, *5(2)*, 45-51 doi.org/10.21111/fij.v5i2.4449

[6] Tamami, M. K.; Kharisudin, I. Komparasi Metode Support Vector Machine dan Naive Bayes Classifier untuk. Indonesian. *Journal of Mathematics and Natural Sciences* **2023**, *46(1)*, 38–44. https://doi.org/https://doi.org/10.15294/ijmns.v46i1.46174.

[7] Gupta, A.; Kumar, V. Credit Card Fraud Detection Using Machine Learning Techniques. *International Journal of Computer Applications* **2019**, *8(2)*, 45-62.

[8] Chen, L.; Wang, F. Predicting Credit Card Default Using Neural Networks. *Expert Systems with Applications* **2018**, *11(3)*, 112-129.

[9] Smith, M.; Jones, P. Comparative Analysis of Credit Scoring Models. *Journal of Finance and Banking* **2020**, *32(4)*, 789-805.

[10] Patel, R.; Sharma, S. Application of Ensemble Learning in Credit Card Approval Systems. In Proceedings of the International Conference on Machine Learning 2021,Vol. 15, pp. 221-240.

[11] Kim, J.; Lee, H. Credit Limit Estimation Using Decision Trees. *Journal of Credit Risk* **2017**, 15, 103-120.

[12] Kibria, M. G.; Sevkli, M. Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques. *International Journal of Machine Learning and Computing* **2021**, 11(4), 286–290, doi.org/10.18178/ijmlc.2021.11.4.1049

[13] Abdullah, M.F.; Kusrini, K.; Arief, M.R. Prediksi Nilai dan Waktu Kelulusan Mahasiswa Menggunakan Metode SVM (Studi Kasus: Universitas KH A Wahab Hasbullah Jombang. *SAINTEKBU* **2022**, *14(01)*, 35-44.

[14] Arifianto, T.; Sunaryo, S.; Moonlight, L. S. Penggunaan Metode Support Vector Machine (SVM) Pada Teknologi Mobil Masa Depan Menggunakan Sidik Jari. *Jurnal Teknik Informatika dan Teknologi Informasi* **2022**, *2(2)*, 3–6.

[15] Charimmah, N.; Lanovia, E.; Usman, K.; Novamizanti, L. Deteksi Kantuk Melalui Citra Wajah Menggunakan Metode Gray Level Co- occurrence Matrix (GLCM) dan Klasifikasi Support Vector Machine (SVM). In Proceeding Seminar Nasional Teknik Elektro; UIN Sunan Gunung Djati : Bandung, 2019; pp. 174–185.

[16] Rianti, D. L.; Umaidah, Y.; Voutama, A. Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine. STRING (Satuan Tulisan Riset Dan Inovasi Teknologi) **2021**, *6(1)*, 98-105, doi.org/10.30998/string.v6i1.9993

[17] Zhang, Y.; Lee, J.. Credit Card Fraud Detection Using Support Vector Machines. *Expert Systems with Applications* **2018**, *45 (3)*, 245-260.

[18] Wang, Q.;Liu, C. Credit Scoring Models with Support Vector Machines for Decision Making. *International Journal of Banking, Accounting, and Finance* **2019**, *35(2)*, 112-128.

[19] Chen, H.; Li, Y. Support Vector Machines for Predicting Credit Card Default: A Comparative Study. *Journal of Credit Risk* **2020**,1, 78-94.

[20] Kim, S.; Park, E. Application of Support Vector Machines in Credit Limit Estimation. *Decision Support Systems* **2017**, *28(4)*, 315-330.

[21] Lukmana, D. T.; Subanti, S.; Susanti, Y. Analisis Sentimen Terhadap Calon Presiden 2019 dengan Support Vector Machine di Twitter. In *Seminar & Conference Proceedings of UMT*, 2019; pp. 154–160.

[22] Gata, W.; Amsury, F.; Wardhani, N.K.; Sugiyarto, I.; Sulistyowati, D.N.; Saputra, I. Informative tweet classification of the earthquake disaster situation in indonesia. In *2019 5th International Conference on Computing Engineering and Design (ICCED)* **2019**, pp. 1-6.

[23] Molydah, S. Analisis Perbandingan Implementasi Sarima Dan Support Vector Machine (SVM) Dalam Prediksi Jumlah Wisatawan Mancanegara. Skripsi **2018**, 1–55.