JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 6, No. 3. June 2024 DOI: https://doi.org/10.34288/jri.v6i3.299
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

INTEGRATION OF ADASYN METHOD WITH DECISION TREE ALGORITHM IN HANDLING IMBALANCE CLASS FOR LOAN STATUS PREDICTION

Ami Rahmawati-1*); Ita Yulianti-2; Tati Mardiana-3; Denny Pribadi-4

Sistem Informasi[1], Sains Data[3]
Universitas Nusa Mandiri[1,3]
ami.amv@nusamandiri.ac.id[1*)], tati.ttm@nusamandiri.ac.id[3]

Sistem Informasi Akuntansi Kampus Kota Sukabumi[2], Ilmu Komputer Kampus Kota Sukabumi[4]
Universitas Bina Sarana Informatika[2,4]
ita.iyi@bsi.ac.id[2], denny.dpi@bsi.ac.id[4]

(*) Corresponding Author

Abstract

Determining the provision of credit is generally carried out based on measuring credibility using credit analysis principles (5C principles). However, this method requires quite a long processing time and is very susceptible to subjective judgments which might influence the final results. This research aims to utilize data mining techniques by developing modeling on loan status prediction datasets. The stages in this research include data preprocessing, modeling and evaluation using accuracy metrics and ROC graphs. In this analysis, it is known that there is a class imbalance in the processed dataset so it is necessary to carry out an oversampling technique. In this research, the ADASYN (Adaptive Synthetic) Oversampling technique is used to ensure the class distribution is more balanced. Then, the ADASYN technique is integrated with the Decision Tree Algorithm to build a prediction model. The research results show that the two methods are able to increase prediction accuracy by 12.22% from 73,91% to 85.22%. This improvement was obtained by comparing the accuracy results before and after using the ADASYN Oversampling technique. This finding is important because it proves that the implementation of such integration modeling can significantly improve the performance of classification models and can provide strong potential for practical application in helping more effective loan status predictions.

Keywords: ADASYN; Decision Tree; Imbalance Class; Loans; Oversampling Techniques

Abstrak

Penentuan pemberian kredit umumnya dilakukan berdasarkan pengukuran kredibilitasnya dengan menggunakan prinsip analisis kredit (prinsip 5C). Namun, cara tersebut membutuhkan waktu pemrosesan yang cukup panjang dan sangat rentan terhadap penilaian subjektif yang mungkin mempengaruhi hasil akhir. Penelitian ini bertujuan untuk memanfaatkan teknik data mining dengan mengembangkan pemodelan pada dataset prediksi status peminjaman. Tahapan dalam penelitian ini meliputi preprocessing data, pemodelan dan evaluasi menggunakan metrik akurasi dan grafik ROC. Dalam analisis ini diketahui bahwa terdapat ketidakseimbangan kelas pada dataset yang diproses sehingga perlu dilakukan teknik Oversampling. Pada penelitian ini, teknik Oversampling ADASYN (Adaptive Synthetic) digunakan untuk memastikan distribusi kelas menjadi lebih seimbang. Kemudian, teknik ADASYN diintegrasikan dengan Algoritma Decision Tree untuk membangun model prediksi. Hasil penelitian menunjukkan bahwa kedua metode tersebut mampu meningkatkan akurasi prediksi sebesar 12,22% dari 73,91% menjadi 85,22%. Peningkatan ini diperoleh dengan membandingkan hasil akurasi sebelum dan sesudah menggunakan teknik Oversampling ADASYN. Temuan ini penting karena membuktikan bahwa implementasi dari pemodelan integrasi tersebut dapat secara signifikan meningkatkan performa model klasifikasi dan dapat memberikan potensi yang kuat untuk diaplikasikan secara praktis dalam membantu prediksi status pinjaman yang lebih efektif.

Kata kunci: ADASYN; Decision Tree; Ketidakseimbangan Kelas; Peminjaman; Teknik Oversampling

INTRODUCTION

Over the last few years, Indonesia has experienced significant changes due to the COVID-19 pandemic which not only has an impact on the health sector, but also has far-reaching consequences for all sectors including the economic sector (Eltania, 2022). he decline in economic activity and income in the community triggers economic instability which encourages them to take out loans to maintain their survival (Yenila et al., 2023). This is in line with survey findings from the Badan Pusat Statistik (BPS) which can be seen in the following image (Badan Pusat Statistik, 2019):

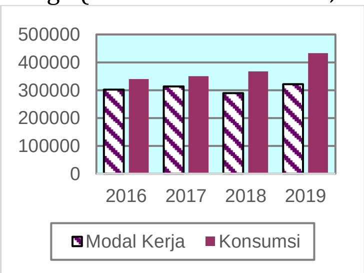

Figure 1. Banking Credit Position by Type of Use in Government Banks in Million Rupiah

Based on this graph, it can be seen that the consumption category dominates the credit increase trend over the last four years, surpassing credit for working capital. The economic downturn in society due to the pandemic has had an impact on increasing lending activity to meet consumer needs. Similar conditions also occur in other countries, one of which is the United States, which shows that loan applications have increased significantly during the pandemic, especially to meet daily consumption needs (GAO Report, 2023).

Loans/credit is defined as an activity where one party provides goods or services to another party in exchange for money that must be returned within a certain period of time based on a loan agreement (Ubaedi & Djaksana, 2022). The institutions/parties that provide these services are banking and non-banking institutions (Yenila et al., 2023). The party who applies for a loan is often referred to as the 'debtor', while the party who provides the loan is known as the 'creditor' (Hermawan & Yamasari, 2022). In practice, creditors apply certain conditions to the provision of credit, such as an agreement to hand over collateral from the debtor to the creditor with the aim of providing security collateral (Prawira et al., 2022) which can be in the form of material or non-material (Tarigan et al., 2021). This is done to minimize the occurrence of credit risks, including the debtor's failure to fulfill its obligations which has the potential to disrupt financial stability and threaten the company's operational continuity (Syafudin et al., 2021). Therefore, to face these challenges, financial institutions and other credit service providers need to review the analytical methods used in determining credit decisions (Pratiwi et al., 2023).

The determination of credit is generally carried out based on the measurement of its credibility using the principles of credit analysis (5C principle) which consists of character, collateral, capital, capacity and condition of Economy (Hermawan & Yamasari, 2022). However, this method has obstacles, including requiring quite a long processing time (Nurdiyanto et al., 2022) because it involves a number of documents and information that need to be observed and analyzed by credit appraisers (Sitepu & Manohar, 2022) so it is very vulnerable to subjective assessments that may affect the final result. To overcome this problem, data mining techniques can be the right solution to help this process by producing new information that can support and assist debtors in determining mitigation policies for each proposed credit decision (Prasojo & Haryatmi, 2021).

Data mining is known as a discipline that can find patterns from data sets that do not represent anything into new knowledge that can be useful in determining a decision (Pratama et al., 2022). In this data mining process, one of the problems that often arises is the class imbalance. This happens, where the sample size of one class is much larger than another, causing minority samples to be treated as noise during the classification process and can affect the accuracy of the data mining algorithm used (Amien et al., 2022). To overcome the problem of class imbalance, oversampling techniques such as ADASYN (Adaptive Synthetic) can be carried out. The ADASYN technique is useful for balancing classes by generating samples adaptively from synthetic samples in the minority class (Monika et al., 2023), so that it can increase the representation of the minority class in the dataset without losing information from the majority class.

There are several studies that have applied oversampling and data mining techniques, including research from (Hidayat et al., 2021) which tested the ADASYN and SMOTE methods on the Airbnb dataset. This research shows that the use of these two oversampling techniques has similarities in improving the performance of the modeling algorithm used, namely the SVM algorithm. There is also research that utilizes the ADASYN and RFECV methods to balance class proportions and feature selection (Pratama et al.,

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v6i3.299
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

2022) which results in an increase in the accuracy performance of the k-NN bagging method used by 2.9%, resulting in a final accuracy of 88%. Then, research from (Amien et al., 2022) tested the dataset to detect malware attacks by implementing the ADASYN method and the Random Forest algorithm. The results of this research prove that ADASYN handles data imbalance well so that the algorithm performance is optimal so that an accuracy rate of 99.86% is obtained.

Referring to the research and explanation above, this research aims to develop an effective predictive model on the loan status prediction dataset by applying the ADASYN oversampling technique and the Decision Tree algorithm. This was done to find out how the oversampling technique can provide better accuracy, precision, recall and F1-score results than the data mining algorithm used, namely Decision Tree. The contribution of this research is increasing the accuracy of the classification model by proving that the ADASYN oversampling technique in the Decision algorithm can increase accuracy in imbalance class cases. This research also provides an analysis of the effectiveness of ADASYN and provides valuable insight into overcoming class imbalance using this technique, which can be adopted to improve the performance of classification algorithms on imbalanced datasets, especially in credit decision making.

RESEARCH METHODS

The stages in the workflow carried out in this study are illustrated through the following figure:

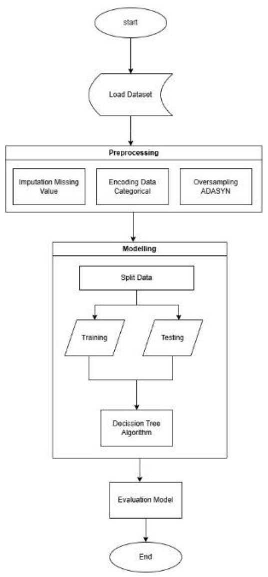

Figure 2. Research Workflow

Figure 2. presents all the processes carried out to achieve analysis results from loan status prediction data. This research first begins by determining the topic to be discussed and the dataset to be processed, namely prediction of loan status. The next stage is to load the dataset to be processed and carry out preprocessing on the dataset. This process includes various stages including filling in empty data using the Imputation Missing Value technique using mean and mode, encoding, feature scaling, and oversampling techniques to overcome class imbalance. During the research process, several oversampling techniques consisting of ADASYN, SMOTE (Synthetic Minority Over-sampling Technique) and Random Oversampling, were carried out to determine the advantages and disadvantages of the ADASYN approach taken. The application of this oversampling technique with ADASYN adaptively produces synthetic samples from minority classes based on the distance from existing minority instances. The parameter selection of the ratio between the number of minority and majority samples in the dataset as well as the distance between synthetic samples is determined based on empirical experiments so as to achieve optimal results in increasing prediction accuracy. The results of this preprocessing stage are then divided into 2 parts, namely Training and Testing with a ratio of 70:30 and further processed at the modeling stage using the Decision Tree Algorithm. Finally, the modeling test results are evaluated using performance metrics which provide output in the form of Accuracy, AUC, Precision, Recall and F1 scores as well as ROC graphs.

Types of research

This research involves historical data from loan applications both approved and rejected previously. The application is measured based on various factors ranging from the borrower's profile, the amount of loans and the property owned. The purpose of data processing in this study is to apply and evaluate a data mining model in order to update the procedures used in determining loan application decisions.

Research Target / Subject

The subject of this research is loan application data from publicly obtained datasets. This data is the main focus of analysis and development of predictive models, so the target of this research is to find a model that can produce predictions for loan applications that are considered optimal.

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v6i3.299

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Data Collection

In this research, data collection was carried out by utilizing secondary data sources that can be accessed publicly via the Kaggle Repository. The dataset chosen is Loan Status Prediction which is obtained in CSV format and has 12 attributes with a target class in the form of loan decision status which consists of the options 'Y' which means approved and 'N' which means rejected. This dataset contains 381 records which can be accessed on the following pages:
https://www.kaggle.com/datasets/bhavikjikadara/loan-status-prediction/

RESULTS AND DISCUSSION

Dataset

The data collection in this study comes from Kaggle and includes 12 key attributes that provide a complete understanding of the applicant's profile, which can be used to predict loan status. These attributes include gender, marital status, number of dependents, education level, independent employment status, applicant's income, loan amount, loan term, credit history, property area, joint applicant's income, and loan status. With this data, in-depth analysis and the development of accurate predictive models can be carried out to determine the likelihood of loan approval or rejection based on various demographic and financial characteristics of the applicant.

Data Exploration and Visualization

Data exploration and visualization aims to understand patterns, trends and anomalies in datasets. Several attributes are explored and visualized in graphical form as follows:

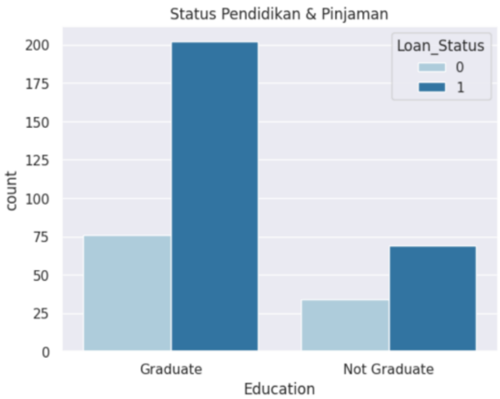

Figure 3. Education & Loan Status attributes

The number of loan applicants based on education status is presented in the form of a bar graph in Figure 3., with additional attributes displaying the approval or rejection status of the loan. This graph illustrates the relationship between education level and loan approval status, showing that applicants with higher education tend to receive more loans.

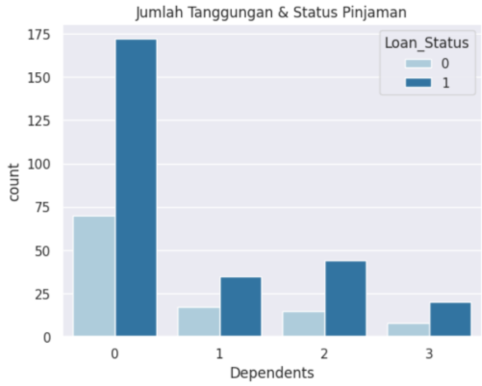

Figure 4. Attributes Number of Dependents and Loan Status

Figure 4 illustrates the correlation between the number of dependents and the status of loan approval. The graph shows that unaccompanied applicants have a higher probability of getting approved. However, even if applicants with dependents have a lower chance, there is still a chance of being approved by considering other additional information.

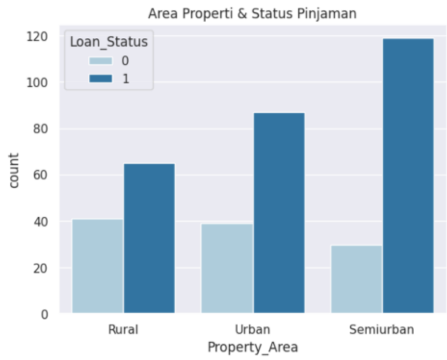

Figure 5. Property Area & Loan Status Attributes

Figure 5. shows the distribution of the number of loan applicants by property area and loan status. From the graph, it can be seen that the number of applicants who successfully obtained approval was higher in Semiurban areas, followed by Urban, and Rural as the lowest. These results indicate that applicants with properties in Semiurban areas have a greater likelihood of being approved for a loan.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v6i3.299
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

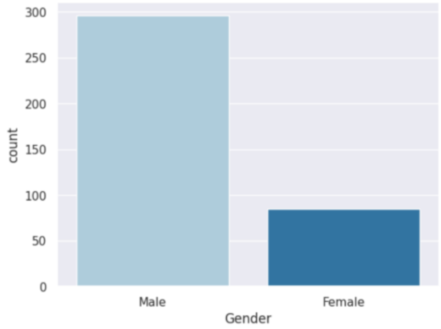

Figure 6. Gender Attribute

Based on Figure 6, it can be seen that the majority of loan applicants are men, amounting to 77%, in accordance with the distribution of the number of loan applicants based on gender attributes. These results provide insight into the characteristics of loan applicants and can serve as a basis for further analysis of the relationship between gender and loan decisions.

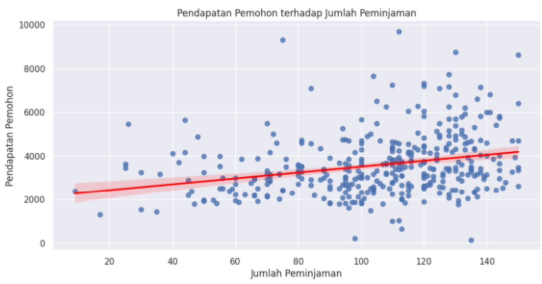

Figure 7. Loan Amount & Applicant Income Attributes

Figure 7. illustrates the distribution of the applicant's income data and the amount of loans to identify patterns or relationships between the two variables. The results of the graph show that there is a positive correlation between the applicant's income and the loan amount, where the line on the scatter plot tends to rise from left to right. This indicates that as the applicant's income increases, the amount of loans applied for also tends to increase.

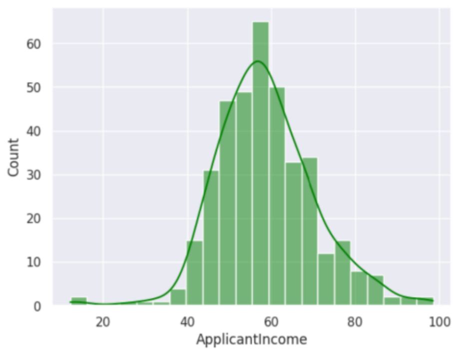

Figure 8. Applicant Income Attributes

Figure 8. displays a histogram depicting the distribution of applicant income, which shows a symmetrical distribution pattern or resembles a normal curve. This pattern indicates that the data tends to be evenly distributed around the middle value, with most values converging around the middle value and then decreasing symmetrically towards the two extremes. These results suggest that the applicant's income has a relatively even or normal distribution.

Preprocessing

a. Imputation Missing Value

In several loan status prediction data attributes, there are empty values, including the gender attribute which is 5, dependent which is 8, self-employed which is 21, loan amount term which is 11, and credit history which is 30. Figure 9 shows the columns that have blank values.

```
Gender               5
Married              0
Dependents           8
Education            0
Self_Employed       21
ApplicantIncome      0
LoanAmount           0
Loan_Amount_Term    11
Credit_History      30
Property_Area        0
CoapplicantIncome    0
Loan_Status          0
dtype: int64
```

Figure 9. Missing Data

To overcome missing or empty values in the dataset, the missing value imputation technique is used, where missing values are filled in based on estimates from the information available in the dataset. In this dataset, attributes such as gender, number of dependents, and independent employment status are categorical data, so that the empty values are filled with the mode value of each attribute. Meanwhile, attributes such as credit history and loan term are included in numerical data, so that empty values are filled with average values. The imputation results can be seen in Figure 10.

```
Gender               0
Married              0
Dependents           0
Education            0
Self_Employed        0
ApplicantIncome      0
LoanAmount           0
Loan_Amount_Term     0
Credit_History       0
Property_Area        0
CoapplicantIncome    0
Loan_Status          0
dtype: int64
```

Figure 10. Imputation Missing Value

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

b. Encoding Data Categorical

The attributes in the dataset for loan status prediction consist of two main types of data, namely categorical data and numerical data. To process data using Python, the data is read and modeled using numerical data, so categorical data such as gender, marital status, number of dependents, education, self-employment status, and property area will be converted into numerical data. Figure 12. shows the results of the process of encoding categorical data into numerical data.

	Gender	Married	Dependents	Education	Self_Employed	Property_Area
0	Male	Yes	1	Graduate	No	Rural
1	Male	Yes	0	Graduate	Yes	Urban
2	Male	Yes	0	Not Graduate	No	Urban
3	Male	No	0	Graduate	No	Urban
4	Male	Yes	0	Not Graduate	No	Urban
...	...	...	...	...	...	...
376	Male	Yes	3+	Graduate	No	Urban
377	Male	Yes	0	Graduate	No	Rural
378	Female	No	0	Graduate	No	Rural
379	Male	Yes	3+	Graduate	No	Rural
380	Female	No	0	Graduate	Yes	Semiurban

Figure 11. Categorical Data

	Gender	Married	Dependents	Education	Self_Employed	Property_Area
0	0	1	1	1	0	0
1	0	1	0	1	1	2
2	0	1	0	0	0	2
3	0	0	0	1	0	2
4	0	1	0	0	0	2
...	...	...	...	...	...	...
376	0	1	4	1	0	2
377	0	1	0	1	0	0
378	1	0	0	1	0	0
379	0	1	4	1	0	0
380	1	0	0	1	1	1

Figure 12. Encoding Data Categorical

c. Over-Sampling ADASYN

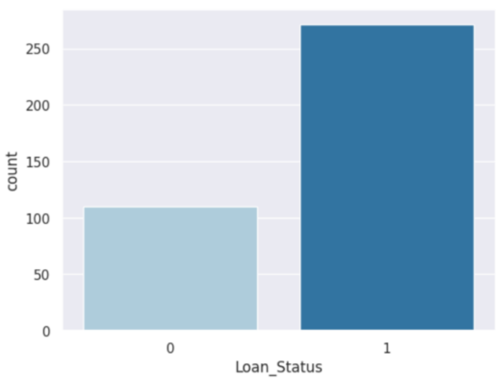

Figure 13. Loan Status Class

Based on Figure 13., there is a significant difference in the distribution of data between classes, where the 'Yes' class has a sample number of 271 data, while the 'No' class only has 110 data. This shows a considerable imbalance between the two. To overcome the problem of data imbalance, the ADASYN (Adaptive Synthetic Sampling) technique was applied in this study. Through ADASYN, points in minority classes that are difficult to separate from the majority are identified, and then new synthetic samples are created around these points by taking into account the level of difficulty in the classification.

After the implementation of the ADACIN Technique, the class distribution has been adjusted significantly. The number of samples between the two classes became more balanced, where the 'No' class had 195 samples, while the 'Yes' class had 190 samples. This indicates that the use of ADASYN successfully improves the representation of minority classes without compromising the important information of the majority class, thus helping to improve the balance and performance of the model on the dataset. The results of ADASYN implementation can be seen in Figure 14.

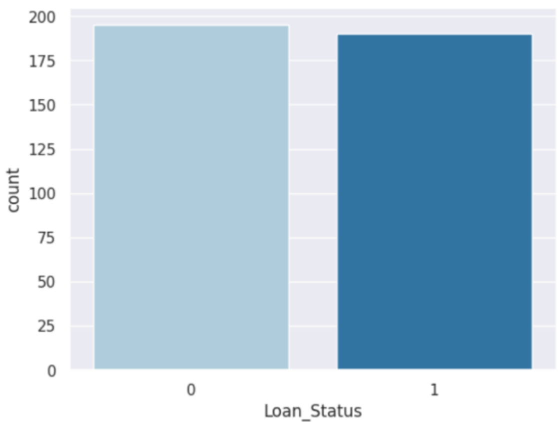

Figure 14. Results of ADASYN Implementation

Modelling Results

To predict loan status, the modeling process is carried out by dividing the dataset into two parts, namely 70% for training data and 30% for testing data. The classification model was built using the Decision Tree Algorithm. Once the modeling is complete, an evaluation is carried out using a variety of performance metrics, including Accuracy, AUC, Precision, Recall, and F1 score. The evaluation results obtained from the model can be seen in Table 1. The table provides a comprehensive understanding of the model's ability to predict the status of loans.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v6i3.299
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Table 1. Evaluation Matrix Results

No.	Matrik Evaluasi	Score without Adasyn	Score with Adasyn
1	Accuracy	73,91%	85,22%
2	AUC	68,68%	76,42%
3	Precision	81,48%	86,25%
4	Recall	81,48%	85,19%
5	F1 Score	81,48%	85,71%

Based on the information from Table 1, it can be seen that the combination of the ADASYN technique and the C4.5 algorithm shows significant results in increasing prediction accuracy on the loan status dataset. From the research results, accuracy increased by 12.22%, from 73.91% to 85.22%. These results show that the ADASYN technique is effective in solving data imbalance problems. In addition, compared to the previous experiment which only used the decision tree algorithm without oversampling techniques, the Area Under the Curve (AUC) of ROC also increased by 7.74%, from 68.68% to 76.42%. This shows a clear improvement in model performance. This shows that the integration of oversampling techniques can substantially improve the performance of prediction models, especially in the context of imbalanced data.

The evaluation results of the loan status prediction model are described using the Confusion Matrix in the following image:

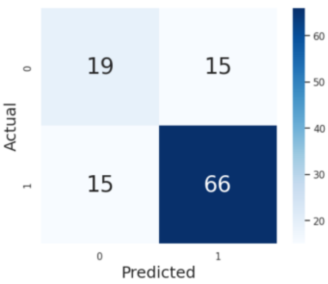

Figure 15. Confusion Matrix of Decision Tree Algorithm Without ADASYN

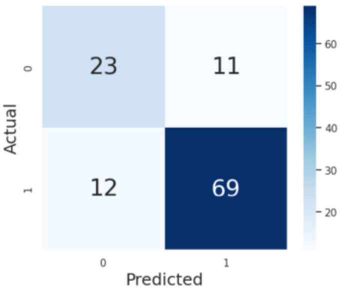

Figure 16. Confusion Matrix of Decision Tree Algorithm With ADASYN

The Confusion Matrix in Figure 16 visualizes the results of the prediction model using the Decision Tree Algorithm with ADASYN which produces 23 data as True Negative (TN), which means that the data was accurately predicted as a loan that was not approved and was indeed not approved. A total of 69 data were identified as True Positive (TP), indicating that the model's predictions for approved loans correspond to actual conditions. However, there are 12 data that fall into the False Negative (FN) category, namely loans that should be approved but are predicted not to be approved, and 11 data are in the False Positive (FP) category, where the data is predicted to be an approved loan even though it should not be approved. Without using ADASYN, the results were 19 TN, 66 TP, 15 FN, and 15 FP, which shows that the use of ADASYN helps increase the number of correct predictions and reduce errors in the predictions of approved and disapproved loans.

Apart from that, the results of evaluating the decision tree algorithm with ADASYN can be seen based on the ROC graph which is visible through the resulting AUC value of 0.76416. This proves that the model has quite good performance in classifying loan status and indicates the probability that the model will give a higher score for loan cases that are approved compared to those that are not approved. The resulting ROC graph can be seen in Figure 15.

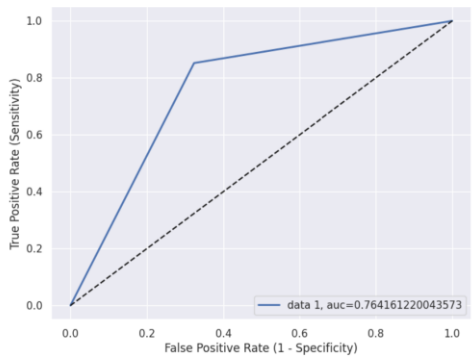

Figure 15. ROC Chart Of Predicts Loan Status

Furthermore, the evaluation results of the loan status prediction model are shown in a graph in Figure 16 with an F1 score of 0.857, which shows that this model has a fairly good ability to balance precision and recall. These values indicate that the model can identify approved and disapproved loans with a high degree of accuracy and completeness. Apart from that, the resulting value also has a good ability to reduce false positives and false negatives.

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

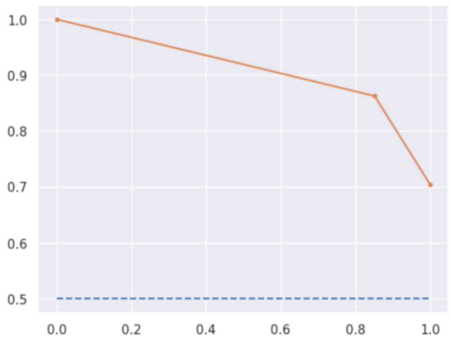

Figure 16. F1 Score Graph

This research also produces a decision tree that provides a visual description of the decision-making process in the model. This decision tree can be used to gain a deeper understanding of how the model classifies loan status based on the various attributes available (See Figure 17).

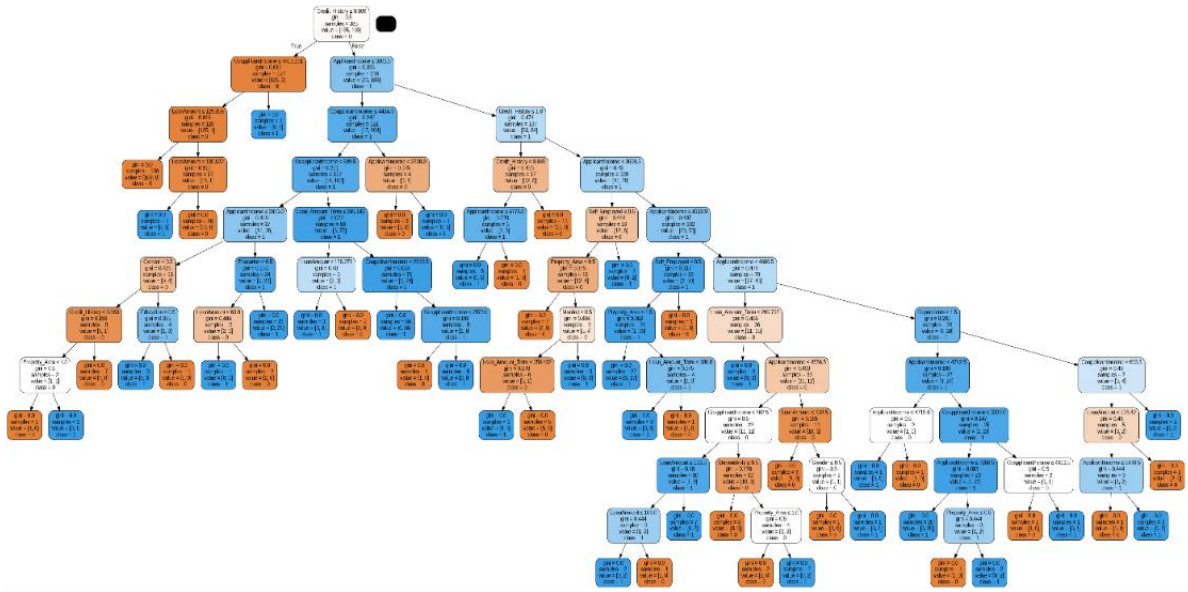
Figure 17. Classification Decision Tree for Loan Status Prediction

Validation of results is done by dividing the data into training data and test data. These validation results show increased accuracy and confirm the effectiveness of the approach used to ensure that the resulting model has good performance in predicting loan status. Further evaluation with more complex validation techniques could be the next research step to ensure the reliability of the model more broadly.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This research reveals that there is a class imbalance in the loan status dataset which can affect the performance of the classification algorithm so that oversampling techniques are needed. In this case, an oversampling technique (Adaptive Synthetic Sampling) is applied which is then integrated with the Decision Tree algorithm and evaluated using several performance metrics to ensure the effectiveness of the technique used. The evaluation results show that the integration of these two methods is able to produce a prediction model with an accuracy value that was initially 73,91%, increasing to 85.22%. Based on these results, it proves that the prediction model tested has good performance in classifying loan status with a fairly high level of accuracy, and provides strong potential for practical application in predicting loan status. However, this research is still limited because it only focuses on one dataset source and does not include an in-depth explanation of comparisons with other methods.

Suggestion

In future research, it is recommended to expand the data preprocessing technique and explore it using other algorithms besides Decision Tree. An in-depth analysis of the modeling results by developing them into a system is also recommended so that this research can provide direct benefits to users.

REFERENCES

Amien, J. Al, Yoze Rizki, & Mukhlis Ali Rahman Nasution. (2022). Implementasi Adasyn Untuk Imbalance Data Pada Dataset UNSW-NB15 Adasyn Implementation For Data Imbalance on UNSW-NB15 Dataset. *Jurnal CoSciTech (Computer Science and Information Technology)*, *3*(3), 242–248. https://doi.org/10.37859/coscitech.v3i3.4339

Badan Pusat Statistik. (2019). *Posisi Kredit Perbankan Menurut Jenis Penggunaan Pada Bank Pemerintah (Juta Rupiah)*. BPS. https://ppukab.bps.go.id/indicator/105/158/1/posisi-kredit-perbankan-menurut-jenis-penggunaan-pada-bank-pemerintah-.html

Eltania, M. (2022). Pengaruh Suku Bunga Kredit, Inflasi, Dan Nilai Tukar Terhadap Jenis Penyaluran Kredit. *Contemporary Studies in Economic, Finance and Banking*, *1 NO 1*(1), 25–37.

GAO Report. (2023). *Pandemic Assistance Likely Helped Reduce Balances, and Credit Terms Varied Among Demographic Groups*. U.S. Government Accountability Office. https://www.gao.gov/products/gao-23-105269

Hermawan, F. F., & Yamasari, Y. (2022). Implementasi K-Nearest Neighbor dengan Pemilihan Fitur pada Aplikasi Prediksi Kelayakan Pengajuan Pinjaman. *Journal of Informatics and Computer Science (JINACS)*, *3*(04), 411–424. https://doi.org/10.26740/jinacs.v3n04.p411-424

Hidayat, W., Ardiansyah, M., & Setyanto, A. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Edumatic: Jurnal Pendidikan Informatika*, *5*(1), 11–20. https://doi.org/10.29408/edumatic.v5i1.3125

Monika, A. P., Risti, F. E. P., Binanto, I., & Sianipar, N. F. (2023). Perbandingan Algoritma Klasifikasi Random Forest , Gaussian Naive Bayes , dan K-Nearest Neighbor untuk Data Tidak Seimbang dan Data yang diseimbangkan dengan Metode Adaptive Synthetic pada Dataset LCMS Tanaman Keladi Tikus. *Jurnal Seminar Nasional Teknik Elektro, Informatika & Sistem Informasi (SINTaKS)*, 3–7.

Nurdiyanto, I., Nurdiawan, O., Irma Purnamasari, A., & Ade Kurnia, D. (2022). Penentuan Keputusan Pemberian Pinjaman Kredit Menggunakan Algoritma C.45. *Jurnal Dadta Science Dan Informatika*, *2*(1), 1–5.

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *7*(2), 79–89. https://doi.org/10.25077/teknosi.v7i2.2021.79-89

Pratama, I., Chandra, A. Y., & Presetyaningrum, P. T. (2022). Seleksi Fitur dan Penanganan Imbalanced Data menggunakan RFECV dan ADASYN. *Jurnal Eksplora Informatika*, *11*(1), 38–49. https://doi.org/10.30864/eksplora.v11i1.578

Pratiwi, A. A., Saraswati, W. T., Ardiansyah, R. F., Rouf, E. H., & Pratama, R. A. (2023). Determining The Loan Feasiblity of Bank Customers Using Naïve Bayes, K-Nearest Neighbors And Linear Regression Algorithms. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, *6*(3), 226–236.

Prawira, A., Arisandi, D., & Sutrisno, T. (2022). Penerapan Algoritma Naive Bayes dan Multiple Linear Regression Untuk Prediksi Status dan Plafon Kredit (Studi Kasus: Bank ABC). *Journal on Education*, *5*(1), 1075–1087. https://doi.org/10.31004/joe.v5i1.720

Sitepu, R., & Manohar, M. (2022). Implementasi Algoritma K-Nearest Neigbor Untuk Klasifikasi Pengajuan Kredit. *Jurnal Sistem Informasi, Teknik Informatika Dan Teknologi Pendidikan*, *1*(2), 49–56. https://doi.org/10.55338/justikpen.v1i2.6

Syafudin, S., Nugraha, R. A., Handayani, K., Gata, W., & Linawati, S. (2021). Prediksi Status Pinjaman Bank dengan Deep Learning Neural Network (DNN). *Jurnal Teknik Komputer AMIK BSI*, *7*(2), 130–135.

Tarigan, I. F., Hartama, D., Suhada, Saifullah, & Saragih, I. S. (2021). Penerapan Data Mining Pada Prediksi Kelayakan Pemohon Kredit Mobil Dengan K-Medoids Clustering. *KLIK: Kajian Ilmiah Informatika ...*, *1*(4), 170–179. http://www.djournals.com/klik/article/view/153

Ubaedi, I., & Djaksana, Y. M. (2022). Optimasi Algoritma C4.5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. *JSiI (Jurnal Sistem Informasi)*, *9*(1), 17–26.

https://doi.org/10.30656/jsii.v9i1.3505

Yenila, F., Marfalino, H., & Defit, S. (2023). Model Analisis Machine Learning dengan Pendekatan Deep Learning dalam Penentuan Kolektabilitas. *JST (Jurnal Sains Dan Teknologi)*, *12*(2), 403–414. https://doi.org/10.23887/jstundiksha.v12i2.54035