Brilliance
Research of Artificial Intelligence

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

Application of the Learning Vector Quantization Algorithm for Classification of Students with the Potential to Drop Out

I Gusti Made Wahyu Krisna Widiantara[1])*, Kadek Yota Ernanda Aryanto[2]), I Made Gede Sunarya[3])
[1)2)3)]Department of Computer Science, Postgraduate Program, Universitas Pendidikan Ganesha, Indonesia
[1)]whykrisna@email.com, [2)]yota.ernanda@undiksha.ac.id, [3)]sunarya@undiksha.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 08-11-2023
Accepted: 10-11-2023
Published: 17-11-2023

Keywords:
Artificial Neural Network; Classification; Dropouts; Learning Vector Quantization; Unbalanced Data

ABSTRACT

Universities, as providers of academic education services, are required to provide an optimal educational process for students to produce a generation of quality human beings. Student learning success is seen as the success of universities in implementing the higher education process. One of the problems universities face in maintaining the quality of education is student dropout. The high dropout rate in universities can impact accreditation assessments. As a result, it will affect the level of public trust. The number of dropouts in higher education can be minimized from an early age by analyzing the factors that cause student dropouts using data on students who graduated and those who dropped out. This data can be used to determine student dropout patterns by classifying them using the artificial neural network learning vector quantization (LVQ) approach. The data used in this research was 4053, consisting of 3840 graduate student data and 213 dropout student data. This data is considered unbalanced, an unbalanced dataset can cause errors because the model tends to classify the majority class with a high classification and pays less attention to the minority class. So, it is necessary to apply oversampling techniques to overcome this problem. The research results show that the application of the LVQ method to unbalanced data produces an accuracy value of 95.53%, a precision value of 100%, a recall value of 15.02% and an f1-score of 0.26, while the application of the LVQ method to data that has undergone resampling resulting in an accuracy value of 94.66%, a precision value of 92.22%, a recall value of 97.55%, and an f1-score value of 0.95. The LVQ method can be used to classify dropout students with excellent results.

INTRODUCTION

College as an academic education service provider, we are required to be able to provide an optimal educational process (Ramayasa 2018). The success of the student learning process is often assessed as the success of the university in organizing the educational process. On the other hand, student failure is seen as the inability of universities as organizers of the educational process (Nasrullah 2018). One of the factors that is assessed in higher education is the timeliness of student graduation (Sari, Kusrini, and Sunyoto 2021). Students whose study period exceeds the time determined by the university will be declared a dropout. Dropout is the revocation of a student's student status. Apart from being due to pass the study period, other things determine the application of the status dropout to students who have been determined by the university concerned. High-level dropout students at higher education institutions can impact accreditation assessments. As a result, it will affect public trust in higher education institutions (Ratniasih 2019). Several students dropping out of higher education can be minimized from an early age by analyzing the characteristics or causal factors of students who drop out using student data at universities. This data can be used to determine student patterns or characteristics of dropout. Student data in higher education is an important thing to analyze, especially to find student dropout patterns, so that it can be used as a guide for education administrators to provide direction to students from the start of their education.

Previous research by (Muchamad Taufiq Anwar, Lucky Heriyanto, and Fadhla Fanini 2021) predicted student dropout using the 1st-semester achievement index, 2nd-semester achievement index, 3rd-semester achievement index, and 4th-semester achievement index. It showed that the feature that had the most influence on students graduating or dropping out was graded semester 1 achievement index and semester 2 achievement index values. The expected update in this research is to explore features other than semester achievement index values. Other research related to student dropouts was conducted by (Wahyuni, Saragih, & Angin-Angin, 2018) using the C4.5 algorithm using features of parent's income, parent's occupation, students age, school, and cumulative achievement index. The accuracy value obtained was 59.58%. The update in this research is to use another algorithm that is better in terms of accuracy and processing speed and uses variables that cause student dropout that are appropriate to the research location. The data

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

used in this research is student data drop out and pass with the features used as input parameters are family income, semester 1 achievement index, semester 2 achievement index, semester 3 achievement index, semester 4 achievement index, number of absences semester 1, number of absences semester 2, number of absences semester 3, number of absences semester 4 (Armansyah, 2021; Kusumawati, Factsari, & Redjeki, 2019a; Nasrullah, 2018; Bali State Polytechnic, 2021). The dataset in this study needs to be balanced. Classification carried out on a dataset that needs to be balanced between classes can cause errors in classification.(Indrawati, 2021; Wulan Purnami, Master of Statistics Department, Faculty of Mathematics & Data Science, 2018), so it needs to be handled by carrying out a data resampling process using an oversampling technique with the Synthetic Minority Oversampling Technique (SMOTE) method using the RapidMiner tool. This research aims to utilize student data at universities to classify student dropouts using the learning vector quantization (LVQ) algorithm. LVQ is one of the classification methods of artificial neural networks that applies the principle of competition (Rahmawati 2019). LVQ is a grouping method where the labels for each class have been determined (supervised learning). This algorithm aims to find the right weight value to classify the input vectors into the target class that was initialized when forming the LVQ model. In contrast, the testing algorithm calculates the output value closest to the input vector (Nurpadillah et al. 2019).

LITERATURE REVIEW

(Muchamad Taufiq Anwar et al. 2021) Develop a model to predict students who have the potential to drop out using data mining techniques. The dataset used is student data from the 2014-2019 class. The initial process uses python and continues with modeling with the C4.5 / J48 algorithm in the WEKA (Waikato Environment for Knowledge Analysis) application. Obtained results that show the features that have the most influence on students graduating or dropping out are the achievement index value for semester 1 and the achievement index value for semester 2, and the resulting accuracy is 90.6%. The expected update in this research is to explore other features that influence student dropout to optimize model performance.

(Nasrullah, 2018) Researching students who have the potential to drop out using the C4.5 method produces 17 rules which are used as a reference for identifying students who have the potential to drop out by using 9 features in carrying out classification, namely regional origin, gender, class, religion, age, semester 1 achievement index, semester achievement index. 2, semester 3 achievement index, and semester 4 achievement index. Where the religious feature is not used because the gain value obtained for the religious feature in the results of the second node gain calculation is -4.53, in this research, several features are used based on patterns that have been obtained in the research carried out (Nasrullah, 2018), namely semester 1 achievement index, semester 2 achievement index, semester 3 achievement index, and semester 4 achievement index.

(Wahyuni et al., 2018) Implementing the C4.5 algorithm to research student dropout. The features used are parent’s income, occupation, cumulative achievement index, student age, and school origin. The features of the cumulative achievement index, age, and parent’s income are grouped into one to anticipate many ramifications. The results obtained were an accuracy value of 59.58% with an error rate of 40.42%. The update carried out in this research is to use another better algorithm in terms of accuracy and processing speed and uses causal features dropout students appropriate to the research location.

(Kusumawati et al. 2019a) Develop a model to identify early dropout students using the Dempster-Shafer method. Symptoms or factors that influence students to drop out are obtained from knowledge about dropout on campuses in Yogyakarta, which is divided into 2 factors that have a significant influence on dropout cases in higher education, namely student intelligence and parent’s income (sociodemographics). This research uses socio-demographic factors, namely family income, as an attribute in determining student’s dropout based on research results (Kusumawati, Factsari, & Redjeki, 2019b).

(Pamungkas, Muflikhah, and Wihandika 2019) Grouping recipients of the Program Keluarga Harapan (PKH) using the Learning Vector Quantization algorithm (Case Study of Kedungjati Village) with the parameters used learning rate (α)=0.7, decreasing learning rate (Dec α)=0.3, maxEpoch=2, minimum learning rate (Min α) = 0.01 to get an accuracy result of 100%.

(Setyowati and Mariani 2021) Applying artificial neural network techniques using the Learning Vector Quantization (LVQ) algorithm in classifying Acute Respiratory Tract Infection (ARI) diseases with the parameters used learning rate (α) = 0.02, learning rate decrease value (Dec α) = 0.01, MaxEpoch =20, comparison of training data and test data 80:20. The accuracy results obtained on average reached 96.5% with the highest accuracy obtained at 100%.

(Tantiati, Tanzil Furqon, and Dewi 2019) Implemented the Learning Vector Quantization (LVQ) method to classify labor into 2 target classes, namely normal labor and risk labor. With the input features used, namely hemoglobin, cell levels, blood pressure measurements, pelvic size, age, fetal position, psychological examination results of the prospective mother, interpretation of fetal weight, proteinuria, and upper arm circumference with learning

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

parameters rate (α) = 0.1 and maximum iteration 24 with a ratio of the amount of training data and test data (64:16) to obtain an accuracy value of 93.78%.

METHOD

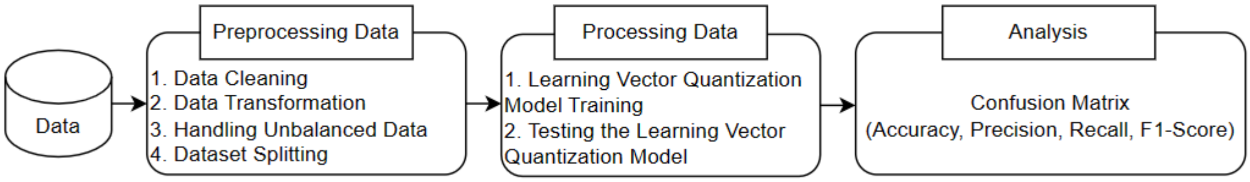

Fig.1 Stages of Research Methods

Datasets

The data used in this research is secondary data obtained from the Bali State Polytechnic. The data obtained is from D3 and D4 education level students from the 2014/2015 academic year to the 2019/2020 academic year.

Data Cleaning

First, an examination of inconsistent data and missing values was carried out by looking directly at the available data using the Microsoft Excel application and looking to see if there were inconsistent values, namely data that did not match the problem boundaries in this research. For example, data with labels other than pass and dropout and checking whether there are empty values for each feature of the dataset used in this research, then deleting the data. Next, feature selection was carried out. In this research, the features used were selected based on the semester graduation requirements listed in the Bali State Polytechnic education guidebook as well as the features used in previous research regarding dropout students (Armansyah, 2021; Kusumawati et al., 2019; Nasrullah, 2018; Bali State Polytechnic, 2021), allows selecting the most relevant and significant features in representing the data used. In this research, this step is taken to ensure that only the necessary features are used in the Learning Vector Quantization (LVQ) model in accordance with the literature study that has been carried out. The next stage is to test the selected features using the chi-square test to ensure that the selected features have a strong relationship with the target variable (Gustantia Annisa et al. 2020; Igo Cahya Negara 2018).

Data Transformation

This stage is a data transformation process that has been selected after going through a data cleaning process. The features that are factors causing student dropout are first transformed into binary scale data to make it easier to compare different categories. The value given to each feature in this study is 0 or 1, a value of 0 is given for descriptions of features that tend to pass, and a value of 1 is given for descriptions of features with the potential to experience dropout.

Handling Unbalanced Data

In this study, unbalanced data was handled using an oversampling technique using the Synthetic Minority Oversampling Technique (SMOTE) method.

Dataset Splitting

Research data is divided into 2 parts, namely training data and test data. Training data is used to train the classification model, while test data is used to evaluate model performance. Dataset division in this study was carried out using k-fold cross validation with a value of k=10. The dataset is divided into ten folds, and at each iteration, one fold is taken as test data, and the other is used as training data. The selection of test data is adjusted to the order of iteration in fold order, namely test data in the 1st iteration using data in fold 1, test data in the 2nd iteration using data in fold 2, and so on.

Learning Vector Quantization (LVQ) Model Training

The training or model building process using the Learning Vector Quantization (LVQ) algorithm aims to find the final weight of each class by calculating the minimum distance between the input vector and the initial weight selected from the dataset using the Euclidean distance technique. In the process of forming the LVQ model, the classification model will produce a vector for each class as a representation of each class, which is the final weight of each feature. The final weight obtained will be used in the testing process.

Brilliance
Research of Artificial Intelligence

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

Testing the Learning Vector Quantization (LVQ) Model

In this process, testing is carried out on the ability of the LVQ algorithm to classify input variables in the form of factors that cause students to drop out of the expected class or target. In the classification process, each new data will be classified based on its closest distance to the weight vector. In other words, the LVQ testing process is to find the minimum distance between the input and the weights to determine the class or target of the input vector. The results obtained from this testing process are the classification of the input vector, which is the class that has the smallest distance to the input vector.

Analysis

After all the test results are obtained, the next stage is to carry out an analysis of the test results using the confusion matrix table by calculating the accuracy, error rate, precision, recall, and f1-score values. The analysis in question includes, among other things, an analysis of model performance using all features before and after handling unbalanced data and then analyzing the influence of each feature on the causes of student dropout.

RESULT

Datasets

The data obtained was 8186 data with features NIM, name, gender, level, major, study program, family income, 1st-semester achievement index, 2nd-semester achievement index, 3rd-semester achievement index, 4th-semester achievement index, 1st-semester attendance, 2nd-semester attendance, 3rd-semester attendance and 4th-semester attendance as well as pass or dropout labels for each data. After carrying out the data cleaning process by deleting inconsistent data and missing values and selecting and testing features, 4053 data were obtained to be used in this research, with 3840 students graduating and 213 students dropping out. There are 9 features used in this research, namely family income, 1st-semester achievement index, 2nd-semester achievement index, 3rd-semester achievement index, 4th-semester achievement index, 1st-semester attendance, 2nd-semester attendance, 3rd-semester attendance, and 4th-semester attendance previously by (Muchamad Taufiq Anwar et al. 2021) which only uses 4 features, namely semester 1 achievement index, semester 2 achievement index, semester 3 achievement index and semester 4 achievement index. Next, the data is transformed into binary form. The transformation results are shown in**Error! Reference source not found.**.

Table 1. Research Data After Transformation to Binary Form

No.	PK	IP1	IP2	IP3	IP4	A1	A2	A3	A4	Label
1	1	0	0	0	1	0	0	0	1	D
2	1	1	0	0	1	0	1	0	0	D
3	0	0	0	0	0	0	0	0	0	L
4	0	0	0	0	1	0	0	0	1	D
5	0	0	0	0	0	0	0	0	0	L
...	0	0	0	0	0	0	0	0	0	L
...	0	0	0	0	0	0	0	0	0	L
4053	1	0	0	1	1	0	0	1	1	D

The data was then oversampled using the Synthetic Minority Oversampling Technique (SMOTE) method. A comparison of the amount of data between classes before and after handling unbalanced data is shown inTable 2.

Table 2. Comparison of the amount of data between classes before and after handling unbalanced data

Information	Amount of data		
	Passed	Dropouts	Total
Before handling unbalanced data	3840	213	4053
After handling unbalanced data	3840	3840	7680

The dataset used in this research after going through the data resampling process was 7680 data, with 3840 students graduating and 3840 students dropping out. From a series of data preparation processes that have been carried out, namely data cleaning, data transformation, and handling of unbalanced data, data has been formed that is ready to be used for classification. The next stage is dividing the dataset using cross validation with a value of k=10. The ratio of the amount of data in training data and test data is 90:10. The amount of data for each fold is 1/10 of the total data, namely 768 data with experiments carried out in 10 iterations, where each iteration uses the kth fold as test data.

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

Iterations are performed 10 times, and each fold is used as test data. The training data in this study uses 6912 data, as shown inTable 3. 90% of the training data can provide enough data to train a classification model. The more data used for training, the better the model will likely learn the patterns in that data (Rahmawati 2019). Meanwhile, the test data in this study used 768 data, as shown inTable 4.

Table 3. Training Data

No.	X1	X2	X3	X4	X5	X6	X7	X8	X9	Label
1	1	0	0	1	0	0	0	0	0	D
2	0	0	0	0	0	0	0	0	0	L
3	1	0	0	1	0	0	1	0	0	D
4	0	0	1	0	0	0	0	1	0	D
5	1	1	0	0	0	1	0	0	0	D
...	0	0	0	0	0	0	0	0	0	L
...	0	0	0	0	0	0	0	0	0	L
900	0	0	0	0	0	0	0	0	0	L

Table 4. Test Data

No.	X1	X2	X3	X4	X5	X6	X7	X8	X9	Label
1	1	0	0	1	0	0	1	0	0	D
2	0	0	1	0	1	0	1	0	0	D
3	0	0	0	0	0	0	0	0	0	L
4	1	0	0	0	0	0	0	0	0	L
5	0	1	0	0	0	0	0	0	0	L
...	1	0	0	0	0	0	1	0	0	D
...	1	0	0	0	0	0	0	0	0	L
100	0	0	0	0	0	0	0	0	0	L

Classification

Classification is carried out using the Learning Vector Quantization (LVQ) method with parameters learning rate (α) = 0.05 and maximum epoch (MaxEpoch) = 1000. Classification in this research was carried out on a dataset that used all the selected features of the dataset before and after handling unbalanced data and compared the performance of both using a confusion matrix with an evaluation matrix of accuracy, precision, recall, and f1-score, then continued by carrying out classification on the dataset after handling unbalanced data by subtracting one feature in turn to analyze the influence of each feature on student dropout. The classification results on the dataset before handling imbalanced data are shown inTable 5 while the classification results on the dataset after handling imbalanced data are shown inTable 6.

Table 5. Confusion Matrix Before Handling Unbalanced Data

		Actual Class	
		Dropouts	Passed
Predicted Class	Dropouts	32	0
	Passed	181	3840

Based onTable 5, information was obtained regarding the performance of the model produced before handling unbalanced data. The total data used was 4053 data. The actual number of students dropout and predicted correctly into the class dropout (TP) totaled 32 data. The number of students who actually passed and were incorrectly predicted to enter the class dropout (FP) amounts to 0 data, the actual number of students dropout and incorrectly predicted to be in the passing class (FN) totaling 181 data, the number of students who actually passed and correctly predicted to be in the passing class (TN) was 3840 data. From 4053 data, 32+3840=3872 data are classified in the correct class. In other words, the accuracy level is 95.53%. The first row shows that 32 data were correctly classified in the dropout class out of 32 data predicted in the dropout class. In other words, the resulting precision value was 100%. The first column shows 32+181=213 data in the dropout class, but only 32 are classified correctly. In other words, the resulting recall value is 15.02%. The resulting f1-score value is 0.26.

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

Table 6. Confusion Matrix After Handling Unbalanced Data

		Actual Class	
		Dropouts	Passed
Predicted Class	Dropouts	3746	316
	Passed	94	3524

Based onTable 6 information was obtained regarding the model's performance produced after handling unbalanced data. The total data used was 7680 data. The actual number of student dropouts was predicted correctly in the class dropout (TP) totaled 3746 data. The number of data on students who actually passed and were incorrectly predicted to enter the class dropout (FP) totaled 316 data. The actual number of students who dropped and incorrectly predicted to be in the passing class (FN) totaled 94 data, and the number of data for students who actually passed and correctly predicted to be in the passing class (TN) was 3524. From 7680 data, 3746+3524=7270 data are classified in the correct class. In other words, the accuracy level is 94.66%. The first row shows 3746+316=4062 data classified in the dropout class, but only 3746 data are classified correctly. In other words, the resulting precision value is 92.22%. The first column shows that there are 3746+94=3840 data in the dropout class, but only 3746 data are classified correctly, in other words, the resulting recall value is 97.55%. The resulting f1-score value is 0.95. The f1-score value ranges between 0 and 1, the higher the value, the better the model performance.

DISCUSSION

Before handling unbalanced data, it produces an accuracy value of 95.53%, a precision of 100%, a recall of 15.02%, and an f1-score of 0.26. After handling unbalanced data, it produces an accuracy value of 94.66%, precision of 92.22%, recall of 97.55%, and f1-score of 0.95. In the classification of students who have the potential to drop out, an evaluation metric that needs attention apart from accuracy is recall. Because recall measures the extent to which the classification model can correctly identify student cases. Dropout of all cases classified as dropout: recall provides information about how accurate the model is in classifying student dropout cases. With high recall, the model can minimize the number of errors and increase the model’s reliability in classifying student cases. The higher the recall value, the fewer dropout students are classified as passed. In other words, this research prefers classification results with cases where students are predicted to drop out because it is true that the student is a dropout. We do not want cases where students are classified as passed, but the truth is dropout. Accuracy and precision values in model testing before handling unbalanced data produced greater values, namely 95.53% and 100%, while model testing after handling unbalanced data produced accuracy and precision values of 94.66% and 92.22%. However, this cannot show that model testing before handling unbalanced data is better than model testing after handling unbalanced data, especially when the target class is unbalanced. Classification carried out on unbalanced data will produce wrong predictions because the model focuses too much on the majority class and tends to ignore the minority class. An evaluation metric that can be used to compare the performance of classification models before and after fairly handling imbalanced data is the f1-score because the f1-score is a combination of precision and recall, which is useful for taking into account both true positive predictions (precision) and the model's ability to find all true positive samples (recall). The F1-score provides a more comprehensive picture of the model's ability to deal with class imbalance, giving equal attention to both classes and keeping the model from focusing only on the majority class. So, the f1-score is very useful in comparing model performance on unbalanced data because it provides a balanced assessment of precision and recall. In this study, the minority class, namely dropout, has a significant value that the model must recognize. Thus, if you compare the overall evaluation metric results in testing before and after handling unbalanced data, the test results on balanced data have better model performance because they produce a higher f1-score value, namely 0.95. This shows that handling unbalanced data at the data preprocessing stage influences model performance in the Learning Vector Quantization (LVQ) algorithm.

After testing the LVQ model using the complete features and without one of the features, the data is obtained, as shown inTable 7.

Table 7. Test Results Using Complete Features and Without One Feature

No	Classification	Accuracy	Precision (%)	Recalls (%)	F1-Score
1	Complete features	94.66	92.22	97.55	0.95
2	No family income feature	89.73	90.05	89.32	0.90
3	Without the IPS 1 feature	93.98	91.89	96.48	0.94
4	Without the IPS 2 feature	93.41	91.01	96.32	0.94
5	Without the IPS 3 feature	91.27	87.32	96.56	0.92

Brilliance
Research of Artificial Intelligence

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

6	Without the IPS 4 feature	93.65	91.06	96.80	0.94
7	No 1st-semester attendance feature	93.98	91.33	97.19	0.94
8	No 2nd-semester attendance feature	94.10	91.54	97.19	0.94
9	No 3rd-semester attendance feature	93.67	90.90	97.06	0.94
10	No 4rd-semester attendance feature	93.39	90.50	96.95	0.94

Based on data from Table 7, information was obtained that all the features used in the classification of dropout students affect model performance. The absence of one of the features will cause a decrease in the level of accuracy, precision, recall, and f1-score. The magnitude of the influence of each feature on the level of accuracy and f1-score is shown inFig.2.

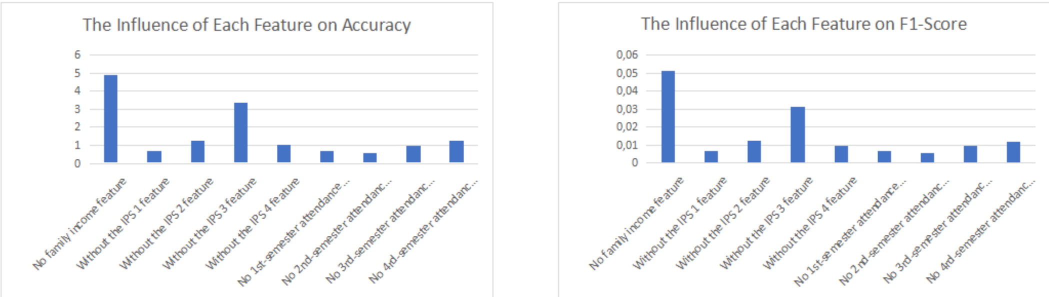

Fig.2 2 The Influence of Each Feature on Accuracy and F1-Score

Based on Fig.2, the most influential feature in this research is the family income feature, with an accuracy contribution of 4.93% and an f1-score value contribution of 0.05.

CONCLUSION

The data used in this research was 7680 data, which was divided into 6912 training data and 768 test data. The features used consist of family income, 1st-semester achievement index, 2nd-semester achievement index, 3rd-semester achievement index, 4rd-semester achievement index, 1st-semester attendance, 2nd-semester attendance, 3rd-semester attendance and 4rd-semester attendance. Features used the most influential influence on students dropping out is family income, 2nd-semester achievement index and 3rd-semester achievement index. Learning Vector Quantization (LVQ) is a good method for classifying students drop out. With the adaptability of weight vectors during the training process, LVQ can effectively handle diverse and complex data, including nonlinear relationships between features. LVQ's ability to nonlinearly map data into groups corresponding to its target classes makes it suitable for complex classification problems. Thus, LVQ is a strong choice to overcome the problem of student drop out classification with high accuracy and extraordinary adaptability. The calculation results are as follows: accuracy value 94.66%, precision 92.22%, recall 97.55%, and f1-score 0.95.

Future research is expected to be able to handle imbalanced data better, such as using a combination of undersampling and oversampling techniques, but still be able to maintain important information contained in the dataset and be able to reduce underfitting and overfitting, as well as taking into account the use of other features that can influence students drop out apart from the features used in this research and other factors causing students drop out such as demographic, psychological and social factors, it has been tested that the features used have a strong relationship as factors causing students drop out using the feature testing method.

REFERENCES

Armansyah. 2021. “Prototipe Jaringan Syaraf Tiruan Multilayer Perceptron Untuk Prediksi Mahasiswa Dropout.” *Jurnal Nasional Komputasi Dan Teknologi Informasi* 4:265–70.

Gustantia Annisa, Nurul, Riswan Efendi, Lisya Chairani, Jurusan Matematika, Fakultas Sains, Dan Teknologi, Jurusan Psikologi, and Fakultas Psikologi. 2020. “Hubungan Sistem Pembelajaran Daring Dengan Kesehatan Mental Mahasiswa Di Era COVID-19 Menggunakan Chi-Square Test Dan Dependency Degree.” Pp. 600–607 in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 12*.

Igo Cahya Negara. 2018. “PENGGUNAAN UJI CHI–SQUARE UNTUK MENGETAHUI PENGARUH TINGKAT PENDIDIKAN DAN UMUR TERHADAP PENGETAHUAN PENASUN MENGENAI HIV–AIDS DI PROVINSI DKI JAKARTA.” Pp. 2–8 in *Prosiding Seminar Nasional Matematika dan Terapannya 2018*.

Indrawati, Ariani. 2021. “PENERAPAN TEKNIK KOMBINASI OVERSAMPLING DAN UNDERSAMPLING

E-ISSN: 2807-9035
Volume 3, **Number** 2, November 2023
https://doi.org/10.47709/brilliance.v3i2.3155

UNTUK MENGATASI PERMASALAHAN IMBALANCED DATASET." *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI* 4(1). doi: 10.33387/jiko.

Kusumawati, Dara, Dini Faktasari, and Sri Redjeki. 2019a. "Model Identifikasi Dini Mahasiswa Drop Out Menggunakan Dempster Shafer." 15–22.

Kusumawati, Dara, Dini Faktasari, and Sri Redjeki. 2019b. "Model Identifikasi Dini Mahasiswa Drop Out Menggunakan Dempster Shafer." 15–22.

Muchamad Taufiq Anwar, Lucky Heriyanto, and Fadhla Fanini. 2021. "Model Prediksi Dropout Mahasiswa Menggunakan Teknik Data Mining." *JURNAL INFORMATIKA UPGRIS* 7:56–60.

Nasrullah, Asmaul Husnah. 2018. "Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out." *ILKOM Jurnal Ilmiah* 10(2):244–50. doi: 10.33096/ilkom.v10i2.300.244-250.

Nurpadillah, Dinda Izmya, Haviluddin Haviluddin, Herman Santoso Pakpahan, Islamiyah Islamiyah, and Hario Jati Setyadi. 2019. "Pengenalan Karakter Tulisan Menggunakan Metode Learning Vector Quantization." *Sains, Aplikasi, Komputasi Dan Teknologi Informasi* 1(2):23. doi: 10.30872/jsakti.v1i2.2602.

Pamungkas, Vidya Capristyan, Lailil Muflikhah, and Randy Cahya Wihandika. 2019. *Klasifikasi Penerimaan Program Keluarga Harapan (PKH) Menggunakan Metode Learning Vector Quantization (Studi Kasus Desa Kedungjati)*. Vol. 3.

Politeknik Negeri Bali. 2021. *Pedoman Pendidikan Politeknik Negeri Bali*. Badung.

Rahmawati, Amalia. 2019. "(Amalia Rahmawati, 2019) Perbandingan Klasifikasi Menggunakan Metode Backpropagation Dan Metode LVQ.Pdf."

Ramayasa, I. Putu. 2018. "Perancangan Sistem Klasifikasi Mahasiswa Drop Out Menggunakan Algoritma K-Nearest Neighbor." *Seminar Nasional Sistem Informasi Dan Teknologi Informasi 2018* 1(1):585–89.

Ratniasih, Ni Luh. 2019. "Penerapan Algoritma K-Nearest Neighbour (K-Nn) Untuk Penentuan Mahasiswa Berpotensi Drop Out." *Jurnal Teknologi Informasi Dan Komputer* 5(3):314–18. doi: 10.36002/jutik.v5i3.804.

Sari, Eka Yulia, Kusrini Kusrini, and Andi Sunyoto. 2021. "Analisis Akurasi Jaringan Syaraf Tiruan Dengan Backpropagation Untuk Prediksi Mahasiswa Dropout." *Creative Information Technology Journal* 6(2):85. doi: 10.24076/citec.2019v6i2.235.

Setyowati, Endang, and Scolastika Mariani. 2021. "Penerapan Jaringan Syaraf Tiruan Dengan Metode Learning Vector Quantization (LVQ) Untuk Klasifikasi Penyakit Infeksi Saluran Pernapasan Akut (ISPA)." *PRISMA, Prosiding Seminar Nasional Matematika* 4.

Tantiati, Romlah, M. Tanzil Furqon, and Candra Dewi. 2019. *Implementasi Metode Learning Vector Quantization (LVQ) Untuk Klasifikasi Persalinan*. Vol. 3.

Wahyuni, Sri, Kana Saputra Saragih, and Mochammad Iswan Perangin-Angin. 2018. *IMPLEMENTASI METODE DECISION TREE C4.5 UNTUK MENGANALISA MAHASISWA DROPOUT IMPLEMENTATION OF DECISION TREE C4.5 METHOD TO ANALYZE THE DROPOUT STUDENTS*.

Wulan Purnami, Santi, Program Magister Departemen Statistika Fakultas Matematika, and Dan Sains Data. 2018. "ENSEMBLE SUPPORT VECTOR MACHINE DENGAN RANDOM UNDERSAMPLING PADA KLASIFIKASI DATA DNA MICROARRAY UNTUK MENANGANI KASUS MULTICLASS IMBALANCE."