CICES (*Cyberpreneurship Innovative and Creative Exact and Social Science*)
DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

Application of Naive Bayes Model, SVM and Deep Learning Predicting

Padeli[*1], Aris Martono[2], Sudaryono[3]
[1]Information Technology Education Study Program, [2]Informatics Engineering Study Program,
[3]Informatics Engineering Master Study Program, Faculty of Science and Technology, Raharja University
E-mail: [*1]**padeli@raharja.info**, [2] aris.martono@raharja.info, [3]sudaryono@raharja.info

Abstract

The college hopes that every semester students are able to pay tuition properly and smoothly. The hope is that the institution will be able to maintain monthly cash flow so that its operational and maintenance costs can be met. Therefore, this study was conducted to predict and fulfill the institution's cash-in from the method of paying tuition fees either by cash, installments, or sometimes late payments every semester. In predicting the method of paying tuition fees, using student profile data (name, name, study program) and achievement index every semester for 5 semesters passed and the method of payment (cash, installments, and late--cash or installments). Using the Naive Bayes (NB) method, Support Vector Machine (SVM), and Deep Learning, this study aims to forecast tuition costs. The Classification Prediction Model with Naive Bayes, SVM, and Deep Learning produces Confusion Matrix Performance NB with an Accuracy of 91.49%, Confusion Matrix Performance SVM with an Accuracy of 85.11%, and Confusion Matrix Performance Deep Learning with an Accuracy of 89.36%, according to the research findings.

Keywords—*Payments, Algorithm, Performance*

1. INTRODUCTION

There are several terms that we often use for the payment process, as follows:

1. Base money, commonly known by the public as building money or Development Development Contribution (SPP)
2. Tuition, the payment of tuition fees used to pay tuition fees. Usually this fee is based on a new student form.
3. This tuition fee payment is made every new semester before the lecture starts.
4. SKS fee The credit fee is usually calculated starting from each course taken by the student.
5. Laboratory costs are usually quite varied, depending on the needs of the laboratory used.

To predict the accuracy of student tuition payments based on these factors, a data mining is needed which is a strong and effective approach by providing a pattern discovery process on large datasets [1]. The method that performs data grouping with a partition system and modeling process without supervision is carried out by data mining [2]. This research was conducted using the classification method.

The data set used in this study totaled 146 students from the data processed in this study. The classification/taxonomy technique involves grouping certain items into a number of groups based on the idea of an item or piece of property [3]. With the use of pre-existing training data, classification creates a model, which is subsequently used to categorize fresh data. The definition of classification is the process of learning target functions that associate each characteristic or feature set with a certain set of class labels [4]. The four fundamental elements of the classification process are predictors, classes, dataset testing, and training sets.

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

The classification process in Naive Bayes is based on the conditional probability from one feature to feature selection using an existing algorithm [5]. It is a simple way to use. SVM is a technique that can create predictions and has superior performance in a variety of applications. SVM is also applicable to high-dimensional data, however it is challenging to utilize for huge data sets.

2. RESEARCH METHOD

In this study, the research method used is with three Naive Bayes models, Support Vector Machine (SVM) and Deep Learning. The literature review in this research is focused on the same research method using data mining, specifically using Naive Bayes, SVM, and Deep Learning which presents several student attributes that contribute to tuition payments based on research to select modeling variables.

2.1. Research Flow

To provide direction or reference in conducting this research, it is necessary to have a clear research flow. The research flow used in this study is illustrated in Figure 1. be optimized in implementing artificial intelligence machine systems that can detect and prevent messages cyber bullying online and filter it [7].

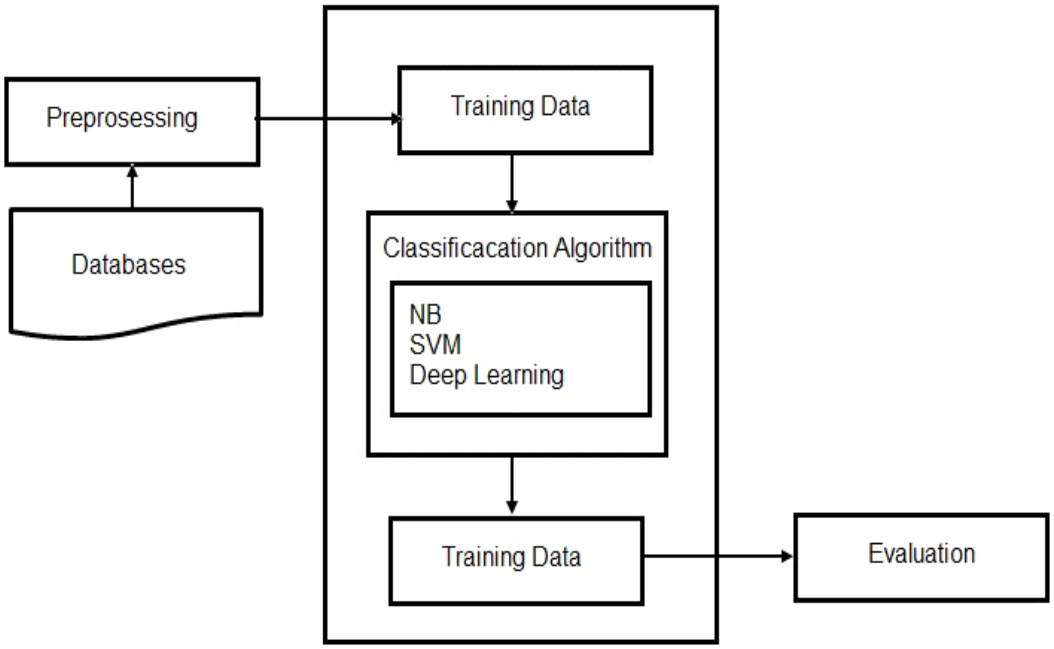

Figure 1. Description of the research flow

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

3. RESEARCH RESULTS AND DISCUSSION

3.1. Dataset

Data, often known as a dataset, is an entity that will be processed in data mining. A dataset is a collection of data, where one dataset corresponds to a database table, or it might be a data matrix where each row corresponds to the quantity of data and each column corresponds to a variable [2].

3.2. Preprocessing

Before starting the data classification stage in data mining, the Preprocessing stage is an important stage, especially if the data we have contains redundant data, noise, and other data problems, then this Preprocessing stage is very necessary before starting mining.

Preprocessing itself is a stage where the data will be filled in empty data,

removing duplicate data, examining data discrepancies, cleansing data, and fixing data mistakes Filling in blanks in data, getting rid of duplicate data, looking for inconsistent data, and fixing mistakes in data are all part of the cleaning process. Empty data is typically brought on by fresh data for which no information is available [12].

The dataset utilized in this study did not contain any missing values or empty data. In order to go on to the next classification step, the research might do so right now.

3.3. Naive Bayes

Based on Bayes' Theorem, Naive Bayes is a probability categorization. In order to make computations simpler, Naive Bayes makes the assumption that the impact on attribute values for certain classes is independent of the impact on other attribute values [1].

Bayesian probability formula:

$$P(Ci|X) = \frac{P(X|Ci)P(Ci)}{P(X)}$$

Information:

P(Ci|X)	: Probability of Ci with proof X P(Ci)
	: Probability of Ci
P(X|Ci)	: The probability that X has an effect on Ci
P(X)	: Probability X

Naïve Bayes is a basic probability classification based on Bayesain theory. The main advantage of this classification is to measure the average of several variables to be classified. Naïve Bayes technique for predicting performance in the field of data mining processing [6], Nave Bayes Classification can be optimized in implementing artificial intelligence machine systems that can detect and prevent messages cyber bullying online and filter it [7]

3.4. Support Vector Machine (SVM)

SVM is a classification algorithm that functions to classify non-linear data and linear data . The principle of SVM is to find a hyper plane that can classify data into 2 classes. The calculation formula is as follows:

Data point : xi = {x1,x2,….,xn} Rn

Data class : yi {-1,+1} Data and class pair :

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

$\{(x_i y_i)\}\ ^{N}_{i}= 1$
Maximize function:

$$Ld = \sum_{i=1}^{N} a_i - \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j y_i y_j\, k(x_i, x_j) syarat: 0 \leq\ a_i c \leq dan \sum_{i=1}^{N} a_i y_i = 0$$

Calculating the values of w and b:
$w = \sum_{i=1}^{N} a_i\ y_i\ x_i\ b = -\frac{1}{2}\ (w.x^{+} + \mathrm{w}.x^{-})$

Classification decision function sign(f(x)) :

$$\int(x) = w.x + b\ atau\ \int(x) = \sum_{i=1}^{m} a_i\ y_i$$

$\mathrm{K(x,}x_i\mathrm{)+b}$

Information :
N : number of data
n : many features/data dimensions
C : constant value
m : number of support vectors/data points that have i > 0
K(x,xi) : kernel function
Ld : duality Lagrange Multiplier
i : the weight value of each data point

The SVM algorithm is a classification algorithm in the machine learning approach to sentiment analysis. This algorithm belongs to the supervised learning algorithm which was first introduced by Cortes [8], the SVM and Naïve Bayes methods based on the level of classification accuracy of this method can assist in analyzing making the best decisions based on samples taken from several information or variables used [9].

3.5. Deep Learning

Deep Learning (DL) is one of the fields of Machine Learning (ML) to solve problems with large datasets that utilize artificial neural networks. DL is a branch of Supervised Learning. The DL model can be seen in Figure 2.

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

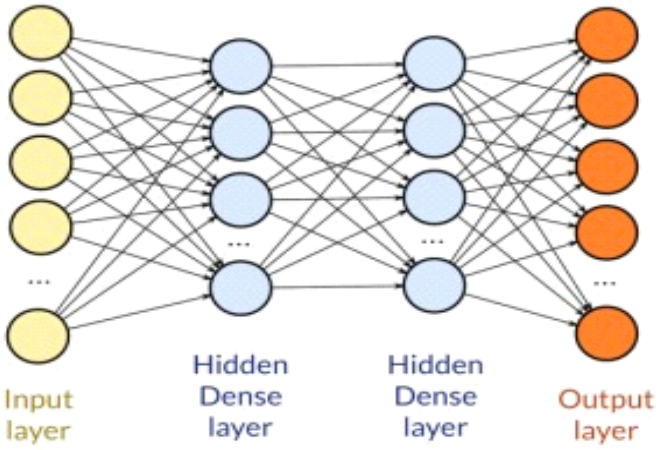

Figure 2. Deep Learning

In this study, a classification learning model is used to determine the accuracy with a confusion matrix and predictive measurements using the Naïve Bayes algorithm, Deep Learning, and SVM. The measurement results of these three algorithms with the highest accuracy value are selected. To produce good predictions, we have compressed the data modalities into one feature that represents information, previous work has found significant cross-correlation between different data types. . By using datasets to test and compare different combinations of modalities [10]. with deep learning, this model is highly recommended in developing a sophisticated deep learning recommendation model (DLRM) and providing its implementation in a research framework [11].

3.6. Results and Discussion

Solutions for analyzing data mining, text mining, and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques to provide users with insights so they can make the best decisions

Table 1. Payment data for semester 5 students

no	nim	name	rtbyrkuliah	ps	ipsm 1	ipsm 2	ipsm 3	ipsm 4	ipsm 5
1	1951427458	Adinda Zahrah Syahadat	Cash	AK	3.85	3.91	3.93	3.95	3.94
2	1951428471	Dini Febriyanti	Cash	AK	3.82	3.88	3.91	3.92	3.9
3	1951423870	Eritya Pingkan	Cash	AK	3.41	3.62	3.7	3.74	3.75
4	1951428304	Jihan Juliana	Cash	AK	3.89	3.93	3.95	3.95	3.94
5	1951426006	Kania Amara Brigita	Instalment	AK	0	1.64	2	0	2.79
6	1951426819	Margareta Mega	Instalment	AK	0	0	0	0	0
7	1951426794	Mochamad Prasetyo	Cash	AK	3.29	3.49	3.42	3.39	3.37
8	1951428402	Monica	Cash	AK	3.72	3.83	3.87	3.86	3.86
9	1951422921	Nabilah Fitri	Cash	AK	3.9	3.88	3.86	3.79	3.83
10	1951428453	Nesa Elvina	Cash	AK	3.63	3.78	3.75	3.74	3.7
11	1951425625	Nurfitri Hairunisa	Cash	AK	3.54	3.67	3.75	3.76	3.79
12	1941428651	Yosep	Instalment	MR	2.92	2.28	2.33	2.52	2.67

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

Attributes, namely nim, the average method of payment as a label is carried out with the attribute select operator and shared with the multiply operator of the three algorithms above. The classification model with naive Bayes algorithm, SVM, and Deep Learning is shown in Figure 3.

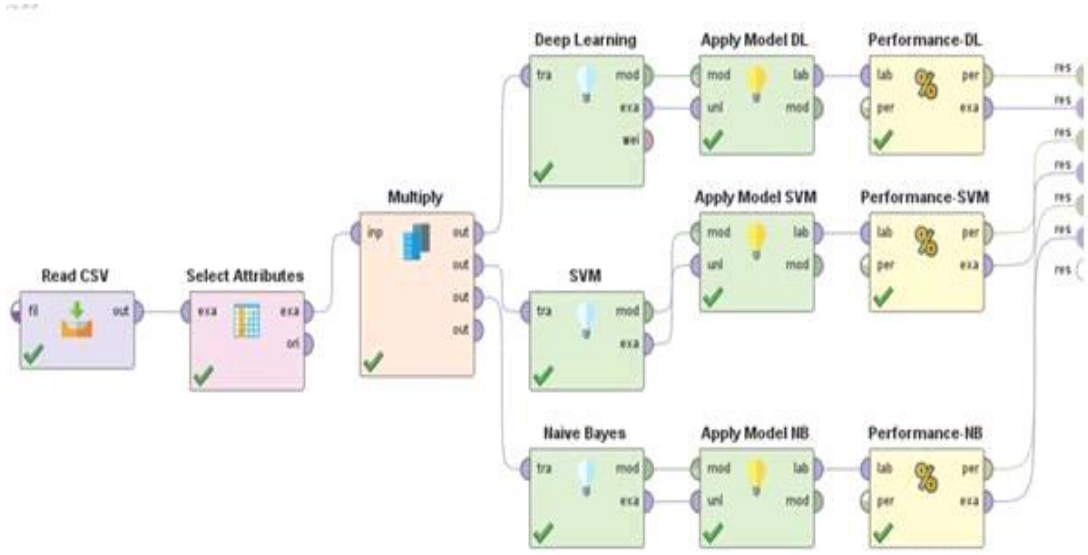

Figure 3. Classification Prediction Model with Naïve Bayes, Deep Learning and SVM

The results of the performance of the Naïve Bayes (NB) algorithm through training and testing are generated as shown in table 2 below.

Table 2. Confusion Matrix Performance NB with 91.49% Accuracy

Class recall	true cash	TRUE installment	True late	class precision
Pred. cash	28	2	0	93.33%
Pred. installment	0	13	0	100.00%
Pred. late	2	0	2	50.00%
Class recall	93.33%	86.57%	100.00%	

The results of the performance of the Deep Learning (DL) algorithm through training and testing are generated as shown in table 3 below.

Table 3. Confusion Matrix Performance DL with 89.36% Accuracy

	TRUE cash	TRUE installment	TRUE late	class precision
Pred. cash	30	3	2	85.71%
Pred. installment	0	12	0	100.00%
Pred. late	0	0	0	0.00%
Class recall	100%	80.00%	0.00%	

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

The results of the Support Vector Machine (SVM) algorithm performance through training and testing are generated as shown in table 4 below.

Table 4. Confusion Matrix Performance SVM with 85.11% Accuracy

	true cash	TRUE installment	True late	class precision
Pred. cash	30	5	2	81.08%
Pred. installment	0	10	0	100.00%
Pred. late	0	0	0	0.00%
Class recall	100.00%	66.67%	0.00%	

Which of the three approaches is more accurate is determined by the results of the tests that have been conducted, while The results of the tests that have been conducted compare which of the three approaches is more accurate, and table 5 below shows a performance comparison of each algorithm model:

Table 5. Results of Comparison of Accuracy Measurements of the Three Algorithms

Algorithms	Accuracy	Description
Naïve Bayes	91,49%	Best
Deep Learning	89,36%	Good
SVM	85,11%	Pretty good

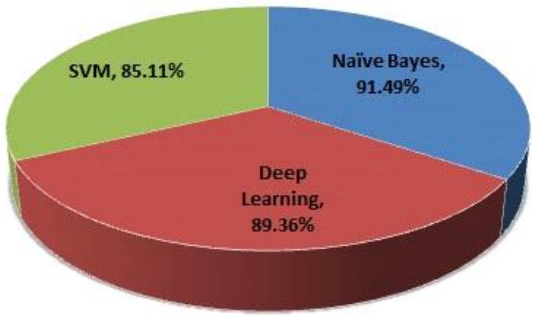

Figure 4. Graph of the results of the comparison of three Naive Bayes models, SVM and Deep Learning

4. CONCLUSION

This study finds three payment characteristics that have an impact on program completion, specifically (order of significance): From the three classification models mentioned above, it can be inferred that the Naive Bayes algorithm has the highest level of accuracy for the method dataset.

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

Based on the results of a comparison between data mining classification algorithms, namely Naïve Bayes, Deep Learning, and Support Vector Machines to predict smooth payments using multivariate data, namely student tuition payments for 5 semesters and this achievement index can be predicted well.

Learning model, the analysis results show that the Naïve Bayes method is an accurate algorithm that produces an accuracy of 91.49%, and this method is also very dominant compared to other methods. Based on the Accuracy value, these three methods are classified as very good in classification or prediction.

The results of this study have a big influence on forecasting how the following lecture process will go. Administrators at the university can help a student succeed in finishing college by using the payment process' predicted outcomes. For instance, in the payment attribute, the university can offer a scholarship program or any other assistance that may be available, such as the opportunity to earn extra money as a teaching assistant in the lab, assist in the library, or be recommended for a scholarship from the public or private sector.The campus party that has data on students who are at risk of not being able to make tuition payments but have a fairly good IPK, must receive immediate attention, students who have identified need assistance, such as guidance and counseling

5. SUGGESTION

The Naive Bayes model is very good for implementation in a computer-based prediction system for paying student tuition every semester to anticipate every payment for student tuition so that payments are made on time and campus cash flow is smooth.

REFERENCES

[1] Budiman and Z. Niqotaini, ―PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK Abstrak,‖ vol. 15, no. 3, pp. 1–5, 2021.

[2] N. Hidayati, A. I. Rizmayanti, C. B. S. Dewi, R. Fatmasari, and W. Gata, ―Penerapan Algoritma Klasterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka,‖ Swabumi, vol. 8, no. 2, pp. 134–142, 2020, doi: 10.31294/swabumi.v8i2.8385.

[3] M. Rahmadi, F. Kaurie, and T. Susanti, ―Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools,‖ JURIKOM (Jurnal Ris. Komputer), vol. 7, no. 1, p. 134, 2020, doi: 10.30865/jurikom.v7i1.1761.

[4] D. P. Utomo and M. Mesran, ―Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,‖ J. Media Inform. Budidarma, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.

[5] W. Zhang and F. Gao, ―An improvement to naive bayes for text classification,‖ Procedia Eng., vol. 15, pp. 2160–2164, 2011, doi: 10.1016/j.proeng.2011.08.404.

[6] J. Galopo Perez and E. S. Perez, ―Predicting Student Program Completion Using Naïve Bayes Classification Algorithm,‖ Int. J. Mod. Educ. Comput. Sci., vol. 13, no. 3, pp. 57–67, 2021, doi: 10.5815/ijmecs.2021.03.05.

DOI: 10.33050/cices.v9i1.2584 ISSN: **2356-5209**
pp. 93~101 *Online* ISSN: **2655-3058**

[7] T. Ige and S. Adewale, ―AI Powered Anti-Cyber Bullying System using Machine Learning Algorithm of Multinomial Naïve Bayes and Optimized Linear Support Vector Machine Interception of Cyberbully Contents in a Messaging System by Machine Learning Algorithm,‖ IJACSA) Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 5, p. 2022, 2022, [Online]. Available: www.ijacsa.thesai.org

[8] M. Resa, A. Yudianto, A. Rahim, P. Sukmasetya, and R. A. Hasani, ―PERBANDINGAN METODE SUPPORT VECTOR MACHINE DENGAN METODE LEXICON DALAM ANALISIS SENTIMEN BAHASA,‖ vol. 6, no. 1, 2022.

[9] R. A. Ariyanto and N. Chamidah, ―Sentiment Analysis for Zoning System Admission Policy Using Support Vector Machine and Naive Bayes Methods,‖ J. Phys. Conf. Ser., vol. 1776, no. 1, pp. 0–9, 2021, doi: 10.1088/1742-6596/1776/1/012058.

[10] A. Cheerla and O. Gevaert, ―Deep learning with multimodal representation for pancancer prognosis prediction,‖ Bioinformatics, vol. 35, no. 14, pp. i446–i454, 2019, doi: 10.1093/bioinformatics/btz342.

[11] M. Naumov et al., ―Deep Learning Recommendation Model for Personalization and Recommendation Systems,‖ 2019, [Online]. Available: http://arxiv.org/abs/1906.00091

[12] Y. Pristyanto, ―Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset,‖ J. Teknoinfo, vol. 13, no. 1, p. 11, 2019, doi: 10.33365/jti.v13i1.184

[13] SigitAdinugroho, Y. A. S, "Implementasi Data Mining Menggunakan WEKA," UniversitasBrawijaya Press, Malang, 2018.

[14] Hendrian, S., "AlgoritmaKlasifikasi Data Mining UntukMemprediksi," 11(3), 266–274, 2018.

[15] Nengsih, W., "Implementasi Data Mining DalamKasusTerapan," Politeknik Caltex, Riau, 2019

[16] Werdiningsih, I., Nuqoba, B., &Muhammadun, Data Mining Menggunakan Android, Weka Dan SPSS, UNAIR, JawaTimur, 2020.

[17] Kusuma, & Purba, D., Machine Learning Teori, Program Dan StudiKasus, Deepublish, Yogyakarta, 2020.