Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Improving the Education Development Contribution Payment Model at SMK Istiqomah Maruyung Using the C4.5 Algorithm

Noviyanti[1]*, Ade Irma Purnamasari[2], Agus Bahtiar[3], Edi Tohidi[4]

[1]*Informatics Engineering,* STMIK IKMI Cirebon, Indonesia
[2]*Information System,* STMIK KMI Cirebon, Indonesia
[3]*Computerized Accounting,* STMIK IKMI Cirebon, Indonesia

novi19341@gmail.com[1*], irma2974@yahool.com[2], agusbahtir038@gmail.com[3], editohidi00@gmail.com[4]

Abstract

Payment of tuition fees is one of the important aspects of school financial management. At SMK Istiqomah Maruyung, the management of SPP payments is still done manually, which causes student non-compliance in paying on time. The purpose of the research is to improve the SPP payment model by using the C4.5 algorithm to classify the level of student compliance and identify the main factors that influence late payments. The method used is the Knowledge Discovery in Databases (KDD) approach which includes the stages of data selection, preprocessing, transformation, data mining, and result evaluation. The research data was taken from 206 students in the 2023/2024 academic year with attributes such as parental income, number of siblings, scholarship status, and academic grade point average. The C4.5 algorithm was applied to build a decision tree model, with evaluation using five-fold cross validation. The result of this study is that the C4.5 algorithm is able to classify student compliance levels with an average accuracy of 93.55%. The main factors that influence late payment are academic grade point average, class, and parental income. Although the model is very good at predicting compliant students (precision 95%, recall 98%), it shows weakness in predicting lateness (precision 67%, recall 40%). It is concluded that the C4.5 algorithm can improve the efficiency of managing tuition payments and provide data-driven insights for policy making. With further implementation, this algorithm is expected to be adopted by other educational institutions to address similar challenges in financial management.

Keywords: *Data mining, student compliance, tuition payment, c4.5 algorithm.*

1. Introduction

Payment of tuition fees is an important aspect of school financial management that affects the smooth running of educational operations. However, manual management of SPP payments often leads to problems such as late payments by students, as is the case at SMK Istiqomah Maruyung. This problem requires a technology-based solution to improve payment compliance and efficiency of school financial management.

Previous research has addressed relevant topics in the context of education payment management. Using the C4.5 algorithm to predict late payment of tuition fees among students [1]. While evaluating late payments with the Support Vector Machine (SVM) algorithm [2]. Evaluating the performance of the C4.5 algorithm in various data classification scenarios [3], while applying the C4.5 algorithm to determine direct cash assistance recipients [4]. Research using the C4.5 algorithm to predict student resignation [5],and utilizing data mining to analyze late work attendance [6]. In addition, using K-Nearest Neighbor to predict late tuition payments [7], and comparing Naive Bayes and SVM in the classification of late tuition payments [8]. Research focuses on optimizing web-based tuition payment system [9], while comparing the C4.5 algorithm and K-Nearest Neighbors in determining scholarship recipients [10].

The difference between this research and previous research is the focus on the application of the C4.5 algorithm to improve the SPP payment model specifically at SMK Istiqomah Maruyung, using more diverse attributes such as academic grade point average and scholarship status, while previous research tends to focus only on predicting late payments without utilizing a more comprehensive analysis framework.The novelty of this research is the application of the C4.5 algorithm with a Knowledge Discovery in Databases (KDD) approach that includes analyzing key factors such as parental income, attendance, and academic grades, to identify patterns of late payments and design a more efficient tuition management model.The purpose of this research is to improve the tuition payment model using the C4.5 algorithm, as well as identify the main factors that influence late payments, so that it can assist schools in designing data-driven policies to improve student compliance.

2. Literature Review

Manually handled tuition payments often lead to delays and poor student compliance. The C4.5 algorithm has demonstrated its efficacy in examining payment patterns and determining important elements including academic performance, scholarship status, and parental income [1], [3]. This method is suitable for handling compliance issues as it creates a decision tree that offers proper classification.

found that the C4.5 algorithm is more accurate than Naive Bayes and Support Vector Machine (SVM) in predicting late payments [8], while successfully using this algorithm to classify socio-economic data Juwita et al. (demonstrated the algorithm's dependability for educational data analysis by showing its superiority over K-Nearest Neighbor (KNN) in the classification of scholarship recipients [10].

Systematic data processing is supported by the Knowledge Discovery in Databases (KDD) framework, which consists of data selection, preprocessing, transformation, mining, and evaluation. Highlighting how digital payment methods can improve tuition fee administration [9]. Demonstrates the role of data mining in analyzing behavioral patterns, including late payments [6]. This research adopts the C4.5 algorithm in a KDD framework to develop an effective model to improve student payment compliance and provide data-driven insights to educational institutions.

3. Research Methods

The Knowledge Discovery in Databases (KDD) process is used to extract important information from SPP data through the stages of selection, processing, transformation, data mining, and evaluation. This method ensures high-quality data to support the development of data-based strategies in administrative work.

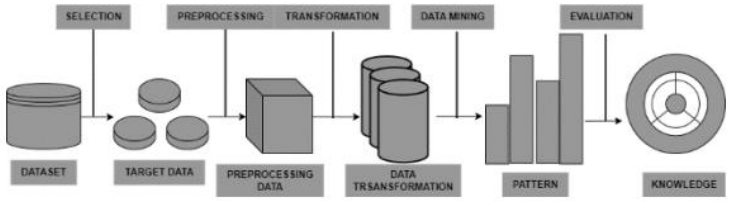

Fig. 1*:* Stages *Knowledge Discovery in Databases (KDD)*[11]

Table 1: Activity Description Research Methods

Stages	Activities	Activity Description
Selection	Data collection.	To improve the prediction of compliance with the payment of coaching donations at SMK Istiqomah Maruyung.
Preprocessing	Data Clealization ning and Norma	Removing irrelevant values, resolving missing data, and normalizing the range of values for the C4.5 algorithm.
Transformation	Converting Data into a Processable Format	Transforming value data into a format suitable for inclusion in the C4.5 algorithm.
Data mining	Application of the C4.5 Algorithm	Use the C4.5 algorithm to classify student compliance in paying Education Development Contributions.
Evaluation	Analysis of Clustering Results	Analyze the clustering results to assess whether the resulting groups are in accordance with the research objectives.

3.1. Data Source,Population and Sample

This research uses primary data obtained directly from SMK Istiqomah Maruyung. This data includes information relevant to the payment of Education Development Contribution (SPP), such as student data, payment history, late payment details, and other related attributes. The data collection process was conducted through direct cooperation with the school, ensuring that the data obtained was accurate, specific, and relevant to the research objectives.

The population of this study includes all student tuition payment data of SMK Istiqomah Maruyung during the 2021 to 2024 academic years, with a total of 206 records. The data consists of 15 attributes, namely: number, name, gender, major, parental income, parental status, number of siblings, payment status, number of tuition arrears, number of student arrears, scholarship status, student attendance, vehicle used, and academic grade point average. These attributes were analyzed to understand payment patterns and factors that influence student delay or compliance in tuition payments.

This population selection was based on the coverage of student payment behavior over a three-year period. With this data, this research can produce a comprehensive and relevant analysis to support the purpose of developing a C4.5 algorithm-based tuition payment model.

3.2. Data Collection Technique

The collection prepare started with the recovery of educational cost installment information from computerized authoritative records that included understudy personality, installment date, delay, and installment strategy. Information was gotten to with authorization from the school to guarantee lawfulness and legitimacy. Following, the information is analyzed independently to get it understudy installment behavior, at that point amassed within the frame of a dataset for encourage examination utilizing the C4.5 calculation. The crude information gotten was handled through a preprocessing prepare, which included erasing unimportant information, filling in spaces, and normalization to guarantee consistency. Information approval was done by cross-checking between authoritative records and data from the

school to guarantee accuracy before being utilized within the examination. This procedure permitted the inquire about to get a clear picture of the students' educational cost installment compliance designs that will be analyzed advance.

3.3. Data Analysis Technique

The information investigation strategy in this ponder employments the C4.5 calculation to construct a choice tree. This calculation distinguishes components that impact understudy educational cost installment compliance and predicts future installment behavior.
Steps of Analysis:

3.3.1. Entropy and Gain Information calculation

Measuring data uncertainty with:

$$E(S) = -\sum_{i=1}^{n} P(i) \log_2 P(i) \quad (1)$$

E(S)= Entropy of the dataset S,P(i) = Probability of category to-i,n = Number of categories.

3.3.2. Getting Information

Measures the decrease in entropy after data division:

$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_V|}{|S|} E(S_v) \quad (2)$$

Sv = Data subset based on attribute value A.

3.3.3. Gain Ratio

Avoid bias of attributes with many unique values:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$
$$SplitInfo(S, A) = -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|} \quad (3)$$

3.3.4. Decision Tree Creation

The information investigation strategy in this ponder employments the C4.5 calculation to construct a choice tree. This calculation distinguishes components that impact understudy educational cost installment compliance and predicts future installment behavior.

4. Results and Discussion

In this study, the classification process was conducted in multiple steps. 206 records with 15 attributes pertaining to tuition payments made by students at SMK Istiqomah Maruyung in 2021–2024 were gathered and put through a preprocessing step that included data cleansing, handling empty values, and converting categorical data into numerical form. In order to identify the attribute that had the most influence on the classification, the C4.5 algorithm was also used in conjunction with entropy calculations to quantify data uncertainty and information gain. Using the highest gain characteristic as the root node, a decision tree was constructed, branching according to other qualities to create a categorization model. A decision tree and classification rules are used to illustrate the categorization results. For instance, a student is classified as "non-compliant" if their tuition arrears exceed three months and they are not eligible for a scholarship, and as "compliant" if their arrears are less than or equal to one month or if they are eligible for a scholarship. The model's correctness was verified through test data, which produced a picture of tuition payment trends that can aid in the school's decision-making.

4.1 Research Results

4.1.1 Data Selection

This think about employments a dataset from SMK Istiqomah Maruyung that incorporates 206 factors related to understudy educational cost installment compliance. The most factors incorporate sexual orientation, course, major, parental salary, family status, number of kin, educational cost expense sum, grant, participation, and normal scholastic score. The center of the examination is on the installment status

variable (1 = on time, = late) to recognize components that impact installment compliance. This dataset incorporates financial, grant, and participation data, giving setting in understanding educational cost installment consistency.

	NO	NAME	G	CLASS	MAJOR	INCOME CATEGORIES	PARENTS' INCOME	ONE PARENT	NUMBER OF SIBLINGS	PAYMENT STATUS	SPP AMOUNT	STUDENT ARRESTS
count	206.000000	206	206	206	206	207.0	206	206	206.000000	206	206.0	206
unique	NaN	206	2	7	1	4.0	9	7	NaN	1	NaN	1
top	NaN	ABDUL HAKIM FATAHILLAH	P	XI F A	PHARMACY	1500000.0	Rp. 1,000,000 - Rp. 1,999,999	Laborer	NaN	On time	NaN	NO
freq	NaN	1	137	36	206	136.0	124	145	NaN	206	NaN	206
mean	103.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.131068	NaN	150000.0	NaN
std	59.611241	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.631192	NaN	0.0	NaN
min	1.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	NaN	150000.0	NaN
25%	52.250000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.000000	NaN	150000.0	NaN
50%	103.500000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.000000	NaN	150000.0	NaN
75%	154.750000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	3.000000	NaN	150000.0	NaN
max	206.000000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	18.000000	NaN	150000.0	NaN

Fig. 2: Results Descriptive statistics

4.1.2 Initial Processing Data

This ponder employments a dataset from SMK Istiqomah Maruyung that incorporates 206 factors related to understudy educational cost installment compliance. The most factors incorporate sex, lesson, major, parental salary, family status, number of kin, educational cost charge sum, grant, participation, and normal scholarly score. The center of the examination is on the installment status variable (1 = on time, = late) to recognize variables that impact installment compliance. This dataset incorporates financial, grant, and participation data, giving setting in understanding educational cost installment consistency.

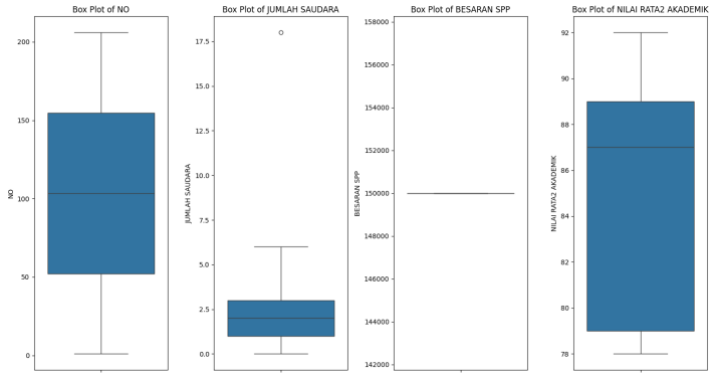

Fig. 3: Box plot results checking for outliers

4.1.3 Data Transformation

The Information Change arrange points to alter the scale of the information to fit the modeling needs, particularly in calculations that are touchy to the scale of highlights. the taking after are the stages of Information Change.

```
Cross-Validation Scores: [0.88095238 0.90243902 0.92682927 0.85365854 0.85365854]
Mean Accuracy: 0.8833075493612079
Standard Deviation of Accuracy: 0.028367686162022357
```

Fig. 4: Training results and model evaluation

4.1.4 Data Mining

Information Mining is the method of extricating information or data from huge unstructured information sets with the point of finding useful patterns or connections. Within the setting of KDD, the information mining handle includes the utilize of measurable methods, calculations, and machine learning to extricate important data from a arranged dataset.

```
['DATA SHEET SMK ISTIQOMAH MARUYUNG.xlsx', 'dataset',
'decision_tree_model.joblib', 'model_random_forest.joblib', 'spp-
c45.ipynb']
Model loaded successfully
Hasil Prediksi:
1. Tepat Waktu
2. Tepat Waktu
```

Fig. 5: Prediction Model Results

4.1.5 Evaluation

Within the KDD prepare, assessment is an critical step that points to decide how valuable and important the models produced from the analyzed information are. At this organize, different assessment methods, such as cross-validation or partitioning information into preparing and testing information, are utilized to guarantee that the information found isn't as it were precise but moreover pertinent to unused information.

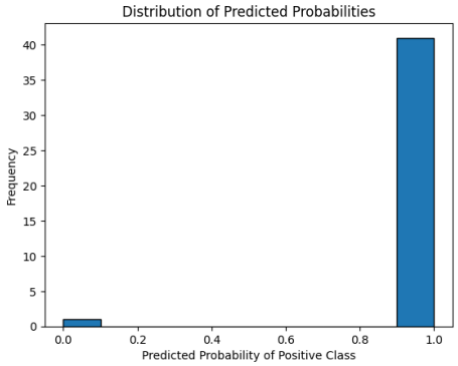

Fig. 6: Probability prediction results

4.2 Discussion

a) Knowing the performance of the C4.5 algorithm in classifying the level of student compliance with the payment of Education Development Contribution at SMK Istiqomah Maruyung.
The C4.5 calculation was utilized to classify the level of understudy compliance with SPP installments at SMK Istiqomah Maruyung based on properties such as parental salary, sexual orientation, SPP sum, grants, and understudy participation. The comes about appear 93.55curacy with cross-validation, as well as tall execution for the compliant understudy category (exactness 95% and review 98%). This calculation is demonstrated to be successful in distinguishing designs of understudy compliance with educational cost installments.

b) The results of analyzing the classification model for the Education Development Contribution payment using the C4.5 algorithm at SMK Istiqomah Maruyung.
The application of the C4.5 calculation makes a difference schools distinguish the most components that impact educational cost installment compliance, such as parental salary, charge sum, and understudy participation. The comes about of the choice tree examination empower expectation of the hazard of late installment so that schools can take early activity, such as giving updates or motivations. This show makes strides the effectiveness of installment administration, diminishes blunders, and gives more noteworthy straightforwardness for guardians, whereas giving a premise for future optimization of technology-based frameworks.

5. Conclusion

This research shows that the C4.5 algorithm successfully improves the model of Education Development Contribution (SPP) payment at SMK Istiqomah Maruyung. Using the Knowledge Discovery in Databases (KDD) approach, this research was able to identify patterns of late payments and factors that influence student compliance, such as academic grade point average, class, and parental income. The resulting model has an accuracy of 93.55%, with excellent performance in predicting students who pay on time.However, this study also revealed the weakness of the C4.5 algorithm in predicting late-paying students, with a precision of 67% and a recall of 40%. These findings have important implications for schools in designing data-driven policies, such as payment reminders or incentives for compliant students. This research is expected to be a reference for further development in the field of education and information technology.

6. Suggestions

Suggestions that can be given for further research are as follows:

1. Schools are advised to integrate the C4.5 algorithm into a digital-based payment management system so that analysis and prediction of late payments can be done automatically and continuously.
2. To improve the accuracy of predicting tardiness in certain categories of students, schools can consider the use of additional algorithms, such as Random Forest or Support Vector Machine, to compare model performance and optimize results.

References

[1] B. Angkoso and I. Irmayansyah, “Penerapan Algoritma C4.5 Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) Santri,” *TeknoIS J. Ilm. Teknol. Inf. dan Sains*, vol. 13, no. 1, pp. 13–23, 2023, doi: 10.36350/jbs.v13i1.166.

[2] L. Rosiana, “Analisis Kemungkinan Keterlambatan Pembayaran SPP Menggunakan Algoritma Support Vector Machine(Studi Kasus Smp Perintis 2 Bandar Lampung),” *Ilmu Data*, vol. 2, no. 9, pp. 1–11, 2022, [Online]. Available: http://repository.teknokrat.ac.id/2837/1/b116311150.pdf

[3] Z. Gustiana, “Performance Evaluation Algoritma C 4.5 Pada Klasifikasi Data,” *Djtechno J. Teknol. Inf.*, vol. 5, no. 2, pp. 289–296, 2024, doi: 10.46576/djtechno.v5i2.4654.

[4] S. Juwita, R. Muliono, and mutahir, “Implementasi Algoritma C4 . 5 Untuk Klasifikasi Penentuan Penerima Bantuan Langsung Tunai Di Desa

Tanjung Rejo The Implementation of the C4 . 5 Algorithm for the Determination Classification of Direct Cash Assistance (BLT) Recipients in Tanjung Rejo Vil," *Jitek:Jurnal Ilmiah Teknik Informatika Dan Elektro*, vol. 1, no. 2. pp. 92–95, 2022. doi: 10.31289/jitek.v1i2.1474.

[5] Y. Steven, Arif Bijaksana Putra Negaraa, "Implementasi Algoritma K-Nearest Neighbor Untuk Mengklasifikasi Masa Studi Mahasiswa Informatika Universitas Tanjungpura," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 311, 2022, doi: 10.26418/justin.v10i3.56804.

[6] S. S. Bahri, "Implementasi Data Mining Untuk Memprediksi Keterlambatan Jam Masuk Kerja Menggunakan Algoritma Klasifikasi," *J. Sist. Inf.*, vol. 1, no. 1, pp. 11–20, 2020, doi: 10.32546/jusin.v1i1.854.

[7] M. R. Akhmad and T. A. Y. Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 185, 2022, doi: 10.35889/progresif.v18i2.921.

[8] H. Umar, R. Kusumawati, M. Imamudin, and M. A. Rohman, "Klasifikasi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *TIN Terap. Inform. Nusant.*, vol. 4, no. 11, pp. 709–718, 2024, doi: 10.47065/tin.v4i11.4969.

[9] P. devi Yuliana, "Optimalisasi Pembayaran Spp Berbasis Web Dengan Framework Codeigniter Pada Smk Terpadu Bina Insan Mandiri Optimization of Web-Based Spp Payments With Codeigniter Framework At Bina Insan Mandiri Integrated Vocational School," *J. Inf. Technol. Comput. Sci.*, vol. 7, no. 4, pp. 1315–1322, 2024.

[10] M. I. Siti Chodijah, "PENENTUAN PENERIMA BEASISWA DI STIK BANI SALEH DENGAN PERBANDINGAN METODE ALGORITMA C4.5 DAN KNEAREST NEIGHBORS," *J. Inf. dan Komput.*, vol. 10, no. 1, p. 15, 2022.

[11] K. Gustipartsani, N. Rahaningsih, R. Danar Dana, and I. Yulia Mustafa, "Data Mining Clustering Menggunakan Algoritma K-Means Pada Data Kunjungan Wisatawan Di Kabupaten Karawang," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3595–3601, 2024, doi: 10.36040/jati.v7i6.8282.