BIOS: Jurnal Informatika dan Sains
Volume 2 Number 02, 2024, DOI 10.54209/bios.v1i02
E.SSN 2988-6910(Online)
https://seaninstitute.or.id/bersinar/index.php/bios/index

Analysis Of Indonesian Daily Test Results Using the C4.5 Decision Tree Algorithm

Bambang Sugito[1], Nelviony Parhusip[2], Ibrahim[3], Maida Indrayani[4]
[1,2,3,4]Magister Teknologi Informasi, Universitas Pembangunan Pancabudi, Medan, Indonesia

Article Info

Corresponding Author:
Bambang Sugito
E-mail:
Bambang.sugito@pertamina.com

ABSTRACT

This study aims to analyse the results of Indonesian daily tests using the Decision Tree C4.5 algorithm in improving the effectiveness of evaluating student learning outcomes. The data used is the daily test results of students of class X SMK Panca Budi in the academic year 2023/2024. The research process involves the stages of data collection, preprocessing to clean the data, and modelling using RapidMiner software. Decision Tree C4.5 algorithm was used to predict grade improvement based on students' STATUS score. The results showed that the higher the STATUS score, the greater the increase in grades obtained by students. This decision tree model can be used to evaluate student progress and provide clearer insights for the development of technology-based learning methods. This research contributes to the utilisation of data mining technology to improve educational evaluation, although the results are limited to a sample of grade X students at SMK Panca Budi. Further research with larger samples and additional variables is expected to provide more representative results.

Keywords:
Decision Tree, C4.5, Data Mining, Learning Evaluation, RapidMiner, Grade Improvement.

INTRODUCTION

Technology 5.0 has brought major changes in various sectors, including education, by connecting the physical and digital worlds more integratively and effectively (Fricticarani et al., 2023) (Sadriani et al., 2023) (Sundari, 2024). Global demands encourage the world of education to utilise technology in improving the quality of learning (Belva Saskia Permana et al., 2024). One approach that supports this is data mining, the process of analysing data to generate valuable insights that support decision making (Jhody, 2024) (Pradiah et al., 2024).

Data mining becomes relevant to overcome challenges in the evaluation of learning outcomes such as daily tests. Daily tests serve to monitor student progress, provide feedback, and improve the effectiveness of the learning process (Sari Ndune & Jahja, 2024). On the other hand, Indonesian language education has an important role in shaping students' character during adolescence, which is a crucial period in the formation of their personality (Tri Diantami et al., 2023). In addition, technology-based learning outcomes data management is now an important aspect in supporting the efficiency of educational

evaluation (Ramdhani & Setiawan, 2024) (Istiqomah et al., 2024). Decision Tree C4.5 algorithm is one of the data mining methods chosen in this study because of its ability to visualise decision making in a structured and easy to understand manner (Fernanda et al., 2023) (Triawan et al., 2024).

This research aims to analyse and compare Indonesian daily test results using this algorithm. With a data-driven approach, it is expected that the evaluation of learning outcomes will be more effective, thus supporting the development of technology-based learning methods and contributing to improving the quality of education (Ramdhani & Setiawan, 2024).

RESEARCH METHODS

Data collection, Data collection techniques are methods used to obtain data needed in research, both from subjects and samples (Naqibuzzahidin & Fatah, 2024a). The data used is class X Indonesian assessment data at Panca Budi school in the academic year 2023/2024.

Preprocessing, preprocessing is a stage to prepare a dataset by cleaning and organising a data content from a raw dataset into a clean and organised dataset (Parsakh Nursyamsyi & Noor Hasan, 2023). The preprocessing stage involves removing data containing zero or empty values and restructuring the data through data cleaning and transformation. This step aims to ensure that the data to be processed is more organised and facilitates the modelling process (Putera, Siahaan, Jabar, Parhusip, et al., 2024).

Modeling and analysis, The processed data is then modelled using RapidMiner, and the modelling results are then analysed (Putera, Siahaan, Jabar, Pranoto, et al., 2024). RapidMiner is software designed with an intuitive visual interface, allowing users to implement machine learning algorithms without the need to have deep programming skills. Using RapidMiner, data analysis becomes easier as the software not only facilitates data processing and visualisation, but also allows users to experiment with various machine learning techniques efficiently (Naqibuzzahidin & Fatah, 2024).

RESULTS AND DISCUSSION

Data collection

	A	B	C	D	E	F	G	H
1	NIS	NAMA LENGKAP	KELAS	UH1	UH2	PTS	PAS	NA
2	23.01.033	ADITYA DWI GUNA	X TKR	80	90	80	90	85
3	23.01.035	Arga Ardiansyah	X TKR	85	91	85	91	88
4	23.01.037	DAFA AL HILAL	X TKR	80	90	80	90	85
5	23.01.038	DICKY RAMADHAN	X TKR	82	92	82	92	87
6	23.01.039	FADHIL SEPTIAN	X TKR	80	90	80	90	85
7	23.01.040	FAIZ NAUFAL MUZAKKI	X TKR	87	93	87	93	90
8	23.01.041	HILAL ABIYU LUBIS	X TKR	80	90	80	90	85
9	23.01.042	M.FAREL LUBIS	X TKR	80	90	80	90	85
10	23.01.043	MHD FACHROZI	X TKR	80	90	80	90	85
11	23.01.044	MUHAMMAD ARIS RIZKY	X TKR	80	90	80	90	85
12	23.01.045	MUHAMMAD DAFFA FIRJATULLAH	X TKR	80	94	80	94	87
13	23.01.046	MUHAMMAD HAIKAL	X TKR	80	90	80	90	85
14	23.01.047	MUHAMMAD RINALDI	X TKR	80	90	80	90	85
15	23.01.048	MUHAMMAD RIZKY	X TKR	80	94	80	94	87
16	23.01.049	MUHAMMAD SARRIZKI UTAMA	X TKR	80	90	80	90	85
17	23.01.051	QOLBY DANU TIRTA	X TKR	80	94	80	94	87
18	23.01.052	RAFAEL JOSI KIBAR PRATAMA	X TKR	80	90	80	90	85
19	23.01.053	REZI NAUFAL DZAKI	X TKR	80	90	80	90	85
20	23.01.054	RIZKY ANANDA	X TKR	83	93	83	93	88
21	23.01.055	TARUNA PRADITYO SETIAWAN	X TKR	80	90	80	90	85
22	23.01.056	UTOMO REZA AHMAD	X TKR	87	95	87	95	91
23	23.01.057	Vino Dwi Nugraha	X TKR	80	90	80	90	85
24	23.01.058	WILPANDI TAMBA	X TKR	87	93	87	93	90
25	23.01.059	YUDHA PRATAMA	X TKR	88	92	88	92	90
26	23.01.060	YUSUF ADITYA WARDANA	X TKR	80	90	80	90	85
27	23.01.061	ANGGA ISLAMI PUTRA	X TKR	85	95	85	95	90

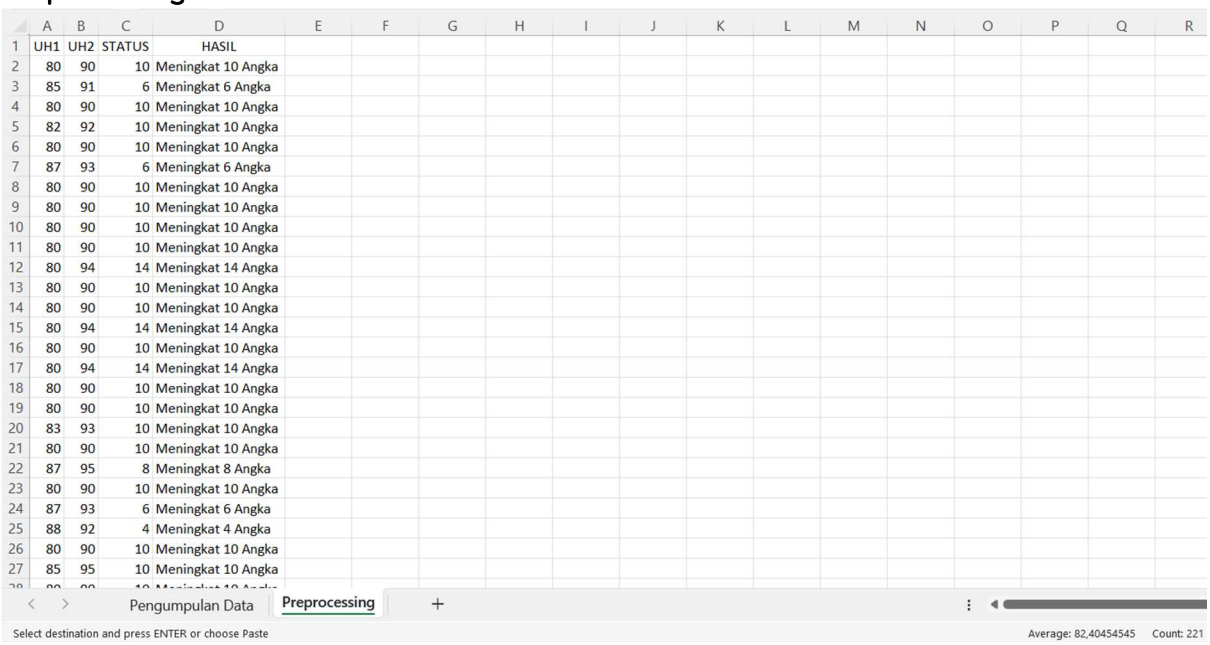
Pengumpulan Data | Preprocessing | +

Ready | Accessibility: Good to go | Count: 221

Figure 1, Data Collection

Based on Figure 1, the data consists of 8 columns, namely (NIS, FULL NAME, CLASS, UH1, UH2, PTS, PAS, NA).

Preprocessing



	A	B	C	D
1	UH1	UH2	STATUS	HASIL
2	80	90	10	Meningkat 10 Angka
3	85	91	6	Meningkat 6 Angka
4	80	90	10	Meningkat 10 Angka
5	82	92	10	Meningkat 10 Angka
6	80	90	10	Meningkat 10 Angka
7	87	93	6	Meningkat 6 Angka
8	80	90	10	Meningkat 10 Angka
9	80	90	10	Meningkat 10 Angka
10	80	90	10	Meningkat 10 Angka
11	80	90	10	Meningkat 10 Angka
12	80	94	14	Meningkat 14 Angka
13	80	90	10	Meningkat 10 Angka
14	80	90	10	Meningkat 10 Angka
15	80	94	14	Meningkat 14 Angka
16	80	90	10	Meningkat 10 Angka
17	80	94	14	Meningkat 14 Angka
18	80	90	10	Meningkat 10 Angka
19	80	90	10	Meningkat 10 Angka
20	83	93	10	Meningkat 10 Angka
21	80	90	10	Meningkat 10 Angka
22	87	95	8	Meningkat 8 Angka
23	80	90	10	Meningkat 10 Angka
24	87	93	6	Meningkat 6 Angka
25	88	92	4	Meningkat 4 Angka
26	80	90	10	Meningkat 10 Angka
27	85	95	10	Meningkat 10 Angka

Figure 2, Preprocessing

In Figure 2, the data has gone through preprocessing so that it only produces 4 columns, namely (UH1, UH2, STATUS, RESULTS). UH1 is the first daily test, UH2 is the second daily test.

Modeling and Analysis

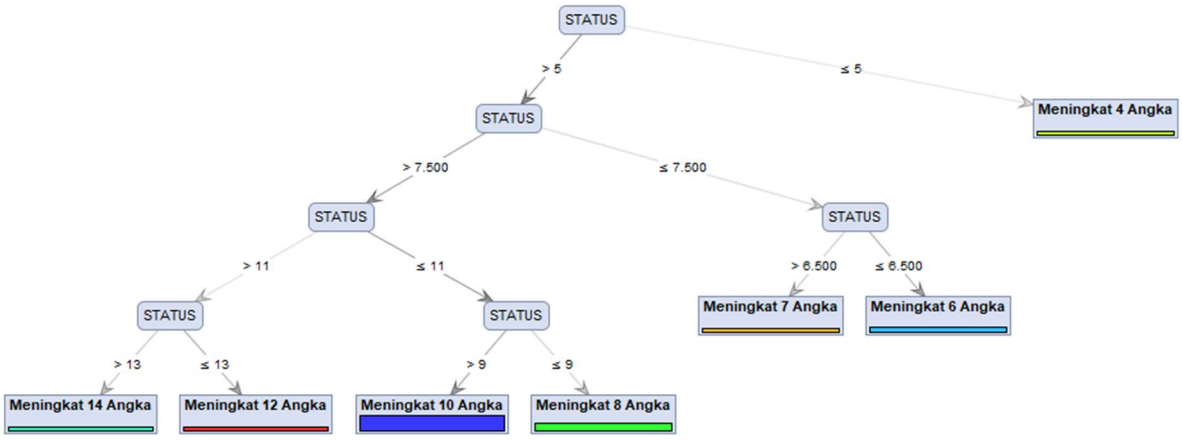


Figure 3, Decision Tree Model

In Figure 3, this decision tree model is used to determine the increase in daily test scores called status. If the status is more than 5, it will be examined further: if the status is more than 7,500 and more than 11, then if the status is more than 13, the grade increase is 14 numbers, while if the status is 13 or less, the grade increase is 12 numbers. If the status is 11 or less, and more than 9, the increase in value is 10 numbers, while if the status is 9 or less, the increase in value is 8 numbers. If the status is 7,500 or less, but more than 6,500, the grade

increase is 7 points, and if the status is 6,500 or less, the grade increase is 6 points. Finally, if the status is 5 or less, the grade increase is 4 points.

Tree

```
STATUS > 5
|   STATUS > 7.500
|   |   STATUS > 11
|   |   |   STATUS > 13: Meningkat 14 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=0, Meningkat 14 Angka=6, Meningkat 8 Angka=0, Meningkat 4 Angka=0, Meningkat 7 Angka=0, Meningkat 12 Angka=0}
|   |   |   STATUS ≤ 13: Meningkat 12 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=0, Meningkat 14 Angka=0, Meningkat 8 Angka=0, Meningkat 4 Angka=0, Meningkat 7 Angka=0, Meningkat 12 Angka=11}
|   |   STATUS ≤ 11
|   |   |   STATUS > 9: Meningkat 10 Angka {Meningkat 10 Angka=123, Meningkat 6 Angka=0, Meningkat 14 Angka=0, Meningkat 8 Angka=0, Meningkat 4 Angka=0, Meningkat 7 Angka=0, Meningkat 12 Angka=0}
|   |   |   STATUS ≤ 9: Meningkat 8 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=0, Meningkat 14 Angka=0, Meningkat 8 Angka=41, Meningkat 4 Angka=0, Meningkat 7 Angka=0, Meningkat 12 Angka=0}
|   STATUS ≤ 7.500
|   |   STATUS > 6.500: Meningkat 7 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=0, Meningkat 14 Angka=0, Meningkat 8 Angka=0, Meningkat 4 Angka=0, Meningkat 7 Angka=11, Meningkat 12 Angka=0}
|   |   STATUS ≤ 6.500: Meningkat 6 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=23, Meningkat 14 Angka=0, Meningkat 8 Angka=0, Meningkat 4 Angka=0, Meningkat 7 Angka=0, Meningkat 12 Angka=0}
STATUS ≤ 5: Meningkat 4 Angka {Meningkat 10 Angka=0, Meningkat 6 Angka=0, Meningkat 14 Angka=0, Meningkat 8 Angka=0, Meningkat 4 Angka=5, Meningkat 7 Angka=0, Meningkat 12 Angka=0}
```

Figure 4, Decision Tree Description

This decision tree illustrates the grouping of data based on the value of 'STATUS' with several interrelated divisions, where at the first level, if STATUS is greater than 5, a division into two groups is made: if STATUS is greater than 7. 500, it is further divided based on STATUS values greater than 11 and STATUS less than or equal to 11, with students experiencing different number increases, such as 6 students with an increase of 14 numbers if STATUS is greater than 11, and 11 students with an increase of 12 numbers if STATUS is less than or equal to 11, as well as 123 students who experienced an increase of 10 numbers if STATUS is greater than 9 and 41 students who experienced an increase of 8 numbers if STATUS is less than or equal to 9, while on the other hand, if STATUS is less than or equal to 7. 500, the division was made based on STATUS greater than 6,500 with 11 students who experienced an increase of 7 numbers, and 23 students who experienced an increase of 6 numbers if STATUS was less than or equal to 6,500, and 5 students who experienced an increase of 4 numbers if STATUS was less than or equal to 5, all of which indicated that the higher the STATUS value.S

CONCLUSION

This study aims to analyse the daily test results of Indonesian language using Decision Tree C4.5 algorithm to improve the effectiveness of student learning outcomes evaluation. Based on the analysis conducted using RapidMiner software, it was found that the Decision Tree C4.5 algorithm can be used to model the improvement of daily test scores based on the STATUS scores obtained. The distribution of test results in the decision tree model shows a clear relationship between STATUS and grade improvement, where the higher the STATUS score, the greater the grade improvement obtained by the student. However, this study is limited to data obtained from grade X students at SMK Panca Budi, and these results may not be fully representative for other schools. Therefore, further research with a wider sample and additional variables such as other external factors may provide more comprehensive insights. This research contributes to the development of technology-based evaluation methods in education and opens up opportunities for the utilisation of data mining in improving a more effective and efficient learning process.

REFERENCES

Belva Saskia Permana, Lutvia Ainun Hazizah, & Yusuf Tri Herlambang. (2024). Teknologi Pendidikan: Efektivitas Penggunaan Media Pembelajaran Berbasis Teknologi Di Era Digitalisasi. *Khatulistiwa: Jurnal Pendidikan Dan Sosial Humaniora*, *4*(1), 19–28. https://doi.org/10.55606/khatulistiwa.v4i1.2702

Fernanda, E., Suranti, D., & Sudarsono, A. (2023). PERANCANGAN APLIKASI DATA MINING UNTUK KLASIFIKASI KELUARGA MISKIN MENGGUNAKAN METODE DECISION TREE. In *Journal of Science and Social Research* (Issue 3). http://jurnal.goretanpena.com/index.php/JSSR

Fricticarani, A., Hayati, A., Hoirunisa, I., & Mutiara Rosdalina, G. (2023). STRATEGI PENDIDIKAN UNTUK SUKSES DI ERA TEKNOLOGI 5.0. *JIPTI (JURNAL INOVASI PENDIDIKAN DAN TEKNOLOGI INFORMASI)*, *4*(1), 56–68.

Istiqomah, N., Ridla, M. A., & Azise, N. (2024). Data Mining: Tingkat Penghuni Kamar Hotel Di Aceh Dari Tahun 2018-2022 Menggunakan Aplikasi Zaitun. *Gudang Jurnal Multidisiplin Ilmu*, *2*(8), 9–12. https://doi.org/10.59435/gjmi.v2i8.671

Jhody, J. R. (2024). Penerapan Teknik Data Mining terhadap Prediksi Pemilihan Jurusan IPA/IPS Siswa Menggunakan Algoritma C4.5. *Vuca: Jurnal Media Teknologi Dan Informasi*, *1*(1), 27–32.

Naqibuzzahidin, & Fatah, Z. (2024a). Analisis Faktor Risiko Obesitas Pada Individu Menggunakan Algoritma Random Forest Dengan Rapidminer. *Gudang Jurnal Multidisiplin Ilmu*, *2*(10), 54–60. https://doi.org/10.59435/gjmi.v2i11.1016

Naqibuzzahidin, & Fatah, Z. (2024b). Analisis Faktor Risiko Obesitas Pada Individu Menggunakan Algoritma Random Forest Dengan Rapidminer. *Gudang Jurnal Multidisiplin Ilmu*, *2*(10), 54–60. https://doi.org/10.59435/gjmi.v2i11.1016

Parsakh Nursyamsyi, F., & Noor Hasan, F. (2023). Klasifikasi Sentimen Terhadap Aplikasi Identitas Kependudukan Digital Menggunakan Algoritma Naïve Bayes dan SVM. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(3), 1788–1798. https://doi.org/10.30865/klik.v4i3.1517

Pradiah, A. R., Shinta Az-Zahra, A., Lintang, A. M., Suci, M. A., Putri, F. S., Pamulang, U., & Selatan, T. (2024). PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI KUALITAS UDARA DI KOTA BANDUNG. *Jurnal Ilmiah Sains Dan Teknologi*, *2*(8), 312–317.

Putera, A., Siahaan, U., Jabar, A. A., Parhusip, N., Indrayani, M., & Barutu, S. (2024). Analysis of User Age Predictions in Public Satisfaction Surveys at Public Service Malls Using Decision Tree C4.5. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(2), 129–133. https://doi.org/10.61306/jitcse.v1i2

Putera, A., Siahaan, U., Jabar, A. A., Pranoto, S., Sutiono, S., & Ramatika, D. (2024). Analysis of Property Tax Bill Classification Using the C4.5 Algorithm. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(3), 181–185. https://doi.org/10.30596/jitcse

Ramdhani, F., & Setiawan, K. (2024). Penerapan Data Mining untuk Prediksi Pelanggan di PT. XYZ Menggunakan Algoritma Linear Regression. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 490–497. https://doi.org/10.57152/malcom.v4i2.1217

Sadriani, A., Ridwan, M., & Arifin, I. (2023). Peran Guru Dalam Perkembangan Teknologi Pendidikan di Era Digital. In A. Sadriani, I. Arifin, & M. R. S. Ahmad (Eds.), *Peran Guru Dalam Perkembangan Teknologi Pendidikan di Era Digital* (pp. 32–37). UNIVERSITAS NEGERI MAKASSAR. https://journal.unm.ac.id/index.php/Semnasdies62/index

Sari Ndune, I., & Jahja, M. (2024). Evaluasi Ulangan Harian Menggunakan Model Context, Input, Process, Product (CIPP) pada Mata Pelajaran Fisika di Sekolah Menengah Atas Se-Provinsi Gorontalo. *Jurnal Jendela Pendidikan*, *4*(02), 198–207.

Sundari, E. (2024). TRANSFORMASI PEMBELAJARAN DI ERA DIGITAL: MENGINTEGRASIKAN TEKNOLOGI DALAM PENDIDIKAN MODERN. *Sindoro CENDIKIA PENDIDIKAN*, *4*, 50–54.

Tri Diantami, Siwi Widura Yuwana, & Eni Nurhayati. (2023). Pentingnya Pendidikan Bahasa Dalam Membangun Karakter Yang Berbudaya Di SMP PGRI 9 Sidoarjo. *Jurnal Riset Rumpun Ilmu Bahasa*, *2*(2), 132–144. https://doi.org/10.55606/jurribah.v2i2.1512

Triawan, B., Lubis, I., & Kadim, L. A. N. (2024). PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN SPANDUK MENGGUNAKAN ALGORITMA C4.5. *Journal of Mathematics and Technology (MATECH)*, *3*(2), 149–157.