Journal of Information Technology, computer science and Electrical Engineering (JITCSE)
Vol. 1, No. 3, October 2024, pp. 181~185
ISSN: 3046-4900, DOI: 10.30596/jitcse

Analysis of Property Tax Bill Classification Using the C4.5 Algorithm

Andysah Putera Utama Siahaan[1], Ami Abdul Jabar[2], Sugeng Pranoto[3], Sulis Sutiono[4], Desy Ramatika[5]
[1,2,3,4,5]Master of Information Technology, Universitas Pembangunan Panca Budi

ABSTRACT

This study analyzes the classification of Property Tax (Pajak Bumi dan Bangunan, PBB) bills in Tebing Tinggi City using the C4.5 algorithm to improve tax management efficiency. The secondary data used consists of 56,332 entries related to PBB for the 2022-2023 tax year. Using data mining methods and decision tree modeling, the C4.5 algorithm successfully classified taxpayers based on their total bill amount into five categories of tax books. The analysis results show that the majority of taxpayers are classified into categories with lower bills (Books I and II), while high-bill taxpayers (Book V) represent only a small portion of the data. These findings can help local governments design more efficient tax collection policies and adjust resource allocation. Although the study is limited to a single tax year and a specific region, these results contribute to data mining-based PBB management and can serve as a foundation for further research.

Keyword : Property Tax; C4.5 Algorithm; Classification.

Corresponding Author:
Andysah Putera Utama Siahaan,
Master of Information Technology
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing 20122, Medan, Indonesia.
Email : andiesiahaan@gmail.com

Article history:
Received Oct 29, 2024
Revised Oct 30, 2024
Accepted Oct 31, 2024

1. INTRODUCTION

The development of information technology has progressed rapidly along with the changing times, impacting various aspects of human life and also influencing the administration and data management systems in the public and government sectors (Harahap et al., 2023). Computerized technology is expected to provide ease in processing data efficiently, thus generating accurate and timely information, particularly in the field of taxation (Irawan et al., 2020). In this context, the use of information technology to support tax management has become crucial, as taxes are the primary source of government revenue used to finance various development programs.

Taxation, as a mandatory contribution to the state, plays a strategic role in improving public welfare. Law No. 7 of 2021 states that tax is a contribution that must be paid by individuals or legal entities to finance state needs without receiving direct benefits from the payment (Kosasi & Laturette, 2024). One type of tax with significant potential is the Land and Building Tax (PBB), which is levied on the ownership of land or buildings and is expected to be a major source of revenue for local governments to support development (Dewi & EDT, 2024) (Agustin et al., 2024).

In an effort to improve the management of PBB, the use of data analysis techniques such as data mining becomes highly relevant. Data mining is the process of analyzing large volumes of data to identify patterns or hidden information that can be used to make better decisions (Pradiah et al., 2024). One of the commonly used techniques in data mining is the C4.5 algorithm, which is employed to create decision trees. The C4.5 algorithm is known for its ability to classify data into specific categories based on existing attributes and its ease of interpretation (Fauzia et al., 2024) (Bore et al., 2024).

In the context of Land and Building Tax analysis, the use of the C4.5 algorithm for classifying total land and building tax bills can provide significant benefits for tax management. By utilizing decision trees, the government or tax authorities can classify taxpayers based on the size of their tax bills. This enables the government to adjust its tax collection methods, giving more attention to taxpayers with larger bills, while handling those with smaller bills in a simpler manner. Thus, this classification analysis can help allocate resources more effectively, design fairer tax policies, and identify potential risks of tax non-compliance. Therefore, this study aims to perform a classification analysis of total land and building

tax bills using the C4.5 algorithm to enhance efficiency in tax management and support better decision-making in the regional taxation system..

2. RESEARCH METHOD

a. Classification, The data will be grouped based on the Total Bill value into specific categories, and then the appropriate decision tree branches will be assigned to each category. This classification aims to simplify the process of identification and mapping based on the recorded range of tax bills.
b. Data Collection, The data collected for this study is secondary data. Secondary data refers to information gathered from existing sources (Warahmah et al., 2023). This secondary data was obtained from the "Land and Building Tax (PBB-P2) Records of Tebing Tinggi City for the 2022-2023 tax year."
c. Preprocessing, Preprocessing is a crucial step in the data mining process (Putera et al., 2024) and involves the preparation of data before modeling with machine learning algorithms is carried out (Rafrastara et al., 2024). The primary goal of preprocessing is to clean and prepare the data to make it easier to analyze and to produce more accurate results (Haq et al., 2024).
d. Modeling and Analysis, The processed data is then modeled using RapidMiner, and the results of the modeling are subsequently analyzed. RapidMiner is an application or software that serves as a tool for learning in the field of data mining (Hera Yuanti et al., 2024). RapidMiner is used as a solution for analyzing data processing (Apriyani et al., 2023) and also generates decision trees that can be viewed visually (Suriani et al., 2024).

3. RESULTS AND DISCUSSION

A. Data Collection

	A	B	C	D	E	F	G	H	I
1	TAX OBJECT NUMBER	LAND TAX OBJECT SALE VALUE	BUILDING TAX OBJECT SALE VALUE	TAX DETERMINATION	BOOK	TOTAL BILL	TOTAL PAID	STATUS	
2	1274202206001024l0	42539000	33033000	81965	I	163930	163930	OBEDIENT	
3	1274204204003002 70	6480000	12672000	11440	I	22880	22880	OBEDIENT	
4	1274202203003054 50	48330000	48195000	120656	II	241312	241312	OBEDIENT	
5	1274204201002042 90	47400000	0	59250	I	118500	118500	OBEDIENT	
6	1274205206002016 50	100536000	233732000	405335	II	810670	818777	OBEDIENT	
7	1274204201004001 80	43560000	0	54450	I	108900	108900	OBEDIENT	
8	1274205202002037 70	75254000	0	94068	I	188136	94068	DISOBEDIENT	
9	1274205204002024 00	32205000	61285000	104363	II	208726	208726	OBEDIENT	
10	1274205201005005 80	36504000	4536000	38800	I	77600	79928	OBEDIENT	
11	1274202201001002 60	98847000	0	123559	II	247118	247118	OBEDIENT	
12	1274202202004008 40	116064000	0	145080	II	290160	290160	OBEDIENT	
13	1274201206009019 60	50985000	125096000	207601	II	415202	415202	OBEDIENT	
14	1274201204007062 40	21920000	39865000	73525	I	147050	147050	OBEDIENT	
15	1274203201007002 90	80000000	59500000	161875	II	323750	323750	OBEDIENT	
16	1274201202002027 70	59328000	0	74160	I	148320	148320	OBEDIENT	
17	1274204204002030 30	17088000	0	21360	I	42720	21360	DISOBEDIENT	
18	1274204206001005 50	82915000	0	103644	II	207288	207288	OBEDIENT	
19	1274201201007035 70	72960000	58800000	165115	II	330230	330230	OBEDIENT	
20	1274201202001013 60	13376000	18600000	27470	I	54940	54940	OBEDIENT	
21	1274202202005010 20	27408000	24240000	57460	I	57460	57460	OBEDIENT	
22	1274201205004006 80	14729000	51480000	70261	I	140522	140522	OBEDIENT	
23	1274201203005014 30	29274000	0	36593	I	73186	73186	OBEDIENT	
24	1274204203004022 40	9102000	0	11378	I	22756	22756	OBEDIENT	
25	1274201205003004 50	61600000	71175000	153469	II	306938	306938	OBEDIENT	
26	1274205205003022 40	38000000	0	47500	I	95000	95950	OBEDIENT	
27	1274204203005006 50	41097000	0	51371	I	102742	102742	OBEDIENT	

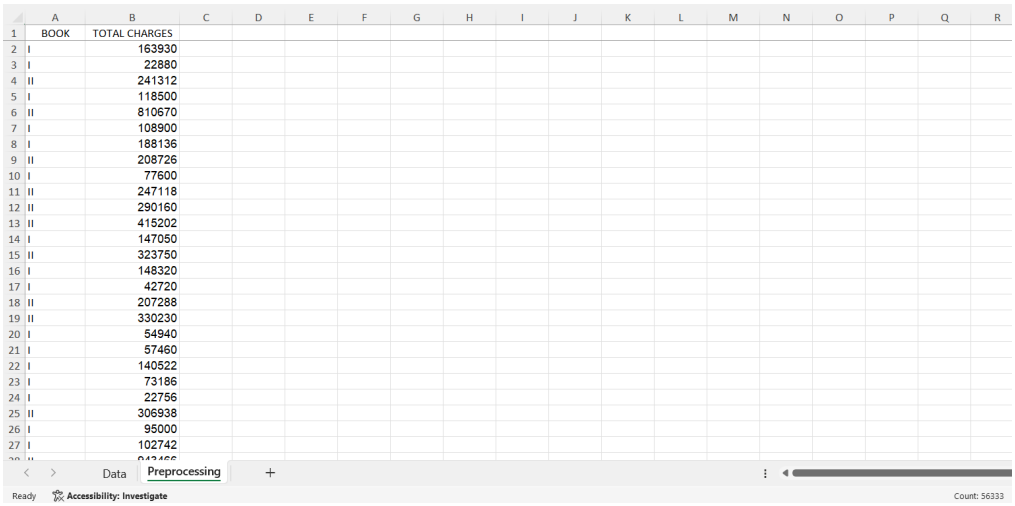
Data Preprocessing +

Ready Accessibility: Investigate Count: 56333

Fig 1. Data Collection

Figure 1 shows that the data from the Land and Building Tax (PBB-P2) Records of Tebing Tinggi City for the 2022-2023 tax year consists of 56,332 rows of data, which include the following attributes: TAX OBJECT NUMBER, LAND TAX OBJECT SALE VALUE, BUILDING TAX OBJECT SALE VALUE, TAX DETERMINATION, BOOK, TOTAL BILL, TOTAL PAID, and STATUS.

B. Preprocessing



	BOOK	TOTAL CHARGES
2	I	163930
3	I	22880
4	II	241312
5	I	118500
6	II	810670
7	I	108900
8	I	188136
9	II	208726
10	I	77600
11	II	247118
12	II	290160
13	II	415202
14	I	147050
15	II	323750
16	I	148320
17	I	42720
18	II	207288
19	II	330230
20	I	54940
21	I	57460
22	I	140522
23	I	73186
24	I	22756
25	II	306938
26	I	95000
27	I	102742

Fig 2. Preprocessing

In Figure 2, the data has undergone preprocessing, leaving only the attributes BOOK and TOTAL CHARGES.

C. Modeling and Analysis

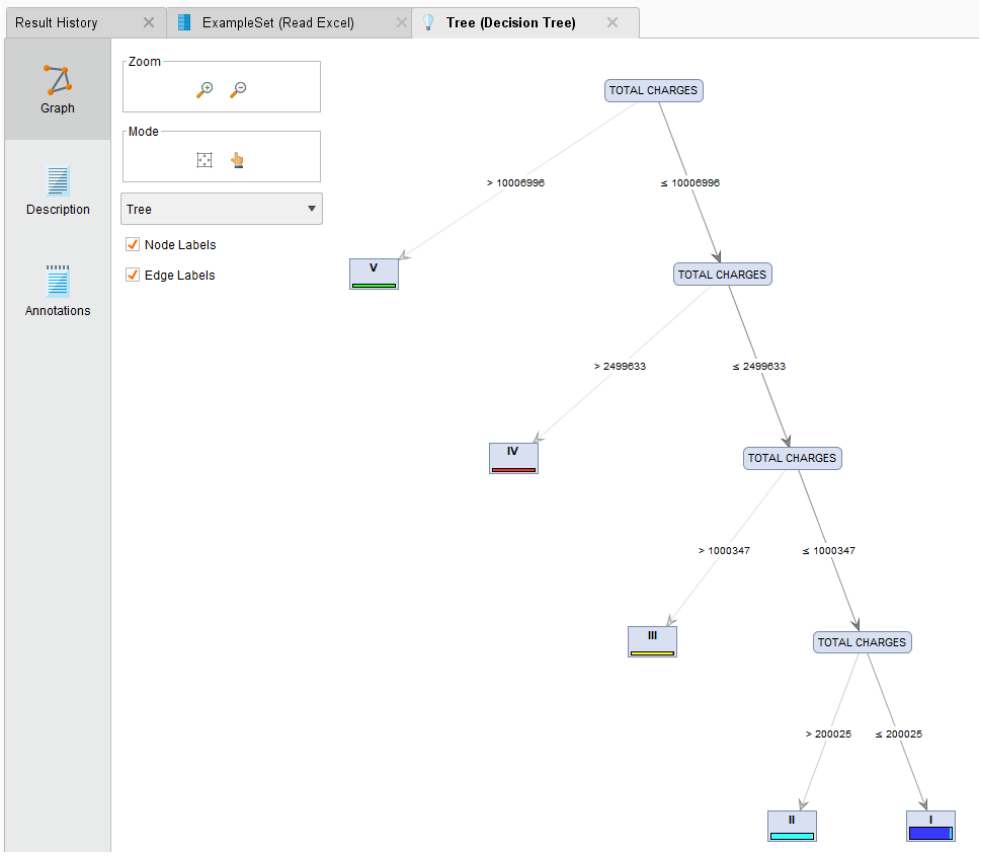


Fig 3. Modeling and Analysis

In Figure 3, it can be observed that this decision tree model groups the data based on the "TOTAL CHARGES" value using a series of progressively smaller thresholds. At the first level, if the TOTAL CHARGES are greater than 10,006,996, the data is classified into Book V. However, if the TOTAL CHARGES are less than or equal to 10,006,996, the model proceeds to the next level. At the second level, if the TOTAL CHARGES are greater than 2,499,633, the result is Book IV, while if they are less than or

equal to that value, the division continues. At the third level, if the TOTAL CHARGES are greater than 1,000,347, the data is classified into Book III, and if less than or equal to that value, the model proceeds to the next level. At the fourth level, if the TOTAL CHARGES are greater than 200,025, the resulting classification is Book II, and if less than or equal to that value, the final classification is Book I. Thus, this decision tree uses a series of increasingly specific divisions to classify the data based on the TOTAL CHARGES value.

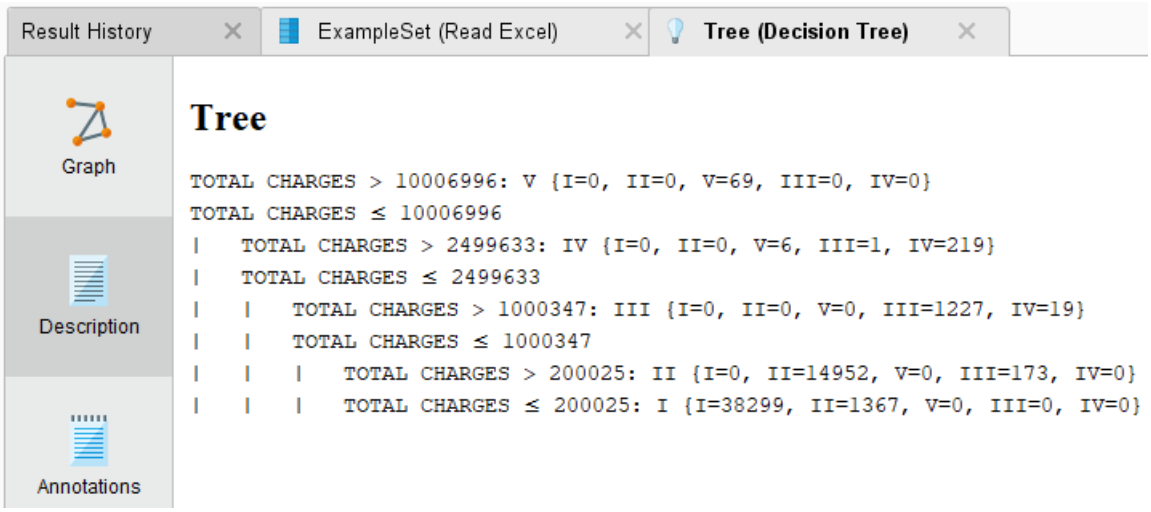


Fig 4. Description Tree

In Figure 4, it can be seen that this decision tree structure illustrates the division of data based on the TOTAL CHARGES value with progressively more specific levels of division. If the TOTAL CHARGES are greater than 10,006,996, the data is classified into Book V (I=0, II=0, V=69, III=0, IV=0). If the TOTAL CHARGES are less than or equal to 10,006,996, the division continues: if greater than 2,499,633, the data is classified into Book IV (I=0, II=0, V=6, III=1, IV=219); if less than or equal to 2,499,633 but greater than 1,000,347, the data is classified into Book III (I=0, II=0, V=0, III=1227, IV=19); if less than or equal to 1,000,347 but greater than 200,025, the data is classified into Book II (I=0, II=14,952, V=0, III=173, IV=0); and if the TOTAL CHARGES are less than or equal to 200,025, the data is classified into Book I (I=38,299, II=1,367, V=0, III=0, IV=0). This analysis shows that the majority of the data is distributed in books with lower TOTAL CHARGES, namely Book I and Book II, while Book V contains only a small number of entries with the highest TOTAL CHARGES.

4. CONCLUSION

This study analyzes the classification of Property Tax (Pajak Bumi dan Bangunan, PBB) bills in Tebing Tinggi City using the C4.5 algorithm to enhance the efficiency of tax management. Based on the decision tree modeling, the bill data were successfully grouped into five categories based on the Total Bill value, which reflects variations in tax bill amounts. The analysis shows that the majority of taxpayers are classified into the books with lower tax bills (Book I and II), while the books with higher tax bills (Book V) contain only a small portion of the data. The use of the C4.5 algorithm has proven effective in classifying taxpayers, helping local governments design more efficient and equitable tax collection policies, as well as assisting in better resource allocation and the identification of tax non-compliance risks. However, this study has limitations, such as the data sample being limited to a single tax year and a specific region, necessitating further research with a broader scope to ensure the validity and sustainability of the findings. Overall, this research makes a significant contribution to the management of PBB in Tebing Tinggi City and can serve as a reference for the application of data analysis in local taxation. It also encourages further research that involves additional variables and different regions to enrich understanding of data mining-based tax management.

REFERENCES

Agustin, P. D., Dinanty, D., & Nasution, J. H. (2024). PAJAK BUMI DAN BANGUNAN (PBB). *MUSYTARI*, *5*(5).

Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer*, *2*(1), 20–33. https://doi.org/10.56211/helloworld.v2i1.230

Bore, A. T., Trisno, & Kurra Titus. (2024). PENENTUAN DAYA LISTRIK MENGGUNAKAN METODE DECISION TREE DI DESA TEBARA. *Multidisciplinary Indonesian Center Journal (MICJO)*, *1*, 396–403.

Dewi, W. K., & EDT, R. W. (2024). Pengaruh Pengetahuan Pajak dan Kesadaran Wajib Pajak Terhadap Kepatuhan Membayar Pajak Bumi dan Bangunan (PBB). *JCA (Jurnal Cendekia Akuntansi)*, *5*(1), 61. https://doi.org/10.32503/akuntansi.v5i1.5343

Fauzia, A., Ndruru, N., & Sindar, A. (2024). Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5. *Katera : Jurnal Sains Dan Teknologi*, *1*(1), 23.

Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, *13*(1), 1079–1084. https://doi.org/10.33395/jmp.v13i1.13937

Harahap, S. N., Simatupang, M., & Atika, L. (2023). Penguatan Learning Management System (LMS) untuk Peningkatan Kualitas Pembelajaran pada Era Society 5.0 di Prodi Pendidikan Teknologi dan Kejuruan. *JAVIT : Jurnal Vokasi Informatika*, *3*(1), 1–10. https://doi.org/10.24036/javit.v3i1.149

Hera Yuanti, A., Bisnis, F., & Teknologi Yogyakarta, U. (2024). Analisis Pengaruh Covid-19 Terhadap Kesehatan Mental Dengan Visualisasi Data Rapidminer. *Gudang Jurnal Multidisiplin Ilmu*, *2*(1), 183–187. https://doi.org/10.59435/gjmi.v2i1.225

Irawan, D., Novianto, Z., Komputer, F., Bina Insan Jalan Jenderal Besar M Soharto Kelurahan Lubuk Kupang Kecamatan Lubuklinggau Selatan, U. H., & Lubuklinggau, K. (2020). PERANCANGAN E-LEARNING PADA SMAN 1 KOTA LUBUKLINGGAU MENGGUNAKAN FRAMEWORK CODEIGNITER (CI). *Jurnal Digital Teknologi Informasi*, *3*(2), 2020.

Kosasi, J., & Laturette, K. (2024). Pengaruh Motivasi, Self-Efficacy, Prospek Kerja, Pengetahuan Perpajakan, Nilai Sosial, dan Persepsi terhadap Minat Mahasiswa Jurusan Akuntansi untuk Menjadi Konsultan Pajak. *JURNAL EKSPLORASI AKUNTANSI*, *6*(3), 946–960. https://doi.org/10.24036/jea.v6i3.1785

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. In *International Journal of Computer Sciences and Mathematics Engineering* (Vol. 3, Issue 1).

Rafrastara, F. A., Ghozi, W., & Wardoyo, A. (2024). Deteksi Serangan berbasis Machine Learning pada Internet of Vehicle. *Seminar Nasional Informatika-FTI UPGRIS*, *2*.

Suriani, U., Palingga Ninditama, I., Syaputra, W., Studi Teknologi Rekayasa Multimedia, P., & Prasetya Mandiri, P. (2024). Pemodelan Prediktif Keterlambatan Bicara pada Balita Terkait dengan Penggunaan Smartphone Menggunakan Data Mining. *Journal of Information Technology Ampera*, *5*(1), 2774–2121. https://doi.org/10.51519/journalita.v5i1.589

Warahmah, M., Risnita, & Jailani, M. S. (2023). Pendekatan Dan Tahapan Penelitian Dalam Kajian Pendidikan Anak Usia Dini. *DZURRIAT : Jurnal Pendidikan Islam Anak Usia Dini*, *1*(2), 72–81.