Journal of Information Technology, computer science and Electrical Engineering (JITCSE)
Vol. 1, No. 2, June 2024, pp. 121~128
ISSN: 3046-4900, DOI: 10.61306/jitcse.v1i2

Comparison of Accuracy between Naïve Bayes and Decision Tree Methods for Property Tax (PBB-P2) Compliance in Tebing Tinggi City

Zulham Sitorus[1], Sugeng Pranoto[2], Sulis Sutiono[3], Sarifuddin[4]
[1,2,3,4]Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

ABSTRACT

This research aims to compare the accuracy of the Naïve Bayes and Decision Tree methods in predicting Land and Building Tax (PBB-P2) compliance in Tebing Tinggi city. The data used includes tax and payment determination for 2022 and 2023. The methods applied include data preprocessing, use of an inconvenience matrix for evaluation, as well as measuring accuracy with various data sharing ratios (80:20, 75:25, 70:30, 60:40, and 50:50). The research results show that the Decision Tree model consistently has much higher accuracy compared to the Naïve Bayes model, with accuracy reaching 99% at all data split ratios, while Naïve Bayes shows accuracy between 54% and 56%. The confusion matrix supports this finding by showing that the Decision Tree model has higher True Positives and True Negatives, and lower False Positives and False Negatives compared to Naïve Bayes. In conclusion, the Decision Tree method is more effective in classifying tax compliance compared to Naïve Bayes so that it is a more optimal choice for a tax compliance classification system based on the accuracy and performance obtained from this research.

Keyword : Naive Bayes, Decision Tree, Tax Compliance, Classification, Confusion Matrix

Corresponding Author:
Zulham Sitorus
Master Of Information Technology
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing 20122, Medan, Indonesia.
Email : zulhamsitorus@dosen.pancabudi.ac.id

Article history:
Received May 25, 2024
Revised May 27, 2024
Accepted May 30, 2024

1. INTRODUCTION

Information technology utilization is recognized as one of the ways organizations support the effectiveness of achieving their goals, including government organizations (Mugiarto et al., 2023). The development of information and communication technology is rapidly advancing and has a significant impact on human daily activities (Khoiriah et al., 2023). One of the technological advancements that has captured attention and transformed the way we interact with the world is artificial intelligence, commonly known as AI (Hanila & Alghaffaru, 2023). Within the scope of AI, machine learning (ML) technology refers to systems developed to learn autonomously without direct guidance from users. (Hanila & Alghaffaru, 2023).

Land and Building Tax is a highly potential source of revenue for regions as one of the direct taxes (Kolatung, 2021). The Rural and Urban Land and Building Tax is levied on the ownership, control, and/or use of rural and urban land and/or buildings. Land refers to the surface of the Earth, including land and inland waters as well as seas within the region. Buildings are engineering constructions that are permanently fixed or attached to the land and/or inland waters and/or seas (Sabu & Tang, 2023). Tax compliance is described as the condition in which all tax obligations are fulfilled and tax rights are exercised. This compliance is a strategic action aimed at increasing state revenue (Coo et al., 2021). To determine taxpayer compliance, classification is necessary. Classification is the process of identifying definitions of similar characteristics within a group or class (Yudana et al., 2023). The goal of classification is to accurately predict the target class for each case in the data (Alfiah, 2021).

In the context of machine learning, there are various classification methods used to analyze and predict data. One simple classification method is Naive Bayes, which utilizes probability theory to find the best likelihood and predict future probabilities based on past information (Wibisono et al., 2020). Additionally, the Decision Tree is one of the most popular classification methods in ML, known for its ability to break down complex decision-making processes into simpler, more understandable structures for humans (Rahmansyah Nur, Ningsih Sari, Lantana Dhieka Avrilia, Suryaningtyas Adisti, Wirawan Putri, Wijaya Sifonne Adi, 2023). To evaluate the performance of classification models such as Naive Bayes and Decision Tree, a Confusion Matrix is used. A Confusion Matrix is a method that illustrates the model's accuracy by presenting a table showing the number of test data correctly classified as well as the number of test data misclassified (Isman et al., 2021). The use of a Confusion Matrix allows for the acquisition of important evaluation metrics such as accuracy, precision, recall, and F1-score, which serve to assess the effectiveness of the applied classification methods (Endang Etriyanti, 2021).

This research can be conducted through a comparative study, which is a type of analysis aimed at comparing the similarities and differences between two or more objects under investigation to discover a new conceptual framework (Rizkita & Supriyanto, 2020). To conduct a comparison, Google Colab and Python are required. Google Colab, also known as Colab, is a free cloud computing service provided by Google that allows users to create or run Python programs in a web browser (Maulana & Noriska, 2023). Python is a programming language that uses an interpreter to execute its code. This interpreter translates the code directly, allowing Python to be run on various platforms, such as Windows, Linux, and others (Rahman et al., 2023). Based on the previous explanation, this research aims to compare the Naive Bayes and Decision Tree methods in predicting the compliance of Land and Building Tax (PBB-P2) taxpayers in Tebing Tinggi city by calculating their accuracy values. Through this comparative study, it is hoped that a more effective classification method for predicting tax compliance can be identified, thereby providing deeper insights into the performance of both methods.

2. RESEARCH METHOD

a. Data Collection, where property tax (PBB-P2) determination and payment data for the tax years 2022 and 2023 in Tebing Tinggi city are used
b. Data preprocessing, performed to obtain high-quality data (Endang Etriyanti, 2021).
c. Confusion Matrix, according to (Yulian et al., 2023), consists of rows and columns. Rows represent the results for actual classes, while columns represent the results for predicted classes. For binary classification problems, a 2x2 matrix is used, displaying values for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True Positives (TP) indicate the number of positive reviews that the model correctly predicted. True Negatives (TN) show the number of negative reviews that the model correctly predicted. False Positives (FP) represent the number of negative reviews that the model incorrectly predicted as positive, and False Negatives (FN) reflect the number of positive reviews that the model incorrectly predicted as negative. This matrix is used to test and compare the accuracy of Naïve Bayes and Decision Tree models.

3. RESULTS AND DISCUSSION

A. Data collection

	NOP	NM_WP_SPPT	THN_PAJAK_SPPT	NJOP_BUMI	NJOP_BNG	KETETAPAN	BAYAR
1	127420320600104180	SUMADI	2022	7680000	12672000	15700	(null)
2	127420320600104190	TINUR BUTAR BUTAR	2022	17496000	11250000	23433	(null)
3	127420320600104200	IRWANSYAH	2022	11200000	9720000	13650	(null)
4	127420320600104210	UMBIANTO SINAGA	2022	3200000	2592000	10000	(null)
5	127420320600104220	AGUSTINA	2022	5120000	7200000	10000	(null)
6	127420320500600180	AMALUDDIN SARAGIH	2022	12874000	0	16093	16093
7	127420320500600190	RUSLI	2022	82000000	43800000	144750	144750
8	127420320500600200	ALI	2022	23780000	0	29725	29725
9	127420320500600210	H RASIMAN	2022	60576000	0	75720	(null)
10	127420320500600220	YUSDIANA	2022	19200000	0	24000	24000
11	127420420100302250	RACHMAT SUPRAPTO	2022	13440000	23166000	33258	(null)
12	127420420100302260	SURIAWAN	2022	13440000	29601000	41301	(null)
13	127420420100302270	MIRA YULIANA NASUTION	2022	13440000	18018000	26823	(null)
14	127420420100302280	SARI DENI PURBA	2022	13440000	20592000	30040	30040
15	127420420100302290	SRI UTAMI ASTUTI	2022	13440000	20592000	30040	(null)
16	127420320500500460	MUCSHIN HASIBUAN	2022	29925000	12672000	40746	(null)
17	127420320500500470	AMAT DALIL	2022	25935000	9450000	31731	31731
18	127420320500500490	AISYAH	2022	17100000	10800000	22375	(null)
19	127420320500500500	MASTUTI, DKK	2022	104310000	0	130388	(null)
20	127420320500500510	ARNISAH	2022	46170000	12672000	61053	(null)
21	127420320500500520	FIRMAN DKK	2022	18810000	24090000	41125	58398
22	127420320500500530	AKIAR	2022	17100000	4860000	14950	14950

Fig 1. Determination and Payment Data

In Figure 1, the data for property tax (PBB-P2) determination and payments for the tax years 2022 and 2023 can be seen, consisting of columns (NOP, NM_WP_SPPT, THN_PAJAK,SPPT, NJOP_BUMI, NJOP_BNG, KETETAPAN, BAYAR)

B. Pre Processing

	NOP	NJOP_BUMI	NJOP_BNG	KETETAPAN	BUKU	TOTAL_TUNGGAKAN	TOTAL_PEMBAYARAN	STATUS
1	127420220600102410	42539000	33033000	81965	1	163930	163930	1
2	127420420400300270	6480000	12672000	11440	1	22880	22880	1
3	127420220300305450	48330000	48195000	120656	2	241312	241312	1
4	127420420100204290	47400000	0	59250	1	118500	118500	1
5	127420520600201650	100536000	233732000	405335	2	810670	818777	1
6	127420420100400180	43560000	0	54450	1	108900	108900	1
7	127420520200203770	75254000	0	94068	1	188136	94068	0
8	127420520400202400	32205000	61285000	104363	2	208726	208726	1
9	127420520100500580	36504000	4536000	38800	1	77600	79928	1
10	127420220100100260	98847000	0	123559	2	247118	247118	1
11	127420220200400840	116064000	0	145080	2	290160	290160	1
12	127420420300101390	36396000	0	45495	1	90990	90990	1
13	127420120600901960	50985000	125096000	207601	2	415202	415202	1
14	127420120400706240	21920000	39865000	73525	1	147050	147050	1
15	127420320100700290	80000000	59500000	161875	2	323750	323750	1
16	127420120200202770	59328000	0	74160	1	148320	148320	1
17	127420420400203030	17088000	0	21360	1	42720	21360	0
18	127420420400306540	19200000	18180000	34225	1	68450	69135	1

Fig 2. Data Preprocessing

In Figure 2, the data for property tax (PBB-P2) determination and payments for the tax years 2022 and 2023, after preprocessing, can be seen. It consists of the following columns (NOP, NJOP_BUMI, NJOP_BNG, KETETAPAN, BUKU, TOTAL_TUNGGAKAN, TOTAL_PEMBAYARAN, STATUS)

C. Confusion Matrix

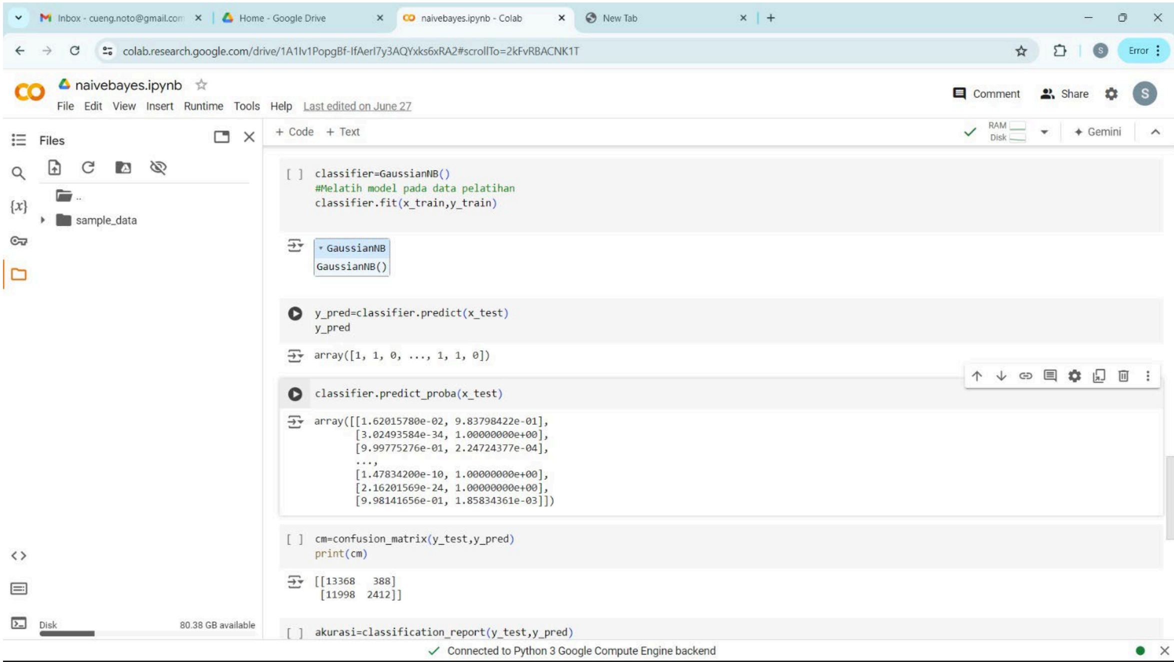
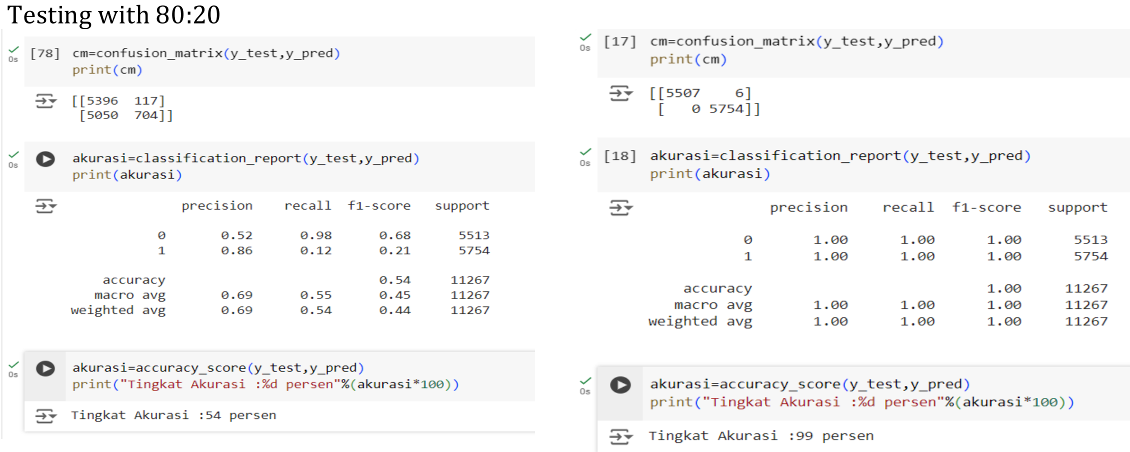


Fig 3. Confusion Matrix

The confusion matrix displayed in Figure 3 shows the following values: [[13368 388], [11998 2412]]. In the confusion matrix, rows represent the results for actual classes, while columns represent the results for predicted classes. According to the theory (Yulian et al., 2023), in this 2x2 matrix for binary classification, the number 13,368 represents True Positives (TP), which indicates the number of correct predictions for the positive class. The number 388 is False Positives (FP), showing the number of incorrect predictions for the positive class. The number 11,998 represents False Negatives (FN), reflecting the number of incorrect predictions for the negative class. Finally, the number 2,412 denotes True Negatives (TN), indicating the number of correct predictions for the negative class. This confusion matrix provides insights into the model's ability to classify data, both in terms of correct predictions and classification errors.

D. Testing and Comparing the Accuracy of Naive Bayes and Decision Tree Methods

1) Testing with 80:20

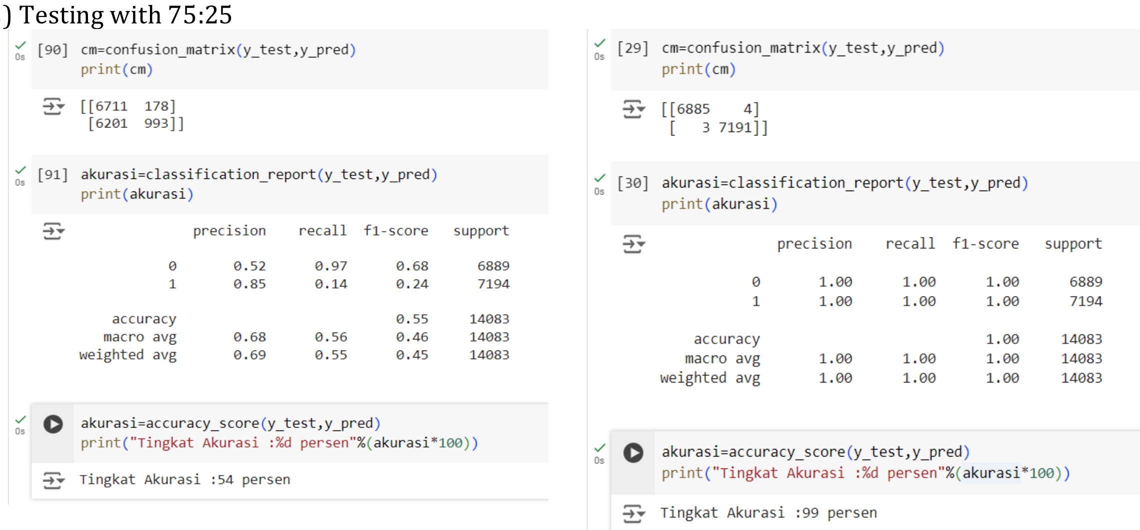


(a) (b)

Fig 4. Results of Testing with 80:20,
(a) Naïve Bayes, (b) Decision Tree

The results of testing with an 80:20 split, as shown in Figure 4, indicate that the Naïve Bayes model achieved an accuracy rate of 54%, while the Decision Tree model achieved an accuracy rate of 99%. Thus, the Decision Tree model demonstrates the highest accuracy.

2) Testing with 75:25

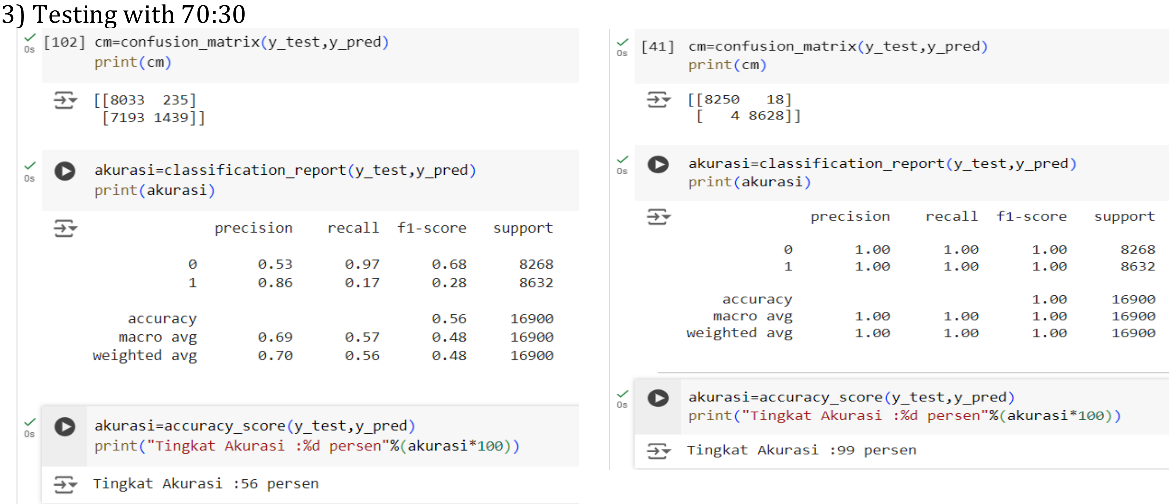


(a) (b)

Fig 5. Results of Testing with 75:25,
(a) Naïve Bayes, (b) Decision Tree

The results of testing with a 75:25 split, as shown in Figure 5, indicate that the Naïve Bayes model achieved an accuracy rate of 54%, while the Decision Tree model achieved an accuracy rate of 99%. Thus, the Decision Tree model demonstrates the highest accuracy.

3) Testing with 70:30


```
[102] cm=confusion_matrix(y_test,y_pred)
      print(cm)

[[8033  235]
 [7193 1439]]

akurasi=classification_report(y_test,y_pred)
print(akurasi)

              precision    recall  f1-score   support

           0       0.53      0.97      0.68      8268
           1       0.86      0.17      0.28      8632

    accuracy                           0.56     16900
   macro avg       0.69      0.57      0.48     16900
weighted avg       0.70      0.56      0.48     16900

akurasi=accuracy_score(y_test,y_pred)
print("Tingkat Akurasi :%d persen"%(akurasi*100))

Tingkat Akurasi :56 persen
```

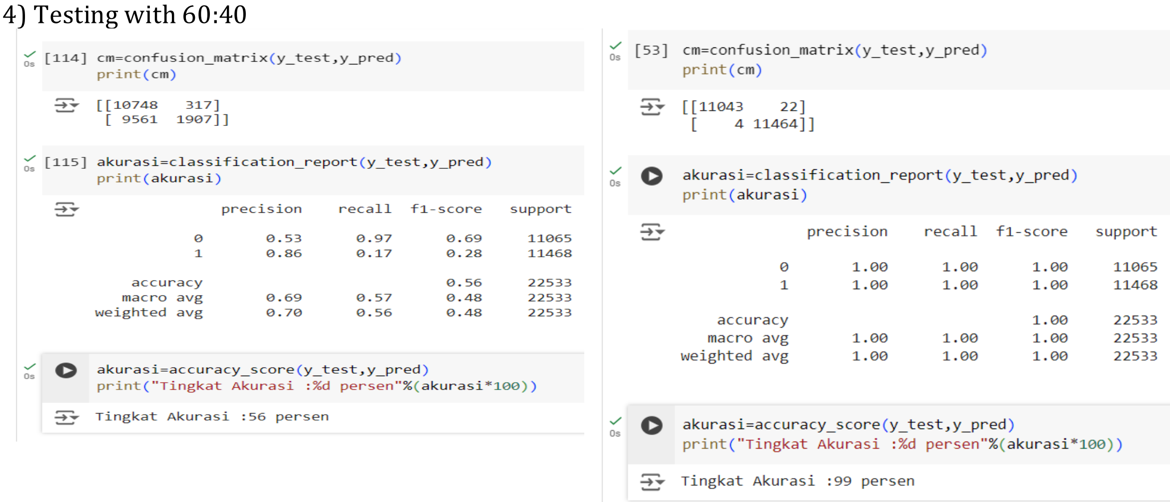
```
[41] cm=confusion_matrix(y_test,y_pred)
     print(cm)

[[8250   18]
 [   4 8628]]

akurasi=classification_report(y_test,y_pred)
print(akurasi)

              precision    recall  f1-score   support

           0       1.00      1.00      1.00      8268
           1       1.00      1.00      1.00      8632

    accuracy                           1.00     16900
   macro avg       1.00      1.00      1.00     16900
weighted avg       1.00      1.00      1.00     16900

akurasi=accuracy_score(y_test,y_pred)
print("Tingkat Akurasi :%d persen"%(akurasi*100))

Tingkat Akurasi :99 persen
```

(a) (b)

Fig 6. Results of Testing with 70:30,
(a) Naïve Bayes, (b) Decision Tree

The results of testing with a 70:30 split, as shown in Figure 6, indicate that the Naïve Bayes model achieved an accuracy rate of 56%, while the Decision Tree model achieved an accuracy rate of 99%. Thus, the Decision Tree model demonstrates the highest accuracy.

4) Testing with 60:40


```
[114] cm=confusion_matrix(y_test,y_pred)
      print(cm)

[[10748   317]
 [ 9561  1907]]

[115] akurasi=classification_report(y_test,y_pred)
      print(akurasi)

              precision    recall  f1-score   support

           0       0.53      0.97      0.69     11065
           1       0.86      0.17      0.28     11468

    accuracy                           0.56     22533
   macro avg       0.69      0.57      0.48     22533
weighted avg       0.70      0.56      0.48     22533

akurasi=accuracy_score(y_test,y_pred)
print("Tingkat Akurasi :%d persen"%(akurasi*100))

Tingkat Akurasi :56 persen
```

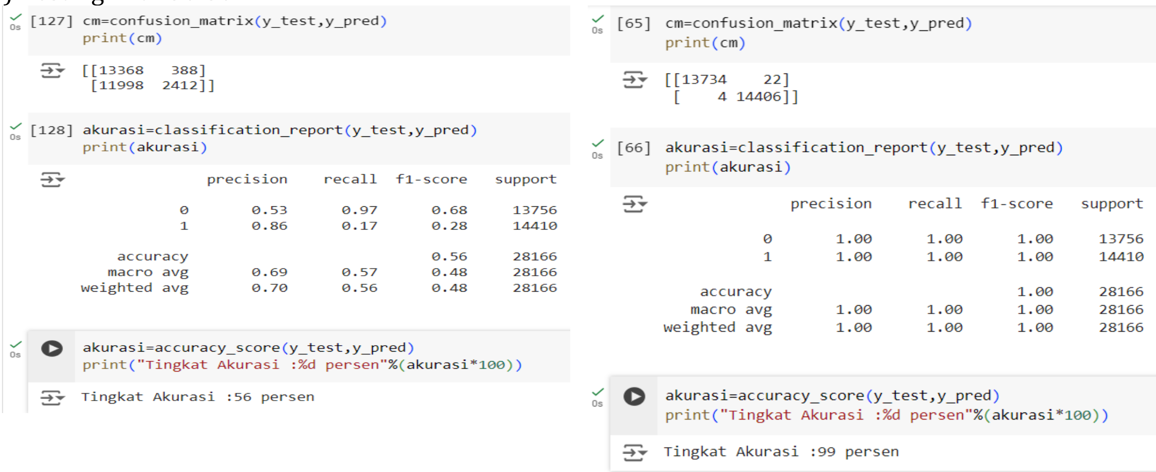
```
[53] cm=confusion_matrix(y_test,y_pred)
     print(cm)

[[11043    22]
 [    4 11464]]

akurasi=classification_report(y_test,y_pred)
print(akurasi)

              precision    recall  f1-score   support

           0       1.00      1.00      1.00     11065
           1       1.00      1.00      1.00     11468

    accuracy                           1.00     22533
   macro avg       1.00      1.00      1.00     22533
weighted avg       1.00      1.00      1.00     22533

akurasi=accuracy_score(y_test,y_pred)
print("Tingkat Akurasi :%d persen"%(akurasi*100))

Tingkat Akurasi :99 persen
```

(a) (b)

Fig 7. Results of Testing with 60:40,
(a) Naïve Bayes, (b) Decision Tree

The results of testing with a 60:40 split, as shown in Figure 7, indicate that the Naïve Bayes model achieved an accuracy rate of 56%, while the Decision Tree model achieved an accuracy rate of 99%. Thus, the Decision Tree model demonstrates the highest accuracy.

5) Testing with 50:50



(a) (b)

Fig 8. Results of Testing with 50:50
(a) Naïve Bayes, (b) Decision Tree

The results of testing with a 50:50 split, as shown in Figure 8, indicate that the Naïve Bayes model achieved an accuracy rate of 56%, while the Decision Tree model achieved an accuracy rate of 99%. Thus, the Decision Tree model demonstrates the highest accuracy.

6. Results of Overall Accuracy Comparison Testing

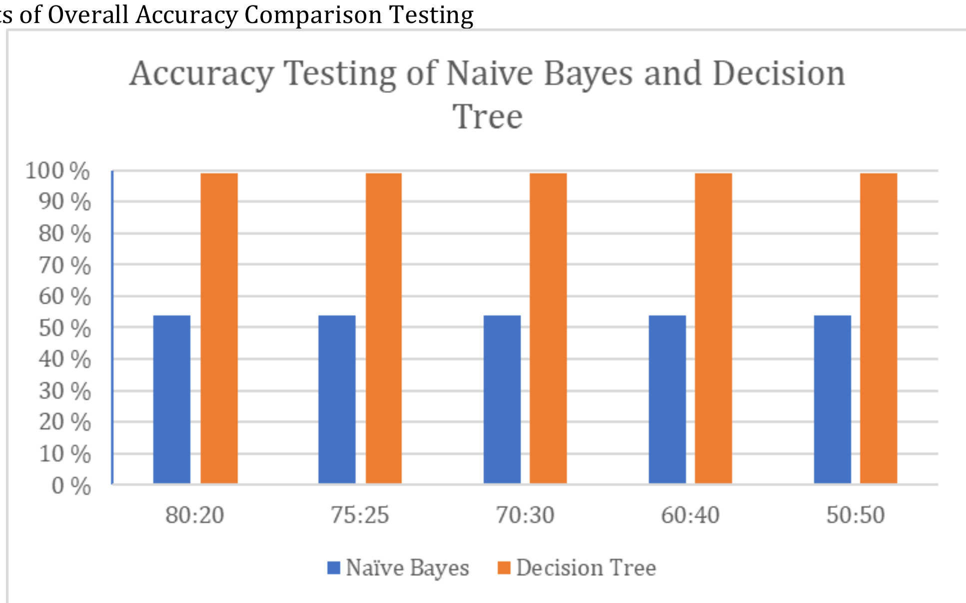


Fig. 9 Overall Accuracy Comparison Results of Naïve Bayes and Decision Tree

From Figure 9, it can be observed that the overall comparison results of Naïve Bayes and Decision Tree across tests with splits of (80:20, 75:25, 70:30, 60:40, 50:50) indicate that the Decision Tree model has the highest accuracy compared to Naïve Bayes.

4. CONCLUSION

Based on the results of testing conducted with various data splits, it was found that the Decision Tree model consistently shows a significantly higher accuracy compared to the Naïve Bayes model. In tests with data splits of 80:20, 75:25, 70:30, 60:40, and 50:50, the Decision Tree model achieved an accuracy rate of 99%, while the Naïve Bayes model had accuracy ranging from 54% to 56%. These results indicate that the Decision Tree is more effective in classifying tax compliance compared to Naïve Bayes. The confusion matrix also supports this finding by showing that the Decision Tree model has higher True Positives and True Negatives, and lower False Positives and False Negatives compared to Naïve Bayes. Overall, choosing the Decision Tree model for the tax compliance classification system is a more optimal decision based on the accuracy and performance levels obtained from this study.

REFERENCES

Alfiah, N. (2021). Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes. *Respati*, *16*(1), 32. https://doi.org/10.35842/jtir.v16i1.386

Coo, N., Suhardiyah, M., & Kurnia, Y. (2021). Pengaruh Kesadaran Wajib Pajak dan Sanksi Pajak Terhadap Kepatuhan Wajib PajaK pbb di Kelurahan Menanggal Surabaya. *Journal of Sustainability Business Research*, *2*(1), 399–407.

Endang Etriyanti. (2021). Perbandingan Tingkat Akurasi Metode Knn Dan Decision Tree Dalam Memprediksi Lama Studi Mahasiswa. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, *3*(1), 6–14. https://doi.org/10.52303/jb.v3i1.40

Hanila, S., & Alghaffaru, M. A. (2023). Pelatihan Penggunaan Artificial Intelligence (AI) Terhadap Perkembangan Teknologi Pada Pembelajaran Siswa Sma 10 Sukarami Kota Bengkulu. *Jurnal Dehasen Mengabdi*, *2*(2), 221–226. https://doi.org/10.37676/jdm.v2i2.4890

Isman, Andani Ahmad, & Abdul Latief. (2021). Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(3), 557–564. https://doi.org/10.29207/resti.v5i3.3006

Khoiriah, S. U., Lubis, L. K. L. U., & Anas, D. K. N. (2023). Analisis Perkembangan Sistem Manajemen Pendidikan di Era Society 5.0. *JISPENDIORA: Jurnal Ilmu Sosial, Pendidikan Dan Humaniora*, *2*(2), 117–132. https://doi.org/10.56910/jispendiora.v2i2.650

Kolatung, J. F. (2021). Analisis Tingkat Kepatuhan Wajib Pajak Bumi Dan Bangunan Di Kota Manado. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, *9*(2), 1006–1014.

Maulana, B. A., & Noriska, N. K. S. (2023). The Use of Machine Learning in the Analysis of Pawnshop Digital Service on Customer Satisfaction. *Indonesian Journal of Contemporary Multidisciplinary Research*, *2*(4), 681–702. https://doi.org/10.55927/modern.v2i4.4612

Mugiarto, M., Agustina, I., & Suryaman, W. (2023). Adoption Of Asset Application Technology And Employee Competence In Improving Employee Performance Of State-Owned Asset Management In Sopd Cimahi City. *Management Studies and Entrepreneurship Journal*, *4*(5), 4823–4835. http://journal.yrpipku.com/index.php/msej

Rahman, S., Sembiring, A., Siregar, D., Khair, H., Gusti Prahmana, I., Puspadini, R., & Zen, M. (2023). Python : Dasar Dan Pemrograman Berorientasi Objek. In *Penerbit Tahta Media*.

Rahmansyah Nur, Ningsih Sari, Lantana Dhieka Avrilia, Suryaningtyas Adisti, Wirawan Putri, Wijaya Sifonne Adi, P. D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *INNOVATIVE: Journal Of Social Science Research Volume*, *3*(2), 13880–13892.

Rizkita, K., & Supriyanto, A. (2020). Komparasi kepemimpinan pendidikan di Indonesia dan Malaysia dalam upaya peningkatan mutu pendidikan. *Jurnal Akuntabilitas Manajemen Pendidikan*, *8*(2), 155–164. https://doi.org/10.21831/jamp.v8i2.32362

Sabu, junius menase sau, & Tang, sefnat aristarkus. (2023). Analisis Pengaruh Pajak Bumi Dan Bangunan Perdesaan Dan Perkotaan, Pajak Hotel Dan Pajak Restoran Terhadap Pendapatan Asli Daerah Kabupaten Alor. *Jurnal Ilmiah Wahana Pendidikan*, *9*(20), 368–382. http://www.nber.org/papers/w16019

Wibisono, A. D., Dadi Rizkiono, S., & Wantoro, A. (2020). Filtering Spam Email Menggunakan Metode Naive Bayes. *TELEFORTECH : Journal of Telematics and Information Technology*, *1*(1). https://doi.org/10.33365/tft.v1i1.685

Yudana, F. R., Suyanto, M., & Nasiri, A. (2023). Model Klasifikasi Untuk Menentukan Kesiapan Kerja Mahasiswa Dan Kelulusan Tepat Waktu Dengan Metode Machine Learning. *IJITECH: Indonesian Journal of Information Technology*, *1*(1), 1–12. https://doi.org/10.37680/ijitech.v1i1.xx

Yulian, S., Noorihsan, F., Setiawan, N. Y., & Saputra, M. C. (2023). Analisis Sentimen Ulasan Google Review New Star Cineplex Pasuruan menggunakan Artificial Neural Network (ANN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *7*(2), 564–573. http://j-ptiik.ub.ac.id