CLASSIFICATION OF TRAFFIC TICKET CASES AT THE PAGAR ALAM DISTRICT ATTORNEY'S OFFICE USING THE C4.5 ALGORITHM

Muntahanah[1], Sabella Azhari[2]*, Sastya Hendri Wibowo[3], Dandi Sunardi[4]
[1,2,3,4]*Informatics Engineering*
[1,2,3,4]*Faculty of Engineering*
[1,2,3,4]*University of Muhammadiyah Bengkulu*
E-mail : *muntahanah@umb.ac.id[1], sabellaazhari96@gmail.com[2]*

Abstract

The increasing number of traffic violations in Pagar Alam City has led to a yearly rise in ticketing case data at the Pagar Alam District Attorney's Office. This accumulation of data creates difficulties in effective data management and hinders the extraction of meaningful insights. The current classification process for ticketing cases remains limited in its accuracy and efficiency, making it difficult to identify patterns or trends. This study aims to address this issue by developing a classification model for traffic ticket cases using data mining techniques, specifically the C4.5 algorithm. The model classifies cases based on attributes such as the relevant article of law, type of vehicle, evidence submitted, and the fine imposed. The CRISP-DM framework is used to guide the process through six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. RapidMiner is used as the primary tool for data processing, and the model is evaluated using the X-Cross Validation technique. The results show that the C4.5 algorithm achieves a high classification accuracy of 99.75%. The "Article" attribute emerged as the most influential factor with the highest gain ratio value. These findings can support law enforcement and policymakers in identifying the most frequent violations and developing more targeted strategies to improve traffic law enforcement and public safety.

Keywords- *Classification, Data Mining, C4.5 Algorithm*

1. INTRODUCTION

Land transportation is fundamental to daily mobility, but its growth in Indonesia has been accompanied by a significant rise in traffic violations and accidents [1], [2]. Common infractions such as failure to wear a helmet, using a mobile phone while driving, and ignoring traffic signs endanger all road users [3], [11]. In Pagar Alam City, this issue is particularly pressing. Data from the "Operasi Keselamatan Musi 2024" revealed 385 traffic tickets issued in just two weeks [4], contributing to a massive annual accumulation of ticketing data at the District Attorney's Office. This raw data, if left unprocessed, becomes an administrative burden rather than a strategic asset. The core problem is that manual analysis is inefficient for such large volumes, preventing law enforcement from identifying critical patterns and forcing them into a reactive, rather than proactive, stance on traffic safety [5][6].

To transform this data into actionable intelligence, data mining techniques are essential. Specifically, classification is required to systematically categorize the vast number of violation cases into distinct, predefined groups based on their attributes. This process is crucial for uncovering hidden relationships between variables (e.g., vehicle type, location, and type of offense), which is impossible to achieve through manual review. Previous research has consistently demonstrated the power of classification in similar domains, such as identifying crime hotspots and predicting accident severity, proving it to be a valuable tool for strategic law enforcement. By classifying ticketing data, authorities can move beyond simple statistics to understand the underlying drivers of traffic violations [9][14].

Several algorithms can be used for classification, but for this study, the C4.5 algorithm was chosen due to its distinct advantages in this specific context. While other methods like Naïve Bayes are computationally fast, they operate on a strict assumption of feature independence, which is often not true for traffic data where attributes like vehicle type and violated article are correlated. Similarly, powerful algorithms like Support Vector

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

Machines (SVM) can yield high accuracy but often function as "black boxes," making their results difficult for non-technical stakeholders, like police officers, to interpret and trust [21]. The C4.5 algorithm, in contrast, builds a decision tree that generates transparent, human-readable IF-THEN rules. This high level of interpretability is a key requirement, as it allows law enforcement officials to understand precisely why a particular decision was made by the model. Furthermore, C4.5 is robust in handling both numerical and categorical data and can manage missing values, which are common in real-world administrative datasets [7], [8].

Therefore, this research aims to implement the C4.5 algorithm to develop a highly accurate and interpretable classification model for traffic ticket data from the Pagar Alam District Attorney's Office. The objective is not merely to classify data, but to create a data-driven decision support tool[12][13]. This model will identify the most significant factors influencing different types of violations, providing actionable insights. The ultimate goal is to equip law enforcement and related institutions with the intelligence needed to design targeted and proactive strategies such as optimizing patrol deployments or launching focused public awareness campaigns to improve traffic law enforcement and enhance public safety.

2. METHODOLOGY

This research applies data mining techniques following the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework as its methodological foundation [10][15].

This structured approach ensures a comprehensive and systematic process, consisting of six distinct phases:

a. Business Understanding Phase

The primary objective of this phase was to define the research goals from a practical perspective. The core problem identified was the inefficiency of manually processing a large volume of traffic ticket data at the Pagar Alam District Attorney’s Office, which hindered strategic law enforcement. Therefore, the goal was set to design and build a classification model capable of automatically identifying the most frequent patterns of traffic violations. The success criterion for this project is the development of a high-accuracy model whose outputs (violation patterns) can serve as a data-driven foundation for policy-making.

b. Data Understanding Phase

This phase began with data collection. The dataset used is secondary data, comprising 5,188 records of traffic violation cases from 2024, obtained directly from the Pagar Alam District Attorney’s Office [16]. An initial data exploration was conducted to familiarize ourselves with the data's structure and content. This involved analyzing the key attributes: Pasal (Article of Law), Barang Bukti (Evidence), Jenis Kendaraan (Vehicle Type), and Denda (Fine), which serves as the target label for classification. Data quality was assessed by checking for missing values and inconsistencies to understand the scope of data preparation needed.

c. Data Preparation Phase

This phase focused on preparing the final dataset for modeling, with all tasks conducted using the built-in operators within the RapidMiner Studio software. The process involved data cleaning, where records with incomplete or inconsistent entries were reviewed to ensure data integrity. Following this, attribute selection was performed to designate Pasal, Barang Bukti, Jenis Kendaraan, and Denda as the final features for model training, while excluding irrelevant information. Lastly, no significant data transformation was necessary, as the C4.5 algorithm effectively handles both categorical and numerical data.

d. Modeling Phase

In this phase, the classification model was built and tested. The C4.5 algorithm (implemented as the "Decision Tree" operator in RapidMiner) was selected as the modeling technique. The Gain Ratio was chosen as the splitting criterion to prevent bias towards attributes with a large number of unique values. To ensure the model's performance is robust and generalizable, a 10-fold Cross-Validation technique was applied[17]. This method partitions the dataset into ten subsets, using nine for training and one for testing, rotating this process ten times to yield a reliable performance average. The entire modeling

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

and validation process was executed within the RapidMiner environment.

e. Evaluation Phase

The model's performance was evaluated based on quantitative metrics derived from the cross-validation process. The primary metric was accuracy, which measures the overall correctness of the classifications. A confusion matrix was also generated to provide a detailed breakdown of the model's performance for each class (fine amount), allowing for the calculation of class precision and recall. The model was deemed successful as it achieved an exceptionally high accuracy of 99.75%, aligning with the initial project objectives.

f. Deployment Phase

In the context of this research, the "deployment" is the delivery of the validated classification model and the actionable insights it generates. The final output is not a live system but a comprehensive report detailing the model's structure, performance, and the interpretable IF-THEN rules derived from the decision tree. These findings, which highlight the most significant factors contributing to traffic violations, are presented to the Pagar Alam District Attorney's Office as a proof-of-concept for a future decision support tool.

3. RESULTS AND DISCUSSIONS

Using the C4.5 algorithm and the CRISP-DM approach, the traffic violation cases at the Pagar Alam District Attorney's Office were classified based on the data that had been cleaned and prepared [18].

3.1 Process of the C4.5 Algorithm

The C4.5 algorithm was applied in this study to build a classification model through data mining. A decision tree was constructed from the tabular data.

The process began by calculating the number of cases based on features such as article of violation, evidence, type of vehicle, and imposed fine. Next, the entropy and gain values for each feature were calculated. Because the gain ratio method was used, the calculation continued to determine the highest gain ratio in accordance with the steps of the C4.5 algorithm.

Step 1: Calculate the total entropy from the traffic violation case data at the Pagar Alam District Attorney's Office.

EntropyTotal

$(99000^{2880}79000^{1968}119000^{222}139000^{118})$

=(-2880/5188*log2((2880/5188))+(1967/5188*log2(1967/5188))+(223/5188*log2(223/5188))+(-118/5188*log2 (118/5188)))

= 1,321152977

Step 2: The entropy is calculated for each attribute under Article.

Entropy 291(1) JO PSL 106 (8)

$(99000^{2352}79000^{1679}119000^{0}139000^{0})$

=(-2352/4031*log2((2352/4031))+(-1679/4031*log2(1679/4031))+(-0/4031*log2(0/4031))+(-0/4031*log2 (0/4031)))

= 0

Entropy 291(2) JO PSL 106 (8)

$(99000^{366}79000^{288}119000^{0}139000^{0})$

=(-366/654*log2((366/654))+(-288/654*log2(288/654))+(-0/654*log2(0/654))+(-0/654*log2 (0/654)))

= 0

Entropy 291(1)(2) JO PSL 106 (8)

$(99000^{162}79000^{0}119000^{0}139000^{0})$

=(-162/162*log2((162/162))+(-0/162*log2 (0/162))+(-0/162*log2 (0/162))+(-0/162*log2(0/162)))

= 0

Entropy 281 JO PSL 77 (1)

$(99000^{0}79000^{0}119000^{123}139000^{0})$

=(-0/123*log2((0/123))+(-0/123*log2(0/123))+(-123/123*log2(123/123))+(-0/123*log2 (0/123)))

= 0

Entropy 288(1) JO 106 (5)a

$(99000^{0}79000^{0}119000^{0}139000^{59})$

= (-0/59*log2((0/59))+(-0/59*log2(0/59))+(-0/59*log2 (0/59))+(-59/59*log2(59/59)))

= 0

Entropy 285(1) JO PSL 106 (3)

$(99000^{0}79000^{0}119000^{42}139000^{0})$

= (-0/42*log2((0/42))+(-0/42*log2(0/42))+(-42/42*log2 (42/42))+(-0/42*log2(0/42)))

= 0

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

Entropy 287(1) JO PSL 106 (4)
$(99000^{0}79000^{0}119000^{34}139000^{0})$
= (-0/34*log2((0/34))+(-0/34*log2(0/34))+(-34/34*log2 (34/34))+(-0/34*log2(0/34)))
= 0

Entropy 288(2) JO PSL 106 (5)
$(99000^{0}79000^{0}119000^{0}139000^{34})$
= (-0/34*log2((0/34))+(-0/34*log2(0/34))+(-0/34*log2 (0/34))+(-34/34*log2(34/34)))
= 0

Entropy 288(3) JO 106 (5c
$(99000^{0}79000^{0}119000^{0}139000^{22})$
= (-0/22*log2((0/22))+(-0/22*log2(0/22))+(-0/22*log2 (0/22))+(-22/22*log2(22/22)))
= 0

Entropy 289(1) JO PSL 106 (3)
$(99000^{0}79000^{0}119000^{11}139000^{0})$
= (-0/11*log2((0/11))+(-0/11*log2(0/11))+(-11/11*log2 (11/11))+(-0/11*log2(0/11)))
= 0

Entropy 280 JO 68 (1)
$(99000^{0}79000^{0}119000^{5}139000^{0})$
=(-0/5*log2((0/5))+(-0/5*log2(0/5))+(-5/5*log2 (5/5))+(-0/5*log2(0/5)))
= 0

Entropy 286 JO PSL 106
$(99000^{0}79000^{0}119000^{4}139000^{0})$
=(-0/4*log2((0/4))+(-0/4*log2(0/4))+(-4/4*log2 (4/4))+(-0/4*log2(0/4)))
= 0

Entropy 307 JO PSL 169 (1)
$(99000^{0}79000^{0}119000^{3}139000^{0})$
=(-0/3*log2((0/3))+(-0/3*log2(0/3))+(-3/3*log2 (3/3))+(-0/3*log2(0/3)))
= 0

Entropy 293 JO PSL 77 (1)
$(99000^{0}79000^{0}119000^{1}139000^{0})$
=(-0/1*log2((0/1))+(-0/1*log2(0/1))+(-1/1*log2 (1/1))+(-0/1*log2(0/1)))
= 0

Entropy 300 JO PSL 124 (1)
$(99000^{0}79000^{0}119000^{0}139000^{1})$
=(-0/1*log2((0/1))+(-0/1*log2(0/1))+(-0/1*log2 (0/1))+(-1/1*log2(1/1)))
= 0

Entropy 305 JO 165
$(99000^{0}79000^{0}119000^{0}139000^{1})$
=(-0/1*log2((0/1))+(-0/1*log2(0/1))+(-0/1*log2 (0/1))+(-1/1*log2(1/1)))
= 0

Entropy 308 a JO 173 (1)a
$(99000^{0}79000^{0}119000^{0}139000^{1})$
=(-0/1*log2((0/1))+(-0/1*log2(0/1))+(-0/1*log2 (0/1))+(-1/1*log2(1/1)))
= 0

Step 3: Calculate the gain value for the Article attribute. Gain for Article
=1,321152977-((4031/5188*0)+(654/5188*0)+(162/5188*0)+(123/5188*0)+(59/5188*0)+(42/5188*0)+(34/5188*0)+(34/5188*0)+(22/5188*0)+(11/5188*0)+(5/5188*0)+(4/5188*0)+(3/5188*0)+(1/5188*0)+(1/5188*0)+(1/5188*0)+(1/518*0))
= 1,321152977

Step 4: Calculate the Split Information value for the Article attribute.
Split Information (Total Article)
= (-4031/5188*log2(4031/5188))+(654/5188*log2(654/5188))+(-162/5188*log2(162/5188))+(-123/5188*log2(123/5188))+(-59/5188*log2(59/5188))+(42/5188*log2(42/5188))+(-34/5188*log2(34/5188))+(-34/5188*log2(34/5188))+(22/5188*log2(22/5188))+(-11/5188*log2(11/5188))+(-5/5188*log2(5/5188))+(4/5188*log2(4/5188))+(3/5188*log2(3/5188))+(1/5188*log2(1/5188))+(1/5188*log2(1/5188))+(1/5188*log2(1/5188))+(-1/5188*log2(1/5188))
= 1,25404249

Step 5, calculate the Gain Ratio for Article
Gain Ratio (Total Article)
= 1,321152977 / 1,25404249
= 1,053515321

For each attribute used, recalculate according to the previous steps until you obtain the gain ratio value as shown in Table 1.

Table 1. Calculation Results of Root/Node 1

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

Atribut	Nilai	Jumlah	99000	79000	119000	139000	Entropy	Gain	SplitInfo	Gainratio
Total		5188	2880	1967	223	118	1,321152977			
ARTICLE	291(1) JO PSL 106 (8)	4031	2352	1679	0	0	0	1,32115297	1,254704249	1,053515321
	291(2) JO PSL 106 (8)	654	366	288	0	0	0			
	291(1)(2) JO PSL 106 (8)	162	162	0	0	0	0			
	281 JO PSL 77 (1)	123	0	0	123	0	0			
	288(1) JO 106 (5)a	59	0	0	0	59	0			
	285(1) JO PSL 106 (3)	42	0	0	42	0	0			
	287(1) JO PSL 106 (4)	34	0	0	34	0	0			
	288(2) JO PSL 106 (5)	34	0	0	0	34	0			
	288(3) JO 16 (5c	22	0	0	0	22	0			
	289(1) JO PSL 106 (3)	11	0	0	11	0	0			
	280 JO 68 (1)	5	0	0	5	0	0			
	286 JO PSL 106	4	0	0	4	0	0			
	307 JO PSL 169 (1)	3	0	0	3	0	0			
	293 JO PSL 77 (1)	1	0	0	1	0	0			
	300 JO PSL 124 (1)	1	0	0	0	1	0			
	305 JO 165	1	0	0	0	1	0			
	308 a JO 173 (1)a	1	0	0	0	1	0			
Evidence	Kendaraan	2956	2834	0	101	21	0	1,049417774	1,323836216	0,792709673
	Stnk	1834	41	1622	111	60	0,685628265			
	Sim C	348	5	326	9	8	0,437701129			
	Sim A	44	0	20	0	24	0			
	Sim B1umum	6	0	0	1	5	0			
Type of Vehicle	Sepeda motor	5014	2880	1925	163	46	1,212478089	0,14934013	0,2581599764	0,577495164
	Mobil	118	0	33	40	45	0			
	Pick up	24	0	4	6	14	0			
	Bus	21	0	3	10	8	0			
	Truck	11	0	3	3	5	0			

The results of the manual calculation show that the attribute “article” has the highest gain ratio of 1.0535 and was selected as the root of the decision tree, as presented in the following figure:

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

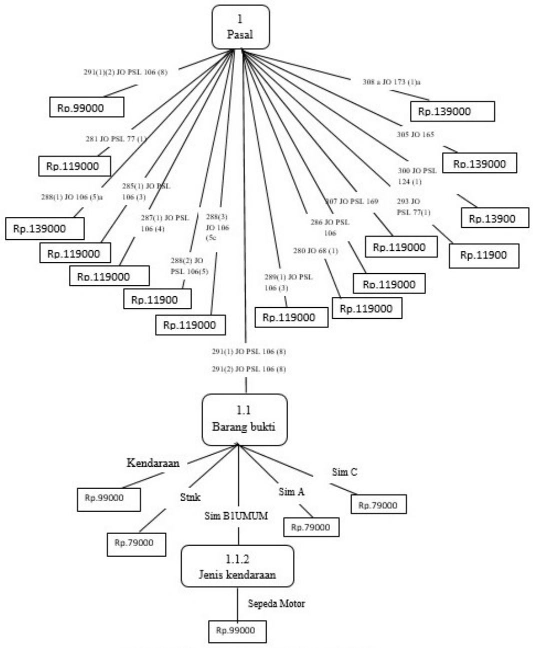

Figure 1. Decision Tree from Manual Calculation

3.2 Results of C4.5 Algorithm Measurement Using RapidMiner

The RapidMiner application, which implements the C4.5 algorithm, was then used to test the data that had undergone preprocessing. The purpose of the testing was to obtain the most accurate results by using the X-Cross Validation operator model and the gain percentage calculation method [19]. The scheme for the X-Cross Validation operator used is presented in Figure 2.

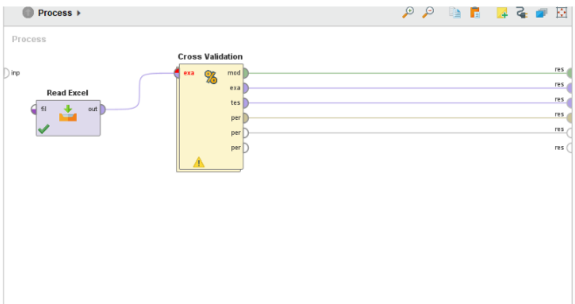

Figure 2. Display of Read Excel and Validation on the main process

Search for "X-Validation" in the Operators column. Then, drag it to the Process area and double-click to display the training and testing process, as shown in Figure 3.

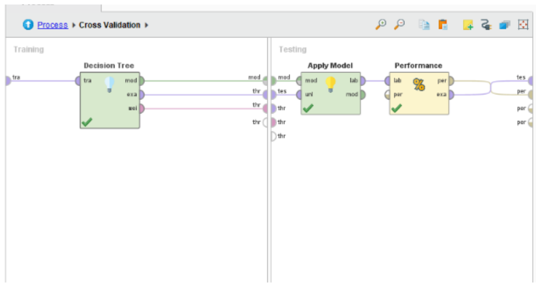

Figure 3. Display of validation process

The test data processed using C4.5 in RapidMiner shows an accuracy of 99.75%, with the following confusion matrix:

Table 2. Confusion Matrix Results

	True 139000	True 79000	True 119000	True 99000	Class precision
Pred. 139000	115	0	2	0	98.29%
Pred. 79000	2	1967	3	0	99.75%
Pred. 119000	1	1	213	0	99.07%
Pred. 99000	0	0	4	2880	99.86%
Class recall	97.46 %	99.95 %	95.95 %	100.00 %	

The test results using the C4.5 algorithm model in RapidMiner produced an accuracy value as shown in Figure 4

1. Accuracy

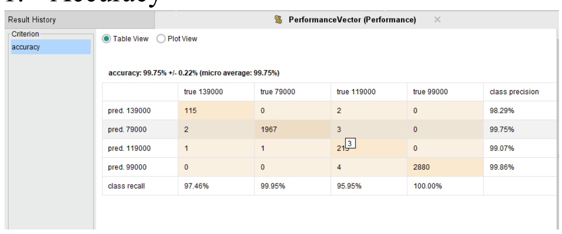

Figure 4. Accuracy Result

The test results using the C4.5 algorithm model in RapidMiner produced an accuracy value as shown in Figure 4. The measurement results show an accuracy of 99.75% [20]. The prediction values are as follows: For prediction 139.000: true 139.000 became 115 predictions, true 79.000 became 0 predictions, true 119.000 became 2 predictions, true 99.000 became 0 predictions, with a precision value of 98.29%. For prediction 79.000: true 139.000 became 2 predictions, true 79.000 became 1967 predictions, true 119.000 became 3 predictions, true 99.000 became 0 predictions, with a

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

precision value of 99.75%. For prediction 119.000: true 139.000 became 1 prediction, true 79.000 became 1 prediction, true 119.000 became 213 predictions, true 99.000 became 0 predictions, with a precision value of 99.07%. For prediction 99.000: true 139.000 became 0 predictions, true 79.000 became 0 predictions, true 119.000 became 4 predictions, true 99.000 became 2880 predictions, with a precision value of 99.86%. The highest recall is for class 99.000 reaching 100.00%. The recall for class 139.000 is 97.46%, recall for class 79.000 is 99.95%, and recall for class 119.000 is 95.95%. The test results carried out using RapidMiner with the C4.5 algorithm are shown in Figure 5 below.

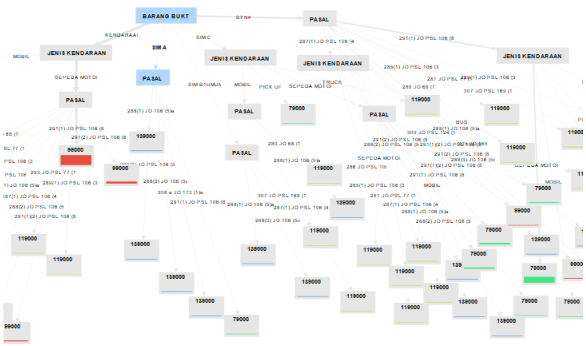

Figure 5. Decision Tree Results C4.5

Based on the decision tree results above, testing using RapidMiner produced the rules presented in Figure 6 below.

Tree

```
BARANG BUKTI = KENDARAAN
|   JENIS KENDARAAN = MOBIL
|   |   PASAL = 281 JO PSL 77 (1): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 288(1) JO 106 (5)a: 139000 {139000=1, 79000=0, 119000=0, 99000=0}
|   JENIS KENDARAAN = SEPEDA MOTOR
|   |   PASAL = 280 JO 68 (1): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 281 JO PSL 77 (1): 119000 {139000=0, 79000=0, 119000=48, 99000=0}
|   |   PASAL = 285(1) JO PSL 106 (3): 119000 {139000=0, 79000=0, 119000=30, 99000=0}
|   |   PASAL = 286 JO PSL 106: 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 287(1) JO PSL 106 (4): 119000 {139000=0, 79000=0, 119000=18, 99000=0}
|   |   PASAL = 288(1) JO 106 (5)a: 139000 {139000=18, 79000=0, 119000=0, 99000=0}
|   |   PASAL = 288(2) JO PSL 106 (5): 139000 {139000=2, 79000=0, 119000=0, 99000=0}
|   |   PASAL = 289(1) JO PSL 106 (3): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 291(1) JO PSL 106 (8): 99000 {139000=0, 79000=0, 119000=0, 99000=2352}
|   |   PASAL = 291(1)(2) JO PSL 106 (8): 99000 {139000=0, 79000=0, 119000=0, 99000=116}
|   |   PASAL = 291(2) JO PSL 106 (8): 99000 {139000=0, 79000=0, 119000=0, 99000=366}
|   |   PASAL = 293 JO PSL 77 (1): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
BARANG BUKTI = SIM A
|   PASAL = 288(1) JO 106 (5)a: 139000 {139000=16, 79000=0, 119000=0, 99000=0}
|   PASAL = 288(2) JO PSL 106 (5): 139000 {139000=1, 79000=0, 119000=0, 99000=0}
|   PASAL = 288(3) JO 106 (5c: 139000 {139000=6, 79000=0, 119000=0, 99000=0}
|   PASAL = 291(1) JO PSL 106 (8): 79000 {139000=0, 79000=20, 119000=0, 99000=0}
|   PASAL = 308 a JO 173 (1)a: 139000 {139000=1, 79000=0, 119000=0, 99000=0}
BARANG BUKTI = SIM B1UMUM
|   PASAL = 288(1) JO 106 (5)a: 139000 {139000=2, 79000=0, 119000=0, 99000=0}
|   PASAL = 288(3) JO 106 (5c: 139000 {139000=3, 79000=0, 119000=0, 99000=0}
|   PASAL = 307 JO PSL 169 (1): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
```

```
BARANG BUKTI = SIM C
|   JENIS KENDARAAN = MOBIL
|   |   PASAL = 280 JO 68 (1): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 287(1) JO PSL 106 (4): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 288(1) JO 106 (5)a: 139000 {139000=1, 79000=0, 119000=0, 99000=0}
|   JENIS KENDARAAN = PICK UP: 79000 {139000=0, 79000=1, 119000=0, 99000=0}
|   JENIS KENDARAAN = SEPEDA MOTOR
|   |   PASAL = 281 JO PSL 77 (1): 119000 {139000=0, 79000=0, 119000=2, 99000=0}
|   |   PASAL = 285(1) JO PSL 106 (3): 119000 {139000=0, 79000=0, 119000=3, 99000=0}
|   |   PASAL = 287(1) JO PSL 106 (4): 119000 {139000=0, 79000=0, 119000=1, 99000=0}
|   |   PASAL = 288(1) JO 106 (5)a: 139000 {139000=5, 79000=0, 119000=0, 99000=0}
|   |   PASAL = 288(2) JO PSL 106 (5): 139000 {139000=2, 79000=0, 119000=0, 99000=0}
|   |   PASAL = 291(1) JO PSL 106 (8): 79000 {139000=0, 79000=271, 119000=0, 99000=0}
|   |   PASAL = 291(1)(2) JO PSL 106 (8): 99000 {139000=0, 79000=0, 119000=0, 99000=5}
|   |   PASAL = 291(2) JO PSL 106 (8): 79000 {139000=0, 79000=54, 119000=0, 99000=0}
|   JENIS KENDARAAN = TRUCK: 119000 {139000=0, 79000=0, 119000=1, 99000=0}
BARANG BUKTI = STNK
|   PASAL = 280 JO 68 (1): 119000 {139000=0, 79000=0, 119000=3, 99000=0}
|   PASAL = 281 JO PSL 77 (1): 119000 {139000=0, 79000=0, 119000=72, 99000=0}
|   PASAL = 285(1) JO PSL 106 (3): 119000 {139000=0, 79000=0, 119000=9, 99000=0}
|   PASAL = 286 JO PSL 106: 119000 {139000=0, 79000=0, 119000=3, 99000=0}
|   PASAL = 287(1) JO PSL 106 (4)
|   |   JENIS KENDARAAN = BUS: 119000 {139000=0, 79000=0, 119000=2, 99000=0}
|   |   JENIS KENDARAAN = MOBIL: 79000 {139000=0, 79000=1, 119000=0, 99000=0}
|   |   JENIS KENDARAAN = SEPEDA MOTOR: 119000 {139000=0, 79000=0, 119000=10, 99000=0}
|   PASAL = 288(1) JO 106 (5)a: 139000 {139000=16, 79000=0, 119000=0, 99000=0}
|   PASAL = 288(2) JO PSL 106 (5): 139000 {139000=29, 79000=0, 119000=0, 99000=0}
|   PASAL = 288(3) JO 106 (5c: 139000 {139000=13, 79000=0, 119000=0, 99000=0}
|   PASAL = 289(1) JO PSL 106 (3): 119000 {139000=0, 79000=0, 119000=10, 99000=0}
|   PASAL = 291(1) JO PSL 106 (8): 79000 {139000=0, 79000=1388, 119000=0, 99000=0}
|   PASAL = 291(1)(2) JO PSL 106 (8): 99000 {139000=0, 79000=0, 119000=0, 99000=41}
|   PASAL = 291(2) JO PSL 106 (8): 79000 {139000=0, 79000=233, 119000=0, 99000=0}
|   PASAL = 300 JO PSL 124 (1): 139000 {139000=1, 79000=0, 119000=0, 99000=0}
|   PASAL = 305 JO 165: 139000 {139000=1, 79000=0, 119000=0, 99000=0}
|   PASAL = 307 JO PSL 169 (1): 119000 {139000=0, 79000=0, 119000=2, 99000=0}
```

Figure 6. Decision Tree Rules

4. CONCLUSIONS

The implementation of the C4.5 algorithm to classify ticket case data at the Pagar Alam District Attorney's Office can be carried out by building a decision tree model based on the attributes of article, type of vehicle, evidence, and fine. For classification, entropy and gain ratio are calculated. According to the test results, the C4.5 algorithm has the ability to classify ticket data with an accuracy of 99.75%. From the analysis results, it is known that the most frequent traffic violation is Article 291 (1) JO PSL 106 (8). This information can be an important reference for the relevant authorities to take more precise preventive and enforcement measures against the most common types of violations. From the available test results and conclusions, there are several suggestions for further development. Determining the correct labels, especially between 'ticketed' and 'not ticketed,' is important to improve the accuracy of the results. Accurate label determination is very important so that the classification process can run more optimally. Second, it is recommended to conduct testing using other classification methods as a comparison. This aims to evaluate the performance of the C4.5 algorithm and find out whether there are alternative methods with better accuracy,

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

precision, and efficiency. Thus, future system development can be further improved in terms of both accuracy and reliability.

REFERENCES

[1] E. Dewi, S. Mulyani, A. M. Yusup, A. K. Tisna, F. A. Fauzi, I. B. Seta, R. Khairunas, and W. Ardiansyah, "Clustering wilayah dan pelanggaran berkendaraan menggunakan algoritma K-Means pada data Satlantas Polres Tasikmalaya Kota," *e-Jurnal JUSITI (Jurnal Sistem Informasi dan Teknologi Informasi)*, vol. 8-1, no. 1, pp. 1–11, 2019. doi: 10.36774/jusiti.v8i1.595.

[2] N. L. Mauliddiyah, "Analisis cluster dan pemodelan pelanggaran lalu lintas menggunakan metode K-Means clustering studi kasus Kota Bandung," p. 6, 2021.

[3] M. Suttri, "Pelaksanaan Electronic Traffic Law Enforcement (ETLE) Mobile di Kota Pekanbaru (Studi pada Ditlantas Polda Riau)," *J. Public Adm. Rev.*, vol. 1, no. 1, pp. 389–406, 2024.

[4] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi melalui pohon keputusan berbasis algoritme C4.5," *Prisma: Prosiding Seminar Nasional Matematika*, vol. 3, pp. 64–71, 2020.

[5] R. D. B. Febriyanto, L. Handoko, and H. Aisyah, "Klasifikasi penyebaran Covid-19 menggunakan pendahuluan studi literatur," vol. 4, pp. 23–35, 2021.

[6] F. F. Nugraha, I. Sunandar, and C. Julian, "Penerapan data mining melalui metode klasifikasi menggunakan algoritma C4.5," *Teknologi*, vol. 7, no. Mar., pp. 10–20, 2022. doi: 10.35957/jatisi.v9i4.2399.

[7] A. Lazuardy, F. A. Homaidi, and A. Lutfi, "Perbandingan metode klasifikasi pengujian kendaraan bermotor melalui algoritma C4.5 dan Naïve Bayes," vol. 19, no. Sep., pp. 140–154, 2024. doi: 10.30587/e-link.v19i2.8113

[8] T. R. Matondang, Y. R. Nasution, Armansyah, and M. Furqan, "Penerapan algoritma C4.5 pada klasifikasi status gizi balita," *J. Fasilkom*, vol. 14, no. 1, pp. 216–225, 2024. doi: 10.37859/jf.v14i1.6941.

[9] M. Soleh, "Algoritma C4.5 dalam mengklasifikasi tindak kejahatan yang sering terjadi di Kecamatan Siantar Barat Pematangsiantar," *TIN: Terapan Informatika Nusantara*, vol. 2, no. 5, pp. 335–340, 2021.

[10] Sutisna and N. M. Yuniar, "Klasifikasi kualitas air bersih menggunakan metode Naïve Bayes," *J. Sains dan Teknologi*, vol. 5, no. 1, pp. 243–246, 2023.

[11] M. S. Othman, R. A. Rahman, and A. R. Hamdan, "A data mining approach for classification of traffic violations types," *Int. J. Adv. Intell. Inform.*, vol. 7, no. 3, pp. 282–291, 2021, doi: 10.26555/ijain.v7i3.708.

[12] W. Xuanyuan, Y. Zhang, and C. Liu, "Application of C4.5 algorithm in insurance and financial services using data mining methods," *Mobile Information Systems*, 2022, Article ID 5670784, doi: 10.1155/2022/5670784.

[13] G. I. Salama, M. B. Abdelhalim, and M. A. E. Zeid, "Very fast C4.5 decision tree algorithm," *Applied Artificial Intelligence*, vol. 32, no. 2, pp. 119–137, 2018, doi: 10.1080/08839514.2018.1447479.

[14] F. Martínez-Plumed et al., "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[15] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications – A holistic extension to the CRISP-DM model," *Procedia CIRP*, vol. 79, pp. 403–408, 2019, doi: 10.1016/j.procir.2019.02.106.

[16] W. Y. Ayele, "Adapting CRISP-DM for idea mining: A data mining process for generating ideas using a textual dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 6, pp. 23–32, 2020, doi: 10.14569/IJACSA.2020.0110603.

[17] P. Chen, "Design of intelligent traffic simulation system based on decision tree algorithm," in *Proc. 2nd Int. Conf. Artificial Intelligence and Information Systems*, 2021, pp. 1–3, doi: 10.1145/3469213.3472778.

[18] M. Al-Barrak and M. Al-Razgan, "Predicting students final GPA using decision trees: A case study," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 528–533, 2016, doi: 10.7763/IJIET.2016.V6.745.

[19] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65

vector machines," in *Proc. Int. MultiConf. Engineers and Computer Scientists*, vol. 1, pp. 442–447, 2011, doi: 10.1109/WISA.2009.36.

[20] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011, doi: 10.1002/widm.8.

[21] B. S. D. Nugraha, S. Defit, and G. W. Nurcahyo, "Algoritma C4.5 untuk prediksi bimbingan siswa berdasarkan tipologi Hippocrates-Galenus," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 11, no. 1, pp. 1–8, 2023, doi: 10.21063/jtif.2023.V11.1.1-8.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.57-65