JURNAL INFORMATIKA
Volume 13 No. 1 April 2026, pp. 54-61
ISSN (Online) 2528-2247 | ISSN (Print) 2355-6579
DOI: https://doi.org/10.31294/informatika.v13i1.11944

Comparison of Logistic Regression and Random Forest Methods in Predicting Vehicle Tax Payment Compliance

Khairul Fuadi [1*], Taufiq[1], Arnawan Hasibuan[1], Dahlan Abdullah[1], Nurdin[1]

[1]Teknologi Informasi, Universitas Malikussaleh
Jl. Kampus Unimal Bukit Indah, Blang Pulo, Muara Satu, Kabupaten Aceh Utara, Indonesia

Correspondence e-mail: khairul.237110201007@mhs.unimal.ac.id

Submission:	Revision:	Acceptance:	Available Online:
26-01-2026	05-02-2026	19-02-2026	06-03-2026

Abstract - Motor vehicle tax is a major source of Regional Original Income (PAD). However, the level of motor vehicle tax payment compliance in North Aceh Regency is still suboptimal, particularly related to late payments. A data-driven approach is needed to predict and understand taxpayer compliance patterns more accurately. This study aims to compare the performance of the Logistic Regression and Random Forest methods in predicting motor vehicle tax payment compliance, as well as to identify factors that influence taxpayer compliance behavior at the North Aceh Samsat (Sat). This study uses secondary data in the form of motor vehicle tax payment transactions at the North Aceh Samsat for the 2022–2024 period, totaling 100,000 observations. The response variable is the tax payment compliance status (compliant and non-compliant), while the predictor variables include vehicle age, type of ownership, vehicle type, and vehicle brand. The data is divided into 70% training data and 30% testing data. The performance evaluation model is conducted using accuracy, precision, recall, and Area Under Curve (AUC) metrics. The analysis results show that Random Forest has better predictive performance than Logistic Regression, with higher accuracy and AUC values. Vehicle age and type of ownership are the most influential variables in predicting tax payment compliance, while vehicle brand has a relatively smaller influence. Logistic Regression provides a clear interpretation of the variable relationship, but has lower discrimination ability than Random Forest. Random Forest has proven to be more effective as a prediction model for motor vehicle tax payment compliance at the North Aceh Samsat. The application of machine learning-based predictive models has the potential to support more targeted policy making in an effort to improve motor vehicle tax payment compliance, especially in reducing late payments.

Keywords: Tax; Motor Vehicle Tax; Logistic Regression; Random Forest; Classification

1. Introduction

Motor vehicle tax is a key component of Regional Original Income (PAD), playing a crucial role in infrastructure development, public services, and regional government administration. However, in North Aceh Regency, the level of motor vehicle taxpayer compliance is still far from optimal (Nisa, 2024). Taxes are a major source of regional revenue, so clear targets are needed to achieve them (Abd Madjid et al., 2024). Motor vehicle tax is a primary source of Regional Original Revenue (PAD), supporting the fiscal independence of local governments. However, in practice, there remains a tax gap, the difference between the potential tax revenues and the actual amount paid by taxpayers (Wasana et al., 2021). Regional tax authorities through Samsat are trying to minimize this gap through education and the use of information technology (Maulidya, 2025). One type of regional tax that contributes significantly to Regional Original Income (PAD) is the Motor Vehicle Tax (PKB) (Damayanti et al., 2023).

Logistic regression is a statistical method that is often used to predict binary variables, such as taxpayer compliance or non-compliance (Wasana et al., 2021). This method has advantages in interpreting results and identifying the influence of independent variables on the dependent variable. Classification with Logistic Regression is carried out using a mathematical model that connects the response variable with the independent variable (Yuniarsyih R.A et al., 2025).

Random Forest offers advantages in handling complex data and producing predictions with high accuracy, precision, recall, and AUC. Random Forest is capable of processing data with a large number of variables and detecting interactions between those variables. Random Forest classifies objects by combining multiple decision trees (Hulaifah Al Abrori & Subhiyakto, 2025). This algorithm works by building a large number of decision trees during the training process and combining the results to produce more accurate and stable predictions (Nugroho & Harini, 2024). Random Forest offers advantages in handling complex data and producing predictions with high accuracy, precision, recall, and AUC. Random Forest is capable of processing data with a large number of variables and detecting interactions between those variables. Random Forest classifies objects by combining multiple decision trees. Random Forest was first introduced by Leo Breiman in 2001 as a development of the bagging and decision tree methods (Suci Amaliah et al., 2022). With its main advantages of predictive stability and high accuracy, Random Forest is often used as the algorithm of choice in various data-driven projects (Purnama et al., 2024)

Based on Law Number 28 of 2009 and Aceh Qanun Number 2 of 2012, Aceh Tax is a mandatory contribution to the Aceh Government owed by individuals or bodies that is mandatory based on statutory regulations, without receiving direct compensation and is used for the needs of Aceh for the greatest prosperity of the people. Motor vehicle tax is a tax on ownership and/or control of motor vehicles. Based on Law Number 28 of 2009 and Aceh Qanun Number 2 of 2012, Aceh Tax is a mandatory contribution to the Aceh Government owed by individuals or bodies that is mandatory based on statutory regulations, without receiving direct compensation and is used for the needs of Aceh for the greatest prosperity of the people. Motor vehicle tax is a tax on ownership and/or control of motor vehicles (Baj et al., 2023). The increasing number of vehicles each year will increase taxpayer compliance and local revenue sources through tax collection on motor vehicle owners. Logistic regression is a method that can be used to find the relationship between response variables that are dichotomous (ordinal or nominal scale) or polychotomous (having a nominal or ordinal scale with more than two categories). When the dependent variable has a polychotomous or multinomial scale, multinomial logistic regression can be used (Rahmadani et al., 2023).

The tax administration system at Samsat (State-Owned Tax Office) uses a one-stop system (Satuan Administrasi Manunggal Satu Atap) concept, which aims to speed up and simplify public services (Hasibuan et al., 2025a). However, many motorized vehicles still have outstanding taxes, resulting in a decline in regional revenue (Wati & Saepuloh, 2025). Data mining techniques are one solution for analyzing large amounts of data to obtain relevant information. This study uses the random forest algorithm (Hasibuan et al., 2025). Motor Vehicle Tax (PKB) is a type of regional tax imposed on the ownership and/or control of motorized vehicles, both two-wheeled and four-wheeled, operating on the road. PKB falls into the objective tax category, where the amount is determined based on the taxable object, namely the motorized vehicle, without regard to the subjective circumstances of the owner (Avidaniar Bintary, 2020). Samsat (One-Stop Integrated Administration System) is an administrative system designed to streamline and expedite public services in the payment of motor vehicle taxes. SAMSAT North Aceh is a service unit that provides various related services (Nurahaliza & Mulyadi, 2022). Samsat North Aceh Regency (Apriansyah et al., 2023) is an integrated cooperation system between the National Police, the Aceh Financial Management Agency, and PT. Jasa Raharja, in services to issue STNK (Motor Vehicle Registration Certificate), Motor Vehicle Registration Numbers which are linked to the income of money into the State Treasury either through Motor Vehicle Tax (PKB), Motor Vehicle Transfer Fee, and Mandatory Road Traffic Fund Contribution (SWDKLJJ), at one office called the Joint Samsat Office.

Previous research has shown that applying undersampling, oversampling, and combined sampling to a Logistic Regression model increases sensitivity but decreases accuracy, precision, recall, specificity, and AUC. In the Random Forest model, this effect is only seen with the undersampling scheme. The feature selection process tends to reduce the stability of model performance in certain schemes. Overall, the best model is Logistic Regression with a combined sampling scheme without feature selection (Purwa, 2019).

Further research compared the two methods using weather data in Central Java divided into three different training data proportions: 60%, 70%, and 80%, and evaluated the models using the area under the curve (AUC) value. The average AUC of the Logistic Regression and Random Forest methods were 0.6923 and 0.7419, respectively. Based on the analysis of the two methods, the highest AUC value was obtained from the classification results using the Random Forest method (Larasati, 2023).

The low level of compliance with motor vehicle tax payments at the North Aceh Samsat, which has an impact on optimizing the receipt of Regional Original Income (PAD). Although various efforts have been made to improve taxpayer compliance, there is no data-based predictive approach that is comprehensively able to identify patterns and factors that influence taxpayer behavior. However, to date there has been no comprehensive research in the North Aceh region that directly compares the performance of these two methods in the context of predicting motor vehicle tax compliance. Samsat can implement a data-based early detection system that helps in more strategic and efficient decision making (Awal et al., 2025). This research is crucial because the low level of motor vehicle taxpayer compliance in North Aceh Regency has not been effectively addressed through conventional approaches. Therefore, a data-driven prediction method is needed as a more targeted and evidence-based alternative solution for mapping taxpayer compliance patterns.

2. Research Methods

2.1 Type of Research

This study uses a quantitative approach with a comparative method to compare the performance of the Logistic Regression and Random Forest algorithms in predicting motor vehicle tax compliance. Model performance was evaluated using several classification metrics: accuracy, precision, recall, and area under the

ISSN (Online) 2528-2247 | ISSN (Print) 2355-6579

curve (AUC).

2.2 Research Stages

The research entitled Comparison of Logistic Regression and Random Forest Methods in Predicting Vehicle Tax Payment Compliance at North Aceh Samsat with the research steps carried out can be seen in Figure 1 below.

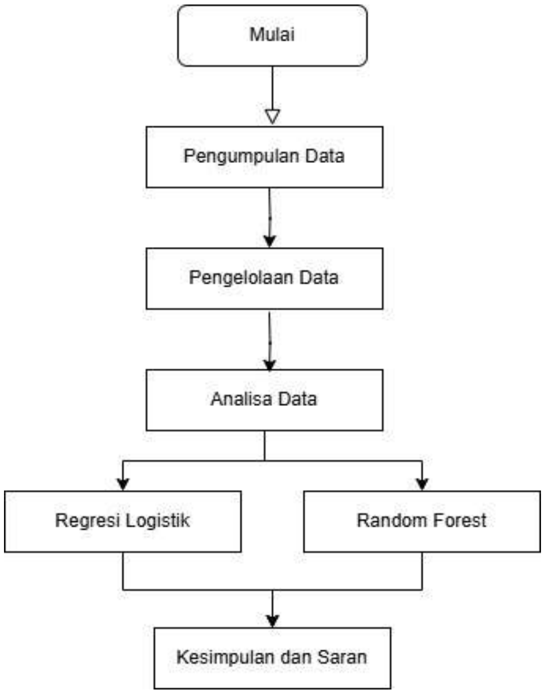


Figure 1. Research Flowchart

2.3 Data Sources and Types

The data used in this study is secondary data obtained from the digital archives of the North Aceh Samsat (Village and Administration Agency). The dataset includes information on motor vehicle tax payment transactions for the 2022–2024 period. The available data includes several attributes such as tax payment date, tax validity period, vehicle type, vehicle year of manufacture, fuel type, vehicle make, and vehicle ownership type. Taxpayer compliance status is determined based on the difference between the payment date and the vehicle tax expiration date. In this study, taxpayers are categorized as compliant if they make payments before or on time during the tax expiration period, while payments after the expiration period are categorized as non-compliant.

2.4 Research Variables

a. Independent Variable

The independent variable is the data used in this study, namely motor vehicle tax payment transaction data obtained from the North Aceh Samsat.

Table 1. Variable

No.	Atribut	Information
1.	Vehicle Age	Identification age based on the year of manufacture of the vehicle, used to analyze how long the vehicle was manufactured and the likelihood of the owner paying taxes.
2.	Vehicle Ownership	Identification type of vehicle ownership (e.g., government, corporate, and private)
3.	Type	The type of identification of the vehicle type (e.g., minibus, pick up, motorbike) can affect the amount of tax to be paid.
4.	Vehicle Type	The type of identification of the vehicle type (e.g., mini bus, pick up, motorbike) can affect the amount of tax to be paid.
5.	Vehicle Brand	Type of identification of vehicle brands (Honda, Toyota, Suzuki)

b. Dependent Variable

The dependent variable is the data used in this study, namely motor vehicle tax payment transaction data obtained from the North Aceh Samsat.

Table 2. Dependen Variable

No.	Attribute	Description
1.	Tax Payment Compliance Status	"Compliant" (if payment is made before or on the expiry date of the tax period) "Non-Compliant" (if vehicle tax has not been paid).

2.5 Data Analysis Methods

a. Logistic Regression

Logistic regression is used to model the relationship between independent variables and a binary dependent variable (compliant or non-compliant). This model uses the logit function to estimate the probability of taxpayer compliance based on the variables used in the study.

b. Random Forest

Random Forest is an ensemble-based machine learning method consisting of multiple decision trees. This model works by combining predictions from multiple decision trees to improve accuracy and reduce the risk of overfitting. This method is also capable of handling complex and nonlinear data relationships.

2.6 Research Stages

Data collection was conducted from the North Aceh Samsat digital archive containing motor vehicle tax payment data. The data then underwent a preprocessing stage that included checking for missing values, consistency checking, and data cleaning. The data was then divided into training and test data for the modeling process using Logistic Regression and Random Forest methods. The performance of both models was evaluated using accuracy, precision, recall, and AUC metrics, then compared to determine the method with the best performance in predicting motor vehicle tax payment compliance.

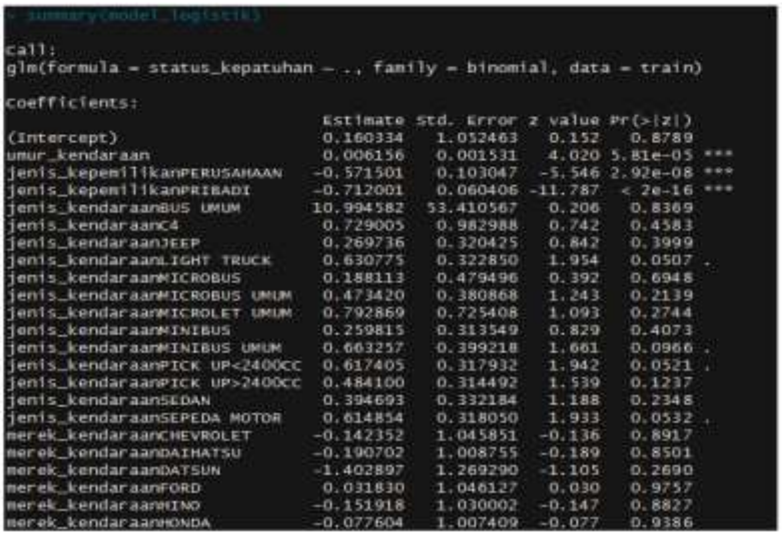
3. Result and Discussion

3.1 Research Data Overview

The data used in this study is secondary data derived from the motor vehicle tax payment transaction database at the North Aceh Samsat (Vehicle Tax Office) for the 2022–2024 period. The research dataset consists of approximately 100,000 observations representing vehicle tax payment transactions and vehicle characteristics. The variables used in the modeling include vehicle age, vehicle ownership type, vehicle type, and vehicle brand. The target variable is tax payment compliance status, categorized as compliant and non-compliant.

3.2 Random Forest Model Evaluation Results

A Random Forest model was built using 500 decision trees with training data representing approximately 70% of the total dataset. Based on testing results, this model achieved an accuracy of 54.71%, a precision of 55.36%, a recall of 49.66%, and an AUC of 0.5572. These results indicate that Random Forest has better classification capabilities than Logistic Regression in predicting motor vehicle tax compliance.



```
> summary(model_logistik)

Call:
glm(formula = status_kepatuhan ~ ., family = binomial, data = train)

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.160334   1.052463   0.152   0.8789
umur_kendaraan                0.006156   0.001531   4.020 5.81e-05 ***
jenis_kepemilikanPERUSAHAAN  -0.571501   0.103047  -5.546 2.92e-08 ***
jenis_kepemilikanPRIBADI     -0.712001   0.060406 -11.787  < 2e-16 ***
jenis_kendaraanBUS UMUM      10.994582  53.410567   0.206   0.8369
jenis_kendaraanC4             0.729005   0.982988   0.742   0.4583
jenis_kendaraanJEEP           0.269736   0.320425   0.842   0.3999
jenis_kendaraanLIGHT TRUCK    0.630775   0.322850   1.954   0.0507 .
jenis_kendaraanMICROBUS       0.188113   0.479496   0.392   0.6948
jenis_kendaraanMICROBUS UMUM  0.473420   0.380868   1.243   0.2139
jenis_kendaraanMICROLET UMUM  0.792869   0.725408   1.093   0.2744
jenis_kendaraanMINIBUS        0.259815   0.313549   0.829   0.4073
jenis_kendaraanMINIBUS UMUM   0.663257   0.399218   1.661   0.0966 .
jenis_kendaraanPICK UP<2400CC 0.617405   0.317932   1.942   0.0521 .
jenis_kendaraanPICK UP>2400CC 0.484100   0.314492   1.539   0.1237
jenis_kendaraanSEDAN          0.394693   0.332184   1.188   0.2348
jenis_kendaraanSEPEDA MOTOR   0.614854   0.318050   1.933   0.0532 .
merek_kendaraanCHEVROLET     -0.142352   1.045851  -0.136   0.8917
merek_kendaraanDAIHATSU      -0.190702   1.008755  -0.189   0.8501
merek_kendaraanDATSUN        -1.402897   1.269290  -1.105   0.2690
merek_kendaraanFORD           0.031830   1.046127   0.030   0.9757
merek_kendaraanHINO          -0.151918   1.030002  -0.147   0.8827
merek_kendaraanHONDA         -0.077604   1.007409  -0.077   0.9386
```

Figure 2. *Regresi Logistik*

Figure 2 shows that vehicle age has a positive and significant effect on taxpayer compliance, with older vehicles tending to be more compliant. Ownership type also has a significant effect, with government-owned vehicles having a higher compliance rate than private and corporate ownership. Most vehicle types and brands do not show a significant effect. While logistic regression performs better than models without predictors, its limitations in capturing nonlinear relationships warrant comparison with the Random Forest method.

3.3 Random Forest Model Construction

In addition to logistic regression, this study used the Random Forest model as a decision tree-based ensemble method. The model was built with 500 trees using the randomForest() function in R and standard parameters. Random Forest applied bootstrapping techniques, random variable selection at each node, and majority voting for classification. Initial model performance was evaluated using the Out-of-Bag (OOB) error value, while the influence of variables was analyzed using the variable importance measure in predicting compliance status.

```
Call:
 randomForest(formula = status_kepatuhan ~ ., data = train, ntree = 500,
mtry = floor(sqrt(ncol(train) - 1)), importance = TRUE)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 45.33%
Confusion matrix:
            PATUH TIDAK PATUH class.error
PATUH       16993       17929   0.5134013
TIDAK PATUH 13804       21274   0.3935230
```

Figure 3. Random Forest

Figure 3 shows that a Random Forest model with 500 trees and mtry = 2 produces an OOB error of 45.33%, reflecting moderate accuracy. The confusion matrix indicates that the model is better at classifying non-compliant taxpayers than compliant ones. This indicates that Random Forest is more sensitive in detecting non-compliance, making it potentially effective as a risk detection tool, although it still has limitations in accurately identifying compliant taxpayers.

3.4 Evaluation of Logistic Regression Model Performance

After the training process, model performance was evaluated using 30% of the total testing data. The evaluation was conducted by comparing the predicted results to the actual compliance status using accuracy, precision, recall, and area under the curve (AUC) metrics. These metrics describe the level of prediction accuracy, the model's ability to identify positive classes, and its ability to distinguish between compliant and non-compliant classes at various thresholds, as presented in Table 3.

Table 3. Results of Logistic Regression Calculation

Akurasi	Precision	Recall	AUC
0,4629 (46,29%)	0,4743 (47,43%)	0,6611 (66,11%)	0,4676

3.5 Application of Logistic Regression Method

Based on research data with a total of 30,000 test data observations, the calculation is as follows:

1. Accuracy (Total correct predictions / Total data)
   The proportion of all correct predictions (both compliant and non-compliant) to the total observations in the testing data. Accuracy is calculated using the equation:
   $$Accuracy = \frac{TP + TN}{Total}$$
   $$Accuracy = \frac{9.940 + 3.947}{30.000} = \frac{13.887}{30.000} = 0{,}4629 = \mathbf{46{,}29}\%$$
2. Precision (Accuracy of positive predictions)
   The proportion of positive predictions (e.g., NON-COMPLIANCE) that accurately reflect the

actual situation. Precision is calculated using the equation:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{9.940}{9.940 + 11.017} = \frac{9.940}{20.957} = 0{,}4743 = \mathbf{47{,}43}\%$$

3. Recall (Ability to detect true positive cases)
   The proportion of all true positive cases that the model successfully identifies as positive. Recall is calculated using the equation:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{9.940}{9.940 + 5.096} = \frac{9.940}{15.036} = 0{,}6611 = \mathbf{66{,}11}\%$$

4. AUC is the area under the ROC curve. The AUC (Area Under Curve) value in the Logistic Regression model is calculated based on the ROC curve formed from the probability values of the model's predicted results. The ROC curve represents the relationship between the True Positive Rate and False Positive Rate at various threshold values.

$$\mathrm{AUC} = 2\ x\ \frac{\mathrm{FP + FN}}{TP + FP + FN + TN} = \mathbf{0{,}4676}$$

The high recall value indicates that Logistic Regression is very effective in detecting compliant taxpayers, although the classification error rate is relatively larger than Random Forest.

3.6 Application of the Random Forest Method

The calculation of random forest in the evaluation metrics is as follows:

1. Accuracy (Total correct predictions / Total data) The proportion of all correct predictions (both compliant and non-compliant) to the total observations in the testing data. Accuracy is calculated using the equation.

$$Accuracy = \frac{TP + TN}{Total}$$

$$Accuracy = \frac{16.993 + 21.274}{70.000} = \frac{38.267}{70.000} = 0{,}5467 = \mathbf{54{,}67}\%$$

2. Precision (Accuracy of positive predictions)
   The proportion of positive predictions (e.g., NON-COMPLIANCE) that accurately reflect the actual situation. Precision is calculated using the equation:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{16.993}{16.993 + 13.804} = \frac{16.993}{30.797} = 0{,}5517 = \mathbf{55{,}17}\%$$

3. Recall (Ability to detect true positive cases)
   The proportion of all true positive cases that the model successfully identifies as positive. Recall is calculated using the equation:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{16.993}{16.993 + 17.929} = \frac{16.993}{34.922} = 0{,}4868 = \mathbf{48{,}68}\%$$

4. AUC is the area under the ROC curve. The AUC (Area Under Curve) value in the Logistic Regression model is calculated based on the ROC curve formed from the probability values of the model's predicted results. The ROC curve represents the relationship between the True Positive Rate and False Positive Rate at various threshold values.

$$\mathrm{AUC} = 2\ x\ \frac{\mathrm{FP} + \mathrm{FN}}{TP + FP + FN + TN} = \ 0{,}4549$$

A lower error rate value indicates that Random Forest has a smaller classification error rate and more stable prediction performance than Logistic Regression.

3.7 Confusion Matrix Hasil Klasifikasi

A confusion matrix is used to evaluate the performance of a classification model by comparing the model's predictions to actual conditions. This matrix consists of four main components: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), as shown in Table 4 below.

Table 4. Confusion Matrix Random Forest and Logistic Regression

Actual Compliant Non	Compliant	Actual Compliant Non
Compliant	TP	FN
Disobey	FP	TN

Table 4 describes the confusion matrix, which consists of four categories: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). The evaluation results show that the Random Forest method has a precision value of 0.5389 and an error rate of 0.4549, indicating a lower classification error rate than Logistic Regression. However, the recall value of 0.6065 indicates that Random Forest's ability to detect compliant taxpayers is still lower than Logistic Regression.

3.8 Comparison of Logistic Regression and Random Forest Performance

After each model was built and evaluated, the next step was to compare the performance of Logistic Regression and Random Forest to determine the most effective method for predicting motor vehicle tax compliance at the North Aceh Samsat. The comparison was conducted using accuracy, precision, recall, and AUC metrics, which are summarized in Table 5.

Tabel 5.Comparison of Logistic Regression and Random Forest Performance

Model	Accuracy	Precision	Recall	AUC
Regresi Logistik	0,4629	0,4743	0,6611	0,4583
Random Forest	0,5471	0,5536	0,4966	0,5572

3.9 Factors Influencing Tax Compliance

The results of the variable analysis using the Random Forest model indicate that vehicle age is the most dominant factor influencing motor vehicle tax compliance. The older the vehicle, the higher the propensity for late tax payments. Furthermore, vehicle brand and type also contribute significantly to the classification process, while vehicle ownership type has a relatively smaller impact.

4. Conclusion

Future research is recommended to use a larger dataset and include other variables that could potentially influence taxpayer compliance, such as economic factors, region, or income levels. Furthermore, future research could test other machine learning methods, such as Support Vector Machines, Gradient Boosting, or Neural Networks, to obtain more accurate prediction models.

References

Abd Madjid, Z., Sabijono, H., & Mintalangi, S. S. E. (2024). Evaluasi pemungutan pajak restoran berdasarkan Peraturan Daerah Kota Ternate Nomor 15 Tahun 2014 tentang pajak restoran di Badan Pengelolaan Pajak dan Retribusi Daerah Kota Ternate. *Riset Akuntansi dan Portofolio Investasi*, *2*(2), 84–92. https://doi.org/10.58784/rapi.132

Apriansyah, R., Hasibuan, A., Fahmi, B. L., & Munawaroh, N. L. (2023). *Sosialisasi Pemberdayaan kaum Perempuan Sebagai Upaya Penghasilan Tambahan dari Hasil Panen Nelayan di Bantayan, Kecamatan Seunuddon, Kabupaten Aceh Utara*.

Avidaniar Bintary, A. (2020). Analisis Kepatuhan Wajib Pajak Kendaraan Bermotor dalam upaya Meningkatkan Penerimaan Pajak Daerah pada Kantor Bersama Samsat Jakarta Timur Tahun 2015-2018. *Jurnal Pajak Vokasi (JUPASI)*, *1*(2), 86–101. https://doi.org/10.31334/jupasi.v1i2.816

Awal, F. S., Sirat, A. H., & Hadady, H. (2025). Pengelolaan Pajak Kendaraan Bermotor Di Provinsi Maluku Utara. *Sibatik Journal: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, Dan Pendidikan*, *4*(7), 1505-1530.

Baj, T. D. W., Baj, D. R., Baj, T. T., & Baj, M. H. A. (2023). Pengaruh Kesadaran Pajak, Pengetahuan Pajak, Sanksi Pajak Terhadap Kepatuhan Wajib Pajak Kendaraan Bermotor (Studi Kasus Kantor Bersama Samsat Surabaya Selatan). *Behavioral Accounting Journal*, *2*(1), 41–53. https://doi.org/10.33005/baj.v2i1.38

Damayanti, A. Y., Afifah, A. N., & Sunaningsih, S. N. (2023). *Analisis Kontribusi Pemungutan Pajak Kendaraan Bermotor Terhadap Peningkatan Pendapatan Asli Daerah (Pad) Di Kota Magelang Tahun 2018—202*. *12*(2).

Hasibuan, M. R., Rahman, S., & Chiuloto, K. (2025). *Implementasi Metode Random Forest Dalam Prediksi Penyebab Tunggakan Pembayaran Pajak Kendaraan Bermotor Di Samsat Medan Utara*. *1*(2).

Hulaifah Al Abrori, Z. Z., & Subhiyakto, E. R. (2025). Analisis Komparatif Akurasi Prediksi Kanker Payudara Menggunakan Algoritma Random Forest dan Logistic Regression. *Jurnal Algoritma*, *22*(1), 300–311. https://doi.org/10.33364/algoritma/v.22-1.2164

Larasati, N. (2023). Perbandingan Regresi Logistik dan Random Forest pada Klasifikasi Cuaca Wilayah Jawa Tengah. *AKSIOMA : Jurnal Matematika dan Pendidikan Matematika*, *14*(2), 172–181. https://doi.org/10.26877/aks.v14i2.15985

Maulidya, H. (2025). Pemodelan Regresi Logistik Biner Untuk Mengetahui Faktor-Faktor Yang Berpengaruh Terhadap Stunting. *Jurnal Matematika*, *5*.

Nisa, M. (2024). *Analisis Pengaruh Pengenaan Pajak Kendaraan Bermotor, Bea Balik Nama Kendaraan Bermotor, Dan Pemungutan Denda Pajak Kendaraan Bermotor Terhadap Peningkatan Pendapatan Asli Daerah Kota Lhokseumawe Pada Tahun 2018-2022*. *3*(4).

Nugroho, A., & Harini, D. (2024). Teknik Random Forest untuk Meningkatan Akurasi Data Tidak Seimbang. *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, *2*(2), 128–140. https://doi.org/10.53624/jsitik.v2i2.379

Nurahaliza, F., & Mulyadi, N. (2022). Aplikasi Sistem Informasi Dan Pemetaan Daerah Pariwisata Berbasis Web. *Jurnal Tika*, *7*(1), 48–54. https://doi.org/10.51179/tika.v7i1.1085

Purnama, M. A., Ramadhani, J., Anugraha, Y. S., Efrizoni, L., & Rahmaddeni, R. (2024). Perbandingan Performa Algoritma Random Forest dan Gradient Boosting dalam Mengklasifikasi Churn Telco. *Techno.Com*, *23*(3), 645–657. https://doi.org/10.62411/tc.v23i3.11278

Purwa, T. (2019). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017). *Jurnal Matematika, Statistika dan Komputasi*, *16*(1), 58. https://doi.org/10.20956/jmsk.v16i1.6494

Rahmadani, A. A., Putri, A. A., Dala, D., Angka, M. T., & Rafiq, M. (2023). *Analisis Regresi Logistik Biner Untuk Memprediksi Faktor-Faktor Internal Yang Memengaruhi Keharmonisan Rumah Tangga Menurut Provinsi Di Indonesia Pada Tahun 202*.

Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/variansiunm31

Wasana, W. S., Adityo, R. D., & Herulambang, W. (2021). Implementation of Intelligent Parking System Using IoT-Based Devices (case Study of Galaxy Mall Surabaya). *JEECS (Journal of Electrical Engineering and Computer Sciences)*, *6*(2), 1135–1158. https://doi.org/10.54732/jeecs.v6i2.207

Wati, R., & Saepuloh, C. (2025). *Pengaruh Program Pemutihan Pajak Kendaraan Bermotor Terhadap Kepatuhan Wajib Pajak Kendaraan Bermotor Pada Kantor SAMSAT Kabupaten Bandung II Soreang*. *11*.

Yuniarsyih R.A, R. D., Muhadi, R. A., Fitrianto, A., & Silvianti, P. (2025). Analisis Regresi Logistik Biner dan Random Forest untuk Prediksi Faktor-Faktor Stunting di Pulau Jawa. *Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi*, *13*(2), 147–156. https://doi.org/10.37905/euler.v13i2.31680