Journal of Soft Computing Exploration
Homepage: htttps://shmpublisher.com/index.php/joscex
p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

Predicting willingness to pay for urban rail transit using machine learning : Evidence from jakarta MRT

Wisnu Wardan Kusuma[1], Ade Irfan Efendi[2], Dandun Prakosa[3], M. Popik Montanasyah[4], Yus Rizal[5], M. Adil Wanadi[6]
[1]Road Transportation Management Study Program, Politeknik Transportasi Darat Indonesia (PTDI-STTD), Indonesia
[2]Computer Science Study Program, Universitas Nusa Mandiri, Indonesia
[3,5,6]Land Transportation Study Program, Politeknik Transportasi Darat Indonesia (PTDI-STTD), Indonesia
[4]Railway Transportation Management Study Program, Politeknik Transportasi Darat Indonesia (PTDI-STTD), Indonesia

Article Info

Article history:

Received Mar 14, 2026
Revised April 2, 2026
Accepted April 4, 2026

Keywords:

Willingnes to pay
Ability to pay
MRT
Machine learning
Public transportation
Fare policy

ABSTRACT

The development of urban transportation requires an efficient, reliable and sustainable system, so fare determination is an important factor in the success of the Jakarta MRT service. In this context, understanding the user's Willingness to Pay (WTP) is crucial because it is not only influenced by economic ability, but also perception and preference for services. This study aims to analyze and predict the WTP of MRT users by integrating transportation economics approaches and machine learning methods. The research data is in the form of primary data from a survey of 296 MRT users which includes socio-economic characteristics, transportation costs, frequency of use and Ability to Pay (ATP). The methodology used includes descriptive analysis and regression modeling using various algorithms, namely Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regression (SVR) and XGBoost. Model evaluation was carried out using MAE, RMSE and determination coefficient (R^2).
The results showed that the value of WTP was relatively homogeneous compared to variations in income and transportation costs, which indicated that willingness to pay was not entirely determined by economic ability. The performance of the model shows that no algorithm is consistently superior, with R^2 values that tend to be low. The feature importance analysis identified income, transportation costs and ATP as the main factors. This research contributes through the application of a multi-model machine learning framework and policy implications that MRT fare determination needs to consider economic aspects and user preferences in a balanced manner.

Corresponding Author:

Wisnu Wardana Kusuma,
Road Transportation Management Study Program,
Indonesian Land Transport Polytechnic-STTD,
PTDI-STTD, 17520, Bekasi, Indonesia.
Email: wisnuwardanakusuma85@gmail.com
https://doi.org/10.52465/joscex.v7i1.18

1. INTRODUCTION

The rapid development of community mobility in urban areas has increased the need for an efficient, reliable and sustainable public transportation system [1]. Big cities face various problems such as chronic congestion, uncertainty of travel time, and environmental pressure due to the dominance of private vehicle users. This condition requires a mass transportation solution that is able to improve the efficiency of movement and the quality of urban transportation services. Jakarta as one of the metropolitan cities in Southeast Asia faces this problem significantly. In an effort to overcome this, the Mass Rapid Transit (MRT) system was developed as a mode of transportation that is expected to be able to increase urban mobility, expand the accessibility of public transportation services, and encourage the switch from the use of private vehicles to mass transportation. However, the success of the MRT system is not only determined by the infrastructure and operational aspects, but also by the fare policies implemented.

Fare planning is one of the key factors in the success of the mass transportation system [2]. The proper fare structure not only affects people's interest in using public transportation, but also determines the operational sustainability of transit system operators [3]. Therefore, an approach is needed that not only considers the growth in passenger demand but also the ability and willingness to pay users [4]. In transportation economics studies, Ability to Pay (ATP) and Willingness to Pay (WTP) are the two main concepts used to evaluate the affordability of price such as the value perceived by users for certain services [5]. ATP reflects the user's objective economic ability to pay for transportation costs based on income and expenses, while WTP reflects the user's willingness to pay a certain fare in exchange for assessed benefits such as convenience, speed, security and reliability of the service [6]. ATP and WTP estimates are very important in designing transportation price policies that are responsive to the social and economic conditions of the community [7].

Research on public transportation services such as the Commuter Line in South Tangerang has shown that service quality factors such as availability and travel time have a significant impact on users' WTP [8]. Various previous studies have examined ATP and WTP in the context of public transportation. A study on Line commuter services in South Tangerang shows that service quality such as availability and travel time have a significant effect on user WTP [9]. Another study used ATP and WTP to evaluate the suitability of public transport fares, and found differences between affordability and willingness to pay across different income groups [10]. In addition, a study on bus services in Dhaka identified comfort and safety factors as the main determinants of WTP [11]. Research on Jakarta LRT users also showed that the value of WTP tends to be lower than WTP, which indicates that the perception of service benefits affects the Willingness to Pay [12].

However, most of these studies still use conventional methods such as the Contingent Valuation Method (CVM), travel cost method and discrete choice, which tend to be descriptive and have limitations in capturing the complex relationships between variables [13]. As technology has evolved, machine learning methods have shown significant potential in modeling travel behavior and predicting transportation demand with greater accuracy than traditional methods [14].

However, the application of machine learning in ATP and WTP analysis, especially in the context of MRT Jakarta, is still relatively limited. Most of the research focuses more on demand modeling or mode selection and has not integrated ATP and WTP as a basis for determining fair and sustainable tariff policies. In addition, in the context of developing countries such as Indonesia, studies that combine transportation economics and machine learning approaches are still very limited. Based on these gaps, this study aims to analyze and predict the Willingness to Pay (WTP) of MRT Jakarta users using machine learning methods based on data on socioeconomic characteristics, travel patterns and user perceptions. This research is expected to contribute to the development of a more comprehensive analysis model and become the basis for the formulation of MRT pricing policies that are fair, sustainable and responsive to user preferences.

2. METHOD

Research Design

This study uses a data-based quantitative approach to analyze and predict the Willingness to Pay (WTP) of MRT Jakarta users, The method used is machine learning regression to identify the relationship between socioeconomic characteristics, travel behavior and WTP values [15].

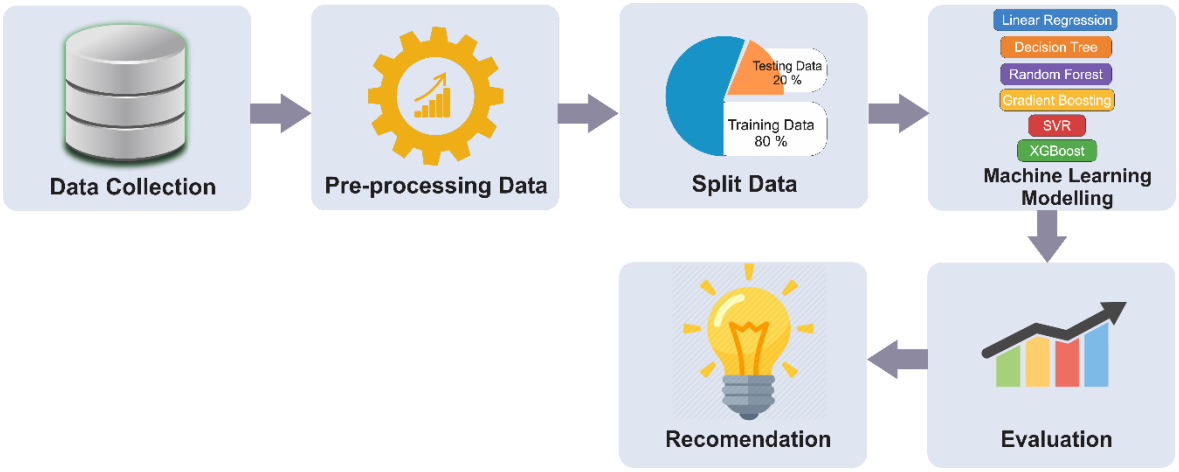

Figure 1. Research framework of WTP using machine learning

Data Collection

The data used in this study is primary data obtained through a questionnaire survey to Jakarta MRT users at Lebak Bulus Station, Dukuh Atas station and HI Roundabout station. This survey was conducted to collect information related to the demographic characteristics of respondents, travel patterns and transportation costs incurred. Data collection was carried out by distributing questionnaires directly to MRT users at several main stations. The number of respondents who were successfully collected in this study was 296 respondents which were then used as research samples. Although the sample size of only 296 respondents was limited to the application of complex ensemble algorithms such as Gradient Boosting and XGBoost, this size was still considered adequate for exploratory analysis and initial model development. The dataset used is able to present the main variations in the socioeconomic characteristics and travel behavior of users at MRT stations with high levels of mobility, thus allowing the identification of common patterns in the WTP. However, these limitations have the potential to affect the generalization ability of the model as well as the stability of more complex algorithms, so it is necessary to be concerned in interpreting the results of the research.

Research

The variables of this study consist of :

1) Variable dependen : Willingness to Pay (WTP)
2) Variable independent: monthly income, monthly transportation costs, frequency of use and Ability to Pay (ATP)
3) Categorical variable: gender and frequency of travel.

These variables are used to present socioeconomic conditions and user behavior in influencing the value of WTP.

Statistics Description

This stage is used in the early process in data analysis to describe the characteristics of the respondents and key variables before being further modeled [16]. This analysis is useful for assessing distribution, central tendencies and variations of variables such as respondents' monthly income, monthly transportation costs, MRT travel frequency, Ability to Pay (ATP) value and Willingness to Pay (WTP) value. This approach is commonly used in the study of public transport behavior and WTP to understand the data base that is the input of further modeling [17].

In looking for central size and variability using values Mean, Minimum and maximum as well as standard deviation.

1) Mean to describe the central tendency of a variable.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \bar{x}_i \, [18] \tag{1}$$

2) Minimum and Maxim are used to indicate the observed range [19].
3) Standard Deviation is used to measure the distribution of data around the mean:

$$SD = \sqrt{\frac{1}{n-1}}\sum_{i=1}^{n}(x_i - \bar{x})^2 [20] \tag{2}$$

Data Preprocessing

The data pre-processing process is carried out through several stages to ensure data quality and readiness in modeling. The first stage is data cleansing, which is removing irrelevant data and handling empty values, especially on target variables so as not to affect the results of the analysis [21]. Next, data transformation is carried out by converting categorical variables into numerical forms using the one-hot encoding method, so that the data can be processed by machine learning algorithms [21].

The next stage is feature selection where all variables are grouped into two categories, namely numerical and categorical variables. This grouping aims to simplify the advanced pre-processing process as well as the integration of data into prediction models [22]. After that, data splitting is carried out, which is the division of the dataset into training data of 80% and test data of 20%. This division aims to ensure that model evaluation is carried out objectively and can measure the model's generalization ability against data that has never been seen before.

Machine Learning Modeling

This study applies several machine learning-based ejection algorithms to compare its performance in predicting the value of Willingness to Pay (WTP). The models used in this study include Linear Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Regression (SVR) and XGBoost. The entire model is implemented using a pipeline approach that integrates the pre-processing and modeling stages into one unified workflow. This approach aims to ensure that the model training process is carried out consistently, systematically and reproducibly.

Model Evaluation

The evaluation stage of the performance of the model is carried out using three main indicators, namely:

1) Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{i-} \widehat{y_i}| \quad [4] \tag{3}$$

2) Root Mean Square Error (RMSE)

$$MAE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i-} \widehat{y_i})^2} \, [4] \tag{4}$$

3) Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_i^n (y_{i-}\widehat{y_i})^2}{\sum_i^n (y_{i-}\bar{y})^2} \, [4] \tag{5}$$

Where y_i is the actual value of WTP and, $\widehat{y_i}$ is the predicted value of the WTP, $\bar{y}$ is the average value of WTP and n is the amount of observation. The best models are determined based on the highest R^2 values as well as the lowest MAE and RMSE values [23].

Feature Important Analysis

Analysis of feature importance was used to identify the contribution level of each independent variable to the model's performance in predicting WTP values. In this study, feature importance is calculated based on features on the reduction of prediction errors in each model formation process. In general, the value of feature importance for the variable can be stated: i

$$FI_i = \frac{1}{T}\sum_{t=1}^{T} \Delta I_{i,t} \quad [24] \tag{6}$$

Where FI_i, is the feature importance value of the variable to-i, T expresses the total number of trees or iterations in the model and $\Delta I_{i,t}$ indicates the magnitude of the contribution of the variable to-i to a decrease in error function or an improvement in the quality of data separation on iterations or trees to-t.

Greater feature importance values indicate that the variable has a more significant influence on the model's decision-making process [24]. This analysis provides a deeper understanding of the dominant factors

that affect WTP so that it is used as a basis for understanding the results and formulating public transportation fare policy recommendations.

Visualization and Interpretation

The results of the study are visualized in several forms to facilitate the interpretation and evaluation of the model. The visualization used includes a graph of the relationship between the actual value and the predicted value, a feature importance graph, and a comparison table of performance between models [25]. This visual presentation aims to provide a clearer picture of the model's ability to make predictions, as well as validate the performance of the model used more comprehensively. [26].

Research Outputs

The outputs produced in this study have several main aspects, namely the best machine learning model in predicting the value of Willingnes to Pay (WTP), identification of dominant factors that affect WTP, and policy recommendations in determining public transportation fares. These outputs are expected to contribute to the development of data-driven analysis and become a basis for consideration in the formulation of more effective and sustainable transportation policies.

3. RESULTS AND DISCUSSIONS

Descriptive Data and Statistics Characteristics

This study uses survey data from 296 respondents who use MRT Jakarta at Lebak Bulus Station, Dukuh Atas and HI Roundabout. This data was collected through a distributed questionnaire containing the socio-economic status of the respondents as well as information related to transportation expenses and the value of Willingness to Pay (WTP). All data used has been cleaned and validated so that it is suitable for use in the modeling stage.

The main variables analyzed in this study include respondents' monthly income, monthly transportation costs and the average WTP value to MRT fares. The distribution of respondents' income shows a wide enough variation to reflect the diversity of economic backgrounds of MRT users. This is important to ensure that the prediction model is able to capture differences in ability and Willingness to Pay between groups of users.

Tabel 1. Respondent characteristics

Variable	Category	Quantity	Percentage
Gender	Male	130	43.9 %
	Women	166	56.1 %
Age	<20 Years	15	4.7 %
	20-30 Years	173	58.4 %
	30-40 Years	99	33.4 %
	40-50 Years	9	3.4 %
Education	Bachelor's Degree	180	60.8 %
	Diploma	69	23.3 %
	Vocational High School	44	14.9 %
	Magister Degree	3	1 %
Occupation	Private Employees	156	50 %
	Civil Servant / Military / Police	24	8 %
	State-Owned Enterprise Employee	51	17 %
	Self-employed	19	6 %
	Student	46	16 %
Origin_station– destination_station	Lebak Bulus - Bundaran HI	77	26 %
	Bundaran HI - Lebak Bulus	76	26 %
	Lebak Bulus - Dukuh Atas	72	24 %
	Dukuh Atas - Lebak Bulus	71	24 %
Trip Purpose	Work	227	77 %
	Education	39	13 %
	Social	9	3 %
	Recreation	11	4 %
	Education	2	0 %
	Shopping	8	3 %
Mode of Transportation	Motorcycle	125	52 %
	Online Motorcycle Taxi	39	16 %
	Bus	35	15 %
	Commuter Train	33	14 %
	Car	7	3 %
Travel Frequency	1-2 times	24	8 %
	3-4 times	96	32 %
	5-6 times	176	60 %
Reason for Using MRT	Travel Comfort	44	15 %
	Shorter Travel Time	173	58 %
	Ease of Access	27	9 %

	Affordable Fare	25	9 %
	Travel Safety	27	9 %

The table above presents the socio-demographic characteristics and travel patterns of the 296 MRT respondents analyzed in this study. Based on gender distribution, female respondents accounted for 56.1% of the total sample, while male respondents accounted for 43.99%. This composition shows that MRT users are relatively balanced between genders with a slightly higher participation tendency than female users. In terms of age, the majority of respondents are in the productive age group. The 20-30 year old age group dominated the sample with a proportion of 58.4%, followed by the 30-40 year old age group at 33.4%. The proportion of respondents in the age group under 20 years and above 40 years is relatively small, which indicates that the MRT is mainly used by young age groups and early adults with a high level of mobility.

Based on education level, most of the respondents were undergraduate graduates with a percentage rate of 68%, followed by a diploma graduate rate of 23.3% and vocational high school graduates of 14.9%. The dominance of respondents with secondary to higher education levels shows that MRT users in this study have a relatively good educational background which has the potential to affect the perception of service quality and willingness to pay public transportation fares. In terms of employment, respondents were dominated by private employees with a percentage level of 50%, followed by employees of State-Owned Enterprises at 17% and students or students at 16%. This distribution indicates that the MRT plays a role as the main mode of transportation for groups of workers and users with regular travel activities.

Judging from the travel pattern, the Lebak Bulus-Bundaran HI and the HI-Lebak Bulus Roundabout route were each used by 26% of respondents, while travel to and from Dukuh Atas accounted for 48% overall. This indicates that the main corridor of the MRT is used in a relatively balanced manner in both directions of travel. The respondents' travel goals were dominated by work by 77%, followed by educational goals at 13%. These findings confirm the role of the MRT as a mode of commuter transportation that supports daily economic activities.

Before using the MRT, the majority of respondents used motorcycles by 52%, followed by online motorcycle taxis at 16%, buses at 15% and commuter trains at 14%, indicating a shift from private transportation and public transportation to the MRT. Meanwhile, in terms of usage intensity, most respondents use MRT 5-6 times per week with a percentage rate of 60% which reflects a high level of dependence on MRT services. The main reason for using MRT is shorter travel time with a percentage rate of 58%, followed by travel comfort of 15% as well as safety factors, ease of access and affordability of fares. Overall, respondents' characteristics showed that MRT users were dominated by productive age groups with high travel frequency and primary travel destinations for work. This condition provides a strong basis for the analysis of Willingness To Pay (WTP) predictions, because respondents are directly related to price policies and the intensity of MRT service use.

Tabel 2. Research variables

Variabel	Unit	Deskripsi
Gender	Category	Numerics Categorical
Travel Frequency	Category	Numerics Categorical
Monthly Income	IDR	Variable Target Numerics
Monthly Transportation Cost	IDR	Variable Target Numerics
WTP	IDR	Variable Target Numerics

The table above summarizes the research variables used in the analysis which include numerical and categorical variables that present the socio-economic characteristics of respondents as well as MRT user behavior. The variables of monthly income and monthly transportation costs were used to describe the economic capabilities of the respondents, while the actual WTP value acted as the target variable in the modeling. Categorical variables, namely gender and frequency of MRT use, were used to capture differences in individual characteristics and travel patterns that have the potential to affect the willingness to pay MRT fares. Descriptive analysis was carried out on three main variables that were the focus of the study, namely respondents' income, monthly transportation costs, and actual WTP value. These three variables represent the socio-economic conditions of MRT users and their perception of the affordability of service price.

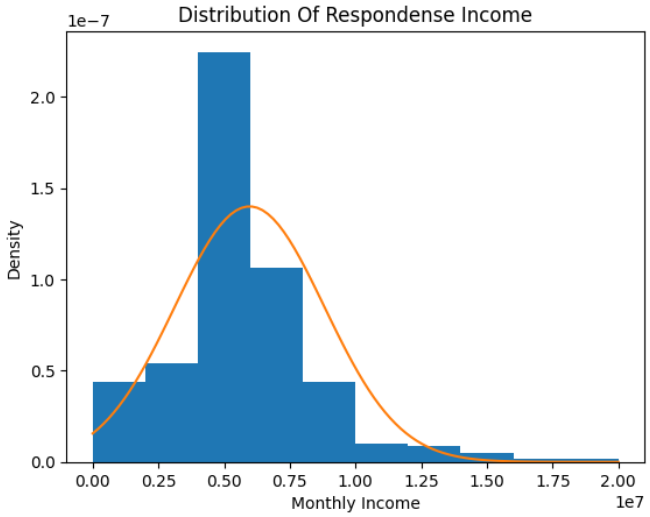

Figure 2. Distribution of respondents' monthly income

Figure 2 above shows that the distribution of respondents' income shows a wide enough variation so that it reflects the diversity of economic backgrounds of MRT users which shows that most respondents have income in the middle range around the average income. This indicates that the MRT is used by various segments of society. This variation is an important basis for analyzing the difference in ability and willingness to pay public transportation fares.

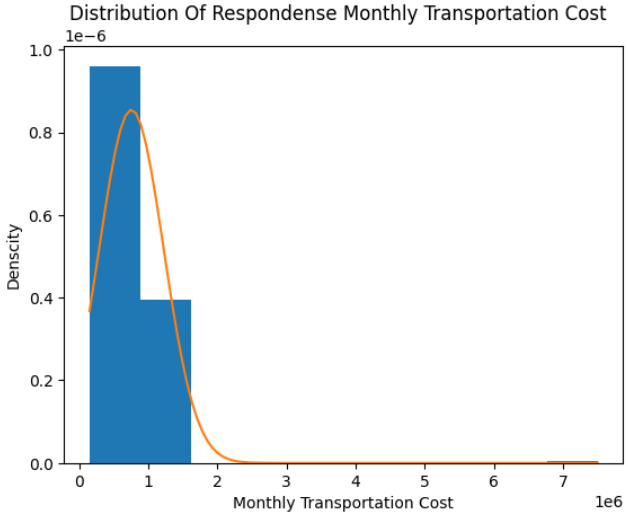

Figure 3. Distribution of respondents' monthly transportation costs

Figure 3 in the distribution of monthly transportation costs also shows various distribution patterns. Most of the respondents have transportation expenses in the middle range, while there are a small number of respondents with relatively high transportation costs. This condition can be influenced by the frequency of travel, distance traveled and the combination of transportation modes used.

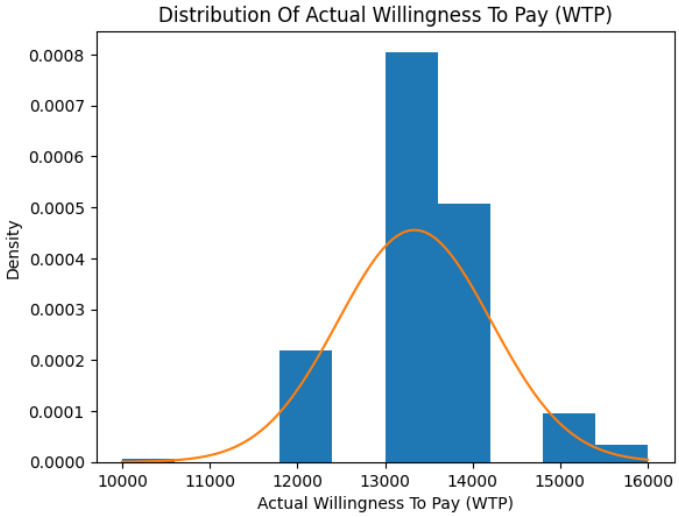

Figure 4. Actual distribution willingness to pay (WTP)

Meanwhile, in Figure 4, the actual WTP value shows a relatively concentrated distribution in a certain range of values. This indicates that there is a fairly homogeneous perception of the value of MRT fares that are considered reasonable by users even though their income background and transportation costs are different. In the descriptive statistical analysis of the main variables in providing a comprehensive quantitative picture, this study presents statistical values in the form of average, minimum, maximum and standard deviation values for each main variable, namely monthly income, monthly transportation costs and actual ATP values as shown in table 3 below.

Tabel 3. Descriptive statistics

Variabel	Mean (IDR)	Min (IDR)	Max (IDR)	Standar Deviasi (IDR)
Monthly Income	5,967,917	400,000	20,000,000	2,849,361
Monthly Transportation Cost	756,507	150,000	7,500,000	466,777
WTP	13,338	10,000	16,000	876

The average monthly income of the respondents of IDR 5,967,917 shows that MRT users in this study are dominated by the middle-income group, but the income range is very wide from IDR 400,000 to IDR 20,000,000 which reflects a significant level of economic inequality among respondents. This is strengthened by the relatively large standard deviation value of IDR 2,849,361 which indicates the heterogeneity of the economic conditions of MRT users. These findings confirm that the MRT functions as a mode of public transportation that serves various socioeconomic strata starting from low to high income groups.

In the variable monthly transportation costs, the average expenditure of the respondents is IDR 756,507 which shows a significant proportion of the income allocated for daily mobility.the wide cost range is between IDR 150,000 to IDR 7,500,000, indicating a large variation in travel patterns and the intensity of transportation users. The standard deviation of IDR 466,777 confirms that although most respondents have relatively balanced transportation expenses, there are groups of users with much higher transportation costs that are possible due to long mileage, high travel frequency or the use of a combination of transportation modes.

On the other hand, the actual ATP value shows different characteristics. With an average value of IDR 13,338 with a minimum range of IDR 10,000 and a maximum of IDR 16,000 and a relatively small standard deviation of IDR 876, it can be concluded that the willingness to pay MRT users tends to be relatively the same. The stability of the WTP value indicates a common perception among respondents regarding MRT fares that are considered feasible and acceptable regardless of differences in income levels and transportation costs.

Overall, the difference in statistical characteristics between economic variables (income and transportation costs) and WTP variables shows that the economic capabilities of users are very diverse, the willingness to pay for MRT services is relatively concentrated in a certain value range. This finding has important implications for price policy so that the price determination falam needs to consider the limitations

of the low-income group bearing in mind that the WTP value is relatively uniform and does not increase in proportion to income. Thus, the descriptive statistical hail provides a strong field evidence basis for further analysis of WTP modeling and the formulation of fair and sustainable MRT fare policies.

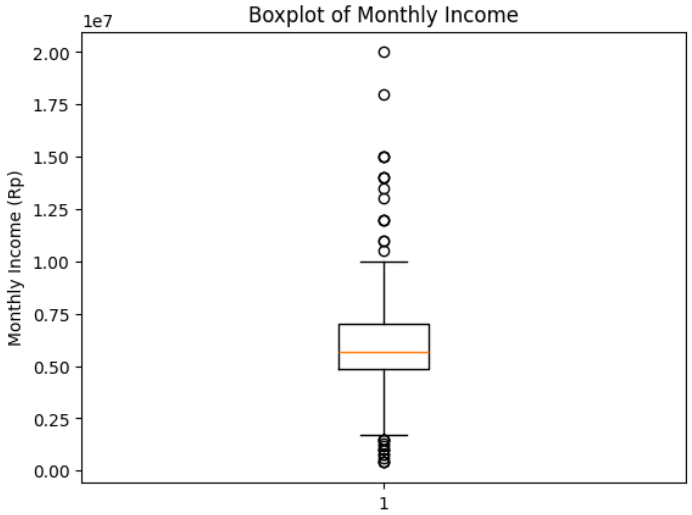

Figure 5. Distribution of respondents' monthly income

Figure 5 illustrates the distribution of data based on the value of monthly income which shows the median, quartile and outlier values, which represent the median income value and reflect the median income level of the respondents. It can be seen in the box range of the interquartile range / IQR which describes the income distribution of 50% of respondents around the median which shows that the main focus of income is in the middle range.

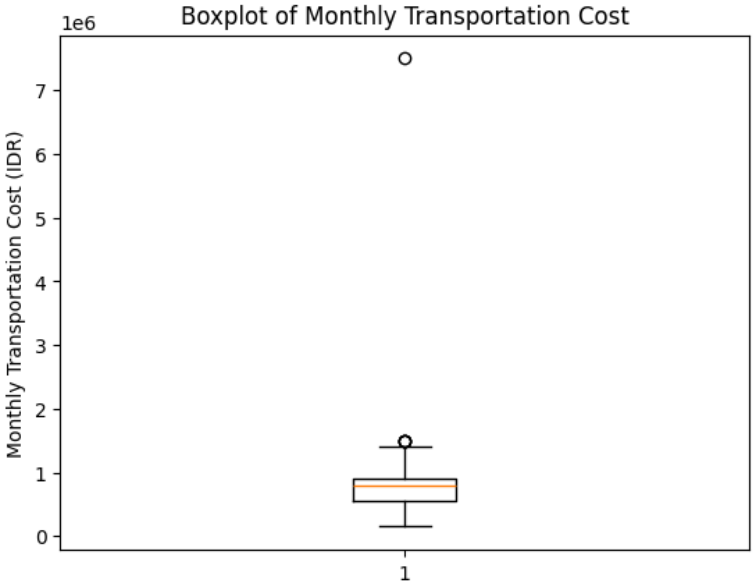

Figure 6. Distribution of respondents' monthly transportation costs

The distribution of monthly transportation cost data shows the characteristics of the distribution of respondents' transportation expenditures based on the median value, quartile and the existence of outliers that depict the red line in the box representing the median value of monthly transportation costs. The red line reflects the amount of public transportation expenses experienced by some respondents which varies, especially for 50% of respondents around the median.

Predictive Performance Model

The performance of the model was evaluated using the main metrics, namely, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and determination coefficient (R^2). This evaluation aims to assess the accuracy of the Willingness To Pay (WTP) value and the ability of each da model; am explaining the variation of the data.

Tabel 4. Comparison of performance models of WTP prediction results

	Model	MAE	RMSE	R^2
1	Linear Regression	580.265234	760.547754	0.051317
2	Decision Tree	725.000000	1134.680572	-1.111617
3	Random Forest	651.301389	869.947774	-0.241236
4	Gradient Boosting	622.135865	847.156813	-0.177052
5	SVR	583.324115	884.911852	-0.284304
6	XGBoost	598.266907	812.086395	-0.081614

Based on table 4 above, Linear Regression shows the highest R^2 tilapia of 0.051 even though the value is relatively low. This indicates that linear models are only able to account for about 5% of the WTP variation, but still provide more stable performance than other nonlinear models. This low R^2 value illustrates that the relationship between independent variables (monthly income, transportation costs and ATP) and WTP is weak and not completely linear.

Figure 7. Comparison result MAE, RMSE, R^2

Figure 7 shows a comparison of the performance of the regression model based on MAE, RMSE and R^2 values. The results of the evaluation showed that there was no significant difference in performance in each model. Although ensemble-based models such as Random Forest, Gradient Boosting and XGBoost tend to produce more stable error values than the default Decision Tree. The relatively low R^2 value across the model indicates that the variation in Willingness to Pay (WTP) is not only influenced by the economic factors used in the model, but is also very likely to be influenced by latent variables such as perception of service quality and individual preferences. Therefore, the ensemble approach and feature importance analysis are important to obtain a more substantive interpretation rather than relying solely on prediction accuracy values.

In addition, the argument that WTP analysis in the context of public transport requires a holistic approach, such as the integration of conceptual variables, quality of service, or the use of discrete choice models that are explicitly designed to capture individual preferences. This analysis is an important basis for discussing the implications of price policy and developing a more representative model in future research.

Feature Importance Analysis

Feature importance analysis was conducted to identify the contribution of each independent variable in predicting the Willingness to Pay (WTP) value of MRT users using various modeling approaches, including linear models, tree-based models, ensemble methods and kernel-based methods. The results of the analysis of each model are presented in table 5.

Table 5. Comparison of performance models of WTP prediction results

Variable	LR	RF	GB	DT	XGB	SVR

Monthly Income (IDR)	0.000003	0.353	0.405	0.383	0.262	0.000201
Monthly Transportation Cost (IDR)	0.000019	0.334	0.337	0.362	0.295	0.000013
Travel Frequency	0.999119	0.04	0.018	0.035	0.161	0.000117
ATP	0.000858	0.269	0.239	0.22	0.282	0.000105

Table 5 shows the comparison of the relative contribution of each independent variable to WTP prediction based on the six regression models tested, namely Linear Regression (LR), Random Forest (RF), Gradient Boosting (GB), Decisin Tree (DT), XGBoost (XGB) and Support Vector Regression (SVR). These results show a striking contrast in the variable contribution patterns between linear and nonlinear models. In the Linear Regression model, the variable Travel Frewquency dominated almost all the contribution with a net of 0.999, while economic variables such as income, transportation costs and Ability to Pay (ATP) had a very small contribution. This condition indicates a distortion in the linear model where complex relationships are oversimplified into one dominant linear relationship, resulting in interpretations that do not reflect the actual economic conditions. In contrast, tree- and ensemble-based models (RF, GB, DT and XGB) show a more balanced and realistic distribution of feature importance. Monthly income variables and transportation costs consistently emerged as the most influential factors, with contribution values ranging from 0.26 to 0.41. These findings are in line with the theory of transportation economics which states that WTP is greatly influenced by the economic capacity of users, which is reflected in the concept of Ability to Pay (ATP). The ATP variable also showed a stable and significant contribution to the entire nonlinear model, ranging from 0.22 to 0.28, which reinforces its role as an important factor in explaining the behavior of WTP.

Meanwhile, the Travel Frequency variable only provided a moderate contribution in the XGBoost model and relatively small in the other models, suggesting that this variable plays more of a supporting behavioral factor than a key determinant of WTP. In the SVR model, the feature importance value generated is relatively small for all variables because the kernel-based approach used reflects more predictive sensitivity than the structural contribution of the variables in the model. Overall, these results suggest that although modern ensembles do not always produce the highest prediction accuracy, they are able to provide more realistic and economically consistent interpretations than linear models. Therefore, the ensemble approach is considered more appropriate to be used to understand the structure of the WTP formation as well as as a basis for the formulation of MRT pricing policies that are oriented towards fairness and affordability.

Economic Interpretation and Policy Implications

The results of WTP predictions obtained from various regression models can provide a quantitative picture of the limit of the willingness of MRT Jakarta users to pay service rates based on economic conditions and travel characteristics. Economically, the predicted WTP value reflects the balance between the benefits felt by users and the limithed purchasing power so that it can be used as an indicator of price affordability.

The WTP value which is relatively close to the WTP value indicates that most MRT users are in a condition of limited elasticity to fare changes. So that if the price is increased that exceeds the WTP threshold, it will have the potential to significantly reduce the level of MRT use, especially in the lower middle income group. Based on the results of the WTP prediction, the policy implication that can be formulated is the need for a fair price strategy. MRT fare adjustments should be made gradually and data-based, taking into account the average value of WTP as the upper limit of user tolerance.

In addition, these results support the implementation of income-segment-based subsidies, where low-income groups receive price protection through cross-subsidies or differential price mechanisms. This approach allows operators to maintain financial sustainability without sacrificing the accessibility of public transportation. Overall, the integration of machine learning results in WTP analysis provides a strong empirical basis for the formulation of MRT fare policies that are not only economically efficient, but also socially just. This approach strengthens the role of data analytics in bridging the technical needs of transportation planning and public policy objectives oriented towards community welfare.

Comparison with Previous Research

The results of this study show the heterogeneous characteristics of WTP and the limitations of the ability of economic variables to explain WTP variations, which is in line with the findings of other studies in the public transportation literature. In his research [27] WTP distributions are for various travel attributes and find that WTP distributions are heterogeneous and multimodal. As a result, traditional parametric models often fail to explain WTP variations thoroughly without including user preferential factors such as service preferences and travel attributes. Another study in the context of urban transportation showed that various factors such as service quality, travel time and comfort significantly affected respondents' WTP scores which confirmed that the relationship between ATP and WTP is not linear and is influenced by individual perceptions [28]. In addition, previous research has conducted research on WTP that is often lower than current rates or

average travel costs, so fare policies need to consider user preferences so as not to exceed their actual WTP. In a study related to integrated bus fares in the Italian region, it was found that the average WTP for annual tickets was lower than the prevailing rate, particularly among elderly homeless people, which indicates a market limitation to fare increases without improving the quality of service [29].

In terms of the use of machine learning methods in this study, transportation literature also shows an increasing trend in the use of this technique to predict mobility behavior and transportation demand. A systematic study of the application of machine learning in mode selection models states that ensemble and deep learning models have been widely used for more accurate predictions than traditional counterfactual models such as multinomial logit, although their interpretation is often a challenge [30]. Overall, this research is in line with the previous findings, this research provides additional contributions through the implementation of machine learning that is integrated with transportation economic analysis. These findings affirm the importance of developing a more comprehensive and policy-oriented WTP model by considering the interaction between economic and non-economic factors to support the formulation of fairer, more adequate, and evidence-based public transportation fare policies.

Although this study provides an empirical contribution and analysis method of Willingnes to Pay (WTP) of MRT users using machine learning, there are still some limitations that need to be openly acknowledged. The sample size is relatively limited, so it has the potential to affect the model's ability to capture variations in WTP behavior as a whole. The data used is cell reported and obtained through surveys, so it is inseparable from the potential for response bias, recall, bias and the tendency of respondents to give normative answers. This study also did not explicitly include psychological variables and service perceptions, such as satisfaction levels, perception of service quality, comfort and trust in transpotation operators. This research is also limited to the coverage area of the research area that has not been generalized directly to other regions with different social, economic and transpotation system characteristics. Therefore, further research is recommended to use larger and more diverse data, as well as integrate behavioral and psychological factors to improve model accuracy and policy relevance.

Overall, this study shows that the Willingness to Pay (WTP) of Jakarta MRT users is a complex phenomenon and is not solely determined by economic factors. Although variables such as revenue, transportation costs and Ability to Pay (ATP) have important contributions to the model, the results of the analysis show that the ability of these variables to explain WTP variation is still limited. This is reflected in the relatively low value of the determination coefficient (R^2) and the distribution of WTP which tends to be homogeneous amid variations in the economic conditions of the respondents. These findings indicate that there are other factors at play, especially non-economic factors such as perceptions of service quality, convenience, security, and user preferences and experience. In addition, the use of the machine learning approach in this study provides an idea that although the method is able to identify the patterns and contributions of variables more flexibly than conventional methods, the results still depend on the completeness of the variables used. Thus, the limitation in including psychological variables and service perception is one of the factors that affect the performance of the model. In a broader context, these findings confirm the importance of a multidimensional approach in the analysis of public transport, which considers not only economic aspects, but also aspects of user behaviour and perception. This synthesis is an important basis for formulating conclusions and providing direction for further research and the development of a more comprehensive transportation policy.

4. CONCLUSION

This study aims to analyze the Wiliingness to Pay (WTP) of MRT users by integrating transportation economics approaches and machine learning methods. The results of the descriptive analysis show that the value of WTP is relatively homogeneous compared to variations in income and transportation costs, which indicates that willingness to pay is not fully determined by economic ability. These findings reinforce that WTP is the result of an interaction between economic factors and users' subjective preferences for public transportation services. Based on the model evaluation, Linear Regression shows the highest R^2 value, while ensemble models such as Random Forest and Gradient Boosting provide more stable error performance. Although some models produce negative R^2 values, this reflects the complexity of WTP behavior that is difficult to explain with economic variables alone. The feature importance analysis identified monthly income, transportation costs and ATP as the dominant factors, while the frequency of MRT use represented contextual behavioral aspects.

The main contribution of this research lies in the application of a multi-model machine learning approach to understand the structure of WTP formation as well as its integration with the implications of fair pricing policy. However, this study has limitations, including not including psychological variables and service perceptions such as satisfaction level, service quality, comfort and trust as well as limited scope of the research area that affects the generalization of results. In terms of policy, the results of this study show that the determination of MRT pricing needs to consider not only the ability to pay, but also the perception of service value through quality improvements such as convenience, reliability and accessibility as well as the potential for the implementation of differentiated pricing schemes or subsidies based on user groups. Therefore, further research is recommended to use broader and diverse data and integrate behavioral and perceptual variables and develop more comprehensive modeling methods to improve model accuracy and policy relevance.

ACKNOWLEDGEMENTS

The author would like to thank the respondents and PT. MRT Jakarta for the support of the data and information provided. Appreciation was also conveyed to colleagues and for the constructive input that improved the quality of this research.

REFERENCES

[1] R. Ranjan and S. Sinha, "A systematic review of mode choice behavior in urban transportation with emphasis on individual preferences and influencing factor," *Discover Cities*, vol. 2, no. 1, p. 98, Oct. 2025, doi: 10.1007/s44327-025-00138-3.

[2] E. Eriskin, "Collaborative Game-Theoretic Optimization of Public Transport Fare Policies: A Global Framework for Sustainable Urban Mobility," *Sustainability (Switzerland)*, vol. 16, no. 24, Dec. 2024, doi: 10.3390/su162411199.

[3] A. Schöbel and R. Urban, "Fare Structure Design in Public Transport," Feb. 2025, [Online]. Available: http://arxiv.org/abs/2502.08228

[4] R. Shahangian and N. Ghasemi, "Es timation of the Willingness to Pay for Public Transport Improvements," vol. 29, pp. 73–93, Aug. 2024, doi: 10.52547/jpbud.29.2.75.

[5] M. Bruno, M. Kouwenberg, and N. van Oort, "Equity in transit fare policy: a literature review," *Transp. Rev.*, 2025, doi: 10.1080/01441647.2025.2601689.

[6] V. Vichiensan, V. Wasuntarasook, S. Malaitham, A. Fukuda, and W. Rujopakarn, "Willingness to Pay for Station Access Transport: A Mixed Logit Model with Heterogeneous Travel Time Valuation," *Sustainability (Switzerland)*, vol. 17, no. 15, Aug. 2025, doi: 10.3390/su17156715.

[7] A. Leung, C. Burlando, and T. Pavanini, "Market segmentation and willingness to pay for public transport annual passes among older adults: insights from Genoa, Italy," *Research in Transportation Business and Management*, vol. 58, Jan. 2025, doi: 10.1016/j.rtbm.2024.101243.

[8] E. Saptutyningsih and L. Karimah, "Valuing Public Transportation Using Contingent Valuation Method: A Study Of Commuter Line (Krl) In Tangerang Selatan, West Java Indonesia," *Jurnal Ekonomi & Studi Pembangunan*, vol. 20, no. 1, 2019, doi: 10.18196/jesp.20.1.5017.

[9] M. Filgueiras, M. Amorim, A. Lobo, A. Couto, M. Kern, and S. Ferreira, "Adapting to the new normal: Understanding public transport use and willingness-to-pay for social distancing during a pandemic context," *Journal of Urban Mobility*, vol. 5, Jun. 2024, doi: 10.1016/j.urbmob.2024.100079.

[10] P. Prasetya and A. Tjitra Handayani, "Analisis Ability To Pay (ATP) dan Willingness To Pay (WTP) Kereta Kompartemen (Studi Kasus: KA Semeru Relasi Yogya-Gambir PP)," vol. 9, pp. 3737–3743, 2025, [Online]. Available: https://www.kompas.tv/ekonomi/462119/ini-

[11] A. Umme, K. Aya, and K. Hisashi, "Exploring the Commuters' Willingness-to-Pay and Its Influencing Factors for an Improved Public Bus Service in Dhaka City," *J. Transp. Technol.*, vol. 13, no. 02, pp. 139–157, 2023, doi: 10.4236/jtts.2023.132007.

[12] H. Ayan, M. Bell, and D. Dissanayake, "Investigating the Factors That Influence the Ridership of Light Rail Transit Systems Using Thematic Analysis of Academic Literature," *Future Transportation*, vol. 5, no. 1, Mar. 2025, doi: 10.3390/futuretransp5010022.

[13] A. Daly, S. Hess, and J. de Dios Ortúzar, "Estimating willingness-to-pay from discrete choice models: Setting the record straight," *Transp. Res. Part A Policy Pract.*, vol. 176, p. 103828, 2023, doi: https://doi.org/10.1016/j.tra.2023.103828.

[14] M. Movaghar, E. Jenelius, and D. Hunter, "Machine learning framework to estimate ridership loss in public transport during external crises: case study of bus network in Stockholm," *European Transport Research Review*, vol. 17, no. 1, Dec. 2025, doi: 10.1186/s12544-025-00722-z.

[15] H. Zhang, L. Zhang, Y. Liu, and L. Zhang, "Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method," *Sustainability (Switzerland)*, vol. 15, no. 14, Jul. 2023, doi: 10.3390/su151411414.

[16] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.

[17] M. Tabasi, J. M. Rose, A. Pellegrini, and T. Hossein Rashidi, "An empirical investigation of the distribution of travellers' willingness-to-pay: A comparison between a parametric and nonparametric approach," *Transp. Policy (Oxf).*, vol. 146, pp. 312–321, Feb. 2024, doi: 10.1016/j.tranpol.2023.12.006.

[18] M. Philip, "Measurement and analysis of change in research scholars' knowledge and attitudes toward statistics after PhD coursework," *BMC Med. Educ.*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12909-024-05487-y.

[19] Y. Dong, "Descriptive Statistics and Its Applications," 2023.

[20] S. McGrath *et al.*, "Estimating the sample mean and standard deviation from commonly reported quantiles in meta-analysis," *Stat. Methods Med. Res.*, vol. 29, no. 9, pp. 2520–2537, Sep. 2020, doi: 10.1177/0962280219889080.

[21] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y. I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, Aug. 2024, doi: 10.3390/math12162553.

[22] P. Koukaras and C. Tjortjis, "Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices," Oct. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ai6100257.

[23] A. Durap, "Data-driven models for significant wave height forecasting: Comparative analysis of machine learning techniques," *Results in Engineering*, vol. 24, Dec. 2024, doi: 10.1016/j.rineng.2024.103573.

[24] R. G. Pensa, A. Crombach, S. Peignier, and C. Rigotti, "Explaining Random Forest and XGBoost with Shallow Decision Trees by Co-clustering Feature Importance," *Mach. Learn.*, vol. 114, no. 12, Dec. 2025, doi: 10.1007/s10994-025-06932-9.

[25] "Exploratory Data Analysis of International Student Demographics: Trends, Insights, and Implications," *Journal of Artificial Intelligence and Metaheuristics*, vol. 7, no. 2, 2024, doi: 10.54216/jaim.070204.

[26] R. Guevara-Reyes, I. Ortiz-Garcés, R. Andrade, F. Cox-Riquetti, and W. Villegas-Ch, "Machine learning models for academic performance prediction: interpretability and application in educational decision-making," *Front. Educ. (Lausanne).*, vol. 10, 2025, doi: 10.3389/feduc.2025.1632315.

[27] M. Tabasi, J. M. Rose, A. Pellegrini, and T. Hossein Rashidi, "An empirical investigation of the distribution of travellers' willingness-to-pay: A comparison between a parametric and nonparametric approach," *Transp. Policy (Oxf).*, vol. 146, pp. 312–321, Feb. 2024, doi: 10.1016/j.tranpol.2023.12.006.

[28] V. Vichiensan, V. Wasuntarasook, S. Malaitham, A. Fukuda, and W. Rujopakarn, "Willingness to Pay for Station Access Transport: A Mixed Logit Model with Heterogeneous Travel Time Valuation," *Sustainability (Switzerland)*, vol. 17, no. 15, Aug. 2025, doi: 10.3390/su17156715.

[29] A. Leung, C. Burlando, and T. Pavanini, "Market segmentation and willingness to pay for public transport annual passes among older adults: insights from Genoa, Italy," *Research in Transportation Business and Management*, vol. 58, Jan. 2025, doi: 10.1016/j.rtbm.2024.101243.

[30] M. Hassan, M. E. Kabir, S. T. Akter, S. S. Shraban, K. S. Basaruddin, and M. A. Islam, "Machine learning in travel mode choice studies: A systematic literature review of applications, methods, and challenges," Dec. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.rineng.2025.108140.

.