



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

Machine Learning Based Prediction of Passenger Demand, Dynamic Pricing and Seat Availability in Bus Transportation

Ambika U, Dr. Hemanth Kumar

P.G. Student, Department of Master of Computer Applications, JNN College of Engineering, Shivamogga,
Karnataka, India

Associate Professor, Department of Master of Computer Applications, JNN College of Engineering, Shivamogga,
Karnataka, India

ABSTRACT: Public transportation plays significant role in connecting people across many cities. Passenger may face struggle with fluctuating price mainly during weekends, holiday and link holidays. Bus operators find it difficult to forecast percentage of seat occupancy. It supports various bus types such as Ac sleeper, Non Ac sleeper, Ac seater, Semi-sleeper bus type etc. Goal of this works is to assist the users by forecasting future fare fluctuations that facilitate them to plan their travel. System also assists bus operators to earn more revenue by changing fare. Past data is processed to train Random Forest model. Festival, holidays, route, available seats and so on features included. System provides bus ticket booking after forecasting. User can reserve ticket by selecting bus travels from bus list according to their convenience. System include discount scheme. Discounts is applied automatically while booking tickets. Discounts such Rs.100 off on more ticket price then Rs.1000, Rs.50 off on 5+ seats etc. while booking tickets, passengers have to provide details such as email, phone number, name, gender. After payment is made passenger receive booking confirmation. Among Decision Tree and Random Forest Regressor, in which Random Forest Regressor provide better accuracy prediction. Overall, this system offers a user friendly interface that facilitate faster booking.

KEYWORDS: Random Forest Regressor, Dynamic Pricing, Ticket Booking.

I. INTRODUCTION

Public Transportation plays major role in society. Travellers face various difficulties when they are planning a trip. Travelling during festivals, holidays, weekends or during rush hours, passengers miss the opportunity to get tickets due to the changing ticket prices and sitting availability. There are also cases of Bus operators and organizations that have trouble maximizing seat occupation. These are some of the reasons that lead to under booking hence loss of income to bus companies and also disappointments to passengers who end up missing their travel schedules. As opposed to a fixed price of tickets, this system allows operators to adjust the price of tickets dynamically based on demand. Using historical data, the system can predict pricing on a dynamic manner and availability of seats. Historical data used parameters like previous dates, times, prices, bus types, routes, festivals, days, etc. These parameters are major influencers of ticket prices and seats occupied. Availability of seats is also employed as an input feature for forecasting ticket prices.

In this work, we apply Machine Learning techniques such as Random Forest Regressor and Decision Tree models to predict outcomes. Random Forest Regressor works better by producing several decision trees and predicting the outcome based on the combining of all the decision trees. This model good at dealing with numerous data.

In this work, the predictions are presented by an interface developed using the customtkinter library. Travel details have to be given by the passengers to make predictions. Django is used to develop booking system. Overall the system integrate Machine Learning technique and Django. Passenger can book tickets easily. Booking details are stored into database.

II. LITERATURE SURVEY

M. Nithin Varma, Manikandhan developed a machine learning model to make prediction of ticket price dynamically instead of fixed pricing [1]. They compared Random Forest and Support Vector Machine (SVM) algorithm in which Random Forest predict price with more accurate as compared to SVM. Their aim is to help Bus Company to set ticket price and increase income. Authors in [2] developed smart bus fare ticketing system using cloud computing. System suggest ticket price based on passenger demand. All the predictions and suggestions are stored in cloud. Passenger buy the ticket online which generates unique QR code. System updates the passenger count in cloud in real time when they scan QR code. Nandini Nagaraj et al. used a deep learning techniques to forecasting how many passenger will travel by buses [3]. Their objective is to reduce overcrowd and help bus companies to schedule bus properly based on passenger demand. Greedy Layer Wise, Long-Short Term Memory, recurrent neural network are used. Authors in [4] developed mathematical model called deterministic linear programming or DLP to set the prices dynamically. They applied dynamic pricing on Bus Company rather than only on one bus called network revenue management. Francesco Branda et al. designed a methodology called DA4PT (Data Analytics for Public Transport) which analyses past booking logs to predict passenger demand and adjust ticket price [5]. Methodology followed by process mining, discover purchase factor, prediction model. Work by Tatsuya Yamamura used Deep Learning technique that is Multiple-Long Short Term Memory to forecast bus ridership [6]. He collected past ridership information, weather condition, time etc. Each Long-Short Term Memory layer learns different past ridership data from different stops. Authors in [7] developed a Multilayer Perceptron Regression model to analyse data. Their aim is to forecast passenger count accurate even before 2 hours of departure. They collected past details like passenger count, holidays, time, day of week and also weather information like temperature, rain to predict passenger count. Challenges faced by authors are unexpected strikes, choosing weather information, handling missing data. Pasqual Marti et al. designed Machine Learning to forecast passenger demand and used Multi Agent Simulation which helps to test their different planning scenarios to find out efficient solution [8]. They collected real- world data from urban bus system and methodology followed by pre-processing, model training and evaluation. Random Forest, Support Vector Regression, Neural Network models are used. Authors in [9] developed model uses real data from public transits system. Next-stop, Origin-designation-pair techniques are used for prediction. This work include Linear Regression and Random Forest model. Joko Siswanto et al. used real time data for forecasting the number of passenger travel [10]. Data collection, preprocessing, train the model, then test model to make forecasting passenger flow. Dayakar, N. Venkatesh proposed passenger flow prediction system that can handle multi-source data [11]. They used stacked models which are different simple models for preprocessing, feature extraction. Then Gradient Boosting Decision Tree model is used to make prediction. Challenges faced by authors are low prediction accuracy with simple model, Features selection and weighting. Results archived by authors are Bus scheduling, Route planning, Passenger convenient. Work by Sindu D developed a model that employs price multiplier which dynamically modify price of ticket [12]. They manipulate bus ticket fare using number which is called price multiplier. They consider aspects such as forecasting such as seat availability, Booking rush and last minute booking and other considerations. In [13] authors developed hybrid approach. Two methods that have been used in training and adjusting the model include Nadam optimizer and Stochastic Gradient Descent optimizer. It also uses sliding window technique in recalling recent passenger demand and to improve the prediction accuracy. Absalom Ezugwu in [14] proposed a model that can be used to examine passenger demand. This project involved cloud platform in order to save whole data and forecasting. They developed web dashboard to track buses, predicted passenger data and plan buses. Authors in [15] used data such as date, route info, weather condition, time. Aim is used to improve bus transport efficiency. Data is preprocessed using Wavelet Packet Decomposition, Attention mechanism is used to focus on important information and Bi- Long Short Term Memory for prediction.

III. METHODOLOGY

Methodology used in this work is Machine Learning. Machine Learning is a way that the computers learns from the data and make the prediction on their own not being definitely programmed.

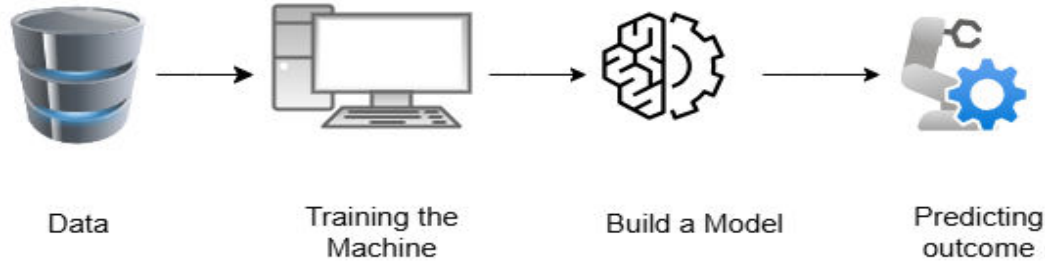


Fig. 1 Machine Learning

Fig. 1 describes working principle of Machine Learning. Machine Learning involves three types, they are: Supervised, Unsupervised and Reinforcement Learning

- **Supervised Learning:** In Supervised Learning, model is trained using input data with similar output. Then we tested the model by asking it to predict the output using new dataset.
- **Unsupervised Learning:** In Unsupervised Learning, the model is trained using the unlabelled dataset to find patterns and group similar data.
- **Reinforcement Learning:** In Reinforcement Learning, by interacting with environment to optimize rewards, an agent gains decision-making skills.

In this work, Random Forest Regressor and Decision Tree models are used to implement to make prediction. However Decision Tree is simple and frequently suffering as a result of overfitting. Whereas Random Forest algorithm reduce overfitting and increased stability resulting in higher accuracy. Random Forest make decision by combining multiple decision tree as compared to the single Decision Tree. Other techniques such as Support Vector Machine, K- Nearest Neighbour are not capable for forecasting continuous value on huge dataset. Long Short Term Memory requires more computational power and vast dataset.

3.1 Random Forest

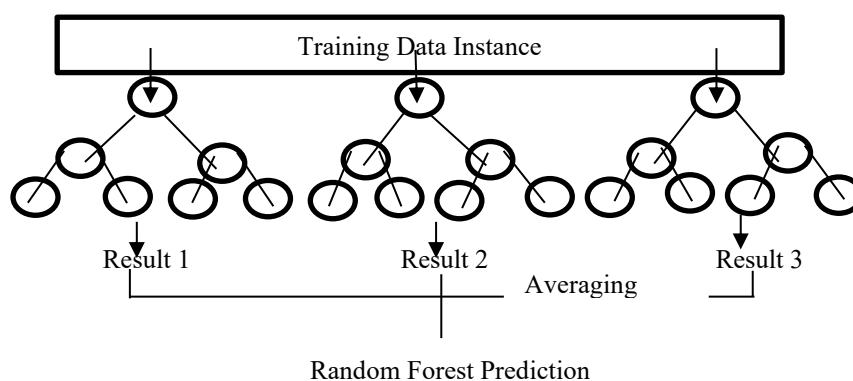


Fig. 2 Random Forest

Fig. 2 describes working of Random Forest. Random Forest is developed by forming many decision trees. Random Forest make the forecasting by all results of many decision trees. Random Forest works better as compared to Decision Tree because in Decision Tree only one tree will be formed so that Random Forest minimize overfitting. In Random Forest as name indicate the random rows with replacement is selected to build many decision trees. Here each decision tree may does not learn same data.

3.2 Decision Tree

Fig. 3 illustrate working of Decision Tree. Decision Tree algorithm is belongs to Supervised Learning. In this model works by splitting the data into subtrees to form tree like structure. Splitting data is done based on features. For each split it calculate Mean Square Error. It compares MSE of each feature, if Means Square Error is lower it will consider that feature as best split.

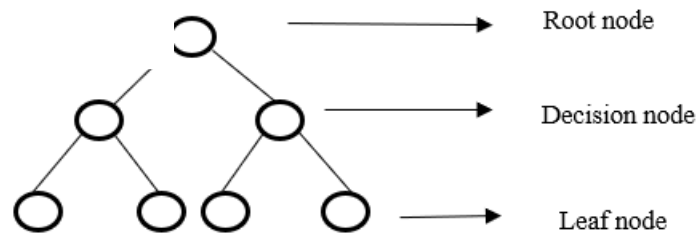


Fig. 3 Decision Tree

3.3 Django

Django framework is used to design web based application. It follows MVT (Model View Temple) architecture

3.4 Data flow Diagram:

Fig. 4 describes Data flow diagram.

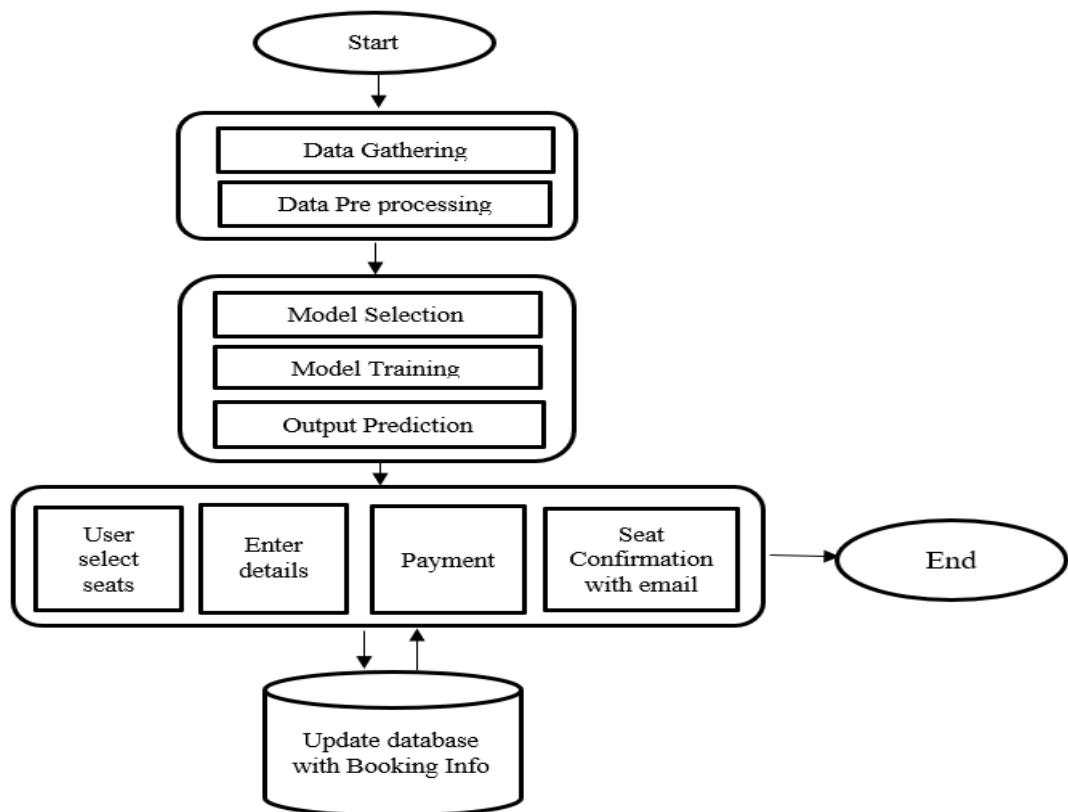


Fig. 4 DFD

The system starts with data gathering, data pre-processing, model development, training model, output prediction, booking module as follows:

Data Collection: The dataset is used in this work includes historical data about passenger flow and seat availability. It contain date, time, route, price, bus type, price, total seats, available seats, day of week, week days, weekends, festivals, linking holidays and other factors are considered.

Data Pre-processing: Data pre-processing involves converting text categories that is route, festivals, holidays etc. into numbers. Label encoder technique is used to convert categorical values into number. Systemizing formats for date and time.

Train and Test model: Random Forest Regressor is used for training and testing the model. In this work dataset is divided into 80% for training the model and 20% for testing the model. Based on trained data, model make prediction of unseen or testing dataset.

Output Prediction: In this interface, user allows to enter date, time, route, bus- type and day of week will be automatically generated based on date. This interface shows prediction of dynamic price and number of seat available by analysing past data.

Booking module: Django application is designed for allowing the user to book the bus tickets. Here user can able to choose bus travels as per their convenient departure time and price. Seat Layout displays different arrangement for different bus type. User can select comfortable seats and continue booking process. In additional, first row of each buses can be able to booked by females only. User needs to be provide details while booking. Discounts are automatically applied to make booking faster. Ticket price is vary based on seat availability. Booking process is followed by initiate payment and booking confirmation will send to user through email.

Database: SQLite 3 is used in this work. Booking information is stored in database.

3.5 Tools and Libraries:

Table 1 illustrate tools and libraries are utilized in this work illustrated in the below:

Table 1 Tools and Libraries

Tools and Libraries	Purpose
Python 3.8	Scripting and Backend logic
Pandas	To upload dataset
Scikit-learn	Training, testing and evaluating model
Customtkinter	GUI
Webbrowser	To handle URLs directly from python
Joblib	To save model

IV. NUMERICAL ANALYSIS

The Performance of model evaluated using standard matrices: R^2 (Coefficient of Determination), MSE (mean square error) and RMSE (Root Mean Square Error) for both Random Forest Regressor and Decision Tree as shown below:

MSE (Mean Square Error): It measure variation between actual value and prediction values.

$$mse = \frac{1}{N} \sum_{i=1}^n (A_i - \hat{P}_i)^2 \quad \text{eq. (1)}$$

In eq. (1),

- N - Taking average gives overall error.
- A_i - Actual value.
- \hat{P}_i - predicted value.

RMSE (Root Mean Square Error): It is square root of mean square error.

$$rmse = \sqrt{mse} \quad \text{eq. (2)}$$

In eq. (2),

mse – Mean Square Error.

R² (Coefficient of Determination): Measure how well model describe the difference in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^n (A_i - \hat{P}_i)^2}{\sum_{i=1}^n (A_i - \bar{A})^2} \quad \text{eq. (3)}$$

In eq. (3),

- A_i - Actual value.
- \hat{P}_i - Predicted value.
- \bar{A} - Mean of actual value.
- 1 - Perfect prediction.
- 0 - Model does not better than mean.

Table 2 shows model performance evaluation between Decision Tree and Random Forest.

Table 2 Performance evaluation

Models	MSE	RMSE	R ²
Random Forest(Price)	25967.9	161.15	0.72
Decision Tree(Price)	48949.4	221.2	0.48
Random Forest(Seats)	108.34	10.41	-
Decision Tree(Seats)	210.93	14.52	-

V. EXPERIMENTAL RESULTS

System integrates Machine Learning and Django framework for prediction. Random Forest works accurate as compared to Decision Tree. In some instance, prediction price varied slight because of sudden demand or other factor that may not be include. Seat availability prediction faced slight challenges because of randomness in booking.

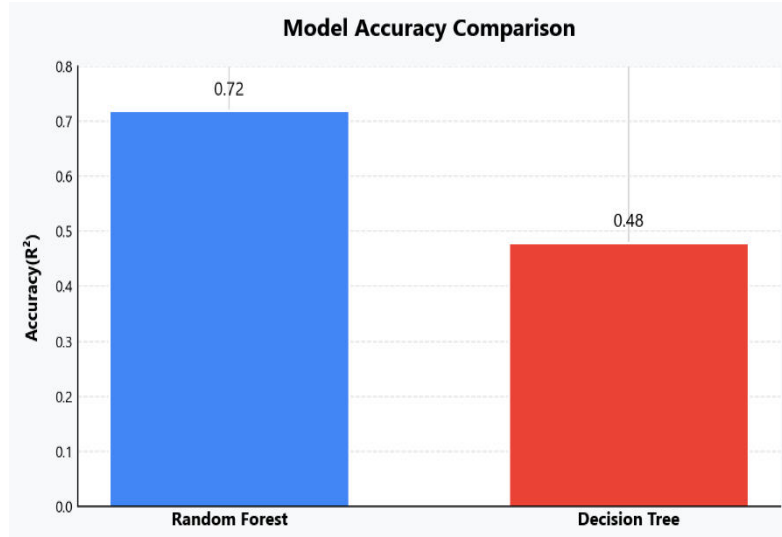


Fig. 5 Accuracy Comparison

Fig. 5 illustrates accuracy comparison between Random Forest and Decision Tree but applying R^2 . In this work Random Forest and Decision Tree models are used. Random Forest Regressor archived 72% accuracy and Decision Tree archived 48% accuracy.

Fig. 6 displays actual price versus prediction price by Random Forest and Decision Tree. Dotted black line represents perfect prediction that matches actual price and prediction price. Blue dot represent Random Forest which indicated that it gives more accurate prediction of price as compared to Decision Tree. Decision Tree is represented by red dots.



Fig. 6 Actual vs Prediction Price

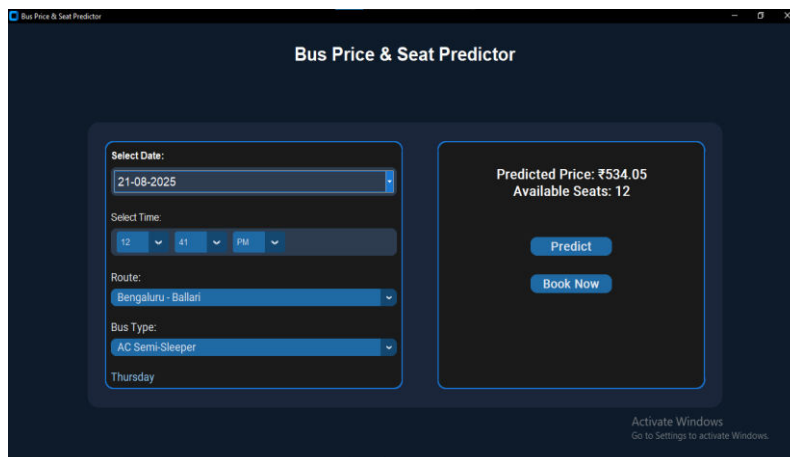


Fig. 7 Predictive interface

Fig. 7 show the interface where user enter details of route, date, time, bus type and it make prediction of number of seats available and price prediction.

VI. CONCLUSION

Bus transportation is most commonly used and plays major role in public transportation. In this work, before actual booking done will predicts price and seat availability which make booking process efficiently. Booking system is designed and developed by combining Machine Learning and Django. System makes prediction and displayed in interface. System allow user to check price and seat availability and can book ticket easily. This system is helpful for both bus companies and as well as passenger for booking bus tickets.

REFERENCES

1. M. Nithin Varma, and Manikandhan “Predictive analysis of bus ticket price by using random forest comparing with support vector machine”, Hybrid and Advanced Technologies, Taylor & Francis, pp. 487–492, Jan. 2025.
2. M. Vadivel, V. Bini Marein, S. Balasubramani, S. Hemalatha, S. Murugan and S. Velmurugan, “Cloud-Based Passenger Experience Management in Bus Fare Ticketing Systems using Random Forest Algorithm”, IEE Int. Conf. Reliability, Infocorn Technologies and Optimization(ICRITO), pp. 1-6, Mar. 2024.
3. N. Nagaraj, H. L. Gururaj, B. H. Swathi, and Y.-C. Hu, “Passenger flow prediction in bus transportation system using deep learning,” Multimedia Tools and Applications, Vol. 81, pp. 12519–12542, Apr. 2022.
4. C. Barz, S. Laumer, M. Freyschmidt, and J. Martínez Blanco, “Discrete dynamic pricing and application of network revenue management for FlixBus,” Journal of Revenue and Pricing Management, Vol. 22, pp. 16–33, Feb. 2023.
5. F. Branda, F. Marozzo, and D. Talia, “Ticket Sales Prediction and Dynamic Pricing Strategies in Public Transport,” Big Data and Cognitive Computing, Vol. 4, Article 36, Nov. 2020.
6. T. Yamamura, “Bus Ridership Prediction with Time Section, Weather, and Ridership Trend Aware Multiple LSTM”, Pervasive Computing and Communications, pp. 509–514, Mar. 2023.
7. T. Fontes, R. Correia, J. Ribeiro, and J. L. Borges, “A Deep Learning Approach For Predicting Bus Passenger Demand Based On Weather Conditions”, Transport and Telecommunication Journal, Vol. 21, pp. 255–264, Dec. 2020.
8. P. Martí, A. Ibáñez, V. Julián, P. Novais, and J. Jordán, “Bus Ridership Prediction and Scenario Analysis through ML and Multi-Agent Simulations”, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, Vol. 13, Dec. 2024.
9. J. Wood, Z. Yu, and V. V. Gayah, “Development and evaluation of frameworks for real-time bus passenger occupancy prediction”, International Journal of Transportation Science and Technology, Vol. 12, Mar. 2022.
10. J. Siswanto, D. Manongga, I. Sembiring, and S. Wiyono, “Deep Learning Based LSTM Model for Predicting the Number of Passengers for Public Transport Bus Operators,” JOIN (Jurnal Online Informatika), Vol. 9, pp. 18–28, Jun. 2024.

11. C. H. Dayakar and N. Venkatesh, "Machine Learning Based Bus Passenger Flow Prediction System," Journal of Engineering Sciences, Vol. 15, pp. 638–643, Aug. 2024.
12. D. Sindu, "Dynamic pricing model using price multipliers for online bus ticket booking," Kongunadu Research Journal, Vol. 1, pp. 23–30, Dec. 2014.
13. Y. Han, C. Wang, Y. Ren, S. Wang, H. Zheng, and G. Chen, "Short Term Prediction of Bus Passenger Flow Based on a Hybrid Optimized LSTM Network," ISPRS International Journal of Geo Information, Vol. 8, Article 366, Sep. 2019.
14. M. Skhosana and A. E. Ezugwu, "A Real-Time Machine Learning-Based Public Transport Bus-Passenger Information System", International Journal of Sustainable Development and Planning, pp. 1221-1238, Vol. 16, Nov. 2021.
15. Y. Pei, S. Ran, W. Wang, and C. Dong, "Bus Passenger Flow Prediction Model Based on WPD, Attention Mechanism, and Bi LSTM," Sustainability, Vol. 15, Oct. 2023.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details