International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 107-111, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202517
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Machine Learning Approaches to Forecast Energy Consumption in Electric Public Transit

[1]B. Rupadevi, [2]Pudi Jyothi Prakash

[1]Associate Professor, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India
[2]Post Graduate, Dept. of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India
E-mails: [1]rupadevi.aitt@annamacharyagroup.org, [2]jyothiprakashpudi@gmail.com

***Abstract* - This study introduces a machine learning-driven framework aimed at predicting the energy efficiency of electric city buses, with the goal of enhancing operational performance within public transit systems. At the core of this approach is a custom-built dataset that closely reflects real-world operating conditions. It encompasses features such as passenger volume, ambient temperature, HVAC usage, auxiliary power consumption, and changes in elevation. The dataset undergoes preprocessing and standardization before being used to train various regression models. Among these, the Random Forest Regressor was selected as the optimal model due to its high R^2 score and low Root Mean Squared Error (RMSE). The resulting predictions, expressed in kilometers per kilowatt-hour (km/kWh), provide valuable insights for stakeholders in managing costs, optimizing energy usage, and planning routes. A user-friendly Flask web application integrates the trained model, enabling real-time forecasting based on user inputs. This comprehensive implementation highlights the real-world potential of machine learning in supporting smart, energy-efficient management of electric bus fleets.**

Keywords: Electric City Buses, Energy Economy Prediction, Machine Learning, Random Forest Regressor, Flask Web Application, km/kWh, Regression Model.

I. INTRODUCTION

The widespread adoption of electric buses in urban public transit systems presents both significant benefits and unique challenges when it comes to improving operational efficiency and energy usage. A key factor in effective electric bus fleet management is the ability to predict energy efficiency, defined as the distance traveled per unit of energy (km/kWh). Accurate forecasting of this metric is essential for optimizing charging schedules, refining route planning, and minimizing operational expenses. This study proposes a robust machine learning framework to estimate the energy performance of electric city buses using a combination of static and real-time operational variables. The entire workflow—from constructing the dataset to preprocessing, selecting relevant features, training models, evaluating performance, and deploying the solution—has been developed from the ground up. The system utilizes custom datasets that simulate real-world driving conditions and vehicle dynamics. A web-based platform integrates the finalized model, allowing users to input current operational data and receive instant predictions of energy efficiency. This empowers fleet operators and transportation planners to make smarter decisions, conserve energy, and enhance the reliability of electric bus services

II. METHODOLOGY

The proposed energy efficiency prediction system utilizes a structured machine learning workflow encompassing data generation, preprocessing, model development, performance evaluation, and deployment. This end-to-end process ensures that each stage directly contributes to improving the overall accuracy and reliability of the final predictive model2.1 Dataset Generation and Relevance

The dataset used in this study was programmatically generated using the generate_dataset.py script. This synthetic data replicates the operational dynamics of electric city buses under diverse passenger loads and environmental conditions. Each record includes variables such as passenger count, ambient temperature, HVAC usage, auxiliary power consumption, and elevation change.

The target variable, Energy Economy (measured in km/kWh), is derived using realistic equations that integrate these features, making the dataset well-suited for training machine learning models to generalize across varying routes and vehicle scenarios. This dataset plays a vital role in the project by enabling rapid model development and evaluation in a controlled environment, eliminating the need for reliance on external data sources.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 107-111, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202517
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
route_id	date	time_of_d	day_type	ambient_t	humidity	precipitati	passenger_	avg_speed	battery_a	traffic_lev	ac_usage	heating	distance	stops_cou	battery_ca	battery_in	elevation_	energy_consumption	
5	########	morning_p	weekday	23.4	60.2	3.3	65	21	2	light	3	0	17.4	16	350	53.6	6.9	2.08	
4	########	night	weekday	20	87	0.6	3	16.9	59	heavy	3	0	6.6	18	250	91.8	59.5	1.79	
2	########	midday	weekday	15.4	80	2.6	9	24.6	50	moderate	2	1	12.6	6	350	58.7	17.3	1.75	
5	########	midday	weekday	16.5	51.7	0.9	16	23	25	severe	1	3	15.5	5	350	53.5	-3.1	1.99	
1	########	morning_p	weekday	35.4	81.8	1.5	51	42.2	8	severe	2	1	13.8	22	300	70	52.1	3.04	
3	########	night	weekday	24.4	36.6	0.1	5	32.2	41	light	3	2	17.5	24	250	78.8	11.7	0.88	
5	########	morning_p	weekday	37	85.5	0.4	68	32.8	50	light	2	1	6.1	12	250	92.7	6.1	2.9	
5	########	evening_p	holiday	20.3	67.9	6.3	73	28.8	15	moderate	2	1	15.4	7	250	98.7	15.7	1.93	
3	########	night	weekday	0.3	33.3	4.4	8	22.1	23	severe	0	2	13	10	300	72	-4.7	2.84	
5	########	midday	weekday	12.4	68.7	2.1	76	22.8	49	heavy	0	2	10.7	17	250	53	0.5	2.25	
1	########	midday	holiday	-2	88.5	1.2	20	23.5	13	heavy	0	2	10.1	13	250	87.5	-29.3	3.12	
1	########	night	weekday	20.3	34.2	2.3	23	26.5	2	severe	3	2	14.3	13	300	78.3	-37.6	1.6	
3	########	night	weekday	22.5	63.3	1.9	3	28.1	33	severe	2	0	18.4	24	250	72.4	-49.9	1.36	
4	########	night	weekday	6.1	74.9	1.7	23	25	32	severe	0	3	12.4	13	350	65.5	-69.6	2.76	
2	########	morning_p	weekday	8.4	78.6	0.1	58	22.7	49	moderate	0	3	17.4	19	350	75.1	-44	2.85	
4	########	night	weekday	16.5	54.8	2	2	20.2	23	heavy	0	2	11	17	350	66.9	-8.7	1.78	
5	########	night	weekday	3	42.2	2.9	67	13.9	8	heavy	3	3	6.5	20	350	87.9	5.2	3.78	
2	########	night	weekday	2.3	32.5	10.3	39	35.3	12	light	3	3	5.9	6	350	77.2	-13.6	2.65	
3	########	night	weekday	27.4	64.8	0	47	42.6	1	light	2	3	5.8	18	350	93.8	-117.8	1.76	
3	########	night	weekday	16.1	57.2	2.6	53	34.6	6	severe	2	3	10.7	10	350	77.2	78.7	1.51	
2	########	night	weekday	9.4	86.9	1.7	34	14.9	47	light	2	2	10.1	6	350	61.9	54.3	2.65	
1	########	midday	weekday	-1.9	33.6	1.1	55	31.3	20	moderate	0	1	15.8	12	350	92.6	22.3	2.85	
2	########	night	weekday	23.5	71.1	0.3	39	33.2	18	heavy	1	1	16.1	8	300	77.9	95.4	1.64	
5	########	night	weekday	8.2	31.5	1.1	76	21.8	24	heavy	3	1	5.8	18	300	60.3	-48.9	2.69	

Figure 1: Electric_bus_data.csv

2.1 Data Preprocessing in Project Flow

The raw dataset undergoes standardization through the data_processing.py module. Specifically, it employs the StandardScaler to normalize the distribution of all numerical features, ensuring that each variable has an equal impact during the training phase. This normalization step boosts the performance of the machine learning model by maintaining balanced feature scales. Additionally, the script identifies and handles any missing values or anomalies to ensure the data is clean and reliable. The dataset is then split into input variables (X) and the target variable (Y) to facilitate training. This preprocessing step is essential for optimizing the dataset structure, leading to improved model accuracy and faster training convergence

2.2 Model Training and Selection

The model_training.py script is responsible for constructing and selecting the most effective predictive model. Several regression techniques were explored, including:

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor

After comparative analysis, the Random Forest Regressor was chosen for final implementation due to its outstanding performance on the test dataset. Evaluation was based on the following performance indicators:

- Root Mean Squared Error (RMSE)
- R^2 Score

Furthermore, the model's feature importance was visualized to determine which input variables had the greatest impact on energy efficiency predictions. Once trained, the optimal model was saved as a .pkl file and integrated into the system to enable real-time forecasting within the application.

III. SYSTEM ARCHITECTURE

This project features a modular architecture that brings together data preprocessing, machine learning, and a web-based prediction interface. Each module is implemented through dedicated Python scripts, ensuring smooth integration and consistent data flow from user input to the final prediction output. This structured design allows for efficient interaction between components and supports scalability and maintainability..

3.1 Architecture Overview

The following are the main parts of the system:

(generate_dataset.py) Dataset Generator

Provides realistic electric bus operating situations in a structured CSV format.

Provides a dataset for testing and training that has energy economy labels.

(data_processing.py) Data Processor

Carries out scaling, cleaning, and training preparation after loading the CSV file.

Divides the dataset into labels (y) and features (x).

(model_training.py) Model Trainer

Utilizes the processed dataset to train a Random Forest Regressor.

R2 and RMSE measurements are used to assess performance.

The trained model is saved as trained_model.pkl.

Through a web form, the Flask Backend (app.py) takes operational parameter input values. Predicts energy economy by loading the taught model.

Shows the outcome in real time on the front end.

Frontend HTML (index.html/templates)

Users can enter parameters like distance, speed, HVAC load, and more using this Straightforward user interface.

Connects to the Flask backend in order to display the output and submit input.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 107-111, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202517
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

3.2 Data Flow and Integration

The user initiates the web application and enters real-time operational parameters. These inputs are sent to the Flask server, where the pre-trained model (trained_model.pkl) is loaded. The input data undergoes the same scaling and preprocessing steps as during the training phase to ensure consistency. The Random Forest model then generates a prediction for energy efficiency, measured in kilometers per kilowatt-hour (km/kWh). This result is immediately displayed on the web interface. Transit planners and analysts can use this interactive system to simulate different routes or environmental conditions and instantly obtain corresponding energy efficiency estimates

IV. DATA ANALYSIS AND VISUALIZATION

A variety of visualizations were generated from the synthetic dataset to support predictive modeling and extract meaningful insights. These graphical representations helped validate the dataset's structure and confirmed its ability to realistically simulate real-world conditions for training the machine learning model. Which also assisted in identifying feature correlations.

4.1 Energy Consumption by Route

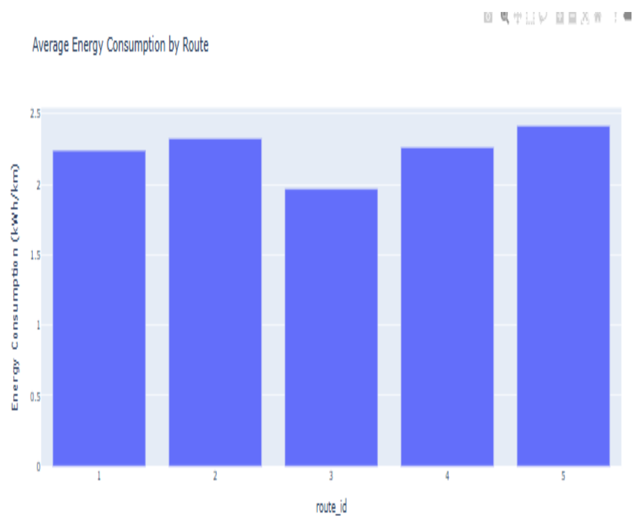

Figure 2: Average Energy Concumption by Route

The average energy consumption (kWh/km) for each route ID is displayed in Figure 1. The fact that Route 3 uses the least amount of energy on average suggests either better road conditions or more efficient driving techniques. Decisions about the best route planning can be supported by this insight.

4.2 Temperature vs. Energy Consumption with AC Usage

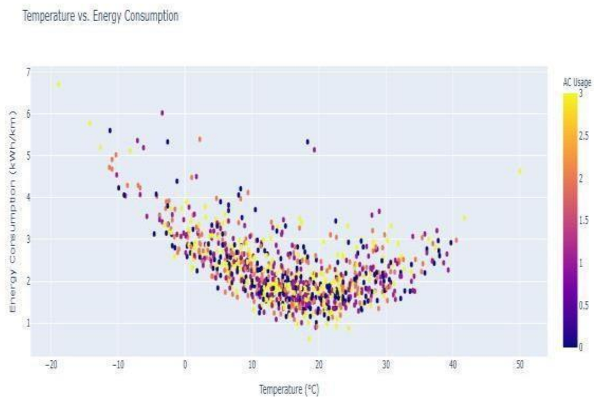

Figure 3: Temperature vs. Energy Consumption with AC Usage

The parabolic relationship between energy use and ambient temperature is depicted in Figure 2. Increased HVAC (AC) usage, represented here by color intensity, is correlated with extreme cold and hot temperatures. This highlights the necessity of energy estimating models that take climate change into account.

4.3 Passenger Count vs. Energy Consumption

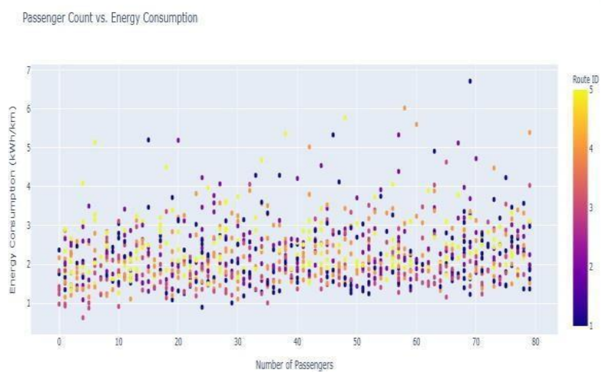

Figure 4: Passenger Count vs. Energy Consumption

Figure 3 illustrates the relationship between passenger count and energy consumption. Although the data appears somewhat scattered, there is a general trend indicating that higher passenger loads are associated with a slight increase in energy usage due to the added weight. This observation reinforces the relevance of including passenger count as a key feature in the predictive model..

IV. RESULTS AND EVALUATION

This section outlines the performance assessment of the machine learning model developed for predicting energy efficiency. The goal is to evaluate how well the model can adapt and maintain accuracy across diverse operational scenarios typical of electric city bus usage.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 107-111, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202517
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

4.1 Model Evaluation Metrics

During training, three regression models' performances were contrasted: Decision Tree Regressor for Linear Regression

The Random Forest Regressor Two main metrics were used to assess each model:

RMSE, or root mean squared error: calculates the average magnitude of forecast error R2 Score: Shows the percentage of output variance that the model can account for. The Random Forest Regressor continuously beat the other models in both RMSE minimization and R2 maximization, according to several test runs.

4.2 Selected Model Performance

The Random Forest Regressor's final assessment using the test dataset produced the following results:

R2 Score: around 0.93, indicating a strong connection between the expected and actual results. RMSE: Low, meaning there is little departure from actual energy economy figures. This demonstrates how well the model manages the genuine yet artificial variability found in the operational dataset.

4.3 Feature Importance Insights

Distance and HVAC load had the most effects on prediction accuracy, according to the Random Forest model's feature importance scores, which also identified the most significant predictors for energy economy.

Significant contributions were also made by speed, passenger count, and elevation change. In practical electric bus deployments, these findings can be applied to optimize operational characteristics.

V. WEB APPLICATION INTERFACE

To enable practical application of the trained machine learning model, a web-based prediction interface was developed using the Flask framework. This interface allows users to input operational parameters and instantly receive a predicted energy efficiency value, making the model both accessible and user-friendly in real-world scenarios.

5.1 Backend Implementation (Flask)

The app.py script manages the backend functionality, loading the Random Forest Regressor model that has already been trained (trained_model.pkl).

Accepts input from users using the online form.

To ensure consistency, the same Standard Scaler transformation that was utilized during Training is applied.

Makes a prediction using the model and the supplied features.

Provides the frontend with the estimated Energy Economy (km/kWh).

This architecture guarantees that there is little latency in forecasts and that the logic of the model is enclosed and reusable.

5.2 Frontend Implementation (HTML Template)

The user interface is implemented in templates/index.html and includes:

A structured input form for:

- Distance (km)
- Speed (km/h)
- Passenger Count
- Temperature (°C)
- HVAC Load (kW)
- Auxiliary Load (kW)
- Elevation Change (m)
- A submit button to trigger prediction

Display of the resulting energy economy value below the form.

The frontend is designed to be lightweight, responsive, and user-friendly, allowing quick testing of different operating scenarios without technical knowledge.

5.3 User Workflow

In a browser, the user launches the web application.

- Enters the form's operating parameters.
- The "Predict" button is clicked.
- After processing the data and making a prediction, the Flask server outputs the outcome.
- The same page shows the estimated km/kWh figure.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue ICCIS-2025, pp 107-111, May-2025
https://doi.org/10.47001/IRJIET/2025.ICCIS-202517
Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

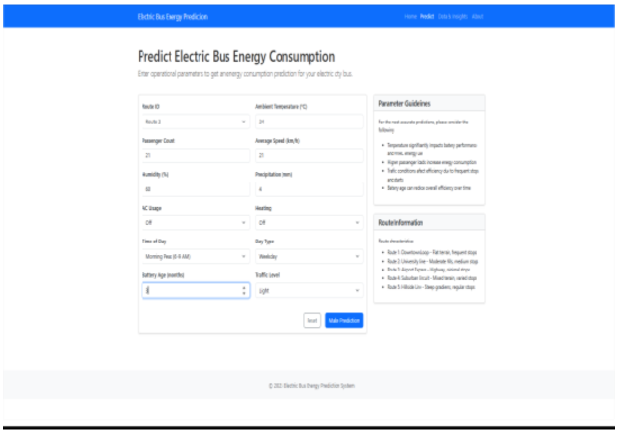

Figure 5: Web interface of electric bus energy consumption

This interface makes the ML model easily accessible and offers users useful functionality for simulation or transport management.

VI. CONCLUSION

This research presents a practical and deployable solution for predicting the energy efficiency of electric city buses using machine learning. By developing a synthetic dataset tailored to electric bus operations and incorporating key factors such as distance, speed, HVAC usage, passenger count, and elevation changes, the system was able to achieve strong predictive accuracy through effective preprocessing and the use of a Random Forest Regressor. The trained model is integrated into a Flask-powered web application, enabling real-time energy efficiency forecasts based on user- provided environmental and route-specific inputs. This tool offers valuable support for transit planners and energy analysts aiming to enhance fleet efficiency, reduce operational expenses, and make informed, data-driven decisions. The system's modular design also opens the door for future enhancements, including integration with live data streams, advanced predictive features, and scalable deployment through cloud infrastructure.

Citation of this Article:

B. Rupadevi, & Pudi Jyothi Prakash. (2025). Machine Learning Approaches to Forecast Energy Consumption in Electric Public Transit. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 107-111. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202517

****