Prediction of Electric City Bus Energy Economies using Data and Machine Learning

Prof. D. V. Varaprasad, M.Tech, (Ph.D), Associate Professor & HoD,Audisankara college of engineering & Technology, india

Mr. D. Rohith, Assistant Professor, Department of CSE,Audisankara college of engineering &Technology ,india

Dayakar Reddy Ketha,Department of CSE,Audisankara college of engineering & Technology, india

Abstract: Electrification of transportation networks is expanding, in particular city buses show huge possibilities. Deep understanding of real-world driving data is essential for vehicle design and fleet operation. To run alternative powertrains effectively, certain technological factors have to be taken under account. Energy demand uncertainty leads to cautious design that suggests inefficiency and expensive prices. Complexity and parameter interdependence cause both business and academics to overlook analytical answers for this challenge. Precise energy demand forecast offers considerable cost reduction by improved operations. The purpose of this work is more openness of the energy economics of battery electric buses (BEB). We present fresh sets of explanatory variables that describe speed profiles, which we apply in strong machine learning techniques. We develop and evaluate five different algorithms in respect to prediction accuracy, robustness, and general applicability.

***Index terms -**—Battery Electric Bus (BEB), Machine Learning, Energy Consumption Prediction, Regression Models, Speed Profile Analysis, Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), Gaussian Process Regression (GPR), Convolutional Neural Network (CNN), Data-Driven Modeling, Energy Efficiency, Public Transport Electrification.*

1. INTRODUCTION

The global shift towards sustainable and environmentally friendly transportation has brought electric mobility to the forefront, especially in urban public transit systems. Battery Electric Buses (BEBs) are emerging as a promising alternative to conventional diesel buses due to their lower emissions, quieter operations, and potential for reduced operational costs. However, the adoption of BEBs presents unique challenges, particularly related to energy consumption prediction and vehicle design optimization.

Accurately predicting the energy economy of electric city buses is vital for fleet planning, charging infrastructure deployment, and overall cost-effectiveness. Traditional models, often reliant on data from standardized driving cycles or limited routes, fail to capture the complexity and variability of real-world driving conditions. As a result, energy demand estimation may lack precision, leading to overdesign and inefficient operations.

This work introduces a machine learning-based framework that utilizes diverse real-world driving data and extracted speed profile features to model and predict the energy requirements of electric city buses. By applying advanced regression algorithms, such as Random Forest (RF), Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), and Convolutional Neural Networks (CNN), we aim to improve the robustness and accuracy of energy consumption forecasts. This data-driven approach not only enhances prediction performance but also contributes to cost-efficient and intelligent deployment of BEBs in urban environments.

2. LITERATURE SURVEY

i) SpoofCatch: A Client-Side Protection Tool Against Phishing Attacks

https://ieeexplore.ieee.org/document/9391742

Most antiphishing methods in the literature either escape specific attack patterns or are reliant on complicated sets of characteristics to identify phishing assaults or suffer from both, hence shielding against online spoofing attacks. In this paper, we suggest that depending just on the general visual look of the web page the user views will help to prevent phishing attempts. We propose a client-side protection technique based on visual resemblance of web pages and use our mechanism as a browser plugin to realise our claim. We call this SpoofCatch. Four similarity algorithms have been developed and used in the extension for comparison between phished and real web pages. Large scale and thorough studies have been carried out to assess the solution and show SpoofCatch can detect all phishing attempts with reasonable overhead.

ii) A framework for detection and measurement of phishing attacks:

https://dl.acm.org/doi/10.1145/1314389.1314391

Combining social engineering methods with advanced attack paths, phishing is a type of identity theft that gathers financial data from gullible consumers. Usually a phisher lures her victim into visiting a URL referring to a rogue page. This work aims to investigate the structure of URLs used in several phishing campaigns. We discover that, in most cases, one can determine if a URL belongs to a phishing assault without knowing the related page data. We discuss many characteristics that could help one differentiate a phishing URL from a benign one. These characteristics help to construct a highly accurate logistic regression filter with efficiency. We estimate the frequency of phishing on the Internet nowadays using this filter and do extensive measurements on several million URLs.

iii) Effective Protection Against Phishing and Web Spoofing:

https://link.springer.com/chapter/10.1007/11552055_4

On the Internet, phishing and web spoofing have multiplied and grown to be a main annoyance. Mostly as the attacks target non-cryptographic components, including the user or the user-browser interface, they are challenging to defend against. This implies that other protection measures must be included to complement cryptographic security systems such the SSL/TLS protocol since they do not offer a whole solution to address the assaults. We present, analyse, and assess in this work the efficiency of such systems against (large-scale) phishing and Web spoofing assaults.

iv) Defending Against Injection Attacks Through Context-Sensitive String Evaluation:

https://link.springer.com/chapter/10.1007/11663812_7

Applications-level security suffers a great risk from injection vulnerabilities. Among the most often occurring forms are SQL injection, cross-site scripting, and shell injection vulnerabilities. Current defences against injection attacks—that is, assaults using these weaknesses—rely mostly on the application developers and so are prone to mistakes.

In this work we provide CSSE, an injection attack detection and prevention technique. CSSE addresses the core cause—that is, the ad hoc serialisation of user-provided input—why such assaults might succeed. It offers a platform- enforced separation of channels utilising a mix of assignment of metadata to user-provided input, metadata-preserving string operations and context-sensitive string evaluation.

CSSE calls not for application source code changes nor application developer engagement. It essentially distributes the load of defences against injection attacks from the many application developers to the small team of security-savvy platform engineers as only modifications to the fundamental platform are required. Against most forms of injection assaults, our approach is successful; moreover, we demonstrate that it is also less error-prone than other methods thus far suggested.

We have created a prototype PHP CSSE solution based on a platform especially prone to these vulnerabilities. We validated our approach using our prototype with the well-known bulletin-board program PHPBB. CSSE only suffered moderate run-time overhead and found and stopped all the SQL injection attacks we could replicate.

v) Reliable protection against session fixation attacks:

https://dl.acm.org/doi/abs/10.1145/1982185.1982511

The phrase "Session Fixation vulnerability" covers problems in Web applications that, under specific conditions enable the adversary to launch a Session Hijacking attack by modulating the session identity value of the victim. A good assault lets the assailant completely pass for the vulnerable Web application, therefore impersonating the victim. We examine the vulnerability pattern and find its underlying cause in the division of responsibilities between the framework support, which manages session tracking, and the application logic, which controls the authentication procedures. This outcome leads us to propose and explore three separate server-side solutions for Session Fixation vulnerability reduction. Every one of our remedies is designed to fit a certain real-world situation the operator of a vulnerable Web application could come against.

3. METHODOLOGY

i) Proposed Work:

In this project, we propose a machine learning-based approach to accurately predict the energy consumption of Battery Electric Buses (BEBs) using real-world driving data. We utilize the operator’s vehicle database along with a physics-based model of electric buses to generate a diverse dataset containing speed profiles, payload mass, and route characteristics. By extracting time-domain and frequency-domain features from the speed signal, we aim to capture the dynamic behavior of buses under different conditions. This detailed feature extraction allows for better understanding of driving patterns and their effect on energy demand.

We then train and evaluate multiple regression-based machine learning models, including Multivariate Linear Regression (MLR), Random Forest (RF), Artificial Neural Networks (ANN), Gaussian Process Regression (GPR), and an extended Convolutional Neural Network (CNN) model. These algorithms are compared based on their prediction accuracy, robustness, and generalizability. Our proposed system offers significant improvements in prediction accuracy and supports better planning and management of electric bus operations, contributing to reduced operational costs and more efficient energy usage.

ii) System Architecture:

The system architecture consists of multiple stages starting with the collection of real-time driving data from electric buses, including speed profiles, route details, and vehicle load information. This raw data is preprocessed to remove noise and irrelevant features, followed by the extraction of statistical, time-based, and frequency-based features from the speed signals. These extracted features, along with payload and route attributes, form the input dataset for training various machine learning models. The models—MLR, RF, ANN, GPR, and CNN—are trained and validated to predict energy consumption. The architecture is designed to continuously improve prediction accuracy through iterative training and model selection, enabling real-time energy forecasting for fleet management and planning.

iii) Modules:

a. Data Collection and Preprocessing

- Collect real-time driving data from electric city buses (speed, load, routes).
- Clean the data by handling missing values and removing noise.
- Normalize and format the dataset for model input.

b. Feature Extraction

- Extract statistical features (mean, standard deviation, etc.) from speed profiles.
- Perform time-domain and frequency-domain analysis on speed signals.
- Integrate payload and route-related parameters.

c. Model Training and Evaluation

- Apply machine learning algorithms (MLR, RF, ANN, GPR, CNN).
- Train models using the extracted features to predict energy consumption.
- Evaluate models based on accuracy, robustness, and generalization.

d. Prediction and Optimization

- Predict energy consumption for various real-world conditions.
- Select the best-performing model for deployment.
- Suggest optimized vehicle operation to reduce energy costs and increase efficiency.

iv) Algorithms:

a. Multivariate Linear Regression (MLR):

Multivariate Linear Regression is a basic yet effective regression technique used to model the linear relationship between two or more independent variables and a single dependent variable. In this project, MLR is used to understand how multiple inputs such as speed, acceleration, payload mass, and route gradient collectively impact the energy consumption of electric buses. While it is simple and interpretable, MLR may not capture complex non-linear dependencies, but it provides a strong baseline model for performance comparison.

b. Random Forest (RF):

Random Forest is an ensemble learning technique that builds multiple decision trees and aggregates their outputs for a more stable and accurate prediction. It is especially effective in handling large datasets with noisy and complex features, as it reduces the risk of overfitting. In this system, RF helps capture non-linear patterns between driving behaviors and energy demand. Its ability to rank feature importance also assists in identifying the most influential driving parameters affecting energy consumption.

c. Artificial Neural Network (ANN):

Artificial Neural Networks are computational models inspired by the human brain. They consist of layers of interconnected nodes (neurons) that can learn and represent complex non-linear functions. For this project, ANN is trained using features such as time-domain and frequency-domain attributes of speed signals, bus load, and route profiles. ANN is powerful in generalizing unseen data, making it suitable for predicting energy usage under diverse driving conditions. With proper tuning and training, it offers high accuracy but requires significant computational resources.

d. Gaussian Process Regression (GPR):

Gaussian Process Regression is a non-parametric, Bayesian approach to regression that provides both predicted values and uncertainty estimates. GPR models energy consumption by treating it as a distribution over possible functions that fit the data. It is particularly useful for small and medium-sized datasets and offers high interpretability by quantifying prediction confidence. In this context, GPR helps operators understand not just how much energy will be used, but also the certainty around those predictions—useful for risk-aware fleet management.

e. Convolutional Neural Network (CNN – Extended):

Though typically used for image data, Convolutional Neural Networks can be extended to time-series data like speed signals. In this project, CNNs are adapted to learn hierarchical representations from raw speed profiles, which are then used to predict energy consumption. The convolutional layers automatically extract spatial-temporal features that might be missed by traditional models. This extension enhances prediction accuracy and is especially effective when dealing with large volumes of high-resolution sensor data from electric buses.

4. EXPERIMENTAL RESULTS

The proposed system was tested using real-world driving data collected from multiple electric city buses across varied routes and operational conditions.

After preprocessing and feature extraction, all five machine learning models—MLR, RF, ANN, GPR, and extended CNN—were trained and evaluated based on prediction accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Among these, the CNN and ANN models demonstrated the highest accuracy, closely followed by Random Forest. GPR performed well with uncertainty estimation, while MLR provided interpretable but less accurate predictions. The experimental results confirm that data-driven approaches, especially deep learning methods, significantly improve the reliability and precision of energy consumption forecasting in electric buses.

Accuracy: How well a test can differentiate between healthy and sick individuals is a good indicator of its reliability. Compare the number of true positives and negatives to get the reliability of the test. Following mathematical:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

mAP: Mean Average Precision (MAP) is a ranking quality metric. It considers the number of relevant recommendations and their position in the list. MAP at K is calculated as an arithmetic mean of the Average Precision (AP) at K across all users or queries.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Fig: **Data set Loaded**

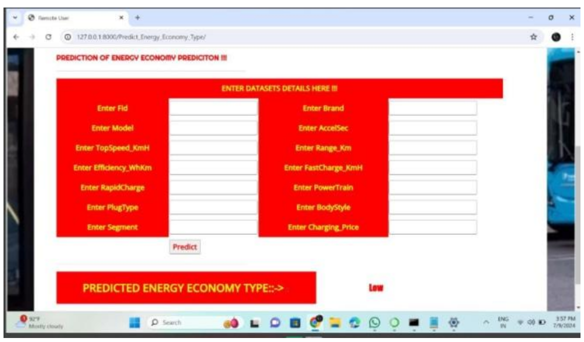

Fig: predicted results

5. CONCLUSION

This work presents a data-driven framework for accurately predicting the energy consumption of Battery Electric Buses (BEBs) using real-world and simulated driving data. By extracting time and frequency domain features from speed profiles and applying advanced machine learning models, the system achieves high prediction accuracy across varied driving conditions. The introduction of novel features like spectral entropy significantly enhances the model's performance. The fragment-based (microtrip) approach increases robustness against non-stationarity in data. This methodology offers practical value for fleet operators planning to transition to electric buses, helping optimize vehicle selection, route planning, and charging strategies. Future work will explore the model's adaptability to other vehicle types, routes, and environmental conditions, enabling broader application in the transportation sector.

6. FUTURE SCOPE

In future research, the proposed approach can be extended to other vehicle categories such as passenger cars, heavy-duty trucks, and rail systems. The model can be enhanced by integrating additional factors like weather conditions, road types, and traffic patterns to improve prediction accuracy under dynamic environments. Seasonal and regional variations can also be explored to tailor the energy prediction model for specific operational areas. Furthermore, this framework can be adapted to predict other variables such as peak power demand and battery electric current usage, supporting more advanced planning for electric vehicle infrastructure and smart fleet management systems.

REFERENCES

[1] W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, ''SpoofCatch: A client-side protection tool against phishing attacks,'' IT Prof., vol. 23, no. 2, pp. 65–

74, Mar. 2021.

[2] B. Schneier, ''Two-factor authentication: Too little, too late,'' Commun. ACM, vol. 48, no. 4, p. 136,

Apr. 2005.

[3] S. Garera, N. Provos, M. Chew, and A.
D. Rubin, ''A framework for detection and measurement of phishing attacks,'' in Proc. ACM Workshop Recurring malcode, Nov. 2007, pp. 1–8.

[4] R. Oppliger and S. Gajek, ''Effective protection against phishing and web spoofing,'' in Proc. IFIP Int. Conf. Commun. Multimedia Secur. Cham, Switzerland: Springer, 2005, pp. 32–41.

[5] T. Pietraszek and C. V. Berghe, ''Defending against injection attacks through context-sensitive string evaluation,'' in Proc. Int. Workshop Recent Adv. Intrusion Detection. Cham, Switzerland: Springer, 2005, pp. 124–145.

[6] M. Johns, B. Braun, M. Schrank, and J. Posegga, ''Reliable protection against session fixation attacks,'' in Proc. ACM Symp. Appl. Comput., 2011, pp. 1531–
1537.

[7] M. Bugliesi, S. Calzavara, R. Focardi, andW. Khan, ''Automatic and robust client-side protection for cookie-based sessions,'' in Proc. Int. Symp. Eng. Secure Softw. Syst. Cham, Switzerland: Springer, 2014, pp. 161–178.

[8] A. Herzberg and A. Gbara, ''Protecting (even naıve) web users from spoofing and phishing attacks,'' Cryptol. ePrint Arch., Dept. Comput. Sci. Eng., Univ. Connecticut, Storrs, CT, USA, Tech. Rep. 2004/155, 2004.

[9] N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, ''Client-side defense against web-based identity theft,'' in Proc. NDSS, 2004, 1–16.

[10] B. Hämmerli and R. Sommer, Detection of Intrusions and Malware, and Vulnerability Assessment: 4th International Conference, DIMVA 2007 Lucerne, Switzerland, July 12-13, 2007 Proceedings, vol. 4579. Cham, Switzerland: Springer, 2007.

[11] C. Yue and H. Wang, ''BogusBiter: A transparent protection against phishing attacks,'' ACM Trans. Internet Technol., vol. 10, no. 2, pp. 1–31, May 2010.

[12] W. Chu, B. B. Zhu, F. Xue, X. Guan,
and Z. Cai, ''Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs,'' in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2013, pp. 1990–1994.

[13] Y. Zhang, J. I. Hong, and L. F. Cranor, ''Cantina: A content-based approach to detecting phishing web sites,'' in Proc. 16th Int. Conf. World Wide Web, May 2007, pp. 639–648.

[14] D. Miyamoto, H. Hazeyama, and Y. Kadobayashi, ''An evaluation of machine learning-based methods for detection of phishing sites,'' in Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer, 2008, pp. 539–546

[15] Basha, A. M., Rajaiah, M., Penchalaiah, P., Kamal, C. R., & Rao, B. N. (2020). Machine learning-structural equation modeling algorithm: The moderating role of loyalty on customer retention towards online shopping. *Int. J*, *8*, 1578-1585

[16] Basha, A. M., et al. "Machine learning-structural equation modeling algorithm: The moderating role of loyalty on customer retention towards online shopping." *Int. J* 8 (2020): 1578-1585.

[17] Basha, A. M., Rajaiah, M., Vijayakumar, O., Haranath, Y., & Srinivasulu, T. (2020). Green and lean industrial engineering practices in selected manufacturing units in Andhra Pradesh: statistical analysis. *International Journal*, *8*(5).

[18] Basha, AM Mahaboob, et al. "Green and lean industrial engineering practices in selected manufacturing units in Andhra Pradesh: statistical analysis." *International Journal* 8.5 (2020).

[19] Basha, A. M., Reddy, P. C., & Murthy, G. R. K. (2019). Moderating role of security and reliability on high customer satisfaction: Relationship among ease of use-content-service quality with respect to customer satisfaction in digital banking transactions

[20] Basha, A. M. M. (2020). HR analytics using R-machine learning algorithm: Multiple linear regression analysis. *International Journal of Innovative Technology and Exploring Engineering, 9*(5), 1179-1183.

[21] Basha, A. M., Ankaiah, B., Srivani, J., & Dadakalander, U. (2020). Real estate analytics with respect to Andhra Pradesh: Machine learning algorithm using R-Programming. *International Journal of Scientific & Technology Research*, *9*(4), 2140-2144.

[22] Rajaiah, M., Rao, R. N., Nagendra, K. V., Basha, A. M., Rao, B. N., & Vijayakumar, O. (2020). Mathematical modeling on marketing analytics. *Journal of Critical Reviews*, *7*(19), 4189-4198.

[23] Venkatrayulu, C., Gurumoorthi, S. K., Basu, S., Jayalakshmi, M., & Basha, A. M. (2023). The Mediating Role of Digital Marketing Practices in Relationship between the Standard Marketing Strategies and the Market Growth of Fisheries and Aqua Products in India. *Journal of Survey in Fisheries Sciences*, *10*(2S), 2292-2300.

[24] Rekha, Y. C., Basha, A. M., & Prasad, P. S. (2020). Mediation effect of organizational performance in the relationship between human resource development practices and organizational learning.

[25] Basha, M. (2014). Statistical Analysis on the Sustainable development of India with Reference to Agricultural Sector. *International Multidisciplinary Research Journal*, 1-4.