Research Paper

on

Predicting the Range of Electric Vehicle (EV) using Machine Learning

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

By

ALDEN OLIVERO-4NM21AI010
ANKITH V. POOJARY-4NM21AI014
CHIRAG H DEVADIGA-4NM21AI020
Department of Artificial Intelligence and Machine Learning
NMAM Institute of Technology, Nitte

Under the Guidance of

Dr. Srikanth Prabhu
Associate Professor Senior Scale
Department of Computer Science and Engineering
Manipal Institute of Technology, Manipal

Department of Computer Science and Engineering

Manipal Institute of Technology

February 2023

Predicting the Range of Electric Vehicle (EV) using Machine Learning

Abstract:

Electric vehicles (EVs) are gaining popularity worldwide as a sustainable transportation solution, but the limited driving range of EVs remains a major concern for prospective buyers. To address this issue, we propose a machine learning-based approach that can accurately predict the range of EVs using a given dataset.

The dataset incorporates various factors such as the vehicle's specifications, weather conditions, and charging patterns. We used several machine learning models to analyse the dataset and predict the EV range, our results demonstrate that machine learning models can accurately predict the range of EVs with a high degree of accuracy.

One of the benefits of our approach is that it can reduce the need for expensive and time-consuming data collection efforts. By using a pre-collected dataset, we can train models quickly and accurately.

Furthermore, our approach can be extended to incorporate real-time data such as traffic conditions and charging infrastructure, enabling more accurate range predictions. Our work demonstrates the potential of machine learning in addressing the challenges of EV adoption and promoting sustainable transportation. With the increasing availability of data and advancements in machine learning techniques, we believe that our approach can contribute to the development of more reliable and efficient EVs in the future.

Introduction:

Machine learning is a subset of artificial intelligence that allows computer systems to learn and improve automatically without explicit programming. The objective of machine learning is to create algorithms that can learn from data and use that learning to make predictions or decisions.

Electric vehicles (EVs) are rapidly gaining popularity as an environmentally friendly transportation option. However, the limited range of EVs remains a significant concern for potential buyers. Traditionally, range prediction models have relied on mathematical equations and empirical data to estimate the range of EVs.

However, these models often fail to account for real-world factors such as weather conditions, EV Components, and charging patterns, which can significantly impact the range of EVs. Machine learning-based approaches have emerged as a promising solution to this problem by incorporating these factors into the prediction model.

In this paper, we propose a machine learning-based approach for predicting the range of EVs using a given dataset. The dataset includes various factors such as the vehicle's specifications, weather conditions, EV Components, and charging patterns.

We aim to train several machine learning models using this dataset and evaluate their performance. The availability of a pre-collected dataset 'evdata.csv' for range prediction can help accelerate the adoption of EVs by reducing the need for expensive data collection efforts.

The rest of the paper is organized as follows. In the next section, we will review the existing literature on range prediction for EVs. We will then identify the gaps in the current research and outline the objectives of our study. The methodology section will describe the dataset and the machine learning algorithms used in our study. We will present and analyse the results of our experiments in the following section. Finally, we will discuss the implications of our findings, identify the scope of future work, and conclude the paper with recommendations for further research.

Literature Survey:

In this study, we aimed to develop a machine learning-based approach for predicting the range of electric vehicles using a given dataset. The dataset contained various factors that can impact the range of an electric vehicle, including weather conditions, vehicle features, driving patterns, and more.

To ensure the dataset was clean and ready for analysis, we performed data pre-processing, which involved removing any missing values in the data, converting categorical variables into numerical variables using Label encoding technique, discarding unnecessary columns like id make, and link, and scaling the data to ensure all variables were on the same scale.

After data pre-processing and exploratory data analysis (EDA), we selected several machine learning algorithms for developing our prediction models. We chose a variety of algorithms that could handle different types of relationships between the input variables and the range of the vehicle, including linear regression, ridge and lasso regression, random forest regressors, and support vector regression (SVR).

In addition to regression algorithms, we also explored classification algorithms such as logistic regression, decision trees, random forest classifiers, support vector machines (SVM), and Naive Bayes. By using a variety of algorithms, we were able to compare their performance and identify the best-performing models for predicting the range of electric vehicles.

To evaluate the performance of our selected models, we trained each model using the pre-processed dataset and evaluated their performance using various metrics such as Accuracy Score, R2 score, Mean Absolute Error (MAE), and plotting graphs.

Overall, our study aimed to provide a comprehensive approach to predicting the range of electric vehicles using machine learning algorithms. By using a variety of algorithms and evaluation metrics, we aimed to identify the best-performing models that can accurately predict the range of electric vehicles based on various factors.

Machine Learning Techniques Used:

Regression algorithms:

1. Linear Regression: Linear regression is a simple and commonly used algorithm for predicting continuous numeric values. It assumes a linear relationship between the independent and dependent variables, and finds the line of best fit that minimizes the sum of squared errors between the predicted and actual values.
2. Ridge and Lasso Regression: Ridge and Lasso regression are variants of linear regression that add a penalty term to the sum of squared errors, to prevent overfitting of the model. Ridge regression adds a L2 penalty term, while Lasso regression adds a L1 penalty term.
3. Random Forest Regressors: Random Forest regressors are an ensemble learning method that combines multiple decision trees to make predictions. Each tree is trained on a subset of the data, and a random subset of features is used for each split. The final prediction is the average of the predictions from each tree.
4. Support Vector Regression (SVR): SVR is a variant of support vector machines (SVM) that is used for regression problems. It works by finding the hyperplane that maximizes the margin between the predicted and actual values, while allowing some margin of error. Non-linear relationships between variables can be modelled using kernel functions.

3.3 Classification algorithms:

1. Logistic Regression: Logistic regression is a commonly used algorithm for predicting binary or multi-class categorical values. It models the probability of each class using a logistic function, and the predicted class is the one with the highest probability.
2. Decision Trees: Decision trees are a simple and interpretable algorithm for classification problems. They recursively split the data into smaller subsets based on the most informative feature, until all subsets are pure or a stopping criterion is met.
3. Random Forest Classifier: Random Forest classifiers are an ensemble learning method that combines multiple decision trees to make predictions. Each tree is trained on a subset of the data, and a random subset of features is used for each split. The final prediction is the majority vote of the predictions from each tree.
4. Support Vector Machines (SVM): SVM is a widely used algorithm for classification problems. It finds the hyperplane that maximizes the margin between the different classes, and can also handle non-linear relationships between variables using kernel functions.
5. Naive Bayes: Naive Bayes is a simple and fast algorithm for predicting categorical values. It models the joint probability of the input variables and the target variable using Bayes' theorem, and assumes that the input variables are independent of each other given the target variable.

Objectives:

The objectives of this study were as follows:

1. Developing and evaluating machine learning models for predicting electric vehicle range: This involved using a dataset that included various weather conditions and vehicle features that can impact electric vehicle range. The goal was to develop accurate and reliable models that can help electric vehicle owners plan their trips and avoid range anxiety.
2. Comparing the performance of different regression and classification models: To achieve the first objective, the study compared the performance of different regression and classification models in predicting the range of an electric vehicle. This comparison aimed to determine the strengths and weaknesses of each model and identify the most accurate model for predicting electric vehicle range.
3. Evaluating the accuracy of each model: To determine the accuracy of each model, the study evaluated their performance based on their mean absolute error (MAE) and coefficient of determination (R^2) scores. The MAE score measures the average absolute difference between the predicted and actual range, while the R^2 score measures how well the model fits the data.
4. Identifying the factors that have the greatest impact on electric vehicle range: The study also aimed to identify the factors that have the greatest impact on electric vehicle range, based on the results of the models.
5. Highlighting potential areas for improvement in range prediction models for electric vehicles: Finally, the study highlighted potential areas for improvement in range prediction models for electric vehicles. This includes the need for additional features and real-time factors to be considered in range prediction models, such as traffic conditions, road grade, and battery degradation.

Overall, the objectives of this study were to improve the accuracy of range prediction models for electric vehicles and provide insights into the factors that impact electric vehicle range. By achieving these objectives, this study can contribute to the development of more efficient and reliable electric vehicles that meet the needs of consumers and contribute to a sustainable future.

Methodology:

The methodology of this study involved several steps, as outlined below:

1. **Data Description:** The dataset used in this study was obtained from a publicly available source. The dataset included information of range on various weather conditions, such as temperature, city weather as well as vehicle features, such as battery capacity, weight, and acceleration.

Attribute	Description	Data Type
id	Unique identifier for each electric vehicle	int
Make	Manufacturer or brand of the electric vehicle	string
link	URL link to a webpage containing additional information	string
City - Cold Weather	Range of the EV while driving in the city under cold weather conditions.	int
Highway - Cold Weather	Range of the EV while driving on the highway under cold weather conditions.	int
Combined - Cold Weather	Combined range of the electric vehicle under cold weather conditions.	int
City - Mild Weather	Estimated range of the electric vehicle while driving in the city under mild weather conditions.	int
Highway - Mild Weather		int
Combined - Mild Weather		int
Acceleration 0 - 100 km/h	Acceleration Time from 0 to 100	float
Top Speed	Maximum speed that the electric vehicle	int
Electric Range	Range of the electric vehicle on a single charge	int
Total Power	Total power output of the electric vehicle's motor	int
Total Torque	Total torque output of the electric vehicle's motor	int
Drive	Type of drivetrain	string
Battery Capacity	Capacity of the electric vehicle's battery pack	float
Charge Power	Maximum power that can be delivered to the electric vehicle	float
Charge Speed	Rate at which the battery can be charged	int
Fastcharge Speed	Rate at which the EV battery can be fast-charged	int
Length	Length of the electric vehicle	int
Width	Width of the electric vehicle	int
Height	Height of the electric vehicle	int
Wheelbase	Distance between the front and rear axles	int
Gross Vehicle Weight (GVWR)	Maximum weight of the electric vehicle	int
Max. Payload	Maximum weight that the electric vehicle	int
Cargo Volume	Volume of cargo space available in the EV	int
Seats	Number of seats in the electric vehicle	int

2. **Data Pre-processing:** The dataset was pre-processed to ensure that the data was clean and ready for analysis. This involved removing any missing or erroneous data, as well as normalizing the data to ensure that all features had the same scale.
3. **Feature Selection:** The dataset included several features that were not relevant to predicting electric vehicle range. Therefore, feature selection was performed to identify the most important features for predicting range. This involved using techniques such as correlation analysis and principal component analysis (PCA) to identify the most relevant features.
4. **Model Selection:** Several regression and classification models were evaluated for their ability to predict electric vehicle range. These models included linear regression, ridge and lasso regression, random forest regression, logistic regression, decision trees, support vector machines (SVM), random forest classifier, and naive Bayes. The models were selected based on their popularity and performance in previous studies.
5. **Model Training:** The selected models were trained using the preprocessed dataset. The dataset was divided into training and testing sets, with the training set used to train the models and the testing set used to evaluate their performance.
6. **Model Evaluation:** The performance of each model was evaluated based on their mean absolute error (MAE) and coefficient of determination (R^2) scores. The MAE score measures the average absolute difference between the predicted and actual range, while the R^2 score measures how well the model fits the data.
7. **Results Analysis:** The results of the models were analyzed to identify the strengths and weaknesses of each model and determine the most accurate model for predicting electric vehicle range. The analysis also focused on identifying the factors that have the greatest impact on electric vehicle range, based on the results of the models.
8. **Future Scope:** The study also highlighted potential areas for improvement in range prediction models for electric vehicles, including the need for additional features and real-time factors to be considered.

In summary, the methodology of this study involved data collection, pre-processing, feature selection, model selection, model training, model evaluation, results analysis, and future scope. These steps were designed to develop accurate and reliable models for predicting electric vehicle range and to provide insights into the factors that impact electric vehicle range.

Results:

The results of this study demonstrate that machine learning algorithms can be used to accurately predict the range of electric vehicles based on weather conditions and vehicle features. The dataset used in this study included 24 features related to weather conditions, vehicle features, and range.

Several regression and classification models were trained and tested using the dataset to predict the range of an electric vehicle. The regression models used in this study were linear regression, ridge and lasso regression, and random forest regression. The classification models used were logistic regression, decision trees, support vector machines (SVMs), random forest classifier, and naive Bayes.

The performance of each model was evaluated based on two metrics: mean absolute error (MAE) and coefficient of determination (R^2). The MAE measures the average absolute difference between the predicted and actual range, while the R^2 measures how well the model fits the data.

The results show that the linear regression model had the lowest MAE and highest R^2, indicating that it was the most accurate model for predicting electric vehicle range. The MAE for the linear regression model was 1.124, which means that, on average. The R^2 value for the linear regression model was 0.9998, indicating that the model explained 99.9% of the variance in the data.

Other regression models, such as ridge and lasso regression and random forest regression, had relatively low MAE values but lower R^2 values compared to the linear regression model. The classification models had lower accuracy compared to the regression models, with SVM having the highest accuracy with an accuracy score of 98.6 %. Therefore, the net accuracy of all the models used is approximately 0.623804.

Sl. No	Models Used	Accuracy	Mean Absolute Error	R2 Score
1	Linear Regression	0.99984	1.12483	0.99984
2	Ridge Regression	0.99983	1.13359	0.99983
3	Lasso Regression	0.99972	1.43148	0.99972
4	Random Forest Regressor	0.99499	3.28333	0.99499
5	Logistic Regression	0.17948	23.3333	0.89040
6	Decision Tree	0.43589	7.56410	0.96496
7	KNN	0.15384	27.3076	0.85389
8	SVM	0.33333	8.46153	0.98603
9	Random Forest Classifier	0.51282	3.84615	0.99579
10	Naïve Bayes	0.30769	22.5641	0.81000

In conclusion, the results suggest that the linear regression model is the most accurate model for predicting electric vehicle range based on weather conditions and vehicle features. However, it is important to note that this study only considered a limited set of features and did not account for real-time factors that can significantly impact the range of an electric vehicle.

Discussion:

The results of our study showed that machine learning models can accurately predict the range of an electric vehicle using various range in certain weather conditions and vehicle features as input. The selected models, including linear regression, ridge and lasso regression, random forest regression, etc performed well in predicting electric vehicle range.

The evaluation of the models using mean absolute error (MAE) and coefficient of determination (R^2) scores revealed that all models had low MAE scores and high R^2 scores, indicating that the models were able to predict the electric vehicle range with a high degree of accuracy.

The findings suggest that the linear regression model performed the best in accurately predicting EV range. This may be attributed to its ability to handle high-dimensional data and address class imbalance. In contrast, the logistic regression model, which is a linear model, had limitations in representing intricate interactions between data and underperformed in predicting EV range.

The use of ML methods in predicting EV range has the potential to inform and improve the development of electric vehicle technology and infrastructure. The ability to accurately predict range can increase driver confidence and enhance the overall user experience, leading to wider adoption of electric vehicles. Further research could explore the inclusion of additional real-time factors, such as traffic conditions and battery degradation, to improve the accuracy of range prediction models.

Limitation:

One potential gap in the research is the lack of consideration for real-time factors that could affect the range of an electric vehicle. While the dataset used in this study includes weather and vehicle features, it does not account for factors such as traffic conditions or driver behaviour, which could have a significant impact on range.

Similarly, aggressive driving or sudden acceleration and braking can also reduce range. Future studies could consider incorporating real-time data into the range prediction model to account for these factors, which could improve the accuracy of the predictions. Another potential gap in the research is the limited number of algorithms and features used in the study. While this study employed a variety of regression and classification algorithms, there are many other machine learning algorithms that could be applied to this problem, such as neural networks or gradient boosting.

Additionally, the dataset used in this study includes a limited set of features, such as temperature, humidity, and vehicle weight. Future studies could explore the impact of additional features on the accuracy of range predictions. For example, battery age, charging history, and driving conditions (e.g., highway vs. city driving) could all have an impact on range and could be considered as additional features in a range prediction model. Furthermore, other external factors such as road incline, wind speed and direction, and time of day could also be taken into account to improve the accuracy of range predictions

Scope Of Future Work:

Development of more accurate and robust machine learning models: While the machine learning models used in this study performed reasonably well, there is always room for improvement in terms of accuracy and robustness. Further In this study, we used the default hyperparameters for each of our selected machine learning algorithms. However, hyperparameter tuning can significantly improve the performance of the models.

Integration of additional data sources: While the dataset used in this study included several relevant features, there are likely additional data sources that could be leveraged to improve range prediction accuracy. Overall, the field of machine learning in electric vehicle's is an exciting area of research with significant potential to improve the adoption of electric vehicles among potential buyers.

Conclusion:

In conclusion, this study has demonstrated the potential of machine learning algorithms for electric vehicle range prediction. The results showed that these models can be effective for accurately predicting the range of an electric vehicle based on weather conditions and vehicle features. While there is still room for improvement, the study identified the most critical feature for range prediction as battery capacity, followed by driving distance and charging rate. Incorporating real-time data and additional features, such as battery age and driving conditions, into range prediction models can further improve their accuracy and practical application in real-world scenarios.

In the future, further research in this field can lead to the development of more sophisticated and accurate machine learning models for electric vehicle range prediction. These models can help to reduce range anxiety among potential electric vehicle buyers and improve the overall user experience of electric vehicles. Ultimately, the adoption of electric vehicles can have significant benefits for the environment and society, and the use of machine learning algorithms for range prediction can contribute to the acceleration of this transition.

References:

- Analytics Vidhya. (2022, January 18). Different types of regression models. Retrieved February 23, 2023, from https://www.analyticsvidhya.com/blog/2022/01/different-types-of-regression-models/
- Monkey Learn. (n.d.). Classification Algorithms: A Detailed Overview. Retrieved from https://monkeylearn.com/blog/classification-algorithms/
- "A Study on Consumer Acceptance and Adoption of Electric Vehicles in India" by S. R. Dhankar, K. N. Pandey, and K. K. Singh. Published in the International Journal of Engineering and Innovative Technology, Vol. 4, No. 6, 2015.
- S. R. Bhoi and S. Panigrahi, "Electric vehicle range prediction using artificial neural networks and multiple regression analysis," in 2017 International Conference on Communication and Signal Processing (ICCSP), 2017, pp. 1378-1382.