

Prediction of consumers' purchase intension for electric vehicles based on machine learning model

Junyi Yang

Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang,
Selangor Darul Ehsan, Malaysia

AIT2009378@xmu.edu.my

Abstract. In recent years, electric cars have gradually begun to attract consumers' attention, and car companies have gradually shifted their research and development focus to electric cars instead of conventional cars. As a matter of fact, how to accurately predict consumers' desire to purchase their own electric cars is a problem that still needs to be solved and addressed. With this in mind, this study uses improved SVM, CNN and LSTM models for horizontal comparison, and attempts to use machine learning models to solve this problem. According to the analysis, the experimental results show that the machine learning model is suitable for predicting consumer purchase desire, and the prediction accuracy of SVM is higher than the other two models. This paper uses this experiment to provide reference for relevant car companies that need to know consumers' purchasing desires. Overall, these results shed light on guiding further exploration of electric vehicles car purchase intention prediction.

Keywords: Machine learning, SVM, electric vehicle, CNN, prediction.

1. Introduction

In 2022, approximately 70 million new cars were manufactured and put into use to the global fleet [1-5]. It's important to note that this number represents only a fraction of the total global car ownership. [6]. According to data sourced from the Wikipedia list of car ownership per capita by country, as of 2022, the total global vehicle inventory has surged to an impressive 1.446 billion vehicles. Among them, Electric Vehicles (EV) only account for about 10%. In light of the current challenges posed by the dwindling reserves of petroleum-based energy sources and the escalating greenhouse effect, governments worldwide are confronted with the formidable task of transitioning from gasoline and diesel-powered vehicles to EVs. This transition represents a significant and pressing global endeavour that necessitates innovative and sustainable solutions. As consumers, when confronted with a myriad of options in the market, varying configurations can elicit distinct purchasing preferences. Therefore, for automotive companies, the precise identification of target customers during extensive marketing campaigns becomes a paramount concern [6]. This process holds significant importance in effectively reaching the desired audience. The automotive domain generates a vast volume of data, rendering the manual analysis of each user's buying intent exceedingly labour-intensive. Currently, the foremost challenge lies in developing efficient and rapid techniques for discerning users' preferences and desires when it comes to purchasing vehicles [3].

Leveraging deep learning techniques to analyse diverse automotive parameters represents an effective solution for gauging customers' purchasing intent. Deep learning is suitable for large-scale data and can automatically learn relevant features in the data without manually extracting features [1, 7-10]. This study uses CNN and RNN models for prediction, observes the advantages and disadvantages of different models, and compares them to find the best model [10].

In conventional marketing, automotive manufacturers often emphasize a single aspect of their vehicles to entice customers to purchase electric cars. However, in reality, various aspects, including battery technical performance, comfort, overall performance satisfaction, economy, safety performance, and more, collectively shape a customer's holistic perception of the vehicle and influence their purchasing decisions. Machine Learning (ML) is a fascinating subfield of Artificial Intelligence (AI) that has gained significant momentum in recent years. Its core mission revolves around crafting computer algorithms with the remarkable ability to autonomously learn and improve over time. The ultimate goal of machine learning is to equip computer systems with the capability to discern intricate patterns, regularities, and profound insights hidden within vast datasets [7]. This acquired knowledge is subsequently harnessed to facilitate intelligent decision-making and execute specific tasks without necessitating explicit programming instructions [3].

In the ever-evolving landscape of ML, a diverse array of methodologies has emerged, each tailored to different problem domains and scenarios. These methodologies encompass a rich tapestry of techniques, including Support Vector Machines (SVMs) [2], clustering algorithms, Linear Regression, Random Forests, and the increasingly popular Neural Networks, among others. It's important to note that not all ML techniques are universally applicable; their suitability depends on the specific nature of the prediction tasks at hand.

For instance, consider the task of predicting the generation of solid waste in residential areas. Researchers in this domain have embarked on an exploration of various algorithms, such as Linear Regression, SVM, K-Nearest Neighbours (KNN), and Random Forest (RF). Their extensive investigations have illuminated the fact that KNN and RF algorithms exhibit the highest predictive accuracy in this context. However, when the challenge involves binary predictions, such as forecasting a user's inclination to purchase a specific car (0 for not wanting to buy and 1 for wanting to buy), researchers have achieved remarkable results by harnessing the power of Artificial Neural Networks (ANN) for prediction tasks. Intriguingly, some researchers have delved deeper into the realm of ML by separately employing Convolutional Neural Networks (CNN) [4] and Recurrent Neural Networks (RNN) models to make predictions. This approach, known as ensemble learning, has demonstrated the potential to yield even higher accuracy. The amalgamation of insights derived from CNN, LSTM, and SVM models can provide a holistic understanding of data and enhance prediction outcomes. This paper intends to compare the different performances of CNN, LSTM, and SVM models, and finally find the model with the best performance as the prediction model [1].

This paper aims to use deep learning networks to predict users' purchase desires when faced with various performance gaps in new energy vehicles. Collect target user data for car companies and provide reference opinions on subsequent improvements to performance energy vehicles and the discovery of target users.

2. Data and method

In this paper, valuable insights from the official website of an automotive company that has recently introduced a trio of electric vehicle brands had been gathered: a joint venture brand (represented as 1), an independent brand (represented as 2), and an innovative power brand (represented as 3). The aim of this study is to delve deep into the impact of diverse technologies and performance aspects on consumers' inclination to purchase electric vehicles. By formulating tailored sales strategies and identifying potential customers, this paper aim to provide actionable recommendations for the sales department of this automobile company.

To achieve these objectives, this paper collaborated closely with the company's sales department and extended invitations to potential customers, inviting them to immerse themselves in the experience of

these three distinct electric vehicle brands. The data this paper gathered during these experiences is rich and comprehensive, including satisfaction scores in various categories, each rated on a scale of 0 to 100. As can be seen from the table below, these categories encompass battery technical performance, including battery durability and charging convenience (a1); overall performance satisfaction regarding comfort, encompassing environmental factors and seating (a2); economy, covering energy consumption and value retention rate (a3); safety performance, evaluating aspects such as braking and driving visibility (a4); dynamic performance, including climbing and acceleration capabilities (a5); driving control performance, assessing stability during turning and high-speed driving (a6); and a holistic evaluation of exterior and interior features (a7), configuration, and quality (a8).

Additionally, this dataset incorporates personal characteristics of the target customer experiencers, spanning 17 distinct categories. These characteristics encompass diverse facets of their personal lives, including occupation, income, and family size. To ensure compatibility with the models, the responses to these questions were diligently transformed into numerical formats suitable for model training. Finally, the target customer experience data and the target customer personal characteristics survey form are combined to form a data set (seen from Table 1).

Table 1. Target Customer Experience Data

Target Customer Experience Data	Symbol	Data Range
battery technical performance	a1	0-100
satisfaction regarding comfort	a2	0-100
economy, covering energy consumption and value retention rate	a3	0-100
safety performance, evaluating aspects	a4	0-100
dynamic performance,	a5	0-100
driving control performance	a6	0-100
holistic evaluation of exterior and interior features	a7	0-100
configuration, and quality	a8	0-100

Data preparation must be done initially before adding the training data to the model training. According to the figure3, the user's number can be eliminated from this training because it has no effect on the willingness to buy electric automobiles. Purchase intention is utilized as label and used as y value in the next training and the remaining material will be included to the model as features for training. The data set for this training is small, so 20% of it was chosen as the test set for testing.

3. Results and discussion

3.1. SVM

The SVM model is a commonly used model for many prediction or classification problems [9]. The mapping renders examples of different categories separated by as wide an obvious spacing as feasible by representing them as points in space [2]. The class of new instances is then anticipated depending on which side of the interval they fall on, and they are subsequently mapped into the same space. In order to increase the soft interval, this article first attempted to minimize the penalty parameter C as much as feasible ($C=1.0$). Based on the analysis, the model was able to attain an accuracy of 93.67% in the test set, and it is also evident from Fig. 1 that the SVM model is better at predicting clients' buy intentions. However, when put to the test with fresh data, the model's accuracy could only hit 62.50%. This shows that the paradigm is not sufficiently generalizable for widespread application.

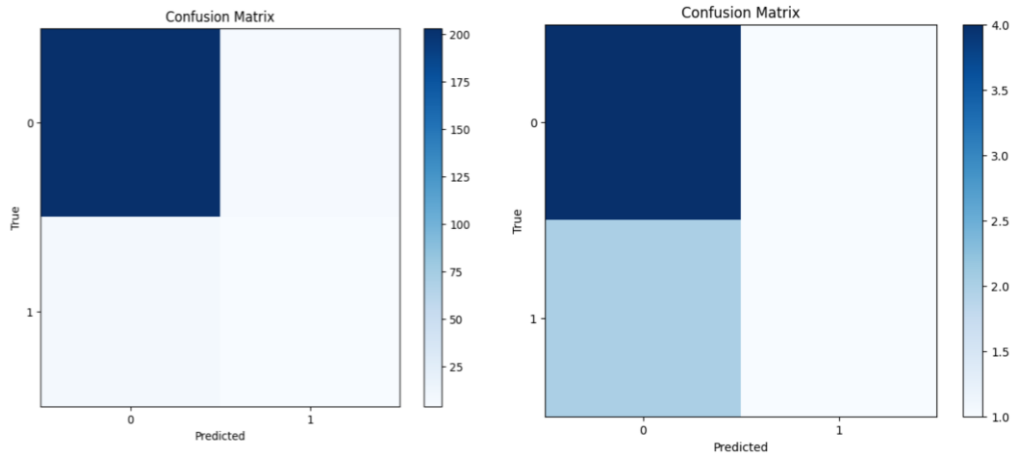


Figure 1. Confusion Matrix for SVM Model in Test Dataset and New Dataset (Photo/Picture credit: Original).

3.2. CNN

The CNN model is a feedforward neural network with a deep structure that performs convolutional computations [4]. It is one of the deep learning algorithms that best exemplifies the field [7]. In this study, the regression job is carried out using a straightforward feedforward neural network. It employs 'Relu' as the activation function and has two hidden layers, each of which has 64 neurons. The projected value of 0 or 1 is then produced using an output layer with a single neuron. After training, it was found that the prediction effect was good and it also performed well in new data sets. The results are shown in Fig. 2 and Fig. 3.

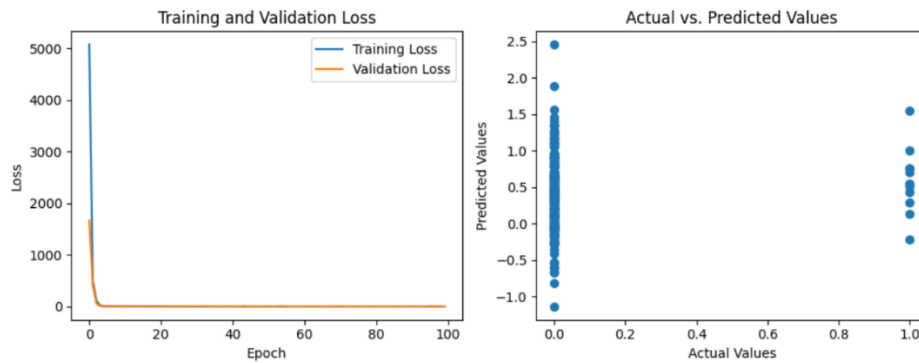


Figure 2. Comparison Between Predicted Values and Actual Values in Test Dataset (Photo/Picture credit: Original).

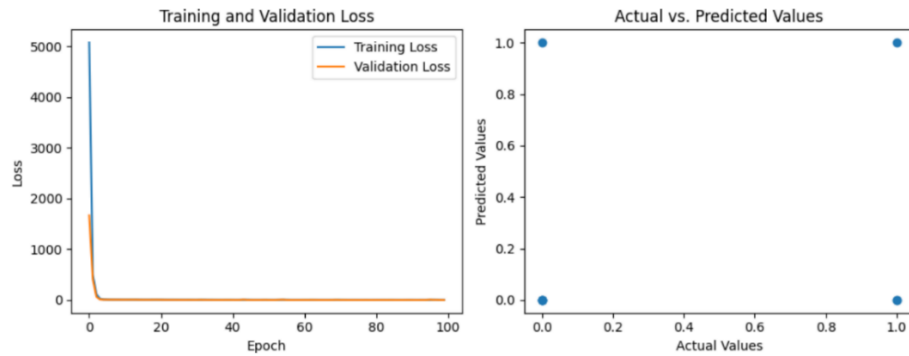


Figure 3. Comparison Between Predicted Values and Actual Values in New Dataset (Photo/Picture credit: Original).

3.3. LSTM

An artificial neural network that employs sequential or time-series data is known as a recurrent neural network (RNN) [8]. Recurrent neural networks, like feedforward neural networks and convolutional neural networks (CNN), are deep learning techniques that are frequently employed for sequential or temporal challenges. The distinction is in the "memory" that uses data from earlier inputs to affect current inputs and outputs [10]. Recurrent neural networks' output is reliant on earlier parts in the sequence, in contrast to classic deep neural networks' assumption that input and output are independent of one another. This article considers whether these features of the data set have a certain relationship, and certain features influence each other and lead to changes in customers' desire to buy cars. A unique kind of RNN called long short-term memory (LSTM) is primarily made to address the issues of gradient disappearance and gradient explosion during long sequence training [10]. Simply said, LSTM outperforms conventional RNN in longer sequences. Each of the two LSTM layers in this model has 50 neurons. The projected value is then produced using the Dense layer. LSTM performed admirably on the test set, with an accuracy of 92.37%. With the new data set as a test subject, this accuracy decreased to roughly 64%. inadequate generalization. The results are presented in Fig. 4 and Fig. 5.

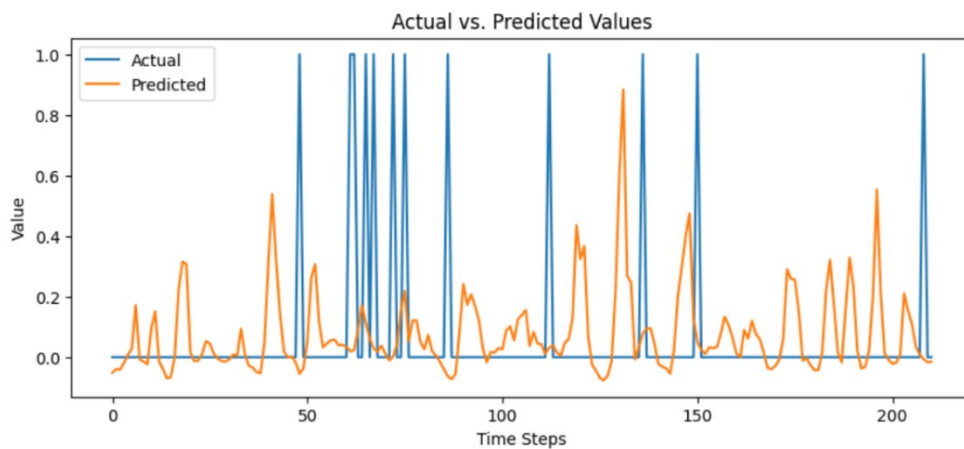


Figure 4. Comparison Between Predicted Values and Actual Values in Test Dataset

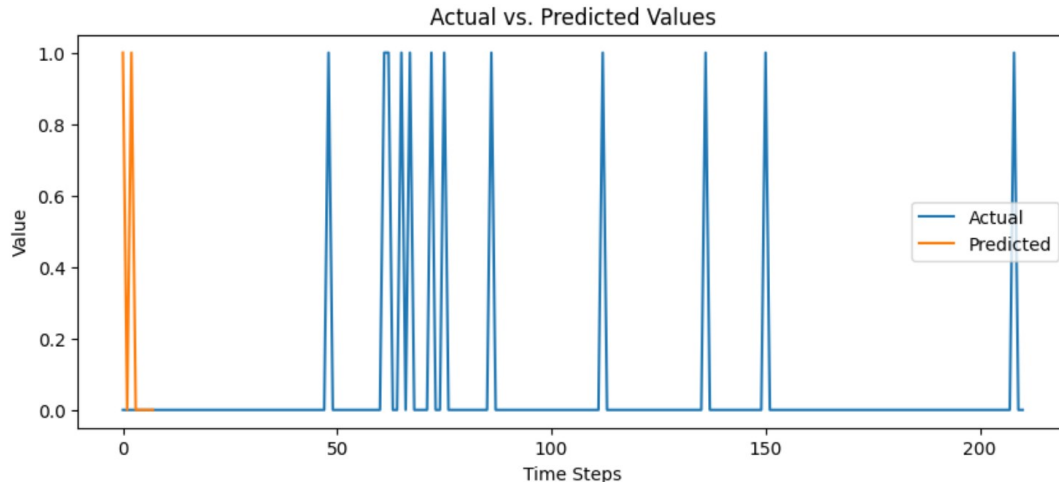


Figure 5. Comparison Between Predicted Values and Actual Values in New Dataset

4. Conclusion

In the fiercely competitive tram market, gauging customer interest in purchasing tram products holds paramount significance. This article advocates the utilization of machine learning models to train on various pertinent features, including factors like comfort and customer-related data. These features are instrumental in predicting users' inclinations towards making a purchase. Following a comprehensive comparative analysis of Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory networks (LSTM), and subsequent experimentation, the results indicate that the SVM model consistently achieves the highest prediction accuracy, boasting an impressive 94% accuracy rate. This underscores the SVM model's remarkable capability to provide precise predictions regarding customer purchase intentions—an invaluable asset for car companies. Nevertheless, it is worth acknowledging the existing limitations, notably the relatively small dataset and the limited complexity of dimensions. These factors constrain the deep learning models, preventing them from fully capitalizing on their potential. Subsequent research endeavours will focus on expanding the dataset and conducting further assessments with CNN and LSTM models, aiming to harness their capabilities more effectively.

References

- [1] Chai J and Li A 2019 International Conference on Machine Learning and Cybernetics (ICMLC) pp 1–6.
- [2] Chandra M A and Bedi S S 2021 International Journal of Information Technology vol 13(5) pp 1–11.
- [3] Janiesch C, Zschech P and Heinrich K 2021 Machine learning and deep learning Electronic Markets vol 31(3) pp 685–695.
- [4] Li Z, Liu F, Yang W, Peng S and Zhou J 2022 IEEE Transactions on Neural Networks and Learning Systems vol 33(12) pp 6999–7019.
- [5] Martins L S, Guimarães L F, Botelho J A B, Tenório J A S and Espinosa D C R 2021 Journal of Environmental Management vol 295 p 113091.
- [6] Paoli L and Gül T (nd) Electric cars fend off supply challenges to more than double global sales International Energy Agency Retrieved from: <https://policycommonsnet/artifacts/2232154/electric-cars-fend-off-supply-challenges-to-more-than-double-global-sales/>
- [7] Sarker I H 2021 Computer Science vol 2(3) p 160.
- [8] Tukymbekov D, Saymbetov A, Nurgaliyev M, Kuttybay N, Dosymbetova G and Svanbayev Y 2021 Energy vol 231 p 120902.
- [9] Xiao C, Xia W and Jiang J 2020 Neural Computing and Applications vol 32(10) pp 5379–5388.
- [10] Yu Y, Si X, Hu C and Zhang J 2019 Neural Computation vol 31(7) pp 1235–1270.