IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

A Comprehensive Study of Machine Learning Algorithms for Predicting car Purchasing based on Customers Demands

Vaishnaw G. Kale[1], Nitin Khedkar[2], Bhosale Prathviraj[3], Avinash Salve[4], Yogini Gajbhiye[5]
Students, Department of Computer Engineering[1,2,3,4,5]
Sinhgad Institute of Technology, Lonavala, Maharashtra, India

Abstract: *The automobile industry is one of the prominent industries for the national economy. Day by day car is getting popular for the private transport system. The customer needs review when he wants to buy the right vehicle, especially the car. Because it is a very costly vehicle. There are many conditions and factors matter before buying a new car like spare parts, cylinder volume, headlight and especially price. So deciding everything, it is important for the customer to make the right choice of purchase which can satisfy all the criteria. Our goal is to help the customer to make the right decision whether he will buy a car or not. Therefore we wanted to build a technique for decision making in-car buy system. That's why we propose some well known algorithms to get better accuracy for a car purchase in our paper. We applied those algorithms in our dataset which contains 50 data. Among them, Support Vector Machine(SVM) gives the best result with 86.7% accuracy of prediction. In this paper, we have also revealed the comparative results using different algorithms precision, recall and F1 score for all data samples.*

Keywords: Supervised Machine Learning, Naive Bayes, Random Forest tree, Support Vector Machine, KNN, Accuracy, Cosine Similarity

I. INTRODUCTION

In this smart era of technology, people like to make those ideas and decisions which are not only for their current benefits as well as price but also for their future advantage. For example, if a person desires to choose a job or if he is planning for where he will stay or planning for a fine v a cation, all are important in his life decisions. Because these are needed to consider about the utility that will increase in the future. People always like to make those types of decisions which maximizes the utility [1]. Since the utility is linked with the financial s ystem i n o ur d ay t o d ay life. At present, the automobile industry is one of the most important businesses in the world. Though Bangladesh is asmall country in South Asia, the demand for automobiles isincreasing with each passing day. People are using private vehicles to move one location to another location. The fourwheeled private car is good and flexible a mong t hose. A good portion of the economic development of a country depends on the transportation system. Because when the transport system is efficient, it affords economic as well as a social opportunity which gives a positive effect to markets [2]. So earlier than buying a new car, customers prefer to be assured of the money they spend to be worthy. Because to buy a new car, it is the matter of a good amount of money in the perspective of the Bangladeshi economy. Thats why it is important to get the information about cars which are good or bad based on customers experiences, who bought those before. The lifecycle
of a modern car depends on so many different parts. In this paper, we want to predict the probability of buying a car based on price, spare part, customer review, cylinder volume, resale price. Predicting the likelihood or probability of purchase for vehicles is a superb and much-needed problem [3]. We applied four popular algorithms Naive Bayes, SVM, Random Forest Tree and K- nearest neighbour to do the comparison that, which algorithm gives better accuracy for predicting purpose.
The rest of the paper is assembled as follows. Section II of the paper analysis related works. Section III describes the particulars of the proposed method for finding better accuracy. Section IV evaluates the experiment and displays the experiment results phase-wise. Section V winds up the paper in a nutshell and highlights some future work that can be done.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

II. LITERATURE SURVEY

1) The Psychological Effect of Weather on Car Purchases

When buying durable goods, consumers must forecast how much utility they will derive from future consumption, including consumption in different states of the world. This can be complicated for consumers because making intertemporal evaluations may expose them to a variety of psychological biases such as present bias, projection bias, and salience effects. We investigate whether consumers are affected by such intertemporal biases when they purchase automobiles. Using data for more than 40 million vehicle transactions, we explore the impact of weather on purchasing decisions. We find that the choice to purchase a convertible or a four-wheel-drive is highly dependent on the weather at the time of purchase in a way that is inconsistent with classical utility theory. We consider a range of rational explanations for the empirical effects we find, but none can explain fully the effects we estimate. We then discuss and explore projection bias and salience as two primary psychological mechanisms that are consistent with our results.

2) Development of Data Mining System to Analyze Cars using TkNN Clustering Algorithm

Now a day's customers are required comfort and their loving brand & color. With the advent of the Internet and Data Mining Algorithms has undoubtedly contributed to the shift of marketing focus. Conventional way of business is a challenging in car market due to many competitors are there around the world for providing competitive products. The car manufacturers categorizes the car users and have to invent a suitable car; the seller correctly groups the buyers and he sells a right car; and the customers selects best car by analyzing more brands of cars with 'N' number of sellers. These three cases they spent too much of time for analyzing old or statistical data for choosing a right product. In this paper, we proposed TkNN algorithm for best car market analysis. We have executed the same in Weka Tool with Java code. We analyzed the graphical performance analysis between KNN and our novel TkNN clustering algorithms with Classes to Clusters evaluation purchase, safety, luggage booting, persons (seating capacity), doors, maintenance and buying attributes of customer's requirements for unacceptable/acceptable/good/very good ratings of a car to purchase. Moving on to more advanced topics, the author explains how to achieve better performance through ensemble pruning and how to generate better clustering results by combining multiple clusterings. In addition, he describes developments of ensemble methods in semi- supervised learning, active learning, cost-sensitive learning, class-imbalance learning, and comprehensibility enhancement.

3) Implementation of Naïve Bayes Classification Method for Predicting Purchase

To choose the right vehicle according to the needs and funds owned by consumers, requires a careful analysis that takes into account many criteria and factors. The criteria used as a benchmark in choosing a vehicle, among others, price, spare parts, cylinder volume, the power of the vehicle. To process all these criteria required a system that can select and classify criteria chosen by consumer, so that can assist consumer in choosing the most appropriate vehicle, therefore needed a system for decision making in making car purchase. The Naive Bayes algorithm is a simple probabilistic classifier that computes a set of probabilities by summing the frequency and value combinations of the given dataset. Application of Naïve Bayes method is expected to be able to predict car purchases. Of the 20 car purchase data used in the test by the Naïve Bayes method, then obtained a percentage of 75% for the accuracy of prediction, where from 20 car purchase data tested there are 15 data purchase car successfully classified correctly.

4) Data classification using Support vector Machine (SVM), a simplified approach

In all our day to day activities we will be classifying things based on situations and on our needs. Human beings do classification of any kind by their natural perception. Classifying data is a common task in machine learning which requires artificial intelligence. Support vector Machine (SVM) is a new technique suitable for binary classification tasks. SVMs are a set of supervised learning methods used for classification, regression and outliers detection. The SVM classifiers work for both linear and nonlinear class of data through Kernel tricks. A Support Vector Machine is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data, the algorithm outputs an optimal hyperplane which categorizes new samples. In this paper, use of SVM for data classification is presented in a simplified way. Discussions are justified with illustrative practical examples. An effective algorithm is

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

developed for data classification on python platform using sklearn tool kit. The results are exhibited both symbolically and graphically. This paper is expected to be an insight for desired readers and researchers in implementing their ideas of item classification using SVM.

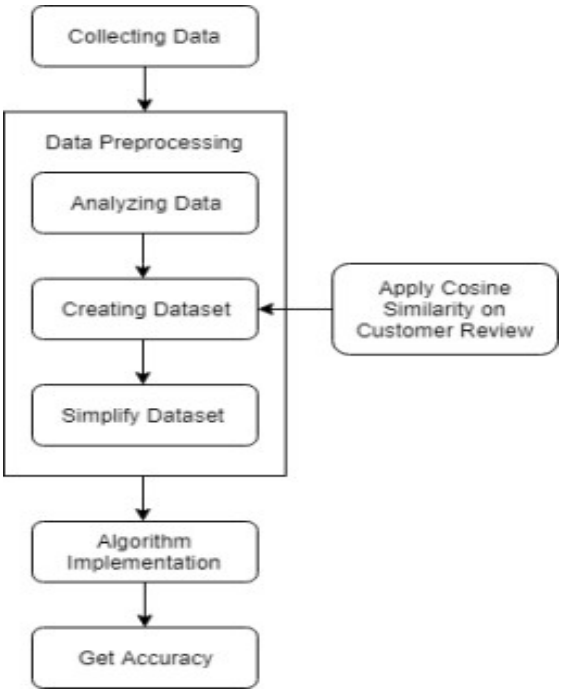
5) Comparative analysis of machine learning algorithms on social media test
Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in main text. It mainly refers to a text classification. Social media is generating a vast amount of sentiment rich data in the form of tweets, blog posts, comments, status updates, news etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the public. Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this paper, Machine learning approach has been used for the sentiment analysis of movie review dataset and is analysed by Naïve Bayes, Decision tree, KNN, and SVM classifiers. Commencing the most efficient classification technique is the moto of the paper. Efficiency of the classifier is decided based on some regular parameters that are outputs of the classification techniques.

III. PROPOSED METHODOLOGY

The aim of this research is to analyze the accuracy of different predictive algorithms that can predict the probability of car purchasing. We focused on customer's review and used cosine similarity to analyze customer's review. Support vector machine gave the highest accuracy among all the algorithms. Data Collection: We collected a diverse dataset containing information on customer demographics, historical car purchase records, preferred car features, and other relevant attributes. The dataset was sourced from various sources, including surveys, online car marketplaces, and customer databases of car dealerships. Feature Engineering: To prepare the data for analysis, we conducted feature engineering tasks such as data cleaning, data transformation, and feature selection. This step involved handling missing values, encoding categorical variables, and normalizing numerical features.

IV. SYSTEM DESIGN

System Architecture



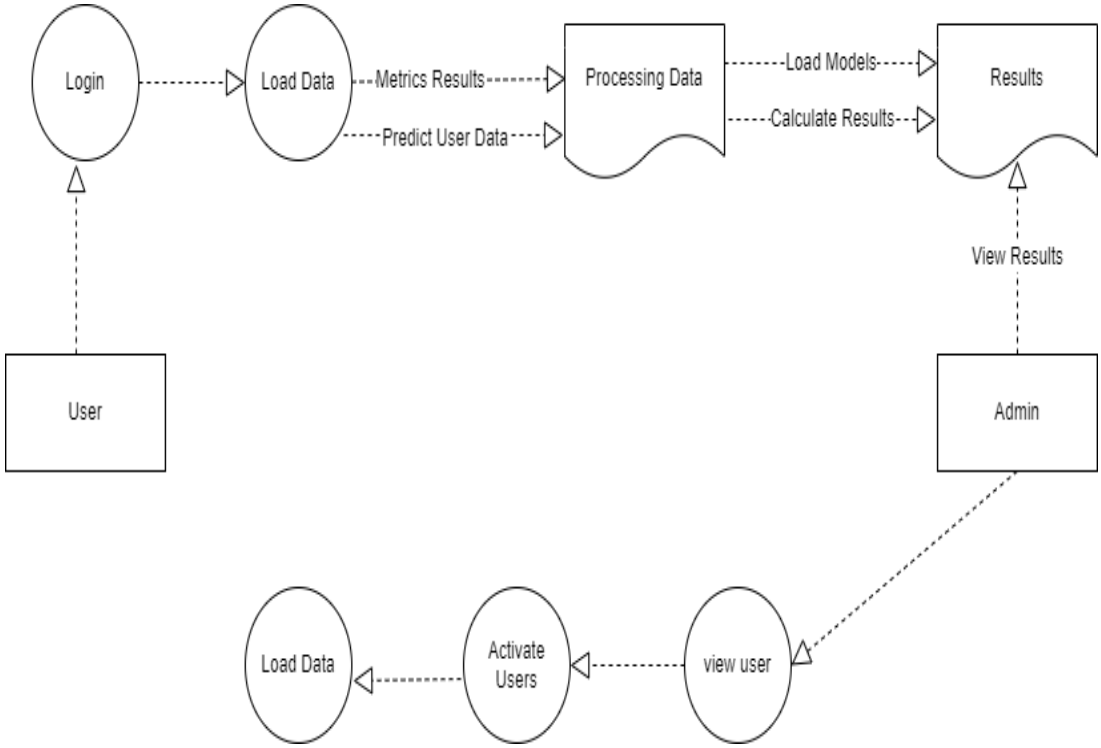
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

Data Flow Diagram

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.

DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



UML Diagrams

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

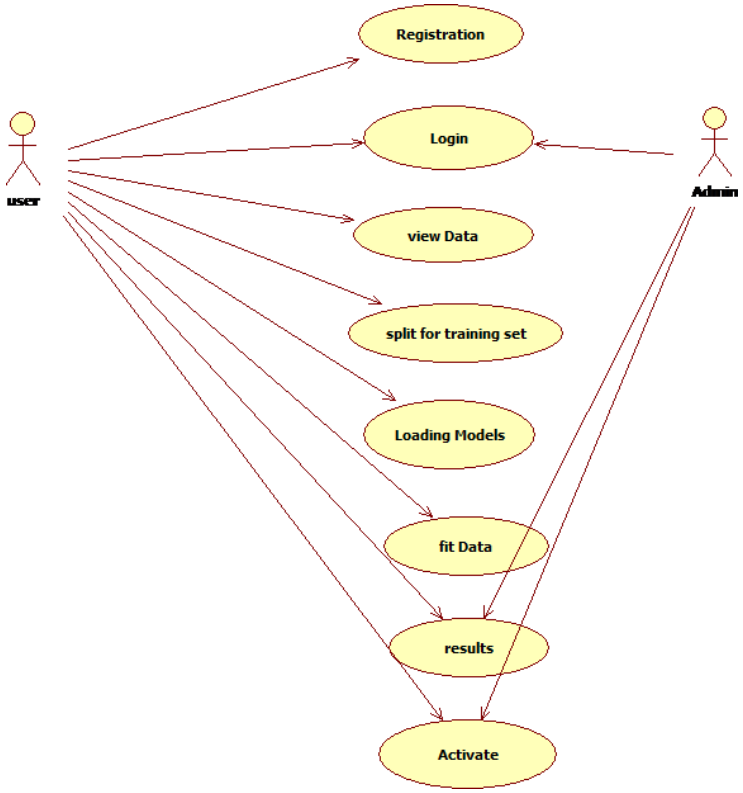
The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



IV. RESULT AND DISCUSSION

For predicating car purchase based on customer demand machine learning provides several benefits. In machine learning svm plays an important role we proposed TkNN for car analysis . We applied four popular algorithms Naive Bayes, SVM, Random Forest Tree and K-nearest neighbour to do the comparison that, which algorithm gives better accuracy for predicting purpose. Performance Comparison: After evaluating the different machine learning algorithms, we found that the random forest algorithm outperformed the other models in terms of accuracy, precision, recall, and F1-score. It exhibited a high level of predictive accuracy and robustness in capturing complex relationships between customer demands and car purchases. Important Features: Through feature importance analysis provided by the random forest algorithm, we identified the key factors that significantly influenced car purchasing decisions. These included factors such as price range, car model, fuel efficiency, safety features, brand reputation, and customer preferences for specific features like navigation systems, advanced driver assistance systems (ADAS), and entertainment options.

Practical Implications: The findings of this study have practical implications for car manufacturers and dealerships. By leveraging the predictive models developed through machine learning algorithms, businesses can gain insights into customer demands and tailor their marketing strategies, inventory management, and product development efforts accordingly. This can lead to improved customer satisfaction, increased sales, and enhanced competitiveness in the automotive market

V. Conclusion

In our study, we mainly focused on Customer's review and we have used Cosine Similarity to analyze customer's review. Then we applied several algorithms on our dataset. We have compared those algorithms according to their accuracy. Support Vector Machine has given the highest accuracy among all the algorithms.Our main goal is To help the customer to make the right decision whether he will buy a or or not and proposed some well known algorithm to geta better

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 13, May 2023

accuracy for car purchase The findings of this study can contribute to more informed decision- making processes for car manufacturers and dealerships, ultimately leading to improved customer satisfaction and business performance in the automotive industry.

REFERENCES

[1]. J. C. Pope and J. Silva-Risso, "The psychological effect of weather oncar purchases* meghan r. busse devin g. pope," The Quarterly Journalof Economics, vol. 1, no. 44, p. 44, 2014.

[2]. M. Jayakameswaraiah and S. Ramakrishna, "Development of datamining system to analyze cars using tknn clustering algorithm," InternationalJournal of Advanced Research in Computer Engineering TECHNOLOGY, VOL. 3, NO. 7, 2014.

[3]. F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah,and C. B. Harahap, "Implementation of na¨ive bayes classificationmethod for predicting purchase," in 2018 6th International Conferenceon Cyber and IT Service Management (CITSM). IEEE, 2018, pp. 1–5.

[4]. K. S. Durgesh and B. Lekha, "Data classification using support vectormachine," Journal of theoretical and applied information technology, vol. 12, no. 1, pp. 1–7, 2010.

[5]. R. Ragupathy and L. Phaneendra Maguluri, "Comparative analysis ofmachine learning algorithms on social media test," International Journal of Engineering and Technology(UAE), vol. 7, pp. 284–290, 03 2018.

[6]. K. Noor and S. Jan, "Vehicle price prediction system using machinelearning techniques," International Journal of Computer Applications, vol. 167, no. 9, pp. 27–31, 2017.

[7]. N. Pal, P. Arora, P. Kohli, D. Sundararaman, and S. S. Palakurthy, "How much is my car worth? a methodology for predicting used cars prices using random forest," in Future of Information and Communication Conference. Springer, 2018, pp. 413–422.

[8]. S. Pudaruth, "Predicting the price of used cars using machine learningtechniques," Int. J. Inf. Comput. Technol, vol. 4, no. 7, pp. 753–764,2014.

[9]. F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, andJ. Akinjobi, "Supervised machine learning algorithms: classification andcomparison," International Journal of Computer Trends and TECHNOLOGY(IJCTT), VOL. 48, NO. 3, PP. 128–138, 2017.

[10]. M. R. Busse, D. G. Pope, J. C. Pope, and J. Silva-Risso, "Thepsychological effect of weather on car purchases," The Quarterly Journalof Economics, vol. 130, no. 1, pp. 371–414, 2015.

[11]. S. Veni and A. Srinivasan, "Defect classification using na¨ive bayesclassification," Interbational Journal of Applied Engineering Research,vol. 12, no. 22, pp. 12 693–12 700, 2017.

[12]. E. Gegic, B. Isakovic, D. Keco, Z. Masetic, and J. Kevric, "Car priceprediction using machine learning techniques," 2019.

[13]. M. Jabbar, "Prediction of heart disease using k-nearest neighbor andparticle swarm optimization," Biomed. Res, vol. 28, no. 9, pp. 4154–4158, 2017.

[14]. M. C. Sorkun, "Secondhand car price estimation using artificial neuralnetwork."