International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-01, Apr 2023

ML Algo for Predicting Car Purchase Based on Customer Demand

[1]Prof. Namrata Ganesh Daware,[2]Ms. Vedanti Sunil Shinde,[3]Ms. Rashi Ravindra Vekhande,
[4]Mr. Sanjyot Uttam Amale.
[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhale College of Engineering & Technology, Asangaon, Maharashtra, India.
[1]namratashinde9881@gmail.com,[2]shinde.ved124@gmail.com,[3]rashivekhande@gmail.com,
[4]sanjyotamale356@gmail.com

Abstract - One of the most important contributions to the nation's economy is the sedulity of machines. Motorcars are becoming more and more admired for their use in private transportation. When he wants to buy the right vehicle, especially a bus, the customers need to review it. because it is a car that is truly priceless. Before purchasing a new bus, there are numerous considerations, including price, spare corridor, cylinder volume, and headlight. As a result, the customer must make the best purchase that meets all requirements for everything to be decided. The job is to assist the customer in making an informed decision about whether or not to purchase a bus. Consequently, there was a need to find a way to make decisions about bus steal systems. Some proposals were made to make a few notable calculations in the paper in order to improve the precision of a transportation purchase. These algorithms were applied to a 50-dataset dataset. With 86.7 delicacies of prophecy, Support Vector Machine (SVM) comes out on top among them. For each data sample, there were also relative results obtained with various algorithms for perfection, recall, and F1 score in this paper. Index Terms include Accuracy, Cosine Similarity, Supervised Machine Knowledge, Naive Bayes, Random Forest Tree, Support Vector Machine, and KNN.[1]

Keywords- Support Vector Machine, Machine Learning, KNN, Random Forest Tree, Naive Bayes.

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.[1] For example, if a person desires to choose a job or is planning for where to stay or planning for a fine vacation, all are important in life decisions. Because these are needed to consider the utility that will increase in the future. People always like to make those types of decisions that maximize utility.[1] Since utility is linked with the financial system in day-to-day life. At present, the automobile industry is one of the most important businesses in the world. Though Bangladesh is a small country in South Asia, the demand for automobiles is increasing with each passing day. [1] People are using private vehicles to move from one location to another location. The four-wheeled private car is good and flexible among those. A good portion of the economic development of a country depends on the transportation system.[1] Because when the transport system is efficient, it affords economic as well as a social opportunities which gives a positive effect on markets. So earlier than buying a new car, customers prefer to be assured of the money they spend to be worthy. [1] Because buying a new car is a matter of a good amount of money from the perspective of the Bangladeshi economy. That is why it is important to get information about cars which are good or bad based on customer's experiences, who bought those before. The lifecycle of a modern car depends on so many different parts. [1] In this paper, the aim is to predict the likelihood of purchasing a vehicle based on price, spare part availability, reviews from customers, cylinder volume, and the resale price. Foreseeing the probability or likelihood of procurement for vehicles is a magnificent and genuinely necessary issue. The appliance of four popular algorithms Naive Bayes, SVM, Random Forest Tree, and K-nearest neighbor to do the comparison, which algorithm gives better accuracy for predicting purposes.

II. AIMS AND OBJECTIVE

a) Aim

One of the key sectors of the national economy is the auto business. Cars are becoming more and more common for individual transportation. When a client wants to purchase the ideal vehicle, particularly a car, a review is necessary.

 DOI : 10.5281/zenodo.7863730 © 2023, IJREAM All Rights Reserved.

because the car is very expensive. Before purchasing a new vehicle, there are a variety of conditions and factors to consider, including price, price in particular, spare parts, cylinder volume, and headlights. To satisfy all the requirements, the customer must make the best purchase decision. The goal is to help the customer make the best decision about whether or not to buy a car. So, the goal was to create a process for choosing an in-car buy.[1]

b) Objective

• This study aims to estimate the likelihood of purchasing a vehicle based on factors such as price, customer feedback, cylinder volume, and resale value. A fantastic and urgent issue is estimating the likelihood or probability of purchase for vehicles.

• It also revealed the comparative results using different algorithms precision, recall, and F1 score for all data samples to compare the performance of popular algorithms Naive Bayes, SVM, Random Forest Tree, and K-nearest Neighbor.[6]

III. LITERATURE SURVEY

1) The Psychological Effect Of Weather On Car Purchases:

AUTHORS: Devin G. Pope, JarenC. Pope, Meghan R. Busse, and Jorge Silva- Risso In "The psychological effect of weather on car purchases" by J. C. Pope and J. Silva-Risso, the authors investigate the relationship between weather conditions and car purchases. They examine data from car dealerships and find that customers are more likely to purchase cars on sunny days compared to cloudy or rainy days. The study suggests that weather conditions can have a significant impact on consumer behavior.[5]

2) Implementation of Naive Bayes Classification Method for Predicting Purchase

AUTHORS: Fitriana Harahap; Ahir Yugo Nugroho Harahap; Evri Ekadiansyah

In "Implementation of Naïve Bayes Classification Method for Predicting Purchase" by F. Harahap et al., the authors propose the use of the Naïve Bayes classification method for predicting purchases. They apply the method to a dataset of customer purchases and evaluate its performance using metrics such as accuracy, precision, and recall. The study concludes that Naïve Bayes is an effective method for predicting purchases.[2][3]

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

AUTHORS: Dr. S V Sathyanarayana In "Data classification using support vector machine" by K. S. Durgesh and B. Lekha, the authors investigate the use of support vector machines (SVMs) for data classification. They evaluate the performance of SVMs using various datasets and compare it to other classification methods such as decision trees and neural networks. The study concludes that SVMs are an effective method for data classification, particularly for datasets with a large number of features.[7]

4) Comparative analysis of machine learning algorithms on social media test :

AUTHORS: R. Ragupathy

In "Comparative analysis of machine learning algorithms on social media test" by R. Ragupathy and L. Phaneendra Maguluri, the authors compare the performance of various machine learning algorithms on social media data. They evaluate algorithms such as k-nearest neighbors, decision trees. and support vector machines and compare their performance based on metrics such as accuracy and precision. The study concludes that decision trees and support vector machines perform the best on social media data.[3]

IV. EXISTING SYSTEM

Some people prefer good parts, some are high or low price with all of their needed features, and some are weak for famous brands of the car only. [1] Selecting the perfect car is still a difficult task though some parameters like color, comfort, seating capacity, etc.[1] In the existing system,it used a logistic Regression algorithm for predicting the car purchase accuracy.

V. COMPARATIVE STUDY

Sr. No	Paper Name	Author/ Publication	Technology	Advantage	Disadvantage
1.	THE PSYCHOLOGICAL EFFECT OF WEATHER ON CAR PURCHASE	J. C. POPE AND J. SILVA-RISSO ELSEVIER, 2019	ML, R-Studio, Rapid	The study uses a large and diverse sample of car buyers (over 9 million transactions) across multiple states in the US, which increases the generalizability of the findings.	The study focuses solely on the impact of weather on car purchase, without considering other factors such as advertising, pricing, and promotions that could also influence consumer behavior.

 DOI : 10.5281/zenodo.7863730 © 2023, IJREAM All Rights Reserved.

2	DATA CLASSIFICATION USING SUPPORT VECTOR	K. S. DURGESH AND B. LEKHA IEEE,2019	ML	Support Vector Machines (SVMs) are a popular and powerful classification algorithm that can be used for both linear and nonlinear classification problems.	The paper focuses solely on SVMs, which means that other classification algorithms and techniques are not covered in the study.
3.	VEHICLE PRICE PREDICTION SYSTEM USING MACHINE LEARNING TECHNIQUES	K. NOOR AND S. JAN SpringerOpen,2 016	MLDPS	The authors use a variety of machine learning techniques, including linear regression, decision trees, and random forests, which allows for a comprehensive comparison of different algorithms.	The paper does not provide a detailed explanation of the features used in the predictive model, which makes it difficult to assess the relevance and significance of the variables.

Table 1: Comparative Analysis

VI. PROBLEM STATEMENT

The challenge for a bus dealership when buying a used vehicle is the risk of purchasing a "kicked" bus, which can have serious issues preventing its resale. This can be costly due to transportation, repair, and lost profits. The study focuses on data analysis, preprocessing, and the use of machine learning models such as Linear Regression and KNN to predict and compare the effectiveness of the models in identifying "kicked" buses.

VII. PROPOSED SYSTEM

The end of this disquisition is to anatomize the delicacy of different predictive algorithms that can predict the probability of copping a bus. A relative analysis of machine knowledge algorithms was proposed by Ragupathy. In this paper, they tried to identify and classify the sentiment, conveyed in the main text. They have collected their data from social media like Twitter, commentary, blog posts, news, status updates, etc.[4] They also applied Naive Bayes, K-Nearest Neighbor, Random timber, and Support Vector Machine classifiers for their comparison purpose. The data was collected from social media. After successfully creating the dataset, estimated algorithms by ramifying the dataset. Used 70 of the data as training and 30 of the data as testing. To estimate the results, Used the perfection- recall, execution time, and delicacy dimension of the algorithms.[1]

VIII. ALGORITHM

Step 1: Start

Step 2: Import necessary libraries: pandas, DecisionTreeClassifier, RandomForestClassifier, LogisticRegression, GaussianNB, KNeighborsClassifier, and SVC.

Step 3: Load the car purchase dataset from a CSV file using the pandas library.

```
df = pd.read_csv(path)
X = df.iloc[:, :-1].values
y = df.iloc[:, -1].values
```

Step 4: Split the dataset into training and testing data.

```
X_train, X_test, y_train, y_test =
train_test_split(X, y, train_size=0.75)
```

Step 5: Create a dictionary to store the different classifier models.

```
classifiers = {
'Logistic Regression': LogisticRegression(),
'Decision Tree': DecisionTreeClassifier(),
'Random Forest': RandomForestClassifier(),
'Naive Bayes': GaussianNB(),
'KNN': KNeighborsClassifier()
}
```

Step 6: For each classifier in the dictionary, train the model on the training data using the fit method.

```
results = {}
for name, classifier in classifiers.items():
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
results[name] = {'accuracy': accuracy, 'precision':
precision, 'recall': recall, 'f1_score': f1}
```

Step 7: Print the evaluation metrics for each classifier and display the result.

```
accuracy = accuracy_score(y_test, y_pred)
print('ANN Accuracy:', accuracy)
cm = confusion_matrix(y_test, y_pred)
precision = precision_score(y_test, y_pred)
print('ANN Precision Score:', precision)
recall = recall_score(y_test, y_pred)
print('ANN Recall Score:', recall)
```

 DOI : 10.5281/zenodo.7863730 © 2023, IJREAM All Rights Reserved.

```
f1score = f1_score(y_test, y_pred)
print('ANN F1-Score:', f1score)
return accuracy, precision, recall, f1score
```

Step 8: Exit

IX. MATHEMATICAL MODEL

Let X be the matrix of input features with dimensions (m, n), where m is the number of samples and n is the number of features. Let y be the target variable with dimensions (m, 1), where y=1 indicates that the customer has purchased a car, and y=0 indicates that the customer has not purchased a car. Let T be the number of trees in the Random Forest ensemble, and let t be the index of the individual tree. Let h_t(x) be the prediction of the t-th tree for input feature vector x. The Random Forest algorithm works by averaging the predictions of T individual decision trees, each trained on a subset of the data. The prediction for a new input feature vector x can be calculated as follows:

y_pred(x) = (1/T) * sum(h_t(x))[6]

X. SYSTEM ARCHITECTURE

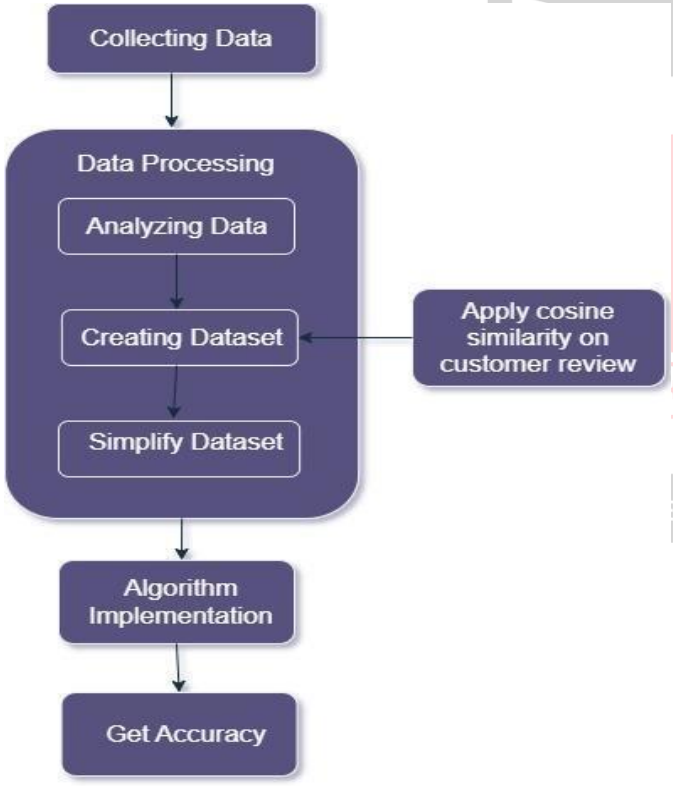

Fig.1: System Architecture

The ML algorithm for predicting car purchase based on customer demand involves collecting and processing data, applying cosine similarity on customer reviews, implementing various algorithms, and determining accuracy through performance evaluation. The accuracy of the algorithm is an important factor in predicting the probability of car purchase based on customer demand.

XI. ADVANTAGES

• Helps break complex real-world problems with several constraints.

• Ease of transferring knowledge from one model to another grounded on disciplines and tasks.

• Provides a path towards achieving Artificial General Intelligence someday in the future.

• It automatically detects important features without any mortal supervision.

• Helps in achieving even more delicacy with a smaller amount of data.

XII. DESIGN DETAILS

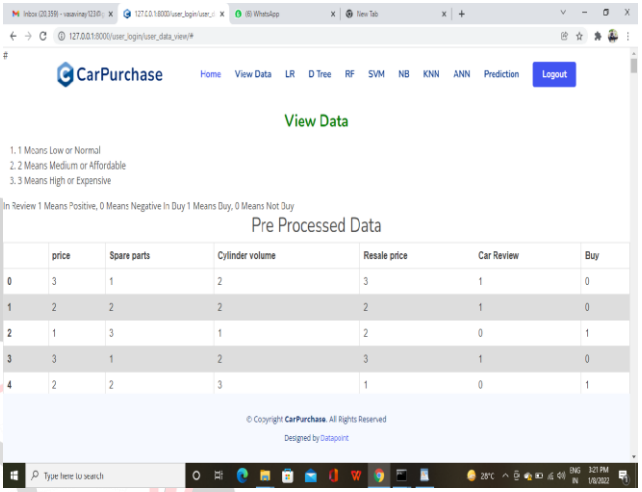

Fig 2: Result

Classification Algorithm	Accuracy	Precision	Recall	F1 - measure
Logistic Regression	84.31%	85.18%	85.18%	85.18%
Decision Tree	84.31%	85.18%	85.18%	85.18%
Random Forest	84.31%	85.18%	85.18%	85.18%
Naive Bayes	84.31%	85.18%	85.18%	85.18%
SVM	84.31%	85.18%	85.18%	85.18%
KNN	84.31%	85.18%	85.18%	85.18%

Table 2: Resultant Analysis

 DOI : 10.5281/zenodo.7863730 © 2023, IJREAM All Rights Reserved.

XIII. CONCLUSION

Thus, We have tried to implement the paper "Anamika Das Mou; Protap Kumar Saha; Sumiya Akter Nisher; Anirban Saha ", "A Comprehensive Study of Machine Learning algorithms for Predicting Car Purchase Based on Customers Demands", IEEE 2021.The paper emphasizes the significance of customer reviews in forecasting car purchases and using machine learning algorithms to identify key features customers consider. The study suggests that Support Vector Machine outperforms other algorithms and can be used to improve products and marketing strategies. However, the research has limitations and future studies could investigate advanced algorithms and external factors affecting customer demand. This study provides important insights into predicting car purchases based on customer demands using machine learning.

REFERENCE

[1] Anamika Das Mou; Protap Kumar Saha; Sumiya Akter Nisher; Anirban Saha "A Comprehensive Study of Machine Learning Algorithms for Predicting Car Purchase Based on Customers' Demands".

[2] Prof. Vishal R. Shinde, "Ingredients Based Food Recommendation System Using Naive Bayes And KNN Algorithm" in IJREAM, ISSN : 2454-9150, Vol-06, Special Issue, June 2020F.

[3] Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," In 2018 6th International Conference On Cyber And IT Service Management(CITSM). IEEE, 2018, PP. 1–5.

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

[5] J. C. Pope and J. Silva-Risso, "The psychological effect of weather on car purchases" Meghan r. Busse, Devin g. pope," The Quarterly Journal of Economics, VOL. 1, NO. 44, P. 44, 2014.

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

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

 DOI : 10.5281/zenodo.7863730 © 2023, IJREAM All Rights Reserved.