Utilizing AI Techniques for Vehicle Information Extraction and Analysis to Predict Trends in the Automotive Market: A Comprehensive Study

J Madhubala

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— *The paper explores the application of data mining in car manufacturing, aiming to develop a reliable classifier for future product classifications. Using datasets from the UCI Machine Learning repository, the study focuses on establishing a classifier capable of accurately predicting vehicle attributes such as type, body-style, horsepower, and fuel type. By leveraging AI techniques, the research aims to improve the accuracy of vehicle market forecasts, aiding in efficient decision-making and sales promotion strategies. Three classification algorithms—Naive Bayes, Logistic Regression, and Multilayer Perceptron—are employed to analyze a vehicle evaluation database, with the extracted insights facilitating informed decisions in vehicle sales promotion. Ultimately, the study demonstrates the potential of data mining and AI methodologies in enhancing predictive modeling and optimizing automotive industry operations.*

I. INTRODUCTION

Vehicle selection is influenced by a range of factors including model preferences, manufacturer reputation, cost, safety features, and luxury amenities. These attributes collectively contribute to accident prevention and overall satisfaction with the purchased vehicle. Buyers also consider standard equipment such as performance enhancements, convenience features, and safety mechanisms when making their decisions. Safety features are particularly crucial, along with convenience attributes like maintenance ease, door configurations, and storage capacity. Cost-effectiveness is essential to ensure that the vehicle purchased provides value for money. Additionally, owning a vehicle entails financial responsibility for its maintenance and upkeep.

This research assesses the perceived value of vehicle cost compared to other attributes such as convenience features, storage space, individual preferences, and safety measures. Data mining, a branch of Artificial Intelligence, proves instrumental in exploring uncharted relationships within vast datasets. Its applications extend across diverse fields including manufacturing, healthcare, business, and education. Various data mining techniques, such as Artificial Neural Networks, Support Vector Machines, and Naive Bayes, offer unique insights into data patterns.

The primary focus of this study is to evaluate the performance of three prominent algorithms—Naive Bayes, Logistic Regression, and Multilayer Perceptron—in terms of speed and accuracy. By analyzing a dataset comprising vehicle attributes, the study aims to determine the efficacy of these algorithms in predicting key aspects of vehicle selection.

II. CLASSIFICATION

Approach is the way toward tracking down a model or a cutoff that portrays and sees information classes and contemplations, to utilize the model to foresee the classes of things whose class mark isn't known. Information solicitation can be seen as a two-stage measure: learning step in which a classifier is created portraying a destined blueprint of classes or musings by isolating the status set contained enlightening list tuples and their associated names [2]. In the subsequent development model is utilized for demand by first assessing the prudent precision of classifier worked during the fundamental development. It is finished utilizing the test information. The accuracy of classifier on a given test set tuples is level of tuples that are correctly mentioned by the classifier. On the off chance that the precision is over some good level, the classifier can be utilized to expect future tuples whose class mark isn't known.

Depiction is a sort of information appraisal that can be utilized to make models depicting gigantic information classes. Strategy is an information mining approach used to foresee pack revenue for information models. It is one of the basic frameworks in information mining and is utilized in different applications, for example, plan attestation, affliction confirmation, client relationship the pioneers, and allocated showing. The objective of the depiction assessments is to gather a model from a ton of preparing information whose target class names are known and therefore this model is utilized to pack covered cases [1] [3].

Plan is the most ordinary and most prestigious information mining strategies. Strategy maps information into predefined social events or classes. It is run of the mill proposed as overseen getting the hang of considering how the classes are settled going before looking at the information. Strategy is the way toward tracking down a model that sees information classes, to utilize the model to anticipate the class of things whose class name is dull. The chose model depends upon the appraisal of a ton of preparing information. Enlightening assortments are rich with disguised data that can be utilized for vigilant dynamic.

Building unequivocal and valuable classifiers for gigantic information bases is one of the essential undertakings of information mining and AI research. Building productive solicitation frameworks is one of the focal errands of information mining.

III. METHODOLOGY

At this moment clarified about administered learning methods like Naive Bayes, Multilayer Perceptron and Logistic Regression system models for our vehicle order issue.

3.1 Naïve Bayes Algorithm

The ML Naïve Bayes calculation is utilized for grouping learning assignments in which examples of the dataset are separated dependent on the predefined include. The calculation is probabilistic in nature and simultaneously depends on Bayes Theorem [3]. The Eq. (1) beneath shows Bayes Theorem:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad (1)$$

Were,

P(A|B) = Probability of A happening given occasion B has effectively happened.

P(B|A) = Probability of A happening given occasion B has effectively happened.

P(A) = Probability of occasion A happening.

P(B) = Probability of occasion B happening.

Let, 'X' is another information point, discovered P(A|X) and P(B|X). Then, at that point our classifier analyzes those two and chooses X has a place with 'A' or 'B'.

3.2 Logistic Regression Algorithm

Strategic Regression ML calculation is utilized for characterization learning errands in which the affiliation versus absolute ward highlights against autonomous highlights are resolved [2]. The learning calculation is utilized when the reliant highlights have parallel qualities like 0 and 1, valid or bogus, negative or positive, and no or yes [3][5]. The following is the calculated relapse calculation numerical Eq. (2) used to compute the relationship between subordinate highlights and autonomous properties or highlights of the dataset:

$$I = \text{Logistic regression}(p) = \ln\left(\frac{P}{1-p}\right) \quad (2)$$

3.3 Multilayer Perceptron (MLP)

A MLP is a legend among the most all around saw Neural Network plan that has been used for various applications. The MLP figure out is customarily made out of different concentrations or overseeing units, and it is figured out into an improvement of no under two layers [4]. The essential layer (or the most decreased layer) is named as an information layer where it gets the external information while the last layer (or the most shocking layer) is a yield layer where the response for the issue is gotten. The secret layer is the comprehensively captivating layer in the data layer and the yield layer, and may layout with some place close to one layer. The arrangement of MLP could be bestowed as a nonlinear improvement issue. The objective of MLP learning is to find the best loads that limit the partition between the information and the yield. The most predominant getting ready evaluation used in NN is Back inducing (BP), and it has been used in overseeing various issues in model declaration and portrayal [6]. This estimation depends a few cutoff points, for instance, extraordinary covered center concentrations at the hid layers learning rate, energy rate, approval work and the measure of wanting to happen.

IV. EXPLORATORY RESULTS

In this work, a true vehicle assessment data set was taken from the UCI storehouse of AI data set [7]. It contains 1728 examples and grouped into four classes, there is no missing worth in the dataset. The vehicle assessment data set contains six credits

models with a vehicle (Buying, Main, Doors, Persons, Lug boot and Safety). Among the 7 ascribes, initial six credits about the purchasing value, support cost, number of entryways, number of people to convey, size of the baggage boot, and the wellbeing level are utilized as information credits. The seventh characteristic class is utilized as the characterization yield to address 4 vehicle classes, i.e., unacc, acc, vgood and great dependent on the different vehicle highlights reasonable for a specific gathering of clients. The class insightful circulation of names is displayed in the figure-1. We have utilized the weka to explore our proposed calculations. Weka is a state-of-the-art office for making ML techniques and their application to genuine data mining issues.

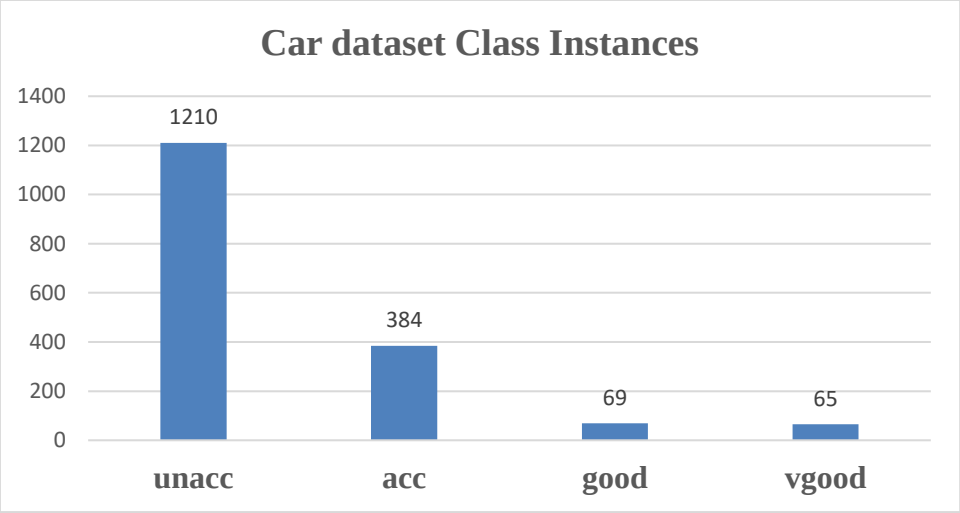

Figure-1: Frequency of class labels of the car dataset

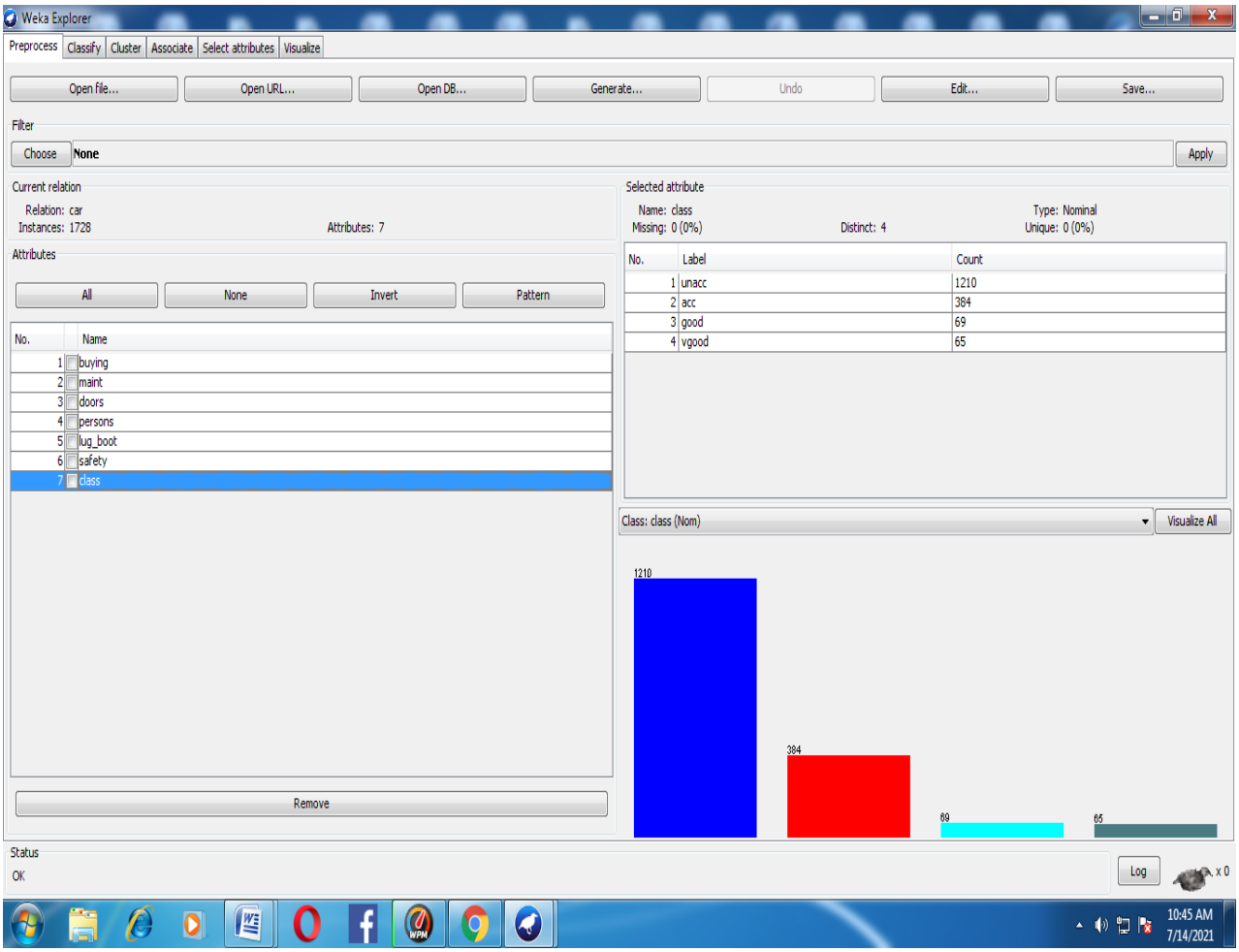

Figure-2: Summary of the dataset

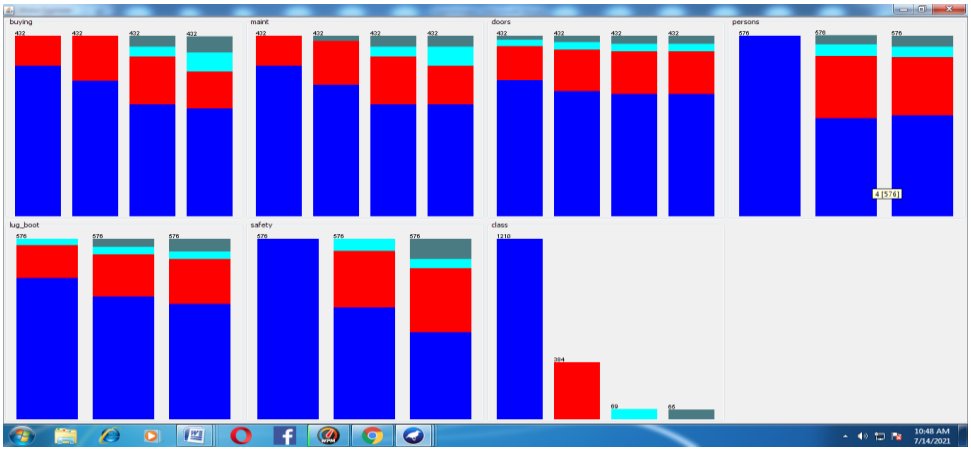

Figure-3: Statistical summary of the dataset

4.1 Results

This segment presents aftereffects of the experimentation arrangement. The interaction is as per the following; it is regulated learning strategy. We have prepared the model using ascribes comprehensive of class credits. As it's anything but an administered model, the model is constructed basing on the class esteems in correspondence to the upsides of traits independently. Weka is utilized for recreation reason. The outcomes accomplished by different experimentation arrangement in Naïve Bayesian, Logistic Regression and multi-facet perceptron are expounded in Table-1 and figure-4 separately.

TABLE-1
EXPERIMENTAL RESULTS

Algorithm	Accuracy	Precision	Recall
Naïve Bayes	93	93	93
Logistic Regression	94	93	93
MLP	99.5	99.5	99.5

Classification performance, classification accuracy by class, and precision of various classifiers are shown using bar charts for each of the 4 car classes for comparative analysis, which are graphically shown and explained below.

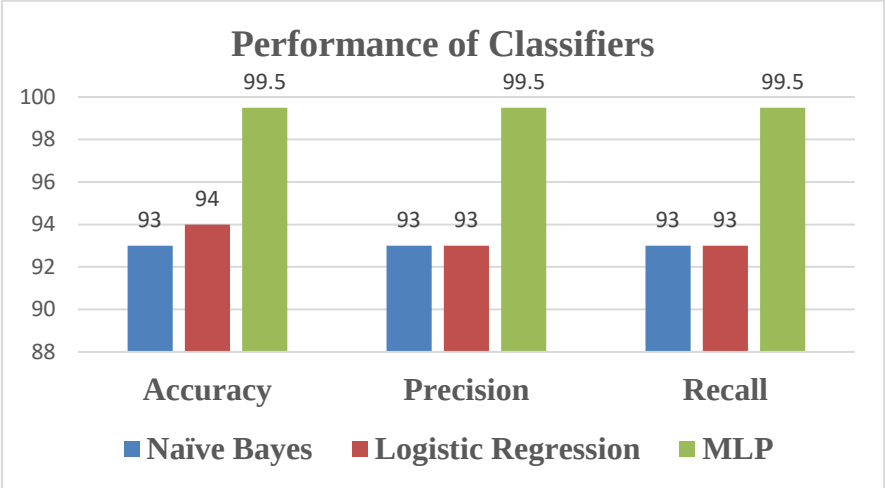

Figure-4: Performance of three algorithms

From the figure-4, we notice the exhibition of classification for Naïve Bayes 93% of Accuracy, Logistic Regression has achieved 94% of accuracy and the Multilayer Perceptron (MLP) has achieved the accuracy of 99.5%. So, the Multilayer Perceptron (MLP) classification has got highest accuracy when compared to Naïve Bayes and Logistic Regression. The screen shots of experimental results are shown in the figure-5, figure-6 and figure-7.

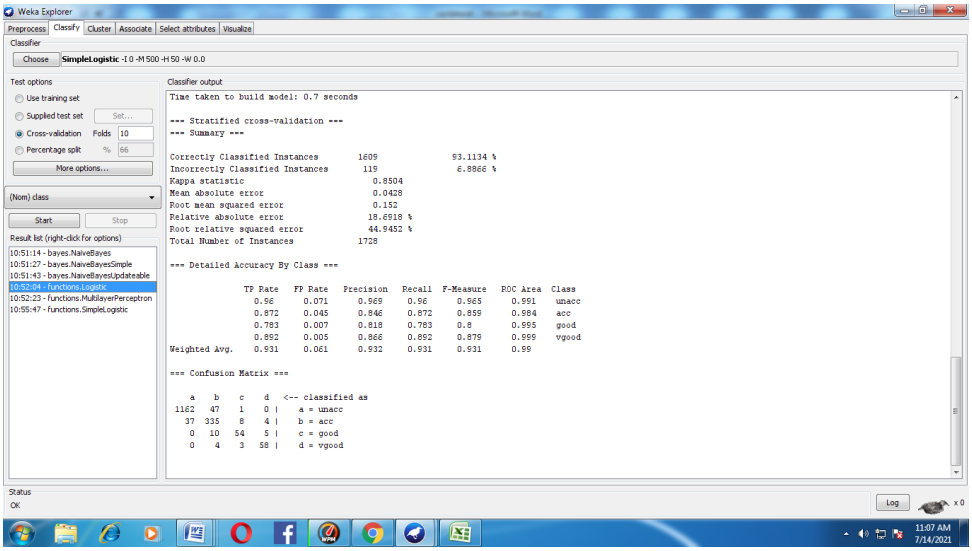

Figure-5: Screen shots of Experimental results

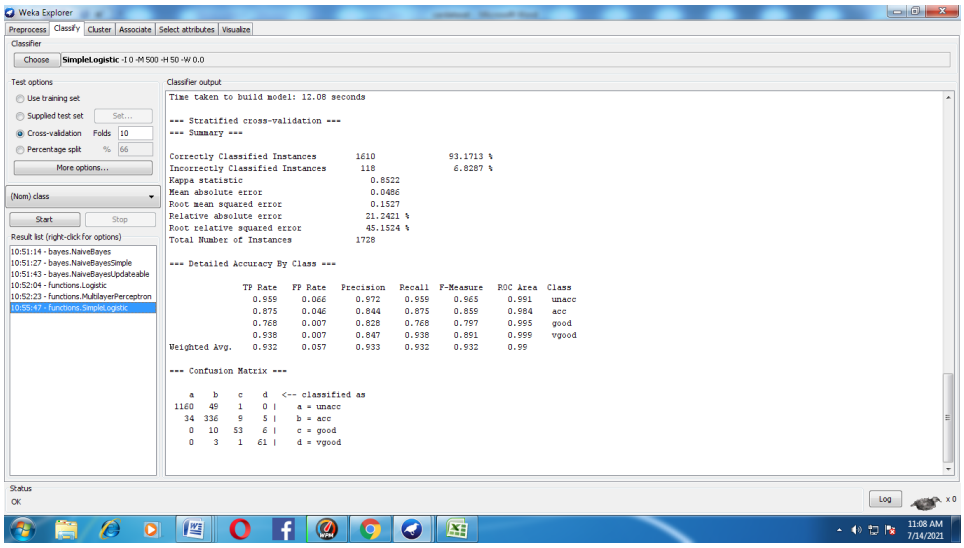

Figure-6: Screen shots of Experimental results

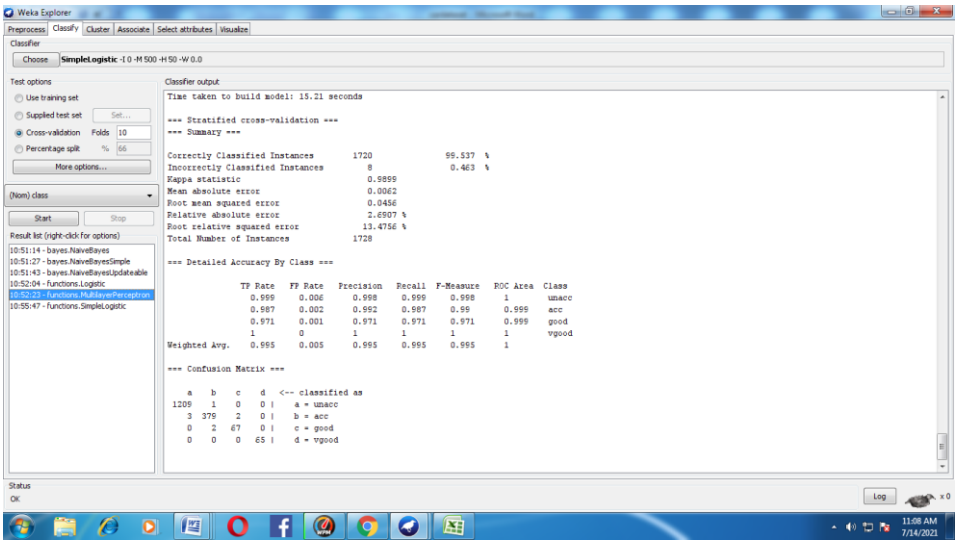

Figure-7: Screen shots of Experimental results

V. CONCLUSION

The research showcases three machine learning classifiers along with their comparative performance. Evaluating three models for experimental vehicle assessment, the results indicate that all three models exhibit similarly strong predictive capabilities. Among them, the Multilayer Perceptron (MLP) model demonstrates the highest accuracy and predictive ability, making it ideal for vehicle suitability assessment. In summary, leveraging data mining modeling techniques accurately predicts vehicle effectiveness, bridging the gap between consumers and businesses, thereby enhancing both enterprise profitability and consumer satisfaction.

REFERENCES

[1] G. Ravi Kumar, Venkata Sheshanna Kongara & Dr. G. A. Ramachandra, "An Efficient Ensemble Based Classification Techniques for Medical Diagnosis", International Journal of Latest Technology in Engineering, Management and Applied Sciences, Volume II, Issue VIII, Pages: 5-9, ISSN-2278-2540, August-2013

[2] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[3] J. Han and M. Kamber," Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.

[4] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems", 2nd edition, Addison Wesley, 2005.

[5] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.

[6] S. Haykin, Neural Networks – A Comprehensive Foundation, 2nd Edition, Pearson Education, Inc., Upper Saddle River, New Jersey 07458, 2000

[7] UCI machine learning repository. http://archive.ics.uci. edu/ml/