Car Evaluation Based on the Car Attributes and Personal Opinion

Gidanz Rafterjude Bibal, and Christian Villamera
Department of Computer Science
College of Engineering
University of the Philippines, Diliman

***Abstract*—The consumer's acceptability on your product is crucial if you want to sell something. With our program, it helps people who are planning to sell their car. There are 3 categories that are considered on the evaluation of the car: (1) price, (2) technicals, and (3) comfort. Based on the results, we conclude that safety of the car and maximum passengers capacity is the main factor on car acceptability to the buyer. Our program also accepts the owner's personal opinion to its car on which it can also affect the car acceptability.**

I. Introduction

A good reputation for quality vehicles is essential if one is thinking of selling used cars[7]. The decision of selling your very first or pre-loved car is definitely not an easy task. Several factors are to be considered especially the cars condition and specifications if you want to get the most out of your car.

This project aims to help and motivate the owner/seller of the car decide whether the car can be sold at a reasonable price or if the car isnt suited for selling at all. This project is organized under 6 main car attributes/specifications namely: Buying Price, Maintenance Price, Number of Doors, Passenger Capacity, Luggage Capacity, and the Overall Safety of the car. The following attributes plus the owners comments about his own car determine the cars evaluation of whether the car to be sold is either Unacceptable, Acceptable, Good, or Very Good.

II. Short of Review of Related Studies

A. Consumer Acceptability

Consumer Acceptability as defined, results from determining the feasibility of whether a product or service will be acceptable to the consumer requires tests, surveys, pretests and even prototypes[1]. Consumer Acceptability is an important factor in the product design process. For this project, we demonstrate the application of the proposed model, car evaluation to be exact, to show that the consumer acceptability problem can be evaluated and predicted using the proposed model[**?**].

B. Machine Learning Models

1) Hierarchical Decision Model(Decision Tree): A Hierarchical Decision Model such as a Decision Tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements[5]. For our project, the Decision Tree branches out but falls shortly only to the only possible outcomes: (positive, unacceptable), (positive, acceptable), (positive, good), (positive, very good), (negative, unacceptable), (negative, acceptable), (negative, good), and (negative, very good).

2) Sentiment Analysis: Sentiment analysis otherwise known as opinion mining is a much bandied about but often misunderstood term. In essence, it is the process of determining the emotional tone behind a series of words, used to gain an understanding of the the attitudes, opinions and emotions expressed within an online mention[3].

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative, or neutral[4]. For this project, we have applied sentiment analysis to gauge the comments and opinions of the owner to his car as input and determine it either as positive or negative one.

III. Methodology and Results

A. Methodology

For both the evaluation for the consumer acceptability on cars and the sentiment analysis on reviews,we used the datasets on the archive of UCI Machine Learning Repository.

The dataset for the consumer acceptability, Car Evaluation Data Set, evaluates cars according to six attributes: Buying Price, Maintenance Price, Number of Doors, Passenger Capacity, Luggage Capacity, and the Overall Safety of the car. There are four class values: unacceptable, acceptable, good, and very good. This dataset can be useful for structure discovery methods and decision analysis. It has 1728 number of instances.

The dataset for the sentiment analysis, Sentiment Labelled Sentences Data Set, determined if a review is positive or negative. It has 3000 reviews from IMDB, Amazon and Yelp.

For preprocessing Car Evaluation Data Set, we converted the values of the attributes of the cars into integer and store it in an array.

On preprocessing Sentiment Labelled Sentences Data Set, we did not remove the stopwords because it affects the accuracy of the model. We turn the words into numerical feature vector. After that, we transform it to a "Term Frequency times Inverse Document Frequency" (tf-idf) representation.

We used Decision Trees for the consumer acceptability and Multinomial Naive Bayes for the sentiment analysis.

		Prediction			
Actual		Unacceptable	Acceptable	Good	Very Good
	Unacceptable	478	2	0	0
	Acceptable	1	143	9	3
	Good	0	0	26	3
	Very Good	0	3	2	22

Fig. 1. Confusion Matrix for Decision Tree

		Prediction	
Actual		Negative	Positive
	Negative	504	113
	Positive	102	481

Fig. 2. Confusion Matrix for Multinomial Naive Bayes

For both of the models, we split the datasets into random subsets. The train set is 80% while the test set is 20%.

The accuracy of the decision tree is 96.68% while the accuracy of the Multinomial Naive Bayes is 82.08%.

B. Results

Our program aims to help the owner of the car to decide whether the car can be sold or if it is still consumer acceptable. There are two parts of our program: (1) objective part to evaluate the car acceptability based on its attributes, and (2) subjective part for the owner to have his/her personal opinion and review on the car.

The program have 8 possible outputs (refer to Fig. 3).

On several runs, we noticed that changing the values of safety of the car and maximum passengers capacity affects of the result of the car evaluation. With that, we build a decision tree with safety and passengers capacity as a variable (refer to Fig. 4).

C. Discussion

There are no input validations on our program. The program expects the user to only input on what the program says in the parethesis on the prompt.

The attributes we used on the car evaluation test is limited to 3 categories: (1) Overall Price (buying price and maintenance price), (2) Technical Characteristics, and (3) Comfort (number of doors, maximum passengers capacity, size of luggage boot and safety of the car). There are some attributes, categories, and characteristics of a car that can still affect the car acceptability to the buyer.

On sentiment analysis, the program might have a wrong output, whether it should be positive or the other way around, depending on the choice of the words of the user. The opinion of the owner can contain both positive and negative words and since our machine learning model does not look the content of the sentence, we might not get the result we wanted.

Car Evaluation	Sentiment Analysis	Output
Unacceptable	Negative	"Your car is not in good condition and customers will not like it. I'm sorry to say but your car isn't suited for selling."
Acceptable	Negative	"Your car has some issues but it can still be sold for a very cheap price."
Good	Negative	"Your car is a little bit rusty but can still do wonders on the road. It can be sold for a reasonable price."
Very Good	Negative	"You think that your car is in poor condition but some customers might like it and definitely will pay for it."
Unacceptable	Positive	"You might think that your car is still good but the customer might not like it. Sorry but it looks like no one would buy it."
Acceptable	Positive	"Your car is in good shape and works fine. It can be sold for a reasonable price."
Good	Positive	"You have a well maintained car. Buyers would surely be satisfied with it."
Very Good	Positive	"You have an amazing car!! It's still in a very good condition. Customers will definitely like it and it is certain that it can be sold at a valuable price."

Fig. 3. Output Table of the Program

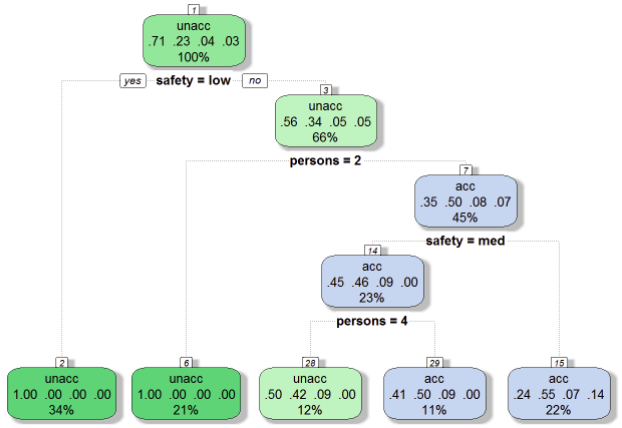

Fig. 4. Decision Tree with Safety and Maximum Passenger Capacity as a Variable

IV. Conclusion

Based on the results we get, all attributes of a car contribute a certain factor to the consumer's acceptability. Of all the attributes, safety and maximum passenger capacity have a great impact to the result meanwhile, the number of doors are the least important variable in deciding the class value of the car.

The owners positive comment does not guarantee to a customers acceptability in a car because it might get an opposite result in the car evaluation, and vice versa.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.