Navigating the Indian Auto Market: Unveiling Insights into Customer Car Buying Behavior for Optimal Sales Growth

By,
Anushka Poddaturi

ABSTRACT
India has one of the largest automobile industries in the world, and it is expanding consistently year after year. Because car loans are so readily available. There is an increase in the number of cars on the road due to wage patterns and the middle class's greater purchasing power. A questionnaire survey was undertaken in addition to a review of the relevant literature to better understand the different aspects that affect people's decisions to buy cars in India. A survey questionnaire was created in Google Forms and distributed to participants by email, Facebook, WhatsApp, and other channels. A paper copy was available for people who were not familiar with these technologies. In the current work, the buying habits of cars in India are studied The paper focuses on the car buying behavior of Indian customers based on some attributes like price, make, income, wife working yes or no etc.

ABOUT THE DATA

The objective here is improve the cars sales from the dealer outlets using machine learning techniques.

In the current state, ML algorithms are applied at online portals like carwala.com other websites which shows recommendations for the selected category of car with the best price and offers. However, In Dealer outlet, ML techniques are not applied to improve the sales. In the dealer outlet, only tele sales are done, that does not boost the sales conversion much. By applying ML techniques, dealer sales can be improved.

INTRODUCTION
India is one of the world's largest marketplaces for the automotive sector. Although it formerly had one of the fastest growth rates in the world, it now has flat or even declining growth rates. The manufacturing of passenger and commercial vehicles in India is the sixth largest in the world, producing more than 3.9 million units annually in 2011. Any person today who has climbed the ladder will choose to buy a car first of his earnings. Car ownership is no longer considered a luxury. Its demand is given the proper priority. Increased foreign investment in India because of FDI policy liberalization changed the market's appetite for automobiles in India. Customers today take engine performance into account in addition to other factors. They look for the parameters that set a product apart from others so they can choose one brand over another. According to Financial Times, India is now the fifth largest market for passenger vehicles, following China, the United States, Japan, and Germany. The offer of traveler vehicles also became 9.24% to 3.04 million during 2016-17, the quickest development rate seen beginning around 2010-11, when they became 28.2%. With the powerful worldwide standards as respects to both business and natural security, the vehicle business must stay up with the evolving situations. Automated and electronically controlled automobiles are on the rise. Ambient intelligence systems are to blame for this, and automakers must take this into account when designing new vehicles. Automobile manufacturers must also keep a lot of data about consumer

preferences and incorporate the results of previous marketing research because of the growing level of competition. They will undoubtedly benefit from this massive data management in determining consumer preference trends and developing future strategies. The purpose of this paper is to investigate the rationale behind consumer preferences regarding car purchasing behavior. The paper mainly focuses on the car buying behavior of Indian customer based on the attributes such as income, price, etc. We have future cleaned the data and removed to missing variables. We tried to analyze which customer prefers to buy which car considering these attributes.

LITERATURE REVIEW

A.K. Tiwari's Manish Kumar Srivastava research customer habits for A3 class cars like the Honda City and SX4 in the city of Jaipur. 100 individuals were surveyed, 50 of whom drove Honda City and Maruti SX4 vehicles. Respondents from a range of backgrounds, including gender, occupation, and income class, were considered. Price, Safety, Comfort, Power & Pickup, Mileage, Max Speed, Styling, After Sales Service, Brand Name, and Spare Parts Cost are other client buying criteria considered for the study. Based on what was just spoken about a certain brand of automobile and the elements that could influence a potential purchase. The notion of gauging customer satisfaction will similarly serve the function of establishing consumer perception. Consequently, by gauging a car's enthusiastic consumers' propensity to share it.

Rahul Singh and Shiny Raizada (2020) examine the consumer behavior, thinking of those who have purchased a mid-range SUV in the last 10 months before the pandemic period or are planning to do so in the next 10 months during the pandemic period. , and usage patterns. Buy panda time. The research focuses on four main factors: media consumption, unassisted recall, brand preference, and price perception and decision making.

Tejaswi Vellampalli (2017) investigated the factors that influence the repurchase behavior of Hyundai Motor customers to understand satisfaction and its influencing factors. The authors examined variables that influence post-purchase behavior of car buyers. Kusuma P (2015) conducted research in Karnataka to identify factors that influence consumer purchasing behavior among car owners. The researcher also developed a theoretical model that influences consumer purchasing behavior for passenger cars, allowing further research to be conducted based on the theoretical model he developed.

Arpita Srivastava and Mitu Matta (2010) investigated consumer behavior towards passenger cars in Delhi NCR. This study focuses on important behavioral aspects such as information retrieval and evaluation, brand preference and brand loyalty, and motivational factors. The study concludes that automakers need to understand what drives customer satisfaction when developing products that meet customer needs and designing marketing programs and strategies.

Sangeeta Gupta (2013) investigated the role of reference groups in influencing the buying behavior of car owners in the city of New Delhi. The results showed that there was a strong correlation between the reference group's attitudes toward attributes such as fuel economy, mileage, and price in making purchase decisions and the target market's purchase decisions.

Brown et al. (2010) analyzed consumer attitudes toward US, European, and Japanese automobiles. This study shows that country of origin plays an important role in consumer behavior. Brand image, reasonable prices, and dealer reputation have a significant impact on passenger car sales.

METHODOLOGY

Methodology means the way the sample and sample size are chosen is for data collection. As this is the backbone of research, various tools are used to study a particular object (or) problems with objects. The study design conducted in this study is exploratory in nature.

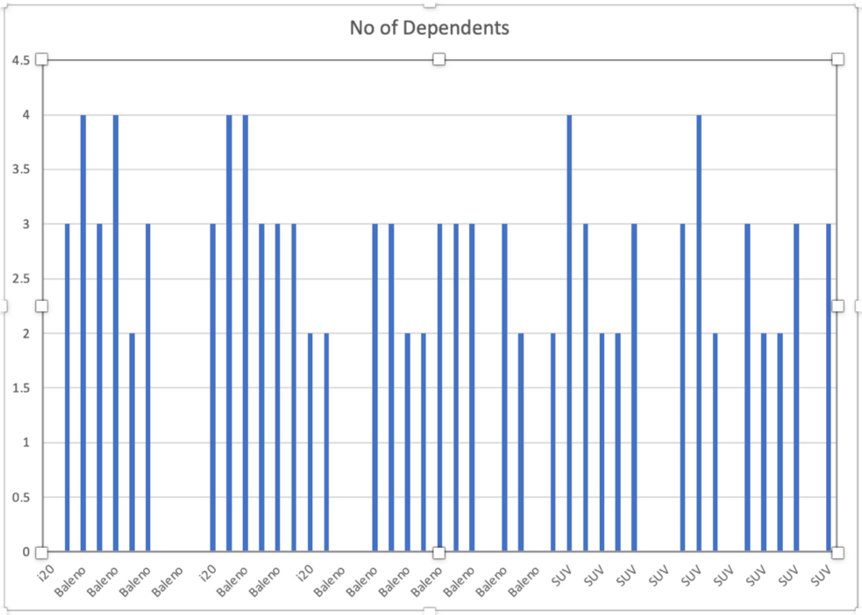
We have used secondary data in this paper. We have taken a data set from Kaggle.com. We took a dataset that contains various attributes considered by Indian customers' while buying a car thinking about it. We have then changed and modified the attributes accordingly.

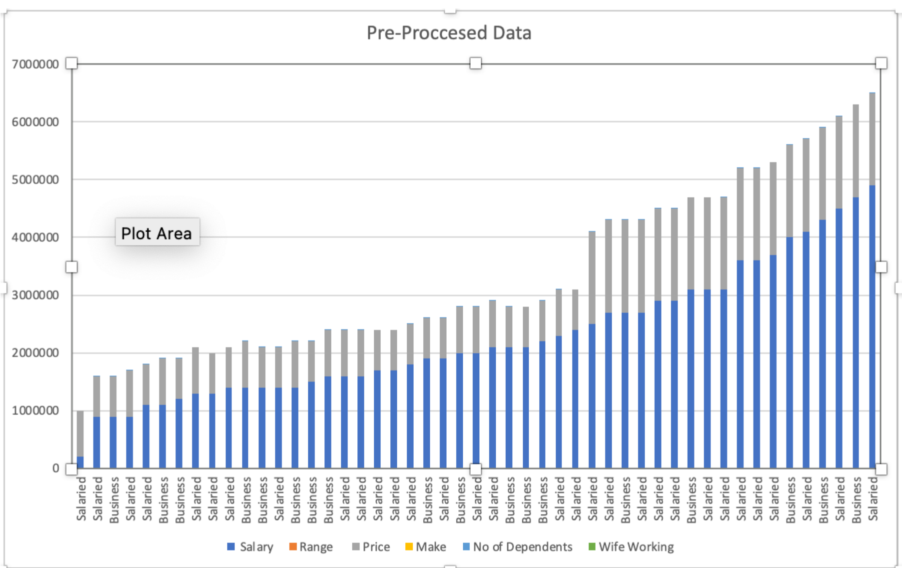
DATA PRE-PROCCESING

The rata data contains eleven (11) attributes such as age, profession, marital status, education, no of dependents, personal loan, house loan, wife working status, wife salary, total salary, make and price. The first step in pre-processing is to be filtering the important attributes and selecting the main ones for running the analysis. We had selected the following attributes

1. Profession – Salaried/ Business
2. Range
3. Price of the car
4. No of dependents – 0/2/3/4
5. Wife working – Yes/No
6. Make i.e., model of the car.

However, since the salary had too many variations and it wouldn’t give us accurate pattern on our data, we added a new attribute ‘range’. We divided the salary into “Low, medium and high” range.
Low: (200000 – 1800000)
Medium: (1900000 – 2700000)
High: 2900000 – 4900000).
The data also contains many models of car such as i20, Baleno, Ciaz, Duster, City, SUV, Creta, Verna and Luxuray. Keeping in mind the occurrence of each car and in accordance with the salary ranges we have chosen 3 cars for the analysis – i20, Baleno and SUV. In the raw data set i20 is bought by 11 customers, Baleno and SUV by 18 customers.





The data Preprocessing is done in manually and starts with removing the irrelevant letters, punctuation marks etc. It contains unnecessary words such as ‘m’, ‘s’ etc. in various fields which do not add any value to the analysis have been removed and only necessary ones were kept.

Format of the data set: Our data set was in .xls format. We had first converted the data in .csv format and used created a notepad code to change it into .arff format so that it could be used in Weka.

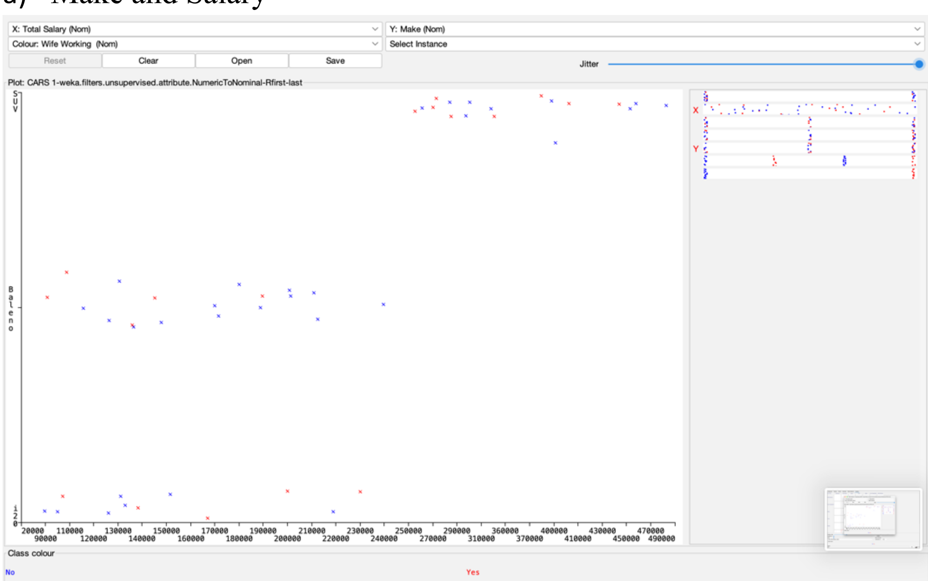
```
@relation cars

@attribute profession {Salaried   Business}
@attribute price { 700000    800000    1600000}
@attribute make {i20    Baleno    SUV}
@attribute dependents {0    2    3    4 }
@attribute wife working { Yes    No}


@data
Salaried   Low    800000   i20      0   No
Salaried   Low    700000   Baleno   3   No
Business   Low    700000   Baleno   4   Yes
Salaried  Low    800000   i20      3   No
Salaried  Low    700000   Baleno   4   Yes
Business  Low    800000   i20      2   Yes
Business  Low    700000   Baleno   3   No
Salaried  Low    800000   i20      0   No
Salaried   Low    700000   Baleno   0   No
Salaried   Low    700000   Baleno   0   No
Business   Low    800000   i20      3   No
Business   Low    700000   Baleno   4   Yes
Salaried   Low    700000   Baleno   4   Yes
Business   Low    800000   i20      3   No
Business   Low    700000   Baleno   3   No
Business   Low    800000   i20      3   No
Salaried   Low    800000   i20      2   Yes
Salaried   Low    800000   i20      2   Yes
Salaried   Low    700000   Baleno   0   No
Salaried   Low    700000   Baleno   0   No
Salaried   Low    700000   Baleno   3   No
Business   Medium 700000   Baleno   3   No
Salaried   Medium 700000   Baleno   2   Yes
Business   Medium 800000   i20      2   Yes
Salaried   Medium 800000   Baleno   3   No
Salaried   Medium 800000   i20      3   No
Business   Medium 700000   Baleno   3   No
Business   Medium 700000   Baleno   0   No
Business   Medium 700000   Baleno   3   No
Salaried   Medium 800000   i20      2   Yes
Salaried   Medium 700000   Baleno   0   No
Salaried   Medium 1600000   SUV     2   Yes
Salaried   Medium 1600000   SUV     4   Yes
Business   Medium 1600000   SUV     3   No
Salaried   Medium 1600000   SUV     2   Yes
Salaried   High    1600000   SUV    2   Yes
Salaried   High    1600000   SUV    3   No
Business   High    1600000   SUV    0   No
Salaried   High    1600000   SUV    0   No
Salaried   High    1600000   SUV    3   No
Salaried   High    1600000   SUV    4   Yes
Salaried   High    1600000   SUV    2   Yes
Salaried   High    1600000   SUV    0   No
Business   High    1600000   SUV    3   No
Salaried   High    1600000   SUV    2   Yes
Business   High    1600000   SUV    2   Yes
Salaried   High    1600000   SUV    3   No
Business   High    1600000   SUV    0   No
Salaried   High    1600000   SUV    3   No
```

The above image is saved as “Cars.arff” to be converted into that format and run in Weka.

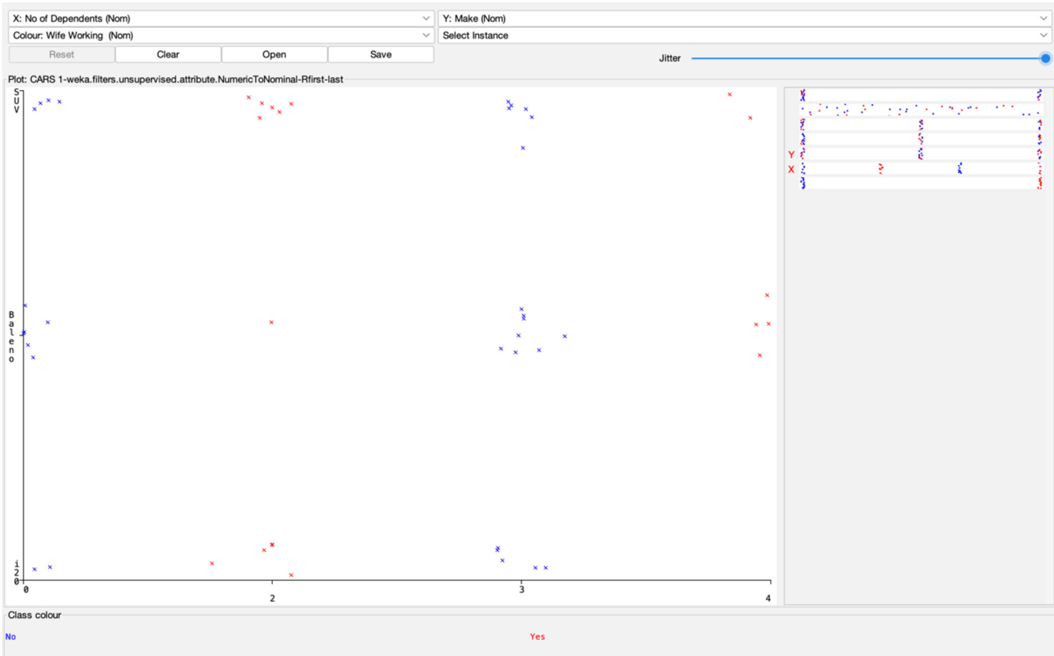
Numerical to Nominal: From the above selected attributes, few attributes such as range, salary were in numerical format. We had load the data set in Weka and converted them into nominal by using the filters option.

Visualization

a) Make and Salary



b) Make and No of Dependents



DATA MINING AND RESULTS

Association Rule Mining

The processed data mining set contains only the relevant fields and attributes. Our analysis is to find out what Indian customers consider/take in account while buying a car and discover patterns on their buying behavior. We are using Association rule mining to find out patterns in the data set. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.[1] In any given transaction with a variety of items, association rules are meant to discover the rules that determine how or why certain items are connected. Association rules are "if-then" statements, that help to show the probability of relationships between data items, within large data sets in various types of databases.

For example: If a customer gets salary in high range, then most likely he is to purchase an SUV rather than a Baleno. We have performed the association rule mining in apriori algorithm analysis with different parameters, following below.

- Support: The frequency of the items to occur together.
- Confidence: That the rule is true
- Lift: Confidence/Expected confidence

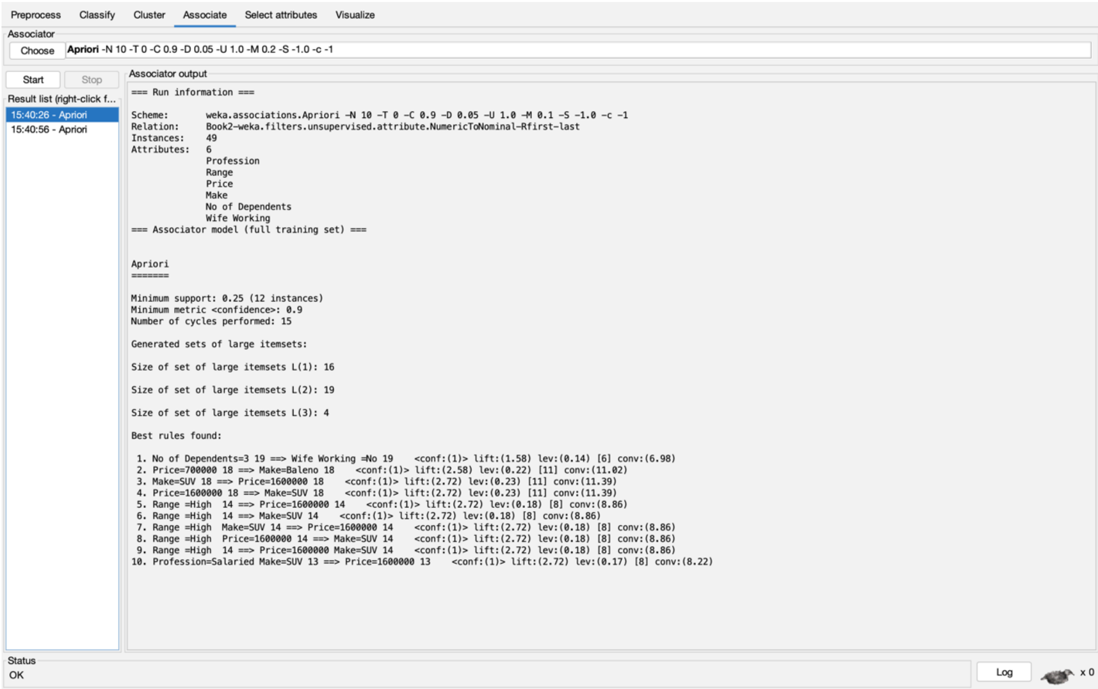
Larger the lift, the stronger the rule. We have adjusted the parameters of support and confidence and tried finding out patterns for the same.

a) The default options of apriori

   Support : 0.1

   Confidence 0.9

When we run the analysis, the following rules are generated.



The strongest rules generated and making a clear pattern are as follows:

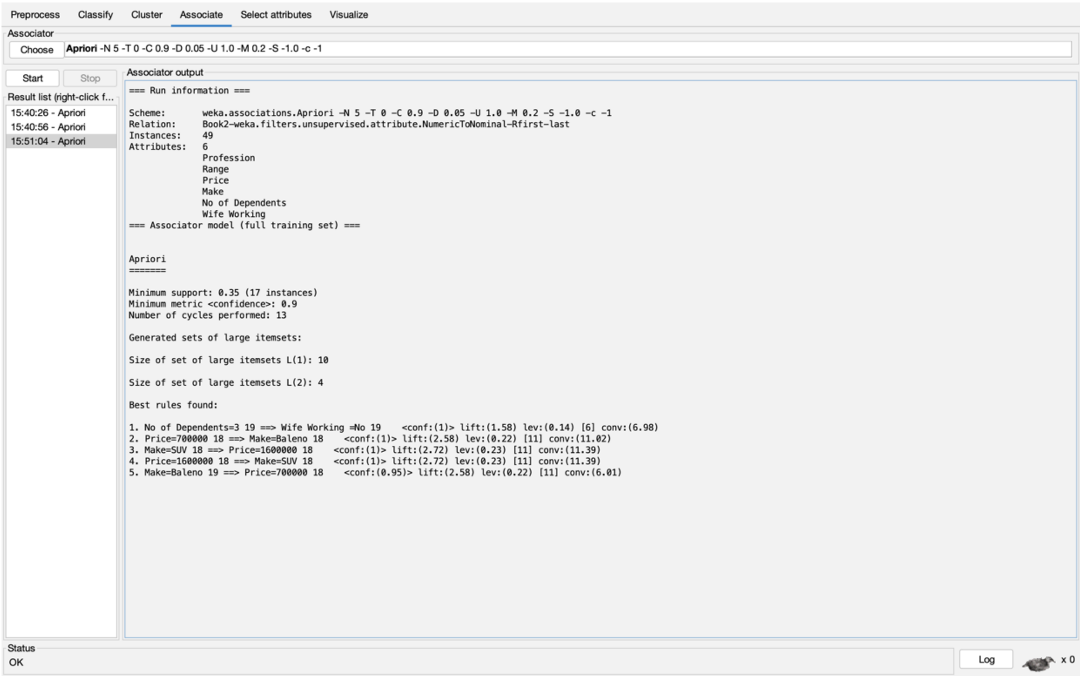
1. For every customer who has e dependents in the house, the status of wife working is No, probably to take care of the household.
2. All the customers' whose salary range is high, are most likely to buy SUV’s.
3. The customers whose profession is salaried are likely to buy SUV’s too.

Here in the above rules, the confidence is ≥ 1 and the lift is higher than 1.

b) Support: 0.2

Confidence:0.9

No of rules: 5



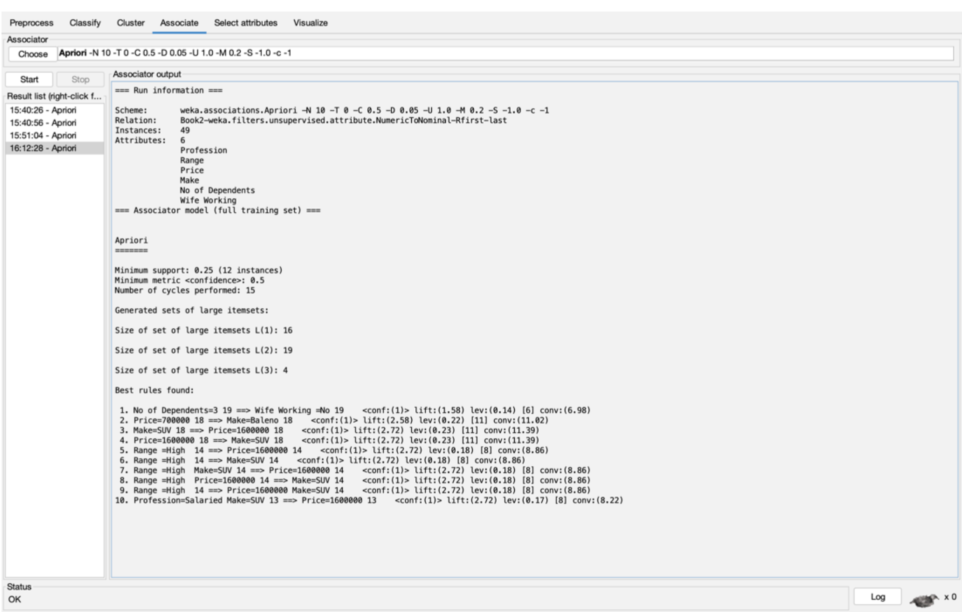
The rules generated here show us the price and make of the car and how they go hand in hand. Like for example Baleno comes in the medium range of salary and SUV in the high.

However, these above rules are very strong.

c) Support : 0.2

Confidence:0.5

Rules generated:10

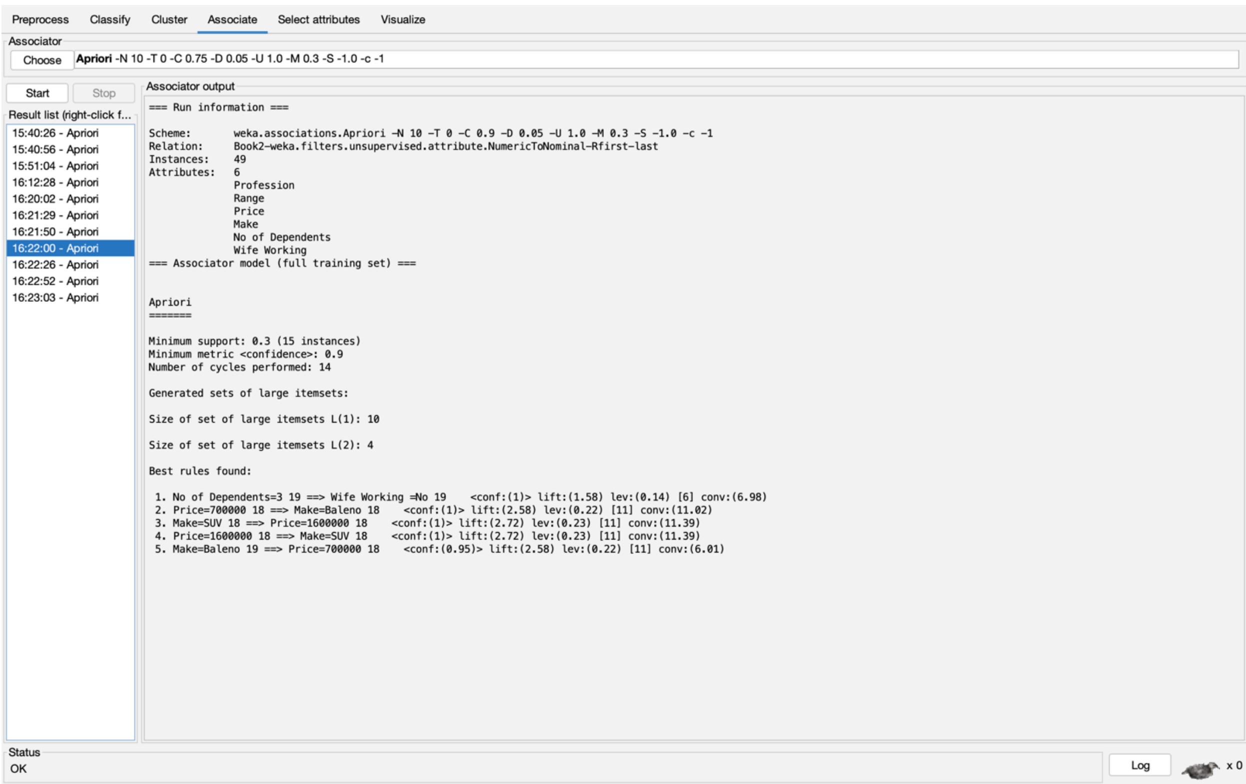


The rules generated indicate that

1. Customers with higher no of dependents and wife not being working tend to avoid buying a car.
2. Customers' who are in the higher range of the salary generally earn more than 170000 as they are opting to buy a SUV.

These rules are strong and are proved to be right.

d) Support: 0.3
   Confidence:0.75
   No of Rules: 10



With the above parameters, it is observed that

1. Households with less number of dependents and wife working as yes tend to buy cars more frequently.
2. The price of the car Baleno suggest that it can be affordable by customers' salary ranging in low and medium.

The rules above are strong and show confidence being 1 and lift greater than 1.

Results from all the above interpreted data are:

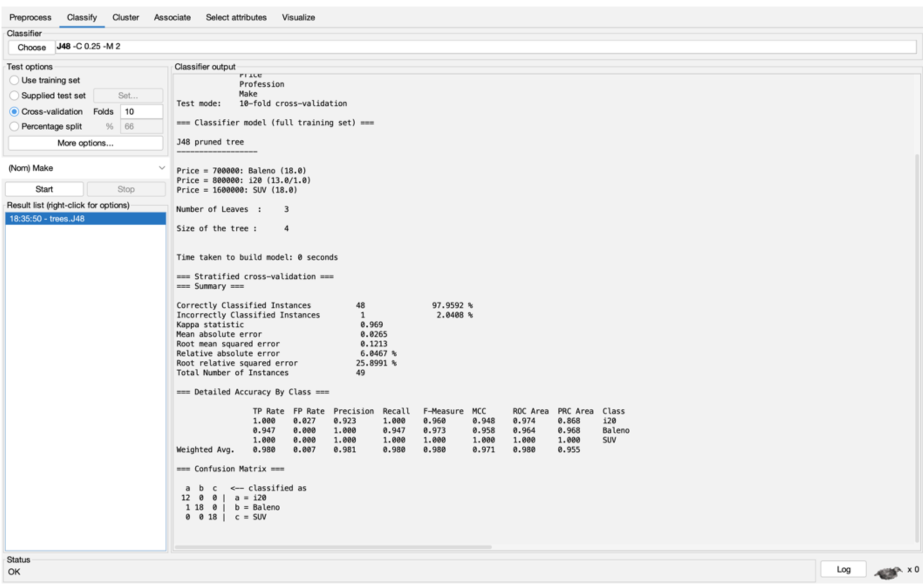
1. Households with no of dependents high are mostly the ones where the wife is not working.
2. Customers whose salary is ranging in ‘high’ are mostly likely to buy SUV’s.
3. Customers whose profession is both salaried and business can afford an SUV.

4. Customers' with 0/1 dependents and wife working tend to buy cars more frequently
5. The high salary range is greater than 170,000
6. Customers with higher no of dependents and wife not being working tend to avoid buying a car.
7. Professionals whose salary ranges in low and medium can buy a Baleno.

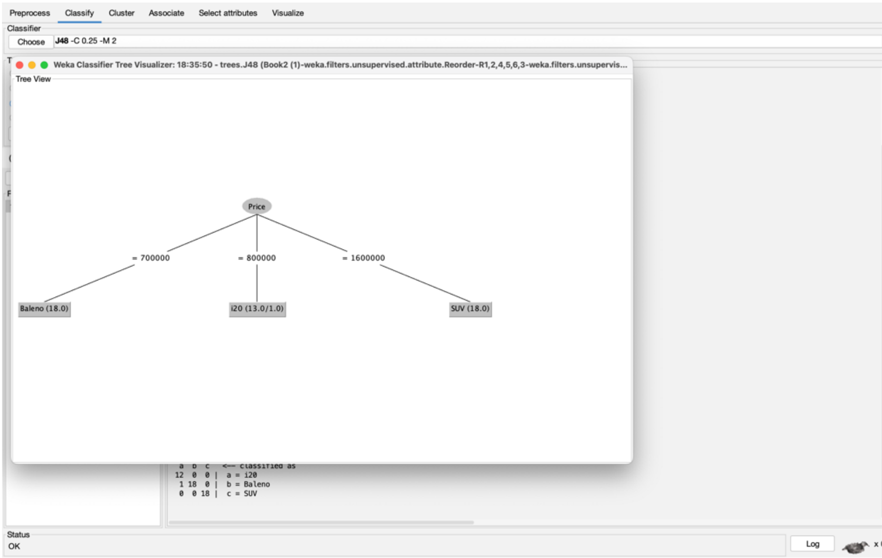
Classification

It is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. We have used the classification model to predict the target class/attribute in the data set so that it can help target customers' in a better way.

a) We are using the J-48 tree model taking our class as: Make

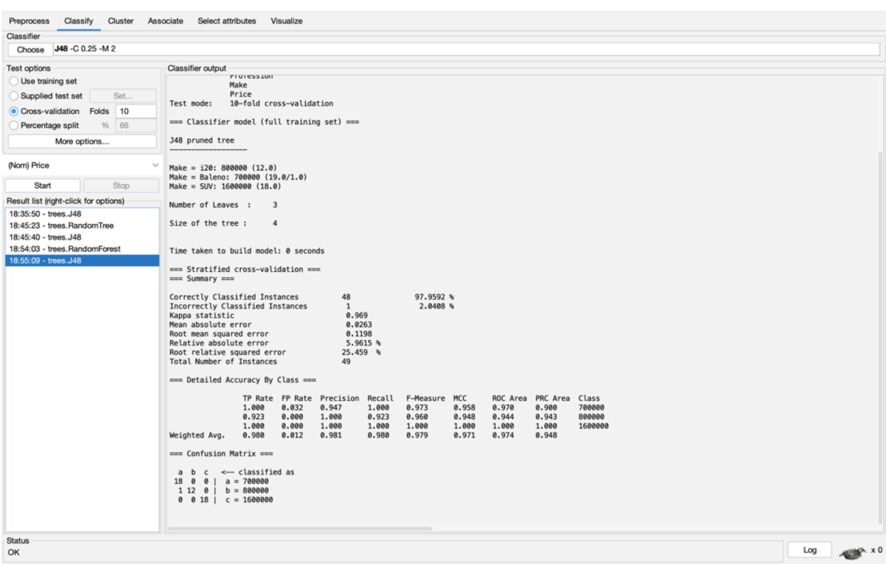


Visualizing the tree:

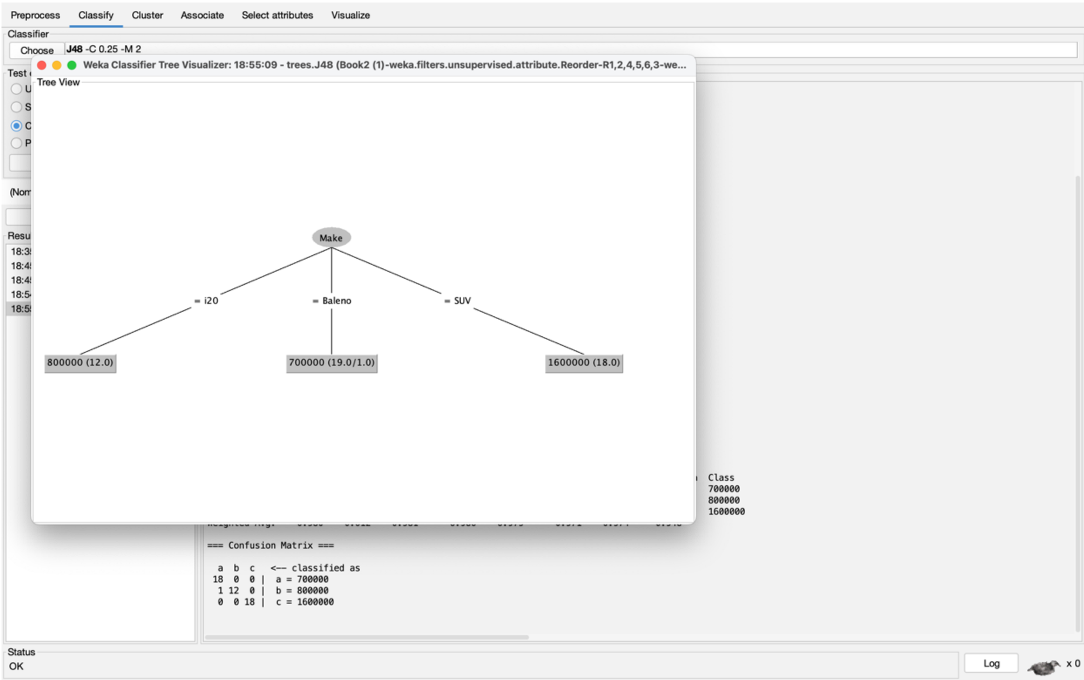


Interpreting the data above, we have found that the correctly classified instances comes up to 97.95% and incorrect instances to 2.04%. When we use the J-48 tree the confusion matrix clearly shows that only one instance was gone wrong, where it was supposed to be Baleno but classified as i20. We can interpret by saying that when used make as our class the most affected attribute is the price of the car. The visualization tree makes it clear and concludes the same.

b) We are using the J-48 tree model taking our class as price



Visualizing the tree:



Interpreting the data above, we have found that the correctly classified instances come up to 97.95% and incorrect instances to 2.04%. When we use the J-48 tree the confusion matrix clearly shows that only one instance was gone wrong. We can interpret by saying that when used price as our class the most affected attribute is the make of the car. The visualization tree makes it clear and concludes the same.

Results:

We have seen that taking make or price as our class value, it gives us mostly correct instances. The attributes that needed to target while buying a car is make/price highly. Customers are likely to first choose the model and then consider other factors.

CONCLUSION

In this analysis we have studied the buying behavior of Indian customers in cars. We used a data set with different attributes to analyze which attributes are considered while a customer decides to purchase car or while considering it. We also tried to find patterns and the most targeted attributes to understand the customers thought process better. The main goal for doing this study was to increase the sales of cars in dealer outlets. Based on our findings and results the company make plan it's marketing strategies better, find a target group and understand what kind of cars need to be advertised to which kind of customers. For example: With the above results it's clear that if the company is trying to promote an SUV to a customer with high No of dependents then the chance of sales are close to null.

However, we can conclude that the most effecting attributes while buying car for and Indian customer are price, make, salary and no of dependents in the household.

REFERENCES

https://www.kaggle.com/datasets/karivedha/indian-consumers-cars-purchasing-behaviour
file:///Users/apple/Downloads/AN+ANALYSIS+ON+NEW+NORMAL+CONSUMER+BEHAVIOUR+FOR+BUYING+CARS+IN+INDIAN+AUTOMOTIVE+INDUSTRIES+IN%255B1968%255D%20(1).pdf
https://www.researchgate.net/publication/327070602_Consumer_Buying_Behaviour_of_Cars_in_India_-_A_Survey
https://www.ripublication.com/gjfm-spl/gjfmv6n6_16.pdf
https://katba-caroline.com/automobile-options-association-rule-mining-data-mining/
https://medium.com/edureka/apriori-algorithm-d7cc648d4f1e
https://www.softwaretestinghelp.com/data-mining-techniques/#1_Frequent_Pattern_MiningAssociation_Analysis
https://www.softwaretestinghelp.com/data-mining-techniques/#1_Frequent_Pattern_MiningAssociation_Analysis