

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 4, April 2019, pg.177 – 181

The Data Mining Model that Predicts the Customer Decision-Making in Buying a Car

Kaveri Kar¹; Rashmi Patel²

Department Of Computer Science and Engineering^{1,2}

Shri Shankaracharya Institute of Professional Management and Technology

¹k.kar@ssipmt.com; ²rashmi.patel@ssipmt.com

Abstract: For the last few of decades, India has become one of the largest country in the world in the adaptation and establishment of the automobile industry which growing steadily each year. That is why the decision-making and behavior of a customer become a point of interest. The decision of buying a car constitutes a unique type of buying behavior. This is something different in comparison to the buying or shopping of the essential needs of the life. The main objective to this prediction is to advance the knowledge of what are the factors that influences the customer's choices of the car stores. More specifically, the main focus has been on how different situations (e.g. type of buying) influence the choice of the car stores. The frequent item sets are mined and associated from the market basket database using the efficient algorithm and hence the association rules are generated. The decision tree can be constructed using Linear Regression algorithm and Random Forest Classification algorithm. Then at last the prediction made by the both linear regression algorithm and random forest classification algorithm is compared for the best results.

Keywords: Customer behavior, Market Analysis, Decision Making, Linear Regression, Association rules, Random Forest Classification, Comparison.

Introduction:

The automobile industry of India is one of the largest in the world, contributing 7.1% to the Gross Domestic Product (GDP). Nowadays it is becoming a necessity of a person to own a car than a luxury.

One of the challenges for the customer that have heavily invested in decision making is how to extract the important information from the market databases and car features databases, in order to gain suitable knowledge about the car that will help the customer in decision making for buying the car.

Association rules are derived as threshold levels from the frequent item sets using support and confidence. Frequent Item set is the sets of items which have minimum support are known as. The proportion of transactions in the data set which contain the item set is called the support count of an item set.

Literature Review:

System Architecture:

A. Customer Dataset:

The customer dataset is provided by the Kaggle Machine Learning Competitions Platform for the analysis of this work. The dataset has 12 continuous and 1 numeric input attributes namely age, profession, marital status, education, no. of dependents, personal loan, house loan, wife working, salary, wife salary, total salary, car type, price of the car and buying type whether the car is affordable, economical or expensive. The linear regression predicts the price of the car and random forest classification predicts the buying type of car whether-affordable, economical or expensive.

B. Association Rules:

If X then Y, then Association rules are of the form. Patterns that appear in a data set frequently is referred to as frequent patterns. In a given data set, frequent pattern mining searches for recurring relationships. It is not necessary that the association rules are not always useful, even if they have high support, confidence. Items often fall into natural hierarchies

C. Regression Analysis:

Regression analysis is a technique used in statistics for investigating and modeling the relationship between variables.

There are different types of regression methods are available such as- Linear Regression, Random Forest Regressor, XGBoost Regression etc. The best prediction is determined by using the linear regression in the data mining model that predicts the customer decision making in the buying of a car. By using the linear regression technique on the customer-car dataset, the predictive model provides the outcome as in the form of an amount such as-Rs.12, 00,000 etc.

D. Classification Analysis:

Classification analysis is used to map data sets into predefined groups and classes. As the classes are determined before examining the datasets, it is considered to be the supervised learning.

There are various classification algorithm used for the implementation of the classification analysis such as- Logistic Regression, RandomForest classifiers, XGBClassifier, Decision tree classifier etc. I have used random forest classifier algorithm.

The random forest classifier classifies the customer into three groups as- customer can buy the affordable, economical or expensive car on the basis of the characteristics mentioned in the datasets.

Simulation Results:

The whole dataset was given to the data mining tool like regression analysis or classification analysis. Then frequent item set is found using appropriate algorithm in the association technique. This paper is mainly focused to find out that which predictive model generates higher accuracy and prediction using the customer-car dataset. For this, it uses regression and classification algorithm. The final result will be the comparison between the accuracy of the predictive model generated for the dataset.

Table1: Statistical analysis of wholesale customer data

	Age	No. of dependents	Salary	Wife Salary	Total Salary	Price
count	99.000000	99.000000	9.900000e+01	9.900000e+01	9.900000e+01	9.900000e+01
mean	36.313131	2.181818	1.736364e+06	5.343434e+05	2.270707e+06	1.204040e+06
std	6.246054	1.335265	6.736217e+05	6.054450e+05	1.050777e+06	4.237633e+05
min	26.000000	0.000000	2.000000e+05	0.000000e+00	2.000000e+05	7.000000e+05
25%	31.000000	2.000000	1.300000e+06	0.000000e+00	1.550000e+06	8.000000e+05
50%	36.000000	2.000000	1.600000e+06	5.000000e+05	2.100000e+06	1.200000e+06
75%	41.000000	3.000000	2.200000e+06	9.000000e+05	2.700000e+06	1.500000e+06

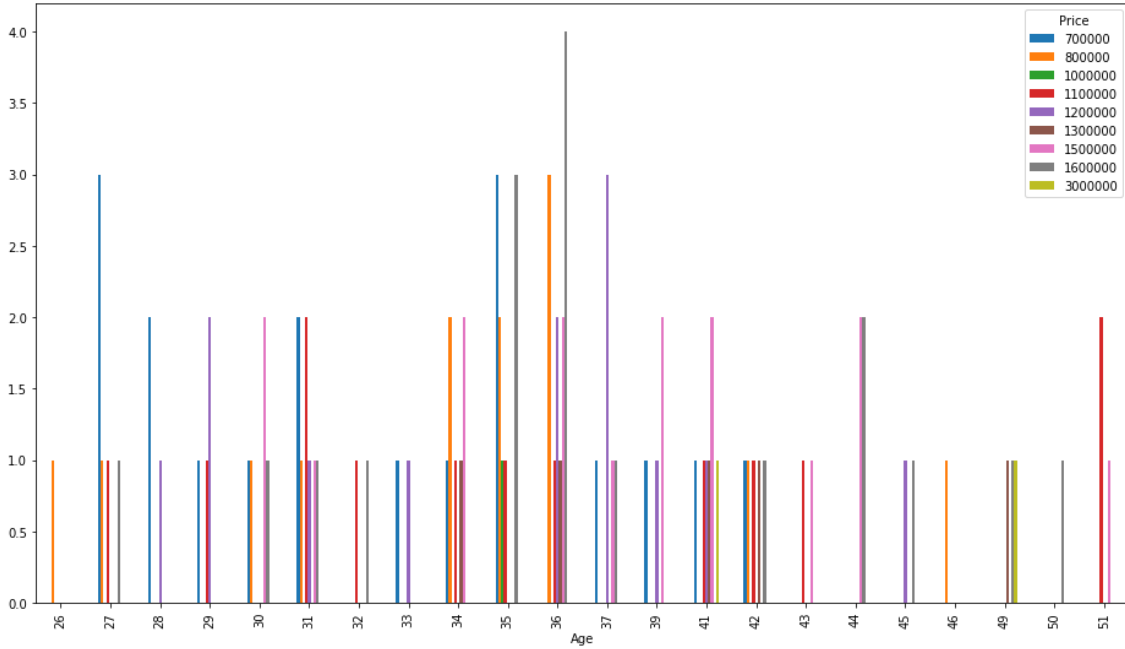


Fig: Graph for customer-car dataset after regression

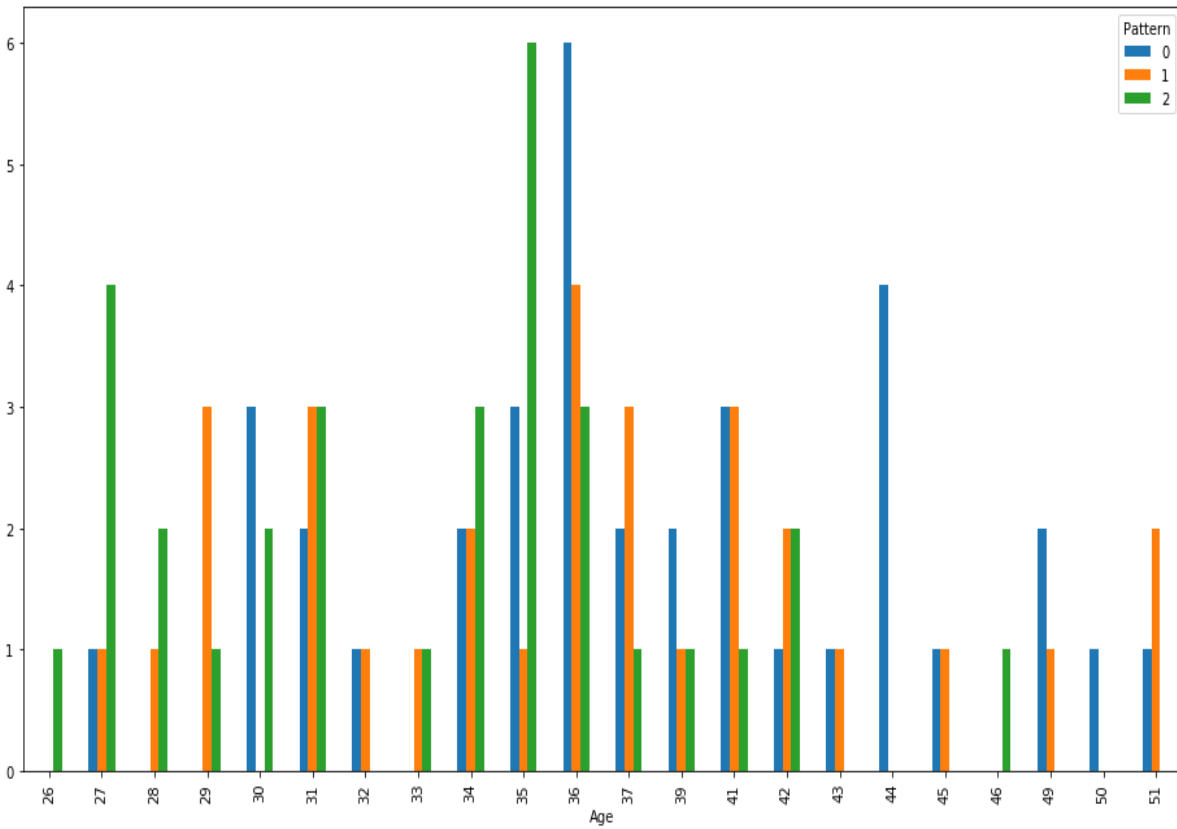


Fig: Graph for customer-car dataset after classification

CONCLUSION:

In this paper, a comparison for Regression and classification is shown on the frequent item set is developed for the analysis of customer-car data. The customer-car dataset is taken and analyzed to know which analysis will be better for prediction. The data in the dataset is preprocessed to make it suitable for regression and classification as per the requirement basis. The preprocessed data is used for regression and classification and tried to obtain high regression accuracy as well as the classification accuracy. The classification technique is considered to be better than the regression analysis because it simplifies the complexity in analysis and makes it easier in making the decision by the customer.

REFERENCES:

1. Nikhil Monga, Saurabh Tripathi, Bhuvnender Chaudhary, Car market and buying behavior-a study of consumer perception, . International Journal of Research in Management, Economics and Commerce, Volume2, Issue 2(February 2012), ISSN: 2250-057X
2. Dhruv Mathur, Avdesh Bhardawaj , Aditi Pandey , Ashish Oberoi , Divya Rani, Consumer Buying Behaviour of Cars in India - A Survey, NFEST/2018/R-115
3. Monga N, Chaudhary B, Tripathi S. Car market and buying behavior: A study of consumer perception. International Journal of Research in Management, Economics and Commerce. 2012;2(2):44-63.
4. Modi S, Jhulka T. Consumer buying behaviour: Changing shopping patterns. Int. J. Bus. Manag. Eco. Res. 2012; 3(3):527-530.
5. Kotler et al. (2004) Chapter 7, pp. 242-244, 'Consumer behaviour' and 'A model of consumer behaviour'.
6.) David L. Albert J. Della Bitta, " Consumer Behavior " Fourth edition, Tata McGraw Hill, pp - 22 -25
7. Vikram Shende, Analysis of Research in Consumer Behavior of Automobile Passenger Car Customer, International Journal of Scientific and Research Publications, Volume 4, Issue 2, February 2014 1 ISSN 2250-3153
8. M.Prasanna Mohan Raj, Jishnu Sasikumar, S.Sriram , A Study of Customers Brand Preference in SUVs and MUVs: Effect on Marketing Mix Variables, International Referred Research Journal Vol.- IV, Issue-1, pp. 48-58, Jan 2013.
9. Samin Rezvani, Goodarz Javadian Dehkordi, Muhammad Sabbir Rahman, A Conceptual study on the country of origin effect on consumer purchase intention, Asian Social Science Vol. 8, No.12, 2012 pp. 205-215.