Maximizing Profits through Customer Segmentation and by finding Top-K Profitable Products for the Automobile Industry

Prateek V. Jain
Research Scholar
VIIT, University of Pune
Maharashtra, India

Sahil Patani
Research Scholar
VIIT, University of Pune
Maharashtra, India

Sumesh Kadam
Research Scholar
VIIT, University of Pune
Maharashtra, India

Swati Patil
Assistant Professor
VIIT, University of Pune
Maharashtra, India

Mahesh Mule
Research Scholar
VIIT, University of Pune
Maharashtra, India

ABSTRACT

Automobile Industry is facing numerous challenges like plunging automobile sale, rising inventory, lowering margins and unsatisfied customers among others. Wrong understanding of customer needs and demands is badly affecting product designs and global ambitions of the auto industry. The auto industry needs an IT solution that will help them in understanding customer's needs in a better way. Hence, the need for adequate performance measures has become more important than ever. In this paper problems like Customer Segmentation and finding top-k Cars are addressed. In Customer Classification Prediction Model, based on the customer's profile and preferences a probabilistic result is provided. This result will predict a set of cars that the customer should most likely purchase. The advantage of this approach is that it assists companies with selecting the classifier which maximizes the profit. The other problem is finding Top-k Profitable Products. The products are compared to other products of different brands. Then a set of features which could be added to the products such that they are not dominated by the products in the existing market are found. In this problem, the prices of these products are set such that the total profit is maximized. Such products are referred as top-k profitable products. The algorithm will also calculate the profits generated by various segments of cars. Precise solutions to find the top-k profitable products efficiently are proposed. An extensive performance study using both synthetic and real datasets is reported to verify its effectiveness and efficiency.

Keywords
Customer classification, Top-k, prediction, dominance.

1. INTRODUCTION
The main aim of companies in the automobile industry is to earn customer's trust towards their brand by providing quality service. In automobile industry, Quality of Service plays an important role in maintaining a customer's loyalty towards the company. Customers buy the automobile by giving preferences to the brand name irrespective of the other features that are provided. Automobile companies always prefer to have dealers that generate huge amount of sales. Nowadays, automobile companies acknowledge the importance of customer relationship, hence they are trying to maintain a good relationship with them. Automobile companies and their dealers are more interested in their profit margins. Automobile companies are now operating in a really stiff competition which has compelled them to deliver the best services and its products at very reasonable prices [3]. Due to globalization, this competition has also become stiffer among the companies. Now the success or survival of any organization is only dependent on the satisfaction of the customers which is expected from the product or service offered [4]. Organizations are giving much more attention on their customer satisfaction which will help them in retention of the existing customers. The retention of existing customers and attracting new customers will show how effectively an organization can manages its customers [5].

The application will allow Automobile dealers to offer a set of probabilistic cars according to customer's profile and preferences. Also it will assist the company in maximizing profits by finding a set of features that should be added to the product in order to increase its sales in a competitive market. The application will provide a pictorial representation of the profits generated by various segments of cars of a particular brand. Overall the product provides an Automobile company not only with ideas to expand their hold over the market but also helps them to decide the future strategies such as what facilities to be offered to the customer which would eventually build a strong market place for the company.

2. CUSTOMER CLASSIFICATION PREDICTION MODEL

2.1 Customer classification
In Customer classification prediction model, generation of an accurate probabilistic result based on the customer's profile and preferences of buying a car takes place. The automobile dataset has been taken from UCI machine learning repository. The customer dataset has been provided by Achilles Solutions, a private IT firm based in Pune, India.

A train and a test dataset has been used to implement this algorithm. The train dataset provides a generic behavior about the values to be predicted. The test dataset is used to measure the strength of a particular relationship based on the behavior provided by the train dataset.

The dataset has a list of independent variables and a set of dependent variables. By applying a decision tree such as J48 would predict the target variable of a new dataset record. A decision tree is a predictive machine-learning model that predicts the target value of a new sample based on various attribute values of the available data [6]. The interior nodes of a decision tree denote the different aspect, the branches between the nodes tell us the possible values that these aspects can have in the observed samples, while the terminal nodes tell us the final value of the dependent variable.

The attribute that is to be predicted is known as the dependent variable, since its value depends upon the values of all the other attributes. These another attributes, which help in predicting the value of the dependent variable, are known as the independent variables in the dataset [7].

The J48 Decision tree classifier follows the simple algorithm. In order to classify a new item, it first needs to create a decision tree based on the attribute values of the available training data [8]. So, whenever it encounters a training set it identifies the attribute that discriminates the various instances most clearly [8]. This element is able to tell us most about the data instances so that to classify them the best is said to have the major information gain. Now, among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, so to terminate that branch and assign to it the target value that has been obtained.

For the other cases, observe another attribute that gives us the highest information gain. Hence a continuation in this manner will either provide a clear decision of what combination of attributes gives a particular target value, or the attributes will be precisely over. In the event of attributes getting over, or if an unambiguous result from the available information is not attained, then assign this branch a target value that the majority of the items under this branch possess.
With the decision tree, the order of attribute selection as for the tree is followed. By checking all the respective attributes and their values with those seen in the decision tree model so as, to assign or predict the target value of this new instance.

2.2 Algorithm
Algorithm-Compute Prediction

Input- RTD(Real Time Data) & UP(User Preferences).

Output- Classification of the Customers

Accept RTD.

If Train set not Available then

Randomly pick 1/10th of RTD.

Insert into Train Data Set

Train this data set according to users need.

Give this Train data to Classifier as Input File.

Construction of Data Set.

A linear set of R linear discriminant functions that are used collectively to assign an instance to one of the R classes. Here p is an instance description that consists 1 and the n features that describe the instance. Discriminant function gi(p) for multivariate has the form as-

W$\mathbf{i}^{\Sigma}$RTD

Where is W$\mathbf{i}$ are instance of $n + 1$ coefficients .The sates that instance p will belong to class i.

(Wi, $i \neq j$)= g i(p) >g j(p)

Attribute(RTD(p| j)=|P $_J$ |log |P|

Gain(RTD, j)=(RTD-(Attribute(RTD(p| j)))

Here gain is calculated through evaluation of the multiple attributes array through which the set g of multi valued attributes will be compared and the result with greatest value will be inserted. Attribute value after logarithmic function will be added to whole set and gain is obtained.

Algorithm for Prediction-

UP := empty set of rules

while not j is empty

UP := best single rule of j

while not g is empty

RTD := add g to R

remove those instances from j that are covered by g.

2.3 Results

Model No.	Car Name	Price
1	Alto 800	400000
2	WagonR Vxi	500000
3	i 10	650000
4	Micra	600000
5	i 20	800000
6	Swift Grand	600000
7	Polo	700000

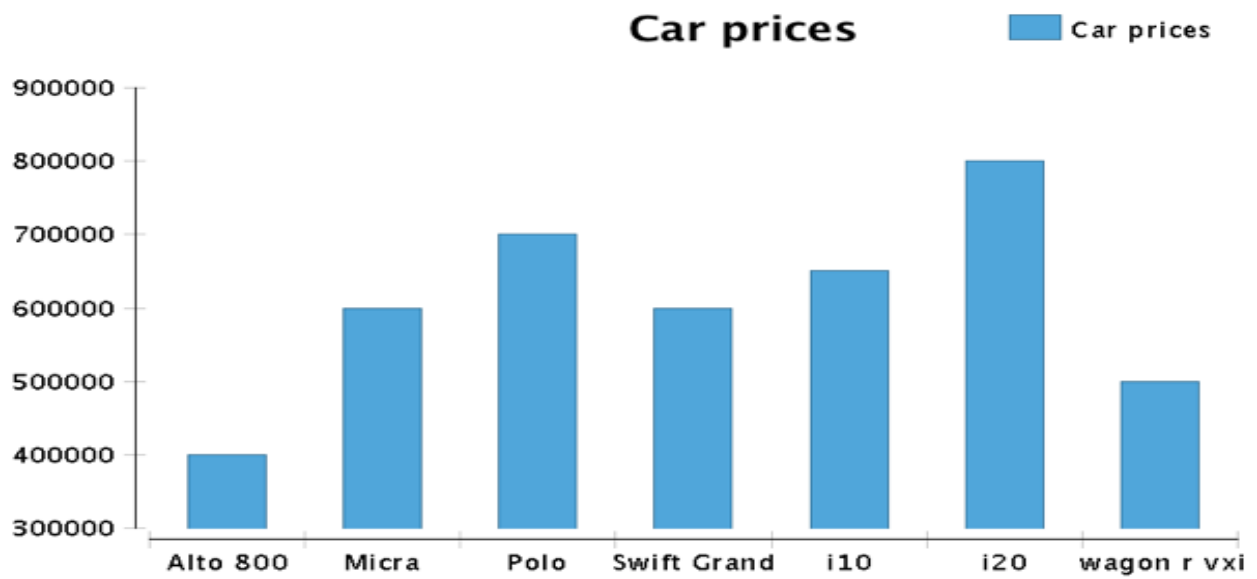**Fig 1:** A set of cars have been segmented according to their price. A bar chart represents the desired results.
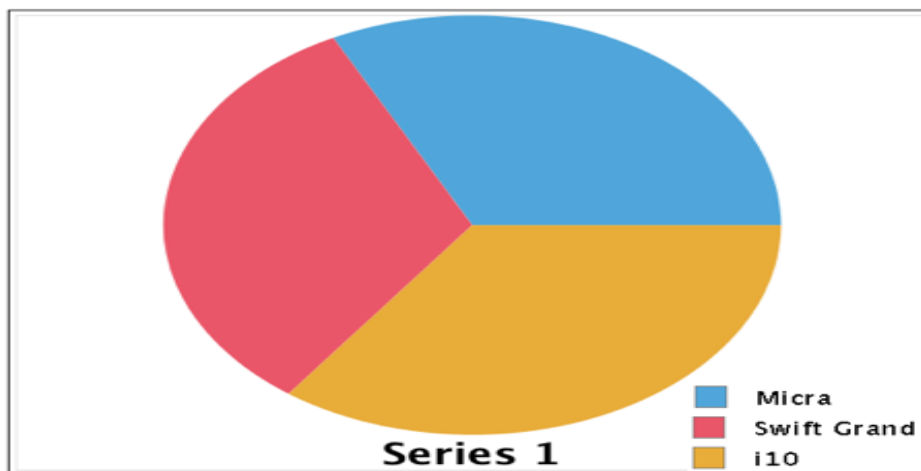

Fig 2: A Pie chart represents the cars that the customer is most likely to purchase based on customer profile and preferences.

3. TOP-K PROFITABLE CARS

3.1 Methodology

In finding Top-k Profitable Cars, Dominance analysis has been used for making various multi criteria decision. In this a set of cars in the existing market is taken and depending upon the best features in them, suggest a methodology that will be beneficial while developing a completely new car. In this approach, special care has been taken of the sum of the various packages that do not overcast the profit and is not dominated by any other car that are in market. Dynamic Programming approach has been used to produce an optimal solution. Two algorithms have been proposed which will assist in generating profits and return a set of features respectively. For analysis multiple features are molded in partial order.

Definition-Consider multiple car C_I in Segment S and F_N are features in Car. Car C_T is generating maximum profit using set of feature f so as to find various combination of feature fin set of F_N such that such that it will dominate or equalize profits generated by car C_T.

Finding Best Features – A company want to develop new Car so there are company can choose various set of feature f to generate the profits. This algorithm help in suggesting set feature will provide optimal profit. Consider table which is represented below with terms Feature, Quality and Price.

Feature	Quality	Price
Bluetooth	B1	2000
	B2	5000
	B3	8000
Parking Assist	P1	10000
	P2	15000
Audio System	A1	2000
	A2	5000
	A3	15000
	A4	20000
GPS	G1	10000
	G2	25000
Lamps	L1	1000
	L2	2500

If Car C_1 is providing set of features F_1 (B2, P1, A3, G2, and L1) which will sum cost [56,000] to provide Car C_2 which will provide feature F_2 (B3, P2, A4, G1, and L2) which have high quality as compared to car C_1 and also will generate profit of Rs.500.

3.2 Algorithm

Input: S, C, F and f

Output: the final selection f of set F and the optimal price

1: **for** S= 1 to n **do**

2: A (S, 0) 0

3: B(S, 0) ;

4: **for** C = 1 to k **do**

5: A(1, C) is (c1, ;, (0, 0, ..., 0))

6: B(1, C) {c1}

7: **for** F = 1 to k **do**

8: **for** f = 1 to n **do**

9: vselect is(C, F(f_1 − 1, f_2, f_3+ 1, f_4+2)

10: F the n-th entry in vselect

11: FselectF (f − 1, f − 1) + n

12: FnotSelectF (i − 1, n)

13: **if** Fselect _ FnotSelect **then**

14: // Case 1: ni is selected

15: n(i, n) _(S, C(c_1,c_2….c_n), F(f_1, f_2))

16: n (i, n) Fselect

17: C(F) is F(f)

18: **else**

19: // Case 2: ni is not selected

20: n(i, t) is n(f, t)

21: C (i, t) FnotSelect

22: S(i, t) is C (i, t)

23: **return** S,C,F and n

3.3 Results

Car Features

Cars/Features	Sunroof	Automatic transmission	Tow Hitch	Cruise Control	power windows	Safety Air Bags	2 DIN Radio+CD+MP3 audio
Swift Grand	✔	✔	✖	✔	✔	✖	✔
i10	✖	✖	✖	✔	✔	✖	✔
Micra	✖	✔	✖	✔	✔	✖	✔
i20	✖	✖	✔	✔	✖	✖	✔
Polo	✔	✔	✔	✔	✖	✖	✔
wagon r vxi	✖	✔	✖	✔	✔	✖	✔
Alto 800	✖	✖	✖	✔	✔	✖	✔

Fig 3: A report shows the availability of various features in different cars. This result will assist the automobile companies to provide a better product than the current products in the market.

Car TopK Result

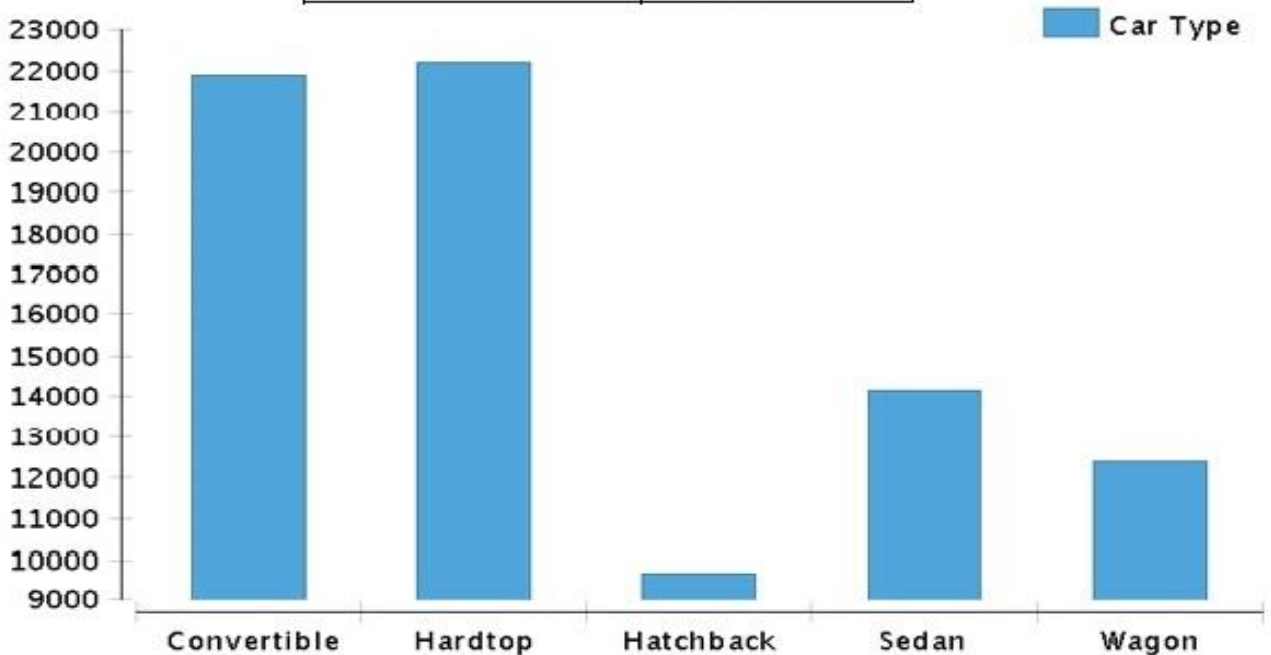| Profit	Car Type
14158.5104166667	Sedan
9672.94285714286	Hatchback
12371.96	Wagon
21890.5	Convertible
22208.5	Hardtop

Topk Result

☐ Car Type

Fig 4: A bar chart represents the profits generated by various segments of cars in million dollars.

4. CONCLUSION

The automobile industry is facing some serious problems. The proposed algorithms of customer segmentation prediction model and top-k profitable cars will assist the companies in maximizing profits. Customer segmentation can be used by the companies to help the customer in selecting the right car. Since the result returns a probabilistic prediction of the cars, the company can try and convince the customer in selecting the car that will not only be perfect for him but also maximize the company profits. Top-k profitable cars represent a set of features present in the cars of the same segment. This can be used by the automobile companies to compare their cars to different cars of different brands. The comparison will be helpful such that the company can then decide the features that could be added to their cars keeping operational costs in mind. This in turn will increase car sales, thus maximizing profits.

5. REFERENCES

[1] Thomas Verbraken, A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 5, MAY 2013.

[2] Qian Wan, Raymond Chi-Wing Wong ,G 2011 Yu Peng, Finding Top-k Profitable Products, University of Science and Technology Clear Water Bay, Kowloon, Hong Kong

[3] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[4] Christopher Bull (2003), Strategic Issues in customer relationship management (CRM) implementation, Business Process Management Journal, Volume 9, pp. 592-602 .

[5] Leo Y.M. Sin, Alan C.B. Tse and Frederick H.K. Yim (2004), CRM: conceptualization and scale development, Journal of Marketing, Vol 39, pp. 1264-1290.

[6] Jeffrey H. Dyer and Kentaro Nobeoka, (2000), Strategic Management Journal, Vol. 21, No. 3, Special Issue: Strategic Networks (Mar., 2000), pp. 345-367.

[7] Mitra, V., Wang, C.J., and Banerjee, S. 2007. Text Classification: A least square support vector machine approach. Journal of Applied

[8] Song, F., Liu, S., and Yang, J. 2005. A comparative study on text representation schemes in text categorization, Journal of Pattern

[9] Hotho, A., Nürnberger, A., and Paaß, G. 2005. A Brief Survey of Text Mining. Journal for Computational Linguistics and Language

[10] Yang, Y., Slattery, S., and Ghani, R. 2002. A study of approaches to hypertext categorization. Journal of Intelligent Information.