ISSN: 2366-1313

A COMPREHENSIVE STUDY OF MACHINE LEARNING ALGORITHMS FOR PREDICTING CAR PURCHASE BASED ON CUSTOMERS DEMANDS

[1]SABBINENI VENKATESWARA RAO, [2]P V RAMA GOPALA RAO

[1] Assistant Professor & HOD of CSE, Megha Institute Of Engineering And Technology For Women, Hyderabad, India. sabbineniv@gmail.com

svrcse@meghaengg.ac.in.

[2]Assistant Professor, Dept of CSE, Teegala Krishna Reddy Engineering College, Hyderabad, India. ramagopal.cse@tkrec.ac.in.

Abstract: *Pricing in the online world is highly transparent & can be a primary driver for online purchase. While dynamic pricing is not new & used by many to increase sales and margins, its benefit to online retailers is immense. The proposed study is a result of ongoing project that aims to develop a generic framework and applicable techniques by applying sound machine learning algorithms to enhance right price purchase (not cheapest price) by customers on e-commerce platform. The automobile industry is one of the prominent industries for the national economy. Day by day car is getting popular for the private transport system. The customer needs review when he wants to buy the right vehicle, especially the car. Because it is a very costly vehicle. There are many conditions and factors matter before buying a new car like spare parts, cylinder volume, headlight and especially price. So, deciding everything, it is important for the customer to make the right choice of purchase which can satisfy all the criteria. Our goal is to help the customer to make the right decision whether he will buy a car or not. Therefore, we wanted to build a technique for decision making in-car buy system. That's why we propose some well-known algorithms to get better accuracy for a car purchase in our paper. We applied those algorithms in our dataset which contains 50 data. Among them, Support Vector Machine (SVM) gives the best result with 86.7% accuracy of prediction.*

Keywords: *SVM, KNN, Data base, auto mobile.*

ISSN: 2366-1313

1. INTRODUCTION

Dynamic pricing or price optimization is the concept of offering goods at different prices which varies according to the customer's demand. The pricing of the commodity can be done based on competitor's pricing, supply, demand and conversion rates and sales goals [1]. The art of dynamic pricing is sometimes also referred to as individual level price discrimination [2], revenue management [3] and yield management. Further, adjustment of the prices attributed towards customer's will [4] is another way of defining dynamic pricing. Additionally, customization of the inventory goods by segmenting the customers on the basis of product choice and thus proffering different prices to them is dynamic pricing [5]. It is also referred as real time pricing, wherein value of a product is determined by the current market conditions under commercial transactions. It is a blanket term for shopping experience which defines the prices of the products according to the competitive environment among the suppliers, time of the day and the weather conditions [6]. Dynamic pricing is a wide spread phenomenon whose influence can be seen in industries like retail, automotive, mobile communication, electricity, air ticket and a lot many. The rise in the retail industry [7] has been due to the increased availability of demand data of the customers, the new technologies helping determine prices more efficiently by studying the consumer pattern and the decision support tools due to the new emerging technology. The influence, experienced in the mobile communication sector [8], can be attributed towards the decreased call rates, higher competition level and improved network infrastructure. Moreover, the impact felt in the automotive industry is the result of enhanced coordination among the production processes and the inventory decisions, building a direct-to-costumer business prototype [9]. Further ahead, this phenomenal concept of dynamic pricing has been at leads because of the intensified network connections [10]. This has helped both the customers and the

ISSN: 2366-1313

sellers under two factors, lower menu cost and integrated customer information as a complete database. The internet accessibility has helped buyers or consumers acting as a self-service facility and thus saving time. The vendors are also benefitted by this concept of dynamic pricing with the amalgamation of web integration and automation in many ways. It eliminates the physical presence of the vendor [12], lowers the input cost, and integrates the customer information under one database and reduces the cost of printing new catalogs [3]. Also it acts not as a one way street between consumers and seller instead an explicit platform to discuss and exchange reviews for better services. Dynamic pricing as an application can be functional under certain factors, that is, customers' willingness to pay different prices, segmented market availability, less arbitrage potential, fair play rules and revenue cost to be higher than the segmenting and policing capital [4]. Further, it can be implemented in the industries with fixed high cost and low variable cost.

2. LITERATURE SURVEY

Some people preferred good parts, some are high or low price with all of their needed features, some are only weak for famous brands of the car only. To select the perfect car is still a difficult task though some parameters like color, comfort, seating capacity, etc are known [2]. Thats why we tried to compare some algorithm for predicting car buying purpose that which one gives better accuracy. An implementation of Nave Bayes Classification method is proposed by Fitrina et. al [3]. Naive Bayes is known as a simple probabilistic classifier. They applied this method for predicting purchase. They used a dataset on 20 car purchas data and got 75% accuracy. Srivasta et. al [4] applied the powerful learning method, Support Vector Machinen(SVM) on different types of data like Diabetes Data, Heart Data, Satellite Data and Shuttle Data. Those datasets have multi classes. They are also proven the analysis of the comparative consequences the use of divers kernel functions on their paper. A comparative analysis of machine learning algorithm was proposed by Ragupathy et. al [5]. In their paper, they tried to identify and

ISSN: 2366-1313

classify sentiment, conveyed in main text. They have collected their data from social media like Twitter, comments, blog posts, news, status updates etc. They also applied Naive Bayes, Decision Tree, K-Nearest Neighbour and Support Vector Machine classifiers for their comparison pur pose. Their goal was to find out the most efficient classification technique and SVM came out with 72.7% which was the best accuracy. Another prediction system using supervised machine learning technique was proposed by Noor et. al [6]. They used multiple linear regression method and predicted vehicle price. They got 98% accuracy on their system. Pal et. al [7] proposed a methodology for predicting used cars costs. In their paper, they used Random Forest classifier to predict the costs of used cars. To train the data, they created a Random Forest with 500 Decision trees. Finally, they got 95.82% as training accuracy and 83.63% as testing accuracy. Pudaruth et. al [8] proposed another methodology for predicting used cars prices. In that paper, he applied multiple linear regression analysis, k-nearest neighbours, Naive Bayes and Decision Tree which were used to make the predictions. Osisanwo F.Y. et. al [9] proposed Supervised machine learning technique. They compared seven different Supervised learning algorithms and described those. They also found out the most effective classification algorithm established on dataset. A different work on car purchase was proposed by R.Busse et. al [10]. In their paper, they prioritized the psychological effect of weather. They applied projection bias and salience as two major psychological mechanism. A new defect classification technique was proposed by veni et. al [11] to predict the class label of ”severity” tuple. Those data tuples were described by various attribute like Phase attribute, Defect, Phase Defected, Impact and Weight. They applied Naive Bayes classifier for prediction purpose. Jayakameswaraiah et. al [2] developed a data mining system to analyze cars. They proposed TkNN clustering algorithm to predict the right car. They also shown the comparison of KNN and their proposed novel TkNN clustering. Another car price prediction technique was proposed by Gegic et. al [2] where they used three

ISSN: 2366-1313

machine learning techniques. They got 92.38% accuracy on combination of all ML methods. Another medical work was proposed by Jabbar et. al [3] to predict heart disease in diagnosis system. They used K nearest neighbour(KNN) algorithm to predict it. The algorithm performs tremendously with 100% accuracy. Peerun et. al [4] presented a technique to predict rice of second-hand cars. In their paper, they used Artificial Neural Networks. They applied it on dataset of 200 records cars and compared different kinds of machine learning algorithm. Yuan et. al [5] offered a study on prediction. He tried to predict the car sales based on some web search records. In spite of these well-known works, there also exist some more challenging works. As a result, we focus on the comparison of four types of well-known machine learning algorithms and try to find out which algorithm gives the best accuracy for our dataset.

3. AN OVERVIEW OF PROPOSED SYSTEM

The proposed model considers the amalgamation of three different techniques - to identify the customer segments, appropriate pricing for them, and the prediction for their likely purchase within that price range. The is the first and foremost step in the process of the framework. It involves the collection of data from various data points under an integrated database. For the research purpose, we used a subset of an online marketplace data the schema of the two data sets.

Algorithm:

To predict something, first of all, we have to learn our machine. Those machines can learn with the proper algorithm. There are three types of machine learning algorithms. They are supervised learning, unsupervised learning, Semi-supervised learning. Among those, we choose supervised learning algorithms. Those are Nave Bayes, Support Vector Machine (SVM), K-nearest neighbour algorithm(KNN) and Random Forest tree.

1) Naive Bayes: Naive Bayes is known as an arithmetical classification method based on the Bayes Theorem for classification problems. It is a simple learning algorithm which is not only known for its easiness but also its effectiveness. It is regarded as nave

ISSN: 2366-1313

because of its assumption shortens calculation.

The KNN algorithm deals with similar things which are near or close to each other. That's mean, it assumes to compute all those similar things which occur in close nearness area4 .K-NN are also known as a lazy learner. Because it only memorizes the training dataset easily, doesn't want to learn a discriminative function system from the training data.

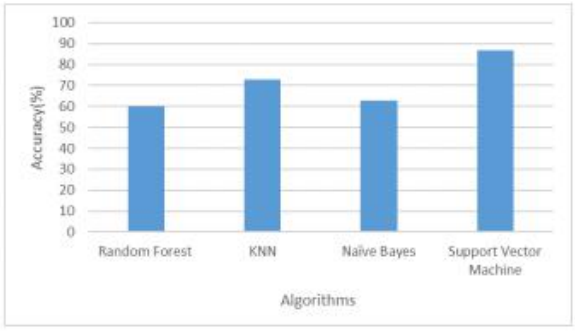

Fig. Accuracy of Several algorithms.

We can see Support Vector Machine gives highest accuracy(87.6%) than Random Forest, K-nearest Neighbour(KNN) and Naive Bayes. That means Support Vec tor Machine can classify approximately 44 car purchase data of 50 dataset.

There are many study about car price prediction from many years. Several study use several machine learning techniques. A methodology for predicting purchase used Naive Bayes algorithm and get 75% accuracy [3]. Another work for predicting used cars prices used Random Forest and get test accuracy 83.63% [7] On the other hand, our proposed method used Cosine Distance for review analysis and after that using Support Vector Machine got 86.7% accuracy. There are some obstacles determined in different classifiers like Naive Bayes classifier cannot cope with more amounts of data with ease

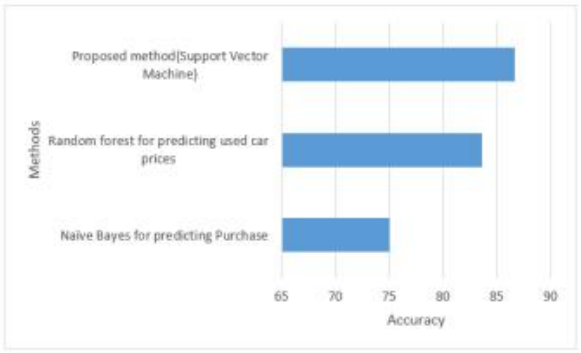

Fig. Comparison with Other Methods

4. CONCLUSION

The proposed framework has been designed using the powerful techniques of Machine Learning, Data Mining and Statistical Methods to predict the purchase behavior of an online customer by selecting an appropriate price range for him based on Dynamic Pricing. This framework has been tested on a large

ISSN: 2366-1313

dataset for an e-commerce firm and results are encouraging enough to implement the framework completely. The error rate is reduced and much better price range, which is appropriate for both customer and organization, is being determined. The general framework can be applied in the various industries working in online mode and can be customized to specific applications. The results of the work-in-progress are likely to be discussed in the extension of this study.

REFERENCES

1. Shpanya, A. (2013) "5 Trends To Anticipate In Dynamic Pricing". Retail Touch Points. Accessed on April 1, 2014.

2. Garbarino, E., & Lee, O. F. (2003). Dynamic pricing in internet retail: effects on consumer trust. Psychology & Marketing, 20(6), pp. 495-513.

3. McAfee, R. P., & Te Velde, V. (2006). Dynamic pricing in the airline industry.forthcoming in Handbook on Economics and Information Systems, Ed: TJ Hendershott, Elsevier.

4. Reinartz, W. (2002). Customizing prices in online markets. Management (www. unimib. it/symphonya), (1), pp. 55-65.

5. Strauss, J., Frost, R., & Ansary, A. I. (2009). E-marketing. Pearson Prentice Hall.

6. Cummings, T. (2013). Csmonitor.com, Everything you need to know about dynamic pricing, accessed from http://www.csmonitor.com/Business/Saving-Money/2013/1104/Everything-you-need-to-know-about-dynamic-pricing on 19th June, 2014.

7. Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. Management Science, 49(10), pp. 1287-1309.

8. Fitkov-Norris, E. D., & Khanifar, A. H. M. A. D. (2000). Dynamic pricing in mobile communication systems.

9. Biller, S., Chan, L. M. A., Simchi-Levi, D., & Swann, J. (2005). Dynamic pricing and the direct-to-customer model in the automotive industry. Electronic Commerce Research, 5(2), pp. 309-334.

10. Weiss, R. M., & Mehrotra, A. K. (2001). Online dynamic pricing: efficiency, equity and the future of e-commerce. Va. JL & Tech., 6, pp. 11-17.

11. Dolan, R. J., & Moon, Y. (2000). Pricing and market making on the Internet.Journal of Interactive Marketing, 14(2), pp. 56-73.

12. Prasadu Peddi (2019), Data Pull out and facts unearthing in biological Databases, International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

13. Rutherford, D. (1995). Routledge dictionary of economics. Taylor & Francis.

14. Randel, S. (2009). Dynamic Pricing: An old practice is making a comeback online, iProspect.com accessed from http://www.iprospect.com/blog/featured/dynamic-pricing-an-old-practice-is-making-a-comeback-online.html on 12th May, 2014.

ISSN: 2366-1313

15. Varian, H. R., & Shapiro, C. (1999). Information rules: a strategic guide to the network economy. Harvard Business School Press, Cambridge.

16. Prasadu Peddi (2021), "Deeper Image Segmentation using Lloyd's Algorithm", ISSN: 2366-1313, Vol 5, issue 2, pp:22-34.