Predicting the Likeliest Customers; Minimizing Losses on Product Trials Using Business Analytics

Tushar Nigam(✉) and Rashmi Agarwal

RACE, REVA University, Bangalore 560064, India
tushar.ba07@race.reva.edu.in, rashmi.agarwal@reva.edu.in

Abstract. A product trial is a great way of marketing. It offers a first-hand experience that a customer can use to evaluate the product and decide whether to buy it. However, there are expenses involved in every trial. If the buyer decides not to purchase the product after the trial, this expense gets wasted. The lower the conversion rate, the higher the loss that the company must face. To reduce these losses, a strategy should be developed on which studies should be conducted. One such business case is the subject of this paper. A software development company has launched a new product and given its current customers access to the product's trial version. A loss of $8.5M resulted from 56% of trial customers not buying the product. The business is therefore looking for strategies to reduce this loss. This paper aims to present the work which is done to solve this business problem. The goal is accomplished by employing analytical methods to determine the most likely clients. The data provided by the company is used to build binary classification models. Another way to find the likeliest customer is based on the purchase pattern of the customers. Affinity analysis is done on the data to identify the set of products with which the new product is frequently sold to target the customers who have purchased those products for the sale of the new product. Among the binary classification models that were built, the Random Forest model outperformed other models with an accuracy of 83%. This enables the business to take calculated decisions while extending the trial to the customers. Alternatively, Market Basket Analysis, with an accuracy of 88%, discovered a set of two products, the existing buyers of which are more likely to buy the newly launched product. This information not only helped find the right customers but also paved the path for cross-selling of the new product.

Keywords: Likeliest Customers · Binary Classification · Market Basket Analysis · Cross Selling

1 Introduction

Product trial is a unique form of advertising as it provides a direct experience to the customer. It is much better than documentation, PPT, and theories as it's the working model of the solution the product provides. With the advancement of technology, new

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2023
Published by Springer Nature Switzerland AG 2023. All Rights Reserved
S. Nandan Mohanty et al. (Eds.): ICISML 2022, LNICST 470, pp. 499–506, 2023.
https://doi.org/10.1007/978-3-031-35078-8_41

products are emerging constantly and rapidly. But not all of them become popular and profitable. Studies mention that 95% of new products fail [1]. Not enough exposure to the customers could be one of the major reasons for this as the customers are not familiar with the capability and potential of the product [2]. Therefore, a trial becomes very important for customers to evaluate the product and make purchase decisions.

Although the trial is an effective way of promotion, it is also a very expensive way of marketing. There are a lot of resources involved in the trials and there is a cost associated with every resource. These resources could be user training, operations, technical support, cloud computation and storage, licenses, etc. The company loses revenue if the conversion rate of customers after trials is low.

Therefore, there should be a strategy based on which the trials are given to the customer so that losses are minimal. Purchase intent is one of the main factors that need to be looked at before giving any trial to the customer [3]. It is the probability that a customer will buy a product. Once the purchase intent is evaluated, the decision of giving a trial or not can be taken. Purchase Intent can be evaluated with the help of predictive analytics.

A software development company, RLP Software Private Limited launched a new product 'Strategy Builder' and has extended the trial version of the tool to its existing customers. The conversion ratio after the trial is just 44%. Based on the data provided, the business has lost $8,450,200 on unsuccessful trials. The objective of the study is to minimize these losses. Predictive Analytics would be used to identify the customers with higher purchase intentions. It would also be interesting to know if the new product is purchased more by the customers who own a specific set of existing products. Affinity analysis is a powerful tool to get this information.

2 Literature Review

Existing published works were reviewed to identify the right set of solutions for this problem. The Literature review done for this case is divided into 4 parts. The first part focuses on the importance of trials after the launch of a product, the second one explains the concept of Purchase Intentions, and the last 2 parts are dedicated to various machine learning algorithms to improve the trial conversion.

2.1 Importance of Product Trial

Whenever a new product is launched, there is always an expectation of great sales by the company. Most of the time the results are not as expected and from there, the retrospect and improvisation start. For customers who have not purchased the product, the company can sell products through the improvement of product satisfaction. However, the more product satisfaction is improved, the more services need to be paid for, which increases the cost of the company. The company wants to sell its products successfully with minimal cost. To maximize the benefits for the company, a data mining system must be developed to help companies focus on unpurchased customers with the strongest purchase intentions [4].

2.2 Purchase Intention

It is the willingness of the customer to buy a certain product. It's a dependent variable based on several independent factors and is a measure of a potential customer's attitude towards purchasing a product. It is an important metric in designing marketing strategies [5]. Price had been an important variable in influencing purchase intentions but other variables such as knowledge, experience, and quality are important in the process of customer's purchase decisions too [6]. For that matter, the duration of association of customers with the company also plays an important role. Another factor is the perceived value, which is considered directly proportional to the purchase intent [7]. With these vast possibilities, the first step is to identify the factors affecting customers' purchase intention.

2.3 Binary Classification

It is a powerful technique that could be used to predict purchase intention. From the business point of view, two prediction models are needed [8]. The prediction model which uses some or all behavioral data could only be applied to the company's existing client base, for whom all data is available - the so-called cross sale [9]. On the other hand, it makes sense to build a separate prediction model using only the socio-demographic data if the company were interested in acquiring customers that have not yet established a business relationship with the company. Since the modeling objective became clear, the next step was to identify the type of classification.

Binary classifiers are a widely used option, however, there is an alternate method available called one-class classification, which works only with a single class of data [10]. This technique is used when class distribution is imbalanced i.e., one of the classes is much higher in data than the other. In such cases, the former may not perform well. Considering the data in this study the classes are divided in a ratio of 44:56, which makes it a balanced dataset. Therefore, it was decided to go with the binary classifiers only for this study.

2.4 Market Basket Analysis

Once the purchase intent is evaluated, several other important patterns can be discovered from data collected from the business. Affinity analysis is one such technique for achieving that. For identifying the affinity between the products there have been multiple pieces of research done. Most of them involved using Market Basket Analysis. A prior study [11] shows that this way is useful in finding out interesting patterns from a large amount of data, predicting future association rules as well as the right methodology to find out outliers. Among various methods for affinity analysis, the apriori algorithm is found to be better for association rule mining. There are various complexities involved in the apriori algorithm like the exhaustive scan of the database multiple times and as the input becomes larger, the computation time increases significantly [12]. Still, this is a popular choice among data scientists. The data collected from the business is moderate in size, hence these limitations will have negligible effects on our study.

Some studies tried a completely new way of improvising the MBA. One such study used the Map/Reduce of Cloud Computing [13]. The algorithm has been executed on EC2 small instances of AWS with nodes 2, 5, 10, 15, and 20. The execution times of the experiments show that the proposed algorithm gets better performance while running on a large number of nodes to a certain point. However, from a certain point, Map/Reduce does not guarantee to increase the performance even though we add more nodes because there is a bottleneck for distributing, aggregating, and reducing the data set among nodes against the computing powers of additional nodes. Hence, the scope of the study is limited to the basic algorithm for finding associations.

3 Data Description

The data was scattered into multiple entities in the SQL database. While most of the sales data were available as read-only to internal employees but there were a couple of entities that had restricted access as they contained Confidential information. So, this study is done on data that does not violate the confidentiality clauses.

The sales team provided the list of customers who were granted trial versions of the new product. The data had two columns 'Customer Id' and 'Purchased' indicating whether the customer bought the product or not after the trial. The rest of the data was extracted from the Sales database based on the id of the customer.

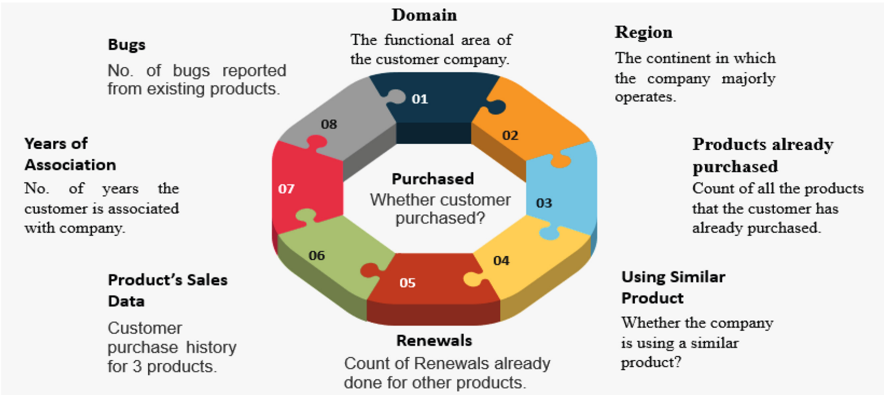

Fig. 1. Description of Variables

Figure 1 shows the variables and their significance that were involved in the study. The data extracted had 6560 observations of the customers on which trials were done. There are 12 variables out of which 10 are independent, 1 is dependent and 1 is a customer id column which would not be considered in the study further. There were no missing values found in the dataset. The dependent Variable 'Purchased' had a balanced no of observations for each class with ratio of 44:56.

4 Data Preparation

Data Preparation is an extremely important stage because the model and its accuracy depend on this a lot. This involves cleaning data, deriving new features, and formatting data. The model usually understands numeric data; hence the categorical variables must be converted to numeric ones.

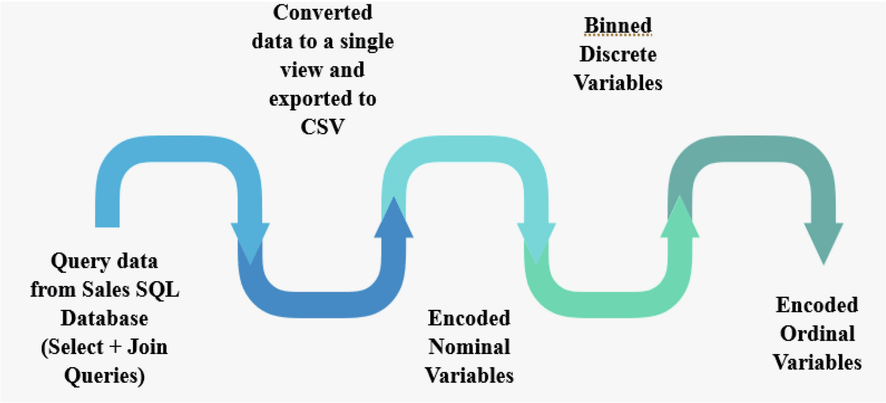

Fig. 2. Data Preparation Flow

Figure 2 shows different techniques that were applied to preparing data for modeling. Encoding and Binning are the major ones among them followed by replacing the binned columns with numeric ones.

One-Hot Encoding: It is the process of creating dummy variables for categorical variables where order does not matter. For every categorical feature, a new variable is created.

Binning: It is a technique for reducing the cardinality of continuous and discrete data. Binning groups related values together in bins to reduce the number of distinct values.

With the application of these preprocessing techniques, the data was ready for the modeling stage.

5 Model Building

For achieving the objective of the study, there were 2 techniques identified. One was to build a predictive model, and another was to find the affinity of the new product with the set of existing products based on purchase history. The solution for each objective is explained below:

5.1 Identify the Most Probable Customer for the New Product

Method: Binary Classification

Dependent Variable: Purchased.
Independent Variables: 10 (All columns - Purchased)
Train and Test data ratio: 75:25
Metrics: Accuracy
Techniques:

Decision Trees,
Logistic Regression,
Naïve Bayes,
Random Forest,
XGBoost.

5.2 Identify the Set of Products with Which the New Product Gets Sold Frequently

Method: Market Basket Analysis

Variables: Individual existing product purchase data for each customer.
Technique: Apriori algorithm for finding Association Rules
Metrics: Lift, Confidence and Support.
Train and Test data ratio: 75:25

6 Evaluation

After application of different machine learning algorithm, the next step is to evaluate them so that the most appropriate technique can be used for solving the business problem. Below are the evaluation results for the techniques applied:

6.1 Predictive Analytics

In the last step, multiple binary classification models were built, but not all can be used by the customer, hence, the most accurate one should be picked for predicting the likelihood. Table 1 shows the comparison of accuracy.

6.2 Market Basket Analysis

The combination of two products showed a high association with the new product in the training data. This result gave 88.28% accuracy when checked on the testing data.

Table 1. Accuracy Comparison

Modeling Technique	Accuracy
Decision Trees	77.9%
Logistic Regression	70.6%
Naïve Bayes	72.8%
Random Forest	**83%**
XGBoost	82%

7 Results

After evaluating the models and algorithms based on their respective metrics, like Lift, Confidence, and Support for Market Basket Analysis and Accuracy score for Binary Classification, below results and observations were found:

7.1 Predictive Analytics

After Applying multiple machine learning algorithms to the dataset, Random Forest came out to be the best model with the highest accuracy. The objective of finding the most probable customer is achieved with the model finalized.

7.2 Market Basket Analysis

Another objective of the study was to find the association between the products. This information is now available for the customer to improvise its sales strategy. Based on the lift and confidence values found from association rules, the customers who have purchased the products 'MLC' and 'QPC' both are more likely to purchase the new product. Therefore, the customers who have already purchased these products should be targeted for the trials of the new product. Additionally, this combination of products can be considered for cross selling the new product.

8 Conclusion

This paper gives an overview of data mining and modeling techniques applied to minimize the losses in product trials. Predictive analytics with the help of the Random Forest model gave an accuracy of 83% which is quite decent. There are no traces of underfitting and overfitting in the model. Therefore, the trials should be given to customers only based on the response of the model. For the customers, whose likelihood is less, should be given a demo/video of the product for marketing instead of a separate trial version, as they are the company's customers too and can be potential customers for the new product. Affinity analysis with the help of the Apriori algorithm discovered the fact that the new product gets sold more often by the customers who already own the existing 'MLC' and 'QPC' products of the company. Hence, the new product should be considered for cross-selling with these products.

9 Scope for Future Work

The solution for the problem that the business is facing could also be achieved by clustering, which could help in understanding similarities between the customers who are not purchasing the new product. This technique would divide the customers into a certain number of clusters based on similar traits. The idea is to take advantage of these common traits to identify the measures to onboard these customers.

References

1. Emmer, M.: Inc.com (2018). https://www.inc.com/marc-emmer/95-percent-of-new-products-fail-here-are-6-steps-to-make-sure-yours-dont.html. Accessed 10 Aug 2022
2. Sun, K., Zuo, M., Kong, D.: What can product trial offer?: the influence of product trial on chinese consumers' attitude towards IT product, vol. 8, no. 1, pp. 24–37 (2017). https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJABIM. https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJABIM.2017010102, doi: https://doi.org/10.4018/IJABIM.2017010102
3. Contributor, T.: What is purchase intent? - Definition from WhatIs.com (2017). https://www.techtarget.com/whatis/definition/purchase-intent. Accessed 10 Aug 2022
4. Mai, W.: A data mining system for potential customers based on one-class support vector machine. J. Phys. Conf. Ser. **2031**(1) (2021). https://doi.org/10.1088/1742-6596/2031/1/012066
5. Team, B.C.: MBA Skool (2021). https://www.mbaskool.com/business-concepts/marketing-and-strategy-terms/10976-purchase-intention.html. Accessed 10 Aug 2022)
6. Mirabi, V.: A study of factors affecting on customers purchase intention case study : the agencies of bono brand tile in Tehran. J. Multidiscip. Eng. Sci. Technol. **2**(1), 267–273 (2015)
7. Fathy, A.: Identifying the factors affecting customer purchase intention. Glob. J. Manag. Bus. Res. Adm. Manag. **15**(2), 1–6 (2015)
8. Jaiswal, S.: Javatpoint (2011). https://www.javatpoint.com/machine-learning-models. Accessed 11 Aug 2022
9. Bole, U., Papa, G.: Who are the likeliest customers: direct mail optimization with data mining. Contemp. Eng. Sci. **4**(6), 259–268 (2011)
10. Bellinger, C., Sharma, S., Japkowicz, N.: One-class versus binary classification: Which and when? In: Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012, vol. 2, pp. 102–106 (2012). https://doi.org/10.1109/ICMLA.2012.212
11. Kaur, M., Kang, S.: Market basket analysis: identify the changing trends of market data using association rule mining. Procedia Comput. Sci. **85**, 78–85 (2016). https://doi.org/10.1016/j.procs.2016.05.180
12. Gupta, S., Mamtora, R.: A survey on association rule mining in market basket analysis. Int. J. Inf. Comput. Technol. **4**(4), 409–414 (2014). http://www.irphouse.com/ijict.htm
13. Woo, J., Xu, Y.: Market basket analysis algorithm with map/reduce of cloud computing. In: 2011 International Conference on Parallel, no. April 2012 (2011). http://www.lidi.info.unlp.edu.ar/WorldComp2011-Mirror/PDP4494.pdf