1st EAI International Conference on Intelligent Systems and Machine Learning (ICISML-2022)
Organized
by
Department of Electronics & Communication Engineering,
Vardhman College of Engineering, (Autonomous), Hyderabad, India, 16th -17th December 2022

Paper Title: **Predicting the Likeliest Customers; Minimizing Losses on Product Trials using Business Analytics**

Authors: Tushar Nigam, Rashmi Agarwal
REVA Academy For Corporate Excellence, REVA University, Bengaluru, India

Paper Id: 325168

Plan of Presentation

- Introduction
- Literature Review
- Methodology
- Results & Discussion
- Future scope
- Conclusion
- References

Introduction

Studies mention that 95% of new products fail. Not enough exposure to the customers could be one of the major reasons for this as they are not familiar with the capability and potential of the product. Therefore, a **Trial** becomes very important for customers to evaluate the product and make purchase decisions.

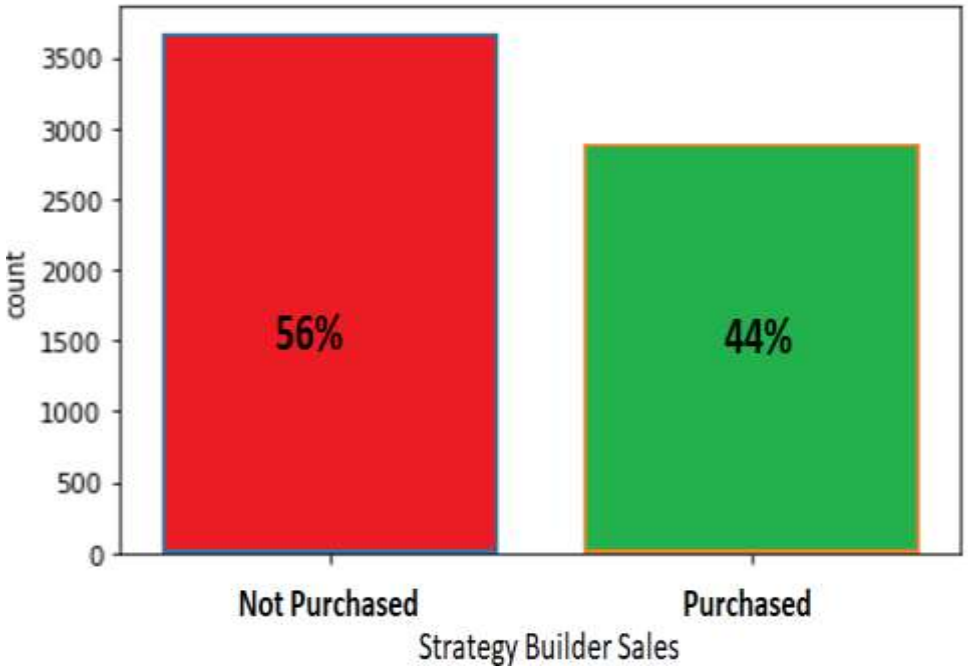
Ref: optinmonster.com

Considering this, **RLP Software Pvt. Ltd***, a software development company, after launching its new product **Strategy Builder*,** extended the tool's trial version to its existing customers. But the problem here is the cost associated with every trial. The resources which contribute to the cost are Dev-Ops, Training, Cloud Storage and Computation, Licensing, Technical Support, etc. If the customer does not purchase the product after the trial, this amount gets wasted.

The **conversion rate** after the trial is just **44%,** leading to a loss of **$8.5 M** on unsuccessful trials. Hence, the company is looking for ways to minimize these losses.

* Names changed for confidentiality

Problem Statement



Cost of each trial: $2300

Customers who did not purchase the product: 3674

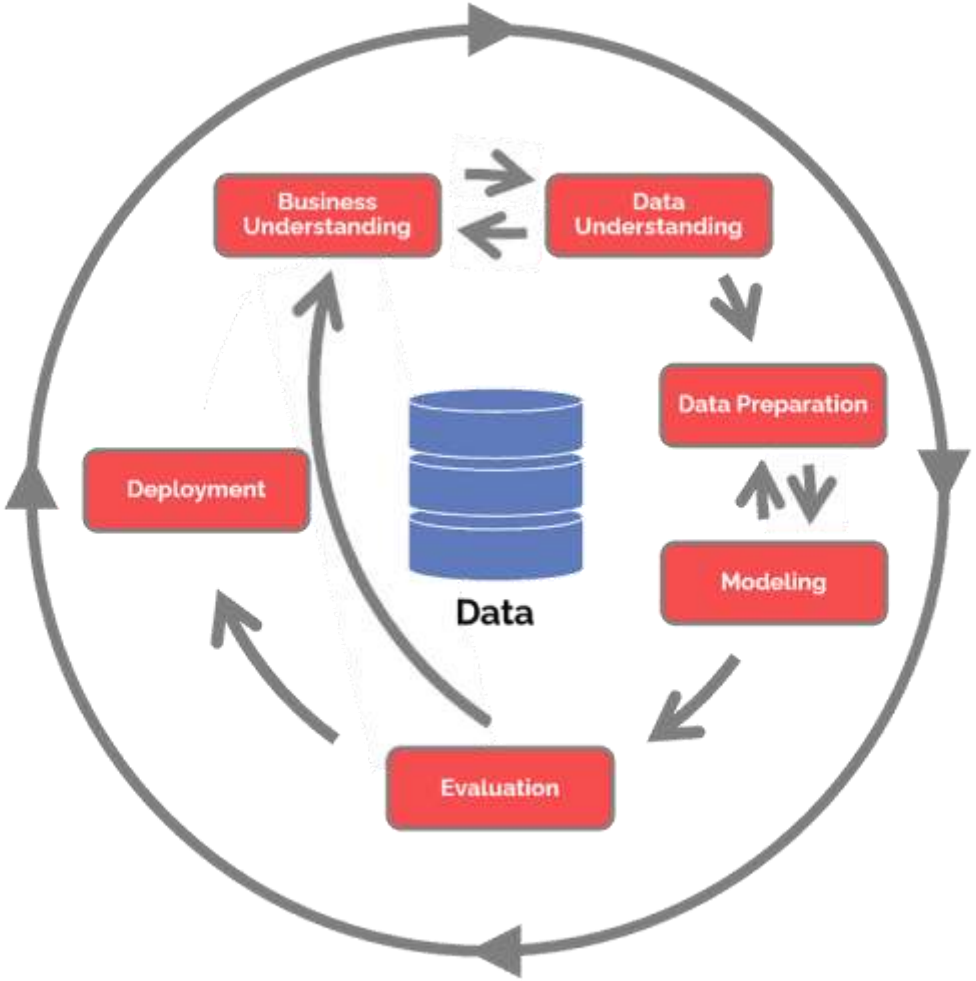
Loss = (3674 * 2300) = **$8,450,200**

The company loses out on Revenue because of this. Hence, the company is looking out for ways to **"Minimize the Losses on Trials of Strategy Builder Tool".**

Literature Review

Title of the Paper	Author	Source	Major Insight
Who are the Likeliest Customers; Direct Mail Optimization with Data Mining	Bole, Uroš Papa, Gregor - 2011	Contemporary Engineering Sciences	This paper paved the path for this study and gives an overview of the data mining work performed with the aim to build a prediction model, which could help save costs for companies that engage in direct marketing activities. Binary Classification and Clustering were the techniques used.
One-class versus binary classification: Which and when?	Bellinger, Colin Sharma, Shiven Japkowicz, Nathalie - 2012	Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012	One-class classification is useful when there is an overabundance of data for a particular class. In such imbalanced cases, binary classifiers may not perform very well. The paper investigates the performance of binary and one-class classifiers as the level of imbalance increases.
Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining	Kaur, Manpreet Kang, Shivani - 2016	Procedia Computer Science	The main aim to provide the information to the retailer to understand the purchase behavior of the buyer for correct decision making. The existing algorithms work on static data, and they do not capture changes in data with time. But proposed algorithm not only mine static data but also provides a new way to consider changes happening in data

Methodology



Ref : www.datascience-pm.com

CRISP-DM: The Cross Industry Standard Process for Data Mining is adopted for this study. It has six sequential phases:

1.Business Understanding

2.Data Understanding

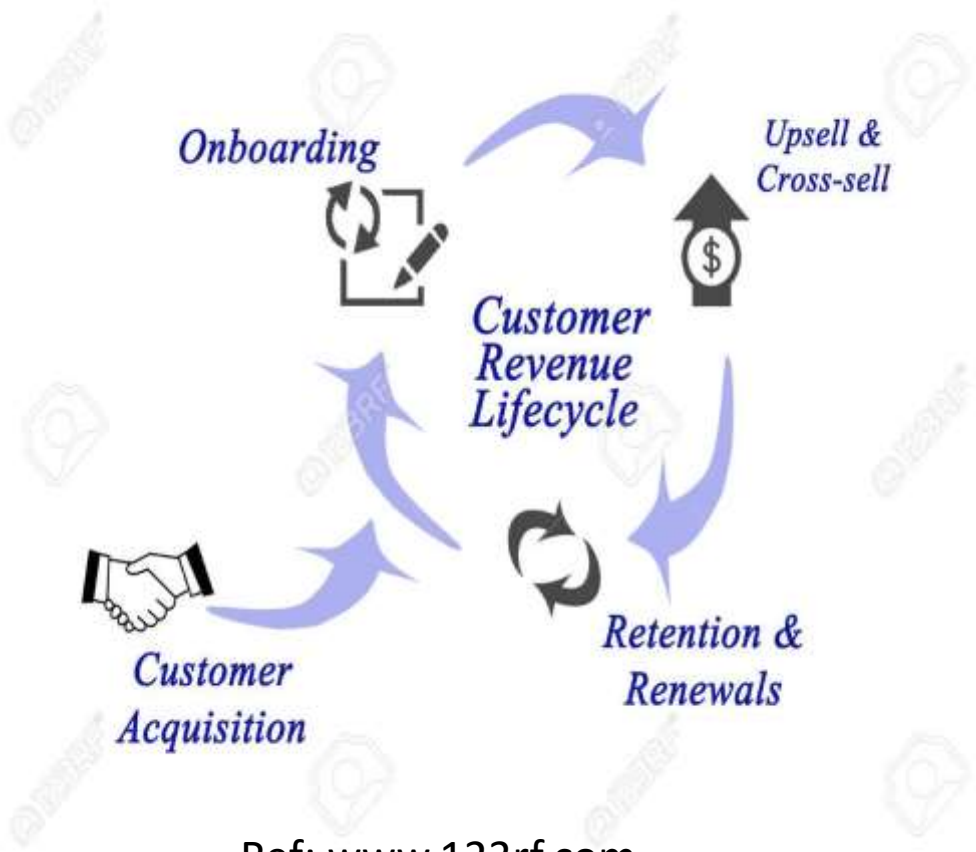
3.Data Preparation

4.Modeling

5.Evaluation

6.Deployment

Business Understanding



Ref: www.123rf.com

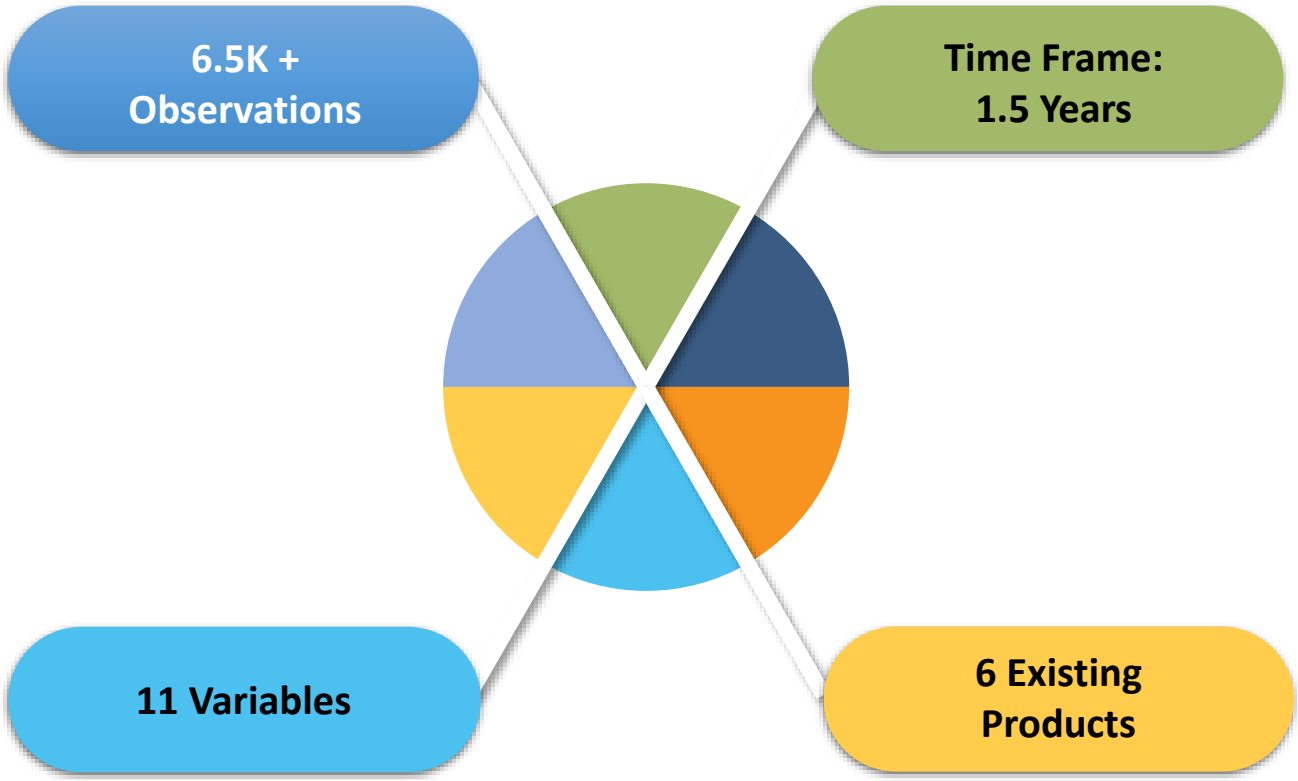
RLP Software Pvt Ltd. headquartered in San Mateo, California is a **B2B** company developing business process automation solutions.

It offers the Commercial Operations Suite, which enables businesses to optimize quotes and digital commerce, manage contracts and documents, as well as automate revenue management. The company caters to the energy, financial services, healthcare, media, retail and other sectors.

Well-known products used by **11000+ customers** of the company are :

1. CPQ (Configure, Price, Quote)
2. CLM (Contract Lifecycle Management)
3. IWA (Intelligent Workflow Approvals), etc.

Data Understanding



Variables

Domain

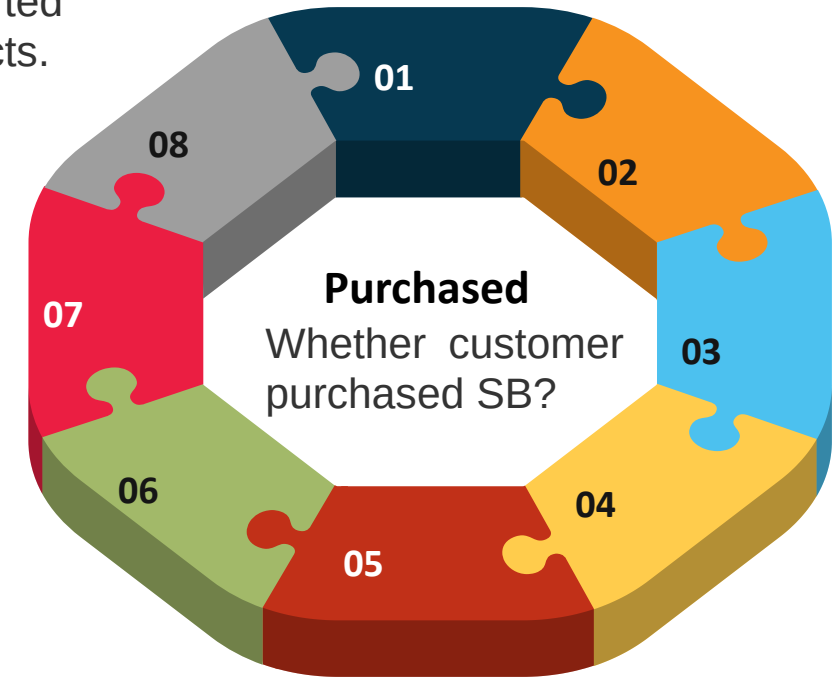
The functional area of the customer company.

Bugs

No. of bugs reported from existing products.

Region

The continent in which the company majorly operates.



Years of Association

No. of years the customer is associated with RLP.

Products already purchased

Count of all the products that the customer has already purchased.

Product's Sales Data

Customer purchase history for 3 products.

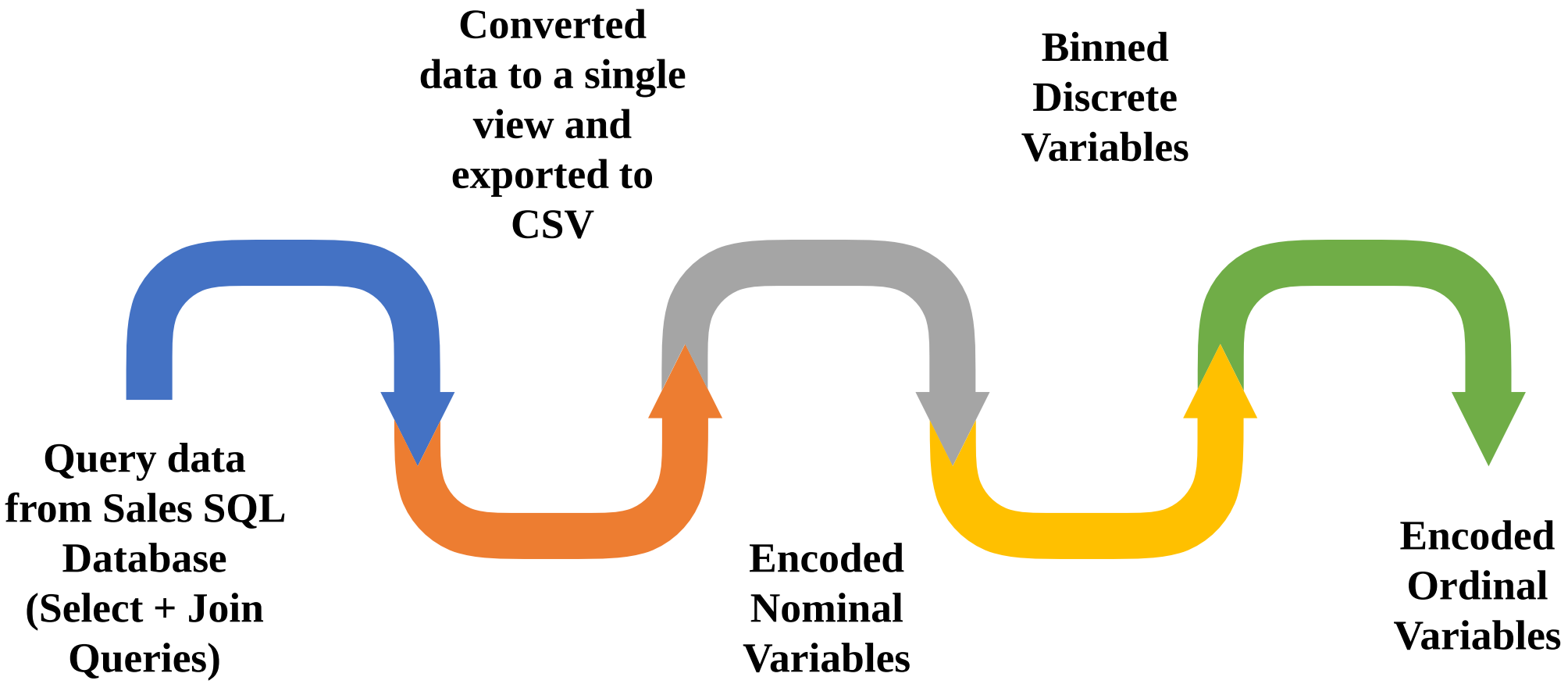
Using Similar Product

Whether the company is using a similar product?

Renewals

Count of Renewals already done for other products.

Data Preparation



Results

Objective 1: Identify the most probable customer for the new product.

Method: Binary Classification

Dependent Variable: Purchased

Independent Variables: 10 (All columns - Purchased)

Ratio of Train and Test Data : 75:25

Result: Random Forest Model performed the best among other models.

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

Results

Objective 2: Identify the affinity of Strategy Builder with the existing products of RLP.

Technique: Market Basket Analysis - Apriori (Association Rules)

Observation: The customers who have purchased CLM and CPQ both are more likely to purchase the new product, followed by the customers who just bought the CPQ product.

Conclusions and Recommendations

Based on the above study, below are the conclusions and recommendations for RLP software:

- 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.

- 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 ‘CLM’ and ‘CPQ’ products of the company. Hence, the new product should be considered for cross-selling with these products.

- 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.

Scope For Future Work - Clustering

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] "95 Percent of New Products Fail. Here Are 6 Steps to Make Sure Yours Don't | Inc.com." https://www.inc.com/marc-emmer/95-percent-of-new-products-fail-here-are-6-steps-to-make-sure-yours-dont.html (accessed Aug. 10, 2022).

[2] K. Sun, M. Zuo, and D. Kong, "What Can Product Trial Offer?: The Influence of Product Trial on Chinese Consumers' Attitude towards IT Product," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJABIM.2017010102*, vol. 8, no. 1, pp. 24–37, Jan. 1AD, doi: 10.4018/IJABIM.2017010102.

[3] "What is purchase intent? - Definition from WhatIs.com." https://www.techtarget.com/whatis/definition/purchase-intent (accessed Aug. 10, 2022).

[4] W. Mai, F. Wu, F. Li, W. Luo, and X. Mai, "A data mining system for potential customers based on one-class support vector machine," *J. Phys. Conf. Ser.*, vol. 2031, no. 1, 2021, doi: 10.1088/1742-6596/2031/1/012066.

[5] "Purchase Intention - Meaning, Importance, Factors & Example | MBA Skool." https://www.mbaskool.com/business-concepts/marketing-and-strategy-terms/10976-purchase-intention.html (accessed Aug. 10, 2022).

[6] V. Mirabi, H. Akbariyeh, and H. Tahmasebifard, "A Study of Factors Affecting on Customers Purchase Intention Case Study : the Agencies of Bono Brand Tile in Tehran," *J. Multidiscip. Eng. Sci. Technol.*, vol. 2, no. 1, pp. 267–273, 2015.

[7] A. Fathy, S. Younus, F. Rasheed, and A. Zia, "Identifying the Factors Affecting Customer Purchase Intention," *Glob. J. Manag. Bus. Res. Adm. Manag.*, vol. 15, no. 2, pp. 1–6, 2015.

[8] "Machine Learning Models - Javatpoint." https://www.javatpoint.com/machine-learning-models (accessed Aug. 11, 2022).

[9] U. Bole and G. Papa, "Who are the Likeliest Customers: Direct Mail Optimization with Data Mining," *Contemp. Eng. Sci.*, vol. 4, no. 6, pp. 259–268, 2011.

[10] C. Bellinger, S. Sharma, and N. Japkowicz, "One-class versus binary classification: Which and when?," *Proc. - 2012 11th Int. Conf. Mach. Learn. Appl. ICMLA 2012*, vol. 2, pp. 102–106, 2012, doi: 10.1109/ICMLA.2012.212.

References

[11] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 78–85, 2016, doi: 10.1016/j.procs.2016.05.180.

[12] S. Gupta and R. Mamtora, “A Survey on Association Rule Mining in Market Basket Analysis,” *Int. J. Inf. Comput. Technol.*, vol. 4, no. 4, pp. 409–414, 2014, [Online]. Available: http://www.irphouse.com/ijict.htm

[13] J. Woo and Y. Xu, “Market Basket Analysis Algorithm with Map/Reduce of Cloud Computing,” *2011 Int. Conf. Parallel …*, no. April 2012, 2011, [Online]. Available: http://www.lidi.info.unlp.edu.ar/WorldComp2011-Mirror/PDP4494.pdf

[14] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, “An overview of clustering methods,” *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.

[15] T. M. Kodinariya and P. R. Makwana, “Review on determining of cluster in K-means,” *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 90–95, 2013, [Online]. Available: https://www.researchgate.net/publication/313554124

[16] “Understanding K-means Clustering in Machine Learning | by Education Ecosystem (LEDU) | Towards Data Science.” https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 (accessed Aug. 11, 2022).

[17] “What is CRISP DM? - Data Science Process Alliance.” https://www.datascience-pm.com/crisp-dm-2/ (accessed Aug. 10, 2022).

[18] “Understanding Random Forest. How the Algorithm Works and Why it Is… | by Tony Yiu | Towards Data Science.” https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed Aug. 10, 2022).

[19] “What is XGBoost? | Data Science | NVIDIA Glossary.” https://www.nvidia.com/en-us/glossary/data-science/xgboost/ (accessed Aug. 10, 2022).

[20] “A Gentle Introduction on Market Basket Analysis — Association Rules | by Susan Li | Towards Data Science.” https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce (accessed Aug. 10, 2022).

Thank you