Customer Analysis Using Machine Learning Algorithms: A Case Study Using Banking Consumer Dataset

R.Siva Subramanian [a,1], D.Prabha [b], B.Maheswari [c], J.Aswini [d]
ªResearch Scholar, Anna University, Chennai-600025,India
ᵇAssociate Professor, Sri Krishna College of Engineering and Technology,
Coimbatore, India
ᶜAssistant Professor, Rajalakshmi Institute of Technology, Chennai, India
ᵈAssistant Professor, Sree Vidyanikethan Engineering College, Tirupati, India

Abstract. The aim of each enterprise is to achieve high revenue from the business and to stay in a high position from their competitors. To archive high revenue and high position from competitors the need of understanding the business consumers is a crucial one. However the firm business is completely dependent on the consumers the efficient analysis of consumers within the enterprises makes to achieve the business to high position. To perform effective consumer analysis, in this study different machine learning is studied and experimented. ML classifiers make to understand in-depth analysis about the consumer data and further enables to plan wise decision strategies to enhance the business revenue and consumer satisfaction intelligently. The use of different ML classifiers is to sort out how the customer prediction outcome changes accordingly to the ML classifier is applied. This makes to find the best ML classifier for the consumer dataset applied in this study. The experimental procedure is performed using different ML classifiers and the outcome achieved is captured and projected using various validity scores. This work applies different ML classifiers like K-NN, C4.5, Random Forest, Random Tree, LR, MLP and NB for customer analysis. The empirical results illustrate the C4.5 model achieves better accuracy prediction compare to other ML classifiers and also compared with the time complexity NB model works efficiently with running time.

Keywords. Artificial Intelligence, Business decision, Customer   Prediction, Machine Learning

1. Introduction

CRM is an efficient tool that helps to build, manage, and analyze a customer relationship with enterprises. The use of CRM makes the enterprises efficiently collect, store and assess the consumer-related data intelligently and make it available across the enterprise business people [1]. The assessment of consumer data efficiently makes the firms understand in-depth patterns of consumer behavior and further enables them to create wise decision strategies to increases the business revenue and customer retention & acquisition wisely. To perform consumer analysis in CRM, Analytical CRM is

---

[1] R.Siva Subramanian, Research Scholar, Anna University, Chennai-600025, India; E-mail: Email:sivasubramanian12@yahoo.com.

carried out. ACRM helps to analyze the stored consumer data in CRM to find interesting customer patterns which the help of machine learning approaches.

Unlike other CRM, ACRM works behind without having any interaction with the customers and predicts the customer analysis wisely. The use of ACRM helps to increases 1. Marketing technique, 2. Retaining potential consumers, 3. Understanding customer needs, 4. increasing consumer satisfaction, 5. improved consumer loyalty, and 6. Efficient customer experience [2]. In ACRM, ML approaches are applied to perform customer analysis.

This work applies different ML classifiers like K-NN, C4.5, Random Forest, Random Tree, LR, MLP and NB for customer analysis in banking enterprises [3]. The work aims to compute the best ML algorithm for the customer analysis problem and in turn, applies the further decision to improve the firm business. Here customer dataset applied is derived from the UCI repository. The experimental procedure is performed using different ML algorithms and the outcome achieved is indicated using the various validity scores. Further time complexity is measured to find out the running time of the ML classifier. The remainder of the work starts with a literature survey, discussion of ML classifier applied, experimental results, and lastly end with conclusions.

2. Literature Survey

The author addresses the problem of customer churn analysis and proposes LLM model to analyze customer churn [4]. The proposed approach is comprised of two phases: one is segmentation & another one is prediction. The experimental procedure is carried out using 14 customer churn datasets and the outcome captured is projected using AUC and TDL parameters. The outcome obtained is compared using LMT, LR, and RF, and results reveal proposed LLM is better. The author in this work assesses the problem of customer churn and to analyze the customer churn the author proposes four approaches based upon RST: one is Genetic algorithm, two is Exhaustive algorithm, third is Covering algorithm and the last one is LEM2 algorithm [5]. The experimental procedure is carried out using a telecom dataset and the results captured reveal GA-RST performs wisely compare to other approaches. [6] The author address the prediction of consumer patterns helps the firms to plan and allocate the resources efficiently and improve the business wisely [8]. To predict the customer analysis the author proposes some consumer-relevant attributes from previous purchases and use three different ML algorithm to analyzes these data by indoor localization and movement prediction [9]. Experimental results reveal gradient model performs well compared to the other two approaches. The experimental procedure is carried out using 10000 customer data

3. Methodology

The methodology first starts with the input customer dataset and then data processing is carried out to remove the missing and nosily data from the customer dataset. Then different ML classifiers are applied to perform customer analysis using NB, KNN, LR, MLP, C4.5, Random Forest, and Random Tree. Further, the outcome captured is recorded using different performance metrics like accuracy, recall, specificity, FNR,

FPR, F-Measure, and Precision are listed in Table 1. Lastly, the model which archives the best performance is projected. Different ML algorithm applied to perform customer analysis with advantages and disadvantages.
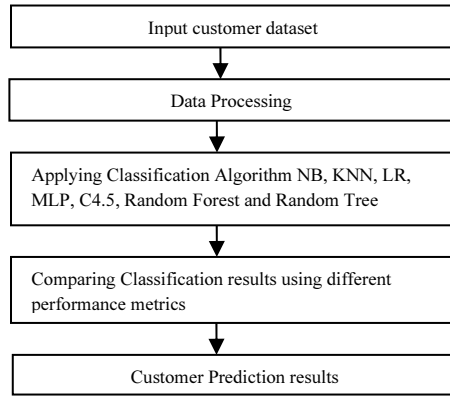
```
┌─────────────────────────────────────────────┐
│            Input customer dataset            │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│               Data Processing                │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│  Applying Classification Algorithm NB, KNN, LR, │
│  MLP, C4.5, Random Forest and Random Tree    │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│  Comparing Classification results using different │
│  performance metrics                         │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│          Customer Prediction results         │
└─────────────────────────────────────────────┘
```

Figure 1. Overall methodology for customer analysis

4. Experimental Results

To analysis about the customer patterns in this study different Ml classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree are studied and experimented.

4.1 Dataset and Validity Scores:

This study applies a Bank dataset derived from UCI for customer analysis purposes. The intent of the consumer data is to predict whether the consumer uses the banking service are not based on the analysis of the customer data. The database comprised of 45211 instances with 17 input predictors [7]. Performance is computed using various metrics like accuracy, recall, precision, f-measure, f-measure, TPR, FPR, and specificity are shown in Figure 2 and Figure 3, respectively.

4.2 Experimental Procedure:

1. The experimental procedure is carried out using the different ML classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree.
2. The empirical outcome captured from the above approaches are represented and compared in terms of accuracy, recall, precision, f-measure, f-measure, TPR, FPR, and specificity.
3. Further the time complexity for the above approaches was also recorded and compared.

Table 1: Represents the outcome of different ML classifiers applied to the problem of customer analysis with a summary of accuracy, recall, precision, of F-measure, FNR, FPR, specificity and time complexity measure.

| S.no	Methodology	Accuracy	Recall	Precision	F-measure	FNR	FPR	Specificity	Time Complexity in seconds
1	NB	88.3789	0.394	0.504	0.442	0.606	0.051	0.9486	0.11
2	LR	89.664	0.319	0.612	0.419	0.681	0.027	0.9732	12.35
3	MLP	89.622	0.427	0.576	0.490	0.573	0.042	0.9584	33.56
4	C4.5	89.8786	0.349	0.620	0.446	0.651	0.028	0.9716	0.64
5	Random Forest	89.1288	0.353	0.556	0.432	0.647	0.037	0.9625	11.6
6	KNN	88.5116	0.237	0.520	0.326	0.763	0.029	0.9709	0.40
7	Random Tree	87.3571	0.352	0.449	0.395	0.648	0.057	0.9426	0.45
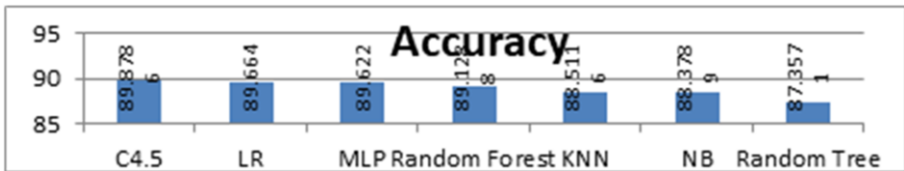
Figure 2. Represents the accuracy comparison obtained from the different ML classifiers
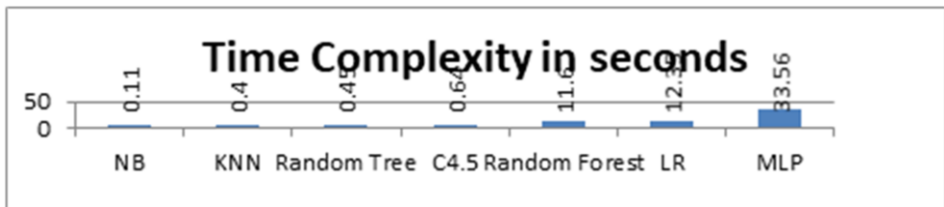

Figure 3. Represents the time complexity comparison obtained from the different ML classifiers

4.3   Result Discussion:

The experimental outcome captured to analyze customer patterns using different ML classifiers is presented in Table 1, it indicates the outcome of different ML classifiers in terms of accuracy, recall, precision, of F-measure, FNR, FPR, specificity and time complexity measure.From the empirical outcome, we can see C4.5 model gets 89.8786 higher accuracy compare to other approaches. The next higher accuracy 89.664, 89.662, and 89.1288 is derived from LR, MLP, and Random Forest approaches. The poor prediction is captured from Random Tree with an accuracy of 87.4588. Next in terms of recall parameter, we can see MLP achieves a high prediction of 0.427. The next higher recall prediction 0.394, 0.391, 0.353 is derived from NB, NBTREE, and Random Forest. Poor recall prediction is captured from KNN with a prediction of 0.237. Next in terms of precision parameter, we can see C4.5 achieves a high prediction of 0.62. The next higher precision prediction 0.612, 0.576, and 0.556 is derived from LR, MLP, and Random Forest. Poor precision prediction is captured from Random Tree with the prediction of 0.449. In terms of F-measure, we can see MLP model 0.49 higher prediction compare to other approaches. The next higher F-Measure prediction 0.446, 0.442, and 0.432 are derived from, C4.5, NB, and the Random Forest approach. The poor prediction is captured from KNN with a 0.326 prediction. Next in terms of the FNR parameter, we can see KNN achieves a high prediction of 0.763. The next higher FNR prediction 0.681, 0.651, 0.68 is derived from LR, C4.5, and Random Tree. Poor FNR prediction is captured from NB and MLP with the prediction of 0.606 and 0.573. Next in terms of the FPR parameter, we can see Random Tree achieves a high prediction of 0.057. The next higher FPR prediction 0.051, 0.042, and 0.037 is derived from NB, MLP, and Random Forest. Poor FPR prediction is captured from LR with the prediction of 0.027. Next in terms of the Specificity parameter, we can see LR achieves a high prediction of 0.9732. The next higher Specificity prediction 0.9716, 0.9709, 0.9625 is derived from C4.5, KNN, and Random Forest. Poor Specificity prediction is captured from Random Tree with the prediction of 0.9426. We can see efficient time complexity is got in NB of 0.11 seconds. The next efficient time complexity is got in KNN, Random Tree, and C4.5 of 0.40, 0.45, and 0.64 seconds. Poor time complexity is got in MLP of 33.56 seconds.

5. Conclusion

The intent of the work is to analyze the customer patterns using different ML classifiers and to sort the wise ML classifier in customer analysis. The analyzes of better customer patterns make to understand the consumer's behavior patterns intelligently and further based upon the analysis the customer satisfaction and business revenue can be increased. To understand customer patterns use of different ML classifier are applied and investigated. This research applies different ML classifiers like NB, LR, MLP, C4.5, Random Forest, KNN, and Random Tree are studied and experimented. The empirical outcomes indicate C4.5 model ends with 89.8786 higher accuracy compare to other approaches. Likewise, poor accuracy is generated in the Random Tree of 87.3571 compare to other approaches. In terms of recall, we can see MLP achieves a high prediction of 0.427 a likewise low recall prediction is captured from KNN with the prediction of 0.237. In terms of precision, we can see C4.5 achieves a high prediction of 0.62 and likewise, less precision prediction is captured from Random Tree with the

prediction of 0.449 precision. In terms of F-measure, we can see the MLP model gets 0.49 high F-measure prediction and likewise less prediction is captured from KNN with 0.326 F-measure predictions. In terms of time complexity, we can see the NB model gets 0.16 seconds best time complexity and likewise less time complexity is captured from MLP with 33.56 Seconds. This research applies only one customer data and in future uses of different customer dataset can be applied.

References

[1] Dalla Pozza, I., Goetz, O., & Sahut, J. M. (2018). Implementation effects in the relationship between CRM and its performance. *Journal of Business Research*, *89*, 391–403. https://doi.org/10.1016/j.jbusres.2018.02.004

[2] Subramanian, R. S., & Prabha, Dr. D. (2017). A Survey on Customer Relationship Management. *2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS 2017*.

[3] Coussement, K., Lessmann, S., & Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*, *95*, 27–36. https://doi.org/10.1016/j.dss.2016.11.007

[4] De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760–772. https://doi.org/10.1016/j.ejor.2018.02.009

[5] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, *237*, 242–254. https://doi.org/10.1016/j.neucom.2016.12.009

[6] Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

[7] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

[8] Ganesh Babu R., Elangovan K., Maurya S., Karthika P. (2021) Multimedia Security and Privacy on Real-Time Behavioral Monitoring in Machine Learning IoT Application Using Big Data Analytics. In: Kumar R., Sharma R., Pattnaik P.K. (eds) Multimedia Technologies in the Internet of Things Environment. Studies in Big Data, vol 79. Springer, Singapore. https://doi.org/10.1007/978-981-15-7965-3_9

[9] C. Jain, G. V. S. Sashank, V. N and S. Markkandan, "Low-cost BLE based Indoor Localization using RSSI Fingerprinting and Machine Learning," Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2021, pp. 363-367, https://doi.org/10.1109/WiSPNET51692.2021.9419388