International Journal of Computing Algorithm
Volume: 9 Issue: 02 12 Dec 2020
ISSN: 2278-2397

Machine Learning Classification Of Active Customers Using Churn

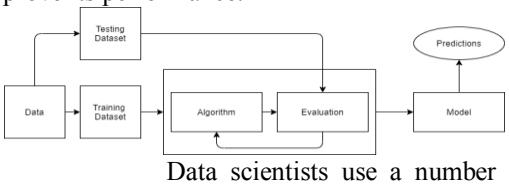
G.GandhiJabakumar[1], R. ArunaDevi[2], Dr.M.RobinsonJoel[3], B.Muthazhagan[4]
[1,2]Department of Computer Science and Engineering, SMK Fomra Institute and Technology,
Tamilnadu.India E-Mail:gandhijabakumar.g@smkfomra.net
E-Mail:arunadevi.cse@smkfomra.net
[3,4]Department of Information Technology,Kings Engineering College,
Tamilnadu.IndiaE-Mail:joelnazareth@gmail.com

Abstract— Client turnover in the banking industry has grown, according to the report. Churn can be classified into a variety of types. It's common knowledge that the cost of acquiring a new client is significantly greater than that of the expense of keeping an existing one. The objective is to find the most accurate machine learning-based churn prediction systems feasible. The entire dataset will be analysed using the supervised machine learning approach (SMLT) to gather a variety of data points, including variable identification, missing value treatments, data validation, data cleaning, and data visualisation.Identify the confusion matrix and categorise data from the supplied credit card dataset, as well as compare and assess multiple machine learning techniques' performance with an evaluation classification report from the given credit card dataset.

Keywords—churn, supervised machine learning technique, data set, credit card.

I.INTRODUCTION

The method of forecasting the future using previous data is known as machine learning. Machine learning (ML) is a type of artificial intelligence that allows machines to understand without having to be explicitly programmed.The building of computer programmes that can adapt to new data, as well as the principles of machine learning, such as the construction of a simple machine learning algorithm in Python, are all covered under machine learning[1]. In the training and prediction phase, specialised algorithms are utilised. It gives the training data to an algorithm[2], which then uses the training data to generate predictions on new test data.Machine learning may be broken down into three categories. Learning may be classified into three categories: supervised, unsupervised, and reinforced. To learn data that must first be labelled by a person, a supervised learning algorithm[3] is given both the input data and the associated labelling. In unsupervised learning, there are no labels. The learning algorithm was given access to it[4]. This technique must figure out how the data in the input is clustered. Finally, reinforcement learning interacts dynamically with its environment and receives positive or negative feedback in order to improve its performance.



Data scientists use a number of machine learning algorithms to find patterns in Python that lead to valuable insights. Based on how they "learn" about data in order to make predictions, these algorithms may be classified into two categories: supervised and unsupervised learning. A approach for predicting the class of provided data objects is classification. Terminologies like objectives, labels, and categories are used to characterise classes. The technique of estimating a transformation matrix from distinct input parameters (X) to discrete independent variables is known as classification predictive modelling (y).Classification [5] is a supervised learning process in statistics and machine learning in which a computer software learns from data input and then applies that learning to categorise fresh observations. It's possible that this data collection is bi-class. Identifying if a person is male or female, for example. As a result, the mail is classified as spam or non-spam. The dataset is likewise dealt with by the multi-class. Choosing whether the individual is male or female, or if the letter is spurious or non-spam, for example. Identifying whether the individual is male or female, or if the letter is phishing or non-spam, for example.A excellent example is speech recognition. Handwritten character recognition, fingerprint recognition, document categorization, and other classification issues are other examples.In the corporate world, churn is a serious issue. The churn rate has been increasing at a quicker rate, and it is the banking department's obligation to regulate and lower it. Because there is such a large amount of turnover data, churn prediction and client identification are key issues in the banking industry. There is a demand for technologies that will allow for speedier case resolution. The situation prompted me to conduct study into how to make a churn scenario easier to address. Machine learning and data science have been shown to make labour easier and faster in several documents and circumstances.

II. RELATED WORK

The latest technological advancements in advanced data analytics and visualisation technologies are assisting society in various ways to examine data of social significance. Churning information of many company sectors is one of these socially significant actions. The churn data analysis will assist decision-making bodies in taking preventative measures to reduce the turnover rate. The findings of several machine learning approaches are shown in the following sections.Asa result, the churn rate is a very high rate in the business and banking sectors. It has created statistical models that use Weighted Moving

Average to predict churn in the next years based on historical customer information from the banking sector. Functional Coefficient Regression is also used to analyse data in the same way. For data analysis, this methodology also employs Arithmetic- Geometric Progression.The accuracy of the suggested methodologies is between 85 and 90 percent based on the gap between actual records and our anticipated values for both years. Furthermore, the approaches of statistical modelling may be used with Machine Learning models to determine weighted accuracy for a specific, making the solution more resilient.

Churn prediction is a significant CRM challenge in today's competitive telecom sector to retain valuable clients by identifying comparable segments of users and providing competitive offers and the services to those categories. For data analysis, Irfan Ullah [6] suggested a customer turnover model. A customer churn model [6] is provided for data analytics in this work, and it is validated using conventional evaluation indicators. Finally, they gave client retention standards for telecom company decision-makers. By using Artificial Intelligence tools for prediction and trend analysis, the research may be expanded to investigate the changing behaviour patterns of churn clients.

Y. Zhang and his team submitted a study [7] that focuses on a novel and fascinating intra-operator customer churn problem. Some customers are switching from 4G to 3G/2G telecoms services. We establish a classification criterion for each customer and make the best choices each moment to accomplish profit-maximizing categorization, which considers the impact of individual variances in monthly payments among clients. Then they award switching ratings to 4G users based on their present switching likelihood. They illustrate the predictability of 4G consumers' switching patterns by constructing a GBDT-Gradient Boosting Decision Tree-based regression model, which lays the groundwork for designing 4G service plan evaluation tests.

A churn analysis was proposed by Eunjo Lee [8]. The goal of churn analysis is to avoid user churn-related losses. As a result, churn prediction is essential to increase forecast accuracy as well as optimise predicted benefits. Their suggested approach has three important characteristics[3]. They begin by defining churn by examining user access behaviours. Second, churn is predicted by identifying long-term committed clients with a high benefit. Finally, they apply cost-benefit analysis to compute the projected profit per user and optimise the prediction model. Finally, they believe that future studies will need to apply the profit estimating approach they employed to study customer attrition in online game businesses.However, the fact that sufficient verification has not been done in reality remains a restriction. They want to carefully test and enhance the proposed approach in future investigations.

They suggest a competitive structure in this paper[9]. Game data was acquired in order to do mining utilising commercial game log data. The researchers supplied a significant time range between the prediction window and the training data, resulting in the shortest time necessary to run the churn avoidance algorithms. Second, to combat idea drift, test sets contain a change in business model. Third, during the competition, they only gave log data from loyal users. They are more difficult to anticipate churn than others, according to our research, but they are more valued in the business.

Using log data, JaehyunAhni [10] presented comparative churn prediction analytic methodologies. Churn analysis is employed in the industries of insurance, gaming, and management, just as it is in the domains of Internet services. The Churn Prediction Models in this research employed deep learning to forecast churn using data timestamps in the seconds or with massive volumes of customer log data. In this scenario, log processing feature engineering approaches have a considerable impact on model performance improvement. By using a layerwise stacked neurons structure, the deep learning model can also learn consumer behavioral patterns from large amounts of data.Given minute timestamps and a large number of observations, combining this data to deep learning algorithms for the development of latent characteristics should outperform traditional churn prediction methods.

P.N.V.V.Prasad Babu Gowd [11] devised a technique that translates voice formatted buyer feedback into text format. The voice recognition module was used to convert the files. The reviews of customers are kept in the cloud and fed into the emotional analysis domain to generate neutral, positive, and negative reviews. These evaluations assist shoppers in locating and deciding whether or not to purchase a product. He primarily advocated a voice or speech-based methodology for gathering speech reviews or comments, with the data being put into a mining algorithm. They delegated the task of determining the worth or weightage of each and every word to machine learning.

Sasikala, P [12] introduced a novel sentiment analysis of online product customer evaluations called DLMNN, which stands for Deep learning modified neural network. They also proposed the IANFIS (Improved adaptive neuro-fuzzy inference system) approach for predicting online items. They compare the results of both approaches and come up with a positive conclusion. These Deep learning customised neural networks were presented for

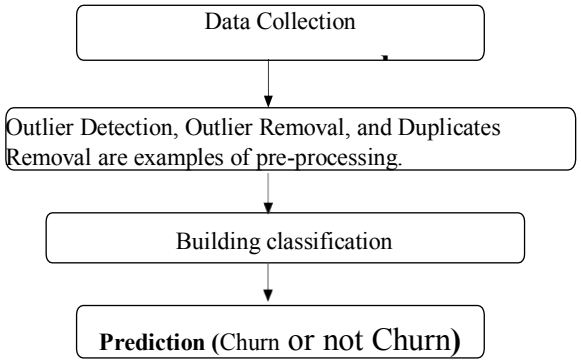
International Journal of Computing Algorithm
Volume: 9 Issue: 02 12 Dec 2020
ISSN: 2278-2397

customer review analysis situations such as Content-based, Grade-based, and Collaboration-based.Singla Z [13] focuses on the many features and evaluations of mobile phones available on e-commerce platforms. They use the dataset to collect both user review data and the manufacturer's perspective on the product. They analyse consumer sentiment for three mobile brands: Samsung, Apple, and BLU.

III. PROPOSED SYSTEM

The dataset is now sent to the machine learning model, which is trained using this data set. Data from previously current datasets from internet sources is obtained in the first phase of the information gathering process. These datasets are combined to generate a single dataset that will be analysed.

The goal would be to put a prediction model to the test. The training data set would be used, and the test dataset would be used to validate the training. Depending on the accuracy, a better algorithm will be used to build the model. Churn prediction will be done using supervised classification and other algorithms. The dataset is shown in order to examine any churns that may have happened in the banking industry. This facilitates the completion of additional formalities by all other departments. By comparing algorithms using python code, it must establish the training dataset's accuracy, the testing dataset's accuracy, the specification, the False Positive rate, precision, and recall, among other things. The following phases in the Involvement process will help you define an issue. Similarly, data preparation for analysis. After that, we must evaluate training methods. It is always enhancing the outcomes. Predicting outcomes works in the same way.



Predictive modelling is a technique for creating a model that can make predictions. A machine learning algorithm is used in the process to learn particular qualities. The prediction of a bright, wet, or snowy day might be a pattern classification challenge in weather forecasting. There are two types of pattern categorization tasks that is supervised and also unsupervised learning.The procedures for creating the data model are outlined below.

Figure 3.1 Proposed System Architecture

The goal of this study is to identify which characteristics are most useful in forecasting churn rate, as well as general patterns that may aid with model and hyper parameter selection. The purpose is to determine if the churn rate is high or low, which will aid in recognising client turnover in banking. To accomplish this, machine learning classification methods were utilised to train a function that can predict the discrete class of fresh churn input data. Every organisation in today's world has ever-increasing issues that must be met swiftly and efficiently. With ever-increasing client numbers, assessing credit card data is a major challenge for banking in terms of making strategic decisions to retain customer growth. This is critical in order to retain clients. The enormous raw data that is created on a regular basis from numerous sources by integrating Data Science with the machine learning idea is the greatest area to look up to find space for development.

The goal of this research is to use machine learning to explore a dataset of customer records from the banking industry. It's more difficult to pinpoint churns in each location. To detect churn issues, machine learning can be employed. In numerous aspects, churn is a big disadvantage, because recruiting new consumers is far more expensive than keeping existing customers.

The suggested approach is capable of detecting churning clients early on. It identifies the causes of churn in order to prevent client loss and suggests ways to keep them. Multiple techniques are utilised to address this restriction, and the best classifier model is chosen for retention. It may assist businesses in identifying, predicting, and retaining churning consumers, as well as aiding decision-making and CRM. On a big dataset from the banking industry, we employed a variety of machine learning methods to classify churn and non-churn. When compared to other machine learning algorithms, the Random Forest (RF) approach exhibited a higher accuracy of 96.4 percent.

The first line imports the iris data set, which is previously specified in the sk learn module, and raw data set, which is just a table containing information on various types. For example, you may use the sk learn and NumPy module to import any algorithm and train test split class for usage in this application. The dataset variable is used to encapsulate the load data () method. Using the train test split approach, further partition the dataset into training and test data. The feature values are denoted by the X prefix in variable, whereas the target values are denoted by the y prefix. This approach divides the dataset into training and test data in a 77:23 / 80:20 ratio at random.

Then any algorithm is encapsulated. We fit our training data into this method in the next line so that the computer may learn from it. The training phase is now complete. Now, the dimensions of new features are stored in a NumPy array named 'n,' and the goal is to forecast the species of these characteristics using the predict function, which accepts this array as input and returns the projected target value as output. As a result, the expected goal value is 0. Finally, to calculate the test score, which is the ratio of the

number of right predictions to the total number of predictions made, and to calculate the accuracy score, which compares the actual values of the test set to the projected values.

IV. CONCLUSION AND FUTURE WORK

Data cleaning and preprocessing, missing assessment, eventually determination, and model creation and assessment were all part of the analytical method. The greatest accuracy score, as well as the best correctness on a public test set, will be identified.This leads to some of the following turnover rate findings. Finding relationships and patterns among varied data have gotten easier. It primarily focuses on anticipating the sort of churn that may occur if we know where it occurred in the credit card department. We created a model utilizing the principles of machine learning and a training data set that has undergone data cleansing and transformation. Data visualisation resulted in several graphs and the discovery of intriguing statistics that aided in the comprehension of customer turnover datasets that may aid in capturing the aspects that can aid in keeping consumers secure. Based on the churn rate of regions, the credit card department wishes to automate the detection of churn from the eligibility procedure (in real-time). This method may be automated by displaying the prediction result in a web or desktop application.This study can be improved in the future by applying Machine Learning models to forecast churn, since the number of data points will be adequate to apply ML models and improve the accuracy of theforecasts.

REFERENCES

[1] A. C. Bahnsen, D. Aouada, and B. Ottersten, “A novel cost-sensitive framework for customer churn predictive modeling,” Decis. Anal., vol. 2, no. 1, pp. 1–15,2015.

[2] C. P. Wei and I.-T. Chiu, “Turning telecommunications call details to churn prediction: A data mining approach,” Expert Syst. Appl., vol. 23, no. 2, pp. 103–112,2002.

[3] D. J. Hand, “Measuring classifier performance: A coherent alternative to the areaundertheROCcurve,”Mach.Learn.,vol.77,no.1,pp.103–123,2009.

[4] Infiniti Research Limited, “Global online gaming market 2014,”2014.

[5] M. S. El-Nasr, A. Drachen, and A. Canossa, Game Analytics. Berlin, Germany:Springer,2016.

[6] Irfan Ullah, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, SailUl Islam, Sung Won Kim, “A Churn Prediction Model Using RandomForest: Analysis of Machine Learning Techniques for Churn Prediction andFactor Identification I Telecom Sector”, Digital Object Identifier 10.1109/ACCESS.2019.2914999.

[7] Zhang, Y., He, S., Li, S., & Chen, J. (2020). Intra-Operator Customer Churn in Telecommunications: A Systematic Perspective. *IEEE Transactions on Vehicular Technology, 69*, 948-957.

[8] Eunjo Lee, Boram Kim, Sungwook Kang, Byungsoo Kang, Yoonjae Jang, Huy Kang KimProfit Optimizing Churn Prediction for Long-term Loyal Customer in Online games *IEEE Transactions on Games* (IF1.851), Pub Date : 2020-03-01, *DOI: 10.1109/tg.2018.2871215*

[9] Eunjo Lee, Yoonjae Jang, Du-Mim Yoon, Jihoon Jeon, Seong-il Yang, Sang-Kwang Lee, Dae-Wook Kim, Pei Pei Chen, Anna Guitart, Paul Bertens, Africa Perianez, Fabian Hadiji, Marc Muller, YoungjunJoo, Jiyeon Lee, Inchon Hwang, Kyung-JoongKim,“Game Data Mining Competition on Churn Prediction and Survival Analysis using Commercial Game Log Data”,*IEEE Transactions on Games* (IF1.851), Pub Date : 2019-09-01, *DOI: 10.1109/tg.2018.2888863*

[10] J. Ahn, J. Hwang, D. Kim, H. Choi and S. Kang, "A Survey on Churn Analysis in Various Business Domains," in IEEE Access, vol. 8, pp. 220816-220839, 2020, doi: 10.1109/ACCESS.2020.3042657.

[11] M.Robinson Joel ,Varaprasad,“User Satisfaction Assessment Extraction From Vocal Reviews Using Speech Recognition”, International Journal Of Information And Computing Science ISSN NO: 0972-1347 Volume 5, Issue 3, March 2018.

[12] Sasikala, P., Mary Immaculate Sheela, L. Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. *J Big Data* **7,** 33 (2020). https://doi.org/10.1186/s40537-020-00308-7

[13] Singla Z, Randhawa S, Jain S. Statistical and sentiment analysis of consumer product reviews. In 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6, 2017.