IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11708

A SURVEY ON PREDICTIVE ANALYSIS FOR CUSTOMER CHURNING

Deekshith Reddy, Idris Malik, Shashank Gattu, Kamalesh N, O Sachin

Dept. of AIML, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka

Abstract: Customer churn is one of the subscription-based business critical tasks that a company needs to make a decision about revenue stream management and to take care of their customers from churning. This research work is an introduction to a machine-learning method for customer churn analysis using predictive models. The process starts with a vast customer transaction dataset, which needs to be transformed into churn labels. Next, the system utilizes several machine learning algorithms, including logistic regression, decision tree, random forest, support vector machines (SVM), and gradient boosting, to process the input data and design predictive models. Carrying out feature selection and feature construction is part of the process. Feature selection is a method used to reduce the input of the dataset that might conflict with the output. Feature construction will unfortunately be a million-dollar question. Accuracy, precision, recall, and F1 scores measure model performances. In addition, the ROC curve can be obtained for a designed model.The findings demonstrate how accurate and efficient the proposed method can be for a customer churn problem. An organization gets an early warning about a customer churn problem using this method. It will put a customer in the retention consideration set. In addition, a design model gives an organization a reason behind a customer churn. This analysis will help organizations understand the cause of Churn and decide what they will do before a customer leaves.

Keywords: Customer Churn Prediction, Machine Learning, Predictive Modeling.

1. INTRODUCTION

1.1. Customer churn

The customer who terminates or stops using some type of service at some point in time is said to be a churner. Customer churn is of two types – one is active churn and the other one is passive churn.

Active churn usually occurs because customers intentionally decide to disconnect the service by themselves, whereas Passive churn occurs because the customer has not removed the connection of the service, but is not using it regardless of the reason and caused by the factor that is outside the control of the customer.

Customer churn prediction is a way of analysis that intends to predict an existing chance of a consumer or a customer who wants/uses it to terminate the services. There could be several reasons for customer churn, the purpose of these predictive models is to help mitigate the possibility.

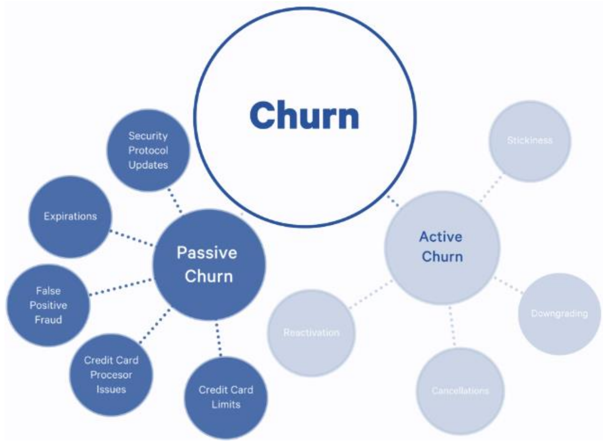

Fig 1.2. Types of Churns

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11708

1.2. Importance of Churn Analysis

Churn Analysis is probably the most important analysis in business in our modern age. It is extremely important in the insurance, telecommunications, or financial industries, for example, which are based on a subscription business model. According to various studies examining the acquisition of new customers, nowadays acquiring a new customer is five to ten times more expensive than retaining an existing one.. It is an analytical method, for example, for identifying existing customer profiles, and customer churn and turnover analysis. The value of a company is largely dependent on the number of active customers. For various reasons, the above mentioned values, which include the costs of companies, profitability, size or investment capacity, and cash flow are dependent on the number of customers and can thus be influenced by their loyalty. A high number of active customers also leads to increased referrals. In addition, studies have shown that long-term customers are more profitable. Methods such as Customer Lifetime Value (CLV/CLTV) are used to understand this return.

There are several reasons as to why Churn analysis is important. The following are a few that have the most impact on maintaining the stability of the business:

1.2.1. Customer Retention

Churn Analysis provides insight into understanding the cause of customer churn which helps the companies to make well-informed decisions and devise strategies to retain customers. By identifying patterns, behaviors, and reasons behind churn, businesses can take proactive measures to prevent it, prompting the company to improve product/service quality, enhance customer support, or offer incentives to stay.

1.2.2. Revenue Impact

Churn has a direct impact on a company's revenue and profits. Customer churn leads to the loss of revenue streams which provide recurring funds for the functionality of a company. Churn analysis helps companies stabilize their customer base, maintain a steady revenue and improve profitability.

1.2.3. Competitive Advantage

Employing churn analysis helps a company have an advantage over its competitors by being more attentive to the needs of a customer and their preferences. This leads to good quality of service which may help the company gain new customers through word of mouth, hence maintaining market leadership.

1.2.4. Predictive Analysis

Predictive analysis is a process of predicting future trends of events using machine learning and statistical algorithms. Predictive analysis models are designed to access historical data and discover patterns. It helps in operation efficiency, improved decision-making, and risk reduction. There are many types of predictive analysis models, here we use classification and regression models.

Fig 1.2.4. Steps for Predictive Modeling

DOI: 10.17148/IARJSET.2024.11708

Churn analysis uses predictive models to forecast the possible customer churn, thereby facilitating the companies to take preventive measures following the patterns and behaviours exhibited by the customers. Using Predictive models in Churn analysis promotes companies to efficiently allocate resources by focusing on customers that are likely to churn. Predictive models enable companies to adopt a customer-centric approach which helps in building trust and forming meaningful relationships with their customers.

2. APPLIED AREAS

Predictive analysis is applied in various industries to reduce customer churning. Here we have elaborated on how predictive analysis is applied in the few areas mentioned below.

2.1. Telecommunication

Customer churn is a major problem in the telecommunication industry. Customers may cancel their subscriptions or change their service provider if they are dissatisfied with their existing service provider. Machine learning models are applied to predict churn on an individual customer basis and take countermeasures by implementing targeted retention strategies such as personalized offerings, various discounts, offers, and better services. Thus, helping them to retain their existing customers and exploring new customers based on their reviews. By understanding the main reasons that cause churn and implementing effective retention techniques, telecom companies can reduce customer attrition and maintain a healthy customer base.

2.2. Hotels and Hospitality

Using this method hotels use historical customer data and predict the reasons why churning might churn and reform them. Predictive analysis is used in the hospitality industry to track the customer's previous purchase behaviour and design the hotel services accordingly thus retaining old customers and attracting new ones which is also the main goal.

2.3. Banking and Finance

Nowadays, the market for credit card services is very competitive due to the vast amount of service providers, especially banks, around the world. One of the biggest challenges that banks must deal with is the change in customer behaviour. Customers are at the heart of all industries, particularly banks where customers are responsible for most of their workings such as deposits, investments, taking loans, and using services provided by the banks. Banks need to avoid losing customers to their competitors to maintain stable income and profit, which is where predictive tools that operate on machine learning techniques and data analysis come into play.

Predictive analysis uses customer behaviour such as durations between transactions, the amount used in these transactions, customer feedback, historical data, and various other means to analyze and predict customer churn. It applies machine learning algorithms to identify the factors and indicators that account for customer churn. These factors include less usage of services, late payments, or changes in spending habits. Using this data, the model assigns a score to each customer which will indicate the probability of churning. The banking system can use these scores to avoid gaining risky customers and improve their existing system by providing suitable services to their customers to maintain or maximize profit.

3. METHODOLOGIES

The objective of utilizing machine learning to predict customer churn is to create a predictive system that can forecast customer attrition. This section will cover the complete process of developing a predictive model and provide detailed explanations for each mentioned applied area in section 2.

There have been various predictive analytic models developed through many studies, here we elaborate on the two types of models that we use to prevent customer churn.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11708

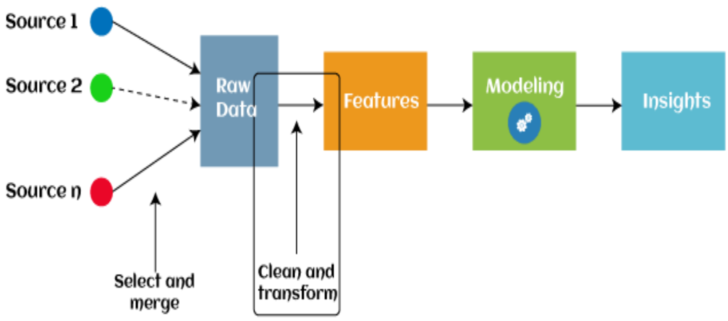

Fig 3. Flow diagram

3.1. Understanding the problem and final goal

A comprehensive grasp of the issue at hand and the establishment of a clear objective are essential for the creation of the predictive model. Grasping the core insights one wishes to gain from the analysis is crucial. It's necessary to identify the right question to pose and determine which variety of machine learning model will yield the desired results.

3.1.1. Classification

The fundamental goal of classification is to ascertain the class or category to which a customer pertains. For such problems, historical data, along with predetermined factors and outcomes, are utilized to identify whether a customer will churn. Through this approach, it becomes possible to address queries such as whether the customer will leave the service or continue using it.

Several algorithms can be used to classify the data, including logistic regression, decision trees, random forests, support vector machines (SVM), k-nearest neighbours (KNN), and neural networks.

Depending on the specific characteristics of the problem and desired performance criteria, the classification models are evaluated based on accuracy, precision, recall, and ROC-AUC.

3.1.2. Regression

Forecasting customer churn is often viewed as a regression task. Regression is employed to examine how a target variable is influenced by other variables, aiming to understand their relationship. Unlike classification, where the result is a category, regression outcomes are numerical values derived from the input data. This approach enables the prediction of not just if, but when a customer might discontinue a service. Techniques such as linear regression, ridge regression, and gradient boosting regressors are commonly used for estimating the likelihood of churn. Additionally, regression can be applied to calculate the churn rate, indicating the percentage of customers anticipated to leave the service within a given period. In this survey, we'll be discussing the mechanisms of predictive models that use regression models to predict churn.

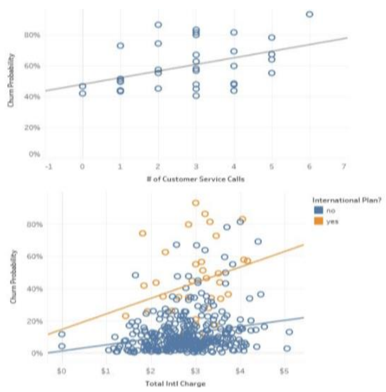

Fig 3.1.2. Regression graph

3.2. Data Collection

Once we have identified and understood the problem, we can decide what data sources are required for predictive modeling. The most common data sources we need to build the predicting model include internal databases, third-party data providers, APIs, web scraping, surveys, or other sources. The acquired data can either be structured (organized in a table) or unstructured (text or images).

In regards to the predictive models developed for the telecom industry, the data can be acquired from various sources, a few such sources are Customer data, Towers and complaints database, Network logs data, Call details records, and Mobile IMEI.

Sources for gathering data regarding Hotels and Hospitality include Hotel booking systems, CRM databases, Customer satisfaction surveys, Guest feedback platforms, Social media interactions, Historical transaction data, Demographic information, Market trends, and competitive analysis.

Potential data sources relevant to the churn prediction model in Banking and Finance include Transactional data, Customer demographic data, Customer interaction data, and External data.

Depending on the sources, collecting data may involve accessing databases, APIs, flat files, or even manual data entry. Thereby integrating data from various sources into a unified dataset for analysis.

3.3. Data preparation and preprocessing

The collected historical data needs to be converted into a format that is compatible with machine learning. The first task with the transformed dataset is to check that the observed values are correct. Incorrect data could lead to larger problems, such as the model inaccurately representing its success rate. A check is done whether the variables are within min-max control and logical. For example, it’s not logically possible that age is negative. When it’s confirmed the observed values are correct, any missing data is handled at first as no mathematical operations are possible without full information. Further, to make variables independent of one another, only one of any related variables is included in the model. Not doing so will lead to a problem called multiple connections, and a model becomes invalid.

3.4. Feature Engineering

In this modern transforming world of artificial intelligence and Machine learning we use machine learning feature engineering which gives us a set of models of tools.

Feature engineering involves selecting and creating relevant variables from existing data to enhance model performance. This process may include transforming raw data, creating new features, and selecting the most predictive ones. Features could encompass customer demographics, behavior patterns, usage frequency, and satisfaction metrics. For instance, variables like customer tenure, purchase frequency, complaints history, and engagement levels can be vital indicators. By crafting meaningful features, predictive models can better discern patterns indicative of potential churn, empowering businesses to proactively address customer retention strategies and mitigate churn risks.

3.5. Model selection

Selecting the appropriate model is paramount in ensuring that the most effective solution is applied to the given task. This involves carefully evaluating various models based on their suitability, performance, and adaptability to the specific requirements of the task. The right choice not only enhances the accuracy and efficiency of the solution but also significantly impacts the success of the project by addressing the problem in the most effective manner possible.

3.5.1. Model Selection Techniques:

We select the model for the given problem, considering factors like complexity, performance metrics, and generalization ability. Techniques include cross-validation, grid search, and information criteria, aiming to strike a balance between bias and variance in predictive modeling. Knowing the techniques like hyperparameters tuning.

3.5.2. Overfitting and Underfitting:

Model selection is influenced by the ideas of overfitting and underfitting. Overfitting occurs when a model learns from the noise rather than the fundamental structure, resulting in suboptimal performance on new data. Underfitting arises

DOI: 10.17148/IARJSET.2024.11708

when a model is overly simplistic, failing to grasp the underlying structure, and thus performing poorly even on the training data. Both issues can diminish the model's ability to generalize.

3.5.3. Ensemble Methods:

In this approach, we combine different models to enhance their prediction accuracy and error tolerance. Methods such as bagging, boosting, and stacking utilize an assortment of models to reduce errors and bolster overall performance, rendering ensemble methods appealing in the realm of machine learning. They stand out due to their capacity to handle intricate data and a variety of situations.

4. CONCLUSION

At present, customer satisfaction is a priority for firms that wish to remain profitable and meaningful. Firms must identify customers who are unhappy to change adaptive systems that align with their preferences as ways of reducing churn factors. Predictive analysis is therefore a vital tool in this endeavour by giving insights about churn rates especially useful in subscription-based businesses.

On the other hand, predictive analysis applies the latest machine learning algorithms and data analytics to delve into customer churn trends surpassing alternative methods through the use of inclusive and accurate datasets. This method not only identifies potential churners but also aids in taking preventive measures aimed at effectively retaining customers.

In essence, blending machine learning, good data, and advanced tools has far-reaching impacts across various fields hence consistently contributing to improvements while transforming research or business as well. Organizations can therefore employ these technologies performantly through which they can be proactive in tackling challenges, and optimizing operations thereby maintaining competitiveness within an ever-changing landscape.

REFERENCES

[1] Ozer Celik and Usame O. Osmanoglu (2019), Comparing Techniques Used in Customer Churn Analysis, Journal of Multidisciplinary Developments. 4(1), 30-38, 2019.
[2] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. Int Res J Eng Technol. 2016;3(4):1065–70.
[3] Abdelrahim Kasem Ahmad, Assef Jafar and Kadan Aljoumaa. Customer churn prediction in telecom using machine learning in a big data platform. Journal of Big Data. (2019) 6:28.
[4] Nettleton, D. (2014). Commercial data mining: processing, analysis, and modeling for predictive analytics projects. Elsevier.
[5] Poel, V. D., Lariviére (2004). Customer Attrition Analysis for Financial Services Using Proportional Hazard Models. European Journal of Operational Research.
[6] Altexsoft, Customer Churn Prediction Using Machine Learning: Main Approaches and Models, By KDnuggets.
[7] Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley, Customer churn prediction in telecommunications.