International Journal of Emerging Technology and Innovative Engineering
Volume 6, Issue 03, March 2020 (ISSN: 2394 – 6598)

A Proficient CRM-Machine Learning Framework for The Prediction of Customer Behaviour

Nithila B[1], Ramya R[1], Vinothini M[1], K. Abinaya[2]
[1]Department of IT, Sri Krishna College of Technology, Coimbatore.
[2]Assistant Professor, Department of IT, Sri Krishna College of Technology, Coimbatore.

ABSTRACT- In CRM (customer relationship management), CRM software is a category of software that covers a broad set of applications designed to help businesses manage many of the following business processes: customer data. customer interaction. access business information. CRM system examples include platforms built to manage marketing, sales, customer service, and support, all connected to help companies work more effectively. With a CRM system, businesses can analyse customer interactions and improve their customer relationships. The information-based prediction models using machine learning techniques have gained massive popularity during the last few decades. Such models have been applied in a number of domains such as medical diagnosis, crime prediction, movies rating, etc. Similar is the trend in telecom industry where prediction models have been applied to predict the dissatisfied customers who are likely to change the service provider. Due to immense financial cost of customer churn in telecom, the companies from all over the world have analysed various factors (such as call cost, call quality, customer service response time, etc.) using several machine learning techniques. This work proposes various ML techniques for customer churn prediction.

KEYWORDS: Telecom, Churn, Data Analytics, Classification

1. INTRODUCTION

Current times are the age of competition among businesses and similar is the trend in communication industry. There is an immense competition among the mobile operators, due to which, the telecommunication industry faces difficulties to retain their current subscribers. Customer churn prevention is a major part of the Customer Relationship Management (CRM). Churn describes the customers who have terminated the relationship with their current service provider on the basis of dissatisfaction. The recent advancement in analysis of information systems methods, particularly in use of the information-based prediction methods, has also inspired the telecommunication industry. Telecom companies invest huge infrastructure and money to implement churn prediction models to find the possible churners before their customers decide to change the service provider. Churn prevention is an important factor as the cost for customer acquisition is much greater than the cost of customer retention. Hence, little improvements in customer retention can lead to major profit for the telecom company. There is a need for accurate churn prediction models to identify the most likely churners and their reasons to churn. In order to prevent customer churn in the telecom industry, it is important to build an effective customer churn prediction model. A number of churn prediction models have been proposed and implemented, which intend to identify customers with a high tendency to leave.

2. LITERATURE SURVEY

Many approaches were applied to predict churn in telecom companies. Most of these approaches have used machine learning and data mining. The majority of related work focused on applying only one method of data mining to extract knowledge, and the others focused on comparing several strategies to predict churn. Gavril et al. presented an advanced methodology of data mining to predict churn for prepaid customers using dataset for call details of 3333 customers with 21 features, and a dependent churn parameter with two values: Yes/No. Some features include information about the number of incoming and outgoing messages and voicemail for each customer. The author applied principal component analysis algorithm "PCA" to reduce data dimensions. Three machine learning algorithms were used: Neural Networks, Support Vector Machine, and Bayes Networks to predict churn factor. The author used AUC to measure the performance of the algorithms. The AUC values were 99.10%, 99.55% and 99.70% for Bayes Networks, Neural networks and support vector machine, respectively. The dataset used in this study is small and no missing values existed. He et al. proposed a model for prediction based on the Neural Network algorithm in order to solve the problem of customer churn in a large Chinese telecom company which contains about 5.23 million customers. The prediction accuracy standard was the overall accuracy rate, and reached 91.1%.

Idris proposed an approach based on genetic programming with AdaBoost to model the churn problem in telecommunications. The model was tested on two standard data sets. One by Orange Telecom and the other by cell2cell, with 89% accuracy for the cell2cell dataset and 63% for the other one. Huang et al. studied the problem of customer churn in the big data platform. The goal of the researchers was to prove that big data greatly enhance the process of predicting the churn depending on the volume, variety, and velocity of the data. Dealing with data from the Operation Support department and Business Support department at China's largest telecommunications company needed a big data platform to engineer the fractures. Random Forest algorithm was used and evaluated using AUC. Makhtar et al. proposed a model for churn prediction using rough set theory in telecom. As mentioned in this paper Rough Set classification algorithm outperformed the other algorithms like Linear Regression, Decision Tree, and Voted Perception Neural Network. Various researchers studied the problem of unbalanced data sets where the churned customer classes are smaller than the active customer classes, as it is a major issue in churn prediction problem. Amin et al. compared six different sampling techniques for oversampling regarding telecom churn prediction problem. The results showed that the algorithms (MTDF and rules-generation based on genetic algorithms) outperformed the other compared oversampling algorithms.

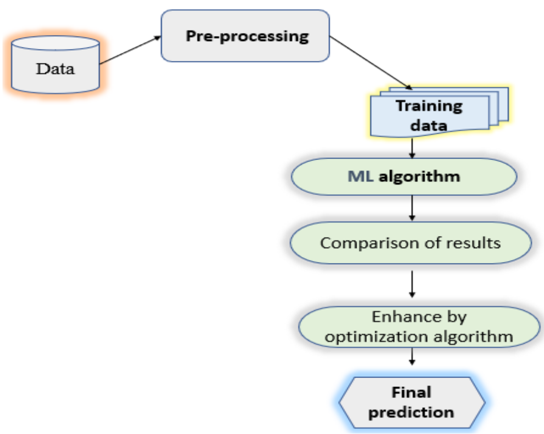

Fig.1.1 Block diagram of proposed framework

3. PRE-PROCESSING

3.1. LABEL ENCODING

In machine learning, we usually deal with datasets which contains multiple labels in one or more than one column. These labels can be in the form of words or numbers. To make the data understandable or in human readable form, the training data is often labelled in words. Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning. Label encoding convert the data in machine readable form, but it assigns a unique number (starting from 0) to each class of data. This may lead to the generation of priority issue in training of data sets. A label with high value may be considered to have high priority than a label having lower value.

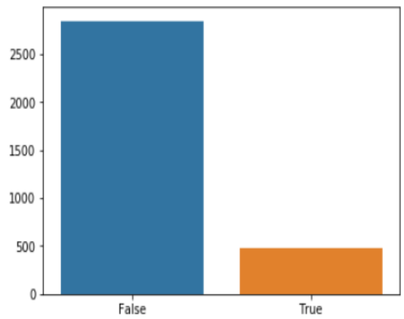

Fig.1.1. Imbalanced Data of Churn

4. MODEL DESIGNING

A **machine learning model** can be a mathematical representation of a real-world process. The **learning** algorithm finds patterns in the training data such that the input parameters correspond to the target. The output of the training process is a **machine learning model** which you can then use to make predictions. **Training** a **model** simply means learning (determining) good values for all the weights and the bias from labelled examples. In supervised learning, a machine learning algorithm builds a **model** by examining many examples and attempting to find a **model** that minimizes loss; this process is called empirical risk minimization.

4.1. SUPPORT-VECTOR MACHINE

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

4.2. LOGISTIC REGRESSION

Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent (*target*) and independent variable (s) (*predictor*). This technique is used for forecasting, time series modelling and finding the causal effect relationship between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression. Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X. Logistic Regression is one of the most popular ways to fit models for categorical data, especially for binary response data in Data Modelling. It is the most important (and probably most used) member of a class of models called generalized linear models. Unlike linear regression, logistic regression can directly predict probabilities (values that are restricted to the (0,1) interval); furthermore, those probabilities are well-calibrated when compared to the probabilities predicted by some other classifiers, such as Naive Bayes. Logistic regression preserves the marginal probabilities of the training data. The coefficients of the model also provide some hint of the relative importance of each input variable. Logistic Regression is used when the dependent variable (target) is categorical. Logistic regression is used in various fields, including machine learning, most medical fields, and social sciences. For e.g., the Trauma and Injury Severity Score (TRISS), which is widely used to predict mortality in injured patients, is developed using logistic regression. Many other medical scales used to assess severity of a patient have been developed using logistic regression. Logistic regression may be used to predict the risk of developing a given disease (e.g. diabetes; coronary heart disease), based on observed characteristics of the patient (age, sex, body mass index, results of various blood tests, etc.).

4.3. K-NEAREST NEIGHBORS

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful. KNN captures the idea of similarity (sometimes called distance, proximity, or closeness) with some mathematics we might have learned in our childhood— calculating the distance between points on a graph. To select the K that's right for your data, we run the KNN algorithm several times with different values of K and choose the K that reduces the number of errors we encounter while maintaining the algorithm's ability to accurately make predictions when it's given data it hasn't seen before. The algorithm is simple and easy to implement. There's no need to build a model, tune several parameters, or make additional assumptions. The algorithm is versatile. It can be used for classification, regression, and search. The algorithm gets significantly slower as the number of examples and/or predictors/independent variables increase.

4.4. RANDOM FOREST

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

It overcomes the problem of overfitting by averaging or combining the results of different decision trees. Random forests work well for a large range of data items than a single decision tree does. Random forest has less variance then single decision tree. Random forests are very flexible and possess very high accuracy. Scaling of data does not require in random forest algorithm. It maintains good accuracy even after providing data without scaling. Random Forest algorithms maintains good accuracy even a large proportion of the data is missing. Complexity is the main disadvantage of Random forest algorithms. Construction of Random forests are much harder and time-consuming than decision trees. More computational resources are required to implement Random Forest algorithm. It is less intuitive in case when we have a large collection of decision trees. The prediction process using random forests is very time-consuming in comparison with other algorithms.

5. RESULTS AND DISCUSSION

Table .1.1. Results and Summary of various classifiers

Classifier	Precision	Recall	F1-score	Accuracy
SVM	0.91	0.92	0.91	92%
Logistic Regression	0.83	0.86	0.83	86%
KNN	0.88	0.89	0.87	89%
Random forest	0.94	0.94	0.94	94%
Gradient Boosting	**0.95**	**0.95**	**0.95**	**95%**

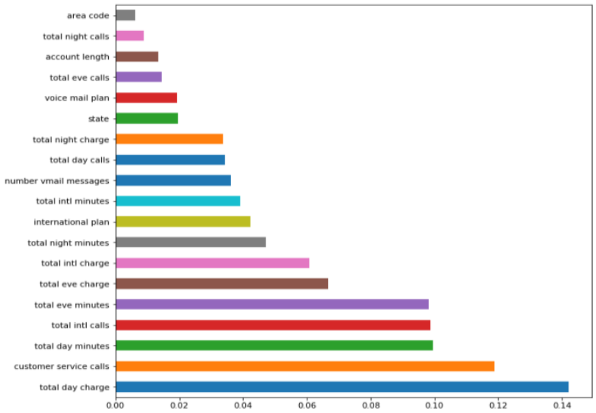

Fig.1.1. Feature Selection by Gradient Boosting

6. CONCLUSION

Data Analytics is the procedure of retrieve a pattern from large data set in connection with machine learning, data base, and statistics. Machine learning can be a very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases. The importance of this type of research in the telecom market is to help companies make more profit. It has become known that predicting churn is one of the most important sources of income to telecom companies. The patterns found from data can be used to make better decisions, identify risks and opportunities for future. The decrease in result could be due to the non-stationary data model phenomenon, so the model needs training each period of time. The use of the Social Network Analysis features enhances the results of predicting the churn in telecom.

REFERENCES

[1] RL Babu, S Vijayan,(2016) “Wrapper based feature selection in semantic medical information retrieval”, Journal of Medical Imaging and Health Informatics, 6 (3), pp.802-805.

[2] RA Suji, DB Anand, RL Babu,(2019) “An Efficient Image Enhancement Technique using

Stationary Wavelet Transform Based Image Fusion", International Journal of Recent Technology and Engineering 8 (1), pp.2277-3878.

[3] RA Suji, DB Anand, RL Babu,(2019) "Polar ice image segmentation using improved estimation and normalization of illumination", Cluster Computing 22 (1), pp.S3811–S3819.

[4] RL Babu, S Vijayan,(2013) "SEMANTIC BASED INFORMATION RETRIEVAL FOR WEB PAGES", Journal of Theoretical & Applied Information Technology 54 (1)

[5] RA Suji, DB Anand, RL Babu,(2019) "An Efficient Image Enhancement Technique using Stationary Wavelet Transform Based Image Fusion", International Journal of Recent Technology and Engineering 8 (1), pp.2277-3878.

[6] RA Suji, DB Anand, RL Babu,(2019) "Polar ice image segmentation using improved estimation and normalization of illumination", Cluster Computing 22 (1), pp.S3811–S3819.

[7] Adaline Suji, R, Saravanan, R & Bright Anand, D 2017, 'ImageFusion-Novel Approach of Contrast Enhancement by Discrete Shearlet Transform (DST)',Journal of Computational and Theoretical Nanoscience, vol. 14, no. 3, pp. 1282-1287

[8] Bright Anand, D, Saravanan, R & Adaline Suji, R 2017, 'Adaptive Batch Mode Active Learning Technique Usingwith an Improved Time Adaptive Support Vector Machine for Classification of Remote Sensing Applications', Journal of Computational and Theoretical Nanoscience, vol. 14, no. 2, pp. 1108-1113.

[9] Adaline Suji, R, Saravanan, R & Bright Anand (2015)"A Contrast Based Image Fusion Technique Using Digital Shearlet Transform and Root Mean Square Contrast"- Middle East Journal of Scientific Research, volume 23,Issue 7,pages 1444- 1449.