International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

[1]S Akhila, [2]R Vineela Reddy, [3]D Supraja Ram, [4]L. Thirupathi

[1,2,3,4]Stanley College of Engineering and Technology for Women, Hyderabad, Telangana.

Corresponding author email: thiru1274@gmail.com

Abstract:

In developed countries, telecommunications has become one of the most important industries. The level of competition has risen as a result of technological advancements and an increase in the number of operators. Companies depend on a variety of techniques to thrive in this competitive industry. Three main strategies have been proposed to generate more revenues: acquire new customers, upsell the existing customers,and increase the retention period of customers. However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable, proves that retaining an existing customer costs much lower than acquiring a new one, in addition to being considered much easier than the upselling strategy. To apply the third strategy, companies have to decrease the potential of customer churn, known as "the customer movement from one provider to another.

Key Words:churn, marketing strategies, telecommunication sector, retaining existing customers, upselling strategy,rate of churning, business models.

1. Introduction:

A customer analysis is an important part of a company's business or marketing strategy. It finds potential clients, determines their wants, and then defines how the product meets those needs. In service industries where services are highly competitive, customer churn is a major problem. Predicting which consumers are likely to depart the company, on the other hand, represents a potentially huge extra revenue source if done early on. Companies develop their marketing strategies by assessing clients depending on churn.

Fig 1: customer churn

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

"Churn Prediction is critical in a variety of industries, including life insurance, finance, and telecommunications. Churn Prediction has become more realistic and precise thanks to recent advances in Machine Learning and Artificial Intelligence. It is critical for detecting clients who are at high risk of quitting the firm or services early on. In this research, Ensemble-based Classifiers such as Bagging, Boosting, and Random Forest were used in the telecom industry for Churn Prediction. Decision Tree, Nave Bayes Classifier, and Support Vector Machine were used to compare Ensemble-based classifiers against well-known classifiers (SVM). When compared to other methods, the testing results demonstrate that Random Forest has a lower error rate, a lower specificity, a higher sensitivity, and a higher accuracy of 91.66 per cent."

2. Literature Survey

Mumin Yildiz et al., [1] recommended the use of the Random Forest Method to improve customer churn prediction performance, and the findings show that the Random Forest Method has a higher specificity than AntMiner+ and C4.5 Decision Tree. The data set used in this study is Larose's data set, which was gathered from a wireless network telecommunication provider and contains 5000 customer records with 21 unique attributes for each.

Tan Yi Fei et al., [2] A unique technique combining K means and a Naive Bayes classifier was given to forecasting client turnover. In terms of accuracy and sensitivity, the K means combined Naive Bayes classifier method outperformed the EWD combined Naive Bayes classifier method.

Saran Kumar et al., [3] proposed an enhanced method such as SVM with AdaBoost Classification using the Feature Discovery-based prediction method which combines classifications of SVM, NBTree and SVM AdaBoost to address the limitation of high dimensional classification

Sebastiaan Hoppner et al., [4] presented a new churn classification method called ProfTree which develops an evolutionary algorithm to optimize the EMPC in the model construction step of a decision tree. [9] The data setsused are the real-life churn data sets taken from various telecommunication service providers which contain 889 customers and 10 explanatory variables

Sepideh Hassankhani Dolatabadi et al., [5] A case study of various Data Mining Techniques and Neural predictorswas provided. The data sets used were gathered over a year and a half and included information on every employeeand consumer.

Chuanqi Wang et al.,[6] The common CART model, customer value means of the training data set for the cost of CARTUS model, customer value means of the subset A for the cost of CARTUS model, and customer value means of the subset B for the cost ofCARTUSS model are examined, and it is concluded that the partition cost- sensitive CART model not only performs well but also reduces the total misclassification cost.

3. Objectives

3.1 The major goal is to keep customers who are at the highest risk of churn by proactively connecting with them. Find an appropriate algorithm for forecasting consumers who are likely to churn for a specific data collection.

3.2 The major goal of the customer churn forecasting model is to proactively engage with customers who are most likely to churn. Offer a gift certificate or special pricing and lock them in for another year or two to increase theirlifetime worth to the company.

4. Proposed Architecture:

The goal of churn prediction is to identify consumers who are planning to depart a service provider. Retaining a single customer costs a company 5 to 10 times more than acquiring a new one. To give a retention solution, predictive models can provide accurate identification of potential churners in the near future.

Clean the data and do exploratory data analysis using a predictive model that covers all phases of the data science life cycle. We will compare five techniques, including SVM, KNN, Logistic regression, RFM, and DT,in order to determine which algorithm provides the highest results and accuracy. The segment of these algorithms may be compared. Select the model that provides the most accuracy and precision. For the model with the highest accuracy, create a confusion matrix.

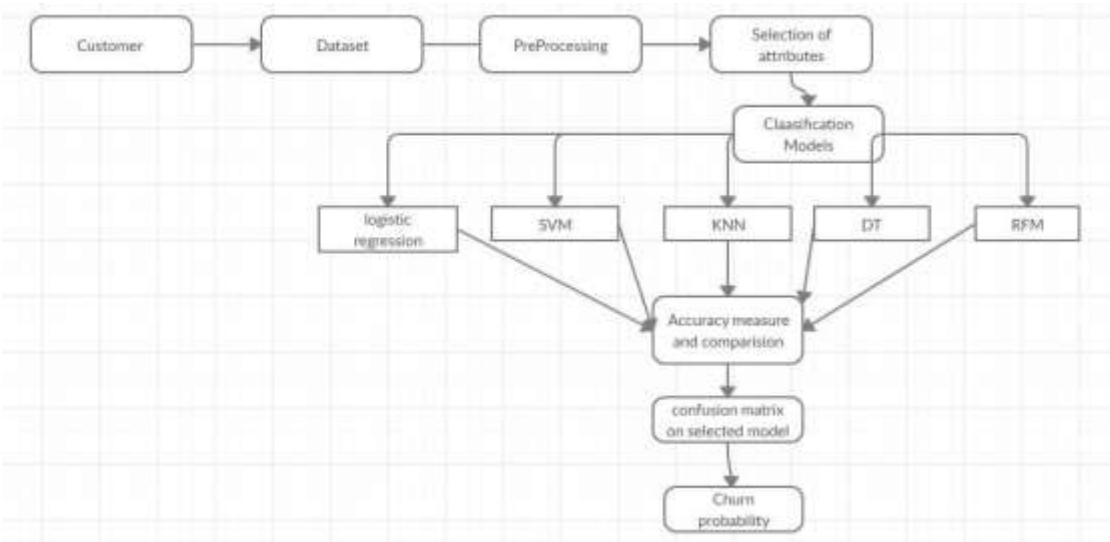

Fig 2: architecture

4.1 Static Analysis:

In today's business world, competition is fierce, and every consumer is priceless. Understanding the consumer iscrucial, and this includes being able to comprehend the customer's behaviour patterns. Customer churn refers to thepace at which a commercial customer (particularly on SaaS platforms) departs and takes their money elsewhere. Understanding customer churn is critical to a company's success, and a churn study is the first step toward doingso.

Customer Churn is a measure that gives you information about an organization's customer attrition rate. It is projected that a 5% reduction in customer churn might result in a significant boost in profitability (25 percent - 125 percent). As a result, forecasting and preventing Customer Churn represents a new revenue opportunity for any interested company

4.2 Dynamic Analysis:

Across prediction horizons, significant even vs. balanced data. The strategy of using independently learned binary classifiers outperforms survival analysis. Horizon particular ranking enables retention efforts to be targeted acrosstime and clients. Allows for the measurement of the impact of environmental factors on the likelihood of churn.

5.Methodologies:

Machine learning is characterised by the creation of systems capable of detecting patterns in data and learning from it without the use of explicit programming. These are online behaviour indicators that signal diminishing consumer satisfaction from using firm services/products in the context of customer churn prediction.

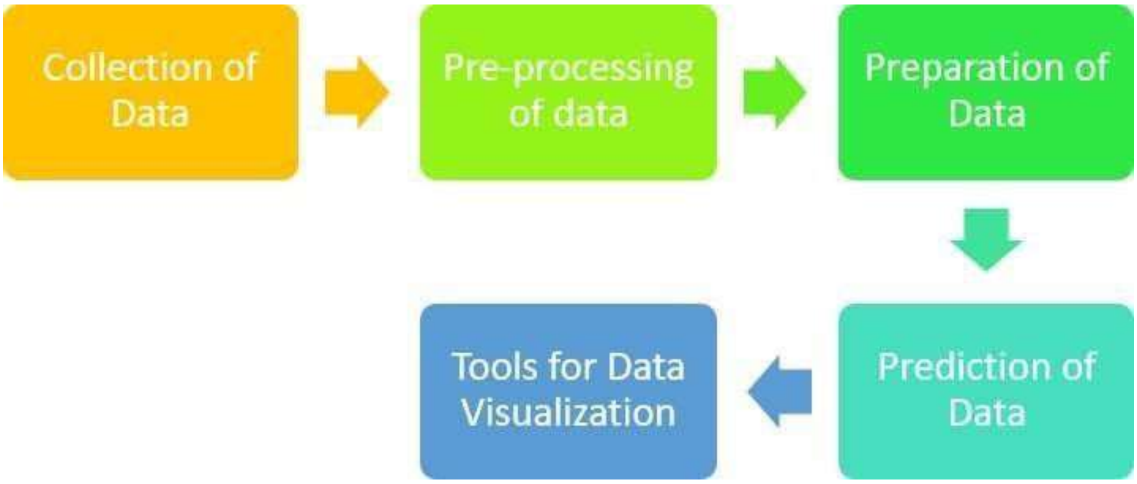

fig 3: Steps used

SUPPORT VECTOR MACHINE:

Support vector machine is a visual representation of coaching expertise as points in a space divided into classesby a transparent gap as broad as possible. New examples of area units were then mapped into the same region and predicted to belong to a class, which supported the gap they fall into.

<u>Parameters of SVM :</u>

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

There are unit 3 main parameters that we tend to may play with once constructing an SVM classifier:

1. Type of kernel
2. Gamma worth
3. C value

RANDOM FOREST:

Random forest is an ensemble model that produces several trees and classifies objects based on all of the trees' "votes." i.e., an object is assigned to the category with the highest number of votes across all trees. As a result, the problem of high bias (overfitting) may be alleviated.

A random forest classifier is a meta estimator that matches a range of call trees on many subsamples of datasetsand utilises an average to improve the model's predictive performance while preventing overfitting. Because of the initial input sample size, the sub-sample size is frequently fixed, but the samples square measurements are redrawn with replacement.

K NEAREST NEIGHBOUR :

Within the area of the input parameter, the kNN classifier associates in the nursing object by a majority vote ofthe item's neighbours. The article is assigned to the category with the most members among its k (a human- defined number) closest neighbours.

It's a lazy, non-parametric formula. It's non-parametric since no assumptions about knowledge distribution are made (knowledge isn't supposed to be regularly distributed).

It's lazy since it doesn't take the time to develop a model and create a generalisation of the data (it doesn't trainsome parameters to perform wherever input X provides output y).

So to be precise, this is often not very a learning formula. It merely classifies objects supported feature similarity (feature = input variables). Classification is computed from an easy majority vote of the k nearest neighbours of every purpose.

DECISION TREE :

Given a set of qualities and their categories, a choice tree generates a set of rules that will be used to classifythe data.

Decision Tree, as the name implies, makes a decision using a tree-like paradigm. It divides the sample into two or more homogeneous groups (leaves) based on the most significant differences in your input variables. Settling on a soul (predictor), the algorithmic rule takes into account all possibilities and performs a binary split on them (for categorical knowledge, split by a cat; for continuous, choose a cut-off threshold). It will thenchoose the one with the smallest quantity value (i.e. highest accuracy) and repeat the process until the information is successfully split into all leaves (or reaches depth).

LOGISTIC REGRESSION :

Logistic regression becomes a classification technique only if a choice threshold is brought into the picture. The setting of the brink price could be a important side of provision regression, depends on classification downside itself.

6. Results and discussions:

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

By forecasting client behaviour, the initial churn prediction will prevent the company from losing money. In thefuture, Reinforcement Learning and Deep Learning will be the two most popular strategies for dealing with the Churn Prediction Problem. Every customer expects intelligent service or reward points from service providers. Providing fast services to authentic consumers is a time-consuming operation, as it is difficult to foresee who will be the company's real clients.

Table-1: The accuracy of logistic regression models is high compared to other models.

	score	model
0	81.14	Logistic regression
1	80.66	Support vector machine
2	79.38	Random forest
3	76.87	k- nearest neighbour
4	73.27	Decision tree

Table-2: Predicted probability of churn rate of the customers:

	customerID	Probability_of_churn
0	7590-VHVEG	0.649225
1	5575-GNVDE	0.043673
2	3668-QPYBK	0.340977
3	7795-CFOCW	0.026396
4	9237-HQITU	0.694569

7. Conclusion:

In today's digital world, having a smart cell phone is extremely important for everyone's survival. As a result, some service providers would like to provide value-added services in order to keep their consumers. Several telecommunications firms are having difficulty predicting when a client-agency may leave their services. Churn Prediction may be a fundamental flaw in the telecommunications industry that has attracted the attention of severalexperts in recent years. Ensemble-based Classifiers were used in a comparative study of customer churn predictionin the telecommunications industry, and they were compared to the current well-known base ClassifiersIn comparison to other models, the experimental results reveal that Logistic regressions are the best Classifier for Churn Prediction drawback in terms of all performance characteristics such as accuracy, sensitivity, specificity, and error rate.

References:

1. J. Hadden, A. Tiwari, R. Roy, D. Ruta “Computer assisted customer churn management: state-of the-art and future trends” Comput. Oper. Res. 34 (10) (2007) 2902–2917.

2. Kotler, P., Keller, K. L. 2009. Marketing Management. Pearson Prentice Hall.

3. N. Kamalraj, A. Malathi "A Survey on Churn Prediction Techniques in Communication Sector" InternationalJournal of Computer Applications (0975 – 8887) Volume 64– No.5, February 2013;

4. Sindhu M E1 and Vijaya M S2 “Predicting Churners in Telecommunication Using Variants of Support VectorMachine” American Journal of Engineering Research Volume-4, Issue-3,pp-11-18

5. Mumin Yıldız, Songul Albayrak “Customer Churn Prediction in Telecommunication with Rotation Forest Method” DBKDA 2017: The Ninth International Conference on Advances in Databases, Knowledge, and Data Applications; 26-29

6. Tan Yi Fei1, Lam Hai Shuan1, Lai Jie Yan1 Guo Xiaoning1, Soo Wooi King. “Prediction on Customer Churnin the Telecommunications Sector Using Discretization and Naïve Bayes Classifier” Int. J. Advance Soft Compu.
7. Appl, Vol.9, No.3,Nov 2017

8. A. Saran Kumar, Dr. D. Chandrakala “An Optimal Churn Prediction Model using Support Vector Machine with Adaboost “International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2017 IJSRCSEIT. Volume 2. Issue 1

9. Sebastiaan Hoppnera, Eugen Striplingb, Bart Baesensbc, Seppe vanden Brouckeb, Tim Verdonck “Profit Driven Decision Trees for Churn Prediction” arXiv:1712.08101v1 ;1-30

10. Verbraken, T., Verbeke, W., Baesens, B., " A novel profit maximizing metric for measuring classification performance of customer churn prediction models. "IEEE Transactions on Knowledge and Data Engineering; 2013, 25 (5), 961-973

11. Sepideh Hassankhani Dolatabadi, Farshid Keynia “Designing of Customer and Employee Churn PredictionModel Based on Data Mining Method and Neural Predictor” The 2nd International Conference on Computer andCommunication Systems 23 October 2017;74-77.

12. Chuanqi Wang, Ruiqi Li, Peng Wang, Zonghai Chen “Partition cost-sensitive CART based on customer value for Telecom Customer churn prediction” Proceedings of the 36th Chinese Control Conference July 26-28, 2017, Dalian, China ;5680-5684

13. Muhammad Azeem1, Muhammad Usman, Fong “A churn prediction model for prepaid customers in telecomusing Fuzzy classifiers” Telecommunication Systems-Modelling, Analysis, Design and Management December 2017, Volume 66, Issue 4, pp 603–614.

14. Long Zhao, Qian Gao, XiangJun Dong, Aimei Dong, Xue Dong "K local maximum margin feature extractionalgorithm for churn prediction in telecom" Cluster Computing June 2017, Volume 20, Issue 2, pp 1401–1409

15. Ruiyun Yu, Xuanmiao An, Bo Jin, Jia Shi, Oguti Ann Move, Yonghe Liu "Particle classification optimization-based BP network for telecommunication customer churn prediction" Neural Computing andApplications February 2018, Volume 29, Issue 3, pp 707–720.

16. Franciska, Swaminathan "Churn Prediction Analysis Using Various Clustering Algorithms in KNIME Analytics Platform" May 2017; 166-170

17. Muhammad Raza Khan, Joshua Manoj , Anikate Singh, Joshua Blumenstock "Behavioral Modeling forChurn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty" December 2015
18. ;1-4

19. Kishor Kumar Reddy C and Vijaya Babu B, "ISPM: Improved Snow Prediction Model to Nowcast the Presence of Snow/No-Snow", International Review on Computers and Software, 2015.
20. (http://www.praiseworthyprize.org/jsm/index.php?journal=irecos&page=article&op=view&path%5B%5D=17055)
21. Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, "SLGAS: Supervised Learning using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow", International Review on Computers and Software, 2015.
22. (http://www.praiseworthyprize.org/jsm/index.php?journal=irecos&page=article&op=view&path%5B%5D=16706)
23. Kishor Kumar Reddy C, Vijaya Babu B, Rupa C H, "SLEAS: Supervised Learning using Entropy as Attribute Selection Measure", International Journal of Engineering and Technology, 2014.
24. (http://www.enggjournals.com/ijet/docs/IJET14-06-05-210.pdf)
25. Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B, "A Pragmatic Methodology to Predict the Presence of Snow/No-Snow using Supervised Learning Methodologies", International Journal of Applied Engineering Research, 2014.
26. (http://www.ripublication.com/Volume/ijaerv9n21.htm)
27. Kishor Kumar Reddy C, Rupa C H and Vijaya Babu, "SPM: A Fast and Scalable Model for Predicting Snow/No-Snow", World Applied Sciences Journal, 2014.
28. (http://www.idosi.org/wasj/wasj32(8)14/14.pdf)
29. Kishor Kumar Reddy C, Anisha P R, Narasimha Prasad L V and Dr. B Vijaya Babu, "Comparison of HAAR, DB, SYM and COIF Wavelet Transforms in the Detection of Earthquakes Using Seismic Signals", International Journal of Applied Engineering Research, 2014, pp. 5439-5452.