



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

“CHURN PREDICTION IN TELECOM INDUSTRY”

Yashraj Bharambe¹, Pranav Karanjawane², Pranav Deshmukh³, Diptesh Chaudhari⁴

^{1,2,3,4}Student at JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

Abstract:

Customers are the base for any business success and that is why firms become aware of the significance of acquiring satisfaction of customers. Customer churn is an essential issue and it is regarded as one of the most essential concerns among firms because of increasing rivalry among firms, increased significance of marketing strategies and customers conscious behaviour in present years. Organizations must develop different strategies to resolve the churn issues relying on the service they offer. Customer churn practise is essential in competitive and rapidly developing telecom sector. The process of migrating from one service provider to another telecom services provider occurs due to good services or rates or due to various advantages which the rivalry firm provides customers when signing up, Due to the greater cost related with acquiring new customers the prediction of customer churn has developed as an indispensable part of planning process and strategic decision making in telecom sector.

Keywords:

Customer churn, Big Data, Machine Learning, KNN, Logistic Regression

INTRODUCTION

Retaining existing customers is a vital part for a company to keep growing without bearing the price of acquiring new customers. Departments likes sales marketing and customers satisfaction have to work constantly to keep providing customers with incentives and offers to reduce churn. Machine Learning helps with reducing churn and analysing the massive amount of customer data. It also helps departments to provide customers with best incentives and at right time to keep them satisfied. Making the best choice like providing sufficient amount of offer to a customer and keeping the company in profit is not possible manually. In today's competitive market we need to keep existing customer to decrease competition. While performing churn prediction we need to know that 1% of monthly churn is equal to 12% of yearly churn. According to Forbes it takes around 5 times the money to find new customers than to keep the one you already have. Less churn is supposed to be a good indication for profitable businesses. Tracking lost customers and growth rates, comparing both metric tells you how much a company is growing over time. But maintaining customer data can often be hard as it is inaccurate and messy. And with the release of every new service and product the metrics change.^[1]

Aims and Objectives:

Primary Objective:

To explore the customer churn prediction in telecom using machine learning in big data platform

Secondary Objective

1. To investigate the impact of customer churn in telecom industry as a whole
2. To discuss the significance of customer churn models in telecom industry
3. To compare algorithms that are effective in reducing churn rate in telecom companies

Related Work:

Azeem and Usman (2018) authors used state of art machine learning like boosting and SVM. High predictive churn models can effectively support retention campaigns to remove unnecessary expenses and refine marketing decisions. Churn customers have imbalanced distribution in the dataset unlike other class. This problem makes the classical accuracy criteria unsuitable for evaluation and selection of the best SVM model. Achieved accuracy in this model was of approx. 88% ^[1]

Ebrah and Elnasir (2019) This model consists of Machine Learning applications like SVM, Naive Bayes, and decision trees. Each model consisted of their own advantages. Decision tree consider all features like monthly Expenditure, Services etc., to decide whether customers will churn or not. Each model had an accuracy rate of 82%, 87% and 77% respectively.^[2]

Sanket Agrawal; Aditya Das, Amit Gaikwad, Sudhir Dhage (2018) This model has used a deep learning approach for predicting churn rate of the telecom industry. Non-linear classification model was built in this case. All features were considered during the churn prediction process. A accuracy of approximately 80% was achieved using deep learning model in this particular model.^[3]

Pushkar Bhuse; Aayushi Gandhi; Parth Meswani; Riya Muni; Neha Katre (2020) In this model machine learning and deep learning concepts were used in order to predict churn rate for the telecom industry. They had done the analysis of lost and gained customers for the company over the period of time.^[4]

To classify whether the customer will churn or not is decided by newer architectures like XGBoost and Deep neural networks. An accuracy of approximately 88% was achieved.

Review of Literature:

Title	Author & Year of Publication	Methodolgy	Advantages / Benefits	Remark
A fuzzy based churn prediction and retention model for prepaid customers in telecom industry .	Azeem M, Usman M. (2018)	Fuzzy Classifiers	Automated and targeted retention strategy .	Managed to accomplish retention accuracy level up to 88% .
Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning.	Sanket Agrawal; Aditya Das; Amit Gaikwad; Sudhir Dhage (2018)	Deep Learning	The model displayed a list of the attributes which were directly and inversely related with the churn rate .	An accuracy of 80.03% was achieved.
Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms.	Ebrah K, Elnasir S. (2019)	Naïve Bayes SVM Decision trees	All churn prediction models described here can be used in other customer response models as well .	Accuracy score of Naïve Bayes : 87% SVM : 82% DT : 77%
Machine Learning Based Telecom-Customer Churn Prediction.	Pushkar Bhuse; Aayushi Gandhi; Parth Meswani; Riya Muni; Neha Katre (2020)	Random Forest Classifiers ,Deep Neural Networks and SVM	Grid search algorithm was implemented on the algorithms to increase its performance .	Prediction accuracy of 88.96% was achieved.

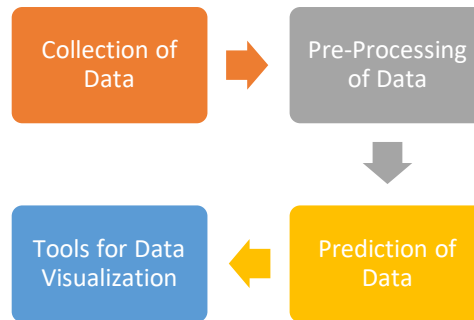
Methodology:

fig 1. proposed system workflow

From the above figure we can see the steps used in proposed system:

1. Collection of data:

In this primary step we need to collect the data to formulate a dataset on which we will further do prediction or analysis in telecommunication.

2. Pre-Processing of data:

There are 3 main steps involved in pre-processing of the data as

Data cleaning, Feature selection and Data transformation.

- The data cleaning step involves manipulation of missing data and filling empty/null values with mean, median or zero accordingly.
- The Feature selection step involves carefully selecting the features from the data set which directly influences the models performance.
- The Data transformation step involves transforming the data-set according to the algorithms needs to suit the algorithm.

3. Prediction of data:

This step refers to the output of an algorithm after it has been trained on the data-set and applied on the new data-set when forecasting the likelihood of a particular algorithm.

4. Tools for data visualization:

Data visualization tools provide designers with an easier way to create visual representations of large data sets. So that the message can be stated clearly and to uncover the essential patterns that would be ignored generally.^[5]

Algorithms Used:1. **KNN:**

KNN (K—Nearest Neighbours) is one of many (supervised learning) algorithms used in data mining and machine learning, it's a classifier algorithm where the learning is based on "how similar" is a data (a vector) from other. It classifies its data point based on how its neighbours are classified. KNN stores all the available cases and classifies new cases based on a similarity measure. 'k' in KNN is a parameter that refers to the number of nearest neighbours to include in the majority voting process. A data point is classified by majority votes from its 'k' nearest neighbours.^[8]

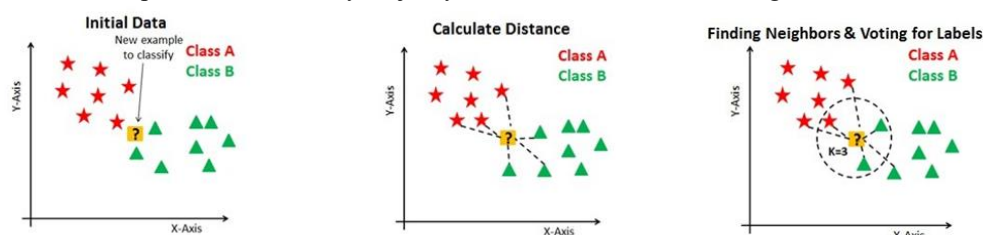


fig 2. basic diagram of knn

2. **Logistic Regression:**

Logistic regression is a supervised learning algorithm. It is used to perform binary classification. Binary classification is used to predict one of the two classes. It is a classification algorithm used to predict binary outcomes for a given set of independent variables. The dependent variable's outcome is discrete. The data fed in the algorithm must be linearly separable. Mainly logistic regression is used to predict the dependent variable from the independent variable when the dependent variable is in categorical format.^[9]

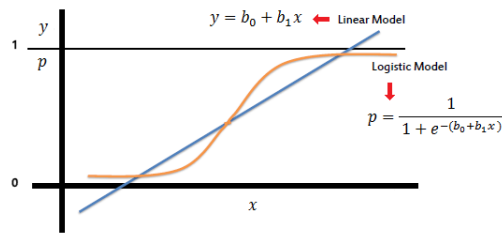


fig 3. basic graph of logistic regression

Advantages:

1. Identifying at risk customers:
2. Optimization product and services:
3. Increased Revenue:
4. Identify pain points:
5. Identify methods to implement:

Conclusion and Future Scope:

The technology industry has gone through drastic changes in last few decades with technology engulfing all of things around us, most people turn to internet when they want to buy something. In order to retain customers company, need to understand and predict customer churn pattern. Comparison between techniques was to find out if the customer will churn or not.

In the coming days, we will further research on various algorithm approaches to have better customer churn prediction. To know the changing behaviour of the customers, the study can be extended by using Artificial Intelligence techniques for trend analysis and customer prediction. Furthermore we can focus on expecting behavioural patterns of customer churn and to employ co-browsing to provide personalized service to customers.

References:

- [1] Azeem, M, Usman, M. (2018). A fuzzy based churn prediction and retention model for prepaid customers in telecom industry. International Journal of Computational Intelligence Systems, 11(1), pp. 66 – 78.
- [2] Ebrah, K, Elnasir, S. (2019). Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms. Journal of Computer and Communications, 7(11).
- [3] Sanket Agrawal; Aditya Das; Amit Gaikwad; Sudhir Dhage (2018) Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning.
- [4] Pushkar Bhuse; Aayushi Gandhi; Parth Meswani; Riya Muni; Neha Katre (2020) Machine Learning Based Telecom-Customer Churn Prediction.
- [5] Ahmad, A.M., Jafar, A, Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(28), pp. 1-24.
- [6] Dwivedi, A, McGaugh, M. (2019). Telecom Industry: Customer Churn Prediction.
- [7] Maroua Bahri; Bruno Veloso; Albert Bifet; João Gama (2021), AutoML for Stream k-Nearest Neighbors Classification.
- [8] Pengbo Wang; Yongqiang Zhang; Wenting Jiang (2021), Application of K-Nearest Neighbor (KNN) Algorithm.
- [9] Xiaonan Zou; Yong Hu; Zhewen Tian; Kaiyuan Shen (2020), Logistic Regression Model Optimization and Case Analysis.
- [10] Zan Yang; Dan Li (2019), Application of Logistic Regression with Filter in Data Classification.