Customer Churn Prediction

**A bstract**- As the client is larger a large number of details are made daily in the field of telecommunications. Decision makers and business analysts say it is way more extravagant to make new customers than to keep existing customers, so it is very important for customer relationship management analysts to know the reason for aggressive customers and behavioral patterns in deceptive customer data. It leads to the salvation of big business regardless of size. This paper proposes a predictive model of churn that will begin its operation by clearing the data initially. As it is very important to have information that is not uncommon it therefore leads to accurate predictions. It uses classification and integration techniques to identify targeted customers and provides the factors that lead to customer withdrawal in the communications sector.

An engineering project will then be conducted to determine which feature plays a key role in the forecast. Feature selection is made using the benefit information and the qualification rating filter. By recognizing key churn values from customer data, CRM can increase productivity, recommend appropriate

promotions to a group of potential customers based on similar behavioral patterns, and radically improve the company's marketing campaigns. The outcome of the model will provide relevant information that will be of great benefit to the sector.

**K eywords -** Prediction, Model, Churn, Telecom.

# INTRODUCTION

A customer churn is called a customer’s tendency to take off a service provider. Customer speculation is the procedure of recognising those customers who may withdraw from the current service provider for a number of reasons. The main purpose of the churn forecast model is to pinpoint those consumers so that the final strategies can be targeted and the company thrives by increasing its total revenue. This has been put up as a popular issue in numerous fields, one of it is Telecommunication. So it will really be so helpful for companies

if they could know who of their customers are going to be churn. That will not only help them to know the stats but also they can work among themselves to prevent it from happening. They can personally provide benefits to specific customers according to their needs. The raw data is taken from an IBM sample containing 7043 records and 20 features.

Now we are going to use the clean data for churn prediction model so as to avoid all the abnormalities. Feature selection will be done so that only those factors will be taken under consideration that really affects the churn of a customer. Feature selection is made using the benefit information and the qualification rating filter. By knowing the key churn values from customer data, CRM can increase productivity, propose appropriate advancement to a group of potential consumers based on alike behavioral designs, and radically enhance corporate marketing campaigns. The result of the model will provide apt information that will be of great help to the sector.

# LITERATURE SURVEY

In paper ‘Customer Churn Prediction Analysis’ published in International Journal of Computer Applications 182(29):15-17, The process of designing the churn prediction mode

l, its application and causes, challenges and problems for designing the model.

Iris Figalist,Christoph Elsner, Jan Bosch, Helena Holmstrom Olsson, in their work introduces a method that enables the

creation of a customer and final user map data based on “customer categories” that allow the prediction model to capture all critical impactful factors.[16]

Bogdan Dumitrescu suggests a state-of-the- art churn prediction, a model based on deep neural models, future models, and using Big Data processing on the same large computer using GPU cells.[19]

Pallav Routh states a method of operation within the framework of an unambiguous survival forest that precisely identifies churn hazards and identifies the relationship between risks and customer behavior. Contrary to existing methods, the proposed model does not rely on a specific functional form to illustrate the relationship between risk and behavior, nor does it have fundamental differing assumptions, both of which are limited to performance.[5]

Malak Fraihat introduces the Selection Ensemble Model (SEM) as a well built forecaster of churn. Within a group of ML models, SEM strongly selects a amalgamtion of models to take part in the creation of the final result.[4]

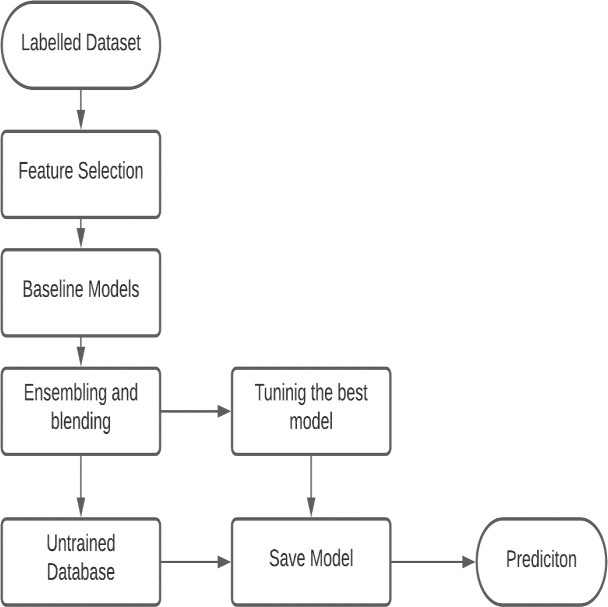
Deepak Gupta states that they try to examine and contrast the performance of more than 100 classifiers in the churn prediction of a telecommunications company. They have made use of popular classifiers from divergent lineage.[1]

S. W. Kim states a churn prediction model that uses segmentation, in addition to aggregation strategies to recognize targeted consumers and provide factors that cause customer's churn in the industry. Feature selection is made using the benefit statistics and the qualification rating filter. It’s output disclose that it produces better churn

prediction classifier using RF.[20]

Asmin Alev Aktaş states that customer data structure is established in order of customer-related information. With sunsequent data, the long-term memory model is designed to measure complex customer segments and is compared to standard classification methods.[6]

# PROPOSED WORK



*Figure 3.1 Architecture Diagram*

The architecture diagram of our project is shown in fig 3.1. We first labelled the dataset and then feature selection is performed to filter out the features that has the most contribution in the prediction. Then ensembling and blending techniques are used for tuning the best model. Now the untrained data is passed through the above made model and prediction is done on that data.

# Methodology

* + - Data Pre-Processing: It is one of the important step in any prediction

model. It is similar to a filter applied to filter out all the unwanted substances. Here we are going to clean our data so as to remove all sorts of abnormalities. Then feature engineering will be performed so as to check which feature plays an important role in the prediction. Feature selection is made using the benefit information and the qualification rating filter. By recognizing key churn values from customer data, CRM can increase productivity, recommend appropriate promotions to a group of potential customers based on alike behavioral patterns, and radically improve the company's marketing campaigns.

* Data: The raw data is taken from an IBM sample containing 7043 records and 20 features. The following data is taken:

- Services each customer has subscribed to - phone, multi-line, internet, online security, device protection, online backup, technical support, and TV and movie streaming.

-Customer account details - tenure of the consumer, contract, method of payment, billing without paper, expenses per month, and total costs.

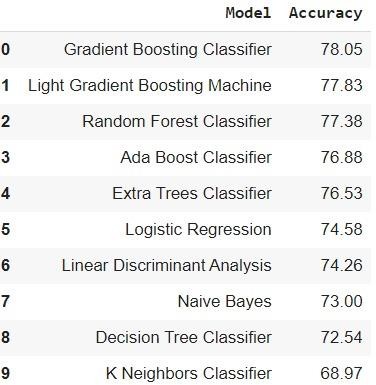
-Demographic information about consumers – their age

lies in which range, gender, and whether they have partners, and dependents contains 7043 rows (customers) and 21 columns.



|  |
| --- |
| Training: First we did Data |
| Cleaning and |
| Preprocessing - Outlier |
| removal/normalisation/lab |
| el encoding/pcs, then |
| Feature Engineering |
| /Selection leading to |
| Training Baseline model. |
| Then Hyper parameter |
| Tuning Grid Search and in |
| the end Evaluation. |

# Algorithms used



*Figure 3.2 Accuracy Table*

Fig3.2 show the accuracy of different algorithms used in the prediction.

* + - Gradient Boosting Classifier - A batch of machine learning algorithms that corporate infirm learning models altogether to produce a solid guessing model called gradient boosting classifiers. Decision trees are often bring up in use when making gradient boost. Because of their effectiveness in separating complex data sets, gradient boosting models are popular and used recently to win most of the data science competitions on Kaggle.
    - Ridge Classifier - Formed on the Ridge regression process, the Ridge classifier converts label data to [-1, 1] and solves the problem in a retrospective manner.

The highest value in prediction is set as the target class and if the data is multiclass then multioutput regression method is used.

* + - Catboost - It is a type of gradient boosting algorithm that can automatically deal with phase variability without pointing out the

that

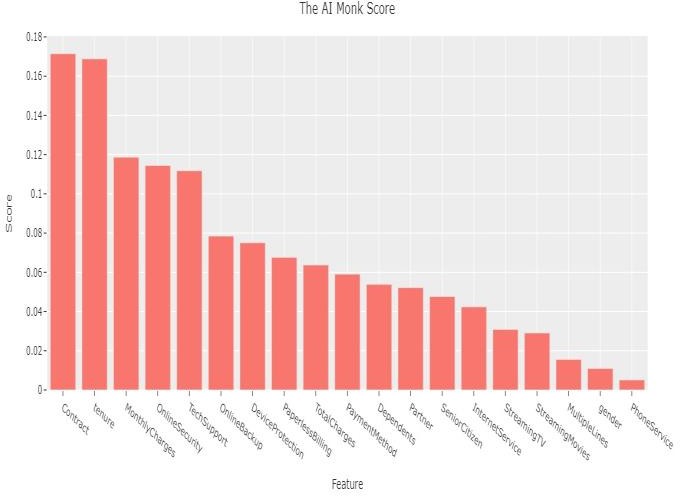
type conversion error, which helps you focus on fixing your model better than editing minor errors.

* Logistic Regression - It is used where response variability is classified and is a classification algorithm. The concept of Logistic Regression is to realize the relationship between the factors and the likelihood of specific results.

issues = p (x) / (1-p (x)) = probabilit y

event occurs / probability of event

* + Feature Selection -it is the uttermost concept in ML that greatly affects the efficiency of the model. The features of the data that are used by you to train your ML models have a notable impact on the efficiency you can achieve. As shown in Fig 3.3, it depicts the impact of different features in the prediction model.

does not occurs

* + - Random Forest - A random forest is a group of trees (forest) and forms many decision-making trees and combines them together to obtain more precise and solid predictions. It can be used for regression as well as classification problems.
    - XGBoost - It is an ensemble ML algorithm which is build on decision- tree. It utilizes a gradient boosting framework. In prediction problems such as that involves unstructured data (images, text, etc.) artificial neural networks be prone to outrun all other algorithms or frameworks. It is an absolute amalgamation of hardware and software accrual techniques to output great results using smaller amount of computing resources in the momentary time.

# IMPLEMENTATION

*Figure 3.3 Feature score*

* Ensembling - Ensemble methods is a technique of machine learning that incorporates several base models to produce one optimal predictive model.
* Blending - It is an algorithm for the ensemble. It is a common term for integrated combinations or mass collections where instead of inserting a meta-model in failing predictions made by the base model, fit into predictions made on the capture

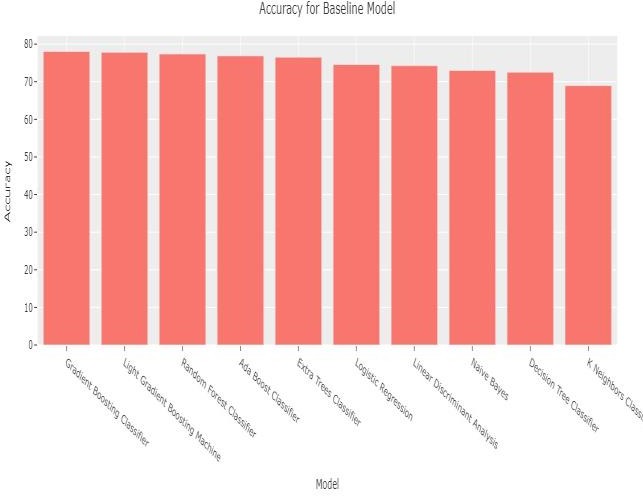
database.

* + Tuning - Hyperparameter is setting values of parameter before the training process. In machine learning it has importance similar to data cleaning and feature extraction. Tuning parameter is very important and interesting. Hyperparameter is very sensitive to a small change in learning rate or the estimators will lead the great change in accuracy of model.

Grid search is the classic way for hyperparameter tuning. We made a grid and search for the best score by joining the values and find the best combination.

Grid search always result to optimal solution but it is time consuming and because of large combination, high computational is required which make it expensive.

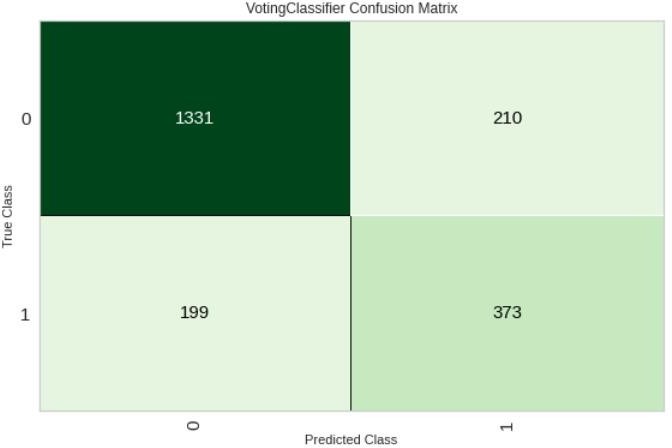
# RESULTS AND DISCUSSION



*Figure 5.1 Accuracy of Baseline Model*

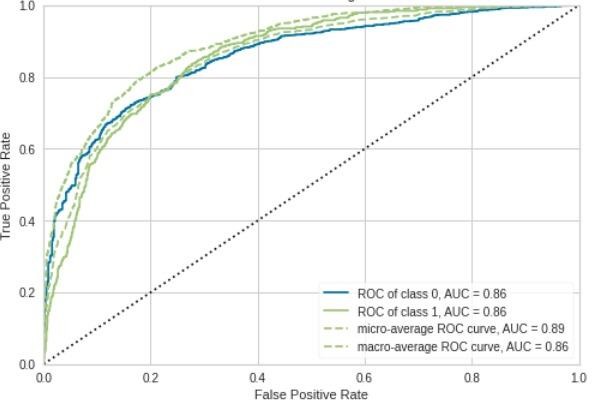
Fig 5.1 is the graph that shows the accuracy of different models. It helps us to determine that which algorithm works the best among all and thus give us the absolute reason for

using it as a prediction algorithm for churn. Table stating the same is given in the fig 3.1 under the algorithms used section.



*Figure 5.2 Voting Classifier confusion matrix*

In fig 5.2, is the measurement of performance for ML classification problem. It states how many times incorrect and correct predictions are made with the count values.



*Figure 5.3 ROC curve*

Fig 5.3 shows the tradeoff between

sensitivity and specificity.

# CONCLUSION AND FUTURE

We have studied through multiple survey papers to better understand the concept and then conducted a literature survey from those papers. After the designing and implementation of our model we can conclude that predictions made by our model were fast and efficient with the real world market. Churn predicted it was really efficient and the customer's recent behavior can tell a lot about his future proceedings with the company. This model deals with the data of the customer’s related to the Telecom companies and predicts whether it will turn out to be a churn or not. This is going to help a lot of companies as maintaining a customer is cheaper than to make new customers therefore telecom companies can make more profits.