**Telecom Customer Churn Prediction Using Watson Auto AI**

**Overview :**

operators to predict customers who are most likely subject to churn Project Requirements: IBM Cloud, IBM Watson Auto AI ,Node-Red ,Node JS.

Functional Requirements: IBM Cloud.

Software Requirements: Watson The main contribution of our work is to develop a churn prediction model which assists telecom Auto Ai, Node-Red.a

Project deliverables, customer churn prediction using Watson Auto AI .

**Purpose:**

Customer churn is the major problem and one of the most important concern for large company. Due to the direct effect on the revenues of the companies ,especially in the telecom fields , companies are developing means to predict potential customer to churn . Therefore finding factors that predict ccustomer churn is important to take neccessary actions to reduce this churn . churn prediction helps in identifying customers who are likely to leave a company. The contribution of our woek is to devolop a churn prediction model which assists telecom operators to predicts the customers.

**LITERATURE SURVEY:**

The study of predicting which persons are going to churn in advance will help the telecommunication industry and the CRM department to identify which persons are going to leave the network. The problem of our work discussed is the classification problem i.e. to classify each subscriber as potential churner or potential non churner

**Existing problem:**

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques .

**Proposed Solution:**

Telecommunication industry always suffers from a very high churn rates when one industry offers a better plan than the previous there is a high possibility of the customer churning from the present due to a better plan in such a scenario it is very difficult to avoid losses but through prediction we can keep it to a minimal level. We are building a Machine Learning model to predict the customer churn using IBM Watson AutoAI Machine Learning Service. The model is deployed on IBM cloud to get scoring end point which can be used as API in mobile app or web app building. We are developing a web application which is built using node red service. We make use of the scoring end point to give user input values to the deployed model.

**1.Churn:** A churn is defined as customer attrition or loss in a telecom industry when customers terminate their contracts or their usage and switch to another service provider.

**2. High Valued Customers**: These are the type of customers that share important service features like higher bills, longdistance transitions, etc. They spend more than he average rate and their expectations from the service providers is high, which make the retention of such customers more important than the rest as they are beneficial to the company. The company must allocate more resources to them to decrease the churn rate.

**3. Customer Relationship Management (CRM):** Customer churn prediction has become important to the telecom industry to protect their customers based on their loyalty to the company for their service given. Hence, it is vital for all companies to improve their Customer Relationship Management. It involves knowing the customer’s attainment so it can retain the most profitable customers and identify those whose churn no longer makes any difference to the company. It aims at better advisement and customized plans for their high valued customers in order to retain them from churning to a competitor.

In the present competitive market of telecom domain, churn prediction is a significant issue of the CRM to retain valuable customers by identifying a similar groups of customers and providing competitive offers/services to the respective groups. Therefore, in this domain, the researchers have been looking at the key factors of churn to retain customers and solve the problems of CRM and decision maker of a company. In this study, a customer churn model is provided for data analytics and validated throughstandard evaluation metrics. The obtained results show that our proposed churn model performed better by using machine learning techniques.Random Forest and J48 produced better F-measure result that is 88%. We identified the main churn factors from the dataset and performed cluster profiling according to their risk of churning. Finally, we provided guidelines on customer retention for decision-makers of the telecom companies.

**Performance analysis:-**

Customers’ churn is a considerable concern in service sectors with high competitive services. On the other hand, predicting the customers who are likely to leave the company will represent potentially large additional revenue source if it is done in the early phase.

Model 1:

Author:Adeniyi, D.A., Wei, Z., Yongquan, YAutomated web usage data mining and recommendation system using K-NearestNeighbor (KNN) classification method. Appl. Compute. Inform. 12(1), 90–108 (2016)

Model 2:

Author:Idris, A., Khan, A., Lee

Adaboost style boosting is used to evolve a number of programs per class. Finally, the predictions are made with the resulting programs using the higher output, from a weighted sum of the outputs of programs per class. The prediction accuracy is evaluated using 10 fold cross validation on standard telecom datasets and a 0.89 score of area under the curve is observed.

Source:

https://ieeexplore.ieee.org/abstract/document/6377917/references#references

Model 3:

Author:Aiswaryadevi V V J

The proposed predictive model identifies the traits that highly influence customer churn, with the help of machine learning techniques like KNN, Random Forest and XG Boost.The proposed model shows that Fiber Optic customers with greater monthlycharges have higher influence for churn.

Source:

https://www.researchgate.net/publication/333650592\_An\_Effective\_Classifier\_for\_Predicting\_Churn\_in\_Telecommunication

Model 4:

Author:IqbalHanif

This research is aimed to prove or disprove that XGBoost algorithm gives better prediction compared with logistic regression algorithm as an existing algorithm. This research was conducted by using customer’s data sample (both churned and stayed customers) and their behaviors recorded for 6 months from October 2017 to March 2018.

Source:

https://www.researchgate.net/publication/338735685

Model 5:

Author: Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A,Hussain A.

Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access. 2016

Source:-

https://ieeexplore.ieee.org/document/7707454

**Our Model:**

we used Gradient Boosting Classifier.Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

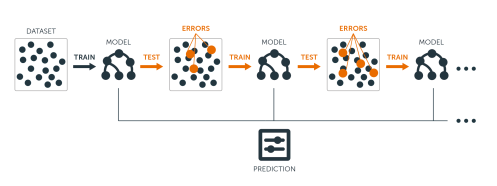


**Abstraction :**

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques Telecommunication industry always suffers from a very high churn rates when one industry offers a better plan than the previous there is a high possibility of the customer churning from the present due to a better plan in such a scenario it is very difficult to avoid losses but through prediction we can keep it to a minimal level.

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting. Gradient boosting models are becoming popular because of their effectiveness at classifying complex datasets

**Steps to Gradient Boosting:**

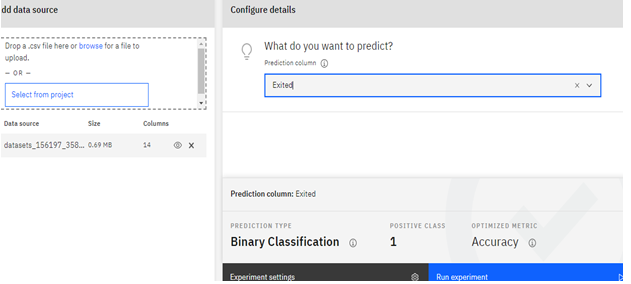


In order to implement a gradient boosting classifier, we'll need to carry out a number of different steps. We'll need to:

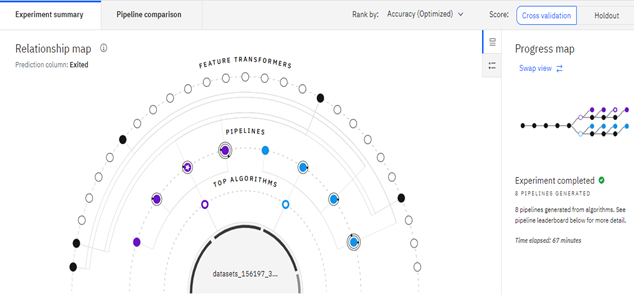
* Fit the model
* Tune the model's parameters and Hyper parameters
* Make predictions
* Interpret the results

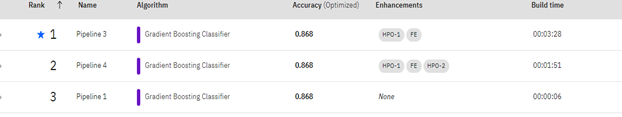
**1.Fit the model**

Upload the training data ﬁle, auto.csv, from your local computer by dragging the ﬁle onto the data panel or by clicking browse and then following the prompts.

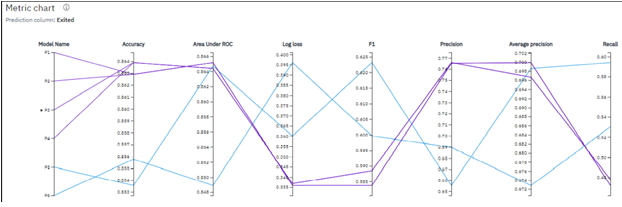


**2.Tune the model's parameters and Hyper parameters**

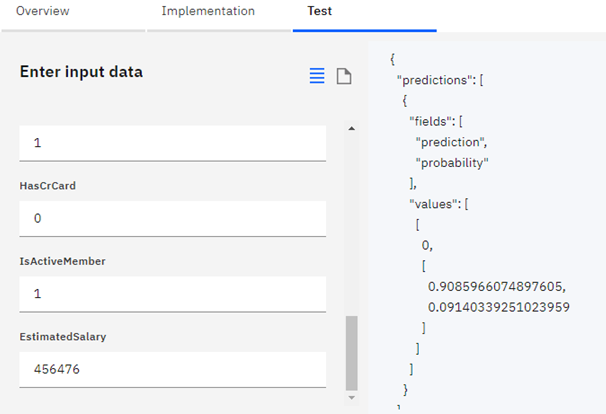


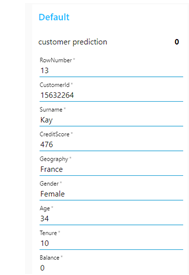


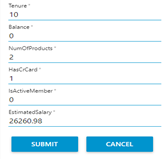
**3.Make predictions**



**4.Interpret the results**







## How Gradient Boosting Works

Gradient boosting involves three elements:

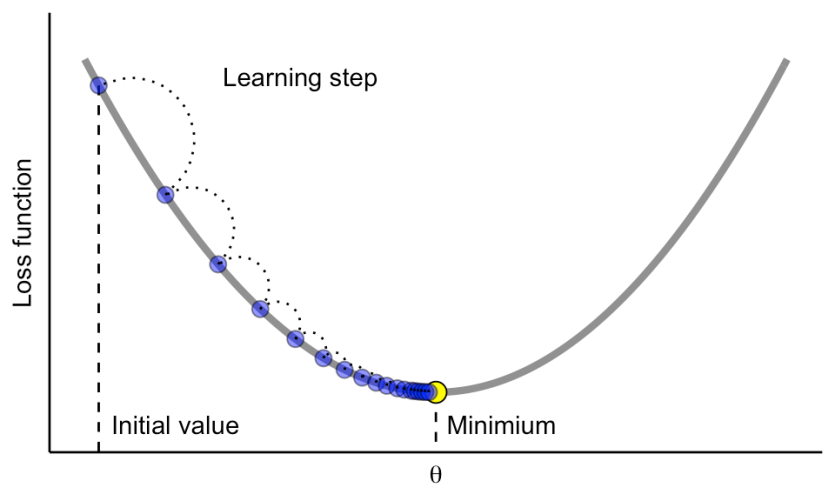
1.A loss function to be optimized.

2.A weak learner to make predictions.

3.An additive model to add weak learners to minimize the loss function.

### 1. Loss Function

The loss function used depends on the type of problem being solved.It must be differentiable, but many standard loss functions are supported and you can define your own.For example, regression may use a squared error and classification may use logarithmic loss



A benefit of the gradient boosting framework is that a new boosting algorithm does not have to be derived for each loss function that may want to be used, instead, it is a generic enough framework that any differentiable loss function can be used.

### 2. Weak Learner

Decision trees are used as the weak learner in gradient boosting.Specifically regression trees are used that output real values for splits and whose output can be added together, allowing subsequent models outputs to be added and “correct” the residuals in the predictions.

Trees are constructed in a greedy manner, choosing the best split points based on purity scores like Gini or to minimize the loss.

### 3. Additive Model

Trees are added one at a time, and existing trees in the model are not changed.

A gradient descent procedure is used to minimize the loss when adding trees.

Traditionally, gradient descent is used to minimize a set of parameters, such as the coefficients in a regression equation or weights in a neural network. After calculating error or loss, the weights are updated to minimize that error.

**Comparative Analysis**

The churn prediction model is created by three well known classifiers namely GBC, RF and XGBoost. The comparison of evaluation metrics such as accuracy score and F score are calculated. We have calculated the accuracy for higher churn prediction model. Experimental results show that Gradient Boost classifier gives higher accuracy score and F score compared to XGB and RF classifiers. Thus, Gradient Boost classifier is used for feature selection in the proposed system.

**Advantages and Disadvantages :**

**Advantages**

predict potential customers

assist telecom operators to predict

**disadvantages**

more customers would churn ,the less is the potential for the business

problems in offers and packages

increase in complexity

**conclusion:**

Telecommunication industry has suffered from high churn rates and immense churning loss. Although the business loss is unavoidable, but still churn can be managed and kept in an acceptable level. Good methods need to be developed and existing methods have to be enhanced to prevent the telecommunication industry to face challenges. In this paper we discussed the various prediction models and also compared the quality measures of prediction models like regression analysis, decision trees. We found that the accuracy achieved with decision tree is far much higher than the logistic regression technique which clearly states that decision tree is an efficient technique.