**REPORT ON**

**Telecom customer churns prediction using Watson Auto AI**

**1.INTRODUCTION**

**1.1. Overview:-**

Rapid increase of telecom providers rises to tremendous competition and leads to churn . Churn is the process of customers switching from one firm to another in given time . The pre-eminent reason for churn is the customer dissatisfaction . Retaining the existing customers is more profitable than fetching the new customers . So, the telecom providers have to maximize concentration to the extant customers to avert churn . Further, long standing satisfied customers provoke more profit because they aren’t going to churn . The customer churn leads to huge loss economically and affect company’s status . An appropriate churn prediction model is essentially needed to predict the churners . Many companies are concentrating on customers induvial for giving best services to them. Churn is the inevitable action in all industry In recent past, various machine learning methods are used to tackle threat in customer churn in telecom .Lu et al.in (2014) suggested Gentle Ada boost algorithm for churn prediction, because it reaches high accuracy and provided good separation of churn models and also a comparison is made with the results of logistic regression and single logistic regression model, which proves that the logistic regression model is the better one. Wenjie et al.in (2016) proposed Semantic-Driven Subtractive Clustering Method (SDSCM) which is proven to have more clustering strength than subtractive clustering method (SCM) and fuzzy c-means (FCM). Hadoop Map Reduce frame work is used to implement this algorithm.. A.keramati et al.in (2014)compared the performance of DT, ANN, KNN, SVM and shown that hybrid of all four performs well, they also proposed a methodology for extracting the influential features from the dataset[18]. Pinar kisioglu et al.in (2010) applied Bayesian belief networks to explain the causal relationships between the features that contributes to customer churn and the important features were found out and suggested promotions to reduce the churn rate [20]. Koen W.De Bock et al.in(2012) Proposed GAMensplus classification algorithm for strong classification and interpretability. The system also demonstrated the relationship between the predictors and probabilities of churn . Muhammad Azeem et al.in (2018) compared the accuracy of several classifiers with fuzzy model and proved that fuzzy classifiers are more accurate in predicting customer churn dataset having noise . Chih-FongTsai et al.in (2009) used neural networks and hybridised data mining technique for predicting customer churn with higher accuracy, this system combines ANN with ANN and ANN with SOM for churn higher prediction accuracy . Wouter Verbeke et al.in (2010) proposed AntMiner+ and ALBA for improving the learnability providing comprehensible and accurate customer churn prediction model . Zhen-YuChen et al.in (2012) proposed Novel approach known as hierarchical multiple kernel support vector machine

(HMK-SVM) combines static customer and longitudinal behavioural data to improve churn prediction, a three-phase training algorithm was implemented. Long Zha et al.in (2017) proposed New K- local maximum margin feature extraction algorithm (KLMM), they followed the fact that extracting the features of the data will reduce the dimensionality of prediction .T.Vafeiadis et al.in (2015) made a comparative study on a public domain dataset, next deals with improving the accuracy by using boosting algorithms. They used Monte Carlo simulation at different settings of each classification method and improved SVM-POLY with Ada boost classifier which gave higher accuracy [26]. Ying Huang et al.in (2013) proposed Novel hybrid model-based learning system, which integrates k-means clustering algorithm (unsupervised) and inductive technique (FOIL)-(supervised) for building an effective predictive model.

**1.2.Purpose:-**

In recent days, telecom industry plays a major role in our daily life. The proliferation of telecommunication industry becomes very difficult for the service providers to survive in the market. To stabilize in this field, the service providers have to be aware of the features that make the customer to churn. 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. IBM Watson dataset has been analysed to forecast the churn. At last a comparative study has been made among the machine learning algorithm to identify the better algorithm of

higher accuracy. The proposed model shows that Fiber Optic customers with greater monthly charges have higher influence for churn.

**2. LITERATURE SURVEY**

**2.1. Exitising Problem:-**

predict customer churn machine learning problems can be either of two types: Classification or Regression.

* Classification:-

This type would need the data scientists to determine to which class or category the customer belongs. It is referred to as a data point.To train the algorithm, they make use of the historical data of the customers and use the predefined target variables. These variables are the labels we give to the subjects of the problems which are ‘churners’ in our case. Classification helps businesses to answer following questions:

Will the customer churn or not?

Would they repurchase the subscription or not?

Will the customer downgrade the subscription plan?

Is there any sign of erratic behavior in customers?

The last question specifically targets a common problem in classification which is called anomaly detection. It identifies the data points that significantly deviate from the usual behaviour.

* Regression:-

Regression analysis is one of the widely used methods in customer churn prediction software. It is a value that is used in statistical analysis to define the relationship between the customer churn and the data values that influence it. This analysis helps in finding exact values for the business prediction. E.g. it can give you the exact time within which a customer is predicted to churn.

**2.2. Proposed Solution:-**

* Collection of Data

After deciding the kind of insights you are going to use, you will need to identify which data sources would give you the best data. You will have to consider all the sources from where you can gather data to create a predictive analysis of your customers. The customer data you have on different portals will give you different data values. And the more you gather from different resources, the more detailed and precise your algorithm would be.

There are multiple sources for gathering the data. They are your CRM software, customer analytics product e.g. Google analytics, CrazyEgg, review comments on social media or any other platform where your customers have their footprint.

* Preparing data

To predict customer churn machine learning algorithms should be able to understand the data you gathered in the previous steps. Hence, you need to convert the data into the required format. For the algorithm to run without errors you have to make sure that all data points you collected have the same logic. There should be no inconsistencies in the datasets.

* Modeling and testing

This is the stage where the churn prediction model is prepared by the specialists. They prepare different models, test them, fine-tune them, and finally settle down with the one that is able to predict the customer churn in the most accurate manner.

The most commonly used models for predicting customer churn is one from the classic machine learning models. The list is long but the few worth mentioning here are logistic regression, decision trees and random forest. It totally depends on each company and their specific business which predictive model they would use.

* Deployment and monitoring

Finally, after the selected model has gone through enough testing they have to be deployed into production. The data scientist can either incorporate that into an existing software or can deploy it into the core of a program.

After a successful deployment, the data scientists need to constantly monitor its performance in production. The manual verification of the predictions made by the software would help realize the need of further improvements. For the companies where the data becomes outdated too soon, they need frequent testing on model performance.

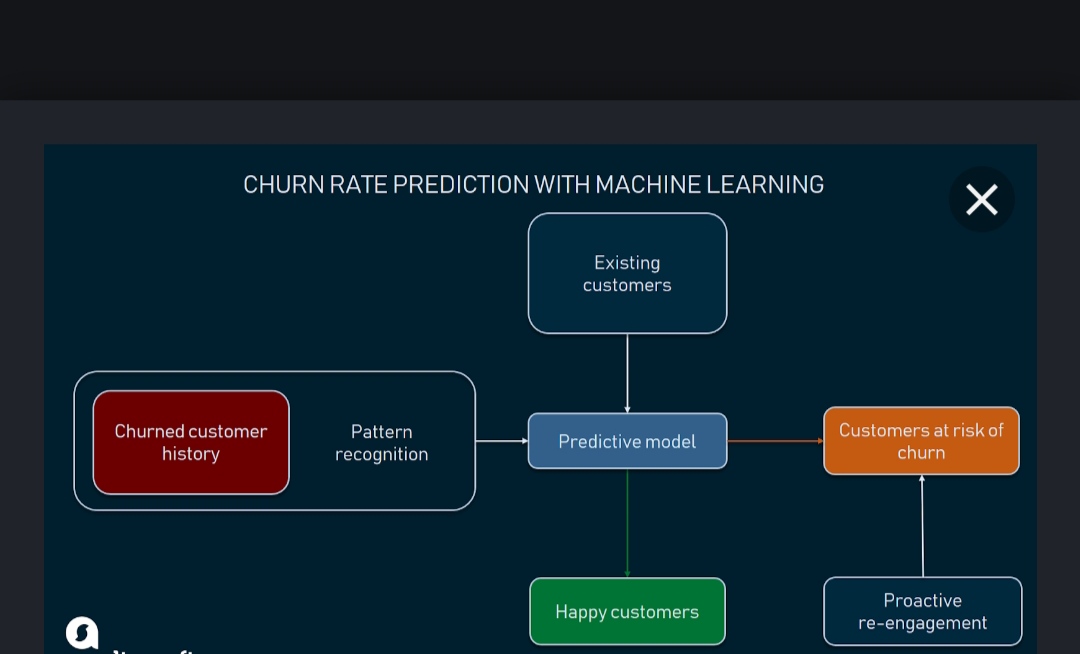
* Wrapping up

Churn prediction is one of the most sought after features for subscription based businesses. Gone are the days when you could depend only on CRM to improve customer retention. With easy access to so much customer data these days, the benefits of customer churn prediction software can be leveraged like anything.

Companies with large customer bases cannot simply rely on traditional CRM or servicing software anymore. A comprehensive customer success platform may fulfil the need of arresting customer churn of large as well as small enterprises. With Machine Learning spanning across wide areas of data analysis in almost all industries, the SaaS businesses are among the top ones who can make the most out of what ML software have to offer.

**3. THEORITICAL ANALYSIS**

**3.1. Block Diagram:-**



**3.2. Hardware/software designing:-**

Hardware

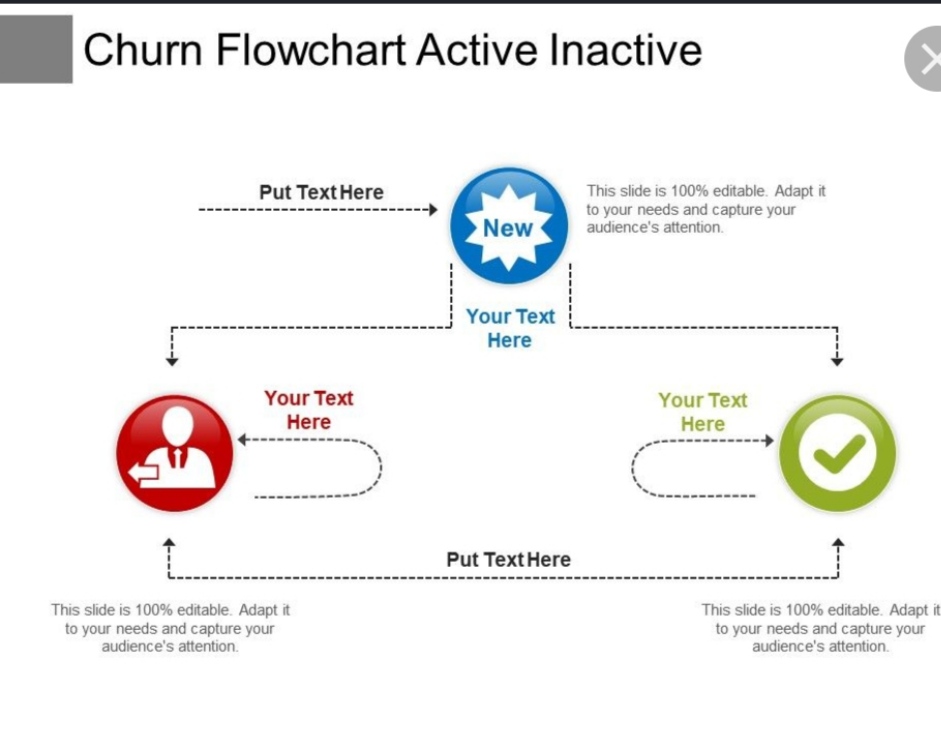
1. OS – Windows 7,8 or 10 (32 or 64 bit)
2. RAM – 4GB

Software

1. Python IDLE
2. Anaconda – Jupiter Notebook

**4.EXPERIMENTAL INVESTIGATIONS:-**

Decision tree, neural network and logistic regression were applied frequently as models of customer churn prediction, but the application of them has been mature and they are difficult to be improved. In this paper, Bayesian Networks, Support Vector Machines, Rough Sets and Survival Analysis were selected for experimental comparison study. An integrated contrast among the four models from the applicability of model in theory and experimental comparison has been processed. Overall, of the four models the Bayesian network model performed best while the Survival analysis did worst.

**5.FLOWCHART**

**6. RESULT:-**

The results are :-

1. Data Mining Techniques
2. Small Data set
3. Medium Data set
4. Large Data set

**7. ADVANTAGES:-**

1). Having the ability to accurately predict future churn rates is necessary because it helps your business gain a better understanding of future expected revenue

2). Predicting churn rates can also help your business identify and improve upon areas where customer service is lacking.

**DISADVANTAGES:-**

1).A customer’s lifetime value and the growth of the business maintain a direct relationship between each other.

2). More chances that the customer would churn

3). There is less in the potential for the business to grow.

**8. APPLICATIONS:-**

1). Churn prediction is used in a variety of different industries and types of business.

2). It is however, most relevant to SaaS companies and membership based businesses that charge an ongoing montly, quarterly, or annual fee for their software or service.

3).Churn prediction can be used within your business.

4). It is one of the key components of determining the lifetime value of customer.

**9.CONCLUSION:-**

The experiment result shows that Fiber Optic customers with greater monthly charges have higher influence for churn. Anticipated directions can be to predict by hybrid of classifiers which gives high accuracy and desirable results.

**10. FUTURE SCOPE:-**

The future scope of this paper will use hybrid classification techniques to point out existing association between churn prediction and customer lifetime value. The retention policies need to beconsidered by selecting appropriate variables from the dataset. The passive and the dynamic nature of the industry ensure that data mining has become increasingly significant aspect in the telecommunication industry prospect.

**11. BIBILOGRAPHY APPENDIX:-**

|  |  |
| --- | --- |
| Analytics | The quantities information inputs that use past data to identify possible trends that may provide valuable insight for future action. |
| Association Rules | Unsupervised learning technique to find similarities in data items based on frequent item. |
| Classification | Supervised leaning technique that identifies to which of a set of categories a observations whose category membership is known |
| Clustering | Unsupervised learning technique to find similarities based on the proximity of features in a dataset. |
| Co linearity | Refers to a linear relationship between two or more independent variables. |
| Correlation | Statistical relationship that involves dependence and is most often used in reference to a linear relationship. |