International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 01 | January - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Customer Churn Prediction by Using Machine Learning

Guide: - Prof S.V. Bodake

Associate Professor, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon bk, Pune, Maharashatra, India.

Students: -

1)Harshal Takade

2)Shreeyash Kotgiri

3)Kartikeyan Charkupalli

4)Vinayak Swami

Student at Sinhgad College of Engineering, Vadgaon bk, Pune, Maharashatra, India

Abstract:-

It becomes a significant challenge to predict customer behaviour and retain an existing customer with the rapid growth of digitization which opens up more opportunities for customers to choose from subscription-based products and services model. Since the cost of acquiring a new customer is five-times higher than retaining an existing customer, henceforth, there is a need to address the customer churn problem which is a major threat across the industries. Considering direct impact on revenues, companies identify the factors that increases the customer churn rate. Here, key objective of the paper is to develop a unique Customer churn prediction model which can help to predict potential customers who are most likely to churn. Here, we evaluated and analysed the performance of various machine learning algorithms and identified the most suitable algorithm for telecommunication dataset .The selection of algorithm changes as per dataset provided, working of telecommunication dataset is considered here .To deal with such real-world problems, Paper emphasizes the Model interpretability which is an important metric to help customers to understand how Churn Prediction Model is making predictions.(accuracy measurement techniques).

Introduction:

I today's world service industry is a big industry in the world. Every business want more and more customers for their services and marketing their services in best possible way. By this more and more cutomers are acquired ,but customers churn rate also plays an important role in revenue of the company and adversely impact the organization.

Customer churn, the bane of the telecom industry, refers to the rate at which subscribers terminate their service and switch providers. It's a critical issue, as acquiring new customers is expensive, while retaining existing ones is much more cost-effective. A high churn rate can significantly impact a company's profitability and market share.

There are various reasons for this like competitive prices from rival companies, customer service, billing issues, customers might feel their current plan doesn't offer enough value for the price, leading them to seek better deals elsewhere.

The impact is significant and affects the organization's revenue ,brand reputation and higher marketing costs in order to acquire new customers.

In order to tackle the problems, organization can offer competitive prices, improve customer support and etc, but for this identification of customer will churn or not is important.

This paper focuses on the identification of churning customers and help the organizations to identify unsatisfied customers and efficiently retain them.

Machine learning can be considered as the effective application of the artificial intelligence, which has been widely used by the telecom industries in evaluating and nullifying the customer churn. Support vector machine learning is one vital machine learning algorithm that efficiently performs the data analysis for predicting the churn

Literature Survey:

Existing papers on churn prediction uses various approaches and diverse datasets. Customer Churn prediction is one of the challenging problem in telecommunication industry.

First paper we studied was customer churn prediction by using particle swarm optimization and machine learning approaches by Samina Kanwal, Muhammad Wasif Nisar, Junaid Rashid, Jungian Kim, Saba Batoo. This paper focuses on PSO for feature selection and four most powerful machine learning machine learning algorithms in customer churn prediction are decision tree, k-nearest neighbour, Gradient Boosted Tree, Naive Bayes. The proposed methodology initially employs classification algorithms to categorize churn customer data, with the Gradient Boosted Tree, Decision Tree, k-NN, and Naive Bayes performing well in

accuracy, achieving 93 percent, 90 percent, 89 percent, and 89 percent, respectively. The experimental findings showed that the Gradient Boosted suggested methodology outperformed by obtaining an overall accuracy of 93 percent and precision of 87 percent, which shows the effectiveness of the proposed method.[1]

The paper proposed by Raja Gopal Kesiraju VLN, P. Deeplakshmi having title, Dynamic Churn Prediction using Machine Learning Algorithms - Predict your customer through customer behaviour. This paper mainly focuses on the QoS (Quality of service) aspect of customer churn prediction. Also, as the title suggests the behaviour intension of customer is influenced by various factors like perceived value, perceived playfulness, performance expectancy etc. This paper suggests to work with support vector machine algorithm because, it encompasses with a series of supervised learning methods for separating the data points. Support vector machine algorithm is one of the powerful prediction method for identifying the churn rate. In contrast to the traditional churn prediction methods, SVM allows the problem solution to depend upon the subsets of the data set, which rovides comparative computational advantages to the technique.[2]

Another paper proposed, Methodologies used for Customer Churn Detection in Customer Relationship Management by Jajam Nagaraju, Vijaya J have focus on CRM(customer relationship model) bu using ,machine learning and deep learning facilities. This paper consists of usage of various machine learning algorithms like XG boost, linear regression, decision trees, random forest, naïve bayes, K-nearest neighbour, support vector machine and ensemble learning. Deep learning methods used are Convolutional Neural Network, General Adversarial Network, Radial Basis Function Network, Multilayer Perception, Self-Organizing Maps, Restricted Boltzmann Machines, Variational Auto Encoder. At last , it is observed that the DT with feature selection technique yields extremely good performance with best accuracy value. The NB approach, on the other hand, just provides average accuracy & ROC Area. The ANN approach yields somewhat better results than the NB technique. However, it is still inferior to the DT approach. We can now see that for those three approaches, test choice 10-folds cross-validation delivers a best outcome than a % split of 70:30.[3]

Methodology:

1. About dataset: -

The sample data tracks a fictional telecommunications company, Telco. It's customer churn data sourced by the IBM Developer Platform. It includes a target label indicating whether the customer left within the last month, and other dependent features that cover demographics, services that each customer has signed up for, and customer account information. It has data for 7043 clients, with 20 features.

These features can also be subdivided into:

1. Demographic customer information:

• Gender, SeniorCitizen, Partner, Dependents

2. Services that each customer has signed up for:

• PhoneService, MultipleLines, InternetService , OnlineSecurity , OnlineBackup , DeviceProtection , TechSupport , StreamingTV , StreamingMovies,

3. Customer account information: • Tenure, Contract, PaperlessBilling , PaymentMethod , MonthlyCharges , TotalCharges

3. Functional Analysis: -

Model Selection: -

The proper algorithm will be selected after tuning of parameters. This helps in increasing performance and efficiency of prediction model.

Data Input and Output: -

The user can give a dataset as input and the model can give the list of customers that are likely to be churn.

4. Project Process Modelling: -

We have selected waterfall model for software development for clear and well defined requirements, predictable timelines and budget, clear accountability and responsibility and comprehensive documentation.

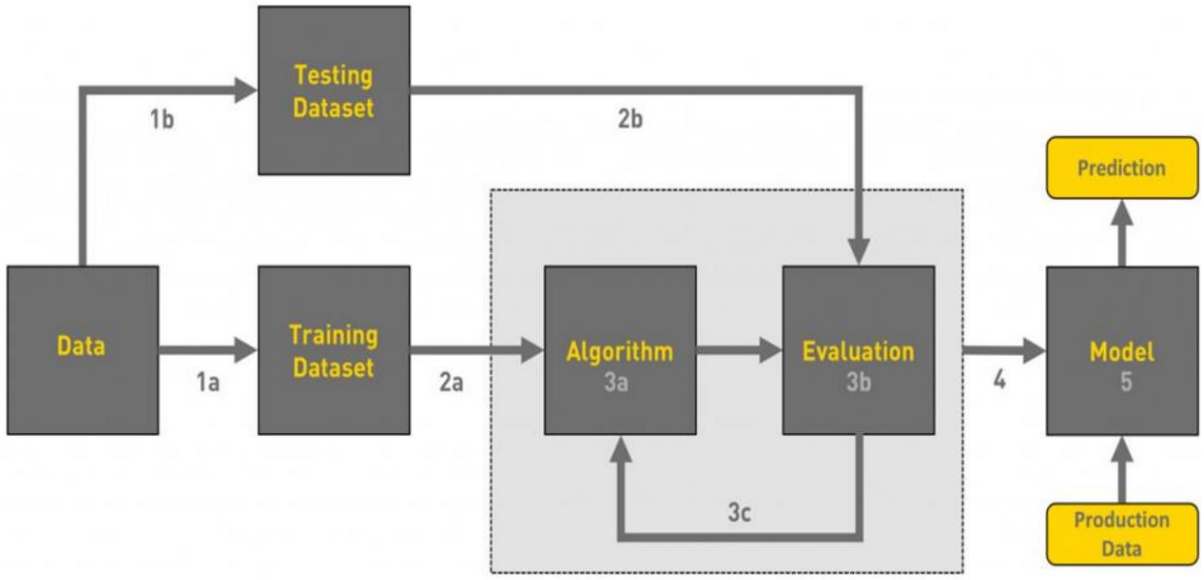
It involves following steps:

1. Requirement Phase: -Where all requirements are gathered
2. System design Phase_ - Involves designing the system that satisfies all the requirements
3. Implementation Phase: -It involves putting everything discussed till now into coding thing.
4. Testing Phase: -Testing of each and every component is done here

5. Deployment Phase: -This is the stage in which the software is deployed to the end user.

6. Maintenance phase: - Once a project is deployed, there may be instances where a new bug is discovered, or a software update is required

Proposed system:



The system is ML based application and the project aims to build a comprehensive system for prediction of customer churn for telecommunication industry by using various machine learning algorithms K-nearest neighbour, decision trees, Naïve bayes, Gradient Boosted, logistic regression, SVM so we compared the metrics and developed a more precise system.

Key components: -

1 User Interface: -

Has a user friendly interface that can interact with user with efficiency. User can give input of data set and result can be seen on the screen after evaluation through backend.

2.Dataset

We have considered the IBM dataset provided on website for default dataset and predicted the result for the same.

3.Train, Test, Split: -

Splitting the dataset int o 70-30 or 75-25 percentage. A larger portion of data for training purpose and smaller portion of data for testing.

4.Algorithms: -

a. K-nearest neighbour: -

K-Nearest Neighbours (K-NN) is a versatile machine learning algorithm used for classification and regression tasks. It works by identifying the 'k' nearest data points in the feature space to make predictions. K-NN is effective for small to medium-sized datasets and serves as a baseline algorithm for quick assessments and exploratory data analysis. Its performance depends on selecting an appropriate 'k' value and distance metric.

b. SVM (Support Vector machine): -

A Support Vector Classifier (SVC), also known as a support vector machine (SVM) for classification, is a powerful machine learning algorithm for solving binary and multi-class classification problems. It works by finding an optimal hyperplane that maximizes the margin between data points of different classes

c. Logistic Regression: -

Logistic regression is a statistical model commonly used for binary classification tasks, such as predicting whether an email is spam or not. It estimates the probability of an event occurring based on input features. The model employs the sigmoid function to squash the output between 0 and 1, representing the probability. A linear combination of feature values is passed through this function, and the resulting probability is compared to a threshold (typically 0.5) to make a binary prediction.

d . Naïve Bayes: -

A Naive Bayes Classifier is a simple and probabilistic machine learning algorithm used for classification and text categorization tasks. It is based on Bayes' theorem, which calculates the probability of a particular event happening given the probability of another event

e. Random Forest: -

combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

5.Evaluation: -

Model evaluation is done by using confusion matrix, accuracy, f1 score, precision, recall.

Conclusion: -

In conclusion, our project on customer churn prediction has provided valuable insights and solutions for businesses aiming to mitigate customer attrition. Through extensive data analysis, feature engineering, and the application of various machine learning algorithms, we have developed an effective churn prediction model. Our model's performance metrics, including accuracy, precision, and recall, indicate its ability to accurately identify potential churners, enabling businesses to take proactive measures to retain these customers. This predictive capability is crucial in reducing revenue loss and maintaining a loyal customer base.

References: -

[1] Samina Kanwal, Muhammad Wasif Nisar, Junaid Rashid, Jungian Kim, Saba Batool, “An Attribute Weight Estimation Using Particle Swarm Optimization and Machine Learning Approaches for Customer Churn Prediction”, DOI: 10.1109/ICIC53490.2021.9693040. 2021

[2] Raja Gopal Kesiraju VLN, P. Deepla kshmi, “Dynamic Churn Prediction using Machine Learning Algorithms - Predict your customer through customer behaviour”, 2021, Jan. 27 – 29, 2021, Coimbatore, INDIA

[3] Lawchak Fadhil Khalid, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, Falah Y. H. Ahmed, “Customer Churn Prediction in Telecommunications Industry Based on Data Mining”, August 2021.

[4] Jajam Nagaraju, Vijaya J., “Methodologies used for Customer Churn Detection in Customer Relationship Management”, November 2021.

[5] Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal, Ahsan Rehman, "Telecommunication Subscribers' Churn Prediction Model Using Machine Learning," IEEE International Conference on Digital Information Management (ICDIM), 2013 Eighth International Conference on, 2013, pp. 131–136