International Journal of Advanced Development in Science and Technology
Volume : 5 Issue : 02 May 2023 www.ijadst.com ISSN No: 2582-1059

TELECOM CUSTOMER CHURN PREDICTION USING SUPERVISED MACHINE LEARNING TECHNIQUES

[1] Anujha SB, [2] Kiran Koushika R, [3] Reehana S, [4] E. Archana,
[1,2,3] UG Scholar, [4] Assistant Professor,
[1,2,3,4] Department of Computer Science and Engineering,
[1,2,3,4] Panimalar Institute Of Technology

ABSTRACT

One of the biggest issues and top concerns for big businesses is customer churn. Companies are attempting to create methods to predict possible customer churn due to the direct impact on the revenues of the companies, especially in the telecom area. In order to reduce customer churn, it is crucial to identify the variables that contribute to this churn. Our work's primary contribution is the creation of a model for predicting customer turnover that helps telecom carriers identify those clients who are most likely to experience churn. The objective of this research is to present a novel method for identifying potential customers who might leave in order to design marketing retention tactics accordingly. The historical dataset is gathered, and the machine learning model is built using that dataset.

1. INTRODUCTION

The goal is to develop a machine learning model for Telecom Customer Churn Prediction, potentially replacing the updatable supervised machine learning classification models by predicting results with the best accuracy by comparing supervised algorithms.

1.1. SCOPE OF PROJECT

The main scope of the project is to predict the Telecom Customer Churn , with the help of supervised machine learning Techniques. The model build will take many attributes that are customer details and give us the status of the customer i.e if the customer is likely to churn or stay in the same Telecom Service Provider.

1.2. PROJECT GOALS

- Exploration data analysis of variable identification
  - Loading the given dataset
  - Import required libraries packages
  - Analyse the general properties
  - Find duplicate and missing values
  - Checking unique and count values
- Uni-variate data analysis
  - Rename, add data and drop the data
  - To specify data type
- Exploration data analysis of bi-variate and multi-variate
  - Plot diagram of pair plot, heatmap, bar chart and Histogram
- Method of Outlier detection with feature engineering
  - Pre-processing the given dataset
  - Splitting the test and training dataset
  - Comparing the Support Vector Machine and Logistic regression model and random forest and Decision Tree Etc.
- Comparing algorithm to predict the result
  - Based on the best accuracy

2. LITERATURE REVIEW

PAPER 1: Churn Prediction in Telecommunications using Logistic Regression and Logit Boost

AUTHORS: Hemalata Jain, Ajay Khunteta, Sumit Srivastava

ABSTRACT: Today in every industry weather, it is ISP, IT products, social network or mobile services there is the problem of customer churn (Customers changing their services from one service provider to another). However, in telecommunication the customers churning very frequently. As the market in telecom is fiercely competitive, in that case, companies proactively have to determine the customers' churn by analyzing their behavior and try to put effort and money in retaining the customers. In this proposed model, two machine-learning techniques were used for predicting customer churn: Logistic regression and Logit Boost. Experiment was carried out in the WEKA Machine-learning tool, along with a real database from an American company Orange. The results were shown in different evaluation measures.

MERIT: Using Logistic Regression makes it easier to implement, interpret, and very efficient to train.

DEMERIT: Using Logistic Regression constructs linear boundaries.

PAPER 2: Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping
AUTHORS: Tianpei Xu, Ying Ma and Kangchul Kim
ABSTRACT: In recent years, the telecom market has been very competitive. The cost of retaining existing telecom customers is lower than attracting new customers. It is necessary for a telecom company to understand customer churn through customer relationship management (CRM). Therefore, CRM analyzers are required to predict which customers will churn. This study proposes a customer-churn prediction system that uses an ensemble-learning technique consisting of stacking models and soft voting. Xgboost, Logistic regression, Decision tree, and Naïve Bayes machine-learning algorithms are selected to build a stacking model with two levels, and the three outputs of the second level are used for soft voting. Feature construction of the churn dataset includes equidistant grouping of customer behavior features to expand the space of features and discover latent information from the churn dataset. The original and new churn datasets are analyzed in the stacking ensemble model with four evaluation metrics. The experimental results show that the proposed customer churn predictions have accuracies of 96.12% and 98.09% for the original and new churn datasets, respectively. These results are better than state-of-the-art churn recognition systems.

MERIT: When the dataset consists of both linear and non-linear data, Ensemble methods are very useful.

DEMERIT: This method is less interpretable and it is hard to predict and explain the output of the ensembled model.

3. EXISTING SYSTEM

User churn is a significant problem for Internet businesses that threaten their future viability and profitability. The majority of earlier studies on churn prediction reduce the issue to a binary classification challenge in which customers are classified as churned or non-churned. Recently, some studies have attempted to translate the forecast of user churn into the prediction of user return time. In this strategy, which is more accurate for real-world online services, the model forecasts the user return time at each time step rather than a churn label. The earlier studies in this field, however, lack generality and demand complicated computations. In the current study, we introduce an oracle that forecasts user churn by simulating user return rates to services using a combination of Temporal Point Processes and Recurrent Neural Networks. In addition, we add latent variables to the suggested recurrent neural network to simulate the hidden system user loyalty.

Disadvantages

- Machine learning prediction methods are not being used.
- Their model's precision is not being obtained.
- The deployment phase is not being implemented.

4. PROPOSED SYSTEM

The proposed method involves applying machine learning to create a prediction of telecom customer churn. We intend to create an AI-based model. Data are required to train our model. Therefore, the dataset for telecom customer churn can be used to train the model. We must comprehend the training intentions to make use of this dataset. The intent is the purpose for which a user interacts with a predictive model or the purpose for which each piece of data that a certain user provides to the model is provided. These intents may differ from one another depending on the domain for which you are designing an AI solution. The aim is to create training samples for each intent, define various intents, and then train the AI model using the intents as model training categories and the training sample data as model training data. The vectorization technique, which uses the vectors to comprehend the data is used to develop the model. We can achieve a better AI model and the best accuracy by using multiple algorithms. Following model construction, the model is assessed using a variety of measures including confusion metrics, precision, recall, sensitivity, F1 score, and others.

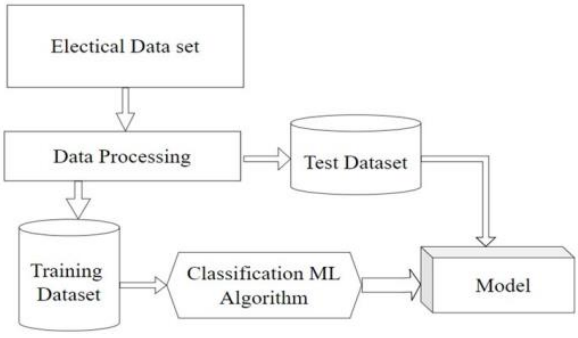

Fig 1 Architecture of Proposed Model

Advantages

- We are developing a predictive AI algorithm for predicting telecom customer churn.
- To increase accuracy, we are implementing machine learning algorithms.
- The outcome estimations can be calculated to improve performance outcomes.

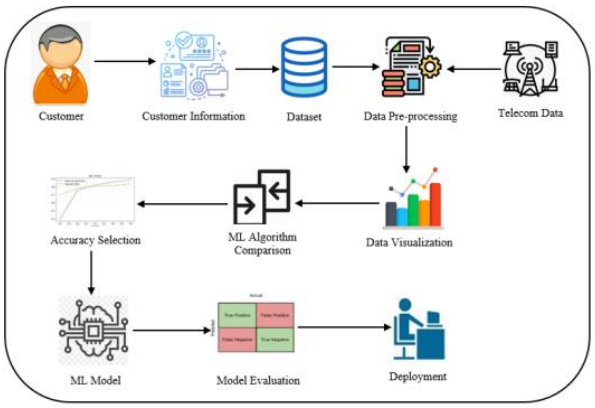

Fig 2 System Architecture

5. CONCLUSION

The analytical process started from data cleaning and processing, missing values analysis, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out. This project can help to find the Prediction of Telecom Customers Churn Details.

6. FUTURE ENHANCEMENT

- ✓ Feature engineering: Reviewing the features that are currently using in this model and adding new features that might be relevant to this prediction task. Incorporating customer demographic data or information about their past interactions with the company.
- ✓ Model tuning: Experiment with different algorithms and hyperparameters to find the best performing model
- ✓ Implementing an online learning system which allows the model to continuously learn from new data as it becomes available.
- ✓ Integration: Integrating the model into other systems and processes within the organization. For example, we might use the predictions to inform marketing campaigns or prioritize customer support tickets.
- ✓ By continuously refining and improving the model, we can ensure that it remains relevant and effective for business needs.

REFERENCES

[1] Jain, H., Khunteta, A. and Srivastava, S., 2020. Churn prediction in telecommunication using logistic regression and logit boost. Procedia Computer Science, 167, pp.101-112.

[2] Xu, T., Ma, Y. and Kim, K., 2021. Telecom churn prediction system based on ensemble learning using feature grouping. Applied Sciences, 11(11), p.4742.v

[3] Ly, T.V. and Son, D.V.T., 2022. Churn prediction in telecommunication industry using kernel Support Vector Machines. Plos one, 17(5), p.e0267935.

[4] Sundararajan, A. and Gursoy, K., 2020. Telecom customer churn prediction.

[5] Jain, H., Khunteta, A. and Srivastava, S., 2022. Telecom Churn Prediction Using an Ensemble Approach with Feature Engineering and Importance. International Journal of Intelligent Systems and Applications in Engineering, 10(3), pp.22-33.

[6] Melian, D.M., Dumitrache, A., Stancu, S. and Nastu, A., 2022. Customer Churn Prediction in Telecommunication Industry. A Data Analysis Techniques Approach. Postmodern Openings, 13(1 Sup1), pp.78-104.

[7] Peng, K. and Peng, Y., 2022. Research on Telecom Customer Churn Prediction Based on GA-XGBoost and SHAP. Journal of Computer and Communications, 10(11), pp.107-120.

[8] Sudharsan, R. and Ganesh, E.N., 2022. A Swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. Connection Science, 34(1), pp.1855-1876.

[9] Ali, M., Rehman, A.U., Hafeez, S. and Ashraf, M.U., 2018, August. Prediction of churning behavior of customers in telecom sector using supervised learning techniques. In 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE) (pp. 1-6). IEEE.

[10] Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S.U. and Kim, S.W., 2019. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE access, 7, pp.60134-60149.

[11] Bhanuprakash, P. and Nagaraja, G.S., 2017, December. A review on churn prediction modeling in telecom environment. In 2017 2nd International Conference on Computational Systems and Information

Technology for Sustainable Solution (CSITSS) (pp. 1-5).
IEEE.