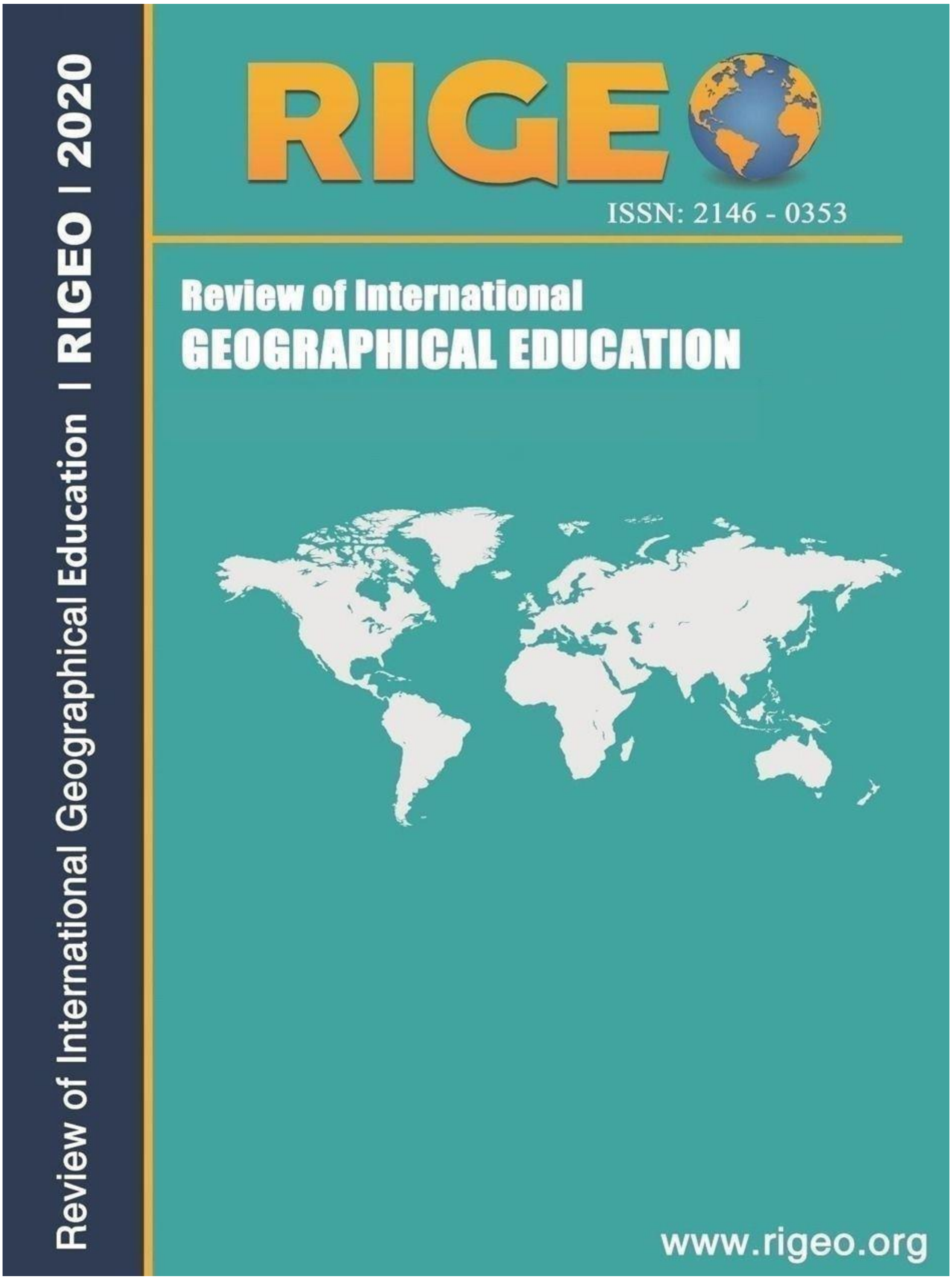

Review of International Geographical Education I RIGEO I 2020
RIGEO
ISSN: 2146 - 0353
Review of International
GEOGRAPHICAL EDUCATION
www.rigeo.org

CUSTOMER CHURN PREDICTION IN TELECOM INDUSTRY USING MACHINE LEARNING

[1] K.Lakshmi Kalpana, [2] G.Sowjanya,[3] M.Vinay kumar, [4] MACHARLA PREM KUMAR
[1,2,3] Assistant Professors,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505
[4]student,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

ABSTRACT

Research into customer churn detection is vital for telecommunications firms as it assists them in retaining their current clientele. Client churn occurs when services offered by competitors are discontinued or when there are problems with the network. The customer has the ability to terminate their membership at any time. Since the churn rate impacts both the length of service and the future revenue of the organization, it has a significant impact on the lifetime value of a customer. Companies want a model that can forecast customer attrition since it has a direct impact on industry revenue. The model in this research was built using machine learning techniques. Using machine learning algorithms, we can determine which customers are most prone to canceling their subscription.

I.INTRODUCTION

Developed nations rely heavily on the telecommunications sector.A major problem for service businesses is customer churn, which occurs when important clients leave to compete with other businesses. The degree of resistance was raised by both technological advancement and the expansion of operators. Companies are using intricate tactics to stay afloat in this challenging economy. A major issue arises when customers go, leading to a dramatic decrease in communication services. There are three primary ways to boost sales: by attracting new consumers, by upselling to existing ones, and by keeping existing clients as

customers. Looking at the return on investment (RoI) for each strategy, the third one shown that retaining current customers is far simpler than finding new ones, and it also indicated that upselling methods are significantly more expensive than maintaining existing customers. For businesses to put the third tactic into action, they must lower customer churn, or "the customer movement from one provider to another."Due to their aggressive and great customer service, service industries frequently have significant customer churn rates. Giving out early predictions of which consumers are going to depart the company may be a very profitable side activity. Several studies have shown that this kind of prediction is well-suited to machine learning. This approach is based on reviewing historical data.

Objective: This study's overarching objective is to develop a machine learning-based model for identifying telecom consumers most likely to terminate their service. The algorithm takes into account consumer demographics, use patterns, contract details, and historical churn data in an effort to accurately predict which consumers would terminate their subscriptions. Helping telecom companies enhance revenue output, boost customer lifetime value, and proactively retain at-risk consumers is the ultimate goal.

II.EXISTING SYSTEM

One approach to predicting client churn is data mining. Discovering, forecasting, and retaining churn customers may be achieved with the use of these tactics.

Industries also benefit from their assistance with CRM and decision-making. Decision trees are often used to gauge client retention rates; nevertheless, they fail miserably when faced with intricate problems. The study found that reducing data improved the accuracy of decision trees.

Data mining algorithms have many applications, such as analyzing historical trends and purchasing patterns to forecast future purchases. Regression tree principles were discussed with other well-known data mining methodologies including neural networks, decision trees, and rule-based learning.

Negative aspects: Decrease in efficiency

Decreased accuracy

III.PROPOSED SYSTEM:

Here, we use a slew of ML algorithms to predict client attrition with pinpoint accuracy. Here, we apply the model to a dataset that has already been trained and tested, which produces the most accurate results. the proposed model for churn prediction and details its methods. As a preliminary step before feature selection, data preparation involves filtering data and transforming it into a similar format.

Classification and prediction are the next steps, implemented using a variety of techniques. During the model's training and testing on the dataset, we assess and monitor the customer's actions. Last but not least, we analyze the data to predict customer churn.We proposed using ML techniques such as Random Forests and SVMs for this. Among these two algorithms, one will provide the most accurate prediction of the test data.Then we'll know for sure whether the customer is churned.

Advantages:

Better Predictions

Greater Accuracy

IV.METHODOLOGY

Data Collection:

The first step in the methodology involves gathering data from the telecom company's customer database. This dataset typically includes customer demographics, service usage details, billing information, customer support interactions, contract details, and historical churn data (i.e., whether a customer has left the service). The data can be sourced from the company’s internal databases, CRM systems, or through data partnerships. Key data sources include internal databases and CRM systems, with attributes such as customer ID, age, gender, location, service tenure, contract type, monthly charges, payment method, and the number of support calls.

Data Preprocessing:

Preprocessing is a critical step to prepare the data for analysis and modeling. This includes data cleaning, which involves handling missing values, correcting data entry errors, and removing duplicates. Data transformation is also performed by converting categorical variables (e.g., contract type, payment method) into numerical formats using techniques like one-hot encoding. Additionally, feature scaling is applied to normalize or standardize numerical features to ensure equal contribution to model performance. If there is a class imbalance (e.g., fewer churn cases than non-churn), techniques such as oversampling (e.g., SMOTE) or undersampling are used to balance the dataset.

Exploratory Data Analysis (EDA):

EDA is conducted to understand the relationships between different features and the target variable (churn). This step includes performing descriptive statistics to summarize the central tendencies, variability, and distribution of the data. Data visualization techniques, such as histograms, box plots, scatter plots, and correlation heatmaps, are used to identify patterns and correlations. Feature selection is then carried out to identify the most relevant features that strongly influence churn, possibly using statistical tests or feature importance scores from models.

Model Selection:

In this step, multiple machine learning algorithms are considered to build predictive models for churn prediction. Common models include Logistic Regression, which provides interpretable results and highlights the influence of different factors on churn; Decision Trees, which split data into branches for decision-making; Random Forest, an ensemble of decision trees that improves prediction accuracy; Gradient Boosting Machines (e.g., XGBoost or LightGBM), which are powerful for capturing complex patterns; Support Vector Machines (SVM), effective for high-dimensional spaces with non-linear decision boundaries; and Neural Networks, deep learning models that can capture intricate patterns in large datasets. These models are trained using the labeled dataset and evaluated to select the best-performing one.

Model Training and Evaluation:

The selected models are trained on the preprocessed dataset. The dataset is typically split into training and testing sets, using an 80-20 or 70-30 split. To ensure robustness and avoid overfitting, k-fold cross-validation is employed. The models are then evaluated using performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic Curve), which help assess how well the model predicts churn versus non-churn. The model with the best performance on the test set is chosen for further tuning.

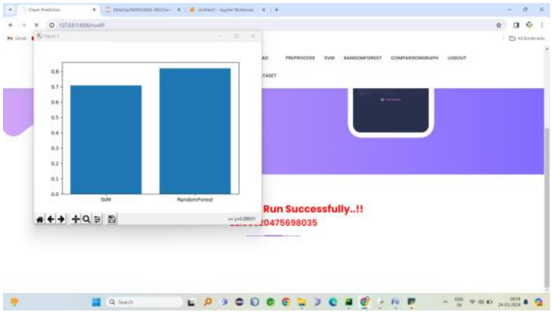
Hyperparameter Tuning:

After selecting the best model, hyperparameters are optimized to enhance performance. Techniques such as Grid Search or Random Search are used to identify the optimal set of hyperparameters that maximize the model’s predictive power. This step ensures the model is fine-tuned to provide the best possible results.

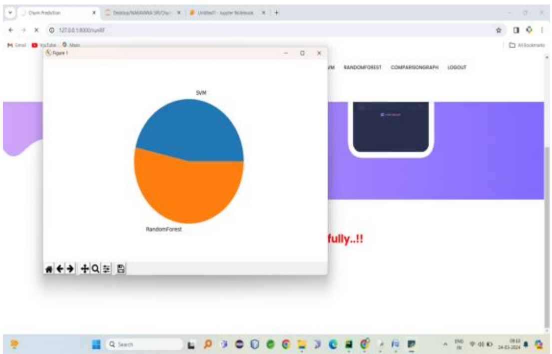
Model Deployment:

Once the model is tuned and validated, it is deployed into the telecom company's production environment. This involves integrating the model into the company's customer management system to provide real-time churn predictions. Continuous monitoring of the model’s performance is essential to ensure it maintains accuracy. Periodic retraining may be necessary as new data becomes available, allowing the model to adapt to changes in customer behavior.

Comparison graph



Comparison in pie graph



IV.CONCLUSION

Churn prediction is a boon to many businesses, especially those in the telecommunications industry, in terms of revenue and profitability. Organizations in the telecom business emphasize customer retention above customer acquisition since predicting client attrition is the industry's biggest difficulty. The versatility and adaptability of three tree-based algorithms made them ideal for this task. Outperforming rival technologies, Random Forest and Support Vector Machines provide more accurate outcomes. Here, we are assessing the worth of hem and producing a precise forecast by looking at a dataset that contains information about specific customers' service plans. This approach may be more effective in identifying customers who are likely to switch to other services offered by the business. The telecom company can see everything plainly and provide them deals to stay with that service.Thanks to the deployment of ML approaches, our suggested churn model surpassed the competition, according to the findings.Random Forest produced the most accurate findings. To improve our customer churn forecast, we will do more research on lazy learning techniques in the next days.Using AI approaches for trend analysis and consumer prediction, the study might be broadened to get a better understanding of customers' growing behavior.

V.FUTURE SCOPE

Natural language processing (NLP) and other sophisticated analytics methods may be included into future versions of churn prediction models to analyze customer input from other sources, including social media, contacts with customer care, and surveys. The accuracy of churn forecasts may be further improved with the help of this integration, which can provide deeper insights into customer mood and preferences.

Using Real-Time Data: With the proliferation of real-time data from sources including Internet of Things (IoT) devices, network logs, and location data, telecoms businesses may improve the accuracy and timeliness of their churn prediction models in the future. Preventing customer churn is possible with proactive intervention tactics made possible by real-time monitoring of consumer behavior and network performance.

Future churn prediction algorithms may zero down on creating individualized plans to keep consumers by taking into account their specific tastes, habits, and point in the customer lifecycle. Telecommunications providers may improve customer retention by creating segmented lists of customers and using machine learning algorithms for recommendation and segmentation systems to provide personalized offers and incentives.

VI.REFERENCES

1. V. Lazarov and M. Capota, ‘‘Churn prediction,’’ Bus. Anal. Course,TUM Comput. Sci, Technische Univ. München, Tech. Rep., 2007.[Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.7201&rep=rep1&type=pdf.

2.Vadakattu, B. Panda, S. Narayan, and H. Godhia, ‘‘Enterprisesubscriptionchurnprediction,’’inProc.IEEEInt.Conf.BigData,Nov.2015, pp. 1317–1321.

3.S.A.Qureshi,A.S.Rehman,A.M.Qamar,A.Kamal,andA.Rehman,‘‘Telecommunication subscribers’ churn prediction model usingmachinelearning,’’inProc.8thInt.Conf.Digit.Inf.Manage.,Sep.2013, pp.131–136

4.A Novel Approach for Churn Prediction Using Deep Learninghttps://www.researchgate.net/publication/328819998_A_Novel_Appoach_for_Churn_Prediction_Using_Deep_Learning

5.S. A Survey on Customer Churn Prediction using Machine LearningTechniques.https://www.researchgate.net/publication/310757545_A_Survey_on_Customer_Churn_Prediction_using_Machine_Learning_Techniques

6.DeepLearninginCustomerChurnPrediction:UnsupervisedFeatureLearning on Abstract Company Independent Feature Vectorshttps://arxiv.org/pdf/1703.03869.pdf

7.AutomatedFeatureSelectionandChurnPredictionusingDeepLearningModels.https://www.irjet.net/archives/V4/i3/IRJET-V4I3422.pdf

8.Effectual Predicting Telecom Customer Churn using Deep NeuralNetwork.https://www.ijeat.org/wp-content/uploads/papers/v8i5/D6745048419.pdf

9.CustomerChurnPredictionin TelecommunicationwithRotationForestMethodhttps://www.researchgate.net/publication/282981765_Customer_churn_prediction_in_telecommunication

10.Customer churn prediction in telecom usingmachinelearninghttps://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6