International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 10 August 2023 **Revised:** 27 September 2023 **Accepted:** 18 October 2023*

A Survey and Implementation of Machine Learning Algorithms for Customer Churn Prediction

Dr. Snehal Rathi[1], Atharva Puranik[2], Vaishnavi Pophale[3], Prajwal Kutwal[4] , Vibhav Kulkarni[5], Shaantanu Pratham[6], Prof. Vikas Maral[7]

[1]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
snehal.rathi@viit.ac.in
[2]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
atharva.22010869@viit.ac.in
[3]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
vaishnavi.22010473@viit.ac.in
[4]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
prajwal.22010845@viit.ac.in
[5]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
vibhav.22010011@viit.ac.in
[6]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
shaantanu.22010051@viit.ac.in
[7]Department of Computer Engineering
Vishwakarma Institute of Information Technology, Pune, India
vikas.maral@viit.ac.in

Abstract— Estimating customer traffic is an important task for businesses because it helps them identify customers who are most likely to leave and take preventative measures to retain them by improving customer satisfaction and further increasing their own revenue. In this article, we focus on developing a machine-learning model for predicting customer churn using historical customer data We performed engineering operations on the data, addressed the missing digits, coded the categorical variables, and preprocessed the data before evaluating it using a variety of performance indicators, including accuracy, precision, recall, f1 score, and ROC AUC_Score. Our feature significance analysis revealed that monthly fees, customer tenure, contract type, and payment method are the factors that have the most impact on forecasting customer churn. Finally, we conclude the best-performing model, the Soft Voting Classifier, implemented on the four best-performing classifiers with a good accuracy of 0.78 and a relatively better ROC AUC_Score of 0.82. Keywords— Customer churn prediction, Machine learning, Feature importance analysis, Gradient boosting, Business revenue.

I. INTRODUCTION

Customer churn prediction is a critical problem for companies across various industries. It refers to the task of identifying customers who are likely to discontinue using a company's products or services. From a business perspective, customer churn poses significant challenges and can have a substantial impact on a company's profitability and growth. Retaining existing customers is the most important task for the survival of the business, which has become common sense in the business world. [15]. Customer acquisition requires substantial marketing and promotional efforts while retaining loyal customers can lead to repeat business and increased customer lifetime value. Therefore, accurately predicting customer churn allows companies to proactively address the underlying issues and take appropriate measures to retain valuable customers. Customer churn prediction relies on analysing historical customer data, such as demographic information, transactional records, service usage patterns, and customer interactions. Different data types have different analysis capabilities. It is necessary to determine the most appropriate data for the type of analysis performed. Different datasets provide better metrics for different problems and services. [11] In order to find trends and signs that assist in identifying high-risk customers, the organization employs sophisticated analytics as well as machine learning approaches.

Some popular betting methods include:

1. Machine-Learning Algorithms: Gradient boosting, decision trees, random forests, logistic regression, etc. The likelihood of

additional incoming losses can frequently be predicted by monitoring machine learning systems as such.

2. Survival Analysis: Survival analysis is a statistical technique used to analyze event-time data, such as the time before losing a customer. It considers the varying time periods during which customers remain active and enables the prediction of the probability of churn over time.

3. Neural Networks: Deep learning techniques, specifically neural networks that can learn complex patterns and relationships from large volumes of data, can be used for customer churn prediction.[16],[17],[18]

4. Ensemble Methods: The integrated system combines multiple models using techniques such as packaging, promotion, and configuration to increase the accuracy of loss estimation.

By applying these techniques, companies can gain valuable insights into customer behaviour, identify early warning signs of churn, and implement proactive strategies to mitigate churn risk.

II. MOTIVATION

With the help of historical customer data, our research aims to give a thorough and methodical assessment of numerous machine learning models in order to predict customer attrition. Although there have been previous research works [1],[3],[5],[6] that involved a comparison of the performances of a few machine learning models, we aimed at comparing as many models as possible that one can use for classification. We applied a wide array of traditional(logistic regression, ridge, decision tree, naïve bayes, knn, etc) and ensemble(catboost, adaboost, xgboost, LGBM, bagging, etc) machine-learning models. We wanted to truly explore the potential of Ensemble Learning and we did so by employing a Voting Classifier which considered the predictions of the best-performing classifiers of our research and therefore provided better results. We were also interested in gaining important insights from the dataset by thorough visualization of every numerical and categorical column (via kdeplot, boxplot, and histograms) as well as getting a quantitative value for each feature signifying its importance w.r.t the target column (via the chi-square test). Also, we addressed the issue of an imbalanced target column using SMOTE. We laid greater emphasis on addressing the above specific aspects which drove us to do this research. We believe this paper will surely provide valuable insights and methodologies that can be applied in real-world scenarios as well as prove to be helpful for future researchers.

III. LITERATURE SURVEY

This is a concise outline of related research that has been proposed by notable scholars as well as churn prediction in the telecom business.

On the dataset, Dhangar et al. [1] discovered that SVM and Random Forest had the greatest accuracy rates, at 84 and 87 percent, respectively. SVM classifiers surpass others with an AUC score of 92.1 percent, while Random Forest earns the highest AUC score of 94.5 percent. Working on the Customer DNA website, Saad et al [2] highlighted the usage of the re-sampling strategy to address the issue of class imbalance. Their findings show that decision trees are the most accurate classification system when it comes to detecting losses for data analysis. Authors Adbelrahim et al. [3] Predict users using decision trees, random forests, GBM tree algorithms, and XGBoost as tree-based algorithms. Comparative research shows that XGBoost outperforms its competitors in terms of AUC accuracy. However, feature selection techniques and optimization algorithms can improve accuracy.

Tamaddoni et al. [4] found that both the simple predictive model and the cost-sensitive model were better than the comparison of both the CART model and the multi-model algorithm, and the cost-sensitive learning model obtained the model base only in the CART model but not in many models.

Praveen et al.[5] employed Logistic regression, Support Vector Machine, Pruning Tree, and Naive Bayes in their comparative study of machine-learning techniques for forecasting attrition of customers and looked into how magnification affects accurate classification.

Lalwani et al.[5] found that combining special selection techniques such as randomization can improve classification accuracy.

For predicting customer attrition, Horia Beleiu et al.[6] used three types of machine learning techniques: deep neural networks, support vector models, and Bayesian networks. Principal component analysis (PCA) was considered throughout the feature selection process, which reduced residual data. They used optimization techniques to improve the feature selection process and thereby improved accurate population classification.

The authors of K Coussement et al. [7] used support vector machines, logistic regression (LR), and random forests (RF) to attempt to model the churn prediction problem. SVM initially performed about as well as LR and RF, but when the best parameters were chosen, it surpassed LR and RF with regard to PCC and AUC.

The decision tree and logistic regression model were used in the churn prediction data set by K. Dahiya et al. [8]. WEKA tool was employed during the trial.

Authors Umman et al. [9] used decision tree and logistic regression machine learning models to analyze a large data set, but the accuracy of the results was poor. And proposed that

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 10 August 2023 **Revised:** 27 September 2023 **Accepted:** 18 October 2023*

development was, therefore, necessary before using additional machine learning and feature selection techniques.

J. Hadden et al. [10] showed that decision trees are better than other systems because of their rules. By using the existing feature selection strategy, the acquisition accuracy can be further improved.

J. Hadden et al. [11] reviewed all machine learning models considered and provided a thorough study of the methods currently used for feature selection. They discovered that decision trees outperformed the competition in the prediction models. The improvement of the prediction algorithms in feature selection is greatly aided by optimization techniques.

According to Y. Huang et al. [12], the authors used a variety of classifiers on the churn prediction dataset, and the findings showed that random forest outperforms the competition with regard to AUC and AUC(PR) analysis. However, it is possible to raise accuracy even more by employing feature extraction optimization approaches.

Genetic programming (GP) and the Adaboost machine learning model were combined by researchers working under the direction of A. Idris et al. [13] in order to compare their results with those of other classification algorithms. Adaboost and GP's results were more accurate than those of the competition. However, accuracy can be increased even more by utilizing various optimization strategies, like the gravitational search algorithm, bio-geography-based optimization, and many others.

Authors P. Kisioglu et al. [14] used Bayesian Belief Networks (BBN) to estimate client attrition. Correlation analysis and multicollinearity tests were carried out during the experimental analysis. BBN was proven to be a viable alternative for the prediction of churn. They offered suggestions for future study directions as well.

IV. METHODOLOGY

4.1 SYSTEM ARCHITECTURE:

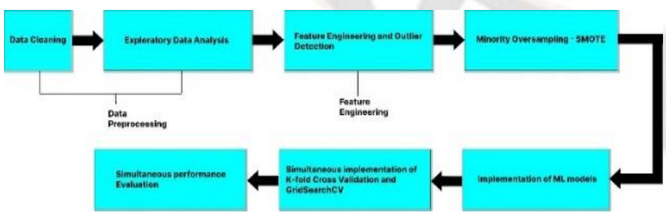

Fig. 4.1: System Architecture

Description of the system architecture:

4.2.1 Stage 1 - Data Cleaning: The initial stage of our research involves data cleaning to ensure the integrity and quality of the Telecom Churn Dataset. Through manual inspection, inconsistencies in columns like 'Total Charges' were identified and rectified by eliminating missing values.

4.2.2. Stage 2 - Exploratory Data Analysis (EDA): In the second stage of our research, we conducted Exploratory Data Analysis (EDA) with a strong emphasis on data visualization techniques. EDA involves a systematic examination of the dataset to uncover patterns, relationships, and trends in the data.

4.2.2.1. Visualization of Numerical Columns:

Probability density distribution (or Kernel Density Estimation) plots were generated for the three numerical columns: tenure,

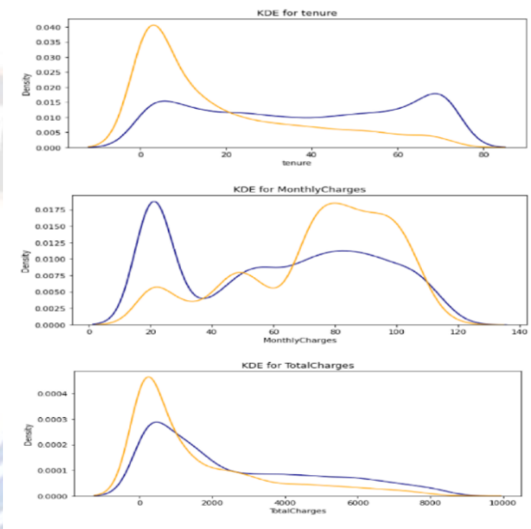

Fig. 4.2: Kernel Density Estimation (KDE) plot of the numerical columns

The conclusions derived from these plots are as follows:
1. The recent clients, indicated by shorter tenure, are more likely to churn.
2. The clients with higher monthly charges are also more likely to churn.
3. Both tenure and monthly charges are likely to be important features in predicting churn, as they show significant variations between churned and non-churned customers.

Furthermore, boxplots were generated for the same three numerical columns in Figure 4.3

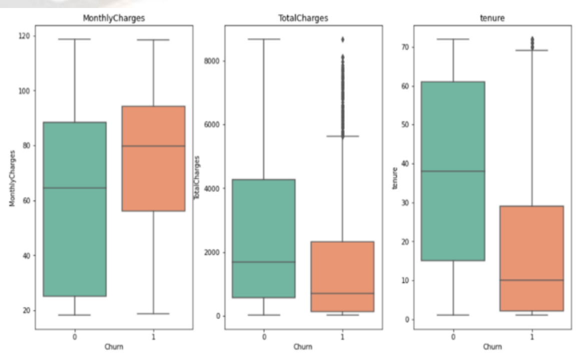

Fig 4.3: Boxplot of the numerical columns

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 10 August 2023 **Revised:** 27 September 2023 **Accepted:** 18 October 2023*

The boxplot analysis revealed additional conclusions:
1. The Monthly plan doesn't seem to be such a big driver of churn, but we can see that 75% of churners pay between 60 and 100 dollars a month
2. One important mark we can see in this plot is that 50% of Churners leave the company before the first year goes by.

4.2.2.2 Visualization of Categorical Columns: Count plots were generated for each categorical column to understand the churn rate with respect to the values within each category in Figure 4.4

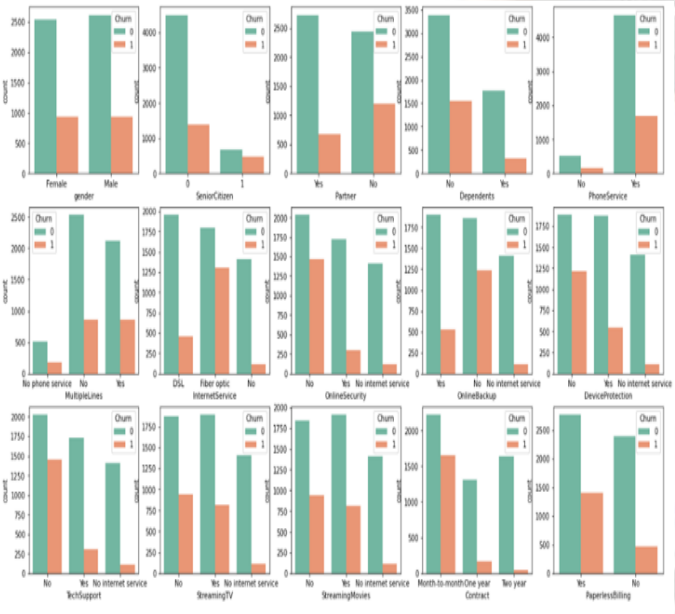

Fig. 4.4: Count plot of 'Churn' w.r.t various categories

From these count plots, several conclusions can be drawn:
1. The analysis of the "gender" column revealed that the churn percentage is relatively equal for both males and females.
2. The percentage of churn is higher in the case of senior citizens.
3. The customers who have partners and dependents showed a lower churn rate compared to those who do not have partners or dependents.
4. Customers with fiber optic internet services have a higher churn rate.
5. Customers who do not have essential services such as online security, online backup, and tech support tend to leave the platform.
6. Customers with monthly subscription contracts have a higher churn rate compared to those with one or two-year contracts.
7. Customers who have opted for paperless billing have a higher churn rate.
8. Customers who use the electronic check payment method tend to leave the platform more frequently

4.2.3 Stage 3 - Feature Engineering and Outlier Detection
Outlier detection was performed using the Interquartile Range (IQR) method and no outliers were found. During the feature engineering phase, it was observed that the "TotalCharges" column exhibited a high correlation and therefore it was dropped from the dataset. The contiguous values of the "tenure" column were grouped into five groups with a fixed interval length of 12 and a new column "tenure groups" was created. Several additional columns were deemed not influential on the "churn" column and were deleted from the dataset. The string values 'yes' and 'no' were replaced with integer values of 1 and 0, respectively. Dummy variables or one-hot encoded representations of the categorical columns were done.

4.2.4 Stage 4 - Feature Selection

In this stage, the training set was subjected to the chi-square test, a statistical technique used in feature selection that helps to identify the categorical variables that are most likely to have an impact on the target variable. F-value and p-value arrays were obtained as a consequence. A higher F-value suggests a stronger association between the feature and the target variable. A lower p-value indicates a higher level of significance of that association. From the results, we can infer that 'MonthlyCharges' and 'tenure_group' are the columns that have the most influence on the target column.

Feature/Column	p-value	F-values
MonthlyCharges	0.000000e+00	2959.466655
tenure_group	4.986197e-173	786.359563
Contract_Two year	7.367296e-88	394.825814
PaymentMethod_Electronic check	7.235235e-82	367.302852
InternetService_Fiber optic	2.550623e-67	300.510062
DeviceProtection_No internet service	1.965566e-50	223.039150
TechSupport_No internet service	1.965566e-50	223.039150
StreamingTV_No internet service	1.965566e-50	223.039150
InternetService_No	1.965566e-50	223.039150
StreamingMovies_No internet service	1.965566e-50	223.039150
OnlineBackup_No internet service	1.965566e-50	223.039150
OnlineSecurity_No internet service	1.965566e-50	223.039150
Contract_One year	1.858853e-32	140.713251
SeniorCitizen	1.619079e-28	122.703715
Dependents	1.686336e-26	113.488771
OnlineSecurity_Yes	4.165335e-25	107.131767
TechSupport_Yes	1.232245e-24	104.982417
PaymentMethod_Credit card (automatic)	2.218786e-22	94.696970
PaperlessBilling	1.663882e-20	86.154793
Partner	1.840257e-16	67.766762
PaymentMethod_Mailed check	7.647630e-10	37.848011
OnlineBackup_Yes	2.682098e-07	26.466057
DeviceProtection_Yes	1.736350e-04	14.096843
StreamingMovies_Yes	5.029738e-04	12.104608
StreamingTV_Yes	7.290385e-04	11.413716

Fig. 4.5: F-value and P-value of every column obtained via the Chi-Square test.

To mitigate the issue of multicollinearity and to ensure the independence of features, pairs of columns with a correlation coefficient greater than 0.9 were considered highly correlated and were removed.

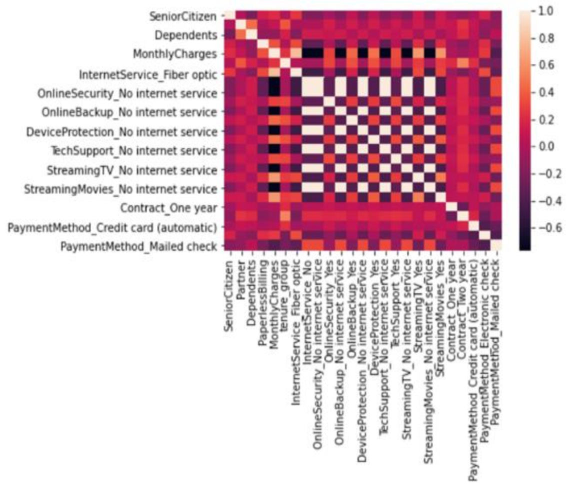

Fig 4.6: Heat Map of the Correlations in the Dataset.

4.2.5 Stage 5 - Minority Oversampling - SMOTE: Analysis of the 'churn' column revealed that 26.6% of the responses indicated churn as 'yes', while the remaining 73.4% indicated churn as 'no'. Synthetic Minority High Sampling Technique (SMOTE) was a dopted for solving the class mismatch issue.

The SMOTE algorithm follows these steps:
1) For each minority class instance, find its k nearest neighbours.
2) Select a random neighbour and interpolate between the chosen instance and the neighbour to generate a new synthetic instance.
3) Repeat the process for a desired number of synthetic instances.

Applying SMOTE resulted in an increased representation of the minority class in the dataset.

```
The num of classes before fit Counter({0: 4125, 1: 1500})
the num of classes after fit Counter({0: 3803, 1: 2978})
```

4.2.6 Stage 6 - Implementation of Machine Learning Models: After data cleaning and feature engineering, we divide the datase t into training and test sets, use 80% for training and 20% for test ing, and evaluate the features between 0 and 1 and believe that th ey are compatible.

1) Logistic Regression: Utilizes the logistic function f(z) = 1 / (1 + e^(-z)) (where z represents the linear combination of the feature values and their respective weights) and the logistic loss function.
2) Ridge Classifier: Involves solving the following equation: w = (X^T * X + alpha * I)^-1 * X^T * y. Here, X is the matrix of input features, I is the identity matrix, and y is the target variable.
3) SVM: Locates the hyperplane with the greatest margin, defined as the distance between the hyperplane and the nearest data points for each class.
4) KNN: Identifies the K nearest neighbors based on the chosen distance metric and determines the majority class among them.
5) Naive Bayes: Assumes that the features are conditionally independent of each other, learns the class probabilities (P(y)), and calculates the likelihood probabilities (P(X|y)) based on the observed features in the training data.
6) Multi-Layer Perceptron: Uses a process called backpropagation to optimize the weights of the connections between neurons iteratively. Uses optimization algorithms like gradient descent.
7) Decision Tree: The algorithm continues splitting the data at each node, creating child nodes and branches, until a termination condition is met. Gini Impurity and Information are the popular criteria for splitting data.
8) Random Forest: Each decision tree is trained independently using the random subsets of data and features, and the combined forecasts of all the decision trees are used to arrive at the final prediction.
9) Bagging Classifier: Based on a distinct bootstrapped subset of the training data, each base classifier is independently trained. The sum of the forecasts from each model is the final prediction.
10) AdaBoost Classifier: A weighted voting system is used to decide the final prediction after each weak classifier is trained on a subset of the training data and given a weight depending on classification accuracy.
11) Gradient Boosting Classifier: Follows a boosting framework where weak learners are trained sequentially to correct the errors of previous models by utilizing gradient descent optimization.
12) XGBoost Classifier: Follows the gradient boosting framework and uses decision trees as weak learners, which are constructed in a greedy manner.
13) CatBoost Classifier: Follows the gradient boosting framework and handles categorical variables by encoding them based on the target variable's values within each category.
14) LGBM Classifier: Follows the gradient boosting framework and provides efficient handling of categorical features by using 'Optimized Exclusive Feature Bundling' (OEFB).
15) Voting Classifier: In hard voting, the majority class is chosen as the final guess. In soft voting, the class with the highest average performance or score is chosen as the final estimate.

4.2.7 Stage 7 - Simultaneous implementation of K-fold Cross Validation and GridSearchCV: While training every model, we implemented K-Fold Cross Validation and GridSearchCV simultaneously. K-fold Cross Validation is primarily used to evaluate a model's performance and generalizability, whereas GridSearchCV is used for hyperparameter tuning. A parameter grid is passed to the GridSearchCV along with a cross-validation engine like the K-fold CV. It tests all possible combinations of hyperparameters and identifies the optimal set of values that yield the best performance.

4.2.8 Stage 8 - Simultaneous performance evaluation: Right after hyperparameter tuning and training, predictions of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 10 August 2023 **Revised:** 27 September 2023 **Accepted:** 18 October 2023*

model were generated over the testing set. We thoroughly evaluated every model on the basis of various performance indicators like accuracy, precision, and recall as well as f1, AUC(ROC), and AUC(PR) scores. Also, plotted necessary graphs like the ROC and PR curves along with the Confusion Matrix.

V. RESULTS AND DISCUSSION:

5.1 Analysis of the ROC curve:

It visualizes the model's performance in distinguishing between positive and negative instances across various threshold settings.

Here's a comparative analysis of the ROC Curves of different models implemented in Figure 5.1:

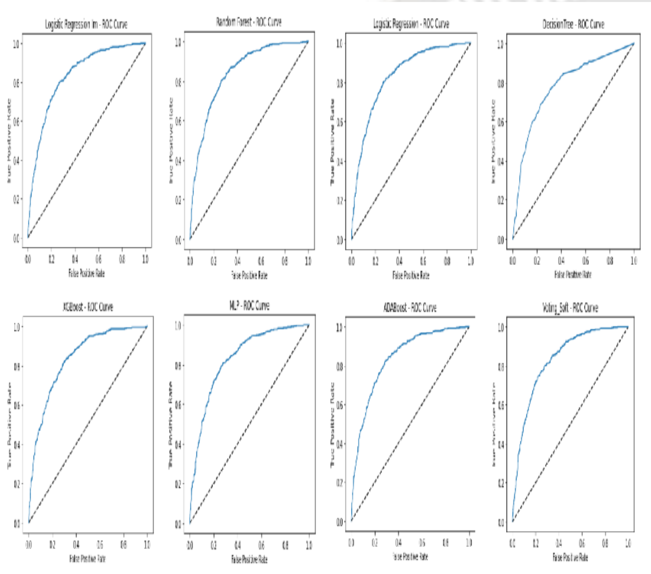

Fig. 5.1: ROC Curves of the models

5.2 Analysis of the Precision-Recall Curve:

It provides insights into how well the model performs in correctly classifying positive instances (precision) and capturing all positive instances (recall). Here's a comparative analysis of the Precision vs Recall Curves of various models implemented in Figure 5.2:

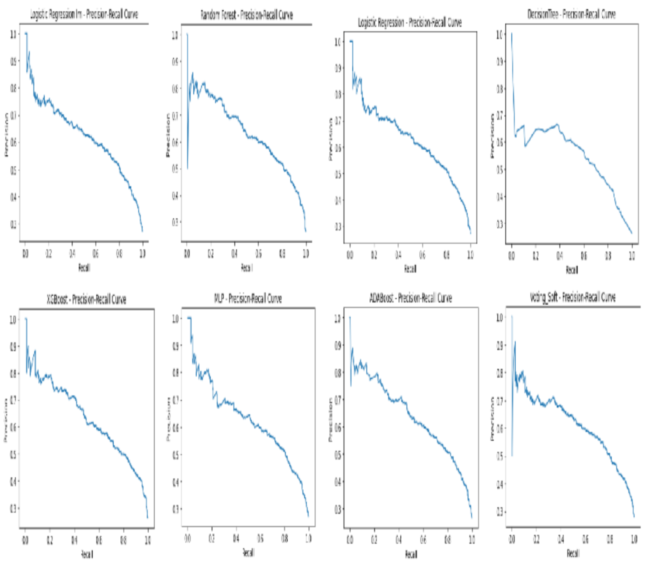

Fig. 5.2: Precision vs Recall Curves of the models.

5.3 Comparison of Every Model:

Figure 5.3 shows a detailed comparison of all the models implemented in this research work.

	Classifier	Accuracy	Precision	Recall	F1 Score	ROC AUC_Score
10	XGBClassifier	0.79	0.61	0.55	0.78	0.71
1	Imbalanced Logistic Regression	0.79	0.62	0.50	0.78	0.70
15	Voting Classifier Soft	0.78	0.61	0.63	0.74	0.82
11	MLPClassifier	0.78	0.58	0.65	0.78	0.74
9	AdaBoostClassifier	0.78	0.69	0.63	0.79	0.73
8	GradientBoostingClassifier	0.78	0.59	0.59	0.78	0.72
7	CatBoostClassifier	0.78	0.59	0.53	0.77	0.70
12	LGBMClassifier	0.78	0.58	0.54	0.77	0.70
5	Decision Tree	0.77	0.58	0.55	0.77	0.70
0	Logistic Regression	0.76	0.54	0.71	0.77	0.75
14	Bagging Classifier	0.76	0.56	0.47	0.76	0.67
2	SVM	0.75	0.54	0.55	0.76	0.69
13	Ridge Classifier	0.74	0.50	0.79	0.75	0.76
6	Random Forest	0.72	0.48	0.84	0.74	0.76
3	KNN	0.72	0.47	0.51	0.72	0.65
4	Naive Bayes	0.66	0.42	0.87	0.68	0.72

Fig. 5.3: Values of evaluation metrics of all the models

The graph in Figure 5.4 visualizes how "accuracy", "f1 score", and "ROC AUC_Score" with respect to every classifier.

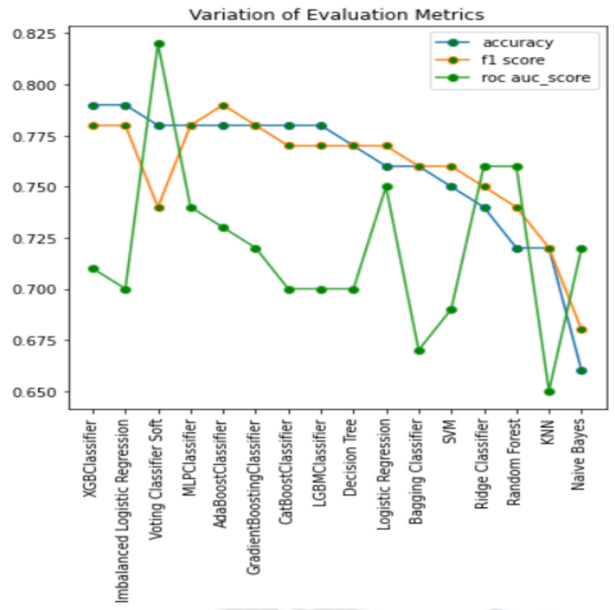

Fig. 5.4: Comparison of Accuracy, F1, and ROC(AUC) Scores of all the models

From Figure 5.4, we inferred that the following classifiers are the best-performing ones:

XGBoost- It boasts the highest accuracy (0.79), a respectable f1-score (0.78), which is nearly as high as the previous record-holder (0.79 of AdaBoost), and a respectable ROC AUC_Score (0.71).

Multi-Layer Perceptron(MLP) – With a superior ROC AUC_Score of 0.74, its accuracy and f1 score of 0.78 are both quite respectable and close to the greatest values thus far.

AdaBoost - It also has a respectable ROC AUC_Score (0.73), the highest f1 score (0.79), and the second-best accuracy (0.78).

Logistic Regression - The ROC AUC_Score achieves a local maximum of 0.75 for logistic regression with an f1-score of 0.77, close to the highest value, and a decent accuracy of 0.76.

Based on the above 4 best-performing classifiers according to our assessment, we implemented the voting classifier (soft as well as hard). The Soft Voting Classifier showed the best performance out of all. It showed an accuracy of 0.78 which is at par with the highest accuracy achieved in our research so far (i.e. 0.79 of XGBoost) and a decent f1-score of 0.74. The difference-maker here is the ROC AUC_Score of 0.82 which is significantly more than the highest ROC AUC_Score we've achieved so far (i.e. 0.76 of Random Forest and Ridge Classifier). Therefore, we infer that the Soft Voting Classifier is the best-performing classifier in our research.

VI. CONCLUSION

In this research, we did a comparative analysis of the effectiveness of various machine learning models for user loss prediction in the telecom industry. We used qualitative selection techniques i.e., chi-square test and correlation analysis to select the best features. We also used SMOTE on the data and created a new synthetic model with a small number of classes to sample them. We employed numerous learning and machine learning techniques, including Logistic Regression, Support Vector Machine, K Neighbors, Naive Bayes, Pruning Tree, Random Forest, Gradient Boosted Classifier, AdaBoost Classifier, XGBoost Classifier, Light Gradient Accelerator Machine Classifier, Ridge Classifier, Bagging Classifier. We evaluate each model using performance indicat ors such as accuracy, precision, and recall along with f1 and AUC(ROC) scores.

After comparing the performance of each model used, we conc luded that XGBoost, AdaBoost, Logistic Regression, and Multilayer Perceptron (MLP) performed best with regard to accuracy, f1, and ROC(AUC) scores. Based on the top four performances, we continue to use the survey. When compared to other classifiers, the **voting classifier** had a much better AUC(ROC) score of 0.82 in addition to an accuracy of 0.78 and an f1-score of 0.74.

Consequently, a loss forecasting model can be used for market ing communications to identify customers who will switch to o ther service providers and provide their own incentives. Projec t requests can help businesses stay ahead of the competition an d achieve their goals by leading to more information for decisi on-making.

REFERENCES

[1] Radosavljevik D, van der Putten P, Larsen KK (2010) The impact of experimental setup in prepaidchurn prediction for mobile telecommunications: What to predict, for whom and does the customerexperience matter? Trans. MLDM 3(2):80–99

[2] Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A., Rehman, A.: Telecommunication subscribers'churn prediction model using machine learning. In: Eighth International Conference on Digital Infor-mation Management (ICDIM 2013), pp. 131–136. IEEE (2013)

[3] Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data 6(1):28

[4] Tamaddoni Jahromi, A.: Predicting customer churn in telecommunications service providers (2009)

[5] Lalwani, Praveen & Mishra, Manas & Chadha, Jasroop & Sethi, Pratyush. (2022). Customer churn prediction system: a machine learning approach. Computing. 104. 1-24. 10.1007/s00607-021-00908-y.

[6] Brându¸soiu, I., Toderean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications

industry. In: 2016 International conference on communications (COMM), pp. 97–100.IEEE (2016)

[7] Coussement K, Van den Poel D (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications 34(1):313–327

[8] Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Confer-ence on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions),pp. 1–6 (2015)

[9] Gürsoy U ¸S (2010) Customer churn analysis in telecommunication sector. Istanbul Üniversitesi ˙I¸sletme Fakültesi Dergisi 39(1):35–49

[10] Hadden J, Tiwari A, Roy R, Ruta D (2006) Churn prediction: Does technology matter. International Journal of Intelligent Technology 1(2):104–110

[11] Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research 34(10):2902–2917

[12] Huang, Y., Zhu, F.,Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., Zeng, J.: Telco churn prediction with big data. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607–618 (2015)

[13] Idris, A., Khan, A., Lee, Y.S.: Genetic programming and adaboosting based churn prediction for telecom. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328–1332. IEEE (2012)

[14] Kisioglu P, Topcu YI (2011) Applying bayesian belief network approach to customer churn analysis: A case study on the telecom industry of turkey. Expert Systems with Applications 38(6):7151–7157

[15] Saleh, Sarkaft & Saha, Subrata. (2023). Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university. SN Applied Sciences. 5. 10.1007/s42452-023-05389-6.

[16] Rathi, S., Kant Hiran, K., & Sakhare, S. (2023). Affective state prediction of E-learner using SS-ROA based deep LSTM. Array, 19, 100315.

[17] Rathi, S.R. and Deshpande, Y.D. (2022), "Course complexity in engineering education using E-learner's affective-state prediction", Kybernetes, Vol. ahead-of-print No. ahead-of-print.

[18] S. R. Rathi and Y. D. Deshpande, "Embedding Affect Awareness into Online Learning Environment using Deep Neural Network," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2019, pp. 1-6, doi: 10.1109/ICCUBEA47591.2019.9128811