ML FINAL REVIEW PRESENTATION

Customer Churn Prediction in Telecom Industry Using ML Algorithm

05 December 2023

MUHAMMED SHIBIL C V (CB.SC.P2DSC23007)
SUDHARSHAN R MOHAN (CB.SC.P2DSC23022)

INTRODUCTION

- Customer churn, or the ***loss of customers***, is a critical concern for businesses as it directly impacts revenue. Early detection of client churn enables businesses to retain customers and minimize revenue loss proactively.
- This presentation will guide the optimum machine-learning strategy for early ***client churn prediction***. It explores different machine-learning algorithms and evaluates their accuracy in predicting customer churn
- By implementing effective churn prediction models, businesses can enhance customer retention initiatives, reduce costs, and improve overall customer satisfaction.

PROBLEM STATEMENT

- Implementing a ***machine learning predictive model*** to prevent or identify customer churning in the ***telecom industry***
- What will be the ***inputs*** to the model?
- What are the ***outputs*** of the model?
- What do we do with the predictions?

OBJECTIVES

- To guide the optimum machine-learning strategy for early client churn prediction
- ***Analysis of customer data insights*** to understand the problem
- Implementation of different machine learning algorithms
- ***Comparison*** between each machine learning algorithm to find the most efficient

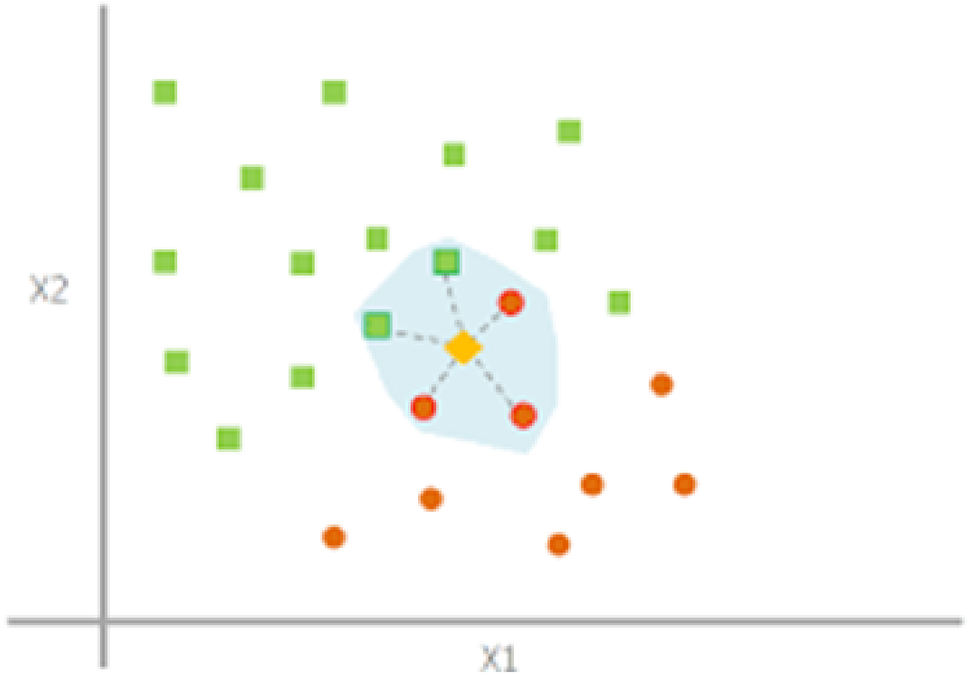
METHODOLOGY

Implementation of ML algorithms

1. K-Nearest Neighbors
2. Decision tree
3. Random Forest model

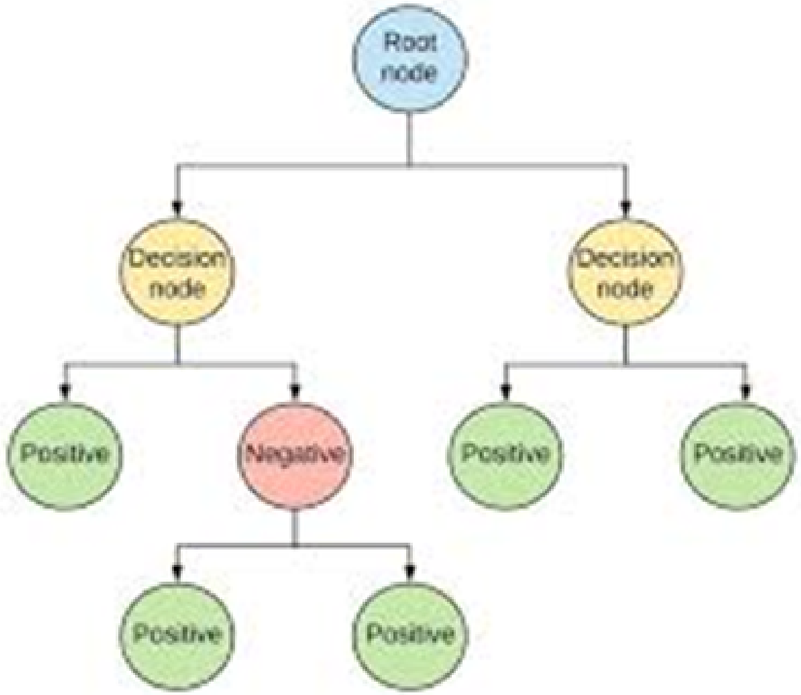
1. K-Nearest Neighbors

In the KNN model, ***we initialize the K value find the distance to each data point, and sort the arranged data in ascending order,*** after checking for the labels of K instances we finalize the target. Usually, it is very easy to implement because of few hyperparameters and it gives a satisfactory accuracy



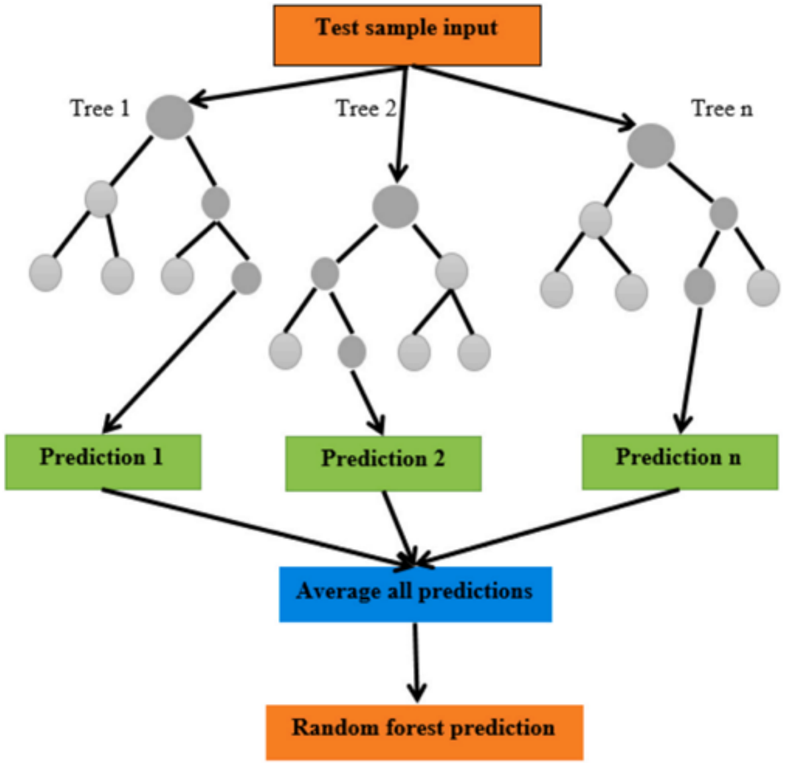
2. Decision tree

Decision trees are another supervised learning algorithm and ***it is a graphical representation of getting all possible solutions to a problem based on the given condition.*** Compared to KNN it gives much more accuracy to the given problem and the logic behind the decision tree can be easily understood because it can refer to the decision-making capability of humans



3. Random Forest model

The Random Forest approach is an ***Ensemble learning approach*** it includes several classifiers to build the solution ***by getting predictions from every model(combining a lot of decision tree models) and making a final decision on majority voting or averaging.***it is usually efficient for large datasets. Since it is a much time-consuming process however it can provide better accuracy compared to the other implemented models

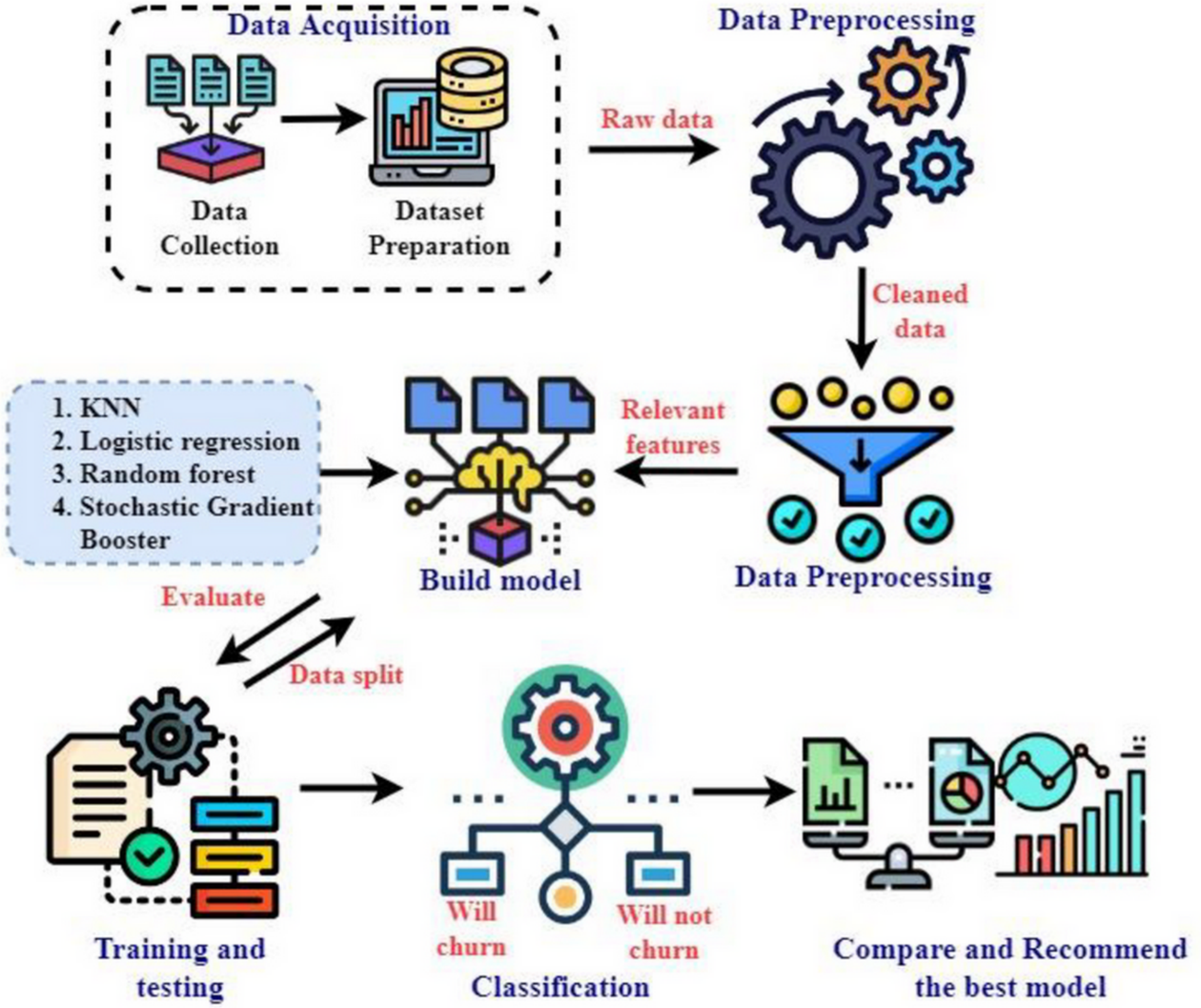


DATASET DESCRIPTION

7043 customers telco data with 21 features

customerID	tenure	OnlineBackup	Contract	
gender	PhoneService	DeviceProtection	PaperlessBilling	
SeniorCitizen	MultipleLines	TechSupport	PaymentMethod	Churn
Partner	InternetService	StreamingTV	MonthlyCharges	
Dependents	OnlineSecurity	StreamingMovies	TotalCharges	

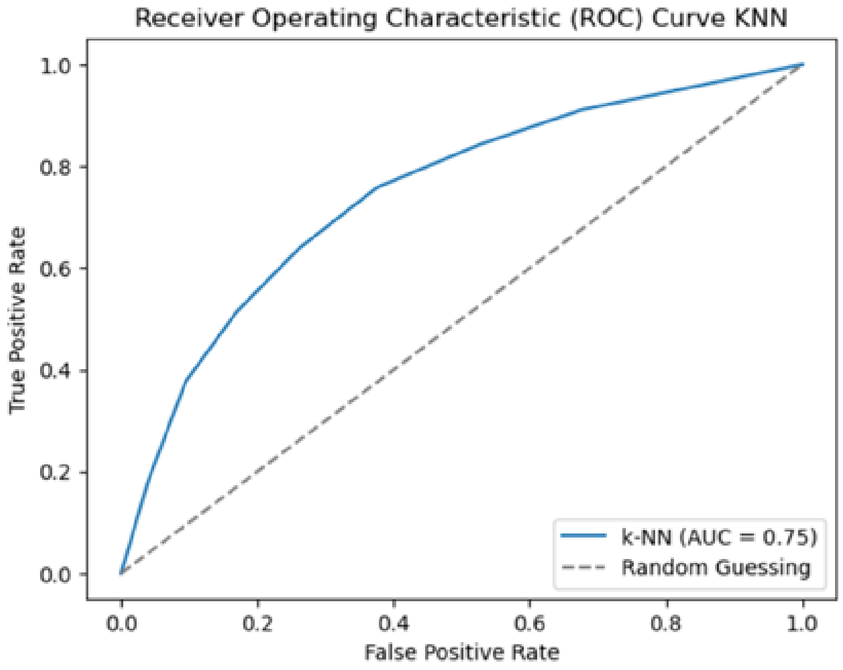
EXPERIMENTAL DESIGN



PERFORMANCE MEASURES

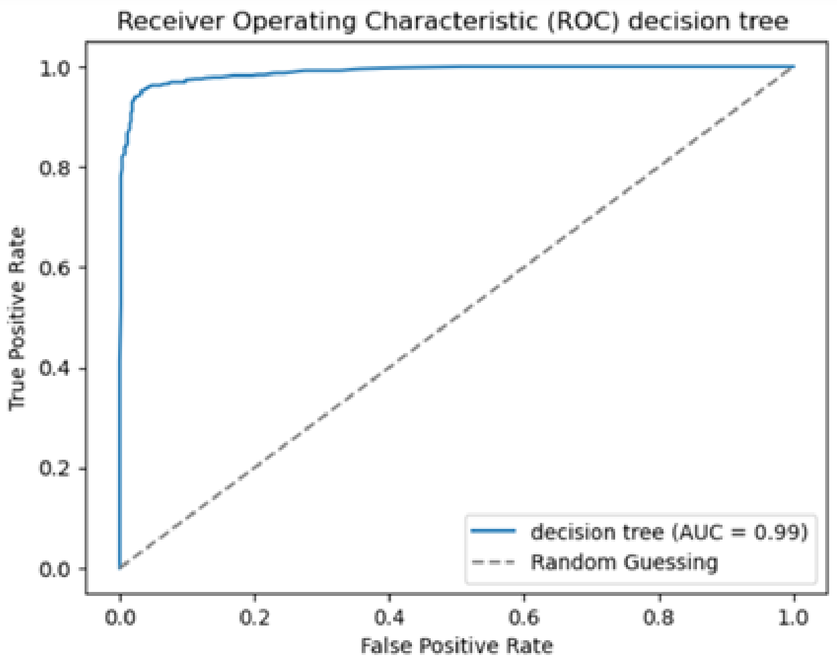
1. K-Nearest Neighbors:

748	301
146	871



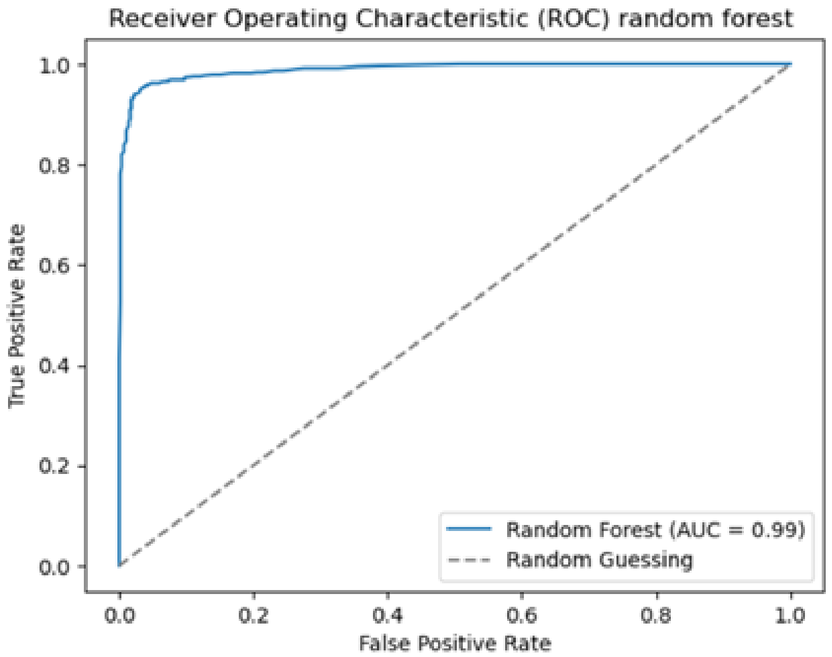
2. Decision tree:

489	36
26	586



3. Random Forest model

489	**36**
40	**601**



RESULTS

1. K-Nearest Neighbors:

The KNN Algorithm gives a model with an accuracy of ***78.6 after applying SMOTE and the AUROC obtained is 0.75***

2. Decision tree:

The prediction results of Random Forest algorithms give a ***reasonable accuracy of 95.76*** The corresponding area under ***the ROC curve is obtained as 0.99***

3. Random Forest model

The prediction results of the decision tree also give an almost ***near accuracy of 93.6***. is almost equal to the accuracy of the ***random forest model and the AUROC is .99***

CONCLUSION

During the study, we selected 3 algorithms to do a comparative study which are K-Nearest Neighbors, Decision tree, and Random Forest classifier. The application of SMOTE played a crucial role in mitigating the impact of class imbalance within the dataset.

So as a result out of 3 different machine learning models the random forest Classifiers have the largest performance measures and AUROC of 95.76 and 0.99 respectively

POSSIBLE FUTURE DIRECTIONS

- ***Enhancing Model Effectiveness:*** Using more Machine learning models and analyzing each one's performance will improve the effectiveness of the machine learning models with the customer churn prediction dataset
- ***Extracting Deeper Insights:*** Extract more useful insights from the customer churn dataset
- ***Leveraging DSA Algorithms for Feature Selection:*** Use DSA algorithms like priority queue in the feature selection and improve the accuracy of the model

REFERENCES

- R. Shalini, B.R. Kavitha Customer churning analysis using machine learning algorithms B. Prabadevi International Journal of Intelligent Networks 4 (2023) 145–15
- Youngjung Suh Suh learning-based customer churn prediction in-home appliance rental business, Journal of Big Data (2023) 10:41
- Kimura T. Customer churn prediction with hybrid resampling and ensemble learning. J Manage Inform Decis Sci. 2022;25(1):1–2