

# **TELECOM CUSTOMER CHURN PREDICTION USING VARIOUS MACHINE LEARNING ALGORITHMS**



**STUDENT ID: D3127027**

**FACULTY: SCHOOL OF COMPUTING,  
ENGINEERING & DIGITAL TECHNOLOGIES**

# Abstract

This study uses various metrics to assess different classification models and predict telecom customer churn. The effectiveness of Logistic Regression, Random Forest, KNN, and Naive Bayes models after hyper-parameter tuning was implored in this analysis. The result shows that Random Forest and Logistic Regression are the best-performing models, with excellent precision and accuracy. While Random Forest displays balanced performance across various criteria, Logistic Regression excels in decreasing false positive predictions. Naive Bayes performs competitively despite its simplified assumptions. Although KNN outperformed Logistic Regression in recall, its performance across various metrics was the lowest as compared to others. It is recommended that Random Forest or Logistic Regression should be used depending on business needs.

## Introduction

In present times, one of the most significant industries is the telecommunications sector, as it plays a significant role in connecting individuals, businesses, and devices worldwide, enabling communication, facilitating e-commerce, and collaboration on a large scale. The increase in the number of operators under each telecommunication company has increased the level of competitions, due to this, the issue of customer churn continues to be a major concern for service providers [1]. To estimate customer attrition, telecom churn prediction becomes a crucial responsibility for the industry. Churn prediction helps telecom companies determine which customers are likely to cancel their subscriptions by using machine learning approaches.

The anticipated average monthly churn rate for mobile telecommunications is 2.2%, based on a previous study [3]. This suggests that one in every fifty consumers of a particular company cancels their subscription monthly. Telecom churn prediction uses a variety of machine learning techniques to build predictive models that accurately forecast client attrition. Neural networks [4][2], Logistic Regression [2][3][11], Decision Trees [12], Random Forest [12], Gradient Boosting Machines (GBM) [12], and Support Vector Machines (SVM) [4] are common machine-learning (ML) methods used for churn prediction. These algorithms use variables like billing information, and customer interactions in identifying churn predictors.

# Literature

Highlighting existing research, novel approaches, and new trends, this review of the literature offers an overview of the literature on telecom churn prediction.

A study by [2] included several new features designed to predict customer attrition. To predict customer attrition, seven prediction models, including logistic regression, linear classifications, and Naive Bayes were used. Comparative experiments showed that the modelling approaches and the addition of these additional features produced better performance than the earlier approaches.

[1] in their study used data from SyriaTel to assess many algorithms, like Decision Tree, Gradient Boosted Machine Tree (GBM), Random Forest and Extreme Gradient Boosting (XGBoost). When consumer social network analysis (SNA) data were included, the accuracy of the model increased dramatically from 84% to an astounding 93.3%, as determined by the Area Under Curve (AUC) criterion.

In a study by [3], boosting algorithms and clustering techniques were used in a novel way to predict client attrition. The purpose of the study was to group customers according to weights supplied by the boosting algorithm to identify high-risk consumer clusters. To create churn prediction models for each cluster, logistic regression was used as the primary learner. The efficacy of these cluster-specific models was compared to a single logistic regression model in the study.

## Methodology

The methodology includes phases like data collection, pre-processing, exploration, feature engineering, modelling, and performance review.

### Dataset

This study uses data from the Telecom Churn Dataset, which was sourced from [Kaggle](#). This dataset offers a wide range of features that are related to customers in the telecom industry. It comprises of 38 columns of distinctive features and 7043 instances. The features in the dataset can be seen in the appendix below.

	Customer ID	Gender	Age	Married	Number of Dependents	City	Zip Code	Latitude	Longitude	Number of Referrals	...	Payment Method	Monthly Charge	Total Charges	Total Refunds	Total Extra Data Charges	Total Long Distance Charges	I
0	0002-ORFBO	Female	37	Yes	0	Frazier Park	93225	34.827662	-118.999073	2	...	Credit Card	65.60	593.30	0.00	0	381.51	
1	0003-MKNFE	Male	46	No	0	Glendale	91206	34.162515	-118.203869	0	...	Credit Card	-4.00	542.40	38.33	10	96.21	
2	0004-TLHLJ	Male	50	No	0	Costa Mesa	92627	33.645672	-117.922613	0	...	Bank Withdrawal	73.90	280.85	0.00	0	134.60	
3	0011-IGKFF	Male	78	Yes	0	Martinez	94553	38.014457	-122.115432	1	...	Bank Withdrawal	98.00	1237.85	0.00	0	361.66	
4	0013-EXCHZ	Female	75	Yes	0	Camarillo	93010	34.227846	-119.079903	3	...	Credit Card	83.90	267.40	0.00	0	22.14	
5	0013-MHZWF	Female	23	No	3	Midpines	95345	37.581496	-119.972762	0	...	Credit Card	69.40	571.45	0.00	0	150.93	
6	0013-SMEOE	Female	67	Yes	0	Lompoc	93437	34.757477	-120.550507	1	...	Bank Withdrawal	109.70	7904.25	0.00	0	707.16	
7	0014-BMAQU	Male	52	Yes	0	Napa	94558	38.489789	-122.270110	8	...	Credit Card	84.65	5377.80	0.00	20	816.48	
8	0015-UOCOJ	Female	68	No	0	Simi Valley	93063	34.296813	-118.685703	0	...	Bank Withdrawal	48.20	340.35	0.00	0	73.71	
9	0016-QLJIS	Female	43	Yes	1	Sheridan	95681	38.984756	-121.345074	3	...	Credit Card	90.45	5957.90	0.00	0	1849.90	

Fig1: Dataset Overview.

## Pre-Processing

Data preprocessing, which includes operations like data reduction and transformation to guarantee the efficiency of learning algorithms, is a crucial stage in information discovery activities [6]. By choosing the right algorithms, proper preprocessing helps to accurately examine gathered information by preventing the conversion of raw data into low-quality data [11].

## Data Visualisation

Data Visualisation is done to get a quick overview of features in the dataset. To visualise the distribution of features in the dataset, a count plot & box plot was used.

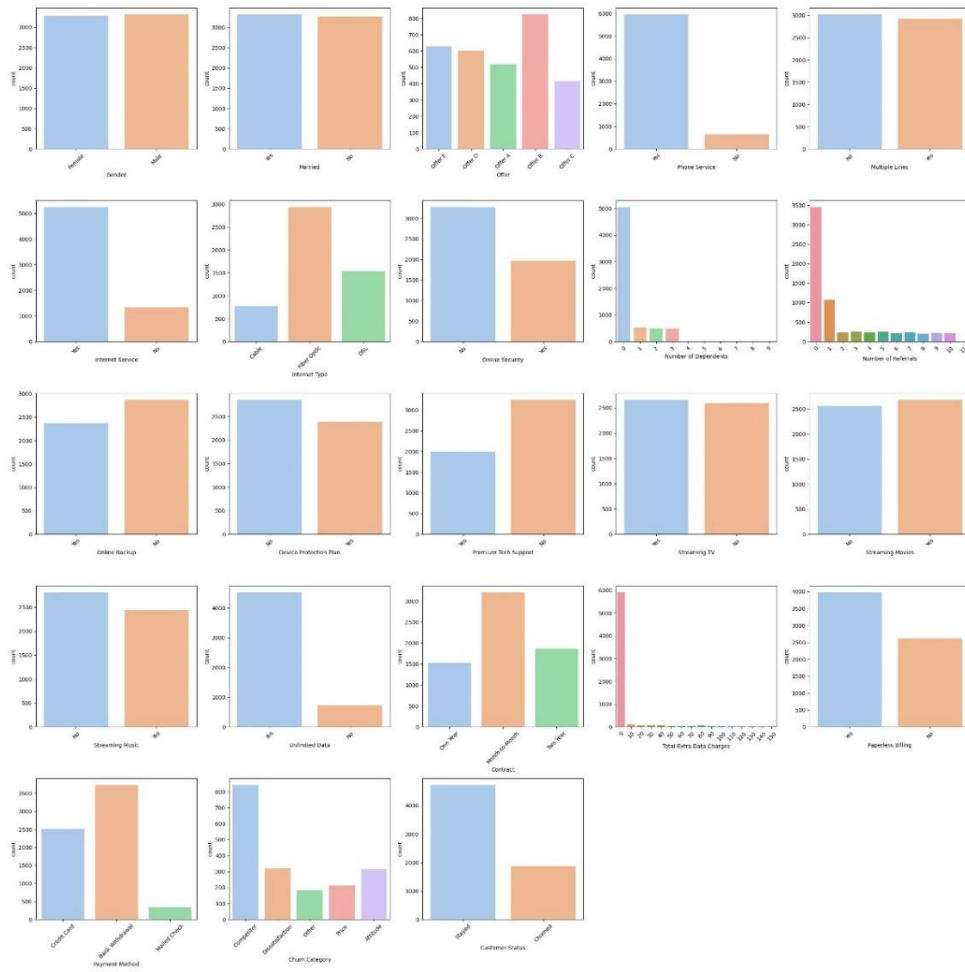


Fig 2: Categorical Features Distribution.

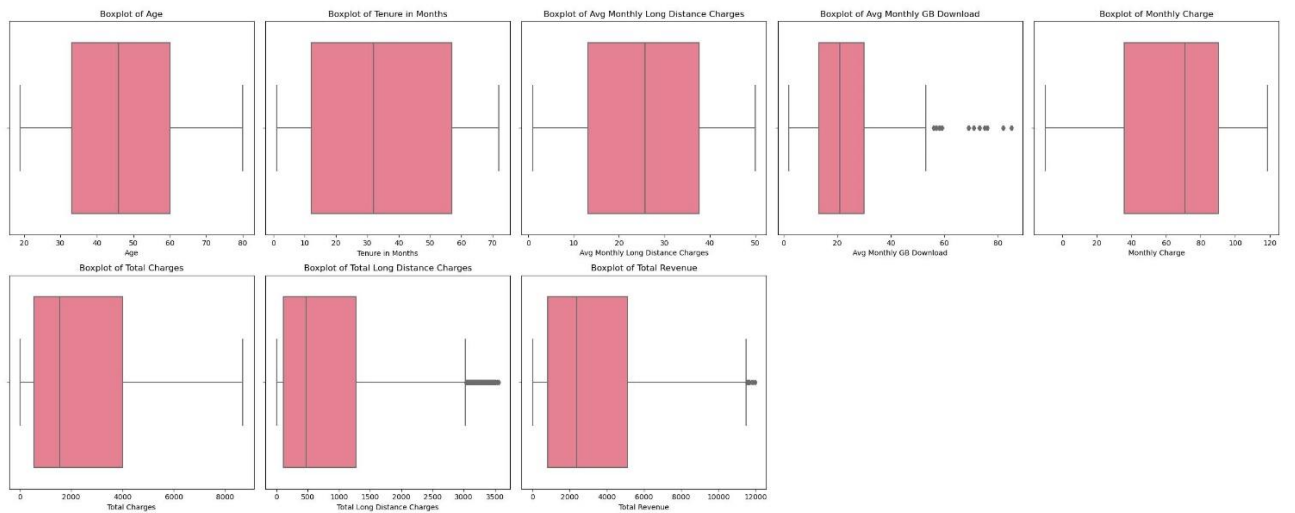


Fig 3: Numerical Features Distribution.

## Removal of Columns

The removal of columns such as ZIP Code, City, Latitude, and Longitude would occur throughout this process. Also eliminated from the dataset were the Churn Reason and Churn Category columns, as well as the Customer ID column, which was deemed unnecessary for the purpose of this research.

## Missing Values

According to [7], [13], it is advised to eliminate missing data if they exceed the 30% cutoff, otherwise use the residual mean or median for numerical values and the most frequent for categorical values.

Upon checking, “Offer” column had missing values of 54.6%, so it was removed from the dataset. The other columns had missing values ranging from 9%-20%, therefore, the categorical columns were replaced with the mode while the numerical columns were replaced with the median rather than mean. The median was used because its less vulnerable to outliers than the mean [15].

## Data Transformation

### *Label-Encoding*

A method for mapping categories as continuous numbers is label-encoding (LE). When dealing with categorical variables that have only two possible values, like "yes" or "no", the LE approach is more helpful [16]. The dataset includes 15 categorical features with 2 unique values, so these features were encoded using label encoder.

### *One-Hot-Encoding*

One-Hot Encoding is a widely used for encoding when a feature has more than 2 distinct values [16]. There are 3 features in the dataset with more than 2 unique values they were encoded using one-hot encoder.

## Normalisation

The dataset was normalised using “MinMaxScaler” which rescales variables into the range 0 & 1 and only the high variance features were subjected to these scaling [17].

## Feature Selection

Feature selection is an approach that is commonly employed in the pre-stage of machine learning [16]. This process involves choosing a subset of features from a set of features.



# Multi-Collinearity Assessment

Correlation analysis is used in machine learning to evaluates multi-collinearity between independent variables in a dataset. Examining correlations can help identify characteristics that are redundant or highly linked, which can help with feature selection procedures and reduce the chance of overfitting [18]. A threshold of 0.7 was set for high correlation, features higher than the set threshold was removed.

## Heatmap

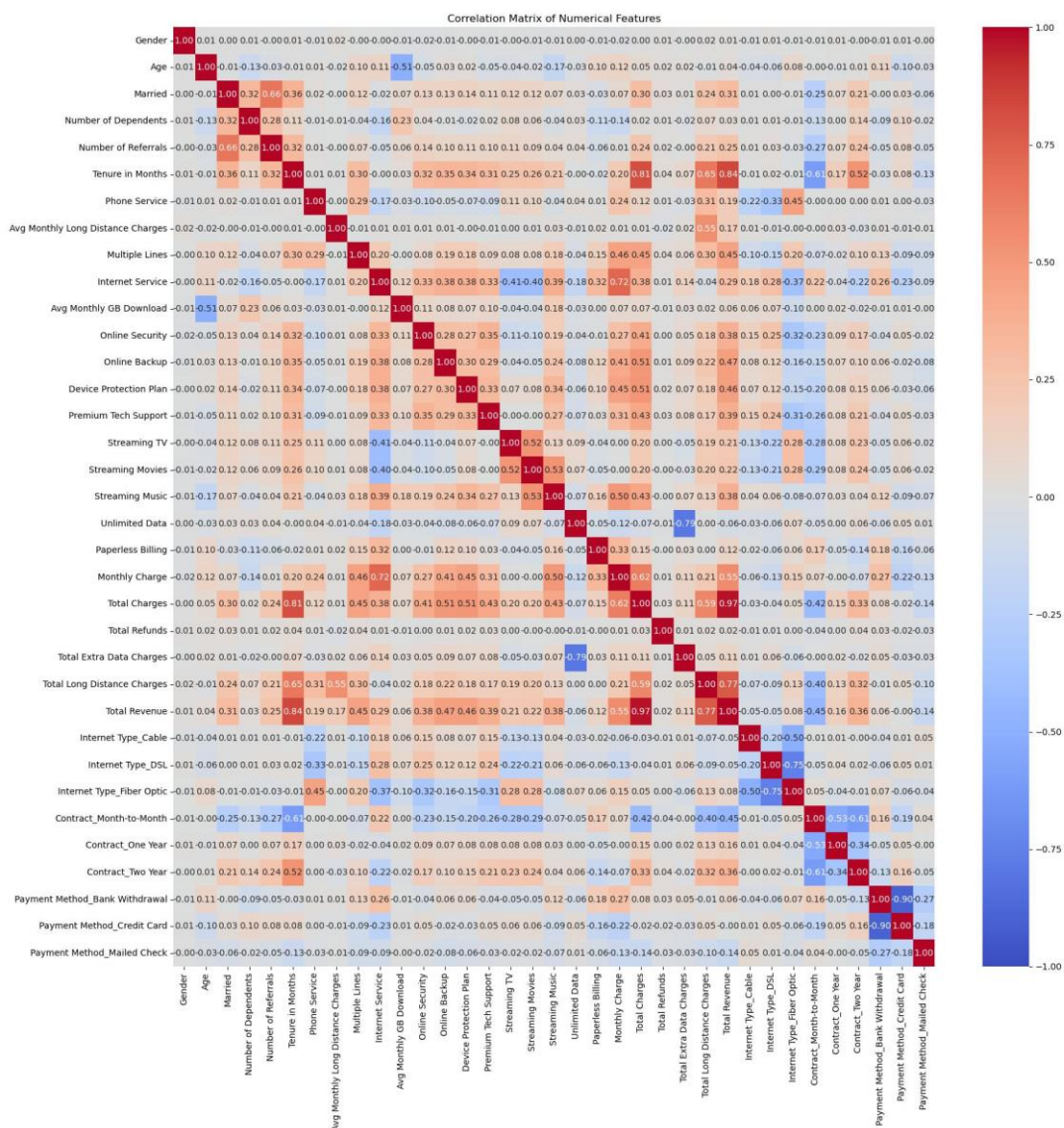


Fig 4: Correlation Plot.

## **Variance Inflation Factor (VIF)**

The VIF is an essential metric in feature selection for numerical variables because it can identify multicollinearity [19]. Features with VIF value greater than 10 usually indicates multicollinearity-induced inflation [19], so they are frequently eliminated from the dataset to improve model stability. The VIF conducted on the dataset showed that all the features have variance less than 10, so they were maintained.

## **Chi-Square Test**

Chi-square test is a process of finding statistically significant correlations between predictors and the target feature. The features that yielded statistical significance in this test are retained since they offer valuable information about the model's prediction ability otherwise, they are removed [20]. After the test was performed it showed some features were statistically insignificant to the target variable, so they were dropped.

## **Backward Stepwise Regression**

Regression modelling relies on a feature selection method called backward stepwise regression, especially when many predictors are involved. To eliminate the least significant predictors, this strategy entails removing variables with p-values higher than a threshold, usually set at 0.05 [21], this simplifies the model and keeps it from overfitting. Following these concepts, guarantees that the final model includes only the most relevant predictors, improving its interpretability and performance.

## **Data Splitting**

The dataset was split into training and testing set using the 70/30 splitting, making the training set 70% and the test set 30%, ensuring that the model is given enough samples to train and improve the performance.

## **Target Variable Up-Sampling**

Plotting counts to visualise the target variable's distribution provides important insights into the balance of classes in a dataset as predictions may be skewed due to class imbalance.



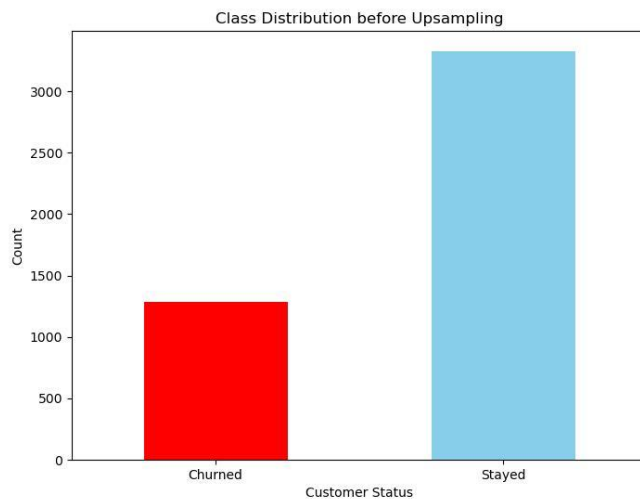


Fig 5: Target Variable Distribution before SMOTE.

Creating artificial samples for the minority and balancing the dataset using SMOTE is essential in resolving class imbalance, this improves classification performance and makes more reliable predictions.

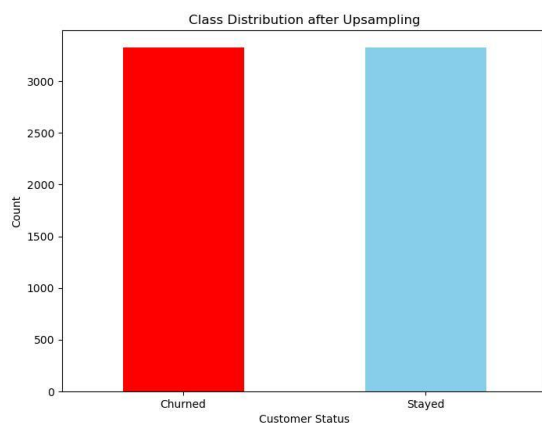


Fig 6: Target Variable Distribution after SMOTE.

## Machine Learning Algorithms

Machine-Learning (ML) use data-driven methodologies to create prediction models for telecom churn, making it possible to analyse past customer data and find patterns, trends linked to churn behaviour. ML lays the groundwork for developing predictive models that can efficiently foresee and reduce client attrition, improving long-term profitability and customer pleasure.

## Modelling Approach

### Model Justification

The machine learning techniques for used for churn prediction includes:

1. **Logistic Regression** is a statistical model that helps identify the factors driving churn by predicting the probability of a binary outcome using predictor variables making it easier to identify the elements that influence customer-churn.
2. **Random Forest** is an ensemble learning technique that boosts performance in churn prediction tasks by using several decision trees. It achieves strong generalisation to test data and prevents overfitting by combining the predictions of individual trees.
3. **Guassian Naives Bayes (GNB)** is a simple probabilistic classifier that outperforms other algorithms in prediction rates by assuming strong feature independence. By using this assumption, it exhibits strong performance in churn prediction tasks [\[22\]](#).
4. **K-Nearest Neighbour (KNN)** is non-parametric and makes predictions by comparing a new data point to its k closest neighbours in the training set. Because it can handle a variety of data distributions without making assumptions, it is preferred for churn prediction and excels at capturing complicated patterns in the feature space.

The motive behind the choice of these algorithms for churn prediction stems from their unique advantages and appropriateness for the given task.

### Model Optimisation

Optimisation increases the model's accuracy while reducing the possibility of errors [\[23\]](#). The optimisation technique used in this analysis is Hyperparameter Tuning which plays a critical role in improving the performance capacity of a model [\[24\]](#).

## Model Performance Analysis

The model's performance is evaluated using a range of measures. The Classification Report offers a thorough analysis of each class's accuracy, precision, recall, F1-score. Accuracy is a key indicator of how well the model is working overall. Precision is essential for evaluating the reliability of positive predictions [\[12\]](#), While recall shows how well the model recognises positive occurrence among the real positives. The F1-score takes into consideration unequal class distributions [\[25\]](#), while ROC-AUC provides an evaluation of the model's performance across a range of classification thresholds. These strategies

guarantee a strong evaluation, allowing for well-informed choices about the model's implementation.

## Classification Report

The classification report of each model after optimisation is displayed below. From the figures below, we can see the outcomes of the analysis that was done. In comparison to the other methods, Random Forest (Fig 8) and Logistic Regression (Fig 7) produced better performance and accuracy. We can see also the outcomes of applying the GNB (Fig 10) and KNN (Fig 9). It is also observed from the classification report that the KNN model produced the lowest accuracy.

```
Logistic Regression Classification Report:
              precision    recall  f1-score   support

     0           0.80       0.87       0.83       1393
     1           0.85       0.78       0.81       1393

 accuracy                   0.82       2786
 macro avg           0.83       0.82       0.82       2786
 weighted avg           0.83       0.82       0.82       2786
```

Fig 7: Classification Report of Logistic Regression Model.

```
Random Forest Classification Report:
              precision    recall  f1-score   support

     0           0.84       0.82       0.83       1393
     1           0.82       0.84       0.83       1393

 accuracy                   0.83       2786
 macro avg           0.83       0.83       0.83       2786
 weighted avg           0.83       0.83       0.83       2786
```

Fig 8: Classification Report of Random Forest Model.

```
KNN Classification Report:
              precision    recall  f1-score   support

     0           0.77       0.74       0.76       1393
     1           0.75       0.78       0.77       1393

 accuracy                   0.76       2786
 macro avg           0.76       0.76       0.76       2786
 weighted avg           0.76       0.76       0.76       2786
```

Fig 9: Classification Report of KNN Model.

```

GNB Classification Report:
              precision    recall  f1-score   support

     0       0.79       0.81       0.80     1393
     1       0.81       0.78       0.80     1393

 accuracy              0.80     2786
 macro avg       0.80       0.80       0.80     2786
 weighted avg    0.80       0.80       0.80     2786

```

Fig 10: Classification Report of GNB Model.

## Findings

Given their capacity to identify variety of patterns in the data, Random Forest and Logistic Regression models demonstrated the highest accuracy (82.20% and 82.30%, respectively) in predicting churn. Random Forest handled complex, nonlinear patterns well, while Logistic Regression was superior at modelling linear relationships. The preprocessing techniques like feature selection etc carried out greatly improved these models' predictive capabilities. On the other hand, KNN showed the lowest accuracy (76.13%), probably due to its vulnerability to high-dimensional feature spaces and difficulties with parameter selection. KNN's performance may have been limited by its reliance on distance measures for classification, despite preprocessing efforts. Overall, the robust performance of Random Forest and Logistic Regression highlights how well they can recognise complex churn patterns.

Model	Accuracy %	Precision %	Recall %	F1-Score
Logistic Regression	82.30%	85.43%	77.89%	0.81
Random Forest	82.20%	79.96%	85.93%	0.83
KNN	76.13%	75.07%	78.25%	0.77
Naive Bayes	79.83%	80.71%	78.39%	0.80

Fig 11: Performance Metrics Table.

According to the results, Random Forest and Logistic Regression had the greatest accuracy rates (82%), followed by Naive Bayes (79.83%) and KNN (76.13%). The results from the performance metrics table are supported by this visualisation, which shows how Random Forest and Logistic Regression outperform Naive Bayes and KNN. The plot aids the interpretation and comprehension of each model's predictive capabilities by providing a summary of each model's comparative performance.

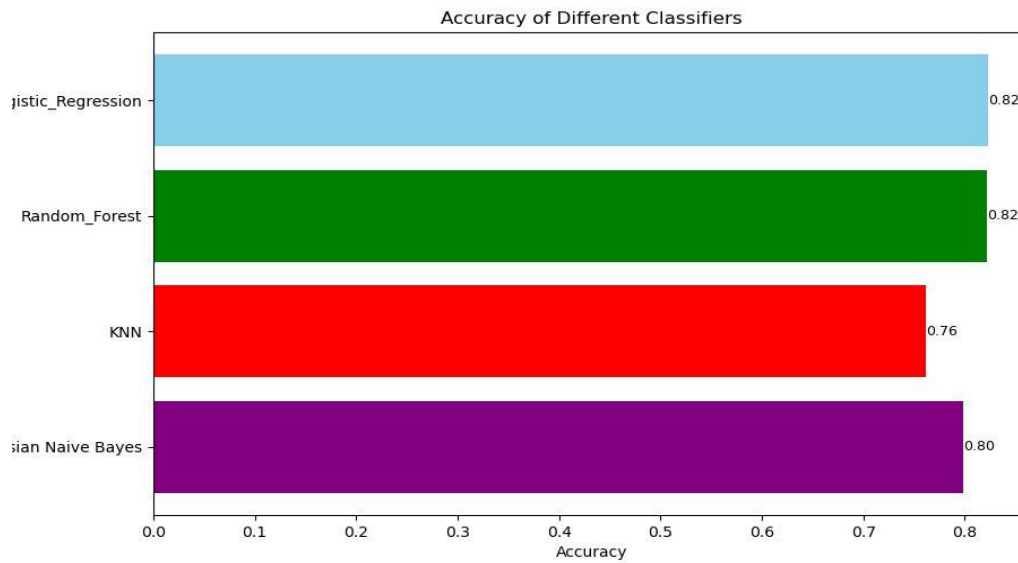


Fig 12: Performance Accuracy.

There are noticeable differences in performance amongst the models when comparing their recall, F1-score, and precision. Because it captures a higher proportion of real positive events, Random Forest has the highest recall (85.93%), while Logistic Regression has the lowest recall (77.89%), possibly because it assumes linear connections between attributes and the target variable. Because of the ensemble learning technique, Random Forest has a better recall than Logistic Regression, but its precision (79.96%) is significantly lower than Logistic Regression's (85.43%), which could result in more false positives and swings in predictions. Nonetheless, Random Forest is effective at identifying actual positive cases while reducing false positives, as evidenced by its balanced F1-score (0.83).

Overall, KNN performs poorly, but due to its instance similarity approach which finds local patterns in the data, it outperforms Logistic Regression in recall, this approach might lead to more false positives, which would decrease accuracy, precision, and F1-score. When compared against KNN, Naive Bayes performs competitively in terms of accuracy, precision, recall, and F1-score, however, it is not as strong as Random Forest and Logistic Regression. Its performance in high-dimensional or nonlinear dataset scenarios, such as the churn prediction, may be limited due to its dependence on feature independence assumptions and incapacity to capture complex relationships.

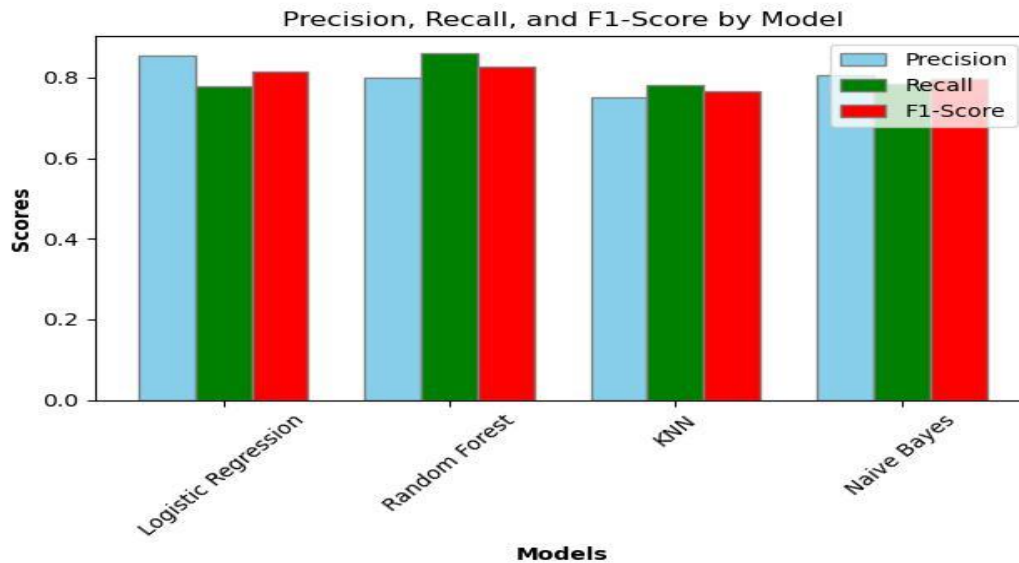


Fig 13: Precision, Recall & F1-Score.

Relevant information on the general performance of the classification model is provided by the ROC-AUC. The model that performed best in the prediction was Random Forest, with a ROC-AUC of 83%. This shows that the model has a good discriminatory capacity to differentiate between positive and negative cases. Despite its linear assumptions, Logistic Regression demonstrated effective ranking of positive cases, as seen by its ROC-AUC of 82%. KNN, on the other hand, had a lower ROC-AUC of 76%, indicating difficulties in correctly differentiating between positive and negative events because of its dependence on local similarity-based categorization. Compared to Random Forest and Logistic Regression, Naive Bayes performed competitively in other measures but may have had weaker discriminatory power (ROC-AUC of 80%) when it came to identifying positive from negative events. Overall, the ROC-AUC observations correspond to performance rankings across multiple evaluation criteria, offering more details about each model's ability to discriminate.

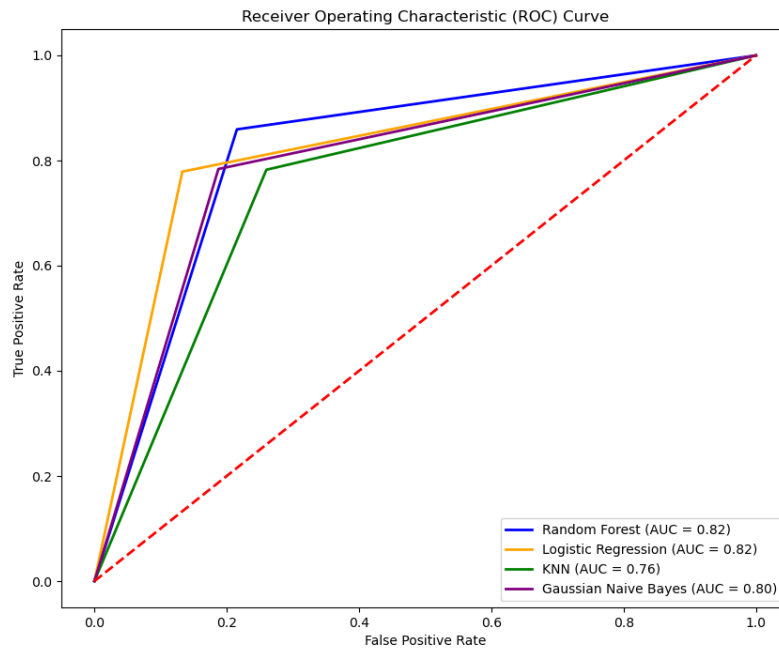


Fig 14: ROC-AUC Curve.

## Conclusion and Recommendation

In conclusion, Random Forest and Logistic Regression were shown to be the best models for predicting churn. They demonstrated good performance in terms of ROC AUC, accuracy, precision, recall, and F1-score, across other metrics. Random Forest performed quite well in detecting true positive situations and avoiding false positives, but Logistic Regression was more precise, which made it a good choice for reducing false positive predictions. Naive Bayes also outperformed the competition despite its naive assumptions, but with slightly lower results in several areas.

For churn prediction analysis, Random Forest or Logistic Regression are recommended based on specific needs and business factors. When accuracy is critical, logistic regression is the better choice, particularly for reducing false positive predictions. However, because Random Forest can handle nonlinear relationships and recognise a variety of patterns in the data, it is useful for attaining balanced performance in terms of precision, recall, and ROC-AUC.

To respond to changing data patterns and business requirements, the implemented churn prediction model must be continuously monitored and adjusted. Over time, optimal accuracy and efficiency are ensured through regular evaluation and parameter adjustment. To improve model resilience and manage complicated data relationships, ensemble techniques and advanced



machine-learning strategies can be explored to further improve performance and produce more accurate and dependable churn forecasts.

## References

1. Ahmad, A.K., Jafar, A. and Aljoumaa, K., 2019. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), pp.1-24.
2. Huang, B., Kechadi, M.T. and Buckley, B., 2012. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1), pp.1414-1425.
3. Lu, N., Lin, H., Lu, J. and Zhang, G., 2012. A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, 10(2), pp.1659-1665.
4. Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C., 2015. A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, pp.1-9.
5. Xie, J., Li, Y., Wang, N., Xin, L., Fang, Y. and Liu, J., 2020. Feature selection and syndrome classification for rheumatoid arthritis patients with Traditional Chinese Medicine treatment. *European Journal of Integrative Medicine*, 34, p.101059.
6. Ahmed, U., Khan, A., Khan, S.H., Basit, A., Haq, I.U. and Lee, Y.S., 2019. Transfer learning and meta classification based deep churn prediction system for telecom industry. arXiv preprint arXiv:1901.06091.
7. Karanovic, M., Popovac, M., Sladojevic, S., Arsenovic, M. and Stefanovic, D., 2018, November. Telecommunication services churn prediction-deep learning approach. In 2018 26th Telecommunications Forum (TELFOR) (pp. 420-425). IEEE.
8. Pires, I.M., Hussain, F., Garcia, N.M., Lameski, P. and Zdravevski, E., 2020. Homogeneous data normalisation and deep learning: A case study in human activity classification. *Future Internet*, 12(11), p.194.
9. Mohammad, N.I., Ismail, S.A., Kama, M.N., Yusop, O.M. and Azmi, A., 2019, August. Customer churn prediction in telecommunication industry using machine learning classifiers. In Proceedings of the 3rd international conference on vision, image, and signal processing (pp. 1-7).
10. Coussement, K. and 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), pp.313-327.

11. Kraljević, G. and Gotovac, S., 2010. Modelling data mining applications for prediction of prepaid churn in telecommunication services. *Automatika*, 51(3), pp.275-283.

12. Kavitha, V., Kumar, G.H., Kumar, S.M. and Harish, M., 2020. Churn prediction of customer in telecom industry using machine learning algorithms. *International Journal of Engineering Research & Technology* (2278-0181), 9(05), pp.181-184.

13. Wadikar, D., 2020. Customer Churn Prediction.

14. Fujo, S.W., Subramanian, S. and Khder, M.A., 2022. Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1), p.24.

15. Gromski, P.S., Xu, Y., Kotze, H.L., Correa, E., Ellis, D.I., Armitage, E.G., Turner, M.L. and Goodacre, R., 2014. Influence of missing values substitutes on multivariate analysis of metabolomics data. *Metabolites*, 4(2), pp.433-452.

16. Fujo, S.W., Subramanian, S. and Khder, M.A., 2022. Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1), p.24.

17. Pires, I.M., Hussain, F., Garcia, N.M., Lameski, P. and Zdravevski, E., 2020. Homogeneous data normalisation and deep learning: A case study in human activity classification. *Future Internet*, 12(11), p.194.

18. Chan, J.Y.L., Leow, S.M.H., Bea, K.T., Cheng, W.K., Phoong, S.W., Hong, Z.W. and Chen, Y.L., 2022. Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8), p.1283.

19. Thompson, C.G., Kim, R.S., Aloe, A.M., and Becker, B.J., 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, 39(2), pp.81-90.

20. Mengash, H.A., Hussain, L., Mahgoub, H., Al-Qarafi, A., Nour, M.K., Marzouk, R., Qureshi, S.A. and Hilal, A.M., 2022. Smart cities-based improving atmospheric particulate matters prediction using chi-square feature selection methods by employing machine learning techniques. *Applied Artificial Intelligence*, 36(1), p.2067647.

21. Wang, Q., Koval, J.J., Mills, C.A., and Lee, K.I.D., 2007. Determination of the selection statistics and best significance level in backward stepwise

logistic regression. *Communications in Statistics-Simulation and Computation*, 37(1), pp.62-72.

22. Kirui, C., Hong, L., Cheruiyot, W. and Kirui, H., 2013. Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. *International Journal of Computer Science Issues (IJCSI)*, 10(2 Part 1), p.165.

23. Sun, S., Cao, Z., Zhu, H. and Zhao, J., 2019. A survey of optimisation methods from a machine learning perspective. *IEEE transactions on cybernetics*, 50(8), pp.3668-3681.

24. Elgeldawi, E., Sayed, A., Galal, A.R. and Zaki, A.M., 2021, November. Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. In *Informatics (Vol. 8, No. 4, p. 79)*. MDPI.

25. Rijnen, M., Hazen, B. and Vletter, C., 2018. Predicting Churn using Hybrid Supervised-Unsupervised Models (Doctoral dissertation, master's thesis. Leiden University).

## Appendix

**TABLE OF FEATURES**

INDEX	FEATURES	DESCRIPTION
1	Customer ID	Customer's unique identification
2	Gender	Customer's gender
3	Age	Customer's age
4	Married	Is customer married or not?
5	Number of Dependents	How many dependents does the customer have?
6	City	Customer's city of primary residence in California
7	Zip Code	Customer's zip code of primary residence
8	Latitude	The customer's primary residence's latitude
9	Longitude	The customer's primary residence's longitude
10	Number of Referrals	Customer's total number of referrals
11	Tenure in Months	The total number of months the client has been with the business by the end of the specified quarter.
12	Offer	Customer's most recent promotional offer
13	Phone Service	Has the customer subscribed to home phone service?
14	Avg Monthly Long-Distance Charges	Customer's Average long-distance charges
15	Multiple Lines	Has the customer subscribed to multiple lines?

16	Internet Service	Did the customer subscribe to internet service?
17	Internet Type	What's the customer's internet type?
18	Avg Monthly GB Download	Customer's average monthly GB download
19	Online Security	Did the customer subscribe to online security?
20	Online Backup	Did the customer subscribe to online backup?
21	Device Protection Plan	Did the customer subscribe to device protection?
22	Premium Tech Support	Did the customer subscribe to tech support?
23	Streaming TV	Did the customer use their internet to stream tv?
24	Streaming Movies	Did the customer use their internet to stream movie?
25	Streaming Music	Did the customer use their internet to stream music?
26	Unlimited Data	Did the customer subscribe to unlimited data?
27	Contract	Customer's Contract
28	Paperless Billings	Did the customer chose paperless billing?
29	Payment Methods	Customer's Payment method
30	Monthly Charges	Customer's total monthly charges
31	Total Charges	Customer's total charges till specified quarter
32	Total Refunds	Customer's total refunds
33	Total Extra Data Charges	Customer's extra data charges
34	Total Long-Distance Charges	Customer's total long-distance charges
35	Total Revenue	Customer's total revenue
36	Customer Status	Customer's status at the end of quarter
37	Churn Category	Customer's churn category
38	Churn Reason	Customer's churn reason