Enhancing Customer Retention in Telecommunications: Leveraging Predictive Algorithms to Forecast Customer Churn in Telco

Project Final Report

Irene Sunny
CentraleSupélec
Gif-sur-Yvette, France
irene.sunny@student-cs.fr

Ru Yi
CentraleSupélec
Gif-sur-Yvette, France
ru.yi@student-cs.fr

Shuqi Deng
CentraleSupélec
Gif-sur-Yvette, France
shuqi.deng@student-cs.fr

Wenjing Zhao
CentraleSupélec
Gif-sur-Yvette, France
wenjing.zhao@student-cs.fr

1. ABSTRACT

The study addresses the issue of customer churn prediction in the telecommunications industry. The study first makes use of exploratory data analysis procedures to obtain insights on how the telecommunication company can decrease churn rates by increasing contract length and decreasing monthly charges. The project then leverages on advanced predictive machine learning algorithms that aims to forecast churn rates. The methodology encompasses five machine learning algorithms: decision trees, logistic regression, support vector machine ("SVM"), Extreme Gradient Boosting ("XGBoost"), and artificial neural networks ("ANN"). Results revealed that the logistic regression, SVM (with hyperparameter tuning), and ANN (with hyperparameter tuning) models performed the best. Accuracy score was used as the main metric for comparison with other qualitative and quantitative metric used in the event where accuracy score is the same across models.

This study contributes to existing literature by comparing various predictive algorithms in the context of telecommunication industry and offers a pragmatic approach for companies in the sector. By adopting the most suitable predictive algorithm model, telecommunication companies like Telco can identify customers at risk of churn and tailor their retention efforts to tackle these customers to reduce churn rates.

On the other hand, this project might also be limited to the size and type of dataset for other telecommunications company as the size, number of variables, and interdependency of variables in the dataset can impact predictions.

2. INTRODUCTION & MOTIVATION

Telecommunications companies seek to understand and predict customer churn (the loss of customers) as it directly impacts the company's revenue and long-term growth. This is because the acquisition of new customers is often more costly than retaining existing customers. Large amount of funds must be invested in marketing strategies to attract new customers. In addition to marketing strategies, customer acquisition in this highly saturated industry where almost everyone has a current telecommunications provider also involves trust building. Companies must invest significantly in discounts, service, and other incentives to successfully convince customers.

The purpose of this project is to determine the main causes for customer churn and compare the use of five common prediction models to identify the most suitable machine learning model for customer churn in the telecommunication industry. The project focuses on the use of decision trees, logistic regression, SVM, XGBoost, and ANN. These models were chosen for their proven capabilities in prior predictive analytics models. The models will be evaluated based on their accuracy in providing reliable predictions of customer churn rate on the sample data obtained from Kaggle that tracks a fictional telecommunications company, Telco.

Results from the study provides potential applications to telecommunications companies where decision makers can make more strategic decisions in business functions like marketing, customer service, product pricing, etc to increase customer acquisition and retention.

3. PROBLEM DEFINITION

The objective is to predict and address the problem of customer churn in the telecommunications industry, In the dataset used, our target variable, customer churn, is a binary variable indicating whether the customer has churned (represented as 1) or remained (represented as 0). The result of the project is limited by the completeness and accuracy of the given dataset.

3.1 Predictive variables

The dataset consists of various information on customer information and plans subscribed that are hypothesized to influence churn rate. There are 19 initial features in the dataset comprising of ‘customerID’, ‘gender’, ‘SeniorCitizen’, ‘Partner’, ‘Dependents’, ‘tenure’, ‘PhoneService’, ‘MultipleLines’, ‘InternetService’, ‘OnlineSecurity’, ‘OnlineBackup’, ‘DeviceProtection’, ‘TechSupport’, ‘StreamingTV’, ‘StreamingMovies’, ‘Contract’, ‘PaperlessBilling’ ‘PaymentMethod’, ‘MonthlyCharges’, and ‘TotalCharges’, with a total of 7043 rows of data. Apart from ‘CustomerID’, the other variables were used in our models. Detailed handling of data would be explored in the following sections.

3.2 Optimization criteria

As mentioned in the introduction, the main goals of this project are to understand the factors involved in decreasing customer churn, as well as to optimise predictive accuracy of churn predictions.

Here is the mathematical notation for accuracy (number of correct predictions over the total number of predictions made):

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}}$$

Additionally, a confusion matrix was plotted to visualise the performance of each algorithm.

4. RELATED WORK

This project used a combination of machine learning methods taken from various research papers dedicated to churn prediction. The paper “Customer churning analysis using machine learning algorithms” explores the use of different machine learning techniques of which logistic regression was selected. SVM, XGBoost, and decision tree models, as well as performance metrics used to evaluate and compare different models were inspired by the paper titled “Telecom churn prediction and used techniques, datasets and performance measures: a review”. Next, newer algorithms like ANN was selected upon analysis of the paper titled “Simulation Modelling Practice and Theory”. Lastly, the paper titled “A training algorithm for optimal margin classifiers” provided deeper insights into the understanding of SVM algorithms used in the project.

Other methods used in predicting customer churn in the telecommunications industry include the use of Naïve Bayes, K-means analysis, CNN, stochastic gradient booster, and random forest.

This paper includes a combination of the best algorithms from each of the reference papers with older and newer machine learning models. It adds value in the form of providing comparison on a range of algorithms in a practical setting.

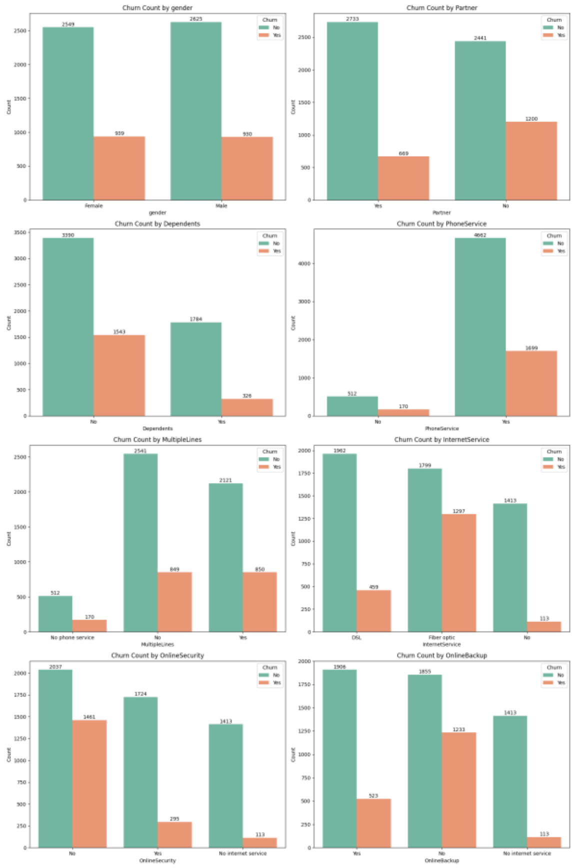
5. METHODOLOGY

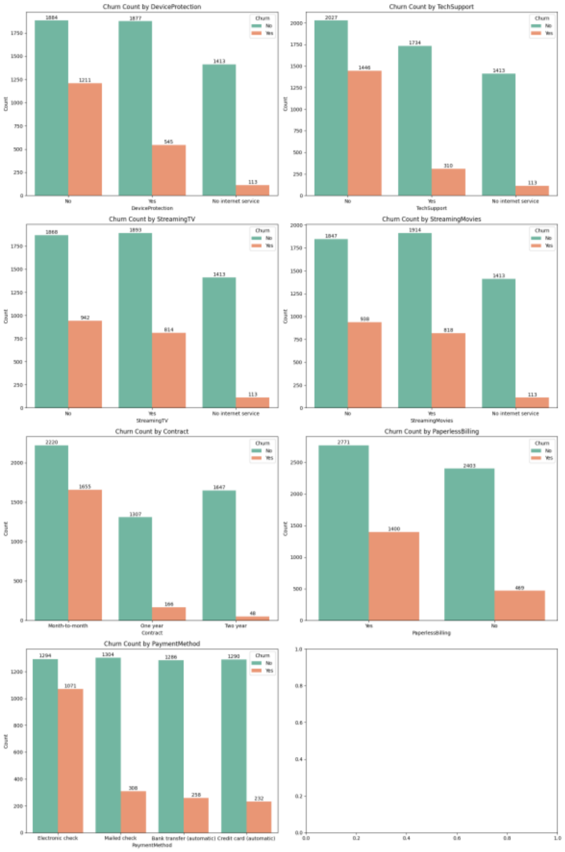
This section outlines the methodology used in this project. The approach involves several key steps: exploratory data analysis, data preprocessing, as well as model comparison and selection.

5.1 Exploratory data analysis (“EDA”)

Firstly, EDA was carried out to facilitate deeper understanding and smooth pre-processing of the data. This stage composed of several steps as listed below:

- Visualising the distribution of data for each variable
- Churn analysis on each variable to determine if there are specific categories with noticeably higher or lower churn rates to identify trends in churn across variables
- Visualising churn rate in relation to customer tenure where it was observed that churn rate decreases significantly as customer tenure increases
- Visualising the distribution of monthly and total charges of customers who have churned compared to those who have remained, observing that churn rate is higher among customers with high monthly charges indicating potential price sensitivity





Delving deeper into each variable, it can be noted that:

- Gender has minimal impact on customer churn potential
- Single customers and those without dependents are more inclined to churn
- Customers with monthly contracts have the potential for customers to churn
- Senior citizens have a higher churn potential
- Customers with fibre optic internet service are more likely to churn
- Customers without online security, online backup, device protection, and technical support have a higher churn potential
- If we look at it based on the Payment Method, customers have a high potential for Churn are customers use Electronic Check.

Further analysis of these results can be found in the evaluation section of this report.

5.2 Data preprocessing

In the data preprocessing stage, data is prepared for modelling through the following steps:

- Converting binary variables to 1 and 0
- One-hot encoding of non-binary variables to turn them into binary (except charges and tenure)
- Obtain correlation matrix to observe the relationship of each variable to churn rate

Evaluations of these observations would be explored in the following evaluations section.

5.3 Model comparison

This section would explore the use of different machine learning modelling techniques on our predictions. The dataset is first split into training and testing sets, with 20% of the data used for testing.

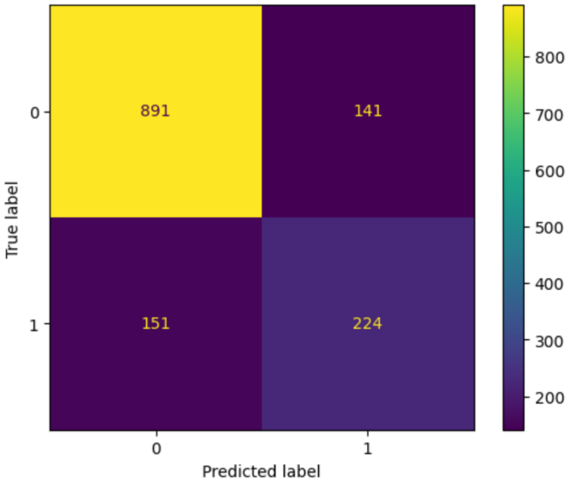
Out of the five algorithms (decision tree, logistic regression, SVM, XGBoost, and ANN), logistic regression, SVM and ANN performed the best with an accuracy score of 0.81 followed by XGBoost and decision tree with 0.80 and 0.79 respectively.

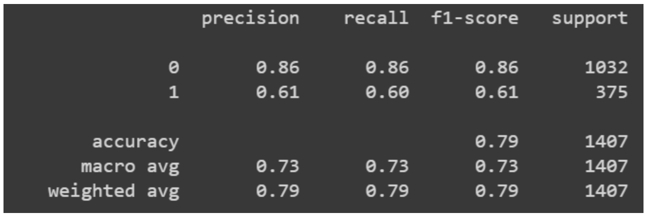
5.3.1 Decision tree

Decision trees are tree-like structures used to establish sets of decisions that can generate classification rules for the dataset. Terminal nodes (leaves) represents class labels, branches represent the combination of features that lead to those class labels, and the root node (topmost decision node) is where the splitting of data starts.

Selection of the best attribute uses measures like information gain, Gini impurity, gain ratio, and chi-square where the tree selects the best attribute at each step to split data into subsets that contain the best homogeneity.

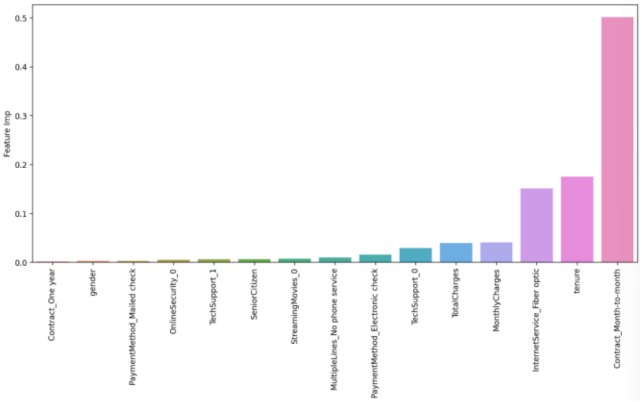
Our decision tree model obtained an accuracy of 0.79 on the test set with the results as shown below:

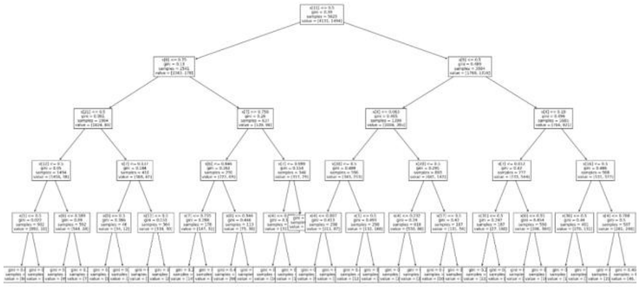




```
              precision    recall  f1-score   support

           0       0.86      0.86      0.86      1032
           1       0.61      0.60      0.61       375

    accuracy                           0.79      1407
   macro avg       0.73      0.73      0.73      1407
weighted avg       0.79      0.79      0.79      1407
```

Next, feature importance was performed to evaluate the importance of each variable used by a trained decision tree classifier. Feature importance is calculated based on the Gini impurity and a higher value means that a specific feature is more important to the prediction. The most important features for this dataset are the presence of monthly contract, tenure, the presence of fibre optic, monthly, and total charges.



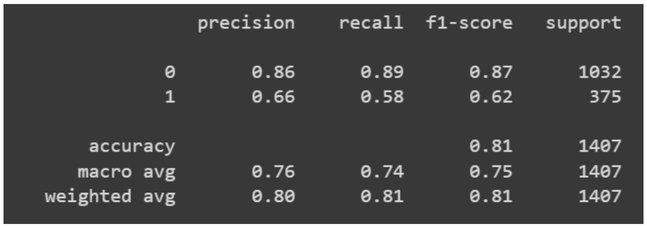


5.3.2 Logistic regression

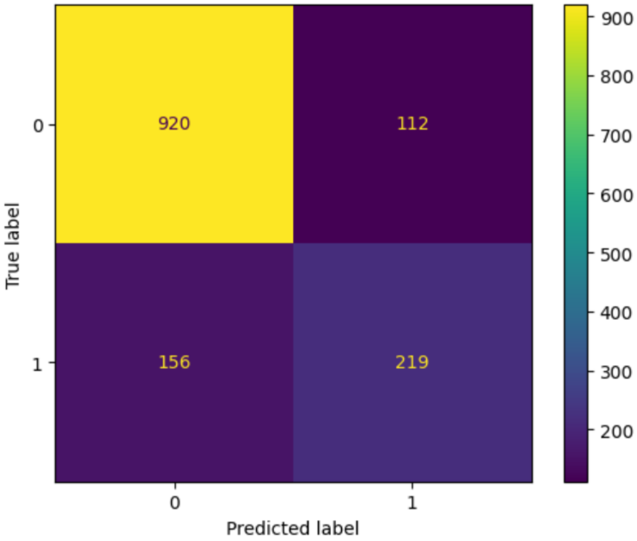
Logistic regressions are used for analysing dataset in which one or more independent variables determine the final output that is binary. This is suitable for this project as the outcome (churn or remain) is a dichotomous variable. The algorithm fits data into a logistic curve to predict the probability of an occurrence.

Apart from being suitable for binary classification, the model also allows for the quantification of the impact of each coefficient on churn probability to allow for the selective focus on important features that are suitable to the dataset. Potential drawbacks include overfitting when applied to small datasets, and linearity assumptions which may limit its ability to handle complex interactions between dependent and independent variables.

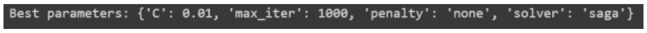
Our logistic regression model obtained an accuracy score of 0.81 which is better than that of the decision tree model with higher number of true positives and slightly lower number of true negatives on the test set.



```
              precision    recall  f1-score   support

           0       0.86      0.89      0.87      1032
           1       0.66      0.58      0.62       375

    accuracy                           0.81      1407
   macro avg       0.76      0.74      0.75      1407
weighted avg       0.80      0.81      0.81      1407
```

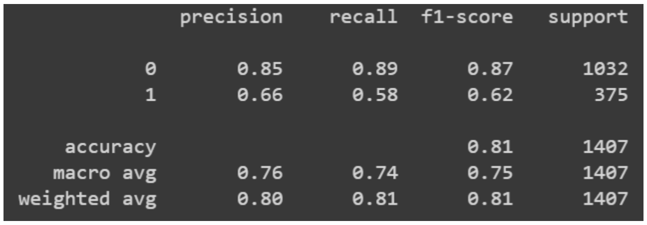


‘GridSearchCV’ from scikit-learn was used to optimise parameters and increase accuracy score with the following results.



```
Best parameters: {'C': 0.01, 'max_iter': 1000, 'penalty': 'none', 'solver': 'saga'}
```

Using this in our updated logistic regression model, the same accuracy score of 0.81 was obtained on the test set. This could be because the default parameters are optimal for the dataset and hyperparameter tuning might not lead to significant improvements, or because the dataset has complex patterns that cannot be captured by the linear model and tuning the model would have limited impact on accuracy.

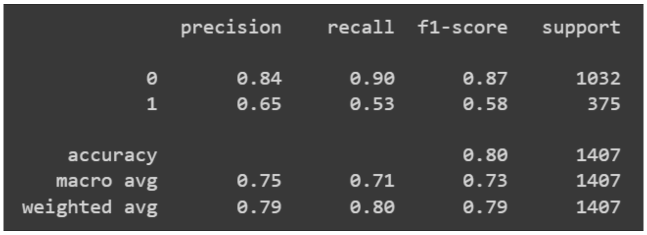


```
              precision    recall  f1-score   support

           0       0.85      0.89      0.87      1032
           1       0.66      0.58      0.62       375

    accuracy                           0.81      1407
   macro avg       0.76      0.74      0.75      1407
weighted avg       0.80      0.81      0.81      1407
```

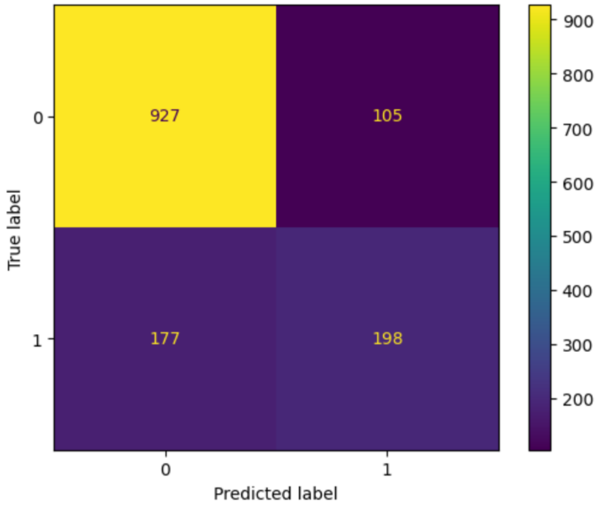
5.3.3 SVM

SVM are used for classification and regression analysis based on structural risk minimisation. Kernel functions are employed to enhance their performance, and ongoing research explores the best kernel selection. SVM is effective in high-dimensional spaces where the number of dimensions exceeds the number of samples and for small/medium sized datasets.

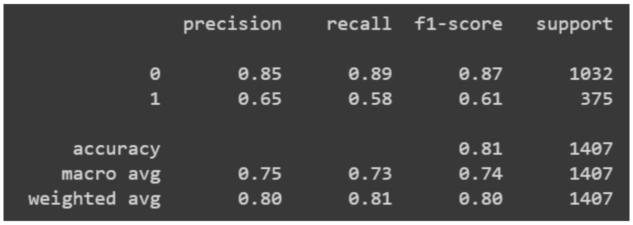
The accuracy score of the test set recorded at 0.80 with the highest number of true positive samples recorded as compared to previous algorithms.



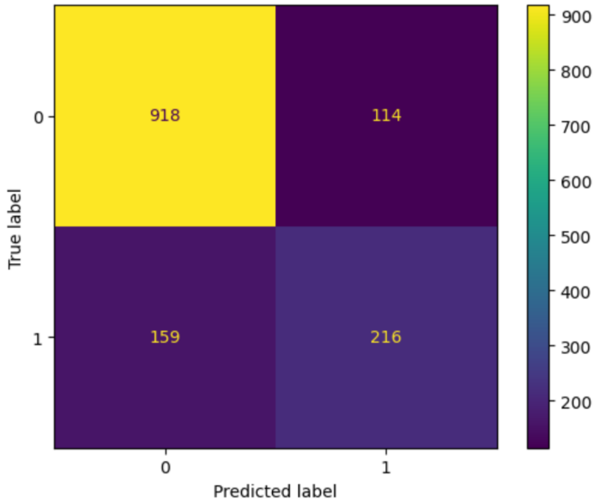
```
              precision    recall  f1-score   support

           0       0.84      0.90      0.87      1032
           1       0.65      0.53      0.58       375

    accuracy                           0.80      1407
   macro avg       0.75      0.71      0.73      1407
weighted avg       0.79      0.80      0.79      1407
```



With the addition of parameters (kernel = 'linear', C = 0.1), a higher accuracy score of 0.81 was obtained with a larger number of true negatives but slightly lower true positives.



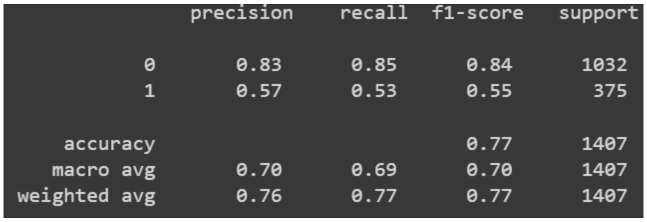
```
              precision    recall  f1-score   support

           0       0.85      0.89      0.87      1032
           1       0.65      0.58      0.61       375

    accuracy                           0.81      1407
   macro avg       0.75      0.73      0.74      1407
weighted avg       0.80      0.81      0.80      1407
```



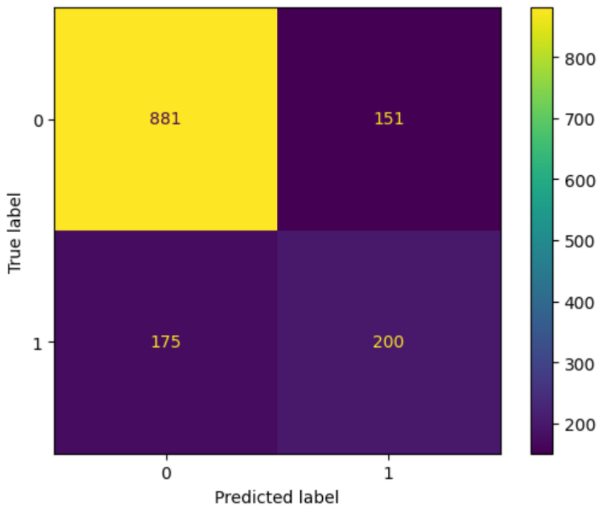
5.3.4 XGBoost

XGBoost is an implementation of gradient boosted decision tree and is highly efficient for classification tasks such as churn prediction. It includes both L1 and L2 regularization to prevent overfitting which is particularly important in churn prediction where the model might overfit to noise in the training dataset. XGBoost handles imbalance in datasets that is suitable for this project as churn numbers are often lower than that of non-churn numbers. A drawback of this algorithm is that it requires parameter tuning to perform well which might be computationally intensive for large datasets. This is however not a problem for this project consisting of a relatively smaller number of samples.

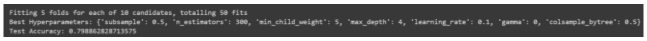
An accuracy score of 0.77 was obtained on the test set.



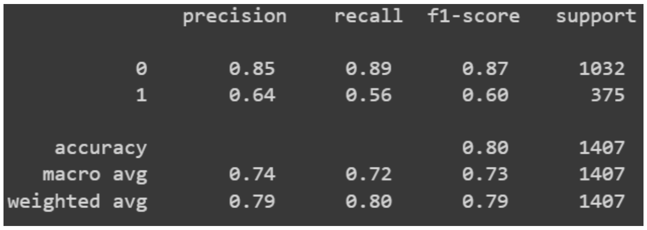
```
              precision    recall  f1-score   support

           0       0.83      0.85      0.84      1032
           1       0.57      0.53      0.55       375

    accuracy                           0.77      1407
   macro avg       0.70      0.69      0.70      1407
weighted avg       0.76      0.77      0.77      1407
```



Next, hyperparameter tuning was performed to obtain the best hyperparameters for the algorithm which increased accuracy score to 0.80.



```
Fitting 5 folds for each of 10 candidates, totalling 50 fits
Best Hyperparameters: {'subsample': 0.5, 'n_estimators': 300, 'min_child_weight': 5, 'max_depth': 4, 'learning_rate': 0.1, 'gamma': 0, 'colsample_bytree': 0.5}
Test Accuracy: 0.798862828713575
```

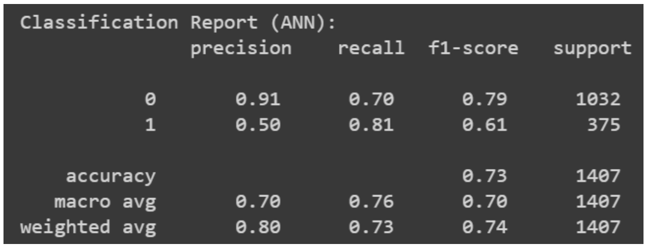


```
              precision    recall  f1-score   support

           0       0.85      0.89      0.87      1032
           1       0.64      0.56      0.60       375

    accuracy                           0.80      1407
   macro avg       0.74      0.72      0.73      1407
weighted avg       0.79      0.80      0.79      1407
```

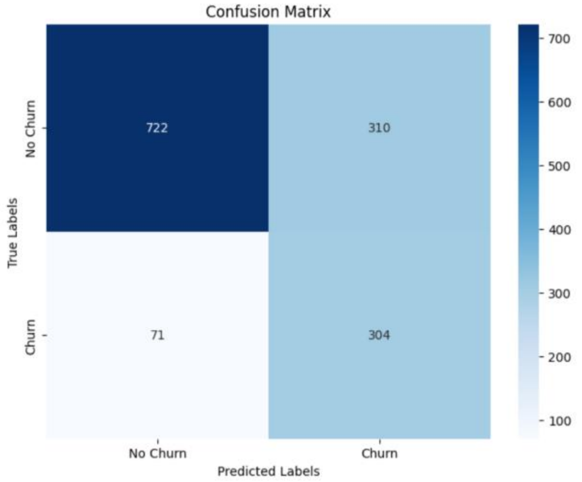
5.3.5 ANN

Lastly, ANN is a machine learning model suitable for learning non-linear patterns and relationships between features in complex datasets. It can be used for classification, regression analysis, and unsupervised learning. ANN consists of layers of interconnected nodes (neuron) where each neuron in one leayer is connected to neurons in the next layer through weights. These networks have an input layer with hidden layers and an output layer. The network can learn complex patterns and relationships. Moreover, ANN can automatically learn relevant features from small datasets to eliminate the need for manual quantifying of the importance of a variable on churn rate. One drawback of ANN is that it requires significant computational resources ad a substantial amount of data to perform well.

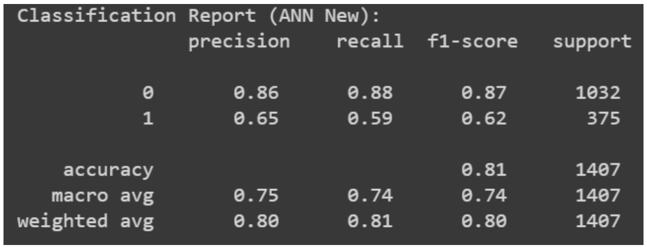
An initial accuracy score of 0.73 was obtained for this model after training the model over 100 epochs.



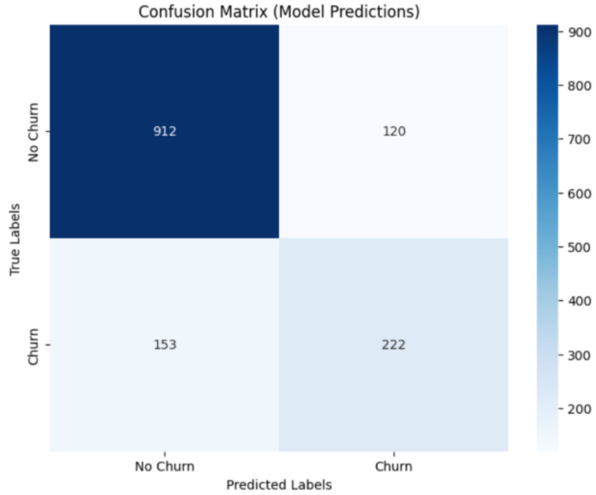
```
Classification Report (ANN):
              precision    recall  f1-score   support

           0       0.91      0.70      0.79      1032
           1       0.50      0.81      0.61       375

    accuracy                           0.73      1407
   macro avg       0.70      0.76      0.70      1407
weighted avg       0.80      0.73      0.74      1407
```



Upon the use of hyperparameter tuning with customised hyperparameters, accuracy score increased to 0.81.



```
Classification Report (ANN New):
              precision    recall  f1-score   support

           0       0.86      0.88      0.87      1032
           1       0.65      0.59      0.62       375

    accuracy                           0.81      1407
   macro avg       0.75      0.74      0.74      1407
weighted avg       0.80      0.81      0.80      1407
```



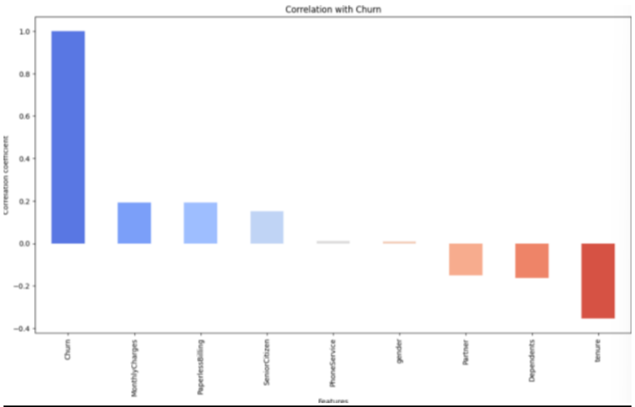
6. EVALUATION

This section explores the results findings of the project in relation to the two main objectives: to find out the main reasons for customer churn, and the best model for this dataset.

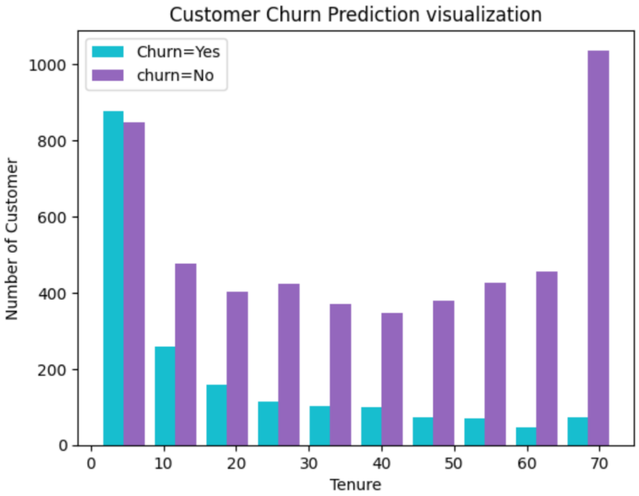
6.1 Main reasons for customer churn

Results from the EDA stage was used to identify the features that impact customer churn.

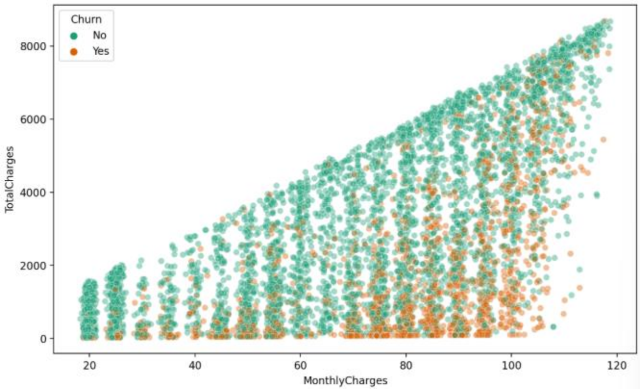
The correlation matrix identified monthly charges and churn as the top two most significant features predicting customer churn.



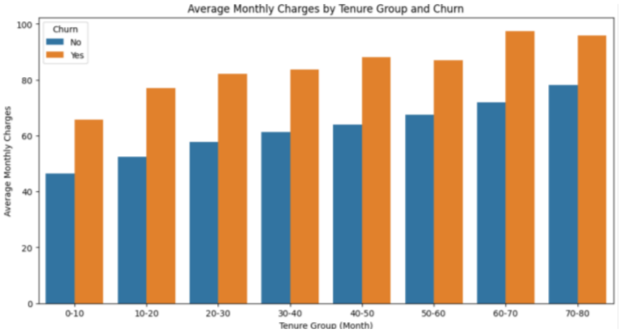
Next, customer churn rate was plotted in relation to tenure and it was found that customers with a higher tenure had a lower tendency to churn.



Moving on, plotting churn in relation to monthly charges showed that customers with higher monthly charges had a higher tendency to churn.



Therefore, the company should reduce monthly charges and extend the duration of contracts to retain customers as seen from the graph below where customers are more inclined to churn when tenure is shorter and monthly charges are high. Apart from revising pricing strategies, the company can also provide targeted incentives like loyalty discounts, enhanced customer service, complimentary add-ons to retain customers at the early stages of the customer-company relationship.



Additionally, using the results of variable analysis, the company can consider providing complimentary security related services such as online security, online backup, device protection, and technical support to retain customers. The company should also investigate their fibre optic service as customers with fibre optic internet service have a much higher potential to churn as compared to customers with DSL internet or no internet service. This might indicate service problems with the fibre optic service causing customer dissatisfaction. The same can be said for the electronic check payment method where the difficulty of payment could be a reason for large churn. Next, to tackle senior customer churn rates, the company can seek feedback from this group of customers so as to better tailor their services to serve their differing needs.

In conclusion, the company has to also provide better service in addition to changing the pricing strategy so as to retain customers.

6.2 Model selection

Moving on to model selection, it can be observed that logistic regression, SVM (with hyperparameter tuning), and ANN (with hyperparameter tuning) obtained the highest accuracy. Therefore, other factors must be considered when determining the most efficient model for this dataset.

- Computational efficiency: logistic regression is the most computationally efficient amongst the three models as SVMs can have complex hyperparameters and kernel functions while ANNs have multi-layer networks that can require significant computational resources.
- Implementation: logistic regression is straightforward in implementation. SVMs and ANNs require hyperparameter tuning to provide more accurate predictions.
- F1 score: F1 score is the combination of precision and recall where a higher F1 score indicates good precision and recall scores. If accuracy is the same, F1 score can be used as another metric of comparison. In relation to this dataset, the F1 score for logistic regression is 0.87 and 0.62 on non-churn and churn data respectively, while that of SVM is 0.87 and 0.61, and ANN 0.87 and 0.62. Therefore, they have similar F1 score with logistic regression and ANN performing slightly better than SVM.

$$\boldsymbol{Fscore = 2 \times \frac{Precision \times Recall}{Precision + Recall}}$$

Where $\boldsymbol{Precision = \frac{TP}{TP+FP}}$ & $\boldsymbol{Recall = \frac{TP}{TP+FN}}$

Considering the above factors, for this dataset, logistic regression seems to be the most suitable machine learning model as it is computational efficient, easy to implement, and returns the highest F1 score.

7. CONCLUSION

The above results are only applicable to this dataset or similar datasets. For other datasets, the results might be different based on many factors such as the size of the dataset, the number of features in the dataset, whether the dependent variable is binary or nonbinary, and the relationship complexity between variables in the dataset.

Future research could explore the integration of real time data and customer feedback to ensure relevance and accuracy of predictions, dynamic learning of the model that could improve predictions over time, and the use of customer feedback as a qualitative insight into reasons for churn to refine churn prediction. Furthermore, continuous data can also aid the model in identifying anomalies in a shorter amount of time. Additionally, due to one hot-encoding of non-binary variables, these variables do not appear on the correlation matrix resulting in a loss of accuracy. Techniques such as the use of feature importance using tree-based machine learning models such as the decision tree model (previously used) can provide insights into which features are most influential in predicting the target variable. Other visualization methods such as the mosaic plot could also be utilised to understand variable relationships in categorical data.

8. REFERENCES

[1] H. Jain, A. Khunteta, and S. Srivastava. 2020. Telecom churn prediction and used techniques, datasets and performance measures: a review. Telecommun Syst 76, 613–630 (2021). https://doi.org/10.1007/s11235-020-00727-0

[2] Jia Wertz. 2018. Don't Spend 5 Times More Attracting New Customers, Nurture The Existing Ones. Forbes. https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new-customers-nurture-the-existing-ones/?sh=badf6205a8e0

[3] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K.Ch. Chatzisavvas. 2015. A comparison of machine learning techniques for customer churn prediction, Simulation Modelling Practice and Theory, Volume 55, 2015, Pages 1-9, ISSN 1569-190X https://www-sciencedirect-com.essec.idm.oclc.org/science/article/pii/S1569190X15000386

[4] B.E. Boser, I.M. Guyon, V.N. Vapnik. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM (1992), pp. 144-152. https://dl.acm.org/doi/10.1145/130385.130401

[5] Prabadevi, B., Shalini, R. and Kavitha, B.R. 2023. Customer churning analysis using machine learning algorithms. https://doi.org/10.1016/j.ijin.2023.05.005