Executive Summary

The capstone project focuses on predicting customer churn for a telecommunications company. Churn prediction is essential for the company to proactively address customer retention and mitigate revenue loss. The project utilizes a combination of conventional machine learning and neural network approaches to build robust predictive models. The key steps undertaken in this project include data understanding and preparation, model development, evaluation, and hyperparameter tuning.

Data Understanding and Preparation

The dataset contains around 100,000 observations with 225 features, capturing various aspects of customer behavior, such as recharge amounts, call usage, internet usage, and service schemes. We began by understanding and exploring the data, handling missing values, and conducting feature engineering. Exploratory data analysis is conducted to identify key features influencing churn, such as contract type, tenure, and monthly charges.

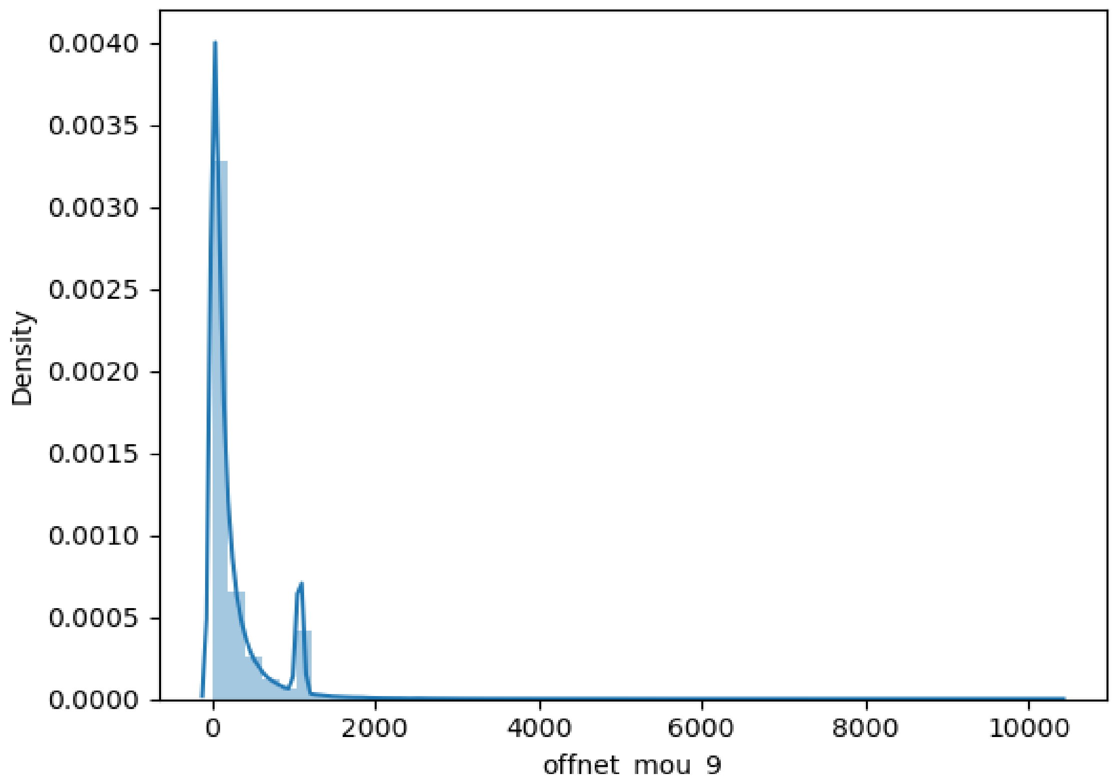
Preprocessing Steps: -

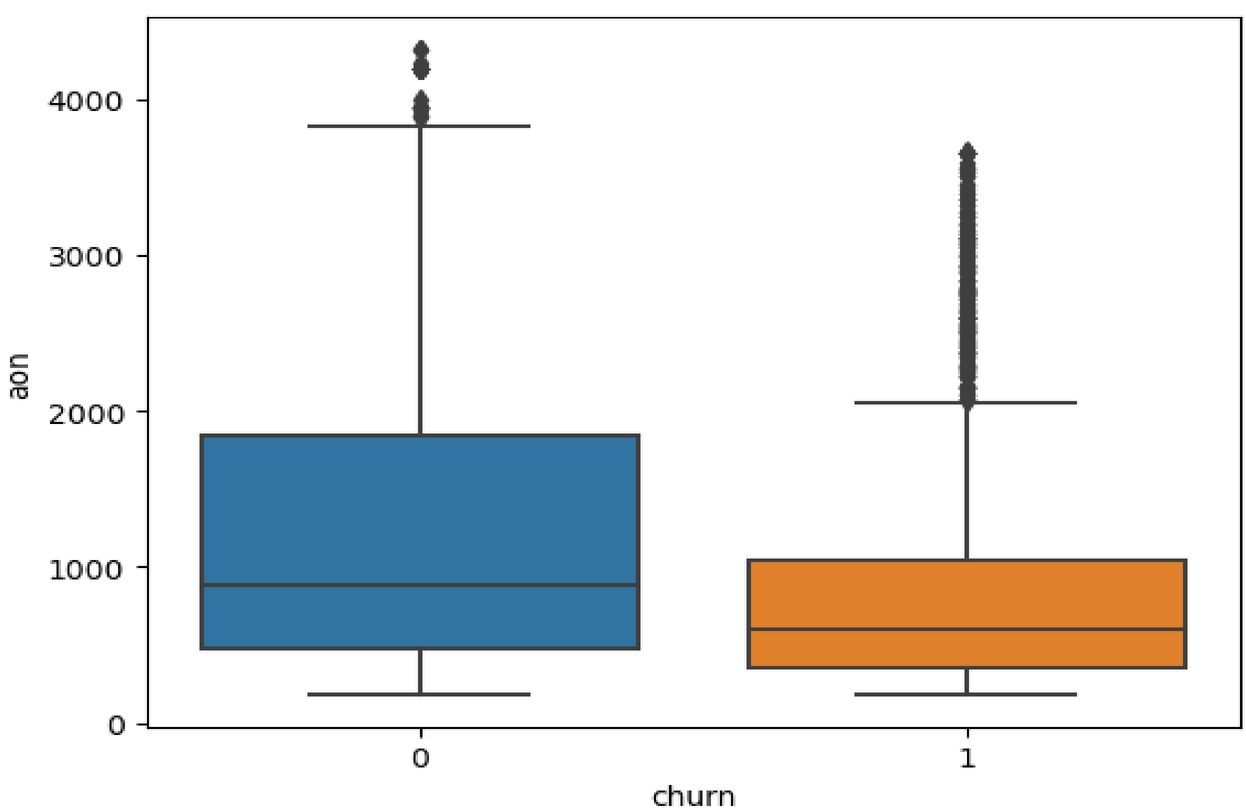
- Missing Value Ratio.
- Drop variables with more than a given threshold.
- Handling missing values using MICE.
- Categorical encoding.
- Feature Engineering.
- Derive Churn

Data Visualization: -

Performed the univariate and bivariate EDA to visualize the relationship between different features in the data set. Which includes: -

- Average Revenue per User
- Minute usage of Local / Std
- Usage of calls within the same operator.
- Usage of calls outside the operator


0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Density
0
2000
4000
6000
8000
10000
offnet_mou_9


4000
3000
2000
1000
0
aon
0
1
churn

Implemented the outlier capping function to cap the outliers present in all the numeric columns as shown in the above boxplot.

Handling Class Imbalance

Classification tasks often involve datasets with class imbalances, where the number of samples in one class significantly outweighs the other(s). Class imbalance can pose significant challenges to learning algorithms, as they tend to favor the majority class and struggle to accurately predict the minority class. Data augmentation is one technique to address this issue; however, this project explores class imbalance techniques as a primary approach or in conjunction with augmentation methods.

In this dataset, the churn class ('1') is approximately 1/10th of the non-churn class ('0'). This imbalance means a simple model could achieve 90% accuracy by predicting all customers as non-churners, but the focus should be on accurately predicting churners.

To address this, we use sampling methods to balance the dataset:

- Random Under-Sampling: This method involves removing data to balance the classes, resulting in a 50/50 ratio. For instance, if there are 1221 '0' entries, the method ensures 1221 '1' entries by removing excess '0' entries. The downside is potential information loss.
- Random Over-Sampling: This method adds data to balance the classes, again achieving a 50/50 ratio. If there are 13780 '1' entries, the method ensures 13780 '0' entries by duplicating '1' entries.

Conventional Machine Learning Model

Various machine learning algorithms, including logistic regression, decision trees, KNN, and random forests are employed. These models are evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The performance of each model is compared, and the best-performing model is **Decision Tree with Random Oversampling** selected for further tuning.

	Model Name	Training Score	Testing Score	F1 Score	Precision	Recall
0	Logistic Regression - without balancing	0.940071	0.937133	0.930030	0.686406	0.418033
1	Logistic Regression - Random Undersampling	0.840758	0.838446	0.838442	0.834008	0.844262
2	Logistic Regression - Random Oversampling	0.837310	0.844884	0.844884	0.844259	0.845791
3	Decision Tree - without balancing	1.000000	0.908867	0.912170	0.448709	0.527049
4	Decision Tree - Random Undersampling	1.000000	0.773006	0.773000	0.769231	0.778689
5	Decision Tree - Random Oversampling	1.000000	0.973512	0.973494	0.949690	1.000000
6	kNN - without balancing	0.936338	0.932867	0.918666	0.730022	0.277049
7	kNN - Random Undersampling	0.826421	0.809816	0.808919	0.857820	0.741803
8	kNN - Random Oversampling	0.895682	0.877903	0.877607	0.844070	0.927068
9	Random Forest - without balancing	0.911339	0.898467	0.910332	0.428165	0.740164
10	Random Forest - Random Undersampling	0.876088	0.852761	0.852108	0.905660	0.786885
11	Random Forest - Random Oversampling	0.868197	0.868106	0.867819	0.905962	0.821480

Model Evaluation and Hyperparameter Tuning

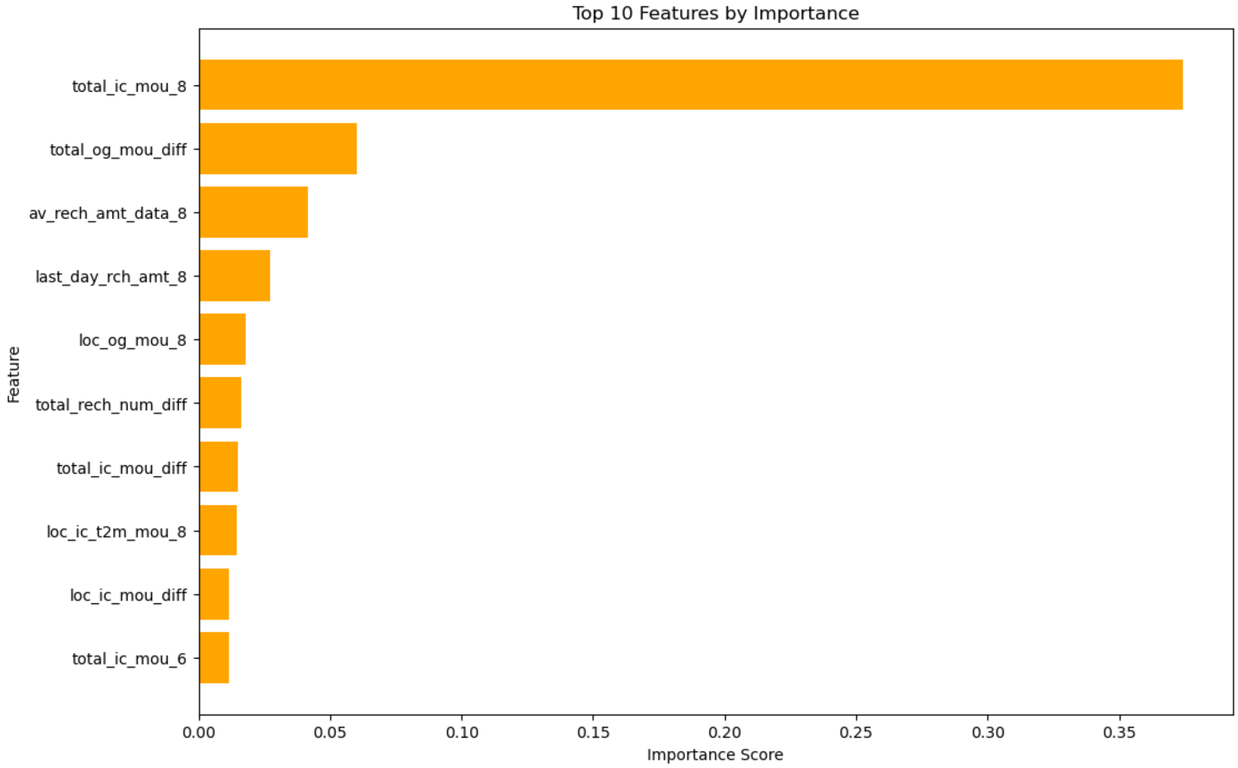
The Decision Tree model undergoes hyperparameter tuning using GridSearchCV to optimize its performance. The tuned model is evaluated on the validation set to ensure its generalizability. And the model is retrained on the basis of best parameters found that is max_depth = 30and random state = 0.

Model Performance:

- Training Accuracy: 0.9981
- Testing Accuracy: 0.9735
- Test F1 Score: 0.9734
- Test Precision: 0.9496
- Test Recall: 1.0

Feature Importance

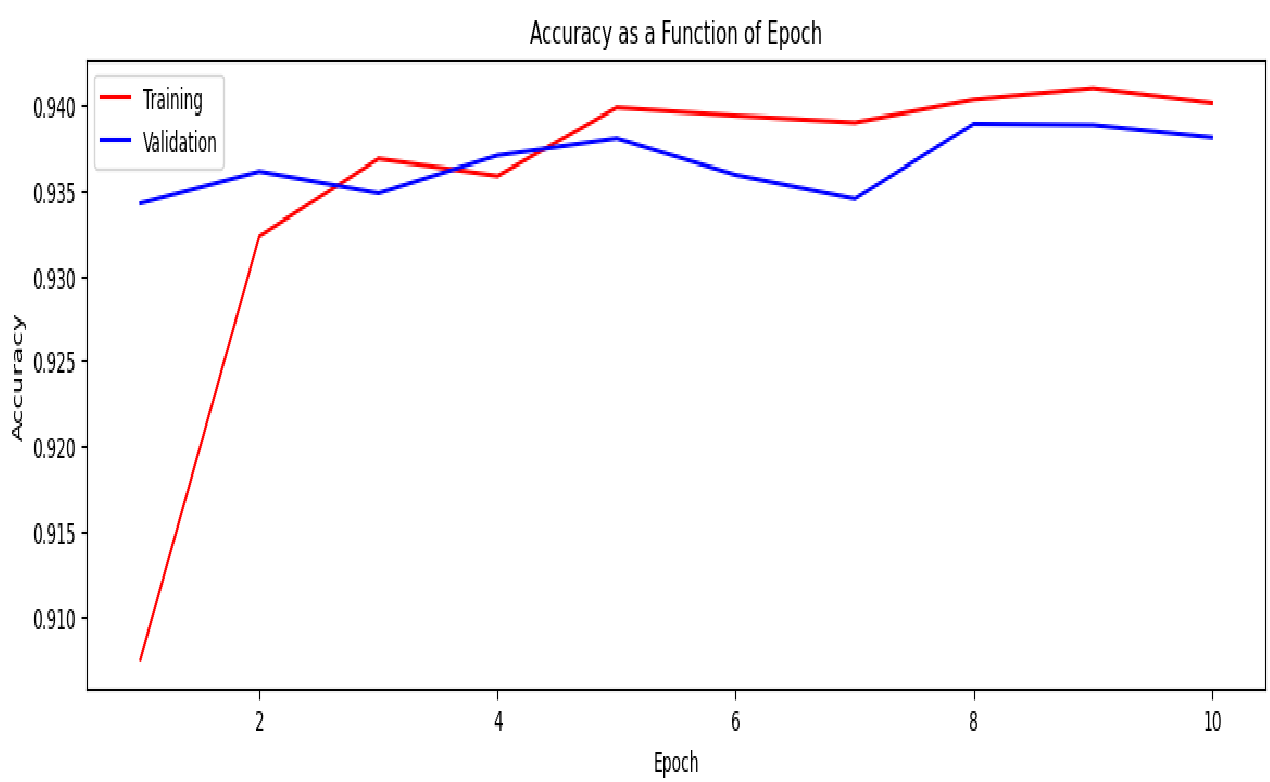
The top 10 features with highest importance are:



Neural Network Model

A neural network model is developed to capture complex patterns in the data. The model was built using the Keras library within TensorFlow. Input Layer: The input layer was defined to match the number of features in the dataset. Hidden Layers: The first hidden layer's number of neurons was a variable hyperparameter. The second hidden layer had a fixed number of 64 neurons. Output Layer: The output layer consisted of a single neuron with a 'sigmoid' activation function to perform binary classification.

	accuracy	loss	val_accuracy	val_loss
epoch				
1	0.907473	1.880255	0.934267	0.209761
2	0.932338	0.223424	0.936133	0.234840
3	0.936871	0.234422	0.934867	0.324643
4	0.935871	0.221587	0.937067	0.200982
5	0.939871	0.229442	0.938067	0.225687
6	0.939404	0.244529	0.935933	0.199389
7	0.939004	0.224048	0.934533	0.196693
8	0.940337	0.226778	0.938933	0.228453
9	0.941004	0.212312	0.938867	0.203645
10	0.940137	0.217520	0.938133	0.806105



The loss value of the model on the validation data is 0.8061050772666931
The accuracy of the model on the validation data is 0.9381333589553833

Hyperparameter Tuning

GridSearchCV: Parameters tuned: Activation function and number of neurons in the first hidden layer. Optimal parameters found: Activation function: 'sigmoid' Hidden layer neurons: 512. Accuracy achieved with optimal parameters: 0.9396

RandomizedSearchCV: Parameters tuned: Activation function and number of neurons in the first hidden layer. Optimal parameters found: Activation function: ‘sigmoid’ , Hidden layer neurons: 448 Accuracy achieved with optimal parameters: 0.9330

Misclassification Cost

Define Costs:

- False Negative: $500
- False Positive: $300

Model Performance:

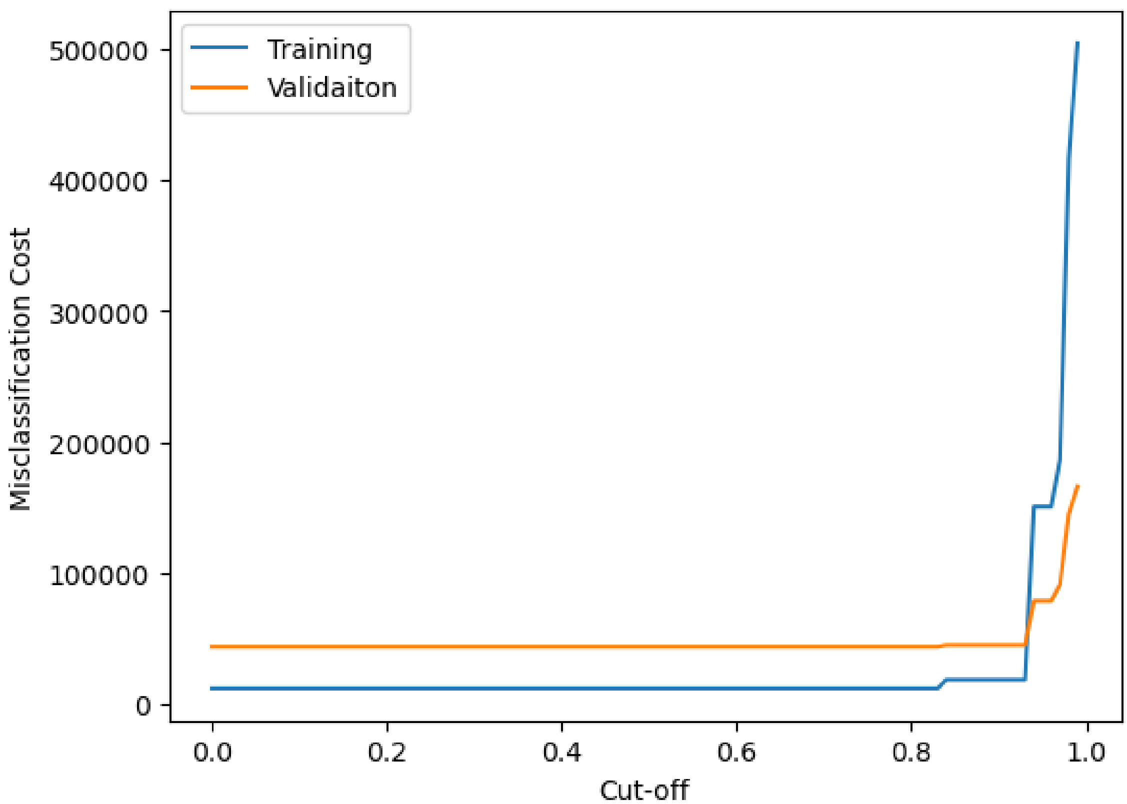
- Trained Decision Tree Classifier (max depth = 30)
- Training Accuracy: 99.8%
- Validation Accuracy: 97.4%
- F1 Score: 0.9735
- Precision: 0.9497
- Recall: 1.0

Current Misclassification Cost:

- False Positives: 146
- False Negatives: 0
- Cost: $43,800

Optimize Cut-off Value:

- Tested cut-off values from 0 to 1.
- Best cut-off on training data: 0.000
- Training Cost at Best Cut-off: $12,000
- Validation Cost with Best Cut-off: $43,8



Business Insights and Recommendations

Insights into Profitability of the telecommunication service program

For telecom firms, customer attrition is a crucial indicator since it has an immediate effect on profits. Proactively predicting churn allows the company to implement retention strategies to keep customers and maintain steady revenue streams. The high accuracy and recall rates indicate that the model is highly effective in predicting churn, ensuring that nearly all churners are correctly identified. This is crucial for minimizing revenue loss. By effectively addressing class imbalance, the model improved its ability to predict churn accurately, further enhancing profitability through better-targeted retention efforts.

The model's performance led to $43,800 in misclassification costs with 146 false positives and 0 false negatives. Nonetheless, training costs were considerably decreased by optimising the cut-off value for predictions, suggesting that continual optimisation may be able to further reduce costs.

Assessing Potential Impact of Misclassification Costs

1. Economic Impact: The costs of misclassification have an immediate impact on the business's revenue. False negatives, also known as missed churners, can result in a large loss of income when these users quit the service without any help. Conversely, false positives—erroneously identified as churners—require needless retention expenses. Because no prospective churner is overlooked, the model's ability to reduce false negatives to zero suggests a large positive impact on profitability.

2. Model Optimisation: Further lowering training costs considerably involved improving the cut-off value for predictions. Testing was used to find the ideal cut-off value, which balanced the trade-off between false positives and false negatives. As a result, training data had an optimised cost of $12,000, while validation costs stayed at **$43,800**. This cut-off can be continuously improved.

Recommendations for Leveraging NLP Techniques, LLMs, and Generative AI

1. Examining Data on Customer Interaction:
An analysis of text data from customer interactions, including support tickets, call transcripts, and social media feedback, can be done with Natural Language Processing (NLP) approaches. To find patterns and sentiments that are associated with churn, important domains to focus on are sentiment analysis, topic modelling, and keyword extraction.

2. Applying New Developments in Generative AI and LLMs:
Advances in Generative AI and Large Language Models (LLMs) have made it possible to generate and understand human language with greater nuance. Using these models, one can:

Predict Churn: By examining the context of the conversation and spotting signs of discontent.
Personalised Interventions: Using information from past interactions, develop customised communication plans to keep at-risk clients.
Automated help: Increase customer happiness and possibly lower attrition by providing intelligent, context-aware automated help responses.

3. Concerns about data privacy and resource limitations:
Adopting these cutting-edge methods calls for substantial processing power and sound data management procedures. In order to overcome resource constraints:

Scalable Cloud Solutions: For scalable processing and storage, make use of cloud services.
Gradual Implementation: Prior to a full-scale implementation, begin with pilot projects to show value. Data privacy is crucial, particularly when it comes to private client information. To guarantee adherence to laws governing data privacy. Anonymise data if it is practical to do so in order to safeguard client identities. Enable access restrictions and end-to-end encryption for secure data handling. Transparency: Get client consent and uphold clear data usage regulations.

4. Implementation Difficulties:
Implementing these methods presents the following challenges:
Technical expertise: Needs knowledgeable AI experts and data scientists.
Integration with Current Systems: Robust integration with current analytics and CRM platforms.
Constant Monitoring: To keep models accurate and relevant, update and monitor them frequently.

5. Creative Solutions: Take into account the following ideas to get beyond these obstacles:
Collaborative Partnerships: For knowledge and assistance, team up with AI and IT companies.
Training Programs: Spend money on NLP and AI technology upskilling for internal staff.
Agile Development: Quickly iterate and modify solutions in response to feedback and performance indicators by utilising agile approaches.

Conclusion

In conclusion, the telecommunication service program's capacity to anticipate and reduce customer attrition can be greatly improved by utilising cutting-edge machine learning and AI techniques. Through the implementation of LLMs and Generative AI, addressing misclassification costs, and using NLP for deeper insights, the organisation can enhance profitability while simultaneously providing a more customised and fulfilling client experience.