# Telecom Customer Churn Prediction – Final Report (Ullas Manjunatha)

## 1. Introduction

This project aims to predict customer churn in a telecommunications company using machine learning. Customer churn refers to the loss of clients who discontinue their service. By predicting churn early, the company can implement targeted strategies to retain customers and improve revenue stability.

## 2. Models Used

Several machine learning models were tested to determine the most effective approach for churn prediction. The following models were implemented and compared:

* 1. Logistic Regression – A baseline model for binary classification.
* 2. Decision Tree – Captures non-linear patterns in customer data.
* 3. Random Forest – Ensemble model that reduces variance and overfitting.
* 4. Gradient Boosting – Sequential learning to improve weak predictions.
* 5. XGBoost – Optimized gradient boosting model known for accuracy and speed.
* 6. K-Nearest Neighbors (KNN) – Simpler instance-based learning method for comparison.

## 3. Model Performance Summary

Each model was evaluated based on Accuracy, Precision, Recall, and F1 Score on both validation and test datasets. The performance summary is presented below:

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Validation Accuracy | Test Accuracy | F1 Score (Test) |
| Logistic Regression | 79% | 80% | 0.60 |
| Decision Tree | 80% | 79% | 0.59 |
| Random Forest | 80% | 79% | 0.59 |
| Gradient Boosting | 82% | 80% | 0.60 |
| XGBoost (Base) | 93% | 81% | 0.64 |
| XGBoost (Model 3) | 78% | 76% | 0.65 |
| KNN | 83% | 78% | 0.58 |

## 4. Final Model Selection – XGBoost (Model 3)

After extensive experimentation, Model 3 (Tuned XGBoost) was selected as the final model. It achieved a validation accuracy of 78.3% and a test accuracy of 76.1%, with an F1 score of 0.65. Model 3 provides an excellent balance between precision and recall, ensuring reliable identification of churners while avoiding overfitting seen in earlier versions.

## 5. Recommendations for Business Use

1. Identify high-risk customers (churn probability > 0.55) and target them with personalized offers or loyalty rewards.
2. Analyze top churn-driving factors (e.g., contract type, tenure, monthly charges) to guide retention strategies.
3. Use churn predictions to prioritize proactive customer engagement and service improvements.
4. Integrate this model into the CRM system for automated churn monitoring and alerts.
5. Continuously retrain the model with new customer data to maintain predictive performance.

## 6. Conclusion

The Tuned XGBoost (Model 3) outperformed other models by maintaining consistent results across validation and test data. It generalizes well and effectively balances recall and precision, making it suitable for real-world deployment. By implementing this model, the telecom company can reduce churn rates, improve customer satisfaction, and enhance long-term business growth.