Project Report: Telecom Customer Churn Prediction

Team
Abdul Kuddus (M24DE3002)(G23AI2044)
Ritik Sharma (M24DE3065)(G23AI2024)
Mohit Mathur (M24DE3050)(G23AI2034)
Jojo Joseph (M24DE3041)(G23AI2100)

1. Executive Summary

The project focuses on predicting customer churn in the telecom sector using machine learning. Churn prediction is vital for customer retention and business sustainability. This project leverages structured telecom data to build models that can identify customers likely to discontinue their service.. This project successfully developed, trained, and compared deep learning models (ANN and 1D CNN) to predict customer churn in the telecom sector using US telecom data (653,753 rows $\times$ 74 columns). The goal was to identify customers likely to discontinue service to enable targeted retention campaigns. Despite a severe class imbalance (4.57% churn rate), the Artificial Neural Network (ANN) achieved the best overall performance, with 95.9% accuracy and a Recall of 0.1441, indicating superior ability to detect actual churners compared to the CNN. Key business insights highlight high ARPU customers and specific data usage patterns as major churn risk factors.

2. Problem Definition and Business Context

The objective of this project is to develop a predictive model capable of identifying customers who are most likely to churn, enabling the business to take proactive measures to retain them. By analyzing a comprehensive dataset containing demographic information, service usage patterns, account details, and customer behavior, the model aims to uncover patterns and factors influencing churn.

Therefore, the problem can be defined as follows: <u>“Given historical customer data, predict the probability that a customer will discontinue using the telecom service (churn) within a defined future period.”</u> This problem requires a combination of data preprocessing, feature engineering, and Deep learning modeling to produce accurate, interpretable, and actionable churn predictions that drive business value.

3. Project Details

Data Source and Preprocessing

Aspect	Details
Dataset	US Telecom Data
Shape	653,753 rows, 74 columns
Target Variable	Churn Value (Binary: 0 = Retained, 1 = Churned)
Churn Rate	4.57%
Missing Values	total_rech_data filled with 0 or mean (≈ 4.85); Internet Type filled with "Not Applicable."
Outliers	Detected and replaced in arpu_4g, arpu_5g, and vol_5g (extreme values replaced with 0).
Feature Engineering	Created total_recharge (total_rech_amt + total_rech_data) and derived Quarter and Quarter of Joining.
Scaling	StandardScaler applied to all numerical features.
Encoding	One-hot encoding for 18 categorical columns.

Missing values:

- total_rech_data: Filled with 0 when ARPU was “Not Applicable,” else filled with mean (≈ 4.85).
- Internet Type: Filled with “Not Applicable.”
- Outliers: Detected and replaced in arpu_4g, arpu_5g, vol_5g (values like 254,687 and 87,978 replaced with 0).

Feature Engineering:

- Created total_recharge = total_rech_amt + total_rech_data
- Derived Quarter and Quarter of Joining from month columns.
- Categorical Encoding: One-hot encoding for 18 categorical columns.
- Scaling: StandardScaler applied to numerical features.
- Churn rate: 4.57%

4. Model Building and Evaluation and Comparison of Approaches

Four deep learning models were implemented and evaluated on the test set

This document compares the modeling approaches used in the churn prediction project across methodology, strengths/weaknesses, and evaluation metrics.

4.1 Qualitative Comparison

★ **Neural Network (ANN) :**
  - Strengths: Flexible function approximator; can model complex interactions.
  - Weaknesses: Requires careful preprocessing; sensitive to class imbalance; longer training.
  - Use when: You can invest in architecture/tuning and want a deep learning benchmark.

★ **1D CNN :**
  - Strengths: Can learn local patterns if features are ordered meaningfully.
  - Weaknesses: Tabular features do not always have spatial locality; may underperform tree models.
  - Use when: Feature ordering conveys locality or you engineer embeddings/blocks for structure.

★ **Transformer (Self-Attention) :**
  - Strengths: Captures long-range feature interactions via attention; flexible architecture.
  - Weaknesses: Requires careful design for tabular data; may need more data/tuning; compute heavier than ANN.
  - Use when: You want to model global interactions beyond local patterns with a modern deep architecture.
  -

★ **TabNet (Google AI, 2021) :**
  - Strengths: Designed specifically for tabular data; built-in feature selection via sequential attention; interpretable feature importance masks; competitive with gradient boosting; handles both categorical and continuous features natively.
  - Weaknesses: More hyperparameters to tune (decision steps, attention dimensions); slower training than tree-based models; requires careful regularization (sparsity coefficient).
  - Use when: You need interpretability with deep learning performance on tabular data; want to understand which features drive predictions; have mixed feature types.

4.2 Quantitative Comparison (Metrics)

Model	Accuracy	Precision	Recall	F1	ROC AUC
ANN (Keras)	0.9590	0.7589	0.1047	0.2422	0.5710
1D CNN	0.9584	0.8383	0.1047	0.1862	0.5519
Transformer	0.95	0.75	0.09	0.1733	0.90
TabNet	0.9591	0.7839	0.1374	0.2338	0.8816

4.3 Class Imbalance Considerations

Churn rate is ~4.57% (highly imbalanced). Consider:

Threshold tuning on predicted probabilities to raise recall.
Class weights or focal loss (for NN) to emphasize minority class.
Calibrating probabilities (Platt/Isotonic) for decisioning.
Using recall/PR AUC as primary optimization criteria for retention targeting.

5. Team Members and Task Distribution

https://docs.google.com/document/d/1xkK_FKSQtzT8boLcsmdZTgzCmYhi8UEgKCqQ7d2wq28/edit?usp=sharing