Predicting Customer Churn in Nigeria's Telecom Industry: A Machine Learning Approach with MTN Data

By Team Ben Okri

Problem Statement

Customer churn–the loss of clients to competitors–remains one of the most pressing challenges in Nigeria's telecom industry [1]. With increased competition among providers (MTN, Airtel, Glo, 9mobile), retaining existing customers is often more cost-effective than acquiring new ones. MTN Nigeria, the largest telecom operator in the country, faces high churn rates due to factors such as poor network quality, relocation, pricing, and customer dissatisfaction.

This project seeks to build a machine learning model to predict customer churn based on demographic, behavioural, and service usage data. Accurately identifying customers at risk of leaving will help telecom operators implement targeted retention strategies, improve service quality, and optimize marketing budget.

Existing Solutions

Globally, customer churn prediction is widely applied in telecom analytics. International firms employ ML-based solutions for retention modeling [2]. In Nigeria, however, churn management strategies are often reactive rather than predictive, relying on customer complaints or feedback instead of data-driven insights.

Our project differentiates itself by leveraging a recent Kaggle dataset specifically designed for MTN Nigeria, applying robust ML techniques, and developing an interactive deployment tool for telecom analysts and strategists.

Objectives

- Clean and preprocess the MTN churn dataset for accurate analysis
- Perform Exploratory Data Analysis (EDA) to uncover patterns in churn drivers
- Build classification models to predict customer churn features such as age, tenure, satisfaction, plan type, and usage behaviour
- Evaluate model performance using metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC
- Deploy the final model on a web interface (Streamlit) for interactive use by stakeholders

Proposed Dataset

We will use the MTN Nigeria Customer Churn & Usage Dataset (2025) from Kaggle [3].

Why This Dataset?

- Contains 974 customer records with rich features such as demographics, device usage, subscription plans, revenue patterns, and churn status.
- Captures real-world challenges like multiple churn reasons, imbalanced classes (churn vs active), and categorical/numeric feature combinations.
- Aligns directly with business problems in Nigerian telecom, making findings highly relevant.

Proposed Methodology

1. **Data Sourcing & Cleaning**: Handle missing values, duplicate entries, and inconsistent labels.
2. **Exploratory Data Analysis (EDA)**: Visualize churn distribution across demographics (age, gender, state) and other categories (device, plan, etc.). Analyze service usage patterns and their impact on churn. Visualize correlation between features.
3. **Featuring Engineering**: Encode categorical variables (e.g., state, device, subscription plan) using techniques like one-hot encoding. Create new features that might be indicative of churn, such as, "Revenue per Month" or "Usage Intensity (GB per ₦)". Scale numerical features as necessary.
4. **Model & Evaluation**: Train multiple classification algorithms: Logistic Regression, Support Vector Machine, Random Forest, XGBoost/LightGBM. Evaluate performance with Accuracy, Precision, Recall, F1-score, and ROC-AUC. Apply SMOTE or class weighting to handle imbalance. Perform hyperparameter tuning with cross-validation.
5. **Deployment**: Develop an interactive Streamlit web app that allows stakeholders to input customer data or view overall churn insights and predictions.

Modeling Plan

We will compare multiple models:

- **Logistic Regression**: for a simple, interpretable, baseline model
- **Support Vector Machine**: for high-dimensional classification
- **Random Forest**: for non-linear relationships and feature importance insights
- **Gradient Boosting (XGBoost/LightGBM)**: for high predictive accuracy and performance

Models will be compared using the suite of metrics mentioned above with primary focus on Precision and Recall, since false negatives (failing to predict churn) are especially costly in telecom.

Deployment Plan

The final model will be deployed as an interactive web application via Streamlit. The app will have two main functionalities:

- **Analytical Dashboard**: Displaying key insights and visualizations from our EDA
- **Prediction Interface**: Allowing users to input customer features to receive an instant churn risk prediction (e.g., "High Risk", "Low Risk")

The app will be hosted on Streamlit Community Cloud.

Expected Outcome

- A predictive ML model that identifies at-risk MTN customers
- Interactive web tool for analysts, managers, and strategists
- Data-driven insights into key churn drivers

Community Impact

By providing a framework for predicting churn, this project can help telecom companies reduce customer acquisition costs, improve loyalty, and ultimately offer better, more personalized services to their users. The methodology developed can also be applied to other industries facing similar retention challenges.

Team Members

- Izundu Dan-Ekeh
- Daud Abolade
- Edeh Ndubuisi Michael
- Ekaette Samuel
- Akanbi Abiodun Olawale

References

[1] Kavitha, V. & Kumar, G. & Kumar, S. & Harish, M. (2020). Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms. *International Journal of Engineering Research and Technology, 9*(5), 10.17577/IJERTV9IS050022

[2] Glory Sam, Philip Asuquo & Bliss Stephen (2024). Customer Churn Prediction using Machine Learning Models. *Journal of Engineering Research and Reports, 26*(2), 181–193. https://doi.org/10.9734/jerr/2024/v26i21081

[3] Kaggle MTN Nigeria Churn Dataset - Link

[4] MTN Nigeria Official Website - Link