TELECOM CUSTOMER CHURN PREDICTION: A MACHINE LEARNING APPROACH TO CUSTOMER RETENTION FOR THE TELECOM INDUSTRY IN GHANA

**Abstract**

This paper presents a machine learning-based approach to predicting customer churn in the Ghanaian telecom sector. Using a labeled dataset covering major operators – MTN, Vodafone, AirtelTigo, and Glo – several classification models were trained and evaluated. The Random Forest model demonstrated the highest predictive performance with an F1-score of 0.829. The study analyzes churn's impact on individual businesses and the industry as a whole and recommends actionable steps to mitigate churn. A dashboard and model deployment plan are also proposed for real-time monitoring and prediction.

# 1 Introduction

Customer churn is a critical issue in the telecom industry, where intense competition and low switching costs contribute to customer attrition. Predictive analytics offers telecom companies a proactive way to identify and retain at-risk customers.

# 2 Dataset and Methodology

The dataset included customer demographic, behavioral, and usage variables. Key features included AgeGroup, Gender, Education, EmploymentStatus, TelecomCompany, ReasonForChoosing, DurationWithCompany, PlanType, MonthlyCharges, and derived churn indicators.

Four machine learning models were trained:

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost

The models were evaluated using precision, recall, F1-score, accuracy, macro and weighted averages.

# 3 Model Evaluation

## 3.1 Random Forest Performance (Best)

```
Accuracy: 0.8293
Precision: 0.8320
Recall: 0.8296
F1 Score: 0.8290
```

| Class | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Churned (0) | 0.8000 | 0.8743 | 0.8355 | 183 |
| Not Churned (1) | 0.8639 | 0.7849 | 0.8225 | 186 |
| **Macro Avg** | 0.8320 | 0.8296 | 0.8290 | 369 |
| **Weighted Avg** | 0.8322 | 0.8293 | 0.8290 | 369 |

Table 1: Random Forest Classification Report

# 4 Impact Analysis

## 4.1 Business-Level Impact

- **MTN**: Likely to lose high-value customers. Retention plans should target customers with high monthly charges.
- **Vodafone**: High churn due to pricing dissatisfaction. Introduce flexible plans and bundles.
- **AirtelTigo**: Churn driven by poor network coverage. Invest in infrastructure in underserved areas.
- **Glo**: Smaller market share, but vulnerable to even minimal churn. Improve customer service and digital engagement.

## 4.2 Industry-Wide Impact

Consistent churn across all players weakens the industry. It results in high customer acquisition costs and reduced brand loyalty. Churn insights can help the industry collaborate on standards for network quality and fair pricing.

# 5 Recommendations

- Develop predictive dashboards to track churn drivers.
- Personalize customer retention campaigns based on risk scores.
- Monitor and benchmark KPIs like churn rate, average revenue per user (ARPU), and satisfaction.
- Collaborate with regulators on fair customer data portability policies.

# 6 Risks and Limitations

- Bias in the dataset may skew predictions.
- Customer behavior may shift over time, requiring model retraining.
- Data privacy must be maintained in real-time deployments.

# 7 Deployment and Dashboard

A Streamlit-based application will be developed and deployed on Render. It will:

- Predict churn from uploaded customer data
- Display metrics and business insights per telecom
- Enable download/export of predictions

# 8 Conclusion

This study highlights the potential of machine learning to predict and mitigate customer churn in Ghana's telecom industry. The Random Forest model achieved the highest performance, and business-specific insights provide a roadmap for targeted retention strategies.

# References


- Brownlee, J. (2020). *Machine Learning Mastery With Python*. Machine Learning Mastery.
- Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- Ghana Chamber of Telecommunications. (2023). *Annual Telecom Industry Report*.
- Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.