Telco Customer Churn Prediction

Model Card

# Model Overview

This machine learning solution predicts customer churn for a telecommunications company, enabling proactive retention efforts. Multiple classifiers were evaluated, with the **k-Nearest Neighbors (KNN)** model selected for deployment due to its superior recall and balanced prediction errors, aligned with business priorities.

# Intended Use

* **Primary Users:** Marketing analysts, CRM teams, call center agents
* **Purpose:** Accurately identify customers likely to churn, allowing targeted retention offers and improving revenue retention
* **Scope:** Customer churn risk prediction only; not for credit decisions or automated contract actions

# Data Summary

* **Source:** IBM Telco Customer Churn dataset (7,043 records)
* **Features:** Customer demographics, contract info, services, billing details
* **Target Variable:** Churn (1 = left, 0 = stayed)
* **Class Imbalance:** Approximately 26% churn rate
* **Preprocessing:** Imputed missing values, standard scaling for numerics, one-hot encoding for categorical variables

# Model Architecture & Selection

* Evaluated algorithms include Logistic Regression, Decision Tree, Random Forest, Support Vector Classifier, Gradient Boosting, and KNN
* **KNN was chosen due to:**
  + Highest recall score ensuring most churners are detected
  + Balanced false positive and false negative rates, minimizing wasted retention incentives and overlooked churners
  + Competitive performance supporting efficient resource allocation

# Performance Metrics

* **Key Metrics:** Recall, ROC-AUC, F-beta (β=1.2), and confusion matrices
* **Validation:** Stratified 80/20 train-test split with cross-validation during tuning
* **Model Efficiency:** Training and prediction times recorded to support deployment decisions
* **Results:**
  + **Models results:**

Logistic Regression - Time: 0.1067s, ROC AUC: 0.8419, Recall: 0.5588

Decision Tree - Time: 0.0338s, ROC AUC: 0.6573, Recall: 0.5053

KNN - Time: 0.0000s, ROC AUC: 0.7899, Recall: 0.5749

Random Forest - Time: 0.8208s, ROC AUC: 0.8164, Recall: 0.4759

SVC - Time: 6.6919s, ROC AUC: 0.7905, Recall: 0.4840

Gradient Boosting - Time: 0.9787s, ROC AUC: 0.8432, Recall: 0.5241

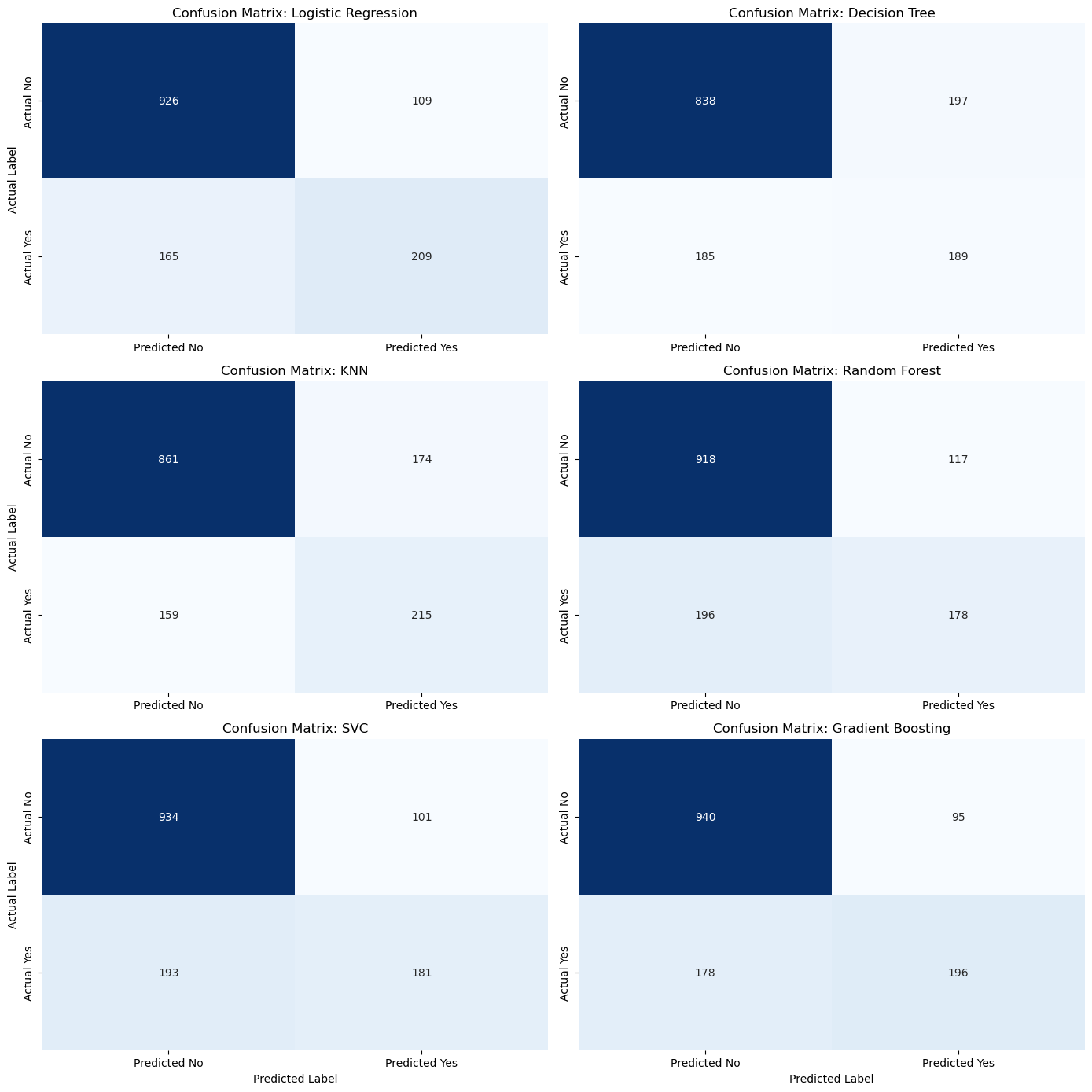
Best Recall Model (Before Tuning): KNN with Recall: 0.5749

Best ROC AUC Model (Before Tuning): Gradient Boosting with ROC AUC: 0.8432

Best params for KNN: {'n\_neighbors': 11, 'weights': 'uniform'} | Tuning Time: 5.8820s

Best params for Gradient Boosting: {'learning\_rate': 0.1, 'max\_depth': 7, 'n\_estimators': 50} | Tuning Time: 262.8054s

* + **Model Default Confusion Matrix results:**

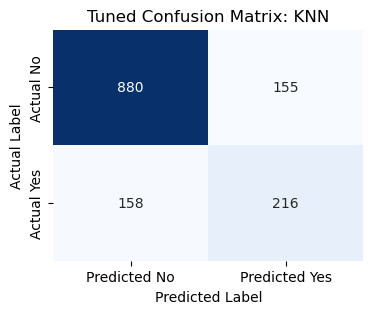


Best balanced Pre-tuned model between Recall and False Negatives is KNN.

* + **Retained Models Tuned Confusion Matrix results:**

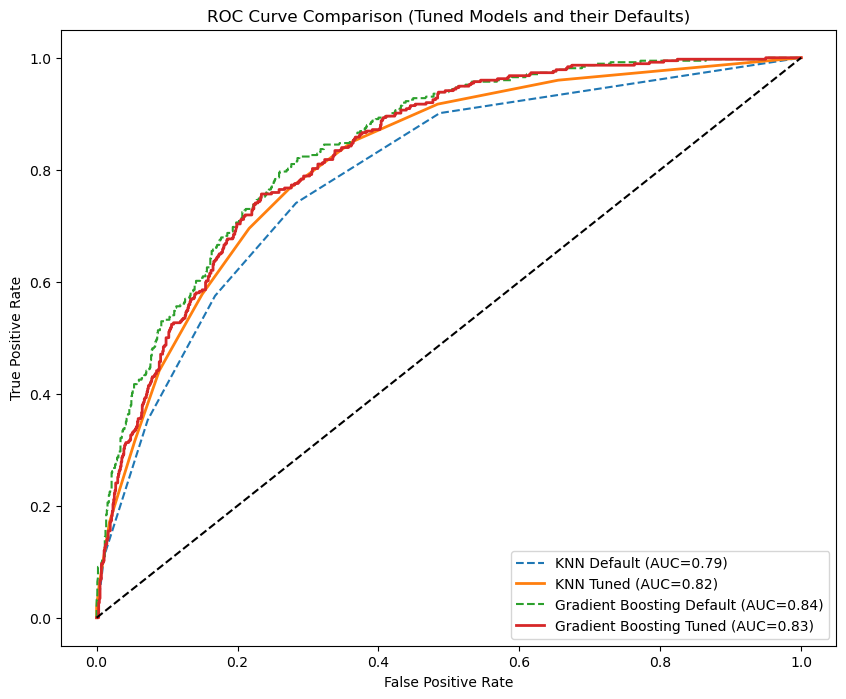
Tuned KNN ROC AUC: 0.8166

Tuned KNN Recall: 0.5775



* + **Retained Model AUC ROC Curve:**

I compared the two best recall model retained (KNN) to the best ROC AUC model (Gradient Boosting) to identify the tuning enhancements.

****

Recall Comparison:

Best Recall Model (Default) KNN: 0.5749

Best Recall Model (Tuned) KNN: 0.5775

Best AUC Model (Default) Gradient Boosting: 0.5241

Best AUC Model (Tuned) Gradient Boosting: 0.5267

# Business Impact

* **Customer Retention:** KNN’s high recall reduces missed churn cases, ensuring retention efforts reach the right customers
* **Cost Optimization:** Balanced errors help avoid unnecessary discounts to customers unlikely to churn, reducing operational costs
* **Resource Allocation:** Enables focused marketing and support efforts, improving campaign ROI and customer satisfaction
* **Competitive Edge:** Proactive churn mitigation strengthens brand reputation and market position
* **Financial Stability:** Retaining customers supports reliable revenue streams and long-term growth

# Ethical Considerations

* Regular fairness audits recommended identifying and addressing any bias across customer subgroups
* Transparent reporting of predictions and probabilities allows informed human decision-making
* Compliance with GDPR, CCPA, and other privacy regulations enforced

# Limitations

* Dataset is a public sample; model performance may vary with real-world data
* External factors like market competition and seasonality are not modeled
* Class imbalance challenges remain, with focus on recall to mitigate risk of missed churn

# Deployment & Maintenance

* Designed for integration within CRM and customer retention workflows
* Periodic retraining advised to adapt to population changes and maintain performance
* Logging and monitoring planned to track drift and alert when intervention is needed

# Contact & Support

* For inquiries, incidents, or retraining requests, contact: **sabri.bensaber@gmail.com**