Customer Churn Prediction & Explainability Report

Project Title: Understanding Customer Churn Through AI (2024)

Candidate Name: Kamalu Chioma

Tools Used: Python, SHAP, Gradient Boosting, XGBoost, Logistic Regression, Random Forest

1. Business Problem

Customer churn is a critical issue for subscription-based businesses such as telecom, banking, and SaaS platforms. Losing customers means loss of revenue and increased acquisition costs.

 ◆ **Objective:**
This project aims to **predict customer churn** and **explain the reasons behind churn using SHAP (SHapley Additive Explanations)**.

 ◆ **Impact:**
Understanding churn drivers helps businesses:
✓ Reduce churn by targeting high-risk customers
✓ Improve customer satisfaction through better service offerings
✓ Optimize marketing campaigns to retain users

2. Dataset Overview

Dataset: Telco Customer Churn (Kaggle)
Size: 7043 customers
Columns Used: 10 key features selected using feature importance analysis.

Key Features:

- **Tenure**: How long a customer has been with the company

- **Contract Type**: Month-to-month, one year, or two years

- **Monthly & Total Charges**: Cost factors influencing churn

- **Online Security, Tech Support, Online Backup**: Service-related features

- **Internet Service Type**: DSL, Fiber-optic, or No service

3. Models Tested

Model	Accuracy
XGBoost	77.39%
RandomForest	77.04%
Gradient Boosting	**79.17% (Best Model)**
Logistic Regression	78.39%

Best Model Selected: Gradient Boosting (79.17% Accuracy)
✓ **Balanced performance** between precision and recall
✓ **Handles categorical & numerical features well**

4. Explainability Analysis (SHAP)

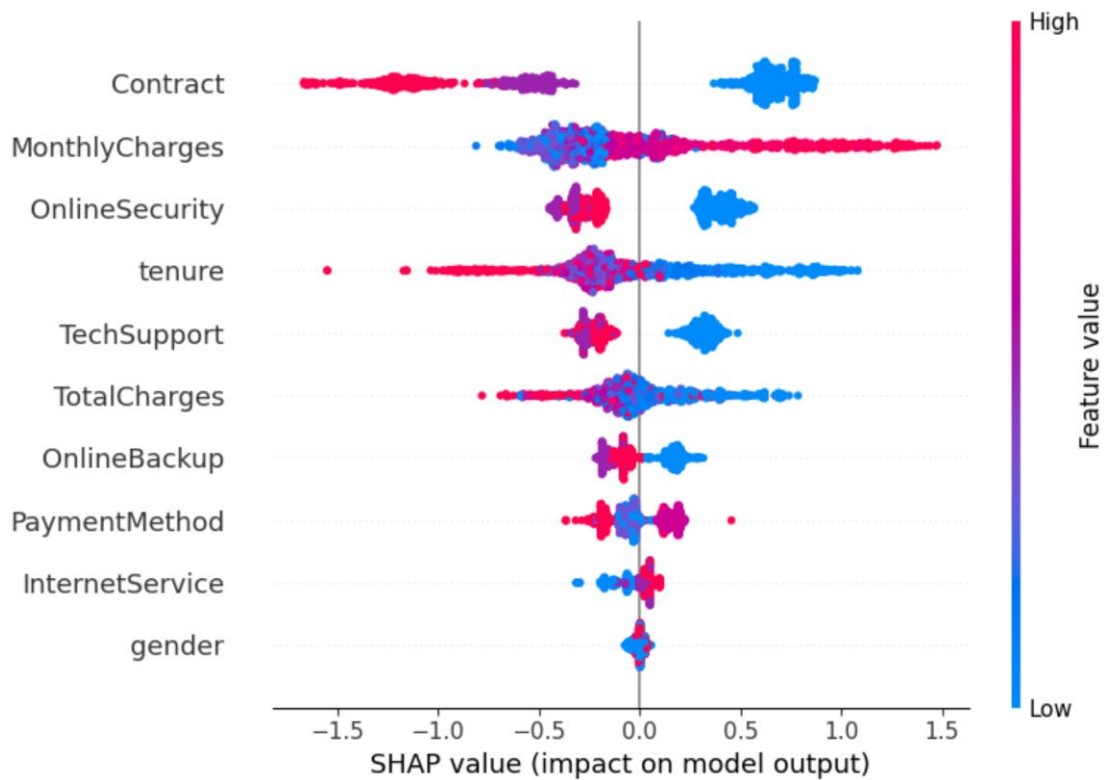SHAP helps us **understand why a customer is predicted to churn.**

◆ **SHAP Summary Plot: Feature Importance**

✓ **Contract Type & Tenure** are the biggest churn drivers
✓ Customers with **higher Monthly Charges** have higher churn risk
✓ Lack of **Online Security & Tech Support** increases churn probability
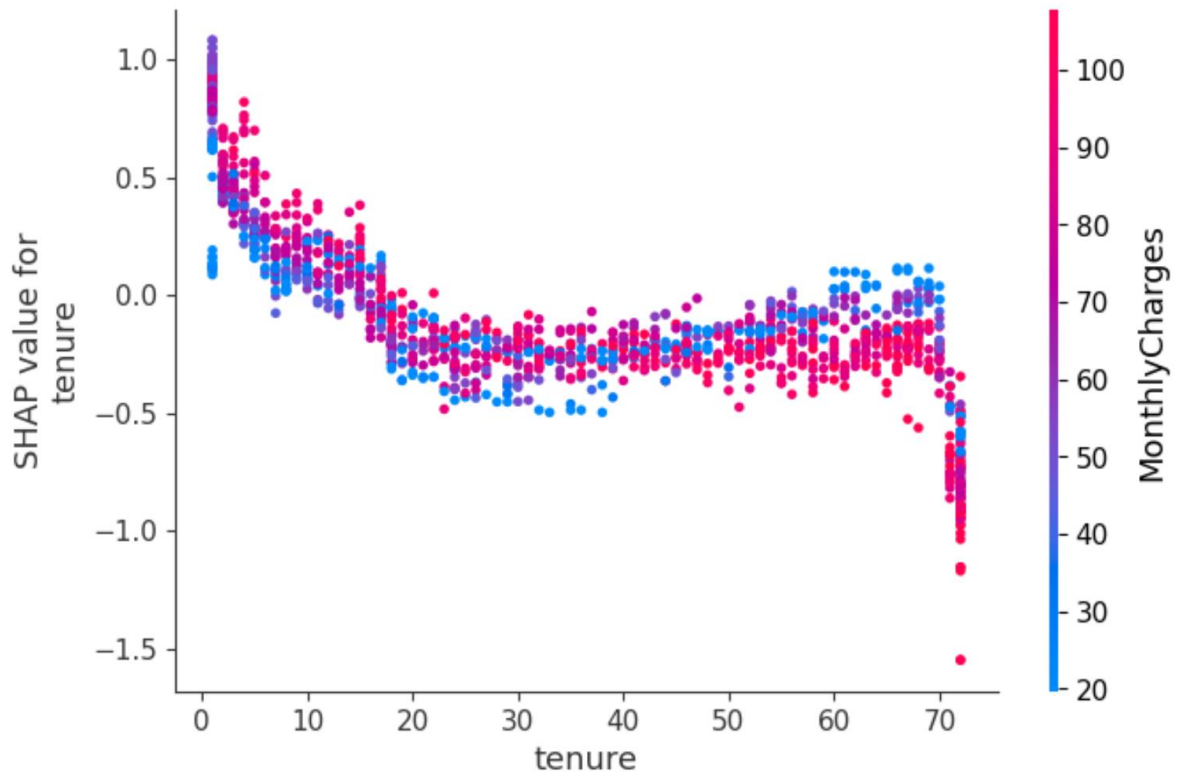
Visual Insight:
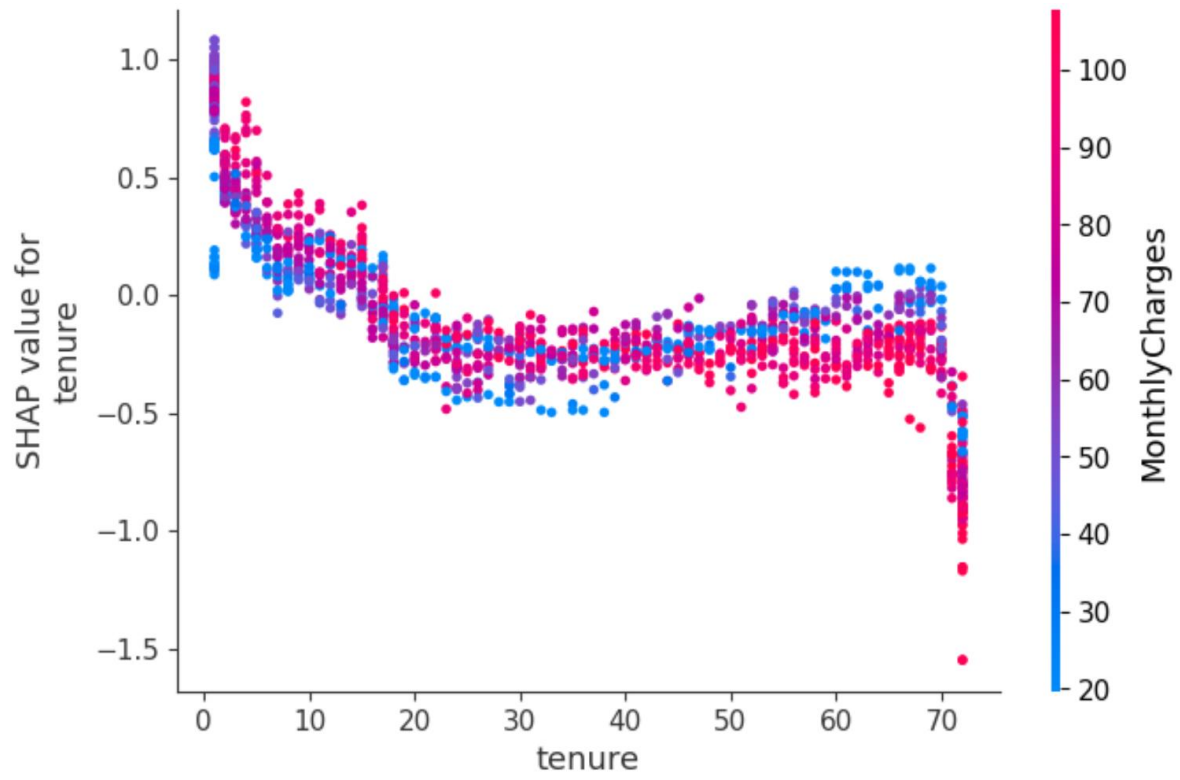
◆ **SHAP Dependence Plot: Tenure vs Churn**

✓ Short-tenure customers (new users) are at the highest risk of churn

✓ Long-term customers tend to stay

✓ High-churn customers often have **higher Monthly Charges**
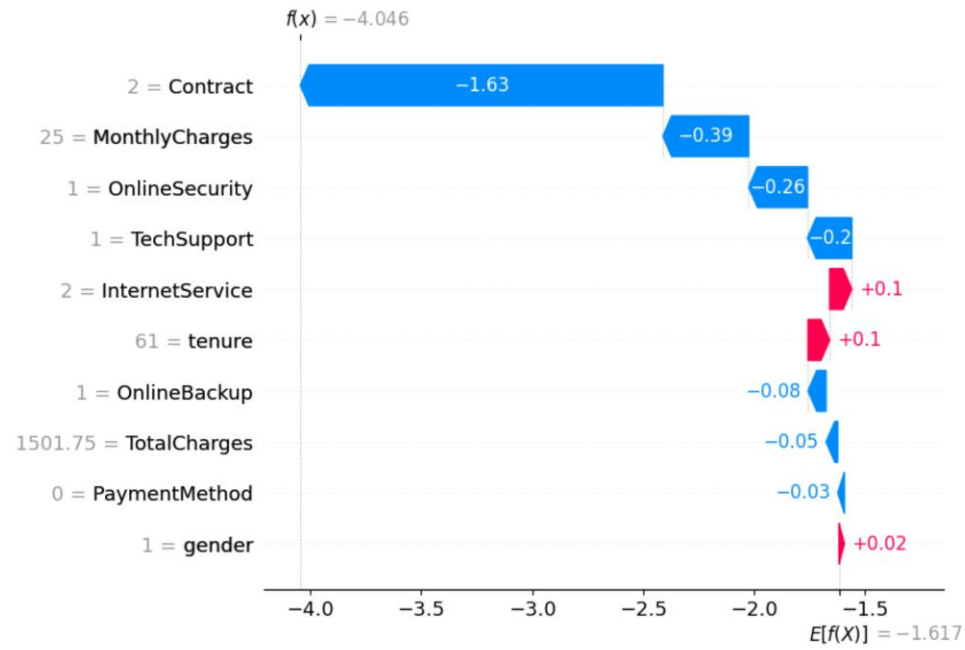
Visual Insight:

◆ **SHAP Waterfall Plot: Individual Customer Prediction**

✓ This plot explains why a **specific customer churned**
✓ Key contributing factors:

- High **Monthly Charges**

- **Month-to-month contract**

- Lack of **Online Security**

Visual Insight:

$f(x) = -4.046$

2 = Contract	−1.63
25 = MonthlyCharges	−0.39
1 = OnlineSecurity	−0.26
1 = TechSupport	−0.2
2 = InternetService	+0.1
61 = tenure	+0.1
1 = OnlineBackup	−0.08
1501.75 = TotalCharges	−0.05
0 = PaymentMethod	−0.03
1 = gender	+0.02

$E[f(X)] = -1.617$

5. Business Recommendations

1. Reduce churn by offering longer contracts

- Customers with month-to-month contracts churn the most.

- **Solution:** Offer discounts for yearly subscriptions.

2. Target new customers early

- Customers with **low tenure (<6 months)** churn the most.

- **Solution:** Provide better onboarding & engagement strategies.

3. Optimize pricing for high-risk customers

- Customers with **high monthly charges** have a higher churn rate.

- **Solution:** Offer personalized discounts or tiered pricing.

4. Improve customer support & security features

- Lack of **Tech Support & Online Security** is a strong churn predictor.

- **Solution:** Improve service offerings for these customers.

6. Final Takeaways

✓ **Gradient Boosting achieved the best accuracy (79.17%)**
✓ **SHAP provided clear explanations of churn drivers**
✓ **Business strategies can now be data-driven for customer retention**