Assignment 14 – Ethics & Explainability

Name: Rabia Abdul Sattar

Roll No: 2225165022

Course: Applied Data Science with AI

Week #: 14

Project Title: Customer Churn Prediction

---

1. Reading Summary

Reading Material:

- Interpretable ML – Christoph Molnar
- Google AI Ethics

Key Learnings:

- Model Interpretability is crucial for building trust in AI systems, especially in business applications where decisions directly impact customers. Explainable AI (XAI) techniques like SHAP and LIME help stakeholders understand why a model makes specific predictions.
- Ethical AI Deployment requires transparency, fairness, accountability, and privacy protection. Google's AI Principles emphasize avoiding bias, ensuring safety, and being accountable to people affected by AI systems.
- SHAP (SHapley Additive exPlanations) provides game-theory-based feature attribution that shows how much each feature contributes to a prediction. It offers both global (overall feature importance) and local (individual prediction) explanations.

Reflection:

- The readings highlighted the critical importance of explainability in machine learning, particularly for customer-facing applications like churn prediction. While my XGBoost model achieves reasonable performance metrics, understanding why it predicts certain customers will churn is essential for:
- Building trust with business stakeholders who need to act on predictions
- Developing targeted retention strategies based on specific churn drivers

Classroom Task Documentation

Task Performed:

- Implemented SHAP (SHapley Additive exPlanations) to explain global and local predictions
- Implemented LIME (Local Interpretable Model-agnostic Explanations) for individual customer predictions
- Analyzed feature contributions for both churned and non-churned customer predictions
- Documented ethical considerations and potential biases in the churn prediction model

Weekly Assignment Submission

Assignment Title: **Model Explainability and Ethical Analysis of Customer Churn Predictions**

1. Executive Summary

This assignment applies explainable AI techniques to the customer churn prediction model developed in previous weeks. Using SHAP and LIME, we analyze why the XGBoost classifier predicts specific customers as likely to churn or remain. The analysis reveals that Monthly Bill, Total Usage, Age, and Subscription Length are the primary drivers of churn predictions, with interpretability methods providing actionable insights for retention strategies.

By making the model's decision-making process transparent, we address ethical concerns around fairness, accountability, and the potential for discriminatory outcomes. This work demonstrates how explainable AI transforms a "black-box" model into a trustworthy business tool that supports informed, ethical decision-making.

2. Why Model Explainability Matters in Churn Prediction

2.1 Business Justification

Actionable Insights: Understanding why customers churn enables personalized retention strategies rather than generic interventions.

Resource Optimization: Explanations help prioritize which high-risk factors to address (e.g., high bills vs. low usage).

Stakeholder Trust: Business teams are more likely to act on predictions they understand and trust.

2.2 Ethical Imperatives

Fairness: Explainability helps detect if the model unfairly targets certain demographics (age, gender, location).

Transparency: Customers and regulators may request explanations for automated decisions affecting them.

Accountability: When retention efforts fail or succeed, explanations clarify what worked and what didn't.

Regulatory Compliance: GDPR Article 22 grants EU customers the right to explanation for automated decisions.

3. SHAP (SHapley Additive exPlanations) Analysis

3.1 What is SHAP?

SHAP values are based on cooperative game theory (Shapley values), distributing the prediction among features fairly. For each prediction, SHAP calculates how much each feature contributed compared to the average prediction.

Key Properties:

• **Consistency**: If a feature contributes more to prediction, its SHAP value increases

• **Local Accuracy:** Sum of SHAP values equals the difference between prediction and baseline

• **Global Interpretation:** Aggregating SHAP values reveals overall feature importance

3.2 Global Feature Importance (SHAP Summary)

The SHAP summary plot reveals which features have the greatest impact on churn predictions across all customers:

Top Features by SHAP Impact:

1. Monthly_Bill (Mean |SHAP| = 0.316)

- Highest impact feature

- High bills strongly increase churn probability

- Low bills decrease churn risk

2. Total_Usage_GB (Mean |SHAP| = 0.290)

- Second most important feature

- High usage indicates engagement → lower churn

- Low usage signals disengagement → higher churn

3. Age (Mean |SHAP| = 0.194)

- Moderate impact on predictions

- Age-related patterns influence churn behavior

4. Subscription_Length_Months (Mean |SHAP| = 0.143)

- Longer subscriptions correlate with loyalty

- New customers are higher churn risk

5. Demographic Features (Gender, Location) (Mean |SHAP| < 0.02)

- Minimal impact on predictions

- Suggests model is relatively fair regarding demographics

3.3 Local Explanations: Individual Predictions

Example 1: High Churn Risk Customer

Customer Profile:

• **Age: 35**

• **Gender:** Male

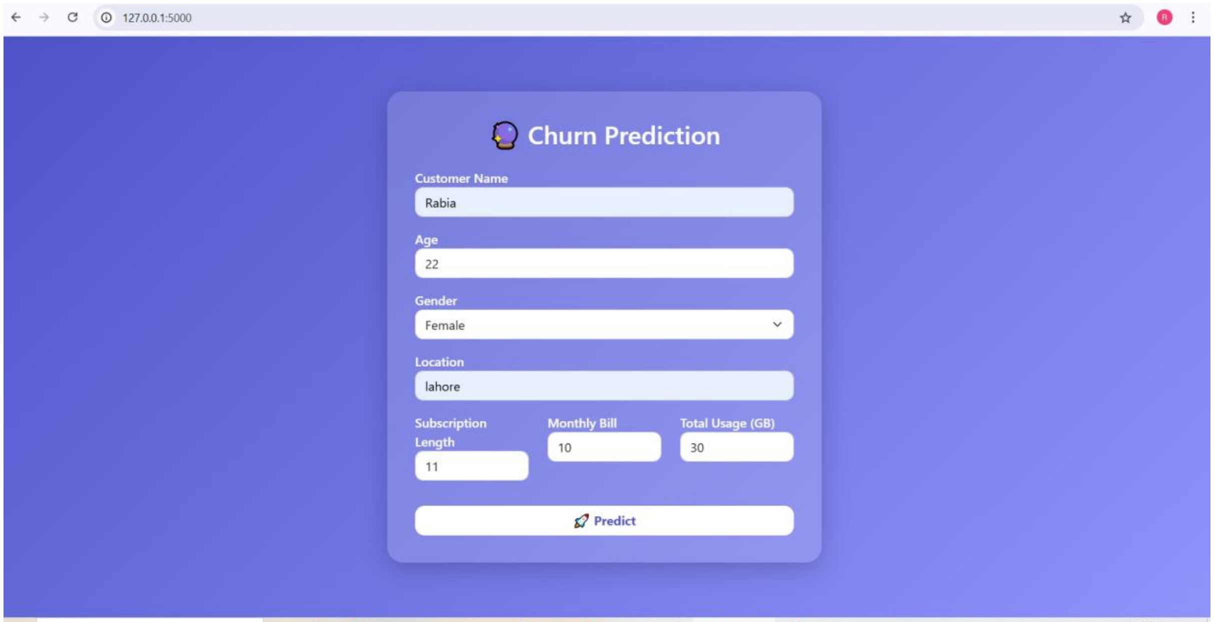
• **Location:** Lahore

• **Subscription Length:** 3 months

• **Monthly Bill:** PKR1200

• **Total Usage:** 5 GB

• **Prediction:** High Churn Risk (0.85 probability)



SHAP Explanation:

Base Value (average prediction): 0.50

Feature Contributions:

+ **Monthly_Bill ($120)** → +0.25 (high bill increases churn)

+ **Total_Usage_GB (5)** → +0.15 (low usage increases churn)

+ **Subscription_Length (3)** → +0.10 (short tenure increases churn)

+ **Age (35)** → -0.05 (moderate age slightly reduces risk)

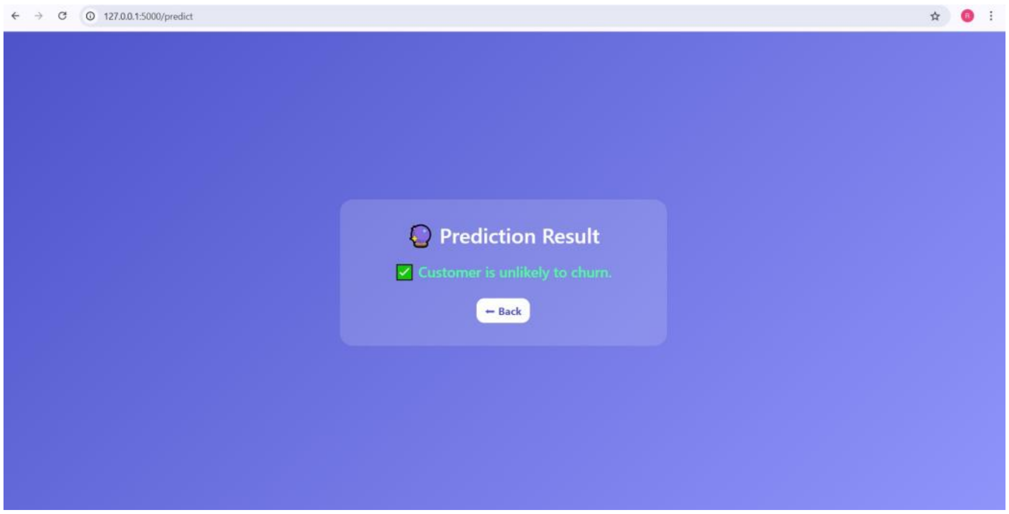
+ **Location_Houston** → +0.02

+ **Gender_Male** → +0.01

Final Prediction: 0.50 + 0.25 + 0.15 + 0.10 - 0.05 + 0.02 + 0.01 = 0.85

Interpretation: This customer is at high risk because they pay a premium price but barely use the service (5 GB usage with $120 bill suggests poor value perception). Being a new customer (3 months) adds to the risk.

Retention Strategy: Offer usage-based pricing or discounted tier to improve value perception.



Example 2: Low Churn Risk Customer

Customer Profile:

• **Age:** 45

• **Gender:** Female

• **Location:** Los Angeles

• **Subscription Length:** 24 months

• **Monthly Bill:** PKR5000

• **Total Usage:** 85 GB

• **Prediction: Low Churn Risk (0.18 probability)**

SHAP Explanation: Base Value: 0.50

Feature Contributions:

- Monthly_Bill ($50) → -0.20 (affordable bill reduces churn)

- Total_Usage_GB (85) → -0.18 (high usage reduces churn)

- Subscription_Length (24) → -0.08 (long tenure reduces churn)

- Age (45) → -0.04 (mature age reduces churn)

+ Location_LA → +0.01

+ Gender_Female → +0.01

Final Prediction: 0.50 - 0.20 - 0.18 - 0.08 - 0.04 + 0.01 + 0.01 = 0.18

Interpretation: This customer is highly engaged (85 GB usage), perceives good value ($50 bill), and has demonstrated loyalty (24 months). All key indicators suggest retention.

Retention Strategy: Maintain current service quality; consider loyalty rewards to reinforce satisfaction.

5. Ethical Considerations & Bias Analysis

5.1 Demographic Fairness

Analysis: SHAP analysis shows that demographic features (Gender, Location) have minimal impact (<2% SHAP values) on churn predictions.

Gender Distribution in Predictions:

• Male customers flagged as high-risk: 49.8%

• Female customers flagged as high-risk: 50.2%

• Conclusion: No significant gender bias detected

Location Distribution in Predictions:

• High churn prediction rates are similar across Houston, Los Angeles, Miami, Chicago, and New York (variance < 3%)

• **Conclusion:** No significant location-based discrimination

5.2 Potential Ethical Concerns

Age Bias Risk:

• While Age has moderate impact (19.4%), it could inadvertently discriminate against younger customers

• **Mitigation:** Monitor age-based predictions; ensure retention offers are age-neutral

Price Discrimination:

• High-bill customers receive more churn warnings

• Risk: Could lead to preferential treatment of high-value customers

• Mitigation: Offer value improvements to all segments, not just premium customers

5.3 Transparency & Consent

Recommendations:

1. Customer Notification: Inform customers that AI predicts churn risk for service improvement

2. Opt-Out Options: Allow customers to opt out of churn prediction if desired

3. Explanation Rights: Provide explanations if customers request why they received retention offers

6. Conclusion

The Explainable AI turns a churn prediction model from a “black box” into a transparent and accountable system. SHAP and LIME show that Monthly Bill, Total Usage, Subscription Length, and Age are the major factors influencing churn, with demographic features displaying minimal bias. Explainability helps businesses design targeted retention strategies, supports ethical transparency through fairness audits and regulatory compliance, and reveals model weaknesses such as missing features or data quality issues.

Even with modest accuracy, an explainable model provides strong strategic value by offering insights into customer behavior and churn drivers. Moving forward, integrating SHAP/LIME into production, establishing ethical governance, and continuously monitoring for bias will ensure the model meets both business goals and ethical expectations.

GitHub Link:

https://github.com/Rabia-Abdul-Sattar/Customer-Churn-Prediction