Customer Churn Prediction Project

Objective

This project predicts which telecom customers are likely to cancel services (churn) using machine learning. The goal is to **reduce revenue loss** by proactively retaining at-risk customers through targeted interventions (e.g., discounts, personalized offers).

Problems Addressed

1. **High Churn Rates**: Identified customers most likely to leave, focusing on key segments (e.g., short-tenure, month-to-month contracts).
2. **Business Blind Spots**: Revealed hidden patterns (e.g., customers with dependents and low monthly charges churn more).
3. **Actionable Insights**: Translated model outputs into business strategies.

Key Steps & Techniques

1. Data Preparation

- **Cleaned data**: Handled missing values (e.g., imputed TotalCharges).
- **Engineered features**: Created a High Risk flag for customers with:
    - Tenure <6 months
    - Month-to-month contracts
    - Dependents
    - Monthly charges <$150

2. Model Development

- **Compared algorithms**: Tested Logistic Regression, Random Forest, XGBoost, and **LightGBM** (best performance).
- **Optimized hyperparameters**: Used GridSearchCV to maximize F1 score (handled class imbalance).
- **Key Metrics**:

- **F1 Score (0.72)**: Balanced precision/recall for churn predictions.
- **ROC AUC (0.83)**: Strong class separation.

3. Explainability (SHAP)

- Identified top churn drivers:
  1. **Short tenure** (<6 months)
  2. **Month-to-month contracts**
  3. **Lack of tech support**
- Quantified feature impact (e.g., month-to-month contracts **3× higher** churn risk).

4. Business Impact

- **Targeted Retention**: High-risk customers flagged for personalized offers.
- **Reduced Costs**: Focused resources on the **20% most at-risk** customers.
- **Strategic Insights**: Recommended long-term contract incentives and improved tech support.

KPIs Tracked

KPI	Formula	Value
Churn Rate	(Churned Customers / Total Customers) × 100	26.5%
Avg. Tenure (Churned)	Mean Tenure Months for churned customers	17.8 months
High-Risk Churn Rate	Churn rate in engineered High Risk segment	48%

Conclusion

This project transformed raw customer data into **actionable retention strategies** by:

1. **Predicting churn** accurately (LightGBM model).
2. **Explaining why** customers churn (SHAP).
3. **Guiding business decisions** (e.g., contract redesign for high-risk segments).

Tools Used: Python, Pandas, Scikit-learn, LightGBM, SHAP.
Skills Demonstrated: Data cleaning, feature engineering, model tuning, explainable AI, business storytelling.