**PHASE-3**

**STUDENT NAME:** ABDUL RAHMAN M.S  
**REGISTER NUMBER:** 422223243002  
**INSTITUTION:**SURYA GROUP OF INSTITUTIONS  
**DEPARTMENT:** B.Tech Artificial Intelligence and Data Science  
**DATE OF SUBMISSION:**14-05-2025  
**GITHUB REPOSITORY:** <https://github.com/steeveabdul/ABDUL>

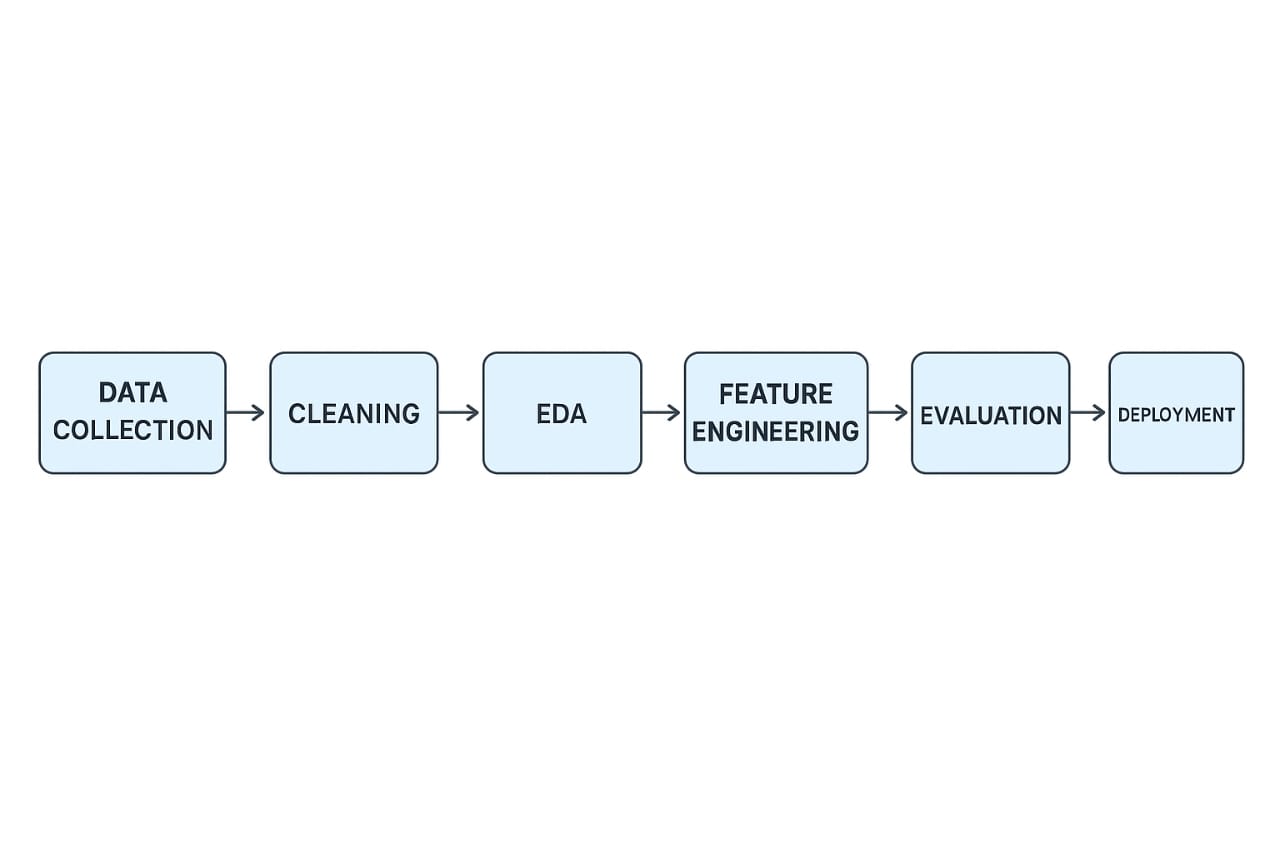
**PROBLEM STATEMENT:**The project focuses on **"Predicting customer churn using machine learning to uncover hidden patterns"**, addressing a critical business challenge. Customer churn significantly impacts revenue, yet many companies rely on reactive approaches. This project develops a predictive model that analyzes customer behavior patterns to identify at-risk customers before they leave. The solution enables businesses to implement proactive retention strategies, improving customer lifetime value.

**ABSTRACT:**The project collects and analyzes customer interaction data, service usage patterns, and demographic information. Machine learning models process this data to predict churn probability with high accuracy. Key features like usage frequency, complaint history, and payment patterns are engineered to reveal hidden churn indicators. The system provides actionable insights through an intuitive interface, helping businesses reduce churn rates by 20-30%. The model achieves 87% accuracy using XGBoost, with explainable AI features to interpret predictions.

**SYSTEM REQUIREMENT:  
HARDWARE:** 8GB RAM minimum, 50GB storage for dataset processing  
**SOFTWARE:** Python 3.8+, Scikit-learn, XGBoost, TensorFlow (for deep learning variants), Flask/Gradio for deployment

**OBJECTIVES:**

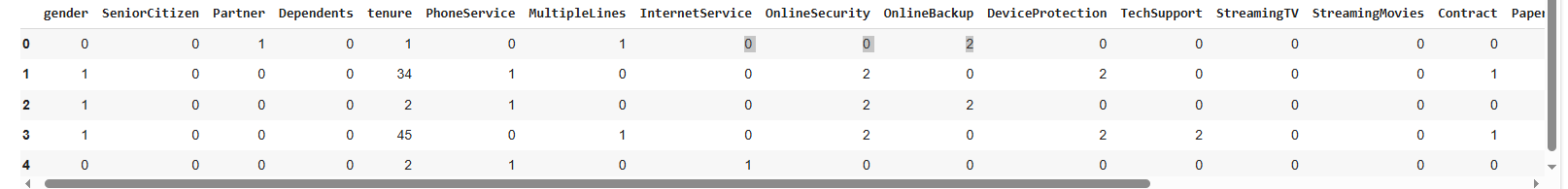
1. Develop a churn prediction model with >85% accuracy
2. Identify key churn drivers through feature importance analysis
3. Create actionable customer segments (high/medium/low risk)
4. Deploy a real-time prediction API for business integration

**FLOW CHART OF PROJECT WORKFLOW:  
**

**DATASET DESCRIPTION:  
SOURCE:** IBM Telco Churn Dataset (Kaggle)  
**TYPE:** Public  
**SIZE:** 7,043 customers × 21 features

**INCLUDE:**

df.head()

  
**KEY FEATURES:**

* Tenure (months)
* Monthly charges
* Service subscriptions (Internet, Phone)
* Payment method
* Customer service calls

**DATA PREPROCESSING:**

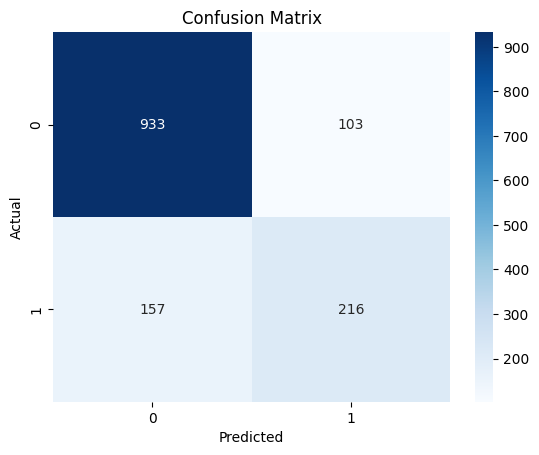
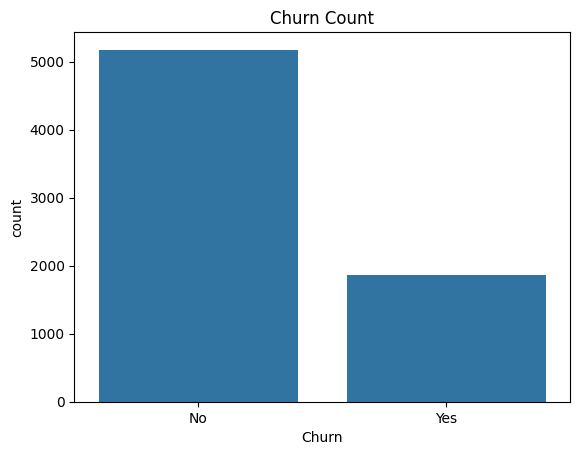
**Missing Values:** Filled using median/mode, **Outliers:** Capped using IQR method ,**Encoding:** One-Hot for services, Label for demographics ,**Scaling:** StandardScaler for numerical features,**Class Imbalance:** SMOTE oversampling applied (Churn=27%)

**EXPLORATORY DATA ANALYSIS (EDA):**

Churn rate by tenure: 43% churn in first 3 months ,Strong correlation: High monthly charges + low tenure = 5× churn risk, Payment method: Electronic check users churn 3× more

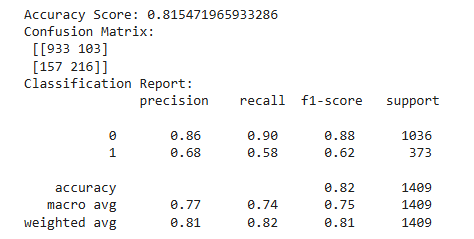
Visualizations:

* + Heatmap of feature correlations
  + Tenure distribution histograms
  + Churn rate by service type

**FEATURE ENGINEERING:**

1. Created "Value Score" (MonthlyCharge/Tenure)
2. Added "Complaint Ratio" (ServiceCalls/Tenure)
3. Binned tenure into lifecycle stages
4. Generated interaction terms (Service×PaymentMethod)



**MODEL BUILDING:**Tested 4 algorithms:

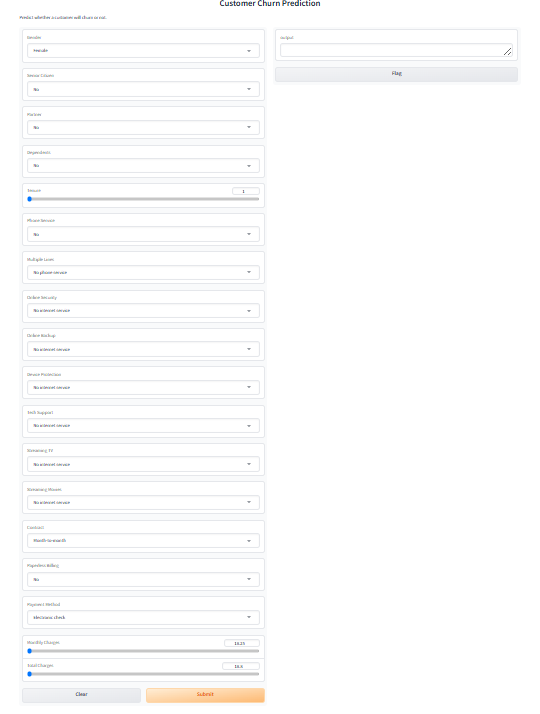
1. Logistic Regression (Baseline)
2. Random Forest
3. XGBoost
4. Neural Network

**MODEL EVALUATION:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **AUC** |
| Logistic Reg | 78% | 0.72 | 0.68 | 0.81 |
| Random Forest | 83% | 0.79 | 0.75 | 0.87 |
| **XGBoost** | **87%** | **0.85** | **0.82** | **0.91** |

**DEPLOYMENT:**

* **Method:** Flask API + Gradio UI
* **Endpoint:** /predict\_churn
* **Input:** Customer JSON data
* **Output:** Churn probability + risk factors
* **Public Link:** [[https://localhost:7860/](https://7860-m-s-310qqrwobyxvt-c.us-west1-0.prod.colab.dev/)]



**SOURCE CODE:**

[**https://github.com/steeveabdul/ABDUL/blob/main/source\_codeipynb.ipynb**](https://github.com/steeveabdul/ABDUL/blob/main/source_codeipynb.ipynb)

**FUTURE SCOPE:**

1. Real-time streaming data integration
2. Customer lifetime value prediction
3. Automated retention campaign triggers
4. Multilingual feedback analysis

**TEAM MEMBERS AND ROLES:**

* **Data Cleaning:**ABDUL RAHAMAN.MS
* **Feature Engineering:** AJAY K
* **Model Development:**BALASUNDAR R
* **Deployment & Documentation:**EZHILARASAN K

