Low Level Design (LLD)

NoMoreChurn – Telco Risk Intelligence

Date: 25/5/2025

Submitted By :-

Soumen Baidya

Gourisankar Maity

Disha Jaiswal

Contain

Introduction

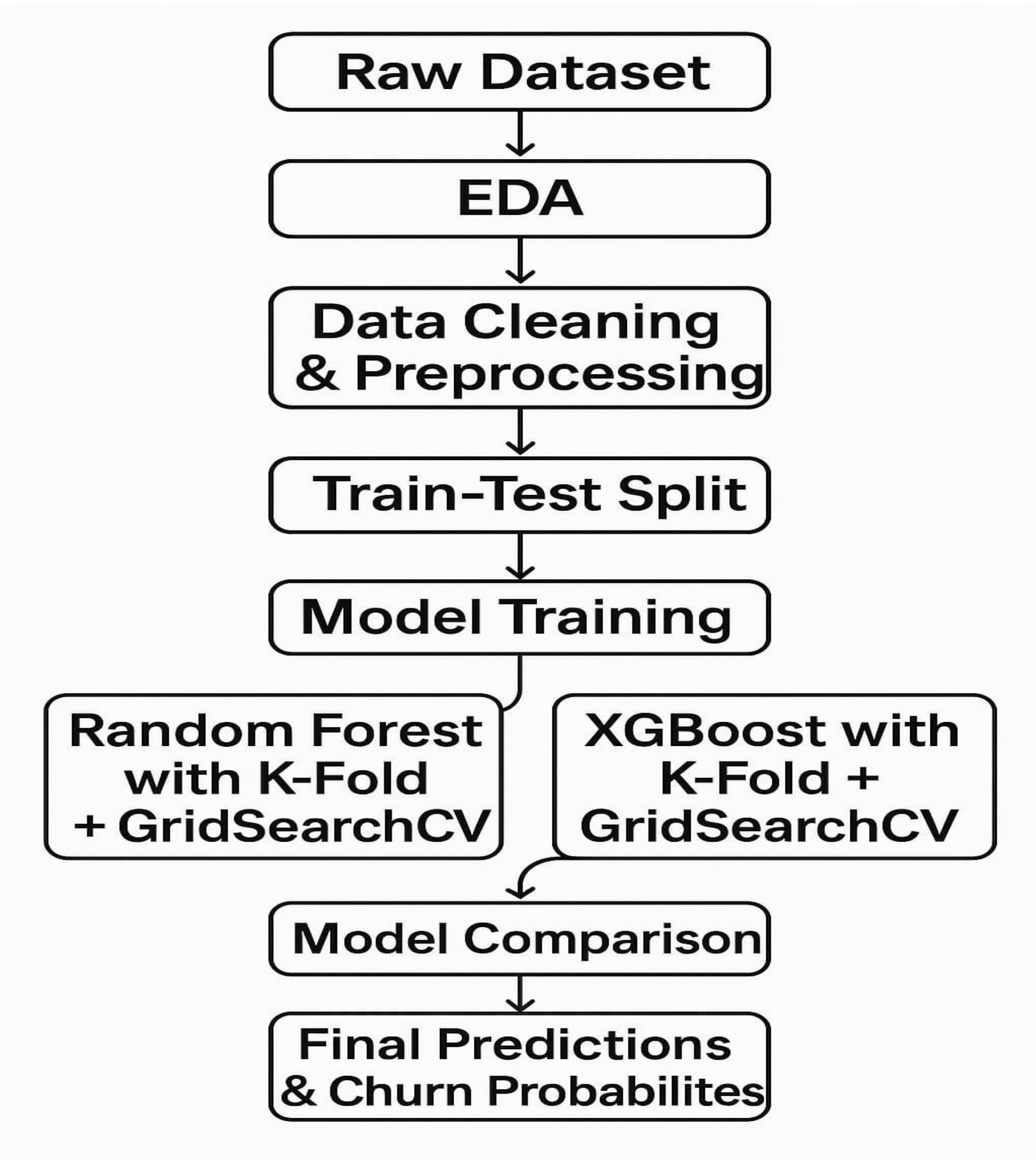
Why this Low-Level Design Document?

- Purpose of LLD: to outline architectural and functional overview for non-technical and semi-technical audiences.

Scope

- What the system will and won't cover.
- Only test dataset in Power BI.
- Focus on customer churn modelling, predictions, insights.

Architecture



Architecture Description

Data Description

Raw Dataset Columns:

Customer ID: Unique customer identifier

Gender Customer: Gender

Senior Citizen: Binary indicator

Partner, Dependents: Relationship attributes

tenure: Duration with the company

Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, StreamingTV, Streaming Movies: Service usage features

Contract, Paperless Billing, Payment Method: Subscription and payment info

Monthly Charges, Total Charges: Financial attributes

Num Admin Tickets, num Tech Tickets: Support interaction metrics

Actual Churn, Predicted Churn, churn probability: Final modelling outcomes

Data Transformation

1. Missing Value Handling: Convert Total Charges to numeric and fill missing with 0.

2. Feature Engineering: Binned tenure into ranges (e.g., 0–12, 13–24 months).Created binary flags for high-risk contract types.

3. Encoding: One-hot encoding for multiclass columns (Payment Method, Internet Service, Contract). Label Encoding for binary categories.

4. Scaling: Standard Scaler for Monthly Charges, Total Charges.

Model Building & Evaluation

Random Forest Classifier:

Used K-Fold Cross Validation (K=5)

Grid SearchCV parameters: n_estimators, max_depth, min_samples_split

Output: accuracy \~82%, precision \~0.79

XGBoost Classifier:

Used Grid SearchCV to tune learning_rate, n_estimators, max_depth

Performed better on AUC-ROC

Voting Classifier (Soft Voting):

Combined RF and XGBoost using probability-based aggregation

Final AUC-ROC: \~0.86

Evaluation Metrics:

Confusion matrix generated using sklearn

Precision, Recall calculated manually and confirmed via DAX in Power BI

Export Data from Model

Created data frame with the following columns:

Customer ID, tenure, Monthly Charges, Actual Churn, Predicted Churn, churn probability

Saved as CSV: final_powerbi_churn_data.csv

Deployment

- Final CSV exported from Python
- Power BI dashboard built on top of it
- Local Power BI file, future plan for Power BI Cloud or embedding into web app (optional mention)

Unit test Cases

Test Case ID	Component	Description	Input	Expected Output
TC001	Data Load	Check file import	CSV	Data Frame with correct columns
TC002	Feature Engineering	Check tenure binning	tenure = 15	"13–24"
TC003	Encoding	One-hot Payment Method	Payment Method	Dummies created
TC004	Random Forest	Model accuracy test	X _test	Accuracy > 80%
TC005	Voting Classifier	Churn probability range test	All rows	$0 \le prob \le 1$
TC006	Export	Output CSV format check	Final Data Frame	CSV with 6 key columns