Customer Churn Prediction using ML

Predicting telecom customer churn

Problem Definition

High customer churn = revenue loss

Retaining is cheaper than acquiring

Goal: Predict churn early

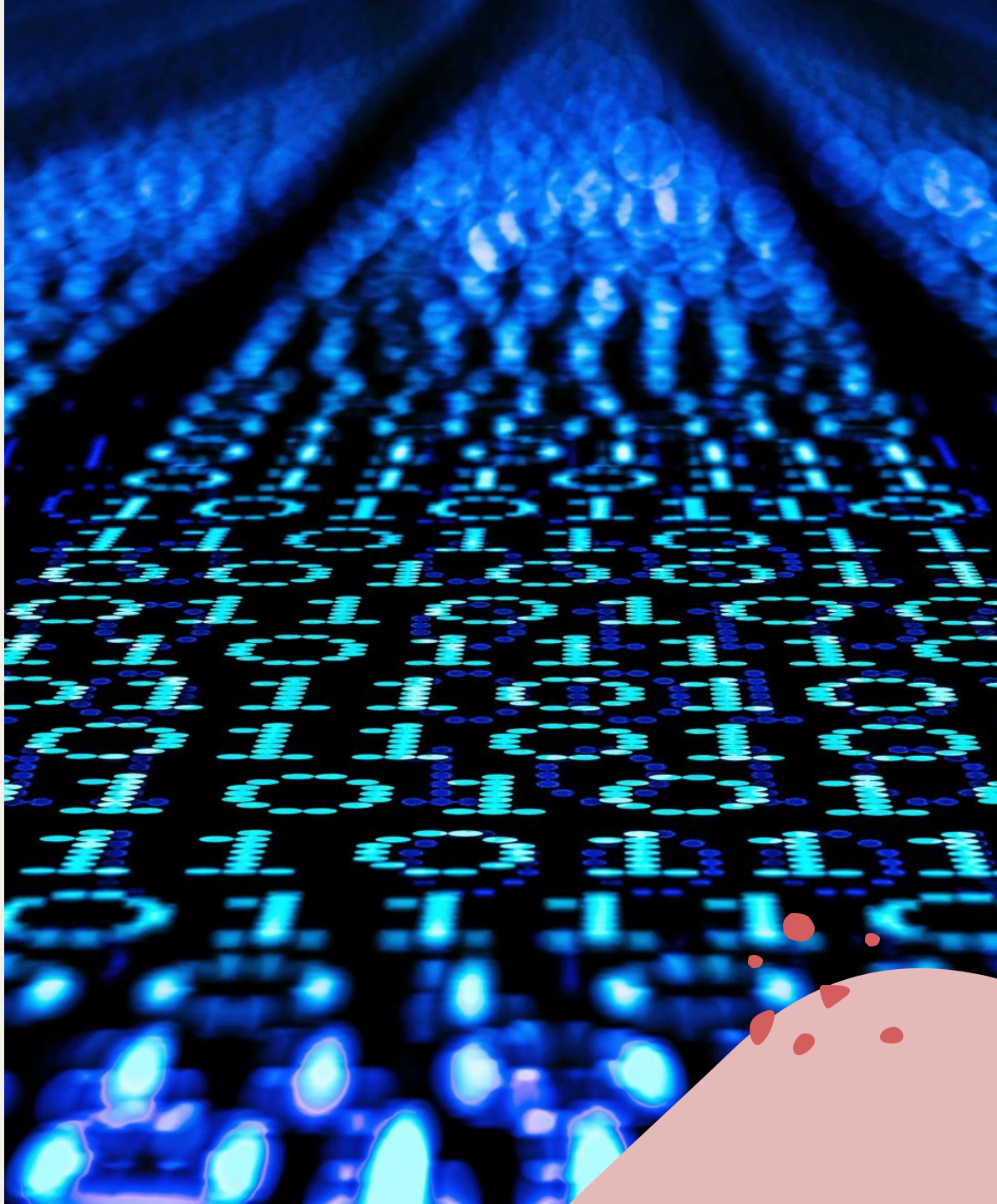
Helps improve customer retention

Dataset

- Source: Kaggle Telco Churn dataset
- ~7,000 customers, 21 features
- Features: demographics, services, account info
- Target: **Churn (Yes / No)**

Data Cleaning

- Fixed missing values (*TotalCharges*)
- Encoded categorical variables
- Standardized numerical features
- Dropped irrelevant IDs



Data Insights

~26% customers churned

Month-to-Month contracts → highest churn

Higher monthly charges → higher churn risk

Long-term customers churn less

Model Building

- Models used:
  - Logistic Regression
  - Random Forest
  - Gradient Boosting
- Evaluation: Accuracy & ROC-AUC

Model Performance

- Logistic Regression → 78%
- Random Forest → 85%
- Gradient Boosting → 88%
- Best Model: Gradient Boosting

Key Insights

01

Month-to-Month contract = churn risk

02

High monthly charges drive churn

03

Electronic check users churn more

04

Tech support reduces churn

Conclusion

- Achieved **~90% churn prediction accuracy**
- Gradient Boosting best performer
- Clear business strategies identified
- Future scope: real-time deployment & deep learning