Telecom Revenue Optimization Model

Complete Project Walkthrough

Project Overview

This is an end-to-end data science project focused on optimizing revenue for telecom companies through advanced analytics and machine learning. The project addresses key business challenges in the telecom industry:

- Increasing Average Revenue Per User (ARPU)
- Reducing customer churn
- Identifying cross-sell/up-sell opportunities
- Improving campaign effectiveness

The project is designed for Dentsu's strategy focusing on algorithmic media, identity solutions, and privacy-first analytics.

Project Structure

```
Telecom Revenue Optimization Model/
■■■ data/
■ ■■■ raw/ # Generated synthetic datasets
■ ■■■ processed/ # Cleaned data for modeling
■ ■■■ models/ # Saved ML models
■■■ notebooks/
■ ■■■ 01_data_generation.ipynb # Synthetic data creation
■ ■■■ 02_data_preprocessing.ipynb # Data cleaning and feature engineering
■ ■■■ 03_exploratory_analysis.ipynb # EDA and visualizations
■ ■■■ 04_churn_prediction.ipynb # Churn prediction modeling
■ ■■■ 05_arpu_forecasting.ipynb # Time-series ARPU prediction
■ ■■■ 06_model_explainability.ipynb # SHAP analysis
■■■ src/
■ ■■■ data_generation.py # Data creation utilities
■ ■■■ preprocessing.py # Data cleaning functions
■ ■■■ models.py # ML model implementations
■ ■■■ evaluation.py # Model evaluation metrics
■■■ dashboard/
■ ■■■ app.py # Streamlit dashboard application
■■■ reports/
■ ■■■ business_report.md # Executive business report
■■■ requirements.txt # Python dependencies
■■■ README.md # Project documentation
```

Step-by-Step Project Execution

1. Data Generation

Key Component: src/data_generation.py

- Generates a synthetic dataset of 10,000 telecom customers with realistic patterns

- Ensures privacy compliance using hashed customer IDs
- Creates 7 interconnected datasets:
  - Customer demographics (age, gender, location, income, tenure)
  - Usage patterns (data consumption, voice minutes, OTT streaming)
  - Billing information (plan types, ARPU, payment history)
  - CRM data (satisfaction scores, complaints, support interactions)
  - Campaign data (exposures, impressions, conversions)
  - Digital engagement (web/app sessions, self-service usage)
  - Churn labels (based on multiple factors)

Key Features:

- Privacy-first approach with synthetic data
- Realistic customer behaviors and patterns
- Hashed customer identifiers for anonymity
- GDPR-compliant data handling

2. Data Preprocessing & Feature Engineering

Key Component: notebooks/02_data_preprocessing.ipynb

- Cleans and prepares data for modeling
- Creates advanced features like:
  - RFM scores (Recency, Frequency, Monetary value)
  - Usage efficiency metrics
  - Risk indicators for churn prediction
  - Customer lifetime value scores

3. Exploratory Data Analysis

Key Component: notebooks/03_exploratory_analysis.ipynb

- Analyzes customer segments and behaviors
- Identifies key business metrics:
  - Total monthly revenue: $1.88M
  - Overall churn rate: 30.31%
  - Average ARPU: $188.29
  - Revenue at risk from churn: $563K monthly

4. Machine Learning Models

A. Churn Prediction Model

Key Component: notebooks/04_churn_prediction.ipynb and src/models.py

- Uses LightGBM gradient boosting algorithm
- Achieves 84.7% AUC score
- Identifies top 20% at-risk customers (2,000 customers)
- Key features influencing churn:
  - Satisfaction score (strongest predictor)
  - Number of complaints
  - Late payments
  - Tenure (negative correlation)
  - Usage efficiency

Business Impact: Potential to save $140K monthly revenue through targeted retention efforts

B. ARPU Forecasting Model

Key Component: notebooks/05_arpu_forecasting.ipynb and src/models.py

- Uses Facebook Prophet for time-series forecasting
- Performance metrics:
  - RMSE: $23.45
  - MAE: $18.20
  - MAPE: 12.3%

Business Impact: Projects 5% ARPU increase, adding $94K in additional monthly revenue

C. Uplift Modeling

Key Component: src/models.py

- Two-model approach for cross-sell/up-sell targeting
- Identifies customers most likely to respond to campaigns
- Expected lift: 15% improvement in campaign conversion rates
- ROI enhancement: 3.2x better than random targeting

5. Model Explainability

Key Component: notebooks/06_model_explainability.ipynb

- Uses SHAP (SHapley Additive exPlanations) values
- Provides transparent, interpretable model decisions
- Key benefits:
  - Regulatory compliance (GDPR, CCPA)
  - Stakeholder trust in AI decisions
  - Actionable business insights

6. Interactive Dashboard

Key Component: dashboard/app.py

- Built with Streamlit for interactive visualizations
- Six dashboard sections:
  1. Executive Overview
  2. Revenue Analysis
  3. Churn Analysis
  4. Campaign Effectiveness
  5. Customer Insights
  6. Predictive Insights

To run the dashboard:

```
python -m streamlit run dashboard/app.py
```

7. Business Report

Key Component: reports/business_report.md

- Comprehensive 8-page business report
- Projects $4.7M annual savings through churn prevention
- Shows 1,395% ROI in first year
- Provides strategic recommendations aligned with Dentsu's capabilities

Key Technologies Used

- Programming Language: Python
- Data Processing: pandas, NumPy
- Machine Learning: LightGBM, XGBoost, scikit-learn
- Time Series Analysis: Prophet
- Model Explainability: SHAP
- Visualization: Matplotlib, Seaborn, Plotly
- Dashboard: Streamlit
- Statistical Modeling: statsmodels

Business Impact & Results

1. Churn Reduction: 7% improvement in retention
2. Revenue Growth: 5% ARPU increase
3. Campaign Efficiency: 15% improvement in ROI
4. Annual Savings: $4.7M projected
5. ROI: 1,395% first-year return

Privacy & Compliance Features

- Synthetic data generation (no real customer information)
- Hashed customer identifiers for privacy protection
- GDPR-compliant data handling
- Privacy-by-design implementation

How to Run the Project

1. Install dependencies:

```
pip install -r requirements.txt
```

2. Generate synthetic data:

```
cd src
python data_generation.py
```

3. Run Jupyter notebooks in order (01-06)

4. Launch the dashboard:

```
python -m streamlit run dashboard/app.py
```

Interview Talking Points

Technical Depth

- Explain the choice of LightGBM for churn prediction (handles categorical features well, fast training)
- Discuss feature engineering approaches (RFM scores, risk indicators)
- Describe time-series forecasting with Prophet (seasonality, trend detection)

- Explain SHAP values and their business importance

Business Impact

- Quantify the financial benefits ($4.7M annual savings)
- Discuss customer segmentation strategies
- Explain how the models translate to actionable business insights
- Highlight the ROI projections (1,395%)

Problem-Solving Approach

- Emphasize the privacy-first approach to data handling
- Discuss how you balanced model performance with interpretability
- Explain your methodology for feature selection
- Describe how you validated model results

Challenges & Solutions

- Handling imbalanced datasets in churn prediction
- Dealing with seasonality in ARPU forecasting
- Ensuring model explainability for business stakeholders
- Creating realistic synthetic data that mimics real-world patterns

This project demonstrates enterprise-level data science capabilities, showcasing advanced machine learning, privacy-conscious analytics, and measurable business impact. It's perfect for demonstrating your skills in end-to-end data science projects during interviews.