

E-Commerce Customer Churn Prediction

Sravani Karnataka





Understanding Customer Churn in E-Commerce

What is churn? When customers stop buying from your business. It's a silent threat that can erode your customer base and profitability.

Why it hurts: High churn rates mean lost revenue, decreased customer lifetime value, and increased customer acquisition costs. It's a cycle that can hinder growth.

Who benefits: This solution empowers Marketing, Retention Teams, and Management to make informed decisions.

Our Goal: Proactively identify at-risk customers, allowing us to intervene and retain them before they churn.



Dataset Overview: UCI Online Retail II



Rich Dataset

Utilized the comprehensive **UCI Online Retail II dataset** for this project.



Extensive Records

Initially comprised **541,910 rows** of transaction-level purchase history.



UK Customer Base

Focused on customer transactions from an e-commerce business primarily operating across the **United Kingdom**.



Real-World Data

Presented the key challenge of working with **raw, messy data** requiring significant preparation.

Navigating Data Cleaning Challenges

1

Inconsistent IDs

Handling missing or malformed customer identifiers.

2

Cancelled Orders

Filtering out transactions marked with "C" invoices.

3

Negative Quantities

Addressing erroneous entries with negative purchase quantities.

4

Outlier Detection

Identifying and managing extreme values that could skew results.

5

Date Formatting

Standardizing various date formats for consistent analysis.

Clean Data Result: After rigorous cleaning, the dataset was reduced to 392,733 rows.

Calculated Retention: This yielded an approximate customer retention rate of 72.47%.

Transforming Transactions into Insights: Feature Engineering

To predict churn, we converted raw transaction data into meaningful customer-level metrics.

→ **Recency**

Days since the last purchase.

→ **Frequency**

Total number of purchases.

→ **Monetary Value**

Total spend by the customer.

→ **Customer Tenure**

Length of time as an active customer.

→ **Unique Products**

Variety of items purchased.

→ **Purchase Span**

Duration between first and last order.

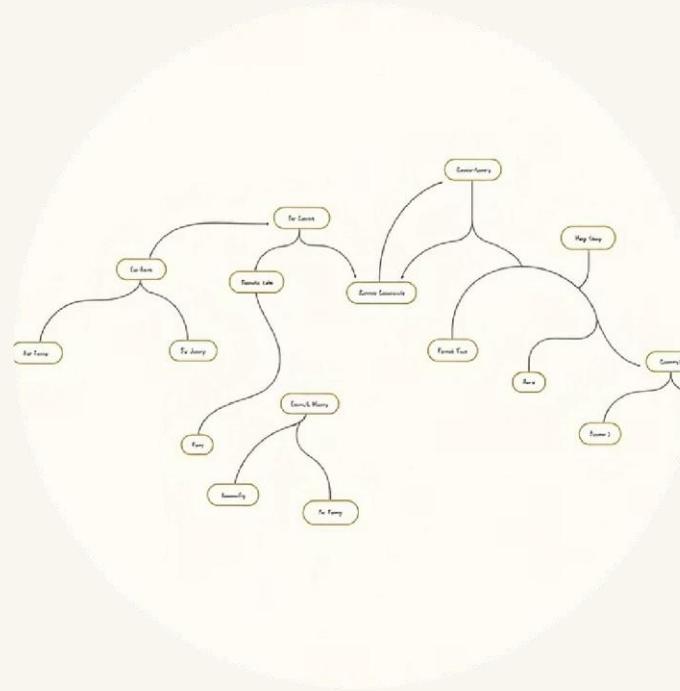
Selecting the Optimal Prediction Model

We rigorously evaluated several machine learning models to find the best fit for churn prediction.



Logistic Regression

A strong baseline for binary classification.



Decision Tree

Intuitive, rule-based approach.



Random Forest

Ensemble power, chosen for its robust performance.

Our selection was based on **ROC-AUC performance**, **model stability**, and ease of **business interpretability**

The Chosen Solution: Random Forest Classifier

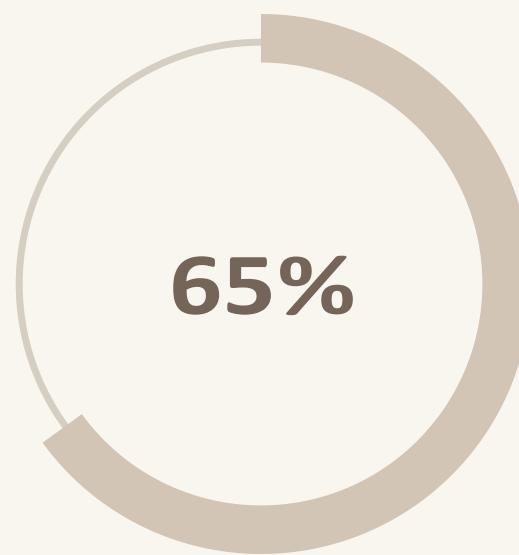
Our final model is a **Random Forest Classifier**, known for its accuracy and ability to handle complex datasets.

Model Output: It provides not only a clear **churn label** (yes/no) but also a **probability of churn**, giving deeper insights into customer risk.

Versatile Application: This model is designed for both **single customer predictions** and **batch processing** of large customer lists.

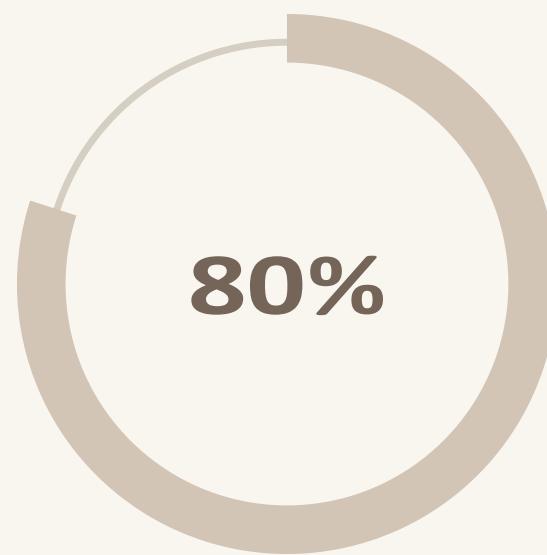


Model Performance & Business Metrics



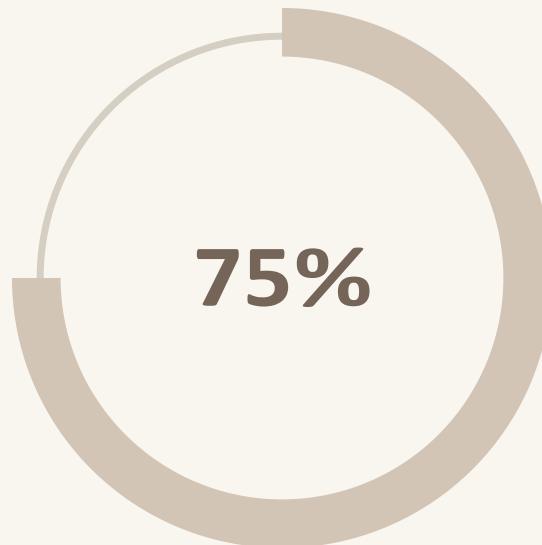
ROC-AUC Score

Achieved a robust score of >0.65 , indicating strong predictive power.



Churn Separation

The model successfully differentiates between churning and non-churning customers.



Balanced Accuracy

Strikes a good balance between identifying at-risk customers and overall accuracy.

What does ROC-AUC mean? It measures the model's ability to distinguish between churners and non-churners. A higher score means better separation, allowing us to balance the risk of false positives against the benefits of early detection.

Driving Growth: The Business Impact



Early Churn Detection

Proactively identify at-risk customers before they disengage.



Targeted Campaigns

Enable personalized retention strategies for maximum impact.



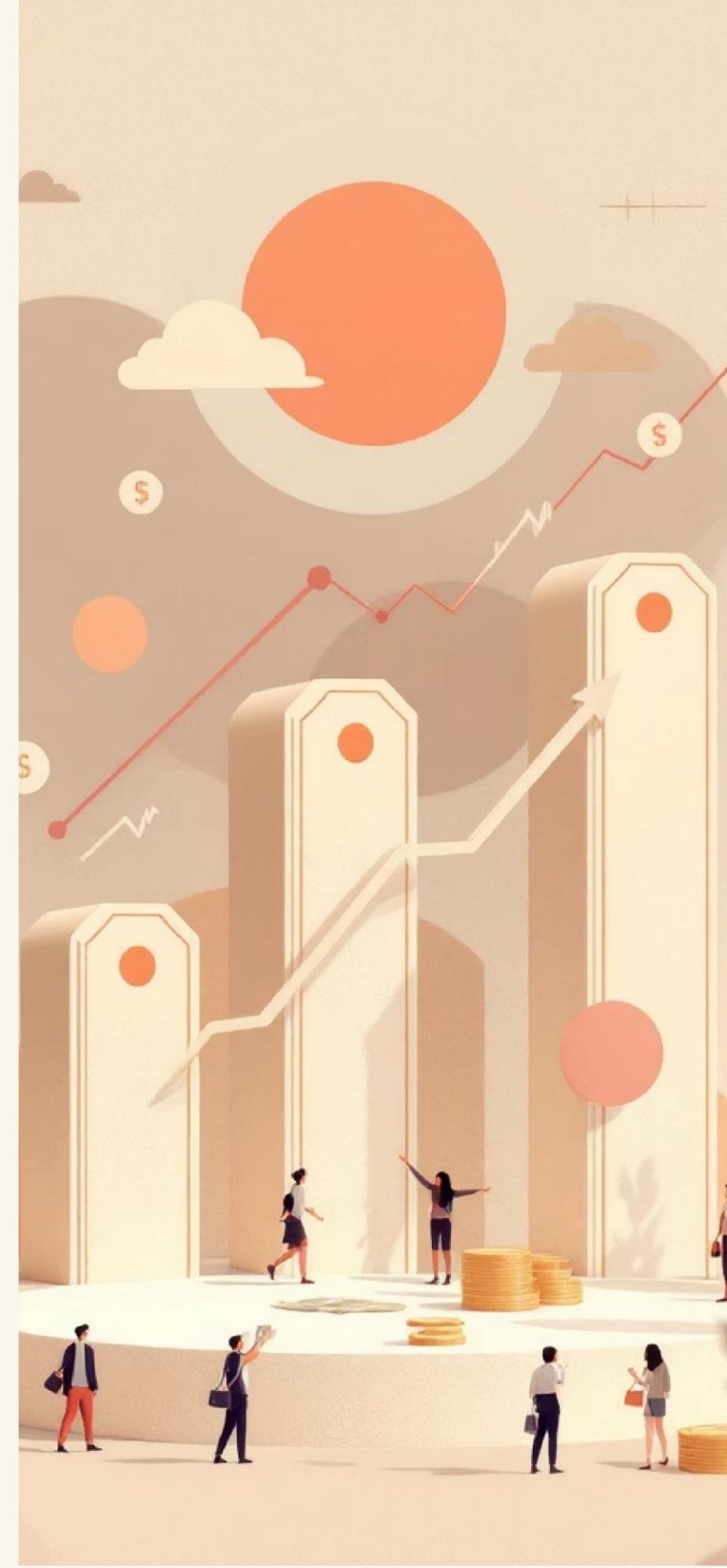
Cost Savings

Retaining customers is significantly cheaper than acquiring new ones.



Revenue & LTV Growth

Increase customer lifetime value and overall business revenue.



Deployment: Interactive Streamlit Web App

We've deployed the churn prediction model into an intuitive **Streamlit Web App**, making it accessible to all stakeholders.

[Access the Web App](#)

Key Features:

- Single Customer Prediction:** Quickly check individual customer churn risk.
- CSV Batch Prediction:** Upload a list of customers for bulk analysis.
- Model Dashboard:** Visualize key model insights and performance.



This application provides a user-friendly interface to leverage the power of our churn prediction model.

Key Learnings & Future Directions

Key Learnings

Practical ML Pipeline

Designing an end-to-end machine learning workflow.

Data Quality Challenges

Mastering cleaning and preparation of real-world datasets.

Feature Engineering Value

Transforming raw transactions into powerful predictive signals.

Model Evaluation Thinking

Focusing beyond accuracy to relevant business metrics.

Deployment & MLOps

Understanding steps to productionize ML solutions.

Future Improvements

Advanced Models

Exploring Gradient Boosting methods (e.g., XGBoost, LightGBM).

Automated Monitoring

Implementing systems for data drift and model decay detection.

Real-time API

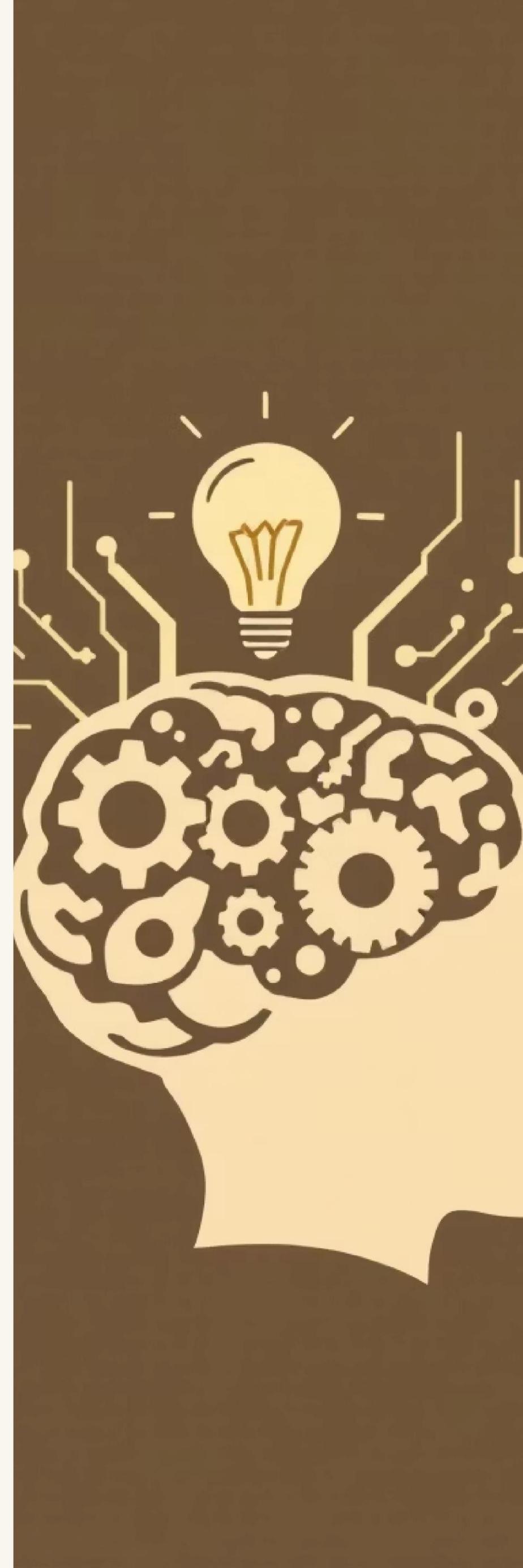
Developing an API for instant churn prediction on new customer data.

Explainable AI Dashboard

Building tools to interpret model decisions and feature importance.

Customer Segmentation

Integrating with segmentation for highly targeted retention campaigns.



Thank You

Questions welcome

Sravani Karnataka

Thank You