Portfolio Project: Customer Segmentation Using K-Means Clustering

Project Overview

This project focused on predicting customer churn for a telecom company using supervised machine learning models. The goal was to help the business proactively identify at-risk customers and improve retention strategies.

Tools & Technologies Used

- Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
- Logistic Regression, Random Forest, Decision Tree
- SMOTE for class balancing
- Jupyter Notebook, Google Colab

Project Steps

1. Data Cleaning & Preprocessing
2. Exploratory Data Analysis (EDA)
3. Feature Engineering & Encoding
4. Model Training & Tuning
5. Model Evaluation (Confusion Matrix, AUC Score)
6. Feature Importance Analysis

Key Insights

- The Random Forest model achieved 87% accuracy and an AUC score of 0.91
- Top churn drivers included contract type, monthly charges, and tenure
- Business recommended to improve engagement with month-to-month customers

Impact

The model helped the company identify high-risk customers and develop targeted retention campaigns, contributing to a measurable reduction in churn rate over the following quarter.

Portfolio Project: Customer Segmentation Using K-Means Clustering

Contact & Availability

If you're looking to reduce customer churn and drive loyalty using data, I'm available to help build predictive models tailored to your business goals.

Let's put your data to work!