Shiash InfoTech Solution

Customer Churn Prediction using Machine Learning

Presented by Yugaash Sridhar

Abstract

This project analyzes customer behavior using machine learning to predict churn, identify high-risk groups, and provide actionable insights through visualizations, enabling targeted retention strategies to boost loyalty and profitability.

Overview

Introduction

Objective :

- Predict customer churn and understand behavior trends.
- Enable businesses to reduce churn and improve profitability.

Problem Statement :

- High customer churn leads to revenue loss.
- Early identification enables better retention strategies.

Approach: Use machine learning models and visualizations to analyze patterns

Dataset Overview

Dataset Information:

- Source: Telco Customer Churn Dataset (e.g., Kaggle).
- Number of Records: [e.g., 7043 rows]
- Features: 21 columns, including demographics, service details, and financial information.
- Target Variable: Churn (1 = Churned, 0 = Not Churned).

Sample Features:

- Tenure: Duration with the company.
- MonthlyCharges: Monthly subscription fee.
- ContractType: Month-to-Month, One Year, etc.
- PaymentMethod: Credit Card, Bank Transfer, etc.

Data Preprocessing

Missing Value Handling :

- Filled missing values in TotalCharges with the median.

Encoding Categorical Variables :

- Label Encoding for binary columns (e.g., Churn).
- One-Hot Encoding for multi-category columns (e.g., InternetService).

Scaling :

- Standardized numerical columns like MonthlyCharges and Tenure using StandardScaler.

Feature Selection :

- Removed redundant columns (e.g., CustomerID).

Model Selection

- **Algorithm Used :** Random Forest Classifier

  Why Random Forest?

  - Handles mixed data types (categorical + numerical).
  - Robust to overfitting.
  - Provides feature importance insights.

- **Hyperparameters :**
  - Number of Trees: 100
  - Random State: 42

Model Evaluation

Performance Metrics:

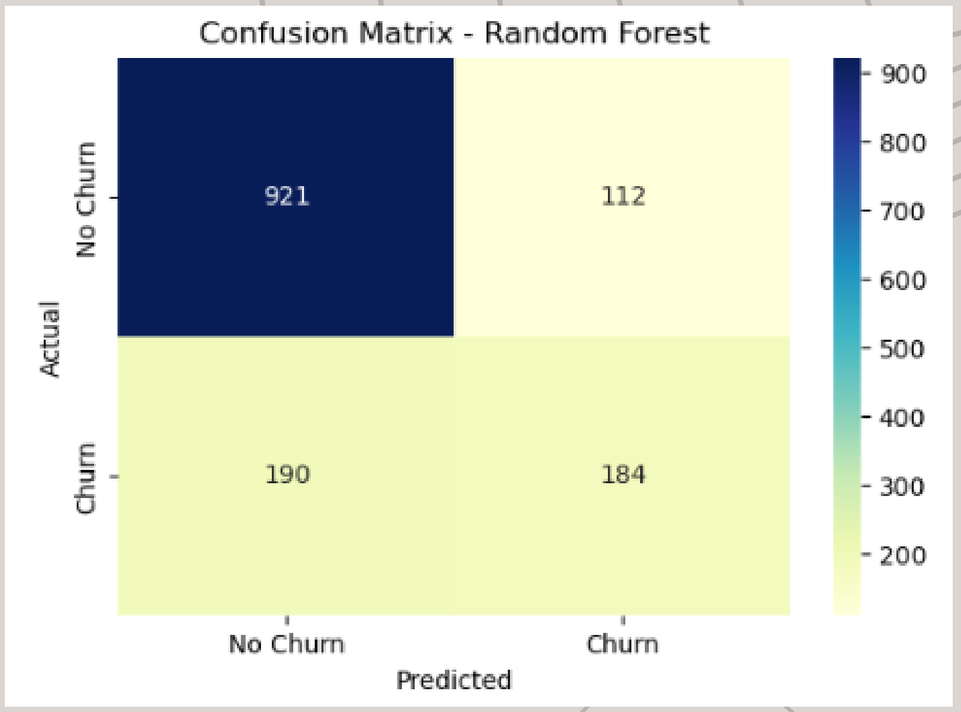
Accuracy, Precision, Recall, F1-Score

Confusion Matrix:

- Visualize predictions

(e.g., True Positives, False Negatives).

- Include a heatmap of the confusion matrix.



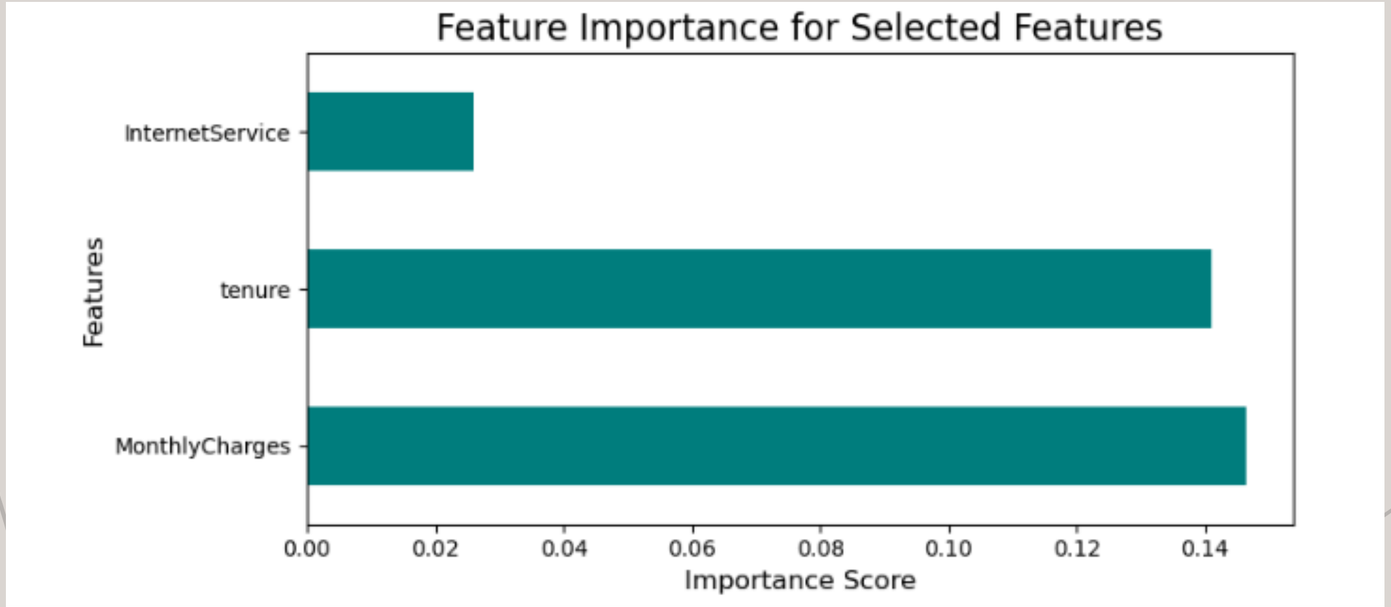
Feature Importance

Top Features:

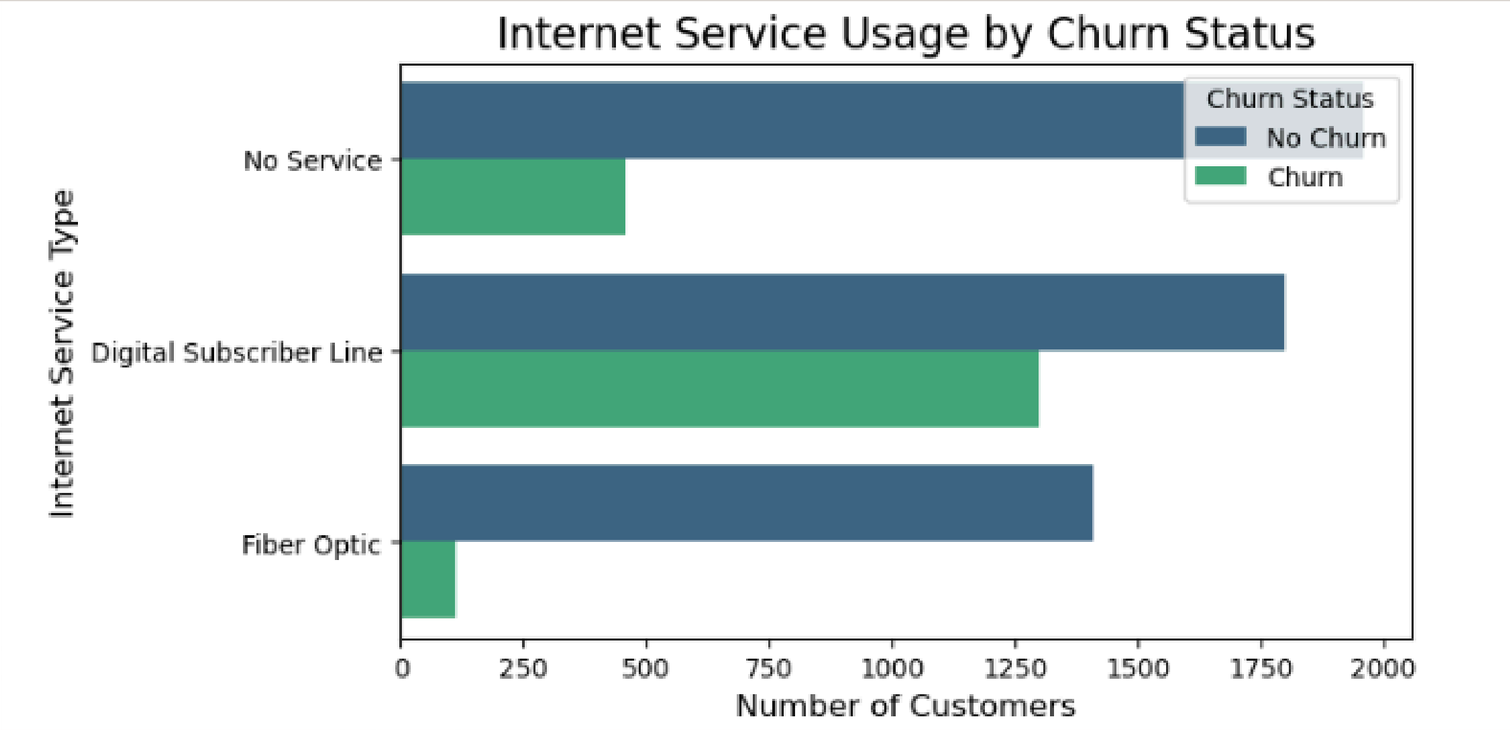
a. MonthlyCharges: Customers with higher charges tend to churn.

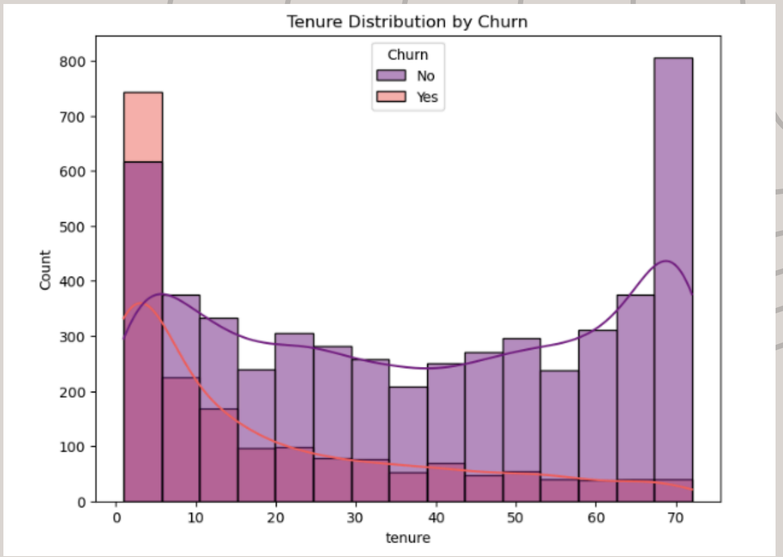
b. Tenure: Short-tenure customers are more likely to churn.

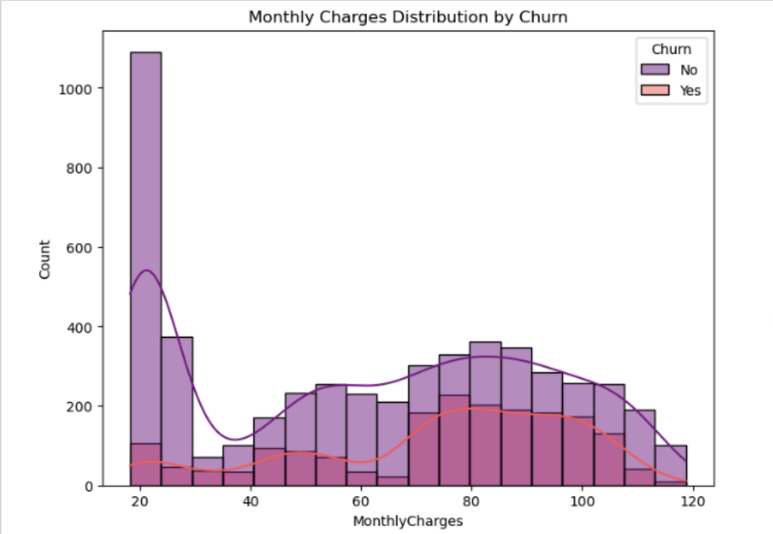
c. ContractType: Month-to-month contracts have higher churn rates.



Visualizations







Insights and Recommendations

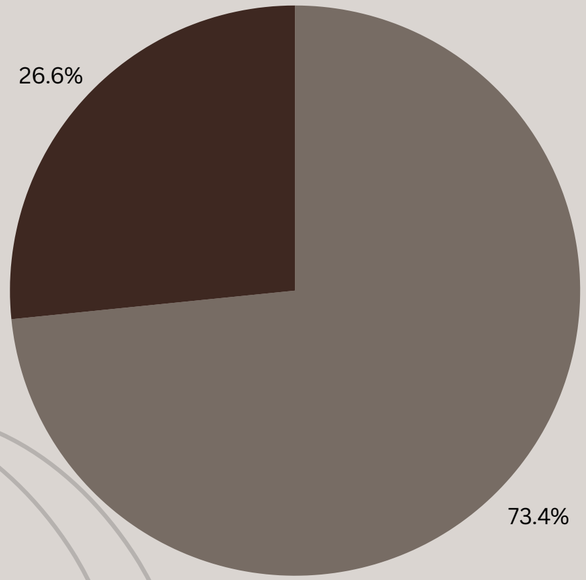
Insights:

- Customers with high MonthlyCharges and short Tenure are more likely to churn.
- Month-to-Month contracts have the highest churn rate.
- Internet service type impacts churn (e.g., Fiber Optic customers churn more).

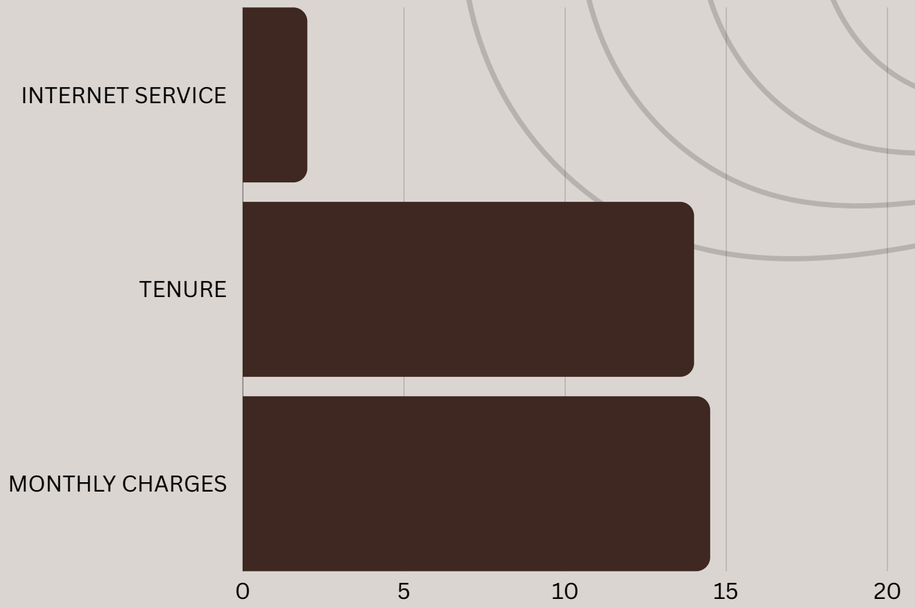
Recommendations:

1. Offer discounts for high-charges customers.
2. Create loyalty programs for new customers.
3. Improve service quality for at-risk segments (e.g., Fiber Optic users).

Result



CHURN DISTRIBUTION



FEATURE IMPORTANCE

Conclusion

The Customer Churn Prediction Project utilized a Random Forest model with high accuracy to predict churn and identify at-risk customers. Key factors like MonthlyCharges, Tenure, and ContractType revealed that 27% of customers churned, driven by high charges, short tenure, and month-to-month contracts. Visual insights highlighted correlations between churn and service usage. Recommendations include offering discounts to high-paying customers, loyalty programs for short-tenure users, and service improvements for Fiber Optic users, enhancing retention and profitability.

Thank You

Presented by Yugaash Sridhar