Title: "Customer Churn Prediction: A Machine Learning Approach"

Subtitle: Predicting Customer Attrition with Classification Algorithms

PRESENTER NAME:SAMUEL YASHUA
DATE:

Introduction

What is Customer Churn?

- Churn occurs when customers stop doing business with a company.
- It significantly impacts revenue, making churn prediction critical.

Objective:

- Identify at-risk customers.
- Predict churn using machine learning and the Telco Customer Churn dataset.
- Implement retention strategies for identified segments.

Tools:

- Python: `pandas`, `matplotlib`, `scikit-learn`, etc.

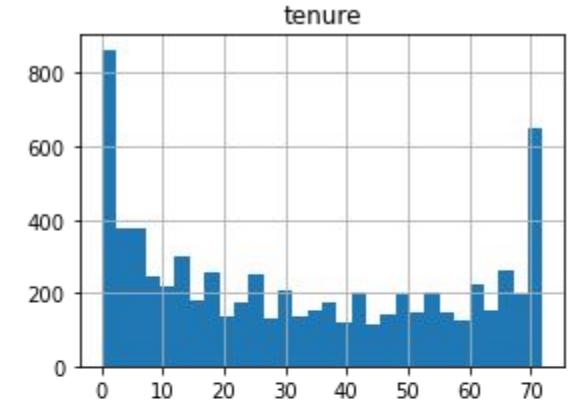
Dataset and Workflow Overview

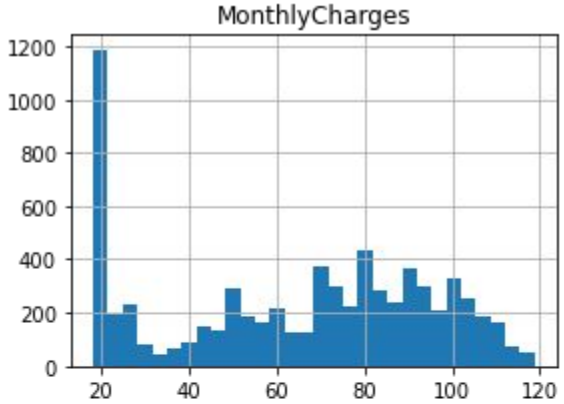
Dataset: Telco Customer Churn dataset with features like tenure, monthly charges, and churn status.

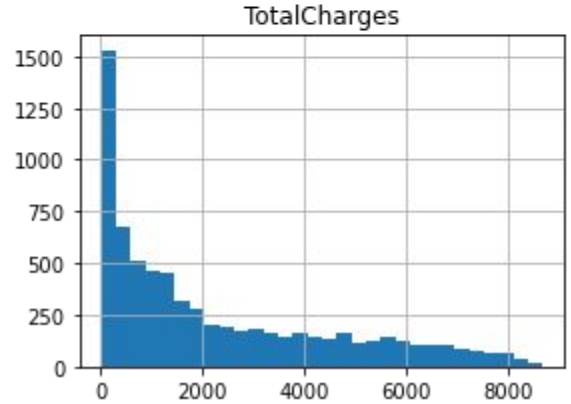
Workflow:

1. Import Libraries
2. Load Dataset
3. Exploratory Data Analysis (EDA)
4. Handle Outliers (IQR Method)
5. Data Cleaning and Transformation
6. One-hot Encoding
7. Feature Scaling
8. Model Training and Evaluation:
    - Logistic Regression
    - Support Vector Classifier
    - Decision Tree
    - K-Nearest Neighbors (KNN)

Exploratory Data Analysis (EDA) - Numerical Features





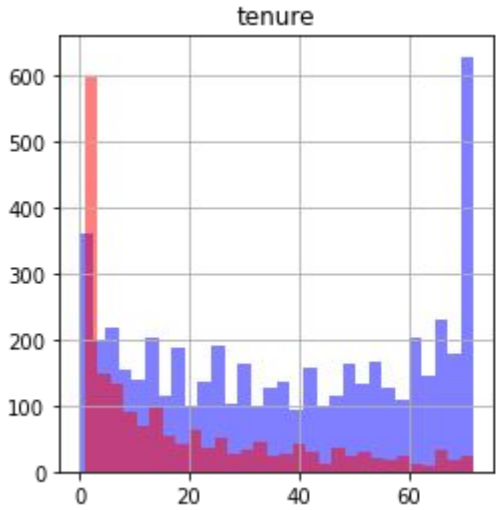


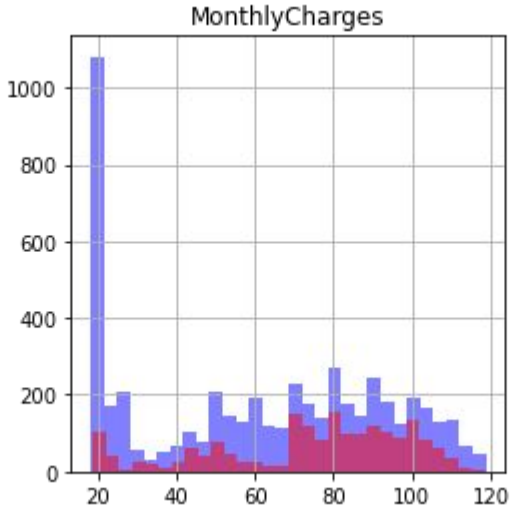
Content:

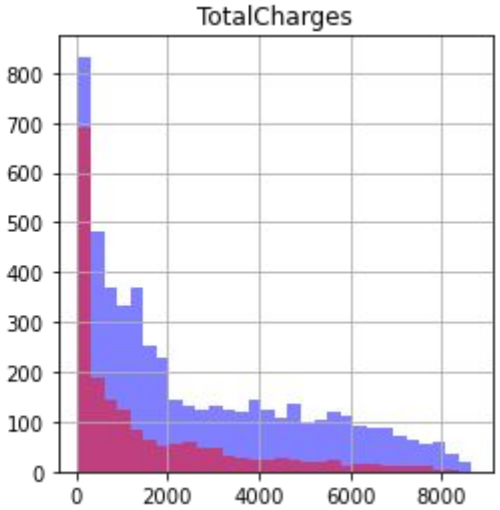
- Features: Tenure, MonthlyCharges, TotalCharges.
- Longer tenure and higher TotalCharges indicate lower churn probability.

Side-by-side comparison of churn vs. non-churn customers for numerical features.

EDA - Categorical Features



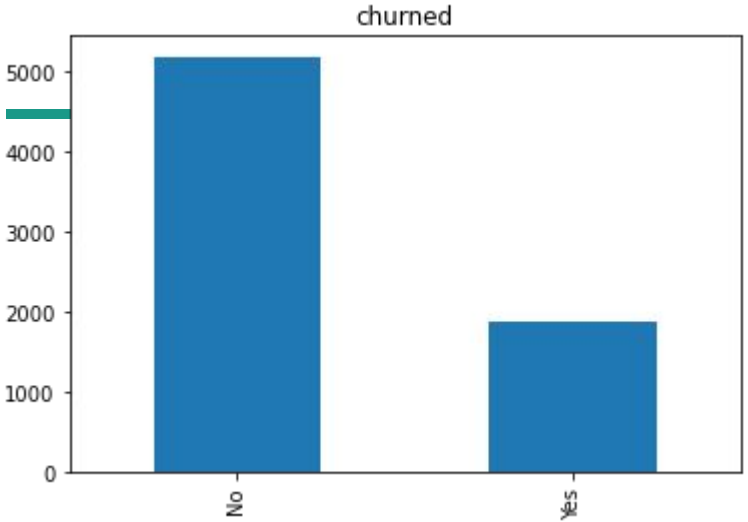




Content:

- Contract Type: Month-to-month contracts see higher churn.
- Senior Citizens and customers without phone service are less represented.
- Heatmap for categorical features in relation to churn.

EDA - Target Variable Distribution



Content:

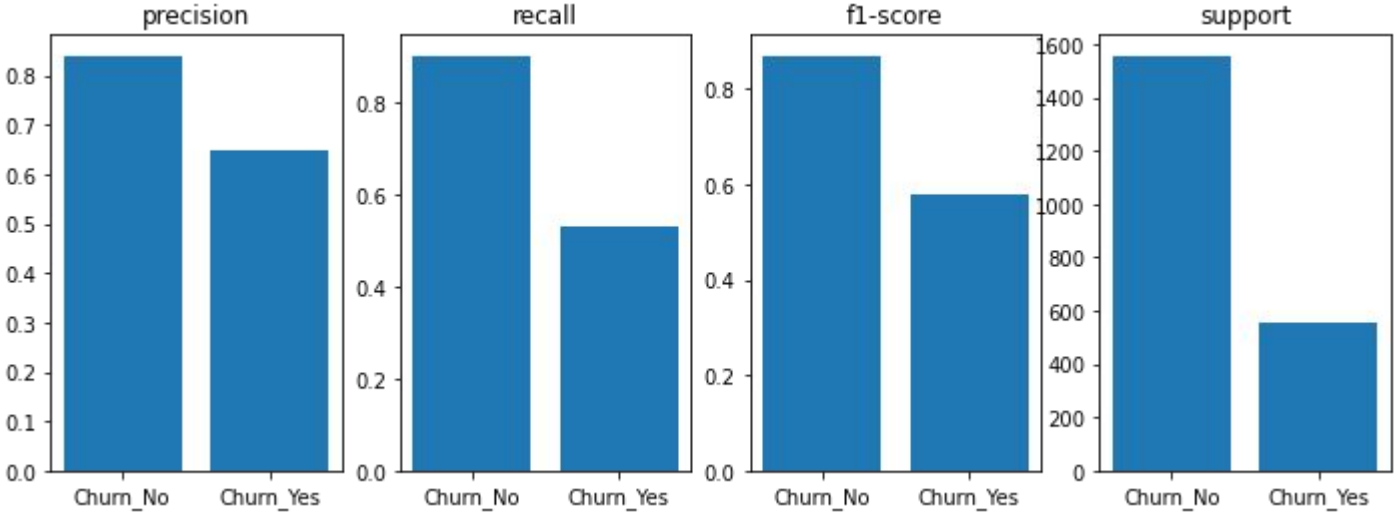
- Non-churned customers are significantly higher, leading to an imbalanced dataset.
- This imbalance could bias predictions toward the majority class.
- Proposed Solutions:
  - Resample the dataset.
  - Focus on metrics like precision and recall.

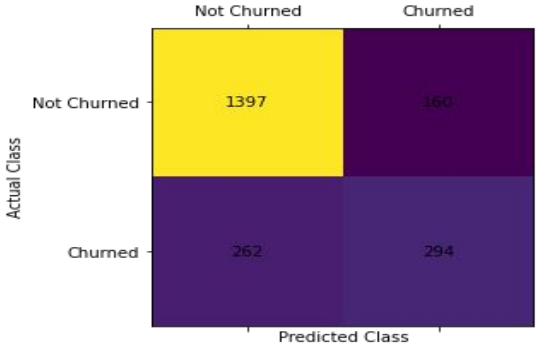
Model Building - Overview

Content:

- **Feature Scaling**: StandardScaler applied to numerical data.
- **Feature Selection**: Split data into train and test sets.
- Models Used:
  - Logistic Regression
  - Support Vector Classifier
  - Decision Tree Classifier
  - KNN Classifier

Logistic Regression





Content:

- **Accuracy**: 80%
- **Precision**: 84% (non-churned), 65% (churned)
- **Confusion Matrix**:
  - True Positives: Customers correctly identified as churned.
  - True Negatives: Customers correctly identified as non-churned.

Other Models and Comparisons

Content:

- **Support Vector Classifier**:
  - Accuracy: 80%, Precision: 83% (non-churned), 67% (churned)
- **Decision Tree Classifier**:
  - Accuracy: 72%, Overfits training data.
- **KNN Classifier**:
  - Accuracy: 79%, Best results at k=30.

Insights and Recommendations

Content:

- Churn is influenced by tenure, TotalCharges, and contract type.
- Month-to-month contracts have the highest churn rate.
- The imbalanced dataset affects model performance.

Recommendations:

- Implement retention strategies for at-risk groups.
- Collect additional churn data to balance the dataset.
- Use precision-recall-focused metrics in future analysis.

Thank You

Content:

- Thank the audience for their attention.
- Invite questions or discussions.