**Project Report:**

**Telecom Churn Prediction and Analysis**

**Introduction**

Customer churn is a critical issue in the telecom industry, directly impacting revenue and operational costs. This project aims to predict customer churn using machine learning techniques, identify the factors driving it, and recommend actionable strategies to improve customer retention and service quality. By leveraging data-driven insights, telecom companies can proactively address churn, reduce customer loss, and optimize their operations.

**Data Collection and Quality Enhancement**

**Data Collection:** A synthetic dataset was created to simulate real-world telecom data, consisting of 10,000 records with attributes such as customer demographics, contract details, usage patterns, and churn status.

**Data Quality Enhancement Steps:**

1. **Handling Missing Values:**
   * Checked for missing values and imputed them using statistical methods (mean/median for numerical attributes and mode for categorical attributes).
   * Code:

python

# Handling missing values

telecom\_data.fillna(telecom\_data.mean(), inplace=True)

telecom\_data['contract\_type'].fillna(telecom\_data['contract\_type'].mode()[0], inplace=True)

1. **Outlier Removal:**
   * Identified outliers in numerical columns like monthly\_charges and removed extreme values using IQR-based filtering.

python

Q1 = telecom\_data['monthly\_charges'].quantile(0.25)

Q3 = telecom\_data['monthly\_charges'].quantile(0.75)

IQR = Q3 - Q1

telecom\_data = telecom\_data[~((telecom\_data['monthly\_charges'] < (Q1 - 1.5 \* IQR)) |

(telecom\_data['monthly\_charges'] > (Q3 + 1.5 \* IQR)))]

1. **Feature Encoding:**
   * Categorical attributes such as gender and contract\_type were converted into numerical formats using one-hot encoding.

**Data Analysis and Insights**

**Analysis Results:**

* Customers with **"Month-to-Month" contracts** showed higher churn rates compared to those with longer-term contracts.
* High churn rates were observed among customers with frequent **customer support calls** or low tenure.

**Key Visualizations:**

1. **Churn Rate by Contract Type:**

python

import seaborn as sns

import matplotlib.pyplot as plt

sns.barplot(x="contract\_type", y="churn", data=telecom\_data, estimator=lambda x: sum(x)/len(x))

plt.title("Churn Rate by Contract Type")

plt.show()

1. **Churn by Tenure:**

python

sns.histplot(data=telecom\_data, x="tenure", hue="churn", multiple="stack", bins=20)

plt.title("Churn Distribution by Tenure")

plt.show()

**Modeling and Recommendations**

**Modeling Approach:**

* A **Random Forest Classifier** was used to predict churn. The model achieved an overall accuracy of 76%, with scope for improving recall for churned customers.
* **Feature Importance:** The most influential features were tenure, monthly charges, and customer support calls.

**Code for Model Training:**

python

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import train\_test\_split

# Splitting the data

X = telecom\_data.drop(columns=["customer\_id", "churn", "contract\_type", "payment\_method", "gender"])

X = pd.get\_dummies(X, drop\_first=True)

y = telecom\_data["churn"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Training the model

rf\_model = RandomForestClassifier(random\_state=42)

rf\_model.fit(X\_train, y\_train)

**Recommendations:**

1. Target high-risk customers with proactive retention strategies (e.g., personalized offers for "Month-to-Month" customers).
2. Automate predictive insights into the CRM system to flag customers with a high churn probability for immediate intervention.

**GitHub Repository**

All project files, including the dataset, code, and documentation, are available on GitHub: https://github.com/deekshithanadikattu/telecom\_churn

**Conclusion**

This project highlights the importance of predictive modeling in addressing customer churn in the telecom industry. Key findings emphasize the influence of contract type, tenure, and customer support interactions on churn behavior. The model enables actionable insights, supporting targeted retention strategies that can significantly reduce churn rates.

**Future Directions:**

1. Incorporate real-time data streams for ongoing churn analysis.
2. Enhance model performance using advanced techniques like XGBoost or deep learning.
3. Integrate additional data sources such as network quality and customer feedback.