**Telecom Company Customer Churn Prediction: Summary Report**

**Source code - Notebook**

Link - https://github.com/ahmedpas/TELECOM-CUSTOMER-CHURN

PREDICTION/blob/main/Telecom\_cust\_churn\_project%20(1).ipynb

**1. Introduction**

Dataset, Features and Target value

Source : https://www.kaggle.com/blastchar/telco-customer-churn ( IBM Sample dataset)

Here, IBM provided customer data for Telco industry to predict churn customer based on demographic, usage and account based information. Main objective is that to analyze churn customer behaviors and develop strategies for customer retention.

Assumption - Here, data source has not provided any information related to time; So I have assumed that records are specific to the particular month.

Dataset has information related to,

**2.Demographic:**

∙ Gender - Male / Female

∙ Age range - In terms of Partner, Dependent and Senior Citizen

∙ Services:

∙ Phone service - If customer has Phone service, then services related to Phone like; ∙ Multiline Phone service

∙ Internet Service - If customer has Internet service, then services related to Internet like ∙ Online security

∙ Online backup

∙ Device protection

∙ Tech support

∙ Streaming TV

∙ Streaming Movies

∙ Account type:

∙ Tenure - How long customer is with the company?

∙ Contract type - What kind of contract they have with a company? Like

∙ Monthly bases

∙ On going bases - If on going bases, then One month contract or Two year contract ∙ Paperless billing - Customer is paperless billion option or not?

∙ Payment method - What kind of payment method customer has?

∙ Mailed check

∙ Electronic check

∙ Credit card (Automatic)

∙ Bank transfer (Automatic)

**Usage:**

∙ Monthly charges

∙ Total charges

**Target:**

∙ Churn - Whether customer left the company or still with the company?

∙ New\_customer- Here predicting the churn of the new customer

**Problem Description:-**

Predict whether a customer will leave a telecom company (churn) based on their usage patterns and demographic data. This project introduces classification techniques and business applications

The reasons behind the customer leaving company could be-

∙ High charges

∙ Better offer from competitor

∙ Poor customer service

∙ Some unknown reasons

∙ How to detect the churn customer?

∙ Monitoring usage

∙ Analysing complains

∙ Analyzing competitors offers

∙ How to prevent customers from leaving a company?

∙ Once you detect high risk customers, apply

∙ Customer Service

∙ Retention plans

**3. Data Preprocessing**

Several preprocessing steps were taken to prepare the dataset for model training:

∙ **Categorical Variables**: Mapped to numerical values for easier model interpretation (e.g., Yes to 1 and No to 0).

∙ **Normalization**: Tenure and Monthly Charges were normalized to bring all features to a similar scale.

∙ **Feature Selection**: Selected key features like Tenure, MonthlyCharges, Partner, Dependents, and internet services for model input.

**4. Exploratory Data Analysis (EDA)**

∙ **Churn Rate**: A significant proportion of customers churned (about 26%). ∙ **Feature Analysis**: Higher churn was noticed among customers with month-to-month contracts and higher monthly charges.

**5. Model Training and Evaluation**

Three machine learning models were implemented: Decision Tree, Random Forest, and Logistic Regression.

1. **Decision Tree Classifier**:

o **Accuracy**: Around 78%.

o **Performance**: Provides a clear view of the decision process but can overfit the data without tuning.

2. **Random Forest Classifier**:

o **Accuracy**: Around 80% – the highest of all models.

o **Performance**: Excellent at reducing overfitting by averaging multiple decision trees, providing robust predictions.

3. **Logistic Regression**:

o **Accuracy**: Around 76%.

o **Performance**: Offers good interpretability but may not capture complex patterns compared to tree-based models.

**6. Predictions on New Customers**

The model was tested on new customer data to predict churn likelihood:

∙ **Customer 1**: High risk of churn (Predicted to Churn).

∙ **Customer 2**: Low risk of churn (Predicted to Stay).

**7. Conclusion**

∙ **Best Model**: Random Forest outperformed the other models in terms of accuracy and robustness, making it the most reliable for predicting customer churn.

∙ **Business Insight**: Customers with higher monthly charges and short tenures, especially those on month-to-month contracts, are more likely to churn. The company can focus retention efforts on these high-risk customers by offering better service plans or incentives.

**8. Recommendations**

∙ **Next Steps**: Fine-tune the Random Forest model by using hyperparameter tuning and feature engineering.

∙ **Retention Strategies**: Target high-risk customers with personalized retention offers, such as discounts or contract upgrades.