Telco Customer Churn Prediction Report

# Dataset

- Source: IBM Sample Dataset  
- File: Telco Customer Churn.csv  
- Rows: 7,043  
- Target: Churn (Yes/No)

# Project Steps

1. Data Cleaning  
 - Converted TotalCharges to numeric  
 - Dropped missing values (~11 rows)  
 - Removed customerID as it's non-informative  
  
2. Feature Engineering  
 - One-hot encoding for categorical columns  
 - Converted Churn to binary  
  
3. Train/Test Split  
 - 80/20 split using train\_test\_split

# Models Used

1. Logistic Regression  
 - Accuracy: ~78.7%  
 - Good precision for non-churners, low recall for churners  
  
2. Random Forest  
 - Accuracy: ~78.5%  
 - More balanced performance  
  
3. Random Forest with class\_weight='balanced'  
 - Improved churn recall  
  
4. SMOTE (Synthetic Minority Oversampling)  
 - Improved recall on churn class (minority)

# Evaluation Metrics

- Accuracy  
- Precision / Recall / F1-Score  
- Confusion Matrix  
- ROC Curve and AUC Score

# Top Predictors of Churn

1. Contract\_Month-to-month  
2. tenure  
3. OnlineSecurity\_No  
4. TechSupport\_No  
5. InternetService\_Fiber optic  
6. MonthlyCharges  
7. PaymentMethod\_Electronic check

# Optional Enhancements

- [x] SMOTE for class balancing  
- [x] Hyperparameter tuning (GridSearchCV)  
- [ ] Export model (joblib / pickle)  
- [ ] Streamlit / Power BI Deployment

# How to Run

pip install pandas scikit-learn imbalanced-learn matplotlib seaborn  
  
Run the Jupyter Notebook: Telco\_Customer\_Churn.ipynb

# Conclusion

We used data preprocessing, Random Forest, class balancing, and evaluation to build a churn prediction model. It helps telecom providers retain customers by identifying those likely to churn.

**📊 Key Insights from EDA (Exploratory Data Analysis)**

**1. 🔄 Overall Churn Rate**

* Around **26.5%** of customers have churned (industry benchmark based on this dataset).

**2. 📉 Tenure**

* **Tenure is negatively correlated with churn**.
  + Customers with short tenure (new users) churn more frequently.
  + Long-tenure customers are much more loyal.

**3. 💳 Contract Type**

* Churn rate is highest for:
  + **Month-to-month** contracts (~43% churn).
* Very low churn for:
  + **Two-year contracts** (~3% churn).
* Insight: **Longer contracts reduce churn.**

**4. 🌐 Internet Service**

* Customers using **Fiber optic** internet have a much higher churn rate than DSL.
  + Possibly due to higher costs or poor experience.

**5. 🛡️ Online Security / Tech Support**

* Customers **without security or tech support** churn more.
* These services may make customers feel more supported and "locked-in".

**6. 💸 Monthly & Total Charges**

* **Higher Monthly Charges** → higher churn.
* But **Total Charges** (lifetime value) is higher for retained customers.
  + New, high-paying customers are more likely to leave.

**7. 💰 Payment Method**

* Churn is highest among users paying via **Electronic Check**.
* Lowest among those using **automatic bank transfers**.

**🤖 Key Insights from ML Models**

**✅ Best Model: Random Forest + SMOTE + Class Weights**

* **Accuracy**: ~78.4%
* **Precision (Churners)**: Balanced
* **Recall (Churners)**: Improved from 52% to ~65% using SMOTE

**🔍 Most Important Features (from feature importance)**

1. **Contract\_Month-to-month** – strongest churn indicator
2. **tenure** – higher tenure = less churn
3. **OnlineSecurity\_No**
4. **TechSupport\_No**
5. **InternetService\_Fiber optic**
6. **MonthlyCharges**
7. **PaymentMethod\_Electronic check**

**🧠 Strategic Takeaways**

* **Target month-to-month users** with offers to convert to longer plans.
* **Encourage auto-pay options** (lower churn).
* **Offer onboarding support** and tech help to new customers.
* **Monitor fiber users' satisfaction** and pricing perception.
* **Use ML predictions** in a dashboard to flag at-risk users weekly/monthly.