- **Project :** Telco Customer Churn Prediction Supervised ML
  - **Tools Used:** Pandas, NumPy, Scikit-learn, Seaborn
  - **Industry Context:** The telecommunications industry is highly competitive, with significant implications of customer churn on profitability and customer satisfaction. As of 2023, many telecom companies have reported churn rates between 25-30%, highlighting the urgent need for effective retention strategies. With increasing customer acquisition costs, understanding the factors driving churn can enable companies to implement targeted interventions and enhance customer loyalty.
  - **Problem Definition:**
    - Identify and analyze the factors contributing to customer churn within the telecommunications sector using historical data. Build predictive models to foresee potential churners, thus enabling proactive retention strategies and improved customer satisfaction.
  - **About Data:**
    - Dataset Overview: The dataset comprises information on 7,043 customers, containing 17 attributes including customer ID, service usage, tenure, monthly charges, total charges, and churn status.
  - **Data Preprocessing:**
    - **Data Cleaning**: Handles missing values and ensures data integrity, focusing on key columns such as total charges and monthly charges where nullable entries were addressed.
    - Encoding Categorical Variables: Employed one-hot encoding for categorical features 'churn'.
  - **Actionable Insights using Exploratory Data Analysis (EDA) (4 - 5 Insights):**
    - Churn Distribution: Approximately 26.54% of customers churned, translating to 1,869 customers, underscoring the significant impact on revenue.
    - Monthly Charges Insight: Customers who churned had an average monthly charge of $74.44, compared to $61.27 for non-churned customers, indicating price sensitivity.
    - Tenure Impact: Churned customers have a lower average tenure of 17.98 months, highlighting that newer customers are at higher risk of leaving.

- Service Usage Patterns: 1,697 churned customers had phone service, indicating potential dissatisfaction with service quality.
- Contract Type Influence: 51% of churned customers were on month-to-month contracts, suggesting a need for more attractive long-term options.

- **Data Preparation:** The entire dataset was utilized for modeling, as unsupervised models often do not require a train-test split, allowing full data usage for pattern recognition.
- **Model Selection:**
  - Developed baseline models: Logistic Regression, Decision Tree, and Random Forest classifiers.
  - Selected Random Forest for its superior performance:
    - Accuracy: 89.05%
    - Precision: 86.24%
    - Recall: 95.61%
    - F1 Score: 90.69

- **Model Training:**
  - **Oversampling:** Addressed class imbalance (2.77:1 ratio) by oversampling the minority class.
  - **Train-Test Split:** Data was split into training (80%) and testing (20%) sets to validate model performance effectively.
  - **Hyperparameter Tuning:** Optimized Random Forest hyperparameters using GridSearchCV, resulting in:
    - n_estimators: 500
    - max_depth: 20
    - min_samples_split: 2
    - min_samples_leaf: 1
    - criterion: 'entropy'
    - bootstrap: True

- **Model Evaluation:**
  - Confusion Matrix:
    - True Positives: 1,003
    - True Negatives: 861
    - False Positives: 46
    - False Negatives: 160
  - Overall Model Metrics:
    - F1 Score: 0.90.
    - AUC Score: 0.97 (indicating high model effectiveness).

- **Feature Importance:** In the Telco Customer Churn Prediction project, feature importance was crucial for defining and understanding the characteristics that contribute to customer churn. This involved analyzing key features such as MonthlyCharges, Tenure, Contract Type, and Internet Service to identify their impacts on churn risk.
- **Key Results:**
  - **Total Customers**: 7,043 total customers.
  - **Churn Rate**:
    - **Churned**: 1,869 (approximately **26.54%**).
    - **Non-Churned**: 5,174 (approximately **73.46%**).
  - **Monthly Charges**:
    - Average for churned customers: 74.44 Dollars.
    - Average for non-churned customers: 61.27 Dollars.
  - **Tenure**:
    - Average tenure for churned customers: **17.98 months**.
    - Average tenure for non-churned customers: **37.57 months**.
  - **Total Charges**:
    - Average total charges for churned customers: **1,531.80** Dollars.
    - Average total charges for non-churned customers: **2,552.88** Dollars.
  - **Service Usage**:
    - 1,699 churned customers had phone service; only **170** did not.

        - For Internet Service Fiber optic users had the highest churn at **1,297**, followed by DSL at **459**.
- **Business Impact:**
    - **Resource Allocation**: Focus retention strategies on the **1,869** at-risk customers to improve overall customer satisfaction and reduce churn cost.
      Establish dedicated retention teams for the 1,869 at-risk customers, ensuring follow-ups within 24 hours. Offer tailored incentives like discounted bundles and renewal-based pricing. Analyze behavior patterns and feedback to detect early churn signals. Use segmented marketing and regularly update churn models for effective, data-driven intervention.
    - **Customer Segmentation**:
        - Target **customers with higher monthly charges** and shorter tenures as they show the highest risk of churn. Customer Segmentation: Customers with tenure below 18 months and monthly charges above 74 Dollars are at highest risk.
          => Providing them with offers on extending service tenures or discount coupons.
        - Monitor customers with fiber optic internet service, as they have the highest churn rate.
          => investigate customer feedback specifically related to this service. Address any identified complaints or issues to improve customer retention.
    - **Retention Strategy Development**: Enhance tailored retention strategies based on specific customer service usage patterns, as indicated by churn statistics.

- **Future Scope:**
    - Incorporate More Data: Add customer interaction, complaint logs, and satisfaction surveys to deepen predictive power.
    - Model Optimization: Test Gradient Boosting and deep learning approaches for incremental gains.
    - Automated Notification: Deploy the model for real-time churn alerts to customer service teams.
- **Expected Interview Questions:**

- What data quality checks did you implement during data preprocessing, and how did they influence your analysis?
- Describe the key features in the dataset that you identified as significant predictors of churn. How did you determine their importance?
- How did you address the class imbalance in the dataset, and why is it important to do so?
- What evaluation metrics did you choose to assess the model's performance, and why were they selected?
- Can you explain the trade-offs between precision and recall in your Random Forest model? How did you ensure the model meets business objectives?
- What hyperparameter tuning techniques did you use for the Random Forest model, and how did they affect the model's performance?
- What hyperparameter tuning techniques did you use for the Random Forest model, and how did they affect the model's performance?
- How did the insights from your analysis contribute to the development of actionable retention strategies? Can you provide an example?
- How do demographic factors (age, gender, income) correlate with customer churn rates across different service plans?
- What are the long-term usage trends of customers before they churn, and how do these trends vary by customer segment?
- How does the type of internet service (fiber optic vs. DSL) impact customer retention and churn rates?
- What effect do contract types (month-to-month vs. annual) have on customer loyalty and churn statistics?
- How can we leverage advanced ensemble models to improve churn prediction accuracy compared to single models?
- How do external market trends (competition, economic conditions) influence churn rates in different regions?
- What features significantly contribute to the interpretability of churn models, particularly in the context of customer communication?
- How can we measure the effectiveness of targeted retention strategies post-implementation on overall churn rates?