**Predictive Modeling for Telco Customer Churn Mitigation**

**1. Executive Summary**

This report outlines the end-to-end development of a machine learning model to predict customer churn for a telecommunications company. The primary business objective was to proactively identify customers at high risk of leaving, enabling targeted retention efforts. Through a comprehensive process of data analysis, feature engineering, and iterative model optimization, we developed a highly effective solution.

The final recommended model, a **tuned Logistic Regression classifier**, achieved an outstanding **Recall score of 78%** for the churn class. This means the model successfully identifies **78 out of every 100 customers** who are genuinely at risk of churning, a significant improvement over baseline models. This powerful predictive tool can now be integrated into business operations to drastically reduce customer attrition and protect revenue.

**2. Introduction & Business Problem**

Customer churn is a critical challenge in the competitive telecommunications sector. Acquiring new customers is significantly more expensive than retaining existing ones. Therefore, the ability to accurately predict which customers are likely to churn is of immense strategic value. This project aimed to leverage customer data—including demographics, account information, and subscribed services—to build a robust classification model that addresses this challenge head-on.

**3. Data Analysis & Preparation**

The initial phase involved a thorough Exploratory Data Analysis (EDA) of the provided dataset, which contained 7,043 customer records.

**Key Insights from EDA:**

* **Contract Type:** Customers on a **month-to-month contract** were overwhelmingly more likely to churn compared to those on one or two-year contracts.
* **Tenure:** New customers with low tenure represented a higher churn risk.
* **Internet Service:** Customers with Fiber optic internet showed a higher churn rate.

**Data Preprocessing Pipeline:**

* **Data Cleaning:** Handled missing values found in the TotalCharges column.
* **Encoding:** Converted all categorical features (e.g., gender, Contract) into a numerical format using **One-Hot Encoding**.
* **Scaling:** Standardized numerical features (tenure, MonthlyCharges, TotalCharges) using StandardScaler to ensure they were evaluated fairly by the models.

**4. Modeling & Iterative Improvement**

The model development was an iterative process focused on systematically enhancing predictive performance, with a primary focus on the **Recall** metric for the "Churn" class.

**I. Baseline Models:** We began by training several standard models (Logistic Regression, Random Forest). The initial Logistic Regression model achieved a **Recall of 58%**, serving as our performance baseline.

**II. Addressing Class Imbalance:** The most significant challenge was the severe class imbalance in the dataset (more non-churners than churners). This was addressed using the **Synthetic Minority Over-sampling Technique (SMOTE)**. By applying SMOTE to the training data, we created a balanced dataset, which immediately boosted the Logistic Regression model's **Recall to 64%**.

**III. Hyperparameter Tuning:** To extract the maximum performance from each model, we employed **GridSearchCV** to systematically search for the optimal hyperparameter settings. This step proved crucial, further improving model performance by finding the best configuration for each algorithm.

**IV. Feature Engineering:** The final optimization step involved creating a new, insightful feature: charge\_per\_tenure (MonthlyCharges / tenure). This feature was engineered to capture the relationship between the monthly cost and customer loyalty, providing the model with a powerful new predictive signal.

**5. Final Model Selection & Results**

We systematically evaluated nine different modeling strategies. The **final, winning model** was a **Logistic Regression classifier that incorporated all advanced techniques**: SMOTE for data balancing, GridSearchCV for hyperparameter tuning, and the custom-engineered feature.

**Final Model Performance:**

* **Recall (Churn): 78%** 🏆
* **Precision (Churn): 52%**
* **F1-Score (Churn): 62%**
* **Overall Accuracy: 75%**

This model was selected because it decisively achieved the project's primary goal: maximizing the identification of at-risk customers. While its precision is moderate, this is an acceptable trade-off for a business that prioritizes minimizing missed churn opportunities.

**6. Conclusion & Recommendations**

This project has successfully produced a high-performance machine learning model capable of predicting customer churn with a **78% success rate** in identifying at-risk customers. This represents a powerful tool for proactive customer retention.

**Recommendations for Implementation:**

1. **Deploy the Model:** Integrate the final Logistic Regression model into the company's CRM or operational systems.
2. **Generate Proactive Alerts:** Use the model to score the entire customer base monthly and generate a "risk list" of customers with the highest probability of churning.
3. **Targeted Retention Campaigns:** Empower the customer retention team to use this list to engage at-risk customers with targeted offers, such as discounts for switching to annual contracts, personalized service check-ins, or loyalty bonuses.

By operationalizing this model, the company can transition from a reactive to a proactive customer retention strategy, leading to increased customer loyalty, reduced churn, and a stronger bottom line.