**Employee Attrition Prediction – Final Report**

**Problem Statement**

The objective of this case study is to develop a predictive model to identify employees who are likely to leave the organization. This enables the HR team to take proactive actions and improve employee retention strategies, potentially saving costs and retaining critical talent.

**Methodology & Approach**

1. **Data Understanding & Preparation**
   * The dataset includes demographic, work-related, and satisfaction-based factors along with the target variable: Attrition (Stayed/Left).
   * Checked column types, missing values, and correcting encoding issues.
   * Handled missing data for numerical features (Distance from Home and Company Tenure) using median imputation.
   * Dropped identifier columns like Employee ID as they do not contribute to prediction.
2. **Train-Test Split**
   * Split the data into **training (70%)** and **validation (30%)** sets with stratification to maintain the target distribution.
3. **Feature Engineering**
   * Dummy variables were created for all relevant categorical fields.
   * Numerical features were standardized using Standard Scaler.
   * Feature selection was performed using **Recursive Feature Elimination (RFE)** to choose the top 15 features.
   * All selected features were verified for multicollinearity using VIF, and no issues were found.
4. **Model Building**
   * Built a **logistic regression** model using the selected features.
   * The model was trained and evaluated on the training set using the default classification threshold of **0.5**.
   * While cutoff optimization (Step 7.3) was not performed in this iteration, the current model provides a strong starting point.

***Continued below ..***

**Model Performance (Training Set – Cutoff = 0.5)**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| **Accuracy** | **26.17%** |
| **Sensitivity** (Recall) | **28.22%** |
| **Specificity** | **24.32%** |
| **Precision** | **25.21%** |
| **Recall** | **28.22%** |

**Key Insights**

* The model was built using a robust and structured approach: data cleaning, feature engineering, and model training steps were handled effectively.
* The current results indicate that, with a default cutoff of 0.5, the model's predictive performance is limited.
* Despite that, the foundational work done in this version sets up the model well for iterative improvements.

**Key Takeaways / What the Organization Should Consider**

Even though the current version of the model is not production-ready, several important insights emerged that HR should start thinking about:

* **Overtime (Yes)** : Strongly associated with attrition. Employees working overtime are more likely to leave. Consider managing workload or offering incentives.
* **Job Satisfaction (Low)** : Low satisfaction correlates with leaving. Conduct pulse surveys and address concerns quickly.
* **Work-Life Balance (Poor/Fair)** : Imbalance is a driver of attrition. Promote flexible work, wellness programs, and boundary-setting.
* **Performance Rating (Low)** : Low performers are at risk — possibly due to disengagement. Offer coaching before performance drops.
* **Job Level (Senior/Mid)** : Attrition at senior levels is especially costly — focus on leadership engagement and career growth.
* **Marital Status (Single) :** Singles may have different lifestyle needs or flexibility — offer personalized retention approaches.
* **Remote Work (Yes)** : Having the option to work remotely is indicating less attrition.
* **Company reputation also matters**.

***Continued…***

**Attaching some key graphs**









