



## Project Overview

### Title: Predicting Employee Attrition at Salifort Motors

#### Business Problem & Objectives

Salifort Motors is experiencing a high employee turnover rate, which includes voluntary resignations and terminations. This turnover is costly due to the expenses associated with recruiting, training, and upskilling employees. Additionally, it negatively impacts corporate culture and employee morale. The company needs a way to predict employee attrition and understand the key factors contributing to turnover so that it can implement strategies to improve retention and job satisfaction.

#### Solution

To address high employee turnover, this project will develop predictive models to identify key factors contributing to attrition. Using machine learning techniques such as Logistic Regression, Decision Tree, Random Forest, and XGBoost, the analysis will help HR and leadership understand turnover drivers. The project will also provide business recommendations to improve retention, based on data-driven insights.

#### Stakeholders

- **HR:** Interested in actionable insights for retention strategies.
- **Senior Leadership:** Concerned with strategic decisions and high-level findings.
- **Data Team (Myself):** Responsible for execution, model development, and reporting.

## PACE Framework Breakdown

### Plan Stage: Understanding the Problem and Data Context

The primary goal is to identify key drivers of employee attrition and develop predictive models to forecast turnover risk. This involves analyzing employee survey data to determine factors that contribute most to turnover, predicting which employees are at risk of leaving, and formulating strategies to improve retention.

#### Scope & Boundaries:

- The analysis will focus on employee attributes such as department, tenure, number of projects, and working hours.
- External factors (e.g., economic conditions, job market trends) will not be included in this analysis.

**Key Questions to Address:**

- What factors most contribute to employee turnover?
- How accurately can we predict which employees are at risk of leaving?
- What strategies can HR implement to improve retention?

**Resources Required:**

- **Tools:** Python, Jupyter Notebook, Tableau, machine learning libraries (Scikit-Learn, XGBoost)
- **Data:** Employee survey dataset ([Kaggle Source](#))

**Risks & Mitigation Strategies:**

- **Data Bias:** Ensure diverse representation in the dataset and perform fairness analysis.
- **Overfitting Models:** Use cross-validation techniques and test on unseen data.
- **Stakeholder Misinterpretation:** Present findings in a business-friendly format, ensuring clarity.

## Analyse Stage: Exploratory Data Analysis (EDA) and Model Selection

The EDA process will involve summarizing key statistics of each variable, identifying missing values, and handling inconsistencies. Correlations between employee attributes and turnover will be visualized to extract meaningful insights.

**Data Quality Assessment:**

- **Completeness:** Assess missing values and impute where necessary.
- **Consistency:** Check for anomalies or inconsistencies in categorical variables.
- **Integrity:** Validate that all attributes align with expected business logic.

**Key Assumptions & Hypothesis:**

- High workloads and long hours may contribute to higher turnover.
- Some departments may have higher attrition rates than others.
- Employees with limited career growth opportunities might have a higher likelihood of leaving.

## Construct Stage: Model Development & Evaluation

The predictive modelling phase will involve Logistic Regression, Decision Tree, Random Forest, and XGBoost.

**Model Selection Criteria:**

- **Logistic Regression:** Chosen for baseline interpretability and benchmarking.
- **Decision Tree:** Provides clear decision paths for understanding turnover drivers.
- **Random Forest:** Enhances prediction reliability through ensemble learning.
- **XGBoost:** Optimized for high performance with boosting techniques.

**Feature Engineering & Selection:**

- Encode categorical variables (e.g., department).
- Scale numerical features for consistency in models that require it.
- Use feature importance analysis to refine inputs.

**Performance Metrics:**

- Accuracy, Precision, Recall, F1-score, and AUC-ROC will be used for model evaluation.
- Cross-validation will ensure robustness of results.

## Execute Stage: Model Interpretation & Business Impact

The analysis will assess hypotheses related to department, workload, and working hours to determine their influence on employee turnover.

**Business Recommendations:**

- Implement workload balancing strategies to prevent burnout.
- Strengthen professional development programs to enhance retention.
- Improve employee engagement and support initiatives.

**Final Deliverables**

1. **Cleaned Dataset** – A processed dataset ready for analysis and modelling.
2. **Tableau EDA Dashboard** – Interactive visualizations showcasing key trends.
3. **Predictive Models** – Four trained models (Logistic Regression, Decision Tree, Random Forest, XGBoost) with evaluation metrics.
4. **Feature Importance Analysis** – Identification of top turnover drivers.
5. **Risk & Mitigation Report** – A document outlining potential biases, risks, and mitigation strategies.
6. **Final Report & Business Recommendations** – A structured document for HR and leadership.

## **Conclusion**

This document provides a structured approach to planning, analyzing, constructing, and executing a data-driven strategy to predict employee turnover at Salifort Motors. It ensures alignment with business objectives, facilitates data-driven decision-making, and provides actionable insights to improve employee retention and reduce attrition-related costs.