**Proposed Project Name:** AttritionAI: Employee Churn Prediction

**Problem Statement:**

The problem at hand is to develop an accurate and reliable machine learning model to predict employee churn within our organization. By analyzing various factors such as employee demographics, job satisfaction, performance metrics, and historical retention data, etc. the goal is to forecast which employees are likely to leave the company in the near future.

The uniqueness of this project lies in its dual functionality – accurately predicting employee churn and providing HR with a ranked list of potential churners. Additionally, the model integrates predefined rules to recommend tailored retention strategies, offering a proactive approach to employee retention. This combination of predictive modeling, ranking mechanism, and rule-based recommendations distinguishes this project from conventional churn prediction approaches.

**Source(s) of Datasets:**

The IBM HR Analytics Attrition Dataset available on Kaggle is a comprehensive collection of employee data that provides valuable insights into factors influencing employee attrition. This dataset is particularly valuable for human resources professionals, data analysts, and data scientists aiming to understand the dynamics of employee attrition within organizations.

* <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/>

**Implementation Plan:**

* Data Collection: Gather comprehensive and relevant historical data on employees, including personal details, job-related metrics, and tenure information.
* Data Preprocessing: Cleanse, transform, and preprocess the data to make it suitable for machine learning algorithms. Handle missing values, outliers, and categorical variables appropriately.
* Feature Selection: Identify the most influential features that significantly contribute to predicting churn. This might include factors like job satisfaction, salary, work-life balance, and performance ratings.
* Model Development: Explore various machine learning algorithms such as logistic regression, random forests, or neural networks to build a predictive model. Evaluate and compare their performance using appropriate metrics like accuracy, precision, recall, and F1-score.
* Model Validation: Validate the chosen model using a separate test dataset to ensure its generalizability and reliability in real-world scenarios.
* Interpretation: Provide insights into the factors that contribute most to employee churn, helping HR and management teams make informed decisions and formulate retention strategies.

**Team Members and Task Allocation:**

**Group – 6**

**Ankit Vikas Agrawal**

**Rohan Sharma**

**Srini Narayanrao Vemuri**

The following tasks will be equally split between the team members:

* Literature review - Ankit
* Research - Rohan
* Individual EDA to analyze more insights (EDA and preprocessing would be done by each person again)
* Coding algorithm (1 Algorithm each person along with the implementation plan process)
* Project report write-up - Srini