# Analysis of Two-Pot using Logistic Regression

## 1. Introduction

This document details the analysis conducted on a dataset concerning employee attrition. The aim of this analysis is to predict employee turnover using various machine learning techniques, primarily focusing on Logistic Regression. This document outlines the steps taken during the analysis, including data preprocessing, model selection, evaluation, and handling encountered errors.

## 2. Data Preprocessing

Initially, the dataset was loaded, and the first step was to check for missing values. The command 'data.isnull().sum()' was used to identify any columns with missing values. It was found that there were no missing values in the dataset, as confirmed by the output:

Missing values in each column:  
Series([], dtype: int64)

Cleaned data shape: (14900, 24)

## 3. Defining Features and Target Variable

The features and target variable were defined as follows:  
Features (X): Age, Monthly Income, Years at Company, Job Satisfaction, Performance Rating  
Target Variable (y): Attrition (whether the employee stayed or left)

## 4. Encoding Categorical Variables

Categorical variables such as 'Job Satisfaction' and 'Performance Rating' were encoded into numerical values using 'LabelEncoder'. This step was crucial as machine learning models require numerical input.

## 5. Train-Test Split and Feature Scaling

The dataset was split into training and testing sets using 'train\_test\_split'. 80% of the data was used for training and 20% for testing. Subsequently, feature scaling was performed using 'StandardScaler' to standardize the features for better model performance.

## 6. Logistic Regression Model

A Logistic Regression model was initialized and fitted to the training data. The model's maximum number of iterations was set to 200 to prevent convergence warnings.

## 7. Model Evaluation

The model was evaluated using accuracy and confusion matrix. The accuracy score achieved was satisfactory, indicating a good predictive capability.

## 8. Errors Encountered and Solutions

Several errors were encountered during the analysis:

1. ModuleNotFoundError for sklearn: This was resolved by installing the scikit-learn library using 'pip install scikit-learn' in the command line.

2. ValueError during Feature Scaling: This error occurred due to the presence of non-numeric values in the feature set. It was resolved by encoding categorical variables before scaling the features.

3. Convergence Warning: A warning indicating that the model failed to converge was received. To address this, the maximum number of iterations was increased to 200, which allowed the model to converge properly.

## 9. Why Logistic regression Model.

Choosing Logistic Regression aligns with the goals of clarity, simplicity, and effective prediction for the employee attrition problem, making it a suitable choice for our analysis

## 10. Conclusion

This analysis aimed to use machine learning to anticipate how employees will act when taking money from their retirement fund and predict the future growth of the locked fund while also analyzing the sentiment of employee responses.. The process involved data cleaning, encoding, model fitting, and evaluation. Future work will focus on refining the model and exploring additional features.