The research question for the code will be addressed by initially identifying and comprehending the impacts of variables associated with employee attrition. This will be followed by the creation of a prediction model (classification model) that can reliably identify employees who might be at risk.

The model utilizes Decision Trees for variable importance identification and Naïve Bayes for the final classification. Factor Analysis was conducted for potential dimension reduction but was ultimately excluded from the final model due to its inferior performance compared to variable selection. Additionally, the code incorporates a data drift test using statistical analysis.

The model evaluation phase involves two crucial criteria. Firstly, it involves pre-determining the measure to qualify the model and using unseen data for final evaluation. In our case, we are focusing on predicting the likelihood of attrition with an accuracy target of over 70%. This choice stems from dealing with imbalanced data where model accuracy alone may not provide a comprehensive view. Therefore, we aim for high Sensitivity to predict as many positives as possible when treating "Yes" as the positive class. Only after meeting the criteria on unseen data can, we proceed to the next step.