```
*****************************************************************

GOAL : To predict employee attrition

TEAM MEMBERS:

1) Pooja Sastry
2) Sindhuri Punyamurthula

PROJECT FILE : FinalProject.ipynb

DATASET FILE : AttritionData.csv



*****************************************************************


I). DATASET DESCRIPTION

We used the dataset from Kaggle website :

https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-
attritiondataset/data

We renamed it to AttritionData.csv and it is located in the same
folder as our project notebook.

This dataset has 34 features and a target label 'Attrition'. Some of
the important features are: Employee_Age, Years_Of_Service, Gender,
Distance_From_Home, Job_Level, Current_Salary, Performance_Rating,
etc.


II). MODEL

The prediction of employee attrition requires a Binary
classification model which decides the Attrition variable as
'Yes/No'. We used 6
different models : Naive Bayes,Decision Tree,Logistic
Regression,Gradient Boosting ,Random Forest and Linear Support
Vector Classifier.
```

We evaluated these models and compared the accuracy and time taken by each of them.ConfisionMatrix was computed for all the models.We also performed cross-validation over a grid of parameters for a few models and evaluated them using the above methods.

III). Environment Setup

We used Jupyter notebook with Apache-Toree Scala kernel.
Name of the notebook file : FinalProject.ipynb

Note : In another notebook file Vizualizations.ipynb we used Python 3 kernel to show the bar graph comparing the accuracies and time taken by the models in one run of the FinalProject.ipynb. We used matplotlib for this.Anaconda and python 3 need to be installed for this.