# THE ATTRITION PROJECT:

THE CODE:

import pandas as pd

import numpy as np

from sklearn import tree

from sklearn import preprocessing

df = pd.read\_excel(r"C:\letsupgrage assignment\general data.xlsx", sheet\_name=0)

from sklearn.ensemble import RandomForestClassifier

label\_encoder = preprocessing.LabelEncoder()

df["BusinessTravel"] = label\_encoder.fit\_transform(df["BusinessTravel"])

df["Department"] = label\_encoder.fit\_transform(df["Department"])

df["EducationField"] = label\_encoder.fit\_transform(df["EducationField"])

df["Gender"] = label\_encoder.fit\_transform(df["Gender"])

df["JobRole"] = label\_encoder.fit\_transform(df["JobRole"])

df["MaritalStatus"] = label\_encoder.fit\_transform(df["MaritalStatus"])

df["Over18"] = label\_encoder.fit\_transform(df["Over18"])

rf\_model = RandomForestClassifier(n\_estimators=1000, max\_features=2, oob\_score=True)

features = ["Age", "BusinessTravel", "Department", "DistanceFromHome",

"Education", "EducationField", "EmployeeCount",

"EmployeeID", "Gender", "JobLevel", "JobRole", "MaritalStatus",

"MonthlyIncome", "NumCompaniesWorked", "Over18", "PercentSalaryHike",

"StandardHours", "StockOptionLevel", "TotalWorkingYears",

"TrainingTimesLastYear", "YearsAtCompany", "YearsSinceLastPromotion",

"YearsWithCurrMa0ger"]

rf\_model.fit(X=df[features], y=df["Attrition"])

print("\noob accuracy:")

print(rf\_model.oob\_score\_);

print("\n")

for features, imp in zip(features, rf\_model.feature\_importances\_):

print(features, imp);

predictors = pd.DataFrame([df["Age"], df["DistanceFromHome"], df["MonthlyIncome"],

df["PercentSalaryHike"], df["TotalWorkingYears"],

df["YearsAtCompany"]]).T

tree\_model = tree.DecisionTreeClassifier(max\_depth=12)

tree\_model.fit(X = predictors, y = df["Attrition"])

with open("Dtree1.dot", 'w')as f:

f = tree.export\_graphviz(tree\_model, feature\_names=["Age", "DistanceFromHome",

"MonthlyIncome", "PercentSalaryHike",

"TotalWorkingYears", "YearsAtCompany"], out\_file=f);

print("\nthe accuracy of the model")

print(tree\_model.score(X = predictors, y = df["Attrition"]))

THE OUTPUT:

oob accuracy:

0.9997732426303855

Age 0.09212878982411786

BusinessTravel 0.027569148686565282

Department 0.02484970663632998

DistanceFromHome 0.06636238593754325

Education 0.039604995753899544

EducationField 0.04091244215292928

EmployeeCount 0.0

EmployeeID 0.03780409804809866

Gender 0.016955586639173816

JobLevel 0.03670928372625833

JobRole 0.05357639099876977

MaritalStatus 0.03904678567836094

MonthlyIncome 0.08891810206572527

NumCompaniesWorked 0.052735793001979385

Over18 0.0

PercentSalaryHike 0.06246023283857052

StandardHours 0.0

StockOptionLevel 0.03251035581243175

TotalWorkingYears 0.08190592877828838

TrainingTimesLastYear 0.042988931322351216

YearsAtCompany 0.0681757537057532

YearsSinceLastPromotion 0.04154913246042474

YearsWithCurrMa0ger 0.05323615593242869

the accuracy of the model

0.945124716553288

In this project the important independent variable are the “AGE” , “DISTANCE FROM HOME”, “MONTHLY INCOME”, “TOTAL WORKING HOURS”, “YEARS AT COMPANY”. These are the most important independent variable of the project. This important variable is calculated by the code:

” for features, imp in zip(features, rf\_model.feature\_importances\_):

print(features, imp);”

By using this features we can calculate the accuracy of the project. By calculating the accuracy of this project is “0.945”. This accuracy is the highest form, this tells the project has a high accuracy rate.