DATA PREPARATIONS

```
#Import dataset
import pandas as pd
import numpy as np
df_train = pd.read_csv('train.csv')
df_test = pd.read_csv('test.csv')
```

```
#Data preparations 1- Clean data, eliminate NAs
df_train.isnull().sum()
```

```
df_test.isnull().sum()
```

```
employee_id                0
department                 0
region                     0
education               1034
gender                     0
recruitment_channel        0
no_of_trainings            0
age                        0
previous_year_rating    1812
length_of_service          0
KPIs_met >80%              0
awards_won?                0
avg_training_score         0
dtype: int64
```

```
#Replace numerical column's NAs with mean
train_mean_previous_year_rating = df_train['previous_year_rating'].mean()
test_mean_previous_year_rating = df_test['previous_year_rating'].mean()
df_train['previous_year_rating']= df_train['previous_year_rating'].fillna(train_mean_previous_year_rating,inplace=False)
df_test['previous_year_rating']= df_test['previous_year_rating'].fillna(test_mean_previous_year_rating,inplace=False)
#Drop catergorical column's NAs
df_train.dropna(axis=0,how='any',inplace=True)
df_test['education']= df_test['education'].fillna("Bachelor's",inplace=False)
#df_test.dropna(axis=0,how='any',inplace=True)
```

```
#Data preparations 2- Convert dummies
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
#Turn string variables into dummies for both test and train dataset
dummies_train = pd.get_dummies(df_train[['department','region','education','gender','recruitment_channel']])
dummies_test = pd.get_dummies(df_test[['department','region','education','gender','recruitment_channel']])
#Pull out the numerical columns in both test and train dataset
numerical_train = df_train.drop((['employee_id','department','region','education','gender','recruitment_channel']), axis='columns')
numerical_test = df_test.drop((['employee_id','department','region','education','gender','recruitment_channel']), axis='columns')
#Pull out the first column 'id' of both test and train dataset
id_train = df_train['employee_id']
id_test = df_test['employee_id']
#Concat the dataset in the same order
train = pd.concat([id_train, dummies_train, numerical_train], axis = "columns")
test = pd.concat([id_test, dummies_test, numerical_test], axis = "columns") # test will have 1 column less than train dataset
```

```
X = train.iloc[:,1:59]
y = train.iloc[:,59:60]
```

```
#Split train dataset into train & test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3, random_state=10)
```

MODEL COMPARISON

```
#Logistic Regression(Binary Classification)
from sklearn.linear_model import LogisticRegression
model_logistic = LogisticRegression()
model_logistic.fit(X_train,y_train)
model_logistic.predict(X_test)
```

```
model_logistic.score(X_test, y_test)
```

```
0.9173664122137405
```

```
#Support Vector Machine(SVM)
from sklearn.svm import SVC
```

```
model_svm = SVC()
model_svm.fit(X_train,y_train)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
SVC()
```

```
model_svm.score(X_test,y_test)
```

```
0.9115139949109414
```

```
#Decision Tree
from sklearn import tree
model_tree = tree.DecisionTreeClassifier()
model_tree.fit(X_train, y_train)
```

```
DecisionTreeClassifier()
```

```
model_tree.score(X_test, y_test)
```

```
0.9000636132315522
```

```
#Random Forest
from sklearn.ensemble import RandomForestClassifier
model_rf = RandomForestClassifier()
model_rf.fit(X_train,y_train)
```

```
<ipython-input-14-50c68e15e7a8>:4: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change
  model_rf.fit(X_train,y_train)
RandomForestClassifier()
```

```
model_rf.score(X_test, y_test)
```

```
0.928117048346056
```

```
X_array = X.values
y_array = y.values
```

```
#k-fold cross validation
from sklearn.model_selection import StratifiedKFold #this method can split data in different classification uniformly
folds = StratifiedKFold(n_splits=10)

logistic_list = []
svm_list = []
tree_list = []
rf_list = []

def get_score(model, X_train, X_test, y_train, y_test):
    model.fit(X_train, y_train)
    return model.score(X_test, y_test)



for (train_index, test_index) in folds.split(X_array,y_array):
    X_train, X_test = X_array[train_index], X_array[test_index]
    y_train, y_test = y_array[train_index], y_array[test_index]

    logistic_list.append(get_score(LogisticRegression(solver='liblinear'), X_train, X_test, y_train, y_test))
    svm_list.append(get_score(SVC(), X_train, X_test, y_train, y_test))
    tree_list.append(get_score(tree.DecisionTreeClassifier(), X_train, X_test, y_train, y_test))
    rf_list.append(get_score(RandomForestClassifier(), X_train, X_test, y_train, y_test))
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversi
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversi
  y = column_or_1d(y, warn=True)
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
<ipython-input-17-089d6ac7d386> in <module>
```

```
logistic_list
```

```
[0.9305343511450381,
 0.933206106870229,
 0.9290076335877863,
 0.9297709923664123,
 0.9299618320610687,
 0.9288167938931298,
 0.9330152671755725,
 0.9314885496183206,
 0.9318702290076336,
 0.9272761977476618]
```

```
    314             self.fit_status_ .
```

```
svm_list
```

```
[0.9133587786259542,
 0.9133587786259542,
 0.9133587786259542,
 0.9131679389312977,
 0.9131679389312977,
 0.9131679389312977,
 0.9131679389312977,
 0.9131679389312977,
 0.9131679389312977,
 0.9133422408856652]
```

```
tree_list
```

```
[0.8965648854961832,
 0.8980916030534352,
 0.8982824427480917,
 0.8923664122137405,
 0.8969465648854962,
 0.8977099236641222,
 0.900381679389313,
 0.8977099236641222,
 0.8958015267175573,
 0.8906279824393968]
```

```
rf_list
```

```
[0.9301526717557251,
 0.933969465648855,
 0.9301526717557251,
 0.9333969465648855,
 0.9337786259541985,
 0.9345419847328245,
 0.9320610687022901,
 0.9353053435114503,
 0.9297709923664123,
 0.9322389769039893]
```

```
#Cross-validation using a package in Python
from sklearn.model_selection import cross_val_score
cross_val_score(LogisticRegression(solver='liblinear'), X, y,cv=10)
cross_val_score(SVC(), X, y, cv=10)
cross_val_score(tree.DecisionTreeClassifier(), X, y, cv=10)
cross_val_score(RandomForestClassifier(), X, y, cv=10)
```

```
score_1 = cross_val_score(LogisticRegression(solver='liblinear'), X, y,cv=10)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
```

```
average(score_1)
```

```
0.9283559026749441
```

```
score_2 = cross_val_score(SVC(), X, y, cv=10)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
```

```
average(score_2)
```

```
0.9133416258387701
```

```
score_3 = cross_val_score(tree.DecisionTreeClassifier(), X, y, cv=10)
```

```
average(score_3)
```

```
0.8971610391393094
```

```
score_4 = cross_val_score(RandomForestClassifier(), X, y, cv=10)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
```

```
average(score_4)
```

```
0.9270526536487494
```

```
#Hyper parameter Tuning
from sklearn.model_selection import GridSearchCV
```

```
clf1 = GridSearchCV(LogisticRegression(solver='liblinear'), {
    'C':[1,5,10]}, cv=10)
clf1.fit(X_train, y_train)
clf1.best_params_
```

```
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
{'C': 10}
```

```
clf1.best_score_
```

```
0.9349657198824681
```

```
clf2 = GridSearchCV(SVC(gamma='auto'), {
    'C':[1,10,20], 'kernel':['rbf','linear']}, cv=10)
clf2.fit(X_train, y_train)
clf2.best_params_
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
```

```
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
 /usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
   y = column_or_1d(y, warn=True)
```

```
clf2.best_score_
```

```
0.9405484818805092
```

```
clf3 = GridSearchCV(tree.DecisionTreeClassifier(), {
    'max_depth':[1,2,3,4,5,6,7,8,9,10,15,20]}, cv=10)
clf3.fit(X_train, y_train)
clf3.best_params_
```

```
{'max_depth': 10}
```

```
clf3.best_score_
```

```
0.9309500489715964
```

```
clf4 = GridSearchCV(RandomForestClassifier(),{
    'n_estimators':[1,5,10]}, cv=10)
clf4.fit(X_train, y_train)
clf4.best_params_
```

```
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
   estimator.fit(X_train, y_train, **fit_params)
 /usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_validation.py:680: DataConversionWarning: A column-vector y was pa
  estimator.fit(X_train, y_train, **fit_params)
/usr/local/lib/python3.8/dist-packages/sklearn/model_selection/_search.py:926: DataConversionWarning: A column-vector y was passed
  self.best_estimator_.fit(X, y, **fit_params)
{'n_estimators': 10}
```

```
clf4.best_score_
```

```
0.924975514201763
```

```
model_svm_best = SVC(C=10, kernel='rbf')
```

```
#Prediction for test data
model_svm_best.fit(X_train,y_train)
y_predicted = model_svm_best.predict(test.iloc[:,1:59])
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
```

```
df_prediction = pd.DataFrame(y_predicted)
df_prediction.columns = ['is_promoted']
df_id = test['employee_id']
id = df_id.values.tolist()
df_id_final = pd.DataFrame(id)
df_id_final.columns = ['employee_id']
df_id_final
```

	employee_id
0	8724
1	74430
2	72255
3	38562
4	64486
...	...
23485	53478
23486	25600
23487	45409
23488	1186
23489	5973

23490 rows × 1 columns

```
df_id_final.to_csv('Submission_final.csv',index=False)
```

```
df_prediction.to_csv('Prediciton_final.csv',index=False)
```