DETAILS FOR THE PROJECT OF HR ANALYTICS.

**Group Information:**

Group ID (GID): **CN8iR5JuY**  
Group Lead : **shimlee ghosh**  
Selected Project: **HR Analytics (Case Study)**

**ABOUT THE PROJECT:**

An HR analytics firm to understand what factors they should focus on, in order to curb attrition

I have used various models and techniques listed below and get inference from it.

1. Created feature OVER\_Time from 2 datasets that is in\_time and out\_time.

2.DATA PREPROCESSING-Removed every outliers,null values and imputed it.

3.EXPLORATORY DATA ANALYSIS:UNIVARIATE AND BI-VARIATE.

4.Performed feature enginnering

5. Selection of variables done with the help of Random forest and correlation top10 features are selected . and few features I also dropped because of their very high correlation.

6.Then normalise the dataset aswell.

7.In train and test dataset I applied stratified k-fold cross validation as it was an imbalanced dataset.

8.In building model I used: Below models have been evaluated:  
· Gaussian Naive Bayes  
· Bernoulli Naive Bayes  
· Logistic Regression  
· K Nearest Neighbour  
· Decision Tree Classifier  
· Random Forest Classifier  
· Support Vector Classification (SVC)  
· Linear SVC

9.This two-class dataset is imbalanced. As a result, there is a possibility that the model built might be biased towards to the majority and over-represented class. After applying Synthetic Minority Oversampling Technique (SMOTE) to over-sample the minority class, some improvement in both F1-score & Recall is observed.

10.To improve futher I applied hypertuning parameter to modelling with the help of random search.and gave good results to Random Forest.

11. This are the top 10 features(in decreasing order) selected by feature selection:

1.YearsAtCompany-we can infer that total working years of 1 year tends to switch job mostly.

2.Age-so the mean age of people leaving is 33 and mostly leaving are the younger people 20-40 specially age of 33.

3.CurrMgr\_Coy-negative correlated,tends to leave at initial years of working years.

4.OverTime-It is a key factor may be because of monetary issue.

5.CurrRole\_CurrMgr-negative correlated,leaves at starting years only if manager is not good for the employee.

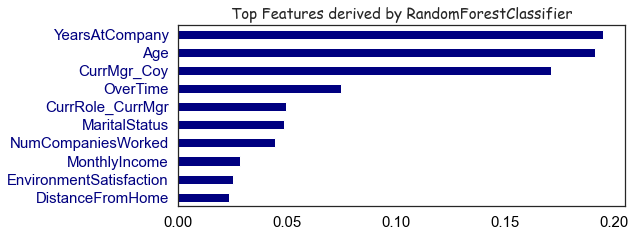
6.MaritalStatus-we can infer that singles tends to leave the job more often i.e 27.5% approx.

7.NumCompaniesWorked- people tends to leave job who have more or equal to 5 number of companies worked.

8.MonthlyIncome- monetary factor for the employee.

9.EnvironmentSatisfaction- level lower than 1 tends to leave the job so one should focus on environment of the employee.

10.DistanceFromHome-Travel can be one factor.



11.We see Random forest have highest F1 score And Recall after feature engeneering and SMOTE and after I did hypertuning parameter.

