Machine Learning

Analysis of Supervised and Unsupervised Data

NEETHU SIDHARDHAN

8/9/2020

Case Study 1

You are hired by one of the leading news channel CNBE who wants to analyse recent elections. This survey was conducted on 1525 voters with 9 variables. You have to build a model, to predict which party a voter will vote for based on the given information, to create an exit poll that will help in predicting overall win and seats covered by a particular party.

Dataset for Problem: Election_Data.xlsx

Data Ingestion: 12 marks

1. Read the dataset. Do the descriptive statistics and do null value condition check. Write an inference on it. (5 Marks)
2. Perform Univariate and Bivariate Analysis. Do exploratory data analysis. Check for Outliers. (7 Marks)

Data Preparation: 5 marks

1. Encode the data (having string values) for Modelling. Is Scaling necessary here or not? Data Split: Split the data into train and test (70:30). (5 Marks)

Modelling: 26 marks

1. Apply Logistic Regression and LDA (linear discriminant analysis). (5 marks)
2. Apply KNN Model, Naïve Bayes Model and support vector machine (SVM) model. Interpret the results. (7 marks)
3. Model Tuning, Bagging (Random Forest should be applied for Bagging) and Boosting. (7 marks)
4. Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model. Final Model: Compare the models and write inference which model is best/optimized. (7 marks)

Inference: 5 marks

1. Based on these predictions, what are the insights? (5 marks)

Data Ingestion: (1) Read the dataset. Do the descriptive statistics and do null value condition check. Write an inference on it

Load the Libraries:

We have to load the relevant liabraries that are required for us to analyse the casestudy.

```
import pandas as pd
import numpy as np
from sklearn import metrics
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
#from sklearn.feature_extraction.text import CountVectorizer  #DT does not take strings as input for the model fit step....
from IPython.display import Image
#import pydotplus as pydot
from sklearn import tree
from os import system
import warnings
warnings.filterwarnings("ignore")
```

We require the libraries such as numpy, pandas, matplotlib and seaborn for reading and preceding the EDA and other basis analysis of our dataset

The dataset shared to us is in a spread sheet. Hence we will be using the option "pd.read_excel" to read the data.

```
df=pd.read_excel("Election_Data.xlsx", sheet_name='Election_Dataset_Two Classes')
```

The loaded dataset will be as below:

	Unnamed: 0	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gender
0	1	Labour	43	3	3	4	1	2	2	female
1	2	Labour	36	4	4	4	4	5	2	male
2	3	Labour	35	4	4	5	2	3	2	male
3	4	Labour	24	4	2	2	1	4	0	female
4	5	Labour	41	2	2	1	1	6	2	male

The dataset is having the details like the "Vote" related to the parties involved in the election. There are two parties involved in election that are Labour and Conservative party. The age of the voters are also shared to us. Then we have data related to the assessment of current economic condition and current household economic condition given in a five-point scale. The exit poll that we are predicting here is about Tony Blair and William Hague that happened on 1997. Tony Blair is leader of Labour party and William Hague is leader of Conservative party. We have a five-point scale of evaluation on the leaders in the dataset. Europe mentioned in the dataset is a 11 point scaled that are measured on the attitude of European integrations. Knowledge of parties' position also shared in three-point scale. We have the gender of the voters shared in the dataset.

- The dataset is having 1525 rows and 9 columns
- The dataset consist of objects and numeric values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1525 entries, 0 to 1524
Data columns (total 9 columns):
vote                       1525 non-null object
age                        1525 non-null int64
economic.cond.national     1525 non-null int64
economic.cond.household    1525 non-null int64
Blair                      1525 non-null int64
Hague                      1525 non-null int64
Europe                     1525 non-null int64
political.knowledge        1525 non-null int64
gender                     1525 non-null object
dtypes: int64(7), object(2)
memory usage: 107.4+ KB
```

- The description of the dataset is as below:

	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge
count	1525.000000	1525.000000	1525.000000	1525.000000	1525.000000	1525.000000	1525.000000
mean	54.182295	3.245902	3.140328	3.334426	2.746885	6.728525	1.542295
std	15.711209	0.880969	0.929951	1.174824	1.230703	3.297538	1.083315
min	24.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
25%	41.000000	3.000000	3.000000	2.000000	2.000000	4.000000	0.000000
50%	53.000000	3.000000	3.000000	4.000000	2.000000	6.000000	2.000000
75%	67.000000	4.000000	4.000000	4.000000	4.000000	10.000000	2.000000
max	93.000000	5.000000	5.000000	5.000000	5.000000	11.000000	3.000000

- The dataset is not standardised hence we have to standardise the data before processing the train and split of the dataset.
- The dataset is not having any missing values.

```
vote                       0
age                        0
economic.cond.national     0
economic.cond.household    0
Blair                      0
Hague                      0
Europe                     0
political.knowledge        0
gender                     0
dtype: int64
```

- The dataset is having "Vote" and "Gender" as categorical variable and rest all columns as numeric variable.

- The dataset was having duplicate of 8 rows. Those rows were removed from the dataset for further analysis.

```
Number of rows before discarding duplicates = 1525
Number of rows after discarding duplicates = 1517
```

- We have to convert the dataset from categorical to numerical. Then the descriptive analysis of the dataset will be as below

	count	mean	std	min	25%	50%	75%	max
vote	1517.0	0.696770	0.459805	0.0	0.0	1.0	1.0	1.0
age	1517.0	54.241266	15.701741	24.0	41.0	53.0	67.0	93.0
economic.cond.national	1517.0	3.245221	0.881792	1.0	3.0	3.0	4.0	5.0
economic.cond.household	1517.0	3.137772	0.931069	1.0	3.0	3.0	4.0	5.0
Blair	1517.0	3.335531	1.174772	1.0	2.0	4.0	4.0	5.0
Hague	1517.0	2.749506	1.232479	1.0	2.0	2.0	4.0	5.0
Europe	1517.0	6.740277	3.299043	1.0	4.0	6.0	10.0	11.0
political.knowledge	1517.0	1.540541	1.084417	0.0	0.0	2.0	2.0	3.0
gender	1517.0	0.467370	0.499099	0.0	0.0	0.0	1.0	1.0

After removing the duplicates, we have 1517 rows and 9 rows. The age of the voters are showing mean average of 54.24. The minimum age of voter is 24 and maximum is 93. The whole dataset we will be having only age and Europe as the one variable that is shows the dataset requires scaling. The rest of the data are standardised with a five-point scale, hence their standard deviation is under 1.

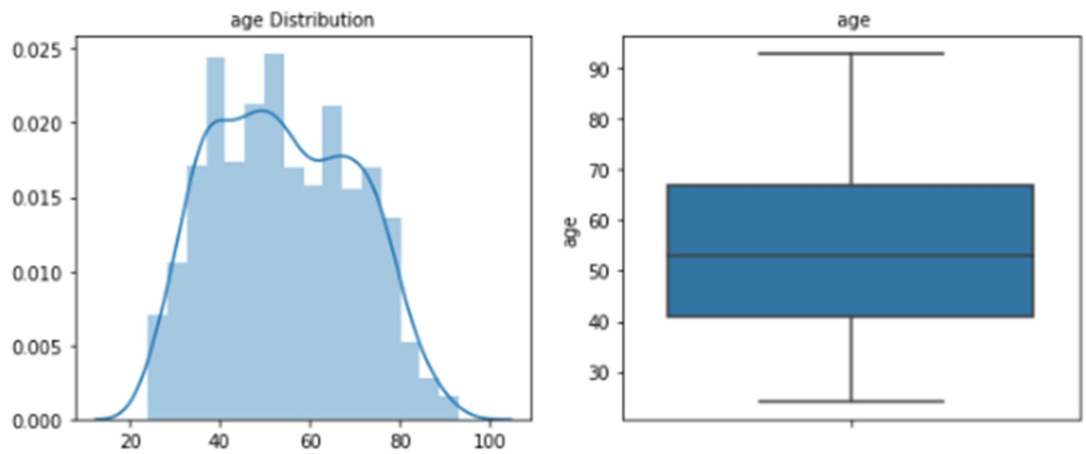
- Interference: We have to consider the target variable as "Vote" as we are predicting the exit poll for which party is getting through the election.

Data Ingestion: (2) Perform Univariate and Bivariate Analysis. Do exploratory data analysis. Check for Outliers

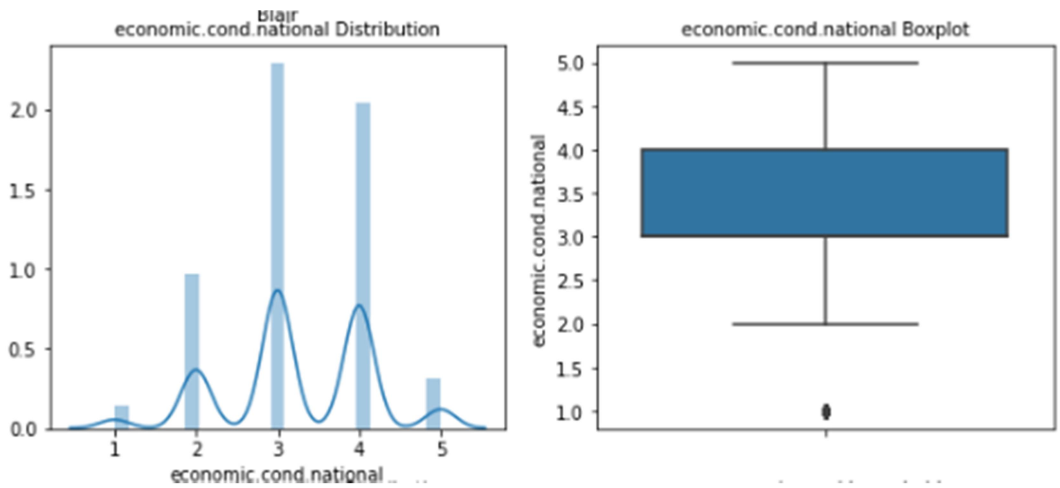
Perform Univariate Analysis and Bivariate Analysis:

Univariate Analysis: This type of data consists of only one variable. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it.

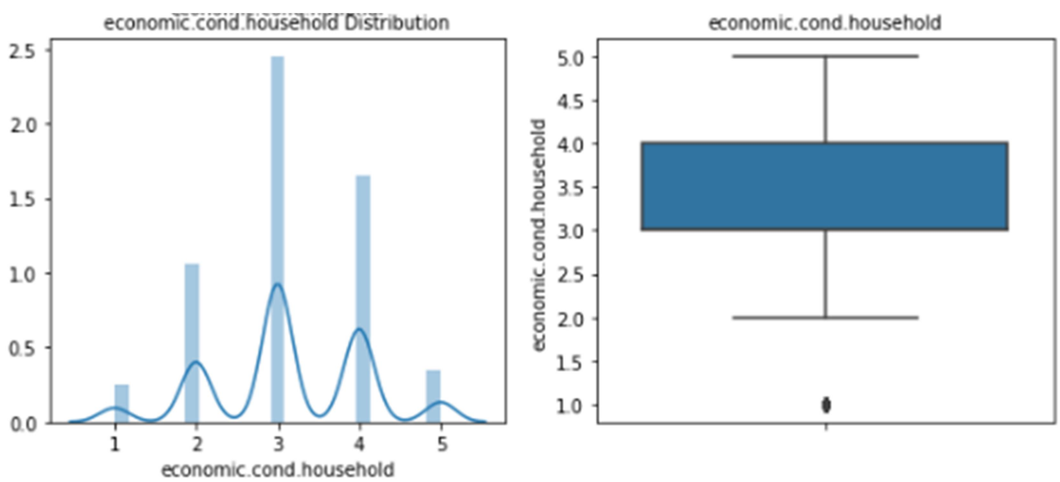
Age: - The distribution for age shows it is normally distributed and there is no outliers present in the dataset.



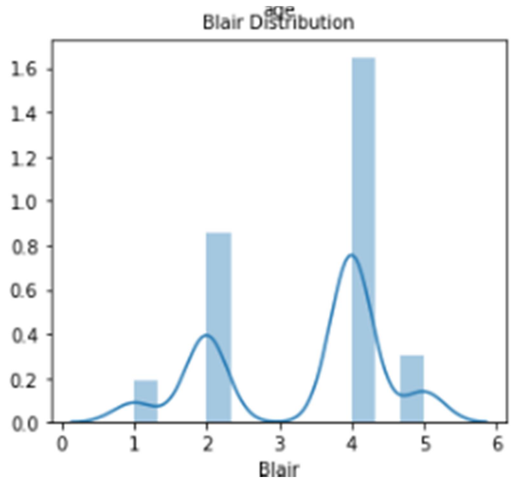
Economic Condition National Distribution:- The data is a five-point scale dataset which is not normally distributed and has outlier present in the dataset.

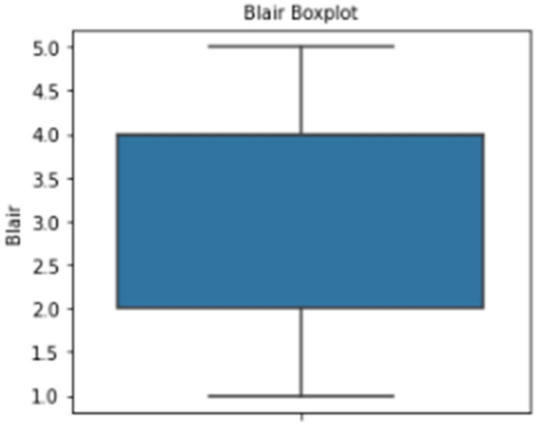


Economic Condition Household:-

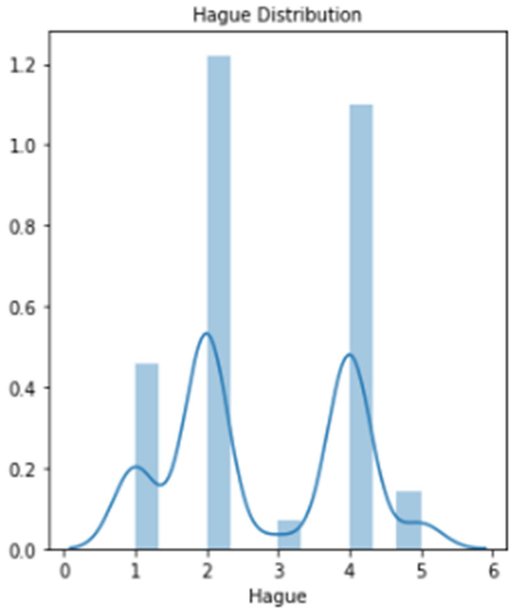


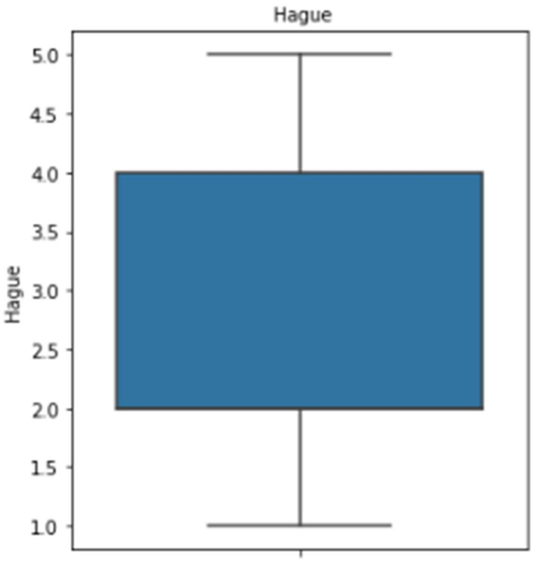
Blair:-The distribution is based on five-point scale where there is no values present in under third scale. There are no outliers present in the dataset.



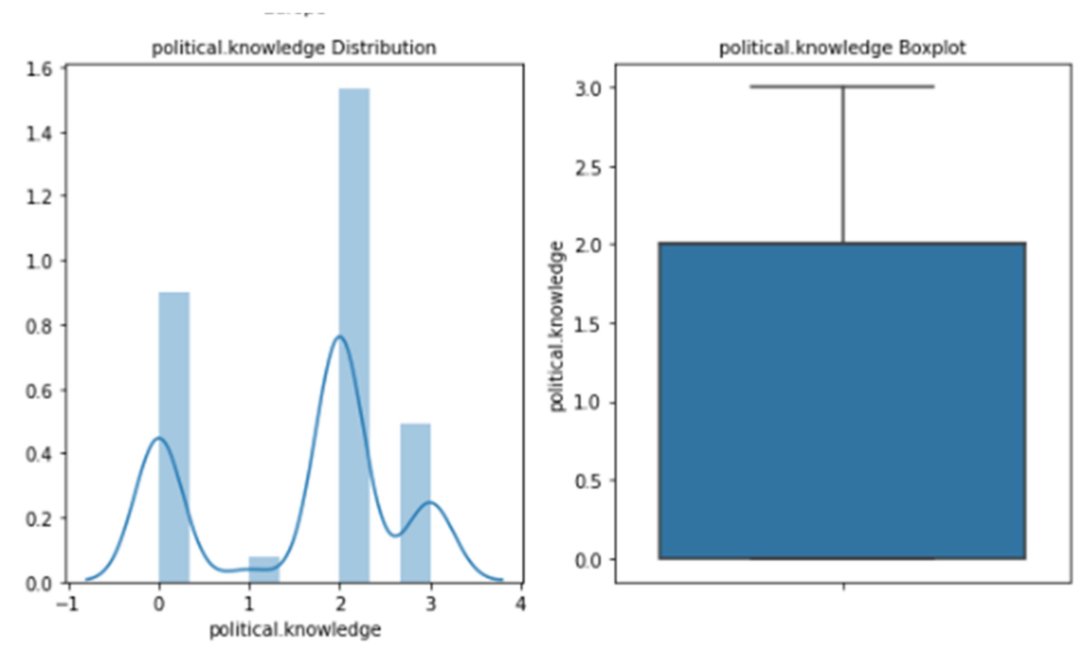


Hague: - The assessment of conservative leader is rated in a five-point scale and they are not evenly distributed. There are no outliers present in the dataset.





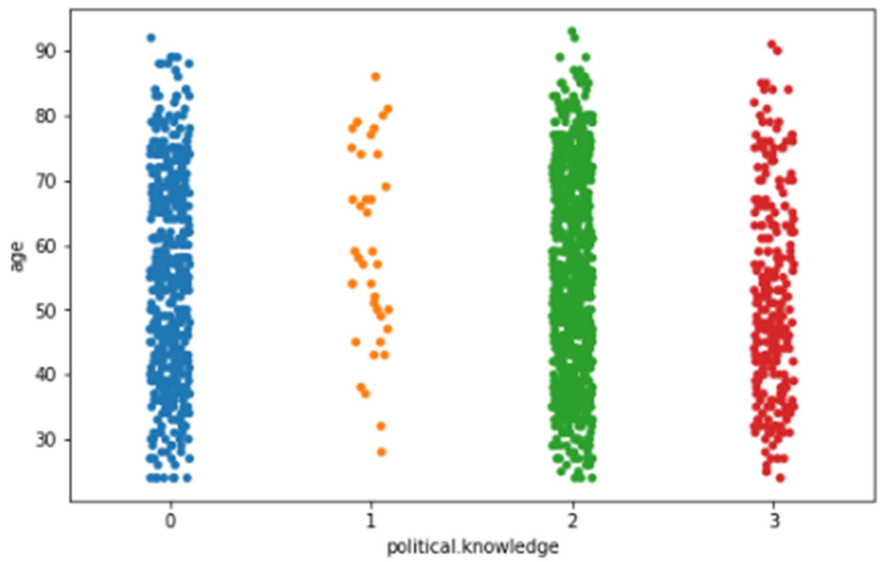
Political knowledge:



Bivariate Analysis: This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables.

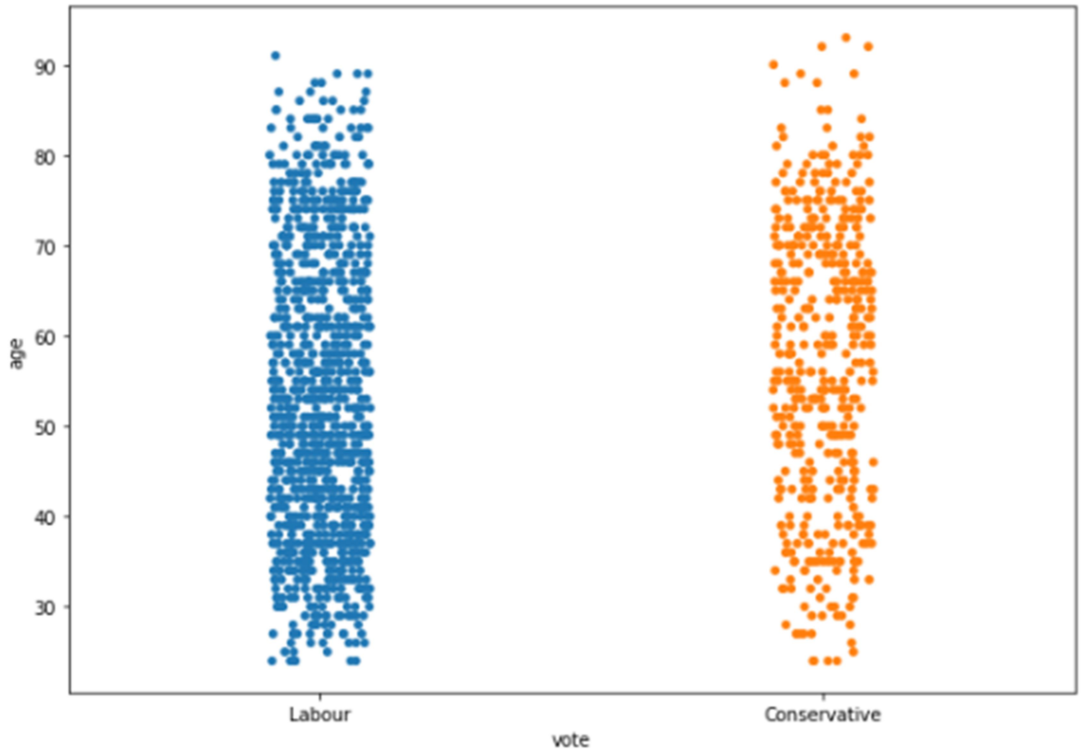
I. Political Knowledge Vs. Age

The political knowledge over the European integrations was analysed as three levels. The data set has scaled it from 0 to 3 levels. The minimum age of a voter is 24 years as per the EDA and maximum age is 90 years. Most of the voters preference is for the level 2 as compared other levels.



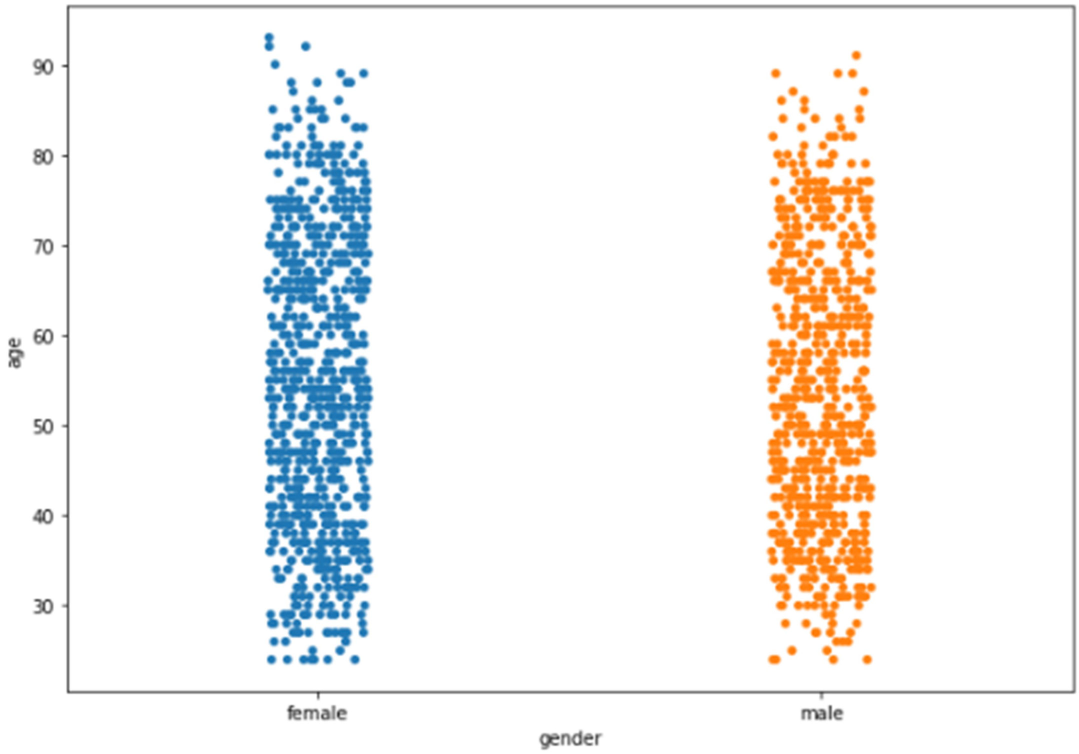
II. Votes Vs. Age:

There are two active parties involved in the election. They are Labour party leas by Tony Blair and Conservative party lead by William Hague. Majority of the voters prefer to vote for labour party

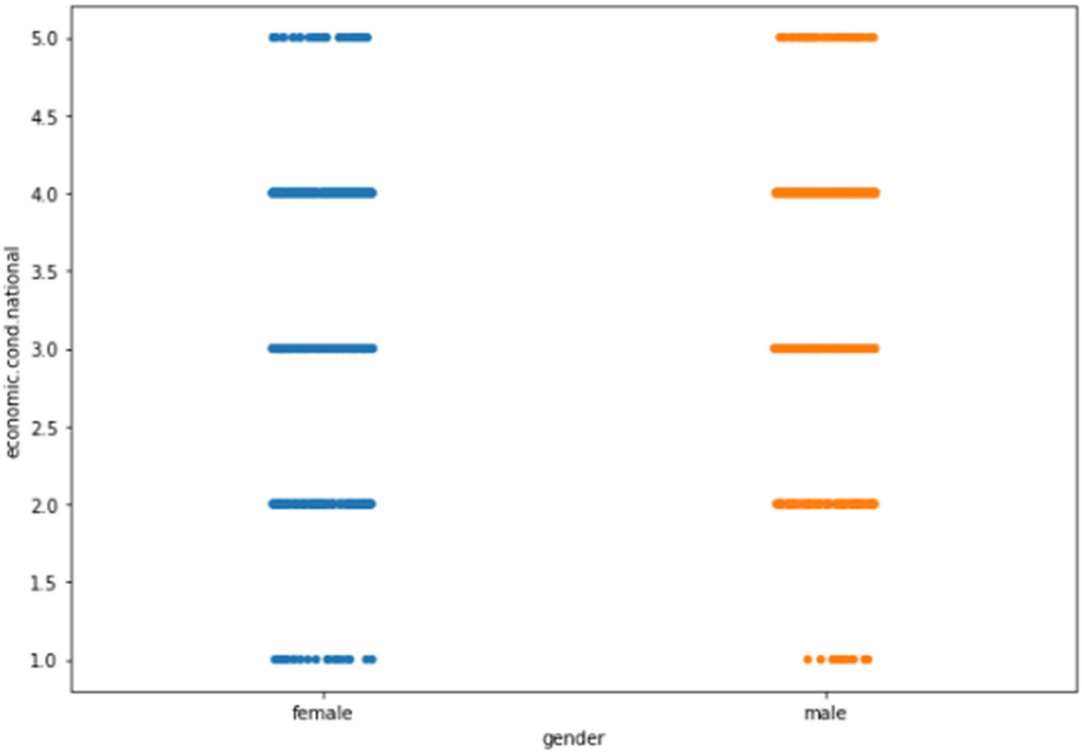


III. Gender vs. Age

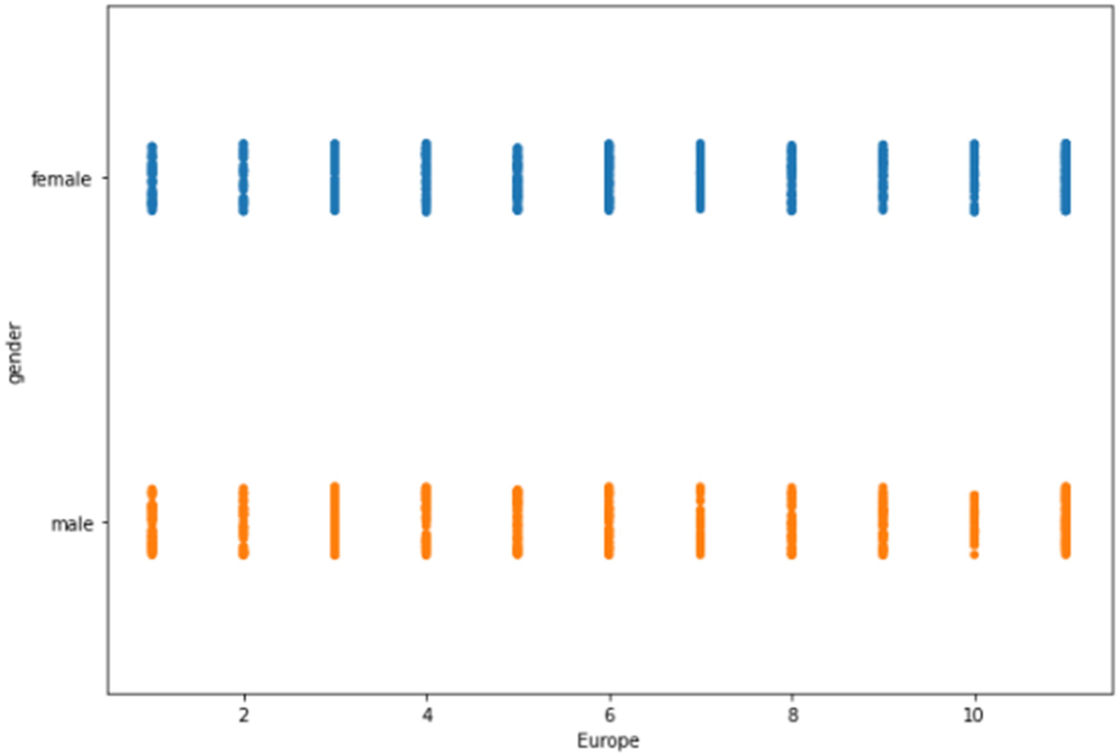
For the election dataset, the female count is more than the male count here.



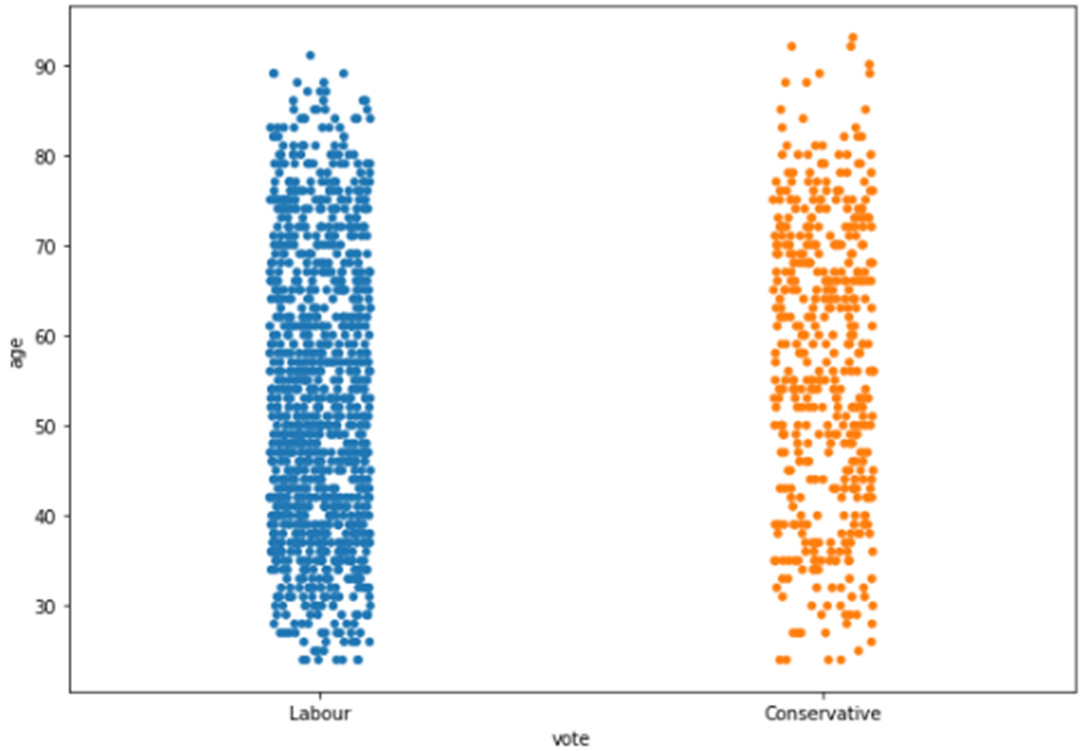
IV. Gender vs. Economic condition national



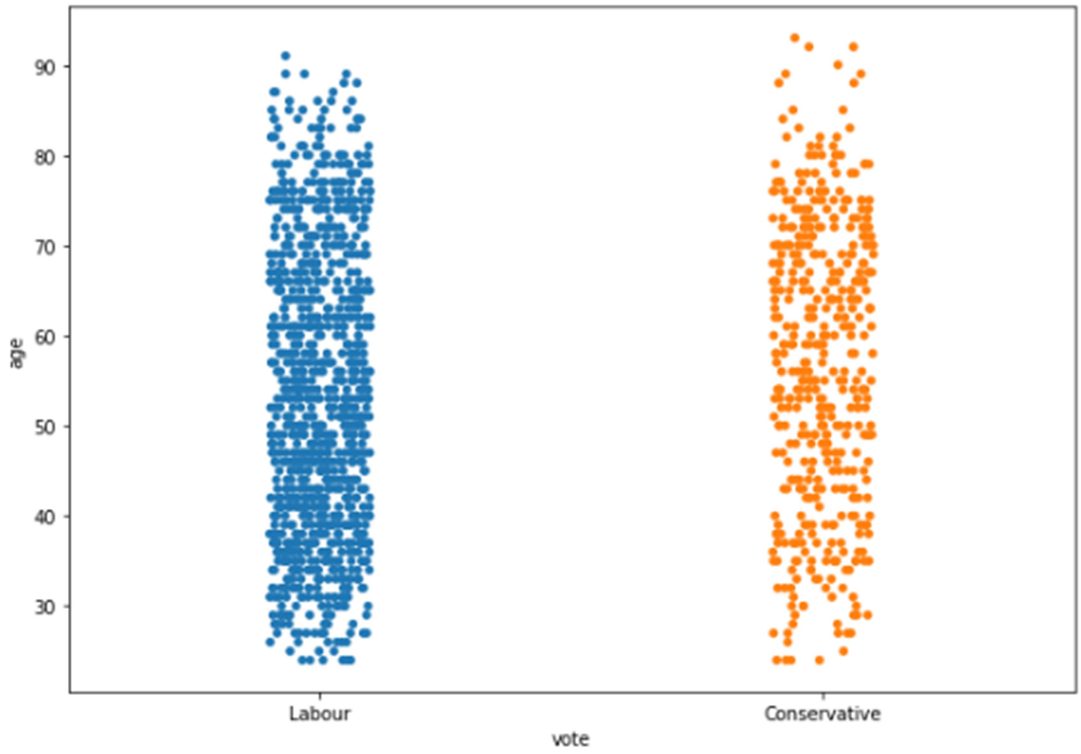
V. Europe vs. Gender



VI. Vote vs. Age (Hue = Hague)



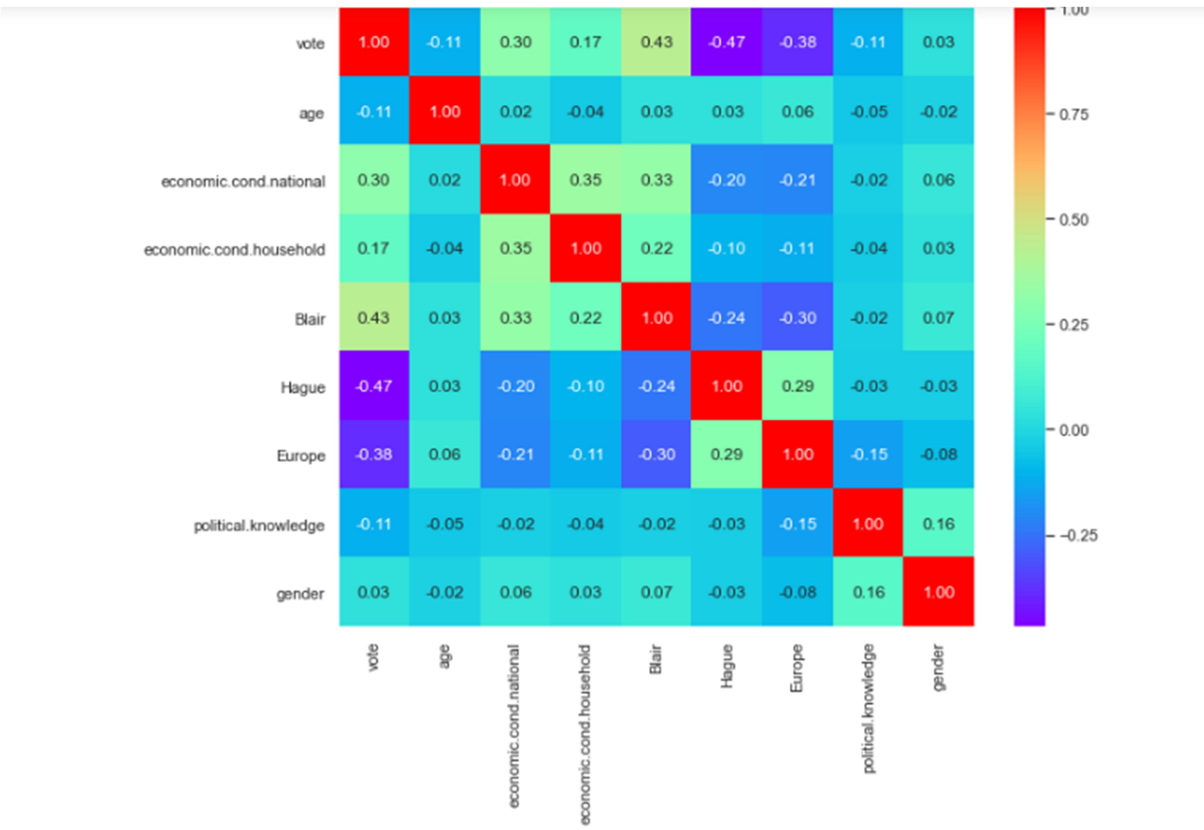
VII. Vote vs. Age (Hue= Blair)



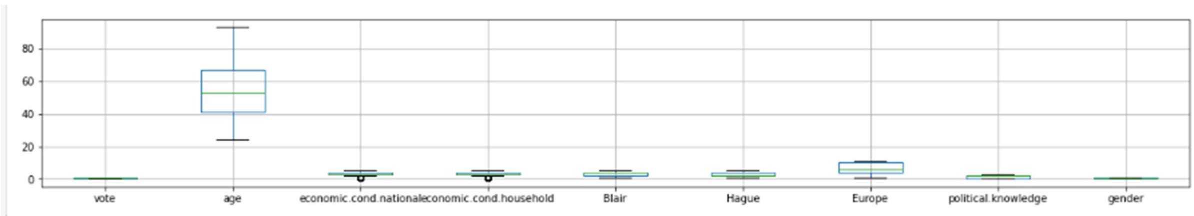
Heat Map:

The heat map is related to the correlation of the dataset. Deeper the colours in the heat map indicated the stronger the correlation between the data and lighter the colours of the lesser correlation in the data.

- The correlation between the voting party and the age of the votes are negatively correlated.
- The correlation between the vote and Blaire and Hague are also negatively correlated
- The age and economic condition nation is also negatively correlated.
- The factor "Political Knowledge" are negatively correlated with the whole data.
- The economic condition and the Blair are positively correlated.



Check for Outliers:



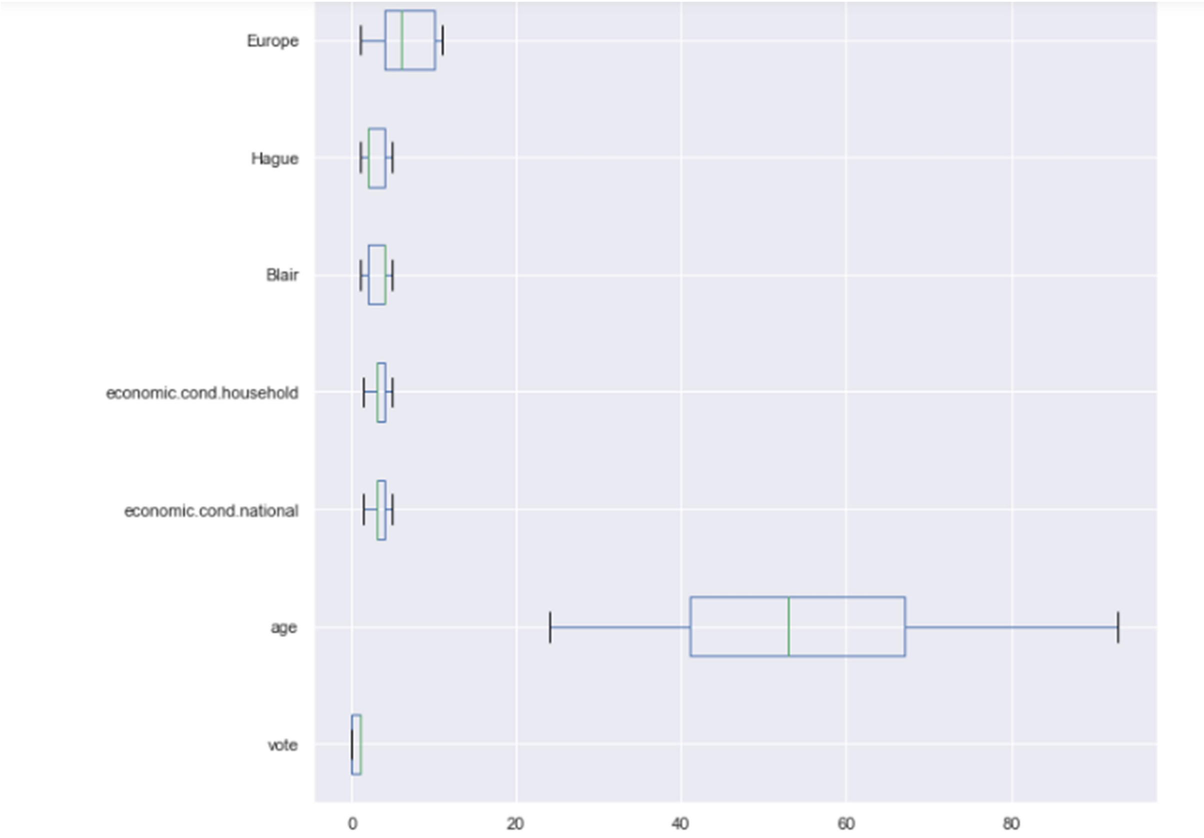
There are outliers present in the dataset for the factor "economic condition national" and "economic condition household" that need to be treated. Rest all variables are not having outliers.

Treatment of Outliers:

```
def remove_outlier(col):
    sorted(col)
    Q1,Q3=np.percentile(col,[25,75])
    IQR=Q3-Q1
    lower_range= Q1-(1.5 * IQR)
    upper_range= Q3+(1.5 * IQR)
    return lower_range, upper_range
```

```
for column in df.iloc[:, 2:4].columns:
    lr,ur=remove_outlier(df[column])
    df[column]=np.where(df[column]>ur,ur,df[column])
    df[column]=np.where(df[column]<lr,lr,df[column])
```

Now there are no outliers present in the dataset once the treatment of the dataset is done.



EDA:

The description of the data once after processing the treatment of outliers will be as below:

	count	mean	std	min	25%	50%	75%	max
vote	1517.0	0.696770	0.459805	0.0	0.0	1.0	1.0	1.0
age	1517.0	54.241266	15.701741	24.0	41.0	53.0	67.0	93.0
economic.cond.national	1517.0	3.257416	0.853647	1.5	3.0	3.0	4.0	5.0
economic.cond.household	1517.0	3.159196	0.886279	1.5	3.0	3.0	4.0	5.0
Blair	1517.0	3.335531	1.174772	1.0	2.0	4.0	4.0	5.0
Hague	1517.0	2.749506	1.232479	1.0	2.0	2.0	4.0	5.0
Europe	1517.0	6.740277	3.299043	1.0	4.0	6.0	10.0	11.0
political.knowledge	1517.0	1.540541	1.084417	0.0	0.0	2.0	2.0	3.0
gender_1	1517.0	0.467370	0.499099	0.0	0.0	0.0	1.0	1.0

(3) Encode the data (having string values) for Modelling. Is Scaling necessary here or not? Data Split: Split the data into train and test (70:30).

Data Encoding:

Here we use the target variable for predicting the exit polling as “vote”. The data encoding are processed by adding dummy variables in the dataset. There is a categorical variable “Gender” which we need to code it as dummy variable. Even we can use the dummy variable for “Blaire” and “Hague” in the dataset in order to get better prediction.

Initially the data is first read out by using only one dummy variable as “Gender”

```
df = pd.get_dummies(df, columns=['gender'],drop_first=True)
```

	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gender_1
0	1	43	3	3	4	1	2	2	0
1	1	36	4	4	4	4	5	2	1
2	1	35	4	4	5	2	3	2	1
3	1	24	4	2	2	1	4	0	0
4	1	41	2	2	1	1	6	2	1

Scaling the dataset:

The dataset is not evenly formed as the variable Age is showing high values and the rest all data are on smaller variables, which are updated on a 3-point scale, 5-point scale and 11-point scale. Hence, the data should be scaled before proceeding further.

To scale the data we use the function:

```
df[num1] = df[num1].apply(lambda x:(x-x.min()) / (x.max()-x.min()))
```

	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gender_1
0	1	0.275362	0.428571	0.428571	0.75	0.00	0.1	0.666667	0
1	1	0.173913	0.714286	0.714286	0.75	0.75	0.4	0.666667	1
2	1	0.159420	0.714286	0.714286	1.00	0.25	0.2	0.666667	1
3	1	0.000000	0.714286	0.142857	0.25	0.00	0.3	0.000000	0
4	1	0.246377	0.142857	0.142857	0.00	0.00	0.5	0.666667	1

Now the data is scaled and is even in shape.

Data Split: Split the data into train and test (70:30)

The target variable is taken as "Vote" to predict which party will be winning the election.

To split the data we use the package from sklearn_train and split package

```
# Copy all the predictor variables into X dataframe
X = df.drop('vote', axis=1)

# Copy target into the y dataframe.
y = df['vote']
```

```
# Split X and y into training and test set in 75:25 ratio
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30 , random_state=1)
```

Modelling

(4) Apply Logistic Regression and LDA (linear discriminant analysis).

Logistic Regression

Import the library from sklearn for Logistic Regression.

```
from sklearn.linear_model import LogisticRegression
```

```
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
y_predict = model.predict(X_train)
model_score = model.score(X_train, y_train)
print(model_score)
print(metrics.confusion_matrix(y_train, y_predict))
print(metrics.classification_report(y_train, y_predict))
```

The model results for the train are as below:

```
0.8369462770970783
[[193 114]
 [ 59 695]]
              precision    recall  f1-score   support

           0       0.77      0.63      0.69       307
           1       0.86      0.92      0.89       754

    accuracy                           0.84      1061
   macro avg       0.81      0.78      0.79      1061
weighted avg       0.83      0.84      0.83      1061
```

- The Accuracy of the model is 84%
- The F1 score for the model is 86%
- The precision of the train data is 92%
- The recall for the train data is 86%

The model results for the logistic regression for the test data is as below:

```
0.8289473684210527
[[110  43]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.76      0.72      0.74       153
           1       0.86      0.88      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.81       456
weighted avg       0.83      0.83      0.83       456
```

- The Accuracy of the test model is 83%
- The precision for the test model is 88%
- The recall for the test model is 87%
- The F1 score for the test data is 86%

Linear Discriminant Analysis [LDA]

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. As the name implies dimensionality, reduction techniques reduce the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible.

The training data:

```
0.8341187558906692
[[200 107]
 [ 69 685]]
              precision    recall  f1-score   support

           0       0.74      0.65      0.69       307
           1       0.86      0.91      0.89       754

    accuracy                           0.83      1061
   macro avg       0.80      0.78      0.79      1061
weighted avg       0.83      0.83      0.83      1061
```

- The accuracy of the model is 83%
- The precision of the model is 91%

- The recall of the model is 89%
- The F1 score of the model is 86%

The LDA analysis for the test model is as below:

```
0.831140350877193
[[111  42]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.76      0.73      0.74       153
           1       0.86      0.88      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.81       456
weighted avg       0.83      0.83      0.83       456
```

- The Accuracy for the test data is 83%
- The F1 score for test data is 85%
- The recall for the test data is 87%

(5) Apply KNN Model, Naïve Bayes Model and Support Vector Machine (SVM) model.

Naïve Bayes Model:

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x1 through xn:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$

GaussianNB implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The parameters σy and μy are estimated using maximum likelihood.

To do the analysis of Navies Bayes we will be using the package from Sklearn.

```
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
```

The training model for the Navies Bayes theorem is as below:

```
0.8341187558906692
[[212  95]
 [ 81 673]]
              precision    recall  f1-score   support

           0       0.72      0.69      0.71       307
           1       0.88      0.89      0.88       754

    accuracy                           0.83      1061
   macro avg       0.80      0.79      0.80      1061
weighted avg       0.83      0.83      0.83      1061
```

- The Accuracy of the training dataset is 83%
- The recall for the dataset is 88%
- The F1 score for the dataset is 88%

The test model for the Navies Bayes theorem is as below:

```
0.8223684210526315
[[112  41]
 [ 40 263]]
              precision    recall  f1-score   support

           0       0.74      0.73      0.73       153
           1       0.87      0.87      0.87       303

    accuracy                           0.82       456
   macro avg       0.80      0.80      0.80       456
weighted avg       0.82      0.82      0.82       456
```

- The Accuracy for the test model is 82%
- The recall for the test model is 87%
- The f1 score for the test model is 85%
- The precision for the test model is 90%

KNN Model:

K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.

Distance functions

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

The KNN model uses the package from SK learn .neighors.

```
from sklearn.neighbors import KNeighborsClassifier
```

The analysis results for the training data are as below:

```
0.8576814326107446
click to scroll output; double click to hide
              precision    recall  f1-score   support

           0       0.77      0.72      0.75       307
           1       0.89      0.91      0.90       754

    accuracy                           0.86      1061
   macro avg       0.83      0.82      0.82      1061
weighted avg       0.86      0.86      0.86      1061
```

- The accuracy of the training model is 86%
- The recall of the training model is 90%
- The precision for the model is 91%
- The F1 score of the model is 89%

The analysis for the test data is as below:

```
0.8267543859649122
[[108  45]
 [ 34 269]]
              precision    recall  f1-score   support

           0       0.76      0.71      0.73       153
           1       0.86      0.89      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.80       456
weighted avg       0.82      0.83      0.83       456
```

- The accuracy of the test model is 82%
- The precision for the test model is 89%
- The recall for the test model is 87%
- The F1 score for the test model is 86%

Support Vector Machines (SVM):

The objective of the support vector machine algorithm is to find a hyper plane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, many possible hyperplanes can be chooser. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classifying with more confidence.

To process the SVM we will be using the package from sklearn

```
from sklearn import svm
```

The model is been split into training and test data where the analysis is been shared as below:

The analysis if training model for SVM is as below:

```
0.8360037700282752
[[194 113]
 [ 61 693]]
              precision    recall  f1-score   support

           0       0.76      0.63      0.69       307
           1       0.86      0.92      0.89       754

    accuracy                           0.84      1061
   macro avg       0.81      0.78      0.79      1061
weighted avg       0.83      0.84      0.83      1061
```

- The accuracy for the training model is 84%
- The precision for the model is 92%
- The recall for the model is 89%
- The F1 score for the model is 86%

The analysis for the test model is as below:

```
0.8421052631578947
[[109  44]
 [ 28 275]]
              precision    recall  f1-score   support

           0       0.80      0.71      0.75       153
           1       0.86      0.91      0.88       303

    accuracy                           0.84       456
   macro avg       0.83      0.81      0.82       456
weighted avg       0.84      0.84      0.84       456
```

- The accuracy for the test model is 84%
- The precision for the model is 91%
- The recall for the model is 88%
- The F1 score for the model is 86%

(6) Model Tuning, Bagging and Boosting.

Bagging: is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data.

The bagging technique is been used from the package from SK Learn.

```
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.ensemble import RandomForestClassifier
RF = RandomForestClassifier()
Bagging_model=BaggingClassifier(base_estimator=RF,n_estimators=100,random_state=1)
Bagging_model.fit(X_train, y_train)
```

```
BaggingClassifier(base_estimator=RandomForestClassifier(bootstrap=True,
                                                        class_weight=None,
                                                        criterion='gini',
                                                        max_depth=None,
                                                        max_features='auto',
                                                        max_leaf_nodes=None,
                                                        min_impurity_decrease=0.0,
                                                        min_impurity_split=None,
                                                        min_samples_leaf=1,
                                                        min_samples_split=2,
                                                        min_weight_fraction_leaf=0.0,
                                                        n_estimators='warn',
                                                        n_jobs=None,
                                                        oob_score=False,
                                                        random_state=None,
                                                        verbose=0,
                                                        warm_start=False),
                  bootstrap=True, bootstrap_features=False, max_features=1.0,
                  max_samples=1.0, n_estimators=100, n_jobs=None,
                  oob_score=False, random_state=1, verbose=0, warm_start=False)
```

The bagging is been performed with Random forest technique for this dataset. The dataset is been spliced into test and training model for identifying the best model

The analysis of training data is been shared as below:

```
0.9670122525918945
[[278  29]
 [  6 748]]
              precision    recall  f1-score   support

           0       0.98      0.91      0.94       307
           1       0.96      0.99      0.98       754

    accuracy                           0.97      1061
   macro avg       0.97      0.95      0.96      1061
weighted avg       0.97      0.97      0.97      1061
```

- The accuracy of the training model is 97%
- The recall for the model is 98%
- The precision for the model is 99%
- The F1 score for the model is 96%

The analysis for the test data is as below:

```
0.8267543859649122
[[103  50]
 [ 29 274]]
              precision    recall  f1-score   support

           0       0.78      0.67      0.72       153
           1       0.85      0.90      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.79      0.80       456
weighted avg       0.82      0.83      0.82       456
```

- The accuracy for the test model is 83%
- The precision for the model is 90%
- The f1 score for the model is 85%
- The recall for the model is 87%

Boosting: is a generic algorithm rather than a specific model. Boosting needs you to specify a weak model (e.g. regression, shallow decision trees, etc.) and then improves it.

The boosting techniques used are as below:

1. Ada Boost: Boosting algorithm developed for classification problems (also called discrete AdaBoost). The weakness is identified by the weak estimator's error rate:

$$\text{for a weak classifier } C$$
$$X: n \times d, Y: n \times k \text{ with sample weights } W: n \times 1$$
$$Error = \frac{\sum_{j=1}^{n} w_j I(C(x_j) \neq Y_j)}{\sum_{j=1}^{n} w_j}, I(x) = \begin{cases} 1, & \text{if } x \text{ is True} \\ 0, & \text{if } x \text{ is False} \end{cases}$$

To use the concept of Ada Booster we need to use the package from SK Learn

```
from sklearn.ensemble import AdaBoostClassifier
```

The dataset is been segregated into test and training where the analysis is as below:

Train Dataset:

```
0.8501413760603205
[[214  93]
 [ 66 688]]
              precision    recall  f1-score   support

           0       0.76      0.70      0.73       307
           1       0.88      0.91      0.90       754

    accuracy                           0.85      1061
   macro avg       0.82      0.80      0.81      1061
weighted avg       0.85      0.85      0.85      1061
```

- The Accuracy for the training dataset is 85%
- The precision for the model is 91%
- The recall for the model is 90%
- The F1 score for the model is 88%

Test Dataset:

```
0.8135964912280702
[[103  50]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.75      0.67      0.71       153
           1       0.84      0.88      0.86       303

    accuracy                           0.81       456
   macro avg       0.79      0.78      0.79       456
weighted avg       0.81      0.81      0.81       456
```

- The accuracy for the test dataset is 81%
- The precision for the dataset is 88%
- The recall for the dataset is 86%
- The F1 score for the dataset is 84%

2. Xtreme Gradient Boosting: Gradient boosting requires a differential loss function and works for both regression and classifications. To use the XGBoost we have to use the package from XGBoost.

```
import xgboost as xgb
```

The dataset is analysed based on test and training model

Training Dataset:

```
0.8416588124410933
[[210  97]
 [ 71 683]]
              precision    recall  f1-score   support

           0       0.75      0.68      0.71       307
           1       0.88      0.91      0.89       754

    accuracy                           0.84      1061
   macro avg       0.81      0.79      0.80      1061
weighted avg       0.84      0.84      0.84      1061
```

- The accuracy for the training model is 84%
- The precision for the training model is 91%
- The recall for the model is 89%
- The F1 score for the model is 88%

Test Dataset:

```
0.8377192982456141
[[104  49]
 [ 25 278]]
              precision    recall  f1-score   support

           0       0.81      0.68      0.74       153
           1       0.85      0.92      0.88       303

    accuracy                           0.84       456
   macro avg       0.83      0.80      0.81       456
weighted avg       0.84      0.84      0.83       456
```

- The accuracy of the model is 83%
- The precision of the model is 92%
- The recall for the model is 88%
- The F1 score for the model is 85%

Model Tuning

The tuning for the model is been used by following either one of the below given ways:

1. Cross Validation
2. Grid Search.

Here we have used cross validation for the analysis of the data here.

Cross validation are performed by using the package from SK Learn

```
from sklearn.model_selection import cross_val_score
```

Here for the dataset as the bagging is processed with Random forest the cross validation is also been performed by using the technique random forest classifier.

```
from sklearn.model_selection import cross_val_score
clfCVRF = RandomForestClassifier(n_estimators=100)
scores = cross_val_score(clfCVRF, X_train, y_train, cv=10)
scores
```

```
array([0.8411215 , 0.81308411, 0.82242991, 0.81308411, 0.88679245,
       0.81132075, 0.82075472, 0.80952381, 0.77142857, 0.82857143])
```

We have got the accuracy scores for the 10 folds for the training dataset as "array([0.8411215 , 0.81308411, 0.82242991, 0.81308411, 0.88679245, 0.81132075, 0.82075472, 0.80952381, 0.77142857, 0.82857143])"

```
print("Accuracy: Final mean:%.3f%%, Final standard deviation:(%.3f%%)" % (scores.mean()*100.0, scores.std()*100.0))
print('Accuracies from each of the 5 folds using clfCVR:',scores)
print("Variance of clfCVR accuracies:",scores.var())
```

```
Accuracy: Final mean:82.181%, Final standard deviation:(2.756%)
Accuracies from each of the 5 folds using clfCVR: [0.8411215  0.81308411 0.82242991 0.81308411 0.88679245 0.81132075
 0.82075472 0.80952381 0.77142857 0.82857143]
Variance of clfCVR accuracies: 0.0007594413162868666
```

(7) Performance Metrics: Check the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model. Final Model: Compare the models and write inference which model is best/optimized.

Performance Matrix for all the prediction techniques used:

1. Naïve Bayes Theorem:

   Training data set:

```
0.8341187558906692
[[212  95]
 [ 81 673]]
              precision    recall  f1-score   support

           0       0.72      0.69      0.71       307
           1       0.88      0.89      0.88       754

    accuracy                           0.83      1061
   macro avg       0.80      0.79      0.80      1061
weighted avg       0.83      0.83      0.83      1061
```

Confusion Matrix:

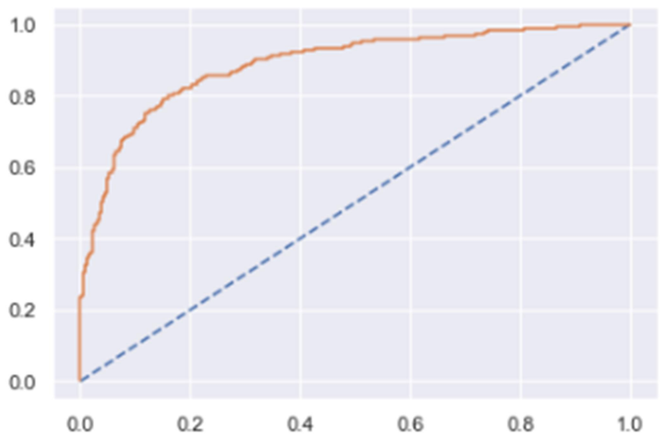
0	212	95
1	81	673

Classification Report:

```
NB_train_precision  0.89
NB_train_recall  0.88
NB_train_f1  0.88
```

ROC & AUC curve:

```
AUC: 0.889
[<matplotlib.lines.Line2D at 0x150afbd4148>]
```



Test Dataset:

```
0.8223684210526315
[[112  41]
 [ 40 263]]
              precision    recall  f1-score   support

           0       0.74      0.73      0.73       153
           1       0.87      0.87      0.87       303

    accuracy                           0.82       456
   macro avg       0.80      0.80      0.80       456
weighted avg       0.82      0.82      0.82       456
```

Confusion Matrix:

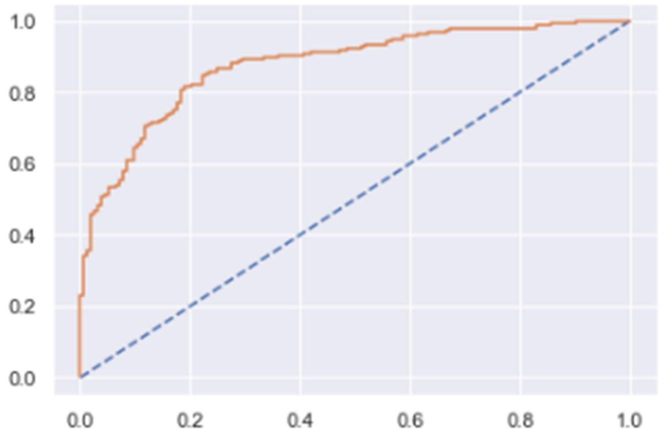
0	112	41
1	40	263

Classification Report:

```
NB_test_precision  0.9
NB_test_recall  0.87
NB_test_f1  0.85
```

ROC & AUC curve:

```
AUC: 0.876
[<matplotlib.lines.Line2D at 0x150afcf4dc8>]
```



2. Logistic Regression:

Training Dataset:

```
0.8369462770970783
[[193 114]
 [ 59 695]]
              precision    recall  f1-score   support

           0       0.77      0.63      0.69       307
           1       0.86      0.92      0.89       754

    accuracy                           0.84      1061
   macro avg       0.81      0.78      0.79      1061
weighted avg       0.83      0.84      0.83      1061
```

Confusion Matrix:

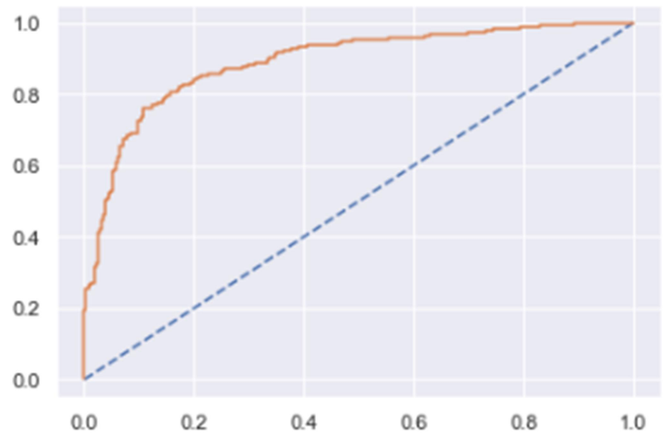
0	193	114
1	59	695

Classification Report:

```
LR_train_precision  0.92
LR_train_recall  0.89
LR_train_f1  0.86
```

ROC & AUC curve:

```
AUC: 0.891

[<matplotlib.lines.Line2D at 0x150af38dc48>]
```



Test Dataset:

```
0.8289473684210527
[[110  43]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.76      0.72      0.74       153
           1       0.86      0.88      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.81       456
weighted avg       0.83      0.83      0.83       456
```

Confusion Matrix:

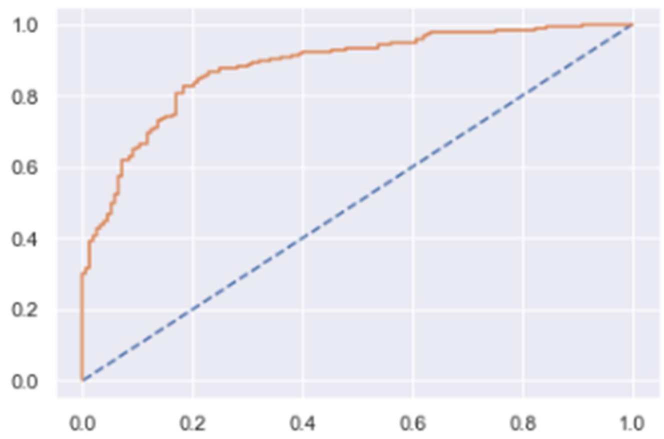
0	110	43
1	35	268

Classification Report:

```
LR_test_precision  0.88
LR_test_recall  0.87
LR_test_f1  0.86
```

ROC & AUC curve:

```
AUC: 0.881
[<matplotlib.lines.Line2D at 0x150afdbef48>]
```



3. Decision Tree

Test Dataset:

```
0.793859649122807
[[103  50]
 [ 44 259]]
              precision    recall  f1-score   support

           0       0.70      0.67      0.69       153
           1       0.84      0.85      0.85       303

    accuracy                           0.79       456
   macro avg       0.77      0.76      0.77       456
weighted avg       0.79      0.79      0.79       456
```

Confusion Matrix:

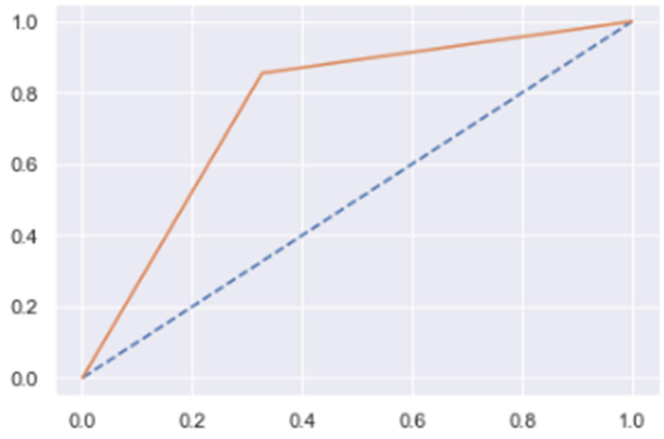
0	103	50
1	44	259

Classification Report:

```
DT_test_precision  0.85
DT_test_recall  0.85
DT_test_f1  0.84
```

ROC & AUC curve:

```
AUC: 0.764
[<matplotlib.lines.Line2D at 0x150afefbe08>]
```



Training Dataset:

```
1.0
[[307   0]
 [  0 754]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       307
           1       1.00      1.00      1.00       754

    accuracy                           1.00      1061
   macro avg       1.00      1.00      1.00      1061
weighted avg       1.00      1.00      1.00      1061
```

Confusion Matrix:

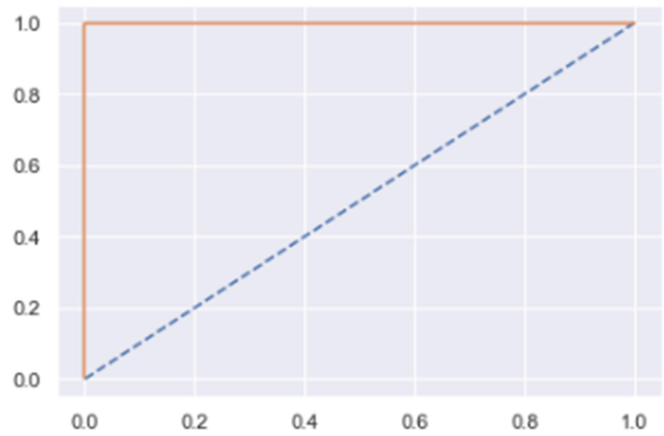
0	307	0
1	0	754

Classification Report:

```
DT_train_precision  1.0
DT_train_recall  1.0
DT_train_f1  1.0
```

ROC & AUC curve:

```
AUC: 1.000

[<matplotlib.lines.Line2D at 0x150afe22708>]
```



4. KNN:

Training Dataset

```
0.8576814326107446
[[221  86]
 [ 65 689]]
              precision    recall  f1-score   support

           0       0.77      0.72      0.75       307
           1       0.89      0.91      0.90       754

    accuracy                           0.86      1061
   macro avg       0.83      0.82      0.82      1061
weighted avg       0.86      0.86      0.86      1061
```

Confusion Matrix:

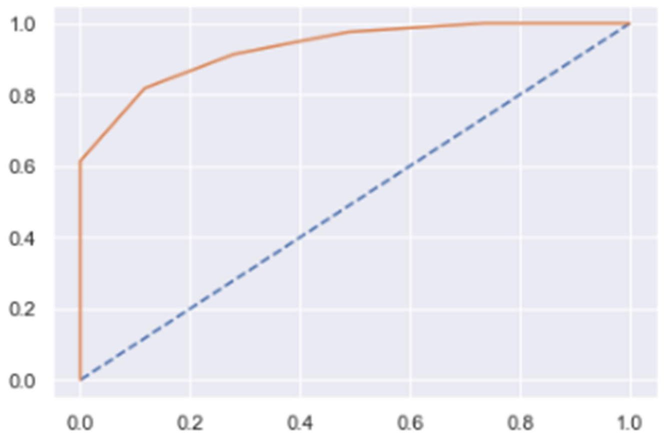
0	221	86
1	65	689

Classification Report:

```
KNN_train_precision  0.91
KNN_train_recall  0.9
KNN_train_f1  0.89
```

ROC & AUC curve:

```
AUC: 0.930
[<matplotlib.lines.Line2D at 0x150af7c9cc8>]
```



Test Dataset

```
0.8267543859649122
[[108  45]
 [ 34 269]]
              precision    recall  f1-score   support

           0       0.76      0.71      0.73       153
           1       0.86      0.89      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.80       456
weighted avg       0.82      0.83      0.83       456
```

Confusion Matrix:

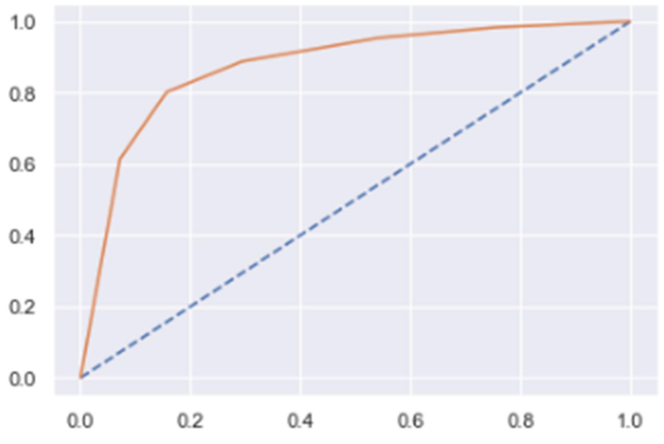
0	108	45
1	34	269

Classification Report:

```
KNN_test_precision  0.89
KNN_test_recall  0.87
KNN_test_f1  0.86
```

ROC & AUC curve:

```
AUC: 0.876

[<matplotlib.lines.Line2D at 0x150af6b3688>]
```



5. SVM

Training Dataset

```
0.8360037700282752
[[194 113]
 [ 61 693]]
              precision    recall  f1-score   support

           0       0.76      0.63      0.69       307
           1       0.86      0.92      0.89       754

    accuracy                           0.84      1061
   macro avg       0.81      0.78      0.79      1061
weighted avg       0.83      0.84      0.83      1061
```

Confusion Matrix:

0	194	113
1	61	693

Classification Report:

```
SVM_train_precision  0.92
SVM_train_recall  0.89
SVM_train_f1  0.86
```

Test Dataset:

```
0.8421052631578947
[[109  44]
 [ 28 275]]
              precision    recall  f1-score   support

           0       0.80      0.71      0.75       153
           1       0.86      0.91      0.88       303

    accuracy                           0.84       456
   macro avg       0.83      0.81      0.82       456
weighted avg       0.84      0.84      0.84       456
```

Confusion Matrix:

0	109	44
1	28	275

Classification Report:

```
SVM_test_precision  0.91
SVM_test_recall  0.88
SVM_test_f1  0.86
```

6. Linear Discriminant Analysis:

Training Model

```
0.8341187558906692
[[200 107]
 [ 69 685]]
              precision    recall  f1-score   support

           0       0.74      0.65      0.69       307
           1       0.86      0.91      0.89       754

    accuracy                           0.83      1061
   macro avg       0.80      0.78      0.79      1061
weighted avg       0.83      0.83      0.83      1061
```

Confusion Matrix:

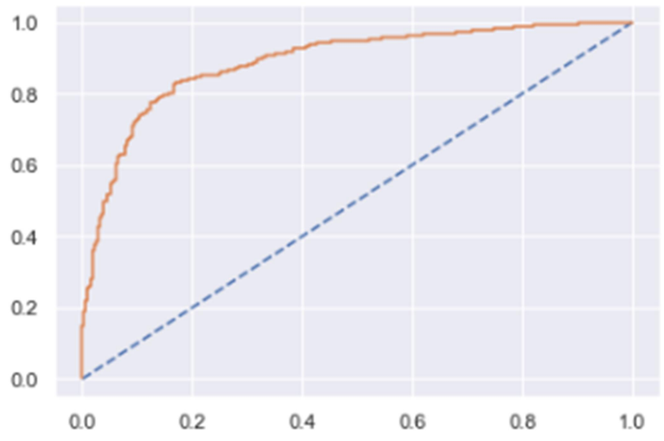
0	200	107
1	69	685

Classification Report:

```
LDA_train_precision  1.0
LDA_train_recall  1.0
LDA_train_f1  1.0
```

ROC & AUC curve:

```
AUC: 0.890

[<matplotlib.lines.Line2D at 0x150af017ec8>]
```



Test Model:

```
0.831140350877193
[[111  42]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.76      0.73      0.74       153
           1       0.86      0.88      0.87       303

    accuracy                           0.83       456
   macro avg       0.81      0.80      0.81       456
weighted avg       0.83      0.83      0.83       456
```

Confusion Matrix:

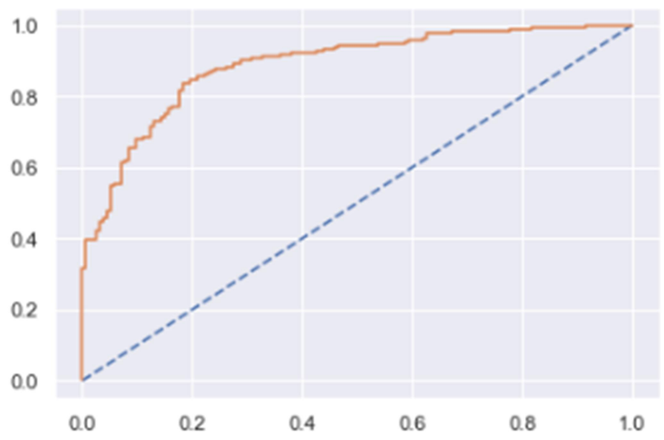
0	111	42
1	35	268

Classification Report:

```
LDA_test_precision  0.9
LDA_test_recall  0.87
LDA_test_f1  0.85
```

ROC & AUC curve:

```
AUC: 0.888

[<matplotlib.lines.Line2D at 0x150b389ffc8>]
```



7. Ada Booster

Training Model:

```
0.8501413760603205
[[214  93]
 [ 66 688]]
              precision    recall  f1-score   support

           0       0.76      0.70      0.73       307
           1       0.88      0.91      0.90       754

    accuracy                           0.85      1061
   macro avg       0.82      0.80      0.81      1061
weighted avg       0.85      0.85      0.85      1061
```

Confusion Matrix:

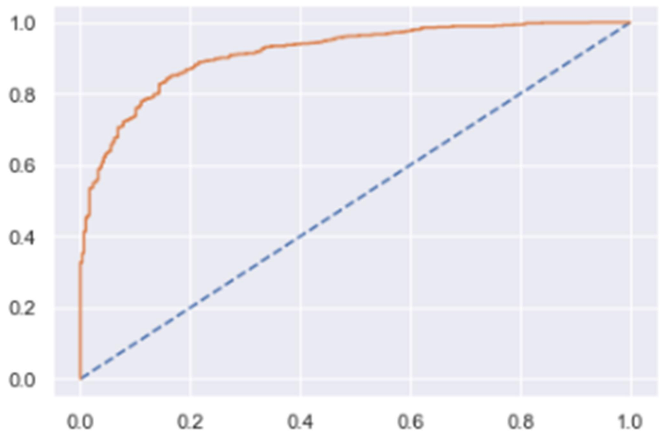
0	214	93
1	66	688

Classification Report:

```
ClfADB_train_precision  0.91
ClfADB_train_recall  0.9
ClfADB_train_f1  0.88
```

ROC & AUC curve:

```
AUC: 0.915
[<matplotlib.lines.Line2D at 0x150b3908ec8>]
```



Test Model:

```
0.8135964912280702
[[103  50]
 [ 35 268]]
              precision    recall  f1-score   support

           0       0.75      0.67      0.71       153
           1       0.84      0.88      0.86       303

    accuracy                           0.81       456
   macro avg       0.79      0.78      0.79       456
weighted avg       0.81      0.81      0.81       456
```

Confusion Matrix:

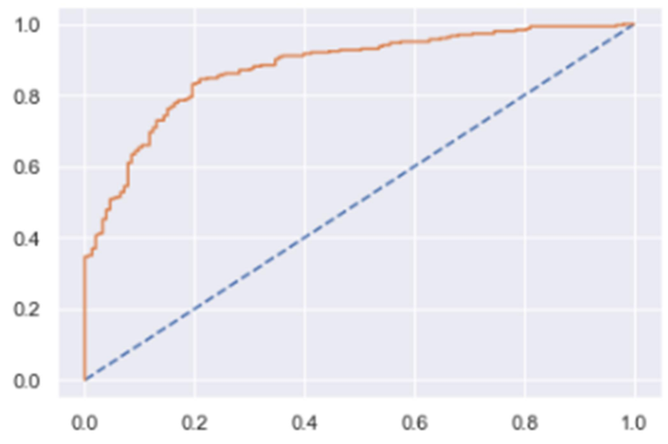
0	103	50
1	35	268

Classification Report:

```
ClfADB_test_precision  0.88
ClfADB_test_recall  0.86
ClfADB_test_f1  0.84
```

ROC & AUC curve:

```
AUC: 0.877

[<matplotlib.lines.Line2D at 0x150affc7ec8>]
```



8. Random Forest:

Training Model

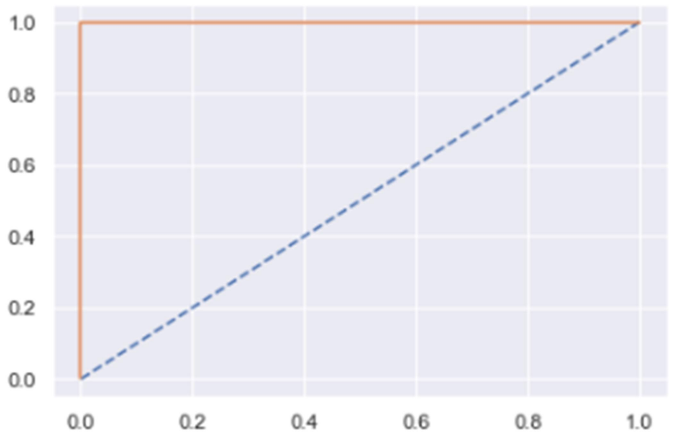
```
1.0
[[307   0]
 [  0 754]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       307
           1       1.00      1.00      1.00       754

    accuracy                           1.00      1061
   macro avg       1.00      1.00      1.00      1061
weighted avg       1.00      1.00      1.00      1061
```

Classification Report:

```
RF_train_precision  1.0
RF_train_recall  1.0
RF_train_f1  1.0
```

ROC & AUC curve:

```
AUC: 1.000
[<matplotlib.lines.Line2D at 0x150b12aec08>]
```



Test Model:

```
0.831140350877193
[[105  48]
 [ 29 274]]
              precision    recall  f1-score   support

           0       0.78      0.69      0.73       153
           1       0.85      0.90      0.88       303

    accuracy                           0.83       456
   macro avg       0.82      0.80      0.80       456
weighted avg       0.83      0.83      0.83       456
```

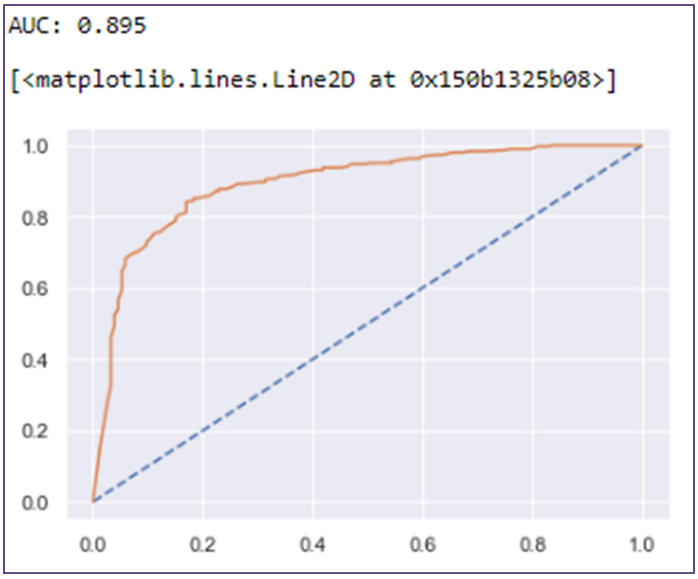
Confusion Matrix:

0	105	48
1	29	274

Classification Report:

```
RF_test_precision  0.9
RF_test_recall  0.87
RF_test_f1  0.85
```

ROC & AUC curve:



Comparison between all the models:

Method	Accuracy	Recall	Precision	F1 Score	AUC
NB Train	83	88	89	88	89
NB Test	82	87	90	85	88
RF Train	100	100	100	100	100
RF Test	83	87	90	85	89
LR Train	84	89	92	86	90
LR Test	83	87	88	86	88
DT Train	100	100	100	100	100
DT Test	79	85	85	84	88
KNN Train	86	90	91	89	93
KNN Test	83	87	89	86	88
LDA Train	83	98	99	96	89
LDA Test	83	87	90	85	88
Ada Boost Train	85	90	91	88	92
Ada Boost Test	81	86	88	84	88
Bagging Train	97	99	98	96	98
Bagging Test	83	87	90	85	90
XG Boost Train	84	89	91	88	90
XG Boost Test	84	88	92	85	89

From the above comparison between the various model predicting we could find out that the bagging training model by using random forest classifier was giving a prediction of 98% where the model accuracy was 97%. Even the random forest training model is giving an output for 100%.

(8) Based on these predictions, what are the insights?

All our models are pointing out for Labour party that is been led by Tony Blair to win the election. The exit polling therefore predicts that this time the Labour party will be winning the election are with 85 to 90% from our analysis.

Case Study 2

In this particular project, we are going to work on the inaugural corpora from the nltk in Python. We will be looking at the following speeches of the Presidents of the United States of America:

1. President Franklin D. Roosevelt in 1941
2. President John F. Kennedy in 1961
3. President Richard Nixon in 1973

- Find the number of characters, words and sentences for the mentioned documents. – 3 Marks
- Remove all the stop words from all the three speeches. – 3 Marks
- Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stopwords) – 3 Marks
- Plot the word cloud of each of the speeches of the variable. (after removing the stopwords) – 3 Marks [refer to the End-to-End Case Study done in the Mentored Learning Session]

(1) Find the number of characters, words and sentences for the mentioned documents.

Load the libraries:

```
import numpy as np
import pandas as pd
import nltk
import random
import string
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS #calling WordCloud and Stopwords
from nltk.corpus import stopwords  #stopwords
from nltk.tokenize import word_tokenize #word tokenizer
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

Read the Text:

```
nltk.download('inaugural')
```

In nltk packages is having an inbuild packages under “Corpus” reader , where we have to download the inaugural speeches saved under the corpus reader.

```
from nltk.corpus import inaugural
```

We could find that there are 58 inaugural speeches saved under the "inaugural" packages in nltk.corpus.

```
names[0],names[57]
```

```
('1789-Washington.txt', '2017-Trump.txt')
```

The first speech saved is 1789 Washington's speech and the last speech saved is by Donald Trump in 2017.

By using the below coding all the three speeches are uploaded to python.

- Inaugural. Raw('1941-Roosevelt.txt')
- Inaugural. Raw('1961-Kennedy.txt')
- Inaugural. Raw('1973-Nixon.txt')

Brief of Speeches from Corpus-Inaugural:

1941-Roosevelt Speech:

```
'On each national day of inauguration since 1789, the people have renew
ed their sense of dedication to the United States.\n\nIn Washington\'s
day the task of the people was to create and weld together a nation.\n\
nIn Lincoln\'s day the task of the people was to preserve that Nation f
rom disruption from within.\n\nIn this day the task of the people is to
save that Nation and its institutions from disruption from without.\n\n
To us there has come a time, in the midst of swift happenings, to pause
for a moment and take stock -- to recall what our place in history has
been, and to rediscover what we are and what we may be. If we do not, w
e risk the real peril of inaction.\n\nLives of nations are determined n
ot by the count of years, but by the lifetime of the human spirit. The
life of a man is three-score years and ten: a little more, a little les
s. The life of a nation is the fullness of the measure of its will to l
ive.\n\nThere are men who doubt this. There are men who believe that de
mocracy, as a form of Government and a frame of life, is limited or mea
sured by a kind of mystical and artificial fate that, for some unexplai
ned reason, tyranny and slavery have become the surging wave of the fut
ure -- and that freedom is an ebbing tide.\n\nBut we Americans know tha
t this is not true.\n\nEight years ago, when the life of this Republic
seemed frozen by a fatalistic terror, we proved that this is not true.
We were in the midst of shock -- but we acted. We acted quickly, boldly
, decisively.\n\nThese later years have been living years -- fruitful y
ears for the people of this democracy. For they have brought to us grea
ter security and, I hope, a better understanding that life\'s ideals ar
e to be measured in other than material things.\n\nMost vital to our pr
esent and our future is this experience of a democracy which successful
ly survived crisis at home; put away many evil things; built new struct
ures on enduring lines; and, through it all, maintained the fact of its
democracy.\n\nFor action has been taken within the three-way framework
```

of the Constitution of the United States. The coordinate branches of th
e Government continue freely to function. The Bill of Rights remains in
violate. The freedom of elections is wholly maintained. Prophets of the
downfall of American democracy have seen their dire predictions come to
naught.\n\nDemocracy is not dying.\n\nWe know it because we have seen i
t revive--and grow.\n\nWe know it cannot die -- because it is built on
the unhampered initiative of individual men and women joined together i
n a common enterprise -- an enterprise undertaken and carried through b
y the free expression of a free majority.\n\nWe know it because democra
cy alone, of all forms of government, enlists the full force of men\'s
enlightened will.\n\nWe know it because democracy alone has constructed
an unlimited civilization capable of infinite progress in the improveme
nt of human life.\n\nWe know it because, if we look below the surface,
we sense it still spreading on every continent -- for it is the most hu
mane, the most advanced, and in the end the most unconquerable of all f
orms of human society.\n\nA nation, like a person, has a body--a body t
hat must be fed and clothed and housed, invigorated and rested, in a ma
nner that measures up to the objectives of our time.\n\nA nation, like
a person, has a mind -- a mind that must be kept informed and alert, th
at must know itself, that understands the hopes and the needs of its ne
ighbors -- all the other nations that live within the narrowing circle
of the world.\n\nAnd a nation, like a person, has something deeper, som
ething more permanent, something larger than the sum of all its parts.
It is that something which matters most to its future -- which calls fo
rth the most sacred guarding of its present.\n\nIt is a thing for which
we find it difficult -- even impossible -- to hit upon a single, simple
word.\n\nAnd yet we all understand what it is -- the spirit -- the fait
h of America. It is the product of centuries. It was born in the multit
udes of those who came from many lands -- some of high degree, but most
ly plain people, who sought here, early and late, to find freedom more
freely.\n\nThe democratic aspiration is no mere recent phase in human h
istory. It is human history. It permeated the ancient life of early peo
ples. It blazed anew in the middle ages. It was written in Magna Charta
.\n\nIn the Americas its impact has been irresistible. America has been
the New World in all tongues, to all peoples, not because this continen
t was a new-found land, but because all those who came here believed th
ey could create upon this continent a new life -- a life that should be
new in freedom.\n\nIts vitality was written into our own Mayflower Comp
act, into the Declaration of Independence, into the Constitution of the
United States, into the Gettysburg Address.\n\nThose who first came her
e to carry out the longings of their spirit, and the millions who follo
wed, and the stock that sprang from them -- all have moved forward cons
tantly and consistently toward an ideal which in itself has gained stat
ure and clarity with each generation.\n\nThe hopes of the Republic cann
ot forever tolerate either undeserved poverty or self-serving wealth.\n
\nWe know that we still have far to go; that we must more greatly build
the security and the opportunity and the knowledge of every citizen, in
the measure justified by the resources and the capacity of the land.\n\
nBut it is not enough to achieve these purposes alone. It is not enough
to clothe and feed the body of this Nation, and instruct and inform its
mind. For there is also the spirit. And of the three, the greatest is t
he spirit.\n\nWithout the body and the mind, as all men know, the Natio
n could not live.\n\nBut if the spirit of America were killed, even tho
ugh the Nation\'s body and mind, constricted in an alien world, lived o
n, the America we know would have perished.\n\nThat spirit -- that fait
h -- speaks to us in our daily lives in ways often unnoticed, because t
hey seem so obvious. It speaks to us here in the Capital of the Nation.
It speaks to us through the processes of governing in the sovereignties

of 48 States. It speaks to us in our counties, in our cities, in our to wns, and in our villages. It speaks to us from the other nations of the hemisphere, and from those across the seas -- the enslaved, as well as the free. Sometimes we fail to hear or heed these voices of freedom bec ause to us the privilege of our freedom is such an old, old story.\n\nT he destiny of America was proclaimed in words of prophecy spoken by our first President in his first inaugural in 1789 -- words almost directed , it would seem, to this year of 1941: "The preservation of the sacred fire of liberty and the destiny of the republican model of government a re justly considered deeply, finally, staked on the experiment intruste d to the hands of the American people."\n\nIf we lose that sacred fire- -if we let it be smothered with doubt and fear -- then we shall reject the destiny which Washington strove so valiantly and so triumphantly to establish. The preservation of the spirit and faith of the Nation does, and will, furnish the highest justification for every sacrifice that we may make in the cause of national defense.\n\nIn the face of great peri ls never before encountered, our strong purpose is to protect and to pe rpetuate the integrity of democracy.\n\nFor this we muster the spirit o f America, and the faith of America.\n\nWe do not retreat. We are not c ontent to stand still. As Americans, we go forward, in the service of o ur country, by the will of God.\n'

1961 Kennedy Speech:

'Vice President Johnson, Mr. Speaker, Mr. Chief Justice, President Eise nhower, Vice President Nixon, President Truman, reverend clergy, fellow citizens, we observe today not a victory of party, but a celebration of freedom -- symbolizing an end, as well as a beginning -- signifying ren ewal, as well as change. For I have sworn I before you and Almighty God the same solemn oath our forebears l prescribed nearly a century and th ree quarters ago.\n\nThe world is very different now. For man holds in his mortal hands the power to abolish all forms of human poverty and al l forms of human life. And yet the same revolutionary beliefs for which our forebears fought are still at issue around the globe -- the belief that the rights of man come not from the generosity of the state, but f rom the hand of God.\n\nWe dare not forget today that we are the heirs of that first revolution. Let the word go forth from this time and plac e, to friend and foe alike, that the torch has been passed to a new gen eration of Americans -- born in this century, tempered by war, discipli ned by a hard and bitter peace, proud of our ancient heritage -- and un willing to witness or permit the slow undoing of those human rights to which this Nation has always been committed, and to which we are commit ted today at home and around the world.\n\nLet every nation know, wheth er it wishes us well or ill, that we shall pay any price, bear any burd en, meet any hardship, support any friend, oppose any foe, in order to assure the survival and the success of liberty.\n\nThis much we pledge -- and more.\n\nTo those old allies whose cultural and spiritual origin s we share, we pledge the loyalty of faithful friends. United, there is little we cannot do in a host of cooperative ventures. Divided, there i s little we can do -- for we dare not meet a powerful challenge at odds and split asunder.\n\nTo those new States whom we welcome to the ranks of the free, we pledge our word that one form of colonial control shall not have passed away merely to be replaced by a far more iron tyranny. We shall not always expect to find them supporting our view. But we sha ll always hope to find them strongly supporting their own freedom -- an d to remember that, in the past, those who foolishly sought power by ri

ding the back of the tiger ended up inside.\n\nTo those peoples in the
huts and villages across the globe struggling to break the bonds of mas
s misery, we pledge our best efforts to help them help themselves, for
whatever period is required -- not because the Communists may be doing
it, not because we seek their votes, but because it is right. If a free
society cannot help the many who are poor, it cannot save the few who a
re rich.\n\nTo our sister republics south of our border, we offer a spe
cial pledge -- to convert our good words into good deeds -- in a new al
liance for progress -- to assist free men and free governments in casti
ng off the chains of poverty. But this peaceful revolution of hope cann
ot become the prey of hostile powers. Let all our neighbors know that w
e shall join with them to oppose aggression or subversion anywhere in t
he Americas. And let every other power know that this Hemisphere intend
s to remain the master of its own house.\n\nTo that world assembly of s
overeign states, the United Nations, our last best hope in an age where
the instruments of war have far outpaced the instruments of peace, we r
enew our pledge of support--to prevent it from becoming merely a forum
for invective -- to strengthen its shield of the new and the weak -- an
d to enlarge the area in which its writ may run.\n\nFinally, to those n
ations who would make themselves our adversary, we offer not a pledge b
ut a request: that both sides begin anew the quest for peace, before th
e dark powers of destruction unleashed by science engulf all humanity i
n planned or accidental self-destruction.\n\nWe dare not tempt them wit
h weakness. For only when our arms are sufficient beyond doubt can we b
e certain beyond doubt that they will never be employed.\n\nBut neither
can two great and powerful groups of nations take comfort from our pres
ent course -- both sides overburdened by the cost of modern weapons, bo
th rightly alarmed by the steady spread of the deadly atom, yet both ra
cing to alter that uncertain balance of terror that stays the hand of m
ankind\'s final war.\n\nSo let us begin anew -- remembering on both sid
es that civility is not a sign of weakness, and sincerity is always sub
ject to proof. Let us never negotiate out of fear. But let us never fea
r to negotiate.\n\nLet both sides explore what problems unite us instea
d of belaboring those problems which divide us.\n\nLet both sides, for
the first time, formulate serious and precise proposals for the inspect
ion and control of arms -- and bring the absolute power to destroy othe
r nations under the absolute control of all nations.\n\nLet both sides
seek to invoke the wonders of science instead of its terrors. Together
let us explore the stars, conquer the deserts, eradicate disease, tap t
he ocean depths, and encourage the arts and commerce.\n\nLet both sides
unite to heed in all corners of the earth the command of Isaiah -- to "
undo the heavy burdens ... and to let the oppressed go free."\n\nAnd if
a beachhead of cooperation may push back the jungle of suspicion, let b
oth sides join in creating a new endeavor, not a new balance of power,
but a new world of law, where the strong are just and the weak secure a
nd the peace preserved.\n\nAll this will not be finished in the first 1
00 days. Nor will it be finished in the first 1,000 days, nor in the li
fe of this Administration, nor even perhaps in our lifetime on this pla
net. But let us begin.\n\nIn your hands, my fellow citizens, more than
in mine, will rest the final success or failure of our course. Since th
is country was founded, each generation of Americans has been summoned
to give testimony to its national loyalty. The graves of young American
s who answered the call to service surround the globe.\n\nNow the trump
et summons us again -- not as a call to bear arms, though arms we need;
not as a call to battle, though embattled we are -- but a call to bear
the burden of a long twilight struggle, year in and year out, "rejoicin
g in hope, patient in tribulation" -- a struggle against the common ene
mies of man: tyranny, poverty, disease, and war itself.\n\nCan we forge

```
against these enemies a grand and global alliance, North and South, Eas
t and West, that can assure a more fruitful life for all mankind? Will
you join in that historic effort?\n\nIn the long history of the world,
only a few generations have been granted the role of defending freedom
in its hour of maximum danger. I do not shrink from this responsibility
-- I welcome it. I do not believe that any of us would exchange places
with any other people or any other generation. The energy, the faith, t
he devotion which we bring to this endeavor will light our country and
all who serve it -- and the glow from that fire can truly light the wor
ld.\n\nAnd so, my fellow Americans: ask not what your country can do fo
r you -- ask what you can do for your country.\n\nMy fellow citizens of
the world: ask not what America will do for you, but what together we c
an do for the freedom of man.\n\nFinally, whether you are citizens of A
merica or citizens of the world, ask of us the same high standards of s
trength and sacrifice which we ask of you. With a good conscience our o
nly sure reward, with history the final judge of our deeds, let us go f
orth to lead the land we love, asking His blessing and His help, but kn
owing that here on earth God\'s work must truly be our own.\n'
```

1973 Nixon Speech:

```
'Mr. Vice President, Mr. Speaker, Mr. Chief Justice, Senator Cook, Mrs.
Eisenhower, and my fellow citizens of this great and good country we sh
are together:\n\nWhen we met here four years ago, America was bleak in
spirit, depressed by the prospect of seemingly endless war abroad and o
f destructive conflict at home.\n\nAs we meet here today, we stand on t
he threshold of a new era of peace in the world.\n\nThe central questio
n before us is: How shall we use that peace? Let us resolve that this e
ra we are about to enter will not be what other postwar periods have so
often been: a time of retreat and isolation that leads to stagnation at
home and invites new danger abroad.\n\nLet us resolve that this will be
what it can become: a time of great responsibilities greatly borne, in
which we renew the spirit and the promise of America as we enter our th
ird century as a nation.\n\nThis past year saw far-reaching results fro
m our new policies for peace. By continuing to revitalize our tradition
al friendships, and by our missions to Peking and to Moscow, we were ab
le to establish the base for a new and more durable pattern of relation
ships among the nations of the world. Because of America\'s bold initia
tives, 1972 will be long remembered as the year of the greatest progres
s since the end of World War II toward a lasting peace in the world.\n\
nThe peace we seek in the world is not the flimsy peace which is merely
an interlude between wars, but a peace which can endure for generations
to come.\n\nIt is important that we understand both the necessity and t
he limitations of America\'s role in maintaining that peace.\n\nUnless
we in America work to preserve the peace, there will be no peace.\n\nUn
less we in America work to preserve freedom, there will be no freedom.\
n\nBut let us clearly understand the new nature of America\'s role, as
a result of the new policies we have adopted over these past four years
.\n\nWe shall respect our treaty commitments.\n\nWe shall support vigor
ously the principle that no country has the right to impose its will or
rule on another by force.\n\nWe shall continue, in this era of negotiat
ion, to work for the limitation of nuclear arms, and to reduce the dang
er of confrontation between the great powers.\n\nWe shall do our share
in defending peace and freedom in the world. But we shall expect others
to do their share.\n\nThe time has passed when America will make every
other nation\'s conflict our own, or make every other nation\'s future
```

our responsibility, or presume to tell the people of other nations how
to manage their own affairs.\n\nJust as we respect the right of each na
tion to determine its own future, we also recognize the responsibility
of each nation to secure its own future.\n\nJust as America\'s role is
indispensable in preserving the world\'s peace, so is each nation\'s ro
le indispensable in preserving its own peace.\n\nTogether with the rest
of the world, let us resolve to move forward from the beginnings we hav
e made. Let us continue to bring down the walls of hostility which have
divided the world for too long, and to build in their place bridges of
understanding -- so that despite profound differences between systems o
f government, the people of the world can be friends.\n\nLet us build a
structure of peace in the world in which the weak are as safe as the st
rong -- in which each respects the right of the other to live by a diff
erent system -- in which those who would influence others will do so by
the strength of their ideas, and not by the force of their arms.\n\nLet
us accept that high responsibility not as a burden, but gladly -- gladl
y because the chance to build such a peace is the noblest endeavor in w
hich a nation can engage; gladly, also, because only if we act greatly
in meeting our responsibilities abroad will we remain a great Nation, a
nd only if we remain a great Nation will we act greatly in meeting our
challenges at home.\n\nWe have the chance today to do more than ever be
fore in our history to make life better in America -- to ensure better
education, better health, better housing, better transportation, a clea
ner environment -- to restore respect for law, to make our communities
more livable -- and to insure the God-given right of every American to
full and equal opportunity.\n\nBecause the range of our needs is so gre
at -- because the reach of our opportunities is so great -- let us be b
old in our determination to meet those needs in new ways.\n\nJust as bu
ilding a structure of peace abroad has required turning away from old p
olicies that failed, so building a new era of progress at home requires
turning away from old policies that have failed.\n\nAbroad, the shift f
rom old policies to new has not been a retreat from our responsibilitie
s, but a better way to peace.\n\nAnd at home, the shift from old polici
es to new will not be a retreat from our responsibilities, but a better
way to progress.\n\nAbroad and at home, the key to those new responsibi
lities lies in the placing and the division of responsibility. We have
lived too long with the consequences of attempting to gather all power
and responsibility in Washington.\n\nAbroad and at home, the time has c
ome to turn away from the condescending policies of paternalism -- of "
Washington knows best."\n\nA person can be expected to act responsibly
only if he has responsibility. This is human nature. So let us encourag
e individuals at home and nations abroad to do more for themselves, to
decide more for themselves. Let us locate responsibility in more places
. Let us measure what we will do for others by what they will do for th
emselves.\n\nThat is why today I offer no promise of a purely governmen
tal solution for every problem. We have lived too long with that false
promise. In trusting too much in government, we have asked of it more t
han it can deliver. This leads only to inflated expectations, to reduce
d individual effort, and to a disappointment and frustration that erode
confidence both in what government can do and in what people can do.\n\
nGovernment must learn to take less from people so that people an do mo
re for themselves.\n\nLet us remember that America was built not by gov
ernment, but by people -- not by welfare, but by work -- not by shirkin
g responsibility, but by seeking responsibility.\n\nIn our own lives, l
et each of us ask -- not just what will government do for me, but what
can I do for myself?\n\nIn the challenges we face together, let each of
us ask -- not just how can government help, but how can I help?\n\nYour
National Government has a great and vital role to play. And I pledge to

you that where this Government should act, we will act boldly and we wi
ll lead boldly. But just as important is the role that each and every o
ne of us must play, as an individual and as a member of his own communi
ty.\n\nFrom this day forward, let each of us make a solemn commitment i
n his own heart: to bear his responsibility, to do his part, to live hi
s ideals -- so that together, we can see the dawn of a new age of progr
ess for America, and together, as we celebrate our 200th anniversary as
a nation, we can do so proud in the fulfillment of our promise to ourse
lves and to the world.\n\nAs America\'s longest and most difficult war
comes to an end, let us again learn to debate our differences with civi
lity and decency. And let each of us reach out for that one precious qu
ality government cannot provide -- a new level of respect for the right
s and feelings of one another, a new level of respect for the individua
l human dignity which is the cherished birthright of every American.\n\
nAbove all else, the time has come for us to renew our faith in ourselv
es and in America.\n\nIn recent years, that faith has been challenged.\
n\nOur children have been taught to be ashamed of their country, ashame
d of their parents, ashamed of America\'s record at home and of its rol
e in the world.\n\nAt every turn, we have been beset by those who find
everything wrong with America and little that is right. But I am confid
ent that this will not be the judgment of history on these remarkable t
imes in which we are privileged to live.\n\nAmerica\'s record in this c
entury has been unparalleled in the world\'s history for its responsibi
lity, for its generosity, for its creativity and for its progress.\n\nL
et us be proud that our system has produced and provided more freedom a
nd more abundance, more widely shared, than any other system in the his
tory of the world.\n\nLet us be proud that in each of the four wars in
which we have been engaged in this century, including the one we are no
w bringing to an end, we have fought not for our selfish advantage, but
to help others resist aggression.\n\nLet us be proud that by our bold,
new initiatives, and by our steadfastness for peace with honor, we have
made a break-through toward creating in the world what the world has no
t known before -- a structure of peace that can last, not merely for ou
r time, but for generations to come.\n\nWe are embarking here today on
an era that presents challenges great as those any nation, or any gener
ation, has ever faced.\n\nWe shall answer to God, to history, and to ou
r conscience for the way in which we use these years.\n\nAs I stand in
this place, so hallowed by history, I think of others who have stood he
re before me. I think of the dreams they had for America, and I think o
f how each recognized that he needed help far beyond himself in order t
o make those dreams come true.\n\nToday, I ask your prayers that in the
years ahead I may have God\'s help in making decisions that are right f
or America, and I pray for your help so that together we may be worthy
of our challenge.\n\nLet us pledge together to make these next four yea
rs the best four years in America\'s history, so that on its 200th birt
hday America will be as young and as vital as when it began, and as bri
ght a beacon of hope for all the world.\n\nLet us go forward from here
confident in hope, strong in our faith in one another, sustained by our
faith in God who created us, and striving always to serve His purpose.\
n'

Find the number of characters, words and sentences for the mentioned documents.

1. 1941 Roosevelt Speech

- The number of words appearing in the speech for 1941 Roosevelt is 1536 words.
- The number of unique words used in the speech is 698 words
- The number of characters used in the speech is 169

2. 1961 Kennedy Speech

- The number of words appearing is Kennedy's speech is 1546 words
- The number of unique words listed in the speech is 546 words
- The number of characters used in the speech is 161

3. 1973 Nixon Speech:

- The number of words appearing in Nixon speech is 2028 words
- The number of unique words used in the speech is 516 words
- The number of characters used in the speech is 164 characters

(2) Remove all the stop words from the three speeches.

Stop Words: A stop word is a commonly used word (such as "the", "a", "an", "in") that a search engine programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query

To remove the stop words, we have to import the library from NLTK Corpus package for stop words.

```
from nltk.corpus import stopwords
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'y
ourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself',
'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'those',
'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'a
n', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'b
etween', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'of
f', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very',
's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'ar
en', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "have
n't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "should
n't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"]
```

1941- Roosevelt's Speech:

From the speech we will be removing, the special characters first, and then remove the single character. Convert the speech from upper case to lower case. The codes shared as below:

```python
processed_features = []
for sentence in range(0, len(words)):
      # Remove all the special characters
    processed_feature = re.sub(r'\W', ' ', str(words[sentence]))

    # remove all single characters
    processed_feature= re.sub(r'\s+[a-zA-Z]\s+', ' ', processed_feature)

    # Remove single characters from the start
    processed_feature = re.sub(r'\^[a-zA-Z]\s+', ' ', processed_feature)

    # Substituting multiple spaces with single space
    processed_feature = re.sub(r'\s+', ' ', processed_feature, flags=re.I)

    # Removing prefixed 'b'
    processed_feature = re.sub(r'^b\s+', '', processed_feature)

    # Converting to Lowercase
    processed_feature = processed_feature.lower()

    processed_features.append(processed_feature)
```

After removing the stop words:

```
['national', 'day', 'inauguration', '1789', 'people', 'renewed', 'sense', 'dedication', 'united', 'states', 'washington', 's',
'day', 'task', 'people', 'create', 'weld', 'together', 'nation', 'lincoln', 's', 'day', 'task', 'people', 'preserve', 'nation',
'disruption', 'within', 'day', 'task', 'people', 'save', 'nation', 'institutions', 'disruption', 'without', 'us', 'come', 'tim
e', 'midst', 'swift', 'happenings', 'pause', 'moment', 'take', 'stock', 'recall', 'place', 'history', 'rediscover', 'may', 'ris
k', 'real', 'peril', 'inaction', 'lives', 'nations', 'determined', 'count', 'years', 'lifetime', 'human', 'spirit', 'life', 'ma
n', 'three', 'score', 'years', 'ten', 'little', 'little', 'less', 'life', 'nation', 'fullness', 'measure', 'will', 'live', 'me
n', 'doubt', 'men', 'believe', 'democracy', 'form', 'government', 'frame', 'life', 'limited', 'measured', 'kind', 'mystical',
'artificial', 'fate', 'unexplained', 'reason', 'tyranny', 'slavery', 'become', 'surging', 'wave', 'future', 'freedom', 'ebbin
g', 'tide', 'americans', 'know', 'true', 'eight', 'years', 'ago', 'life', 'republic', 'seemed', 'frozen', 'fatalistic', 'terro
r', 'proved', 'true', 'midst', 'shock', 'acted', 'acted', 'quickly', 'boldly', 'decisively', 'later', 'years', 'living', 'year
s', 'fruitful', 'years', 'people', 'democracy', 'brought', 'us', 'greater', 'security', 'hope', 'better', 'understanding', 'lif
e', 's', 'ideals', 'measured', 'material', 'things', 'vital', 'present', 'future', 'experience', 'democracy', 'successfully',
'survived', 'crisis', 'home', 'put', 'away', 'many', 'evil', 'things', 'built', 'new', 'structures', 'enduring', 'lines', 'main
tained', 'fact', 'democracy', 'action', 'taken', 'within', 'three', 'way', 'framework', 'constitution', 'united', 'states', 'co
ordinate', 'branches', 'government', 'continue', 'freely', 'function', 'bill', 'rights', 'remains', 'inviolate', 'freedom', 'el
ections', 'wholly', 'maintained', 'prophets', 'downfall', 'american', 'democracy', 'seen', 'dire', 'predictions', 'come', 'naug
ht', 'democracy', 'dying', 'know', 'seen', 'revive', 'grow', 'know', 'die', 'built', 'unhampered', 'initiative', 'individual',
'men', 'women', 'joined', 'together', 'common', 'enterprise', 'enterprise', 'undertaken', 'carried', 'free', 'expression', 'fre
e', 'majority', 'know', 'democracy', 'alone', 'forms', 'government', 'enlists', 'full', 'force', 'men', 's', 'enlightened', 'wi
ll', 'know', 'democracy', 'alone', 'constructed', 'unlimited', 'civilization', 'capable', 'infinite', 'progress', 'improvemen
t', 'human', 'life', 'know', 'look', 'surface', 'sense', 'still', 'spreading', 'every', 'continent', 'humane', 'advanced', 'en
d', 'unconquerable', 'forms', 'human', 'society', 'nation', 'person', 'body', 'body', 'must', 'fed', 'clothed', 'housed', 'invi
gorated', 'rested', 'manner', 'measures', 'objectives', 'time', 'nation', 'person', 'mind', 'mind', 'must', 'kept', 'informed',
'alert', 'must', 'know', 'understands', 'hopes', 'needs', 'neighbors', 'nations', 'live', 'within', 'narrowing', 'circle', 'wor
ld', 'nation', 'person', 'something', 'deeper', 'something', 'permanent', 'something', 'larger', 'sum', 'parts', 'something',
'matters', 'future', 'calls', 'forth', 'sacred', 'guarding', 'present', 'thing', 'find', 'difficult', 'even', 'impossible', 'hi
t', 'upon', 'single', 'simple', 'word', 'yet', 'understand', 'spirit', 'faith', 'america', 'product', 'centuries', 'born', 'mul
titudes', 'came', 'many', 'lands', 'high', 'degree', 'mostly', 'plain', 'people', 'sought', 'early', 'late', 'find', 'freedom',
'freely', 'democratic', 'aspiration', 'mere', 'recent', 'phase', 'human', 'history', 'human', 'history', 'permeated', 'ancien
```

```
t', 'upon', 'single', 'simple', 'word', 'yet', 'understand', 'spirit', 'faith', 'america', 'product', 'centuries', 'born', 'mul
titudes', 'came', 'many', 'lands', 'high', 'degree', 'mostly', 'plain', 'people', 'sought', 'early', 'late', 'find', 'freedom',
'freely', 'democratic', 'aspiration', 'mere', 'recent', 'phase', 'human', 'history', 'human', 'history', 'permeated', 'ancien
t', 'life', 'early', 'peoples', 'blazed', 'anew', 'middle', 'ages', 'written', 'magna', 'charta', 'americas', 'impact', 'irresi
stible', 'america', 'new', 'world', 'tongues', 'peoples', 'continent', 'new', 'found', 'land', 'came', 'believed', 'create', 'u
pon', 'continent', 'new', 'life', 'life', 'new', 'freedom', 'vitality', 'written', 'mayflower', 'compact', 'declaration', 'inde
pendence', 'constitution', 'united', 'states', 'gettysburg', 'address', 'first', 'came', 'carry', 'longings', 'spirit', 'millio
ns', 'followed', 'stock', 'sprang', 'moved', 'forward', 'constantly', 'consistently', 'toward', 'ideal', 'gained', 'stature',
'clarity', 'generation', 'hopes', 'republic', 'forever', 'tolerate', 'either', 'undeserved', 'poverty', 'self', 'serving', 'wea
lth', 'know', 'still', 'far', 'go', 'must', 'greatly', 'build', 'security', 'opportunity', 'knowledge', 'every', 'citizen', 'me
asure', 'justified', 'resources', 'capacity', 'land', 'enough', 'achieve', 'purposes', 'alone', 'enough', 'clothe', 'feed', 'bo
dy', 'nation', 'instruct', 'inform', 'mind', 'spirit', 'three', 'greatest', 'spirit', 'without', 'body', 'mind', 'men', 'know',
'nation', 'live', 'spirit', 'america', 'killed', 'even', 'though', 'nation', 's', 'body', 'mind', 'constricted', 'alien', 'worl
d', 'lived', 'america', 'know', 'perished', 'spirit', 'faith', 'speaks', 'us', 'daily', 'lives', 'ways', 'often', 'unnoticed',
'seem', 'obvious', 'speaks', 'us', 'capital', 'nation', 'speaks', 'us', 'processes', 'governing', 'sovereignties', '48', 'state
s', 'speaks', 'us', 'counties', 'cities', 'towns', 'villages', 'speaks', 'us', 'nations', 'hemisphere', 'across', 'seas', 'ensl
aved', 'well', 'free', 'sometimes', 'fail', 'hear', 'heed', 'voices', 'freedom', 'us', 'privilege', 'freedom', 'old', 'old', 's
tory', 'destiny', 'america', 'proclaimed', 'words', 'prophecy', 'spoken', 'first', 'president', 'first', 'inaugural', '1789',
'words', 'almost', 'directed', 'seem', 'year', '1941', 'preservation', 'sacred', 'fire', 'liberty', 'destiny', 'republican', 'm
odel', 'government', 'justly', 'considered', 'deeply', 'finally', 'staked', 'experiment', 'intrusted', 'hands', 'american', 'pe
ople', 'lose', 'sacred', 'fire', 'let', 'smothered', 'doubt', 'fear', 'reject', 'destiny', 'washington', 'strove', 'valiantly',
'triumphantly', 'establish', 'preservation', 'spirit', 'faith', 'nation', 'will', 'furnish', 'highest', 'justification', 'ever
y', 'sacrifice', 'may', 'make', 'cause', 'national', 'defense', 'face', 'great', 'perils', 'never', 'encountered', 'strong', 'p
urpose', 'protect', 'perpetuate', 'integrity', 'democracy', 'muster', 'spirit', 'america', 'faith', 'america', 'retreat', 'cont
ent', 'stand', 'still', 'americans', 'go', 'forward', 'service', 'country', 'will', 'god']
```

1961 Kennedy’s Speech

After removing the Stop words:

```
['vice', 'president', 'johnson', 'mr', 'speaker', 'mr', 'chief', 'justice', 'president', 'eisenhower', 'vice', 'president', 'ni
xon', 'president', 'truman', 'reverend', 'clergy', 'fellow', 'citizens', 'observe', 'today', 'victory', 'party', 'celebration',
'freedom', 'symbolizing', 'end', 'well', 'beginning', 'signifying', 'renewal', 'well', 'change', 'sworn', 'almighty', 'god', 's
olemn', 'oath', 'forebears', 'l', 'prescribed', 'nearly', 'century', 'three', 'quarters', 'ago', 'world', 'different', 'now',
'man', 'holds', 'mortal', 'hands', 'power', 'abolish', 'forms', 'human', 'poverty', 'forms', 'human', 'life', 'yet', 'revolutio
nary', 'beliefs', 'forebears', 'fought', 'still', 'issue', 'around', 'globe', 'belief', 'rights', 'man', 'come', 'generosity',
'state', 'hand', 'god', 'dare', 'forget', 'today', 'heirs', 'first', 'revolution', 'let', 'word', 'go', 'forth', 'time', 'plac
e', 'friend', 'foe', 'alike', 'torch', 'passed', 'new', 'generation', 'americans', 'born', 'century', 'tempered', 'war', 'disci
plined', 'hard', 'bitter', 'peace', 'proud', 'ancient', 'heritage', 'unwilling', 'witness', 'permit', 'slow', 'undoing', 'huma
n', 'rights', 'nation', 'always', 'committed', 'committed', 'today', 'home', 'around', 'world', 'let', 'every', 'nation', 'kno
w', 'whether', 'wishes', 'us', 'well', 'ill', 'pay', 'price', 'bear', 'burden', 'meet', 'hardship', 'support', 'friend', 'oppos
e', 'foe', 'order', 'assure', 'survival', 'success', 'liberty', 'much', 'pledge', 'old', 'allies', 'whose', 'cultural', 'spirit
ual', 'origins', 'share', 'pledge', 'loyalty', 'faithful', 'friends', 'united', 'little', 'host', 'cooperative', 'ventures', 'd
ivided', 'little', 'dare', 'meet', 'powerful', 'challenge', 'odds', 'split', 'asunder', 'new', 'states', 'welcome', 'ranks', 'f
ree', 'pledge', 'word', 'one', 'form', 'colonial', 'control', 'passed', 'away', 'merely', 'replaced', 'far', 'iron', 'tyranny',
'always', 'expect', 'find', 'supporting', 'view', 'always', 'hope', 'find', 'strongly', 'supporting', 'freedom', 'remember', 'p
ast', 'foolishly', 'sought', 'power', 'riding', 'back', 'tiger', 'ended', 'inside', 'peoples', 'huts', 'villages', 'across', 'g
lobe', 'struggling', 'break', 'bonds', 'mass', 'misery', 'pledge', 'best', 'efforts', 'help', 'help', 'whatever', 'period', 're
quired', 'communists', 'may', 'seek', 'votes', 'right', 'free', 'society', 'help', 'many', 'poor', 'save', 'rich', 'sister', 'r
epublics', 'south', 'border', 'offer', 'special', 'pledge', 'convert', 'good', 'words', 'good', 'deeds', 'new', 'alliance', 'pr
ogress', 'assist', 'free', 'men', 'free', 'governments', 'casting', 'chains', 'poverty', 'peaceful', 'revolution', 'hope', 'bec
ome', 'prey', 'hostile', 'powers', 'let', 'neighbors', 'know', 'join', 'oppose', 'aggression', 'subversion', 'anywhere', 'ameri
cas', 'let', 'every', 'power', 'know', 'hemisphere', 'intends', 'remain', 'master', 'house', 'world', 'assembly', 'sovereign',
'states', 'united', 'nations', 'last', 'best', 'hope', 'age', 'instruments', 'war', 'far', 'outpaced', 'instruments', 'peace',
'renew', 'pledge', 'support', 'prevent', 'becoming', 'merely', 'forum', 'invective', 'strengthen', 'shield', 'new', 'weak', 'en
large', 'area', 'writ', 'may', 'run', 'finally', 'nations', 'make', 'adversary', 'offer', 'pledge', 'request', 'sides', 'begi
n', 'anew', 'quest', 'peace', 'dark', 'powers', 'destruction', 'unleashed', 'science', 'engulf', 'humanity', 'planned', 'accide
ntal', 'self', 'destruction', 'dare', 'tempt', 'weakness', 'arms', 'sufficient', 'beyond', 'doubt', 'certain', 'beyond', 'doub
t', 'will', 'never', 'employed', 'neither', 'two', 'great', 'powerful', 'groups', 'nations', 'take', 'comfort', 'present', 'cou
```

```
'racing', 'alter', 'uncertain', 'balance', 'terror', 'stays', 'hand', 'mankind', 's', 'final', 'war', 'let', 'us', 'begin', 'an
ew', 'remembering', 'sides', 'civility', 'sign', 'weakness', 'sincerity', 'always', 'subject', 'proof', 'let', 'us', 'never',
'negotiate', 'fear', 'let', 'us', 'never', 'fear', 'negotiate', 'let', 'sides', 'explore', 'problems', 'unite', 'us', 'instea
d', 'belaboring', 'problems', 'divide', 'us', 'let', 'sides', 'first', 'time', 'formulate', 'serious', 'precise', 'proposals',
'inspection', 'control', 'arms', 'bring', 'absolute', 'power', 'destroy', 'nations', 'absolute', 'control', 'nations', 'let',
'sides', 'seek', 'invoke', 'wonders', 'science', 'instead', 'terrors', 'together', 'let', 'us', 'explore', 'stars', 'conquer',
'deserts', 'eradicate', 'disease', 'tap', 'ocean', 'depths', 'encourage', 'arts', 'commerce', 'let', 'sides', 'unite', 'heed',
'corners', 'earth', 'command', 'isaiah', 'undo', 'heavy', 'burdens', 'let', 'oppressed', 'go', 'free', 'beachhead', 'cooperatio
n', 'may', 'push', 'back', 'jungle', 'suspicion', 'let', 'sides', 'join', 'creating', 'new', 'endeavor', 'new', 'balance', 'pow
er', 'new', 'world', 'law', 'strong', 'weak', 'secure', 'peace', 'preserved', 'will', 'finished', 'first', '100', 'days', 'wil
l', 'finished', 'first', '1', '000', 'days', 'life', 'administration', 'even', 'perhaps', 'lifetime', 'planet', 'let', 'us', 'b
egin', 'hands', 'fellow', 'citizens', 'mine', 'will', 'rest', 'final', 'success', 'failure', 'course', 'country', 'founded', 'g
eneration', 'americans', 'summoned', 'give', 'testimony', 'national', 'loyalty', 'graves', 'young', 'americans', 'answered', 'c
all', 'service', 'surround', 'globe', 'now', 'trumpet', 'summons', 'us', 'call', 'bear', 'arms', 'though', 'arms', 'need', 'cal
l', 'battle', 'though', 'embattled', 'call', 'bear', 'burden', 'long', 'twilight', 'struggle', 'year', 'year', 'rejoicing', 'ho
pe', 'patient', 'tribulation', 'struggle', 'common', 'enemies', 'man', 'tyranny', 'poverty', 'disease', 'war', 'forge', 'enemie
s', 'grand', 'global', 'alliance', 'north', 'south', 'east', 'west', 'assure', 'fruitful', 'life', 'mankind', 'will', 'join',
'historic', 'effort', 'long', 'history', 'world', 'generations', 'granted', 'role', 'defending', 'freedom', 'hour', 'maximum',
'danger', 'shrink', 'responsibility', 'welcome', 'believe', 'us', 'exchange', 'places', 'people', 'generation', 'energy', 'fait
h', 'devotion', 'bring', 'endeavor', 'will', 'light', 'country', 'serve', 'glow', 'fire', 'truly', 'light', 'world', 'fellow',
'americans', 'ask', 'country', 'ask', 'country', 'fellow', 'citizens', 'world', 'ask', 'america', 'will', 'together', 'freedo
m', 'man', 'finally', 'whether', 'citizens', 'america', 'citizens', 'world', 'ask', 'us', 'high', 'standards', 'strength', 'sac
rifice', 'ask', 'good', 'conscience', 'sure', 'reward', 'history', 'final', 'judge', 'deeds', 'let', 'us', 'go', 'forth', 'lea
d', 'land', 'love', 'asking', 'blessing', 'help', 'knowing', 'earth', 'god', 's', 'work', 'must', 'truly']
```

1973– Nixon's Speech

After removing the stop words from the speech:

```
['mr', 'vice', 'president', 'mr', 'speaker', 'mr', 'chief', 'justice', 'senator', 'cook', 'mrs', 'eisenhower', 'fellow', 'citiz
ens', 'great', 'good', 'country', 'share', 'together', 'met', 'four', 'years', 'ago', 'america', 'bleak', 'spirit', 'depresse
d', 'prospect', 'seemingly', 'endless', 'war', 'abroad', 'destructive', 'conflict', 'home', 'meet', 'today', 'stand', 'threshol
d', 'new', 'era', 'peace', 'world', 'central', 'question', 'us', 'use', 'peace', 'let', 'us', 'resolve', 'era', 'enter', 'wil
l', 'postwar', 'periods', 'often', 'time', 'retreat', 'isolation', 'leads', 'stagnation', 'home', 'invites', 'new', 'danger',
'abroad', 'let', 'us', 'resolve', 'will', 'become', 'time', 'great', 'responsibilities', 'greatly', 'borne', 'renew', 'spirit',
'promise', 'america', 'enter', 'third', 'century', 'nation', 'past', 'year', 'saw', 'far', 'reaching', 'results', 'new', 'polic
ies', 'peace', 'continuing', 'revitalize', 'traditional', 'friendships', 'missions', 'peking', 'moscow', 'able', 'establish',
'base', 'new', 'durable', 'pattern', 'relationships', 'among', 'nations', 'world', 'america', 's', 'bold', 'initiatives', '197
2', 'will', 'long', 'remembered', 'year', 'greatest', 'progress', 'end', 'world', 'war', 'ii', 'toward', 'lasting', 'peace', 'w
orld', 'peace', 'seek', 'world', 'flimsy', 'peace', 'merely', 'interlude', 'wars', 'peace', 'endure', 'generations', 'come', 'i
mportant', 'understand', 'necessity', 'limitations', 'america', 's', 'role', 'maintaining', 'peace', 'unless', 'america', 'wor
k', 'preserve', 'peace', 'will', 'peace', 'unless', 'america', 'work', 'preserve', 'freedom', 'will', 'freedom', 'let', 'us',
'clearly', 'understand', 'new', 'nature', 'america', 's', 'role', 'result', 'new', 'policies', 'adopted', 'past', 'four', 'year
s', 'respect', 'treaty', 'commitments', 'support', 'vigorously', 'principle', 'country', 'right', 'impose', 'will', 'rule', 'an
other', 'force', 'continue', 'era', 'negotiation', 'work', 'limitation', 'nuclear', 'arms', 'reduce', 'danger', 'confrontatio
n', 'great', 'powers', 'share', 'defending', 'peace', 'freedom', 'world', 'expect', 'others', 'share', 'time', 'passed', 'ameri
ca', 'will', 'make', 'every', 'nation', 's', 'conflict', 'make', 'every', 'nation', 's', 'future', 'responsibility', 'presume',
'tell', 'people', 'nations', 'manage', 'affairs', 'respect', 'right', 'nation', 'determine', 'future', 'recognize', 'responsibi
lity', 'nation', 'secure', 'future', 'america', 's', 'role', 'indispensable', 'preserving', 'world', 's', 'peace', 'nation',
's', 'role', 'indispensable', 'preserving', 'peace', 'together', 'rest', 'world', 'let', 'us', 'resolve', 'move', 'forward', 'b
eginnings', 'made', 'let', 'us', 'continue', 'bring', 'walls', 'hostility', 'divided', 'world', 'long', 'build', 'place', 'brid
ges', 'understanding', 'despite', 'profound', 'differences', 'systems', 'government', 'people', 'world', 'friends', 'let', 'u
s', 'build', 'structure', 'peace', 'world', 'weak', 'safe', 'strong', 'respects', 'right', 'live', 'different', 'system', 'infl
uence', 'others', 'will', 'strength', 'ideas', 'force', 'arms', 'let', 'us', 'accept', 'high', 'responsibility', 'burden', 'gla
dly', 'gladly', 'chance', 'build', 'peace', 'noblest', 'endeavor', 'nation', 'engage', 'gladly', 'act', 'greatly', 'meeting',
'responsibilities', 'abroad', 'will', 'remain', 'great', 'nation', 'remain', 'great', 'nation', 'will', 'act', 'greatly', 'meet
ing', 'challenges', 'home', 'chance', 'today', 'history', 'make', 'life', 'better', 'america', 'ensure', 'better', 'education',
'better', 'health', 'better', 'housing', 'better', 'transportation', 'cleaner', 'environment', 'restore', 'respect', 'law', 'ma
ke', 'communities', 'livable', 'insure', 'god', 'given', 'right', 'every', 'american', 'full', 'equal', 'opportunity', 'range',
```

```
onsibility', 'lived', 'long', 'consequences', 'attempting', 'gather', 'power', 'responsibility', 'washington', 'abroad', 'hom
e', 'time', 'come', 'turn', 'away', 'condescending', 'policies', 'paternalism', 'washington', 'knows', 'best', 'person', 'expec
ted', 'act', 'responsibly', 'responsibility', 'human', 'nature', 'let', 'us', 'encourage', 'individuals', 'home', 'nations', 'a
broad', 'decide', 'let', 'us', 'locate', 'responsibility', 'places', 'let', 'us', 'measure', 'will', 'others', 'will', 'today',
'offer', 'promise', 'purely', 'governmental', 'solution', 'every', 'problem', 'lived', 'long', 'false', 'promise', 'trusting',
'much', 'government', 'asked', 'deliver', 'leads', 'inflated', 'expectations', 'reduced', 'individual', 'effort', 'disappointme
nt', 'frustration', 'erode', 'confidence', 'government', 'people', 'government', 'must', 'learn', 'take', 'less', 'people', 'pe
ople', 'let', 'us', 'remember', 'america', 'built', 'government', 'people', 'welfare', 'work', 'shirking', 'responsibility', 's
eeking', 'responsibility', 'lives', 'let', 'us', 'ask', 'will', 'government', 'challenges', 'face', 'together', 'let', 'us', 'a
sk', 'government', 'help', 'help', 'national', 'government', 'great', 'vital', 'role', 'play', 'pledge', 'government', 'act',
'will', 'act', 'boldly', 'will', 'lead', 'boldly', 'important', 'role', 'every', 'one', 'us', 'must', 'play', 'individual', 'me
mber', 'community', 'day', 'forward', 'let', 'us', 'make', 'solemn', 'commitment', 'heart', 'bear', 'responsibility', 'part',
'live', 'ideals', 'together', 'see', 'dawn', 'new', 'age', 'progress', 'america', 'together', 'celebrate', '200th', 'anniversar
y', 'nation', 'proud', 'fulfillment', 'promise', 'world', 'america', 's', 'longest', 'difficult', 'war', 'comes', 'end', 'let',
'us', 'learn', 'debate', 'differences', 'civility', 'decency', 'let', 'us', 'reach', 'one', 'precious', 'quality', 'governmen
t', 'provide', 'new', 'level', 'respect', 'rights', 'feelings', 'one', 'another', 'new', 'level', 'respect', 'individual', 'hum
an', 'dignity', 'cherished', 'birthright', 'every', 'american', 'time', 'come', 'us', 'renew', 'faith', 'america', 'recent', 'y
ears', 'faith', 'challenged', 'children', 'taught', 'ashamed', 'country', 'ashamed', 'parents', 'ashamed', 'america', 's', 'rec
ord', 'home', 'role', 'world', 'every', 'turn', 'beset', 'find', 'everything', 'wrong', 'america', 'little', 'right', 'confiden
t', 'will', 'judgment', 'history', 'remarkable', 'times', 'privileged', 'live', 'america', 's', 'record', 'century', 'unparalle
led', 'world', 's', 'history', 'responsibility', 'generosity', 'creativity', 'progress', 'let', 'us', 'proud', 'system', 'produ
ced', 'provided', 'freedom', 'abundance', 'widely', 'shared', 'system', 'history', 'world', 'let', 'us', 'proud', 'four', 'war
s', 'engaged', 'century', 'including', 'one', 'now', 'bringing', 'end', 'fought', 'selfish', 'advantage', 'help', 'others', 're
sist', 'aggression', 'let', 'us', 'proud', 'bold', 'new', 'initiatives', 'steadfastness', 'peace', 'honor', 'made', 'break', 't
oward', 'creating', 'world', 'world', 'known', 'structure', 'peace', 'last', 'merely', 'time', 'generations', 'come', 'embarkin
g', 'today', 'era', 'presents', 'challenges', 'great', 'nation', 'generation', 'faced', 'answer', 'god', 'history', 'conscienc
e', 'way', 'use', 'years', 'stand', 'place', 'hallowed', 'history', 'think', 'others', 'stood', 'think', 'dreams', 'america',
'think', 'recognized', 'needed', 'help', 'far', 'beyond', 'order', 'make', 'dreams', 'come', 'true', 'today', 'ask', 'prayers',
'years', 'ahead', 'may', 'god', 's', 'help', 'making', 'decisions', 'right', 'america', 'pray', 'help', 'together', 'may', 'wor
thy', 'challenge', 'let', 'us', 'pledge', 'together', 'make', 'next', 'four', 'years', 'best', 'four', 'years', 'america', 's',
'history', '200th', 'birthday', 'america', 'will', 'young', 'vital', 'began', 'bright', 'beacon', 'hope', 'world', 'let', 'us',
'go', 'forward', 'confident', 'hope', 'strong', 'faith', 'one', 'another', 'sustained', 'faith', 'god', 'created', 'us', 'striv
ing', 'always', 'serve', 'purpose']
```

(3) Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (After removing the stopwords)

The top three words that have appeared in the three speeches are as below:

- Nation, People , Life, Spirit and Democracy are the words that are used most number of times after removing the stop words in 1941 Roosevelt speeches
- Let, World, Sides ,Nation, New and Power are the words that are used most number of times in the speeches of Kennedy in the year 1961
- Peace, America ,Let, Us, and World are the words that are used most number of times in the speeches of Nixon in the year 1973
- If we are looking for a combined version of all the three speeches the most common word will be World, People and Nation

(4) Plot the word cloud of each of the three speeches. (after removing the stop words)

Word Cloud: Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points are been highlighted using a word cloud.

Codes used:

```
stop_words = set(stopwords.words('english'))  #intialise stopwords from English Language

filtered_sentence = [] #empty list
for i in processed_features: # iterating in processes features through each sentence
    word_tokens = word_tokenize(i) # converting each sentence to a token
    for w in word_tokens:#in each token, removing stopwords from english language
        if w not in stop_words:
            filtered_sentence.append(w) #appending non-stopwords to filtered_sentence list
            comment_words = ' ' #empty string

            stop_words = set(STOPWORDS) #stopwords from Wordcloud

for words in filtered_sentence:
    comment_words = comment_words + words + ' ' #converting to string


wordcloud = WordCloud(width = 1000, height = 1000, #wordcloud image creation
                background_color ='white',
                stopwords = stop_words,
                min_font_size = 10).generate(comment_words)
```

1941 Roosevelt's Speech:

1961 Kennedy's speech

1973 Nixon's Speech