# Machine Learning Project

by Anisha Sharma

PGPDSBA.O.Mar23.A

Great Learning

## Problem 1:

You are hired by one of the leading news channels CNBE who wants to analyze 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 on the basis of the given information, to create an exit poll that will help in predicting overall win and seats covered by a particular party.

## Problem 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

coring guide (Rubric) - Machine Learning Project

| Criteria | Points |
| --- | --- |

**1.1)** Read the dataset. Describe the data briefly. Interpret the inferences for each. Initial steps like head() .info(), Data Types, etc . Null value check, Summary stats, Skewness must be discussed.

4

**1.2)** Perform EDA (Check the null values, Data types, shape, Univariate, bivariate analysis). Also check for outliers (4 pts). Interpret the inferences for each (3 pts) Distribution plots(histogram) or similar plots for the continuous columns. Box plots. Appropriate plots for categorical variables. Inferences on each plot. Outliers proportion should be discussed, and inferences from above used plots should be there. There is no restriction on how the learner wishes to implement this but the code should be able to represent the correct output and inferences should be logical and correct.

7

**1.3)** Encode the data (having string values) for Modelling. Is Scaling necessary here or not?( 2 pts), Data Split: Split the data into train and test (70:30) (2 pts). The learner is expected to check and comment about the difference in scale of different features on the bases of appropriate measure for example std dev, variance, etc. Should justify whether there is a necessity for scaling. Object data should be converted into categorical/numerical data to fit in the models. (pd.categorical().codes(), pd.get_dummies(drop_first=True)) Data split, ratio defined for the split, train-test split should be discussed.

4

**1.4) Apply Logistic Regression and LDA (Linear Discriminant Analysis) (2 pts). Interpret the inferences of both model s (2 pts). Successful implementation of each model. Logical reason should be shared if any custom changes are made to the parameters while building the model. Calculate Train and Test Accuracies for each model. Comment on the validness of models (over fitting or under fitting)**

4

**1.5) Apply KNN Model and Naïve Bayes Model (2pts). Interpret the inferences of each model (2 pts). Successful implementation of each model. Logical reason should be shared if any custom changes are made to the parameters while building the model. Calculate Train and Test Accuracies for each model. Comment on the validness of models (over fitting or under fitting)**

4

**1.6) Model Tuning (4 pts) , Bagging ( 1.5 pts) and Boosting (1.5 pts). Apply grid search on each model (include all models) and make models on best_params. Compare and comment on performances of all. Comment on feature importance if applicable. Successful implementation of both algorithms along with inferences and comments on the model performances.**

7

**1.7 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, classification report (4 pts) Final Model - Compare and comment on all models on the basis of the performance metrics in a structured tabular manner. Describe on which model is best/optimized, After comparison which model suits the best for the problem in hand on the basis of different measures. Comment on the final model.(3 pts)**

7

**1.8) Based on your analysis and working on the business problem, detail out appropriate insights and recommendations to help the management solve the business objective. There should be at least 3-4 Recommendations and insights in total. Recommendations should be easily understandable and business specific, students should not give any technical suggestions. Full marks should only be allotted if the recommendations are correct and business specific.**

5

**2.1) Find the number of characters, words and sentences for the mentioned documents. (Hint: use .words(), .raw(), .sent() for extracting counts)**

3

**2.2) Remove all the stopwords from the three speeches. Show the word count before and after the removal of stopwords. Show a sample sentence after the removal of stopwords.**

3

| | |
|---|---|
| **2.3) Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stopwords)** | 3 |
| **2.4) Plot the word cloud of each of the three speeches. (after removing the stopwords)** | 3 |
| **Quality of Business Report (Please refer to the Evaluation Guidelines for Business report checklist. Marks in this criteria are at the moderator's discretion)** | 6 |
| **Please reflect on all that you learnt and fill this reflection: https://docs.google.com/forms/d/e/1FAI pQLSfqHHlmJyUkniiBiejtudIuRFk_TVCLe 843wfX6Iu3QNRPmng/viewform?usp=sf_ link** | 0 |
| Points | 60 |

## Problem 1:

1.1 Read the dataset. Do the descriptive statistics and do the null value condition check.

Solution: Head of data:-

| | 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 |

## Tail of data:-

| | Unnamed: 0 | vote | age | economic.cond.national | economic.cond.household | Blair | Hague | Europe | political.knowledge | gender |
|---|---|---|---|---|---|---|---|---|---|---|
| 1520 | 1521 | Conservative | 67 | 5 | 3 | 2 | 4 | 11 | 3 | male |
| 1521 | 1522 | Conservative | 73 | 2 | 2 | 4 | 4 | 8 | 2 | male |
| 1522 | 1523 | Labour | 37 | 3 | 3 | 5 | 4 | 2 | 2 | male |
| 1523 | 1524 | Conservative | 61 | 3 | 3 | 1 | 4 | 11 | 2 | male |
| 1524 | 1525 | Conservative | 74 | 2 | 3 | 2 | 4 | 11 | 0 | female |

## Data summary:-

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| Unnamed: 0 | 1525.0 | 763.000000 | 440.373894 | 1.0 | 382.0 | 763.0 | 1144.0 | 1525.0 |
| age | 1525.0 | 54.182295 | 15.711209 | 24.0 | 41.0 | 53.0 | 67.0 | 93.0 |
| economic.cond.national | 1525.0 | 3.245902 | 0.880969 | 1.0 | 3.0 | 3.0 | 4.0 | 5.0 |
| economic.cond.household | 1525.0 | 3.140328 | 0.929951 | 1.0 | 3.0 | 3.0 | 4.0 | 5.0 |
| Blair | 1525.0 | 3.334426 | 1.174824 | 1.0 | 2.0 | 4.0 | 4.0 | 5.0 |
| Hague | 1525.0 | 2.746885 | 1.230703 | 1.0 | 2.0 | 2.0 | 4.0 | 5.0 |
| Europe | 1525.0 | 6.728525 | 3.297538 | 1.0 | 4.0 | 6.0 | 10.0 | 11.0 |
| political.knowledge | 1525.0 | 1.542295 | 1.083315 | 0.0 | 0.0 | 2.0 | 2.0 | 3.0 |

## Info of datatype:-

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

## Insights:

- There are 1525 rows and 10 columns in the dataset.

- Vote is the dependent variable which has two attributes labour and conservative.
- Dependent variable includes age, Hague, Blair, economic.cond.national, economic.cond.household, political knowledge, gender.
- There is no null value in the dataset.
- There is no duplicate value in the dataset.
- Vote and gender are object datatype and rest are int64.
- The overall mean of the age of age is 54.18 with the standard deviation of 15.71.
- There is huge age gap of voters, the min is 24 and max is 93.
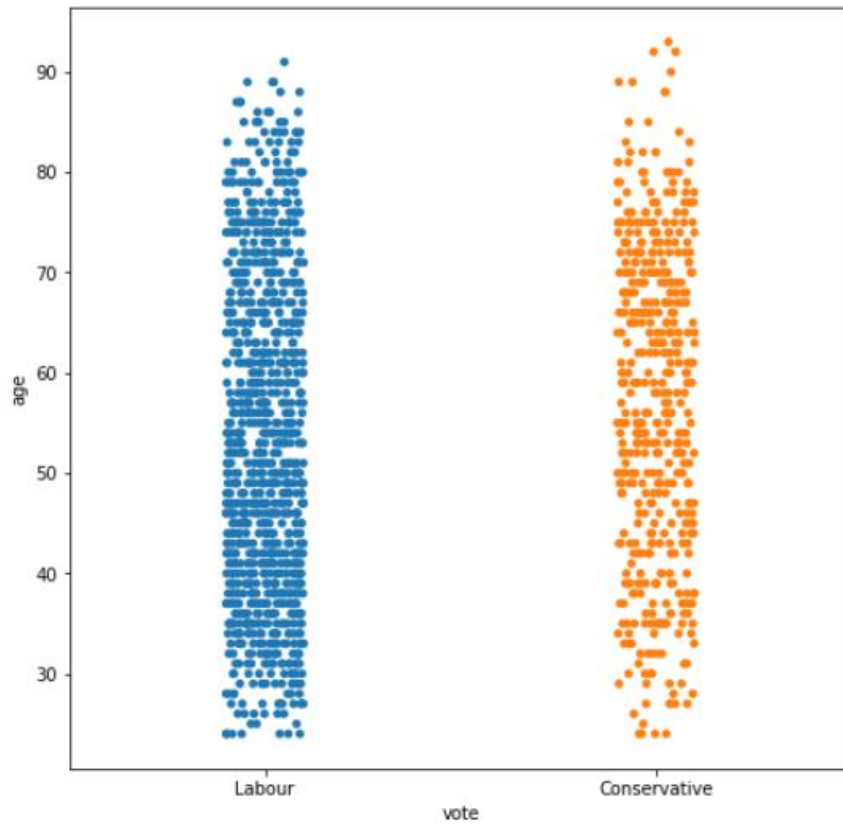
1.2 Perform Univariate and Bivariate Analysis. Do exploratory data analysis.
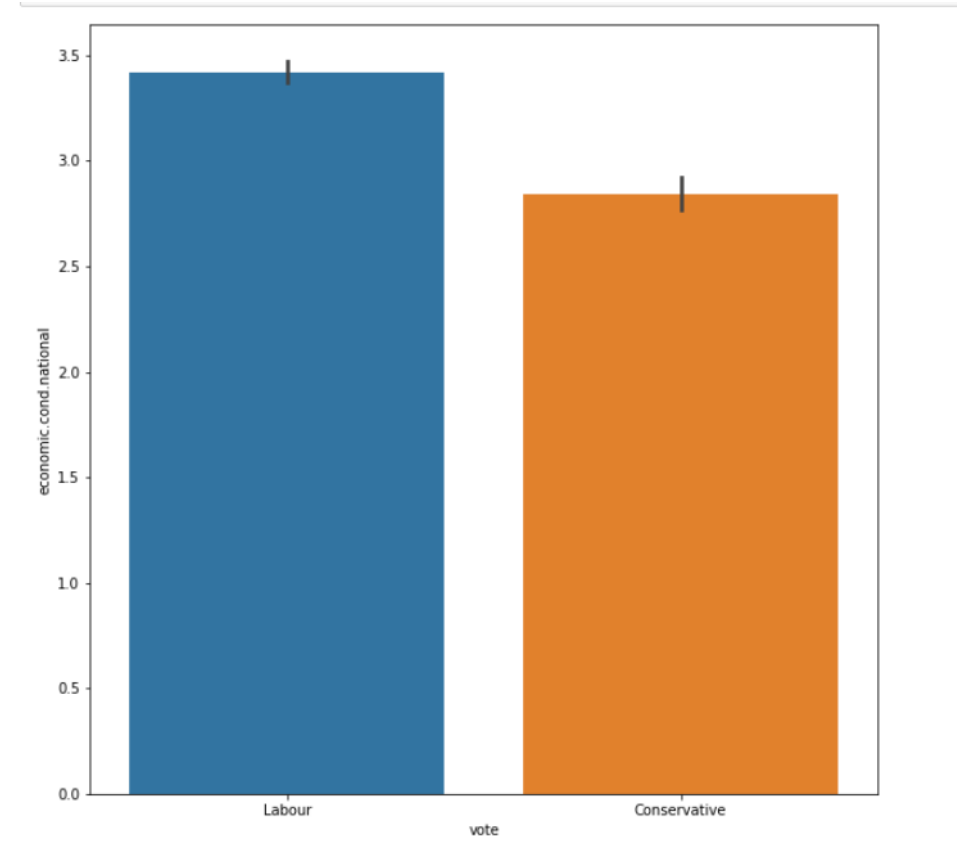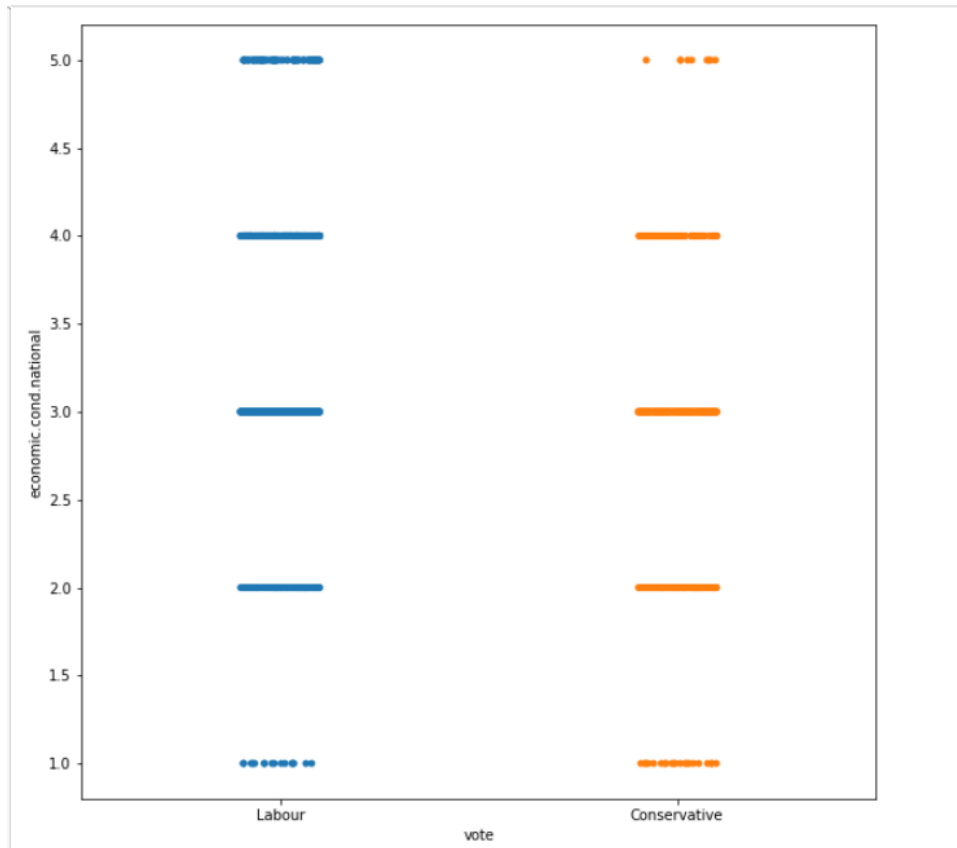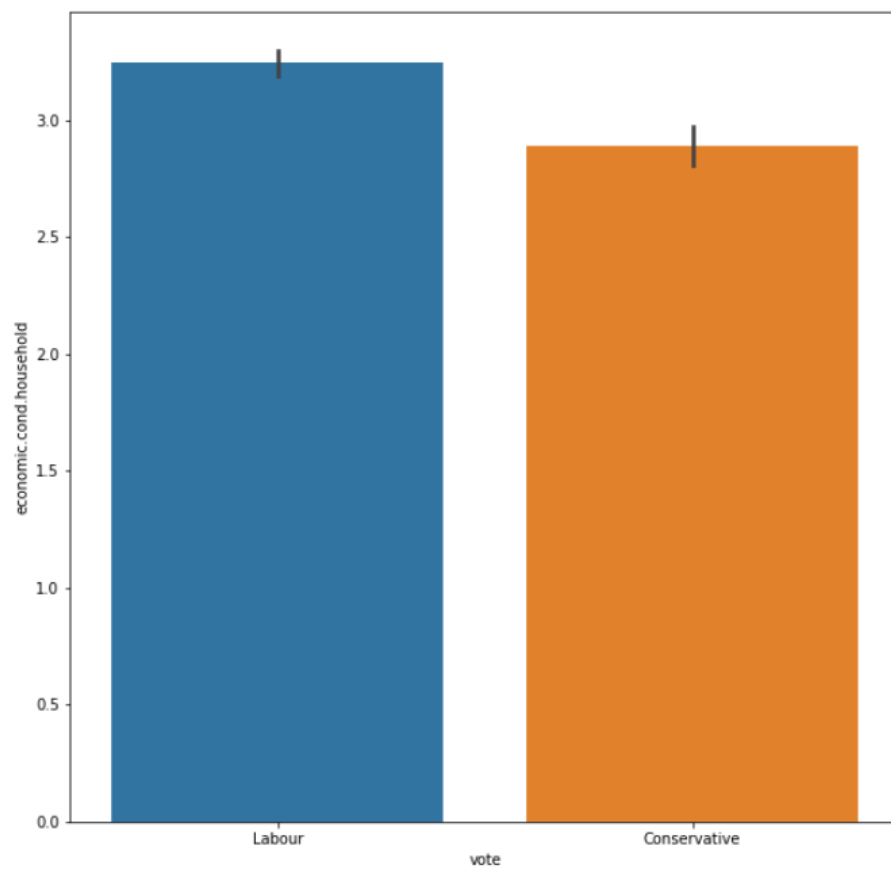
Solution:

Univariate Analysis-
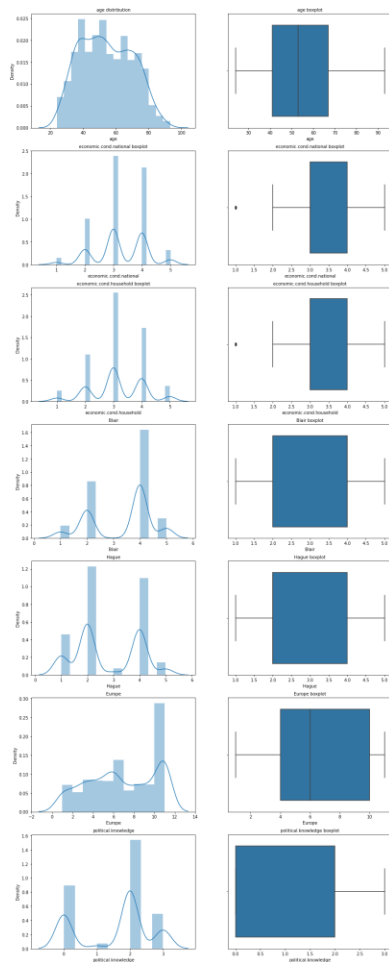
| vote | age | economic.cond.national | economic.cond.household | Blair | Hague | Europe | political.knowledge | gender |
|---|---|---|---|---|---|---|---|---|
| Conservative | 462 | 462 | 462 | 462 | 462 | 462 | 462 | 462 |
| Labour | 1063 | 1063 | 1063 | 1063 | 1063 | 1063 | 1063 | 1063 |

# Bivariate Analysis-
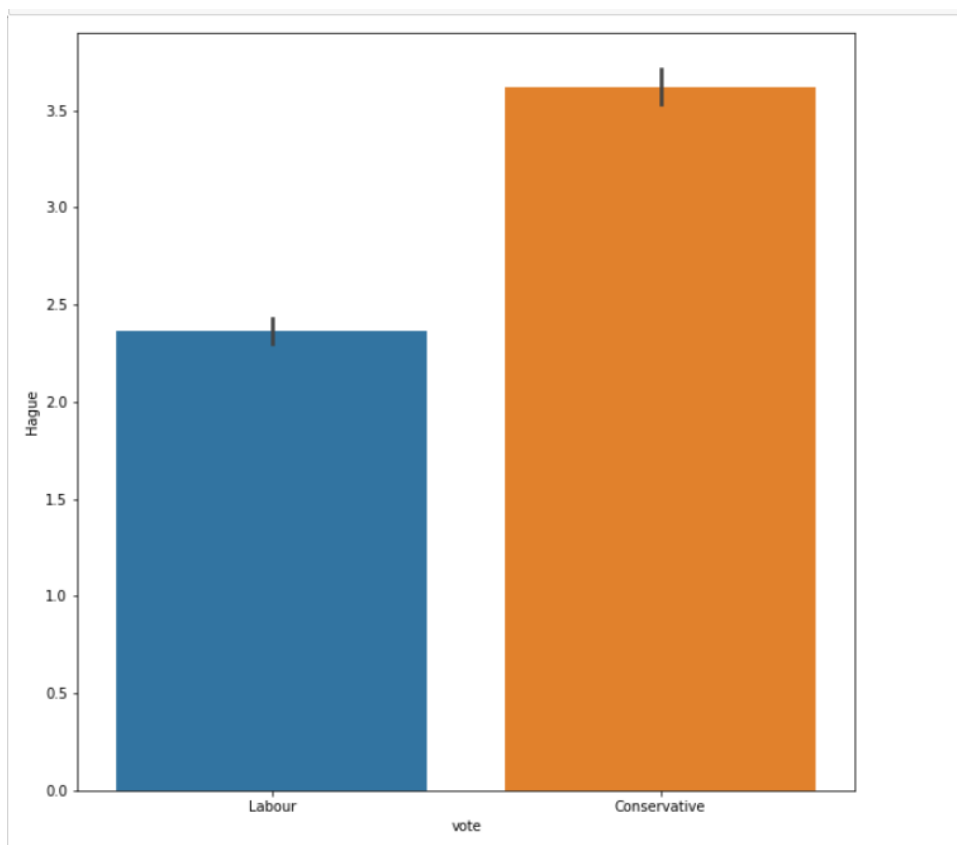
```
details of vote
Labour          1063
Conservative     462
Name: vote, dtype: int64
```
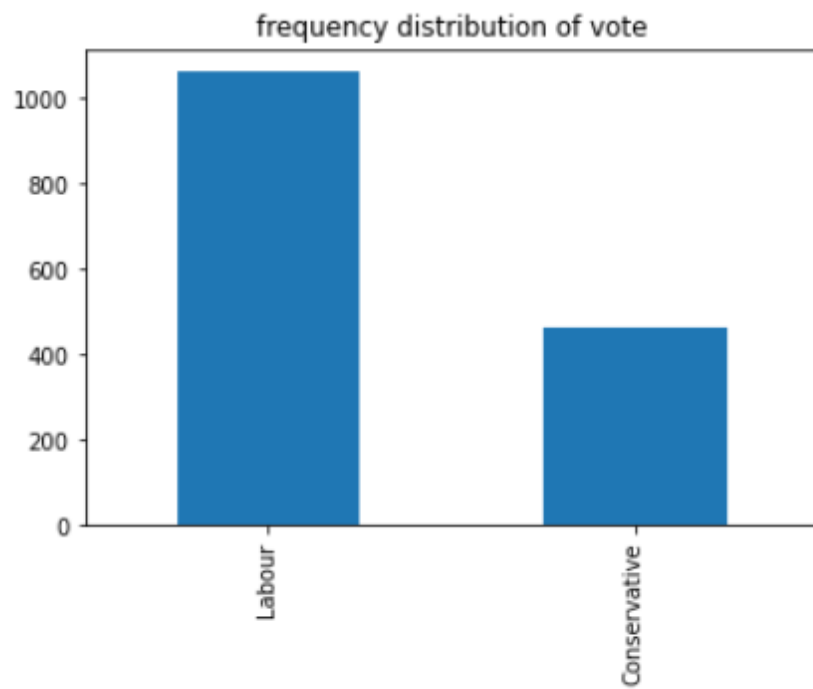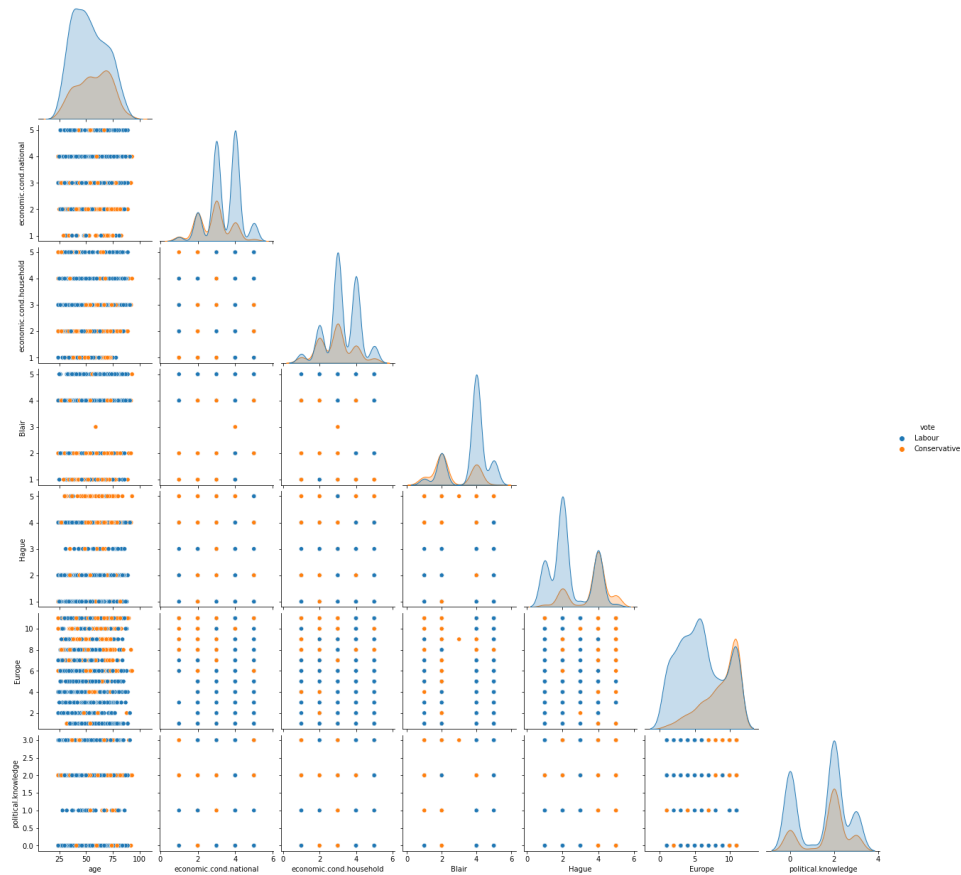
**frequency distribution of vote**

```
details of gender
female    812
male      713
Name: gender, dtype: int64
```



frequency distribution of gender
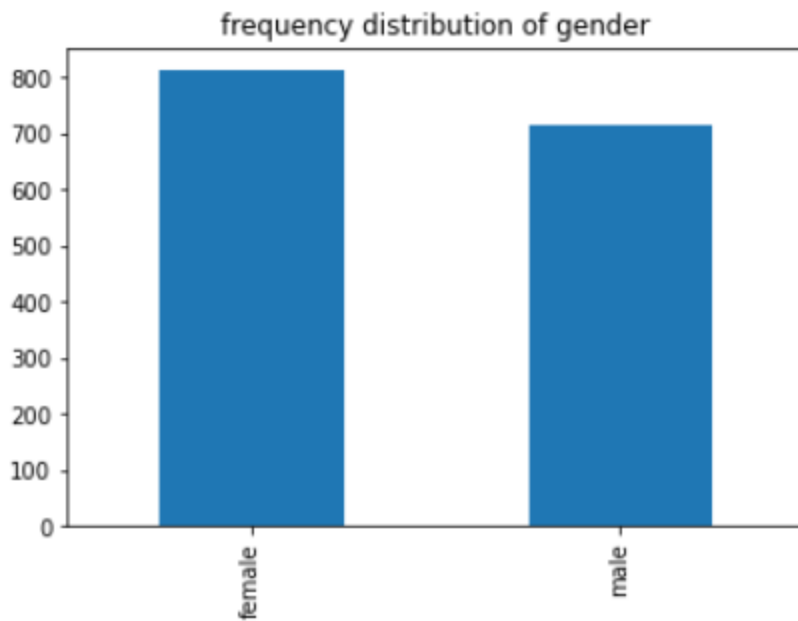
Multivariate Analysis-
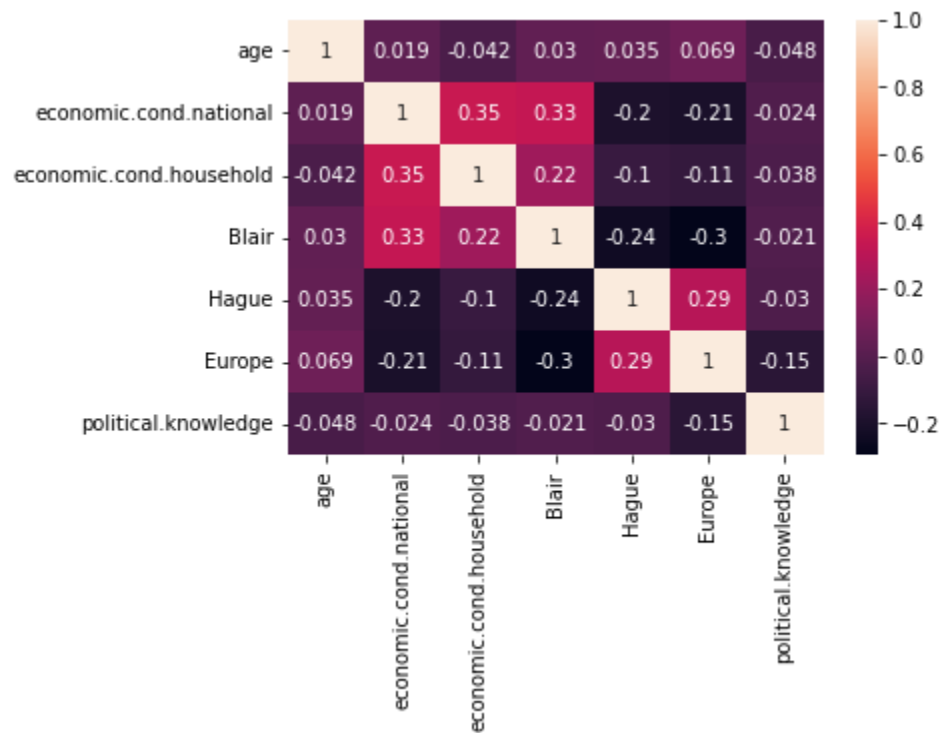
Insights:-

- <u>Univariate Analysis</u>: Major observations i.e 1057 belong to Labour class and only 460 belong to Conservative class, which is even half of the voters achieved by labour party.
- There are one outliers in each economic.cond.household and economic.cond.national, we won't be removing outliers.
- <u>Bivariate Analysis</u>: More frequency can be observed in Labour category.
- It appears that oldest person has voted for conservative party.
- <u>Multivate Analysis:</u> Looking at the distributions it is clearly visible that classes are fairly separted and thus can predict dependent variables efficiently.
- Looking at the correlation plot it is visible that there is no significant correlation between the variables.

1.3 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).

Solution:

- We will be using label encoding for vote which is our dependant variable, we are assigning 1 to Labour and 0 to Conservative.
- We will be using one hot encoding for gender and will not drop any variables.

<u>Train – Test split & Model Building:</u>

- To build the Machine Learning models, we split the entire data set into a ratio of 70:30 into Training dataset and Testing dataset.
- Since there are values ranging from 0 to 11 in the dependent variable, we need to ensure that an equal number of these values are split into both Training and Testing datasets.
- This will ensure a balance in the data and will not cause biasness while Training or Testing the model.
- Therefore, a function stratify=vote is used while splitting.

Confusion matrix on training and testing data
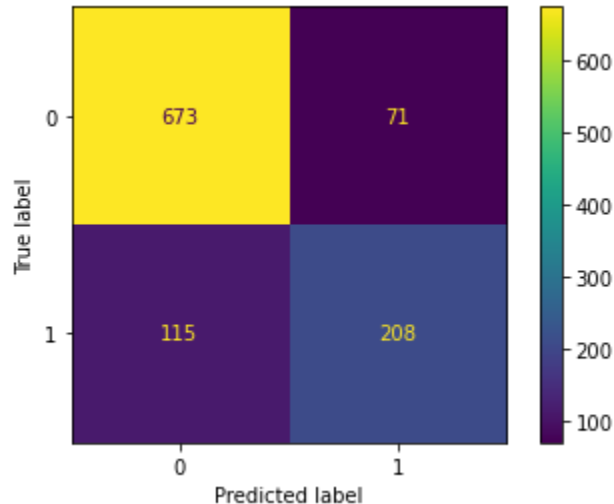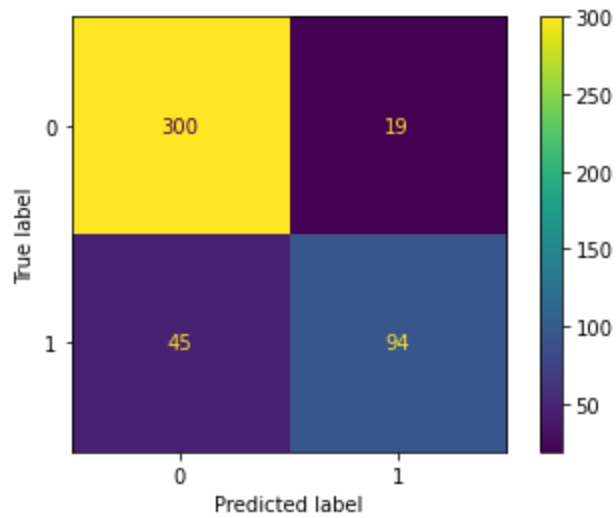
## AUC and ROC on training and testing data

Training AUC: 0.879
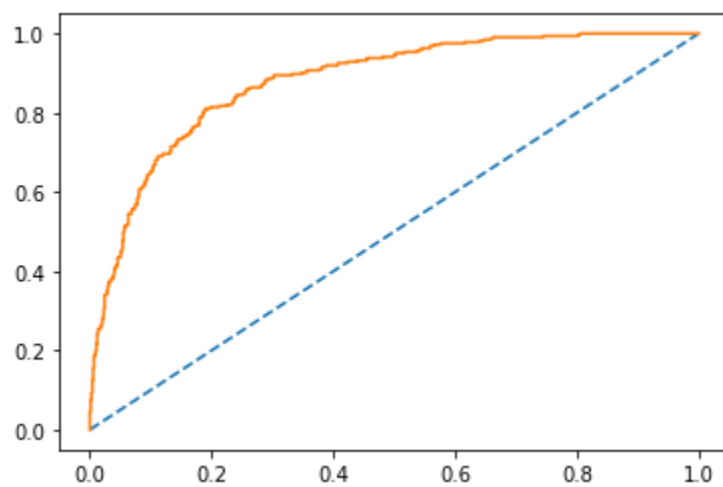


Testing AUC: 0.912
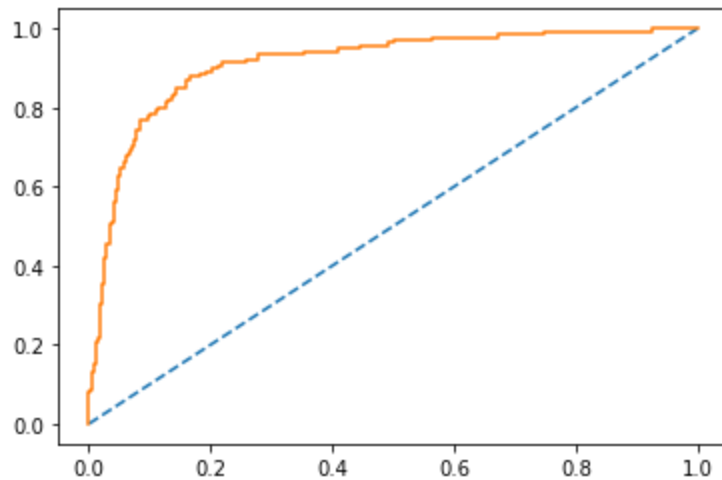
1.4) Apply Logistic Regression and LDA (Linear Discriminant Analysis) (2 pts). Interpret the inferences of both model s (2 pts). Successful implementation of each model. Logical reason should be shared if any custom changes are made to the parameters while building the model. Calculate Train and Test Accuracies for each model. Comment on the validness of models (over fitting or under fitting).

Solution:

Applying logistic regression:

On training data

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.85 | 0.90 | 0.88 | 744 |
| 1 | 0.75 | 0.64 | 0.69 | 323 |
| accuracy | | | 0.83 | 1067 |
| macro avg | 0.80 | 0.77 | 0.78 | 1067 |
| weighted avg | 0.82 | 0.83 | 0.82 | 1067 |

# On testing data

```
              precision    recall  f1-score   support

           0       0.87      0.94      0.90       319
           1       0.83      0.68      0.75       139

    accuracy                           0.86       458
   macro avg       0.85      0.81      0.82       458
weighted avg       0.86      0.86      0.86       458
```

```
Text(0.5, 1.0, 'Testing Data')
```

```
Classification Report of the training data:

                precision    recall  f1-score   support

            0       0.86      0.89      0.88       744
            1       0.72      0.68      0.70       323

     accuracy                           0.82      1067
    macro avg       0.79      0.78      0.79      1067
 weighted avg       0.82      0.82      0.82      1067


Classification Report of the testing data:

                precision    recall  f1-score   support

            0       0.87      0.93      0.90       319
            1       0.81      0.69      0.75       139

     accuracy                           0.86       458
    macro avg       0.84      0.81      0.82       458
 weighted avg       0.86      0.86      0.85       458
```
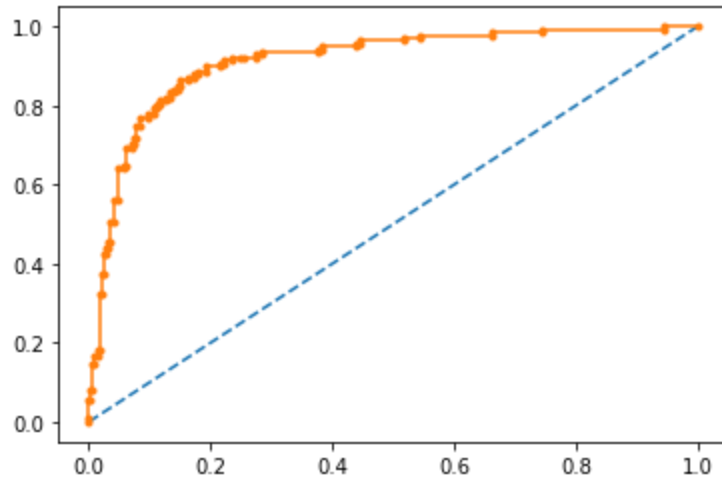
Applying Gaussian naïve bayes:

On training data:

```
0.8200562324273665
[[653  91]
 [101 222]]
              precision    recall  f1-score   support

           0       0.87      0.88      0.87       744
           1       0.71      0.69      0.70       323

    accuracy                           0.82      1067
   macro avg       0.79      0.78      0.78      1067
weighted avg       0.82      0.82      0.82      1067
```
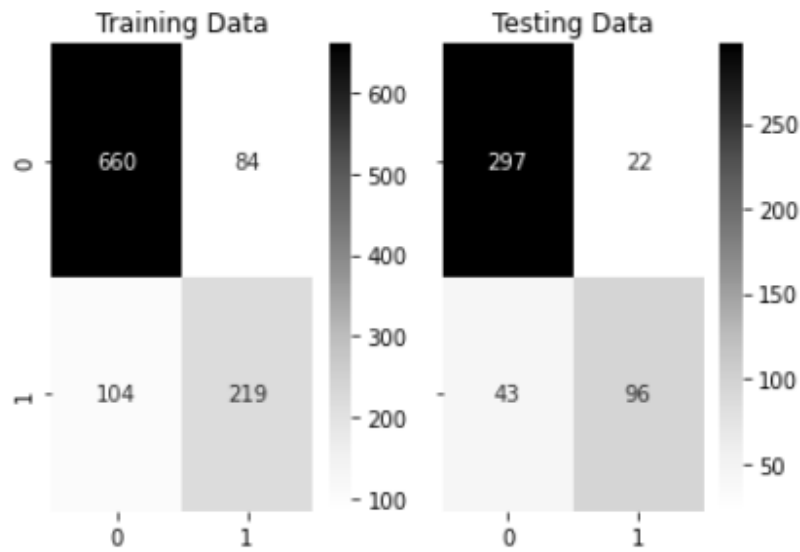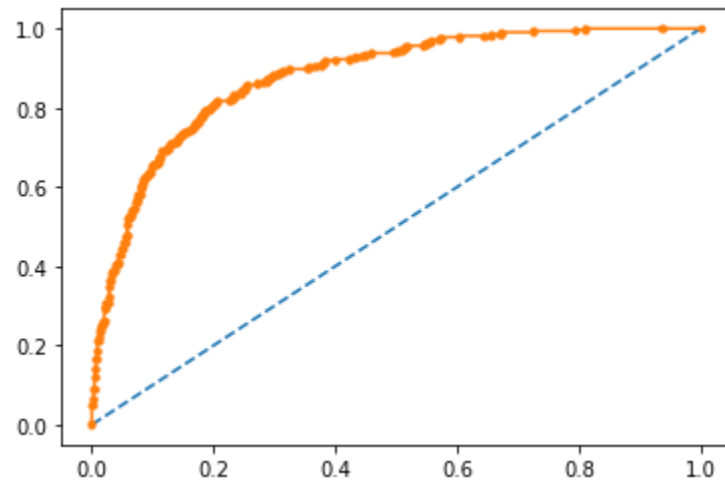
On testing data:

```
0.8558951965065502
[[292  27]
 [ 39 100]]
              precision    recall  f1-score   support

           0       0.88      0.92      0.90       319
           1       0.79      0.72      0.75       139

    accuracy                           0.86       458
   macro avg       0.83      0.82      0.83       458
weighted avg       0.85      0.86      0.85       458
```

```
NB_model.score(X_train, y_train)
```

0.8200562324273665

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

0.8558951965065502

1.5) Apply KNN Model and Naïve Bayes Model (2pts). Interpret the inferences of each model (2 pts). Successful implementation of each model. Logical reason should be shared if any custom changes are made to the parameters while building the model. Calculate Train and Test Accuracies for each model. Comment on the validness of models (over fitting or under fitting).

Solution:

Applying KNN model on training and testing data.

```
0.8444236176194939
[[675  69]
 [ 97 226]]
              precision   recall  f1-score   support

           0       0.87     0.91      0.89       744
           1       0.77     0.70      0.73       323

    accuracy                         0.84      1067
   macro avg       0.82     0.80      0.81      1067
weighted avg       0.84     0.84      0.84      1067



0.7816593886462883
[[277  42]
 [ 58  81]]
              precision   recall  f1-score   support

           0       0.83     0.87      0.85       319
           1       0.66     0.58      0.62       139

    accuracy                         0.78       458
   macro avg       0.74     0.73      0.73       458
weighted avg       0.78     0.78      0.78       458
```

Performance matrix on train data set for kneighborsclassifier(n_neighbors=7)

```
0.8416119962511716
[[682  62]
 [107 216]]
              precision   recall  f1-score   support

           0       0.86     0.92      0.89       744
           1       0.78     0.67      0.72       323

    accuracy                         0.84      1067
   macro avg       0.82     0.79      0.80      1067
weighted avg       0.84     0.84      0.84      1067
```

Performance matrix on test data set for kneighborsclassifier(n_neighbors=7)

```
0.7947598253275109
[[284  35]
 [ 59  80]]
              precision    recall  f1-score   support

           0       0.83      0.89      0.86       319
           1       0.70      0.58      0.63       139

    accuracy                           0.79       458
   macro avg       0.76      0.73      0.74       458
weighted avg       0.79      0.79      0.79       458
```

plot misclassification error vs k with k value on x-axis using matplotlib.
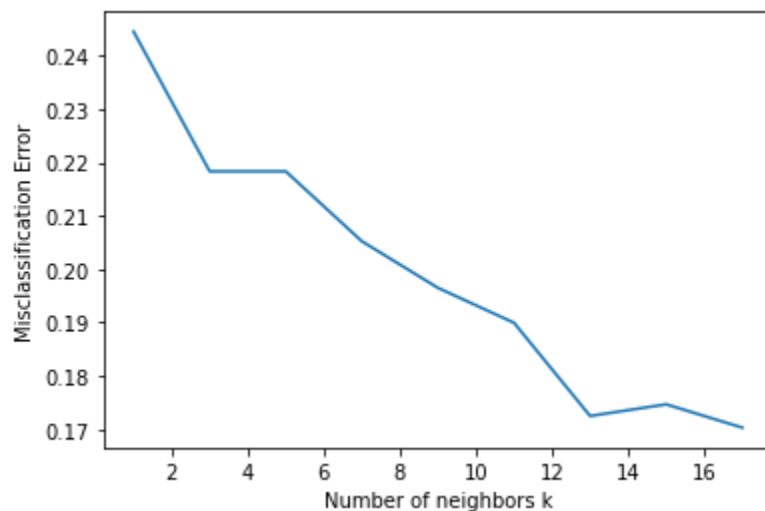


for k=15 it is giving the best test accuracy lets check train and test for k = 17 with other evaluation metrics.

Performance matrix on training data set for kneighborsclassifier(n_neighbors=15)

```
0.823805060918463
[[688  56]
 [132 191]]
              precision    recall  f1-score   support

           0       0.84      0.92      0.88       744
           1       0.77      0.59      0.67       323

    accuracy                           0.82      1067
   macro avg       0.81      0.76      0.77      1067
weighted avg       0.82      0.82      0.82      1067
```

Performance matrix on test data set for kneighborsclassifier(n_neighbors=15)

```
0.8253275109170306
[[299  20]
 [ 60  79]]
              precision    recall  f1-score   support

           0       0.83      0.94      0.88       319
           1       0.80      0.57      0.66       139

    accuracy                           0.83       458
   macro avg       0.82      0.75      0.77       458
weighted avg       0.82      0.83      0.82       458
```

1.6) Model Tuning (4 pts) , Bagging ( 1.5 pts) and Boosting (1.5 pts). Apply grid search on each model (include all models) and make models on best_params. Compare and comment on performances of all. Comment on feature importance if applicable. Successful implementation of both algorithms along with inferences and comments on the model performances. Solution:

Applying ada boosting: Performance matrix on train dataset

```
              precision    recall  f1-score   support

          0       0.78      0.73      0.75       322
          1       0.89      0.91      0.90       739

   accuracy                           0.86      1061
  macro avg       0.83      0.82      0.83      1061
weighted avg       0.85      0.86      0.85      1061
```

Performance Matrix on test data set

```
              precision    recall  f1-score   support

          0       0.70      0.63      0.66       138
          1       0.85      0.88      0.86       318

   accuracy                           0.80       456
  macro avg       0.77      0.76      0.76       456
weighted avg       0.80      0.80      0.80       456
```

## COMPARISON OF DIFFERENT MODELS

Interest Classes are both 1 and 0

Let's look at the performance of all the models on the Train Data set

For class 0 and 1

| Model | Accuracy | F1 score | Recall |
|-------|----------|----------|--------|
| Logistic Regression | 84% | 72%, 89% | 68%, 91% |
| Logistic Regression | 84% | 72%, 89% | 67%, 91% |

| | | | |
|---|---|---|---|
| after hyper tuning | | | |
| LDA | 84% | 72%, 89% | 69%,90% |
| Gaussian Naïve Bayes | 83% | 72%, 88% | 72%, 88% |
| KNN Model | 86% | 76%, 90% | 73%, 91% |
| Naïve Bayes Model | 83% | 72%, 88% | 72%, 88% |
| Ada Boost | 86% | 75%, 90% | 73%, 91% |
| Gradient Boosting | 89% | 82%, 92% | 80%, 93% |
| Random Forest | 100% | 100%, 100% | 100%,100% |

So as per the train data,

Worst performing models are - Linear Discriminant Analysis

Best Performing models are -KNN, lda, Gaussian Naïve, Random forest.

Although random forest is overfitted, we apply SMOTE on some best performing models to see if the model has improved further.

Problem 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

Inaugural is having following speeches

```
1789-Washington.txt',
 '1793-Washington.txt',
 '1797-Adams.txt',
 '1801-Jefferson.txt',
 '1805-Jefferson.txt',
 '1809-Madison.txt',
 '1813-Madison.txt',
 '1817-Monroe.txt',
 '1821-Monroe.txt',
 '1825-Adams.txt',
 '1829-Jackson.txt',
 '1833-Jackson.txt',
 '1837-VanBuren.txt',
 '1841-Harrison.txt',
 '1845-Polk.txt',
 '1849-Taylor.txt',
 '1853-Pierce.txt',
 '1857-Buchanan.txt',
 '1861-Lincoln.txt',
 '1865-Lincoln.txt',
 '1869-Grant.txt',
 '1873-Grant.txt',
 '1877-Hayes.txt',
 '1881-Garfield.txt',
 '1885-Cleveland.txt',
 '1889-Harrison.txt',
 '1893-Cleveland.txt',
 '1897-McKinley.txt',
 '1901-McKinley.txt',
 '1905-Roosevelt.txt',
 '1909-Taft.txt',
 '1913-Wilson.txt',
 '1917-Wilson.txt',
 '1921-Harding.txt',
 '1925-Coolidge.txt',
 '1929-Hoover.txt',
 '1933-Roosevelt.txt',
 '1937-Roosevelt.txt',
 '1941-Roosevelt.txt',
 '1945-Roosevelt.txt',
 '1949-Truman.txt',
 '1953-Eisenhower.txt',
```

## Speech 1-1941-Roosevelt

'On each national day of inauguration since 1789, the people have renewed 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 from 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\nTo 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, we risk the real peril of inaction.\n\nLives of nations are determined not 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 less. The life of a nation is the fullness of the measure of its will to live.\n\nThere are men who doubt this. There are men who believe that democracy, as a form of Government and a frame of life, is limited or measured by a kind of mystical and artificial fate that, for some unexplained reason, tyranny and slavery have become the surging wave of the future -- and that freedom is an ebbing tide.\n\nBut we Americans know that 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 years for the people of this democracy. For they have brought to us greater security and, I hope, a better understanding that life\'s ideals are to be measured in other than material things.\n\nMost vital to our present and our future is this experience of a democracy which successfully survived crisis at home; put away many evil things; built new structures 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 the Government continue freely to function. The Bill of Rights remains inviolate. 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 it revive--and grow.\n\nWe know it cannot die -- because it is built on the unhampered initiative of individual men and women joined together in a common enterprise -- an enterprise undertaken and carried through by the free expression of a free majority.\n\nWe know it because democracy 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 improvement 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 humane, the most advanced, and in the end the most unconquerable of all forms of human society.\n\nA nation, like a person, has a body--a body that must be fed and clothed and housed, invigorated and rested, in a manner 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, that must know itself, that understands the hopes and the needs of its neighbors -- all the other nations that live within the narrowing circle of the world.\n\nAnd a nation, like a person, has something deeper, something more permanent, something larger than the sum of all its parts. It is that something which matters most to its future -- which calls forth 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 faith of America. It is the product of centuries. It was born in the multitudes of those who came from many lands -- some of high degree, but mostly plain people, who sought here, early and late, to find freedom more freely.\n\nThe democratic aspiration is no mere recent phase in human history. It is human history. It permeated the ancient life of early peoples. 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 continent was a new-found land, but because all those who came here believed they 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 Compact, into the Declaration of Independence, into the Constitution of the United States, into the Gettysburg Address.\n\nThose who first came here to carry out the longings of their spirit, and the millions who followed, and the stock that sprang from them -- all have moved forward constantly and consistently toward an ideal which in itself has gained stature and clarity with each generation.\n\nThe hopes of the Republic cannot 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 the spirit.\n\nWithout the body and the mind, as all men know, the Nation could not live.\n\nBut if the spirit of America were killed, even though the Nation\'s body and mind, constricted in an alien world, lived on, the America we know would have perished.\n\nThat spirit -- that faith -- speaks to us in our daily lives in ways often unnoticed, because they 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 towns, 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 because to us the privilege of our freedom is such an old, old story.\n\nThe 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 are justly considered deeply, finally, staked on the experiment intrusted 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 perils never before encountered, our strong purpose is to protect and to perpetuate the integrity of democracy.\n\nFor this we muster the spirit of America, and the faith of America.\n\nWe do not retreat. We are not content to stand still. As Americans, we go forward, in the service of our country, by the will of God.\n'

## Speech 2- 1961-Kennedy

Vice President Johnson, Mr. Speaker, Mr. Chief Justice, President Eisenhower, 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 renewal, 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 three 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 all 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 from 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 place, to friend and foe alike, that the torch has been passed to a new generation of Americans -- born in this century, tempered by war, disciplined by a hard and bitter peace, proud of our ancient heritage -- and unwilling to witness or permit the slow undoing of those human rights to which this Nation has always been committed, and to which we are committed today at home and around the world.\n\nLet every nation know, whether it wishes us well or ill, that we shall pay any price, bear any burden, 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 origins we share, we pledge the loyalty of faithful friends. United, there is little we cannot do in a host of cooperative ventures. Divided, there is 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 shall always hope to find them strongly supporting their own freedom -- and to remember that, in the past, those who foolishly sought power by riding 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 mass 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 are rich.\n\nTo our sister republics south of our border, we offer a special pledge -- to convert our good words into good deeds -- in a new alliance for progress -- to assist free men and free governments in casting off the chains of poverty. But this peaceful revolution of hope cannot become the prey of hostile powers. Let all our neighbors know that we shall join with them to oppose aggression or subversion anywhere in the Americas. And let every other power know that this Hemisphere intends to remain the master of its own house.\n\nTo that world assembly of sovereign states, the United Nations, our last best hope in an age where the

instruments of war have far outpaced the instruments of peace, we renew our pledge of support--to prevent it from becoming merely a forum for invective -- to strengthen its shield of the new and the weak -- and to enlarge the area in which its writ may run.\n\nFinally, to those nations who would make themselves our adversary, we offer not a pledge but a request: that both sides begin anew the quest for peace, before the dark powers of destruction unleashed by science engulf all humanity in planned or accidental self-destruction.\n\nWe dare not tempt them with weakness. For only when our arms are sufficient beyond doubt can we be certain beyond doubt that they will never be employed.\n\nBut neither can two great and powerful groups of nations take comfort from our present course -- both sides overburdened by the cost of modern weapons, both rightly alarmed by the steady spread of the deadly atom, yet both racing to alter that uncertain balance of terror that stays the hand of mankind\'s final war.\n\nSo let us begin anew -- remembering on both sides that civility is not a sign of weakness, and sincerity is always subject to proof. Let us never negotiate out of fear. But let us never fear to negotiate.\n\nLet both sides explore what problems unite us instead of belaboring those problems which divide us.\n\nLet both sides, for the first time, formulate serious and precise proposals for the inspection and control of arms -- and bring the absolute power to destroy other 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 the 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 both 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 and the peace preserved.\n\nAll this will not be finished in the first 100 days. Nor will it be finished in the first 1,000 days, nor in the life of this Administration, nor even perhaps in our lifetime on this planet. 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 this country was founded, each generation of Americans has been summoned to give testimony to its national loyalty. The graves of young Americans who answered the call to service surround the globe.\n\nNow the trumpet 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, "rejoicing in hope, patient in tribulation" -- a struggle against the common enemies of man: tyranny, poverty, disease, and war itself.\n\nCan we forge against these enemies a grand and global alliance, North and South, East 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, the 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 world.\n\nAnd so, my fellow Americans: ask not what your country can do for 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 can do for the freedom of man.\n\nFinally, whether you are citizens of America or citizens of the world, ask of us the same high standards of strength and sacrifice which we ask of you. With a good conscience our only sure reward, with history the final judge of our deeds, let us go forth to lead the land we love, asking His blessing and His help, but knowing that here on earth God\'s work must truly be our own.\n'

## Speech 3-1973-Nixon

'Mr. Vice President, Mr. Speaker, Mr. Chief Justice, Senator Cook, Mrs. Eisenhower, and my fellow citizens of this great and good country we share together:\n\nWhen we met here four years ago, America was bleak in spirit, depressed by the prospect of seemingly endless war abroad and of destructive conflict at home.\n\nAs we meet here today, we stand on the threshold of a new era of peace in the world.\n\nThe central question before us is: How shall we use that peace? Let us resolve that this era 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 third century as a nation.\n\nThis past year saw far-reaching results from our new policies for peace. By continuing to revitalize our traditional friendships, and by our missions to Peking and to Moscow, we were able to establish the base for a new and more durable pattern of relationships among the nations of the world. Because of America\'s bold initiatives, 1972 will be long remembered as the year of the greatest progress 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 the 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\nUnless 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 vigorously 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 negotiation, to work for the limitation of nuclear arms, and to

reduce the danger 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 nation 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 role indispensable in preserving its own peace.\n\nTogether with the rest of the world, let us resolve to move forward from the beginnings we have 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 of 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 strong -- in which each respects the right of the other to live by a different 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 -- gladly because the chance to build such a peace is the noblest endeavor in which a nation can engage; gladly, also, because only if we act greatly in meeting our responsibilities abroad will we remain a great Nation, and 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 before in our history to make life better in America -- to ensure better education, better health, better housing, better transportation, a cleaner 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 great -- because the reach of our opportunities is so great -- let us be bold in our determination to meet those needs in new ways.\n\nJust as building a structure of peace abroad has required turning away from old policies that failed, so building a new era of progress at home requires turning away from old policies that have failed.\n\nAbroad, the shift from old policies to new has not been a retreat from our responsibilities, but a better way to peace.\n\nAnd at home, the shift from old policies 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 responsibilities 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 come 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 encourage 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 themselves.\n\nThat is why today I offer no promise of a purely governmental 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 than it can deliver. This leads only to inflated expectations, to reduced 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 more for themselves.\n\nLet us remember that America was built not by government, but by people -- not by welfare, but by work -- not by shirking responsibility, but by seeking responsibility.\n\nIn our own lives, let 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 will lead boldly. But just as important is the role that each and every one of us must play, as an individual and as a member of his own community.\n\nFrom this day forward, let each of us make a solemn commitment in his own heart: to bear his responsibility, to do his part, to live his ideals -- so that together, we can see the dawn of a new age of progress 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 ourselves 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 civility and decency. And let each of us reach out for that one precious quality government cannot provide -- a new level of respect for the rights and feelings of one another, a new level of respect for the individual 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 ourselves and in America.\n\nIn recent years, that faith has been challenged.\n\nOur children have been taught to be ashamed of their country, ashamed of their parents, ashamed of America\'s record at home and of its role 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 confident that this will not be the judgment of history on these remarkable times in which we are privileged to live.\n\nAmerica\'s record in this century has been unparalleled in the world\'s history for its responsibility, for its generosity, for its creativity and for its progress.\n\nLet us be proud that our system has produced and provided more freedom and more abundance, more widely shared, than any other system in the history 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 now 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 not known before -- a structure of peace that can last, not merely for our 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 generation, has ever faced.\n\nWe shall answer to God, to history, and to our 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 here before me. I think of the dreams they had for America, and I think of

how each recognized that he needed help far beyond himself in order to 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 for 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 years the best four years in America\'s history, so that on its 200th birthday America will be as young and as vital as when it began, and as bright 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'

2.1 Find the number of characters, words, and sentences for the mentioned documents. – 3 Marks

Number of characters -

Speech 1 is 6174,

Speech 2 is 6202,

Speech 3 is 2028.

Total Number of characters for all the three speeches is 20498.

Number of Words

Speech 1 is 1536

Speech 2 is 1546

Speech 3 is 8122

Total Number of Words for all the three speeches is 5110.

Number of Sentences

Speech 1 is 68

Speech2 is 52

Speech3 is 69.

Total Number of Sentences for all the three speeches is 189.

2.2 Remove all the stopwords from all three speeches. – 3 Marks

We have removed all the stopwords from all the three speeches and included some extended stopwords like It, For, day, must ,know, In, And, But, like, came, still, find, let, Let, To, ask, shall, always, go, make,In. Which were used in the speeches

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

For speech1 the most common words are **The and We**

For speech2 the most common words are **US, world, and sides**

For Speech3 the most common words are **US, and Peace**

2.4 Plot the word cloud of each of the speeches of the variable. (after removing the stopwords) – 3 Marks

SPEECH 1

SPEECH 2

SPEECH 3