Project

Karthiheswar

Table of Contents

1 Project Objective

The objective of the report is to explore all the projects data set in Python and generate insights about the data set. This exploration report will consist of the following:

- Importing the dataset in Python
- Understanding the structure of dataset
- Checking null values and performing descriptive statistics
- Graphical exploration
- Univariate and Bivariate Analysis
- Encode the data for Modelling
- Applying Logistic Regression, LDA, K-NN, Naive Bayes, SVM
- Model tuning, Bagging and Boosting
- Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and ROC_AUC score
- Text analysis
- Insights from the dataset

2 Analysis on Election_Data for news channel CNBE

2.1 Reading the data, descriptive statistics and do null value condition check

Reading the dataset (head)

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

Describing the data

	count	mean	std	min	25%	50%	75%	max
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

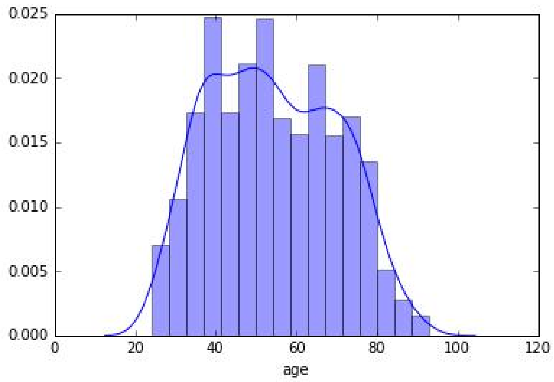
Checking the null values

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

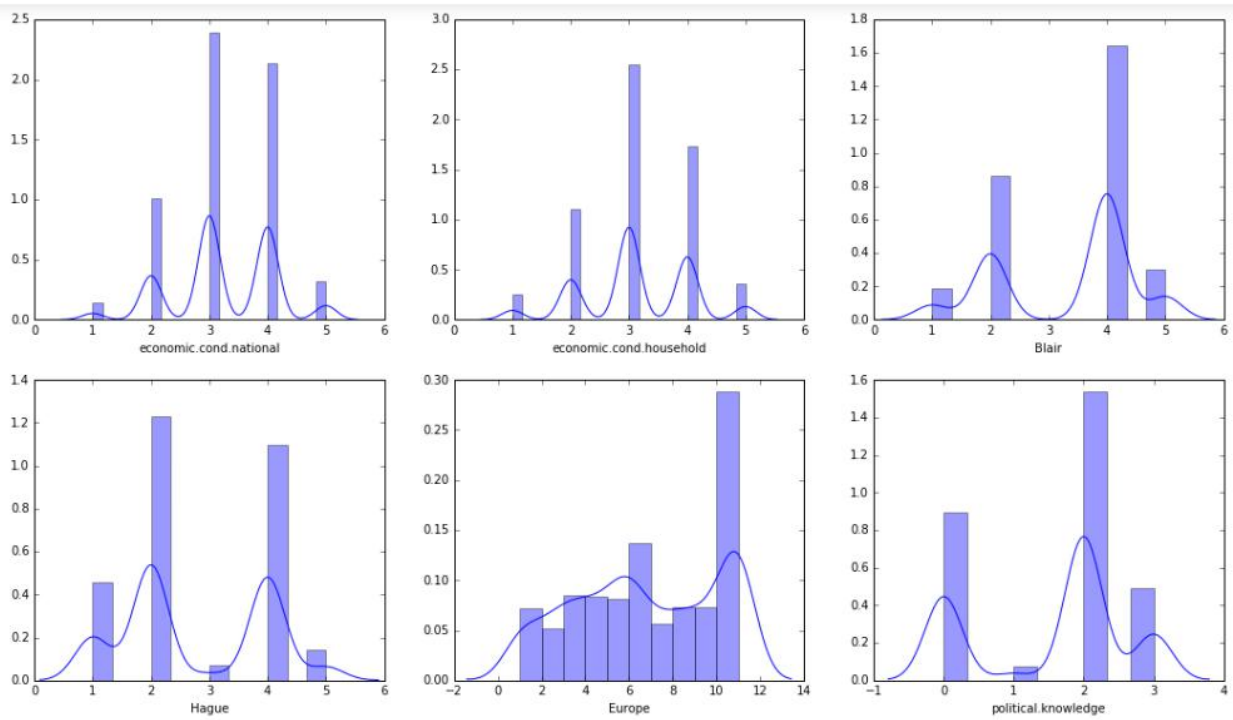
2.2 Performing Univariate and Bivariate Analysis, Exploratory Data Analysis and checking for Outliers

Univariate Analysis

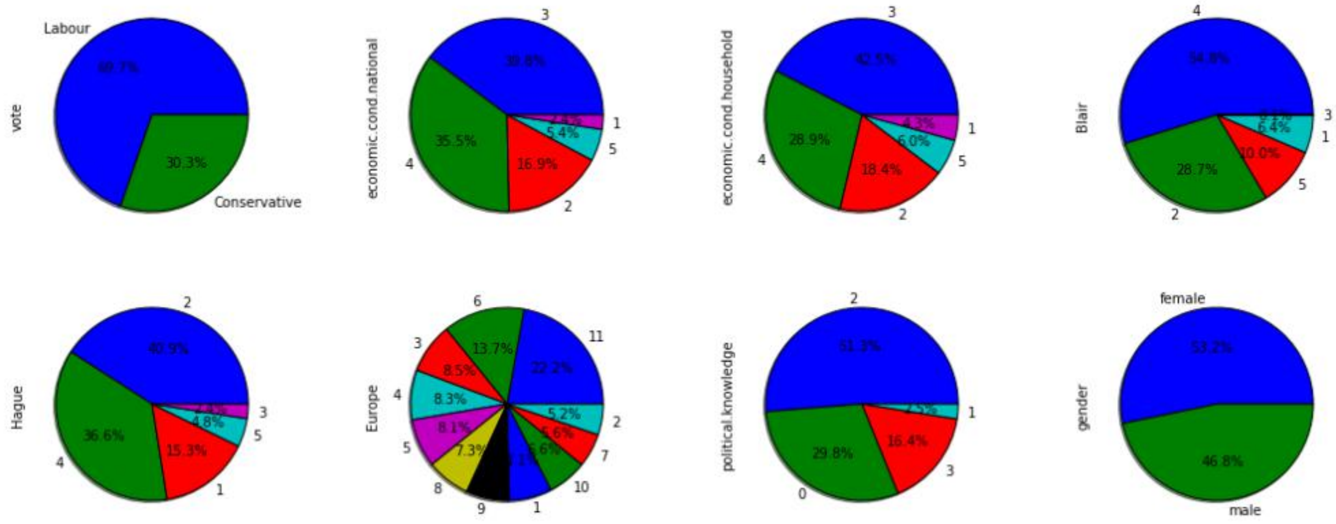
Variable - Age



Variable - Economic.cond.national, Economic.cond.household, Blair, Hague, Europe, Political.knowledge

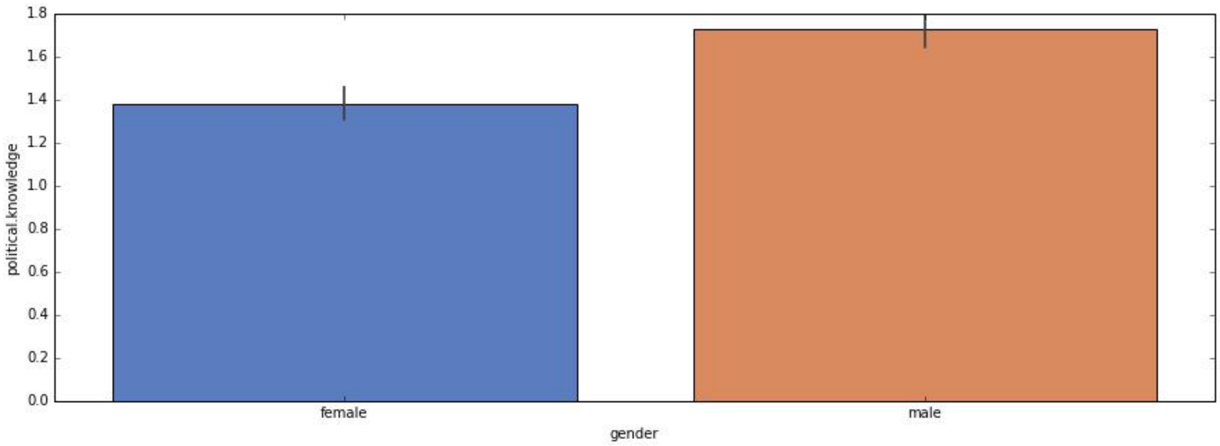


Variable - Vote, Economic.cond.national, Economic.cond.household, Blair, Hague, Europe, Political.knowledge, Gender

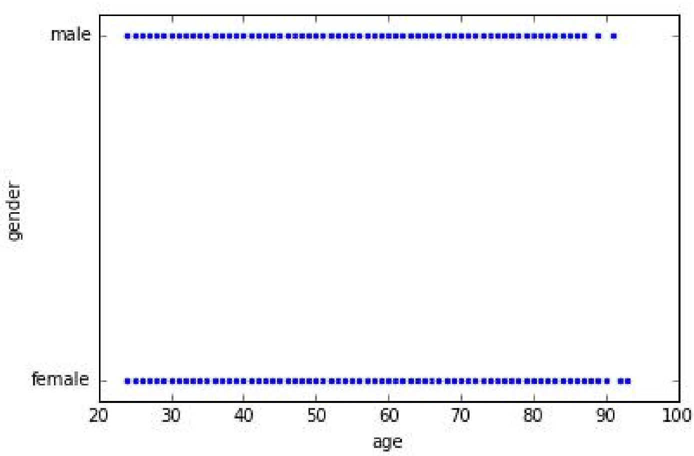


Bivariate Analysis

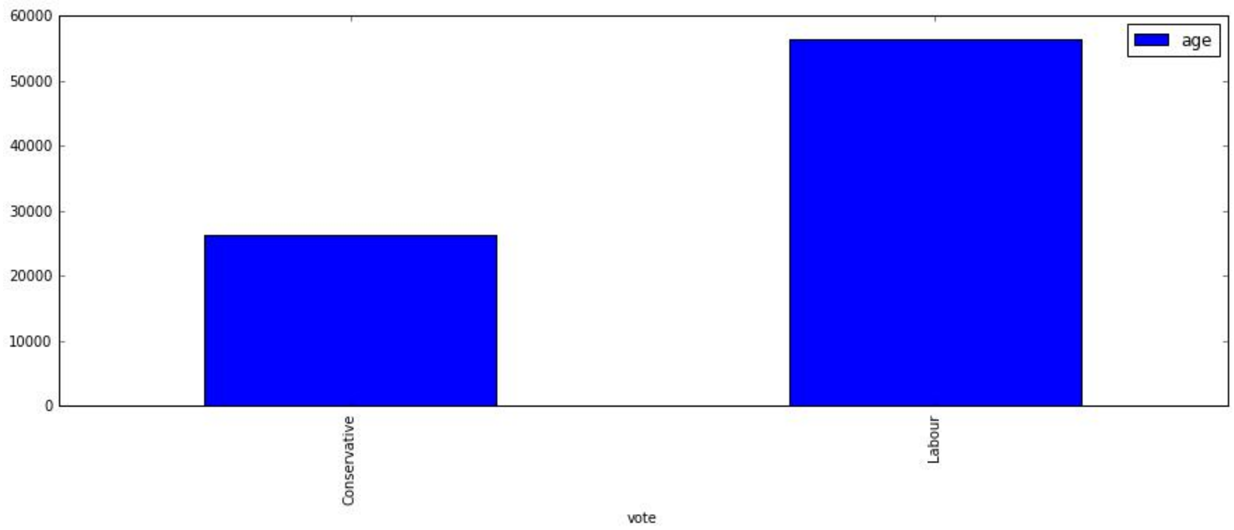
Variable- Gender vs Political.knowledge



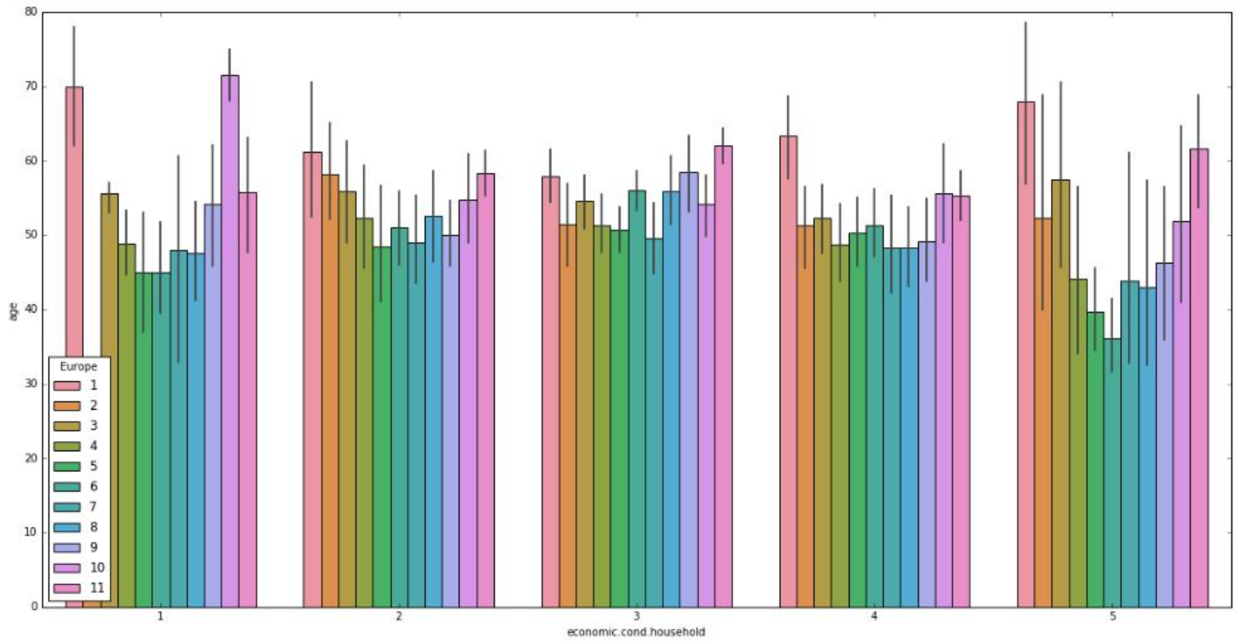
Variable- Age vs Gender



Variable- Vote vs Age



Variable- Age vs Economic.cond.household and Europe



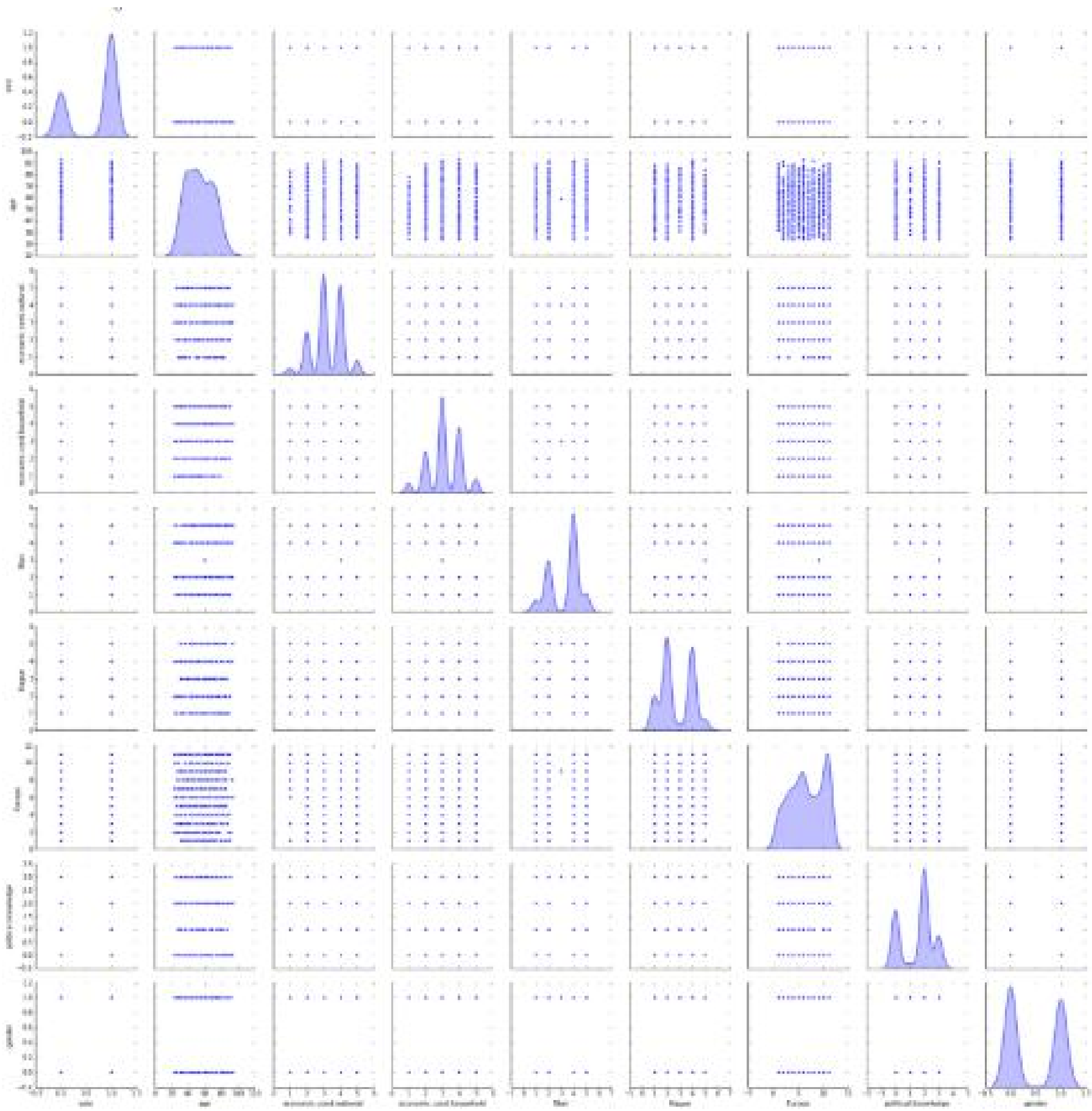
Exploratory Data Analysis

S.no	Description	IQR values for all attributes
1	age	26.0
2	economic.cond.national	1.0
3	economic.cond.household	1.0
4	Blair	2.0
5	Hague	2.0
6	Europe	6.0
7	political.knowledge	2.0

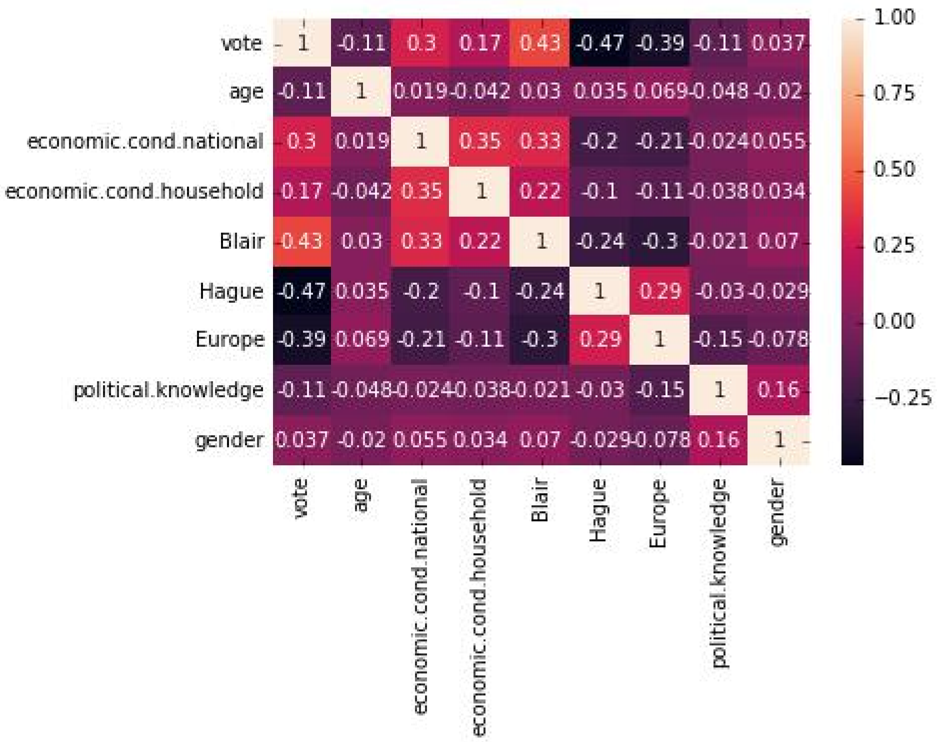
Covariance of each attribute against every other attribute

	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gende
vote	0.211310	-0.814816	0.121789	0.074693	0.230646	-0.265052	-0.584266	-0.054107	0.00853
age	-0.814816	246.842075	0.256981	-0.607619	0.557762	0.669531	3.568550	-0.825301	-0.15484
economic.cond.national	0.121789	0.256981	0.776107	0.283712	0.338314	-0.216589	-0.608397	-0.022546	0.02406
conomic.cond.household	0.074693	-0.607619	0.283712	0.864810	0.235192	-0.116689	-0.352299	-0.038091	0.01571
Blair	0.230646	0.557762	0.338314	0.235192	1.380212	-0.351648	-1.147341	-0.026621	0.04104
Hague	-0.265052	0.669531	-0.216589	-0.116689	-0.351648	1.514631	1.166149	-0.040469	-0.01806
Europe	-0.584266	3.568550	-0.608397	-0.352299	-1.147341	1.166149	10.873759	-0.544285	-0.12758
political.knowledge	-0.054107	-0.825301	-0.022546	-0.038091	-0.026621	-0.040469	-0.544285	1.173571	0.08552
gender	0.008533	-0.154840	0.024063	0.015713	0.041046	-0.018064	-0.127584	0.085527	0.24911

Scatter plot



Heatmap



Skeweness

S.no	Description	Skeweness of every attribute
1	age	0.144
2	economic.cond.national	-0.240
3	economic.cond.household	-0.149
4	Blair	-0.535
5	Hague	0.152
6	Europe	-0.135
7	political.knowledge	-0.426

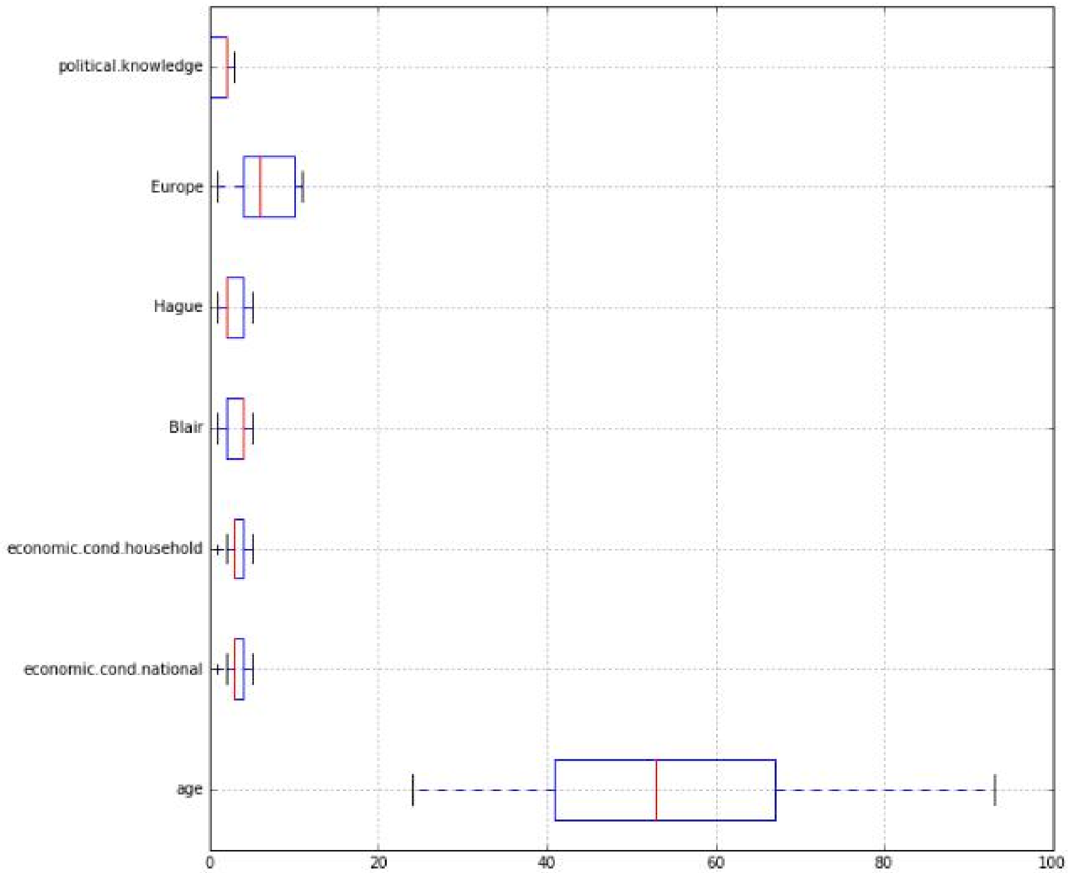
Checking the Data types

```
<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: 107.3+ KB
```

Checking the Shape

(1525, 9)

Outliers



2.3 Encoding the data

Encoding the data

The dataset is encoded using label encoder. As one-hot encoding increases the no. of variables, this method is not used.

	vote	age	economic.cond.national	economic.cond.household	Blair	Hague	Europe	political.knowledge	gender
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 of data

Scaling is not necessary as each columns contains same range of magnitudes except age apart from categorical variables .

Data Split

The variable vote is taken as dependent variable and the data is split into train and test (70:30).

2.4 Applying Logistic Regression and LDA (Linear Discriminant Analysis)

Logistic Regression

```
LogisticRegression(C=1.0, class_weight='balanced', dual=False,
                   fit_intercept=True, intercept_scaling=1, l1_ratio=None,
                   max_iter=100, multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
```

Logistic regression is applied on train data.

Linear Discriminant Analysis

```
LinearDiscriminantAnalysis(n_components=None, priors=None, shrinkage=None,
                           solver='svd', store_covariance=False, tol=0.0001)
```

Linear Discriminant Analysis is applied on train data.

2.5 Applying KNN Model, Naïve Bayes Model and Support Vector Machine (SVM) model

KNN model

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                     metric_params=None, n_jobs=None, n_neighbors=5, p=2,
                     weights='uniform')
```

KNN Analysis is applied on train data.

Naive Bayes model

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

Naive Bayes is applied on train data.

Support Vector Machine (SVM)

```
SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
    max_iter=-1, probability=True, random_state=1, shrinking=True, tol=0.001,
    verbose=False)
```

SVM is applied on train data.

2.6 Model Tuning, Bagging and Boosting

In order to tune the model, SMOTE and cross validation are applied to all models.

Logistic Regression with SMOTE

```
LogisticRegression(C=1.0, class_weight='balanced', dual=False,
        fit_intercept=True, intercept_scaling=1, l1_ratio=None,
        max_iter=100, multi_class='auto', n_jobs=None, penalty='l2',
        random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
        warm_start=False)
```

Linear Discriminant Analysis with SMOTE

```
LinearDiscriminantAnalysis(n_components=None, priors=None, shrinkage=None,
        solver='svd', store_covariance=False, tol=0.0001)
```

KNN with SMOTE

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
        metric_params=None, n_jobs=None, n_neighbors=5, p=2,
        weights='uniform')
```

Naive Bayes with SMOTE

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

SVM with SMOTE

```
SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
  max_iter=-1, probability=True, random_state=1, shrinking=True, tol=0.001,
  verbose=False)
```

S. NO	DESCRIPTION OF MODEL	CROSS VALIDATION OF TRAIN DATA	CROSS VALIDATION OF TEST DATA
1	Logistic Regression	array([0.81308411, 0.81308411, 0.8411215 , 0.81308411, 0.82242991, 0.8411215, 0.81308411, 0.85849057, 0.78301887, 0.79245283])	array([0.76086957, 0.80434783, 0.7826087 , 0.80434783, 0.76086957, 0.7826087, 0.89130435, 0.80434783, 0.88888889, 0.75555556])
2	Linear Discriminant Analysis	array([0.78504673, 0.82242991, 0.85981308, 0.8317757, 0.8411215, 0.82242991, 0.80373832, 0.9245283 , 0.82075472, 0.80188679])	array([0.82608696, 0.84782609, 0.80434783, 0.80434783, 0.82608696, 0.80434783, 0.84782609, 0.84782609, 0.91111111, 0.75555556])
3	KNN	array([0.77570093, 0.74766355, 0.80373832, 0.79439252, 0.78504673, 0.75700935, 0.78504673, 0.87735849, 0.80188679, 0.82075472])	array([0.84782609, 0.84782609, 0.82608696, 0.7826087 , 0.82608696, 0.7826087, 0.93478261, 0.82608696, 0.88888889, 0.8])

4	Naive Bayes	array([0.81308411, 0.8317757 , 0.82242991, 0.85046729, 0.82242991, 0.81308411, 0.81308411, 0.88679245, 0.82075472, 0.81132075])	array([0.82608696, 0.84782609, 0.82608696, 0.80434783, 0.76086957, 0.80434783, 0.84782609, 0.91304348, 0.88888889, 0.82222222])
5	SVM	array([0.73831776, 0.73831776, 0.77570093, 0.77570093, 0.76635514, 0.74766355, 0.71962617, 0.79245283, 0.79245283, 0.79245283])	array([0.7173913 , 0.7173913 , 0.7173913 , 0.7173913 , 0.7173913, 0.7173913, 0.7173913, 0.7173913 , 0.71111111, 0.71111111])

Bagging

The bagging is done using random forest classifier, grid search is applied in it to tune the model.

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=7, max_features=6,
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=25, min_samples_split=50,
                       min_weight_fraction_leaf=0.0, n_estimators=501,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False)
```

Boosting

S. NO	ADA BOOST	XGBOOST
1	AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None, learning_rate=1.0, n_estimators=100, random_state=1)	XGBClassifier(base_score=0.5, booster=None, colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1, importance_type='gain', interaction_constraints=None, learning_rate=0.01, max_delta_step=0, max_depth=6, min_child_weight=1, missing=nan, monotone_constraints=None, n_estimators=100, n_jobs=0, num_parallel_tree=1, objective='binary:logistic', random_state=1, reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1, tree_method=None, validate_parameters=False, verbosity=None)

2.7 Performance Metrics: Checking the performance of Predictions on Train and Test sets using Accuracy, Confusion Matrix, Plot ROC curve and get ROC_AUC score for each model. Comparing the models and inference which model is best/optimized

PREDICTIONS, ACCURACY AND CONFUSION MATRIX

S. NO	**DESCRIPTION**	**LOGISTIC REGRESSION**	**LDA**	**KNN**	**NAIVE BAYES**	**SVM**
1	Predictions on Train data	array ([0, 1, 1,..., 1, 1, 0])	array ([0, 1, 1,.., 1, 1, 0])	array ([1, 0, 1, ..., 1, 1, 1])	array ([0, 1, 1, ..., 1, 1, 1])	array ([0, 1, 1,., 1, 1, 1])
2	Predictions on Test data	array ([0, 0, 0,..., 0, 1, 1])	array ([0, 0, 0,.., 0, 1, 1])	array ([0, 0, 1, ..., 0, 1, 0])	array ([0, 0, 1, ..., 0, 1, 1])	array ([0, 1, 1,., 0, 1, 1])
3	Accuracy on Train data	0.8247	0.8341	0.8537	0.8331	0.7835
4	Accuracy on Test data	0.79694	0.8034	0.7860	0.8253	0.7860
5	Confusion Matrix on Train data	[[270 62] [125 610]]	[[274 58] [119 616]]	[[242 90] [66 669]]	[[240 92] [86 649]]	[[135 197] [34 701]]
6	Confusion Matrix on Test data	[[103 27] [66 262]]	[[104 26] [64 264]]	[[81 49] [49 279]]	[[94 36] [44 284]]	[[50 80] [18 310]]

S. NO	**DESCRIPTION**	**LOGISTIC REGRESSION WITH SMOTE**	**LDA WITH SMOTE**	**KNN WITH SMOTE**	**NAIVE BAYES WITH SMOTE**	**SVM WITH SMOTE**
1	Predictions on Train data	array ([0, 1, 1,..., 1, 1, 0])	array ([0, 1, 1,.., 1, 1, 0])	array ([0, 0, 1, ..., 1, 1, 1])	array ([0, 0, 1, ..., 1, 1, 0])	array ([0, 1, 1,., 1, 1, 1])
2	Predictions on Test data	array ([0, 0, 0,..., 0, 1, 1])	array ([0, 0, 0,.., 0, 1, 1])	array ([0, 0, 1, ..., 0, 1, 0])	array ([0, 0, 0, ..., 0, 1, 0])	array ([0, 0, 1,., 0, 1, 0])
3	Accuracy on Train data	0.831	0.8341	0.8547	0.8181	0.7975
4	Accuracy on Test data	0.7947	0.8034	0.7532	0.7969	0.7816
5	Confusion Matrix on Train data	[[272 60] [120 615]]	[[274 58] [119 616]]	[[299 33] [122 613]]	[[256 76] [118 617]]	[[264 68] [148 587]]
6	Confusion Matrix on Test data	[[102 28] [66 262]]	[[104 26] [64 264]]	[[100 30] [83 245]]	[[103 27] [66 262]]	[[106 24] [76 252]]

S. NO	DESCRIPTION	RANDOM FOREST CLASSIFIER - GRID SEARCH	ADA BOOST	XGBOOST
1	Predictions on Train data	array ([0, 1, 1,..., 1, 1, 1])	array ([0, 1, 1,.., 1, 1, 1])	array ([0, 1, 1, ..., 1, 1, 1])
2	Predictions on Test data	array ([0, 0, 1,..., 0, 1, 1])	array ([0, 0, 1,.., 0, 1, 1])	array ([0, 0, 1, ..., 0, 1, 0])
3	Accuracy on Train data	0.8444	0.8472	0.8865
4	Accuracy on Test data	0.8078	0.8187	0.8122
5	Confusion Matrix on Train data	[[240 92] [74 661]]	[[238 94] [69 666]]	[[262 70] [51 684]]
6	Confusion Matrix on Test data	[[88 42] [46 282]]	[[90 40] [43 285]]	[[93 37] [49 279]]

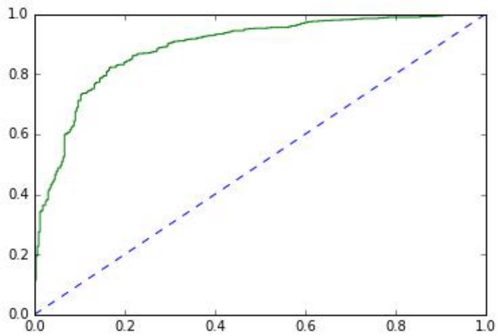
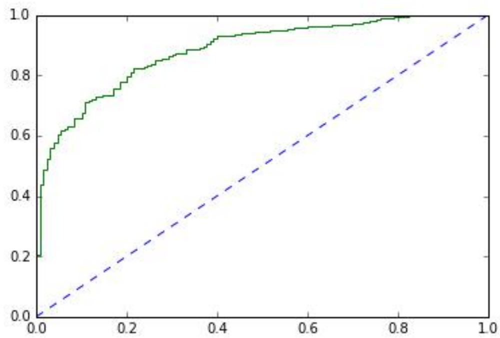
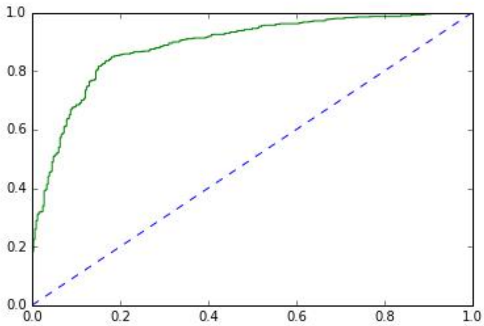
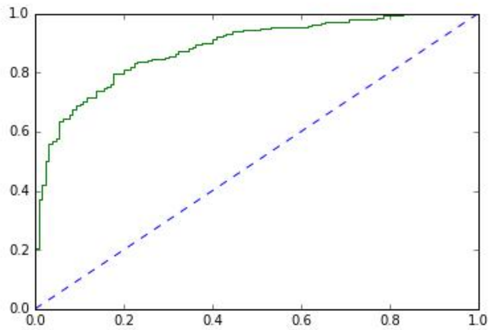
CLASSIFICATION REPORT

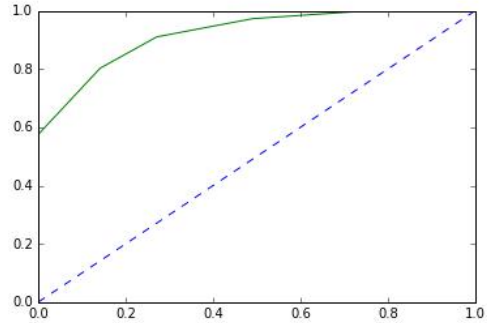
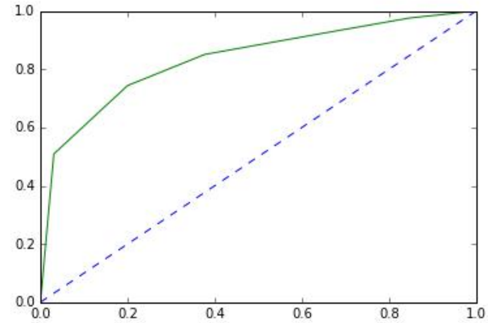
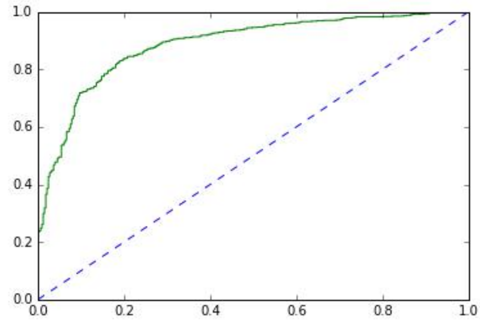
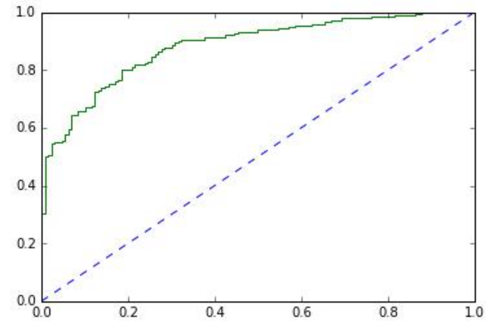
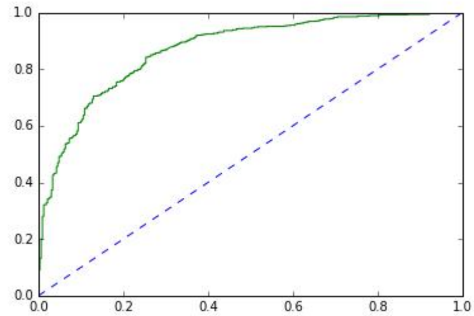
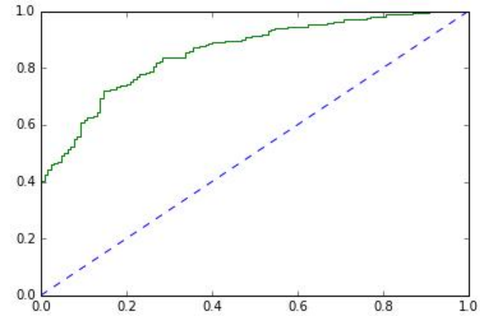
S. NO	DESCRIPTION	CLASSIFICATION REPORT ON TRAIN DATA	CLASSIFICATION REPORT ON TEST DATA
1	Logistic Regression	Classification Report precision recall f1-score support 0 0.68 0.81 0.74 332 1 0.91 0.83 0.87 735 accuracy 0.82 1067 macro avg 0.80 0.82 0.80 1067 weighted avg 0.84 0.82 0.83 1067	Classification Report precision recall f1-score support 0 0.61 0.79 0.69 130 1 0.91 0.80 0.85 328 accuracy 0.80 458 macro avg 0.76 0.80 0.77 458 weighted avg 0.82 0.80 0.80 458
2	LDA	Classification Report precision recall f1-score support 0 0.70 0.83 0.76 332 1 0.91 0.84 0.87 735 accuracy 0.83 1067 macro avg 0.81 0.83 0.82 1067 weighted avg 0.85 0.83 0.84 1067	Classification Report precision recall f1-score support 0 0.62 0.80 0.70 130 1 0.91 0.80 0.85 328 accuracy 0.80 458 macro avg 0.76 0.80 0.78 458 weighted avg 0.83 0.80 0.81 458
3	KNN	Classification Report precision recall f1-score support 0 0.79 0.73 0.76 332 1 0.88 0.91 0.90 735 accuracy 0.85 1067 macro avg 0.83 0.82 0.83 1067 weighted avg 0.85 0.85 0.85 1067	Classification Report precision recall f1-score support 0 0.62 0.62 0.62 130 1 0.85 0.85 0.85 328 accuracy 0.79 458 macro avg 0.74 0.74 0.74 458 weighted avg 0.79 0.79 0.79 458
4	Naive Bayes	Classification Report precision recall f1-score support 0 0.74 0.72 0.73 332 1 0.88 0.88 0.88 735 accuracy 0.83 1067 macro avg 0.81 0.80 0.80 1067 weighted avg 0.83 0.83 0.83 1067	Classification Report precision recall f1-score support 0 0.68 0.72 0.70 130 1 0.89 0.87 0.88 328 accuracy 0.83 458 macro avg 0.78 0.79 0.79 458 weighted avg 0.83 0.83 0.83 458
5	SVM	Classification Report precision recall f1-score support 0 0.80 0.41 0.54 332 1 0.78 0.95 0.86 735 accuracy 0.78 1067 macro avg 0.79 0.68 0.70 1067 weighted avg 0.79 0.78 0.76 1067	Classification Report precision recall f1-score support 0 0.74 0.38 0.51 130 1 0.79 0.95 0.86 328 accuracy 0.79 458 macro avg 0.77 0.66 0.68 458 weighted avg 0.78 0.79 0.76 458

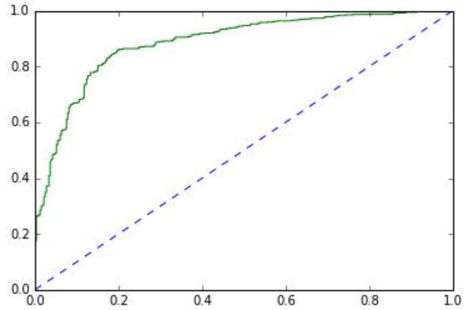
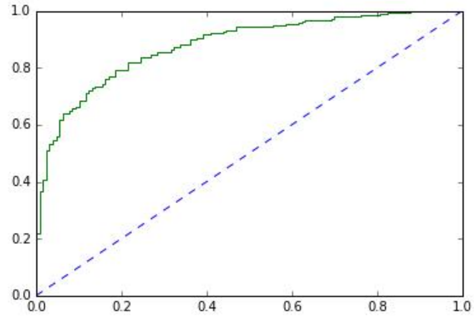
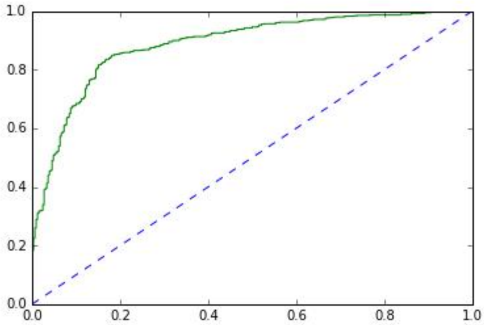
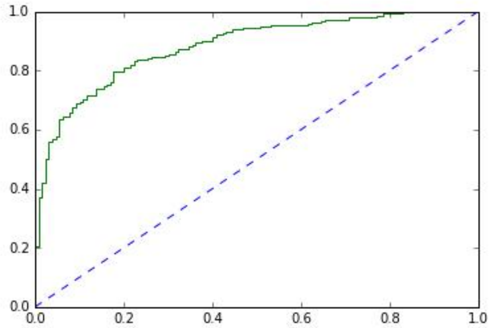
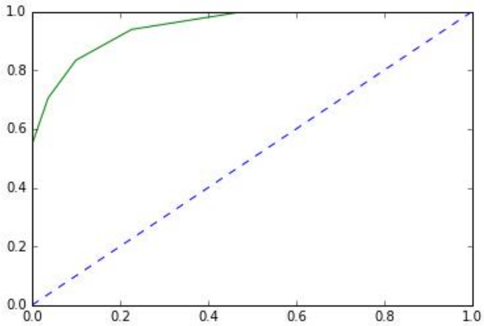
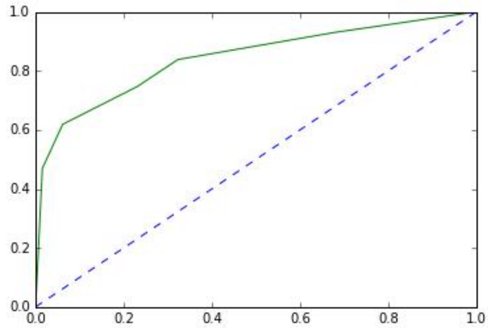
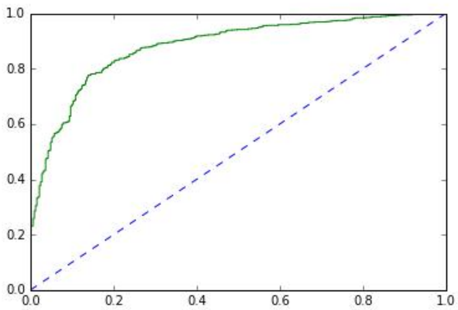
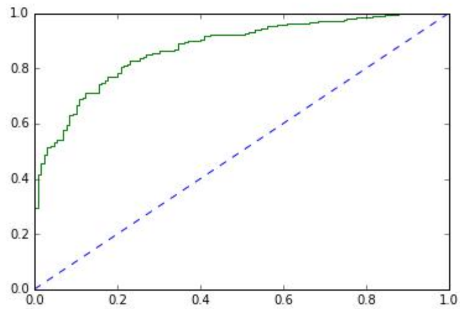
S. NO	DESCRIPTION	CLASSIFICATION REPORT ON TRAIN DATA WITH SMOTE	CLASSIFICATION REPORT ON TEST DATA WITH SMOTE
1	Logistic Regression	Classification Report precision recall f1-score support 0 0.69 0.82 0.75 332 1 0.91 0.84 0.87 735 accuracy 0.83 1067 macro avg 0.80 0.83 0.81 1067 weighted avg 0.84 0.83 0.83 1067	Classification Report precision recall f1-score support 0 0.61 0.78 0.68 130 1 0.90 0.80 0.85 328 accuracy 0.79 458 macro avg 0.76 0.79 0.77 458 weighted avg 0.82 0.79 0.80 458
2	LDA	Classification Report precision recall f1-score support 0 0.70 0.83 0.76 332 1 0.91 0.84 0.87 735 accuracy 0.83 1067 macro avg 0.81 0.83 0.82 1067 weighted avg 0.85 0.83 0.84 1067	Classification Report precision recall f1-score support 0 0.62 0.80 0.70 130 1 0.91 0.80 0.85 328 accuracy 0.80 458 macro avg 0.76 0.80 0.78 458 weighted avg 0.83 0.80 0.81 458
3	KNN	Classification Report precision recall f1-score support 0 0.71 0.90 0.79 332 1 0.95 0.83 0.89 735 accuracy 0.85 1067 macro avg 0.83 0.87 0.84 1067 weighted avg 0.87 0.85 0.86 1067	Classification Report precision recall f1-score support 0 0.55 0.77 0.64 130 1 0.89 0.75 0.81 328 accuracy 0.75 458 macro avg 0.72 0.76 0.73 458 weighted avg 0.79 0.75 0.76 458
4	Naive Bayes	Classification Report precision recall f1-score support 0 0.68 0.77 0.73 332 1 0.89 0.84 0.86 735 accuracy 0.82 1067 macro avg 0.79 0.81 0.79 1067 weighted avg 0.83 0.82 0.82 1067	Classification Report precision recall f1-score support 0 0.61 0.79 0.69 130 1 0.91 0.80 0.85 328 accuracy 0.80 458 macro avg 0.76 0.80 0.77 458 weighted avg 0.82 0.80 0.80 458
5	SVM	Classification Report precision recall f1-score support 0 0.64 0.80 0.71 332 1 0.90 0.80 0.84 735 accuracy 0.80 1067 macro avg 0.77 0.80 0.78 1067 weighted avg 0.82 0.80 0.80 1067	Classification Report precision recall f1-score support 0 0.58 0.82 0.68 130 1 0.91 0.77 0.83 328 accuracy 0.78 458 macro avg 0.75 0.79 0.76 458 weighted avg 0.82 0.78 0.79 458

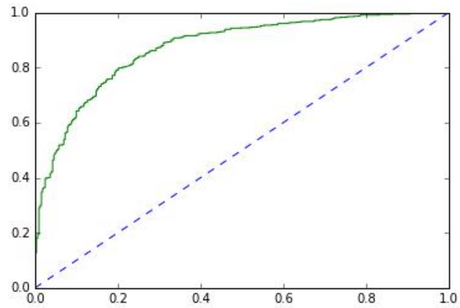
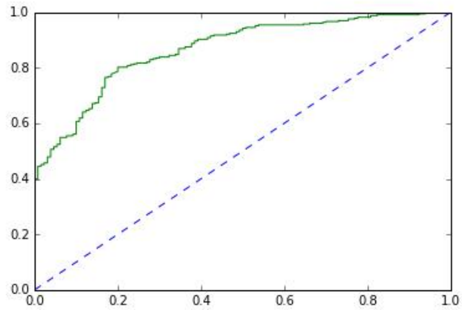
<table>
<tr><th>S. NO</th><th>DESCRIPTION</th><th>CLASSIFICATION REPORT ON TRAIN DATA</th><th>CLASSIFICATION REPORT ON TEST DATA</th></tr>
<tr><td>1</td><td>Random Forest Classifier - Grid Search</td><td>Classification Report
precision recall f1-score support
0 0.76 0.72 0.74 332
1 0.88 0.90 0.89 735
accuracy 0.84 1067
macro avg 0.82 0.81 0.82 1067
weighted avg 0.84 0.84 0.84 1067</td><td>Classification Report
precision recall f1-score support
0 0.66 0.68 0.67 130
1 0.87 0.86 0.87 328
accuracy 0.81 458
macro avg 0.76 0.77 0.77 458
weighted avg 0.81 0.81 0.81 458</td></tr>
<tr><td>2</td><td>Ada Boost</td><td>Classification Report
precision recall f1-score support
0 0.78 0.72 0.74 332
1 0.88 0.91 0.89 735
accuracy 0.85 1067
macro avg 0.83 0.81 0.82 1067
weighted avg 0.84 0.85 0.85 1067</td><td>Classification Report
precision recall f1-score support
0 0.68 0.69 0.68 130
1 0.88 0.87 0.87 328
accuracy 0.82 458
macro avg 0.78 0.78 0.78 458
weighted avg 0.82 0.82 0.82 458</td></tr>
<tr><td>3</td><td>XGBoost</td><td>Classification Report
precision recall f1-score support
0 0.84 0.79 0.81 332
1 0.91 0.93 0.92 735
accuracy 0.89 1067
macro avg 0.87 0.86 0.87 1067
weighted avg 0.89 0.89 0.89 1067</td><td>Classification Report
precision recall f1-score support
0 0.65 0.72 0.68 130
1 0.88 0.85 0.87 328
accuracy 0.81 458
macro avg 0.77 0.78 0.78 458
weighted avg 0.82 0.81 0.81 458</td></tr>
</table>

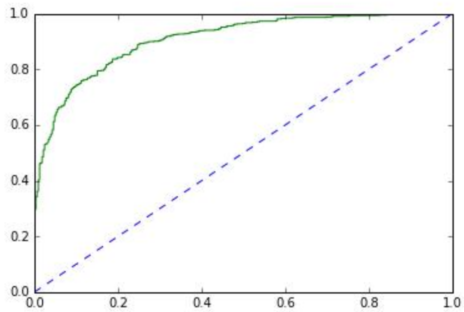
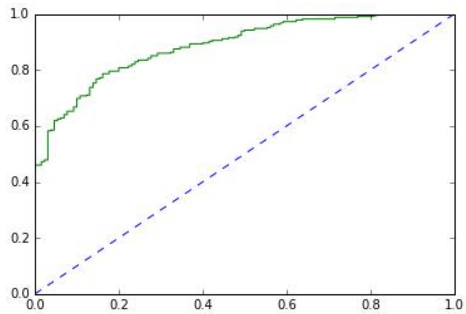
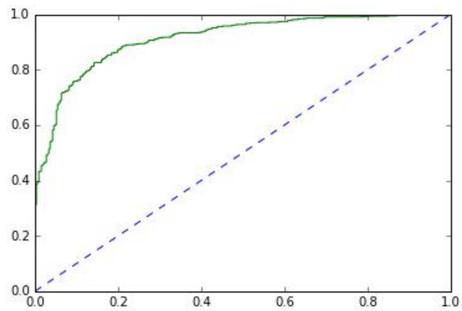
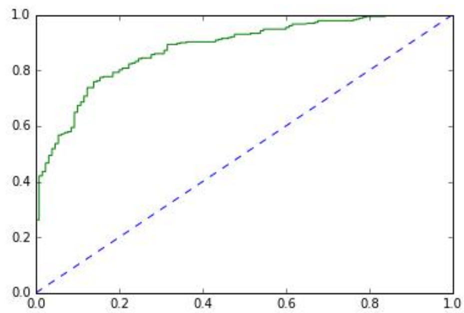
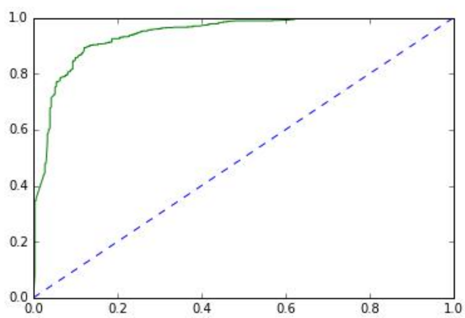
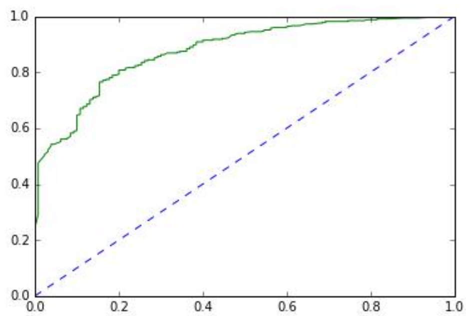
AUC AND ROC CURVE

<table>
<tr><th>S. NO</th><th>DESCRIPTION</th><th>AUC AND ROC ON TRAIN DATA</th><th>AUC AND ROC ON TEST DATA</th></tr>
<tr><td>1</td><td>Logistic Regression</td><td>AUC: 0.890
[<matplotlib.lines.Line2D at 0xdc418d68d0>]
</td><td>AUC: 0.885
[<matplotlib.lines.Line2D at 0xdc4115e7f0>]
</td></tr>
<tr><td>2</td><td>LDA</td><td>AUC: 0.887
[<matplotlib.lines.Line2D at 0xdc420a6240>]
</td><td>AUC: 0.884
[<matplotlib.lines.Line2D at 0xdc4204c828>]
</td></tr>
</table>

3	KNN	AUC: 0.921 [<matplotlib.lines.Line2D at 0xdc42173128>] 	AUC: 0.835 [<matplotlib.lines.Line2D at 0xdc4210ffd0>] 
4	Naive Bayes	AUC: 0.886 [<matplotlib.lines.Line2D at 0xdc42224470>] 	AUC: 0.885 [<matplotlib.lines.Line2D at 0xdc421c6208>] 
5	SVM	AUC: 0.870 [<matplotlib.lines.Line2D at 0xdc42284eb8>] 	AUC: 0.857 [<matplotlib.lines.Line2D at 0xdc42289ac8>] 

S. NO	DESCRIPTION	AUC AND ROC ON TRAIN DATA WITH SMOTE	AUC AND ROC ON TEST DATA WITH SMOTE
1	Logistic Regression	AUC: 0.886 [<matplotlib.lines.Line2D at 0xdc4331c080>] 	AUC: 0.881 [<matplotlib.lines.Line2D at 0xdc4331c9b0>] 
2	LDA	AUC: 0.887 [<matplotlib.lines.Line2D at 0xdc420a6240>] 	AUC: 0.884 [<matplotlib.lines.Line2D at 0xdc4204c828>] 
3	KNN	AUC: 0.950 [<matplotlib.lines.Line2D at 0x922edd5a20>] 	AUC: 0.842 [<matplotlib.lines.Line2D at 0x922ed77278>] 
4	Naive Bayes	AUC: 0.881 [<matplotlib.lines.Line2D at 0x922d16a668>] 	AUC: 0.875 [<matplotlib.lines.Line2D at 0x922ee43080>] 

5	SVM	AUC: 0.873 [<matplotlib.lines.Line2D at 0x922ef065c0>] 	AUC: 0.866 [<matplotlib.lines.Line2D at 0x922ef06ef0>] 

S. NO	DESCRIPTION	AUC AND ROC ON TRAIN DATA	AUC AND ROC ON TEST DATA
1	Random Forest Classifier - Grid Search	AUC: 0.908 [<matplotlib.lines.Line2D at 0x922efbc128>] 	AUC: 0.887 [<matplotlib.lines.Line2D at 0x922edd5320>] 
2	Ada Boost	AUC: 0.913 [<matplotlib.lines.Line2D at 0x922f0a2358>] 	AUC: 0.879 [<matplotlib.lines.Line2D at 0x922f0a20f0>] 
3	XGBoost	AUC: 0.941 [<matplotlib.lines.Line2D at 0x922f16f4a8>] 	AUC: 0.878 [<matplotlib.lines.Line2D at 0x922f16ffd0>] 

INFERENCE

- Logistic regression determines good accuracy and has same and good precision for 1 on both train and test data but for 0 the precision value differs much. The recall value has no much difference on train and test data.
- LDA also determines good accuracy and has same and good precision for 1 on both train and test data but for 0, again the precision value differs much. The recall value has no much difference on train and test data. The recall values are better in LDA then Logistic regression.
- Anyway KNN determines much difference between train and test data, so it is not recommended.
- Naive Bayes determines least difference for recall value on test and train data, also precision value for 1. Naive Bayes determines decent accuracy for train and test data.
- SVM model determines very poor recall score and f1 score but no difference in recall for 1, so it cannot be considered.
- The SMOTE is applied for Logistic regression, LDA, KNN, Naive Bayes and SVM to tune the model, anyway it cannot be used to interpret as this method is used when the minority class is between 1-2%, whereas in this minority class is nearly 30%.
- Random Forest classifier, Ada Boost and XGBoost determines recall values with much difference on train and test data. There is also much difference in precision value for 0 on train and test data.

2.8 Insights

The people with higher in age votes for Conservative where lower age people vote for both Labour and Conservative. With higher Blair the vote is Labour. Vote is highly negative correlated with Hauge and Europe, highly positive correlated with Blair. Vote is very poorly correlated with gender, age and political knowledge. Naive Bayes and SVM shows less difference between test and train data on cross validation. KNN shows much difference between test and train data on cross validation, where Logistic regression and LDA works decently on cross validation. LDA also determines good accuracy and has same and good precision for 1 on both train and test data but for 0, again the precision value differs much. The recall value has no much difference on train and test data. The recall values are better in LDA then Logistic regression. Naive Bayes determines least difference for recall value on test and train data, also precision value for 1.

3 Text analysis on speeches of Presidents of the United States of America

3.1 Determining the number of characters, words and sentences for the mentioned documents

Number of characters of Roosevelt.txt

Number of characters with space in Roosevelt text file : 7571

Number of characters without space in Roosevelt text file : 6249

Number of words of Roosevelt.txt

Number of words in Roosevelt text file : 1360

Number of sentences of Roosevelt.txt

Number of sentences in Roosevelt text file : 68

Number of characters of Kennedy.txt

Number of characters with space in Kennedy text file : 7618

Number of characters without space in Kennedy text file : 6255

Number of words of Kennedy.txt

Number of words in Kennedy text file : 1390

Number of sentences of Kennedy.txt

Number of sentences in Kennedy text file : 52

Number of characters of Nixon.txt

Number of characters with space in Nixon text file : 9991

Number of characters without space in Nixon text file : 8223

Number of words of Nixon.txt

Number of words in Nixon text file : 1819

Number of sentences of Nixon.txt

Number of sentences in Nixon text file : 68

3.2 Removing all the stopwords from the three speeches

Removing stop words from Roosevelt

The whole text are processed using word_tokenize and then stop words are removed from the text file. The following is the filtered sentence:

['On', 'national', 'day', 'inauguration', 'since', '1789', ',', 'people', 'renewed', 'sense', 'dedication', 'United', 'States', '.', 'In', 'Washington', "'s", 'day', 'task', 'people', 'create', 'weld', 'together', 'nation', '.', 'In', 'Lincoln', "'s", 'day', 'task', 'people', 'preserve', 'Nation', 'disruption', 'within', '.', 'In', 'day', 'task', 'people', 'save', 'Nation', 'institutions', 'disruption', 'without', '.', 'To', 'us', 'come', 'time', ',', 'midst', 'swift', 'happenings', ',', 'pause', 'moment', 'take', 'stock', '--', 'recall', 'place', 'history', ',', 'rediscover', 'may', '.', 'If', ',', 'risk', 'real', 'peril', 'inaction', '.', 'Lives', 'nations', 'determined', 'count', 'years', ',', 'lifetime', 'human', 'spirit', '.', 'The', 'life', 'man', 'three-score', 'years', 'ten', ':', 'little', ',', 'little', 'less', '.', 'The', 'life', 'nation', 'fullness', 'measure', 'live', '.', 'There', 'men', 'doubt', '.', 'There', 'men', 'believe', 'democracy', ',', 'form', 'Government', 'frame', 'life', ',', 'limited', 'measured', 'kind', 'mystical', 'artificial', 'fate', ',', 'unexplained', 'reason', ',', 'tyranny', 'slavery', 'become', 'surging', 'wave', 'future', '--', 'freedom', 'ebbing', 'tide', '.', 'But', 'Americans', 'know', 'true', '.', 'Eight', 'years', 'ago', ',', 'life', 'Republic', 'seemed', 'frozen', 'fatalistic', 'terror', ',', 'proved', 'true', '.', 'We', 'midst', 'shock', '--', 'acted', '.', 'We', 'acted', 'quickly', ',', 'boldly', ',', 'decisively', '.', 'These', 'later', 'years', 'living', 'years', '--', 'fruitful', 'years', 'people', 'democracy', '.', 'For', 'brought', 'us', 'greater', 'security', ',', 'I', 'hope', ',', 'better', 'understanding', 'life', "'s", 'ideals', 'measured', 'material', 'things', '.', 'Most', 'vital', 'present', 'future', 'experience', 'democracy', 'successfully', 'survived', 'crisis', 'home', ';', 'put', 'away', 'many', 'evil', 'things', ';', 'built', 'new', 'structures', 'enduring', 'lines', ';', ',', ',', 'maintained', 'fact', 'democracy', '.', 'For', 'action', 'taken', 'within', 'three-way', 'framework', 'Constitution', 'United', 'States', '.', 'The', 'coordinate', 'branches', 'Government', 'continue', 'freely', 'function', '.', 'The', 'Bill', 'Rights', 'remains', 'inviolate', '.', 'The', 'freedom', 'elections', 'wholly', 'maintained', '.', 'Prophets', 'downfall', 'American', 'democracy', 'seen', 'dire', 'predictions', 'come', 'naught', '.', 'Democracy', 'dying', '.', 'We', 'know', 'seen', 'revive', '--', 'grow', '.', 'We', 'know', 'die', '--', 'built', 'unhampered', 'initiative', 'individual', 'men', 'women', 'joined', 'together', 'common', 'enterprise', '--', 'enterprise', 'undertaken', 'carried', 'free', 'expression', 'free', 'majority', '.', 'We', 'know', 'democracy', 'alone', ',', 'forms', 'government', ',', 'enlists', 'full', 'force', 'men', "'s", 'enlightened', '.', 'We', 'know', 'democracy', 'alone', 'constructed', 'unlimited', 'civilization', 'capable', 'infinite', 'progress', 'improvement', 'human', 'life', '.', 'We', 'know', ',', 'look', 'surface', ',', 'sense', 'still', 'spreading', 'every', 'continent', '--', 'humane', ',', 'advanced', ',', 'end', 'unconquerable', 'forms', 'human', 'society', '.', 'A', 'nation', ',', 'like', 'person', ',', 'body', '--', 'body', 'must', 'fed', 'clothed', 'housed', ',', 'invigorated', 'rested', ',', 'manner', 'measures', 'objectives', 'time', '.', 'A', 'nation', ',', 'like', 'person', ',', 'mind', '--', 'mind', 'must', 'kept', 'informed', 'alert', ',', 'must', 'know', ',', 'understands', 'hopes', 'needs', 'neighbors', '--', 'nations', 'live', 'within', 'narrowing', 'circle', 'world', '.', 'And', 'nation', ',', 'like', 'person', ',', 'something', 'deeper', ',', 'something', 'permanent', ',', 'something', 'larger', 'sum', 'parts', '.', 'It', 'something', 'matters', 'future', '--', 'calls', 'forth', 'sacred', 'guarding', 'present', '.', 'It', 'thing', 'find', 'difficult', '--', 'even', 'impossible', '--', 'hit', 'upon', 'single', ',', 'simple', 'word', '.', 'And', 'yet', 'understand', '--', 'spirit', '--', 'faith', 'America', '.', 'It', 'product', 'centuries', '.', 'It', 'born', 'multitudes', 'came', 'many', 'lands', '--', 'high', 'degree', ',', 'mostly', 'plain', 'people', ',', 'sought', ',', 'early', 'late', ',', 'find', 'freedom', 'freely', '.', 'The', 'democratic', 'aspiration', 'mere', 'recent', 'phase', 'human', 'history', '.', 'It', 'human', 'history', '.', 'It', 'permeated', 'ancient', 'life', 'early', 'peoples', '.', 'It', 'blazed', 'anew', 'middle', 'ages', '.', 'It', 'written', 'Magna', 'Charta', '.', 'In', 'Americas', 'impact', 'irresistible', '.', 'America', 'New', 'World', 'tongues', ',', 'peoples', ',', 'continent', 'new-found', 'land', ',', 'came', 'believed', 'could', 'create', 'upon', 'continent', 'new', 'life', '--', 'life', 'new', 'freedom', '.', 'Its', 'vitality', 'written', 'Mayflower', 'Compact', ',', 'Declaration', 'Independence', ',', 'Constitution', 'United', 'States', ',', 'Gettysburg', 'Address', '.', 'Those', 'first', 'came', 'carry', 'longings', 'spirit', ',', 'millions', 'followed', ',', 'stock', 'sprang', '--', 'moved', 'forward', 'constantly', 'consistently', 'toward', 'ideal', 'gained', 'stature', 'clarity', 'generation', '.', 'The', 'hopes', 'Republic', 'forever', 'tolerate', 'either', 'undeserved', 'poverty', 'self-serving', 'wealth', '.', 'We', 'know', 'still', 'far', 'go', ';', 'must', 'greatly', 'build', 'security', 'opportunity', 'knowledge', 'every', 'citizen', ',', 'measure', 'justified', 'resources', 'capacity', 'land', '.', 'But', 'enough', 'achieve', 'purposes', 'alone', '.', 'It', 'enough', 'clothe', 'feed', 'body', 'Nation', ',', 'instruct', 'inform', 'mind', '.', 'For', 'also', 'spirit', '.', 'And', 'three', ',', 'greatest', 'spirit', '.', 'Without', 'body', 'mind', ',', 'men', 'know', ',', 'Nation', 'could', 'live', '.', 'But', 'spirit', 'America', 'killed', ',', 'even', 'though', 'Nation', "'s", 'body', 'mind', ',', 'constricted', 'alien', 'world', ',', 'lived', ',', 'America', 'know', 'would', 'perished', '.', 'That', 'spirit', '--', 'faith', '--', 'speaks', 'us', 'daily', 'lives', 'ways', 'often', 'unnoticed', ',', 'seem', 'obvious', '.', 'It', 'speaks', 'us', 'Capital', 'Nation', '.', 'It', 'speaks', 'us', 'processes', 'governing', 'sovereignties', '48', 'States', '.', 'It', 'speaks', 'us', 'counties', ',', 'cities', ',', 'towns', ',', 'villages', '.', 'It', 'speaks', 'us', 'nations', 'hemisphere', ',', 'across', 'seas', '--', 'enslaved', ',', 'well', 'free', '.', 'Sometimes', 'fail', 'hear', 'heed', 'voices', 'freedom', 'us', 'privilege', 'freedom', 'old', ',', 'old', 'story', '.', 'The', 'destiny', 'America', 'proclaimed', 'words', 'prophecy', 'spoken', 'first', 'President', 'first', 'inaugural', '1789', '--', 'words', 'almost', 'directed', ',', 'would', 'seem', ',', 'year', '1941', ':', '``', 'The', 'preservation', 'sacred', 'fire', 'liberty', 'destiny', 'republican', 'model', 'government', 'justly', 'considered', 'deeply', ',', 'finally', ',', 'staked', 'experiment', 'intrusted', 'hands', 'American', 'people', '.', "''", 'If', 'lose', 'sacred', 'fire', '--', 'let', 'smothered', 'doubt', 'fear', '--', 'shall', 'reject', 'destiny', 'Washington', 'strove', 'valiantly', 'triumphantly', 'establish', '.', 'The', 'preservation', 'spirit', 'faith', 'Nation', ',', ',', 'furnish', 'highest', 'justification', 'every', 'sacrifice', 'may', 'make', 'cause', 'national', 'defense', '.', 'In', 'face', 'great', 'perils', 'never', 'encountered', ',', 'strong', 'purpose', 'protect', 'perpetuate', 'integrity', 'democracy', '.', 'For', 'muster', 'spirit', 'America', ',', 'faith', 'America', '.', 'We', 'retreat', '.', 'We', 'content', 'stand', 'still', '.', 'As', 'Americans', ',', 'go', 'forward', ',', 'service', 'country', ',', 'God', '.']

Removing stop words from Kennedy

The whole text are processed using word_tokenize and then stop words are removed from the text file. The following is the filtered sentence:

['Vice', 'President', 'Johnson', ',', 'Mr.', 'Speaker', ',', 'Mr.', 'Chief', 'Justice', ',', 'President', 'Eisenhower', ',', 'Vice', 'President', 'Nixon', ',', 'President', 'Truman', ',', 'reverend', 'clergy', ',', 'fellow', 'citizens', ',', 'observe', 'today', 'victory', 'party', ',', 'celebration', 'freedom', '--', 'symbolizing', 'end', ',', 'well', 'beginning', '--', 'signifying', 'renewal', ',', 'well', 'change', '.', 'For', 'I', 'sworn', 'I', 'Almighty', 'God', 'solemn', 'oath', 'forebears', 'l', 'prescribed', 'nearly', 'century', 'three', 'quarters', 'ago', '.', 'The', 'world', 'different', '.', 'For', 'man', 'holds', 'mortal', 'hands', 'power', 'abolish', 'forms', 'human', 'poverty', 'forms', 'human', 'life', '.', 'And', 'yet', 'revolutionary', 'beliefs', 'forebears', 'fought', 'still', 'issue', 'around', 'globe', '--', 'belief', 'rights', 'man', 'come', 'generosity', 'state', ',', 'hand', 'God', '.', 'We', 'dare', 'forget', 'today', 'heirs', 'first', 'revolution', '.', 'Let', 'word', 'go', 'forth', 'time', 'place', ',', 'friend', 'foe', 'alike', ',', 'torch', 'passed', 'new', 'generation', 'Americans', '--', 'born', 'century', ',', 'tempered', 'war', ',', 'disciplined', 'hard', 'bitter', 'peace', ',', 'proud', 'ancient', 'heritage', '--', 'unwilling', 'witness', 'permit', 'slow', 'undoing', 'human', 'rights', 'Nation', 'always', 'committed', ',', 'committed', 'today', 'home', 'around', 'world', '.', 'Let', 'every', 'nation', 'know', ',', 'whether', 'wishes', 'us', 'well', 'ill', ',', 'shall', 'pay', 'price', ',', 'bear', 'burden', ',', 'meet', 'hardship', ',', 'support', 'friend', ',', 'oppose', 'foe', ',', 'order', 'assure', 'survival', 'success', 'liberty', '.', 'This', 'much', 'pledge', '--', '.', 'To', 'old', 'allies', 'whose', 'cultural', 'spiritual', 'origins', 'share', ',', 'pledge', 'loyalty', 'faithful', 'friends', '.', 'United', ',', 'little', 'host', 'cooperative', 'ventures', '.', 'Divided', ',', 'little', '--', 'dare', 'meet', 'powerful', 'challenge', 'odds', 'split', 'asunder', '.', 'To', 'new', 'States', 'welcome', 'ranks', 'free', ',', 'pledge', 'word', 'one', 'form', 'colonial', 'control', 'shall', 'passed', 'away', 'merely', 'replaced', 'far', 'iron', 'tyranny', '.', 'We', 'shall', 'always', 'expect', 'find', 'supporting', 'view', '.', 'But', 'shall', 'always', 'hope', 'find', 'strongly', 'supporting', 'freedom', '--', 'remember', ',', 'past', ',', 'foolishly', 'sought', 'power', 'riding', 'back', 'tiger', 'ended', 'inside', '.', 'To', 'peoples', 'huts', 'villages', 'across', 'globe', 'struggling', 'break', 'bonds', 'mass', 'misery', ',', 'pledge', 'best', 'efforts', 'help', 'help', ',', 'whatever', 'period', 'required', '--', 'Communists', 'may', ',', 'seek', 'votes', ',', 'right', '.', 'If', 'free', 'society', 'help', 'many', 'poor', ',', 'save', 'rich', '.', 'To', 'sister', 'republics', 'south', 'border', ',', 'offer', 'special', 'pledge', '--', 'convert', 'good', 'words', 'good', 'deeds', '--', 'new', 'alliance', 'progress', '--', 'assist', 'free', 'men', 'free', 'governments', 'casting', 'chains', 'poverty', '.', 'But', 'peaceful', 'revolution', 'hope', 'become', 'prey', 'hostile', 'powers', '.', 'Let', 'neighbors', 'know', 'shall', 'join', 'oppose', 'aggression', 'subversion', 'anywhere', 'Americas', '.', 'And', 'let', 'every', 'power', 'know', 'Hemisphere', 'intends', 'remain', 'master', 'house', '.', 'To', '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', '--', 'enlarge', 'area', 'writ', 'may', 'run', '.', 'Finally', ',', 'nations', 'would', 'make', 'adversary', ',', 'offer', 'pledge', 'request', ':', 'sides', 'begin', 'anew', 'quest', 'peace', ',', 'dark', 'powers', 'destruction', 'unleashed', 'science', 'engulf', 'humanity', 'planned', 'accidental', 'self-destruction', '.', 'We', 'dare', 'tempt', 'weakness', '.', 'For', 'arms', 'sufficient', 'beyond', 'doubt', 'certain', 'beyond', 'doubt', 'never', 'employed', '.', 'But', 'neither', 'two', 'great', 'powerful', 'groups', 'nations', 'take', 'comfort', 'present', 'course', '--', 'sides', 'overburdened', 'cost', 'modern', 'weapons', ',', 'rightly', 'alarmed', 'steady', 'spread', 'deadly', 'atom', ',', 'yet', 'racing', 'alter', 'uncertain', 'balance', 'terror', 'stays', 'hand', 'mankind', "'s", 'final', 'war', '.', 'So', 'let', 'us', 'begin', 'anew', '--', 'remembering', 'sides', 'civility', 'sign', 'weakness', ',', 'sincerity', 'always', 'subject', 'proof', '.', 'Let', 'us', 'never', 'negotiate', 'fear', '.', 'But', 'let', 'us', 'never', 'fear', 'negotiate', '.', 'Let', 'sides', 'explore', 'problems', 'unite', 'us', 'instead', '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', '.', "''", 'And', 'beachhead', 'cooperation', 'may', 'push', 'back', 'jungle', 'suspicion', ',', 'let', 'sides', 'join', 'creating', 'new', 'endeavor', ',', 'new', 'balance', 'power', ',', 'new', 'world', 'law', ',', 'strong', 'weak', 'secure', 'peace', 'preserved', '.', 'All', 'finished', 'first', '100', 'days', '.', 'Nor', 'finished', 'first', '1,000', 'days', ',', 'life', 'Administration', ',', 'even', 'perhaps', 'lifetime', 'planet', '.', 'But', 'let', 'us', 'begin', '.', 'In', 'hands', ',', 'fellow', 'citizens', ',', 'mine', ',', 'rest', 'final', 'success', 'failure', 'course', '.', 'Since', 'country', 'founded', ',', 'generation', 'Americans', 'summoned', 'give', 'testimony', 'national', 'loyalty', '.', 'The', 'graves', 'young', 'Americans', 'answered', 'call', 'service', 'surround', 'globe', '.', 'Now', 'trumpet', 'summons', 'us', '--', 'call', 'bear', 'arms', ',', 'though', 'arms', 'need', ';', 'call', 'battle', ',', 'though', 'embattled', '--', 'call', 'bear', 'burden', 'long', 'twilight', 'struggle', ',', 'year', 'year', ',', '``', 'rejoicing', 'hope', ',', 'patient', 'tribulation', "''", '--', 'struggle', 'common', 'enemies', 'man', ':', 'tyranny', ',', 'poverty', ',', 'disease', ',', 'war', '.', 'Can', 'forge', 'enemies', 'grand', 'global', 'alliance', ',', 'North', 'South', ',', 'East', 'West', ',', 'assure', 'fruitful', 'life', 'mankind', '?', 'Will', 'join', 'historic', 'effort', '?', 'In', 'long', 'history', 'world', ',', 'generations', 'granted', 'role', 'defending', 'freedom', 'hour', 'maximum', 'danger', '.', 'I', 'shrink', 'responsibility', '--', 'I', 'welcome', '.', 'I', 'believe', 'us', 'would', 'exchange', 'places', 'people', 'generation', '.', 'The', 'energy', ',', 'faith', ',', 'devotion', 'bring', 'endeavor', 'light', 'country', 'serve', '--', 'glow', 'fire', 'truly', 'light', 'world', '.', 'And', ',', 'fellow', 'Americans', ':', 'ask', 'country', '--', 'ask', 'country', '.', 'My', 'fellow', 'citizens', 'world', ':', 'ask', 'America', ',', 'together', 'freedom', 'man', '.', 'Finally', ',', 'whether', 'citizens', 'America', 'citizens', 'world', ',', 'ask', 'us', 'high', 'standards', 'strength', 'sacrifice', 'ask', '.', 'With', 'good', 'conscience', 'sure', 'reward', ',', 'history', 'final', 'judge', 'deeds', ',', 'let', 'us', 'go', 'forth', 'lead', 'land', 'love', ',', 'asking', 'His', 'blessing', 'His', 'help', ',', 'knowing', 'earth', 'God', "'s", 'work', 'must', 'truly', '.']

Removing stop words from Nixon

The whole text are processed using word_tokenize and then stop words are removed from the text file. The following is the filtered sentence:

['Mr.', 'Vice', 'President', ',', 'Mr.', 'Speaker', ',', 'Mr.', 'Chief', 'Justice', ',', 'Senator', 'Cook', ',', 'Mrs.', 'Eisenhower', ',', 'fellow', 'citizens', 'great', 'good', 'country', 'share', 'together', ':', 'When', 'met', 'four', 'years', 'ago', ',', 'America', 'bleak', 'spirit', ',', 'depressed', 'prospect', 'seemingly', 'endless', 'war', 'abroad', 'destructive', 'conflict', 'home', '.', 'As', 'meet', 'today', ',', 'stand', 'threshold', 'new', 'era', 'peace', 'world', '.', 'The', 'central', 'question', 'us', ':', 'How', 'shall', 'use', 'peace', '?', 'Let', 'us', 'resolve', 'era', 'enter', 'postwar', 'periods', 'often', ':', 'time', 'retreat', 'isolation', 'leads', 'stagnation', 'home', 'invites', 'new', 'danger', 'abroad', '.', 'Let', 'us',

'resolve', 'become', ':', 'time', 'great', 'responsibilities', 'greatly', 'borne', ',', 'renew', 'spirit', 'promise', 'America', 'enter', 'third', 'century', 'nation', '.', 'This', 'past', 'year', 'saw', 'far-reaching', 'results', 'new', 'policies', 'peace', '.', 'By', 'continuing', 'revitalize', 'traditional', 'friendships', ',', 'missions', 'Peking', 'Moscow', ',', 'able', 'establish', 'base', 'new', 'durable', 'pattern', 'relationships', 'among', 'nations', 'world', '.', 'Because', 'America', "'s", 'bold', 'initiatives', ',', '1972', 'long', 'remembered', 'year', 'greatest', 'progress', 'since', 'end', 'World', 'War', 'II', 'toward', 'lasting', 'peace', 'world', '.', 'The', 'peace', 'seek', 'world', 'flimsy', 'peace', 'merely', 'interlude', 'wars', ',', 'peace', 'endure', 'generations', 'come', '.', 'It', 'important', 'understand', 'necessity', 'limitations', 'America', "'s", 'role', 'maintaining', 'peace', '.', 'Unless', 'America', 'work', 'preserve', 'peace', ',', 'peace', '.', 'Unless', 'America', 'work', 'preserve', 'freedom', ',', 'freedom', '.', 'But', 'let', 'us', 'clearly', 'understand', 'new', 'nature', 'America', "'s", 'role', ',', 'result', 'new', 'policies', 'adopted', 'past', 'four', 'years', '.', 'We', 'shall', 'respect', 'treaty', 'commitments', '.', 'We', 'shall', 'support', 'vigorously', 'principle', 'country', 'right', 'impose', 'rule', 'another', 'force', '.', 'We', 'shall', 'continue', ',', 'era', 'negotiation', ',', 'work', 'limitation', 'nuclear', 'arms', ',', 'reduce', 'danger', 'confrontation', 'great', 'powers', '.', 'We', 'shall', 'share', 'defending', 'peace', 'freedom', 'world', '.', 'But', 'shall', 'expect', 'others', 'share', '.', 'The', 'time', 'passed', 'America', 'make', 'every', 'nation', "'s", 'conflict', ',', 'make', 'every', 'nation', "'s", 'future', 'responsibility', ',', 'presume', 'tell', 'people', 'nations', 'manage', 'affairs', '.', 'Just', 'respect', 'right', 'nation', 'determine', 'future', ',', 'also', 'recognize', 'responsibility', 'nation', 'secure', 'future', '.', 'Just', 'America', "'s", 'role', 'indispensable', 'preserving', 'world', "'s", 'peace', ',', 'nation', "'s", 'role', 'indispensable', 'preserving', 'peace', '.', 'Together', 'rest', 'world', ',', 'let', 'us', 'resolve', 'move', 'forward', 'beginnings', 'made', '.', 'Let', 'us', 'continue', 'bring', 'walls', 'hostility', 'divided', 'world', 'long', ',', 'build', 'place', 'bridges', 'understanding', '--', 'despite', 'profound', 'differences', 'systems', 'government', ',', 'people', 'world', 'friends', '.', 'Let', 'us', 'build', 'structure', 'peace', 'world', 'weak', 'safe', 'strong', '--', 'respects', 'right', 'live', 'different', 'system', '--', 'would', 'influence', 'others', 'strength', 'ideas', ',', 'force', 'arms', '.', 'Let', 'us', 'accept', 'high', 'responsibility', 'burden', ',', 'gladly', '--', 'gladly', 'chance', 'build', 'peace', 'noblest', 'endeavor', 'nation', 'engage', ';', 'gladly', ',', 'also', ',', 'act', 'greatly', 'meeting', 'responsibilities', 'abroad', 'remain', 'great', 'Nation', ',', 'remain', 'great', 'Nation', 'act', 'greatly', 'meeting', 'challenges', 'home', '.', 'We', 'chance', 'today', 'ever', 'history', 'make', 'life', 'better', 'America', '--', 'ensure', 'better', 'education', ',', 'better', 'health', ',', 'better', 'housing', ',', 'better', 'transportation', ',', 'cleaner', 'environment', '--', 'restore', 'respect', 'law', ',', 'make', 'communities', 'livable', '--', 'insure', 'God-given', 'right', 'every', 'American', 'full', 'equal', 'opportunity', '.', 'Because', 'range', 'needs', 'great', '--', 'reach', 'opportunities', 'great', '--', 'let', 'us', 'bold', 'determination', 'meet', 'needs', 'new', 'ways', '.', 'Just', 'building', 'structure', 'peace', 'abroad', 'required', 'turning', 'away', 'old', 'policies', 'failed', ',', 'building', 'new', 'era', 'progress', 'home', 'requires', 'turning', 'away', 'old', 'policies', 'failed', '.', 'Abroad', ',', 'shift', 'old', 'policies', 'new', 'retreat', 'responsibilities', ',', 'better', 'way', 'peace', '.', 'And', 'home', ',', 'shift', 'old', 'policies', 'new', 'retreat', 'responsibilities', ',', 'better', 'way', 'progress', '.', 'Abroad', 'home', ',', 'key', 'new', 'responsibilities', 'lies', 'placing', 'division', 'responsibility', '.', 'We', 'lived', 'long', 'consequences', 'attempting', 'gather', 'power', 'responsibility', 'Washington', '.', 'Abroad', 'home', ',', 'time', 'come', 'turn', 'away', 'condescending', 'policies', 'paternalism', '--', '``', 'Washington', 'knows', 'best', '.', "''", 'A', 'person', 'expected', 'act', 'responsibly', 'responsibility', '.', 'This', 'human', 'nature', '.', 'So', 'let', 'us', 'encourage', 'individuals', 'home', 'nations', 'abroad', ',', 'decide', '.', 'Let', 'us', 'locate', 'responsibility', 'places', '.', 'Let', 'us', 'measure', 'others', '.', 'That', 'today', 'I', 'offer', 'promise', 'purely', 'governmental', 'solution', 'every', 'problem', '.', 'We', 'lived', 'long', 'false', 'promise', '.', 'In', 'trusting', 'much', 'government', ',', 'asked', 'deliver', '.', 'This', 'leads', 'inflated', 'expectations', ',', 'reduced', 'individual', 'effort', ',', 'disappointment', 'frustration', 'erode', 'confidence', 'government', 'people', '.', 'Government', 'must', 'learn', 'take', 'less', 'people', 'people', '.', 'Let', 'us', 'remember', 'America', 'built', 'government', ',', 'people', '--', 'welfare', ',', 'work', '--', 'shirking', 'responsibility', ',', 'seeking', 'responsibility', '.', 'In', 'lives', ',', 'let', 'us', 'ask', '--', 'government', ',', 'I', '?', 'In', 'challenges', 'face', 'together', ',', 'let', 'us', 'ask', '--', 'government', 'help', ',', 'I', 'help', '?', 'Your', 'National', 'Government', 'great', 'vital', 'role', 'play', '.', 'And', 'I', 'pledge', 'Government', 'act', ',', 'act', 'boldly', 'lead', 'boldly', '.', 'But', 'important', 'role', 'every', 'one', 'us', 'must', 'play', ',', 'individual', 'member', 'community', '.', 'From', 'day', 'forward', ',', 'let', 'us', 'make', 'solemn', 'commitment', 'heart', ':', 'bear', 'responsibility', ',', 'part', ',', 'live', 'ideals', '--', 'together', ',', 'see', 'dawn', 'new', 'age', 'progress', 'America', ',', 'together', ',', 'celebrate', '200th', 'anniversary', 'nation', ',', 'proud', 'fulfillment', 'promise', 'world', '.', 'As', 'America', "'s", 'longest', 'difficult', 'war', 'comes', 'end', ',', 'let', 'us', 'learn', 'debate', 'differences', 'civility', 'decency', '.', 'And', 'let', 'us', 'reach', 'one', 'precious', 'quality', 'government', 'provide', '--', 'new', 'level', 'respect', 'rights', 'feelings', 'one', 'another', ',', 'new', 'level', 'respect',

'individual', 'human', 'dignity', 'cherished', 'birthright', 'every', 'American', '.', 'Above', 'else', ',', 'time', 'come', 'us', 'renew', 'faith', 'America', '.', 'In', 'recent', 'years', ',', 'faith', 'challenged', '.', 'Our', 'children', 'taught', 'ashamed', 'country', ',', 'ashamed', 'parents', ',', 'ashamed', 'America', "'s", 'record', 'home', 'role', 'world', '.', 'At', 'every', 'turn', ',', 'beset', 'find', 'everything', 'wrong', 'America', 'little', 'right', '.', 'But', 'I', 'confident', 'judgment', 'history', 'remarkable', 'times', 'privileged', 'live', '.', 'America', "'s", 'record', 'century', 'unparalleled', 'world', "'s", 'history', 'responsibility', ',', 'generosity', ',', 'creativity', 'progress', '.', 'Let', 'us', 'proud', 'system', 'produced', 'provided', 'freedom', 'abundance', ',', 'widely', 'shared', ',', 'system', 'history', 'world', '.', 'Let', 'us', 'proud', 'four', 'wars', 'engaged', 'century', ',', 'including', 'one', 'bringing', 'end', ',', 'fought', 'selfish', 'advantage', ',', 'help', 'others', 'resist', 'aggression', '.', 'Let', 'us', 'proud', 'bold', ',', 'new', 'initiatives', ',', 'steadfastness', 'peace', 'honor', ',', 'made', 'break-through', 'toward', 'creating', 'world', 'world', 'known', '--', 'structure', 'peace', 'last', ',', 'merely', 'time', ',', 'generations', 'come', '.', 'We', 'embarking', 'today', 'era', 'presents', 'challenges', 'great', 'nation', ',', 'generation', ',', 'ever', 'faced', '.', 'We', 'shall', 'answer', 'God', ',', 'history', ',', 'conscience', 'way', 'use', 'years', '.', 'As', 'I', 'stand', 'place', ',', 'hallowed', 'history', ',', 'I', 'think', 'others', 'stood', '.', 'I', 'think', 'dreams', 'America', ',', 'I', 'think', 'recognized', 'needed', 'help', 'far', 'beyond', 'order', 'make', 'dreams', 'come', 'true', '.', 'Today', ',', 'I', 'ask', 'prayers', 'years', 'ahead', 'I', 'may', 'God', "'s", 'help', 'making', 'decisions', 'right', 'America', ',', 'I', 'pray', 'help', 'together', 'may', 'worthy', 'challenge', '.', 'Let', 'us', 'pledge', 'together', 'make', 'next', 'four', 'years', 'best', 'four', 'years', 'America', "'s", 'history', ',', '200th', 'birthday', 'America', 'young', 'vital', 'began', ',', 'bright', 'beacon', 'hope', 'world', '.', 'Let', 'us', 'go', 'forward', 'confident', 'hope', ',', 'strong', 'faith', 'one', 'another', ',', 'sustained', 'faith', 'God', 'created', 'us', ',', 'striving', 'always', 'serve', 'His', 'purpose', '.']

3.3 Determining the word which occurs the most number of times in his inaugural address for each president and it's top three words. (after removing the stopwords)

Word that occurs most number of times in Roosevelt

[(',', 77),
 ('.', 68),
 ('--', 25),
 ('It', 13),
 ('The', 10),
 ('know', 10),
 ('We', 10),
 ('spirit', 9),
 ('life', 9),
 ('us', 8)]

These are the most occurring words but since the first most and top three words doesn't give any meaning or since it is considered as supporting words, 'We' is considered as the word that occurs most number of times. So the top three words are 'We', 'spirit' and 'life'

Word that occurs most number of times in Kennedy

[(',', 84),
 ('.', 50),
 ('--', 25),
 ('us', 12),
 ('world', 8),
 ('Let', 8),
 ('let', 8),

('sides', 8),
('new', 7),
('pledge', 7)]

These are the most occurring words but since the first most and top three words doesn't give any meaning or since it is considered as supporting words, 'us' is considered as the word that occurs most number of times. So the top three words are 'us', 'world' and 'Let'

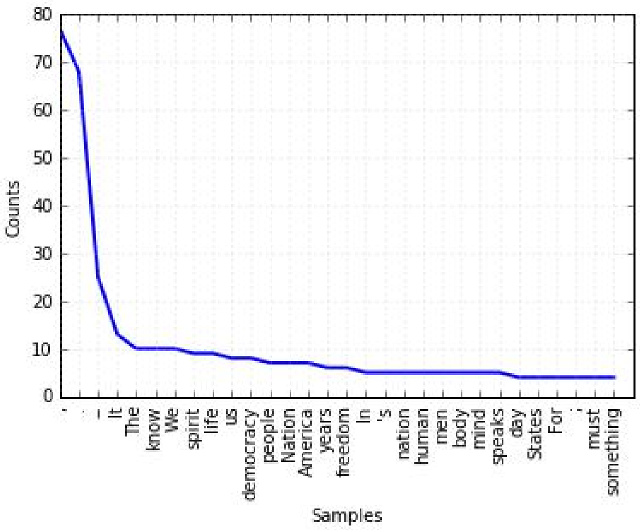
Word that occurs most number of times in Nixon

[(',', 96),
('.', 65),
('us', 26),
('America', 21),
('peace', 19),
('world', 17),
('--', 17),
('new', 15),
("'s", 14),
('Let', 13)]

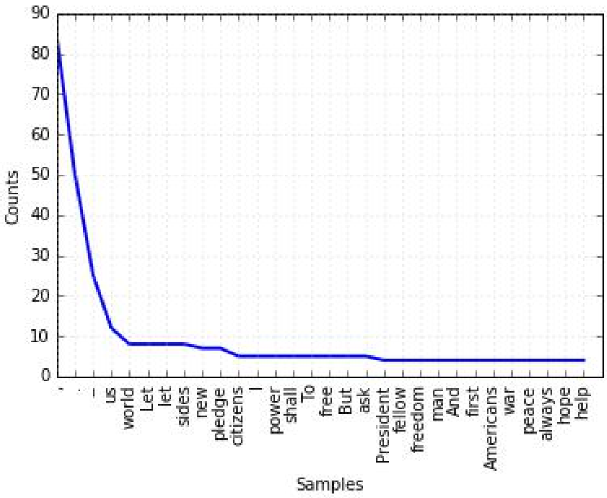
These are the most occurring words but since the first most and top three words doesn't give any meaning or since it is considered as supporting words, 'us' is considered as the word that occurs most number of times. So the top three words are 'us', 'America' and 'peace'

3.4 Plotting the word cloud of each of the three speeches

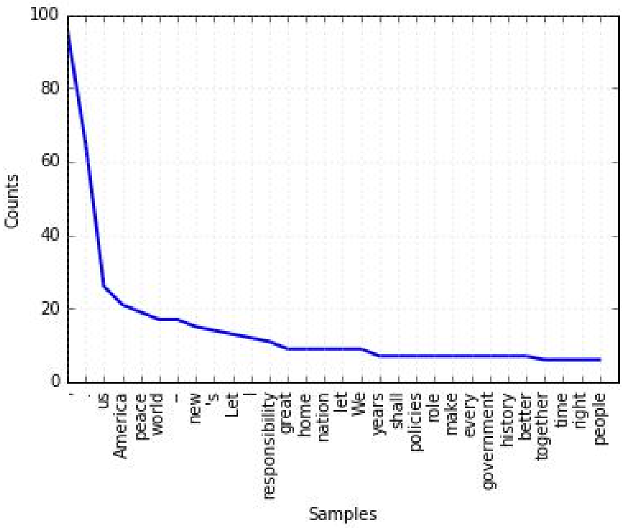
Frequency distribution of words that occurs in Roosevelt



Frequency distribution of words that occurs in Kennedy



Frequency distribution of words that occurs in Nixon



4 Appendix A – Source Code

Karthiheswar_MachineLearning.ipynb