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.

* 1. Data description and Inferences
  2. Data walkthrough (head() .info(), shape)
  3. Data validation (Null check, Summary and Skewness)
  4. Exploratory Data Analysis (Null, Data Types, Shape, Univariate & Bivariate Analysis, Outliers)
  5. Inferences Distribution plots, Box plots, Correlation plots, Plots for Categorical features.
  6. Encode the data for the features with String Values.
  7. Is Scaling necessary or not? Check the 5-point summary to support the Scaling.
  8. Object data should be converted into Categorical/Number type and fit to models.
  9. Ratio for the split and train-test-split should be explained.
  10. Apply Logistic Regression and LDA and Interpret inferences of both models. Logical explanation to select the hyper parameters involved in each model. Comment if model is overfit or underfit.
  11. Apply KNN and Naïve Bayes and explain the choice of Hyper parameters. Calculate Train-Test accuracies for each model. Comment if the model is overfit or underfit.
  12. Model Tuning, Bagging and Boosting.
  13. Apply GridSearchCV on all models and make models on best params. Explain the hyper parameters for GridSearch. Comment on feature importance.
  14. Performance Metrics for train and test sets using Accuracy, Confusion Matrix, ROC curve and ROC\_AUC score, classification report. Compare and comment on all model’s performance metrics in a table. Describe which model is optimized, also explain which model is best for the current problem.
  15. Explain the insights and recommendations.

Problem-2:

In this 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:

President Franklin D. Roosevelt in 1941

President John F. Kennedy in 1961

President Richard Nixon in 1973

* 1. Find the number of characters, words and sentences for the mentioned documents. (Hint: use .words(), .raw(), .sent() for extracting counts).
  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. Which word occurs the most number of times in his inaugural address for each president? Mention the top three words. (after removing the stopwords).
  4. Plot the word cloud of each of the three speeches. (after removing the stopwords).