

STEP 1: DATA PRE-PROCESSING BEFORE NAIVE-BAYES

- In this step, stop words are removed from the tweets. For this, a list of 1628 stop words is downloaded from a GitHub repository on basis of which these words are removed from tweets.
- Bags of words are also created on basis of target values. If the target value is 1: A bag of truthful words is created and for the target value 0: a bag of deceptive words is created.
- The probability of truthful and deceptive sentences is also calculated in this step.

STEP 2: NAIVE-BAYES CLASSIFIER

- In this step, a Naive-Bayes classifier is implemented where the conditional probability of a sentence is calculated given the word.
- Laplace smoothing is also applied in this step.

STEP 3: DATA PRE-PROCESSING AFTER NAIVE-BAYES

- All the unnecessary and less important columns like addtiionsal_comments, sample.name, create.date, etc. are removed from the test dataset.
- Merged same columns for different experts x and y by taking mean of them. These columns are still.exists, disagree.with, sarcasm, is.about.the.holocaust, sentiment.rating, IHRA.section, and calling.out.
- Removed the above columns with x and y ratings separately.

STEP 4: ONE HOT VECTOR ENCODING

- Performed one hot vector encoding on the key column.

STEP 5: SPLIT TRAIN DATA

- Split train data 80-20 as train-test.

STEP 6: RANDOM FOREST CLASSIFIER

- Predicted the results using random forest classifier.

STEP 7: RESULTS

- Displayed results as accuracy score and classification report for the given data.