**Muhammad Hadi**

**25244**

**MULTINOMIALNB:-**

* **ACCURACY ON VALIDATION DATASET:-**

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7637333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7957333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8214666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8272

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8369333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8425333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7609333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7944

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8213333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8257333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8348

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8409333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.7970666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.8318666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8462666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8525333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8601333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.8622666666666666

Best parameters: {}

**ACCURACY FOR TEST:-**

ACCURACY TEST: 0.8585

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.7969333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.83

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8468

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8482666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8550666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.8616

Best parameters: {}

* **WITH STOP WORD AND STEMMING:**

**ACCURACY ON VALIDATION DATASET:-**

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.82

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.8304

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8398666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8457333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8465333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8498666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.82

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.8304

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8397333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8462666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8444

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.848

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.8257333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.8357333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8437333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8529333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8541333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.8549333333333333

Best parameters: {}

**ACCURACY ON TEST:-** ACCURACY TEST: 0.85395

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.8257333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.8357333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8433333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8510666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8526666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.8532

Best parameters: {}

* **WITH STOP WORD AND LEMMATIZATION:-**

**ACCURACY ON VALIDATION DATASET:-**

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.8202666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.8344

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8434666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8481333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8529333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8541333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.8206666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.8349333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8434666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8481333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8516

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8546666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.8254666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.8389333333333333

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8502666666666666

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8548

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8585333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.86

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.8261333333333334

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.8384

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8504

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.8546666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.8586666666666667

Best parameters: {}

Model Name: MultinomialNB

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.8601333333333333

Best parameters: {}

ACCURACY TEST: 0.81425

**KNN**

* **ACCURACY ON VALIDATION DATA :-**

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.6325333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.6321333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.6314666666666666

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.6297333333333334

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.6230666666666667

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.6257333333333334

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.6338666666666667

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.6324

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.6282666666666666

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.6297333333333334

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.6268

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.6222666666666666

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.6957333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.6992

Best parameters: {'n\_neighbors': 9}

ACCURACY TESt: 0.63

**LEMMATIZED:-**

**Accuracy:-** 0.8045333333333333

**ACCURACY TEST:-** 0.57915

**STEMMED: -**

ACCURACY:- 0.8138666666666666

ACCURACY TEST: 0.7255

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.6885333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.6929333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.6928

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.6944

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.6918666666666666

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.6950666666666667

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.6917333333333333

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.6944

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.6958666666666666

Best parameters: {'n\_neighbors': 9}

Model Name: KNeighborsClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.6961333333333334

Best parameters: {'n\_neighbors': 9}

**DECISION TREE CLASSIFIER :-**

* **ACCURACY ON VALIDATION SET :-**

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7169333333333333

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7197333333333333

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.7272

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.7265333333333334

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.7257333333333333

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.7270666666666666

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7138666666666666

Best parameters: {'max\_depth': 9}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7218666666666667

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.7258666666666667

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.726

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.726

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.7278666666666667

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.702

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.7190666666666666

Best parameters: {'max\_depth': 9}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.7244

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.7212

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.7252

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.7253333333333334

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.7022666666666667

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.7201333333333333

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.7261333333333333

Best parameters: {'max\_depth': 10}

**LEMMATIZED:-**

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.7689333333333334

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: TfidfVectorizer

Accuracy: 0.7212

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: TfidfVectorizer

Accuracy: 0.7209333333333333

Best parameters: {'max\_depth': 10}

Model Name: DecisionTreeClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: TfidfVectorizer

Accuracy: 0.7228

Best parameters: {'max\_depth': 10}

**GRADIENT BOOSTING CLASSIFIER**

**KAGGLE NOTEBOOK GOT STUCK AND GAVE AN OUTPUT FOR ONE VARIATION AFTER 4HOURS.**

* **ACCURACY ON VALIDATION:-**

Model Name: GradientBoostingClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.8162666666666667

Best parameters: {'max\_depth': 5, 'min\_samples\_leaf': 32, 'min\_samples\_split': 3, 'n\_estimators': 200}

**STEMMED:-**

Model Name: GradientBoostingClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.8821333333333333

Best parameters: {'max\_depth': 5, 'min\_samples\_leaf': 32, 'min\_samples\_split': 3, 'n\_estimators': 200}

**ACCURACY TEST: 0.82305**

**LEMMATIZED :-**

Model Name: GradientBoostingClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.878

Best parameters: {'max\_depth': 5, 'min\_samples\_leaf': 32, 'min\_samples\_split': 3, 'n\_estimators': 200}

ACCURACY TEST: 0.8149

I believe that if Kaggle notebook functioned correctly then Gradient boosting should have produced the best score for us as we saw in IML.

**RANDOM FOREST**

* **ACCURACY ON VALIDATION DATASET**

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7794666666666666

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 32, 'min\_samples\_split': 7, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7969333333333334

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 7, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8125333333333333

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 3, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8136

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 32, 'min\_samples\_split': 3, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8068

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 32, 'min\_samples\_split': 5, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8190666666666667

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 3, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 500

Vectorizer: CountVectorizer

Accuracy: 0.7769333333333334

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 1000

Vectorizer: CountVectorizer

Accuracy: 0.7964

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 3, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 2000

Vectorizer: CountVectorizer

Accuracy: 0.8076

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 3, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 3000

Vectorizer: CountVectorizer

Accuracy: 0.8141333333333334

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 3, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 4000

Vectorizer: CountVectorizer

Accuracy: 0.8102666666666667

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 32, 'min\_samples\_split': 7, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8162666666666667

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 200}

**ACCURACY ON TEST:-**

ACCURACY TEST: 0.821

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 500

Vectorizer: TfidfVectorizer

Accuracy: 0.7764

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 150}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 1000

Vectorizer: TfidfVectorizer

Accuracy: 0.7989333333333334

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 32, 'min\_samples\_split': 3, 'n\_estimators': 200}

Model Name: RandomForestClassifier

Ngram Range: (1, 2)

Max Features: 2000

Vectorizer: TfidfVectorizer

Accuracy: 0.8076

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 150}

**STEMMED AND STOP WORD REMOVAL RF: -**

**ACCURACY ON VALIDATION: -**

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8273333333333334

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 200}

ACCURACY TEST: 0.8235

**LEMMATIZED AND STOP WORD REMOVAL RF: -**

**ACCURACY ON VALIDATION: -**

Model Name: RandomForestClassifier

Ngram Range: (1, 3)

Max Features: 5000

Vectorizer: CountVectorizer

Accuracy: 0.8278666666666666

Best parameters: {'max\_depth': 7, 'min\_samples\_leaf': 16, 'min\_samples\_split': 5, 'n\_estimators': 200}

ACCURACY TEST: 0.818

**Took me 2 whole days for completing this part as I applied a grid search and multiple models together at a time. I was simultaneously training different models on Kaggle/Colab and my local Machine. Multinomial NB was trained and evaluated on my local machine models like Gradient Boosting and Random forest was done on cloud. Gradient Boosting got stuck and after 4 hours it outputted for only one model. Since I did GridSearch so can’t tell exact time for each model a lot of models were trained and I lost track of time in it**