Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score: 0.7203038340615803

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 5}, F1 Score: 0.6527792550526057

Linear SVM - Best Hyperparameters: {'C': 10}, F1 Score: 0.7132409297353616

RBF SVM - Best Hyperparameters: {'C': 100, 'gamma': 0.01}, F1 Score: 0.7282374035917585

Decision Tree - Best Hyperparameters: {'max\_depth': None}, F1 Score: 0.6147996937646056

Random Forest - Best Hyperparameters: {'max\_depth': None, 'n\_estimators': 150}, F1 Score: 0.6801940770880587

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 150}, F1 Score: 0.6132034225231539

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7379389562318235

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 5}, F1 Score on Validation Set: 0.6480313279572765

Linear SVM - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7233841236896545

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7429712002723504

Decision Tree - Best Hyperparameters: {'max\_depth': 30}, F1 Score on Validation Set: 0.6315373533129579

Random Forest - Best Hyperparameters: {'max\_depth': 30, 'n\_estimators': 75}, F1 Score on Validation Set: 0.49090214522306896

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6172670220784441

Bernoulli Naive Bayes - Best Hyperparameters: {'alpha': 0.1}, F1 Score on Validation Set: 0.6855556466874683

Logistic Regression - Best Hyperparameters: {'C': 12}, F1 Score on Validation Set: 0.737824797381176

Linear SVM - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.7243871503482533

RBF SVM - Best Hyperparameters: {'C': 12, 'gamma': 0.05}, F1 Score on Validation Set: 0.7367826199730427

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7030731097972011

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 7}, F1 Score on Validation Set: 0.6167766085101652

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.6967010482633262

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7051347131372998

Decision Tree - Best Hyperparameters: {'max\_depth': 40}, F1 Score on Validation Set: 0.6025502749584165

Random Forest - Best Hyperparameters: {'max\_depth': 30, 'n\_estimators': 75}, F1 Score on Validation Set: 0.4975734219944583

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6055817915102081

cleaning\_email\_addresses,

cleaning\_URLs,

cleaning\_punctuations,

cleaning\_repeating\_char,

cleaning\_numbers,

cleaning\_stopwords,

tokenize,

stem,

lemmatize,

Logistic Regression - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.7205234706933995

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 5}, F1 Score on Validation Set: 0.6523162359836508

Linear SVM - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7096186612654589

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7267544657504407

Decision Tree - Best Hyperparameters: {'max\_depth': 40}, F1 Score on Validation Set: 0.6138696963720349

Random Forest - Best Hyperparameters: {'max\_depth': 30, 'n\_estimators': 100}, F1 Score on Validation Set: 0.4712498722359911

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6093266355522642

cleaning\_punctuations,

cleaning\_stopwords,

cleaning\_URLs,

cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_numbers,

tokenize,

stem,

lemmatize,

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.723233265974952

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 5}, F1 Score on Validation Set: 0.6529544040473312

Linear SVM - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7098671090962916

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7279855393624551

Decision Tree - Best Hyperparameters: {'max\_depth': 40}, F1 Score on Validation Set: 0.6184921581677356

Random Forest - Best Hyperparameters: {'max\_depth': 30, 'n\_estimators': 50}, F1 Score on Validation Set: 0.475084714261948

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6078064169773396

cleaning\_punctuations,

cleaning\_stopwords,

cleaning\_URLs,

cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_numbers,

tokenize,

stem,

lemmatize,

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7227284168651961

cleaning\_punctuations,

cleaning\_stopwords,

# cleaning\_URLs,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_numbers,

tokenize,

stem,

lemmatize,

Logistic Regression - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.7173099080091644

cleaning\_punctuations,

# cleaning\_stopwords,

cleaning\_URLs,

cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_numbers,

tokenize,

stem,

lemmatize,

]

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7289717143510865

RBF SVM - Best Hyperparameters: {'C': 100, 'gamma': 0.01}, F1 Score on Validation Set: 0.7274546132864308

cleaning\_punctuations,

cleaning\_numbers,

cleaning\_URLs,

cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_stopwords,

tokenize,

stem,

lemmatize,

Logistic Regression - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.7199323252996099

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7274382794796278

Logistic Regression - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.7309601838669122

TFID

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7281042570519671

Count

Logistic Regression - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.7309601838669122

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 5}, F1 Score on Validation Set: 0.656062031191163

Linear SVM - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7204148492613949

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7297385119328678

Decision Tree - Best Hyperparameters: {'max\_depth': 10}, F1 Score on Validation Set: 0.5290596784405219

Random Forest - Best Hyperparameters: {'max\_depth': 10, 'n\_estimators': 75}, F1 Score on Validation Set: 0.3972787758279014

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.5975591140912313

TFID

# Define classifiers with their respective hyperparameter grids

classifiers = {

'Logistic Regression': (LogisticRegression(max\_iter=1000), {'C': [7, 10, 20]}),

# 'Nearest Neighbors': (KNeighborsClassifier(), {'n\_neighbors': [3, 5, 7]}),

# 'Linear SVM': (SVC(kernel='linear'), {'C': [ 1, 10, 100]}),

'RBF SVM': (SVC(kernel='rbf'), {'C': [7, 10, 20], 'gamma': [ 0.1, 0.5, 10]}),

# 'Decision Tree': (DecisionTreeClassifier(), {'max\_depth': [2, 5, 10]}),

# 'Random Forest': (RandomForestClassifier(), {'n\_estimators': [50, 75, 100], 'max\_depth': [2, 5, 10]}),

# 'AdaBoost': (AdaBoostClassifier(), {'n\_estimators': [50, 75, 100], 'learning\_rate': [ 0.01, 0.1, 1]})

}

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.7263455004828918

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7308295714856845

Tfid

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7300866965669076

RBF SVM - Best Hyperparameters: {'C': 7, 'gamma': 0.1}, F1 Score on Validation Set: 0.7301427331201588

Logistic Regression - Best Hyperparameters: {'C': 10}, F1 Score on Validation Set: 0.7300866965669076

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 3}, F1 Score on Validation Set: 0.53891114747259

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7112369284331748

RBF SVM - Best Hyperparameters: {'C': 7, 'gamma': 0.1}, F1 Score on Validation Set: 0.7301427331201588

Decision Tree - Best Hyperparameters: {'max\_depth': 10}, F1 Score on Validation Set: 0.5430376493448044

Random Forest - Best Hyperparameters: {'max\_depth': 10, 'n\_estimators': 75}, F1 Score on Validation Set: 0.3972787758279014

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.5969810344340787

Execution Time: 1239.616839170456 seconds

No spell check

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.7423096257842011

RBF SVM - Best Hyperparameters: {'C': 7, 'gamma': 0.1}, F1 Score on Validation Set: 0.7121494628500796

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_punctuations,

tokenize,

# stem,

# lemmatize,

# cleaning\_stopwords,

# remove\_accent\_marks,

# cleaning\_ascii,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

# cleaning\_numbers,

# cleaning\_email\_addresses,

# correct\_spelling,

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.745697237273081

RBF SVM - Best Hyperparameters: {'C': 7, 'gamma': 0.1}, F1 Score on Validation Set: 0.7138102118781517

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

# stem,

# lemmatize,

# cleaning\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

# cleaning\_numbers,

# cleaning\_email\_addresses,

# correct\_spelling,

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.744512007504866

RBF SVM - Best Hyperparameters: {'C': 20, 'gamma': 0.1}, F1 Score on Validation Set: 0.7199698230957782

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.744512007504866

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 3}, F1 Score on Validation Set: 0.46522402675598307

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7255230774876992

RBF SVM - Best Hyperparameters: {'C': 20, 'gamma': 0.1}, F1 Score on Validation Set: 0.7199698230957782

Decision Tree - Best Hyperparameters: {'max\_depth': 10}, F1 Score on Validation Set: 0.5657358003489309

Random Forest - Best Hyperparameters: {'max\_depth': 10, 'n\_estimators': 50}, F1 Score on Validation Set: 0.3972787758279014

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6308550667981022

Naive Bayes - Best Hyperparameters: {}, F1 Score on Validation Set: 0.6855179121590459

Classifiers: 8%|█████▊ | 1/12 [00:49<09:00, 49.13s/it]

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.744512007504866

Classifiers: 17%|███████████▋ | 2/12 [01:03<04:49, 28.90s/it]

Nearest Neighbors - Best Hyperparameters: {'n\_neighbors': 3}, F1 Score on Validation Set: 0.46522402675598307

Classifiers: 25%|█████████████████▌ | 3/12 [03:36<12:47, 85.32s/it]

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7255230774876992

Classifiers: 33%|███████████████████████ | 4/12 [13:25<37:54, 284.30s/it]

RBF SVM - Best Hyperparameters: {'C': 20, 'gamma': 0.1}, F1 Score on Validation Set: 0.7199698230957782

Classifiers: 42%|████████████████████████████▊ | 5/12 [13:31<21:28, 184.03s/it]

Decision Tree - Best Hyperparameters: {'max\_depth': 10}, F1 Score on Validation Set: 0.5657358003489309

Classifiers: 50%|██████████████████████████████████▌ | 6/12 [14:04<13:15, 132.50s/it]

Random Forest - Best Hyperparameters: {'max\_depth': 10, 'n\_estimators': 50}, F1 Score on Validation Set: 0.3972787758279014

Classifiers: 58%|████████████████████████████████████████▎ | 7/12 [25:12<25:38, 307.75s/it]

AdaBoost - Best Hyperparameters: {'learning\_rate': 1, 'n\_estimators': 100}, F1 Score on Validation Set: 0.6308550667981022

Naive Bayes - Best Hyperparameters: {}, F1 Score on Validation Set: 0.6855179121590459

Classifiers: 75%|██████████████████████████████████████████████████▎ | 9/12 [1:17:54<25:58, 519.38s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (100,), 'max\_iter': 200}, F1 Score on Validation Set: 0.7302939190662537

Classifiers: 50%|███████████████████████████████████ | 1/2 [09:14<09:14, 554.13s/it]

Poly SVM - Best Hyperparameters: {'C': 20, 'degree': 2}, F1 Score on Validation Set: 0.7367438672827139

The following was done on only 25% of the test set:

Classifiers: 25%|█████████████████▊ | 1/4 [00:08<00:26, 8.91s/it]

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.6810620972321432

Classifiers: 50%|███████████████████████████████████▌ | 2/4 [00:21<00:21, 10.89s/it]

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.6716128036795944

Classifiers: 75%|█████████████████████████████████████████████████████▎ | 3/4 [00:58<00:22, 22.74s/it]

RBF SVM - Best Hyperparameters: {'C': 7, 'gamma': 0.1}, F1 Score on Validation Set: 0.6642466608844986

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 4/4 [07:42<00:00, 115.56s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (50,), 'max\_iter': 100}, F1 Score on Validation Set: 0.6667628922369758

Execution Time: 462.22634625434875 seconds

The following was done on only 25% of the test set: With tfid

Classifiers: 25%|█████████████████▊ | 1/4 [00:08<00:25, 8.62s/it]

Logistic Regression - Best Hyperparameters: {'C': 20}, F1 Score on Validation Set: 0.673703709209626

Classifiers: 50%|███████████████████████████████████▌ | 2/4 [00:20<00:21, 10.81s/it]

Linear SVM - Best Hyperparameters: {'C': 100}, F1 Score on Validation Set: 0.6635775272039536

Classifiers: 75%|█████████████████████████████████████████████████████▎ | 3/4 [00:57<00:22, 22.38s/it]

RBF SVM - Best Hyperparameters: {'C': 20, 'gamma': 0.1}, F1 Score on Validation Set: 0.6652804994031088

C:\Users\Andries\AppData\Roaming\Python\Python310\site-packages\sklearn\neural\_network\\_multilayer\_perceptron.py:691: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.

warnings.warn(

C:\Users\Andries\AppData\Roaming\Python\Python310\site-packages\sklearn\neural\_network\\_multilayer\_perceptron.py:691: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.

warnings.warn(

C:\Users\Andries\AppData\Roaming\Python\Python310\site-packages\sklearn\neural\_network\\_multilayer\_perceptron.py:691: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (100) reached and the optimization hasn't converged yet.

warnings.warn(

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 4/4 [09:08<00:00, 137.01s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (50,), 'max\_iter': 200}, F1 Score on Validation Set: 0.6701929185216606

Execution Time: 548.0557217597961 seconds

Classifiers: 25%|█████████████████▊ | 1/4 [00:53<02:40, 53.63s/it]

Logistic Regression - Best Hyperparameters: {'C': 7}, F1 Score on Validation Set: 0.7505561488029496

Classifiers: 50%|███████████████████████████████████ | 2/4 [04:09<04:34, 137.14s/it]

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7235445720536947

Classifiers: 75%|████████████████████████████████████████████████████▌ | 3/4 [17:01<07:07, 427.35s/it]

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.1}, F1 Score on Validation Set: 0.7235501833206003

Classifiers: 100%|███████████████████████████████████████████████████████████████████| 4/4 [1:07:30<00:00, 1012.73s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (50,), 'max\_iter': 200}, F1 Score on Validation Set: 0.7439949150385011

Execution Time: 4050.908079624176 seconds

Classifiers: 25%|█████████████████▊ | 1/4 [00:45<02:17, 45.83s/it]

Logistic Regression - Best Hyperparameters: {'C': 3}, F1 Score on Validation Set: 0.7503402731438381

Classifiers: 50%|███████████████████████████████████ | 2/4 [04:45<05:20, 160.12s/it]

Linear SVM - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7235445720536947

Classifiers: 75%|████████████████████████████████████████████████████▌ | 3/4 [25:36<10:57, 657.88s/it]

RBF SVM - Best Hyperparameters: {'C': 12, 'gamma': 0.05}, F1 Score on Validation Set: 0.7487309556143642

Classifiers: 100%|███████████████████████████████████████████████████████████████████| 4/4 [1:38:36<00:00, 1479.19s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (125,), 'max\_iter': 225}, F1 Score on Validation Set: 0.7421523692051492

Execution Time: 5916.778694629669 seconds

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [00:45<00:00, 45.95s/it]

Logistic Regression - Best Hyperparameters: {'C': 3}, F1 Score on Validation Set: 0.7503402731438381

Execution Time: 45.95137286186218 seconds – using count vector

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [00:30<00:00, 30.11s/it]

Logistic Regression - Best Hyperparameters: {'C': 8}, F1 Score on Validation Set: 0.7434657681979199

Execution Time: 30.113798141479492 seconds using tfid and stopwords

Classifiers: 33%|███████████████████████▋ | 1/3 [01:11<02:23, 71.61s/it]

Logistic Regression - Best Hyperparameters: {'C': 8}, F1 Score on Validation Set: 0.7324487851088335

Classifiers: 67%|██████████████████████████████████████████████▋ | 2/3 [17:10<09:53, 593.38s/it]

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.05}, F1 Score on Validation Set: 0.7414400537018884

Classifiers: 100%|███████████████████████████████████████████████████████████████████| 3/3 [1:19:23<00:00, 1587.93s/it]

Neural Network - Best Hyperparameters: {'hidden\_layer\_sizes': (100,), 'max\_iter': 200}, F1 Score on Validation Set: 0.7163085738676297

Execution Time: 4763.808254003525 seconds using count and stopwords

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [01:13<01:13, 73.60s/it]

Logistic Regression - Best Hyperparameters: {'C': 3}, F1 Score on Validation Set: 0.7407028868770379

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [21:54<00:00, 657.03s/it]

RBF SVM - Best Hyperparameters: {'C': 8, 'gamma': 0.05}, F1 Score on Validation Set: 0.7442248514998773

Execution Time: 1314.05664396286 seconds

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

stem,

# lemmatize,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

# cleaning\_numbers,

# cleaning\_email\_addresses,

# correct\_spelling, count

def preprocess\_text(text):

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

# stem,

lemmatize,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

# cleaning\_numbers,

# cleaning\_email\_addresses,

# correct\_spelling,

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [00:58<00:58, 58.13s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7484853223654423

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [06:46<00:00, 203.06s/it]

RBF SVM - Best Hyperparameters: {'C': 8, 'gamma': 0.05}, F1 Score on Validation Set: 0.7492939534186378

Execution Time: 406.11467957496643 seconds

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [00:56<00:56, 56.13s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7455545610538852

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [06:42<00:00, 201.48s/it]

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.05}, F1 Score on Validation Set: 0.7495912117858395

Execution Time: 402.9569661617279 seconds

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

cleaning\_numbers,

# cleaning\_email\_addresses,

# correct\_spelling,

lemmatize,

Count

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [00:50<00:50, 50.89s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7386323396747873

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [06:14<00:00, 187.10s/it]

RBF SVM - Best Hyperparameters: {'C': 8, 'gamma': 0.05}, F1 Score on Validation Set: 0.7487765884685578

Execution Time: 374.2024748325348 seconds

cleaning\_functions = [

convert\_to\_lowercase,

cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

tokenize,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

# cleaning\_URLs,

cleaning\_numbers,

# correct\_spelling,

lemmatize, count

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [00:55<00:55, 55.76s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.746157885274713

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [06:41<00:00, 200.80s/it]

RBF SVM - Best Hyperparameters: {'C': 10, 'gamma': 0.05}, F1 Score on Validation Set: 0.7510232494755161

Execution Time: 401.6116623878479 seconds

cleaning\_functions = [

convert\_to\_lowercase,

# cleaning\_email\_addresses,

cleaning\_repeating\_char,

cleaning\_ascii,

cleaning\_punctuations,

cleaning\_URLs,

tokenize,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize, counts

Classifiers: 50%|███████████████████████████████████▌ | 1/2 [01:04<01:04, 64.98s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7656491084787354

Classifiers: 100%|██████████████████████████████████████████████████████████████████████| 2/2 [07:17<00:00, 218.60s/it]

RBF SVM - Best Hyperparameters: {'C': 8, 'gamma': 0.05}, F1 Score on Validation Set: 0.7449584851181018

Execution Time: 437.1976957321167 seconds

convert\_to\_lowercase,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

# cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_repeating\_char,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Count

lassifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [01:10<00:00, 70.31s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7619247214998346

Execution Time: 70.309885263443 seconds

cleaning\_functions = [

# convert\_to\_lowercase,

cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

# cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [01:09<00:00, 69.12s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7637748620364431

Execution Time: 69.12660503387451 seconds

# convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

cleaning\_ascii,

# cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [01:05<00:00, 65.18s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7468047470758004

Execution Time: 65.17971324920654 seconds

# convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

Lemmatize

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [00:58<00:00, 58.09s/it]

Logistic Regression - Best Hyperparameters: {'C': 0.5}, F1 Score on Validation Set: 0.7388860394102124

Execution Time: 58.092113971710205 seconds

# convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [00:56<00:00, 56.55s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.748020266364124

Execution Time: 56.55376648902893 seconds

# convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

# cleaning\_punctuations,

cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [01:08<00:00, 68.57s/it]

Logistic Regression - Best Hyperparameters: {'C': 1}, F1 Score on Validation Set: 0.7666624476216484

Execution Time: 68.56847929954529 seconds

convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

# cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

Classifiers: 100%|███████████████████████████████████████████████████████████████████████| 1/1 [00:49<00:00, 49.04s/it]

Logistic Regression - Best Hyperparameters: {'C': 8}, F1 Score on Validation Set: 0.7683624891430085

Execution Time: 49.043070793151855 seconds

convert\_to\_lowercase,

# cleaning\_repeating\_char,

# cleaning\_email\_addresses,

# cleaning\_repeating\_char,

# cleaning\_ascii,

# cleaning\_punctuations,

# cleaning\_URLs,

tokenize,

remove\_twitter\_handles,

# stem,

# remove\_stopwords,

# remove\_accent\_marks,

# cleaning\_non\_alphanumeric,

cleaning\_numbers,

# correct\_spelling,

lemmatize,

tfid