**Console outputs from test dataset Dummy\_Data\_After.csv**

Includes test print statements which are usually absent from the program.  
  
**Console output from bag of words only test – uses bagOfWordsOnlyTest():**  
  
No of records uploaded: 20

TEST PRINT. Vocabulary list (NB: stop words are included): ['aboard', 'abstain', 'annoyed', 'ask', 'away', 'book', 'class', 'clear', 'come', 'didn', 'die', 'disappeared', 'does', 'don', 'drinking', 'eat', 'faster', 'father', 'getting', 'guess', 'halt', 'helen', 'japan', 'let', 'll', 'maybe', 'minute', 'near', 'notebook', 'nut', 'pushes', 'quickly', 'religious', 'right', 'room', 'selfishness', 'shoot', 'sick', 'stay', 'step', 'stolen', 'study', 'time', 'tom', 'train', 'tries', 'try', 'view', 'want', 'watch', 'yesterday']

TEST PRINT. Vector shape (first figure is number of sentences, second is number of words in corpus): (20, 51)

Feature tested: N-gram only test

Classifier: Support Vector Machine

Total predictions: 4

TRUE POSITIVES: 1

FALSE POSITIVES: 0

TRUE NEGATIVES: 2

FALSE NEGATIVES: 1

Accuracy: 0.75

Precision: 1.00

Recall: 0.50

False positive rate: N/A

AUC: 0.25

Balanced accuracy: 0.75

TEST PRINT. List of predictions in order: [False False False True]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Feature tested: N-gram only test

Classifier: Logistic Regression

Total predictions: 4

TRUE POSITIVES: 0

FALSE POSITIVES: 0

TRUE NEGATIVES: 2

FALSE NEGATIVES: 2

Accuracy: 0.50

Precision: N/A

Recall: N/A

False positive rate: N/A

AUC: 0.75

Balanced accuracy: 0.50

TEST PRINT. List of predictions in order: [False False False False]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Feature tested: N-gram only test

Classifier: Multinomial Bayes

Total predictions: 4

TRUE POSITIVES: 1

FALSE POSITIVES: 0

TRUE NEGATIVES: 2

FALSE NEGATIVES: 1

Accuracy: 0.75

Precision: 1.00

Recall: 0.50

False positive rate: N/A

AUC: 0.75

Balanced accuracy: 0.75

TEST PRINT. List of predictions in order: [False False False True]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Feature tested: N-gram only test

Classifier: Random forest

Total predictions: 4

TRUE POSITIVES: 0

FALSE POSITIVES: 0

TRUE NEGATIVES: 2

FALSE NEGATIVES: 2

Accuracy: 0.50

Precision: N/A

Recall: N/A

False positive rate: N/A

AUC: 0.75

Balanced accuracy: 0.50

TEST PRINT. List of predictions in order: [False False False False]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Process finished with exit code 0

**Console output from test not employing n-grams - uses testFeaturesNoBagOfWords()**

No of records uploaded: 20

TEST PRINT. Features to test: ['Number of adjectives', 'Number of nouns']

TEST PRINT. List of all formality classifications in dataset, in sentence order: [False, True, False, False, True, False, False, True, False, True, True, True, False, False, False, True, True, True, False, False]

TEST PRINT. Feature data, in sentence order: [[1.0, 1.0], [0.0, 1.0], [0.0, 2.0], [1.0, 0.0], [0.0, 2.0], [0.0, 0.0], [0.0, 1.0], [0.0, 1.0], [2.0, 1.0], [0.0, 1.0], [0.0, 2.0], [1.0, 2.0], [1.0, 1.0], [0.0, 1.0], [0.0, 2.0], [0.0, 3.0], [0.0, 2.0], [0.0, 2.0], [1.0, 1.0], [0.0, 0.0]]

Feature tested: Number of adjectives and number of nouns

Classifier: Support Vector Machine

Total predictions: 4

TRUE POSITIVES: 2

FALSE POSITIVES: 1

TRUE NEGATIVES: 1

FALSE NEGATIVES: 0

Accuracy: 0.75

Precision: 0.67

Recall: 1.00

False positive rate: 0.50

AUC: 0.88

Balanced accuracy: 0.75

TEST PRINT. List of predictions in order: [ True False True True]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Process finished with exit code 0

**Console output test covering n-grams and non n-gram features – uses testFeaturesIncBagOfWords()**

No of records uploaded: 20

TEST PRINT. Non n-gram features to test: ['Number of verbs', 'Number of adjectives', 'Number of prepositions']

TEST PRINT. Vocabulary list (note stop words excluded): ['aboard', 'abstain', 'annoyed', 'ask', 'away', 'book', 'class', 'clear', 'come', 'didn', 'die', 'disappeared', 'does', 'don', 'drinking', 'eat', 'faster', 'father', 'getting', 'guess', 'halt', 'helen', 'japan', 'let', 'll', 'maybe', 'minute', 'near', 'notebook', 'nut', 'pushes', 'quickly', 'religious', 'right', 'room', 'selfishness', 'shoot', 'sick', 'stay', 'step', 'stolen', 'study', 'time', 'tom', 'train', 'tries', 'try', 'view', 'want', 'watch', 'yesterday']

TEST PRINT. Vector shape (first figure is number of sentences, second is number of words in corpus): (20, 51)

TEST PRINT. List of all formality classifications in dataset, in order: [False, True, False, False, True, False, False, True, False, True, True, True, False, False, False, True, True, True, False, False]

TEST PRINT. List of feature data for the first sentence. Last three entries relate to the non n-gram features:

[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.

0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.

0. 0. 0. 2. 1. 0.]

Feature tested: Number of verbs, adjectives and preposition plus binary unigram test with stop words excluded

Classifier: Support Vector Machine

Total predictions: 4

TRUE POSITIVES: 1

FALSE POSITIVES: 1

TRUE NEGATIVES: 1

FALSE NEGATIVES: 1

Accuracy: 0.50

Precision: 0.50

Recall: 0.50

False positive rate: 0.50

AUC: 0.50

Balanced accuracy: 0.50

TEST PRINT. List of predictions in order: [ True False True False]

TEST PRINT. List of actual whether sentence is formal in order: [False False True True]

Process finished with exit code 0