Unbalanced Dataset

Unigram

MultinomialNB

Accuracy: 66.3783%

precision recall f1-score support

positive 0.94 0.12 0.21 718

negative 0.65 1.00 0.79 2742

neutral 0.82 0.10 0.18 932

micro avg 0.66 0.66 0.66 4392

macro avg 0.80 0.41 0.39 4392

weighted avg 0.74 0.66 0.57 4392

==============================

SVC

Accuracy: 62.4175%

precision recall f1-score support

positive 0.00 0.00 0.00 718

negative 0.62 1.00 0.77 2742

neutral 0.00 0.00 0.00 932

micro avg 0.62 0.62 0.62 4392

macro avg 0.21 0.33 0.26 4392

weighted avg 0.39 0.62 0.48 4392

==============================

DecisionTreeClassifier

Accuracy: 67.1523%

precision recall f1-score support

positive 0.56 0.54 0.55 718

negative 0.78 0.79 0.78 2742

neutral 0.43 0.42 0.42 932

micro avg 0.67 0.67 0.67 4392

macro avg 0.59 0.58 0.59 4392

weighted avg 0.67 0.67 0.67 4392

==============================

RandomForestClassifier

Accuracy: 72.4107%

precision recall f1-score support

positive 0.74 0.37 0.49 718

negative 0.74 0.94 0.83 2742

neutral 0.62 0.35 0.45 932

micro avg 0.72 0.72 0.72 4392

macro avg 0.70 0.55 0.59 4392

weighted avg 0.71 0.72 0.69 4392

==============================

LogisticRegression

Accuracy: 78.6478%

precision recall f1-score support

positive 0.83 0.56 0.67 718

negative 0.79 0.96 0.87 2742

neutral 0.72 0.44 0.55 932

micro avg 0.79 0.79 0.79 4392

macro avg 0.78 0.66 0.70 4392

weighted avg 0.78 0.79 0.77 4392

==============================



Balanced Dataset:

neutral 2363

positive 2363

negative 2363



**Results:**

**NGRAM =1**

MultinomialNB

Accuracy: 71.9323%

precision recall f1-score support

positive 0.81 0.71 0.76 712

negative 0.65 0.88 0.75 699

neutral 0.73 0.57 0.64 715

micro avg 0.72 0.72 0.72 2126

macro avg 0.73 0.72 0.72 2126

weighted avg 0.73 0.72 0.72 2126

==============================

SVC

Accuracy: 32.8632%

precision recall f1-score support

positive 0.00 0.00 0.00 712

negative 0.33 1.00 0.49 699

neutral 0.00 0.00 0.00 715

micro avg 0.33 0.33 0.33 2126

macro avg 0.11 0.33 0.16 2126

weighted avg 0.11 0.33 0.16 2126

==============================

DecisionTreeClassifier

Accuracy: 60.3667%

precision recall f1-score support

positive 0.67 0.60 0.64 712

negative 0.59 0.60 0.60 699

neutral 0.56 0.61 0.58 715

micro avg 0.60 0.60 0.60 2126

macro avg 0.61 0.60 0.60 2126

weighted avg 0.61 0.60 0.60 2126

==============================

RandomForestClassifier

Accuracy: 64.4100%

precision recall f1-score support

positive 0.75 0.58 0.66 712

negative 0.64 0.71 0.67 699

neutral 0.58 0.64 0.61 715

micro avg 0.64 0.64 0.64 2126

macro avg 0.65 0.64 0.64 2126

weighted avg 0.65 0.64 0.64 2126

==============================

LogisticRegression

Accuracy: 74.4241%

precision recall f1-score support

positive 0.83 0.73 0.77 712

negative 0.73 0.78 0.76 699

neutral 0.69 0.73 0.71 715

micro avg 0.74 0.74 0.74 2126

macro avg 0.75 0.74 0.75 2126

weighted avg 0.75 0.74 0.75 2126

==============================

Ngram =2

* **MultinomialNB :** Accuracy: 67.7950%
* **SVC:** Accuracy: 33.2393%
* **DecisionTreeClassifier:** Accuracy: 58.8152%
* **RandomForestClassifier:** Accuracy: 58.7212%
* **LogisticRegression:** Accuracy: 65.5853%

Ngram =1,2

MultinomialNB

Accuracy: 70.0517%

==============================

SVC

Accuracy: 32.3460%

==============================

DecisionTreeClassifier

Accuracy: 57.6869%

==============================

RandomForestClassifier

Accuracy: 62.1063%

==============================

LogisticRegression

Accuracy: 68.7353%

==============================

* **SelectKBest:**

**Ngram=1, k=1000**

MultinomialNB

Accuracy: 71.6032%

precision recall f1-score support

positive 0.80 0.72 0.75 732

negative 0.68 0.78 0.73 690

neutral 0.68 0.65 0.66 705

micro avg 0.72 0.72 0.72 2127

macro avg 0.72 0.72 0.72 2127

weighted avg 0.72 0.72 0.72 2127

==============================

SVC

Accuracy: 32.4401%

precision recall f1-score support

positive 0.00 0.00 0.00 732

negative 0.32 1.00 0.49 690

neutral 0.00 0.00 0.00 705

micro avg 0.32 0.32 0.32 2127

macro avg 0.11 0.33 0.16 2127

weighted avg 0.11 0.32 0.16 2127

DecisionTreeClassifier

Accuracy: 57.4518%

precision recall f1-score support

positive 0.63 0.61 0.62 732

negative 0.57 0.54 0.55 690

neutral 0.53 0.57 0.55 705

micro avg 0.57 0.57 0.57 2127

macro avg 0.58 0.57 0.57 2127

weighted avg 0.58 0.57 0.57 2127

==============================

RandomForestClassifier

Accuracy: 64.7391%

precision recall f1-score support

positive 0.73 0.61 0.67 732

negative 0.63 0.68 0.66 690

neutral 0.59 0.65 0.62 705

micro avg 0.65 0.65 0.65 2127

macro avg 0.65 0.65 0.65 2127

weighted avg 0.65 0.65 0.65 2127

==============================

LogisticRegression

Accuracy: 72.9666%

precision recall f1-score support

positive 0.82 0.71 0.76 732

negative 0.71 0.78 0.74 690

neutral 0.67 0.71 0.69 705

micro avg 0.73 0.73 0.73 2127

macro avg 0.73 0.73 0.73 2127

weighted avg 0.74 0.73 0.73 2127

==============================

****

**ngram= 2**

MultinomialNB

Accuracy: 64.1749%

==============================

SVC

Accuracy: 31.4998%

==============================

DecisionTreeClassifier

Accuracy: 55.7123%

==============================

RandomForestClassifier

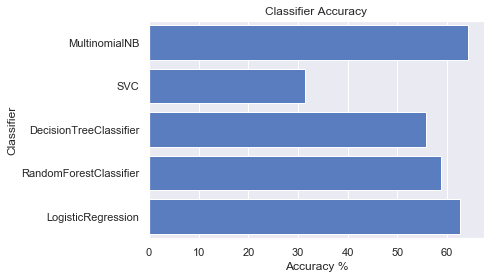
Accuracy: 58.8622%

==============================

LogisticRegression

Accuracy: 62.5294%

==============================



SelectFromModel using LinearSVC as featureSelector

==========/n

Classifier: SVC

['neutral' 'neutral' 'neutral' ... 'neutral' 'neutral' 'neutral']

Accuracy: 32.7692%

==========/n

Classifier: LinearSVC

['neutral' 'neutral' 'neutral' ... 'neutral' 'negative' 'neutral']

Accuracy: 64.6921%

==========/n

Classifier: MultinomialNB

['neutral' 'neutral' 'neutral' ... 'neutral' 'negative' 'neutral']

Accuracy: 64.5980%

==========/n

Classifier: DecisionTreeClassifier

['neutral' 'neutral' 'neutral' ... 'neutral' 'negative' 'neutral']

Accuracy: 57.0757%

==========/n

Classifier: RandomForestClassifier

['neutral' 'neutral' 'neutral' ... 'neutral' 'positive' 'neutral']

Accuracy: 58.9093%

==========/n

Classifier: LogisticRegression

['neutral' 'neutral' 'negative' ... 'negative' 'negative' 'neutral']

Accuracy: 60.9779%

SelectFromModel using LinearSVC as featureSelector