**Results – Initial Model Evaluation**

Initial Model Results (Naïve-Bayes and Logistic Regression)

**Model: Naive Bayes (TF-IDF)**

Accuracy: 0.671960727692752

Classification Report:

precision recall f1-score support

1.0 0.26 0.36 0.30 86

2.0 0.12 0.08 0.10 84

3.0 0.24 0.26 0.25 302

4.0 0.43 0.32 0.37 1731

5.0 0.78 0.84 0.81 4723

accuracy 0.67 6926

macro avg 0.37 0.37 0.37 6926

weighted avg 0.65 0.67 0.66 6926

Confusion Matrix:

[[ 31 4 18 8 25]

[ 11 7 30 14 22]

[ 16 12 79 123 72]

[ 20 14 125 559 1013]

[ 40 19 80 606 3978]]

**Model: Naive Bayes (Count)**

Accuracy: 0.671960727692752

Classification Report:

precision recall f1-score support

1.0 0.26 0.36 0.30 86

2.0 0.12 0.08 0.10 84

3.0 0.24 0.26 0.25 302

4.0 0.43 0.32 0.37 1731

5.0 0.78 0.84 0.81 4723

accuracy 0.67 6926

macro avg 0.37 0.37 0.37 6926

weighted avg 0.65 0.67 0.66 6926

Confusion Matrix:

[[ 31 4 18 8 25]

[ 11 7 30 14 22]

[ 16 12 79 123 72]

[ 20 14 125 559 1013]

[ 40 19 80 606 3978]]

**Model: Logistic Regression (TF-IDF)**

Accuracy: 0.6959283857926654

Classification Report:

precision recall f1-score support

1.0 0.33 0.21 0.26 86

2.0 0.20 0.11 0.14 84

3.0 0.31 0.17 0.22 302

4.0 0.46 0.23 0.30 1731

5.0 0.75 0.92 0.83 4723

accuracy 0.70 6926

macro avg 0.41 0.33 0.35 6926

weighted avg 0.65 0.70 0.65 6926

Confusion Matrix:

[[ 18 6 11 15 36]

[ 7 9 18 18 32]

[ 9 12 52 102 127]

[ 8 9 65 396 1253]

[ 13 10 20 335 4345]]

**Model: Logistic Regression (Count)**

Accuracy: 0.6959283857926654

Classification Report:

precision recall f1-score support

1.0 0.33 0.21 0.26 86

2.0 0.20 0.11 0.14 84

3.0 0.31 0.17 0.22 302

4.0 0.46 0.23 0.30 1731

5.0 0.75 0.92 0.83 4723

accuracy 0.70 6926

macro avg 0.41 0.33 0.35 6926

weighted avg 0.65 0.70 0.65 6926

Confusion Matrix:

[[ 18 6 11 15 36]

[ 7 9 18 18 32]

[ 9 12 52 102 127]

[ 8 9 65 396 1253]

[ 13 10 20 335 4345]]

**Results After “Optimization”**

**Optimized Model: Naive Bayes (TF-IDF)**

Accuracy: 0.11897198960438926

Classification Report:

precision recall f1-score support

1.0 0.00 0.14 0.01 86

2.0 0.03 0.54 0.06 84

3.0 0.09 0.04 0.05 302

4.0 0.32 0.04 0.08 1731

5.0 0.63 0.14 0.23 4723

accuracy 0.12 6926

macro avg 0.22 0.18 0.09 6926

weighted avg 0.51 0.12 0.18 6926

Confusion Matrix:

[[ 12 42 17 13 2]

[ 18 45 6 12 3]

[ 122 81 12 44 43]

[ 922 354 29 74 352]

[2981 901 75 85 681]]

**Optimized Model: Naive Bayes (Count)**

Accuracy: 0.11897198960438926

Classification Report:

precision recall f1-score support

1.0 0.00 0.14 0.01 86

2.0 0.03 0.54 0.06 84

3.0 0.09 0.04 0.05 302

4.0 0.32 0.04 0.08 1731

5.0 0.63 0.14 0.23 4723

accuracy 0.12 6926

macro avg 0.22 0.18 0.09 6926

weighted avg 0.51 0.12 0.18 6926

Confusion Matrix:

[[ 12 42 17 13 2]

[ 18 45 6 12 3]

[ 122 81 12 44 43]

[ 922 354 29 74 352]

[2981 901 75 85 681]]

**Optimized Model: Logistic Regression (TF-IDF)**

Accuracy: 0.0619405140051978

Classification Report:

precision recall f1-score support

1.0 0.01 0.30 0.01 86

2.0 0.02 0.37 0.03 84

3.0 0.08 0.02 0.03 302

4.0 0.28 0.02 0.04 1731

5.0 0.66 0.07 0.13 4723

accuracy 0.06 6926

macro avg 0.21 0.16 0.05 6926

weighted avg 0.53 0.06 0.10 6926

Confusion Matrix:

[[ 26 48 1 9 2]

[ 39 31 4 5 5]

[ 177 81 6 21 17]

[1038 498 15 36 144]

[3055 1232 48 58 330]]

**Optimized Model: Logistic Regression (Count)**

Accuracy: 0.0619405140051978

Classification Report:

precision recall f1-score support

1.0 0.01 0.30 0.01 86

2.0 0.02 0.37 0.03 84

3.0 0.08 0.02 0.03 302

4.0 0.28 0.02 0.04 1731

5.0 0.66 0.07 0.13 4723

accuracy 0.06 6926

macro avg 0.21 0.16 0.05 6926

weighted avg 0.53 0.06 0.10 6926

Confusion Matrix:

[[ 26 48 1 9 2]

[ 39 31 4 5 5]

[ 177 81 6 21 17]

[1038 498 15 36 144]

[3055 1232 48 58 330]]

**Optimized Model: Random Forest (TF-IDF)**

Accuracy: 0.15073635576090094

Classification Report:

precision recall f1-score support

1.0 0.01 0.62 0.02 86

2.0 0.02 0.06 0.03 84

3.0 0.07 0.02 0.03 302

4.0 0.28 0.03 0.06 1731

5.0 0.64 0.20 0.30 4723

accuracy 0.15 6926

macro avg 0.20 0.18 0.09 6926

weighted avg 0.51 0.15 0.22 6926

Confusion Matrix:

[[ 53 6 5 10 12]

[ 53 5 2 11 13]

[ 189 13 6 21 73]

[1154 73 22 59 423]

[3460 174 57 111 921]]

**Optimized Model: Random Forest (Count)**

Accuracy: 0.15521224371931852

Classification Report:

precision recall f1-score support

1.0 0.01 0.64 0.02 86

2.0 0.02 0.07 0.03 84

3.0 0.06 0.02 0.03 302

4.0 0.30 0.04 0.06 1731

5.0 0.64 0.20 0.31 4723

accuracy 0.16 6926

macro avg 0.21 0.19 0.09 6926

weighted avg 0.52 0.16 0.23 6926

Confusion Matrix:

[[ 55 6 4 8 13]

[ 51 6 3 10 14]

[ 181 16 6 18 81]

[1158 75 21 63 414]

[3433 173 59 113 945]]