CSE4062 Spring Group Number: 3

Delivery #2

ProjectTitle: ISSUECATEGORY DETERMINER

CONTRIBUTORS:

Şeyma Kaya - Electrical and Electronics Engineering 150719013 – skaya19@marun.edu.tr

Ataberk Gedik - Electrical and Electronics Engineering - 150719058 - ataberkgedik@marun.edu.tr

Muhammet Eren Atala - Computer Engineering - 150119904 - muhammeterenatala@gmail.com

Emrullah Sevmiş - Computer Engineering - 150119016 – emrullahsevmis@marun.edu.tr

Muhammet Eren Ünlü - Computer Engineering - 150117027 - muhammet.eren@marun.edu.tr

Can Bayrakdar -Industrial Engineering - 150319569 - canbayrakdar@gmail.com

1) Our Class attribute is nominal, so we did a Chi Square analysis to determine the relationship between the nominal columns.

```
Relationship between Class Attribute(Nominal) with Nominal values

Analyzing relationship between ISSUE_CATEGORY and COMPONENT
Chi-Square Test Results:
Chi-Square statistic: 14866.377762623426
P-value: 0.0
There is a significant relationship between ISSUE_CATEGORY and COMPONENT.

Analyzing relationship between ISSUE_CATEGORY and IMPACT
Chi-Square Test Results:
Chi-Square statistic: 6612.004533185184
P-value: 0.0
There is a significant relationship between ISSUE_CATEGORY and IMPACT.

Analyzing relationship between ISSUE_CATEGORY and ASSIGNEE
Chi-Square Test Results:
Chi-Square statistic: 159171.21522854612
P-value: 0.0
There is a significant relationship between ISSUE_CATEGORY and ASSIGNEE.

Analyzing relationship between ISSUE_CATEGORY and ISSUE_TYPE
Chi-Square Test Results:
Chi-Square statistic: 5415.837875120984
P-value: 0.0
```

2) The columns we use and the first 10 rows of those columns.(Before Preprocessing)

```
Original data
Index(['ID', 'PRIORITY', 'COMPONENT', 'URGENCY', 'IMPACT', 'ISSUE_CATEGORY',
       'WORKLOG_ASSIGNEE1', 'WORKLOG_ASSIGNEE2', 'WORKLOG_ASSIGNEE3',
       'WORKLOG_ASSIGNEE4', 'WORKLOG_ASSIGNEE5', 'POZISYON', 'IS_BILGISI',
       'CREATED', 'RESOLUTIONDATE', 'Complition Time', 'ASSIGNEE', 'ASSIGNEE1',
       'ASSIGNEE2', 'ASSIGNEE3', 'ASSIGNEE4', 'ASSIGNEE5', 'Total_Assignee',
       'Total_Worklog_Assginee', 'Total_Log_Hours_Assignee', 'COMMENTOR_COUNT',
       'COMMENT_COUNT', 'ISSUE_TYPE'],
      dtype='object')

First 10 rows of the transformed dataset:
   ID  PRIORITY   COMPONENT URGENCY                          IMPACT ISSUE_CATEGORY  ... Total_Assignee Total_Worklog_Assginee Total_Log_Hours_Assignee COMMENTOR_COUNT COMMENT_COUNT      ISSUE_TYPE
0   1     Minor  Functional     Low  Some Processes Stop on Production            PO  ...              2                     0                      0.0                 0             0        Incident
1   2     Minor  Functional     Low  Some Processes Stop on Production            AP  ...              1                     0                      0.0                 0             0        Incident
2   3     Major  Functional  Medium  Some Processes Stop on Production        Custom  ...              2                     0                      0.0                 0             0        Incident
3   4     Major  Functional  Medium  Some Processes Stop on Production        Custom  ...              1                     0                      0.0                 1             1        Incident
4   5     Minor  Functional     Low  Some Processes Stop on Production            AP  ...              3                     0                      0.0                 1             1        Incident
5   6     Minor  Functional     Low  Some Processes Stop on Production        Custom  ...              1                     0                      0.0                 1             1        Incident
6   7     Minor  Functional     NaN                               NaN            AP  ...              1                     0                      0.0                 1             1  Change Request
7   8     Minor  Functional     Low  Some Processes Stop on Production            AP  ...              1                     0                      0.0                 2             2        Incident
8   9     Minor  Functional     NaN                               NaN            GL  ...              3                     0                      0.0                 2             3  Change Request
9  10  Critical  Functional    High  Some Processes Stop on Production            GL  ...              5                     0                      0.0                 4            10        Incident

[10 rows x 28 columns]

Number of rows and columns in the original data: (40008, 28)
```

3) The columns we use and the first 10 rows of those columns.(After Preprocessing)

```
Preprocessed data
Index(['PRIORITY', 'URGENCY', 'COMPONENT', 'IMPACT', 'ISSUE_CATEGORY',
       'ASSIGNEE', 'Total_Assignee', 'Total_Worklog_Assginee',
       'Total_Log_Hours_Assignee', 'COMMENTOR_COUNT', 'COMMENT_COUNT',
       'ISSUE_TYPE'],
      dtype='object')

First 10 rows of the transformed dataset:
   PRIORITY   URGENCY  COMPONENT  IMPACT  ISSUE_CATEGORY  ASSIGNEE  Total_Assignee  Total_Worklog_Assginee  Total_Log_Hours_Assignee  COMMENTOR_COUNT  COMMENT_COUNT  ISSUE_TYPE
0       1.0  0.000000          0       3              13         0               2                       0                       0.0                0              0           1
1       1.0  0.000000          0       3               0        15               1                       0                       0.0                0              0           1
2       2.0  1.000000          0       3               2         0               2                       0                       0.0                0              0           1
3       2.0  1.000000          0       3               2         0               1                       0                       0.0                1              1           1
4       1.0  0.000000          0       3               0         0               3                       0                       0.0                1              1           1
5       1.0  0.000000          0       3               2         3               1                       0                       0.0                1              1           1
6       1.0  0.571652          0       4               0         0               1                       0                       0.0                1              1           0
7       1.0  0.000000          0       3               0         7               1                       0                       0.0                2              2           1
8       1.0  0.571652          0       4               6         3               3                       0                       0.0                2              3           0
9       3.0  2.000000          0       3               6         0               5                       0                       0.0                4             10           1

Number of rows and columns in the preprocessed data: (23304, 12)
```

4) DECISION TREE CLASSIFIER

a)Decision Tree Using Information Gain Results

```
Test set results:
Accuracy: 0.42005148741418763
Recall: 0.42005148741418763
Precision: 0.4063618125913167
F1 Score: 0.4115812054764304
Macro F1 Score: 0.2733901904930238
Micro F1 Score: 0.42005148741418763

Test Set Confusion Matrix:
[[148  33  80   3  12  16  92  66  27   2   5   7   3  93   5  14]
 [ 36  34  39   1   7   6  35  17  16   2   1   3   0  38   4   3]
 [ 97  30 221   7  15  18  96  90  35   4  10  11  10 154  18  29]
 [  2   0   4   6   0   0   5   9   5   0   3   3   0   9   3   8]
 [  7   4  10   1  27   3  11  13  14   1   1   0   1  32   5   3]
 [ 10   4  13   0   5   4  18  13   6   0   0   3   2  13   3   4]
 [ 95  22  99   6  10   6 139  87  33   0   3   9   4 114  11  23]
 [ 40  11  48   5   3  11  59 615  56   7   3   6   3 131  21  22]
 [ 21   6  40   2  19   6  49  67  85   2   1   4   7 135   9   8]
 [  1   1   5   2   1   0   2   4   2  10   1   4   0  12   2  10]
 [  4   3   6   0   0   2   6   8   5   1   1   2   0  16   3   2]
 [  8   4  15   0   2   3   8  12  10   3   2  26   1   8   5   2]
 [  3   0   6   0   0   0   3   9   3   0   0   2  23   5   2   0]
 [ 91  14 102   2  17   8  95 125  84   2  10  13   4 567  14  35]
 [ 10   4  15   1   0   6  12  20   5   1   1   4   0  18 992   6]
 [ 22   9  29   7   5   1  33  54  11   4   3   9   4  51   8  39]]

Error Rate (Training): 0.6000490436488475
Error Rate (Test): 0.5799485125858124
```


```
Test set results:
Accuracy: 0.42005148741418763
Recall: 0.42005148741418763
Precision: 0.4063618125913167
F1 Score: 0.4115812054764304
Macro F1 Score: 0.2733901904930238
Micro F1 Score: 0.42005148741418763

Test Set Confusion Matrix:
[[148  33  80   3  12  16  92  66  27   2   5   7   3  93   5  14]
 [ 36  34  39   1   7   6  35  17  16   2   1   3   0  38   4   3]
 [ 97  30 221   7  15  18  96  90  35   4  10  11  10 154  18  29]
 [  2   0   4   6   0   0   5   9   5   0   3   3   0   9   3   8]
 [  7   4  10   1  27   3  11  13  14   1   1   0   1  32   5   3]
 [ 10   4  13   0   5   4  18  13   6   0   0   3   2  13   3   4]
 [ 95  22  99   6  10   6 139  87  33   0   3   9   4 114  11  23]
 [ 40  11  48   5   3  11  59 615  56   7   3   6   3 131  21  22]
 [ 21   6  40   2  19   6  49  67  85   2   1   4   7 135   9   8]
 [  1   1   5   2   1   0   2   4   2  10   1   4   0  12   2  10]
 [  4   3   6   0   0   2   6   8   5   1   1   2   0  16   3   2]
 [  8   4  15   0   2   3   8  12  10   3   2  26   1   8   5   2]
 [  3   0   6   0   0   0   3   9   3   0   0   2  23   5   2   0]
 [ 91  14 102   2  17   8  95 125  84   2  10  13   4 567  14  35]
 [ 10   4  15   1   0   6  12  20   5   1   1   4   0  18 992   6]
 [ 22   9  29   7   5   1  33  54  11   4   3   9   4  51   8  39]]

Error Rate (Training): 0.6000490436488475
Error Rate (Test): 0.5799485125858124
```

b) Decision Tree Using Gini Index Results

```
Decision Tree using Gini Index
------------------------------
Training set results:
Accuracy: 0.3999509563511525
Recall: 0.3999509563511525
Precision: 0.3841883166913004
F1 Score: 0.3909433681792296
Macro F1 Score: 0.254723413484193
Micro F1 Score: 0.3999509563511525

Training Set Confusion Matrix:
[[ 316   69  236    6   33   20  206  154   55    8    5   24    6  222
    32   46]
 [ 104   68  108    4    6    5   82   57   22    2    5    9    1   76
     9   13]
 [ 212   77  476   12   32   27  246  180   89    8   12   32   17  328
    58   56]
 [  12    0   13    4    5    6    8   21   15    0    1    6    1   16
     7   11]
 [  35    7   29    6   47    3   24   33   35    1    2    4    2   56
     3   11]
 [  34    5   45    3    7   34   43   29   15    3    2    6    0   39
    12    7]
 [ 216   75  249    7   28   37  291  192   89    8    9   20    6  281
    22   48]
 [  98   31  139   12   21   20  139 1294  108    9   11   24    9  330
    59   71]
 [  69   35   84   12   34   18  102  155  187    8   12   21    5  313
    14   34]
 [  11    5   10    0    2    2   13   17    6    7    2    2    1   17
     8   15]
 [  11    4   15    1    3    2   11   25   16    0   17    5    5   48
     7    3]
 [  27    8   34    5    3    4   29   45   18    7    3   71    3   32
    16   14]
 [   4    1   25    1    4    1    8   11   12    1    2    2   42   16
     4    5]
 [ 177   51  238   10   39   27  230  328  225    9   21   23    8 1385
    33   66]
 [  27   10   39    2    1    8   16   53   10    5    3   12    2   26
  2200   11]
 [  58   12   67    7    7   11   59  122   42   13    4   18    8   97
    23   85]]
```

```
Test set results:
Accuracy: 0.42005148741418763
Recall: 0.42005148741418763
Precision: 0.4063618125913167
F1 Score: 0.4115812054764304
Macro F1 Score: 0.2733901904930238
Micro F1 Score: 0.42005148741418763

Test Set Confusion Matrix:
[[148  33  80   3  12  16  92  66  27   2   5   7   3  93   5  14]
 [ 36  34  39   1   7   6  35  17  16   2   1   3   0  38   4   3]
 [ 97  30 221   7  15  18  96  90  35   4  10  11  10 154  18  29]
 [  2   0   4   6   0   0   5   9   5   0   3   3   0   9   3   8]
 [  7   4  10   1  27   3  11  13  14   1   1   0   1  32   5   3]
 [ 10   4  13   0   5   4  18  13   6   0   0   3   2  13   3   4]
 [ 95  22  99   6  10   6 139  87  33   0   3   9   4 114  11  23]
 [ 40  11  48   5   3  11  59 615  56   7   3   6   3 131  21  22]
 [ 21   6  40   2  19   6  49  67  85   2   1   4   7 135   9   8]
 [  1   1   5   2   1   0   2   4   2  10   1   4   0  12   2  10]
 [  4   3   6   0   0   2   6   8   5   1   1   2   0  16   3   2]
 [  8   4  15   0   2   3   8  12  10   3   2  26   1   8   5   2]
 [  3   0   6   0   0   0   3   9   3   0   0   2  23   5   2   0]
 [ 91  14 102   2  17   8  95 125  84   2  10  13   4 567  14  35]
 [ 10   4  15   1   0   6  12  20   5   1   1   4   0  18 992   6]
 [ 22   9  29   7   5   1  33  54  11   4   3   9   4  51   8  39]]

Error Rate (Training): 0.6000490436488475
Error Rate (Test): 0.5799485125858124
```

c) Decision Tree Using Gradient Boosting Results

```
Decision Tree using Gradient Boosting
-----------------
Training set results:
Accuracy: 0.4484428641490927
Recall: 0.4484428641490927
Precision: 0.4186022057320158
F1 Score: 0.4234586945062903
Macro F1 Score: 0.2830757108588909
Micro F1 Score: 0.4484428641490927

Training Set Confusion Matrix:
[[ 351   46  234    5   22    8  170  183   35    4    3   21    3  307
    22   24]
 [  94   53  111    0    8    3   71   76   11    2    3    8    0  115
    10    6]
 [ 176   53  537    7   25   16  172  235   40    6    4   21   20  479
    42   29]
 [   7    1   15    7    1    2   10   26   12    1    0    7    0   20
     7   10]
 [  19    1   34    3   49    1   19   41   39    0    1    3    1   85
     2    0]
 [  25    5   40    3    2   27   40   47   16    3    1    6    0   60
     7    2]
 [ 218   39  233    5   22   13  279  220   60    5    3   10    5  426
    19   21]
 [  46    9   81    6   15    9   98 1547   72    9    5    8    9  343
    61   57]
 [  52   12   95    9   26    7   89  176  196    3    9   14    6  381
    12   16]
 [   3    1    7    0    3    2    4   27    9   12    1    5    0   23
     7   14]
 [   6    5   14    0    0    1   11   25    9    3   12    5    4   66
     9    3]
 [  22    5   38    5    0    6   14   62   13    7    1   74    1   41
    18   12]
 [   8    1   23    2    1    1    1   17    9    1    3    1   48   20
     2    1]
 [ 109   24  177    4   19    7  173  351  119   10    9   11   11 1781
    27   38]
 [  11    4   24    2    0    5    9   64    6    4    1    7    0   16
  2264    8]
 [  43    6   70   11    7    7   51  145   27   13    0   10    4  147
    14   78]]
```

```
Test Set Confusion Matrix:
[[ 142   18   78    4   15    9   82   86    9    2    1    3    0  140
     9    8]
 [  31   23   52    0    3    4   39   25    3    1    1    1    0   56
     1    2]
 [  80   23  235    2   15    8   72  107   16    5    4    8   10  217
    22   21]
 [   3    0    6    9    0    0    2   11    3    1    0    3    0    9
     4    6]
 [   5    1   17    2   27    0    8   20   12    1    0    0    0   36
     2    2]
 [  12    2   16    0    2    5   18   17    1    0    0    2    2   13
     4    4]
 [  93   11  107    1   12    4  117  116   18    1    0    6    0  157
     9    9]
 [  27    2   39    2    2    4   29  727   36    8    1    1    2  118
    20   23]
 [  20    3   34    2   14    2   40   67   86    2    2    5    5  174
     4    1]
 [   1    0    4    0    1    0    1    9    2   15    0    6    0   13
     2    3]
 [   3    1    6    0    0    0    6    8    1    0    2    0    1   27
     4    0]
 [   7    2   11    0    3    1   10   13    8    2    4   26    1   15
     4    2]
 [   2    0    7    0    0    0    3    8    3    0    1    0   23    8
     1    0]
 [  44   14   80    0    8    3   57  146   59    1    3    6    6  722
    13   21]
 [   1    3   12    2    0    3    7   27    1    1    0    4    0   10
  1020    4]
 [  18    5   26    5    3    0   18   71   18    4    0    6    1   63
     8   43]]

Error Rate (Training): 0.5515571358509073
Error Rate (Test): 0.539187643020595
```

5) SVM CLASSIFIER

a) Linear SVM

```
Linear SVM
------------------------------
Training set results:
Accuracy: 0.33699117214320745
Recall: 0.3369911721432074S
Precision: 0.4166455430134216
F1 Score: 0.26419547985727004
Macro F1 Score: 0.11876072565894006
Micro F1 Score: 0.33699117214320745

Training Set Confusion Matrix:
[[   5    0  358    0    0    0   76  366    0    0    0    2    0  409
   219    3]
 [   1    0  190    0    0    0   27  135    0    0    0    2    0  150
    66    0]
 [   8    0  486    0    0    0  101  512    0    0    0    4    0  507
   243    1]
 [   0    0   32    0    0    0   14   29    0    0    0    0    0   22
    27    2]
 [   1    0   16    0    0    0   17  173    0    0    0    1    0   73
    17    0]
 [   3    0   58    0    0    0   32   74    0    0    0    1    0   58
    57    1]
 [   4    0  290    0    0    0  121  478    0    0    0    0    0  510
   175    0]
 [   2    0  136    0    0    0  173 1217    0    0    0    0    0  607
   239    1]
 [   2    0  128    0    0    0  142  366    0    0    0    1    0  357
   107    0]
 [   0    0   39    0    0    0    5   35    0    0    0    0    0   14
    24    1]
 [   2    0   43    0    0    0    6   22    0    0    0    1    0   60
    39    0]
 [   6    0   90    0    0    0   15   61    0    0    0   20    0   78
    49    0]
 [   1    0   43    0    0    0   30   17    0    0    0    2    0   14
    32    0]
 [   6    0  250    0    0    0   95  595    0    0    0    5    0 1642
   270    7]
 [   3    0  117    0    0    0   49  191    0    0    0    1    0   62
  2002    0]
 [   2    0  107    0    0    0   44  194    0    0    0    2    0  199
    81    4]]
```

```
Test set results:
Accuracy: 0.34024599542334094
Recall: 0.34024599542334094
Precision: 0.4291779273908714
F1 Score: 0.2683963096434787
Macro F1 Score: 0.12104047212273038
Micro F1 Score: 0.34024599542334094

Test Set Confusion Matrix:
[[  4   0 156   0   0   0  35 156   0   0   0   0   0 176  79   0]
 [  2   0  75   0   0   0  13  58   0   0   0   0   0  67  27   0]
 [  1   0 186   0   0   0  62 208   0   0   0   3   0 252 131   2]
 [  0   0   8   0   0   0   7  13   0   0   0   2   0  14  12   1]
 [  0   0  10   0   0   0   3  84   0   0   0   0   0  27   9   0]
 [  0   0  21   0   0   0   5  28   0   0   0   1   0  25  18   0]
 [  4   0  97   0   0   0  48 228   0   0   0   0   0 201  83   0]
 [  0   0  53   0   0   0  75 533   0   0   0   0   0 274 106   0]
 [  0   0  55   0   0   0  57 148   0   0   0   0   0 153  48   0]
 [  0   0  21   0   0   0   2  20   0   0   0   0   0   6   8   0]
 [  0   0   9   0   0   0   6   8   0   0   0   0   0  17  19   0]
 [  0   0  44   0   0   0   6  15   0   0   0  10   0  21  13   0]
 [  0   0  13   0   0   0  19   9   0   0   0   1   0   1  13   0]
 [  0   0 117   0   0   0  45 213   0   0   0   0   0 686 115   7]
 [  2   0  51   0   0   0  23  77   0   0   0   0   0  32 910   0]
 [  0   0  49   0   0   0  26  93   0   0   0   1   0  90  28   2]]
```

```
Error Rate (Training): 0.6630088278567925
Error Rate (Test): 0.659754004576659
```

6) ARTIFICIAL NEURAL NETWORK CLASSIFIER

a) ANN with 1 Hidden Layer (ADAM Optimizer)

```
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 740us/step
51/51 [==============================] - 0s 640us/step
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 680us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 600us/step
510/510 [==============================] - 0s 580us/step
219/219 [==============================] - 0s 583us/step

ANN with 1 Hidden Layer (ADAM Optimizer)
------------------------
Training set results:
Accuracy: 0.38683178028445314
Recall: 0.38683178028445314
Precision: 0.3456148470620421
F1 Score: 0.3370401692770812
Macro F1 Score: 0.19861012179286583
Micro F1 Score: 0.38683178028445314

Test set results:
Accuracy: 0.3812929061784897
Recall: 0.3812929061784897
Precision: 0.35213372232187695
F1 Score: 0.32959700668319625
Macro F1 Score: 0.19181449641549167
Micro F1 Score: 0.3812929061784897

Error Rate (Training): 0.6131682197155468
Error Rate (Test): 0.6187070938215102
```

b) ANN with 1 hidden layer classification (SGD)

```
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 598us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 580us/step
510/510 [==============================] - 0s 582us/step
219/219 [==============================] - 0s 582us/step

ANN with 1 Hidden Layer (SGD Optimizer)
---------------------------------------
Training set results:
Accuracy: 0.3795365375183914
Recall: 0.3795365375183914
Precision: 0.366707210722811
F1 Score: 0.3186870805469884
Macro F1 Score: 0.16963766264053728
Micro F1 Score: 0.37953653751839145

Test set results:
Accuracy: 0.3805778032036613
Recall: 0.3805778032036613
Precision: 0.349477287161445
F1 Score: 0.3151277508202618
Macro F1 Score: 0.16009101687958327
Micro F1 Score: 0.3805778032036613

Error Rate (Training): 0.6204634624816086
Error Rate (Test): 0.6194221967963387
```

c) ANN with 1 Hidden Layer (RMSprop Optimizer)

```
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 600us/step
51/51 [==============================] - 0s 620us/step
51/51 [==============================] - 0s 580us/step
51/51 [==============================] - 0s 600us/step
510/510 [==============================] - 0s 570us/step
219/219 [==============================] - 0s 577us/step

ANN with 1 Hidden Layer (RMSprop Optimizer)
-------------------------------------------
Training set results:
Accuracy: 0.39093918587542914
Recall: 0.39093918587542914
Precision: 0.36105125281338457
F1 Score: 0.3371471518637557
Macro F1 Score: 0.19222235548464617
Micro F1 Score: 0.39093918587542914

Test set results:
Accuracy: 0.3865846681922197
Recall: 0.3865846681922197
Precision: 0.33987150955488143
F1 Score: 0.32962542180825544
Macro F1 Score: 0.18923698003148953
Micro F1 Score: 0.3865846681922197

Error Rate (Training): 0.6090608141245708
Error Rate (Test): 0.6134153318077803
```

7) NAÏVE BAYES CLASSIFICATION

a)Bernoulli Naïve Bayes

```
Bernoulli Naive Bayes
---------------------
Training set results:
Accuracy: 0.2584600294261893
Recall: 0.2584600294261893
Precision: 0.32481191763664596
F1 Score: 0.1919264450198271
Macro F1 Score: 0.09116516712978052
Micro F1 Score: 0.2584600294261893

Training Set Confusion Matrix:
[[   2    0  112    2    0    0    0  522    0    1    0   15    0  655
   125    4]
 [   0    0   91    0    0    0    1  199    0    1    0    6    0  227
    45    1]
 [   9    0  121    4    0    1    0  608    0    2    0   29    0  914
   172    2]
 [   0    0    7    2    0    0    0   65    0    0    0    6    0   31
    15    0]
 [   0    0    4    0    0    0    0  222    0    0    0    1    0   54
    16    1]
 [   2    0   18    2    0    0    0  151    0    0    0    4    0   53
    53    1]
 [   2    0   95    7    0    0    0  758    0    1    0   19    0  587
   105    4]
 [   5    0   19    0    0    1    4 1237    0    1    0   16    0  876
   204   12]
 [   3    0   30    2    0    0    0  649    0    0    0    9    0  312
    97    1]
 [   0    0   13    0    0    0    0   47    0    0    0    7    0   25
    26    0]
 [   0    0   12    0    0    0    0   47    0    0    0    2    0   65
    45    2]
 [   1    0    9    2    0    0    0  100    0    2    0   35    0   53
    84   33]
 [   0    0    7    0    0    0    1   58    0    0    0    3    0   51
    19    0]
 [   1    0   46    1    0    0    1  993    0    1    0   45    0 1562
   183   37]
 [   3    0   53    2    0    1    0  428    0    0    0    8    0  707
  1223    0]
 [   0    0   24    4    0    0    0  273    0    3    0   27    0  170
    98   34]]
```

```
Test set results:
Accuracy: 0.2595823798627002
Recall: 0.2595823798627002
Precision: 0.3542222434817792
F1 Score: 0.19548213866834757
Macro F1 Score: 0.09318945554728432
Micro F1 Score: 0.2595823798627002

Test Set Confusion Matrix:
[[  1   0  43   0   0   0   0 237   0   0   0   8   0 257  59   1]
 [  0   0  42   0   0   0   0  77   0   0   0   1   0  99  23   0]
 [  4   0  55   1   0   0   0 272   0   0   0  12   0 433  67   1]
 [  0   0   0   1   0   0   0  27   0   0   0   4   0  13  11   1]
 [  0   0   3   0   0   0   0 101   0   0   0   1   0  17  11   0]
 [  0   0   7   1   0   0   0  54   0   0   0   4   0  19  13   0]
 [  3   0  27   3   0   0   0 322   0   0   0   2   0 257  46   1]
 [  1   0  14   0   0   0   1 533   0   0   0   5   0 414  71   2]
 [  1   0   9   0   0   0   0 288   0   0   0   1   0 126  35   1]
 [  0   0   5   0   0   0   0  31   0   0   0   0   0   5  16   0]
 [  1   0   5   0   0   0   0  29   0   0   0   0   0  15   9   0]
 [  0   0   8   1   0   0   0  23   0   0   0  14   0  24  28  11]
 [  0   0   3   0   0   0   0  33   0   0   0   1   0  15   4   0]
 [  0   0  21   0   0   0   0 393   0   0   0  14   0 658  79  18]
 [  1   0  25   0   0   0   0 179   0   0   0   2   0 351 537   0]
 [  2   0  12   4   0   0   0 136   0   0   0   9   0  77  33  16]]

Error Rate (Training): 0.7415399705738107
Error Rate (Test): 0.7404176201372998
```

b)Gaussian Naïve Bayes

```
Gaussian Naive Bayes
--------------------
Training set results:
Accuracy: 0.27568661108386466
Recall: 0.27568661108386466
Precision: 0.2662659493172436
F1 Score: 0.23390756213726954
Macro F1 Score: 0.13362330489255556
Micro F1 Score: 0.27568661108386466

Training Set Confusion Matrix:
[[  26   37   88   13  111    9  268  461   44   12    0   31   15  105
   203   15]
 [  17   27   26    5   59    2   93  167   12    6    0   17    0   34
   102    4]
 [  52   40  103   12  148   17  326  649   33   18    0   43   21  122
   261   17]
 [   1    1   10    3    5    2   18   31    6    8    0   14    0    2
    18    7]
 [   1    3    6    1   53    2   29  165    5    1    0    5    2    7
    17    1]
 [   4    1   12    2   24    6   57   82    6    3    0   14    3   12
    57    1]
 [  30   26   54    8  147    4  304  578   63   22    0   29    6  107
   190   10]
 [  15    9   27    3  243    8  231 1459   42   13    1   27    2   42
   232   21]
 [  12   13   29    1  121    5  225  454   35   10    0   18    1   64
   110    5]
 [   0    6   13    1   10    1   15   38    0    6    0    0    0    2
    23    3]
 [   4    6    9    0   12    1   27   49    1    7    0    2    1   16
    35    3]
 [  10    5   24    6   11    5   50   57    7    8    0   61    5    4
    38   28]
 [   3    3   20    1   13    2   30   23    2    2    0    6    5    3
    26    0]
 [  22   22   78    3  285   11  389 1017   56   23    0   25   11  498
   348   82]
 [  27    2   59    7   86    3  142  121   11    9    0   49   39    4
  1857    9]
 [   7   10   41    8   46    2   84  215   18   19    0   11    6   32
    80   54]]
```

```
Test set results:
Accuracy: 0.30234553775743706
Recall: 0.30234553775743706
Precision: 0.28632734385800684
F1 Score: 0.2501217418235451
Macro F1 Score: 0.14909969295003891
Micro F1 Score: 0.30234553775743706

Test Set Confusion Matrix:
[[ 12  19  28   9  18   2 126 217  17   6   0  21   9  41  78   3]
 [ 11   8  13   2   2   0  30  91   6   3   0   3   0  22  49   2]
 [ 13  14  41   5  10   9 139 336  15  11   0  22  11  73 139   7]
 [  2   0   5   4   0   0   7  20   3   1   0   7   0   0   4   4]
 [  0   1   3   0  15   1   8  83   5   0   0   4   1   3   8   1]
 [  4   4   7   1   4   2  12  31   3   3   0   8   0   2  16   1]
 [  7   7  25   1  20   4 133 288  28   5   0   9   3  38  89   4]
 [  7   4  10   0  27   2 126 732  12   5   0   6   1  11  91   7]
 [  4   5  11   0  26   3  98 211  17   2   0  11   0  28  39   6]
 [  1   3   7   0   0   1   5  22   1   9   0   0   0   0   8   0]
 [  3   4   4   0   2   0   6  16   2   2   0   0   0   2  18   0]
 [  4   4  12   2   0   1  18  13   1   3   0  23   3   2  11  12]
 [  1   0   8   0   1   1  19  14   0   0   0   1   3   0   8   0]
 [  8   4  29   1  30  13 184 443  20   9   0   9   2 235 158  38]
 [  8   1  21   5   3   3  72  62   5   2   0  24  25   4 856   4]
 [  3   3  17   6   6   3  29 117   9  10   0   5   0  23  34  24]]

Error Rate (Training): 0.7243133889161353
Error Rate (Test): 0.6976544622425629
```

c) Multinomial Naïve Bayes

```
Multinomial Naive Bayes
-----------------------
Training set results:
Accuracy: 0.18354585581167238
Recall: 0.18354585581167238
Precision: 0.3539616090243747
F1 Score: 0.1374042006960974
Macro F1 Score: 0.06212173571347
Micro F1 Score: 0.1835458558116724

Training Set Confusion Matrix:
[[  24    0   27    0    0    0    0  220    0    0    0  413    0  267
   430   57]
 [   5    0   10    0    0    0    0   76    0    0    0  147    0  105
   206   22]
 [  32    0   19    0    0    0    1  245    0    0    0  525    1  350
   617   72]
 [   0    0    0    0    0    0    0    5    0    0    0   63    0   19
    37    2]
 [   1    0    0    0    0    0    0   73    0    0    0   52    0   45
   123    4]
 [   1    0    2    0    0    0    0   38    0    0    0  134    0   37
    67    5]
 [  16    0    9    1    0    0    1  247    0    0    0  420    0  323
   493   68]
 [   8    0    5    0    0    0    0  574    0    0    0  691    0  314
   716   67]
 [   7    0    5    0    0    0    0  149    0    0    0  301    0  260
   348   33]
 [   0    0    1    0    0    0    0    2    0    0    0   70    0   15
    26    4]
 [   1    0    1    0    0    0    0    2    0    0    0   66    0   24
    77    2]
 [   7    0    0    0    0    0    0    9    0    0    0  222    0   33
    47    1]
 [   0    0    2    0    0    0    0    1    0    0    0   32    0   21
    83    0]
 [  20    0    7    0    0    0    0  489    0    0    0  628    0  650
   976  100]
 [  49    0   13    0    0    0    0   58    0    0    0  571    0  224
  1495   15]
 [   2    0    9    0    0    0    0   88    0    0    0  312    0   89
   124    9]]
```

```
Test set results:
Accuracy: 0.18135011441647597
Recall: 0.18135011441647597
Precision: 0.3019728754822462
F1 Score: 0.13566613785836384
Macro F1 Score: 0.05914685832426173
Micro F1 Score: 0.18135011441647597

Test Set Confusion Matrix:
[[  8   0  10   0   0   0   0 100   0   0   0 151   0 131 173  33]
 [  5   0   6   0   0   0   0  31   0   0   0  63   0  39  87  11]
 [  4   0   5   0   0   0   0 128   0   0   0 246   0 139 303  20]
 [  0   0   0   0   0   0   0   0   0   0   0  31   0  10  16   0]
 [  0   0   0   0   0   0   0  27   0   0   0  26   0  22  56   2]
 [  1   0   1   0   0   0   0   8   0   0   0  39   0  17  31   1]
 [  0   0   6   0   0   0   0 123   0   0   0 162   0 147 195  28]
 [  7   0   0   0   0   0   0 242   0   0   0 284   0 141 335  32]
 [  6   0   1   0   0   0   1  73   0   0   0  93   0 114 163  10]
 [  0   0   0   0   0   0   0   3   0   0   0  42   0   2  10   0]
 [  1   0   0   0   0   0   0   1   0   0   0  19   0  10  28   0]
 [  3   0   1   0   0   0   0   4   0   0   0  68   0  17  15   1]
 [  0   0   0   0   0   0   0   0   0   0   0  10   0   3  43   0]
 [  9   0   3   0   0   0   0 177   0   0   0 283   0 264 412  35]
 [ 20   0   3   0   0   0   0  18   0   0   0 261   0 108 677   8]
 [  2   0   0   0   0   0   0  37   0   0   0 141   1  34  70   4]]

Error Rate (Training): 0.8164541441883276
Error Rate (Test): 0.818649885583524
```