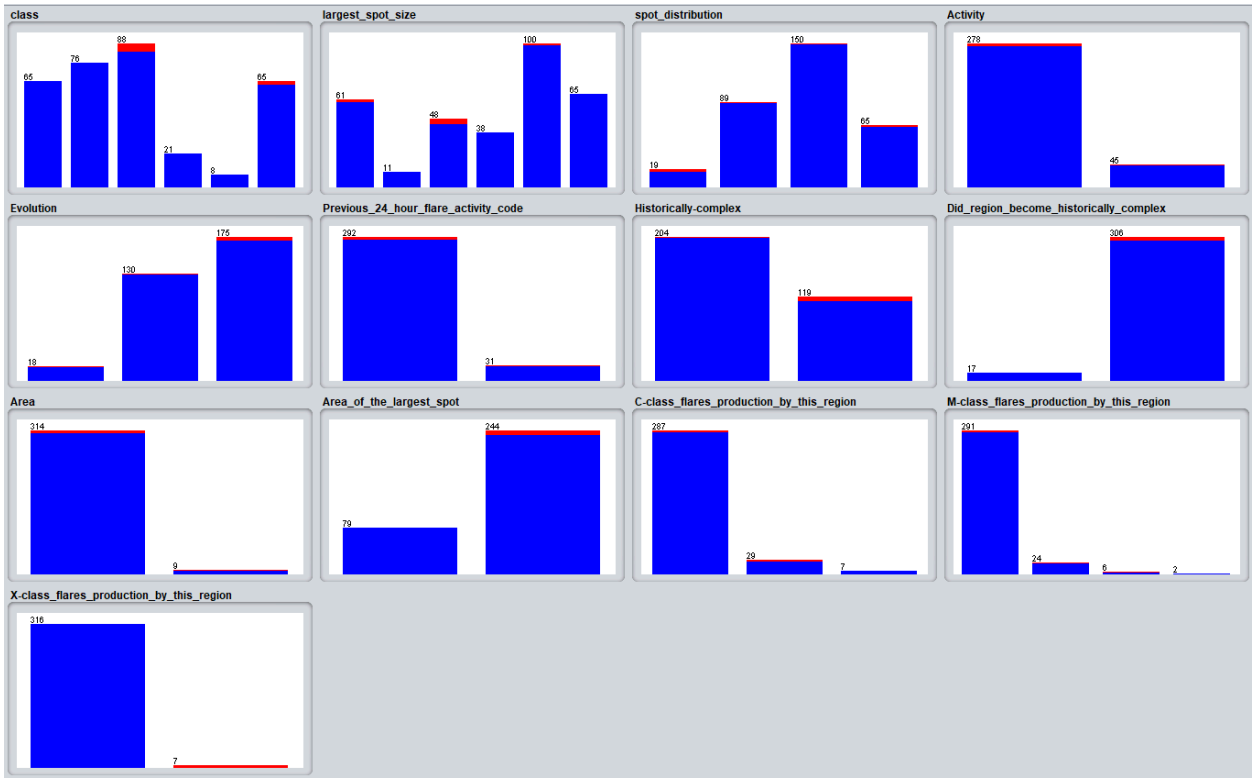

The class difference is huge and therefore our data may be skewed. However in this case, if we do any pre-processing to the data, then our results will also start to vary, therefore we don't need to do any preprocessing. (Besides applied the basic preprocessing techniques, meaning that i removed the missing values but there were none.)

Applied Techniques on the class attribute because that is what we’re trying to predict.

Decision Tree - achieved an accuracy of 72%

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         233               72.1362 %
Incorrectly Classified Instances        90               27.8638 %
Kappa statistic                          0.6379
Mean absolute error                      0.109
Root mean squared error                  0.2494
Relative absolute error                 41.6334 %
Root relative squared error             68.9494 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     B
                 0.671    0.170    0.548      0.671    0.604      0.469    0.838     0.549     C
                 0.580    0.191    0.531      0.580    0.554      0.378    0.807     0.527     D
                 0.048    0.007    0.333      0.048    0.083      0.105    0.730     0.150     E
                 0.000    0.003    0.000      0.000    0.000      -0.009   0.880     0.165     F
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     H
Weighted Avg.    0.721    0.093    0.698      0.721    0.701      0.623    0.889     0.689

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 65  0  0  0  0  0 |  a = B
  0 51 24  1  0  0 |  b = C
  0 35 51  1  1  0 |  c = D
  0  3 17  1  0  0 |  d = E
  0  4  4  0  0  0 |  e = F
  0  0  0  0  0 65 |  f = H
```

Naive Bayes achieved an accuracy of 65%

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         212               65.6347 %
Incorrectly Classified Instances       111               34.3653 %
Kappa statistic                          0.5601
Mean absolute error                      0.1257
Root mean squared error                  0.2631
Relative absolute error                 48.0006 %
Root relative squared error             72.7559 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.969    0.016    0.940      0.969    0.955      0.943    0.998     0.995     B
                 0.671    0.190    0.520      0.671    0.586      0.444    0.860     0.598     C
                 0.386    0.174    0.453      0.386    0.417      0.223    0.777     0.456     D
                 0.095    0.053    0.111      0.095    0.103      0.045    0.810     0.156     E
                 0.000    0.010    0.000      0.000    0.000      -0.015   0.772     0.067     F
                 0.954    0.000    1.000      0.954    0.976      0.971    1.000     1.000     H
Weighted Avg.    0.656    0.099    0.644      0.656    0.647      0.553    0.888     0.678

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 63  2  0  0  0  0 |  a = B
  3 51 21  1  0  0 |  b = C
  1 39 34 13  1  0 |  c = D
  0  3 15  2  1  0 |  d = E
  0  2  5  1  0  0 |  e = F
  0  1  0  1  1 62 |  f = H
```

SVM achieved an accuracy of 70%

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         227               70.2786 %
Incorrectly Classified Instances        96               29.7214 %
Kappa statistic                          0.6143
Mean absolute error                      0.2314
Root mean squared error                  0.3246
Relative absolute error                 88.3464 %
Root relative squared error             89.7551 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     B
               0.645    0.194    0.505      0.645    0.566      0.417    0.762     0.428     C
               0.534    0.187    0.516      0.534    0.525      0.343    0.765     0.455     D
               0.048    0.013    0.200      0.048    0.077      0.069    0.742     0.163     E
               0.000    0.000    ?          0.000    ?          ?        0.646     0.038     F
               1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     H
Weighted Avg.  0.703    0.098    ?          0.703    ?          ?        0.854     0.639

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 65  0  0  0  0  0 |  a = B
  0 49 26  1  0  0 |  b = C
  0 40 47  1  0  0 |  c = D
  0  3 17  1  0  0 |  d = E
  0  5  1  2  0  0 |  e = F
  0  0  0  0  0 65 |  f = H
```

Decision tree achieved the best accuracy of 72 percent because we have multiple classes and the decision tree tends to perform the best when we have multiple classes

Applying on other attributes

Decision Trees

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         176               54.4892 %
Incorrectly Classified Instances       147               45.5108 %
Kappa statistic                          0.4023
Mean absolute error                      0.1721
Root mean squared error                  0.3188
Relative absolute error                 65.2015 %
Root relative squared error             87.7845 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.262     0.126     0.327      0.262     0.291      0.149     0.665     0.288     A
               0.000     0.006     0.000      0.000     0.000     -0.015     0.667     0.104     H
               0.396     0.065     0.514      0.396     0.447      0.369     0.754     0.503     K
               0.079     0.053     0.167      0.079     0.107      0.037     0.678     0.171     R
               0.730     0.354     0.480      0.730     0.579      0.348     0.707     0.433     S
               1.000     0.000     1.000      1.000     1.000      1.000     1.000     1.000     X
Weighted Avg.  0.545     0.150     0.508      0.545     0.515      0.396     0.760     0.488

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0 14  3 28  0 |  a = A
  2  0  0  3  6  0 |  b = H
 17  0 19  0 12  0 |  c = K
  1  1  0  3 33  0 |  d = R
 13  1  4  9 73  0 |  e = S
  0  0  0  0  0 65 |  f = X
```

Naive Bayes

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         185               57.2755 %
Incorrectly Classified Instances       138               42.7245 %
Kappa statistic                          0.4431
Mean absolute error                      0.1611
Root mean squared error                  0.303
Relative absolute error                 61.0285 %
Root relative squared error             83.4364 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.393    0.168    0.353      0.393    0.372      0.216    0.759     0.351     A
                 0.000    0.003    0.000      0.000    0.000      -0.010   0.581     0.097     H
                 0.479    0.058    0.590      0.479    0.529      0.460    0.878     0.658     K
                 0.105    0.039    0.267      0.105    0.151      0.102    0.806     0.259     R
                 0.700    0.287    0.522      0.700    0.598      0.388    0.781     0.568     S
                 0.985    0.008    0.970      0.985    0.977      0.971    0.999     0.997     X
Weighted Avg.    0.573    0.135    0.543      0.573    0.548      0.436    0.831     0.575

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 24  0 11  4 22  0 |  a = A
  3  0  3  0  5  0 |  b = H
 17  0 23  0  8  0 |  c = K
  4  0  0  4 29  1 |  d = R
 19  1  2  7 70  1 |  e = S
  1  0  0  0  0 64 |  f = X
```

.

SVM

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         192               59.4427 %
Incorrectly Classified Instances       131               40.5573 %
Kappa statistic                          0.463
Mean absolute error                      0.238
Root mean squared error                  0.3343
Relative absolute error                 90.1617 %
Root relative squared error             92.0651 %
Total Number of Instances              323

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.492    0.176    0.395      0.492    0.438      0.292    0.727     0.330     A
                 0.000    0.003    0.000      0.000    0.000      -0.010   0.535     0.038     H
                 0.396    0.015    0.826      0.396    0.535      0.527    0.864     0.564     K
                 0.000    0.014    0.000      0.000    0.000      -0.041   0.735     0.208     R
                 0.780    0.341    0.506      0.780    0.614      0.407    0.727     0.473     S
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     X
Weighted Avg.    0.594    0.143    0.555      0.594    0.554      0.455    0.797     0.520

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 30  0  2  1 28  0 |  a = A
  5  0  1  0  5  0 |  b = H
 19  0 19  0 10  0 |  c = K
  5  0  0  0 33  0 |  d = R
 17  1  1  3 78  0 |  e = S
  0  0  0  0  0 65 |  f = X
```