**Problem:**

All the patients suffered heart attacks at some point in the past. Some are still alive and some are not. The survival and still-alive variables, when taken together, indicate whether a patient survived for at least one year following the heart attack.

The problem addressed by past researchers was to predict from the other variables whether or not the patient will survive at least one year.

**Objective:**

* Analyzing and comparing performance of following classifier using ROC Graph:

1. J48
2. JRip
3. KStar
4. Naive Bayes
5. ZeroR

* Finding the most suitable classifier.

**Dataset Prepare:**

In this dataset, there are a lot of missing values. My task was to determine whether the patient will survive one year or not. For that I have chosen alive-at-1 attribute as my decision class. And the missing value was filled up by using attributes survival (number of months), still-alive and age-at-heart-attack as following the attribute information section.

Form the dataset, we can see that for all still-alive’s 0 value, alive-at-1 value is 0, so I filled up all missing value of alive-at-1 with 0 for having still-alive’s 0. For alive-at-1’s value 1, I have seen that survival < 12 months but age-at-heart-attack’s value is >=60, they survived. So, I filled up the missing value with 1 in such cases and for the rest of the cases I have filled up with 0.

**JRip Classifier:**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 128 98.4615 %

Incorrectly Classified Instances 2 1.5385 %

Kappa statistic 0.8669

Mean absolute error 0.0229

Root mean squared error 0.1238

Relative absolute error 16.9678 %

Root relative squared error 48.6851 %

Total Number of Instances 130

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.778 0.000 1.000 0.778 0.875 0.875 0.838 0.793 survive

1.000 0.222 0.984 1.000 0.992 0.875 0.838 0.976 not\_Survive

Weighted Avg. 0.985 0.207 0.985 0.985 0.984 0.875 0.838 0.964

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = survive

0 121 | b = not\_Survive

**ZeroR Classifier:**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 121 93.0769 %

Incorrectly Classified Instances 9 6.9231 %

Kappa statistic 0

Mean absolute error 0.1352

Root mean squared error 0.2542

Relative absolute error 100 %

Root relative squared error 100 %

Total Number of Instances 130

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.000 0.000 ? 0.000 ? ? 0.446 0.069 survive

1.000 1.000 0.931 1.000 0.964 ? 0.446 0.924 not\_Survive

Weighted Avg. 0.931 0.931 ? 0.931 ? ? 0.446 0.865

=== Confusion Matrix ===

a b <-- classified as

0 9 | a = survive

0 121 | b = not\_Survive

**J48 classifier:**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 129 99.2308 %

Incorrectly Classified Instances 1 0.7692 %

Kappa statistic 0.9371

Mean absolute error 0.0118

Root mean squared error 0.0694

Relative absolute error 8.7084 %

Root relative squared error 27.2829 %

Total Number of Instances 130

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.889 0.000 1.000 0.889 0.941 0.939 1.000 1.000 survive

1.000 0.111 0.992 1.000 0.996 0.939 1.000 1.000 not\_Survive

Weighted Avg. 0.992 0.103 0.992 0.992 0.992 0.939 1.000 1.000

=== Confusion Matrix ===

a b <-- classified as

8 1 | a = survive

0 121 | b = not\_Survive

**Naïve Bayes Classifier:**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 122 93.8462 %

Incorrectly Classified Instances 8 6.1538 %

Kappa statistic 0.1888

Mean absolute error 0.0907

Root mean squared error 0.2486

Relative absolute error 67.0604 %

Root relative squared error 97.8135 %

Total Number of Instances 130

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.111 0.000 1.000 0.111 0.200 0.323 0.779 0.296 survive

1.000 0.889 0.938 1.000 0.968 0.323 0.779 0.970 not\_Survive

Weighted Avg. 0.938 0.827 0.942 0.938 0.915 0.323 0.779 0.924

=== Confusion Matrix ===

a b <-- classified as

1 8 | a = survive

0 121 | b = not\_Survive

**KStar Classifier:**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 117 90 %

Incorrectly Classified Instances 13 10 %

Kappa statistic 0.0825

Mean absolute error 0.1016

Root mean squared error 0.2876

Relative absolute error 75.1513 %

Root relative squared error 113.1413 %

Total Number of Instances 130

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.111 0.041 0.167 0.111 0.133 0.084 0.672 0.162 survive

0.959 0.889 0.935 0.959 0.947 0.084 0.672 0.967 not\_Survive

Weighted Avg. 0.900 0.830 0.882 0.900 0.891 0.084 0.672 0.911

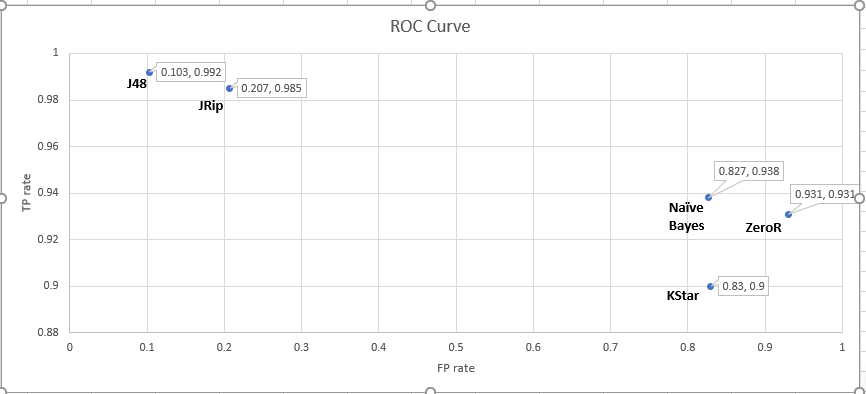
=== Confusion Matrix ===

a b <-- classified as

1 8 | a = survive

5 116 | b = not\_Survive

**ROC Graph:**



**Comments:**

We can see in the ROC graph that J48 is giving the best result for my interest. The FP rate is lower as it gave only 0 output as wrong out of 121 input and TP rate is also higher where only 1 output came as wrong out of 8 input. Other classifiers are also giving near value but KStar gave worst result out of all the classifiers.

So, for the prediction I choose J48 as the classifier.