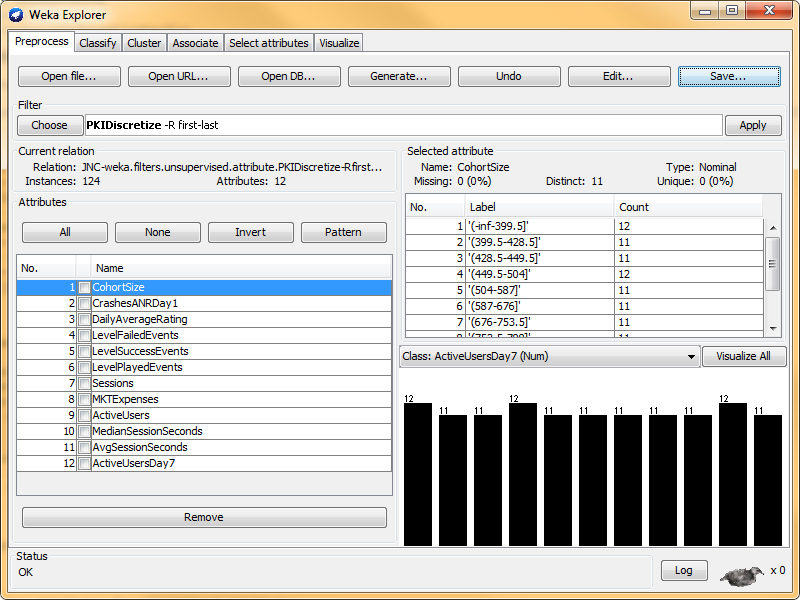
**Data Pre-processing**

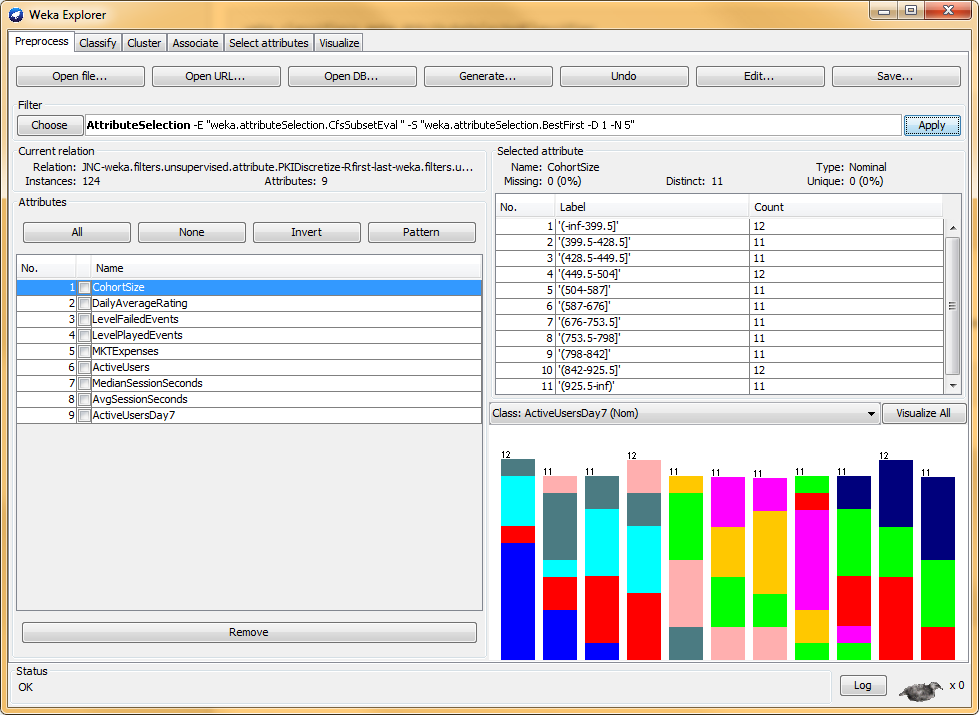
**1. Converting Numeric Attributes to Nominal**

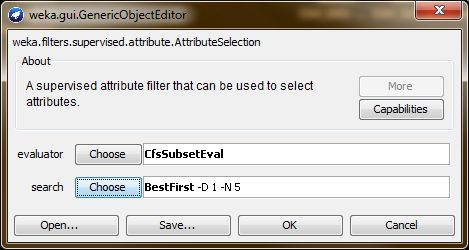
The all features in the test dataset are all numeric. Converting the numeric values to nominal may enable to help a more simple and accurate classification.To convert the numeric attributes to nominal we use the **PKIDiscretize Filter.** This filter Discretizes numeric attributes using equal frequency binning, where the number of bins is equal to the square root of the number of non-missing values.

****

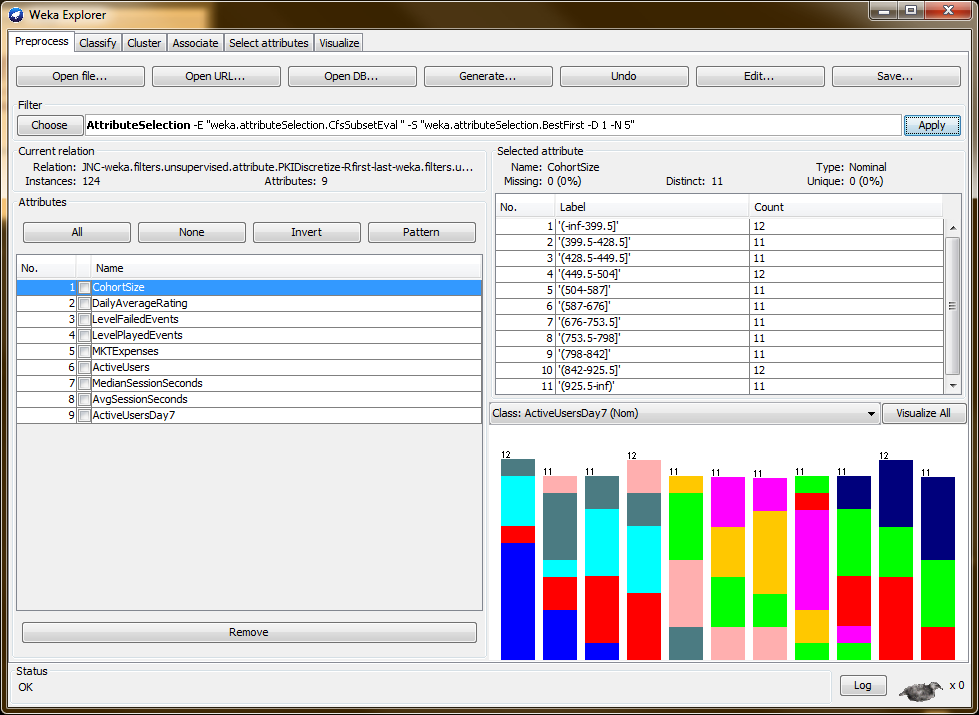
**2. Attribute Selection**

To increase the accuracy of classification, redundant and irrelevant attributes needs to be removed from the dataset. To perform this, **Attribute Selection Filter** is used. The evaluator used is **CfsSubsetEval** that evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Search Method use is **BestFit.**

****



After feature selection the JNC dataset is now reduced to 9 attributes



**Building Classification Model**

**1. Using J48 Classifier**

The first model would be using the J48 classifier with using percentage split of 66% to test build the classification model.

Using All Attributes

|  |  |  |  |
| --- | --- | --- | --- |
| **Seed** | **Accuracy** | **Seed** | **Accuracy** |
| 1 | 38.0952 % | 6 | 38.0952 % |
| 2 | 38.0952 % | 7 | 38.0952 % |
| 3 | 38.0952 % | 8 | 38.0952 % |
| 4 | 38.0952 % | 9 | 38.0952 % |
| 5 | 38.0952 % | 10 | 38.0952 % |

Using Attribute Selection

|  |  |  |  |
| --- | --- | --- | --- |
| **Seed** | **Accuracy** | **Seed** | **Accuracy** |
| 1 | 42.8571 % | 6 | 42.8571 % |
| 2 | 42.8571 % | 7 | 42.8571 % |
| 3 | 42.8571 % | 8 | 42.8571 % |
| 4 | 42.8571 % | 9 | 42.8571 % |
| 5 | 42.8571 % | 10 | 42.8571 % |