Work done so far:

Binned the dataset

Splitting the dataset into training and test set (2/3 of the dataset is training = 732 instances, 1/3 of the dataset is test set = 371 instances ) out of these in order to avoid the class imbalance problem 6.94 % of the total instances should be ‘TRUE’ the remaining would be ‘FALSE’.

Training dataset:

Total number of instance: 732

Out of these ‘TRUE’ instances would be 51

Out of these ‘FALSE’ instance would be 681

Test dataset:

Total number of instance: 371

Out of these ‘TRUE’ instances would be 26

Out of these ‘FALSE’ instance would be 345

Attribute Selection Method

1. **Principle Component**

**Classifier:**

J48 : 92.9919

1. Correlation
2. SymmetricalUncertAttributeEval

* locCodeandComment
* loBlank
* loComment
* uniqueOP
* locuniue  
  \_Opnd