Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE 4142

Course Title: Data Warehouse and Mining Lab

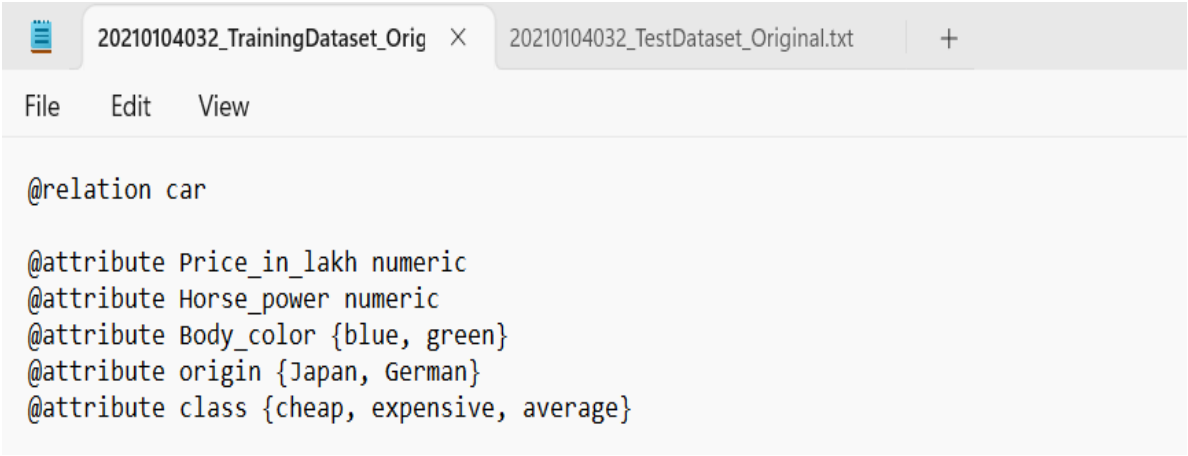
Assignment No: 01

Date of Submission: 24.01.2025

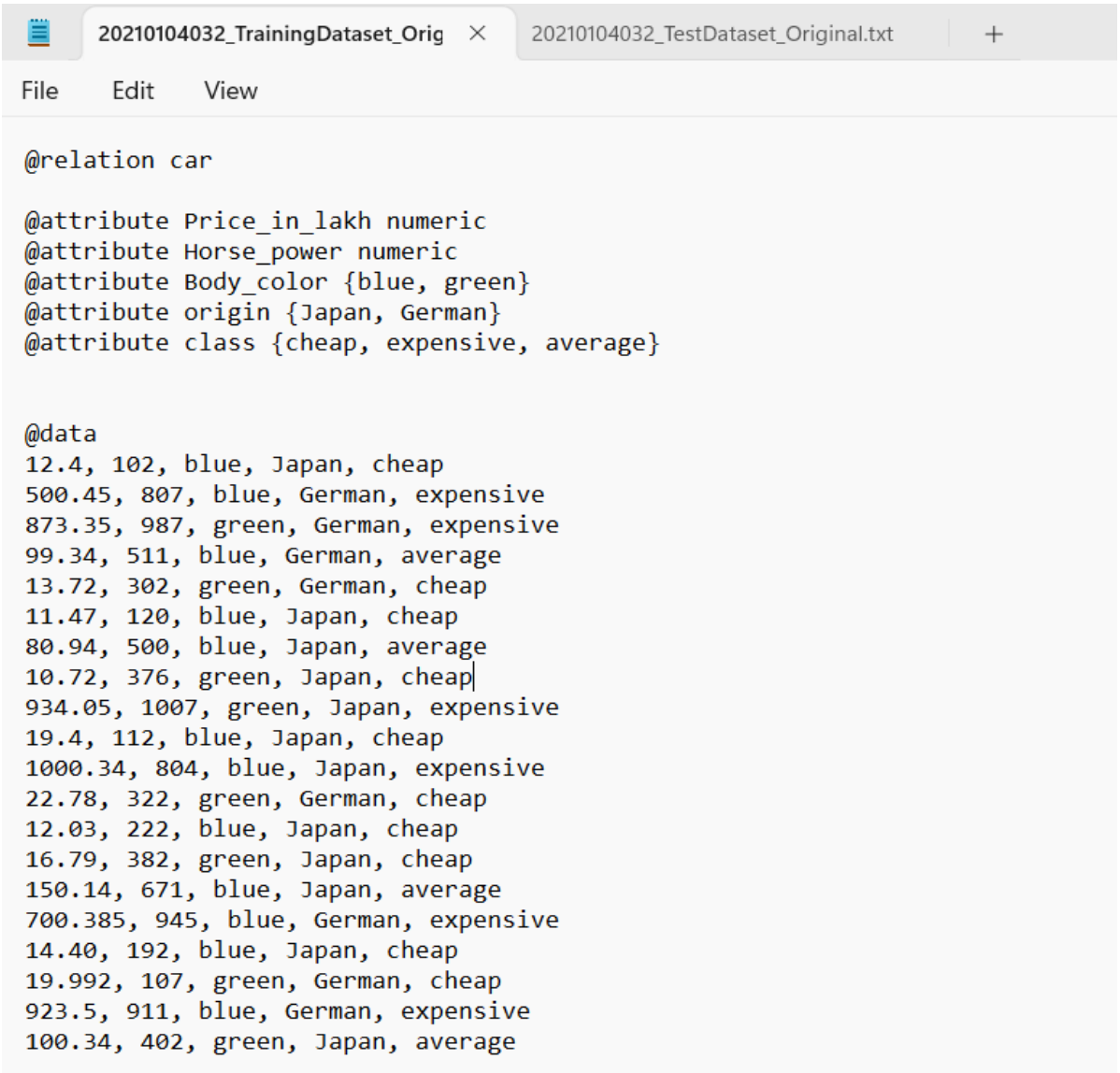
Submitted to:
Mr. Saha Reno
Assistant Professor
Department of CSE, AUST

Submitted by:
Name: Afia Fahmida
Student ID: 20210104032
Group: 02

(i)Customized Dataset:



(ii) 20 instances of Training dataset (10 for 1st class, 6 for 2nd class and rest for 3rd class):



Test Dataset with 5 instances:

20210104032_TrainingDataset_Original.arff | 20210104032_TestDataset_Original.

File Edit View

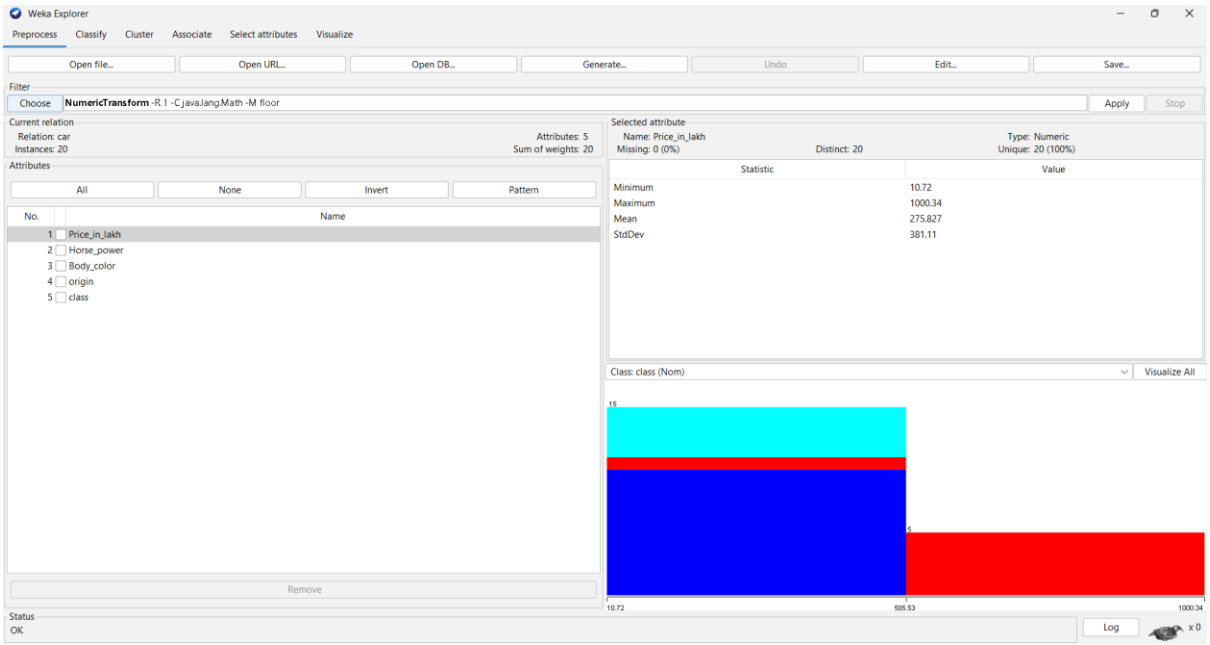
```
@relation car

@attribute Price_in_lakh numeric
@attribute Horse_power numeric
@attribute Body_color {blue, green}
@attribute origin {Japan, German}
@attribute class {cheap, expensive, average}


@data
23.73, 162, green, Japan, cheap
900.34, 704, blue, German, expensive
190.34, 412, green, Japan, average
300.4, 704, blue, German, average
190.34, 412, green, Japan, average
4.23, 120, blue, German, cheap
```

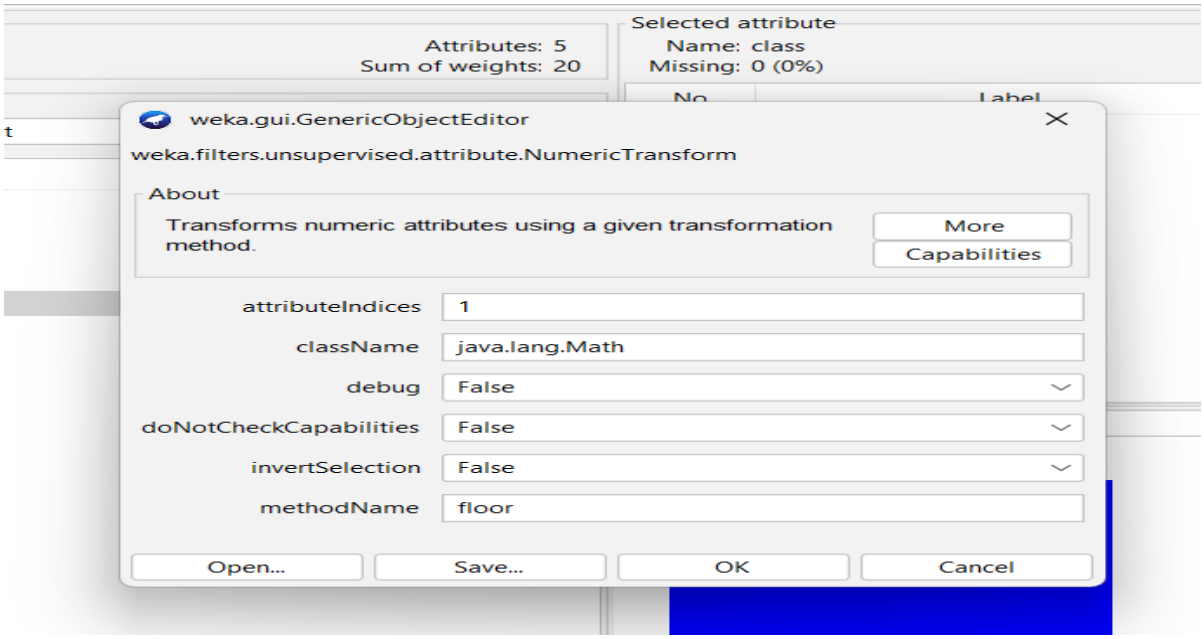
(iii) Convert Any 1 Real Attribute's Values from Float to Integers (which is less than or equal to the original value):

The picture below shows selection of “NumericTransform” filter from unsupervised category of attributes,

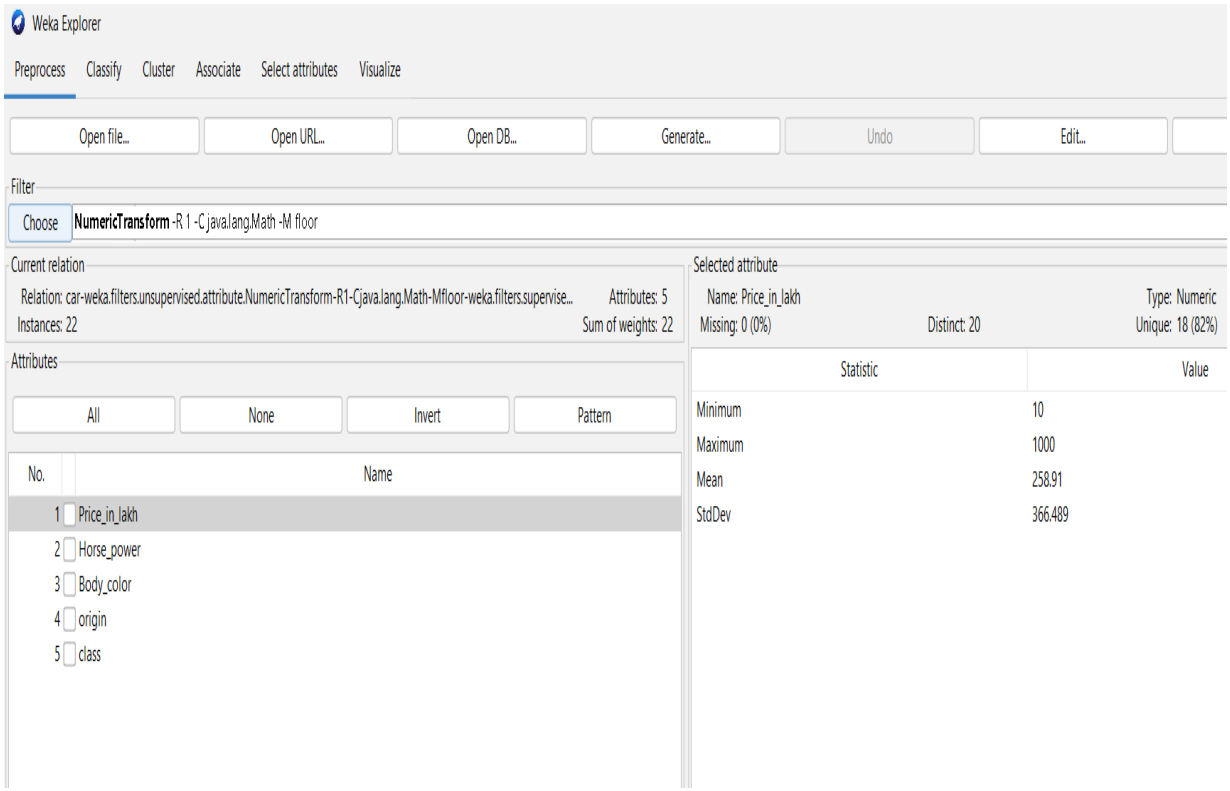


After pressing on the filter box I have changed attributeIndices to 1 and methodName to floor so all the float values of 1 attribute turns into integers

that is smaller or equal to the real value. The step is shown in the screenshot below,

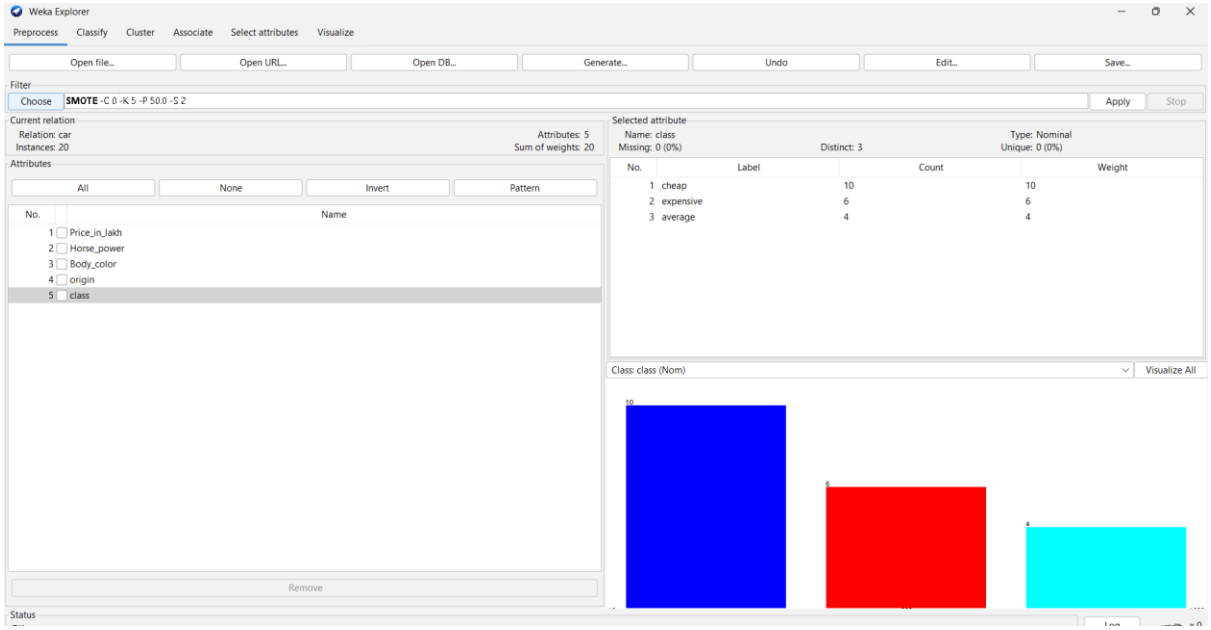


After pressing apply, it can be seen that there's change in minimum and maximum value on left side panel in the given screenshot of weka which means it worked,

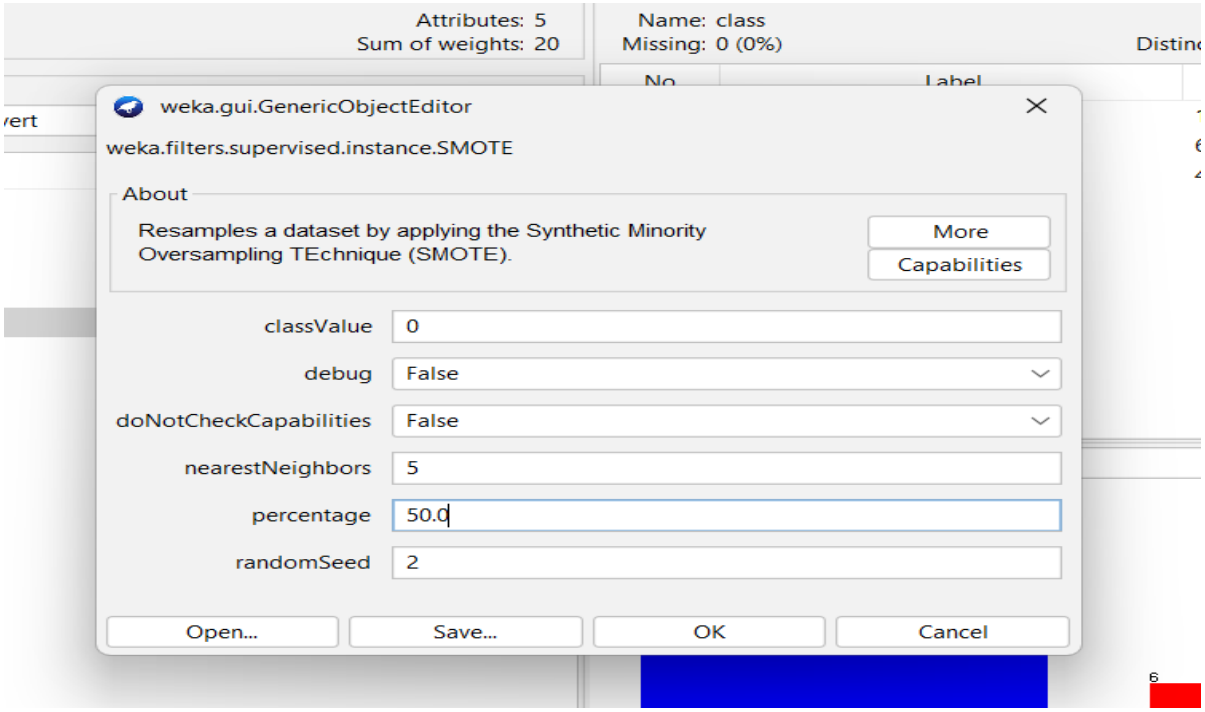


(iv) Fix the Class Imbalance Problem for the 2nd and 3rd Class by Making the Number of Instances for 2nd Class and 3rd Class Equal as the Number of Instances for 1st Class (10):

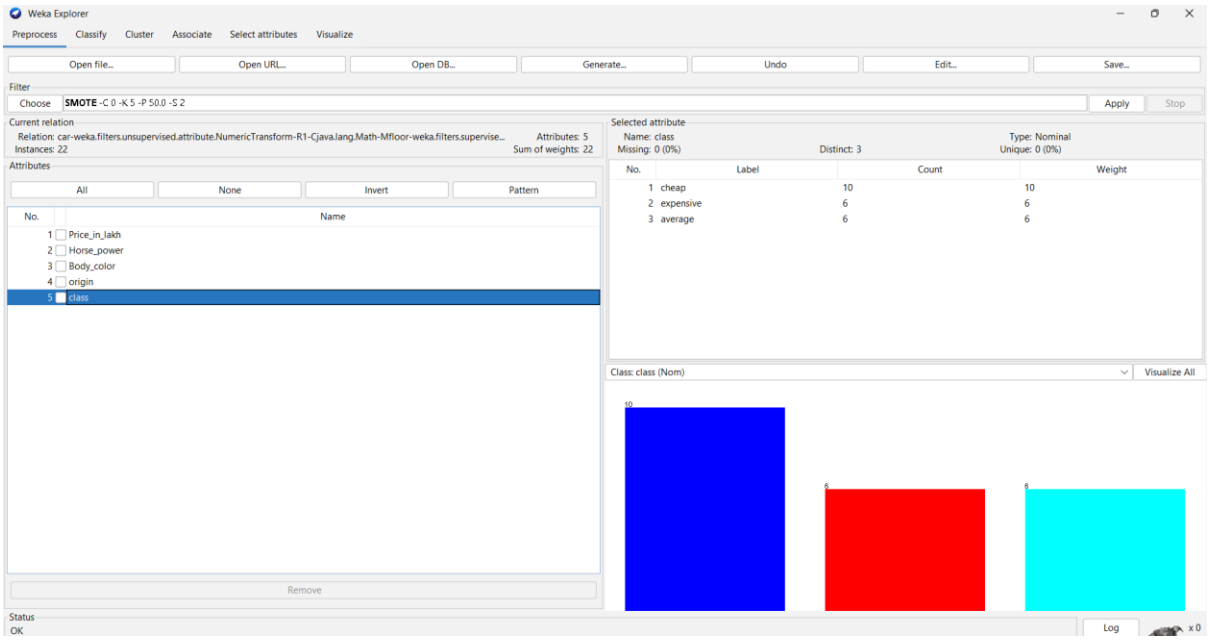
My original training dataset has imbalance classes, so I have chose SMOTE from supervised category of Filter option in the below screenshot,



Then I have changed percentage to 50% and randomSeed to 2 and kepts classValue 0 to perform oversampling on the 2 lesser classes (2nd and 3rd),



After applying SMOTE I got the following result where class 1 stays 10 but the later two classes becomes equal,



(v)Applying Classifier on the Dataset using 5-Fold Cross validation:

I have chosen Naïve Bayes and set the Cross validation fold as required. I have also added a modified Test Dataset that is compatible with the training dataset.