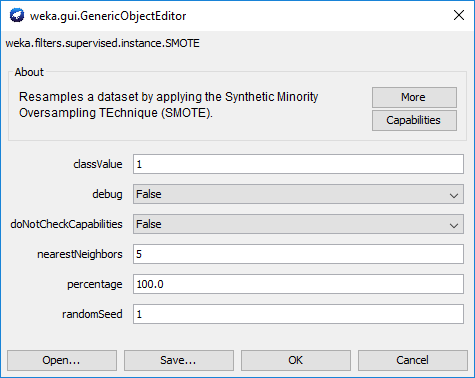
Since our dataset is unbalanced represent more of “None” decision rather than “A” and “B” if we build a model by using that data, the model will be more biased towards “None” decision and will not be able to predict A and B. In order overcome the above-mentioned scenario we have used Synthetic Minority Over-Sampling Technique (SMOTE), an approach to the construction of classifiers from imbalanced datasets. In SMOTE minority class is over-sampled by creating “synthetic” examples. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbours. Depending upon the amount of over-sampling required.



*Fig: Weka - SMOTE Option*

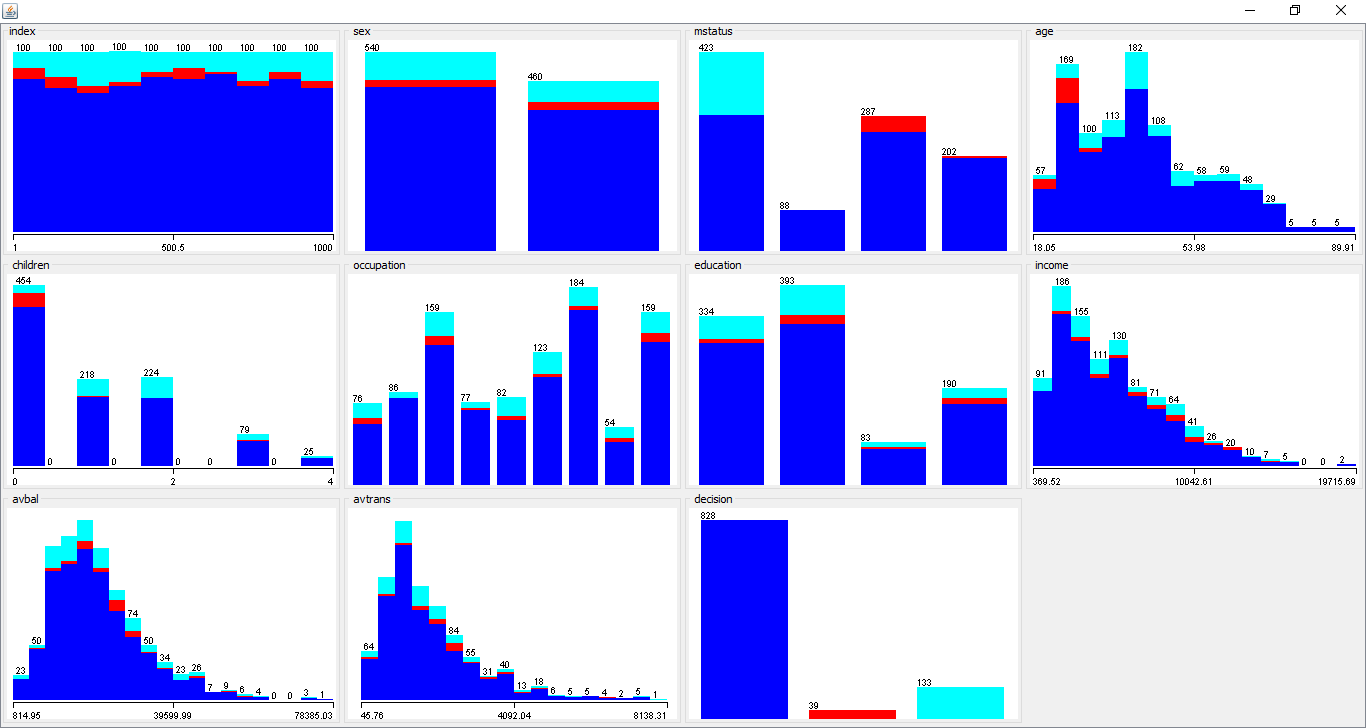
Here the classValue represent the class for which the oversampling technique needs to applied in our case 1- None, 2- A and 3 – B. Nearest neighbors option specifies number of neighbours to be considered for oversampling and the percentage option specifies how much percentage the specified class needs to be oversampled.

By applying SMOTE we have created the synthetic samples as mentioned below, neighbors=5

|  |  |  |
| --- | --- | --- |
| Decision | Before SMOTE | After SMOTE |
| None | 828 | 993 |
| A | 39 | 998 |
| B | 133 | 1064 |

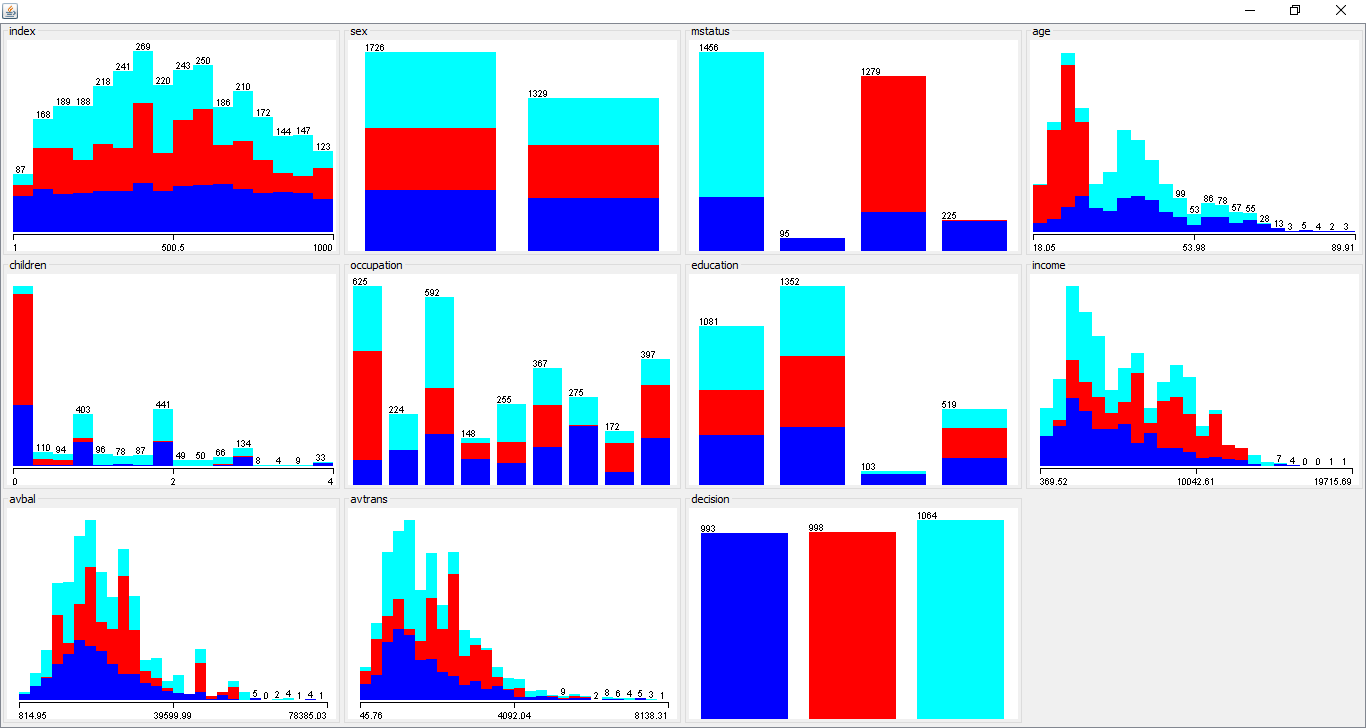
*Table: Dataset after applying SMOTE*

Before SMOTE:



*Fig: Data visualisation*

After applying SMOTE:



*Fig: Data visualisation after applying SMOTE*

References:

SMOTE: Synthetic Minority Over-Sampling Technique, Journal of Artificial Intelligence Research 16 (2002) 321–357