**Assignment – I**

Implement a desktop application by using WEKA library (C# application for WEKA.dll or Java for WEKA.jar) to obtain the suitable dataset content for each classification algorithm.

For example;

* For Naïve Bayes algorithm, the dataset must have only nominal data,
* For Logistic Regression algorithm, the dataset must have only numeric data,
* For K-Nearest Neighbour algorithm (named as IBk in WEKA), the dataset must be numeric.
* For Decision Tree algorithms (named as J48, Random Forest and Random Trees in WEKA), the data type is not important. (Don’t change anything in the dataset)
* For Artificial Neural Network algorithm (named as Multilayer Perceptron in WEKA), the dataset must be numeric
* For Support Vector Machine algorithm (named as SVM in WEKA), the dataset must be numeric

Implement the application for these algorithms.

The dataset must be selected by the user from the interface. After that, the application must give the number of correct classified instances for the best algorithm.

For the data transformation from nominal to numeric, you must use dummy attributes which you can obtain by using NominalToBinary method;

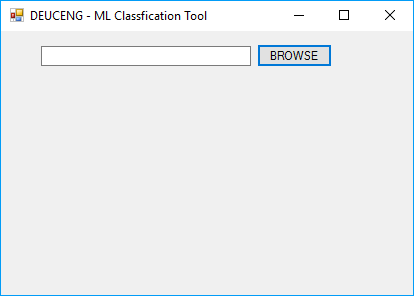
weka.filters.unsupervised.attribute.NominalToBinary.

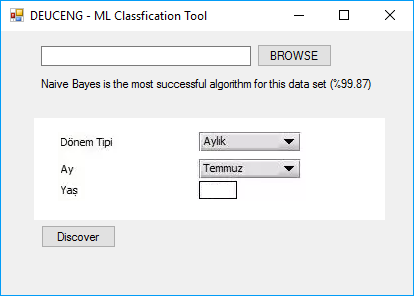
For the data transformation from numeric to nominal, you can use the discretization method in WEKA;

weka.filters.unsupervised.attribute.Discretize.

IMPORTANT:

\*Don’t forget normalization for a numeric dataset before the data mining analysis.





Create dropdownlists or/and textboxes according to nominal and numeric values. You can use grid component for this interface.

In this example dataset, there are 4 attributes as

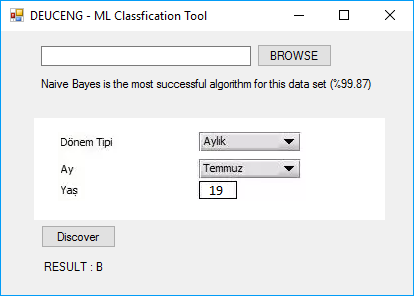
“Dönem Tipi {Günlük, Aylık, Yıllık}”

“Ay {Ocak, Subat, Mart, ....,Aralık}”

“Yaş is numeric”

Class {A, B, C}

Therefore, “Dönem Tipi” and “Ay” attributes have been created as dropdownlist with sub-fields whereas “Yaş” attribute has been created as a textbox.



After pushing “Discover”, the class of the new instance is obtained as “B” according to Naive Bayes which has been found as the most successful algorithm for this dataset previously.