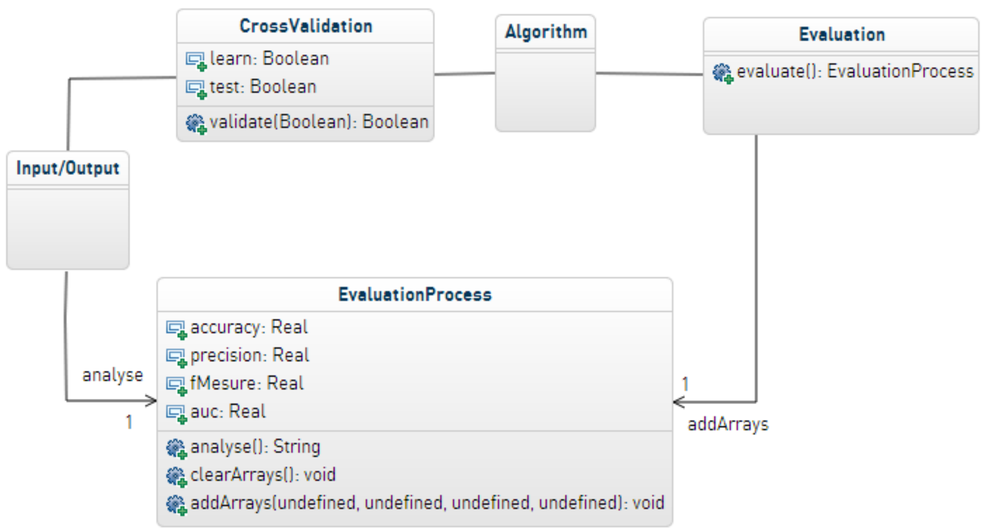
UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING
SEIP LABORATORY

Data Mining - Evaluation

Members: Sandro Babić, Matko Burul, Nikola Domazet, Samir Jugo, Nikola Lacković, Domagoj Pinčić, Filip Stojanac

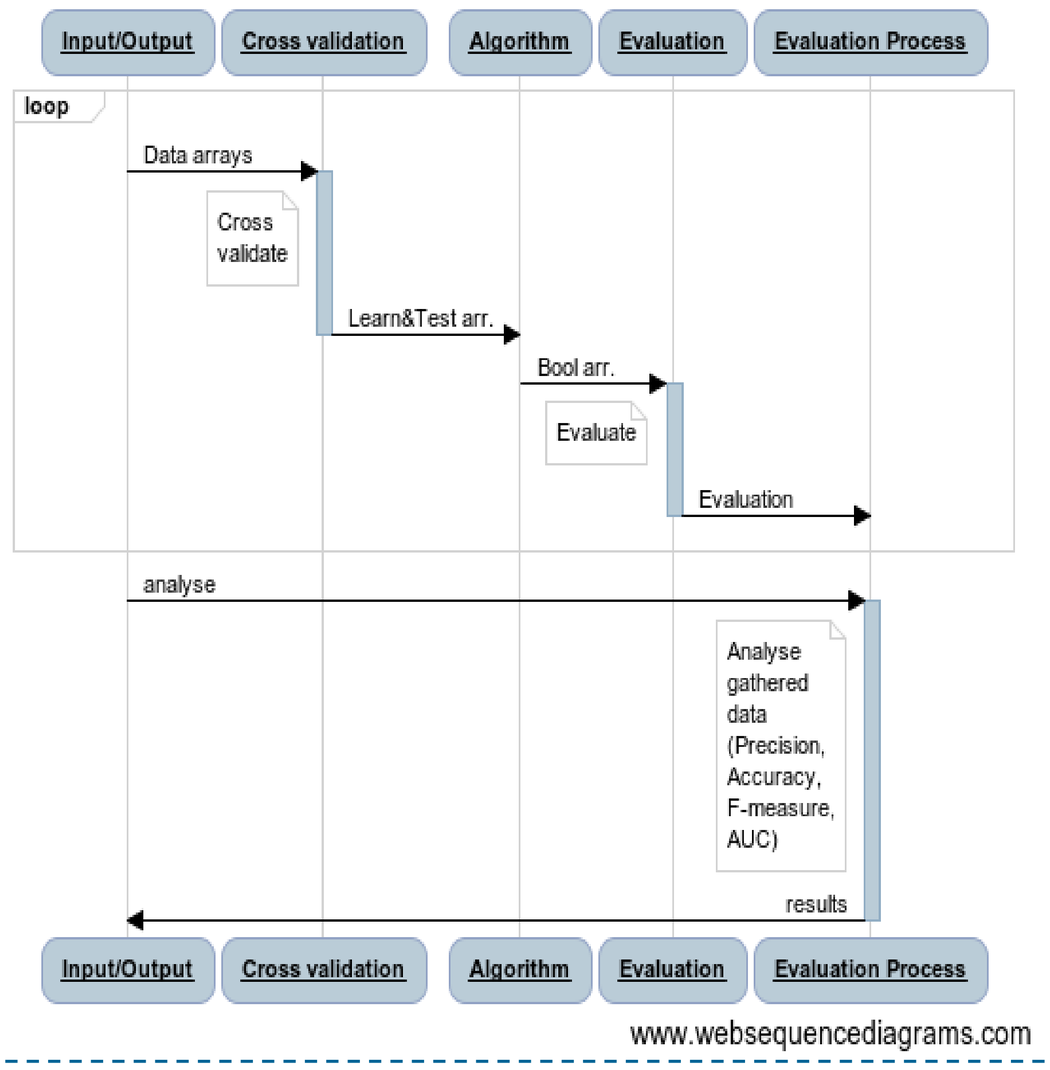
Developed Component

Class diagram



- EvaluationProcess Class
  - Receives information to be analyzed from the CrossValidation class over the addArrays method
  - On analyze method returns a string containing the results of the statistical analysis to the GUI to be shown
- CrossValidation Class
  - From input array creates two arrays – one for learning and one for testing
- Evaluation Class
  - Calculates accuracy, precision. F-measure, AUC

Sequence diagram



Summary

- System Description
  - Creation of test and learning arrays.
  - Computation of accuracy, precision, F-measure and AUC parameters.
  - Statistical analysis of differences between groups of data. The analyzed data is composed of calculated attributes. The difference between groups if different features are selected.
- Project Motivation
  - To determine if data mining alghoritms provide optimum solution.
  - A method to find out if selecting different features results in significantly different calculated attributes.

GUI

GUI for Developed Component

- Feature selection
- Result display

Conclusion

- Experienced problems
  - Integration
  - Interpretation of specifications
  - Interpretation of data implementation
- Learned Concepts
  - Statistical analysis methods and uses
  - Statistical analysis in Python