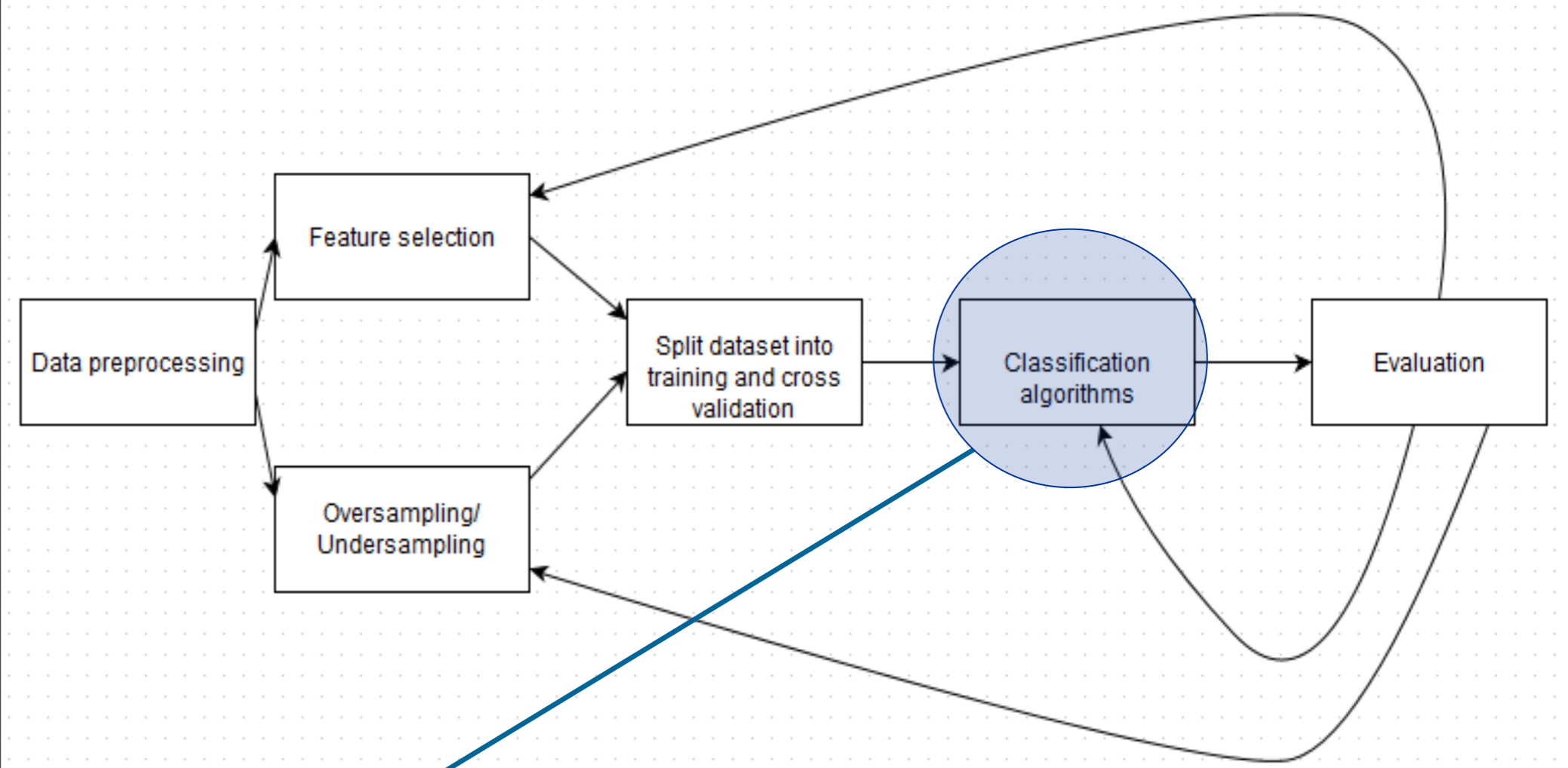


## Summary

- System Description – “DgTool classification” is a part of the DgTool data mining module for bug analysis
- Project Motivation – To automatize the process of finding faulty files
- Project Goals – To apply classification algorithms on software metrics datasets to determine whether a new file contains bugs or not
- Project Requirements – Python programming language, Jython programming language, machine learning libraries

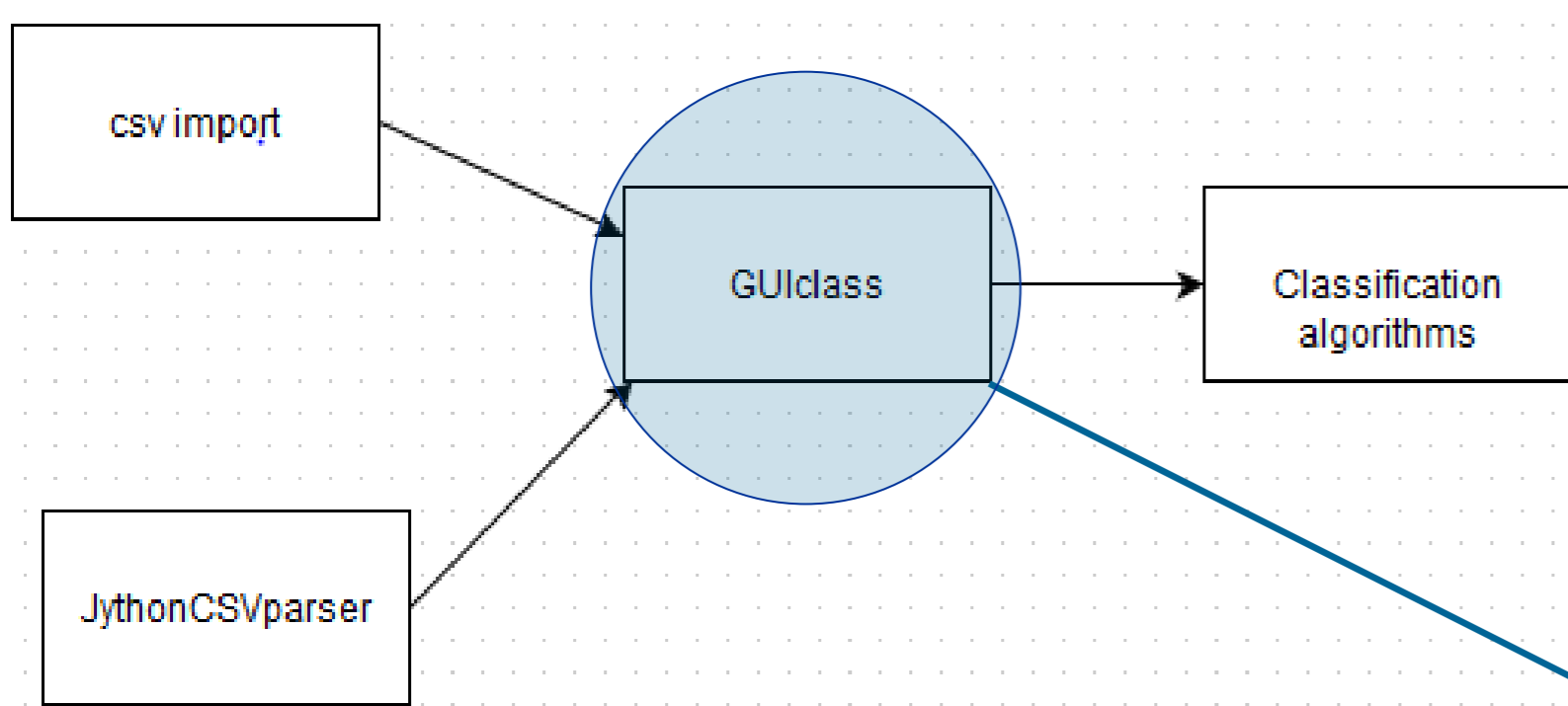
## Overall Architecture



- Architecture components – data preprocessing, oversampling/undersampling, feature selection, classification algorithms, evaluation

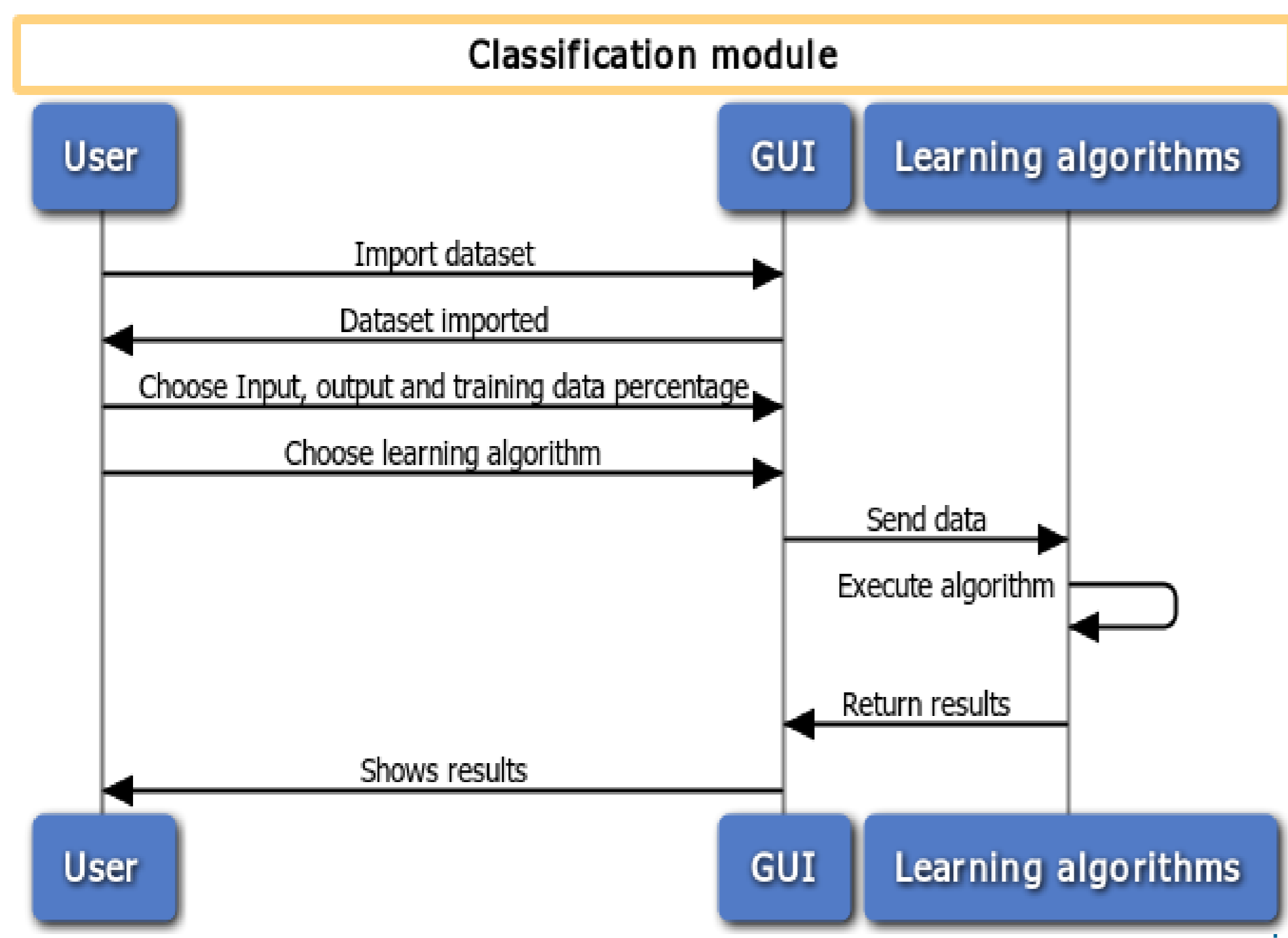
## Developed Component

### Component diagram

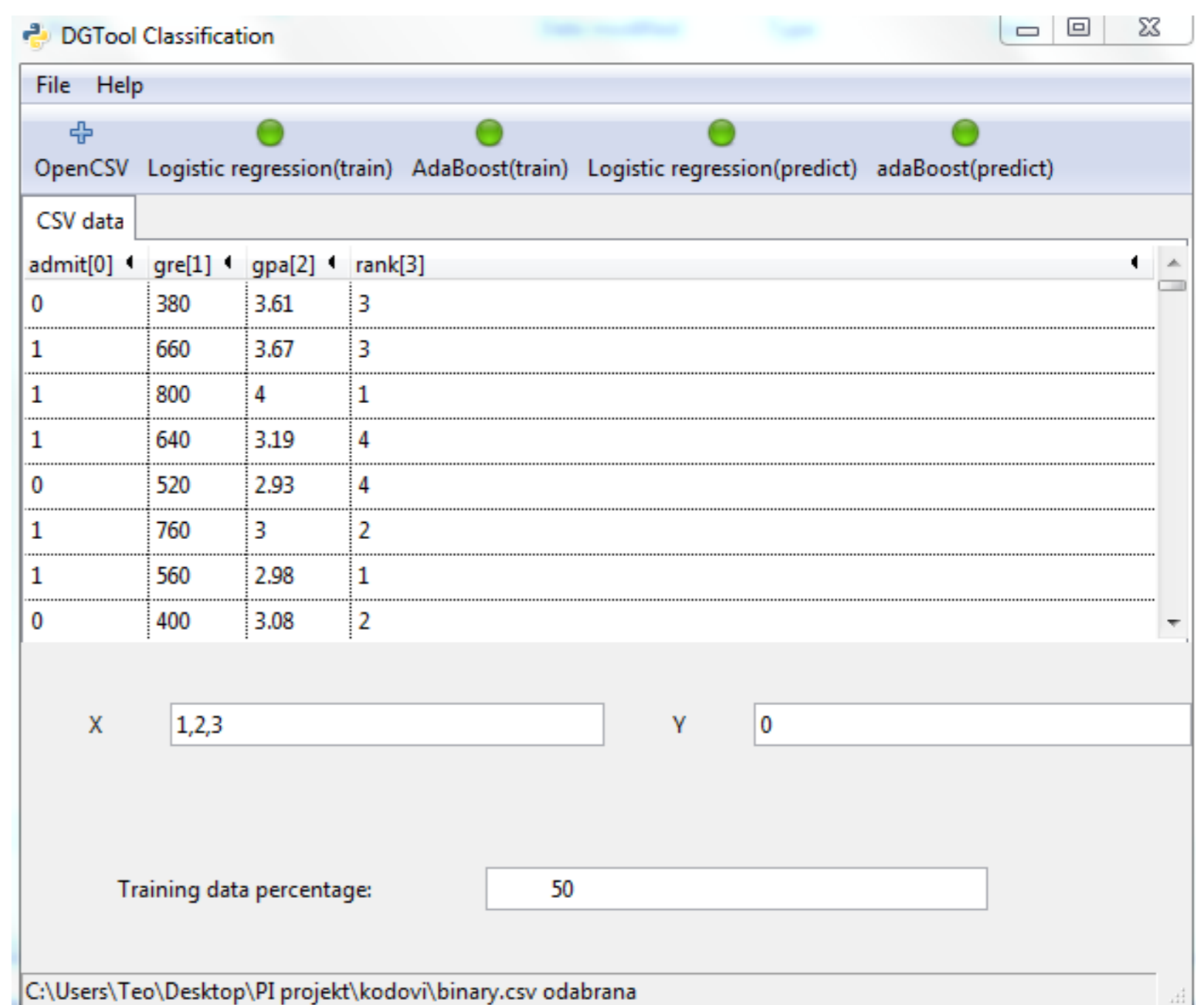


- Classification algorithms - logistic regression, adaboost, rotation forest, nearest neighbours, naive Bayes, random forest, support vector machine

### Sequence diagram



## GUI



- OpenCSV button for importing CSV file
- X – label for input columns
- Y – label for output columns
- Training data percentage label

## Conclusion

- Experienced problems
  - Data importing
  - Importing Java classes into Python
- Learned Concepts
  - Python programming language
  - Classification algorithms (logistic regression, adaboost, rotation forest, nearest neighbours, naive Bayes, random forest, support vector machine)
- Future work improvements
  - Integration