

Summative Assignment

Software Methodologies COMP2231 – Machine Learning 2019/2020

Deadline for submission: 06 March 2020 14:00 (GMT)

Background

- Learning analytics is defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs” [1].
- In this assignment you will investigate the use of two machine learning methods for a learning analytics task using the data from the OULAD [2]. You will write a scientific report detailing the procedure that you complete the task.

Task Specification

- This task aims to predict students’ final result of online courses (4 types: distinction, pass, fail, withdraw). You may decide how many types you will use.
- You will use and compare any 2 suitable machine learning methods (e.g. Decision Tree, Random Forest, Support Vector Machines, etc.) from what you have studied in lectures.

Dataset

- Description and download: https://analyse.kmi.open.ac.uk/open_dataset

Submission (2 files)

- **report.pdf** – the report with a maximum of 1,000 words, and up to 4 A4 pages.
 - Provide tables and charts to summarise and support the comparisons.
 - Discuss results and draw conclusions from your experimentation.
- **classifier.py** – the source code file.
 - Make it clear in the initial comments of your source code how to run your Python script. Your executable must run on one of the lab-based PCs – ensure compatibility before submission.

Marks (total 100)

- Working source code [15 marks]
- Clear, well documented source code [10 marks]
- Experimental procedure (data preparation, performance measurement, parameter search and selection, comparison of methods, etc.) [40 marks]
- Discussions/details of your chosen methods and the experimental procedure [10 marks]
- Evidence of the performance of your chosen methods on the data [15 marks]
- Conclusions from the task [10 marks]

N.B. marks are generally awarded for good experimental procedure that supports the result, NOT for achieving the best prediction performance.

References

1. 1st International Conference on Learning Analytics and Knowledge 2011 | Connecting the Technical, Pedagogical, and Social Dimensions of Learning Analytics, <https://tekri.athabasca.ca/analytics/>, last accessed 2019/06/21.
2. Kuzilek, J., Hlosta, M., Zdrahal, Z.: Open University Learning Analytics dataset. Sci Data. 4, 170171 (2017). <https://doi.org/10.1038/sdata.2017.171>.