A Methodology for Using GitLab for Software Engineering Learning Analytics

Julio César Cortés Ríos, Kamilla Kopec-Harding, Sukru Eraslan, Christopher Page, Robert Haines, Caroline Jay and Suzanne M. Embury

Project Goals
- To explore the way students solve tasks using collaborative development platforms, version control systems & continuous integration systems
- To find patterns and evaluation metrics that can be used to improve course content and provide insight into the most common issues the students are facing

Data Analysis Challenges
- Collecting simple usage metrics from the teaching platforms is not enough in isolation
- We need to consider the human aspects of the experience, such as gradual experimentation, learning from mistakes, and the abilities of each individual

Objective
- To construct a data processing pipeline which:
  - Combines different data sources
  - Ensures that data is collected and used ethically
  - Satisfies data protection policies
  - Minimises the risk of re-identification

Data Processing Pipeline

Acknowledgements
This work is supported by the Institute of Coding which received funding from the Office of Students (OfS) as well as support from Higher Education Funding Council for England (HEFCE).

Generalisation
- Replacement of specific values with generic ones

Permutations
- Permutation of available information

Perturbation
- Replacement of original values with different ones

Suppression
- Removal of sensitive values

Anatomisation
- Creation of groups of sensitive data based on some predefined criteria

Encryption

De-duplication

Analysis