

Data Description

Lucky Mahlangu, Mpinane Mohale, Thulisile Shipyana, Sbusiso Mkhombe

23 August 2018

The Dataset

STUDENTS DATA

- The student dataset contains approximately 5494 artificially generated students information. The information include , each student's ID, Name, Surname and the current year completed.

FIRST_YEAR DATA

- The first year dataset contains approximately 10 real first year Computer Science course codes as well as their descriptions.

SECOND_YEAR DATA

- The second year dataset contains approximately 12 real second year Computer Science course codes as well as their descriptions.

THIRD_YEAR DATA

- The third year dataset contains approximately 21 real third year Computer Science course codes as well as their descriptions.

HONOURS DATA

- The honours dataset contains approximately 14 real honours Computer Science course codes as well as their descriptions.

MASTERS DATA

- The masters dataset contains approximately 17 real master Computer Science course codes as well as their descriptions.

Variable Dictionary

The six datasets (STUDENTS.csv ,FIRST_YEAR.csv ,SECOND_YEAR.csv ,THRID_YEAR.csv ,HONOURS.csv ,MASTERS.csv) contain the following 8 variables.

1. *STUDENT_ID*[Numerical Unique code], Student Numbers.
2. *STUDENT_NAME*[String], Names of Students.
3. *STUDENT_SURNAME*[String], Surnames of Students.
4. *STUDENT_PASSWORD*[String], Students personal passwords of length 8.
5. *STUDENT_CURRENT_LEVEL*[String], Students current level of study, values taken 'YOS3' or 'HONOURS'.
6. *COURSE_CODE*[Unique String], Course code.
7. *COURSE_DESCRIPTION*[String], Description of the course.
8. *COURSE_MARK*[Numerical Value], Mark achieved by a student in a particular course.