

**ITRI626 Large Project, Phase 2**

**31715982 – Daniel Coetzee**

**30949955 – Lesetja Mojapelo**

**Drug consumption classification – Random Forest**

# Data Exploration

As already specified, the drug consumption dataset was used for this project. This dataset contains data from 1885 participants regarding their general characteristics and more importantly, different personality characteristics. These personality characteristics include different measurements such as NEO-FFI-R (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), BIS-11 (impulsivity) and ImpSS (sensation seeking). The general characteristics would be for example level of education and gender. These are set as the features for the drug consumption dataset.

Reaching the labels for this dataset, this is what gives the actual purpose of this machine learning project. There are 18 possible labels in this dataset indicating the usage of different drugs of these 1885 participants.

# Random Forest