**Assignment-1**

**By.: Szabados Dávid**

**Introduction**

While building models for prediction analysis, individuals may unconsciously use too many predictor variables, which may lead to overestimations in the actual data. Using too few predictor variables may lead to underestimations, then which one is better? During this assignment, the *cps-earnings* dataset will be used, to build four basic prediction models, from very simple ones to more complex ones. The models will be compared based on their performance and their complexity, to get a general understanding about the pros and cons of simple and complex models.

**Data & Data Cleaning**

Personal Care and Service occupations were selected from the dataset as the base of the models. Most variables, while being represented as numbers, are actually qualitative