* Got all profiles with industry 'Computer Software' and sorted the educations by date stop and date start
* Replace null values for date stop to 9999 if date start is specified, otherwise to 0. Replace null values for date start to 0
* Get the most common degrees of the last education. These are bachelor, none, master, master of business administration, bachelor of engineering, doctor of philosophy, etc:

[(u'bachelor', 5427),

(u'none', 3988),

(u'master', 3684),

(u'master of business administration', 692),

(u'bachelor of engineering', 688),

(u'doctor of philosophy', 668),

(u'bachelor of technology', 367),

(u'bachelor/master/doctor of engineering', 357),

(u'master of computer application', 311),

(u'bachelor/master/doctor/high school', 176)]

[('110', 883216), # only degree and major specified

('100', 668422), # only degree specified

('000', 450391), # nothing specified

('010', 170314), # only major specified

('001', 44949)] # only degree\_major specified

mvc,blackboard,pipelines

it must be possible to process 1 million of profiles per day

Non-functional

5.1 Performance requirements

The algorithm has to be able to predict missing skills of one profile in 1 second or less. Eventually, the algorithm has to make predictions for millions of profiles and the scraping of profile data will be a continuous process, meaning the algorithm needs to be able to keep up. Additionally, users can provide profile data themselves and the algorithm should be able to predict the missing skills in an instant.

Importance: 5

5.2 Safety requirements

Whenever the user wants to select/specify a new algorithm, data source, dump target, trained algorithm, or start an analysis process, kill an analysis process or import an analysis result, the application must query the user if he's certain he wants to continue to prevent the loss of previous data.

Importance: 2

The user is able to kill an analysis process. When the user starts an analysis process, the user can't perform any actions in the application except from killing the analysis process. If there's a problem with the analysis resulting in the process to never end, that means the user could never perform any actions. By providing the user with the option to kill an analysis process, there's a safeguard to prevent this problem.

Importance: 2

5.3 Security requirements

TBD

5.4 Software quality attributes

Adaptability. Product B1 must be adaptable in different or evolving software products that support Python 2.7.

Importance: 5

Correctness. Product B1 must be able to calculate certainty scores for every prediction with great precision.

Importance: 4

Interoperability. Product B1 and 8vance's Analysis environment as well as product B2 and product B1 must be able to exchange information with each other.

Importance: 5

Modifiability. Both products must be easily modifiable to improve the product quality.

Importance: 4

Installability. Product B2 must be installable and thus usable on any operating system.

Importance: 4

Reusability. Product B1 must be usable in both 8vance's Analysis environment as well as in product B2.

Importance: 4

Testability. Product B2 will mainly be used to test the quality and performance of an algorithm. Multiple test criteria can be established that should measure the quality and performance of an algorithm. The input and output data of an algorithm can also be tested on correctness.

Importance: 5

Learnability. Product B2 must be easy to use. Apart from algorithm developers, users without any understanding of the inner workings of the algorithm and product B2 should be able to use and understand the intended use of the product.

Importance: 1

Modularity. Both products are composed of discrete components. When one component is changed, it should have a minimal impact on the other components. This allows for more effective changes to the system.

Importance: 4

5.5 Business rules

Anyone with access to either product can use all functionality it offers.

Importance: 5