Congratulations for completing the first task successfully! For your next task you will be introduced to machine learning!

Machine learning can be explained simply as the process of finding the correlations in given data automatically. In fact, most of you have already used basic machine learning methods before, one of them being linear regression. (Remember finding the best fitting line in physics labs or linear algebra class.) Nevertheless, used methods for machine learning are not limited to linear regression. For more complicated, non-linear cases deep neural networks are used. The aim for a deep neural network is predicting the outcomes of unseen scenarios by using previous experiences. However, they are only sets of matrix multiplications, meaning they can not process any non-number values. Thus, the "experiences" must be strictly in numerical form. So, in this assignment you are requested to convert categorical information to "one-hot encoded" matrix form, and followingly, normalize the data to fit in range of 0-1. Also, do not forget to delete unrelated data (Such as id and name, which has no effect on the outcome.)

An example scenario:

Input:

Name	Age	Gender	Score
A	16	Female	95
B	21	Male	70
C	25	NonBinary	100

Pre-processed Input:

NormalizedAge	isFemale	isNonBinary	NormalizedScore
0	1	0	0.83
0.55	0	0	0
1	0	1	1

You are asked to work with the following dataset:

https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists

In your submission,

- Explanation for why normalization of data is important.
- Explanation for why one hot encoding is used instead of indexing. (why to use [1, 0,0], [0,1,0] … instead of just 1, 2, 3)
- Pre-processed input as a numpy matrix. (A pandas data frame nor an array object is not accepted!)
- Reconstruction of preprocessed input back to original. (Except for deleted columns.)
- .py or .ipynb file attached. (An explanatory notebook is a plus 😊)

are requested.

The deadline for this task is April 21, 23:59. Your next task is setting up a neural network for this dataset. If you want to make a little research beforehand, attached sources are recommended. Good Luck!

Sources:

Neural Networks:

https://www.youtube.com/watch?v=bVQUSndDllU&list=LL

Neural Networks Detailed:

https://www.youtube.com/playlist?list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

One-hot Encoding:

https://youtu.be/BecEHOVmx9o

Normalization:

https://youtu.be/UIp2CMI0748