

# 04/03/2019

## Weekly Report 1

Hello, we (Safa and Tuğba) are writing you this report in order to inform you about what we have done so far since we met with you.

### What we have done?

We have done research about Levenshtein Algorithm and its applications. As you mentioned, it can be used in order to detect similarities of texts. We have found a website [1] that contains definition and explanation of the algorithm along implementation of it in different languages. We have taken implementation of the algorithm in Python from there [2] and changed somehow that is suitable to use for our project. After that, we have tried to build a simple dataset that consists of small length simple words and tested algorithm with this dataset. We have seen that algorithm is working as it is supposed to be. Then, we started to search available datasets and we have found several datasets, but none of them was exactly suitable for our project. At the end, we had 2 datasets: 1 is consisting of words searched on Wikipedia with the search frequency of it, and the other one that consist of the average rates of the users for different social activity places across 10 different places. First dataset is ready to use for checking word similarities, but second one isn't. So, we have change second one's format and concatenate all ratings into one sentence, separated with white-spaces. All works and related codes is accessible at <https://github.com/ozkalt/Grade-Prediction-System/tree/tugba-home-dev/Levenshtein> . ".ipynb" files are intentionally put in order to allow visitors to see works have done so far and also see visual representations.

As we mentioned above, we have implemented Levenshtein Algorithm following reference page mentioned above, but it is not performing well. So, we have made researches and found that we can use dynamic programming approaches and memoize the results of the some patterns. But there is a better solution, it is already implemented on TensorFlow library. We haven't adopted function served by TensorFlow, but according to its documentation ([https://www.tensorflow.org/api\\_docs/python/tf/edit\\_distance](https://www.tensorflow.org/api_docs/python/tf/edit_distance)), it will not be hard to use it.

On the other side, we have prepared a template to ask Mr. Yaslan to use as a homework template, but haven't sent yet. We will send it today, after to send this mail to you.

That's all for now. We are searching ways to use Levenshtein Distance between words or sentences as a feature in our project.

Best,

Safa Keskin, Tuğba Özkal

## References:

- 1) [https://rosettacode.org/wiki/Levenshtein\\_distance](https://rosettacode.org/wiki/Levenshtein_distance)
- 2) [https://rosettacode.org/wiki/Levenshtein\\_distance#Python](https://rosettacode.org/wiki/Levenshtein_distance#Python)