Assignment 2

Lecturer: Prof. Moshe Sipper, TAs: Shachar Schnapp, Itai Tzruia

Spring 2023

Submission guidelines. **Please read and follow carefully**:

- Submit the assignment in groups of 2 or 3.
- Submit via Moodle.
- The submission should include a Jupyter notebook file that includes all of the experiments you've done, your final model and its results.
- For questions, use the exercise forum, or if they are not of public interest, send them via the course requests system.

Machine Learning Problem

In this exercise we solve a complete Machine Learning problem.
Given a dataset, we divide it into three parts: `train`, `dev` and `test`. The first two parts are used for experiments, and the third part will be used for evaluation (beware of data leaks!).
Given these, your goal is to produce the most accurate classifier.
Your answers for this assignment should be included in Jupyter file: `experiment.ipynb`.
You need to justify by experiments each of the steps that lead to your model, specifically focusing on:

- Show the preliminary data analysis that you did.
- Show all the different models that you tried, and explain how you chose the best one.
- Show the preprocessing that you ran on the dataset and how it affects the model's accuracy.
- Describe the metric(s) you've used to evaluate the model on the `dev` set.
- Describe the hyperparameter search and how each hyperparameter affects the model's accuracy.

Use the `train` set to train your model, and the `dev` set to evaluate it in the experiment phase.
When you finish your experiments, print the `F1 Score` on the test set.
Good luck!