**Running the project:**

1. **Setting the environment**
2. Download the project folders.
3. Download python and the following python packages:
   1. argparse
   2. Numpy
   3. Pandas
   4. Matplotlib
   5. Sklearn
   6. lightGBM
   7. EXGboost
   8. Pydotplus
   9. IPython
   10. Anaconda
   11. Graphviz
4. **Running the Algorithms:**

* In order to run model and hypothesis please run the command:

python src/Main.py --Model <a> --Test\_size <b>

where:

<a> is Model name can be one of the following: KNN, SVM, DT, RF,

Bagging, AdaBoost, GBoost, XGBoost, Voting.

<b> is test size , can be choose between 0.1 to 0.3

* For arguments details and optional values please run the command
* $ python src/Main.py -h

The output will be:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

* For example, for running Model Random Forest and test size 0.2 one should run the command:
* $ python src/Main.py --Model RF --Test\_size 0.2
* After running the above command, the model will start run.
  + First preprocessing stage in running- this is stage of removing irrelevant features that does not contribute to the model performance. (Too many features can decrease the model performance)
  + Second, our model will start training with parameters I already set.

For each model, I investigated and found the best parameters which contribute the model performance subject to time and memory constrains.

Full investigation results can be found under Hila ML Project.docx

1. **Reading Results:**

The model results will appear in the command screen.

Also 2 files will be created automatically in results folder:

* + Txt file named *"log\_<Model\_Name>\_for\_heart\_disease\_prediction.txt*"

(for example: log\_AdaBoost\_for\_heart\_disease\_prediction.txt)

This txt file contains:

* + - 1. General Information

Amount of observations and test size

* + - 1. Preprocessing -Feature Selection

The features that were removed before learning, and does not contribute to the learning process

* + - 1. Model Information

Model name and model parameters

* + - 1. Model Results

Classification Report, confusion matrix and model accuracy

* + - 1. Features Importance

Ranked from high to low

* + csv file named

*"predictedResults <Model\_Name>\_for\_heart\_disease\_prediction.csv"*

(for example: Predicted\_Results\_SVM\_for\_heart\_disease\_prediction)

This file contains the sample I took from the entire dataset (which the learning was execute on it) with one more column of 'predicted\_Class'

* + For Decision tree model png file will be created for tree visualization.

**Good Luck!**