

# DEEP LEARNING & NATURAL LANGUAGE PROCESSING

## PROPSPAL

Using the Cardiovascular Disease dataset from Kaggle (provided in this link: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset> ) we will be trying to solve a classification problem.

**The problem:**   
Gender classification based on person’s height, weight, age, whether he smokes or not.  
How well will we be able to predict if a person is a male or a female based only on his phsycial attributes?

**Divding the Dataset:**We will be splitting the dataset to 70% training and 30% testing as suggested in most common experiments of predicting a classifcation problems.

# Work Progress

תמונה שמכילה שולחן

התיאור נוצר באופן אוטומטיIn this assignment, my partner and I have tried to establish a connection between different physical attributes of a person, to predict his sex. We were using the Cardiovascular Disease dataset from Kaggle (provided in this link: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset> ).

Dataset:

* Our dataset contains 70,000 examples
* Each example holds five features: [age, gender, height, weight, smoke]

Training and Testing

* 50K examples are used as a training set
* 20K are being used for testing

**Logistic Regression Results**

**CNN Results**

**Comparison**