Day 1

Task 1. K - Nearest Neighbor(KNN) algorithm

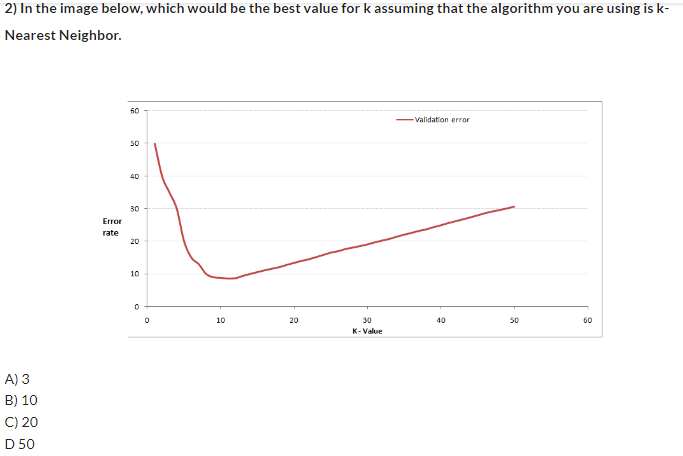
The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, Outcome.

* analyze data and detect anomaly
* create classifier using KNN algorithm
* find best K
* view confusion matrix for predicted data and think about results

Task 2. Quiz about KNN algorithm

**1.** Which statement is true about KNN ?

* k is usually is an even number
* We can only use Euclidean distance for KNN
* All statements all false
* knn is a linear model



**3.** Which of the following distance metrics can not be used in k-NN?

* Manhattan
* Euclidean
* Minkowski
* All can be used

**4.**Which of the following distance measure do we use in case of categorical data in KNN

* Manhattan
* Euclidean
* Hamming

**5.** Which of the following options is true about the k-NN algorithm?

* it can be used for classification
* it can be used for regression
* it can be used in both