

Lab 5- Hagay Ringel

One of my hobbies is climbing. From a young age I have been involved in climbing of all kinds, mountaineering, indoor climbing, etc. Therefore, when I searched for interesting databases in Kaggle and came across the database on climbers I had no doubt that I would do the research on this topic. This database contains data on the leading climbers in the world, from the country of origin, the sex of the climber, age and height / weight to numerical data on how many years of climbing they have, the points accumulated in the cumulative and average. The variable I chose to analyze is whether, based on data of height, weight and climbing years, it is possible to predict the sex of the climber. If it is indeed possible to predict the sex of the climber based on these data, we can conclude that there is a real difference between men and women in relation to the popularity of this industry among the two sexes. Another factor I would like to add is that I know that in this database there are 7.5 times more men than women and it's probably what affect the results in the classification report as below. Below we can see the results according to the model by using to algorithms: K-Nearest Neighbors (KNN) and Naive Bayes. In the visualization graphs I presented to trends: taller climbers are easier to climb- we can see that by their grades. And second, obviously, as more experience you have- .your grades mean will be higher

Accuracy- In KNN the average accuracy after performing 5- fold cross validation is 0.91 (91%) which mean a high value of accuracy, while in GNB the average is a bit higher- 0.92. in both .algorithms the accuracy is high, and this is good for the model

As for the classification report, I would compare the results of both algorithms to see whether .we have a significant difference and to check the difference between the males and females

Precision- in both algorithms the precision of the males is very high with 95%/96% chances that if we predict that the gender of the climber is male, we will be right, while there is only .65%/66& chances to be right if we predict that the climber is a female

Recall- in both algorithms there is a significant difference between the genders. In KNN 95% of the cases the classifier will predict yes for our assuming that the gender is male, while in only 62% of the cases that we predict the gender is female, the classifier will predict yes. In GNB the number under male is 94% (similarly to KNN) but under female is 73%- an increase of 10% .compared to the KNN report

f1 score- In both KNN and GNB the average of the precision and recall for male is 95% while for female is 64% in KNN and 68% in GNB. Since the model for male has a good rate for precision and recall, we can see that the score is high compared to the female which was low in both .precision and recall therefore will be poor performance in the f1 score as well