

Lab Day 3

If you have not done so already, download JASP to your Computer:

<https://jasp-stats.org/download/>

1. Load your dataset. If you do not work out how to do this by yourself, look at this short tutorial video: <https://www.youtube.com/watch?v=GEuBad0DhEc>
2. Make yourself familiar with the user interface of JASP. What do you see? What does JASP offer beyond machine learning?
3. First of all, do some exploratory analysis of your data. Use the *Descriptives* menu point to get some idea what JASP can do for you in exploratory data analysis. F
 - a. For instance, calculate a frequency table for all your predictors and your outcome variable. Also use the *Plots* submenu to make JASP draw some relevant plots for you.
 - b. Also produce correlation plots for all your numerical predictors variables. Looking at collinearity in predictors in machine learning can also help identify relevant prediction variables.
 - c. Play around with the *Descriptives* a little more. Press any buttons you want a see what happens.
4. Klick on the blue + in the top right and add the machine learning module. Now you have a new menu point in the top bar. Klick it an see what algorithms it has in store. Here is an older blog entry that tells you what you can do with the JASP machine learning module: <https://jasp-stats.org/2019/09/24/introducing-jasp-0-11-the-machine-learning-module/>
5. Think about some variable you would want to predict. At best this is a binary variable like we used it during the last two days. However, if you want to be a bit more advanced and classify a multinominal variable, you are free to do so.
6. Before you freely choose a classification algorithm, use the k-nearest neighbour algorithm to classify your outcome variable. Choose any number of relevant (and sensible) variables as predictors. If you need some help look at the following blog article and apply it to your data: <https://jasp-stats.org/2019/10/07/how-to-train-a-machine-learning-model-in-jasp-classification/>
 - a. Change the train-test split from 80-20 to 75-15.
 - b. Produce a confusion matrix.
 - c. Plot the data split.
 - d. Plot a ROC curve.
 - e. Plot a decision boundary matrix.
 - f. Do a 10-fold cross validation.
7. Now try out the Support Vector Machine, Random Forest and Neural Network algorithms to tackle your problem. Produce the same output as in 6 for all of them.