Lab1 Answers

1. **What are the relevant features of the Titanic dataset? Why are they relevant?**  
   Depending on the use case, different features are relevant. For the use case in the tutorial, “dead or not”, the relevant features are: Pclass, SibSp, Sex, Parch, Age.
2. **Can you find a parameter configuration to get a validation score greater than 62%?**No, we did not manage to find such configuration.
3. **What are the advantages/disadvantages of K-Means clustering?**Advantages is that it is easy to implement, scales to large datasets, guarantees convergence and is easy to adapt to new datasets.  
   One disadvantage is that it takes good knowledge of the dataset (because it is a supervised method). Another on is that the method is sensitive to outliers and the order of the data.
4. **How can you address the weaknesses?**One big weakness of K-means is that it does not handle missing values well. To solve for this, we filled the empty columns with the mean values. Additionally, for the model to perform better we selected only the relevant features.