Content of the report (later)

* + Your pipeline: from retrieving data to the actual model
  + Introduction and problem statement
  + Research method & scientific character of the work done
  + Argumentation
  + Results: discussion/ interpretation
  + General conclusion
  + Coherence/ logical composition
  + Originality & creativity
  + References

First, instead of the initially provided datasets, covering a relatively short period (2013-2015) of the Eurovision scores, it was decided to use a dataset with larger number of datapoints, i.e. from 1975 to 2019, considering the nature of the analysis and given that typically, in the context of classification, for instance, large datasets lead to better performance and small datasets may trigger overfitting (Althnian, 2021). Additional variables of consideration and possible influences, such as neighboring countries, time zones, official languages, religion, etc. were added; data preprocessing was performed to work on a coherent data, under the acceptable units.

Second, in order to get an overview of the data at hand, we perform EDA: e.g., countries which were awarded the highest and the lowest scores (individually and in respective pairs) most often, an average point for each country, evolution of the mean score over the years for highest scoring countries overall, as well as when partitioning into finals and semi-final rounds, jury and televoting. Countries with bilateral relationship, such as Greece and Cyprus, Romania and Moldova consistently vote for each other in a reciprocal way, countries, constituting former Yugoslavia, Montenegro-Slovenia, Slovenia-Croatia, also tend to give each other high votes. Geographically, Scandinavian countries, Baltic countries, Southern European countries follow the same trend.

References:

Althnian, A.; AlSaeed, D.; Al-Baity, H.; Samha, A.; Dris, A.B.; Alzakari, N.; Abou Elwafa, A.; Kurdi, H. Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Appl. Sci.* **2021**, *11*, 796. https://doi.org/10.3390/app11020796