Project: Which type of song is the most popular: happy, sad or angry? (mood classification of song lyrics from Billboard Hot 100)

KAROLINA:

1. Modeling:

* adding more sophisticated feature selection (e.g. t-SNE,LSI)
* adding more information about different ML models
* adding a more sophisticated ML model (e.g. SVM, XGBoost) or optimizing traditional

1. Results:

* Table with the performance of each model
* Answer to the research question (which mood had a higher amount of nr 1 songs, which mood was nr 1 for more weeks)

1. Conclusion:

* Was the purpose achieved/Were the assumptions examined?
* Formulate conclusions from obtained results

1. Appendix with source code and data set

Data sources:

* <https://www.kaggle.com/datasets/imuhammad/audio-features-and-lyrics-of-spotify-songs>
* <https://www.kaggle.com/datasets/danield2255/data-on-songs-from-billboard-19992019/>

Code sources:

* <https://medium.com/analytics-vidhya/lyric-mood-identifier-c8e94a49c73>
* <https://www.quantargo.com/help/r/latest/packages/gofastr/NEWS>
* <https://rpubs.com/riazakhan94/naive_bayes_classifier_e1071>
* <https://www.projectpro.io/recipes/create-and-optimize-baseline-decision-tree-model-for-multiclass-classification-r>
* <https://rpubs.com/jkylearmstrong/RF_Imputation_Multi_class>
* <https://www.kaggle.com/code/rtatman/machine-learning-with-xgboost-in-r>