Practice the following rules for writing:

* **Combat wordiness.**
* **Do not express more than one or two ideas in a single sentence.**
* **Start sentences with old information, and end with new information.**

**Exercise: Abstract (thesis proposal)**

This study investigates if the penalized regression methods of relaxed lasso, adaptive lasso or their combination can improve the performance of prediction rule ensembles in terms of model sparsity, prediction accuracy, and stability. In prediction rule ensembles decision trees are built with gradient boosting and a subset of terminal nodes and trees are selected using penalized regression. For the latter step this study compares the performance of ordinary, relaxed, adaptive, and relaxed adaptive lasso regression. In adaptive lasso the regression coefficients in the penalty term are weighted while in relaxed lasso the penalty term lambda is multiplied by a second shrinkage parameter The goal of prediction rule ensembles is to come up with a few sets of rules, based on which interpretable decisions can be made with high accuracy. For the model evaluation, the analyses will be conducted on around 8 different dataset from the area of psychology.

**Exercise: Abstract (thesis)**

Salmon fishing is a major industry in Scotland. At the moment, there are regulations in place to determine whether the farming is not harmful for the macrobenthic organisms subsiding in the oceans. However, to determine the effects on the benthos, samples have to be collected and analysed which may take up to 6 months. Therefore, this research looks into the metabarcoding of environmental DNA as a tool to calculate the benthic status, indicated by the infaunal quality index (IQI). Currently, 228 samples have been collected and a Random Forest algorithm is being trained to predict the derived IQI values based on eDNA. This will be accomplished by using measured features at the sites of the samples. In this project, the main focus will lay on the optimization of this algorithm. Additionally, an R-package will be developed which collects the analysis methods. So, it can be used on future samples of the salmon farming industry.

**Exercise: Introduction paragraph**

To overcome the need for manual verification, MRDM created the automatic data verification which is a scalable AI-driven data verification methodology that can be executed remotely without the need of going on site to verify data. With this methodology, it is possible to verify considerably more data points in just a fraction of the time compared to manual data verification; which makes it feasible to check the entire population of patients in all healthcare institutions instead of just a sample of them.

**Exercise: Introduction paragraph**

However, a drawback of lasso is that it leads to inconsistent results when multicollinearity is present (Zhao & Yu, 2006). Zou (2006) proposes the adaptive lasso, where the coefficients in the penalty term are weighted. By imposing larger weights on relevant base learners, the estimations will become more accurate and consistent. Another method by Meinshausen (2007) is the relaxed lasso where a relaxation parameter φ is introduced to control the shrinkage. The relaxed lasso has a fast convergence rate and leads to more parsimonious models compared to the ordinary lasso (Meinshausen, 2007). A rather novel model is the relaxed adaptive lasso, which combines the two methods (Zhang et al., 2022). Zhang et al. (2022) found higher prediction accuracy compared to the relaxed or adaptive lasso.