The ML model used to solve task1b is ridge regression. In order to figure out an ideal value for the hyperparameter lambda, cross validation with different combinations of number of folds and repetitions for repeated k-fold was run. The lambda with the lowest RMSE was then chosen for model training. Hereby, different ratios of train/test splits were used to analyse the model’s performance.

Values for lambda were varied between 0.1 and 400. The best performances with regards to cross validation could be achieved using lambda=10. Whereas a large number of repeats for RKF reduced the public score in task1a, this was, however, not the case for this problem. RKF was run between 4 and 1,000 times with no significant difference in RMSEs for cross validation results. Similar public scores, which could be observed running the code against the real waits on the project page, confirmed this outcome. A variation of the number of folds from 8 to 14 yielded almost identical public scores (~2.065) as well.

As far as train test splits are concerned, it is to say that the best results were achieved with splits between 70/30 to 80/20. The code was also submitted with a model trained on the whole training data set, i.e. without a train test split, which led to a slightly higher public score (~2.069).