**Data mining project proposal by Gijs van Paridon (s4786858) & Marco Post (1046670)**

**Option: Problem**

**Chosen dataset:** [**https://www.kaggle.com/c/competitive-data-science-predict-future-sales**](https://www.kaggle.com/c/competitive-data-science-predict-future-sales)

**Software:** [**https://scikit-learn.org/stable/supervised\_learning.html#supervised-learning**](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning) **-** [**https://keras.io/**](https://keras.io/)

**Literature:** A. Krishna, A. V, A. Aich and C. Hegde, "Sales-forecasting of Retail Stores using Machine Learning Techniques," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 160-166, doi: 10.1109/CSITSS.2018.8768765.

Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales," 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, 2016, pp. 531-537, doi: 10.1109/ICDMW.2016.0082.

The appended dataset consist of store data and their sales. Using machine learning and previously learned data optimization techniques, we will try to predict future sales for the given data. The database has data of different product that the stores sell. While the descriptions are mostly in Russian (the data was delivered by a Russian company), all the products are corresponded with a unique ID number. The database is already divided in train and test data for us to design our program on.

The software we intend to use is mostly provided by the skicit module of Python and the keras application. Both modules provide a variety in algorithms for machine learning and regression of data. We would like to use regression models such as linear regression or neural networks to try and predict future sales. After that we want to research whether we can optimize the hyper parameters for these models to get more accurate models.

The papers in the literature are similar to our project. Datasets consist of (simplified) stores using deep learning techniques to predict future sales. Though the used methods are more mathematical and less focused on the software that we provide, the papers can still be seen as a good inspiration. One contains an architectural diagram with some of the steps we have to follow.