Project proposal "Algorithm" 2223 Data Mining (KW1 V)
By: Niek Sülter (s110739) and Lean Schoonveld (s1108097)

Boosting is an ensemble method to improve the performance of 'weak' machine learning algorithms. The weak classifiers like decision trees are combined to improve the correlation of the classifier with the actual classes. For our project we would like to implement a tree-based boosting algorithm with the gradient boosting improvement. Opposed to traditional boosting algorithms which use weights to convey the accuracy of a tree. Gradient boosting algorithms try to optimise a loss function of sequentially generated trees to find the 'best' tree. For our dataset we would like to use a dataset that classifies EEG readings (*Temple University EEG Corpus*, n.d.) to determine if a patient has suffered a stroke or not. (Email for data sent no repose yet.) As a starting point we would use the original publication of the gradient boosting proposal (Friedman, 2001) and the sklearn implementation (*Scikit-Learn/Scikit-Learn*, n.d.).Our implementation would be in Python as well. As for the data, it needs a little pre-processing but it has been pre-classified. Several other publications have been made using this data while implementing similar classifiers with different machine learning algorithms like SVM.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Https://Doi.Org/10.1214/Aos/1013203451*, *29*(5), 1189–1232. https://doi.org/10.1214/AOS/1013203451

scikit-learn/scikit-learn. (n.d.). Retrieved November 14, 2022, from https://github.com/scikit-learn/scikit-learn/blob/f3f51f9b6/sklearn/ensemble/_gb.py#L991

Temple University EEG Corpus. (n.d.). Retrieved November 14, 2022, from https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml