The goal of this research is to employ machine learning techniques to improve the quality of market fluctuation (i.e., volatility) forecasts. In order to achieve such result we developed the system depicted here. The purpose of our research is twofold:

- First, we aim to perform a statistical assessment of the relationships among the most used proxies in the volatility literature.
- Second, we explore a NARX (Nonlinear Autoregressive with eXogenous input) approach to estimate multiple steps of the output given the past output and input measurements, where the output and the input are two different proxies.

This setup includes two volatility proxies simultaneously. Two ML techniques (Artificial Neural Networks and K-Nearest Neighbors) are employed in the forecaster. The purpose is to investigate how the combination of proxies affects the forecasting performance of the algorithms.

The figure shows the aggregated correlation (overall the 40 time series) between the proxies, obtained by meta-analysis [4]. The black rectangles indicate the results of an hierarchical clustering using [3] with k=3. We observed that:

- As expected, a correlation clustering phenomenon exists between proxies belonging to the same family, i.e. $\sigma_i$ and $\sigma_{i|N^D}^\alpha$.
- The presence of $\sigma_i$ in the $\sigma_{i|N^D}^\alpha$ cluster can be explained by the fact that the former represents a degenerate case of the latter when $n=1$.
- A significant correlation between the volume and the $\sigma_i$ family.

### References


