Figure 1 displays the results of running each algorithm on a set of different datasets. The figure shows clear dominance of the evolutionary model all the models across all datasets.
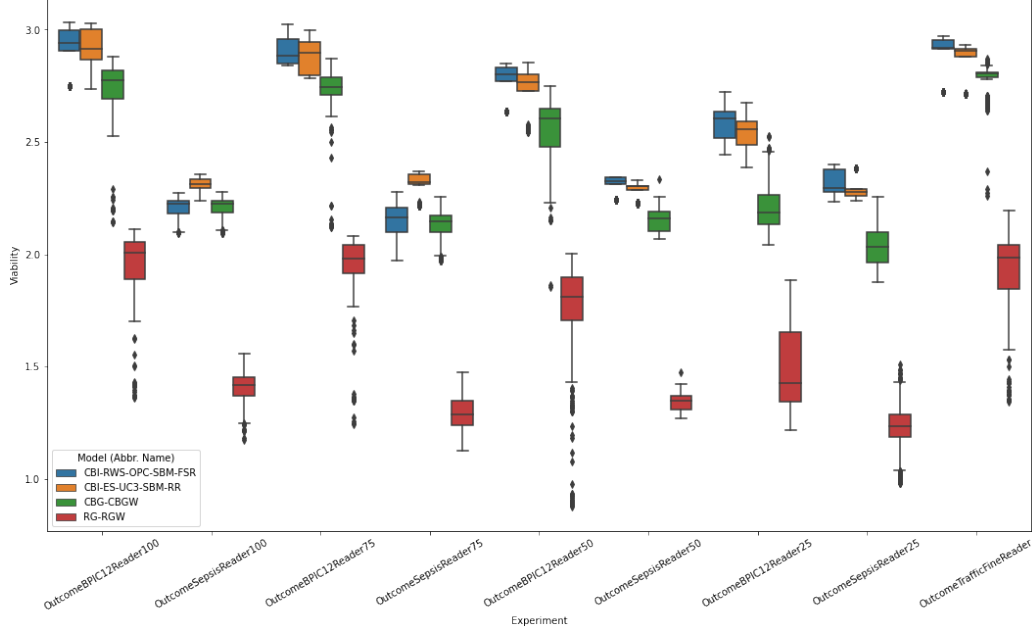

Figure 1: Boxplots of the viability of each models' generated counterfactuals across a herterogeneous collection of datasets.

Here, *CBI-ES-UC3-SBM-RR* and *CBI-RWS-OPC-SBM-FSR* display a higher median of viability across all datasets. This is unsurprising as the evolutionary algorithm use inititiators that are based on the baselines. However, it is surprising that the evolutionary models consistently outperform the Casebased-Search Generator (green) across all datasets. In 6 out of 9 datasets we see an improvement of at least 0.15. From **??** we see that the gap often occurs because of much higher similarity and sparcity scores. The highest median is reached for *CBI-RWS-OPC-SBM-FSR* at 2.94. The Random-Search Generator never manages to come even close to the case based model. Except for the BPIC12-100 dataset, the Random-Search Generator has a median below 2.