**Table S3**. Best models for all training sets

|  |  |  |
| --- | --- | --- |
| ***Training Set*** | ***Model name*** | ***Count*** |
| Anti-infective/hetero | meta.AdditiveRegression | 1 |
| Anti-infective/hetero | lazy.LWL | 1 |
| Anti-infective/hetero | trees.RandomForest | 6 |
| Anti-infective/homnpqc | meta.AdditiveRegression | 1 |
| Anti-infective/homonpqc | lazy.LWL | 1 |
| Anti-infective/homonpqc | meta.RandomCommittee | 1 |
| Anti-infective/homonpqc | functions.SMO | 3 |
| Anti-infective/homonpqc | functions.GaussianProcesses | 2 |
| Anti-infective/homopep | lazy.KStar | 1 |
| Anti-infective/homopep | meta.RandomCommittee | 1 |
| Anti-infective/homopep | trees.RandomForest | 6 |
|  |  |  |
| Anti-gut1/hetero | meta.RandomCommittee | 1 |
| Anti-gut1/hetero | lazy.LWL | 1 |
| Anti-gut1/hetero | trees.RandomForest | 6 |
| Anti-gut1/homnpqc | meta.AdditiveRegression | 2 |
| Anti-gut1/homonpqc | functions.MultilayerPerceptron | 2 |
| Anti-gut1/homonpqc | meta.RandomSubSpace | 1 |
| Anti-gut1/homonpqc | meta.Vote | 1 |
| Anti-gut1/homonpqc | lazy.LWL | 1 |
| Anti-gut1/homonpqc | lazy.IBk | 1 |
| Anti-gut1/homopep | trees.RandomForest | 8 |
|  |  |  |
| Anti-gut4/hetero | meta.AdditiveRegression | 1 |
| Anti-gut4/hetero | meta.AdaBoostM1 | 1 |
| Anti-gut4/hetero | meta.RandomCommittee | 1 |
| Anti-gut4/hetero | trees.RandomForest | 5 |
| Anti-gut4/homnpqc | meta.AdditiveRegression | 1 |
| Anti-gut4/homonpqc | lazy.KStar | 1 |
| Anti-gut4/homonpqc | functions.SMO | 2 |
| Anti-gut4/homonpqc | meta.AdaBoostM1 | 2 |
| Anti-gut4/homonpqc | trees.RandomForest | 1 |
| Anti-gut4/homonpqc | functions.GaussianProcesses | 1 |
| Anti-gut4/homopep | lazy.KStar | 1 |
| Anti-gut4/homopep | trees.RandomForest | 7 |

The complete set of parameters for every model listed in the table is available as supplementary files S43 (anti-infective), S44 (anti-gut1) and S45 (anti-gut4). Count presents the number of times the listed model was found. The model name is derived from the weka.classifiers.