# Example Submission Final Project AutoML Lecture WS 2022/2023

Constantin von Crailsheim<sup>1</sup>

<sup>1</sup>LMU Munich, Institute of Statistics

**Abstract** 

#### 1 Introduction

The AutoML system should be able to deal with missing values and yield good performance as measured by the balanced accuracy on imbalanced datasets. Thus, it optimizes a ML pipeline consisting of five elements. First, a choice of two imputers replaces missing values. Then, an optional sampling method is applied to deal with imbalance in the targets, which consists of both under- and oversampling methods. Next, the pre-processed integer features are rounded such that they do not take values that would not appear in the training data. After optionally normalizing all features, three different models were fitted, i.e., a random forest classifier, a gradient boosting classifier and a SVM classifier. The choice of pre-processing and the hyperparameters of the model were optimized by DEHB, which is a computationally cheap optimizer that works well on discrete search spaces. The final predictions are derived by majority voting of the individual predictions of the best pipelines for each model. This means that the overall predictions should be good of the errors of the model are not too correlated.

#### 2 Method

This section outlines the specific choices of the AutoML system and how it will be optimized and evaluated. The choices of the imputation strategy of sklearn.impute are:

- The SimpleImputer is an univariate imputer, which completes missing values with a descriptive statistic per feature. I chose the median, since it is less sensitive to outliers than the mean.
- The KNNImputer replaces missing values by the mean value of its 5 nearest neighbors as default based on Euclidian distance of their non-missing observations of the same feature.

The choices of data-level sampling method to yield balanced dataset of imblearn are:

- SMOTE as an oversampling approach generates new samples of the minority class by interpolating between existing observations of the minority class, where no distinction is made between easy and hard samples.
- TomekLinks as an undersampling approach removes samples from the majority class, if they
  are nearest neighbors to a minority class sample, thus removing noisy borderline samples of the
  majority class.
- SMOTETomek, which combines SMOTE and Tomek links.
- No sampling method, which would allow algorithmic-level methods to deal with the imbalanced data.

AutoML Lecture 2023

The above pre-processing methods could generate numeric values for features that should contain only integers (e.g., ordinal categorical features), since they impute missing values by means or medians and generate new samples by interpolation. Thus, another layer is added to the pipeline, which rounds all observations of those features to an integer.

The last step of pre-processing is the choice of whether to apply the <u>StandardScaler</u> of sklearn.preprocessing to standardize the features. This will be in particular useful for the SVM, since an RBF kernel assumes features centered around zero and similar variance across features.

Subsequently, the hyperparameters of the three models of the ensemble were optimized, where the search space was defined with the ConfigSpace package. For almost all hyperparameters, the default was set to the default specified for each model. The search space was mostly chosen such that it is centered around the default while accounting for log scale. In those cases where the default would be a the lower end of a reasonable search space, the upper bound was chosen higher.

The first model in the ensemble is a RandomForestClassifier from sklearn.ensemble. It can handle all data types well and generalizes well by having a low variance due to ensembling over relatively uncorrelated models. The hyperparameters, which will all be sampled uniformly, are:

Hyperparameter	Data type	Search space	Default	Other
criterion	Categorical	{Gini, Entropy, Log loss}	Gini	
max_depth	Integer	[5,15]	10	
min_samples_split	Integer	[1, 32]	2	Log scale
min_samples_leaf	Integer	[1, 16]	1	Log scale
max_features	Integer	[0.1, 0.9]	0.5	
class_weight	Categorical	{Balanced, Balanced subsample, None}	None	

The class\_weight is an algorithm-level method that deals with imbalanced data and will only have an effect if no data-level sampling method was used, since in that case the dataset passed to the model will be balanced.

The second model in the stack is a <u>GradientBoostingClassifier</u> from sklearn.ensemble. It has strong predictive performance by iteratively fitting weak learners on the error of the previous learner and has similar advantages as the RandomForestClassifier. The hyperparameters are:

Hyperparameter	Data type	Search space	Default	Other
loss	Categorical	{Log loss, Exponential}	Log loss	
learning_rate	Float	[0.01, 1]	0.1	Log scale
criterion	Categorical	{Friedman MSE, Squared error}	Friedman MSE	
min_samples_split	Integer	[1, 32]	2	Log scale
min_samples_leaf	Integer	[1, 16]	1	Log scale
max_depth	Integer	[2,10]	3	

The third model in the stack is a Support Vector Classifier (SVC) from sklearn.svm. This model works in particular well on easily to separate datasets and in high-dimensional spaces <sup>1</sup>. The hyperparameters are:

<sup>&</sup>lt;sup>1</sup>https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107

Hyperparameter	Data type	Search space	Default	Other
С	Float	[0.1, 10]	1.0	Log scale
kernel	Categorical	{Linear, Polynomial, RBF, Sigmoid}	RBF	
shrinking	Boolean	{True, False}	True	
tol	Float	[1e-5,1e-3]	1e-3	
class_weight	Categorical	{Balanced, None}	None	

To optimize the hyperparameter of the AutoML system, DEHB by Awad et al. (2021) was used, which combines Differential Evolution and Hyperband. Differential evolution generates a new mutant vector from three random parents and then generates the offspring by randomly selecting values from the new mutant vector with probability p and from on of the corresponding parents otherwise. Hyperband allows to search the whole search space with cheap evaluations and only train more expensive models on promising areas of the search space. The algorithm starts by sampling  $N = \eta^{f-1}$  random hyperparameter configurations, which are evaluated at the lowest budget. Then the best  $1/\eta$  of the configurations are evaluated at a  $\eta$  times as higher fidelity and this process is repeated until the highest fidelity (denoted here by f) is reached. After completing one iteration, the algorithm restarts with new instantiations and evaluates these at the second lowest fidelity. DEHB combines both approaches by generating the configurations for the next fidelity by differential evolution from the lower fidelity as parent pool, where the previous evaluations indicate promising regions. The authors state that DEHB is computationally cheap with high speed-up gains compared to BOHB. Furthermore, it has strong final performance for discrete search spaces, which we have for various hyperparameter. Their experiments have shown that DEHB also outperforms SMAC by mean ranks across all chosen benchmarks. For those reasons, DEHB was chose as an efficient optimizer.

For the optimization,  $\eta = 3$  and f = 4 was set, which implies an initial population of  $N = 3^3 = 27$ . The budget for the RandomForestClassifier and the GradientBoostingClassifier, as indicated by the number of trees in the forest, was set to a minimum of 10 and maximum of 270. For SVC, the budget is indicated by the maximum number of iterations and was set to a minimum of 500 and maximum of 13500. However, the runtime between the lowest and largest budget did not differ too much and was relatively cheap compared to the forest based classifiers. Thus, the SVC mostly benefits from the differential evolution and the successive halving element is not as important since many configurations can be tested irrespecitively. Thus, 40% of the maximum cost was allocated to optimizing the forest based models, and 20% was allocated to optimizing the SVC.

To evaluate the performance of the AutoML system, 3-fold external and 4-fold internal cross-validation was used. Given a total budget of 3600 seconds per dataset, a total of 1200 second could be used to optimize the AutoML system in each fold. To evaluate each hyperparameter configuration, 4-fold cross-validation was used. Since the budget is not that larger after accounting for cross-validation, only a selection of hyperparameters of the actual model were optimized and the hyperparameters of the preprocessing functions were kept at their default values.

After the optimization routine, each model is fitted with incumbent configuration and the unbalanced sampling is removed from the pipeline to not change the test set. The final AutoML system is an ensemble of three models with majority voting for the final classification.

#### 3 Experiments

The external cross-validation performance in terms of balanced accuracy of the AutoML system vs. the untuned random forest baseline for each dataset id is shown below:

Model	976	980	1002	1018	1019	1021	1040	1053	1461	41160
Baseline AutoML system					0.990 0.996					
Improvement	0.023	0.046	0.245	0.253	0.006	0.038	0.029	0.069	0.089	0.095

The AutoML system outperforms the baseline across all datasets with an improvement between 0.6% and 25.3%. Include test?

The plots of the trajectories for each dataset are in appendix A. For dataset 976, 980 1019, the SVC is the top performing estimator, followed by gradient boosting classifier and random forest classifier. For the other datasets, the ranking is not as clear. For the last two datasets, the SVM has the worst performance, most likely since SVMs are costlier to fit given the large number of observations and less iterations to tune the hyperparameter were possible. This probably also compromise the overall performance of the AutoML system. The benchmarks are outperformed usually after at most 80 seconds, except for the last two datasets due to the underperforming SVM. The final externally cross-validated performance tends to be a bit lower than the performances of the individual algorithms. This probably arises, since the performance of the individual algorithms is overly optimistic, since it is hyperparameters are tuned on that specific fold.

Write sth about final performance.

The chosen incumbents for each model for all datasets are shown below are shown in the appendix. The key trends are as following: The preferred sampling method for RandomForestClassifier and SVC is quite mixed, but for the GradientBoostingClassifier only oversampling or mixed methods were considered better. As expected, SVC uses the StandardScaler for all datasets except 1019 (Why?). If no sampling method was chosen, the imbalanced was always accounted for by class weights.

#### 4 Conclusion

The AutoML system outperforms the benchmark across all datasets and has thus proven to be useful. However, some improvements are still possible if a larger budget would be available. The hyperparameters of the pre-processing methods could be also tuned to fit better to each dataset and algorithm. Furthermore, a stacking classifier could be trained on the three algorithms, which would allow to find the best combination of the predictions of each individual algorithm for the final prediction of the AutoML system. This would be in particular beneficial to the last two datasets, where the ensemble performance was significantly worse than the two best individual performances. I conducted experiments with stacking, but the training on the final incumbents with the highest budgets took fairly long, thus I decided to use the budget rather to improve the performance of the individual estimators.

#### References

Awad, N., Mallik, N., and Hutter, F. (2021). DEHB: Evolutionary hyberband for scalable, robust and efficient hyperparameter optimization. In Zhou, Z., editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 2147–2153. ijcai.org.

#### **A** Trajectories

The following plots show the trajectories of the incumbent performance for all three models evaluated with internal cross-validation in each fold over runtime. Furthermore, the external cross-validation of the untuned random forest classifier baseline is plotted as grey horizontal line and the external cross-validation of final performance of the AutoML system is plotted as purple dot at the maximum runtime per model in each fold.

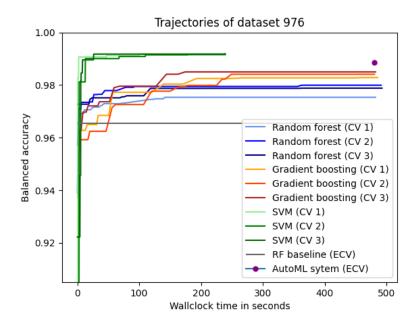


Figure 1: Trajectories for dataset 976 with 9961 observations and 14 features.

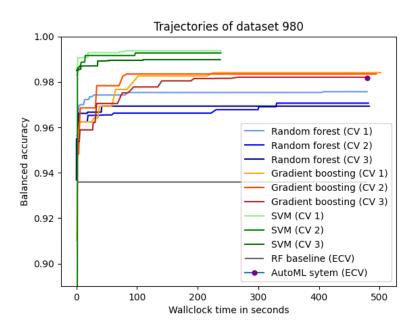


Figure 2: Trajectories for dataset 980 with 5620 observations and 64 features.

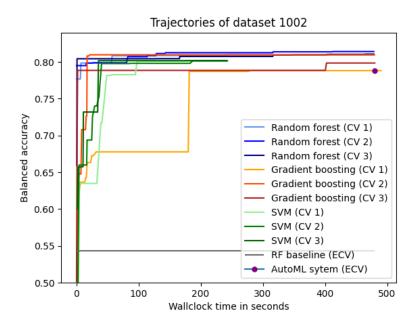


Figure 3: Trajectories for dataset 1002 with 7485 observations and 55 features.

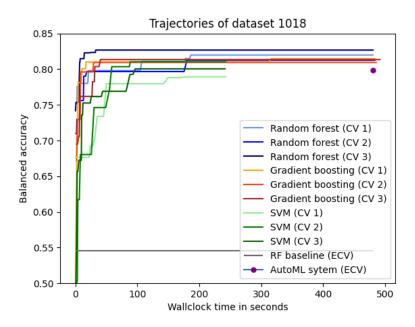


Figure 4: Trajectories for dataset 1018 with 8844 observations and 56 features.

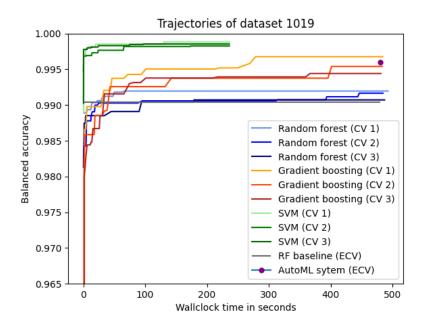


Figure 5: Trajectories for dataset 1019 with 10992 observations with 16 features

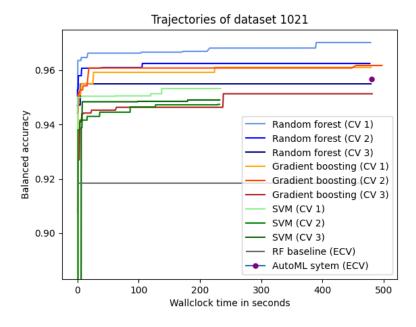


Figure 6: Trajectories for dataset 1021 with 5473 observations and 10 features.

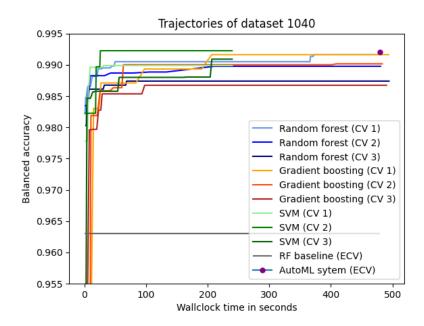


Figure 7: Trajectories for dataset 1040 with 14395 observations and 108 features.

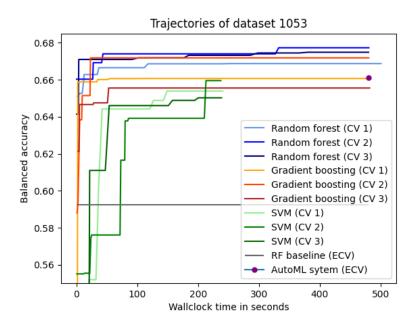


Figure 8: Trajectories for dataset 1053 with 10885 observations and 21 features.

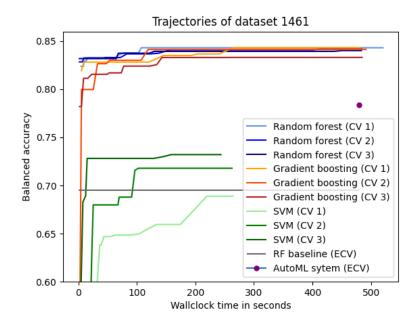


Figure 9: Trajectories for dataset 1461 with 45221 observations and 16 features.

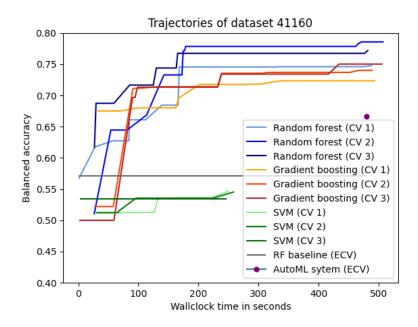


Figure 10: Trajectories for dataset 41660 with 31406 observations and 22 features.

### **B** Incumbent hyperparameter

The following tables shows the incumbent hyperparameter configurations for each algorithm and each dataset in each fold.

Table 1: Incumbent hyperparameter configurations for the random forest classifier

		Cv 101a miputer	Sampler	Scaler	CHICHOH	Max deptn	Min samples per split	min sampies per leaf	Max teatures	Class weight
976 976 976	3 2 1	Simple Simple Simple	SMOTETomek SMOTETomek SMOTE	False True False	entropy log_loss entropy	22 17 16	<b>ਚ ਚ</b> ਚ	1 1 2	0.656466 0.115289 0.262840	None balanced_subsample balanced
086 086 086	3 2 1	Simple KNN Simple	SMOTETomek SMOTETomek SMOTETomek	False True True	entropy entropy log_loss	16 19 11	7 7 1	2 6 9	0.457415 0.282039 0.144065	None balanced None
1002 1002 1002	3 2 1	KNN Simple Simple	Tomek None Tomek	False False True	entropy log_loss gini	5 2 2	8 5 11	2 6 4	0.228523 0.392731 0.192183	balanced_subsample balanced_subsample balanced
1018 1018 1018	3 2 1	KNN KNN KNN	Tomek Tomek Tomek	False False True	entropy log_loss gini	5 5 5	1 25 9	14 3	0.105286 0.497648 0.285309	balanced balanced_subsample balanced
1019 1019 1019	3 2 1	Simple KNN Simple	SMOTETomek SMOTETomek SMOTE	True False True	entropy entropy log_loss	23 19 24	3 3 3 1	1 1 2	0.806836 0.195487 0.726165	balanced_subsample None None
1021 1021 1021	3 2 1	KNN Simple Simple	SMOTE SMOTE SMOTETomek	False False True	gini entropy entropy	20 22 24	19 9 14	4 6 8	0.134323 0.495168 0.208502	balanced_subsample balanced None
1040 1040 1040	3 2 1	Simple Simple KNN	SMOTETomek SMOTETomek Tomek	False False False	gini gini gini	8 23 5	11 2 1	4 5 1	0.577811 0.797167 0.390420	balanced_subsample None balanced
1053 1053 1053	3 2 1	KNN KNN Simple	SMOTETomek Tomek Tomek	False True True	log_loss log_loss entropy	9 7 18	2 6 4	4 4 12	0.393722 0.791928 0.656912	balanced_subsample balanced balanced_subsample
1461 1461 1461	3 3 3	KNN KNN KNN	Tomek SMOTETomek Tomek	False True False	log_loss entropy entropy	20 11 14	1 21 31	11 4 2	0.158737 0.517657 0.221436	balanced_subsample balanced balanced
41160 41160 41160	1 2 3	Simple Simple Simple	None Tomek Tomek	False True True	entropy entropy gini	18 10 12	4 28 3	9 8 4	0.769477 0.783361 0.808427	balanced balanced balanced

Table 2: Incumbent hyperparameter configurations for the gradient boosting classifier

96         1         Simple         SMOTE         True         exponential         634869         friedman_mas         3         2         8           96         2         Simple         SMOTE/Torack         False         exponential         0.348789         friedman_mas         6         8         9           96         2         Simple         SMOTE/Torack         False         exponential         0.348799         friedman_mas         4         9         4           980         2         KNN         SMOTE/Torack         True         exponential         0.358599         squared_error         10         4         4           1002         1         KNN         SMOTE/Torack         True         exponential         0.35859         squared_error         10         4         4           1002         1         KNN         SMOTE/Torack         True         exponential         0.35859         friedman_me         2         1         4           1002         1         KNN         SMOTE/Torack         True         exponential         0.35859         friedman_me         3         1         4           1013         2         KNN         SMOTE/Torack         True	Dataset ID	CV fold	Imputer	Sampler	Scaler	Loss	Learning rate	Criterion	Min samples per split	Min samples per leaf	Max depth
2         Simple SMOTEFormek True         exponential 0.939669         friedman mas 6         8           2         KNN         SMOTEFormek True         exponential 0.939648         friedman mas 4         1           2         KNN         SMOTEFormek True         exponential 0.359648         squared_error 10         1           1         KNN         SMOTEFormek True         exponential 0.359648         squared_error 24         1           2         KNN         SMOTEFormek True         exponential 0.359648         squared_error 24         1           3         KNN         SMOTEFormek True         exponential 0.359648         friedman mas 5         5           4         KNN         SMOTEFormek True         exponential 0.359648         friedman mas 6         3           5         KNN         SMOTEFormek True         exponential 0.050690         friedman mas 9         13           1         KNN         SMOTEFormek True         exponential 0.05759         squared_error 21         1           2         KNN         SMOTEFormek True         exponential 0.056935         friedman mas 11         1           3         KNN         SMOTEFormek True         exponential 0.058376         friedman mas 11         1           4         KNN <td>926</td> <td>1</td> <td>Simple</td> <td>SMOTE</td> <td>True</td> <td>exponential</td> <td>0.814630</td> <td>friedman_mse</td> <td>3</td> <td>2</td> <td>8</td>	926	1	Simple	SMOTE	True	exponential	0.814630	friedman_mse	3	2	8
3         KNM         SMOTETomek         False         exponential         0.948759         friedman_mse         4         11           1         KNN         SMOTETomek         True         logs         0.36042         friedman_mse         4         9           2         KNN         SMOTETomek         True         exponential         0.023344         friedman_mse         2         1           2         KNN         SMOTETomek         True         log_loss         0.06664         friedman_mse         2         5           2         KNN         SMOTETomek         True         log_loss         0.066693         friedman_mse         2         5           2         Simple         SMOTETomek         True         exponential         0.026608         friedman_mse         3         3           3         KNN         SMOTETomek         True         exponential         0.066935         friedman_mse         1         1           4         KNN         SMOTETomek         True         exponential         0.066935         friedman_mse         1         1           5         KNN         SMOTETomek         True         exponential         0.063936         friedman_mse         1<	926	2	Simple	SMOTETomek	True	exponential	6996260	friedman_mse	9	8	10
1         Simple         SMOTE         True         log_loss         0.36042         friedman_mse         4         9           2         KNN         SMOTE/romek         True         exponential         0.36048         squared_error         24         1           2         KNN         SMOTE/romek         True         exponential         0.02560         friedman_mse         5         1           2         KNN         SMOTE Tomek         True         log_loss         0.016664         friedman_mse         9         1           1         KNN         SMOTE Tomek         True         log_loss         0.02600         friedman_mse         9         1           2         KNN         SMOTE Tomek         True         exponential         0.02759         granted_error         13         1           2         KNN         SMOTE         True         exponential         0.05603         friedman_mse         14         1           3         KNN         SMOTE         True         exponential         0.05603         friedman_mse         1         1           4         Simple         SMOTE         True         exponential         0.05603         friedman_mse         1	926	3	KNN	SMOTETomek	False	exponential	0.948759	friedman_mse	4	11	8
2         KNN         SMOTETomek         True         exponential         0.356548         squared_error         14         1           1         KNN         SMOTETomek         True         exponential         0.056544         friedman_mass         5         1           2         KNN         SMOTETomek         True         log_loss         0.026660         friedman_mass         4         3           1         KNN         SMOTETomek         True         log_loss         0.02660         friedman_mass         9         1           2         KNN         SMOTETomek         True         log_loss         0.056935         friedman_mass         9         13           3         KNN         SMOTETomek         True         exponential         0.056935         friedman_mass         13         6           1         KNN         SMOTETomek         True         log_loss         0.056935         friedman_mass         13         1           2         KNN         SMOTETomek         True         log_loss         0.056936         friedman_mass         1         1           3         KNN         SMOTETomek         True         log_loss         0.056936         friedman_mass <t< td=""><td>086</td><td>1</td><td>Simple</td><td>SMOTE</td><td>True</td><td>log_loss</td><td>0.360042</td><td>friedman_mse</td><td>4</td><td>6</td><td>4</td></t<>	086	1	Simple	SMOTE	True	log_loss	0.360042	friedman_mse	4	6	4
3         KNN         SMOTETomek         True         exponential         0.023544         friedman_mse         5         1           1         KNN         SMOTETomek         True         log_loss         0.016664         friedman_mse         5         1           2         KNN         SMOTETomek         True         log_loss         0.016664         friedman_mse         2         5           2         Simple         SMOTETomek         True         exponential         0.025060         friedman_mse         9         13           2         Simple         SMOTETomek         False         exponential         0.025053         friedman_mse         1         13           1         KNN         SMOTETomek         True         exponential         0.056935         friedman_mse         1         13           1         KNN         SMOTETomek         True         exponential         0.056935         friedman_mse         1         1           2         KNN         SMOTETomek         True         exponential         0.056935         friedman_mse         1         1           3         KNN         SMOTETomek         False         exponential         0.0283118         squared_error <td>086</td> <td>2</td> <td>KNN</td> <td>SMOTETomek</td> <td>True</td> <td>exponential</td> <td>0.359648</td> <td>squared_error</td> <td>10</td> <td>1</td> <td>4</td>	086	2	KNN	SMOTETomek	True	exponential	0.359648	squared_error	10	1	4
1         KNN         SMOTE Tomek         True         exponential         0.023344         friedman_mse         2         1           2         Simple         SMOTE Tomek         True         log_loss         0.016664         friedman_mse         2         5           1         Simple         SMOTE Tomek         True         exponential         0.027926         friedman_mse         13         9         13           2         Simple         SMOTE Tomek         False         log_loss         0.065935         friedman_mse         13         6           1         KNN         SMOTE Tomek         False         log_loss         0.086387         friedman_mse         1         6           2         KNN         SMOTE Tomek         True         exponential         0.066936         friedman_mse         1         0           2         KNN         SMOTE Tomek         False         exponential         0.0230746         friedman_mse         1         0           2         Simple         SMOTE Tomek         False         exponential         0.0230746         friedman_mse         1         0           2         Simple         SMOTE Tomek         False         log_loss         0.28307	086	3	KNN	SMOTETomek	True	exponential	0.366579	squared_error	24	12	4
2         KNN         SMOTE         True         log_loss         0.016664         friedman_mse         2         5           1         Simple         SMOTETomek         True         log_loss         0.026600         friedman_mse         9         13           2         Simple         SMOTETomek         True         exponential         0.059326         friedman_mse         9         13           2         KNN         SMOTETomek         True         exponential         0.05935         friedman_mse         1         6           2         KNN         SMOTETomek         True         exponential         0.65637         squared_error         1         6           2         KNN         SMOTE         True         exponential         0.65637         squared_error         7         2           1         KNN         SMOTE         True         exponential         0.65637         squared_error         7         2           2         KNN         SMOTE         True         exponential         0.65637         squared_error         7         2           3         KNN         SMOTETomek         True         exponential         0.65637         squared_error         7	1002	1	KNN	SMOTETomek	True	exponential	0.023344	friedman_mse	5	1	3
3         Simple         SMOTETomek         True         log_loss         0.026060         friedman_mse         4         3           1         Simple         SMOTETomek         Take         exponential         0.029726         friedman_mse         9         13           2         Simple         SMOTETomek         False         exponential         0.065935         friedman_mse         1         5           1         KNN         SMOTETomek         True         exponential         0.065935         friedman_mse         1         5           2         KNN         SMOTETomek         True         exponential         0.065936         friedman_mse         1         5           2         KNN         SMOTETomek         True         exponential         0.02506         friedman_mse         1         2           3         KNN         SMOTETomek         True         exponential         0.650849         friedman_mse         1         4           4         SMOTETomek         True         exponential         0.650849         friedman_mse         1         4           5         Simple         SMOTETomek         True         exponential         0.650849         friedman_mse         1 </td <td>1002</td> <td>2</td> <td>KNN</td> <td>SMOTE</td> <td>True</td> <td>log_loss</td> <td>0.016664</td> <td>friedman_mse</td> <td>2</td> <td>5</td> <td>2</td>	1002	2	KNN	SMOTE	True	log_loss	0.016664	friedman_mse	2	5	2
1         Simple SMOTE Simple Simple SMOTE Simple Simp	1002	3	Simple	SMOTETomek	True	log_loss	0.026060	friedman_mse	4	3	3
2         Simple SMOTETomek False exponential (2002775)         Squared_error (21)         21         1           1         KNN         SMOTETomek True exponential (2005935)         friedman_mse (21)         1         5           2         KNN         SMOTETomek True exponential (2005936)         friedman_mse (21)         2         2           1         Simple SMOTETomek True (200508)         True (200508)         friedman_mse (24)         8         8           2         Simple SMOTETomek False (200508)         friedman_mse (24)         8         8           3         KNN         SMOTETomek False (200508)         friedman_mse (24)         8         8           4         Simple SMOTE (2005)         friedman_mse (24)         8         8         8           5         KNN         SMOTETomek False (200508)         friedman_mse (24)         8         8           6         KNN         SMOTETomek False (200508)         friedman_mse (24)         8         8           1         Simple SMOTETomek True (200508)         friedman_mse (24)         1         4           2         KNN         SMOTETomek True (200508)         friedman_mse (24)         1         4           3         KNN         SMOTETomek True (200508)         friedman_mse	1018	1	Simple	SMOTE	True	exponential	0.039226	friedman_mse	6	13	2
3         KNN         SMOTE         False         log_loss         0.065935         friedman_ms         13         6           1         KNN         SMOTETomek         True         exponential         0.566058         friedman_ms         1         5           2         KNN         SMOTETomek         True         exponential         0.65637         squared_error         7         2           1         Simple         SMOTETomek         True         log_loss         0.320346         friedman_ms         24         8           2         Simple         SMOTE         True         exponential         0.220346         friedman_ms         1         6           2         KNN         SMOTE         True         exponential         0.283118         squared_error         5         2           2         KNN         SMOTE         True         exponential         0.428109         friedman_ms         5         5           2         KNN         SMOTE         True         exponential         0.545570         friedman_ms         5         5           2         Simple         SMOTETomek         False         exponential         0.031338         friedman_ms         5         <	1018	2	Simple	SMOTETomek	False	exponential	0.027759	squared_error	21	1	4
1         KNN         SMOTETomek         True         exponential         0.566058         friedman_mse         11         5           2         KNN         SMOTE         True         exponential         0.686387         squared_error         10         1           1         Simple         SMOTE         True         log_loss         0.238716         friedman_mse         24         8           2         Simple         SMOTE         False         exponential         0.20508         friedman_mse         12         6           1         Simple         SMOTE         True         exponential         0.20508         friedman_mse         14         1           2         KNN         SMOTE         True         exponential         0.20508         friedman_mse         14         4           3         KNN         SMOTE         True         exponential         0.505450         friedman_mse         3         6           1         Simple         SMOTE         True         exponential         0.010315         friedman_mse         3         4           2         KNN         SMOTE         True         exponential         0.010315         friedman_mse         3 <td< td=""><td>1018</td><td>3</td><td>KNN</td><td>SMOTE</td><td>False</td><td>log_loss</td><td>0.065935</td><td>friedman_mse</td><td>13</td><td>9</td><td>3</td></td<>	1018	3	KNN	SMOTE	False	log_loss	0.065935	friedman_mse	13	9	3
2         KNN         SMOTE         True         exponential         0.686387         squared_error         10         1           1         Simple         SMOTETomek         True         log_loss         0.230746         friedman_mse         24         8           2         Simple         SMOTE         True         log_loss         0.230746         friedman_mse         12         6           1         Simple         SMOTE         True         exponential         0.020508         friedman_mse         14         1           2         KNN         SMOTE         True         exponential         0.60849         friedman_mse         1         1           3         KNN         SMOTETomek         True         exponential         0.63849         friedman_mse         1         1           1         Simple         SMOTETomek         True         exponential         0.63849         friedman_mse         3         6           2         Simple         SMOTETomek         True         exponential         0.031338         friedman_mse         3         4           3         KNN         SMOTETomek         True         exponential         0.03294         friedman_mse         3<	1019	1	KNN	SMOTETomek	True	exponential	0.566058	friedman_mse	11	5	4
3         Simple         SMOTETomek         True         log_loss         0.230746         friedman_mse         24         8           2         Simple         SMOTE         False         exponential         0.220746         friedman_mse         12         6           2         Simple         SMOTE         False         exponential         0.283118         squared_error         5         2           1         Simple         SMOTE         True         exponential         0.428100         friedman_mse         14         4           2         KNN         SMOTE         True         exponential         0.59849         friedman_mse         1         4           2         KNN         SMOTETomek         True         exponential         0.545570         friedman_mse         5         5           2         Simple         SMOTETomek         False         exponential         0.0315         friedman_mse         5         4           4         KNN         SMOTETomek         False         exponential         0.129724         friedman_mse         3         4           5         KNN         SMOTE         True         exponential         0.129724         friedman_mse	1019	2	KNN	SMOTE	True	exponential	0.686387	squared_error	10	1	9
1         Simple Simple SMOTE         True False reponential Simple SMOTE         True False reponential Palse synonential Simple SMOTE         True reponential Palse reponential SMOTE         1020508 reponential Palse reponential Simple SMOTE         True reponential Palse reponential SMOTE         11 miles reponential Palse reponential Palse reponential SMOTE         True reponential Palse reponential Palse reponential Palse reponential Simple SMOTE         True reponential Palse reponential Pal	1019	3	Simple	SMOTETomek	True	log_loss	0.358116	squared_error	7	2	4
2         Simple SMOTE         False False         exponential logosometrial logosometrial simple         6.20308         friedman_mse friedman_m	1021	1	Simple	SMOTE	True	log_loss	0.230746	friedman_mse	24	8	5
3         KNN         SMOTETomek         False         log_loss         0.283118         squared_error         5         2           1         Simple         SMOTE         True         exponential         0.428100         friedman_mse         14         1           2         KNN         SMOTE Tomek         True         exponential         0.54570         friedman_mse         5         5           1         Simple         SMOTETomek         True         exponential         0.010315         friedman_mse         3         10           2         Simple         SMOTETomek         False         exponential         0.097863         friedman_mse         3         4           2         KNN         SMOTETomek         False         exponential         0.097863         friedman_mse         3         4           4         KNN         SMOTETomek         True         exponential         0.092295         squared_error         2         4           1         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         3         4           2         Simple         SMOTE         False         log_loss         0.018754         friedman_mse	1021	2	Simple	SMOTE	False	exponential	0.020508	friedman_mse	12	9	11
1         Simple         SMOTE         True         exponential         0.428100         friedman_mse         14         1           2         KNN         SMOTE Tomek         True         exponential         0.590849         friedman_mse         5         4           1         KNN         SMOTE Tomek         True         exponential         0.031338         friedman_mse         8         8           2         Simple         SMOTE Tomek         False         exponential         0.038175         friedman_mse         5         6           1         KNN         SMOTE Tomek         False         exponential         0.129724         friedman_mse         3         4           2         KNN         SMOTE Tomek         True         exponential         0.129724         friedman_mse         23         4           3         KNN         SMOTE Tomek         False         log_loss         0.018754         friedman_mse         23         4           4         KNN         SMOTE Tomek         False         log_loss         0.018754         friedman_mse         2         14           5         Simple         SMOTE Tomek         False         log_loss         0.018754         friedman_m	1021	3	KNN	SMOTETomek	False	log_loss	0.283118	squared_error	5	2	3
2         KNN         SMOTE         True         exponential         0.690849         friedman_mse         11         4           3         KNN         SMOTETomek         True         exponential         0.031338         friedman_mse         3         10           2         Simple         SMOTETomek         False         exponential         0.038175         friedman_mse         5         6           1         KNN         SMOTETomek         False         exponential         0.097863         friedman_mse         3         4           2         KNN         SMOTETomek         True         exponential         0.092295         squared_error         26         14           3         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         2         14           1         KNN         SMOTE         exponential         0.092295         squared_error         26         14           2         Simple         SMOTE         log_loss         0.018754         friedman_mse         3         4           2         Simple         SMOTE         log_loss         0.018754         friedman_mse         3         11           2	1040	1	Simple	SMOTE	True	exponential	0.428100	friedman_mse	14	1	3
3KNNSMOTETomekTrueexponential0.031338friedman_mse552SimpleSMOTETomekTruelog_loss0.010315friedman_mse883SimpleSMOTETomekFalseexponential0.097863friedman_mse341KNNSMOTETomekTrueexponential0.129724friedman_mse2342KNNSMOTETomekTrueexponential0.092295squared_error26141KNNSMOTEFalselog_loss0.018754friedman_mse42SimpleSMOTETomekFalselog_loss0.018754friedman_mse522SimpleSMOTETomekFalselog_loss0.018754friedman_mse523SimpleSMOTETomekFalselog_loss0.018754friedman_mse523SimpleSMOTETomekFalselog_loss0.018754friedman_mse523SimpleSMOTETomekFalseexponential0.011618squared_error516	1040	2	KNN	SMOTE	True	exponential	0.690849	friedman_mse	11	4	4
1         Simple Simple         SMOTETomek SMOTE         True         log_loss         0.003138         friedman_mse         3         10           2         Simple SMOTE         True         log_loss         0.010315         friedman_mse         5         0           1         KNN         SMOTETomek         False         exponential         0.097863         friedman_mse         3         4           2         KNN         SMOTETomek         True         exponential         0.092295         squared_error         26         14           1         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         4         11           2         Simple         SMOTE         False         log_loss         0.018754         friedman_mse         4         14           2         Simple         SMOTETomek         False         log_loss         0.018754         friedman_mse         5         2           2         Simple         SMOTETomek         False         log_loss         0.018754         friedman_mse         5         2           3         Simple         SMOTE         False         log_loss         0.016789         gauared_error         5	1040	3	KNN	SMOTETomek	True	exponential	0.545570	friedman_mse	2	5	2
2         Simple SMOTE SMOTE Mode SMOTE Simple         True In Simple SMOTE SMOTE Simple         True Simple SMOTE SMOTE Simple         True Simple SMOTE SMOTE Simple S	1053	1	Simple	SMOTETomek	True	exponential	0.031338	friedman_mse	3	10	5
3         Simple         SMOTETomek         False         exponential         0.038175         friedman_mse         5         6           1         KNN         SMOTETomek         False         exponential         0.097863         friedman_mse         23         4           2         KNN         SMOTETomek         True         exponential         0.092295         squared_error         26         14           1         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         4         11           2         Simple         SMOTETomek         False         log_loss         0.025380         friedman_mse         5         2           3         Simple         SMOTE         False         exponential         0.01618         squared_error         5         16	1053	2	Simple	SMOTE	True	log_loss	0.010315	friedman_mse	8	8	2
1         KNN         SMOTETomek         False         exponential         0.097863         friedman_mse         3         4           2         KNN         SMOTETomek         True         exponential         0.129724         friedman_mse         23         4           1         KNN         SMOTE         True         exponential         0.092295         squared_error         26         14           2         Simple         SMOTETomek         False         log_loss         0.025380         friedman_mse         5         2           3         Simple         SMOTE         False         exponential         0.011618         squared_error         5         16	1053	3	Simple	SMOTETomek	False	exponential	0.038175	friedman_mse	5	9	2
2         KNN         SMOTETomek         True         exponential         0.129724         friedman_mse         23         4           3         KNN         SMOTE         True         exponential         0.092295         squared_error         26         14           1         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         4         11           2         Simple         SMOTE False         log_loss         0.025380         friedman_mse         5         2           3         Simple         SMOTE         False         exponential         0.011618         squared_error         5         16	1461	1	KNN	SMOTETomek	False	exponential	0.097863	friedman_mse	3	4	5
3         KNN         SMOTE         True         exponential         0.092295         squared_error         26         14           1         KNN         SMOTE         False         log_loss         0.018754         friedman_mse         4         11           2         Simple         SMOTETomek         False         exponential         0.011618         squared_error         5         2           3         Simple         SMOTE         False         exponential         0.011618         squared_error         5         16	1461	2	KNN	SMOTETomek	True	exponential	0.129724	friedman_mse	23	4	7
1KNNSMOTEFalselog_loss0.018754friedman_mse4112SimpleSMOTETomekFalselog_loss0.025380friedman_mse523SimpleSMOTEFalseexponential0.011618squared_error516	1461	3	KNN	SMOTE	True	exponential	0.092295	squared_error	26	14	9
2 Simple SMOTETomek False log_loss 0.025380 friedman_mse 5 2 3 Simple SMOTE False exponential 0.011618 squared_error 5 16	41160	1	KNN	SMOTE	False	log_loss	0.018754	friedman_mse	4	11	14
3 Simple SMOTE False exponential 0.011618 squared_error 5 16	41160	2	Simple	SMOTETomek	False	log_loss	0.025380	friedman_mse	5	2	8
	41160	3	Simple	SMOTE	False	exponential	0.011618	squared_error	5	16	10

Table 3: Incumbent hyperparameter configurations for the SVM classifier  $\,$ 

None SMOTETomek True Tomek SMOTETomek True SMOTETomek True Tomek True T	Dataset ID C	CV fold	Imputer	Sampler	Scaler	C	Kernel	Shrinking	Tolerance	Class weight
2 Simple SMOTETomek True 3 Simple Tomek True 2 Simple SMOTE True 3 KNN Tomek True 4 KNN Tomek True 5 Simple SMOTETomek True 5 Simple SMOTETomek True 6 Simple SMOTETomek True 7 KNN Tomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 KNN Tomek True 1 KNN Tomek True 2 KNN None True 2 KNN SMOTETomek True 3 KNN Tomek True 4 KNN Tomek True 5 KNN Tomek True 6 Simple SMOTETomek True 7 KNN Tomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 Simple SMOTETomek True 2 KNN Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 KNN Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 KNN Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 9 Simple Tomek True 1 True	1		KNN	None	True	6.812195	rbf	True	0.000136	balanced
3 Simple Tomek True 2 Simple SMOTETomek True 3 KNN Tomek True 4 KNN Tomek True 5 Simple SMOTETomek True 5 Simple SMOTETomek True 6 Simple SMOTETomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 KNN Tomek True 1 KNN Tomek True 1 KNN Tomek True 2 KNN SMOTETomek False 1 KNN Tomek True 2 KNN Tomek True 3 KNN SMOTETomek True 5 KNN Tomek True 6 KNN Tomek True 7 KNN Tomek True 7 KNN Tomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 Simple SMOTETomek True 1 Simple SMOTETomek True 2 Simple Tomek True 3 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 KNN Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 True 1 Simple Tomek True 1 True			Simple	SMOTETomek	True	5.942369	rbf	True	0.002453	balanced
1 Simple SMOTE True 2 Simple SMOTE True 3 KNN Tomek False 2 KNN SMOTETomek True 3 Simple SMOTETomek True 5 Simple SMOTETomek True 5 Simple SMOTETomek False 6 Simple SMOTETomek False 7 KNN SMOTETomek False 8 KNN SMOTETomek False 9 KNN SMOTETomek True 1 KNN Tomek True 1 Simple SMOTETomek True 2 KNN SMOTETomek True 3 KNN SMOTETomek True 4 KNN Tomek True 5 KNN Tomek True 6 Simple SMOTETomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 Simple SMOTETomek True 1 Simple SMOTETomek True 2 Simple Tomek True 3 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 True 5 Simple Tomek True 6 Simple Tomek True 7 True			Simple	Tomek	True	5.042524	rbf	False	0.000554	balanced
2 Simple SMOTE True 3 KNN Tomek False 4 KNN Tomek True 5 KNN SMOTETomek True 5 Simple SMOTETomek True 6 Simple SMOTETomek True 7 Simple SMOTETomek True 7 Simple SMOTETomek True 8 KNN Tomek False 9 KNN SMOTETomek False 1 KNN None True 1 KNN Tomek True 2 KNN Tomek True 3 KNN Tomek True 5 KNN Tomek True 6 KNN Tomek True 7 KNN Tomek True 7 KNN Tomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 7 Simple SMOTETomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 KNN Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 True	1		Simple	SMOTETomek	True	2.461050	poly	False	0.004004	balanced
3       KNN       Tomek       False         1       KNN       SMOTETomek       True         2       KNN       SMOTETomek       True         2       Simple       SMOTETomek       True         3       Simple       SMOTETomek       False         4       KNN       Tomek       False         5       KNN       SMOTETomek       False         6       KNN       SMOTETomek       True         7       KNN       None       True         8       KNN       Tomek       True         9       KNN       Tomek       True         1       Simple       SMOTETomek       True         2       KNN       Tomek       True         3       Simple       SMOTETomek       True         4       KNN       SMOTETomek       True         5       Simple       Tomek       True         6       Simple       Tomek       True         7       Simple       Tomek       True         8       Simple       Tomek       True         9       Simple       Tomek       True         2       Sim			Simple	SMOTE	True	0.810882	poly	True	0.002291	balanced
1         KNNN         Tomek         True           2         KNNN         SMOTETomek         True           3         Simple         SMOTETomek         True           2         Simple         SMOTETomek         True           2         Simple         SMOTETomek         False           3         KNN         Tomek         False           4         KNN         SMOTETomek         False           5         KNN         Tomek         True           2         KNN         Tomek         True           3         KNN         Tomek         True           4         KNN         Tomek         True           5         Simple         SMOTETomek         True           6         Simple         SMOTETomek         True           7         Simple         Tomek         True           8         Simple         Tomek         True           9         Simple         Tomek         True           1         Simple         Tomek         True           2         Simple         Tomek         True           3         Simple         Tomek         True     <			KNN	Tomek	False	1.221203	rbf	False	0.003969	balanced
2 KNN SMOTETomek True 3 Simple SMOTETomek True 2 Simple SMOTE True 3 Simple SMOTETomek True 4 Simple SMOTETomek False 5 KNN SMOTETomek False 6 KNN SMOTETomek False 7 KNN None True 7 KNN None True 8 KNN Tomek True 9 KNN Tomek True 7 KNN Tomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 Simple SMOTETomek True 2 Simple SMOTETomek True 3 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True	2 1		KNN	Tomek	True	0.230800	linear	False	0.002982	balanced
3 Simple SMOTETomek True 2 Simple SMOTE True 3 Simple SMOTETomek True 4 KNN Tomek False 5 KNN SMOTETomek False 6 KNN SMOTETomek False 7 KNN None True 7 KNN None True 8 KNN Tomek True 9 KNN Tomek True 7 KNN Tomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 1 Simple SMOTETomek True 2 KNN Tomek True 3 Simple SMOTETomek True 6 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple SMOTETomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 4 Simple Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 1 True 9 Simple Tomek True 1 True 1 Simple Tomek True			KNN	SMOTETomek	True	0.127835	sigmoid	True	0.001461	balanced
1         Simple         SMOTE         True           2         Simple         None         True           3         Simple         SMOTETomek         False           2         KNN         SMOTETomek         False           3         KNN         SMOTETomek         False           4         KNN         None         True           5         KNN         Tomek         True           6         KNN         Tomek         True           7         KNN         Tomek         True           8         Simple         SMOTETomek         True           9         KNN         SMOTETomek         True           1         Simple         Tomek         True           2         Simple         SMOTETomek         True           3         KNN         SMOTETomek         True           4         KNN         SMOTETomek         True           5         Simple         Tomek         True           6         Simple         Tomek         True           7         Simple         Tomek         True           8         Simple         SMOTE         True <td></td> <td></td> <td>Simple</td> <td>SMOTETomek</td> <td>True</td> <td>0.384752</td> <td>linear</td> <td>False</td> <td>0.005138</td> <td>balanced</td>			Simple	SMOTETomek	True	0.384752	linear	False	0.005138	balanced
2 Simple None True 3 Simple SMOTETomek True 2 KNN Tomek False 3 KNN SMOTETomek False 4 KNN None True 5 KNN None True 6 KNN Tomek True 7 KNN Tomek True 7 KNN Tomek True 8 Simple SMOTETomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 7 Simple SMOTETomek True 8 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 2 Simple Tomek True 3 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 9 Simple Tomek True 1 Simple Tomek True 1 Simple Tomek True 1 Simple Tomek True	8 1		Simple	SMOTE	True	0.202185	sigmoid	False	0.000893	None
3 Simple SMOTETomek True  1 KNN Tomek False  2 KNN SMOTETomek False  2 KNN None True  2 KNN None True  3 KNN Tomek True  2 KNN Tomek True  2 KNN Tomek True  3 Simple SMOTETomek True  2 Simple SMOTETomek True  3 Simple SMOTETomek True  4 Simple Tomek True  5 Simple Tomek True  6 Simple Tomek True  7 Simple Tomek True  8 Simple Tomek True  9 Simple Tomek True  1 Simple Tomek True  2 Simple Tomek True  3 KNN SMOTETomek True  4 KNN SMOTETomek True  5 Simple Tomek True  6 Simple Tomek True  7 True  8 Simple Tomek True  9 Simple Tomek True  1 Simple Tomek True  2 Simple Tomek True  3 Simple Tomek True  4 True  5 Simple Tomek True  6 Simple Tomek True  7 True			Simple	None	True	0.245091	linear	True	0.001234	balanced
1         KNN         Tomek         False           2         KNN         SMOTETomek         False           3         KNN         None         True           2         KNN         Tomek         True           3         KNN         Tomek         True           2         KNN         Tomek         True           3         Simple         SMOTETomek         True           4         Simple         SMOTETomek         True           5         Simple         Tomek         True           6         Simple         Tomek         True           7         Simple         Tomek         True           8         Simple         None         True           1         Simple         None         True           2         Simple         None         True           3         Simple         None         True           4         Simple         None         True           5         KNN         Tomek         True           6         Simple         None         True           7         KNN         Tomek         True           8 <td></td> <td></td> <td>Simple</td> <td>SMOTETomek</td> <td>True</td> <td>0.135464</td> <td>linear</td> <td>False</td> <td>0.000205</td> <td>balanced</td>			Simple	SMOTETomek	True	0.135464	linear	False	0.000205	balanced
2 KNN SMOTETomek False 3 KNN SMOTETomek False 2 KNN None True 3 KNN Tomek True 2 KNN Tomek True 3 Simple SMOTETomek True 4 Simple SMOTETomek True 5 Simple SMOTETomek True 6 Simple Tomek True 7 Simple SMOTETomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple SMOTETomek True 7 Simple Tomek True 7 Simple SMOTETomek True 8 Simple SMOTETomek True 9 Simple SMOTETomek True 7 Simple SMOTETOMEK True 8 Simple Tomek True 9 Simple Tomek True 9 Simple None True 9 Simple SMOTETOMEK True 9 Simple Tomek True	9 1		KNN	Tomek	False	9.389382	poly	False	0.000422	None
3         KNN         SMOTETomek         False           1         KNN         None         True           2         KNN         Tomek         True           1         Simple         SMOTETomek         True           2         KNN         Tomek         True           3         Simple         SMOTETomek         True           2         Simple         Tomek         True           3         KNN         SMOTETomek         True           4         Simple         Tomek         True           5         Simple         Tomek         True           6         Simple         None         True           7         KNN         Tomek         True           8         Simple         None         True           9         KNN         SMOTE         True           2         KNN         SMOTE         True           3         Simple         SMOTE         True           4         KNN         SMOTE         True	9 2		KNN	SMOTETomek	False	1.538617	poly	False	0.000794	None
1         KNNN         None         True           2         KNN         None         True           3         KNN         Tomek         True           2         KNN         Tomek         True           3         Simple         SMOTETomek         True           2         Simple         Tomek         True           3         KNN         SMOTETomek         True           4         Simple         Tomek         True           5         Simple         Tomek         True           6         Simple         Tomek         True           7         Simple         Tomek         True           8         KNN         Tomek         True           8         KNN         SMOTE         True           8         KNN         SMOTE         True           9         KNN         SMOTE         True           1         KNN         SMOTE         True			KNN	SMOTETomek	False	9.673782	rbf	True	0.002117	balanced
2 KNN None True 3 KNN Tomek True 2 KNN Tomek True 3 Simple SMOTETomek True 4 Simple SMOTETomek True 5 Simple SMOTETomek True 6 Simple Tomek True 7 Simple SMOTETomek True 7 Simple Tomek True 8 Simple None True 9 Simple SMOTETomek True 7 Simple SMOTETomek True 7 Simple Tomek True 7 Simple SMOTETomek True 7 Simple SMOTETomek True 7 Simple Tomek True 7 Simple Tomek True 7 Simple None True 7 True 7 Simple None True 7 True	1 1		KNN	None	True	7.955140	rbf	False	0.000758	balanced
3 KNN Tomek True 2 KNN Tomek True 3 Simple SMOTETomek True 4 Simple SMOTETomek True 5 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple SMOTETomek True 7 Simple SMOTETomek True 7 Simple Tomek True 7 Simple Tomek True 8 Simple Tomek True 9 Simple Tomek True 7 Simple Tomek True 9 Simple None True 9 Simple None True 9 Simple None True			KNN	None	True	8.527953	rbf	False	0.000206	balanced
1 Simple SMOTETomek True 2 KNN Tomek True 3 Simple SMOTETomek True 2 Simple Tomek True 3 KNN SMOTETomek True 3 Simple Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 7 Simple SMOTETomek True 7 Simple SMOTETOMER True 7 Simple SMOTETOMER True 7 Simple Tomek True 7 Simple Tomek True 7 Simple Tomek True 7 Simple SMOTETOMER True 7 Simple SMOTETOMER True 7 True			KNN	Tomek	True	7.187957	rbf	True	0.000581	balanced
2 KNN Tomek True 3 Simple SMOTETomek True 2 Simple Tomek True 3 KNN SMOTETomek True 4 Simple Tomek True 5 Simple Tomek True 6 Simple Tomek True 7 Simple SMOTETOMER True	0 1		Simple	SMOTETomek	True	0.506277	linear	True	0.001319	None
3 Simple SMOTETomek True 2 Simple Tomek True 3 KNN SMOTETomek True 1 Simple Tomek True 2 Simple SMOTETomek True 2 Simple Tomek True 3 Simple Tomek True 1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTETOMER True 3 Simple SMOTETOMER True 4 Simple SMOTETOMER True 5 KNN Tomek True 6 Simple SMOTETOMER True 7 True	.0 2		KNN	Tomek	True	0.126113	linear	False	0.003028	balanced
1 Simple SMOTETomek True 2 Simple Tomek True 3 KNN SMOTETomek True 2 Simple SMOTETomek True 3 Simple Tomek True 1 KNN Tomek True 2 Simple None True 3 Simple None True 4 Simple SMOTETOMER True 5 Simple SMOTE True 6 Simple SMOTE True 7 True 7 Simple SMOTE True			Simple	SMOTETomek	True	1.165825	sigmoid	True	0.002011	None
2 Simple Tomek True 3 KNN SMOTETomek True 2 Simple SMOTETomek True 3 Simple Tomek True 1 KNN Tomek True 2 KNN SMOTE 3 Simple SMOTE True 3 Simple SMOTE True	3 1		Simple	SMOTETomek	True	2.294233	rbf	False	0.000396	None
3 KNN SMOTETomek True 2 Simple SMOTETomek True 3 Simple None True 1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTE True			Simple	Tomek	True	3.810805	rbf	False	0.009082	balanced
1 Simple SMOTETomek True 2 Simple Tomek True 3 Simple None True 1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTE True			KNN	SMOTETomek	True	1.594506	linear	False	0.001114	balanced
2 Simple Tomek True 3 Simple None True 1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTE True	1 1		Simple	SMOTETomek	True	0.485222	sigmoid	False	0.000393	None
3 Simple None True 1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTE True			Simple	Tomek	True	3.769543	sigmoid	True	0.001322	balanced
1 KNN Tomek True 2 KNN SMOTE True 3 Simple SMOTE True			Simple	None	True	0.533370	sigmoid	False	0.008771	balanced
2 KNN SMOTE True	60 1		KNN	Tomek	True	8.007605	rbf	False	0.000958	None
3 Simple SMOTE True			KNN	SMOTE	True	0.227120	$^{\mathrm{rbf}}$	True	0.001452	balanced
anti di cino admino			Simple	SMOTE	True	9.758566	sigmoid	False	0.000217	balanced

## C Information about datasets

Dataset ID	Observations	Features
976	9961	14
980	5620	64
1002	7485	55
1018	8844	56
1019	10992	16
1021	5473	10
1040	14395	108
1053	10885	21
1461	45221	16
41160	31406	22