

# [Supplementary Document] Configuration Space Reduction in Automated Machine Learning Using Relative Landmarking

Tien-Dung Nguyen<sup>1</sup>, Bogdan Gabrys<sup>2</sup>, and Katarzyna Musial<sup>2</sup>

Advanced Analytics Institute, University of Technology Sydney, Sydney, Australia

`TienDung.Nguyen-2@student.uts.edu.au`<sup>1</sup>

`{Bogdan.Gabrys, Katarzyna.Musial-Gabrys}@uts.edu.au`<sup>2</sup>

This document provides more details about the experiments for the paper entitled *Configuration Space Reduction in Automated Machine Learning Using Relative Landmarking*.

**Table 1.** The best ML pipelines found using different methods to design configuration spaces.

Dataset	r30	baseline	avatar
<b>abalone</b>	CustomReplaceMissingValues → RandomSubset → Resample → Logistic → Bagging	SimpleLogistic	SMO
<b>adult</b>	J48	PART	PART
<b>amazon</b>	CustomReplaceMissingValues → Normalize → RandomSubset → NaiveBayesMultinomial → RandomSubSpace	NaiveBayes	NaiveBayesMultinomial
<b>car</b>	SMO → MultiClassClassifier	SMO	SMO
<b>cifar10small</b>	RandomForest → MultiClassClassifier	DecisionStump	RandomForest
<b>convex</b>	RandomForest → AdaBoostM1	VotedPerceptron	RandomForest
<b>dexter</b>	DecisionStump → AdaBoostM1	NaiveBayesMultinomial	SGD
<b>dorothea</b>	OneR → RandomSubSpace	PART	DecisionStump
<b>gcredit</b>	LMT → Bagging	SMO	SMO
<b>gisette</b>	VotedPerceptron → RandomSubSpace	VotedPerceptron	VotedPerceptron
<b>kddcup</b>	-	DecisionStump	ClassBalancer → RemoveOutliers → InterquartileRange → AttributeSelection → Resample → PART
<b>krvsnp</b>	Jrip → AdaBoostM1	RandomForest	J48
<b>madelon</b>	PrincipalComponents → IBk → LogitBoost	Jrip	Jrip
<b>mnist</b>	CustomReplaceMissingValues → Center → J48 → AdaBoostM1	-	PART
<b>secom</b>	J48 → AdaBoostM1	ClassBalancer → EMImputation → Normalize → PrincipalComponents → Kstar → MultiClassClassifier	SimpleLogistic
<b>semeion</b>	CustomReplaceMissingValues → PrincipalComponents → SMO	SMO	SMO
<b>shuttle</b>	RandomForest → AdaBoostM1	RandomForest	RandomForest
<b>waveform</b>	RemoveOutliers → InterquartileRange → Normalize → SMO → AttributeSelectedClassifier	LMT	SimpleLogistic
<b>winequality</b>	RandomForest → AdaBoostM1	RandomForest	Kstar
<b>yeast</b>	RandomForest → Bagging	RandomForest	RandomForest

**Table 2.** The best ML pipelines found using different methods to design configuration spaces (continued).

Dataset	L-k1	L-k4	L-k8	L-k10	L-k19
<b>abalone</b>	CustomReplaceMissingValues → Normalize → RandomSubset → SimpleLogistic	ClassBalancer → RemoveOutliers → InterquartileRange → Normalize → RandomSubset → SimpleLogistic	REPTree	RandomForest	PART
<b>adult</b>	RemoveOutliers → InterquartileRange → Center → Logistic	Logistic	NaiveBayes	PART	PART
<b>amazon</b>	ClassBalancer → RemoveOutliers → InterquartileRange → Center → Logistic	SMO	NaiveBayesMultinomial	NaiveBayesMultinomial	NaiveBayesMultinomial
<b>car</b>	SMO	SMO	SMO	SMO	SMO
<b>cifar10small</b>	RandomSubset → PeriodicSampling → Logistic	NaiveBayesMultinomial	RandomForest	NaiveBayes	-
<b>convex</b>	-	Jrip	SMO	RandomForest	RandomForest
<b>dexter</b>	SMO	SMO	SMO	SMO	SimpleLogistic
<b>dorothea</b>	-	-	-	DecisionStump	DecisionStump
<b>gcredit</b>	LMT	NaiveBayes	NaiveBayes	SMO	SMO
<b>gisette</b>	Logistic	-	Jrip	RandomForest	Logistic
<b>kddcup</b>	-	-	-	-	-
<b>krvskp</b>	SMO	SMO	J48	J48	RandomForest
<b>madelon</b>	SpreadSubsample → CustomReplaceMissingValues → Normalize → RandomSubset → Resample → VotedPerceptron	Jrip	Jrip	Jrip	Jrip
<b>mnist</b>	-	-	-	SMO	SimpleLogistic
<b>secom</b>	-	ZeroR	ZeroR	DecisionTable	LMT
<b>semeion</b>	SMO	SMO	SMO	SMO	SMO
<b>shuttle</b>	SMO	Jrip	PART	RandomForest	RandomForest
<b>waveform</b>	SMO	SMO	SimpleLogistic	SimpleLogistic	LMT
<b>winequality</b>	SMO	Kstar	SMO	RandomForest	Kstar
<b>yeast</b>	Logistic	SMO	RandomForest	RandomForest	RandomForest

**Table 3.** The best ML pipelines found using different methods to design configuration spaces (continued).

Dataset	O-k1	O-k4	O-k8	O-k10	O-k19
abalone	Logistic	DecisionTable	SimpleLogistic	PART	MultilayerPerceptron
adult	-	PART	J48	J48	PART
amazon	-	NaiveBayesMultinomial	NaiveBayesMultinomial	NaiveBayesMultinomial	NaiveBayesMultinomial
car	LMT	SMO	SMO	SMO	SMO
cifar10small	-	RandomForest	RandomForest	NaiveBayes	RandomForest
convex	-	RandomForest	SMO	RandomForest	RandomForest
dexter	SGD	SGD	VotedPerceptron	SGD	SimpleLogistic
dorothea	SpreadSubsample → CustomReplaceMissingValues → Center → DecisionStump	NaiveBayes	NaiveBayes	NaiveBayes	OneR
gcredit	NaiveBayes	MultilayerPerceptron	SMO	MultilayerPerceptron	MultilayerPerceptron
gisette	VotedPerceptron	VotedPerceptron	VotedPerceptron	VotedPerceptron	VotedPerceptron
kddcup	MultilayerPerceptron	Ibk	DecisionStump	-	DecisionStump
krvsnp	LMT	J48	J48	J48	J48
madelon	Jrip	Jrip	Jrip	Jrip	Jrip
mnist	SMO	-	SMO	RandomForest	SMO
secom	ClassBalancer → CustomReplaceMissingValues → RemoveOutliers → InterquartileRange → Normalize → PeriodicSampling → DecisionStump	SpreadSubsample → Standardize → RandomSubset → PeriodicSampling → ZeroR	Kstar	VotedPerceptron	Kstar
semeion	Logistic	SMO	SMO	SMO	SMO
shuttle	RandomForest	RandomForest	RandomForest	RandomForest	RandomForest
waveform	Logistic	SimpleLogistic	SimpleLogistic	SimpleLogistic	SimpleLogistic
winequality	RandomForest	RandomForest	RandomForest	Kstar	Kstar
yeast	RandomForest	RandomForest	RandomForest	RandomForest	RandomForest