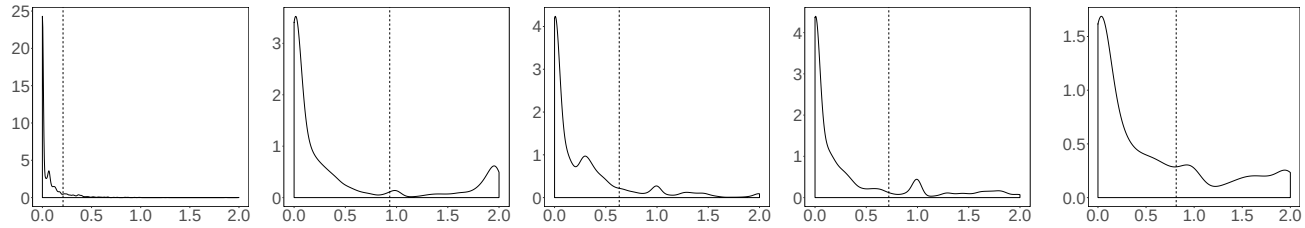
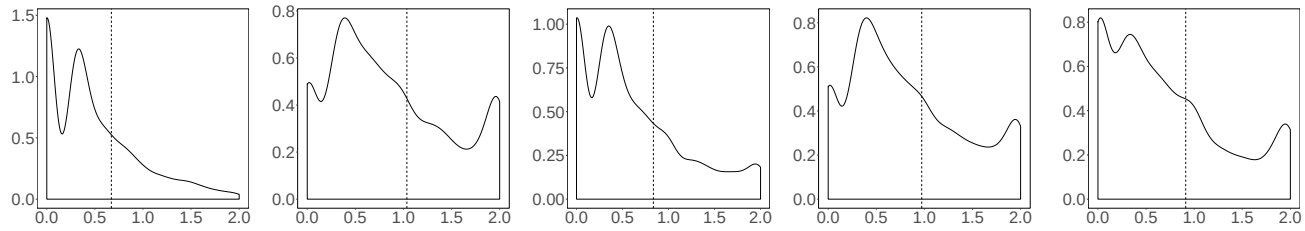


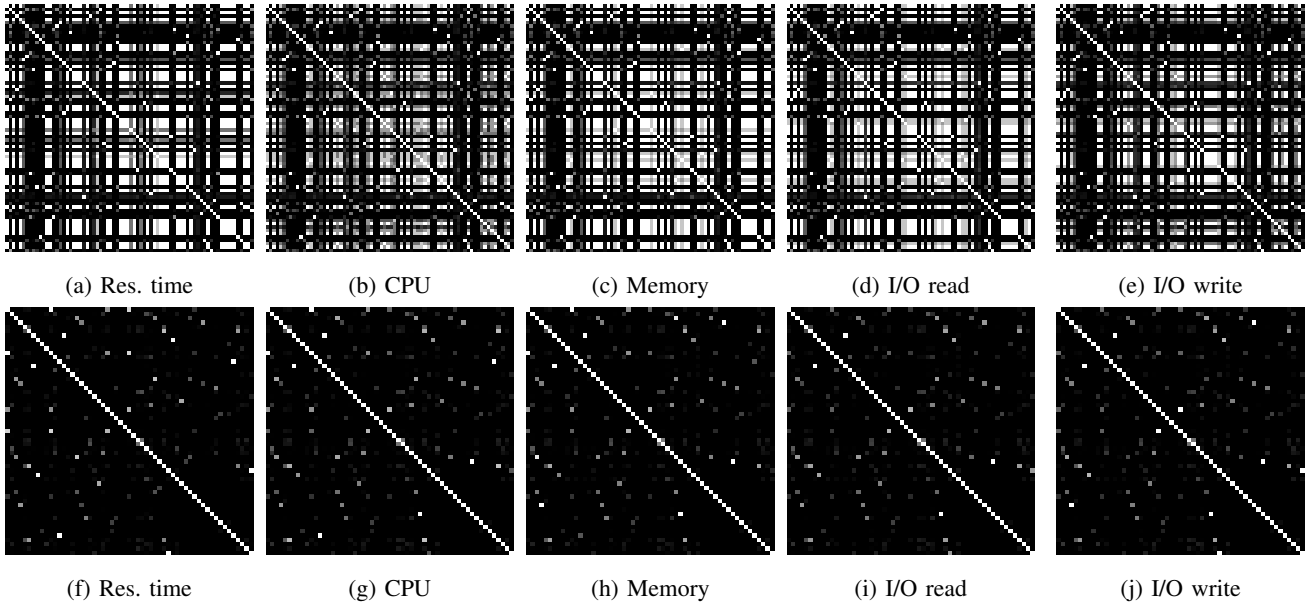
(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 1: The automatically obtained thresholds for splitting option variation into *IoPV* and non-*IoPV* groups for *Hadoop*.



(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

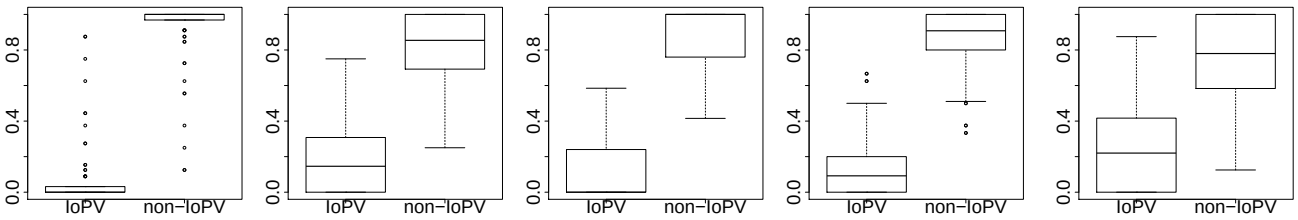
Fig. 2: The automatically obtained thresholds for splitting option variation into *IoPV* and non-*IoPV* groups for *Cassandra*.



(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

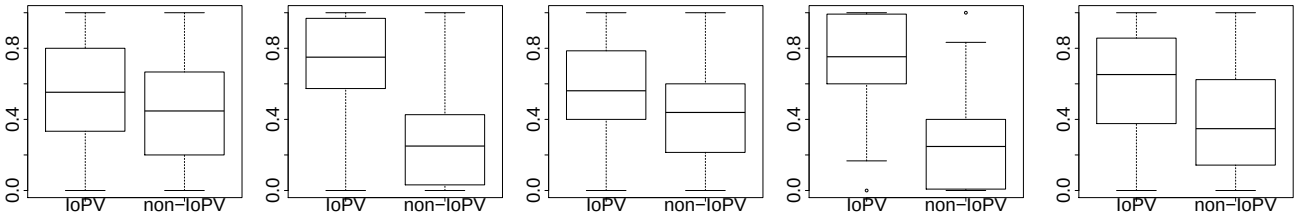
(f) Res. time (g) CPU (h) Memory (i) I/O read (j) I/O write

Fig. 3: Pairwise Jaccard distance between the <test, option, *IoPV* > triplets of the studied commits of the *Hadoop* (top five sub-figures) and *Cassandra* (buttom five sub-figures) system. The x-axis and y-axis show the studied commits, ordered chronologically from left to right on the x-axis and bottom to top on the y-axis. Each cell of the Figure refers to the Jaccard distance of any pair of commits: the darker the color is, the larger the distance is. Most of the commits are with dark color, which means that different commits are unlikely to have the same pairs of tests and options that can lead to *IoPV*.



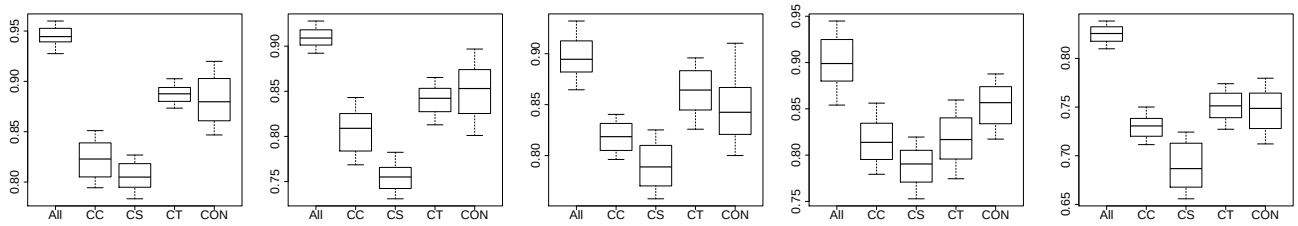
(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 4: Percentage of *IoPV* for each commit of *Hadoop*. Non-*IoPV* is equal to 1-*IoPV*.



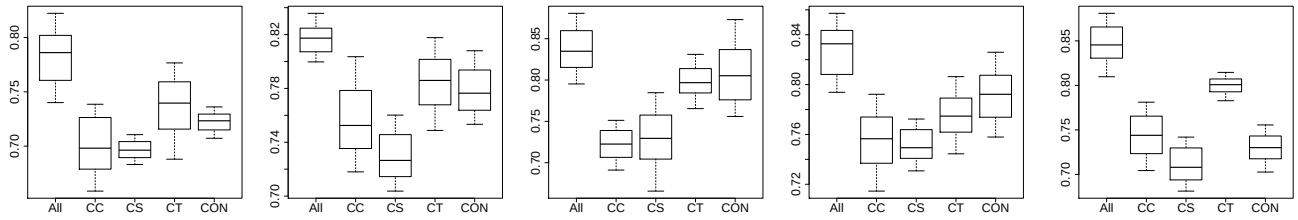
(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 5: Percentage of *IoPV* for each commit of *Cassandra*. Non-*IoPV* is equal to 1-*IoPV*.



(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 6: AUC of RF for *Hadoop* when only keeping one dimension of metrics.

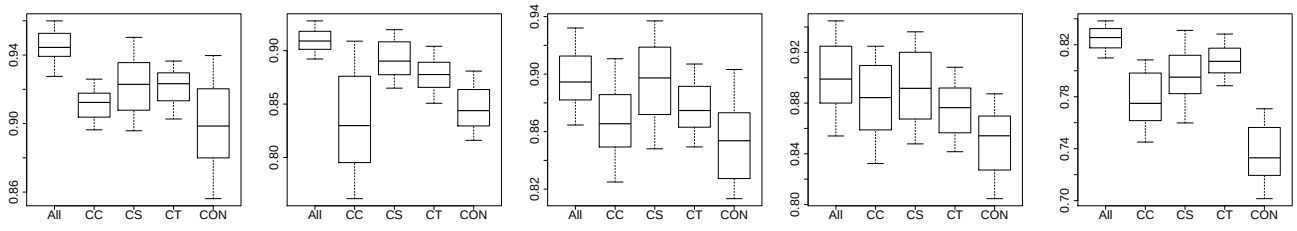


(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 7: AUC of RF for *Cassandra* when only keeping one dimension of metrics.

TABLE I: *Hadoop*'s results of using different models to predict whether configuration options cause the manifesting of *IoPV*. The best results for each performance metric and each model are highlighted in *italic*. The best results for each performance metric across different models are highlighted in ***bold-italic***.

Hadoop									
	RF with tf-idf			RF with PCA			RF with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.68	0.39	*0.93*	0.68	0.39	0.66	0.73	0.33	*0.93*
Cpu	0.70	0.51	0.90	0.55	0.02	0.71	0.77	0.60	***0.92***
Memory	0.64	0.36	0.87	0.48	0.04	0.69	0.75	0.41	***0.91***
I/O Read	0.68	0.54	0.91	0.58	0.02	0.76	0.79	0.56	***0.93***
I/O Write	0.63	0.44	0.82	0.44	0.02	0.59	0.72	0.49	***0.85***
Average	0.67	0.45	0.89	0.55	0.10	0.68	0.75	0.48	***0.91***
	LR with tf-idf			LR with PCA			LR with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.38	0.03	0.67	0.12	0.46	0.54	0.53	0.09	*0.77*
Cpu	0.66	0.06	0.73	0.27	0.29	0.61	0.48	0.14	*0.76*
Memory	0.49	0.04	0.71	0.16	0.40	0.55	0.48	0.10	*0.73*
I/O Read	0.70	0.05	0.71	0.22	0.33	0.57	0.46	0.18	*0.80*
I/O Write	0.50	0.06	0.64	0.33	0.22	0.57	0.50	0.14	*0.66*
Average	0.55	0.05	0.69	0.22	0.34	0.57	0.49	0.13	*0.74*
	XG with tf-idf			XG with PCA			XG with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.65	0.42	***0.94***	1.00	0.05	0.60	0.71	0.31	0.93
Cpu	0.66	0.48	*0.88*	0.32	0.06	0.62	0.67	0.50	*0.88*
Memory	0.66	0.32	*0.87*	0.41	0.04	0.68	0.72	0.32	*0.87*
I/O Read	0.66	0.49	*0.91*	0.49	0.08	0.73	0.73	0.50	*0.91*
I/O Write	0.66	0.38	*0.82*	0.41	0.16	0.58	0.67	0.40	0.80
Average	0.66	0.42	*0.88*	0.52	0.08	0.64	0.70	0.41	0.88
	NN with tf-idf			NN with PCA			NN with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.34	0.54	0.79	0.27	0.83	*0.80*	0.27	0.64	0.75
Cpu	0.53	0.30	0.72	0.63	0.41	*0.73*	0.39	0.33	0.65
Memory	0.43	0.27	*0.67*	0.52	0.34	*0.67*	0.31	0.42	0.66
I/O Read	0.53	0.44	0.73	0.60	0.46	*0.76*	0.48	0.33	0.72
I/O Write	0.50	0.38	*0.68*	0.53	0.32	0.65	0.39	0.41	*0.68*
Average	0.47	0.39	*0.72*	0.51	0.47	*0.72*	0.37	0.43	0.69
	CNN with tf-idf			CNN with PCA			CNN with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.29	0.48	0.75	0.06	0.90	0.73	0.23	0.51	*0.79*
Cpu	0.22	0.68	0.78	0.18	0.78	0.76	0.63	0.25	*0.81*
Memory	0.47	0.25	0.69	0.13	0.87	0.74	0.20	0.57	*0.76*
I/O Read	0.32	0.41	*0.68*	0.27	0.25	*0.68*	0.20	0.38	0.66
I/O Write	0.27	0.31	0.64	0.14	0.64	0.65	0.19	0.60	*0.67*
Average	0.31	0.43	0.71	0.16	0.69	0.71	0.29	0.46	*0.74*

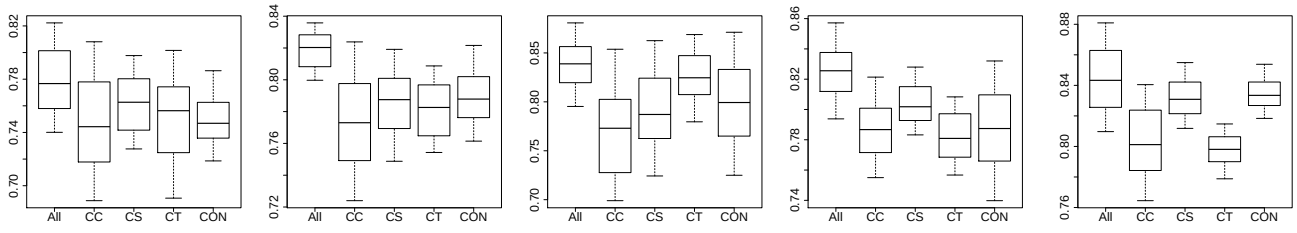


(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 8: AUC of RF for *Hadoop* when removing one dimension of metrics.

TABLE II: *Cassandra*'s results of using different models to predict whether configuration options cause the manifesting of *IoPV*. The best results for each performance metric and each model are highlighted in *italic*. The best results for each performance metric across different models are highlighted in ***bold-italic***.

Cassandra									
	RF with tf-idf			RF with PCA			RF with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.74	0.37	0.74	0.45	0.13	0.62	0.67	0.46	*0.75*
Cpu	0.68	0.39	0.76	0.46	0.15	0.61	0.73	0.59	*0.82*
Memory	0.71	0.37	0.78	0.35	0.04	0.61	0.71	0.58	***0.84***
I/O Read	0.74	0.48	0.79	0.54	0.32	0.67	0.74	0.63	***0.83***
I/O Write	0.76	0.50	0.82	0.58	0.32	0.68	0.77	0.65	***0.86***
Average	0.73	0.42	0.78	0.47	0.19	0.64	0.72	0.58	*0.82*
	LR with tf-idf			LR with PCA			LR with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.38	0.38	0.59	0.28	0.54	0.54	0.38	0.51	*0.63*
Cpu	0.49	0.42	0.64	0.33	0.41	0.53	0.46	0.58	*0.65*
Memory	0.44	0.26	0.62	0.29	0.28	0.55	0.43	0.47	*0.66*
I/O Read	0.50	0.50	0.63	0.35	0.44	0.55	0.49	0.61	*0.67*
I/O Write	0.53	0.51	*0.69*	0.36	0.37	0.54	0.47	0.63	0.68
Average	0.47	0.41	0.64	0.32	0.41	0.54	0.44	0.56	*0.66*
	XG with tf-idf			XG with PCA			XG with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.66	0.38	*0.75*	0.37	0.20	0.60	0.63	0.44	0.74
Cpu	0.65	0.49	0.77	0.49	0.32	0.63	0.70	0.60	*0.82*
Memory	0.65	0.49	0.80	0.45	0.13	0.60	0.70	0.56	*0.83*
I/O Read	0.69	0.55	0.79	0.46	0.35	0.61	0.72	0.62	*0.81*
I/O Write	0.74	0.59	0.84	0.48	0.33	0.65	0.74	0.64	*0.85*
Average	0.68	0.50	0.79	0.45	0.27	0.62	0.70	0.57	*0.81*
	NN with tf-idf			NN with PCA			NN with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.55	0.36	0.71	0.53	0.40	*0.73*	0.47	0.41	0.67
Cpu	0.27	0.94	0.88	0.31	0.96	***0.90***	0.26	0.88	0.84
Memory	0.56	0.51	0.76	0.64	0.57	***0.84***	0.61	0.49	0.76
I/O Read	0.66	0.46	0.75	0.67	0.57	***0.83***	0.60	0.54	0.74
I/O Write	0.67	0.39	0.77	0.70	0.57	*0.84*	0.63	0.55	0.77
Average	0.54	0.53	0.77	0.57	0.62	***0.83***	0.51	0.57	0.76
	CNN with tf-idf			CNN with PCA			CNN with code embedding		
	Pre.	Recall	AUC	Pre.	Recall	AUC	Pre.	Recall	AUC
Res. time	0.30	0.28	0.75	0.37	0.35	***0.79***	0.33	0.34	0.75
Cpu	0.29	0.37	*0.76*	0.25	0.67	0.75	0.11	0.96	*0.76*
Memory	0.37	0.21	*0.77*	0.37	0.21	*0.77*	0.33	0.30	0.74
I/O Read	0.19	0.53	*0.69*	0.30	0.37	0.68	0.24	0.47	*0.69*
I/O Write	0.23	0.40	*0.70*	0.19	0.38	0.67	0.23	0.33	0.69
Average	0.28	0.36	*0.73*	0.30	0.40	*0.73*	0.25	0.48	*0.73*



(a) Res. time (b) CPU (c) Memory (d) I/O read (e) I/O write

Fig. 9: AUC of RF for *Cassandra* when removing one dimension of metrics.