Supplementary Material for
On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study
by G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent and M. E. Houle
Data Mining and Knowledge Discovery 30(4): 891-927, 2016, DOI: 10.1007/s10618-015-0444-8

Pima (10% of outliers version#01)

The data set contains medical data on diabetes. Patients suffering from diabetes were considered outliers.

Download all data set variants used (694.8 kB). You can also access the original data. (pima-indians-diabetes.data)

- Normalized, without duplicates
- Not normalized, without duplicates

Normalized, without duplicates

This version contains 8 attributes, 555 objects, 55 outliers (9.91%)

Download raw algorithm results (4.9 MB) Download raw algorithm evaluation table (53.8 kB)

Best Parameters

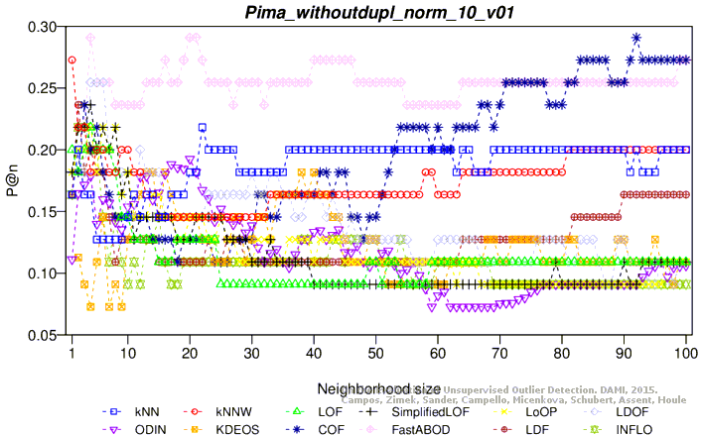
The following table contains the best (overall and per-method) results for each method and evaluation measure (when the same score was achieved twice, only the smallest k is given).
The Maximum F1-Measure is complimentary in addition to the measures in the original publication.

Algorithm	k	P@n	Adj. P@n	AP	Adj. AP	Max-F1	Adj. MF1	ROC AUC
KNN	**1**	0.16364	0.07164	0.22478	0.13950	0.36471	0.29482	0.77233
KNN	**2**	0.20000	0.11200	0.22923	0.14445	0.34826	0.27657	0.78289
KNN	**22**	0.21818	0.13218	0.20534	0.11793	0.33692	0.26398	0.76445
KNNW	**1**	0.27273	0.19273	0.21793	0.13190	0.33480	0.26163	0.76096
KNNW	**3**	0.21818	0.13218	0.23023	0.14555	0.35789	0.28726	0.78273
KNNW	**14**	0.16364	0.07164	0.21089	0.12409	0.36441	0.29449	0.77342
LOF	**4**	0.21818	0.13218	0.17063	0.07940	0.25373	0.17164	0.63767
LOF	**86**	0.10909	0.01109	0.17879	0.08845	0.34028	0.26771	0.74458
LOF	**97**	0.10909	0.01109	0.18193	0.09194	0.32491	0.25065	0.74527
LOF	**98**	0.10909	0.01109	0.18153	0.09150	0.32258	0.24806	0.74535

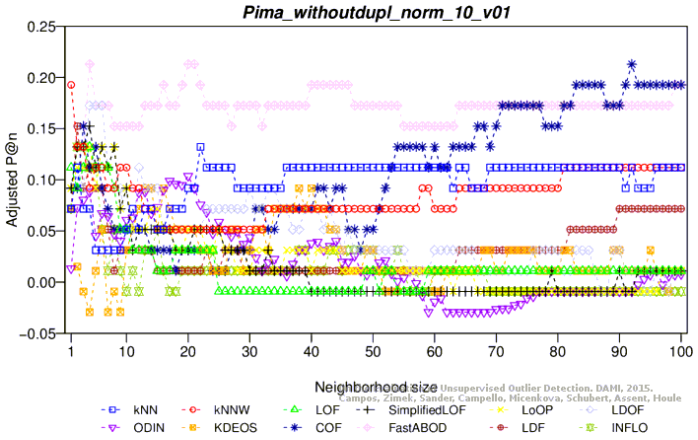
SimplifiedLOF	**4**	0.23636	0.15236	0.18152	0.09149	0.28358	0.20478	0.64564
SimplifiedLOF	**100**	0.10909	0.01109	0.16651	0.07483	0.31250	0.23687	0.70182
LoOP	**4**	0.21818	0.13218	0.17951	0.08926	0.28788	0.20955	0.64669
LoOP	**97**	0.10909	0.01109	0.16271	0.07061	0.30000	0.22300	0.69055
LoOP	**98**	0.10909	0.01109	0.16303	0.07097	0.30705	0.23083	0.68913
LDOF	**4**	0.25455	0.17255	0.19953	0.11147	0.29213	0.21427	0.64215
LDOF	**6**	0.25455	0.17255	0.17398	0.08312	0.27984	0.20062	0.67473
ODIN	**20**	0.19251	0.10369	0.15770	0.06504	0.24432	0.16119	0.64273
ODIN	**80**	0.09091	-0.00909	0.16842	0.07695	0.28682	0.20837	0.69749
ODIN	**99**	0.10545	0.00705	0.16572	0.07395	0.30088	0.22398	0.70193
ODIN	**100**	0.10606	0.00773	0.16606	0.07433	0.29956	0.22251	0.70360
FastABOD	**4**	0.29091	0.21291	0.24822	0.16553	0.36681	0.29716	0.79535
FastABOD	**18**	0.25455	0.17255	0.26549	0.18469	0.43350	0.37118	0.81342
FastABOD	**70**	0.25455	0.17255	0.27233	0.19229	0.41837	0.35439	0.81865
FastABOD	**96**	0.25455	0.17255	0.27088	0.19068	0.41885	0.35492	0.81996
KDEOS	**13**	0.18182	0.09182	0.14512	0.05108	0.22630	0.14119	0.61422
KDEOS	**14**	0.18182	0.09182	0.15092	0.05752	0.21449	0.12809	0.60160
KDEOS	**98**	0.09091	-0.00909	0.12508	0.02884	0.24667	0.16380	0.62353
KDEOS	**100**	0.10909	0.01109	0.12597	0.02983	0.24585	0.16289	0.62538
LDF	**2**	0.23636	0.15236	0.17026	0.07899	0.25714	0.17543	0.59435
LDF	**66**	0.12727	0.03127	0.19121	0.10224	0.35200	0.28072	0.75993
LDF	**76**	0.12727	0.03127	0.19290	0.10412	0.33962	0.26698	0.76167
LDF	**98**	0.16364	0.07164	0.19590	0.10744	0.33213	0.25866	0.75818
INFLO	**2**	0.21818	0.13218	0.16594	0.07419	0.25143	0.16909	0.62516
INFLO	**99**	0.09091	-0.00909	0.17176	0.08065	0.30222	0.22547	0.71673
INFLO	**100**	0.09091	-0.00909	0.17138	0.08023	0.30522	0.22880	0.71575
COF	**92**	0.29091	0.21291	0.24519	0.16216	0.39490	0.32834	0.80444
COF	**100**	0.27273	0.19273	0.25958	0.17814	0.38235	0.31441	0.81167

Plots

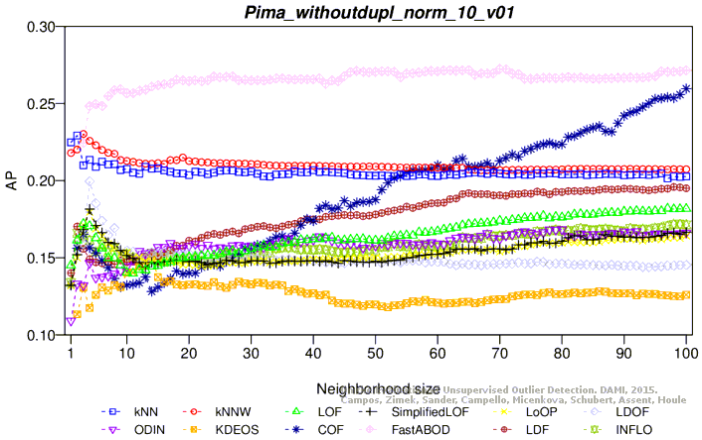
Precision at *n*



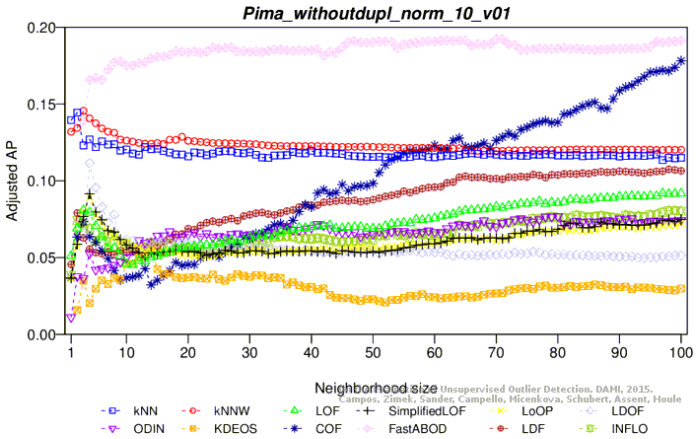
Adjusted precision at *n*



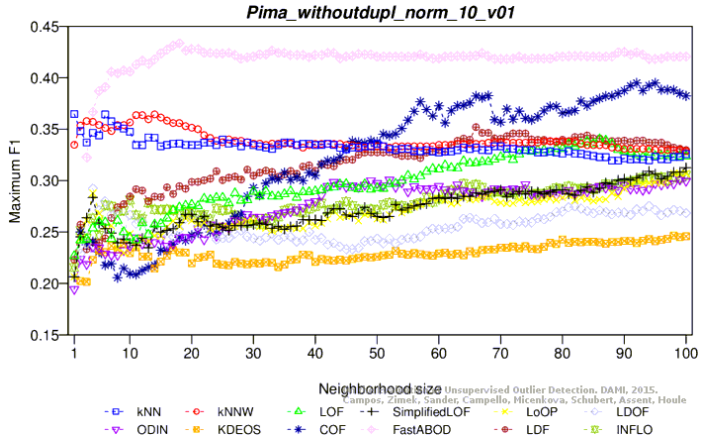
Average precision



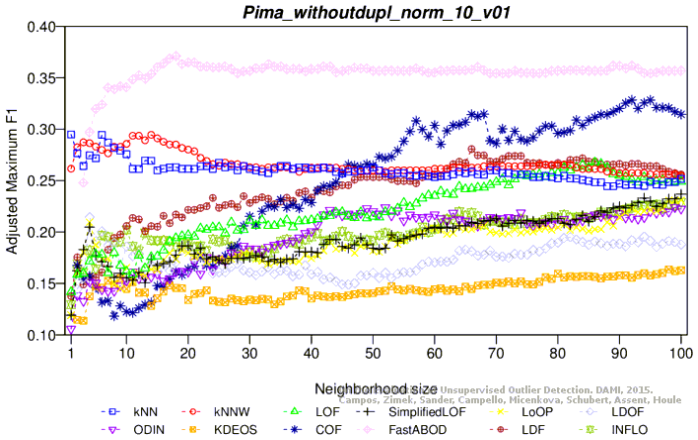
Adjusted average precision



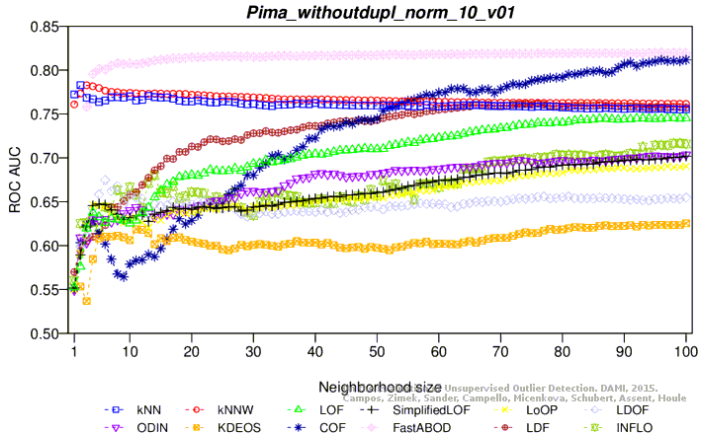
Maximum F1 score



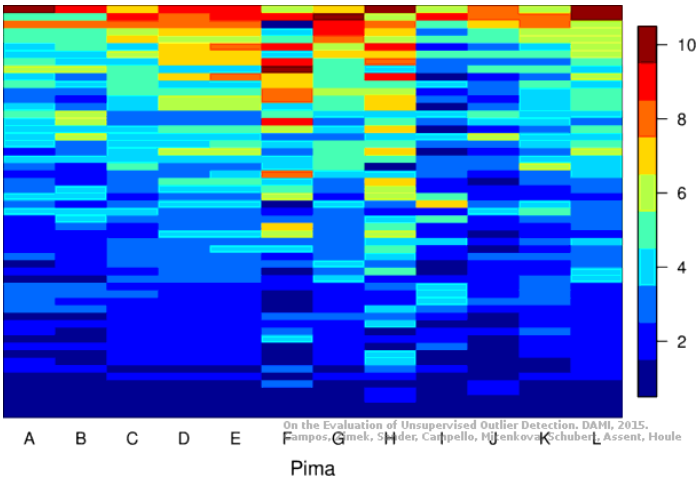
Adjusted maximum F1 score



ROC AUC



Diversity



A: KNN, B: KNNW, C: LOF, D: SimplifiedLOF, E: LoOP, F: LDOF
G: ODIN, H: KDEOS, I: COF, J: FastABOD, K: LDF, L: INFLO

Not normalized, without duplicates

This version contains 8 attributes, 555 objects, 55 outliers (9.91%)

Download raw algorithm results (4.8 MB) Download raw algorithm evaluation table (53.8 kB)

Best Parameters

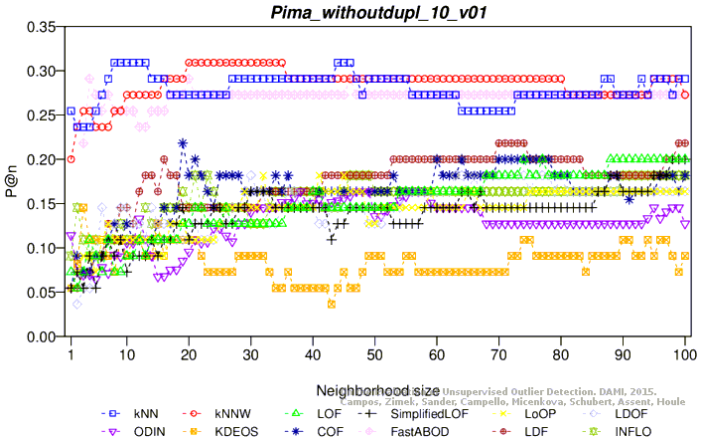
The following table contains the best (overall and per-method) results for each method and evaluation measure (when the same score was achieved twice, only the smallest k is given).
The Maximum F1-Measure is complimentary in addition to the measures in the original publication.

Algorithm	k	P@n	Adj. P@n	AP	Adj. AP	Max-F1	Adj. MF1	ROC AUC
KNN	**8**	0.30909	0.23309	0.21680	0.13065	0.36709	0.29747	0.70756
KNN	**9**	0.30909	0.23309	0.21652	0.13034	0.36025	0.28988	0.70967
KNN	**12**	0.30909	0.23309	0.22097	0.13527	0.35802	0.28741	0.70733
KNNW	**4**	0.25455	0.17255	0.20138	0.11354	0.33533	0.26222	0.70698
KNNW	**14**	0.27273	0.19273	0.21289	0.12631	0.35955	0.28910	0.70680
KNNW	**20**	0.30909	0.23309	0.21645	0.13025	0.35366	0.28256	0.70662
LOF	**74**	0.20000	0.11200	0.18587	0.09631	0.33684	0.26389	0.69600
LOF	**78**	0.18182	0.09182	0.18775	0.09840	0.34043	0.26787	0.69735
LOF	**84**	0.18182	0.09182	0.18894	0.09972	0.33149	0.25796	0.69927
LOF	**95**	0.20000	0.11200	0.18846	0.09919	0.33136	0.25781	0.70156
SimplifiedLOF	**98**	0.20000	0.11200	0.17478	0.08400	0.30000	0.22300	0.66215
SimplifiedLOF	**100**	0.20000	0.11200	0.17479	0.08402	0.30337	0.22674	0.66276
LoOP	**32**	0.18182	0.09182	0.13795	0.04312	0.22581	0.14065	0.60096
LoOP	**81**	0.16364	0.07164	0.16143	0.06919	0.28571	0.20714	0.65425
LoOP	**99**	0.16364	0.07164	0.16504	0.07320	0.29586	0.21840	0.65067
LoOP	**100**	0.16364	0.07164	0.16612	0.07439	0.29586	0.21840	0.65218
LDOF	**30**	0.18182	0.09182	0.13635	0.04135	0.23041	0.14576	0.59204
LDOF	**98**	0.16364	0.07164	0.16764	0.07608	0.28758	0.20922	0.65305
LDOF	**100**	0.16364	0.07164	0.16780	0.07625	0.28571	0.20714	0.65287
ODIN	**38**	0.16364	0.07164	0.13341	0.03809	0.23529	0.15118	0.60515
ODIN	**84**	0.12727	0.03127	0.16271	0.07061	0.28054	0.20140	0.62862
ODIN	**85**	0.12727	0.03127	0.16294	0.07087	0.27149	0.19136	0.62765
FastABOD	**4**	0.29091	0.21291	0.21387	0.12740	0.32836	0.25448	0.72804
FastABOD	**97**	0.27273	0.19273	0.23387	0.14959	0.38857	0.32131	0.73360
FastABOD	**98**	0.27273	0.19273	0.23373	0.14944	0.39080	0.32379	0.73353
FastABOD	**100**	0.27273	0.19273	0.23363	0.14932	0.39080	0.32379	0.73371
KDEOS	**3**	0.14545	0.05145	0.12967	0.03394	0.22826	0.14337	0.55855
KDEOS	**79**	0.09091	-0.00909	0.12073	0.02401	0.24149	0.15805	0.60058
KDEOS	**99**	0.07273	-0.02927	0.12317	0.02672	0.23729	0.15339	0.60709
LDF	**63**	0.20000	0.11200	0.19429	0.10567	0.34682	0.27497	0.71062
LDF	**70**	0.21818	0.13218	0.19465	0.10606	0.33333	0.26000	0.71175
LDF	**92**	0.18182	0.09182	0.19977	0.11175	0.33333	0.26000	0.71949
LDF	**100**	0.21818	0.13218	0.20076	0.11284	0.32692	0.25288	0.71804
INFLO	**21**	0.18182	0.09182	0.13654	0.04156	0.25806	0.17645	0.58744
INFLO	**69**	0.16364	0.07164	0.17031	0.07905	0.31818	0.24318	0.66315
INFLO	**71**	0.16364	0.07164	0.17092	0.07972	0.31638	0.24119	0.67567
INFLO	**90**	0.18182	0.09182	0.17474	0.08396	0.31502	0.23967	0.66629
COF	**19**	0.21818	0.13218	0.14304	0.04878	0.23385	0.14957	0.61909

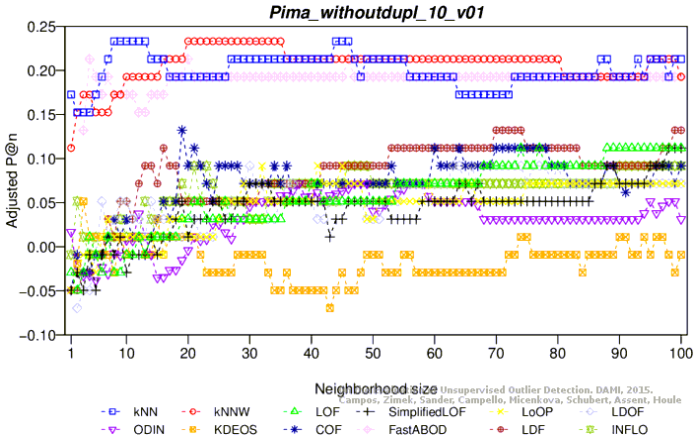
COF	**51**	0.16364	0.07164	0.18513	0.09549	0.27509	0.19535	0.68987
COF	**76**	0.20000	0.11200	0.17456	0.08376	0.30709	0.23087	0.69775
COF	**100**	0.18182	0.09182	0.17768	0.08723	0.30189	0.22509	0.69965

Plots

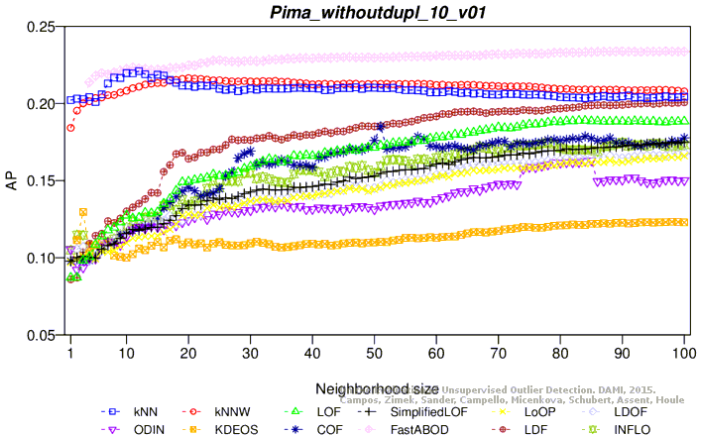
Precision at *n*



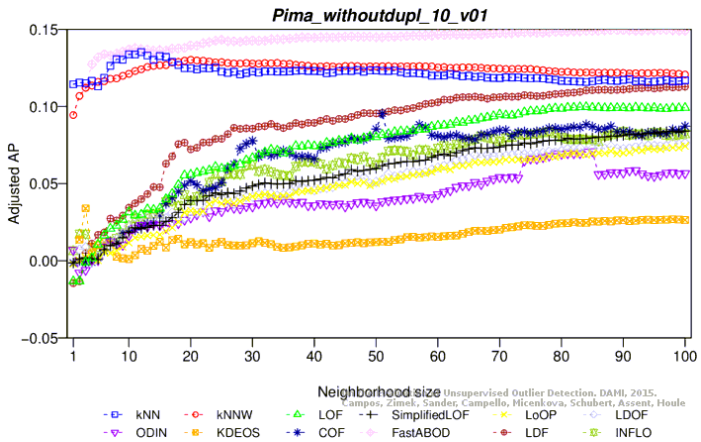
Adjusted precision at *n*



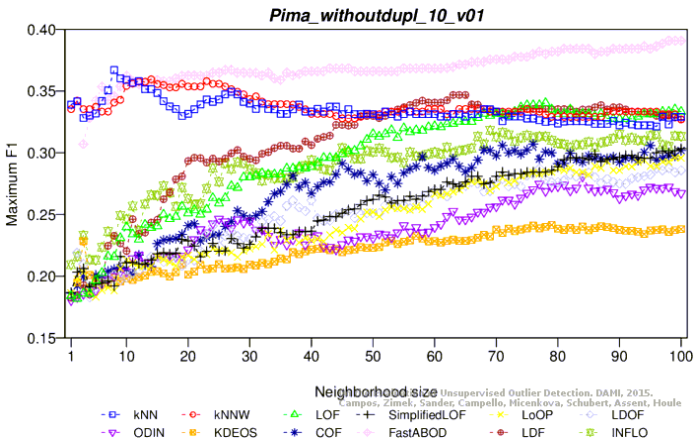
Average precision



Adjusted average precision

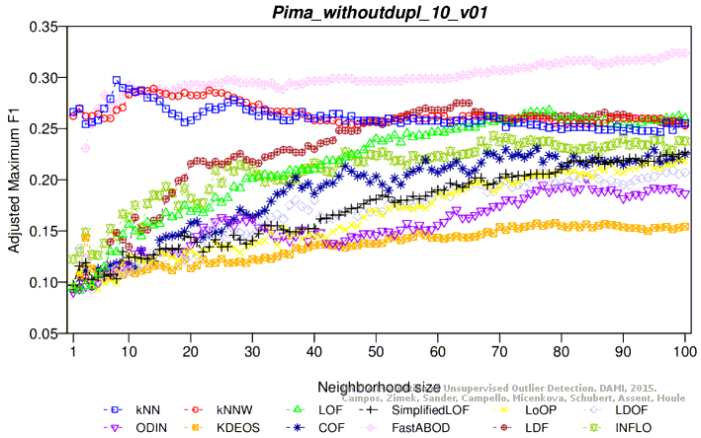


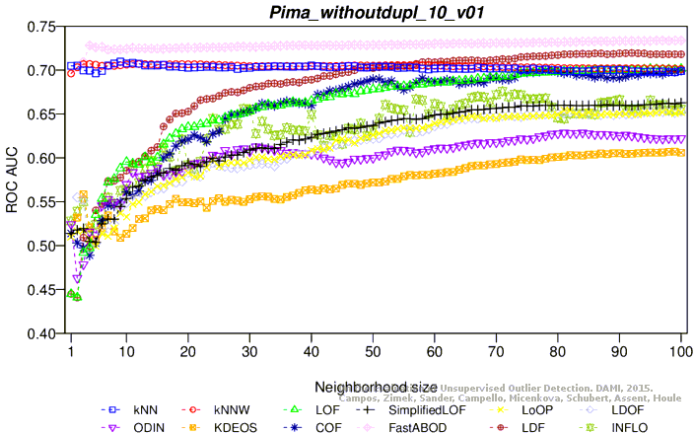
Maximum F1 score



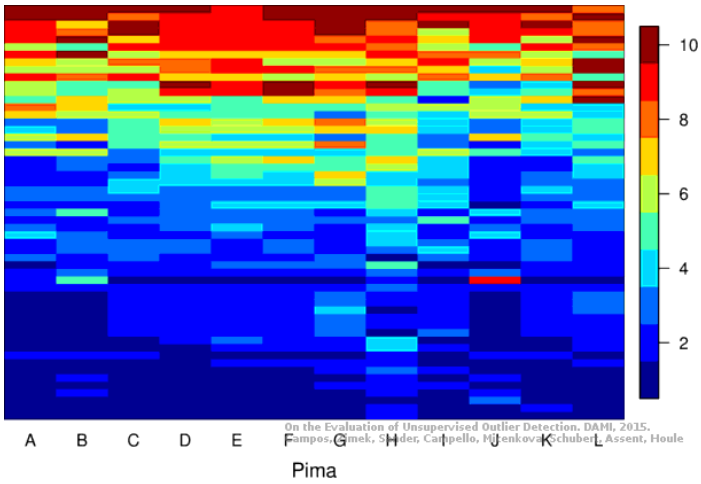
Adjusted maximum F1 score

ROC AUC





Diversity



A: KNN, B: KNNW, C: LOF, D: SimplifiedLOF, E: LoOP, F: LDOF
G: ODIN, H: KDEOS, I: COF, J: FastABOD, K: LDF, L: INFLO

File generated: 2016-07-05T21:48:59