Table 1. Datasets information

	CM1	JM1	KC1	KC2	KC3	MC1	MC2	MW1	PC1	PC2	PC3	PC4	PC5	AR1	AR6
Language	C	C	C++	C++	Java	C++	C	C	C	C	C	C	C++	C	C
LOC	20k	315k	43k	18k	18k	63k	6k	8k	40k	26k	40k	36k	164k	29k	29
Modules	505	10878	2107	522	458	9466	161	403	1107	5589	1563	1458	17186	121	101
Defects	48	2102	325	105	43	68	52	31	76	23	160	178	516	9	15

Table 2. Performance of different machine learning methods with cross validation test mode based on accuracy

	Supervised learning								**Unsupervised learning**		
Datasets	**Naye Bayes**	**MLP**	**SVM**	**Ada Boost**	**Bagging**	**Decision Trees**	**Random Forest**	**J48**	**KNN**	**RBF**	**K-means**
AR1	83.45	89.55	91.97	90.24	92.23	89.32	90.56	90.15	65.92	90.33	90.02
AR6	84.25	84.53	86.00	82.70	85.18	82.88	85.39	83.21	75.13	85.38	83.65
CM1	84.90	89.12	90.52	90.33	89.96	89.22	89.40	88.71	84.24	89.70	86.58
JM1	81.43	89.97	81.73	81.70	82.17	81.78	82.09	80.19	66.89	81.61	77.37
KC1	82.10	85.51	84.47	84.34	85.39	84.88	85.39	84.13	82.06	84.99	84.03
KC2	84.78	83.64	82.30	81.46	83.06	82.65	82.56	81.29	79.03	83.63	80.99
KC3	86.17	90.04	90.80	90.06	89.91	90.83	89.65	89.74	60.59	89.87	87.91
MC1	94.57	99.40	99.26	99.27	99.42	99.27	99.48	99.37	68.58	99.27	99.48
MC2	72.53	67.97	72.00	69.46	71.54	67.21	70.50	69.75	64.49	69.51	69.00
MW1	83.63	91.09	92.19	91.27	92.06	90.97	91.29	91.42	81.77	91.99	87.90
PC1	88.07	93.09	93.09	93.14	93.79	93.36	93.54	93.53	88.22	93.13	92.07
PC2	96.96	99.52	99.59	99.58	99.58	99.58	99.55	99.57	75.25	99.58	99.21
PC3	46.87	87.55	89.83	89.70	89.38	89.60	89.55	88.14	64.07	89.76	87.22
PC4	85.51	89.11	88.45	88.86	89.53	88.53	89.69	88.36	56.88	87.27	86.72
PC5	96.93	97.03	97.23	96.84	97.59	97.01	97.58	97.40	66.77	97.15	97.33
Mean	83.47	89.14	89.29	88.59	89.386	88.47	89.08	88.33	71.99	88.87	87.29

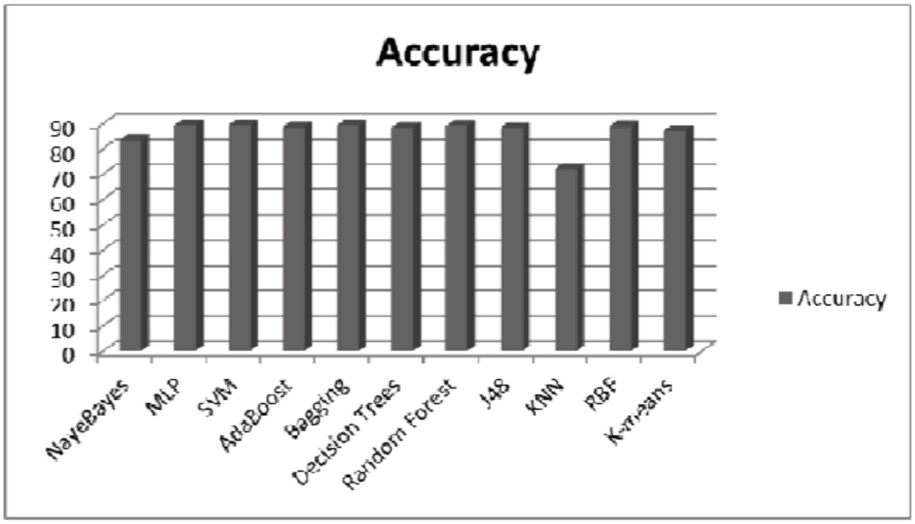

Figure 1. Accuracy results for selected machine learning methods