

(2) przykład wprowadzający

Przykład budowania modeli procesów na przykładzie zioru o irysach oraz glass. Wyniki zapisane do pliku txt oraz przy glass również image saverem.

(3) ocena klasyfikatorów

Dla zbioru klienci_2.arff

Nazwa\klasyfikator	J48	JRip	NaiveBayes	RandomForest	AdaBoostM1	Logistic
Kappa statistic	-0.0013	0.0079	0.0041	0.0239	0.0124	0.0216
Dokładność klasyfikacji	16.6 %	18.7 %	17 %	18.7 %	18.2 %	18.6 %
Mean absolute error	0.2792	0.277	0.2761	0.2759	0.2767	0.2768

Stosując dla zbioru klienci_2 różne klasyfikatory można powiedzieć, że jakość była bardzo niska. Jest to spowodowane zmiennymi jakie były dostępne w tym zbiorze. Opierając się na zbiorze klienci_6 te wyniki są trochę wyższe (np. dokładność ok 44%) ale nadal są to niezadawalające wyniki. Porównując te trzy indeksy można powiedzieć że najlepiej poradził sobie RandomForest ale nadal te wyniki są niezadawalające.

Dla zbioru bank2.arff

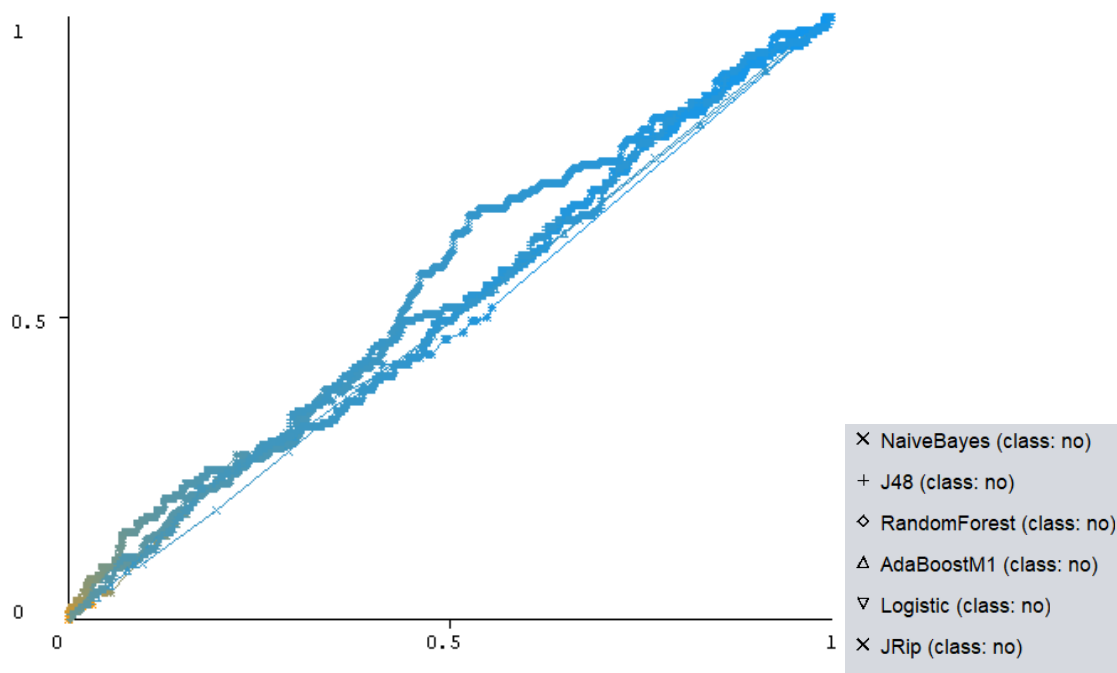
Nazwa\klasyfikator	J48	JRip	NaiveBayes	RandomForest	AdaBoostM1	Logistic
Kappa statistic	0.4592	0.3932	0.4336	0.4437	0.1988	0.3963
Dokładność klasyfikacji	89.8606 %	89.0201 %	85.2505 %	88.8557 %	87.3206 %	89.1799 %
Mean absolute error	0.1553	0.1849	0.1578	0.1524	0.1765	0.1626

Zbiór danych bankowych okazał się zdecydowanie lepszym pod względem analizy zbiorem. Dokładność klasyfikacji jest całkiem wysoka dla danych klasyfikatorów. Najlepiej poradził sobie algorytm J48(najwyższy wynik kappa) oraz największa dokładność. Średni błąd predykcji ma podobną wartość dla klasyfikatorów.

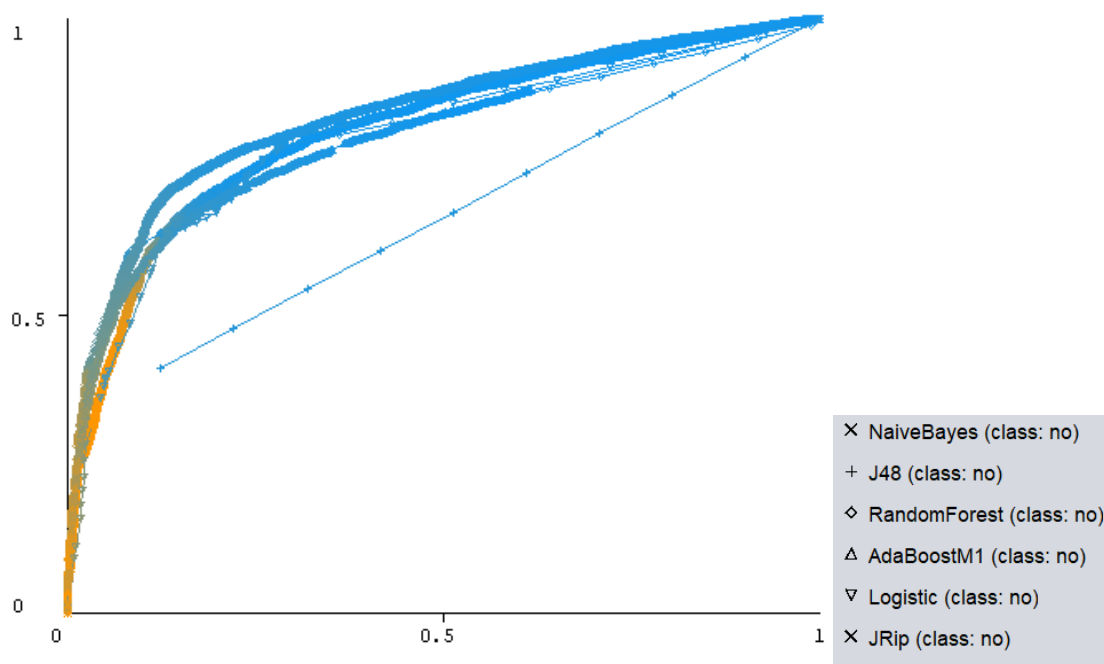
Interpretacja wyników

(4) krzywe ROC, ocena klasyfikatorów

Pierwszy model powstał na zbiorze klienci_2 z wykorzystaniem J48, JRip, NaiveBayes, RandomForest, AdaBoostM1 oraz Logistic. Gdzie jako klasę pozytywną podaliśmy najwyższy przedział wartości ' (190.833333-inf)' (w rzeczywistości od 190.83-250, nie ma zmienionych nazw przedziałów1).

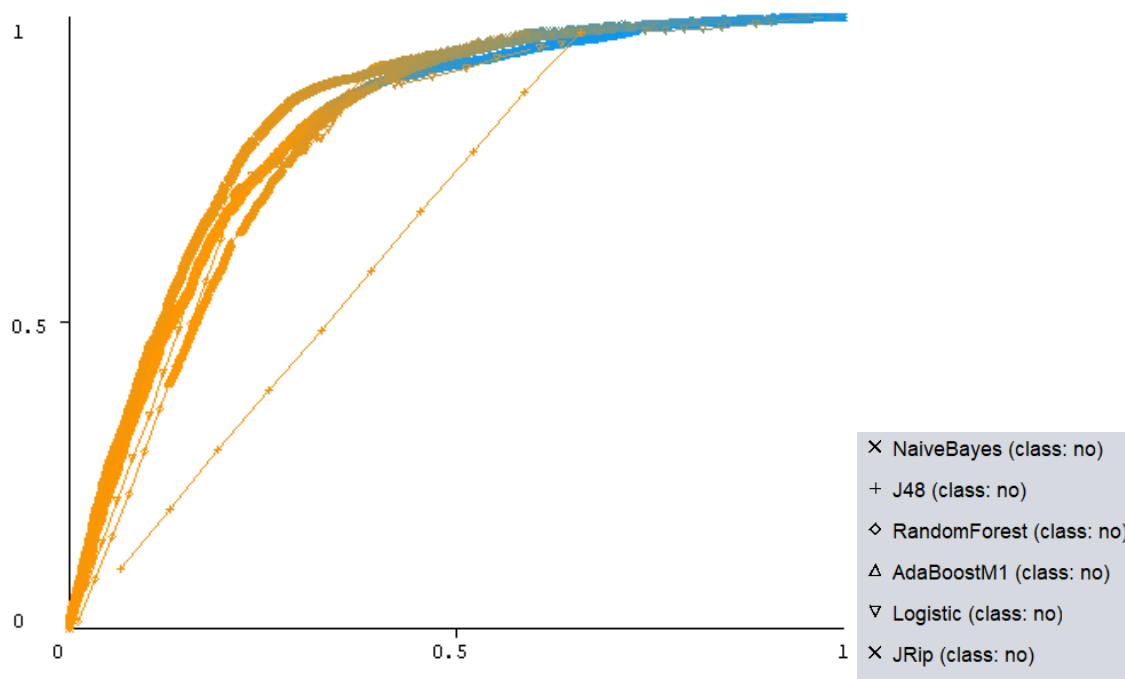


Drugi model powstał na zbiorze bank2 z wykorzystaniem klasyfikatorów J48, JRip, NaiveBayes, RandomForest, AdaBoostM1 oraz Logistic. Gdzie jako klasę pozytywną podaliśmy „yes”.



Krzywa przybiera kształt paraboliczny, wyjątkiem jest algorytm Jrip.

Trzeci model powstał na zbiorze bank2 z wykorzystaniem klasyfikatorów J48, JRip, NaiveBayes, RandomForest, AdaBoostM1 oraz Logistic. Gdzie jako klasę pozytywną podaliśmy „no”.



Tak jak w przypadku drugiego modelu(z yes) algorytm Jrip kształtuje się trochę inaczej od pozostałych. Model z wybraną klasą pozytywną „no” różni się jednak trochę od modelu z klasą wybraną jako „yes”.

(5) inne przykłady Knoledge Flow

Przetwarzanie strumieniowe na zbiorze o irysach z algorytmem IBk. Oraz Z zastosowaniem load template layout - parametrizing job na danych klienci_2 i bank2 z algorytmami (J48, JRip, NaiveBayes, RandomForest, AdaBoostM1 oraz Logistic). Tzn. zbiorami danych napływającymi strumieniowo. Wyniki zapisane do pliku txt.

Mateusz Guściora, 228884, zadanie 2.3