

Dokonując konstrukcji klasyfikatorów przyjdzie nam, że bardziej interesuje nas identyfikacja reklamy. Dane zostały zgromadzone głównie na potrzeby firm, których produkty były promowane w telewizji. Takie firmy mogły być zainteresowane zarówno kontrolą nadawania reklam dotyczących ich wyrobów, jak też badaniem posunięć konkurencji.

W celu wyboru najlepszej metody klasyfikacji przeanalizowaliśmy dokładnie 10 różnych klasyfikatorów i dokonaliśmy ich oceny na podstawie błędów klasyfikacji i innych wskaźników, opisujących stopień dopasowania do danych i efekt predykcyjny. Co się okazuje? Najlepszym klasyfikatorem są lasy losowe. Cechuje go zarówno mały błąd predykcji jak i wysoka precyzja. Nie jest to wielkim zaskoczeniem, gdyż ten klasyfikator jest oparty na tzw. komitetach drzew a i nasz rezultat jest uśredniony. Modele oparte na tzw. *collective wisdom* cechują się wysoką efektywnością.

Klasyfikatorem, który zdecydowanie nie sprawdza w klasyfikacji reklamy jest perceptron. Procent poprawnych klasyfikacji jest tylko na poziomie 70%, natomiast wykres krzywej CAP sugeruje nam, że ten klasyfikator jest niewiele lepszy od modelu losowego. Również niewielka wartość wskaźnika AUC (0.6) mówi nam, że nie mamy do czynienia ze stosunkowo słabym klasyfikatorem.