Zadanie 2.1(4\*), 228884, Mateusz Guściora

- zinterpretuj wyniki: porównaj faktyczne i przewidywane przedziały kwoty zakupu, załóż tabelę i porównaj podstawowe wskaźniki jakości (accuracy, kappa, błąd predykcji) ogółem i dla poszczególnych klas, zapisz i zinterpretuj dwie pierwsze reguły decyzyjne

Badając jak zmienne mają wpływ na naszą zmienną y(chcemy ją określi czy yest/no) algorytmem Ranked.

Search Method: Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 29 y): Information Gain Ranking Filter

Ranked attributes:

0.1245 2 CONS.PRICE.IDX

0.12422 3 CONS.CONF.IDX

0.12142 7 DataRokMiesi¹c

0.11662 13 duration

0.11632 5 NR.EMPLOYED

0.11422 28 przedzial\_nr.employed

0.11001 4 EURIBOR3M

0.09994 21 Rok

0.06602 27 przedzial\_euribor3m

0.06365 18 pdays

0.0629 19 poutcome

0.04542 26 przedzial\_cons.conf.idx

0.044 17 month

0.04391 6 miesiac

0.04169 20 previous

0.03053 25 przedzial\_cons.price.idx

0.02376 9 age

0.02131 11 contact

0.01921 15 job

0.01921 8 kategorie nieliczbowe

0.01698 23 przedzial\_wiek

0.01131 12 default

0.01128 22 przedzial\_czas rozmowy

0.00571 10 campaign

0.00502 14 education

0.00241 16 marital

0.00239 1 EMP.VAR.RATE

0.00217 24 przedzial\_emp.var.rate

Następnie uruchomiłem dwa algorytmy drzewowe(J48 i RandomForesst) i dwa regułowe(JRip i PART).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | J48 | RandomForesst | JRip | PART |
| Correctly Classified Instances | 90.9651 % | 91.5532 % | 89.9347 % | 90.634 % |
| kappa | 0.5794 | 0.5901 | 0.5498 | 0.5686 |
| Średni błąd predyckjii | 0.1169 | 0.1185 | 0.1478 | 0.1105 |

Modele przewidują, że więcej odpowiedzi no (czyli nie założenie lokaty)

//uwaga\*Sama budowa modeli(trees i rules) trwała dosyć długo. Możliwe, że na za dużo zmiennych model był budowany. Model był też budowany na całym zbiorze. Najlepszy modelem (pod względem accuracy i wsk. kappa) okazało się algorytm drzewa RandomForrest