

# Zadanie predykcyjne

*Marta Czerwińska*

*1 maja 2016*

W zadaniu predykcyjnym miałam do przeanalizowania dane dotyczące klientów operatora telekomunikacyjnego. Dane zawierały 21 zmiennych, przy czym ostatnia zmienna była flagą dotyczącą trwania lub rezygnacji danego klienta z usług firmy.

W pierwszym kroku analizy wczytuję dane, sprawdzam czy występują jakiegolwiek braki w danych (w sytuacji zaobserwowania jakiegogoś braku informacji uzupełniłabym go średnią lokalną, globalną lub medianą, w zależności od typu zmiennej) oraz wybieram zmienne najbardziej istotne z punktu widzenia analizy.

Za pomocą korelacji pomiędzy zmiennymi oraz wybraniu najbardziej istotnych spośród nich do modelowania wybieram 12 zmiennych objaśniających plus zmienną objaśnianą (zmienna Churn.).

Zmiennymi objaśniającymi są: Int.l.Plan, Day.Mins, Day.Calls, Day.Charge, Eve.Mins, Eve.Calls, Eve.Charge, Night.Mins, Night.Charge, Intl.Mins, Intl.Charge, CustServ.Calls.

W celu zbadania dokładności, precyzji oraz czułości użytych przeze mnie modeli dzielę w sposób losowy zbiór wejściowy (churn.csv) na zbiór treningowy (800 wierszy) i testowy (866 wierszy).

Przez dokładność rozumiem ile procent obserwacji dany model dobrze sklasyfikował. Precyzją jest stosunek liczy klientów prawidłowo sklasyfikowanych przez model, jako tych którzy zostali (w danych Churn.=“False”) przez wszystkich sklasyfikowanych przez model jako klientów, którzy nie zrezygnowali z usług firmy. Zaś czułością jest stosunek osób dobrze sklasyfikowanych przez model, jako tych którzy nie zrezygnowali z usług firmy do ilości osób, które w rzeczywistości nie zrezygnowały z usług firmy.

Na podstawie tych trzech miar wybiorę model, z najlepszymi wynikami. Do przetestowania wybrałam kilka metod, są nimi: LDA, GLM, naiwny Bayes, drzewa decyzyjne, bagging, lasy losowe.

Po obserwacji wyników wszystkich wyżej podanych metod, okazało się że najlepszą dokładność, precyzję i czułość mają lasy losowe. Z tego powodu do predykcji zadanego przez Państwa zadania użyłam tej metody.

Wyniki predykcji zapisałam w pliku pod nazwą “churn\_wdrozenie\_Marta\_Czerwinska.csv”.