

Zadanie domowe

Modelowanie ryzyka kredytowego

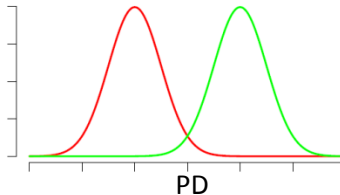
22.04.2020

Wszystkie poniższe zadania proszę wykonać w oparciu o wysymulowany zbiór danych. Rozwiązania wszystkich zadań proszę krótko opisać (max. 3 strony), można pracować w grupach 2-3 osobowych. Proszę o wysłanie raportów do 30.04.2020, 23:59. W raporcie proszę o zapisanie nazwisk wszystkich autorów. W razie jakichkolwiek pytań, zapraszam do kontaktu mailowego. Powodzenia! ☺

1. Przeprowadź wstępną analizę danych. Sprawdź, czy w danych znajdują się brakujące wartości lub wartości odstające. Jeśli tak, to jak można sobie z nimi poradzić?
2. Zbuduj model, przewidyjący prawdopodobieństwo niewypłacalności klienta (PD – probability of default). Dla przypomnienia, przykładowa składnia w języku R:

```
model <- glm(Default ~ ., data = my_data, family = binomial(link="logit"))  
data[, "PD"] <- predict(model, data, type = "response")
```

3. Stwórz krzywą CAP (Cumulative Accuracy Profile) dla powyższego modelu. Jakie jest AR (Accuracy Ratio)?
4. Stwórz wykres, przedstawiający rozkład zmiennej *Default* w zależności od prognozowanego przez model PD.



5*. Zaproponuj inne metody oceny jakości modelu.

Opis zbioru danych

Zbiór danych zawiera 8 543 obserwacje. Zmienną objaśnianą jest zmienna "*Default*".

1. "*age_x*" – wiek pierwszego kredytobiorcy,
2. "*age_y*" – wiek drugiego kredytobiorcy,
3. "*credit_history*" – informacja, czy klient posiada już historię kredytową,
4. "*kids*" – liczba dzieci klienta,
5. "*household_income*" – miesięczny dochód gospodarstwa domowego,
6. "*outstanding_loan_amount*" – pozostała wartość kredytu,
7. "*property_value*" – wartość nieruchomości,
8. "*property_size*" – rozmiar nieruchomości w metrach kwadratowych,
9. "*annual_interest_rate*" – roczna stopa procentowa kredytu,
10. "*monthly_payment*" – miesięczna rata kredytu,
11. "*n*" – liczba miesięcy pozostałych do spłaty kredytu,
12. "*Default*" – przyjmuje wartości TRUE/FALSE, informacja, czy dany klient stał się niespłacalny.