

Walidacja modeli

Adam Wróbel

Risk Modelling & Analytics Specialist



Walidacja modeli

w instytucjach finansowych

- Istnienie walidacji modeli jest wymuszone przez regulacje
- Celem jest zapewnienie, że modele są używane w odpowiedni sposób
- Adresuje to takie elementy jak to, że:
 - wszystkie modele się mylą
 - koncepcyjnie niepoprawny/niestabilny model da niepoprawne wyniki
- Szczególnie istotne obszary:
 - Wyznaczanie wymogów kapitałowych
 - Testy stresu



Cykl życia modelu w banku

Gdzie w to wpisuje się walidacja modeli?

- Model jest budowany
- Model jest walidowany:
 - Zatwierdzony
 - Zatwierdzony warunkowo model może być używany, ale pewne elementy będą musiałby zostać zmienione w przyszłości
 - Odrzucony model musi być zmieniony przed użyciem na produkcji
- Zatwierdzenie przez regulatora (zewnętrzna walidacja)



Na czym polega walidacja modeli?

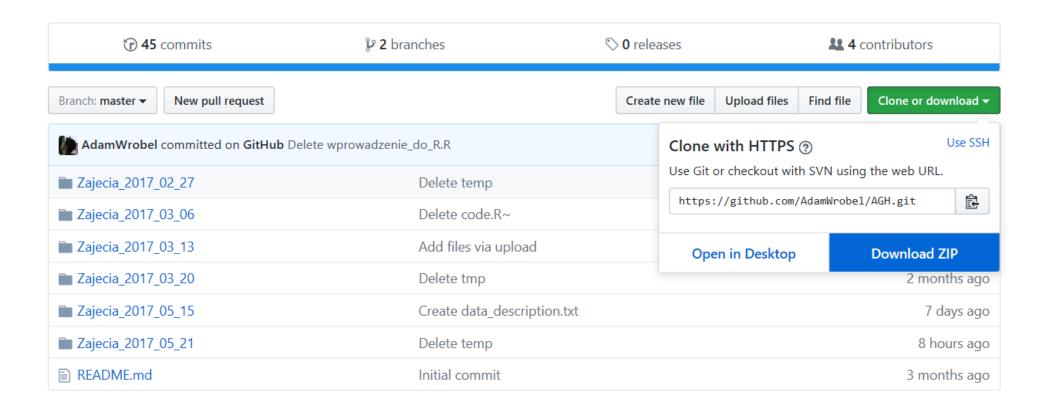
Kluczowe elementy

- Ocena czy model jest koncepcyjnie poprawnie zdefiniowany
- Podważenie wszelkich decyzji modelowych podjętych podczas modelowania:
 - Porównanie wyników modelu przy podjęciu alternatywnych opcji (np. innej funkcji łączącej)
- Sprawdzenie czy założenie stojące za modelem są spełnione:
 - Jeśli nie są spełnione to ocena jaki ma to wpływ na model (jeśli materialny to rozważenie odrzucenia model)
- Ocena czy model jest stabilny
- Budowa alternatywnego modelu (benchmarku)
- Ocena ryzyka modelu i czynników, które podnoszą to ryzyko (np. niska jakość danych)



Materialy

https://github.com/AdamWrobel/AGH/tree/master/Zajecia_2017_05_21





Model zdefiniowany na poprzednich zajęciach

Wyznaczenie prawdopodobieństwo niespłacania karty kredytowej

- Model używa historycznych danych o spłacalności kart kredytowych
- Cześć zmiennych została przekształcona
- Zależność pomiędzy zmiennymi opisującymi danego klienta, a prawdopodobieństwem tego, że
 nie spłaci karty została zdefiniowana przez regresje logistyczną:
 - Regresja logistyczna jest uogólnionym modelem liniowanym:

$$y = funkcja \, \mathrm{i} a \, \mathrm{c} z \, \mathrm{d} c \, \mathrm{d} a \, \mathrm{d} c \, \mathrm{d}$$

z logitową funkcją łączącą:

$$\operatorname{logit}(p) = \ln \frac{p}{1-p} = \ln(p) - \ln(1-p)$$

prawdopodobieństwo zdarzanie jest wtedy zdefiniowane jako:

$$p_i = rac{1}{1+e^{-(eta_1x_{1,i}+\cdots+eta_kx_{k,i})}}$$



Decyzje podjęte podczas modelowania

Wybór zmiennych

- Zmienne ilościowe kroki:
 - Uszeregowanie obserwacji pod kątem wartości
 - Zgrupowanie do 10-ciu koszyków (bins)
 - Wyznaczenie średniego poziomu DR (default rate) dla danej zmiennej w danym koszyku
 - Wygładzenie krzywej DR (loess)
 - Normalizacja wartości wygładzonej krzywej
- Zmienne jakościowe brak transformacji



Skypt: validation.R



Projekty – lista

- Forma: prezentacja na ostatnich zajęciach (12 czerwca)
- Zapisy poprzez maila: piotr-a.morawski@ubs.com
- Aktualna lista projektów: https://github.com/AdamWrobel/AGH/blob/master/Projekt/project.pdf



Informacje kontaktowe

Adam Wróbel

UBS

Risk Modelling & Analytics Specialist

Email: adam.wrobel@ubs.com

ubs.com/polandcareers

