

# Reject inference - analiza wniosków odrzuconych

Mateusz Filarowski

18 listopada 2014

# Agenda

## Pytania badawcze

- 1 Dlaczego analizujemy wnioski odrzucone?
- 2 W jaki sposób się to robi?
- 3 Czy techniki reject inference poprawiają model?

# Dlaczego analizujemy wnioski odrzucone?

Trzy słowa o obciążeniu próby

Model skoringowy uczony jest na próbie, która nie odpowiada całej populacji.

# Dlaczego analizujemy wnioski odrzucone?

Trzy słowa o obciążeniu próby

Model skoringowy uczony jest na próbie, która nie odpowiada całej populacji.

Konsekwencje:

- Wyniki modelu nie mogą zostać uogólnione na całą populację
- Z każdą iteracją modelu zmniejsza się baza klientów, do których ma on zastosowanie
- Konieczność analizy wniosków odrzuconych

# Techniki reject inference

## Ważenie danych

Cel: odzwierciedlenie populacji aplikujących przez zważenie próby zaakceptowanych klientów

### 1. Model A/R

$$\begin{bmatrix} A \\ \bar{R} \end{bmatrix}$$



### 2. Grupowanie i wyznaczenie wag

Score	Liczba Zaakceptowanych	Liczba Odrzuconych	Waga Przedziału (w)
0-149	$A_1$	$R_1$	$\frac{A_1 + R_1}{A_1}$
150-179	$A_2$	$R_2$	$\frac{A_2 + R_2}{A_2}$
180-199	$A_3$	$R_3$	$\frac{A_3 + R_3}{A_3}$
200+	$A_4$	$R_4$	$\frac{A_4 + R_4}{A_4}$



### 3. Model na przeważonych danych

$$\begin{bmatrix} A \\ w \end{bmatrix}$$

# Techniki reject inference

## Ekstrapolacja - parceling

Cel: dołączenie wniosków odrzuconych do próby, na której budowany jest model

1. Ocena aplikacji odrzuconych na bazie populacji zaakceptowanych

$[ A ] \implies [ R ] \implies$

2. Klasyfikacja spłata/default

Score	Procent Złych (Zaakceptowani)	Odrzućeni	Odrzućeni Źli
0-149	30%	150	45
150-179	25%	300	75
180-199	10%	180	18
200+	4%	400	16

$\implies$

3. Model na powiększonej próbie

$\left[ \begin{array}{c} A \\ R \end{array} \right]$

# Techniki reject inference

## Ekstrapolacja - fuzzy augmentation

Cel: dołączenie wniosków odrzuconych do próby, na której budowany jest model

### 1. Ocena aplikacji odrzuconych na bazie populacji zaakceptowanych

$$[ A ] \quad \Rightarrow \quad \begin{bmatrix} R & p \\ R & 1 - p \end{bmatrix}$$

### 2. Modelowanie na rozszerzonych i przeważonych danych

$$\Rightarrow \quad \begin{bmatrix} A & 1 \\ R & p \\ R & 1 - p \end{bmatrix}$$

# Opis badania

- Metodologia
  - Symulacja obciążenia próby
  - Porównanie modeli wykorzystujące *reject inference* z modelami zbudowanymi na pełnej oraz na obciążonej próbie
  - Kryteria porównawcze - *default rate* w grupach *pd*, prognozowana strata finansowa

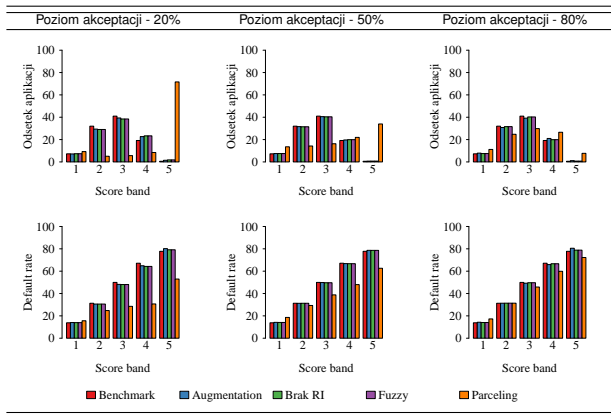


# Opis badania

- Metodologia
  - Symulacja obciążenia próby
  - Porównanie modeli wykorzystujące *reject inference* z modelami zbudowanymi na pełnej oraz na obciążonej próbie
  - Kryteria porównawcze - *default rate* w grupach *pd*, prognozowana strata finansowa
- Zbiór danych
  - Dane losowe - *por. Przanowski (2013)*
- Parametry
  - Liczba profili klienta: 1, 2
  - Poziom akceptacji: 20%, 50%, 80%
  - Parametry poszczególnych technik

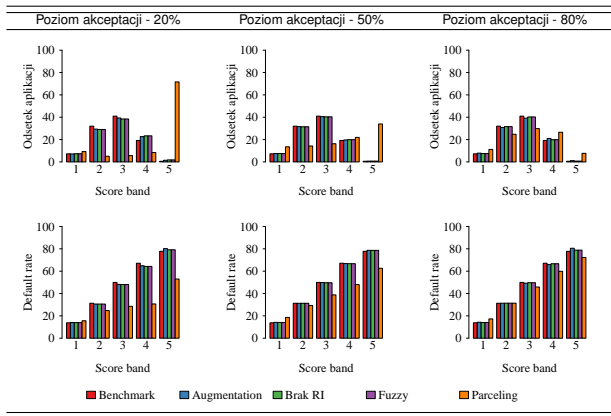
# Default w grupach pd

Jeden profil klienta



# Default w grupach pd

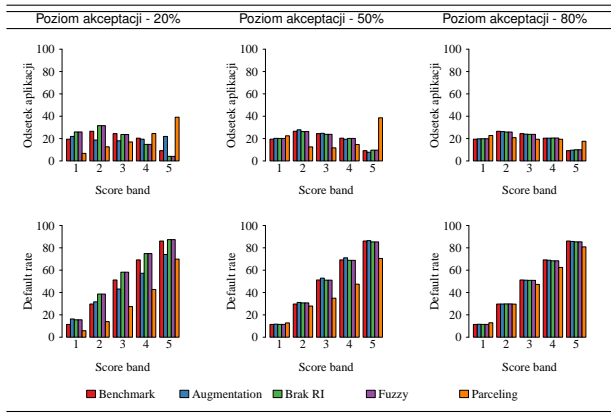
## Jeden profil klienta



- Obciążenie próby niezauważalne
- *Parceling* najbardziej konserwatywną techniką
- Pozostałe techniki *RI* nie doszacowują ryzyka

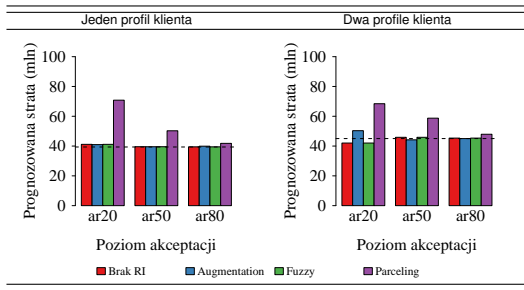
# Default w grupach pd

## Dwa profile klienta



- Obciążenie próby istotne szczególnie przy niskim progu akceptacji
- Techniki *RI* błędnie szacują ryzyko

# Prognozowana strata finansowa



- Obciążenie próby widoczne dopiero przy dwóch profilach klientów
- *RI* przydatny przy niskim poziomie akceptacji
- *Parceling* najbardziej konserwatywną techniką

# Wnioski

- Obciążenie próby i sens analizy wniosków odrzuconych zależy od konkretnego zbioru danych
- Metody *RI* nie zawsze są lepsze od metody naiwnej
- Zastosowanie konkretnej techniki *RI* może istotnie wpłynąć na ocenę ryzyka i zyskowność portfela kredytowego