

Modelowanie dla biznesu : regresja logistyczna, regresja Poissona, Survival Data Mining, CRM, credit scoring / redakcja naukowa Ewa Frątczak. – Wydanie 1. – Warszawa, 2019

Spis treści

Ewa Frątczak

Przedmowa	11
------------------	-----------

Ewa Frątczak

Rozdział 1. Regresja logistyczna - podstawy teorii i modelowania	17
---	-----------

Wprowadzenie	17
--------------	----

1.1. Regresja logistyczna - początki	18
--------------------------------------	----

1.2. Regresja logistyczna - podstawy teorii	21
---	----

1.3. Model regresji logistycznej a model regresji liniowej	26
--	----

1.4. Estymacja parametrów modelu regresji logistycznej (na przykładzie regresji binarnej)	29
---	----

1.5. Interpretacja wyników estymacji	31
--------------------------------------	----

1.6. Model uporządkowany i model o postaci wielomianu	32
---	----

1.6.1. Regresja logistyczna - model uporządkowany	32
---	----

1.6.2. Regresja logistyczna o postaci wielomianu - uogólnione logity	37
--	----

1.7. Znaczenie kodowania zmiennych jakościowych w modelach regresji logistycznej	38
--	----

1.8. Istota wnioskowania statystycznego w modelach regresji logistycznej	40
--	----

1.8.1. Ocena istotności oszacowanych parametrów	41
---	----

1.8.2. Ocena dobroci modelu	42
-----------------------------	----

1.9. Predykcja i modelowanie predykcyjne z wykorzystaniem regresji logistycznej	49
---	----

1.10. Modele predykcyjne - diagnostyka - wybrane miary	52
--	----

1.10.1. Macierz pomyłek	52
-------------------------	----

1.10.2. Krzywa ROC i AUC (pole pod krzywą ROC)	55
--	----

1.10.3. Krzywa wzrostu lift	57
-----------------------------	----

1.10.4. Współczynnik Giniego	60
------------------------------	----

1.10.5. Statystyki wpływu (DFBETAS, DIFDEV, DIFCHISQ, LEVERAGE)	60
---	----

Podsumowanie	61
--------------	----

Bibliografia	61
--------------	----

Aleksandra Iwanicka

Rozdział 2. Analiza tablic kontyngencji - ocena zależności cech jakościowych	67
---	-----------

Wprowadzenie	67
--------------	----

2.1. Testowanie niezależności na podstawie tablic dwudzielczych	71
---	----

2.1.1. Testy chi-kwadrat: Pearsona i ilorazu wiarygodności	71
2.1.2. Testy skorygowane i dokładne	72
2.1.3. Miary zależności oparte na statystyce chi-kwadrat Pearsona	73
2.2. Testowanie niezależności na podstawie zbioru tablic dwudzielczych	74
2.2.1. Statystyka ogólnej asocjacji	76
2.2.2. Statystyka różnicy średnich w wierszach	77
2.2.3. Statystyka korelacji	78
2.3. Pomiar siły asocjacji	78
2.3.1. Miary asocjacji dla cech dychotomicznych	79
2.3.2. Miary asocjacji dla cech porządkowych	81
2.3.3. Miary asocjacji dla cech nominalnych	84
2.4. Przykłady zastosowań	87
2.4.1. Badanie związku pomiędzy cechami dychotomicznymi	88
2.4.2. Badanie zależności pomiędzy cechami nominalnymi	91
2.4.3. Badanie zależności pomiędzy cechami porządkowymi	93
2.4.4. Badanie zależności w przypadku stratyfikacji	96
Podsumowanie	99
Bibliografia	100

Barbara Szewczak, Wojciech Skwirz

Rozdział 3. Regresja logistyczna - model deskryptywny i model predykcyjny	101
Wprowadzenie	101
3.1. Opis procedur SAS	104
3.1.1. Procedury SAS	104
3.1.2. Autorskie makra SAS	106
3.2. Model deskryptywny	109
3.2.1. Metodyka budowy modelu	110
3.2.2. Analiza braków danych występujących w zmiennych objaśniających	110
3.2.3. Analiza korelacji oraz wstępna ocena predykcyjności zmiennych objaśniających	113
3.2.4. Analiza współliniowości zmiennych objaśniających	117
3.2.5. Selekcja zmiennych objaśniających	120
3.2.6. Analiza obserwacji wpływowych i odstających	122
3.2.7. Finalny model - postać i interpretacja	125
3.2.8. Interpretacja parametrów modelu	127
3.2.9. Istotność zmiennych w modelu	128
3.2.10. Ilorazy szans z modelu	128
3.2.11. Istotność czynników w modelu	129
3.2.12. Globalny test na istotność modelu	130
3.2.13. Kryteria informacyjne modelu	131
3.2.14. Analiza współliniowości zmiennych w modelu	131
3.2.15. Statystyki mocy dyskryminacyjnej modelu oraz analizy graficzne	133
3.2.16. Finalny model - podsumowanie	140

3.3. Model predykcyjny	140
3.3.1. Motywacja do budowy modelu predykcyjnego	140
3.3.2. Etapy budowy modelu	143
3.3.3. Finalny model - postać i interpretacja	155
3.3.4. Porównanie modelu deskryptywnego i modelu predykcyjnego	157
Bibliografia	158
Załącznik	159

Mikołaj Tchorzewski

Rozdział 4. Modelowanie wyników ligi piłkarskiej z wykorzystaniem regresji binarnej i regresji uporządkowanej	185
Wprowadzenie	185
4.1. Charakterystyka rozgrywek Premier League	185
4.2. Opis danych	189
4.3. Zmienne do modelowania	194
4.3.1. Zmienne zależne dla modeli ligi angielskiej	194
4.3.2. Wybrane zmienne objaśniające dla modeli ligi angielskiej	195
4.3.3. Wybrane kody z etapu przetwarzania danych	200
4.4. Model regresji binarnej - estymacja i interpretacja wyników	207
4.4.1. Estymacja modelu regresji binarnej	207
4.4.2. Kod finalnego modelu regresji binarnej	207
4.4.3. Model regresji binarnej dla Premier League	209
4.5. Model regresji uporządkowanej - estymacja i interpretacja wyników	219
4.5.1. Estymacja modelu regresji uporządkowanej	219
4.5.2. Kod finalnego modelu regresji uporządkowanej	220
4.5.3. Model regresji uporządkowanej dla Premier League	221
Podsumowanie	228
Bibliografia	229

Piotr Rozenbajgier

Rozdział 5. Model regresji Poissona	231
Wprowadzenie	231
5.1. Estymacja regresji Poissona w SAS	236
5.2. Problem nadmiernej dyspersji (overdispersion)	245
5.3. Rozkład ujemny dwumianowy	249
Bibliografia	259

Justyna Czaja

Rozdział 6. Modelowanie na użytek kampanii marketingu bezpośredniego	261
Wprowadzenie	261
6.1. Marketing bezpośredni - definicje i podstawowe pojęcia	262
6.2. Rola modeli predykcyjnych w marketingu	264
6.3. Znaczenie wskaźnika otwarcia dla sukcesu kampanii marketingu bezpośredniego	266

6.4. Podstawy teoretyczne zastosowanych metod	267
6.4.1. Model regresji logistycznej - metoda karty scoringowej	267
6.4.2. Model sztucznej sieci neuronowej	270
6.4.3. Model drzewa decyzyjnego	272
6.5. Opis zbioru danych	273
6.6. Selekcja zmiennych objaśniających do ostatecznych modeli	281
6.7. Model regresji logistycznej - analiza praktyczna	290
6.8. Model sztucznej sieci neuronowej - analiza praktyczna	297
6.9. Model drzewa decyzyjnego - analiza praktyczna	300
6.10. Porównanie modeli na zbiorze walidacyjnym	303
6.11. Praktyczne zastosowanie wybranego modelu	306
Podsumowanie	310
Bibliografia	311

Adrianna Wołowicz

Rozdział 7. Modelowanie churnu. Regresja logistyczna i <i>survival data mining</i>	313
7.1. Wprowadzenie do modelowania churnu	313
7.1.1. Podstawy teoretyczne zagadnienia churnu	313
7.1.2. Podejścia stosowane w analizie churnu	314
7.2. Opis zbioru danych	316
7.3. Model regresji logistycznej	318
7.3.1. Wstępna analiza i ocena zmiennych objaśniających	318
7.3.2. Wyniki modelu regresji logistycznej	321
7.4. Model <i>survival data mining</i>	326
7.4.1. Podstawy teoretyczne	326
7.4.2. Wyniki analizy	328
7.4.3. Ocena jakości modelu	332
7.5. Podsumowanie wyników	334
Podsumowanie	335
Bibliografia	336

*Marta Jasiewicz-Badowska, Karol Przanowski, Bartosz Staniów,
Krystyna Rup*

Rozdział 8. Dynamiczne segmentacje klientów korporacyjnych i zespołów sprzedażowych metodami credit scoringu	337
Wprowadzenie	337
8.1. Wprowadzenie do credit scoringu	343
8.1.1. Początki credit scoringu	343
8.1.2. Etapy budowy modelu scoringowego	344
8.1.3. Definicja <i>default</i>	345
8.1.4. Struktura danych - ABT i próby modelowe	346
8.1.5. Podział zmiennych na kategorie	347
8.1.6. Selekcja zmiennych i ich raporty	348
8.1.7. Ocena kandydatów i wybór finalnego modelu	349

8.1.8. Wdrożenie modelu i monitorowanie poprawności jego działania	350
8.2. Segmentacja sieci aptecznych metodami credit scoringu	351
8.2.1. Zarys specyfiki rynku farmaceutycznego	351
8.2.2. Wzbogacenie procesu zarządzania sieciami z wykorzystaniem klasycznych modeli scoringowych	357
8.2.3. Obszar aplikacji modelu	359
8.2.4. Modelowanie rankingu efektywności per apteka	361
8.2.5. Interpretacja wybranych zmiennych w modelu	362
8.2.6. Budowa komplementarnego procesu analitycznego - BI i modelowanie predykcyjne	368
8.2.7. Szkolenie z podstaw teoretycznych, prezentacja narzędzia oraz obszary do zastosowania jako newralgiczny punkt we wdrożeniach innowacyjnych rozwiązań	373
8.2.8. Dywersyfikacja strategii biznesowych i komunikacja oparta na porządkowaniu rzeczywistości	376
8.3. Credit scoring dla przedstawicieli medycznych	379
8.3.1. Wprowadzenie do uwarunkowań biznesowych w zarządzaniu zespołem sprzedażowym dedykowanym do rynku farmaceutycznego	379
8.3.2. Credit scoring dla przedstawicieli medycznych - przybliżenie <i>business case</i>	389
8.3.3. Testy pierwszych modeli - wybór właściwego momentu startu	391
8.3.4. Aktualizacja modelu dla przedstawicieli medycznych	398
8.3.5. Interpretacja wybranych zmiennych	400
8.3.6. Segmentacja PM - rola w zarządzaniu zespołami konsultantów medycznych	407
Bibliografia	412
Biogramy	415