INTERPRETOWALNOŚĆ I WYJAŚNIALNOŚĆ UCZENIA MASZYNOWEGO

Dr Robert Małysz

WYKŁAD 2 - AGENDA

1. Obszary wymagające wyjaśnialności i interpretowalności modeli AI
2. Praktyczne przykłady modeli interpretowalnych
   - Karta scoringowa vs model ratingowy
   - Etapy budowy karty scoringowej
   - Selekcja cech
   - Transformacja zmiennych – przykład wykorzystania WoE
   - Kalibracja karty scoringowej

WYKŁAD 2

1. Medycyna i Opieka Zdrowotna

Diagnostyka: Lekarze muszą rozumieć, dlaczego model AI zaleca określoną diagnozę lub plan leczenia, aby móc podejmować świadome decyzje i ewentualnie wyjaśnić je pacjentowi.

Badania Kliniczne: Zrozumienie, jakie cechy wpływają na prognozy modelu, może pomóc naukowcom odkrywać nowe zależności i mechanizmy w danych medycznych.

2. Finanse

Kredyty: Banki i inne instytucje finansowe muszą być w stanie wyjaśnić, dlaczego wniosek kredytowy został zaakceptowany lub odrzucony, aby spełniać wymogi regulacyjne i zapewniać sprawiedliwość (brak dyskryminacji).

Algorytmy Tradingowe/Handlowe: Inwestorzy i regulatorzy mogą wymagać zrozumienia, jak algorytmy handlowe podejmują decyzje, aby zapewnić uczciwość i stabilność rynku.

3. Prawo

Systemy Wykrywania Przestępczości: Zrozumienie, dlaczego system AI flaguje pewne działania jako podejrzane, jest kluczowe dla dalszych śledztw i unikania fałszywych alarmów.

Prognozowanie Ryzyka Przestępczości: Sędziowie i pracownicy systemu sądowego mogą potrzebować wyjaśnień dotyczących ocen ryzyka przestępczości, aby podejmować sprawiedliwe decyzje.

WYKŁAD 2

4. Automatyka Procesów Przemysłowych

Bezpieczeństwo: Inżynierowie muszą rozumieć, dlaczego system AI podejmuje pewne decyzje, zwłaszcza w kontekście bezpieczeństwa i awarii, aby móc interweniować i optymalizować procesy.

Optymalizacja Produkcji: Zrozumienie, jakie zmienne wpływają na prognozy modelu, może pomóc w identyfikacji obszarów do poprawy i optymalizacji.

5. Rekrutacja i Zarządzanie Zasobami Ludzkimi

Selekcja Kandydatów: HR musi być w stanie wyjaśnić, dlaczego pewni kandydaci są preferowani przez system AI, aby unikać dyskryminacji i zapewniać sprawiedliwość w procesie rekrutacji.

Ocena Wydajności: Pracownicy mogą wymagać wyjaśnień dotyczących ocen wydajności generowanych przez modele AI, aby zrozumieć, jak mogą się poprawić.

6. Edukacja

Systemy Rekomendacji: Edukatorzy i studenci mogą chcieć wiedzieć, dlaczego pewne materiały lub kursy są rekomendowane, aby dostosować ścieżki edukacyjne.

Ocena Automatyczna: Nauczyciele i uczniowie mogą potrzebować wyjaśnień dotyczących automatycznie przyznawanych ocen, aby zapewnić sprawiedliwość i jakość edukacji.

WYKŁAD 2

7. Marketing i Reklama

Personalizacja: Marketerzy mogą chcieć zrozumieć, dlaczego pewne produkty lub reklamy są rekomendowane konkretnym klientom, aby optymalizować kampanie.

Segmentacja Klientów: Zrozumienie, jakie cechy wpływają na segmentację klientów, może pomóc w tworzeniu bardziej skutecznych strategii marketingowych.

WYKŁAD 2 – SCORING VS RATING

Model Scoringowy

Model scoringowy to narzędzie używane do oceny ryzyka kredytowego klienta lub innego rodzaju ryzyka finansowego. Model ten opiera się na analizie statystycznej i matematycznej danych historycznych oraz informacji o potencjalnym kliencie, aby przewidzieć prawdopodobieństwo określonego zachowania, takiego jak np. niespłacenie kredytu.

W kontekście kredytów, model scoringowy może uwzględniać różne zmienne, takie jak: Historia kredytowa, Dochód, Wydatki, Zatrudnienie, Wiek, Wykształcenie

Wynik scoringowy, często nazywany wynikiem kredytowym, jest liczbową wartością, która reprezentuje ryzyko związane z konkretnym klientem lub transakcją. Wyższy wynik często oznacza niższe ryzyko.

WYKŁAD 2 – SCORING VS RATING

Model Aplikacyjny Scoringowy

Model aplikacyjny scoringowy, często nazywany po prostu modelem aplikacyjnym, jest rodzajem modelu scoringowego, który jest używany do oceny ryzyka kredytowego klienta w momencie składania wniosku o kredyt. Ten model ocenia prawdopodobieństwo, że klient, który składa wniosek o kredyt, będzie w stanie spłacić zadłużenie zgodnie z ustalonymi warunkami.

Model aplikacyjny bierze pod uwagę różne zmienne, które mogą obejmować:

- Informacje zawarte we wniosku kredytowym (np. dochód, wydatki, zatrudnienie, wiek)
- Dane zewnętrzne, takie jak raporty kredytowe
- Czasami dane demograficzne

Celem modelu aplikacyjnego jest przewidzenie prawdopodobieństwa, że klient, który składa wniosek, będzie "dobrym" kredytobiorcą, co oznacza, że będzie w stanie spłacić kredyt zgodnie z ustalonymi warunkami.

WYKŁAD 2 – SCORING VS RATING

Model Behawioralny Scoringowy

Model behawioralny scoringowy, z kolei, koncentruje się na zachowaniu kredytowym klienta po uzyskaniu kredytu. Jest to model, który jest używany do monitorowania zachowania kredytobiorców i oceny ich zdolności do spłaty kredytu na podstawie ich zachowań transakcyjnych i płatniczych.

Model behawioralny może uwzględniać zmienne takie jak:

- Historia płatności klienta
- Użycie dostępnego kredytu
- Liczba i rodzaj kont kredytowych
- Historia zapytań kredytowych

oraz inne zmienne związane z zachowaniami finansowymi klienta

Celem modelu behawioralnego jest identyfikacja zmian w zachowaniu płatniczym klienta, które mogą sygnalizować zwiększone ryzyko niespłacenia kredytu.

WYKŁAD 2 – SCORING VS RATING

Model Ratingowy

Model ratingowy to system oceny, który jest używany do klasyfikowania różnych jednostek (takich jak kraje, przedsiębiorstwa, czy produkty finansowe) na podstawie ich jakości lub ryzyka. Model ratingowy jest często stosowany w kontekście oceny zdolności kredytowej emitentów papierów wartościowych lub krajów.

W kontekście oceny ryzyka kredytowego, agencje ratingowe, takie jak Standard & Poor's, Moody's i Fitch Ratings, stosują modele ratingowe do przypisania ocen kredytowych, które odzwierciedlają zdolność emitenta do spłaty długu. Ocenę kredytową można przedstawić za pomocą liter i symboli, takich jak "AAA" dla najwyższej jakości kredytowej lub "D" dla defaultu.

WYKŁAD 2 – SCORING VS RATING

Model scoringowy jest częściej używany do oceny indywidualnych konsumentów i jest wyrażony jako liczba.

Model ratingowy jest częściej używany do oceny firm, produktów finansowych, czy krajów i jest wyrażony jako symbol (np. literowy).

Oba modele są używane do oceny ryzyka, ale są stosowane w różnych kontekstach i mają różne formy wyjściowe. Oba modele są kluczowe w zarządzaniu ryzykiem finansowym i są używane przez różne instytucje finansowe, inwestorów i regulatorów.

WYKŁAD 2 – SCORING VS RATING PORÓWNANIE

Aspekt	Model Scoringowy	Model Ratingowy
Zastosowanie	Indywidualni konsumenci	Firmy, kraje, produkty finansowe
Forma wyniku	Liczba (np. 300-850)	Symbol (AAA, BB+, D)
Przykład	Score kredytowy: 720/850	Rating obligacji: AA-
Główni użytkownicy	Banki, firmy pożyczkowe	S&P, Moody's, Fitch
Częstotliwość	Automatyczna, real-time	Okresowa, z analizą ekspercką
Typ modelu	Aplikacyjny / Behawioralny	Kompleksowa ocena

Model Aplikacyjny

Ocena w momencie składania wniosku o kredyt

Dane: wiek, dochód, historia kredytowa

Model Behawioralny

Monitoring zachowania po uzyskaniu kredytu

Dane: historia płatności, użycie kredytu

WYKŁAD 2 – SCORING VS RATING PODSUMOWANIE (1/2)

Cecha	Scoring	Rating
Oceniany podmiot	Osoby fizyczne (klienci indywidualni).	Podmioty o dużej skali działania, takie jak: duże przedsiębiorstwa, instytucje finansowe, a także rządy państw.
Cel oceny	Ocena prawdopodobieństwa spłaty indywidualnego kredytu przez klienta detalicznego.	Ocena ryzyka kredytowego związanego z inwestowaniem w papiery dłużne (np. obligacje) emitowane przez dany podmiot.
Metodologia	Automatyczna i ilościowa. Opiera się na algorytmach, które analizują dane liczbowe, takie jak historia spłat, wysokość zadłużenia, liczba zapytań kredytowych.	Bardziej złożona i jakościowa. Obejmuje analizę wielu czynników, w tym sytuacji finansowej, wyników ekonomicznych, perspektyw rozwoju branży, a nawet ryzyka politycznego i społecznego.
Forma oceny	Wynik punktowy, np. od 0 do 100 punktów (jak w Biurze Informacji Kredytowej - BIK).	System literowy, np. od AAA (najwyższa wiarygodność) do D (niewypłacalność), nadawany przez wyspecjalizowane agencje.

WYKŁAD 2 – SCORING VS RATING
PODSUMOWANIE (2/2)

Cecha	Scoring	Rating
Podmiot oceniający	Najczęściej instytucje finansowe, np. banki, oraz biura informacji kredytowej (w Polsce: BIK).	Wyspecjalizowane agencje ratingowe, takie jak Moody's, Fitch Ratings czy Standard & Poor's.
Zakres oceny	Węższy, skupiony na historii kredytowej i finansach osobistych.	Szeroki, obejmujący całościową analizę kondycji ekonomicznej i perspektyw podmiotu.
Charakter oceny	Ocena wewnętrzna, przeznaczona głównie na użytek instytucji finansowych do podejmowania decyzji kredytowych.	Ocena publiczna, przeznaczona dla szerokiego grona inwestorów i uczestników rynku finansowego.

WYKŁAD 2 – PRZYKŁAD KARTY SCORINGOWEJ

Karta Scoringowa - Kredyt Konsumencki

Zakres punktów: 300-850 | Próg akceptacji: 620 punktów

1. Wiek	Punkty
18-25 lat	+20
26-35 lat	+45
36-50 lat	+60
Powyżej 50 lat	+75

2. Dochód miesięczny	Punkty
Poniżej 3000 zł	+15
3000-5000 zł	+40
5000-8000 zł	+65
Powyżej 8000 zł	+85

3. Historia kredytowa	Punkty
Brak historii	+30
Opóźnienia > 90 dni	+10
Opóźnienia 30-90 dni	+50
Bez opóźnień	+95

4. Rodzaj zatrudnienia	Punkty
Bezrobotny	+5
Umowa zlecenie	+25
Działalność gosp.	+40
Umowa o pracę	+70

5. Liczba aktywnych kredytów	Punkty
Powyżej 4	+10
3-4 kredyty	+35
1-2 kredyty	+60
Brak kredytów	+45

Przykład obliczenia:

Klient: 32 lata, dochód 6500 zł, bez opóźnień, umowa o pracę, 1 kredyt

Wiek (26-35): **+45 pkt**
Dochód (5000-8000): **+65 pkt**
Historia (bez opóźnień): **+95 pkt**
Zatrudnienie (umowa): **+70 pkt**
Liczba kredytów (1-2): **+60 pkt**

WYNIK KOŃCOWY: 335 punktów
❌ **ODRZUCONY (próg: 620)**

Interpretacja progów:

300-450: Bardzo wysokie ryzyko | **450-550:** Wysokie ryzyko | **550-650:** Średnie ryzyko | **650-850:** Niskie ryzyko

WYKŁAD 2 – BUDOWA KARTY SCORINGOWEJ

1. Zbieranie Danych

Zbieranie danych: Zebranie danych historycznych dotyczących klientów oraz zmiennych, które mogą wpływać na zdolność kredytową.

Czyszczenie danych: Usuwanie błędów, brakujących wartości i outlierów.

Podział danych: Podział danych na zbiór treningowy, walidacyjny oraz testowy.

2. Wybór Zmiennych

Analiza zmiennej docelowej: Zrozumienie i definicja zmiennej, którą chcemy przewidzieć.

Selekcja zmiennych: Wybór zmiennych, które będą używane do budowy modelu, na podstawie analizy korelacji, ważności zmiennych i Wartości Informacyjnej (IV).

3. Przekształcenie Zmiennych z Wykorzystaniem WoE

Binning: Podział zmiennej ciągłej na kategorie (binning) lub grupowanie kategorii zmiennej nominalnej.

Obliczenie WoE: Dla każdego binu obliczana jest Waga Dowodów (WoE) według wzoru:

WoE=ln(Prawdopodobienstwo Dobrego/Prawdopodobienstwo Złego)

Zastąpienie zmiennych: Zamiana oryginalnych zmiennych przez odpowiadające im wartości WoE.

WYKŁAD 2 – BUDOWA KARTY SCORINGOWEJ

4. Budowa Modelu

Wybór modelu: Wybór odpowiedniego modelu statystycznego, często jest to regresja logistyczna.

Trenowanie modelu: Uczenie modelu na zbiorze treningowym, używając zmiennych przekształconych przez WoE.

Ocena modelu: Sprawdzenie jakości modelu na zbiorze testowym, często przy użyciu krzywej ROC i AUC/Gini.

5. Tworzenie Karty Scoringowej

Punktacja: Przekształcenie wyników modelu na punkty scoringowe. Każdej zmiennej przypisywana jest określona liczba punktów, które są proporcjonalne do jej wpływu na prawdopodobieństwo zdarzenia niewypłacalności (probability of default).

Kalibracja: Ustalenie progów punktacji i przypisanie kategorii ryzyka (np. niskie, średnie, wysokie ryzyko) na podstawie uzyskanych wyników.

Walidacja: Sprawdzenie, czy karta scoringowa jest stabilna i czy zachowuje swoją moc predykcyjną na różnych próbkach danych.

WYKŁAD 2 – BUDOWA KARTY SCORINGOWEJ

6. Implementacja i Monitorowanie

Implementacja: Wdrożenie modelu i karty scoringowej do systemu decyzyjnego.

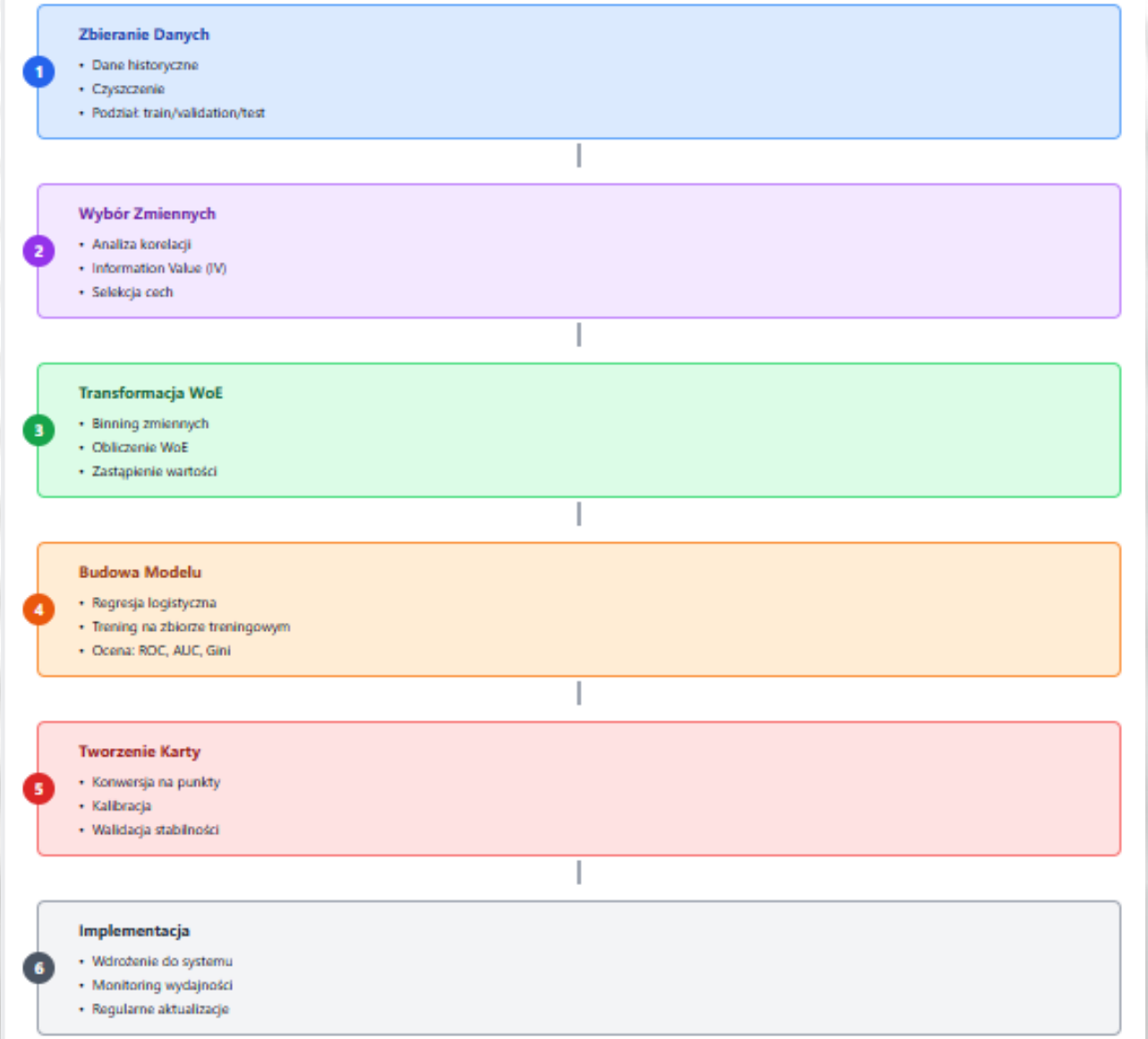
Monitorowanie: Śledzenie wydajności modelu i karty scoringowej w czasie, oraz dokonywanie korekt w razie potrzeby.

7. Utrzymanie Modelu

Aktualizacja: Regularna aktualizacja modelu i karty scoringowej, aby dostosować je do zmieniających się warunków rynkowych i profilu klienta.

Dokumentacja: Utrzymanie dokładnej dokumentacji dotyczącej modelu, zmian i wyników monitorowania.

WYKŁAD 2 – DIAGRAM BUDOWY KARTY



WYKŁAD 2 – SELEKCJA ZMIENNYCH

1. Selekcja Zmiennych (Feature Selection)

Selekcja zmiennych oparta na istotności: Wybór zmiennych na podstawie statystyk (np. test chi-kwadrat, test F) lub ważności cech (np. ważności zmiennych w drzewach decyzyjnych).

Selekcja zmiennych oparta na algorytmach: Użycie algorytmów, takich jak algorytmy zachłanne (np. metoda krokowej selekcji zmiennych) lub algorytmy oparte na regularyzacji (np. LASSO, Ridge).

Selekcja zmiennych oparta na metodach rekurencyjnych: Metody, takie jak rekurencyjna eliminacja zmiennych (RFE), które iteracyjnie usuwają zmienne i budują model.

2. Ekstrakcja Cech (Feature Extraction)

Analiza głównych składowych (PCA): Transformacja zmiennych do nowej przestrzeni, w której nowe zmienne (główne składowe) są liniowymi kombinacjami oryginalnych zmiennych i są od siebie niezależne.

Analiza dyskryminacyjna (LDA): Zmniejszenie wymiarowości danych, maksymalizując rozróżnialność między klasami.

Autoenkodery: Użycie sieci neuronowych (autoenkoderów) do nauczenia się kompresji danych w sposób nieliniowy.

WYKŁAD 2 – SELEKCJA ZMIENNYCH

3. Redukcja Wymiarowości oparta na Modelu

Random Projection: Redukcja wymiarowości przez rzutowanie danych na losowo wygenerowaną niższą wymiarową przestrzeń.

t-SNE lub UMAP: Techniki redukcji wymiarowości, które są szczególnie przydatne w wizualizacji danych wysokowymiarowych.

4. Inżynieria Cech (Feature Engineering)

Tworzenie cech: Tworzenie nowych cech na podstawie istniejących, które mogą lepiej reprezentować problem.

Binning: Grupowanie wartości zmiennych numerycznych w „kosze”.

Kodowanie zmiennych kategorycznych: Przekształcanie zmiennych kategorycznych w formę, którą model może łatwiej przetworzyć, np. one-hot encoding.

WYKŁAD 2 – SELEKCJA ZMIENNYCH

5. Metody Ensemble

Random Forest: Automatycznie wykonuje pewną formę selekcji cech poprzez losowe wybieranie podzbioru cech do rozważenia przy każdym podziale.

XGBoost: Posiada wbudowane mechanizmy do automatycznej selekcji cech poprzez regularyzację.

6. Metody Filtracyjne (Filter Methods)

Korelacja z Zmienną Docelową: Usuwanie zmiennych, które mają niską korelację ze zmienną docelową.

Wariancja: Usuwanie zmiennych, które mają bardzo niską wariancję.

7. Metody Osłonowe (Wrapper Methods)

Metoda Forward Selection: Stopniowe dodawanie zmiennych do modelu na podstawie poprawy jakości modelu.

Metoda Backward Elimination: Stopniowe usuwanie zmiennych z modelu na podstawie pogorszenia jakości modelu.

WYKŁAD 2 – TRANSFORMACJA ZMIENNYCH

WOE

Weight of Evidence (WoE) to technika stosowana głównie w modelowaniu kredytowym i finansowym, która polega na przekształcaniu zmiennych kategorycznych lub dyskretnych w ciągłe. WoE jest szczególnie przydatne w kontekście modeli regresji logistycznej, ale może być używane także w innych typach modeli klasyfikacyjnych.

Definicja WoE

WoE dla danej kategorii zmiennych jest zdefiniowane jako:

$$\text{WoE} = \ln\left(\frac{\text{P}(Y = 1|X = x)}{\text{P}(Y = 0|X = x)}\right)$$

gdzie:

- $\text{P}(Y = 1|X = x)$ to prawdopodobieństwo zdarzenia (np. defaultu kredytu) dla danej kategorii zmiennej,
- $\text{P}(Y = 0|X = x)$ to prawdopodobieństwo niezdarzenia dla tej samej kategorii.

Obliczenia WoE

1. **Dla każdej kategorii zmiennej oblicz:**
   - $\text{P}(Y = 1|X = x)$: Liczba obserwacji z zdarzeniem w danej kategorii podzielona przez całkowitą liczbę zdarzeń.
   - $\text{P}(Y = 0|X = x)$: Liczba obserwacji bez zdarzenia w danej kategorii podzielona przez całkowitą liczbę obserwacji bez zdarzenia.
2. **Oblicz WoE dla każdej kategorii**, stosując powyższą formułę.
3. **Zastąp kategorie zmiennych wartościami WoE**.

WYKŁAD 2 – TRANSFORMACJA ZMIENNYCH

WOE

Zalety WoE

- **Liniowość**: WoE pomaga w utrzymaniu liniowej relacji między zmienną niezależną a zmienną zależną, co jest korzystne w modelach, które zakładają taką relację (np. regresja logistyczna).
- **Stabilność**: WoE jest mniej wrażliwe na zmiany w rozkładzie zmiennej docelowej w porównaniu do kodowania one-hot.
- **Zmniejszenie kategorii**: WoE może być używane do grupowania kategorii, które mają podobne charakterystyki.
- **Interpretowalność**: WoE jest łatwe do zinterpretowania – wyższe wartości WoE wskazują na wyższe prawdopodobieństwo zdarzenia.

Wyzwania i Uwagi

- **Brak Danych**: WoE wymaga uwzględnienia kategorii, dla których nie ma obserwacji lub są one bardzo rzadkie.
- **Nadmierna Granulacja**: Zbyt wiele kategorii może prowadzić do overfittingu, dlatego czasami warto połączyć kategorie o podobnych wartościach WoE.
- **Uwzględnienie Monotoniczności**: Dla pewnych zmiennych (np. wiek, dochód) warto upewnić się, że relacja WoE jest monotoniczna.
- **Uwzględnienie Biznesowego Kontekstu**: Ważne jest, aby analiza WoE była zgodna z intuicją biznesową i ekspertyzą dziedzinową.

Przekształcenie WoE jest szeroko stosowane i uważane za jedną z efektywnych technik przygotowania danych, zwłaszcza w kontekście modelowania ryzyka kredytowego.

WYKŁAD 2 – WOE PRZYKŁAD PRAKTYCZNY (1/3)

Zmienna: Wiek klienta

Cel: Przewidzieć default kredytu (1 = default, 0 = spłata)

Krok 1: Dane oryginalne (próbka 1000 klientów)

Przedział wiekowy	Liczba klientów	Defaulty (Y=1)	Spłaty (Y=0)	% Defaultów
18-25	200	60	140	30%
26-35	300	45	255	15%
36-50	350	35	315	10%
50+	150	10	140	6.7%
SUMA	**1000**	**150**	**850**	**15%**

WYKŁAD 2 – WOE PRZYKŁAD PRAKTYCZNY (2/3)

Krok 2: Obliczenie WoE

```
WoE = ln(% Dobrych / % Złych)
```

gdzie: % Dobrych = (Spłaty w przedziale / Wszystkie spłaty)
% Złych = (Defaulty w przedziale / Wszystkie defaulty)

Przedział	% Dobrych	% Złych	WoE	Interpretacja
18-25	140/850 = 16.5%	60/150 = 40%	-0.89	Wysokie ryzyko
26-35	255/850 = 30%	45/150 = 30%	**0.00**	Neutralne
36-50	315/850 = 37.1%	35/150 = 23.3%	+0.47	Niskie ryzyko
50+	140/850 = 16.5%	10/150 = 6.7%	+0.90	Bardzo niskie ryzyko

Wnioski:

- WoE ujemne → więcej defaultów niż oczekiwano → wyższe ryzyko
- WoE dodatnie → mniej defaultów niż oczekiwano → niższe ryzyko
- WoE = 0 → proporcja defaultów jak w całej populacji

WYKŁAD 2 – WOE PRZYKŁAD PRAKTYCZNY (3/3)

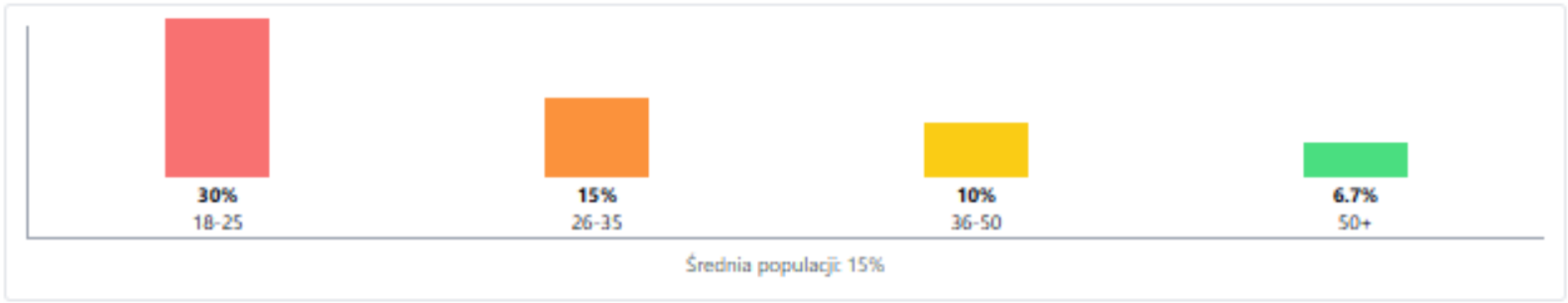
Zmienna: Wiek klienta | Cel: Przewidzieć default kredytu

Dane (1000 klientów)

Przedział	Default	Spłaty	WoE
18-25	60	140	-0.89
26-35	45	255	0.00
36-50	35	315	+0.47
50+	10	140	+0.90

Wizual VoE

Procent defaultów według wieku



Kluczowe wnioski:

- Młodsi klienci (18-25) mają **2x wyższe ryzyko** defaultu (30% vs 15%)
- WoE ujemne → więcej defaultów → model obniży score
- Klienci 50+ to **najlepsza grupa** (tylko 6.7% defaultów)

WYKŁAD 2 – KALIBRACJA (1/4)

1. Metoda Offset (Przesunięcie)

Metoda offset jest jedną z najprostszych technik kalibracji karty scoringowej. Polega na dodaniu lub odjęciu stałej wartości (offsetu) od wszystkich wyników scoringowych, aby dostosować średnią wartość wyników do oczekiwanego poziomu ryzyka.

$$\text{Skalibrowany Wynik} = \text{Wynik Modelu} + \text{Offset}$$

Offset jest często używany, gdy ogólny poziom ryzyka w populacji zmienia się w czasie, ale relatywne różnice ryzyka między jednostkami pozostają stałe.

Obliczanie Offsetu

Offset, czyli przesunięcie, jest wartością, którą dodajemy do wyniku scoringowego w celu dostosowania średniego ryzyka (szans) w próbie modelowania do oczekiwanego poziomu ryzyka w całej populacji. Jeśli znane są szanse na próbie modelowania oraz szanse na całej populacji, offset można obliczyć, korzystając ze wzoru:

$$\text{Offset} = \ln\left(\frac{\text{Szanse na populacji}}{\text{Szanse na próbie}}\right)$$

Gdzie:

- ln oznacza logarytm naturalny,
- Szanse są definiowane jako stosunek prawdopodobieństwa sukcesu (np. dobrego klienta) do prawdopodobieństwa porażki (np. złego klienta) i są obliczane jako:

$$\text{Szanse} = \frac{P(\text{Sukces})}{1 - P(\text{Sukces})}$$

Przykład:

Załóżmy, że szanse na próbie wynoszą 1:2 (czyli prawdopodobieństwo sukcesu na próbie wynosi 0.33) i szanse na całej populacji wynoszą 1:1 (czyli prawdopodobieństwo sukcesu na populacji wynosi 0.5). Wówczas offset można obliczyć jako:

$$\text{Offset} = \ln\left(\frac{1/1}{1/2}\right) = \ln(2)$$

Wartość offsetu jest następnie dodawana do wyniku scoringowego modelu, aby dostosować szanse przewidywane przez model do oczekiwanego poziomu ryzyka w całej populacji.

WYKŁAD 2 – KALIBRACJA (2/4)

2. Estymator Jądrowy

Estymator jądrowy jest techniką nieparametryczną, która może być używana do kalibracji prawdopodobieństw przewidywanych przez model. Metoda ta polega na wygładzaniu rozkładu prawdopodobieństw przy użyciu funkcji jądrowej (kernel function), aby uzyskać bardziej stabilne i wiarygodne przewidywania.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

gdzie:

- $\hat{f}(x)$ to estymowana funkcja gęstości prawdopodobieństwa
- n to liczba obserwacji
- K to funkcja jądrowa
- h to parametr wygładzania (bandwidth)
- X_i to wartości zmiennej objaśniającej

Estymator jądrowy może być używany do kalibracji modelu, dostosowując przewidywane prawdopodobieństwa do obserwowanych częstości zdarzeń w różnych przedziałach prawdopodobieństwa.

WYKŁAD 2 – KALIBRACJA (3/4)

3. Metoda Platt Scaling

Metoda Platt Scaling jest techniką kalibracji, która polega na dopasowaniu modelu regresji logistycznej do przewidywanych prawdopodobieństw modelu.

$$P(Y = 1|f(x)) = \frac{1}{1 + \exp(Af(x) + B)}$$

gdzie:

- $P(Y = 1|f(x))$ to przewidywane prawdopodobieństwo
- A i B to parametry, które są estymowane za pomocą regresji logistycznej
- $f(x)$ to wynik modelu predykcyjnego

WYKŁAD 2 – KALIBRACJA (4/4)

4. Metoda Izotonicznej Regresji

Izotoniczna regresja jest techniką kalibracji, która polega na dopasowaniu niemalejącej funkcji do przewidywanych prawdopodobieństw, zachowując jednocześnie kolejność przewidywań. W kontekście modeli scoringowych, izotoniczna regresja jest używana do dostosowania przewidywanych prawdopodobieństw przez model do rzeczywistych prawdopodobieństw obserwowanych w danych.

Główne Kroki Metody Izotonicznej Regresji:

1. **Sortowanie Danych:** Dane są sortowane na podstawie przewidywanych prawdopodobieństw (lub wyników scoringowych) uzyskanych z modelu.
2. **Budowanie Bloków:** Na początku, każda obserwacja jest traktowana jako osobny blok. Następnie, bloki są łączone w sposób iteracyjny, zaczynając od tych, które są najbardziej podobne pod względem obserwowanych częstości zdarzeń.
3. **Łączenie Bloków:** Bloki są łączone w taki sposób, aby utrzymać własność niemalejącą funkcji kalibracji. W praktyce, bloki są łączone, aż do momentu, gdy dla każdego kolejnego bloku, średnie przewidywane prawdopodobieństwo jest mniejsze lub równe średniemu przewidywanemu prawdopodobieństwu dla poprzedniego bloku.
4. **Obliczanie Skorygowanych Prawdopodobieństw:** Dla każdego bloku, obliczane jest nowe, skorygowane prawdopodobieństwo, które jest średnią ważoną obserwowanych częstości zdarzeń w bloku.
5. **Przypisywanie Skorygowanych Prawdopodobieństw:** Skorygowane prawdopodobieństwa są przypisywane z powrotem do obserwacji, zapewniając, że przewidywania są teraz bardziej zgodne z obserwowanymi prawdopodobieństwami.

Zastosowanie w Modelach Scoringowych:

- **Poprawa Kalibracji:** Izotoniczna regresja poprawia kalibrację modelu, dostosowując przewidywane prawdopodobieństwa do obserwowanych prawdopodobieństw w danych.
- **Zachowanie Rang:** Metoda ta zachowuje kolejność rang przewidywań, co oznacza, że jednostki są nadal rankowane w taki sam sposób, jak przed kalibracją, ale teraz z prawdopodobieństwami, które lepiej odzwierciedlają rzeczywiste prawdopodobieństwa zdarzeń.
- **Zwiększenie Wydajności Modelu:** Poprawa kalibracji może prowadzić do zwiększenia wydajności modelu, szczególnie w kontekście jego zdolności do przewidywania prawdopodobieństw zdarzeń.

Metoda izotonicznej regresji jest często używana w kontekście modeli scoringowych, aby zapewnić, że przewidywane prawdopodobieństwa są nie tylko rangowane poprawnie, ale także są kalibrowane do rzeczywistych prawdopodobieństw obserwowanych w populacji.

WYKŁAD 2 – METODY KALIBRACJI PORÓWNANIE

1. Metoda Offset (Przesunięcie)

Wzór:

```
Score_cal = Score + Offset
```

```
Offset = ln(Szanse_populacja / Szanse_próba)
```

Kiedy stosować:

- Zmieniła się proporcja defaultów w populacji
- Prosta i szybka metoda
- Zachowuje rankingi

Przykład:

Próba: 15% defaultów (1:5.67)
Populacja: 5% defaultów (1:19)
Offset = ln(19/5.67) = 1.21

2. Platt Scaling

Wzór:

```
P(Y=1) = 1/(1+exp(A*f(x)+B))
```

Kiedy stosować:

- Model nie jest dobrze skalibrowany
- Parametry A i B są estymowane
- Dla modeli SVM lub innych

Zalety:

- Elastyczna transformacja
- Działa dla różnych modeli

3. Estymator Jądrowy (KDE)

Charakterystyka:

- Nieparametryczna metoda
- Wygładza prawdopodobieństwa
- Bardziej stabilne predykcje

Kiedy stosować:

- Dane mają złożony rozkład
- Potrzebna większa elastyczność

Uwaga:

Wymaga doboru parametru bandwidth (h)

4. Izotoniczna Regresja

Charakterystyka:

- Niemalejąca funkcja kalibracji
- Zachowuje kolejność predykcji
- Bardziej elastyczna niż Platt

Proces:

1. Sortowanie danych po score
2. Tworzenie bloków
3. Łączenie bloków (PAV algorithm)
4. Przypisanie kalibrowanych P(Y=1)

Kiedy stosować:

- Model wymaga elastycznej kalibracji
- Zachowanie rankingu jest kluczowe

Jak wybrać metodę kalibracji?

Offset → Prosty shift populacji	**Platt / Izotoniczna** → Model źle skalibrowany	**KDE** → Złożony rozkład danych

WYKŁAD 2 – PODSUMOWANIE

Kluczowe Koncepcje

- ✓ **Scoring vs Rating:** Liczby vs symbole, konsumenci vs firmy
- ✓ **WoE:** Transformacja zmiennych zachowująca predykcyjność
- ✓ **Kalibracja:** Dostosowanie prawdopodobieństw do rzeczywistości
- ✓ **Interpretowalność:** Kluczowa w regulowanych sektorach

Etapy Budowy

1. Zbieranie i czyszczenie danych
2. Selekcja zmiennych (IV, korelacja)
3. Transformacja WoE
4. Budowa modelu (regresja log.)
5. Kalibracja i walidacja
6. Implementacja i monitoring

Najważniejsze Wyzwania

Brak Danych
Niektóre kategorie mają za mało obserwacji

Overfitting
Model zbyt dopasowany do danych treningowych

Granulacja
Zbyt wiele kategorii WoE może pogorszyć model

Praktyczne Wskazówki

DO:
- Monitoruj stabilność modelu w czasie
- Dokumentuj wszystkie założenia
- Testuj na danych out-of-time
- Regularnie aktualizuj model

NIE:
- Nie ignoruj zmian w populacji
- Nie używaj zbyt wielu zmiennych
- Nie pomijaj walidacji biznesowej
- Nie zapomnij o aspektach prawnych

WYKŁAD 2 – LITERATURA

1. R. Anderson - Credit Scoring Toolkit https://academic.oup.com/book/53158
2. Ch. Molnar - Interpretable ML https://christophm.github.io/interpretable-ml-book/