



Wydział Inżynierii Lądowej i Transportu PK
Katedra Dróg, Kolei
i Inżynierii Ruchu



INSTYTUT
BADAWCZY
DRÓG I MOSTÓW

SEMINARIUM

Sieci bayesowskie w inżynierii drogowej

Od modelowania ryzyka do szacowania kosztów wypadków drogowych

dr inż. Filip Janowiec, Politechnika Krakowska

27.10.2026, Warszawa

Plan seminarium

1

Wstęp i podstawy sieci bayesowskich

twierdzenie Bayesa, grafy skierowane i tablice CPT, wnioskowanie przy użyciu BN

2

Sieci bayesowskie w inżynierii lądowej

ryzyko, obiekty inżynieryjne, infrastruktura transportu, bezpieczeństwo ruchu

3

Wypadki drogowe: analiza z wykorzystaniem sieci bayesowskich

dane SEWIK 2024, struktura sieci, koszt jako węzeł użyteczności, walidacja

4

Zastosowanie i wnioski

Walidacja danymi, implikacje dla szacowania kosztów wypadków drogowych

Twierdzenie Bayesa

Prawdopodobieństwo warunkowe (definicja)

$$P(A \mid B) = \frac{P(A, B)}{P(B)}, \quad P(B) > 0$$

Prawdopodobieństwo łączne można zapisać na dwa sposoby:

$$P(A, B) = P(A \mid B) P(B) = P(B \mid A) P(A)$$

Twierdzenie Bayesa (Bayes 1763; Laplace 1774) – po podzieleniu przez $P(E)$:

$$P(H \mid E) = \frac{P(E \mid H) P(H)}{P(E)}, \quad P(E) = \sum_j P(E \mid H_j) P(H_j)$$

Uwaga: twierdzenie Bayesa pozwala na **wnioskowanie odwrotne** – ze znanego $P(E \mid H)$, czyli **skutku przy danej przyczynie**, otrzymujemy $P(H \mid E)$, czyli **przyczynę przy zaobserwowanym skutku**.

Przykład: SEWIK 2024

Z	uczestnik zdarzenia zginął	
N	zdarzenie poza terenem zabudowanym	
P(Z)	odsetek zabitych wśród wszystkich uczestników	0,27%
P(N)	odsetek uczestników zdarzeń poza terenem zabudowanym	23,9%
P(N Z)	odsetek zabitych, którzy zginęli poza terenem zabudowanym	63,3%
P(Z N)	szukane: ryzyko zgonu, gdy zdarzenie było poza terenem zabudowanym	?

$$P(Z \mid N) = \frac{P(N \mid Z) P(Z)}{P(N)} = \frac{63,3\% \cdot 0,27\%}{23,9\%} \approx 0,71\%$$

Ryzyko zgonu poza terenem zabudowanym jest **2,6 raza** wyższe niż średnio (0,27%).

Interpretacja bayesowska

Postać dla parametrów modelu θ i danych D

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{p(D)}$$

a posteriori $\propto$ wiarygodność $\times$ a priori

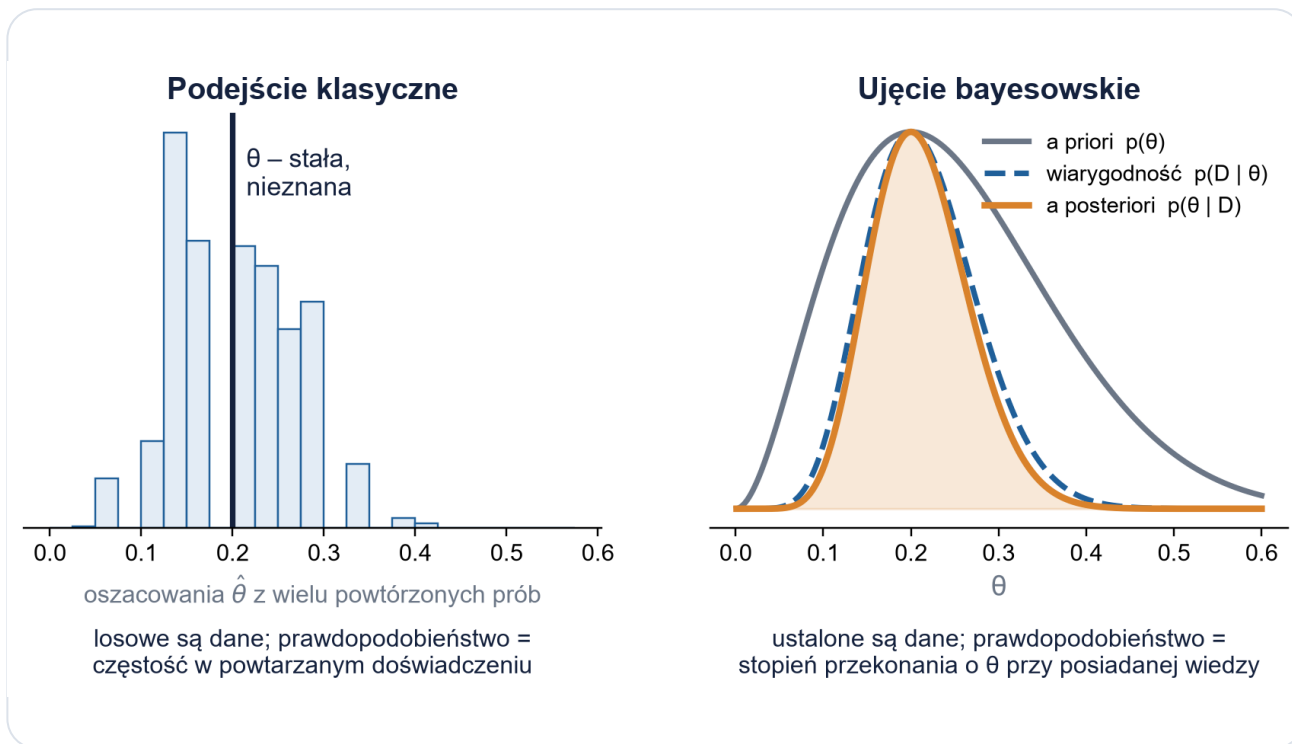
$\propto$ – „jest proporcjonalne do”; $p(D)$ nie zależy od θ

$p(\theta)$ – rozkład a priori: wiedza przed danymi

$p(D | \theta)$ – wiarygodność: jak dobrze θ wyjaśnia dane

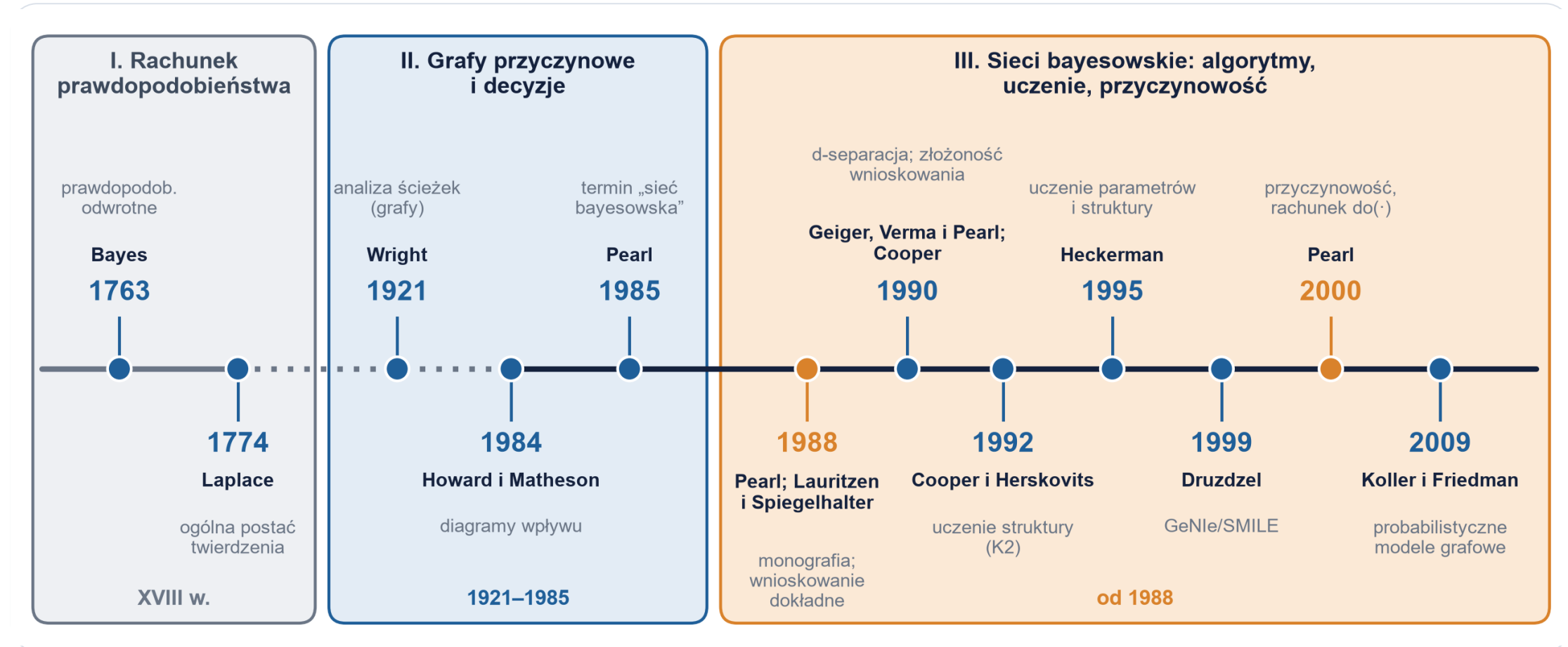
$p(\theta | D)$ – rozkład a posteriori: wiedza po danych

$p(D)$ – wiarygodność brzegowa (stała normująca)



Podjęcie klasyczne pyta, **jak prawdopodobne są dane**, gdy założymy hipotezę; ujęcie bayesowskie pyta, **jak prawdopodobna jest hipoteza** w świetle zebranych danych – i właśnie **tej odpowiedzi potrzebuje inżynier podejmujący decyzję**.

Rozwój metody: chronologia



Wright (1921); Howard i Matheson (1984); Pearl (1985, 1988, 2000); Lauritzen i Spiegelhalter (1988); Druzdzel (1999)

Matematyczne podstawy sieci bayesowskich

Sieć bayesowska to skierowany graf acykliczny: węzły to zmienne losowe $X_1, \dots, X_n$,

a strzałki – bezpośrednie zależności. Każdy węzeł ma tablicę $P(X_i | \text{Pa}(X_i))$ względem swoich rodziców $\text{Pa}(X_i)$.

Lokalna własność Markowa. $X_i \perp \text{ND}(X_i) \setminus \text{Pa}(X_i) \mid \text{Pa}(X_i)$ (ND – niepotomkowie)

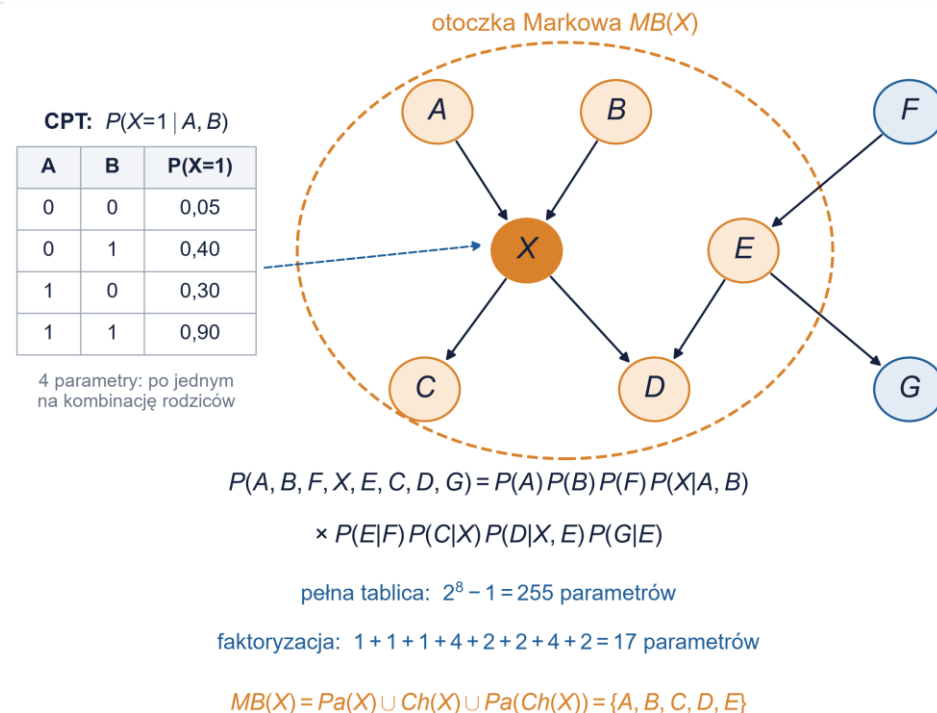
⇒ znając rodziców, dalsze przyczyny pomijamy; brak strzałki jest założeniem.

Reguła łańcuchowa (faktoryzacja). $P(X_1, \dots, X_n) = \prod_i P(X_i | \text{Pa}(X_i))$

⇒ tablice szacuje się osobno; liczba parametrów rośnie z liczbą rodziców, a nie wykładniczo z liczbą zmiennych.

Otoczka Markowa (ang. Markov blanket): $X \perp V \setminus (\{X\} \cup \text{MB}(X)) \mid \text{MB}(X)$

⇒ o X decydują rodzice, dzieci i ich współrodzice; obliczenia są lokalne.



→ **W przykładzie (slajd 20):** sieć v1 ma 22 węzły i 1 241 parametrów zamiast ok. 4,7 biliona kombinacji stanów w pełnej tablicy

Rodzaje węzłów i tablice CPT

Ciężkość

Węzeł szansy. Zmienna losowa; rozkład $P(X \mid \text{rodzice})$ w tablicy CPT. Możliwość zastosowania funkcji gęstości prawdopodobieństwa.

Mechanizm · ekspozycja · ochrona · kruchość	brak	lekkie	ciężkie	zgon
łagodny · niska · dobra · średnia	99,2%	0,7%	0,1%	0,01%
łagodny · niska · brak · wysoka	80,8%	8,7%	9,0%	1,5%
umiarkowany · średnia · słaba · średnia	67,1%	15,9%	12,2%	4,9%

Oświetlić?

Węzeł decyzji. Warianty działania wybiera decydent; rozkładu się nie szacuje.

Wariant	Skutek w sieci
oświetlić odcinek	noc: oświetlenie = tak
nie oświetlać	noc: oświetlenie = nie

Mechanizm

Węzeł deterministyczny. Wartość wyniku z reguły, w CPT tylko 0 i 1 (TAK / NIE).

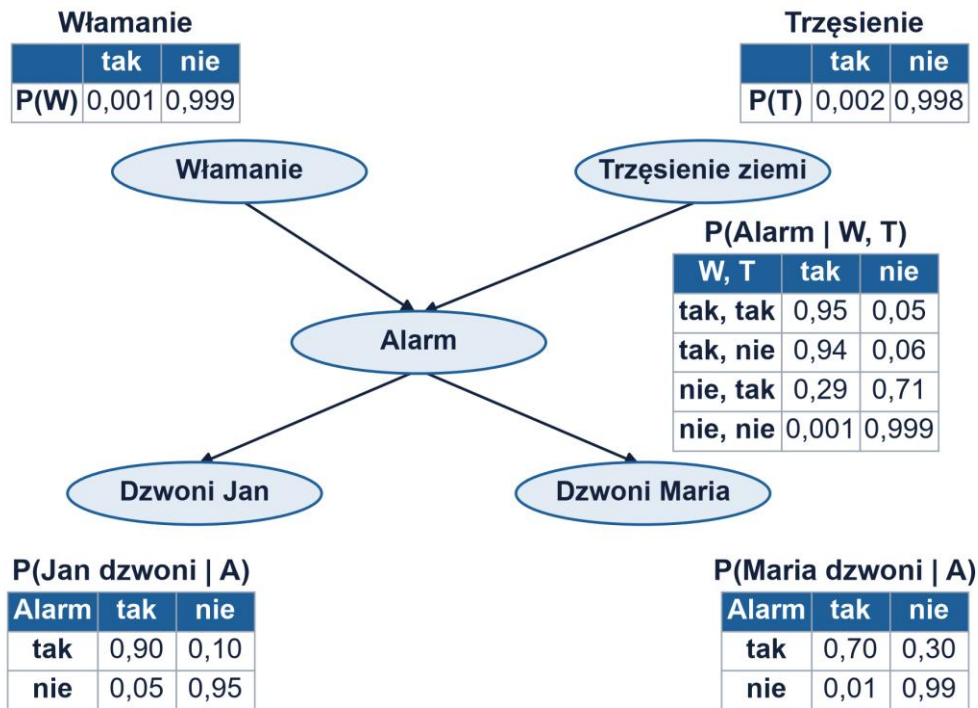
Rodzaj zdarzenia	jednopojazdowe	z ciężarowym
czołowe	ostry	skrajny
tylne	łagodny	umiarkowany
pieszy	umiarkowany	ostry

Koszt

Węzeł użyteczności. Liczba (koszt, korzyść) dla stanów rodziców; sieć liczy wartość oczekiwaną.

Ciężkość obrażeń	zabity	ciężko ranny	lekko ranny	bez obrażeń
Stawka [zł]	2 283 171	3 666 956	49 651	0

$E[\text{koszt}] = \sum P(\text{ciężkość}) \times \text{stawka}$



w każdej tablicy każdy wiersz (jedna konfiguracja rodziców) sumuje się do 1

Scenariusz (Pearl 1988). Alarm w domu włącza się przy włamaniu, ale też przy trzęsieniu ziemi. Sami go nie słyszymy – dzwonią sąsiedzi. Jan dzwoni przy alarmie w 90% przypadków, ale czasem myli go z telefonem (5% fałszywych telefonów). Maria słucha głośnej muzyki: przy alarmie dzwoni w 70% przypadków, bez alarmu prawie nigdy (1%). **Widzimy tylko telefony – z nich wnioskujemy o włamaniu.**

1 · WSTĘP I PODSTAWY SIECI BAYESOWSKICH

Sieć bayesowska jako narzędzie wnioskowania

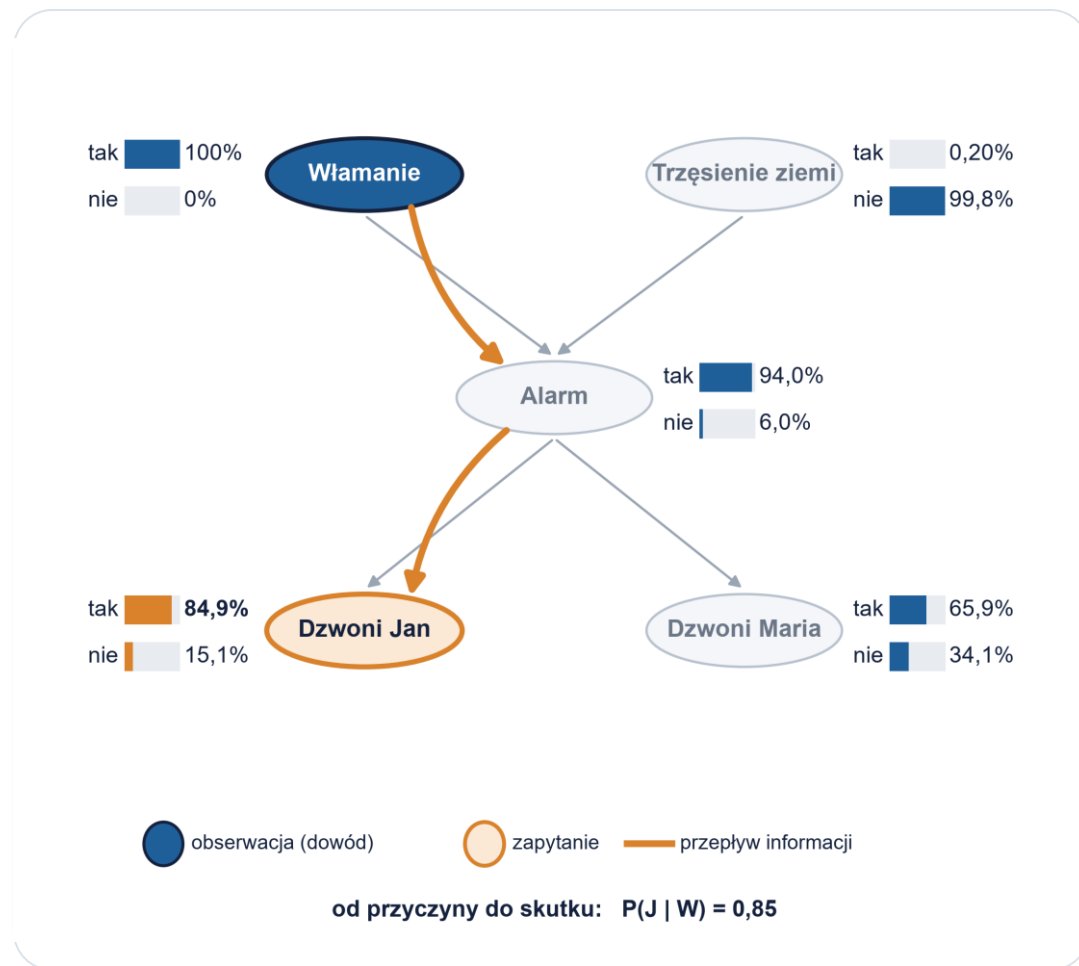
Zapytanie: $P(Q | e) = \sum_h P(Q, h, e) / \sum P(q, h, e)$

Q – zmienne zapytania, e – obserwacje, h – zmienne ukryte

Typ wnioskowania	Zapytanie (sieć alarmowa)	Wynik
predykcyjne	$P(J W)$	0,85
diagnostyczne	$P(W J, M)$	0,284
międzyprzyczynowe	$P(W A) \rightarrow P(W A, T)$	$0,37 \rightarrow 0,003$
a priori	$P(W)$	0,001

W – włamanie, T – trzęsienie ziemi, A – alarm, J, M – telefony Jana i Marii

→ **W przykładzie (slajd 24):** scenariusze „co jeśli” to wnioskowanie predykcyjne: ustawiamy warunki i odczytujemy koszt



1 · WSTĘP I PODSTAWY SIECI BAYESOWSKICH

Wnioskowanie predykcyjne

Pytanie do sieci: Wiemy, że było włamanie. Jak prawdopodobne jest, że Jan zadzwoni?

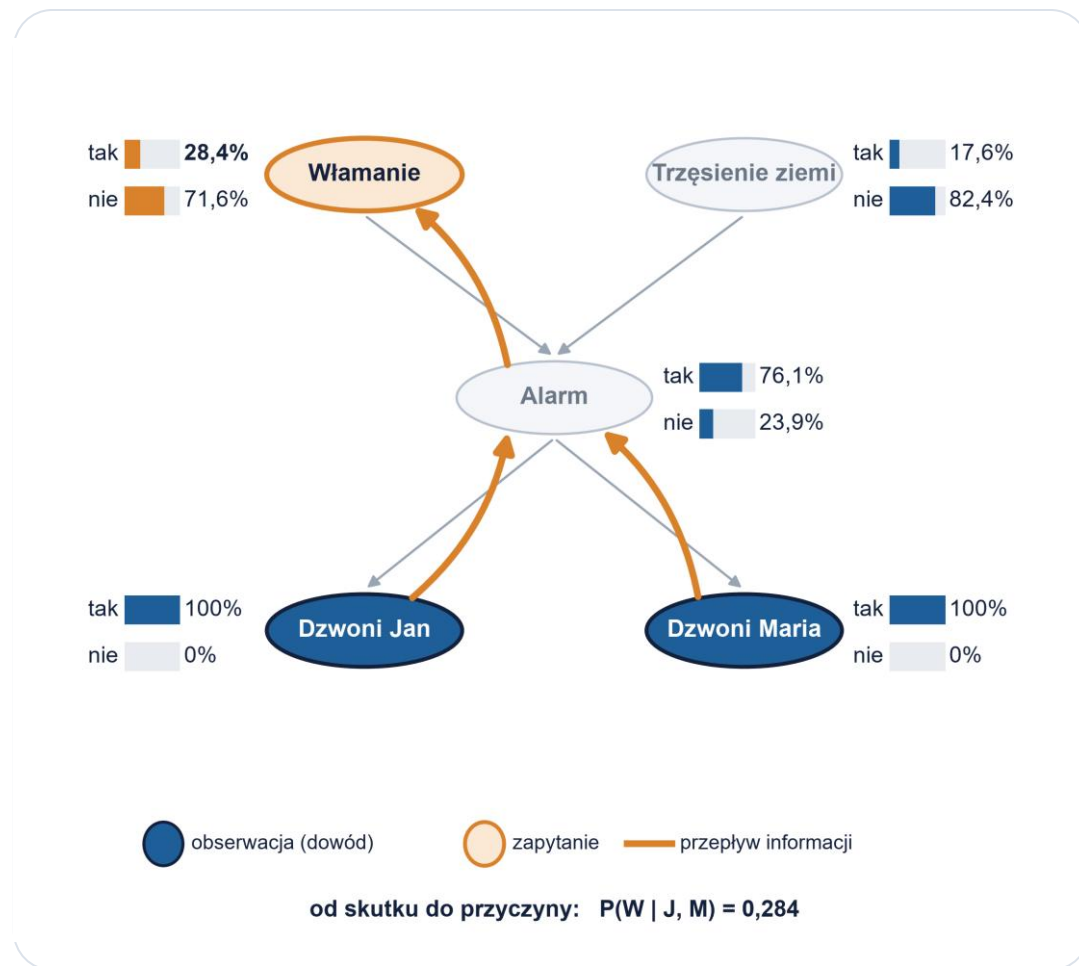
$$P(J | W) = 0,90 \cdot 0,94 + 0,05 \cdot 0,06 \approx 0,85$$

Typ wnioskowania	Zapytanie (sieć alarmowa)	Wynik
predykcyjne	$P(J W)$	0,85
diagnostyczne	$P(W J, M)$	0,284
międzyprzyczynowe	$P(W A) \rightarrow P(W A, T)$	0,37 $\rightarrow$ 0,003
a priori	$P(W)$	0,001

W – włamanie, T – trzęsienie ziemi, A – alarm, J, M – telefony Jana i Marii

Znając przyczynę, przewidujemy skutek.

Informacja płynie zgodnie ze strzałkami: włamanie włącza alarm w 94% przypadków, a przy alarmie Jan dzwoni w 90%. Rosną też szanse telefonu Marii (66%).



1 · WSTĘP I PODSTAWY SIECI BAYESOWSKICH

Wnioskowanie diagnostyczne

Pytanie do sieci: Dzwonią Jan i Maria. Jak prawdopodobne jest, że było włamanie?

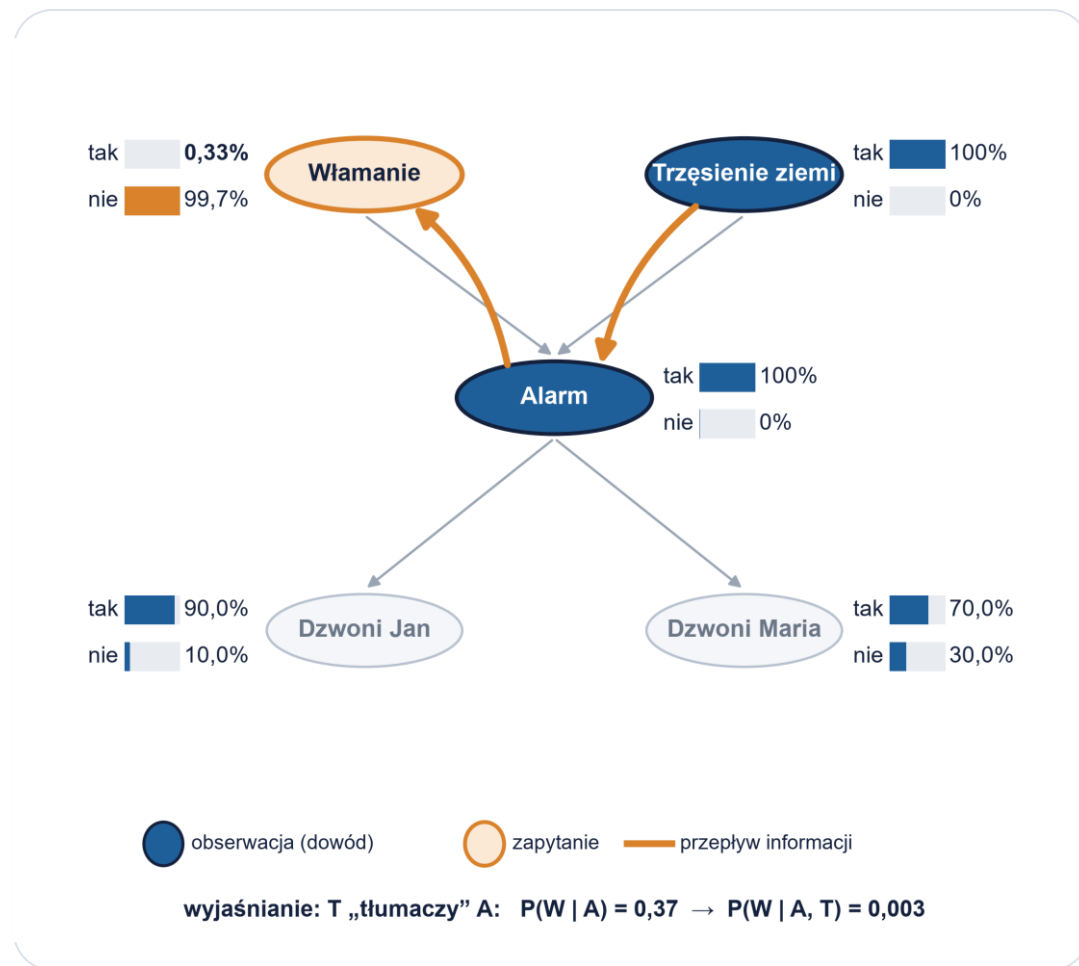
$P(W | J, M) = 0,284$ (bez telefonów: 0,001)

Typ wnioskowania	Zapytanie (sieć alarmowa)	Wynik
predykcyjne	$P(J W)$	0,85
diagnostyczne	$P(W J, M)$	0,284
międzyprzyczynowe	$P(W A) \rightarrow P(W A, T)$	0,37 $\rightarrow$ 0,003
a priori	$P(W)$	0,001

W – włamanie, T – trzęsienie ziemi, A – alarm, J, M – telefony Jana i Marii

Znając skutki, wnioskujemy o przyczynie.

Dwa telefony podnoszą prawdopodobieństwo włamania 284 razy, ale nadal w ok. 72% przypadków przyczyną jest fałszywy alarm albo trzęsienie ziemi (17,6%) – bo włamania są bardzo rzadkie.



1 · WSTĘP I PODSTAWY SIECI BAYESOWSKICH

Wnioskowanie międzyprzyczynowe

Pytanie do sieci: Alarm wyje. Czy było włamanie? A jeśli dowiemy się o trzęsieniu ziemi?

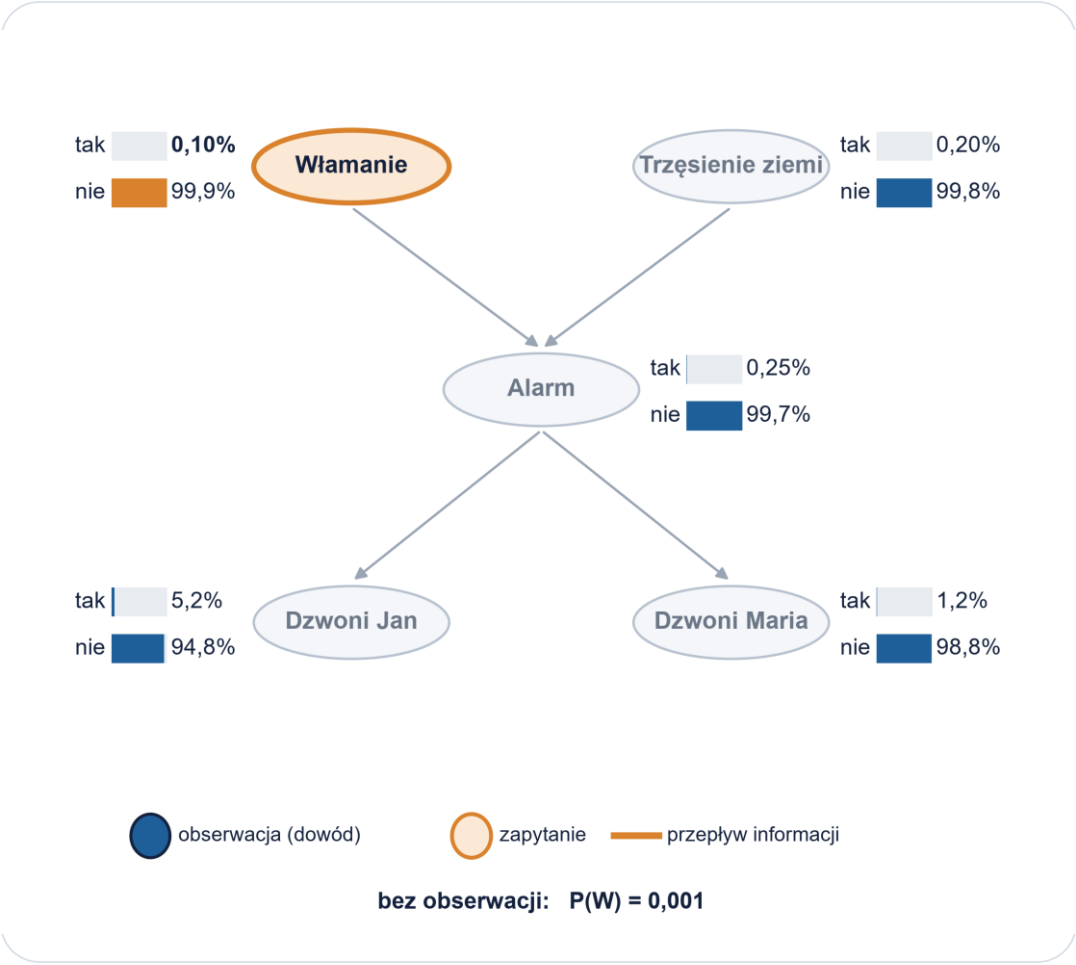
$$P(W | A) = 0,37 \rightarrow P(W | A, T) = 0,003$$

Typ wnioskowania	Zapytanie (sieć alarmowa)	Wynik
predykcyjne	$P(J W)$	0,85
diagnostyczne	$P(W J, M)$	0,284
międzyprzyczynowe	$P(W A) \rightarrow P(W A, T)$	0,37 → 0,003
a priori	$P(W)$	0,001

W – włamanie, T – trzęsienie ziemi, A – alarm, J, M – telefony Jana i Marii

Jedna przyczyna „tłumaczy” skutek i osłabia drugą.

Sam alarm podnosi $P(\text{włamania})$ do 37%; gdy wiemy o trzęsieniu ziemi, alarm jest już wyjaśniony i włamanie spada do 0,3%. Przyczyny niezależne same w sobie stają się zależne, gdy znamy wspólny skutek.



1 · WSTĘP I PODSTAWY SIECI BAYESOWSKICH

Wnioskowanie a priori

Pytanie do sieci: Nic nie wiemy. Jak prawdopodobne jest włamanie?
 $P(W) = 0,001$

Typ wnioskowania	Zapytanie (sieć alarmowa)	Wynik
predykcyjne	$P(J \mid W)$	0,85
diagnostyczne	$P(W \mid J, M)$	0,284
międzyprzyczynowe	$P(W \mid A) \rightarrow P(W \mid A, T)$	0,37 $\rightarrow$ 0,003
a priori	$P(W)$	0,001

W – włamanie, T – trzęsienie ziemi, A – alarm, J, M – telefony Jana i Marii

Bez obserwacji sieć podaje rozkłady brzegowe.

Włamanie 0,1%, alarm 0,25%, telefon Jana 5,2% (głównie pomyłki). To punkt odniesienia: każde wnioskowanie to zmiana tych rozkładów po obserwacjach.

Sieci bayesowskie w inżynierii lądowej

Co daje sieć bayesowska

Problem	Rozwiązanie w sieci
niepewne parametry	rozkłady zamiast wartości
mało danych o awariach	dane + wiedza ekspercka
inspekcje, monitoring	aktualizacja prognozy
zmiany w czasie	sieci dynamiczne (DBN)
decyzje przy ograniczonym budżecie	diagram wpływu

Przykładowe zastosowania

Obszar i zastosowanie	Źródło
mosty i konstrukcje: degradacja, plan inspekcji	Straub (2009); Rafiq i in. (2015)
tunele: czas i ryzyko drążenia	Špačková i Straub (2013)
sieci infrastruktury, wodociągi	Bensi i in. (2013); Kabir i in. (2015)
budowy: upadki z wysokości	Leu i Chang (2015)
transport: ciężkość obrażeń	de Oña i in. (2011)

PN-EN IEC 31010:2020

Zarządzanie ryzykiem – techniki oceny ryzyka

Sieci bayesowskie i diagramy wpływu są jedną z technik oceny ryzyka w procesie ISO 31000.

Identyfikacja ryzyka

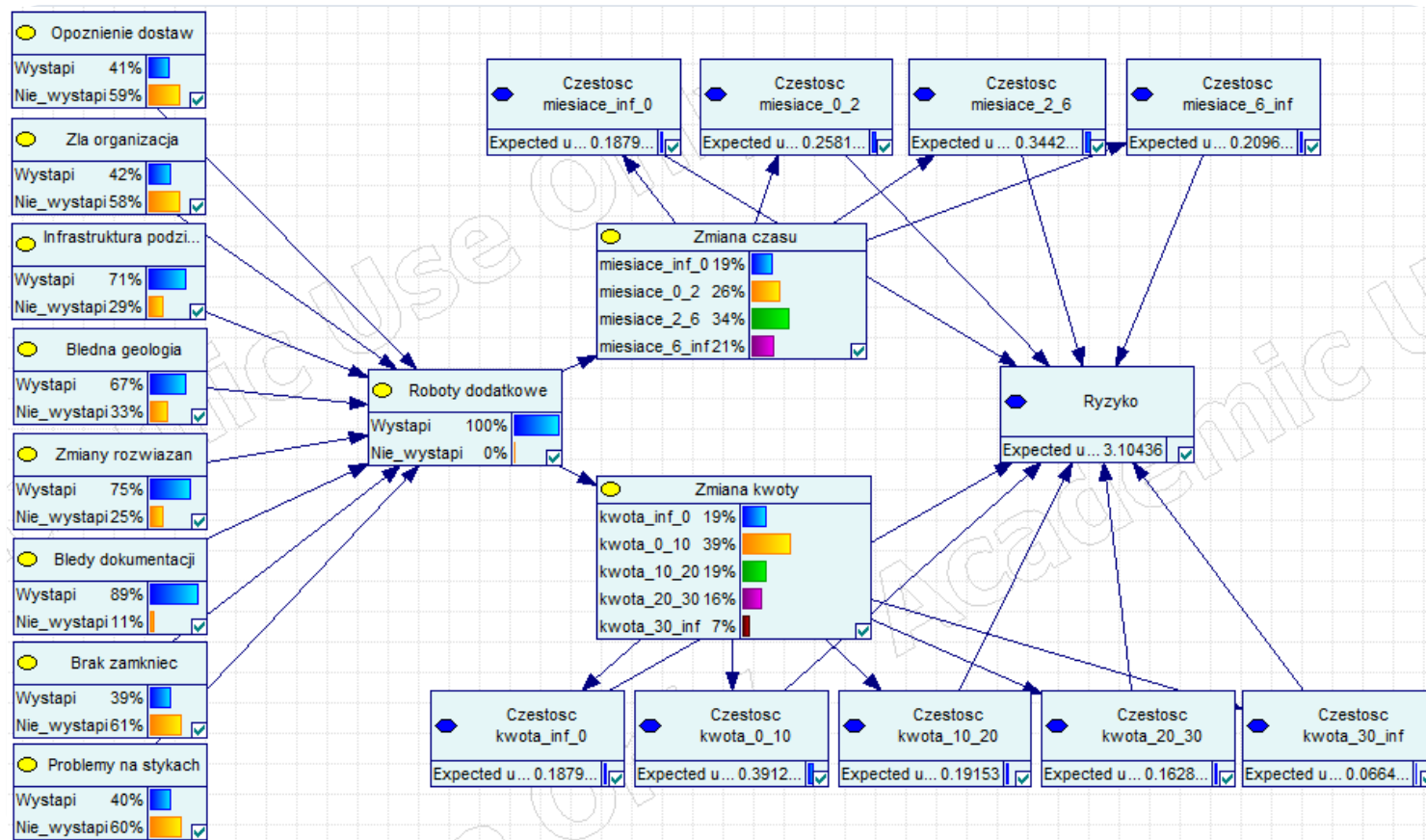
Analiza ryzyka

Ewaluacja ryzyka

Traktowanie ryzyka

■ etapy procesu ISO 31000, w których sieci bayesowskie stosuje się najczęściej

Przykład: ryzyko robót dodatkowych na kolei



Problem: roboty dodatkowe – główne źródło przekroczeń terminów i kosztów inwestycji kolejowych

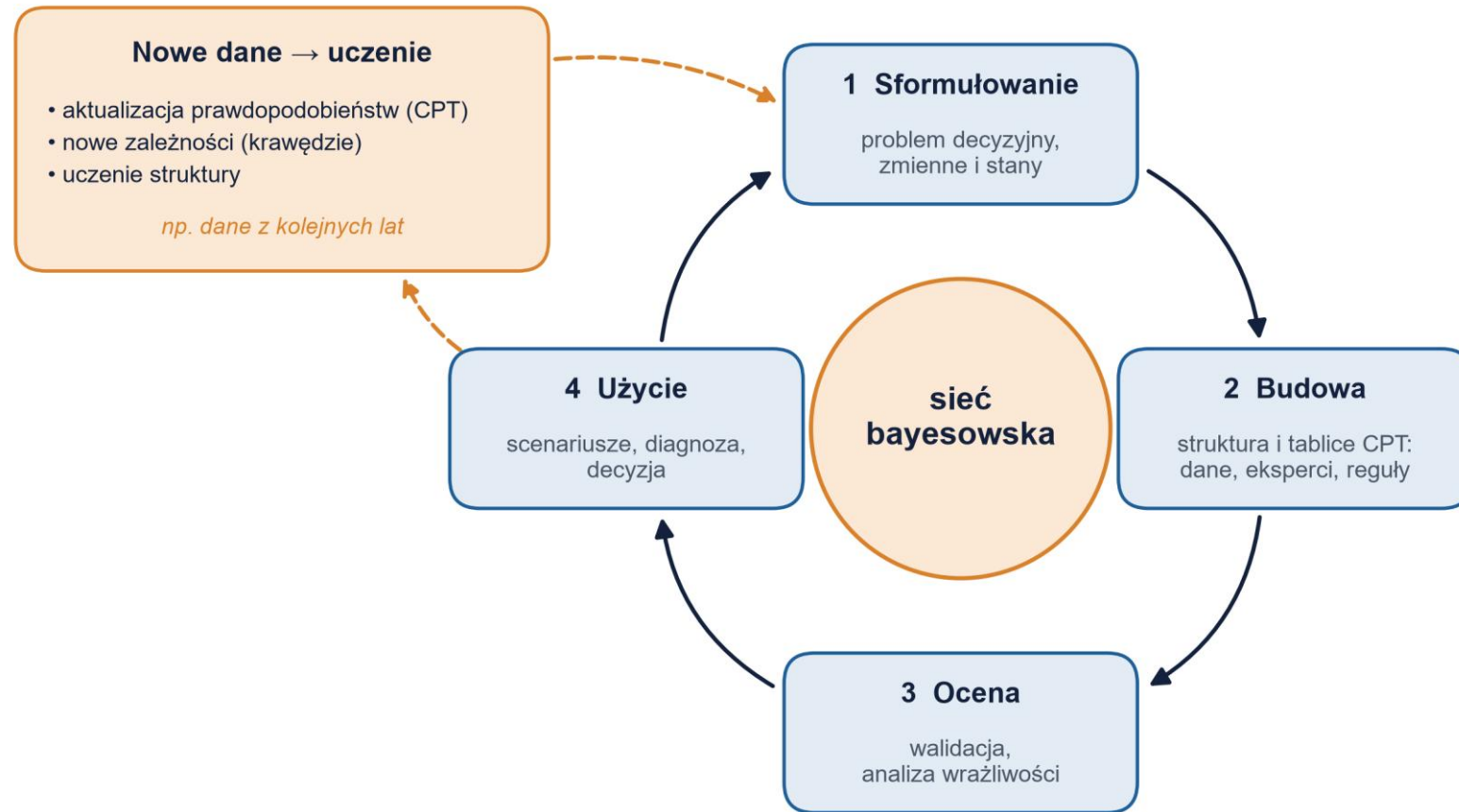
Dane: eksperci, dokumentacja zakończonych inwestycji (16 + 6 testowych), ankieta praktyków

Model: 8 czynników; węzeł „roboty dodatkowe” jako Leaky Noisy-OR – 9 parametrów zamiast 256

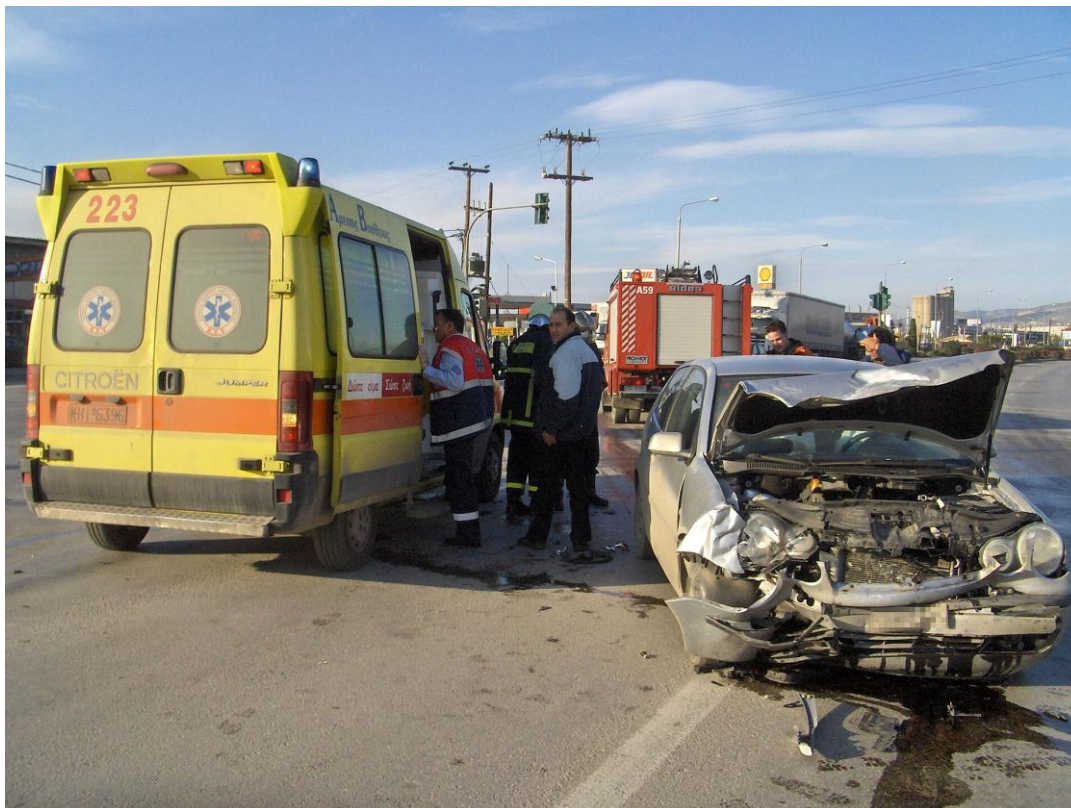
Wynik (scenariusz): gdy roboty wystąpią, wydłużenie o 2–6 mies. ma 34%, wzrost kwoty do 10% – 39%; ryzyko 3,1 w skali 0–7

Sieć w GeNIe po wyznaczeniu ryzyka: 8 czynników → roboty dodatkowe → zmiana czasu i kwoty → ryzyko

Sieć bayesowska jako narzędzie wspomagania decyzji



Sieci bayesowskie a bezpieczeństwo ruchu



Fot. K. Stampoulis, CC BY-SA 3.0 GR, Wikimedia Commons

Obszar	Problem badawczy
Ciężkość obrażeń	Które warunki podnoszą ryzyko ciężkich skutków zdarzenia?
Analiza przyczyn	Co najczęściej poprzedza zdarzenie śmiertelne?
Ryzyko odcinków	Jak geometria, ruch i warunki łączą się w ryzyko w danym miejscu?
Wsparcie decyzji	Który wariant środka poprawy bezpieczeństwa daje najniższy oczekiwany koszt?
Koszty społeczne	Jak warunki zdarzenia przekładają się na koszt w złotych?

Wspólna publikacja w toku: koszt społeczny wypadków drogowych – zespół dr. inż. Marcina Światały (IBDiM)

Problem badawczy - skąd bierze się koszt wypadków?

33,8 mld zł

koszt osobowy ofiar zdarzeń drogowych w Polsce w 2024 r. (bez strat materialnych)

Dziś: koszt = liczba ofiar × stawka jednostkowa (metodyka wyceny kosztów wypadków dla KRBRD)

Brakuje: związku z tym, co dzieje się na drodze: warunki, prędkość, rodzaj zderzenia, kto jest ofiarą

Pytania badawcze: czy sieć bayesowską można nauczyć na danych SEWIK, czy jej wnioskowanie o skutkach i kosztach jest poprawne – i które warunki generują koszt?

Warunki i droga

Zdarzenie

Ofiara i ochrona

Ciężkość obrażeń

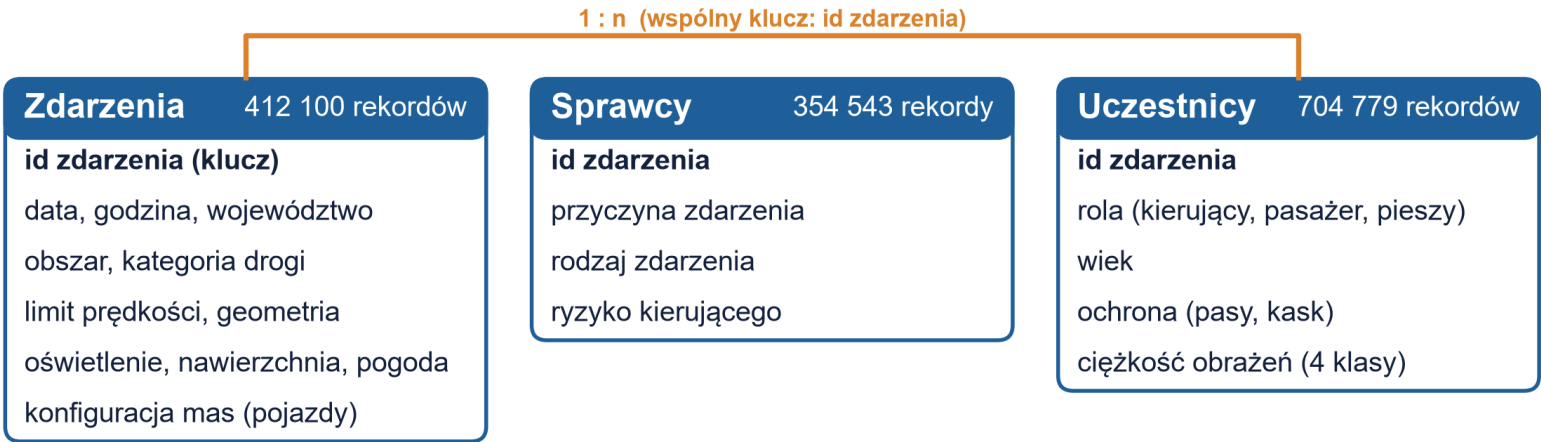
Koszt [zł]

Sieć bayesowska łączy cały łańcuch: od warunków na drodze do kosztu w złotych – w obie strony.

Dane: baza SEWIK 2024



System Ewidencji Wypadków i Kolizji (Policja): relacyjna baza danych – trzy tabele powiązane identyfikatorem zdarzenia



Rekord do sieci = uczestnik + jego zdarzenie + sprawca (wspólny klucz); jednostka analizy: uczestnik (704 779)

Funkcja kosztu jako węzeł użyteczności

$$E[\text{koszt}] = \sum_c P(\text{ciężkość} = c) \cdot \text{stawka}(c)$$

Definicja węzła użyteczności Koszt_na_uczestnika

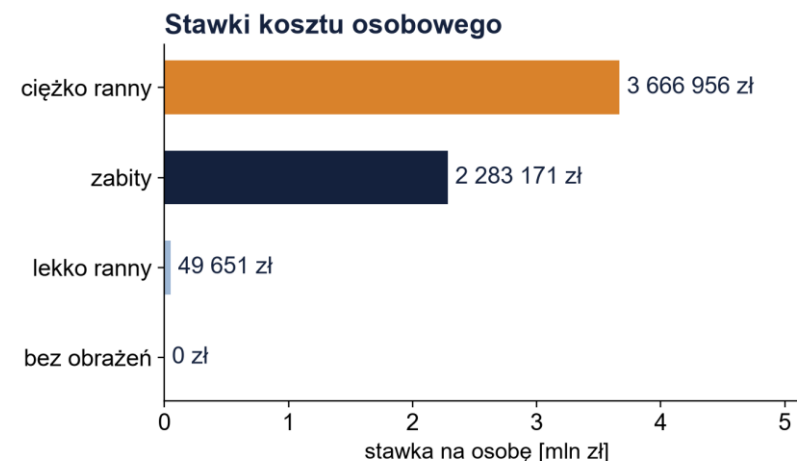
Ciężkość obrażeń	brak	lekkie	ciężkie	śmiertelne
Value	0	49651.21	3666956	2283171

rodzic: Ciężkość obrażeń; wartości wpisane w GeNle (zł na osobę)

Kodowanie: węzeł użyteczności ma jednego rodzica – ciężkość obrażeń – i dla każdego jej stanu stałą stawkę w zł. GeNle liczy wartość oczekiwaną dla dowolnych obserwacji.

Kontrola: funkcja odtwarza koszt 412 099 z 412 100 zdarzeń co do złotówki ($R^2 = 1$).

węzeł z pliku sieci v1 (GeNle)

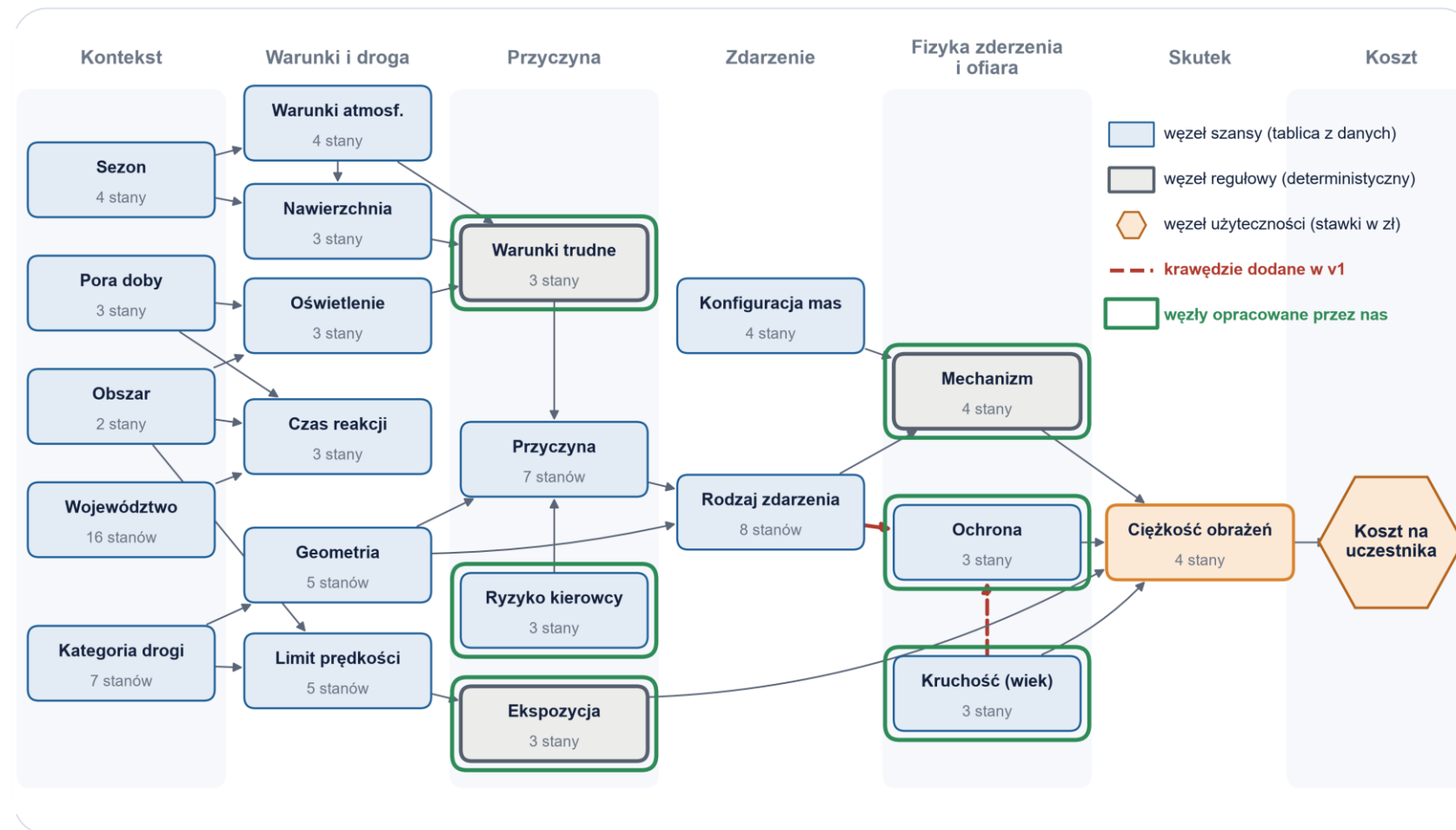


Udział w koszcie osobowym 2024



Wniosek: o jakości szacunku kosztu decyduje trafna prognoza ciężkich obrażeń.

Struktura sieci: łańcuch od warunków do kosztu



22 węzły, 7 warstw

strzałki tylko od przyczyn do skutków

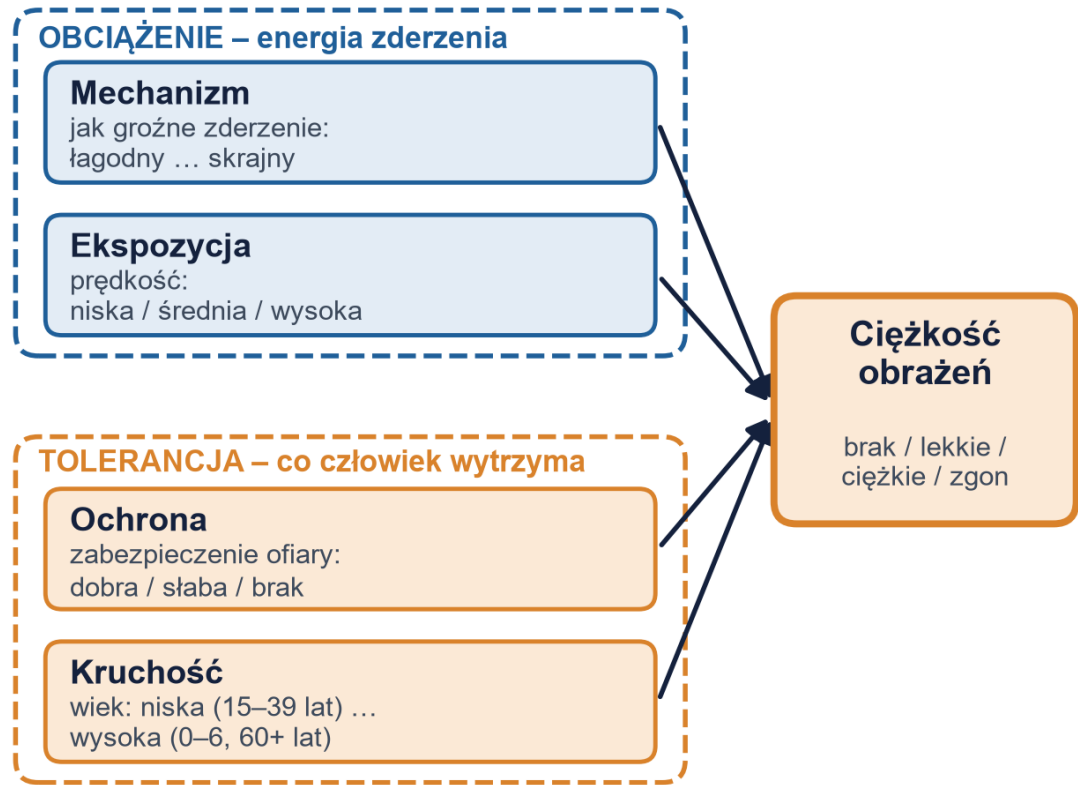
v0 – model bazowy

zielone – węzły opracowane przez nas (przekodowanie SEWIK, reguły). **Cechy ofiary (ochrona, wiek) niezależne od zdarzenia.**

v1 – model główny

Dodane krawędzie: rodzaj zdarzenia → ochrona, kruchość → ochrona. **Lepiej, bo pieszy nie ma ochrony, a ochrona zależy od wieku – sieć wnioskuje o ofierze ze zdarzenia.**

Rozkład warunkowy ciężkości obrażeń



Założenie (Haddon 1973): uraz powstaje, gdy energia zderzenia przekracza tolerancję człowieka.

Tablica oszacowana na 704 779 uczestnikach

Przykład z tablicy: szansa zgonu przy umiarkowanym zderzeniu i prędkości ≥ 80 km/h

Wiek (kruchość)	Ochrona słaba	Brak ochrony
niska (15–39 lat)	4,8%	16,1%
średnia	5,5%	20,8%
wysoka (0–6, 60+ lat)	12,4%	37,4%

Im słabsza ochrona i bardziej wrażliwy wiek, tym większa szansa zgonu
– od 4,8% do 37,4%. Sieć nauczyła się tego z danych.

Skąd te węzły?

W SEWIK ich nie ma – utworzyliśmy je z kilku kolumn danych. Tablica ma dzięki temu 108 kolumn zamiast 46 080. Klasyczny sposób na redukcje kombinacji BN

Mało osób w kolumnie?

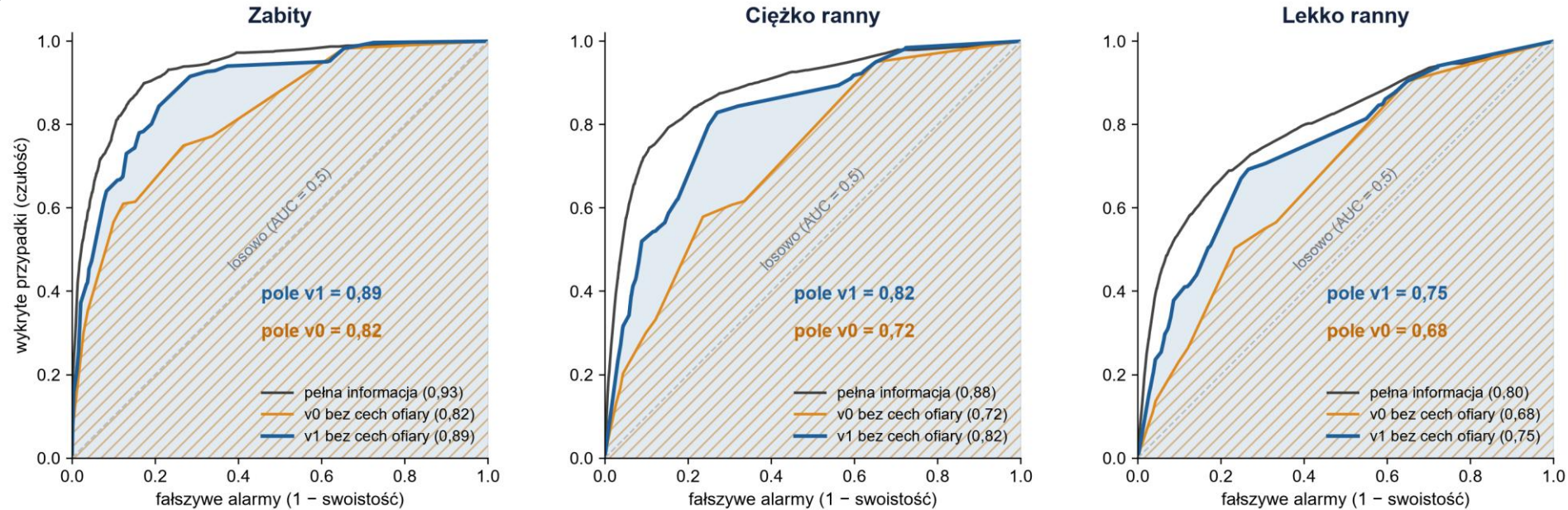
Do każdej liczby osób dodajemy 1 (wygładzenie Dirichleta) – rzadkie kombinacje nie dają zera.

Procedura walidacji - co i dlaczego weryfikujemy



Miara: AUC, bo 96% uczestników jest bez obrażeń – model mówiący zawsze „bez obrażeń” miałby 96% trafności, a AUC ocenia, czy sieć odróżnia ofiary od reszty.

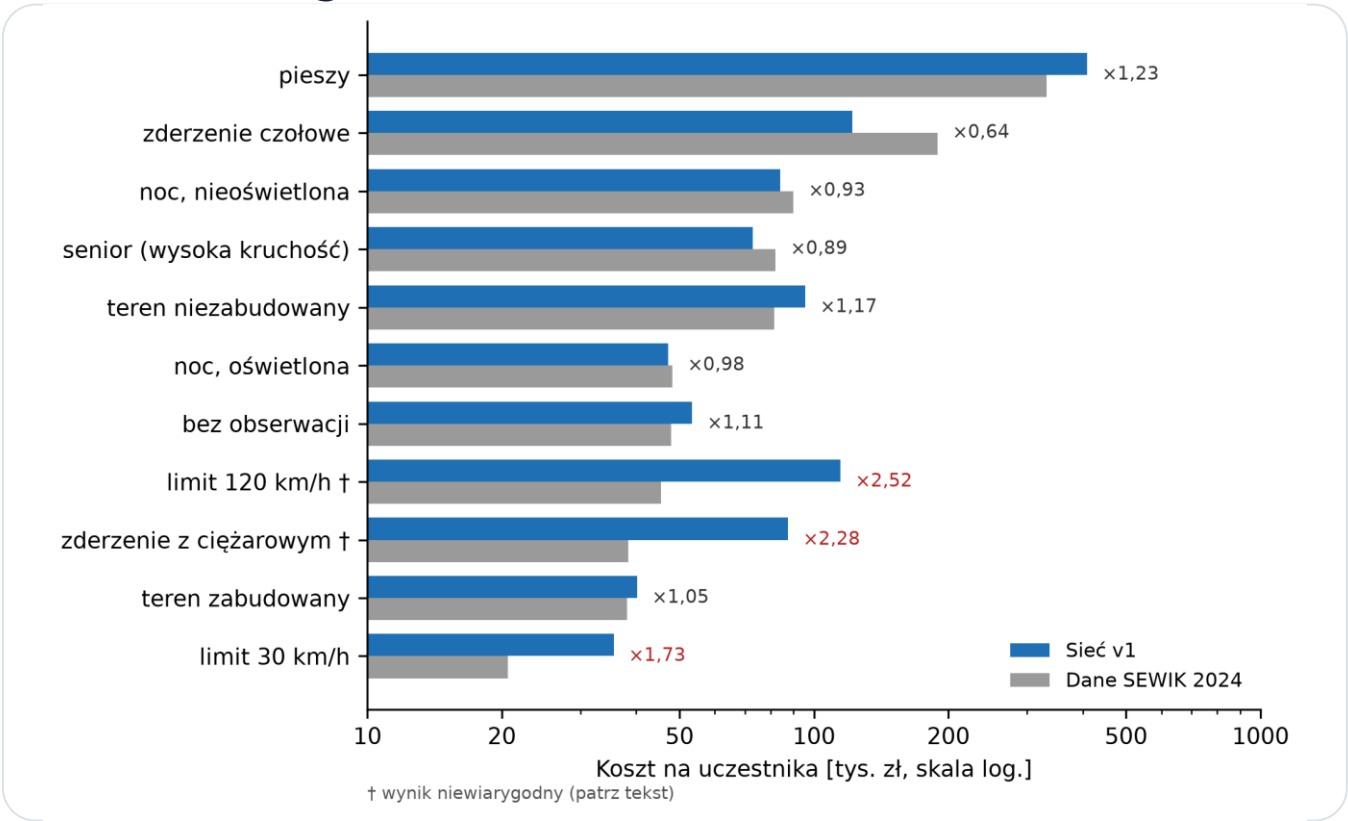
Krzywe ROC i AUC: czy sieć rozróżnia ofiary?



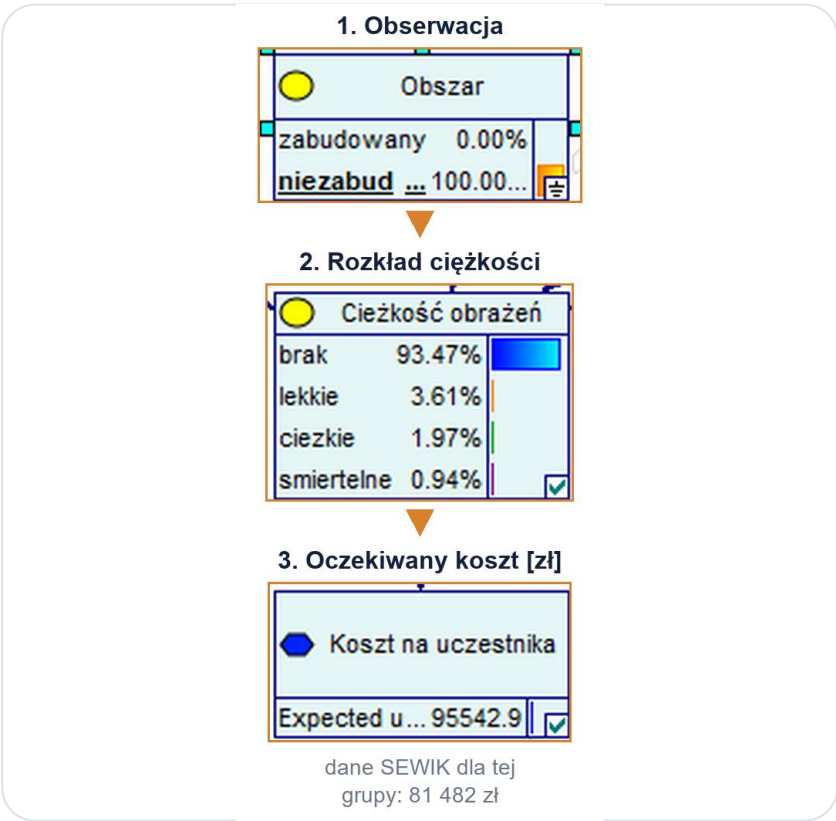
Krzywa	Model	Co sieć wie o uczestniku, gdy przewiduje ciężkość	AUC zgon / ciężkie
— pełna informacja	sieć v0 (v1 daje to samo)	wszystko o zdarzeniu i ofierze: warunki, rodzaj zdarzenia, ochrona, wiek	0,93 / 0,88
— v0 bez cech ofiary	sieć v0	warunki i zdarzenie; ochrona i wiek ukryte – sieć bierze ich średni rozkład	0,82 / 0,72
— v1 bez cech ofiary	sieć v1	jak wyżej, ale sieć wnioskuje ochronę z rodzaju zdarzenia (pieszy → brak ochrony)	0,89 / 0,82

Skala AUC: 0,5 losowo · 0,7–0,8 akceptowalna · 0,8–0,9 doskonała · ≥ 0,9 wybitna (Hosmer i in. 2013)

Walidacja warunkowa: scenariusze w sieci a dane



Trafne: teren, oświetlenie, wiek (±5–17%). **Błędne (†):** limit 120 km/h ×2,5 i pojazd ciężki ×2,3.



GeNie: wycinki z sieci v1: teren niezabudowany → 95 543 zł

Wnioski metodyczne z walidacji

Wynik badań

v0 trafia w średni koszt (+0,5%), ale **dzięki dwóm błędom, które się znoszą**, lepsza v1 zawyża średnią o 11%.

Limit 120 km/h: dane 45 tys. zł, sieć 115 tys. zł. **W sieci rodzaj zdarzenia nie zależy od prędkości**, a na autostradach prawie nie ma pieszych – zgodne z rzeczywistością.

Test na gotowych tablicach **odrzucił 10 477 „niemożliwych” rekordów** i ujawnił błąd w regule mechanizmu przepływu informacji w sieci.

Wniosek

Modelu nie oceniamy po średniej – potrzebne są AUC, kalibracja i scenariusze.

Scenariusz jest wiarygodny tylko w granicach struktury – najpierw sprawdzamy wspólne przyczyny.

Walidacja to też kontrola jakości danych i reguł – przed wnioskami.

Odpowiedzi na pytania badawcze

Czy sieć da się nauczyć na danych SEWIK?	TAK	Średni koszt w sieci 48–53 tys. zł, w danych 47,9 tys. zł.
Czy sieć poprawnie przewiduje skutki?	TAK	Rozpoznaje ofiary: AUC 0,93 dla zgonów, 0,88 dla ciężko rannych.
Czy scenariusze „co jeśli” są wiarygodne?	CZĘŚCIOWO	Teren, oświetlenie, wiek – tak. Limit prędkości i pojazd ciężki – jeszcze nie.
Co najbardziej podnosi koszt?	WYNIK	Teren niezabudowany: +114% wobec zabudowanego; noc bez oświetlenia: +86% wobec oświetlonej; zdarzenie z pieszym: +590% wobec średniej.

Dalsze kierunki badań: rodzaj zdarzenia zależny od prędkości i typu drogi, węzeł decyzji (diagram wpływu), walidacja na kolejnych latach, straty materialne.

Sieć Bayesowska

Wnioskowanie siecią BN



Wydział Inżynierii Lądowej i Transportu PK
Katedra Dróg, Kolei
i Inżynierii Ruchu



INSTYTUT
BADAWCZY
DRÓG I MOSTÓW

Dziękuję.

Pytania i dyskusja

filip.janowiec@pk.edu.pl