

Model procesu dydaktycznego w zakresie Business Intelligence

Zenon Gniazdowski^{1,2)}, Andrzej Ptasznik¹⁾

1) Warszawska Wyższa Szkoła Informatyki, ul. Lewartowskiego 17, Warszawa

2) Instytut Technologii Elektronowej, Al. Lotników 32/46, Warszawa



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Model procesu dydaktycznego w zakresie Business Intelligence

Plan:

- **Problemy w nauczaniu Eksploracji Danych**
 - Co to jest Business Intelligence
 - Miejsce Eksploracji Danych w BI, co to jest Eksploracja Danych
 - Cel nauczania Eksploracji Danych w ramach BI
 - Nieufność jako cel?
 - Przykłady
 - » *Kwartet Anscomba przykład teoretyczny*
 - » *Wnioski z pewnych badań – przykład realny*
 - *Problem komunikacji między uczestnikami procesu BI*



Problemy w nauczaniu Eksploracji Danych

Co to jest Business intelligence?

–Polskie tłumaczenie

- Analityka biznesowa,
- Analiza biznesowa,
- Wywiad gospodarczy?

»Jeśli rzecz dotyczy firmy obcej, to czemu by nie?



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Co to jest Business intelligence?

- **Treść**

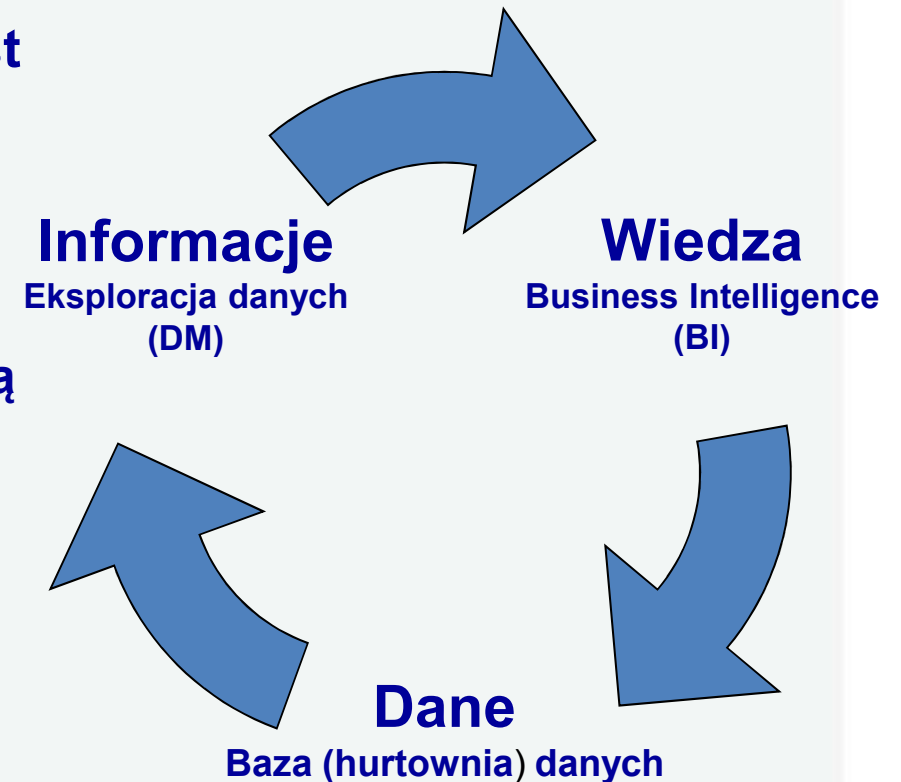
- Business Intelligence można przedstawić jako proces przekształcania danych w informacje, a informacji w wiedzę, która może być wykorzystana do zwiększenia konkurencyjności przedsiębiorstwa³⁾.

3) http://pl.wikipedia.org/wiki/Business_intelligence,
<http://www.controlling.teta.com.pl/business-intelligence>,
<http://www.statsoft.pl/industries/business-intelligence.html>



Problemy w nauczaniu Eksploracji Danych

- **Miejsce Eksploracji Danych w BI⁴⁾**
 - Wiedza pozyskiwana w procesie BI jest zorientowana na końcowego użytkownika (np. właściciela firmy, analityka opracowujący informacje, itp..)
 - Źródłem informacji będących podstawą wiedzy w procesie BI jest Eksploracja Danych (DM).
 - Eksploracja danych jest technologią informatyczną, będącą połączeniem m.in. statystyki oraz technologii baz danych, itp..
 - » Przez odpowiednie algorytmy pozwala wykrywać ukryte wzorce i zależności.



4) R. Łukawiecki. Wstęp do eksploracji danych. Project Botticelli Ltd.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Cel nauczania Eksploracji Danych w kontekście procesu Business intelligence:**
 - **Umiejętność korzystania z metod i narzędzi Eksploracji Danych**
 - **Umiejętność interpretacji otrzymywanych rezultatów**
 - **Nieufność względem otrzymanych wyników**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Np.. Matematyka dyskretna**
 - Pojęcie relacji jako podzbioru iloczynu kartezyjskiego leży u podstaw relacyjnych baz danych
 - Typ relacji konstytuuje rodzaj skali pomiarowej
 - Rodzaj skali determinuje metody analizy danych
 - Mocniejsza skala – często prostsze metody statystyczne
 - Słabsza skala – metody trudniejsze
 - » Różne skale – różne typy danych – różne metody



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Matematyka dyskretna**
 - Typ relacji konstytuuje rodzaj skali pomiarowej
 - Relacja równoważności definiuje skalę nominalną (dane skategoryzowane)
 - Dwa elementy w zbiorze danych są równe, albo różne, należą (bądź nie należą) do wspólnej klasy abstrakcji. Między danymi nie ma mowy o relacji „większe” lub „mniejsze”



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Matematyka dyskretna**
 - Typ relacji konstytuuje rodzaj skali pomiarowej
 - Relacja równoważności definiuje skalę nominalną (dane skategoryzowane)
 - Dane można przydzielać do różnych klas, badać asocjacje pomiędzy klasami itp..
 - W najprostszym przypadku, do badania zależności pomiędzy klasami równoważności można stosować test Chi kwadrat. Przykład:
 - » Czy zachodzi związek pomiędzy klasą kabiny na którą miał bilet pasażer Titanica, a przeżyciem przezeń katastrofy



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Matematyka dyskretna**
 - Typ relacji konstytuuje rodzaj skali pomiarowej
 - Relacja porządku częściowego definiuje skalę porządkową
 - W ramach pewnych podzbiorów można dane porównywać w sensie relacji „>” lub „<”
 - Do badania związków można stosować silniejsze testy niż w przypadku danych skategoryzowanych (nominalnych)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Matematyka dyskretna**

- **Typ relacji konstytuuje rodzaj skali pomiarowej**

- **Relacja porządku liniowego zachodzi dla danych mierzonych w skalach interwałowej lub ilorazowej.**

- Tutaj każde dwa elementy można porównywać w sensie relacji „>”, „>=”, „<”, „<=”,

- Na danych można wykonywać operacje „+” lub „-” (dla skali ilorazowej także „*” lub „/”)

- » **Stosowane metody mogą być bardzo silne – można badać poziom zależności korelacyjnych, budować analityczne modele związków pomiędzy zmiennymi**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- Np. Rachunek prawdopodobieństwa i statystyka matematyczna
 - Fundamentalne jest pojęcie prawdopodobieństwa
 - Brak świadomości, że pewne zjawiska mają charakter zjawisk losowych
 - Stąd mogą istnieć trudności natury psychologicznej



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- Np. Rachunek prawdopodobieństwa i statystyka matematyczna
 - Ważne jest rozróżnianie pomiędzy zależnością przyczynową i korelacyjną. Przykład:
 - Zależność pomiędzy liczbą bocianów, a liczbą urodzeń w Sztokholmie w XIX w. :
 - Dla 73 lat obserwacji uzyskano współczynnik korelacji $R > 0.9$



Problemy w nauczaniu Eksploracji Danych

Niezbędne podstawy matematyczne

- **Rachunek prawdopodobieństwa i statystyka matematyczna**
 - **Problemem jest wnioskowanie statystyczne**
 - **logika testów statystycznych**
 - np. odrzucanie hipotezy zerowej
 - » itp.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Cel nauczania Eksploracji Danych cd.:**
 - Do tej pory była mowa o osiągnięciu celów – w aspekcie pozytywnym, czyli:
 - Co zrobić, aby cel osiągnąć?
 - Teraz o pożądanej nieufności względem otrzymanych wyników
 - Czy nieufność może być celem?
 - Na potwierdzenie przykład teoretyczny:
 - » kwartet Anscomba



Problemy w nauczaniu Eksploracji Danych

- **Kwartet Anscomba**

- **Francis John (Frank) Anscombe (13.V.1918 – 17.X.2001), angielski statystyk.**

- **Wykształcony w Cambridge University.**
 - **Przez dwa lata pracował Rothamsted Experimental Station, jednym z najstarszych rolniczych instytutów badawczych.**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Kwartet Anscomba**

- **Frank Anscombe – angielski statystyk.**

- **Wykładał w Cambridge**

- **W roku 1956 przeniósł się do Princeton University**

- **W 1963 roku zakładał wydział statystyki na Uniwersytecie Yale.**

- **Znany jest tzw. kwartet Anscomba – The American Statistician, Vol. 27, No. 1. (Feb., 1973), pp. 17-21.**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Kwartet Anscomba**

Graphs in Statistical Analysis*

F. J. ANSCOMBE**

Graphs are essential to good statistical analysis. Ordinary scatterplots and “triple” scatterplots are discussed in relation to regression analysis.

1. *Usefulness of graphs*

Most textbooks on statistical methods, and most statistical computer programs, pay too little attention to graphs. Few of us escape being indoctrinated with these notions:

- (1) numerical calculations are exact, but graphs are rough;
- (2) for any particular kind of statistical data there is just one set of calculations constituting a correct statistical analysis;
- (3) performing intricate calculations is virtuous, whereas actually looking at the data is cheating.

through the computer. The analysis should be sensitive both to peculiar features in the given numbers and also to whatever background information is available about the variables. The latter is particularly helpful in suggesting alternative ways of setting up the analysis.

Thought and ingenuity devoted to devising good graphs are likely to pay off. Many ideas can be gleaned from the literature, of which a sampling is listed at the end of this paper. In particular, Tukey [7, 8] has much to say on the topics presented here.

A few simple types of statistical analysis are now considered.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- Kwartet Anscomba**

Lp.	x_1	I	II	III	x_2	IV
1	4	4.26	3.1	5.4	8	6.89
2	5	5.68	4.74	5.7	8	5.25
3	6	7.24	6.13	6.1	8	7.91
4	7	4.82	7.26	6.4	8	5.76
5	8	6.95	8.14	6.8	8	8.84
6	9	8.81	8.77	7.1	8	6.58
7	10	8.04	9.14	7.5	8	8.47
8	11	8.33	9.26	7.8	8	5.56
9	12	10.8	9.13	8.2	8	7.71
10	13	7.58	8.74	13	8	7.04
11	14	9.96	8.1	8.8	19	12.5

9.0	7.5	7.5	7.5
11.0	4.1	4.1	4.1
	0.8	0.8	0.8

9.0	7.5
11.0	4.1
	0.8

średnia
wariancja
wsp. korelacji(x,y)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

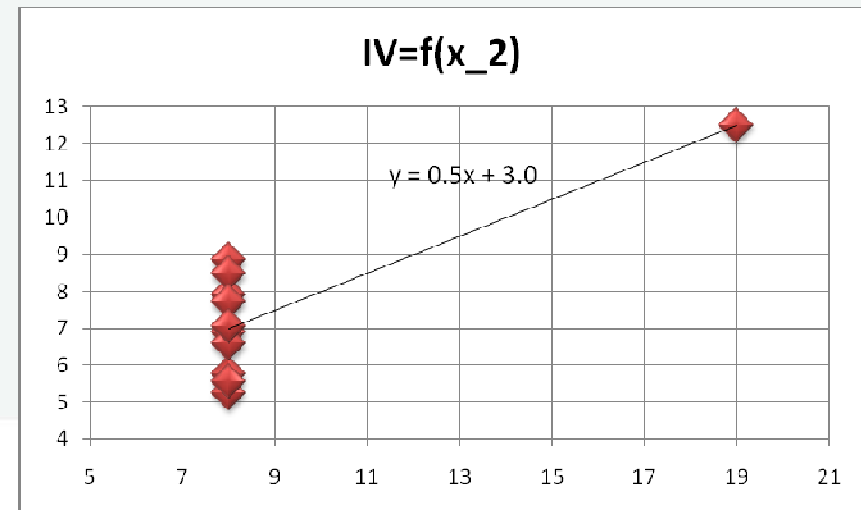
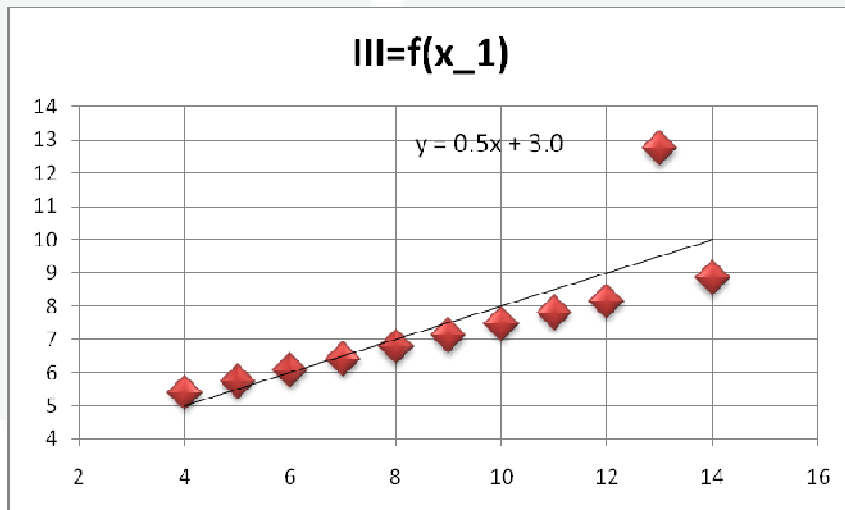
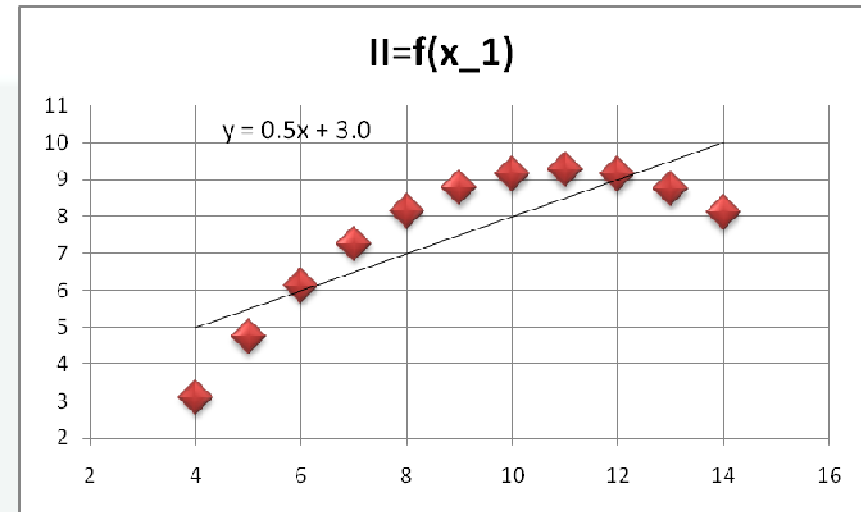
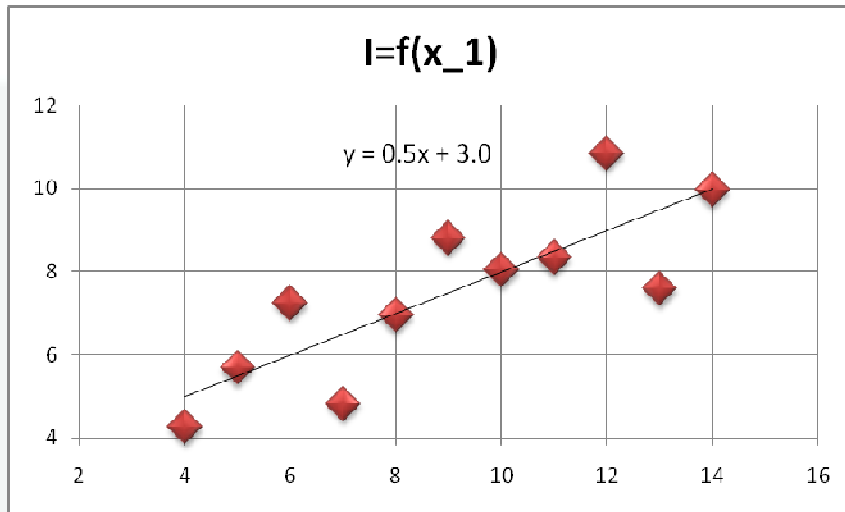


WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Cel nauczania Eksploracji Danych cd.:**
 - **Pierwszy wniosek:**
 - **Wobec uzyskanych wyników trzeba co najmniej nabrać dystansu.**
 - **W pokazanym przykładzie nieufność byłaby uzasadniona**
 - **Dlatego właśnie, celem kształcenia w dziedzinie eksploracji danych powinno być także wpojenie szczypty nieufności!**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Cel nauczania Eksploracji Danych cd.:**
 - **Czego nie robić, aby nie osiągnąć celu niepożądanego (z punktu widzenia zleceniodawcy analizy)**
 - Przykład mało efektowny, ale realny...



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Kontekst: badania naukowe w pewnej instytucji (pierwsza połowa lat dziewięćdziesiątych XX wieku)**
 - **Pytania medyczne**
 - **Związki statystyczne prawie się nie ujawniały**
 - **Pytania socjologiczne**
 - **Zapewne mniej istotne z punktu widzenia medycznego**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

WYNIKI BADAŃ SOCJOLOGICZNYCH

ZWIĄZKI	Grupa A	Grupa B	Grupa A+B
Powrót do pracy a powikłania (lub ich brak)	ns	ns	ns
Stan zdrowia a ocena lekarska zdolności do pracy	ns	ns	ns
Powrót do pracy a charakter pracy	p<0.026	ns	ns
Tryb życia a powikłania		ns	p<0.001
Powrót do pracy a powikłania		p<0.001	ns

ZWIĄZKI	POZIOM ISTOTNOŚCI
Plany po zabiegu, a przydział do grup A i B	p<0.013



Problemy w nauczaniu Eksploracji Danych

WYNIKI BADAŃ SOCJOLOGICZNYCH

ZWIĄZKI	POZIOM ISTOTNOŚCI
Wykształcenie, a przydział do grup A i B	$p < 0.009$



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Problem komunikacji

- **Specjalista od eksploracji danych pracuje na rzecz końcowego odbiorcy informacji**
 - Jeżeli ten odbiorca nie jest specjalistą od DM, może się okazać, że obydwaj posługują się innym językiem. Przykład:
 - Zleceniodawca (lekarz) zażyczył sobie „analizy czynnikowej”
 - Wykonawca zrozumiał, że ma rozwiązać problem znany jako „przekleństwo wielowymiarowości”



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

Problem komunikacji

- **Wykonawca napisał program do analizy czynnikowej.**
 - **W tym celu musiał zaimplementować złożony algorytm rozwiązujący symetryczny problem własny**
 - **Odbiorca z uporem twierdził, że to nie jest analiza czynnikowa**



Problemy w nauczaniu Eksploracji Danych

Problem komunikacji

- Po kilku bezowocnych spotkaniach do rozmów zaproszono matematyka, który przez lata współpracował z lekarzami
 - Okazało się, że odbiorca analiz potrzebował rozwiązać problem powtarzalnych pomiarów



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Podsumowanie**

- **Problemy pozytywne:**

- **Co robić?**

- **Nie unikać studiowania matematyki**

- **Problemy negatywne**

- **Jak nie postępować, albo czego nie robić?**

- **Nie wyciągać pochopnych wniosków**

- **Nie badać tego, o co nas nie proszą?**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Problemy w nauczaniu Eksploracji Danych

- **Podsumowanie**

- **Na koniec dobra rada:**

- **Jeżeli przypuszczasz, że Zleceniodawca wyników analiz może mówić językiem innym niż Ty, postaraj się zaprosić do rozmowy z Nim kogoś kto zna zarówno język Zleceniodawcy , jak i język DM**



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



WARSZAWSKA
WYŻSZA SZKOŁA
INFORMATYKI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

