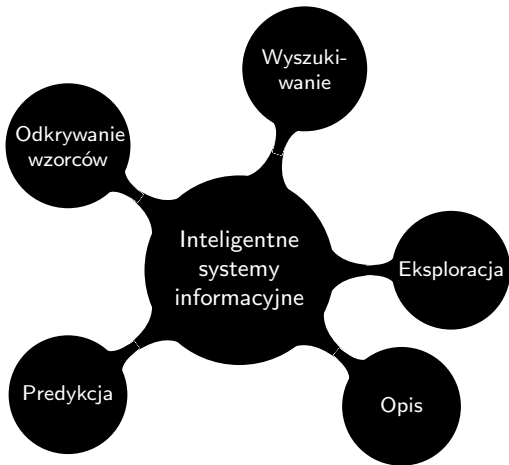


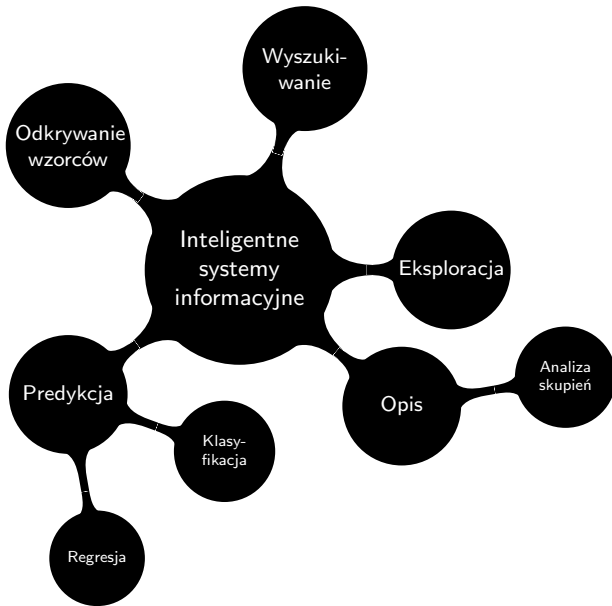
```
graph LR; A[Inteligentne systemy informacyjne] --- B[Filip Graliński]; A --- C[Klasyfikacja]
```

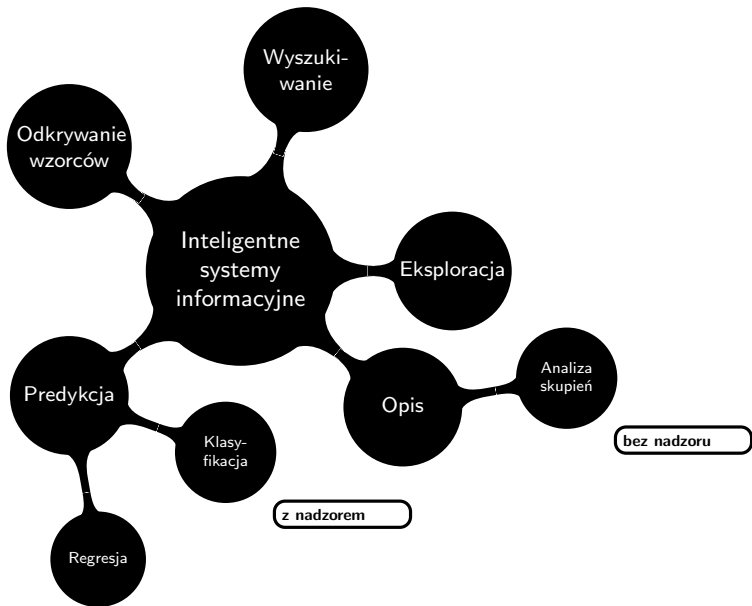
Inteligentne systemy
informacyjne

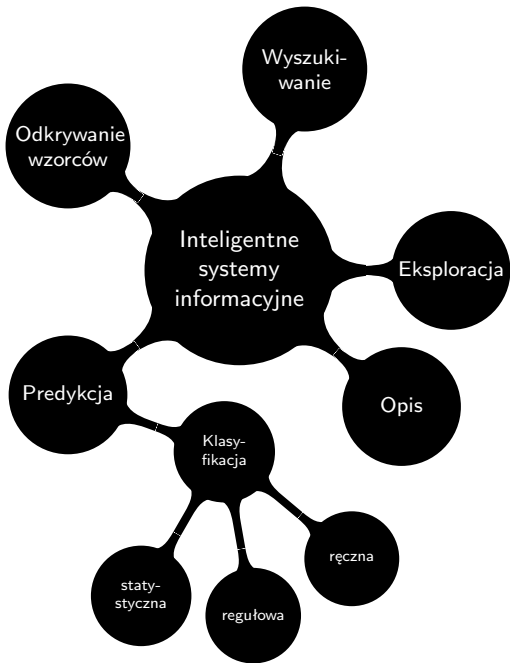
Filip
Graliński

Klasyfikacja



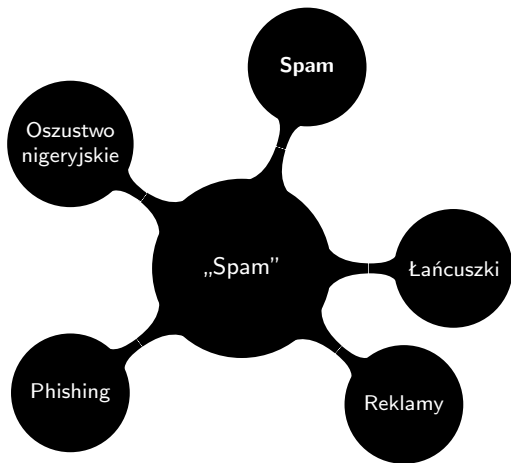


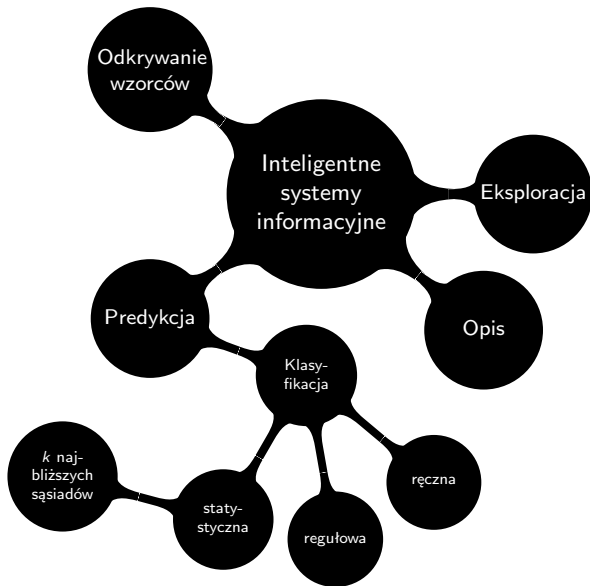






„Spam“

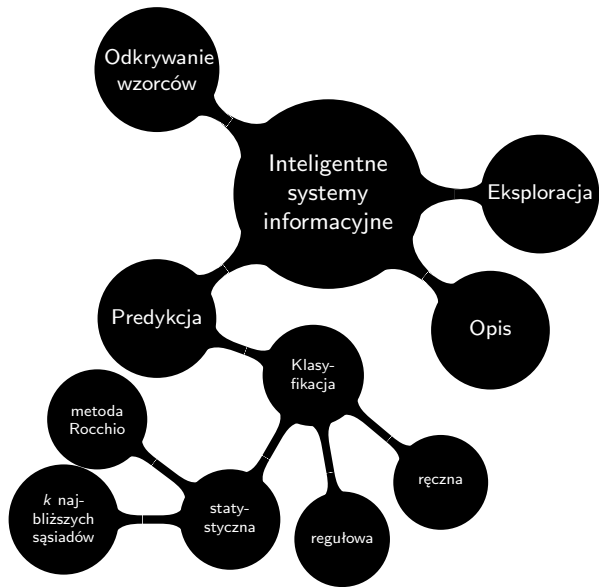


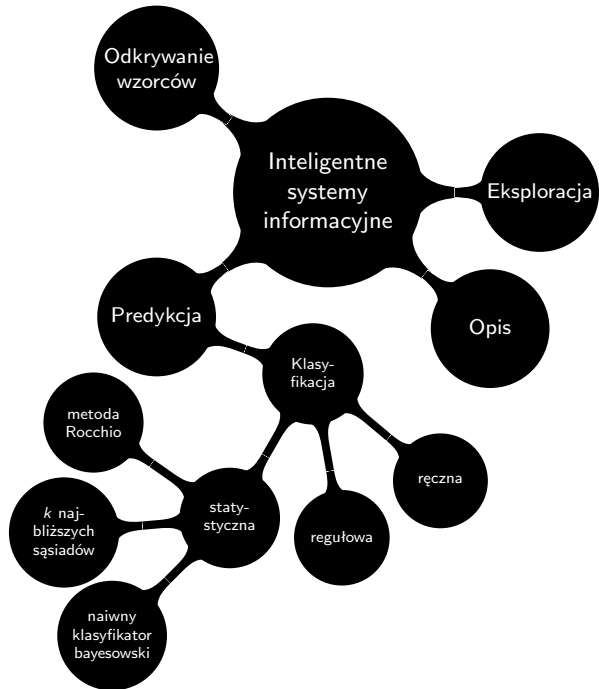


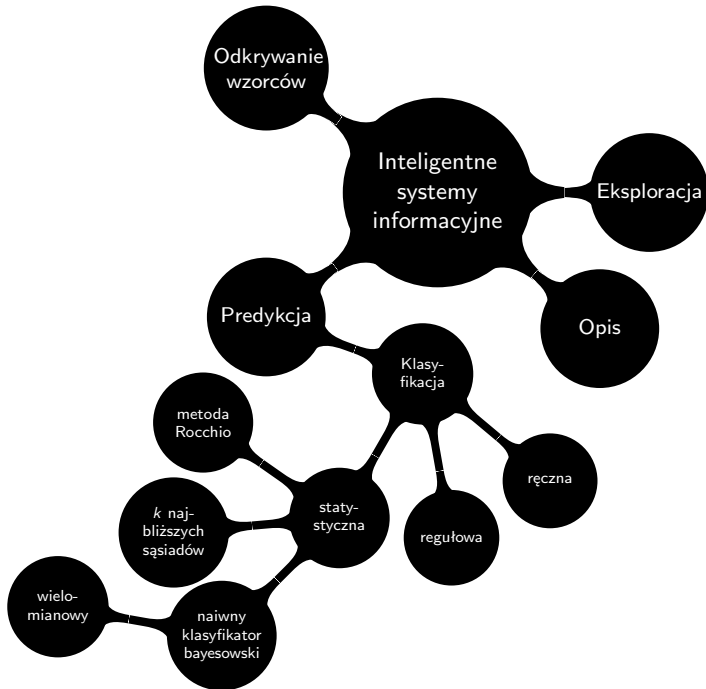
ważone kNN

$$\text{score}(c, d) = \sum_{d' \in S_k(d)} I_c(d') \cos(\vec{d}, \vec{d}')$$

$$I_c(d') = \begin{cases} 1, & \text{jeśli } d' \text{ należy do } c \\ 0, & \text{w przeciwnym razie} \end{cases}$$







$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c)$$

$$\arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c)$$

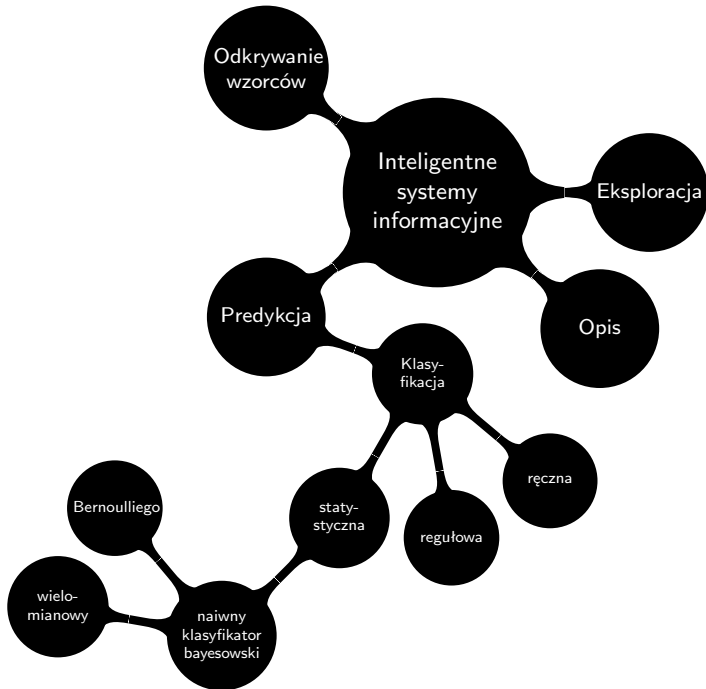
$$\arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c)$$

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c)$$

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(t_k|c) = \frac{T_{ct_k} + 1}{\sum_{t' \in V} (T_{ct'} + 1)}$$



model Bernoulliego

$$\arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{t_i \in V} \hat{P}(U_i = e_i|c)$$

model Bernoulliego

$$\arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{t_i \in V} \hat{P}(U_i = e_i|c)$$

$$\hat{P}(U_i = 1|c) = \frac{N_{t_i} + 1}{N + 2}$$

$$\hat{P}(U_i = 0|c) = \frac{N - N_{t_i} + 1}{N + 2}$$

preview	505	0	0,999983
similar	505	0	0,999983
clients	377	0	0,999977
completely	359	0	0,999976
safer	357	0	0,999976
unsafe	357	0	0,999976
virus	358	0	0,999976
Outlook	104	0	0,999918
MimeOLE	87	0	0,999902
HELO'd	85	0	0,999900
skills	75	0	0,999886
FREEMAIL_FROM	74	0	0,999885
Greetings	52	0	0,999836
drug	51	0	0,999833
Huge	49	0	0,999826
LOTS_OF_MONEY	49	0	0,999826

Wybór cech

- ▶ według frekwencji
- ▶ oczekiwana wspólna informacja (MI)
- ▶ test χ^2
- ▶ test Fishera

Metoda G. Robinsona

$$P = 1 - \sqrt[n]{\prod_{i \in 1, \dots, n} (1 - p_i)}$$

$$Q = 1 - \sqrt[n]{\prod_{i \in 1, \dots, n} p_i}$$

$$S = \frac{P - Q}{P + Q}$$

*Szanowni Państwo,
Szanując Państwa prywatność i zarazem przestrzegając
obowiązujących przepisów (Ustawa z dnia 18 lipca 2002
roku o świadczeniu usług drogą elektroniczną - Dz. U Nr
144, poz.1204) zwracamy się z prośbą o wyrażenie zgody
na przysłanie Państwu naszej oferty dotyczącej
przyczłapów do bulgulatorów.
Wspomnę jedynie, iż jako nowy producent przyczłapów
do bulgulatorów oferujemy atrakcyjne ceny.
Jeżeli jesteście Państwo zainteresowani , bardzo proszę
odpowiedzieć na tę wiadomość, prześlemy wtedy
dokładne informacje dotyczące ww. produktów.
Z poważaniem*

więcej niż słowa – CRM114

Tryb *Markovian*:

Szanując

Szanując Państwa

Szanując [...] prywatność

Szanując Państwa prywatność

Szanując [...] [...] i

Szanując Państwa [...] i

Szanując [...] prywatność i

Szanując Państwa prywatność i

Szanując [...] [...] [...] zarazem

Szanując [...] [...] i zarazem

Szanując [...] prywatność [...] zarazem

Szanując [...] prywatność i zarazem

Szanując Państwa [...] [...] zarazem

Szanując Państwa [...] i zarazem

Szanując Państwa prywatność [...] zarazem

Szanując Państwa prywatność i zarazem

więcej niż słowa – CRM114 (cd.)

Tryb *Markovian*:

Szanując Państwa

Szanując [...] prywatność

Szanując [...] [...] i

Szanując [...] [...] [...] zarazem

CRM114 – tryby trenowania

- ▶ TET (*train everything*)
- ▶ TOE (*train on errors*)
- ▶ SSTTT (*single sided thick threshold training*)