

# Řešení problémů a hry: Spam filter

Naive Bayes spam filter

# Autoři

Zviazdou Yauheni a Trenin Egor

# Spam filter

Spam filter je program, který chrání uživatele před spamem v e-mailu.

Pro hodnocení je potřeba předem trénovat spam filter sadou dat, která vypadá jako buď seznam e-mailů nebo seznam slov.

# Popis použitého principu hodnocení

## Naive Bayes

- 1) Výpočet pravděpodobnosti (Likelihood, Probability) je vytvoření histogramu s informací, která obsahuje pravděpodobnost toho, že můžeme narazit na to slovo v e-mailu.

$\Pr(W|S)$  nebo  $\Pr(W|H)$  = počet kolikrát se slovo objeví v seznamu, který je dělený celkovým počtem slov v seznamu.

- 2) To je potřebné udělat pro obě slovní zásoby: SPAM a HAM.

# Popis použitého principu hodnocení

## Naive Bayes

3) Výpočet počátečního předpokladu, jestli e-mail náleží do spamu nebo do hamu:  $\Pr(S)$  nebo  $\Pr(H)$ : počet SPAM nebo HAM e-mailu, který je dělený počtem všech e-mailů.

4) Násobení všech dat:  $\Pr(S)$  nebo  $\Pr(H)$  a  $\Pr(W|S)$  nebo  $\Pr(W|H)$  pro  $W_1, W_2$  až  $W_n$

# Popis použitého principu hodnocení

## Naive Bayes

5) Použijeme logaritmus, abychom se vyhnuli underflowu:  $\ln(\Pr(S))$  nebo  $\Pr(H) * \Pr(W|S)$  nebo  $\Pr(W|H)$  pro  $W_1, W_2$  až  $W_n$ ).

6) Použití  $SIGMA = 1$  pro  $\Pr(W|S)$  nebo  $\Pr(W|H)$ , které se rovnají nule, tzn. přidání  $SIGMA$  k počtu toho, kolikrát se slovo objeví ve slovníku pro každé slovo, jestli slovo se tam neobjeví nebo se počet rovná nule. To je nutné, abychom nedostali v odpovědi nulu.

# Testování

Pro testování byli použity soubory 1 a 2 z CourseWare Wiki -> Spam Filter -> Formát dat.

# Trénování

Pro trénování byly použity klíče z souboru !truth.txt. Zprvke program cte klíče, potom se vytváří dva listy: jeden se spam e-maily, druhý s ham e-maily. Pak se oba listy rozdělí na jednotlivá slova pomoci metody .split(). Pokud slovo z listu náleží do souboru english\_words.txt\*, přidává se slovo do slovníku.

\*english\_words.txt byl vyroben z knihovny NLTK words.

# 87%

Průměrná účinnost na tréninkových datech



# Rozdělení práce

Společné výzkum způsobů realizace spam filtru a rozhodnutí jakou metodu aplikovat abychom stihli včas

Společné kódování pomoci sdílení obrazovky přes Discord

# Použité nástroje

Telegram (messenger) pro odeslání  
souborů

Discord pro sdílení obrazovky

# Závěr

Naive Bayes není obtížný pro  
implementace a funguje podle  
očekávání, pokud nezapomenete na  
SIGMA

# Použité zdroje

<https://youtu.be/O2L2Uv9pdDA>

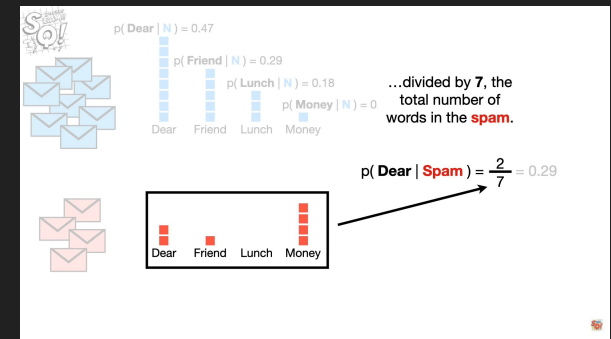
<https://youtu.be/HZGCoVF3YvM>

<https://youtu.be/XQoLVI31ZfQ>

[https://en.wikipedia.org/wiki/Most\\_common\\_words\\_in\\_English](https://en.wikipedia.org/wiki/Most_common_words_in_English)

<https://courses.cs.ut.ee/2004/dm-seminar-spring/uploads/Main/P06.pdf>

[https://en.wikipedia.org/wiki/Naive\\_Bayes\\_spam\\_filtering](https://en.wikipedia.org/wiki/Naive_Bayes_spam_filtering)



Ex/ What is the probability of two girls given at least one girl? GA, GB, BA, BB

$$P(2G | \text{at least } 1G)$$

$$= \frac{P(1G | 2G) \cdot P(2G)}{P(1G)}$$

$$= \frac{1}{2}$$

