НЕТОЛОГИЯ

Юстина Иванова

Программист, data scientist

Кейс-стади 2.
Поиск СПАМа в тексте.
Анализ текста на тональность.
Оценки студентов на экзамене.

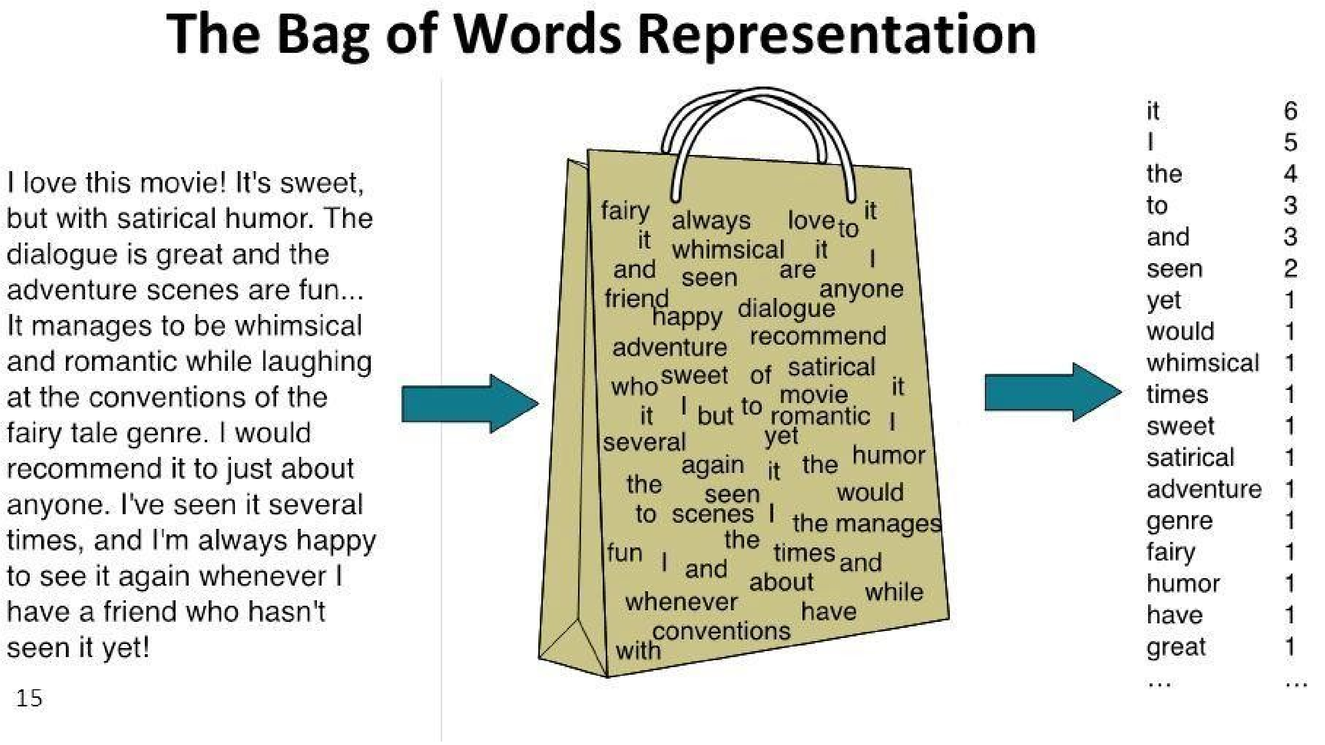
Спикер

НЕТОЛОГИЯ

Юстина Иванова,

- PhD в университете Больцано (Италия)
- Data scientist по компьютерному зрению в компании ОЦРВ, Сочи
- Выпускница МГТУ им. Баумана
- Магистр по Artificial Intelligence В University of Southampton (Англия)

BAG of words - Мешок слов

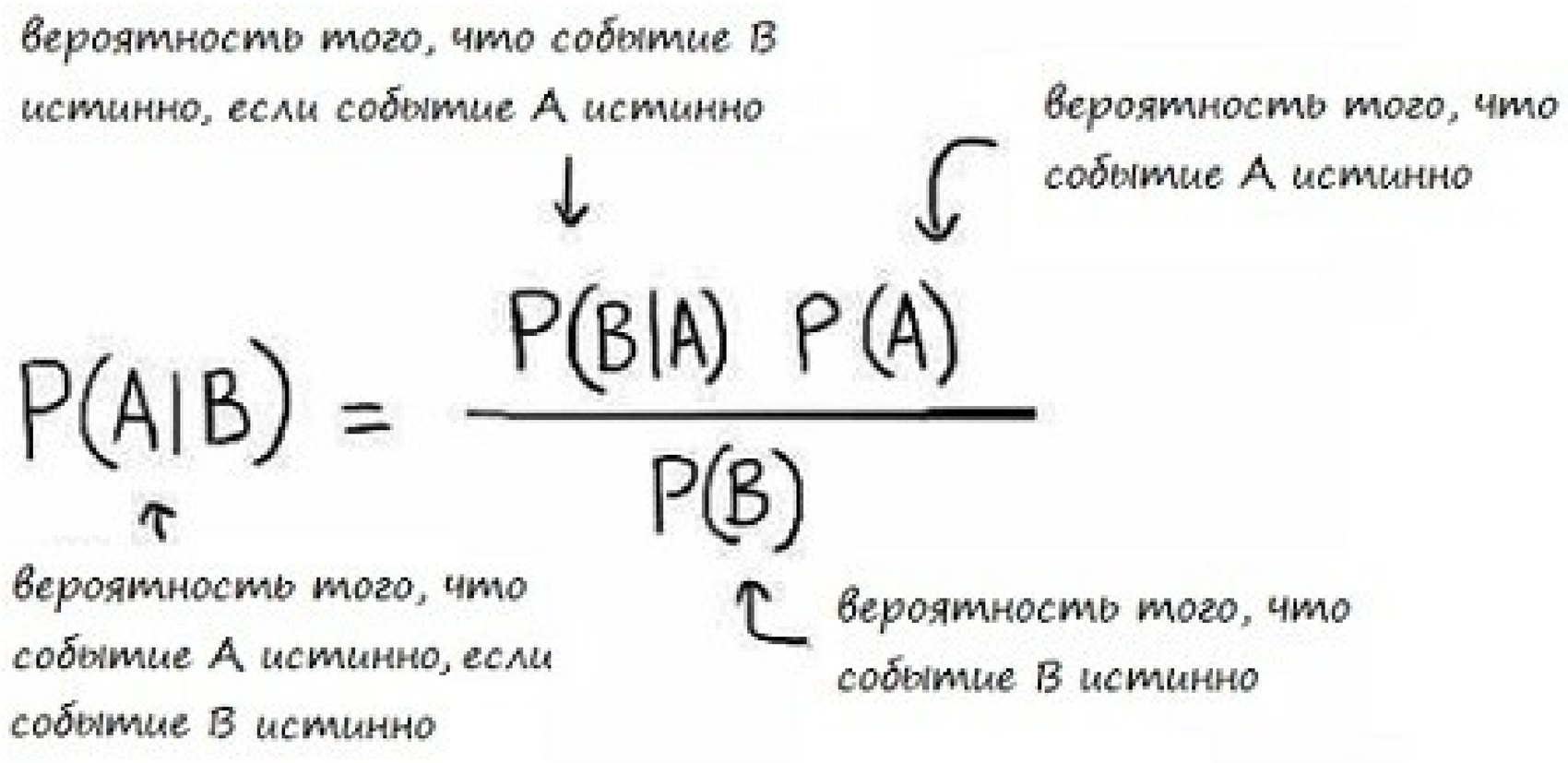


http://www.programmersought.com/article/4304366575/;jsessionid=0187F8E68A22612555B437068028C012

https://en.wikipedia.org/wiki/Bag-of-words_model

НЕТОЛОГИЯ

Теорема Баеса



http://baguzin.ru/wp/den-morris-teorema-bajesa-vizualnoe-vvedenie-dlya-nachinayushhih/

Пример.

1% женщин в возрасте 40 лет, участвовавших в регулярных обследованиях, имеют рак груди. 80% женщин с раком груди имеют положительный результат маммографии. 9.6% здоровых женщин также получают положительный результат (маммография, как любые измерения, не дает 100% результатов). Женщина-пациент из этой возрастной группы получила положительный результат на регулярном обследовании. Какова вероятность того, что она фактически больна раком груди?

http://schegl2g.bget.ru/bayes/YudkowskyBayes.html

Пример.

1% женщин в возрасте 40 лет, участвовавших в регулярных обследованиях, имеют рак груди. 80% женщин с раком груди имеют положительный результат маммографии. 9.6% здоровых женщин также получают положительный результат (маммография, как любые измерения, не дает 100% результатов). Женщина-пациент из этой возрастной группы получила положительный результат на регулярном обследовании. Какова вероятность того, что она фактически больна раком груди?

Правильный ответ - 7.8%, и получается он так: Из 10,000 женщин, 100 больны раком груди; 80 из этих 100 имеет положительные маммограммы. Из тех же 10,000 женщинn, 9,900 не имеют рака груди, и из этих 9,900 женщин 950 тоже получат положительные маммограммы. Таким образом, общее число женщин с положительными маммограммами 950+80 то есть 1,030. Из этих 1,030 женщин с положительными маммограммами, 80 реально больны раком. Вычисляя, пропорцию, мы получаем 80/1,030, или 0.07767, то есть 7.8%.

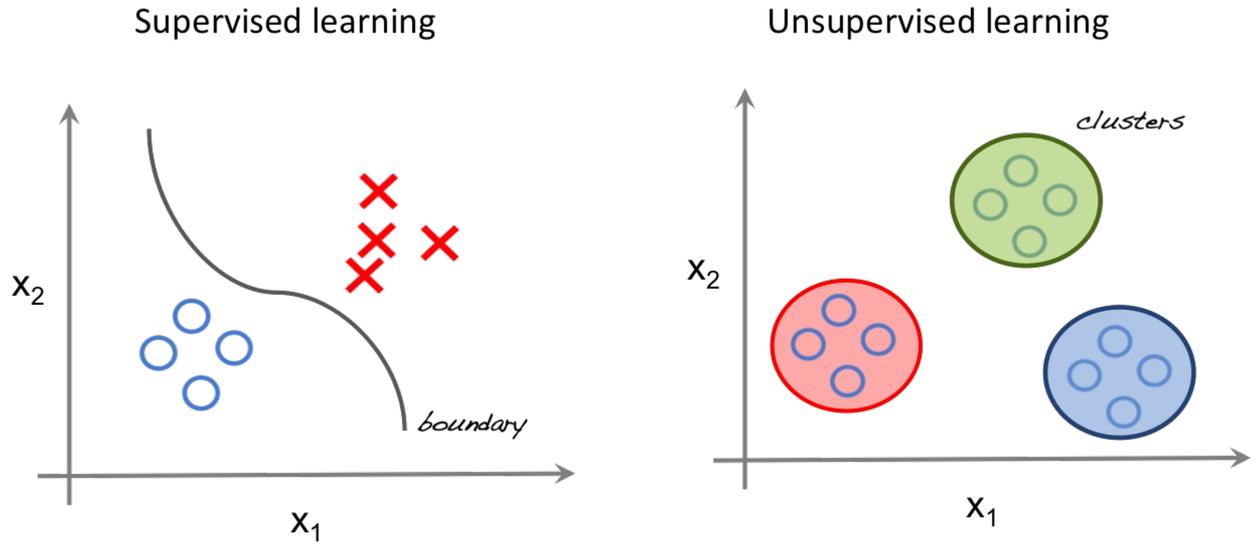
http://schegl2g.bget.ru/bayes/YudkowskyBayes.html

НЕТОЛОГИЯ


A FOREIGN DOG
...
IT PROBABLY
DOESN'T SPEAK
MY LANGUAGE
$\theta_{MLE} = argmax_\theta P(D|\theta)$
$P(\theta|D)$
f pleumier

НЕТОЛОГИЯ

Кластеризация



https://proglib.io/p/unsupervised-ml-with-python/

https://scikit-learn.org/stable/modules/clustering.html

НЕТОЛОГИЯ

Контакты спикера

E-mail: yustiks@gmail.com.