

# Задание 12 от 26.10.20 (Классификация текстов)

Валентин Александров, 620 группа

## Постановка задачи

Даны документы  $D1$ ,  $D2$ ,  $D3$ ,  $D4$  и их классы  $C1$  и  $C2$

Document	Terms	Class
D1	X1, X2, X3	C1
D2	X1, X2, X4	C1
D3	X4, X5, X6	C2
D4	X1, X4, X5	?

Определить класс документа  $D4$  на основе метода наивного Байеса

## Решение

---

Чтобы избавиться от нулевых множителей добавим сглаживание.

Для удобства подсчета введем

	X1	X2	X3	X4	X5	X6	
C1	3/12	3/12	2/12	2/12	1/12	1/12	2/3
C2	1/9	1/9	1/9	2/9	2/9	2/9	1/3

Class	
C1	8.03e-6
C2	5.01e-6

Ответ: **C1**