

Отчет по лабораторной работе №0 по курсу «Искусственный интеллект»

Студент группы 8О-407 Зайцев Никита, № по списку 7.

- **Инструментарий**

Язык программирования Python, Matplotlib для визуализации

- **Текстовый датасет**

Ссылка:

<http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

Датасет содержит отзывы на различные категории товаров(музыка, книги, игрушки, электроника и т.д.), отзывы делятся на положительные и отрицательные, некоторые отзывы вместо положительной/отрицательной оценки имеют пятибалльную систему оценки, которую при желании можно преобразовать в двоичные метки.

Пример отзыва:

```
<review>
<unique_id>
B00004YWGC:what_can_i_say?:ms._aj_"right"
</unique_id>
<unique_id>
1300
</unique_id>
<asin>
B00004YWGC
</asin>
<product_name>
Back For The First Time: Music: Ludacris
</product_name>
<product_type>
music
</product_type>
<product_type>
music
</product_type>
<helpful>
1 of 13
</helpful>
```

<rating>

2.0

</rating>

<title>

What can I say?

</title>

<date>

May 4, 2006

</date>

<reviewer>

Ms. AJ "Right"

</reviewer>

<reviewer_location>

North Carolina, USA

</reviewer_location>

<review_text>

I've always held the philosophy you are what you're entertained by. So, those who like Ludacris are definitely telling on themselves about their views of life and women. And speaking of women...Ludacris never met one who wasn't a h-o.

Ludacris is the epitome of the sexual hypocrisy that plagues men: you demean and look down on women for being h-oes, yet you hope every woman is a h-o so you'll have no problems getting laid. But then, you're so disgusted with the fact the woman behaves too much like you, so you go and find a "nice girl" because you can respect her and marry her and procreate with her. But then you cheat on her with a "bad girl" because h-oes need h-oes; but then you can't respect the "bad girl" because she's a slut, so you go back to the "good girl." But she's not nasty enough for you, so then...around and around we go. Women are basically being punished for men not being able to make up their minds about what the heck they want from us.

He's OBSESSED with sex and h-oes. It makes you wonder if he has hang-ups about it. Maybe Ludacris saw his mother selling herself in a backalley for a dollar and it scarred his outlook on women forever. I guess if she had actually RAISED him and shown him how to respect women-- ALL women--then he probably could have grown up to be a respectable guy instead of the prurient dummy he is.

What's the music like? Well I'll give him 2 stars for mentioning TLC in What's Your Fantasy. That's it.

Addendum: 7/27-07 Some may wonder why I'm "wasting" my time giving Ludacris negative reviews. Well, first of all, if only people who liked the product were the only ones to review it, everything on Amazon would have 5 stars, which would be an inaccurate sign of its quality.

Secondly, I don't like Ludacris. I would give Ludacris more respect if I felt he gave my gender more respect; but he doesn't. My diatribe above shows what my problem with him is, and I feel that argument is valid. There's no way to critique Ludacris' music without critiquing his hypocrisy regarding women. Why should I like him

when he seems to have nothing but contempt for us? I'm biased against him because he's biased against me. So he gets a bad rating out of s-p-i-t-e
</review_text>

Данный отзыв был взят из negative.review, значит его можно будет использовать для классификации негативных отзывов.

Размер датасета:

22972 — положительных отзывов

16576 — негативных отзывов

Возьмём теперь 4 категории из всех в которых по 1000 негативных и положительных отзывов в каждом: dvd, books, electronics, kitchen — и посчитаем в них количество слов

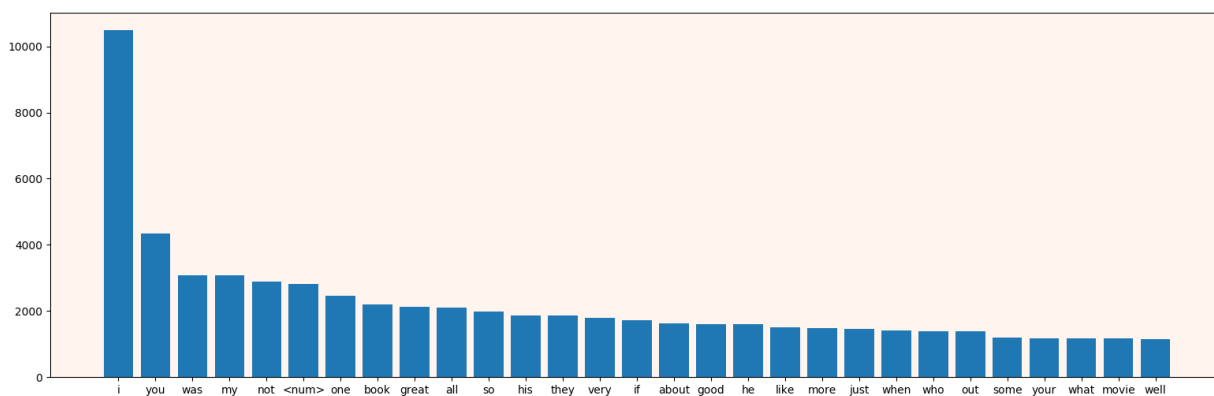
865657 в негативных отзывах

871346 в положительных

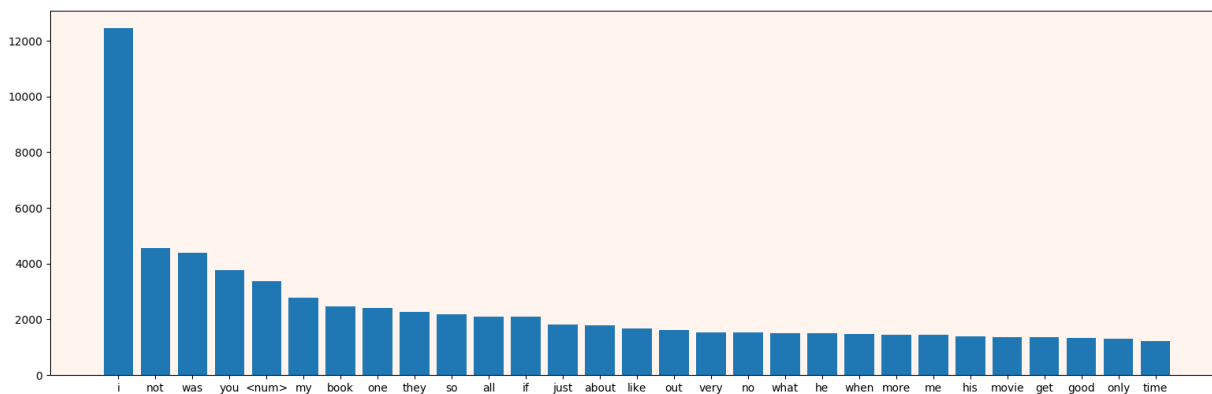
В среднем по 217 слов в обзоре

Анализ данных:

Распределение 30 самых часто встречающихся слов в положительных отзывах в 4 категориях



Распределение 30 самых часто встречающихся слов в негативных отзывах в 4 категориях



По распределению можно понять, что, в основном, чаще всего встречаются слова которые относятся к обоим отзывам примерно одинаково часто, это обычные разговорные слова не влияющие на оценку. Можно построить таблицу таких слов и не учитывать их при обучении, либо же выбрать методы, которые смогут это учитывать.

- **Количественные признаки**

Ссылка:

<https://archive.ics.uci.edu/ml/datasets/2.4+GHZ+Indoor+Channel+Measurements>

- Измерение параметра рассеяния S21 беспроводного канала

Категории классификации:

Замеры проводились в 4 разных местах:

- 1- Lab139 (сильная загромождённость)
- 2- Corridor_rm155 (средняя загромождённость)
- 3- Main_Lobby (низкая загромождённость)
- 4- Sports_Hall (открытое пространство).

Эти места мы будем классифицировать.

Размер:

В этих местах установлены излучатели и приёмники радиосигнала, сигнал будет изменяться в зависимости от загромождённости места.

В каждом месте измерение производится на 10 циклах, с 601 частотной точкой(каждый раз несколько изменялась частота).

Изменяя расположение излучателя было получено 196 положений, учитывая 10 циклов получаем 1960 образцов по 601 точке, в итоге получаем 1177960 точек для каждого из мест.

Атрибуты:

Этот набор данных имеет 5 атрибутов:

- 1- частота
- 2- Действительная часть параметра S11
- 3- Мнимая часть параметра S11
- 4- Действительная часть параметра S21
- 5- Мнимая часть параметра S21

Пример точек:

2.403333333333334E+009;1.655690968036652E-001;6.305278092622757E-002;-
4.648590984288603E-004;-2.615027362480760E-003;

Измерения были разработаны специально для изучения диапазонов WiFi, каналов 1, 6 и 11 в стандарте IEEE 802.11.

Статистика:

Посмотрим как ведут себя величины. В каждом помещении используют одинаковые диапазоны частот, поэтому в зависимости от неё мы можем определить как изменяются остальные аргументы в разных помещениях.

По горизонтали частота излучения(0 аргумент), по вертикали значения аргументов(1-4)

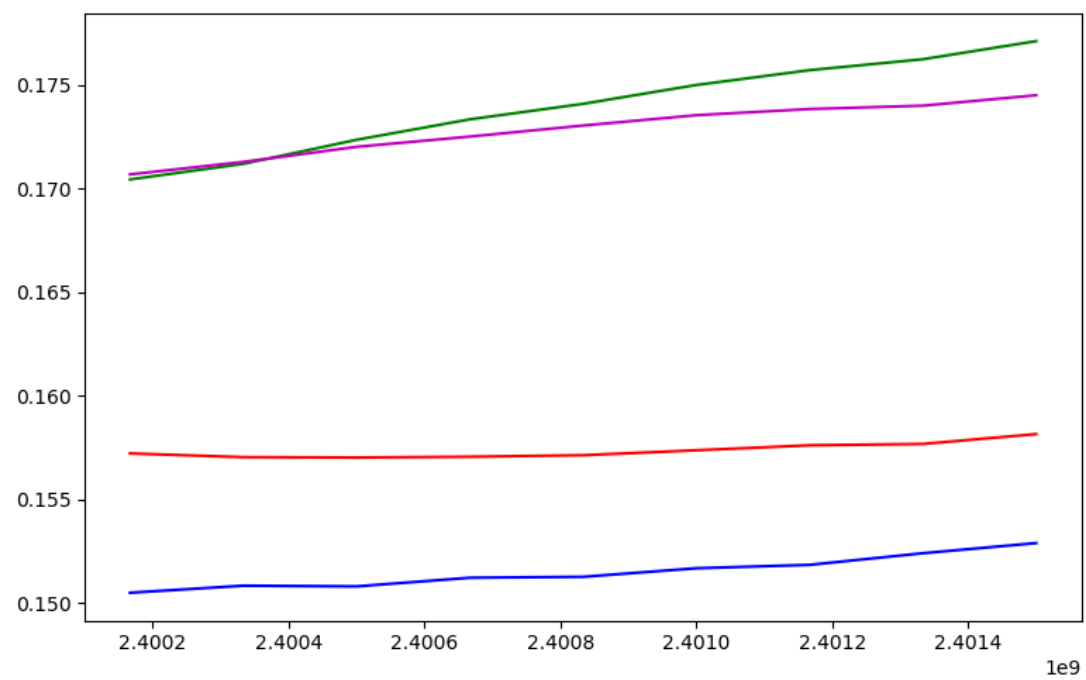
Синий - Corridor_rm155_7.1

Зелёный - Main_Lobby_7.1

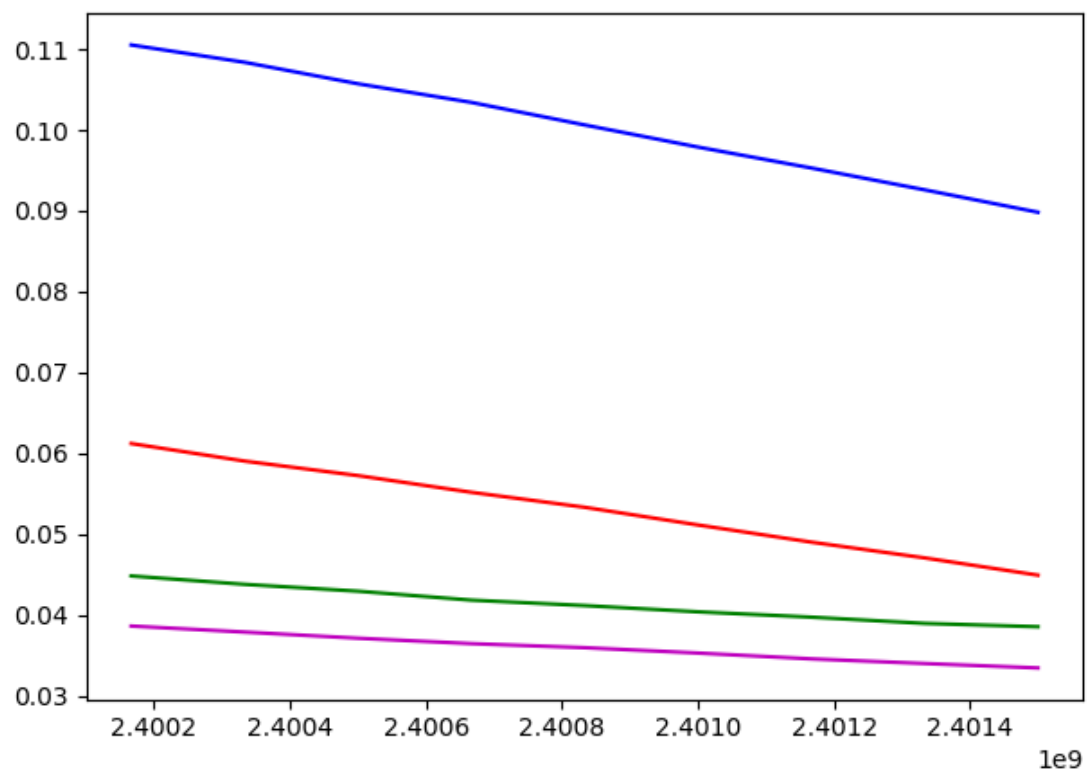
Красный - Lab139_7.1

Пурпурный — Sport_Hall_7.1

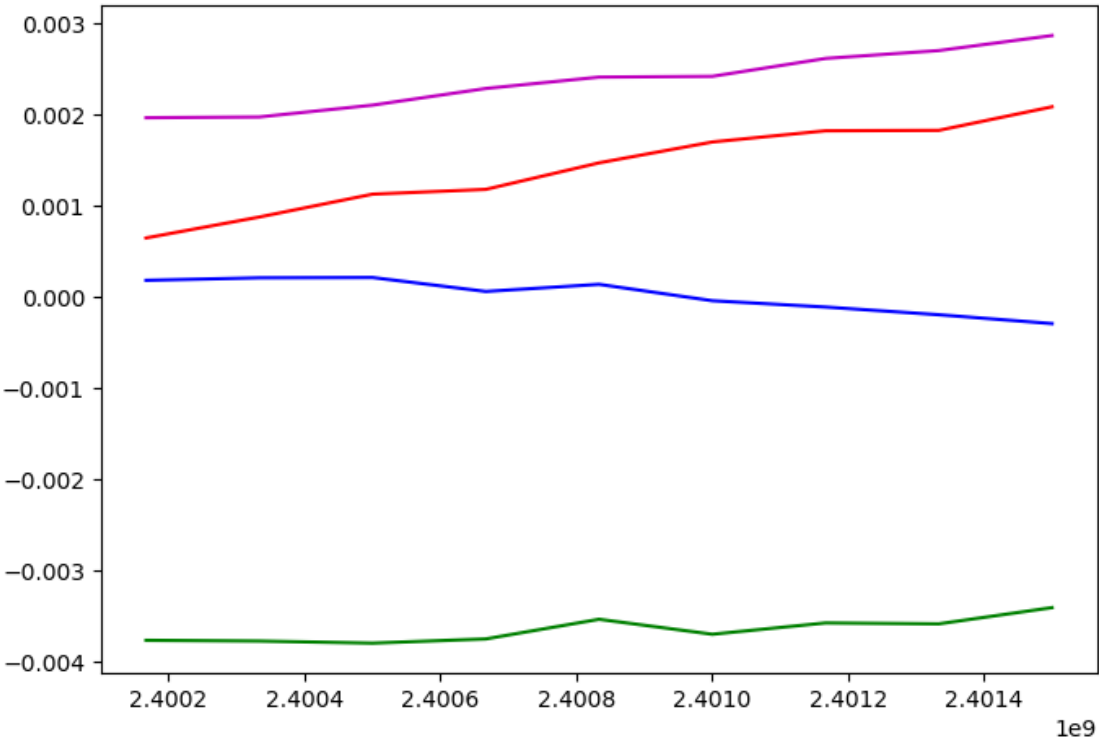
Изменение первого аргумента



Изменение второго аргумента



Изменение третьего аргумента



Изменение четвертого аргумента

