

Отчёт №1
о проделанной работе по проекту

ФИО студента: Кашникова Анна Дмитриевна

Группа: 21-ПМИ-2

Образовательное учреждение: национальный исследовательский университет
«Высшая школа экономики»

1 Анализ датасета: общая информация

DEAM dataset содержит 1802 фрагментов аудио, из которых 1744 - это отрывки из песен, длительностью 45 секунд и с частотой дискретизации 44100Hz, извлеченные из случайной(равномерно распределенной) начальной точки в песне, и 58 - это отдельные полноценные песни. Сами аудио представлены в mp3 формате. Также приведены аннотации для песен, они даны набором измерений аудио для каждого работника, кто занимался аннотированием данных, также они представлены в подсчитанных средних значениях по всем работникам для одной песни. Отдельно в датасете приведены извлеченные признаки с помощью openSMILE для всех аудио. Более подробно в [1].

2 Анализ датасета: аннотации для каждого работника

Общие наблюдения:

- Количество работников, принимавших участие в аннотировании данных, составляет 201 человек.
- В датасете предложены аннотации 2-х видов - динамические, работник оценивал arousal и valence для песни каждые 0.5 секунд(прим. разработчик датасета удалил аннотации для первых 15 сек, так как измерения работников были нестабильны), оценки лежат в интервале $[-1, 1]$; и статистические, одна оценка arousal и valence для всей песни в целом, оценки ставились по 9-бальной шкале.

	WorkerId	sample_15000ms	sample_15500ms	sample_16000ms
0	883449004b1bf2a07a284f59ddae1fd7	0.10	0.10	0.10
1	b09a5957e5d5e47e556d203529a0ae6d	-0.07	-0.08	-0.08
2	a8d14f76676af36b8978406be47c0c38	0.07	0.07	0.07
3	02bf7a99a5e47f4d52939ab7efc8a549	-0.60	-0.60	-0.60
4	467b1ff3ccfff51e2882bd9d39ef2082	-0.03	-0.03	-0.03

Скрин. 1 Пример динамических измерений работников для 2001 аудио.

	workerID	SongId	Valence	Arousal
0	6010bbc8e7ef4b21fa38f9c3a9754ef3	2	5	2
1	3c888e77b992ae3cd2adfe16774e23b9	2	2	3
2	2afd218c3aecb6828d2be327f8b9c46f	2	3	3
3	fd5b08ce362d855ca9152a894348130c	2	4	4
4	9c8073214a052e414811b76012df8847	2	2	2
...

Скрин. 2 Пример статистических измерений работников.

- Также было замечено, что некоторые участники, аннотировали данные больше 1 раза:

	workerID	SongId	Valence	Arousal
2730	6010bbc8e7ef4b21fa38f9c3a9754ef3	323	7	6
2731	bf34e3c5724ce07d29ef12db5f767258	323	4	6
2732	24ebcf86498b0f2793b55c5b9a7756b5	323	5	4
2733	38531641e6c0628757776b0088bcc854	323	6	6
2734	2afd218c3aecb6828d2be327f8b9c46f	323	3	5
2735	fd5b08ce362d855ca9152a894348130c	323	6	4
2736	9c8073214a052e414811b76012df8847	323	6	7
2737	64cae7b86c7dcb1b40d17e43c0c2109e	323	3	3
2738	64cae7b86c7dcb1b40d17e43c0c2109e	323	4	8
2739	259443e9cad56b4f68b664daa20cb323	323	3	7
2740	490951556961e4b88fe1a3ac53b3b186	323	6	7

Скрин. 3 static annotations songs 1 2000.csv.

- в 2013 году не собиралась информация по id работникам для динамических измерений.
- в 2015 году собиралось больше статистической информации о песне, кроме arousal и valence, приводятся также их максимумы и минимумы.

	SongId	WorkerId	Arousal_Average	Valence_Average	Arousal_Maximum
0	2001	523edc3414996c8e52eb14f20c93fe96	7	3	8
1	2001	ae0b751438f08bb474e5cd3e48e0d417	5	3	9
2	2001	fd4cc3ae7434a18286204f133b490502	7	3	8
3	2001	58af5138dd2f00af1f52d2c2c031f4a2	7	2	9
4	2001	3c960e4e5a02213fc82b98b770a0663c	7	5	9

Скрин. 4 static annotations songs 2000 2058.csv.

Valence_Maximum	Arousal_Minimum	Valence_Minimum
7	2	2
3	3	2
5	4	2
4	5	1
6	3	4

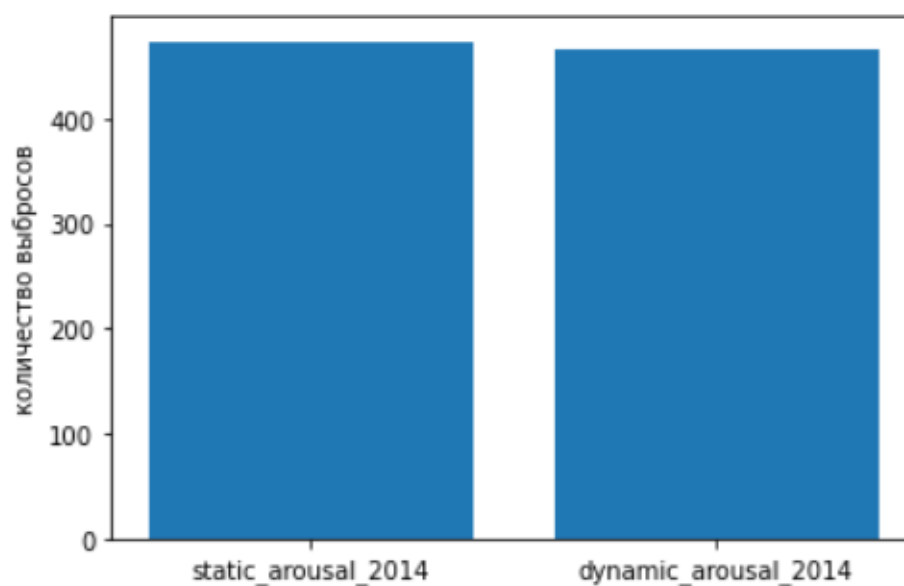
Скрин. 5 продолжение static annotations songs 2000 2058.csv.

- в 2013-2014 году в аннотировании одного аудио принимало участие 10 работников, в 2015 - 5 работников.

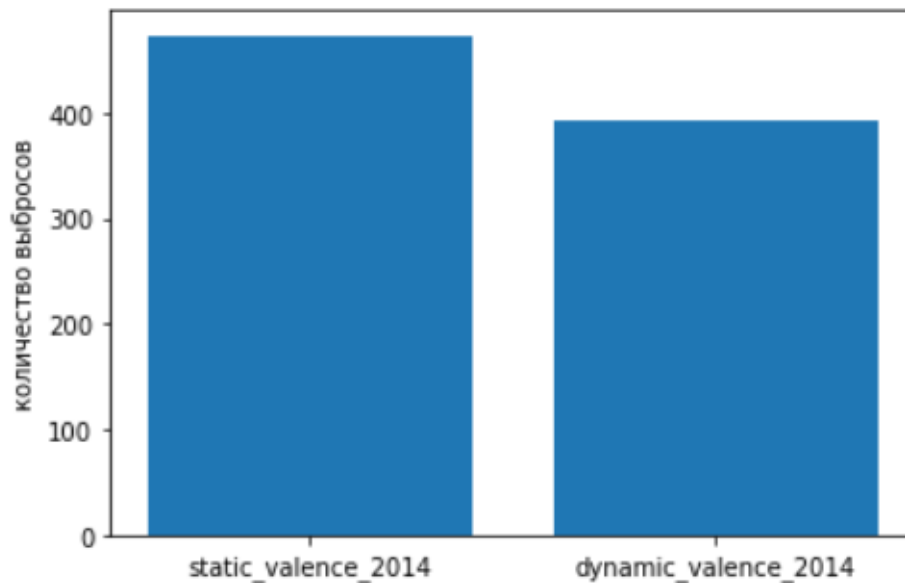
Анализ на выбросы среди работников

Я посчитала количество выбросов для каждого работника для разных измерений и вот какая статистика получилась:

- В нижеприведенных графиках, было подсчитано для каждой песни количество выбросов среди работников, отдельно как для статистических измерений и средних динамических, и просуммировано[3]:



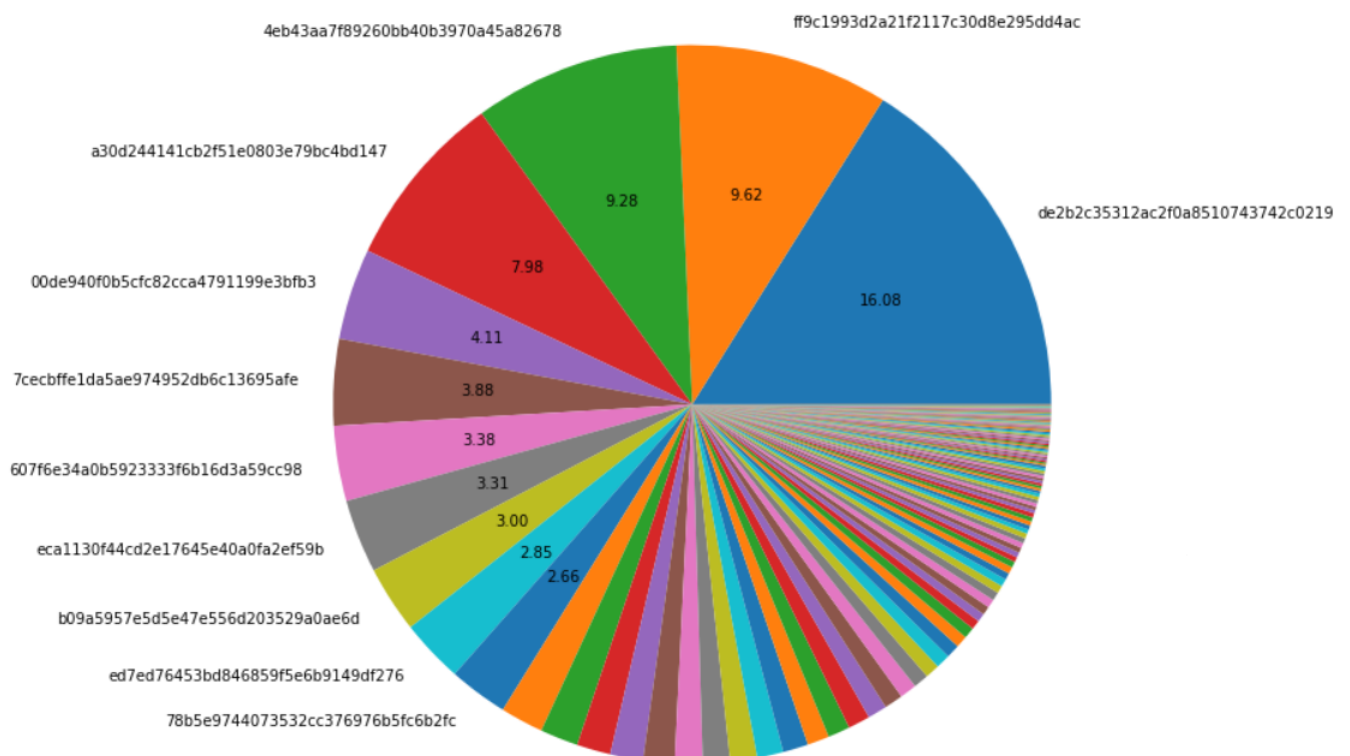
Скрин. 6 Сравнение количества выбросов для arousal в 2014.



Скрин. 7 Сравнение количества выбросов для valence в 2014.

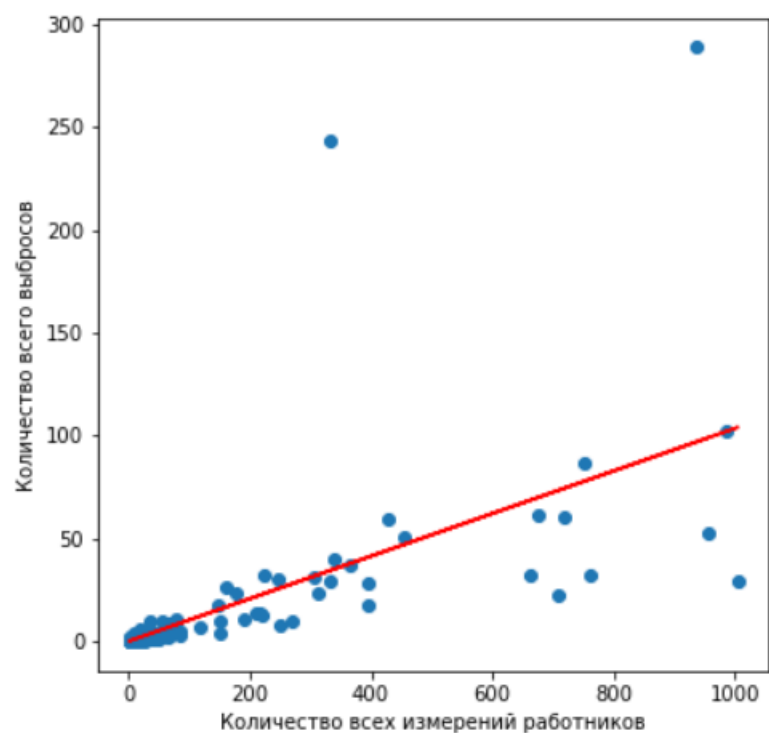
На основе этого можно заметить, что динамические измерения содержат в себе меньше выбросов, по сравнению со статическими, и в теории динамические значения были бы лучшим вариантом в качестве аннотаций.

- Также я собрала статистику по id работникам, и по их выбросам по всем видам измерений в 2013-2014 году[3]:

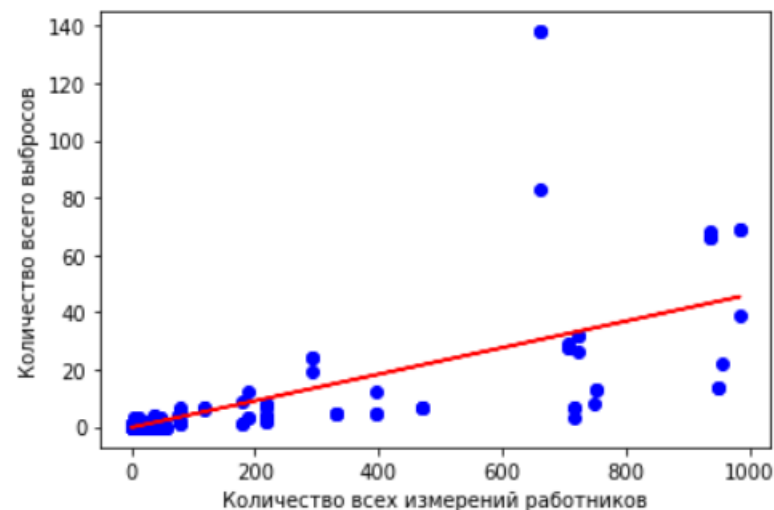


Скрин. 8 Диаграмма, содержащая долю выбросов в данных 2014 года на каждого работника.

Из диаграммы видно, что около 16% всех выбросов приходится на всего одного работника с $id = de2b2c35312ac2f0a8510743742c0219$, что достаточно много, также около 9% всех выбросов принадлежат двум другим работникам. Для более равномерных данных, можно удалить из данных работников с высоким процентом. Однако можно предположить, что количество выбросов растёт вместе с количеством аннотируемых данных работников, рассмотрим эту связь[3]:



Скрин. 9 Связь частоты аннотирования работников и количества выбросов в 2013 и 2014 статика.



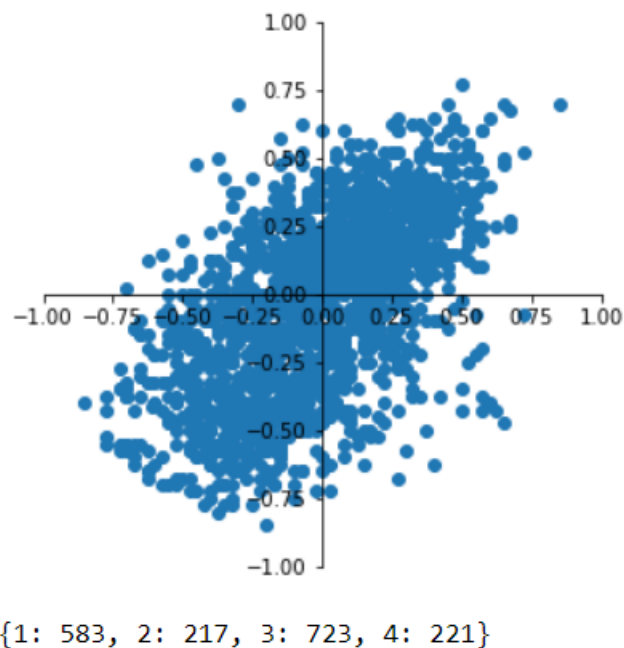
Скрин. 10 Связь частоты аннотирования работников и количества выбросов 2014 динамика.

Из графиков видно явную связь, чем больше работник проаннотирует данных, чем выше количество выбросов для него(такая же тенденция сохраняется для данных 2015 года). Таким образом, если очищать данные, то искать конкретные выбросы работников с высоким процентом.

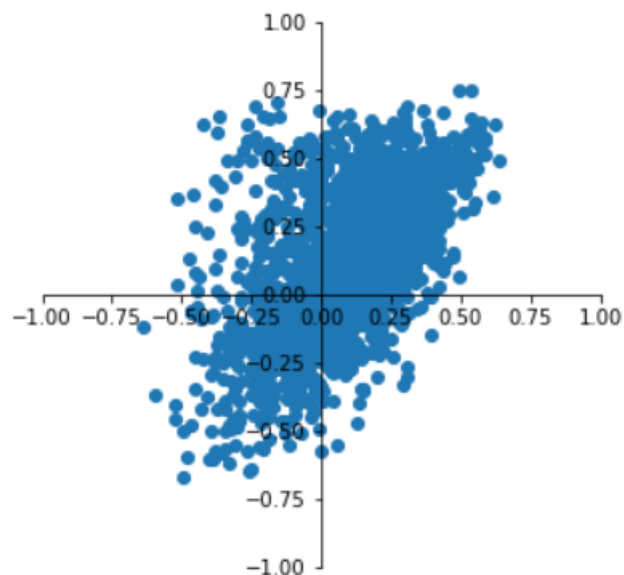
3 Анализ датасета: усредненные аннотации для каждой песни

Усреднённые аннотации для каждой песни были высчитаны на основе оценок работников, также для них приводится стандартное отклонение.

- Рассмотрим связь arousal и valence для всех песен[3]:



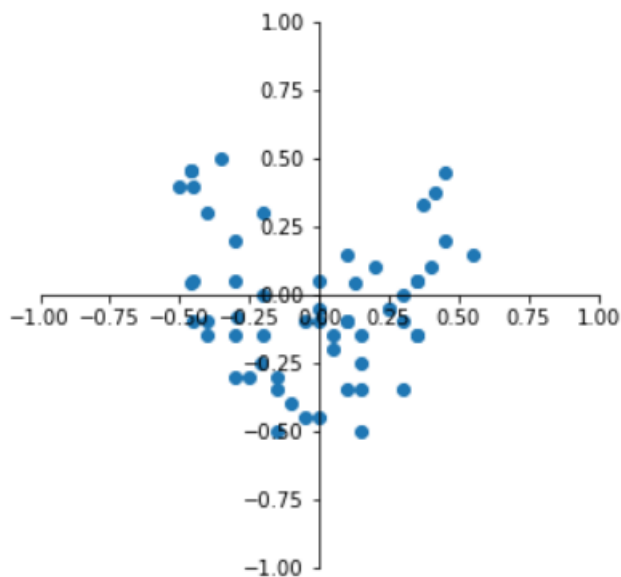
Скрин. 11 График всех песен в 2013-2014 годах построенный по arousal и valence статика, и количество песен, попадающие в соответствующие четверти.



{1: 954, 2: 220, 3: 374, 4: 196}

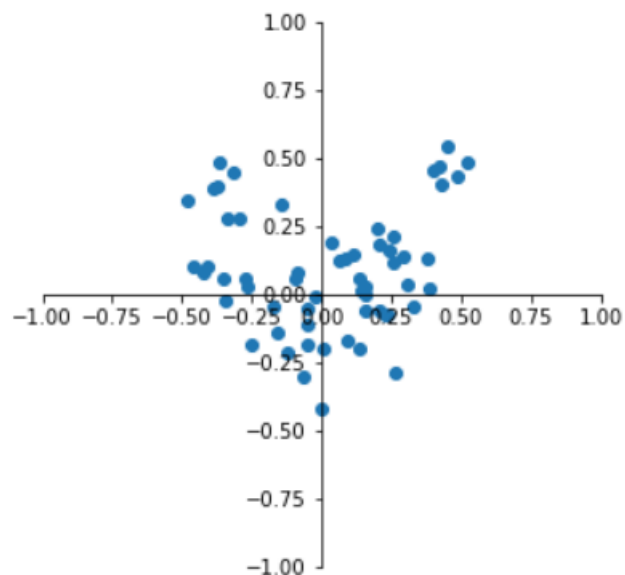
Скрин. 12 График всех песен в 2013-2014 годах построенный по arousal и valence динамика, и количество песен, попадающие в соответствующие четверти.

Из графиков видно прямопропорциональную связь между arousal и valence, также можно отметить, что статистические оценки более "разбросаны по сравнению с динамическими.



{1: 11, 2: 12, 3: 19, 4: 16}

Скрин. 13 График всех песен в 2015 году построенный по arousal и valence статика, и количество песен, попадающие в соответствующие четверти.

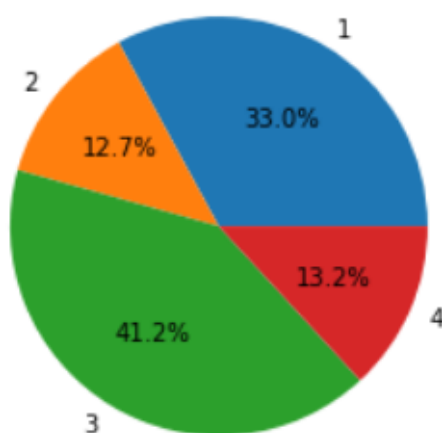


{1: 23, 2: 16, 3: 11, 4: 9}

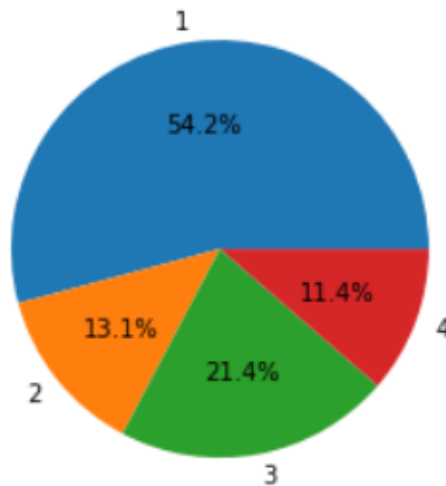
Скрин. 14 График всех песен в 2015 году построенный по arousal и valence динамика, и количество песен, попадающие в соответствующие четверти.

Картина 2015 года отличается, явной связи нет, данные распределены более равномерно по всем четвертям.

- Рассмотрим подробнее распределение по четвертям[3]:



Скрин. 15 Статистическое распределение попаданий в четверти.

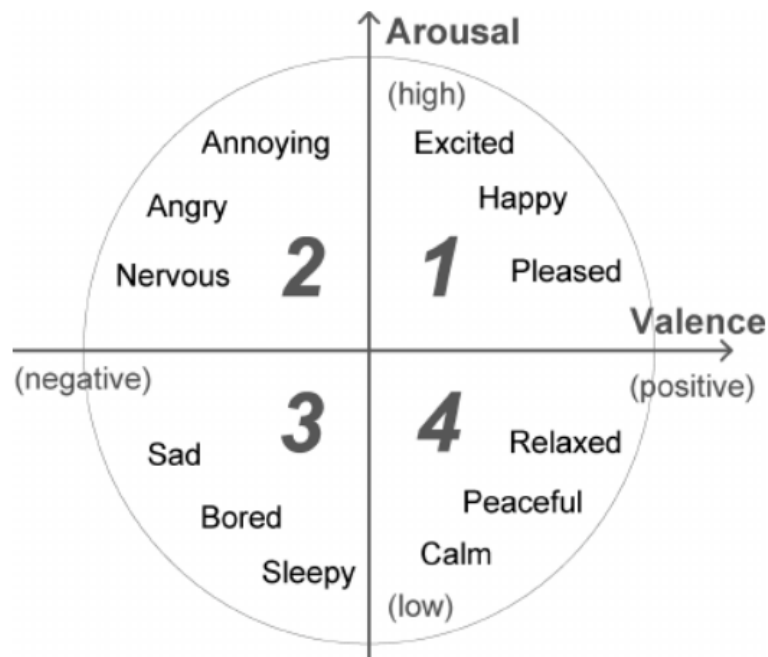


Скрин. 16 Динамическое распределение попаданий в четверти.

На основе диаграмм выше, можно сделать вывод, что более сбалансированы статистические оценки, следовательно имеет смысл классифицировать данные на 4 класса исходя из статистических оценок.

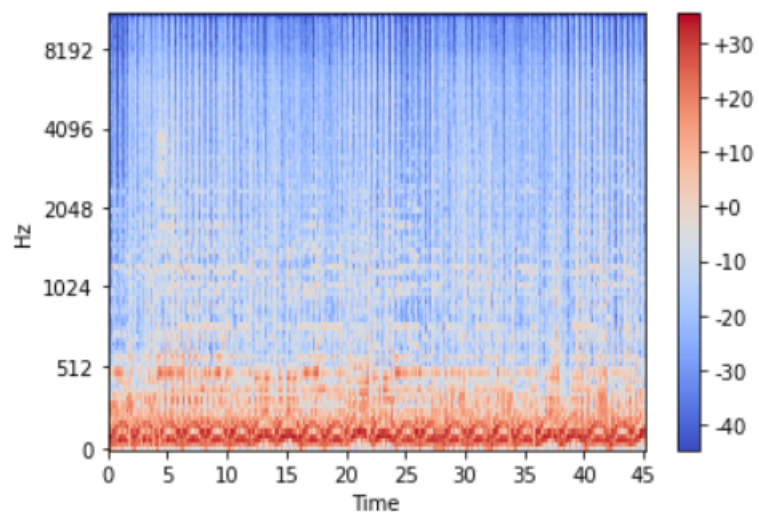
4 Обзор аудио по классам

Исходя из вывода выше, я распределила аудио по классам следующим образом[3]:

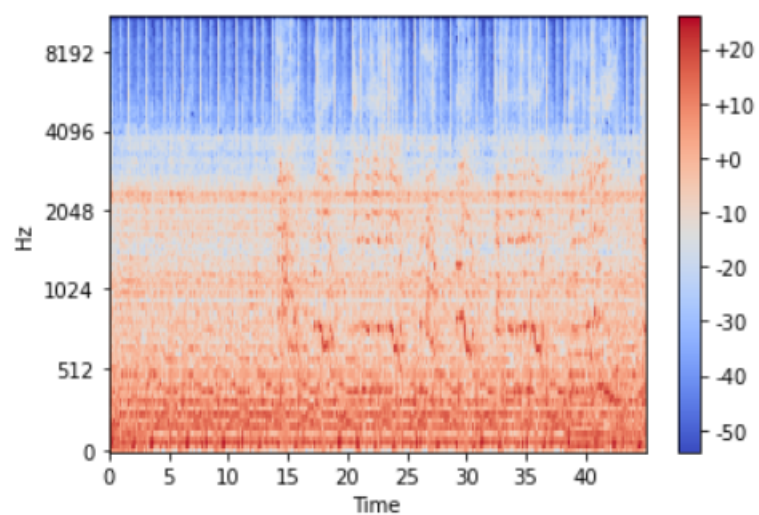


Скрин. 17 Распределение эмоций по классам.

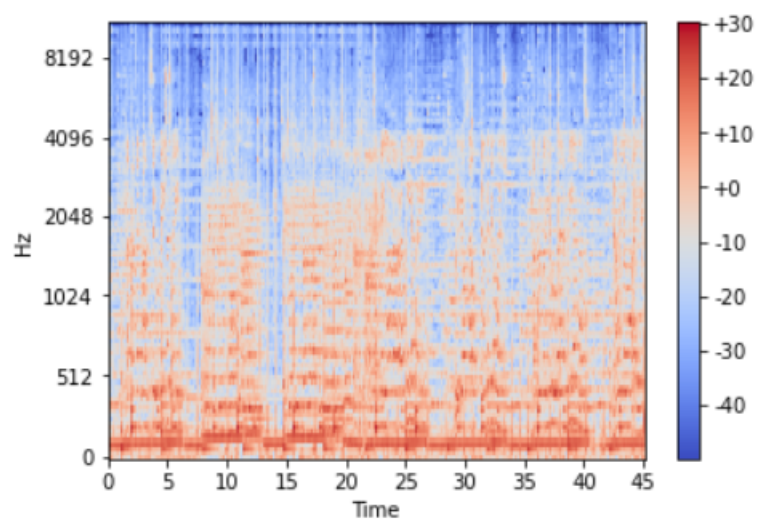
Я взяла по представителю каждого класса и решила посмотреть их mel-спектрограмму:



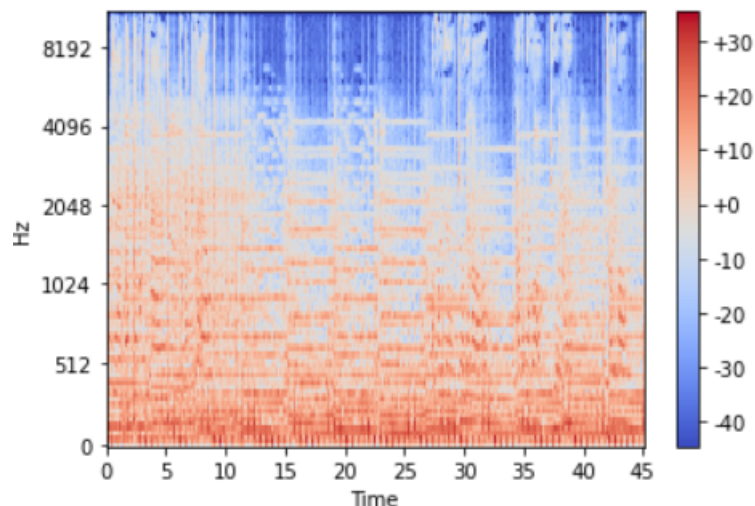
Скрин. 18 mel спектрограмма для представителя 1 класса, 7.mp3.



Скрин. 19 mel спектрограмма для представителя 2 класса, 1996.mp3.



Скрин. 20 mel спектрограмма для представителя 3 класса, 2.mp3.



Скрин. 21 mel спектрограмма для представителя 4 класса, 19.mp3.

Представителя 1 класса от всех остальных отличают мощные низкие частоты, 2 класс имеет градиентную спектрограмму, мощность частот плавно снижается от низких к высоким, 3 класс имеет явные гармоники и много "пустот" в высоких частотах, 4 класс достаточно зашумлен, по сравнению с другими.

5 Инструменты

В ходе анализа использовался jupyter notebook, в качестве библиотеки для построения графиков - matplotlib, для работы с аудио применялась библиотека librosa.

Изученные статьи и код анализа

- [1] Aljanaki A, Yang Y-H, Soleymani M (2017) Developing a benchmark for emotional analysis of music. PLoS ONE 12(3): e0173392. <https://doi.org/10.1371/journal.pone.0173392>
- [2] <https://habr.com/ru/companies/otus/articles/757130/>
- [3] <https://github.com/annkash/MIR.git>