

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**РАЗРАБОТКА МУЗЫКАЛЬНОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ
С УЧЕТОМ НАСТРОЕНИЯ И МУЗЫКАЛЬНОГО ЖАНРА**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 5 курса 551 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Силинской Анны Александровны

Научный руководитель
доцент, к. ф.-м. н.

А. С. Иванова

Заведующий кафедрой
к. ф.-м. н., доцент

С. В. Миронов

Саратов 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Определение музыкального настроения и создание датасета.....	5
1.1 Влияние музыкальных характеристик на настроение	5
1.1.1 Циркумплексная модель Расселла	5
1.1.2 Прочие характеристики	6
1.2 Создание датасета	6
1.2.1 Анализ данных	7
2 Сравнение моделей для предсказания вероятностей	9
3 Структура приложения	10
3.1 Определение музыкальных жанров	10
3.2 Описание пользовательского интерфейса	10
3.3 Описание интерфейса администратора	12
ЗАКЛЮЧЕНИЕ	14

ВВЕДЕНИЕ

В настоящее время цифровое поле пользователя перенасыщено информацией и контентом. Стриминговые видео-, аудио- и киносервисы предлагают такое количество часов контента, для просмотра которого у пользователя не хватит всей жизни. Чтобы облегчить пользователю поиск интересующего контента, стриминговые сервисы вводят рекомендательные технологии, которые анализируют поведение пользователя с целью предсказания его вкусов и предложения наиболее релевантного контента. Музыкальные стриминговые сервисы не являются исключением.

Самый частый принцип работы музыкальных рекомендательных систем заключается в предсказании вкусов пользователя на данных ранее прослушанных композиций и сохраненных плейлистов. Такие алгоритмы используются ключевыми российскими стриминговыми сервисами, такими как Яндекс Музыка, VK Музыка, Звук, а также иностранными Apple Music, Spotify и прочими. Однако, такой алгоритм, анализирующий всю историю прослушиваний и предлагающий на основе этого композиции, не учитывает возможный запрос пользователя подобрать музыку в соответствии с некоторым настроением, которое она передает. Такой запрос может отличаться от типичного поведения пользователя. В наше время некоторые стриминговые сервисы (Яндекс Музыка, Звук) разрабатывают функционал для такого запроса. Однако эти сервисы могут составлять плейлисты не учитывая жанровые предпочтения пользователя.

По этой причине целью работы является разработка такой рекомендательной системы, которая бы генерировала музыкальный плейлист как в соответствии с предпочтительным жанром, так и с выбранным настроением. Для этого необходимо собрать данные, размеченные по настроению и выполнить задачу предсказания вероятностей принадлежности песен к разным категориям настроения с помощью алгоритмов машинного обучения. Также нужно выделить основные жанровые категории и собрать данные о жанровой принадлежности композиций. Это позволит составить базу данных, которая будет использоваться в веб-приложении для генерации рекомендуемых плейлистов.

Выпускная квалификационная работа состоит из трех разделов.

В первом разделе описаны вопросы взаимосвязи различных музыкальных характеристик с настроением и выделены основные категории настроений, которые будут использоваться в задаче предсказания вероятностей и рекомен-

дательной системе. Далее описан метод сбора данных для обучения классификаторов и анализ собранных данных.

Во втором разделе описаны классификаторы для предсказания вероятностей, метрики оценки их качества, методы калибровки вероятностных прогнозов, описаны гиперпараметры моделей и метод их подбора. В конце раздела сравниваются метрики нескольких обученных классификаторов.

В третьем разделе описан метод сбора данных о жанрах композиций, а также выделены основные жанровые категории, используемые в работе. Затем описана структура разработанного веб-приложения, а именно: использованные технологии, функции, доступные пользователям и администраторам приложения и некоторые детали их реализации.

1 Определение музыкального настроения и создание датасета

1.1 Влияние музыкальных характеристик на настроение

1.1.1 Циркумплексная модель Расселла

Циркумплексная модель эмоций Джеймса Расселла – это двумерная модель, которая отображает эмоции по измерениям валентности (положительные эмоции против отрицательных) и энергичности (высокая против низкой, также может называться возбуждением), валентность представляет собой горизонтальную ось, а энергичность – вертикальную. Согласно этой модели, эмоции можно расположить на круговом графике, где положительные эмоции находятся в правой половине круга, а отрицательные – в левой. В верхней половине круга располагаются более энергичные эмоции, а в нижней половине – менее энергичные. Интенсивность эмоции (по обоим измерениям) обозначается расстоянием от эмоции до центра круга, чем ближе эмоция к внешней половине круга, тем большей интенсивностью она обладает. Важно отметить, что согласно этой модели, валентность и энергичность являются независимыми характеристиками.

Для выделения основных категорий настроения, а также первых музыкальных признаков, с помощью которых можно было бы определить принадлежность музыкальных произведений к этим категориям, в данной работе предлагается использовать схожее с моделью Расселла представление эмоций в прямоугольных координатах, где валентность и энергичность принимают вещественные значения от 0.0 до 1.0, а координатные четверти соответствуют категориям настроений (см. рисунок 1). Музыка будет классифицироваться по четырем основным категориям.

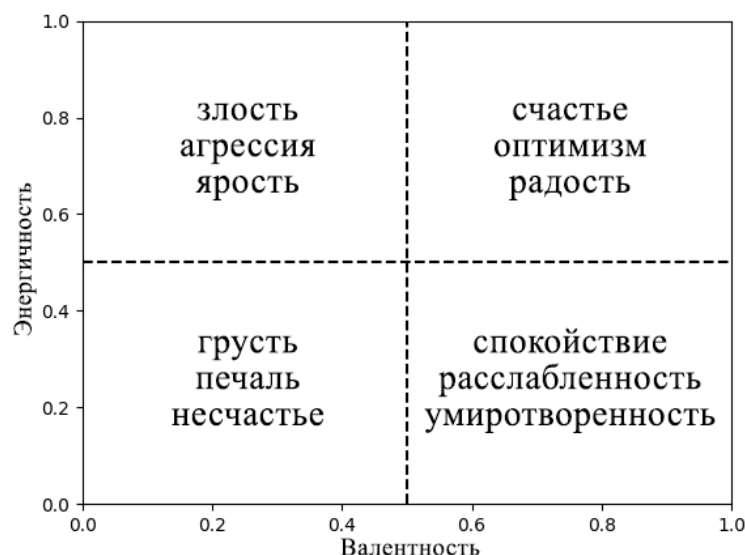


Рисунок 1 – Модель Расселла

1.1.2 Прочие характеристики

Исследования в области музыкальной теории и психологии, проводившиеся с помощью социального опроса, показали, что:

- Высокая громкость и темп ассоциируются с радостью или злостью, а низкая громкость и темп – со спокойствием или печалью.
- Музыку в мажорном ладу чаще ассоциируют с ощущением счастья, а в минорном – с грустью.

1.2 Создание датасета

Для создания датасета использовался поиск по плейлистам сервиса Spotify, названия которых содержали бы ключевые слова для настроений. Для каждого термина искали 10 самых популярных плейлистов на сервисе с условием, что каждый из найденных плейлистов имеет не менее 1000 подписчиков. Предполагается, что если плейлист называется, к примеру, «Грустные песни» и имеет 1000 подписчиков, то хотя бы часть этих подписчиков согласна с тем, что музыку, содержащуюся в этом плейлисте действительно можно охарактеризовать словом «грустная».

Для каждой категории настроения были взяты 14 англоязычных терминов, описывающих настроение. Среди них термины из работы, где авторы выполняли схожую задачу по разметке музыки по настроению, а также синонимы из Кембриджского словаря.

Для всех ключевых слов проверялась релевантность поискового запроса,

термин считался не подходящим, если выполнялось хотя бы одно из следующих условий:

- Термин не явно описывает настроение (например, по термину находятся плейлисты, связанные с конкретным исполнителем, саундтреки к фильму и т.п.)
- Термин редко использовался пользователями для описания настроения музыки (из найденных плейлистов можно было получить менее 500 песен)
- При построении точечного графика в координатах валентности и энергичности точки слабо соответствовали предполагаемой четверти.

Для каждой из четвертей были выбраны 2 наиболее популярных (выдававших плейлисты с наибольшим числом подписчиков) и релевантных термина. В случае с IV четвертью, ни один из терминов не соответствовал ей при построении точечного графика, поэтому в виде исключения были выбраны самые популярные термины, без учета четверти. Категории настроений были названы следующим образом: I - счастливое (happy, upbeat), II - агрессивное (aggressive, rage), III - грустное (sad, melancholy), IV - спокойное (calm, relaxing). Название категории присваивается по самому популярному термину из двух.

Сервис предоставляет следующие характеристики для музыкальных композиций: валентность, энергичность, акустичность, танцевальность, продолжительность, инструментальность, высота звука, живость, громкость, лад, обильность речи, темп, тактовый размер.

Всего датасет содержит 4483 песен, из них 1149 – счастливые, 1139 – агрессивные, 1105 – грустные, 1090 – спокойные. Если песня входила в одну категорию несколько раз, то дубликаты удалялись из датасета, а если песня входила в разные категории, то удалялась из датасета целиком. Также были удалены 67 объектов, где значения тактового размера не соответствовали допустимому диапазону.

1.2.1 Анализ данных

В результате анализа данных были обнаружены выбросы по признакам обильности речи, громкости, живости, темпа и продолжительности песни. Выбросы удаляться не будут, поскольку их нельзя интерпретировать, как ошибки в данных. Между валентностью и энергичностью была обнаружена средняя положительная корреляция, что противоречит модели Расселла. Построение точечного графика в координатах валентности и энергичности показало, что грустные

и спокойные песни оказывались в одной координатной четверти, что позволило предположить недостаточность этих двух характеристик для оценки настроения музыки. Сравнение медиан для всех характеристик позволило предположить, что признаки лада, высоты звука, и тактового размера могут создавать шум в данных, т.к. совпадают для всех категорий настроения. После этого важность признаков была оценена с помощью функции `mutual_info_classif` библиотеки Scikit-Learn. Оценка показала, что наименьшую важность имели признаки лада и высоты звука, после чего они были объединены в новый признак – тональность, который использовался в дальнейшем при обучении моделей машинного обучения. Анализ данных показал, что при предсказании музыкального настроения, следует использовать все имеющиеся признаки, объединив при этом лад и высоту звука в тональность.

2 Сравнение моделей для предсказания вероятностей

В рамках работы были обучены и сравнены 5 моделей машинного обучения, а именно классификаторы: решающее дерево, случайный лес и 3 реализации градиентного бустинга: XGBoost, LightGBM, CatBoost. Гиперпараметры моделей подбирались с использованием алгоритма TPE (Tree-structured Parzen Estimators), реализованного в библиотеке Optuna. К моделям применялись методы калибровки вероятностных прогнозов: сигмоидная (калибровка Платта) и изотоническая, реализованные в библиотеке Scikit-Learn, с использованием кросс-валидации с 10 фолдами. В качестве метрик оценки качества вероятностных прогнозов использовались LogLoss, оценка Бриера (Brier Score), AUC-ROC, ожидаемая ошибка калибровки (ECE, Expected Calibration Error). Наилучшие результаты были получены классификатором CatBoost после изотонической калибровки со значениями метрик: $\text{LogLoss} = 0.485865$; $\text{AUC-ROC} = 0.954951$; $\text{Brier Score} = 0.13451$; $\text{ECE} = 0.022695$.

3 Структура приложения

Список использованных технологий при создании веб-приложения:

- Flask для серверной части;
- SQLAlchemy, Alembic, PostgreSQL для базы данных;
- HTML, CSS, JavaScript для клиентской части;
- Flask-JWT-Extended для JWT-авторизации;
- werkzeug для хэширования паролей;
- RQ и Redis для фоновых задач.

3.1 Определение музыкальных жанров

Для определения музыкальных жанров песен предлагается использовать теги сервиса Last.fm. Пользователи сервиса могут присваивать теги исполнителям, альбомам и отдельным композициям. Используя публичное API сервиса предлагается собирать наиболее популярные теги, присвоенные композициям, но, поскольку теги не обязательно являются жанрами, необходим словарь для фильтрации подходящих тегов.

В качестве словаря будет использовано дерево жанров, предложенное разработчиками плагина LastGenre приложения beets для управления музыкальной библиотекой. Всего в дереве представлено 766 жанров, которые были собраны с помощью анализа статей Википедии о музыкальных жанрах. Так же для определения жанров будет использоваться белый список жанров, т.е. такие наименования, которые могут быть присвоены песням. Песням будут присваиваться наиболее специфичные жанры из белого списка, т.е. те, которые в дереве жанров находятся на наиболее глубоких уровнях. Белый список состоит из 21 наименования и был получен из первого уровня дерева жанров, а также из статистики о самых популярных музыкальных жанрах в музыкальном чарте Billboard 1958-2016 гг.

3.2 Описание пользовательского интерфейса

Первым шагом пользователю предлагается выбрать настроение с помощью двумерного слайдера (см. рисунок 2а), где координаты x, y принимают значения $[0; 1]$, а категории настроений, принимающие вероятность 1, имеют следующие координаты: счастливое $(1; 1)$, агрессивное $(0; 1)$, грустное $(0; 0)$, спокойное $(1; 0)$. Из считанных координат вероятности вычисляются следующим образом:

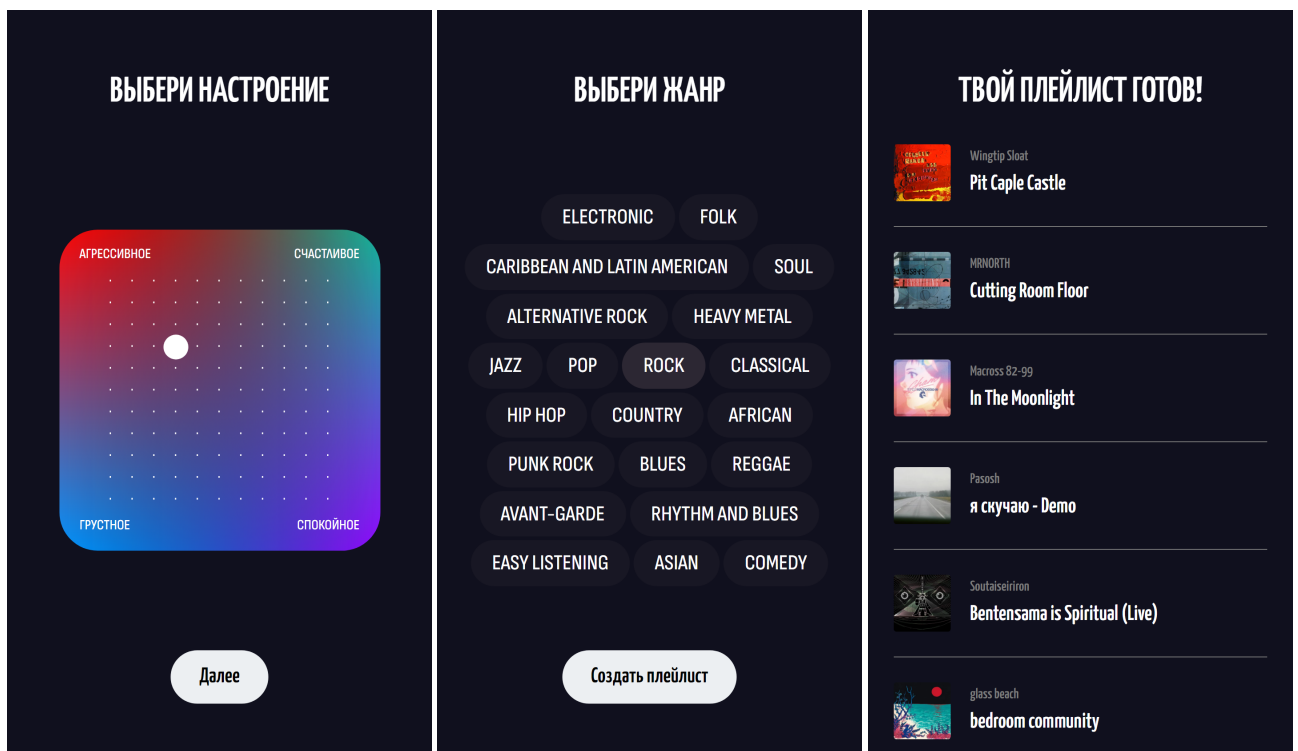
$$p_{happy} = x * y;$$

$$p_{aggressive} = (1 - x) * y;$$

$$p_{sad} = (1 - x) * (1 - y);$$

$$p_{calm} = x * (1 - y)$$

Затем пользователю предлагается выбрать единственный желаемый музыкальный жанр (см. рисунок 2b), после чего совершается запрос к базе данных для поиска песен в соответствии с вычисленными вероятностями настроений и выбранным жанром. Фильтрация выполняется по диапазону, где вероятности настроений находятся в пределах заданного порога $threshold=0.05$. Из найденных песен случайным образом выбирается 20, однако, если количество найденных песен меньше 20, то $threshold$ увеличивается на 0.01 до тех пор, пока не будет найдено 20 песен, или $threshold$ не достигнет значения 1. Найденные песни рендерятся на странице (см. рисунок 2c), пользователь может увидеть имена исполнителей, названия треков и обложки альбомов.



(a) Экран выбора настроения

(b) Экран выбора жанра

(c) Рекомендуемый плейлист

Рисунок 2 – Внешний вид пользовательского интерфейса

3.3 Описание интерфейса администратора

Для добавления администраторов в базу данных реализована CLI-команда, регистрация администратора из интерфейса приложения не предусмотрена. При добавлении администратора в БД, пароль хэшируется по стандарту PBKDF2 с помощью функции `generate_password_hash()` библиотеки `werkzeug`. Для безопасной аутентификации в приложении используются JWT (JSON Web Token) и CSRF (Cross-Site Request Forgery) токены, это позволяет обеспечить защиту от XSS (Cross-Site Scripting, межсайтовый скриптинг) и CSRF (Cross-Site Request Forgery, межсайтовая подделка запроса) атак.

После успешной авторизации администратор попадает на страницу, которую можно увидеть на рисунке 3.

Выберите или перетащите CSV файл

Выберите файл Файл не выбран

Определять вероятности настроений

Определять жанры

Искать обложки альбомов

Загрузить

Статус фоновых задач

Redis job ID to delete Удалить задачу

Количество песен в БД по жанрам

Для добавления треков в БД загрузите CSV файл со следующими столбцами: Строки с пустыми значениями будут проигнорированы.

Имя столбца	Пример значения	Описание
name	Why (feat. Terry Ilous)	Название трека
artists	['Morano', 'Terry Ilous']	Список исполнителей
album	Incognito	Название альбома
aggressive	0.0373189514496802	Вероятность принадлежности к настроению "агрессивное"
calm	0.0191142229735873	Вероятность принадлежности к настроению "спокойное"
happy	0.853636134655076	Вероятность принадлежности к настроению "счастливое"
sad	0.0899306909216564	Вероятность принадлежности к настроению "грустное"
genre	rock	Жанр
album_cover	https://example.com	Ссылка на изображение с обложкой альбома (может быть пустым)

Рисунок 3 – Страница администратора

Первая функция, доступная администратору – это пополнение базы данных песнями из csv файла. Справа от формы загрузки расположена таблица с подсказкой о том, какие столбцы должен содержать файл, она изменяется в

зависимости от того, нужно ли для песен в csv файле определять вероятности настроек, жанры и обложки альбомов (устанавливается с помощью флажков). При успешной загрузке файла, создается фоновая задача и добавляется в очередь, для организации очереди задач используется библиотека RQ (Redis Queue), которая использует Redis в качестве брокера и бэкенда.

Так же администратору доступны функции: узнать статусы всех фоновых задач, удалить фоновую задачу из очереди по ее ID, просмотреть количество песен в БД по жанрам.

ЗАКЛЮЧЕНИЕ

В результате работы были собраны и проанализированы данные для обучения алгоритмов машинного обучения в рамках задачи предсказания вероятностей принадлежности песен к четырем категориям настроения, определенным в соответствии модели Расселла. Были рассмотрены несколько классификаторов, а также метрики оценки их качества и способы калибровки вероятностных прогнозов. Классификатор CatBoost после изотонической калибровки показал наилучшие результаты и в дальнейшем использовался при создании базы данных. Был разработан модуль для сбора информации о жанрах и обложках альбомов с сервиса Last.fm, он также использовался при создании базы данных. Было разработано веб-приложение с использованием Flask для серверной части и JavaScript для клиентской части, которое позволяет пользователю получить плейлист из 20 песен в соответствии выбранному настроению и жанру, а администратору доступен функционал по пополнению базы данных из csv таблиц.