\_\_\_\_\_\_\_\_\_\_\_Національний університет ”Львівська політехніка”\_\_\_\_\_\_\_\_\_\_\_

Інститут комп’ютерних наук та інформаційних технологій

Кафедра\_\_\_\_ Системи штучного інтелекту\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Спеціальність\_\_\_\_\_\_122 «Комп’ютерні науки»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 ПОЯСНЮВАЛЬНА ЗАПИСКА

 до бакалаврської кваліфікаційної роботи на тему:

**Інформаційна система розпізнавання музичного жанру аудіо файлу**

студента групи КН-418 Влощинського В.О.

Керівник проекту \_\_\_\_\_\_\_\_\_\_\_(Шаховська Н.Б.)

Консультанти

(\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Завідувач кафедри     \_\_\_Шаховська Н.Б.\_\_\_\_\_       \_\_\_\_\_\_\_\_\_\_\_

”\_\_\_”\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2020 р.

Національний університет “Львівська політехніка”

**Інститут комп’ютерних наук та інформаційних технологій**

**Кафедра “Систем штучного інтелекту”**

**Спеціальність 122 “Комп’ютерні науки”**

**ЗАТВЕРДЖУЮ:**

Завідувач кафедри СШІ \_\_\_\_\_\_\_\_\_\_\_\_\_

“\_\_\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_р.

ЗАВДАННЯ

НА БАКАЛАВРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТОВІ

Влощинському Віталію Олександровичу

1. Тема роботи: «Інформаційна система розпізнавання музичного жанру аудіо файлу» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 затверджена наказом університету від «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  2020 р. № \_\_\_\_\_\_\_

2. Термін подання закінченої роботи: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Вихідні дані до (проекту) роботи: Набір аудіо файлів з маркуванням жанрів для тренування нейронної мережі.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Зміст розрахунково-пояснювальної записки: (перелік питань, що належить розробити): аналітичний огляд літературних та інших джерел, системний аналіз та обґрунтування проблеми, методи та засоби вирішення проблеми, практична реалізація.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень): діаграма потоків даних, візуалізація архітектури нейронної мережі.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Консультанти по роботі, із зазначенням розділів проекту, що стосується їх

|  |  |  |  |
| --- | --- | --- | --- |
| Розділ | Консультант | Підпис, дата | |
| Завдання видав | Завдання прийняв |
| Економічна частина |  |  |  |
| Консультант з іноземної мови |  |  |  |

7. Дата видачі завдання   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Керівник\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис)

Завдання прийняв до виконання\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис)

КАЛЕНДАРНИЙ ПЛАН

|  |  |  |  |
| --- | --- | --- | --- |
| №  п/п | Назва етапів дипломної роботи | Термін виконання етапів роботи | Примітка |
| 1 | Аналіз літературних джерел | 10.02.2020 – 1.03.2020 |  |
| 2 | Постановка задачі дослідження | 02.03.2020 – 15.03.2020 |  |
| 3 | Розроблення методу аналізу відгуків користувачів | 16.03.2020 – 05.04.2020 |  |
| 4 | Реалізація методу та його апробація | 06.04.2020 – 16.04.2020 |  |
| 5 | Оформлення пояснювальної записки | 17.04.2020 – 13.05.2020 |  |

Студент\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Керівник роботи\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ЗМІСТ

ВСТУП…………………………………………………………….……………

1 ЗАГАЛЬНІ ПОЛОЖЕННЯ……………………………………..………….

* 1. Опис предметного середовища……………………………….………….
  2. Огляд наявних аналогів…………………………………………………..
  3. Постановка задачі…………………………………………………………

АНОТАЦІЯ……………………………………………………….…………….

ВИСНОВКИ ДО РОЗДІЛУ 1………………………………………………….

2 ІНФОРМАЦІЙНЕ ТА МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ…….……….

2.1 Аналіз предметної області……………………………………….………..

2.1.1 Вхідні дані……………………………………………………….……….

2.1.2 Вихідні дані……………………………………………………………….

2.2 Проектування системи………………………………………………..……

2.3 Математичне та алгоритмічне забезпечення………………………..……

ВИСНОВКИ ДО РОЗДІЛУ 2………………………………………….………..

3 ПРОГРАМНЕ ТА ТЕХНІЧНЕ ЗАБЕЗПЕЧЕННЯ…………………..………

3.1 Засоби розробки…………………………………………………….……….

3.2 Вимоги до технічного та програмного забезпечення…………………….

3.3 Опис програмної реалізації………………………………………….……..

3.4 Керівництво користувача……………………………………………..……

ВИСНОВКИ ДО РОЗДІЛУ 3…………………………………………….…….

4 ОЦІНЮВАННЯ РІВНЯ ГОТОВНОСТІ НАУКОВО-ТЕХНІЧНОЇ РОЗРОБКИ ДО КОМЕРЦІАЛІЗАЦІЇ……………………………….……………

ЗАГАЛЬНІ ВИСНОВКИ…………………………………………………….……

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ………………………………………

Додаток А ………………………………………………………………….……

ВСТУП

У наш час цифрових технологій музика стала надзвичайно популярною, а музична індустрія тепер одна із найприбутковіших сфер світу. Враховуючи просто неймовірну кількість створених композицій, використовувати людські ресурси стає просто неефективно. Усі аудіо сервіси для завдань, що пов’язані з аудіо файлами використовують штучний інтелект. В тому числі надзвичайно актуальним стає завдання автоматичного визначення музичного жанру.

Мета дослідження – розробка програмного продукту, що дозволяє швидко визначити музичний жанр аудіо файлу.

Задачі дослідження:

* Аналіз предметної області
* Проектування інформаційної системи
* Вибір інструментарію, методів реалізації програмного продукту

Об’єкт дослідження – це інформаційна система автоматичного визначення музичного жанру аудіо файлу.

Предмет дослідження – способи визначення жанру аудіо файлу за допомогою систем штучного інтелекту.

Отримані результати допоможуть зрозуміти як системи штучного інтелекту сприймають звук та можуть використовуватись для покращення класифікації аудіо файлів у інших завданнях цієї сфери, зокрема: організація особистих колекцій, множинні зв’язки в музичних каталогах, системи рекомендацій, формування списків відтворення, розпізнавання композицій за уривками.

РОЗДІЛ 1

ЗАГАЛЬНІ ПОЛОЖЕННЯ

1.1 Опис предметного середовища

Музична індустрія наразі є однією з найпоширеніших та найприбутковіших сфер світу. Незліченна кількість гуртів, композицій, лейблів, студій існує наразі у цій сфері. В останні роки ця сфера почала набувати шаленої популярності. Для прикладу, навіть найпопулярніші відео на відеохостингу YouTube – це лише відеоряди до найпопулярніших композицій. Вслід за популярністю цієї сфери, масово почали з’являтись найрізноманітніші сервіси для обробки, прослуховування, визначення жанру композицій. Враховуючи просто неймовірну кількість композицій, які наразі існують, а також те, що усім цим аудіо сервісам потрібно ці композиції зберігати та обробляти, використовувати людські ресурси стає дуже неефективно та невигідно. Тут в гру вступають системи штучного інтелекту. Усі найпопулярніші сервіси для більшості завдань, що пов’язані з аудіо файлами використовують штучний інтелект. В тому числі надзвичайно актуальним стає завдання автоматизованого визначення музичного жанру аудіо файлу без використання людських ресурсів. Що допоможе зекономити колосальну кількість фінансів враховуючи неймовірну кількість композицій, що з’являються. Визначення жанру може використовуватись як допоміжний елемент в будь якому аудіо сервісі. Це дозволяє користувачу відсортувати всі свої композиції за музичним жанром. Також на основі жанрів можуть базуватись системи рекомендацій нових композицій на основі вже прослуханих. Інформаційна система визначення музичного жанру аудіо файлу стає дедалі важливішою в наш час, коли на заміну ручної роботи людей приходять автоматизовані системи штучного інтелекту.

Розпізнавання музичного жанру – це підгалузь більшої галузі класифікації аудіо сигналів, що може бути визначена, як отримання релевантних ознак з аудіо і використання цих ознак для того, щоб визначити до якої множини класів відноситься цей аудіо сигнал. Так само, розпізнавання музичного жанру – це отримання інформації з композиції та класифікація її музичного жанру на основі отриманої інформації. Термін «жанр» прийшов до нас від латинського слова «genus», що означає тип або клас. Жанр можна описати як тип чи категорію, визначену певними критеріями. Варто відмітити, що в даному конкретному випадку музичних жанрів, ці критерії не є чітко визначеними а є суб’єктивними та відносними. Класифікація жанру завжди є суб’єктивною відносно конкретного користувача та соціально культурного середовища в якому він знаходиться. В рамках цього обговорення, музичний жанр можна розглядати як певний клас або тип музики з множиною спільних властивостей та ознак, що на слух середньостатистичного користувача відрізняють композиції цього жанру від інших музичних композицій. Кожен конкретний жанр характеризується використаними інструментами, темпом, ритмом, але існує ще чимало інших факторів, які впливають на визначення музичного жанру людиною. Одна з найбільших проблем автоматичного розпізнавання музичного жанру – визначення, які ознаки найбільше впливають на музичний жанр. Обробка цифрових композицій за допомогою комп’ютера в наш час не є проблемою. Композиція в цифровому форматі – це просто певна послідовність біт, які можуть бути легко оброблені машиною, щоб витягнути з них певні дані, які знадобляться для подальшої класифікації композиції. Нижче наведено список завдань, які нам дозволяє вирішити автоматичне розпізнавання жанру композиції:

* Організація особистих колекцій: завантаження композицій з інтернету вже давно стало абсолютно звичайним завданням для більшості меломанів та просто любителів музики. Автоматичне розпізнавання жанру є чудовим фундаментом для автоматичного впорядкування своїх композицій. Класифікатор має дозволяти користувачам впорядковувати композиції відповідно до їхнього жанру і має дозволяти їм додавати нові жанри.
* Множинні зв’язки в музичних каталогах: дуже часто композиція не може бути віднесена повністю лише до одного унікального музичного жанру. Певна оцінка того, що конкретна композиція відноситься до конкретного жанру може бути використана, щоб допомогти користувачам у пошуку композицій. Каталоги можуть формуватися використовуючи цю оцінку певним чином і відповідно одна і та сама композиція може бути знайдена у різних групах.
* Система рекомендацій: в зв’язку з доступністю надзвичайної кількості аудіо файлів в інтернеті, система рекомендацій музики нещодавно набула неймовірної популярності та реалізована практично кожним онлайн сервісом певним чином. Системи рекомендацій в основному базуються на типових техніках обробки аудіо сигналів, в яких розпізнавання жанру грає далеко не останню роль.
* Формування списків відтворення: Дуже часто ми хочемо послухати лише певний тип музики, що залежить від нашого настрою, смаків та просто конкретної ситуації. Система, завдяки прикладам музики, що слухає користувач, має бути спроможна запропонувати йому список відтворення зі схожими композиціями. Знову ж таки, певна оцінка того, що конкретна композиція відноситься до конкретного жанру це дуже цінна інформація в рамках цього завдання.
* Розпізнавання композицій: розпізнавання композицій є дуже важливим завданням не тільки для користувачів, які бажають за маленьким уривком композиції дізнатись всю інформацію про неї, а і для авторів цих композицій для слідкуванням за дотриманням авторських прав. Яскравим прикладом використанням може бути відеохостинг YouTube, який при використанні захищених авторським правом композицій у своєму відео може вимкнути монетизацію та перенаправити увесь прибуток від переглядів відео до автора даної композиції або за вимогою автора видалити ролик повністю з відеохостингу.

1.2 Огляд наявних аналогів

В якості аналогів мною були відібрані найпопулярніші сервіси для прослуховування, розпізнавання, обробки аудіо. Кожен з них в тій чи іншій мірі реалізує завдання визначення жанру композиції за допомогою штучного інтелекту. Але варто відмітити, що ні для одного сервісу, це завдання не є основним, а є лише частиною функціональності цих онлайн гігантів музичної індустрії. Наприклад, є частиною дуже важливої системи рекомендацій композицій чи частиною системи розпізнавання. Серед найпопулярніших аналогів, варто відмітити:

* YouTube Music – це онлайн сервіс для прослуховування аудіо. З’явився відносно недавно, але відразу завоював популярність аудиторії за рахунок свого попередника YouTube. Наявний досить широкий вибір композицій, але він сильно поступається в цьому плані своїм конкурентам. Також деякі з композицій є лише звуковою доріжкою з відео на YouTube і мають посередню якість. Можливості сервісу в безкоштовній версії сильно обмежені. Обліковий запис з розширеною платною підпискою дає можливість для відтворення композицій без реклами між ними, відтворення композицій у фоновому режимі, що актуально для мобільних пристроїв, оскільки при блокуванні мобільного пристрою композиція вимикається в безкоштовній версії та завантажування композицій на цільовий пристрій і подальше їх прослуховування без використання будь-якого виду інтернет з’єднання. Містить функціональність системи рекомендацій та допомагає користувачу підібрати схожі композиції на основі вже прослуханих. Використовує розпізнавання жанру композицій за допомогою штучного інтелекту, щоб покращити систему рекомендацій. Тобто, розпізнавання жанру є лише частиною системи рекомендацій та не є критичним та основним функціоналом. Частково виконує те ж саме завдання, що і мій сервіс.
* Spotify – найпопулярніший онлайн сервіс для прослуховування аудіо. Перший сервіс, що запровадив можливість онлайн прослуховування (тобто прослуховування аудіо файлів без їхнього завантаження на цільовий пристрій). Найбільший вибір композицій серед усіх сервісів для прослуховування. Можливості сервісу в безкоштовній версії також обмежені. Аналогічно, обліковий запис з розширеною платною підпискою дає можливість для відтворення композицій без реклами між ними, відтворення композицій у фоновому режимі та завантажування композицій на цільовий пристрій і подальше їх прослуховування без використання будь-якого виду інтернет з’єднання. Аналогічно містить функціональність системи рекомендацій та допомагає користувачу підібрати схожі композиції на основі вже прослуханих. Також використовує розпізнавання жанру композицій за допомогою штучного інтелекту, щоб покращити систему рекомендацій. Є можливість перегляду всіх композицій певного жанру.
* Shazam – є найпопулярнішим сервісом для розпізнавання композицій. Який, використовуючи мікрофон, за допомогою короткого уривку композиції може здійснити пошук у своїй базі даних та знайти інформацію про цю композицію. Використовує розпізнавання жанру композицій за допомогою штучного інтелекту, щоб покращити систему розпізнавання композиції на основі її уривку. Оскільки, розпізнавання жанру на основі нейронних мереж, однаково добре працює як з цілими композиціями (3хв і більше), так і з їхніми уривками. Тобто, розпізнавання жанру є лише частиною системи розпізнавання та не є критичним та основним функціоналом. Частково виконує те ж саме завдання, що і мій сервіс.
* Sound Cloud – є другим по популярності онлайн сервісом для прослуховування аудіо. Досить широкий вибір композицій. Можливості сервісу в безкоштовній версії також обмежені. Аналогічно, обліковий запис з розширеною платною підпискою дає можливість для відтворення композицій без реклами між ними, відтворення композицій у фоновому режимі та завантажування композицій на цільовий пристрій і подальше їх прослуховування без використання будь-якого виду інтернет з’єднання. Розпізнавання композицій у цьому сервісі виділено як окреме завдання, та він є найбільше наближений до мого сервісу. Кожна композиція при завантаженні та обробці за допомогою штучного інтелекту в автоматичному режимі отримує свій музичний жанр. Який потім можна переглянути. Використовуючи фіксований список жанрів можна проводити пошук схожих по жанру композицій. Також використовується для сортування, фільтрування композицій при пошуку композиції по її назві або по гурту, що її виконує.
* Apple Music – це сервіс для прослуховування композицій від компанії Apple. В порівнянні з іншими сервісами має дуже широкий вибір композицій. Сервіс не має безкоштовного плану та є повністю платним. Функціональність даного сервісу суттєво не відрізняється від поданих вище. Аналогічно містить функціональність системи рекомендацій та допомагає користувачу підібрати схожі композиції, гурти та списки відтворення створені іншими користувачами, функціональність підбору композицій в залежності від настрою, онлайн та офлайн прослуховування, інтеграцію з усіма пристроями Apple. Використовує штучний інтелект для системи рекомендацій, користувачу, а також для групування композицій за їхнім настроєм та жанром.

1.3 Постановка задачі

З появою та стрімким ростом популярності в останні роки інтернет сервісів для потокового прослуховування музики стало неможливим вручну проставляти жанри для композицій. Саме тому неймовірно популярним стало завдання з автоматичного визначення музичного жанру музики.

Як було зазначено вище, розпізнавання музичного жанру – це групування композицій на деякі попередньо задані категорії. Головна ціль – розробити інформаційну систему з поданими надалі властивостями. Вхідні дані це композиція у вигляді цифрового аудіо сигналу. Вихідні дані це інформація про вхідний сигнал, яка описує до яких попередньо заданих музичних жанрів відноситься завантажена користувачем композиція. Коли ми працюємо над складною проблемою, часто бажано побачити як ми можемо розбити її на маленькі незалежні частини. Розпізнавання музичного жанру це завдання класифікації, яке складається з двох базових етапів, які мають бути виконані:

* Виділення ознак
* Класифікація жанру

Головна ціль першого етапу виділення ознак – це отримати базову інформацію з вхідних даних. Основна проблема на першому етапі – це дізнатися які ознаки аудіо сигналу найбільше впливають на те, до якого жанру ця композиція буде віднесена. Детальніше про виділення ознак з аудіо сигналу ми поговоримо в наступних розділах. Головна ціль другого етапу – це визначення, які комбінації певних значень ознак, відповідають яким категоріям. По суті, вихідні дані першого етапу, є вхідними даними для другого етапу класифікації. Розпізнавання музичного жанру є доволі складним завданням для комп’ютера. Є кілька основних причин чому так сталося.

По перше, музика в загальному дуже складна. Колосальна кількість роботи в галузі класифікації аудіо сигналів була пророблена за останні роки і варто відмітити, що успіхи справді вражають. Але, нажаль більшість цієї роботи зосереджено на простих сигналах – звуках, які ми ніколи не зустрінемо в реальному світі і в композиціях в тому числі. Лише останнім часом дослідження звуків, що трапляються в реальному світі (розпізнавання мови, розпізнавання жанру) стали набувати популярності, але дана сфера все ще знаходиться в зародковому стані.

По друге музика це феномен! Хоча це просто коливання повітря, що уловлюються нашими органами чуття і якимось чином інтерпретуються нашим мозком, але як це відбувається насправді, все ще є загадкою для людства. Саме це і не дозволяє нам створити комп’ютерну систему, яка б ідеально імітувала сприйняття звуку людським вухом.

Також, завдання розпізнавання музичних жанрів, навіть людиною є трішки глибшим та складнішим, ніж здається на перший погляд. Розпізнавання музичного жанру, часто здається людям простим завданням, але це не означає, що отримані результати будуть завжди правильними та об’єктивними. Для прикладу, я знайшов однакову композицію (Three Days Grace – Never Too Late) на кількох сайтах онлайн прослуховування музики. На мою особисту думку вона відноситься до жанру – industrial metal. В свою чергу, на сайтах, що спеціалізуються на музиці та на яких жанр відповідно проставляється людьми ця композиція була позначена як: alternative rock, alternative metal. На цьому прикладі, ми можемо чітко побачити, що одна і та ж сама композиція була віднесена до різних жанрів і під жанрів різними людьми. З цього відповідно випливає те, що навіть люди не можуть прийти до спільної думки в області визначення жанрів і тому створити систему, яка б ідеально визначала жанри просто неможливо. При оцінці точності класифікації жанру, варто також враховувати цю особливість, тобто помилки, які також може зробити людина в цій ситуації, наприклад, класифікуючи метал композиції як рок або класифікуючи повільні частини рок симфоній як класичну музику.

АНОТАЦІЯ

Бакалаврська кваліфікаційна робота виконана студентом групи КН-418 Влощинським Віталієм Олександровичем. Тема «Інформаційна система розпізнавання музичного жанру аудіо файлу». Робота направлена на здобуття ступеня бакалавр за спеціальністю 122 «Комп’ютерні науки».

Метою дипломної роботи є створення та розробка програмного продукту, що автоматизує процес розпізнавання музичного жанру аудіо файлу. Об’єктом дослідження є інформаційна система автоматичного визначення музичного жанру аудіо файлу. В результаті виконання дипломної роботи було розроблено програмний продукт, який мінімізує затрати компанії на людські ресурси для визначення музичного жанру аудіо файлу.

Ключові слова: інформаційна система, визначення музичного жанру, програмний продукт, перетворення Фур’є, розпізнавання аудіо, нейронні мережі.

ABSTRACT

Degree work executed by the student of group CS-418 Vloshchynskyi Vitalii Oleksandrovych. The topic is “Information system for recognition of musical genre of an audio file”. Work is directed on reception obtain a bachelor’s degree on a specialty 122 “Computer Science”.

The purpose of the degree work is to create and develop a software product that automates the process of recognition of musical genre of an audio file. The object of research is an information system for recognition of musical genre of an audio file. As a result of the thesis, a software product was developed that minimizes the company’s expenses for human resources to recognize music genre of an audio file.

Keywords: information system, music genre recognition, software product, Fourier transform, audio recognition, neural networks.

ВИСНОВКИ ДО РОЗДІЛУ

У першому розділі даної дипломної роботи було в загальному досліджено музичну сферу в наш час та конкретно область розпізнавання та обробки аудіо файлів за допомогою систем штучного інтелекту, проведено порівняння сервісів аналогів для роботи з аудіо. На основі проведеного дослідження ми можемо зробити такі висновки:

* Абсолютна більшість популярних систем, для завдань, що пов’язані з аудіо файлами використовує штучний інтелект. Кожен з аналогів даної інформаційної системи в тій чи іншій мірі використовує визначення жанру композиції за допомогою штучного інтелекту. Але варто відмітити, що ні для якого сервісу, це завдання не є основним, а є лише частиною іншої більш складної функціональності.
* Визначення жанру можна використовувати для більш складних завдань: організації особистих колекцій, множинні зв’язки в музичних каталогах, системи рекомендацій, формування списків відтворення, розпізнавання композицій за уривками.
* Наразі сфера розпізнавання аудіо зосереджується на простих сигналах та звуках, які досить рідко зустрічаються в реальному житті і в композиціях.
* В наш час все ще є загадкою для людства точний механізм сприйняття звуку органами чуття і подальша його обробка мозком, що і не дозволяє створити дійсно точну імітацію сприйняття звуку людиною.
* Кожна людина сприймає звук по різному.
* Одна і та ж сама композиція може бути віднесена до різних жанрів різними людьми. Оскільки, жанр – це суб’єктивне поняття на яке впливають багато чинників.

РОЗДІЛ 2

ІНФОРМАЦІЙНЕ ТА МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ

2.1 Аналіз предметної області

2.1.1 Вхідні дані

Як було зазначено вище, розпізнавання музичного жанру – це групування композицій на деякі попередньо задані категорії за певними ознаками. Вхідними даними для даної інформаційної системи є композиція у вигляді цифрового аудіо сигналу. Композиція повинна належати до одного з 10 жанрів, що будуть перераховані в наступному розділі та визначення яких забезпечує дана інформаційна система, інакше автоматичне визначення жанру може згенерувати некоректний результат або просто вказати, що неможливо розпізнати дану композицію. Тривалість композиції не має перевищувати 5 хв, оскільки більшість композицій, що ми можемо знайти на просторах світової павутини мають довжину не більше 4 хв і наразі наявне апаратне забезпечення не дозволяє нам досягнути оптимальної швидкодії при більшій тривалості композиції. Це було зроблено для збільшення швидкодії даної інформаційної системи для окремо взятої композиції та для можливості опрацьовування більшої кількості композиції одночасно за рахунок економії апаратних ресурсів, оскільки опрацювання композиції потребує великої кількості оперативної пам’яті та ресурсів процесора або відеокарти через використання нейронних мереж. Композиція має бути завантажена в одному з перерахованих нижче цифрових форматів зберігання аудіо інформації:

* .mp3
* .m4a
* .aac
* .oga
* .flac
* .wav
* .pcm
* .aiff
* .wma

2.1.2 Вихідні дані

Вихідними даними даної інформаційної системи автоматичного визначення музичного жанру певного аудіо файлу – є інформація про поданий вхідний аудіо сигнал, яка показує до яких музичних жанрів відноситься завантажена користувачем музична композиція. Як результат роботи інформаційної системи користувач може побачити 11 варіантів вихідних даних, а точніше це є 10 жанрів і повідомлення, про те що жанр визначити не вдалось, у разі певних технічних проблем на сервері, певних проблем з аудіо файлом яких не було помічено під час попередньої перевірки при його завантаженні або просто якщо завантажена композиція не належить до списку жанрів, які може розпізнавати дана інформаційна система. Часто виникають випадки коли певну композицію не може віднести повністю лише до одного унікального музичного жанру навіть людина, яка напряму задіяна в музичній сфері, оскільки композиція була виконана з поєднанням кількох унікальних музичних жанрів. В такі моменти визначення музичних жанрів стає досить складним завданням як для людини, так і для інформаційної системи. Саме по цій причині кожна завантажена композиція в результаті роботи інформаційної системи може відноситися до трьох жанрів одночасно. Вихідні дані передаються сервером у форматі JSON та відображаються на сторінці як звичайний текст у окремому спливаючому вікні, що було зроблено для простоти розуміння результатів роботи інформаційної системи користувачем. Вихідні дані є звичайним текстом, що представляє собою список жанрів, що відносяться до завантаженої користувачем композиції які є відокремлені комою або є текстом, що містить повідомлення про помилку. Черговим варіантом вихідних даних можна вважати повідомлення з проханням до користувача завантажити іншу музичну композицію у разі, якщо завантажена ним раніше композиція не пройшла попередню перевірку на тривалість, формат або цілісність та не може бути допущена до етапу визначення жанру.

2.2 Проектування системи

Для побудови функціональної моделі даної інформаційної системи та візуалізації її роботи було вирішено використати діаграми потоків даних за методологією Йордана. Контекстну діаграму ми можемо побачити на рис. 2.1.



Рис. 2.1 Контекстна діаграма

Виділимо основні зовнішні сутності даної предметної області:

Зовнішня сутність користувач здійснює запит на автоматичне визначення жанру, передаючи інформаційній системі аудіо файл, жанр якого потрібно визначити.

Має на вході такі потоки:

* Жанри

Має на виході такі потоки:

* Цифровий аудіо файл

Для того, щоб побачити як працює дана інформаційна система більш детально, нам потрібно виконати декомпозицію першого рівня (рис. 2.2).



Рис. 2.2 Декомпозиція першого рівня

Процес 1 – Перевірити аудіо файл. Приймає аудіо файл від користувача та здійснює різні види перевірок над ним, щоб упевнитись що при подальшій обробці аудіо інформаційна система не дасть збій. Перевірений цифровий аудіо файл передає на наступний етап, якщо всі перевірки були пройдені успішно. Деталізацію процесу 1 можна знайти на рис. 2.3.

Має на вході такі потоки:

* Цифровий аудіо файл

Має на виході такі потоки:

* Перевірений цифровий аудіо файл



Рис. 2.3 Деталізація процесу 1

Процес 2 – Поділити аудіо файл. Відбувається розрізання аудіо файлу на менші фрагменти для подальшої його обробки. Фрагменти передаються на наступний етап. Деталізацію процесу 2 можна знайти на рис. 2.4.

Має на вході такі потоки:

* Перевірений цифровий аудіо файл

Має на виході такі потоки:

* Фрагменти аудіо файлу



Рис. 2.4 Деталізація процесу 2

Процес 3 – Виділити ознаки з фрагментів. Процес виділяє ознаки з фрагментів аудіо файлу та здійснює подальшу підготовку виділених ознак для обробки їх нейронною мережею. Деталізацію процесу 3 можна знайти на рис. 2.5.

Має на вході такі потоки:

* Фрагменти аудіо файлу

Має на виході такі потоки:

* Ознаки фрагментів



Рис. 2.5 Деталізація процесу 3

Процес 4 – Визначити жанр. Використовуючи оброблені дані та ознаки визначає жанр кожного з фрагментів та всієї композиції в цілому після чого передає список жанрів кінцевому користувачу. Деталізацію процесу 4 можна знайти на рис. 2.6.

* Ознаки фрагментів

Має на виході такі потоки:

* Жанри



Рис. 2.6 Деталізація процесу 4

2.3 Математичне та алгоритмічне забезпечення

В загальному алгоритм роботи інформаційної системи автоматичного визначення музичного жанру виглядає так:

* Користувач заходить на веб сторінку інформаційної системи.
* Користувач завантажує композицію.
* Сервер зберігає музичну композицію на диску в тимчасовій папці для більш швидкого доступу до неї при подальшій обробці та економії апаратних ресурсів серверу, зокрема оперативної пам’яті, що дасть можливість одночасно обробляти значно більше композицій. Сама тимчасова папка припинить своє існування відразу після закінчення роботи інформаційної системи для даного користувача.
* Відбувається перевірка чи композиція відповідає вимогам вхідних даних, таким як тривалість та формат аудіо файлу. У разі будь якої невідповідності виводиться повідомлення про помилку, а у іншому разі починається обробка композиції.
* Завантажена та перевірена композиція розрізається на фрагменти тривалістю 10 секунд кожен, для збільшення швидкодії та точності класифікації завантаженої композиції. Наприклад, якщо композиція має тривалість 138 секунд, то відрізків буде 13, а залишкові 8 секунд до уваги не беруться.
* З кожного 10 секундного фрагменту обчислюються 98 спеціальних аудіо ознак за допомогою перетворення Фур’є, які буде детально описано в наступному розділі.
* Використовуючи стандартну нормалізацію перетворюємо обчислені раніше ознаки для кращої роботи нейронної мережі.
* Нейронна мережа на основі нормалізованих раніше ознак визначає жанр для кожного фрагменту завантаженої композиції.
* Останнім етапом – є визначення загального жанру композиції за допомогою методу голосування. Тобто кожен фрагмент віддає свій голос за певний жанр. Після чого всі голоси підсумовуються та жанри, що набирають більше 1/3 голосів вважаються жанрами всього аудіо файлу. Якщо такого жанру не знайшлось, виводиться повідомлення про невдалу спробу визначення жанру.

Одна з найбільших проблем у автоматичному розпізнаванні музичних жанрів це визначення, що саме дозволяє нам відрізняти один музичний жанр від іншого. Проблема в тому, що ми хочемо зробити певні висновки про схожість або несхожість двох об’єктів які напряму порівняти неможливо. Щоб зробити порівняння можливим, спочатку ми повинні перетворити дані, щоб отримати доступ до цільової інформації що в них міститься. Цей процес називають виділення ознак. В нашому випадку це обчислення числових представлень, що описують певний аудіо сигнал.

Аналіз Фур’є – це загальна назва для групи математичних технік, що використовуються для розкладання сигналів на синусоїди та косинусоїди. Він виник під час дослідження особливостей рядів Фур’є. Форма хвилі (графік час-амплітуда) звукового сигналу не розповість нам багато про звук. Найцінніша інформація зашифрована в частоті, фазі і амплітуді спектральних компонентів, з яких складається сигнал. Щоб мати змогу побачити цю інформацію, необхідно обчислити спектр частот сигналу, подібно до процесів, що відбуваються у людському слуховому апараті.

Фур’є був першим хто виявив, що будь-який неперервний періодичний сигнал може бути представлений , як сума певним чином підібраних синусоїдів. Синусоїд – це математична функція, що описує прості гармонічні коливання. Може бути описана формулою:

де y(t) – функція від часу;

A – амплітуда хвилі;

f – частота сигналу;

φ – фаза хвилі.

Всі сигнали, неважливо наскільки складними вони є, можуть бути представлені у вигляді синусоїдів з різною частотою, амплітудою і фазою. Це також справедливо і для звуку, який поширюється за допомогою хвиль. Відповідно ми можемо виконувати обернену дію і розкладати сигналу на синусоїди, що і називається аналіз Фур’є. Аналіз Фур’є дозволяє нам побачити які частоти представлені у сигналі і наскільки сильним є вплив кожної з них на спектр частот. Математичне перетворення, яке дозволяє нам виконати аналіз Фур’є називається – перетворення Фур’є і виглядає так:

В той же час обернене перетворення виглядає так:

де x(t) – функція від часу;

X(f) – функція від частоти;

j – квадратний корінь з -1;

e – експонента.

Дуже часто для обчислення перетворення Фур’є використовують формулу Ейлера, що подана нижче, та дозволяє зробити обчислення простішими завдяки зв’язку між тригонометричними функціями та комплексною експонентою.

де j – квадратний корінь з -1;

e – експонента;

– певний кут заданий у радіанах.

Застосування перетворення Фур’є до сигналу, виконує перетворення сигналу з його представлення у часі до його представлення у частотному спектрі. Перетворення Фур’є виконує дії теоретично над сигналом нескінченної довжини. Це дає нам інформацію про частотний спектр цілого сигналу і просто забирає багато часу. Це не є проблемою для періодичних сигналів, але у випадку неперіодичних сигналів (музика) нам потрібно інше рішення. Короткочасне перетворення Фур’є розбиває вхідні дані на маленькі фрагменти і здійснює перетворення Фур’є на кожному з них, аналізуючи як сигнал змінюється в часі. Це дозволяє вирішити проблему врахування тимчасових змін у аудіо сигналі, але в той же час створює нові складнощі. Так як перетворення Фур’є потребує, щоб вхідний сигнал був нескінченний, то кожен фрагмент має бути розширений до нескінченної довжини. Вирішити цю проблему можна шляхом перемноження кожного фрагменту на певну віконну функцію. Артефакти не можна повністю усунути за допомогою цього методу, але вони бути суттєво зменшені якщо обрати правильну віконну функцію. Доволі популярною функцією є вікно Хеммінга, яке пропонує нам чудовий компроміс між ефективністю обчислень та якісними результатами. Варто також відмітити, що використання такого методу може призвести до втрати частини інформації на краях вікна, тому що вікна можуть накладатися. Короткочасне перетворення Фур’є може бути обчислене наступним чином:

де h(t) – віконна функція;

X(f, t) – функція від частоти і часу відповідно;

j – квадратний корінь з -1;

e – експонента.

Нажаль, техніки, що були описані вище працюють лише з неперервними сигналами. Їхнім аналогом у цифровій сфері є дискретне перетворення Фур’є. Дискретне перетворення Фур’є може бути обчислене наступним чином:

А обернене перетворення можна обчислити так:

де – n-ий компонент дискретного перетворення Фур’є;

– k-тий компонент сигналу;

j – квадратний корінь з -1;

e – експонента;

N – кількість фрагментів вхідного сигналу та результатів перетворення.

Варто також відмітити, що частота яка отримана за допомогою дискретного перетворення Фур’є досить часто є комплексним числом, навіть в тому випадку, якщо вхідний сигнал був поданий у вигляді дійсного числа. Насправді є не один шлях обчислення дискретного перетворення Фур’є. Найбільш ефективним і найпопулярнішим у алгоритмах, що використовуються для обробки та перетворення аудіо сигналів вважається швидке перетворення Фур’є, яке і буде використовуватись в програмній реалізації даної інформаційної системи для обчислення ознак із вхідного сигналу завантаженої композиції. Швидке перетворення Фур’є є математичним фундаментом на основі якого і побудована вся сфера обробки та розпізнавання аудіо файлів, в тому числі дана інформаційна система.

ВИСНОВКИ ДО РОЗДІЛУ

У другому розділі даної дипломної роботи було проаналізовано предметну область, побудовано функціональну модель даної інформаційної системи використовуючи діаграми потоків даних за методологією Йордана, досліджено загальний алгоритм роботи програми, досліджено перетворення Фур’є. На основі проведеного дослідження ми можемо зробити такі висновки:

* Для того, щоб зробити порівняння композицій можливим, спочатку ми повинні перетворити дані, щоб отримати доступ до цільової інформації що в них міститься. Перетворення Фур’є використовується для цього у всіх системах пов’язаних з обробкою та розпізнаванням аудіо, зокрема найефективнішим і найпопулярнішим у алгоритмах, що використовуються для обробки та перетворення аудіо сигналів вважається швидке перетворення Фур’є.
* Вхідними даними для даної інформаційної системи є композиція у вигляді цифрового аудіо сигналу. Але часто виникають випадки коли певну композицію не можна віднести повністю лише до одного унікального музичного жанру, саме по цій причині кожна завантажена композиція в результаті роботи інформаційної системи може відноситися до трьох жанрів одночасно.
* В результаті побудови діаграми потоків даних була проведена декомпозиція першого рівня та здійснена деталізація усіх процесів для того, щоб побачити повну картину при реалізації системи та побудувати загальний алгоритм її роботи. Була виділена зовнішня сутність користувач та основні чотири процеси, які будуть виконуватись при обробці кожного аудіо файлу під час роботи інформаційної системи. При деталізації кількість процесів у алгоритмі збільшується до дев’яти, що дало змогу побудувати повноцінний алгоритм роботи програми.

РОЗДІЛ 3

ПРОГРАМНЕ ТА ТЕХНІЧНЕ ЗАБЕЗПЕЧЕННЯ

3.1 Засоби розробки

Python3 – потужна високорівнева мова програмування, є найпопулярнішою в області штучного інтелекту та має неймовірну кількість бібліотек, що нам так потрібні для виконання даного завдання:

* Numpy це одна з найпопулярніших бібліотек для роботи з даними. Обрана мною тому, що вона надзвичайно швидка та оптимізована, оскільки побудована на основі мови C.
* Keras це бібліотека для побудови нейронних мереж. Базується на основі бібліотеки TensorFlow. Була обрана через простоту та швидкість розробки прототипів, що є важливим для будь якого проекту на його початку.
* Librosa одна з найкращих бібліотек для роботи з аудіо сигналами на безмежних просторах інтернету. Була обрана тому, що усі представлені конкуренти цієї бібліотеки мають значно гірший функціонал.
* Scikit-learn це бібліотека, що містить широкий функціонал для попередньої обробки даних для нейронної мережі. Використовувалась мною для нормалізації даних та очистки їх.
* Django це фреймворк для реалізації веб додатків. Використовувався мною для реалізації завантаження та обробки композиції на стороні сервера. Був обраний по причині наявності досвіду роботи з ним та досить широкого функціоналу порівняно з конкурентами.

Для реалізації веб інтерфейсу на стороні користувача було вирішено скористатись HTML5, CSS3 та мовою сценаріїв JavaScript з використанням бібліотеки JQuery. Ці засоби були обрані оскільки вони не мають аналогів та є стандартом, підтримування якого гарантується на даним момент всіма існуючими браузерами. Додаткова бібліотека JQuery була обрана мною для прискорення розробки та підвищення читабельності коду. Оскільки вона реалізує більшість функцій потрібних для розробника та бере на себе більшу частину рутинної роботи.

3.2 Вимоги до технічного та програмного забезпечення

При виконанні даного проекту в першу чергу при виборі технічного та програмного забезпечення я звертав увагу на його популярність серед розробників. Оскільки це дозволяє зробити певні висновки про його функціональність – висока популярність зазвичай означатиме, що дане програмне забезпечення може покрити більшу кількість потреб найрізноманітніших користувачів, що ним користуються. Також висока популярність означатиме велику базу знань щодо цього програмного забезпечення на просторах інтернету, яка допоможе у разі якихось питань та проблем, які завжди виникають в процесі розробки та особливо при знайомстві з новим програмним забезпеченням. Другою вимогою до програмного забезпечення була простота та висока швидкість розробки. Оскільки це дозволяє розробнику не писати типовий код, який вже реалізований певними бібліотеками та завершити розробку в короткі терміни, щоб збільшити час на тестування та налагодження стабільної роботи програми. Третьою вимогою звичайно була висока стабільність та надійність. Що дозволяє забезпечити безперебійну роботи програми з мінімальною кількістю помилок. Останньою ж вимогою, що висувалась до програмного забезпечення було те, щоб воно було вільно розповсюджуваним, тобто таким, що не потребує плати чи ліцензії у разі його використання в цілях не пов’язаних з комерцією. Основною вимогою до технічного забезпечення була стабільність та безперебійність роботи та одночасно з тим висока надійність та швидкість обробки даних, оскільки для роботи нейронної мережі необхідні значно більші технічні ресурси, ніж для звичайних додатків. Для цих цілей був використаний хостинг Heroku як один з найкращих варіантів платформи для хмарних додатків, який і гарантує та забезпечує виконання всіх вимог щодо технічного забезпечення.

3.3 Опис програмної реалізації

В першу чергу для виконання цього завдання потрібно підібрати хороший набір даних, для тренування нейронної мережі. Мною був обраний набір даних GTZAN Genre Collection. Даний набір містить сумарно 1000 композицій, кожна з яких має тривалість рівно 30 секунд. Ці композиції розбиті на 10 жанрів, класифікація яких і буде проводитись, а саме:

* Блюз (Blues)
* Класична (Classical)
* Кантрі (Country)
* Диско (Disco)
* Хіп-хоп (Hip-hop)
* Джаз (Jazz)
* Метал (Metal)
* Поп (Pop)
* Реггі (Reggae)
* Рок (Rock)

Для початку було вирішено працювати з цим набором даних не вносячи до нього ніяких змін. Тому перші спроби побудови нейронної мережі на цих даних були не надто вдалими. Після довгого підбору гіперпараметрів максимальна точність яку вдалося отримати дорівнювала 80%. Тоді було прийнято рішення змінити алгоритм роботи програми та видозмінити дані з цього набору. Було створено скрипт, який розрізав композиції в наборі даних на кілька частин. Кількість частин підбиралась експериментально та після кількох спроб було вирішено ділити 30 секундні композиції на 3 частини кожна з яких тривалістю рівно 10 секунд. Тобто в результаті початковий набір даних видозмінився і тепер ми маємо 3000 композицій (по 300 для кожного жанру) кожна з яких тривалістю 10 рівно секунд. В результаті експериментів з новим набором даних вдалося досягти оптимальної точності з використанням нейронної мережі, що становила близько 90%. Далі слідував етап виділення ознак.

Виділення ознак – це перший з двох етапів підготовки даних до класифікації. Цей етап дозволяє формувати нам нові ознаки з даних, що ми маємо, застосовуючи до них певні перетворення. Другий етап – це вибірка ознак, тобто процес пошуку і підбору певної підмножини ознак, яка найкраще підходить для класифікації. Вибірка ознак може бути застосована до первинних даних або до вихідних даних етапу виділення ознак. Для підготовки даних для класифікації можна використовувати як обидва етапи, так і кожен з них окремо. Теоретично, також можливо використовувати первинні дані, якщо вони вже відформатовані правильно для класифікації. На практиці, це досить важко. Оскільки, розмірність наборів даних часто є занадто високою, дані містять багато невідповідностей або просто не підходять для прямого порівняння. Це особливо стосується сфери класифікації аудіо сигналів, де доводиться мати справу з великими потоками надлишкових та шумних сигналів. Зазвичай, ознаки, що використовуються в класифікації аудіо сигналів поділяються на дві головні категорії: фізичні та ознаки сприйняття (perceptual features). Фізичні ознаки базуються на математичному та статистичному аналізі властивостей певного аудіо сигналу. Як приклад головних фізичних ознак можна навести: основна частота (fundamental frequency), енергію (energy), кількість перетинів нуля функції (zero crossing rate). Ознаки сприйняття основані на тому, як людське вухо сприймає звук, наприклад, тон, тембр, ритм. Доволі очевидним є той факт, що всі ознаки сприйняття якимось певним чином залежать від фізичних ознак, враховуючи те, що наше сприйняття основане на фізичному звуковому сигналі. Деякі з цих зв’язків є надзвичайно очевидними: амплітуда звукового сигналу відповідає за гучність звуку, а від основної частоти напряму залежить тон звуку. В будь-якому випадку детальний та уважний розгляд цих зв’язків покаже нам, що вони є просто грубими наближеннями та насправді все є дещо складніше. Багато ознак сприйняття досить важко описати в математичному плані, оскільки вони основані на досить складному для розуміння поєднанні фізичних ознак. Я не збираюсь дотримуватися суворого розмежування між фізичними та ознаками сприйняття у цьому розділі. По-перше, як було зазначено вище, усі ознаки сприйняття основані на фізичних особливостях сигналу. Таким самим чином, всі фізичні ознаки звукового сигналу є суб’єктивними для людського механізму сприйняття і коли ми чуємо звук він певним чином перетворюється в ознаки сприйняття. Ось головні ознаки, що використовувались мною при автоматичному розпізнаванні музичного жанру.

Спектр частот. Розподіл частот сигналу може бути обчислений за допомогою перетворення Фур’є. Схожі процеси також відбуваються у слуховому апараті людей, що дозволяє нам зробити висновок, що склад спектру частот є для нас найголовнішим джерелом інформації про звук. Спектр частот – це одна з найважливіших ознак для автоматичного розпізнавання музичних жанрів, а також використовується як основа для визначення та розрахунку багатьох інших ознак. Цікавий та інтуїтивний спосіб роботи з інформацією про спектр частот – це графічне зображення результату перетворення Фур’є використовуючи спектрограми. Спектрограма – це графік час – частота результатів перетворення Фур’є. Вісь x показує як збільшується час, в той же час вісь y показує частоту. Колір певної точки на спектрограмі показує амплітуду. Спектрограми використовуються як спроба вирішити проблему автоматичного розпізнавання музичного жанру за допомогою технік аналізу зображень.

Частотні спектральні коефіцієнти (Mel-frequency Cepstral Coefficients). Людське сприйняття частоти звуку не піддається лінійній шкалі, але піддається логарифмічному розподілу. Частотні спектральні коефіцієнти основані на інформації про спектр частот звуку, але змодельовані так, щоб захопити лише сприйнятливі частини слухового спектру людини. Послідовність обробки є такою:

* Обчислити перетворення Фур’є
* Перетворити дані з попереднього кроку у фільтри
* Обчислити логарифм з основою 10
* Обчислити косинус

Фільтри це саме те, що і робить ці коефіцієнти унікальними. За допомогою них ці коефіцієнти дозволяють нам максимально наблизитись до сприйняття звуку людським слуховим апаратом. Зазвичай, число найважливіших коефіцієнтів дорівнює 12. Частотні спектральні коефіцієнти зазвичай використовуються в завданнях розпізнавання мови, і схоже, що вони значно краще ніж будь які інші техніки захоплюють важливу для сприйняття частину спектру.

Середня кількість перетинів нуля (Average Zero-Crossing Rate). Перетин нуля стається тоді, коли послідовні зразки певного цифрового сигналу мають різні знаки. Кількість перетинів нуля, може бути використана, як досить проста міра частоти сигналу. Для простих сигналів, кількість перетинів нуля напряму залежить від основної частоти. Синусоїд перетинає вісь двічі за цикл, отже частота може бути визначена, як кількість перетинів нуля поділена на 2. На перший погляд, це виглядає як дуже простий та ефективний метод знаходження основної частоти звукового сигналу, але це не так. Сигнал може містити частинки з частотою, яка значно вище, ніж основна частота, і відповідно в ті моменти функція сигналу буде перетинати вісь багато разів за цикл. Причина, чому для нас є таким важливим визначення основної частоти є те, що вона надзвичайно важлива для визначення тону, а тон в свою чергу це важлива ознака в системах розпізнавання композицій. В області автоматичного визначення музичного жанру, основна частота не є дуже важливою, по кількох причинах. Досить складно визначити основну частоту для музики, оскільки вона може містити кілька джерел звук, таких як інструменти, голос, спеціальні звукові ефекти. Взагалі, це не тільки складне завдання, а ще і не має ніякого сенсу. Питання в тому, що має розглядатися як головний сигнал у такій композиції. Незважаючи на те, що середня кількість перетинів нуля не може бути використана для визначення основної частоти, вона все ще залишається дуже популярної ознакою в області класифікації аудіо сигналів. Ця ознака дуже швидко обчислюється та може бути використана як статистична міра спектру частот, якщо аналізувати кількість перетинів нуля протягом певного часу.

Ритм і такт. Притупувати в такт музиці є абсолютно звичайним завданням для людей, в той же час це завдання виявилось надзвичайно складним для автоматизованих систем обробки аудіо. Визначення ритму – це досить велика область досліджень, що доволі швидко розвивається. Є незліченна кількість факторів, що роблять визначення ритму складним для комп’ютера, в той же час це не є проблемою для людини. Для прикладу, ми можемо доволі легко визначити ритм, навіть якщо на самому початку пісні він не є чітко поставлений. Також після його зміни протягом композиції, ми можемо легко адаптуватися до цього протягом лічених секунд, в той же час як більшість автоматизованих систем визначення не можуть адаптуватися до зміни ритму. Більше того, системи визначення ритму зазвичай дуже погано працюють з шумами – відхиленнями від очікуваного таймінгу. Всі ці фактори зазвичай присутні в музиці. На щастя, точне визначення ритму не є обов’язковим для автоматичного визначення музичного жанру. Нас не цікавить кожен такт, а більше цікавить темп композиції в загальному. Темп може бути описаний, як відчуття того, що звук є повторюваним через однакові проміжки часу. Варто відмітити, що темп – це суб’єктивна ознака, оскільки звук не має справжнього темпу який може бути напряму виміряний.

Тембр. Тембр – це ознака, що дозволяє нам побачити різницю між звуками одного рівня та гучності, якщо вони були створені різними музичними інструментами або голосами. Тембр залежить від спектру частот по більшій мірі і інших часових характеристик аудіо сигналу. Тембр – це одна з найбільш суб’єктивних ознак і до цього часу не існує ні однієї системи яка б могла визначати його точно. Досить часто для інформації про тембр використовуються лише спектрограми. Якщо вдосконалити визначення тембру, то це стане дійсно дуже важливим покращенням для багатьох підрозділів класифікації аудіо сигналів.

За результатами проведеної роботи в сумі я відібрав 98 найважливіших ознак для коректного визначення музичного жанру композиції та взявся за третій етап – класифікацію. Для коректної класифікації дані спочатку необхідно було нормалізувати. Це може здаватися дивним, якщо знати те, що нейронна мережа може пристосовуватися до будь-яких даних і не потребує нормалізації. Але у випадку такої простої архітектури для підвищення швидкості навчання та класифікації нормалізовані дані просто необхідні. Мною була обрана стандартна нормалізація, що детально описана у попередніх розділах.

Була експериментально підібрана доволі проста архітектура нейронної мережі, яка була компромісом між точністю класифікації та її швидкістю. Оскільки, для користувача є дуже важливим швидкість обробки інформації, я віддав перевагу простій архітектурі нейронної мережі, що дозволила мені досягти високої точності при високій швидкості обробки композиції. Враховуючи те, що нейронна мережа вчилася розпізнавати жанр на композиціях, що мали тривалість 10 секунд, то і завантажений користувачем файл ми будемо ділити на відрізки по 10 секунд. Після чого почергово, за допомогою бібліотеки для обробки аудіо будемо виділяти обрані мною 98 ознак з композиції. Слідом, використовуючи ці ознаки, ми будемо розпізнавати музичний жанр кожного відрізку за допомогою попередньо натренованої нейронної мережі. Заключним етапом нашого алгоритму буде визначення жанру завантаженої композиції методом голосування – якщо певний жанр мають більше третини відрізків то вважається, що композиція відноситься до цього жанру. Таким чином, ми можемо помітити, що кожна композиція може мати до трьох жанрів одночасно. Алгоритм роботи був детально описаний у попередніх розділах.

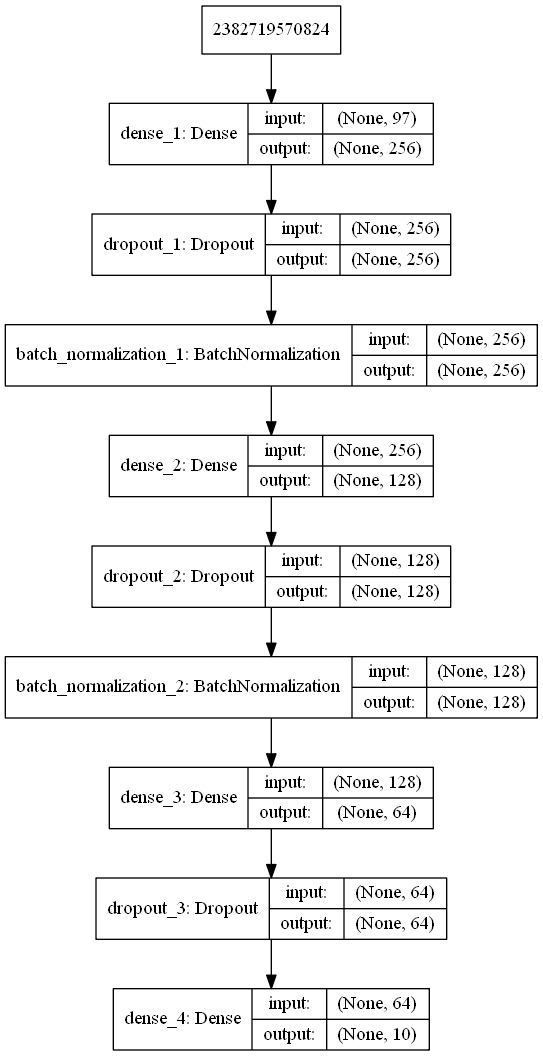


Рис. 3.1 Архітектура нейронної мережі

Для виконання завдання мною була обрана звичайна повнозв’язна нейронна мережа. Було використано 4 повнозв’язних шари з кількістю нейронів 256, 128, 64, 10 відповідно, після кожного з яких використовувалася пакетна нормалізація та техніка Dropout з імовірностями відповідно 20%, 20%, 50% для уникнення перенавчання. Архітектуру нейронної мережі можна переглянути на рис. 3.1.

Файлова структура:

* Constants – файл, що містить константи, що не будуть змінюватися протягом роботи програми.
* Features\_extracting – файл, що містить функції, для прямої роботи з аудіо файлами. Функція get\_features приймає на вхід аудіо в цифровому форматі та повертає список виділених з нього ознак. Функція main розрізає 30 секундні композиції зі стартового набору даних на 10 секундні та зберігає їх на диску. Приймає на вхід шлях до папки з композиціями, результатом роботи є створені нові аудіо файли в цій папці.
* Model – файл, в якому здійснюється попередня обробка даних, нормалізація та створення і тренування нейронної мережі. Містить лише одну функцію main, яка приймає на вхід шлях до набору даних після чого нормалізує його, створює та тренує нейронну мережі на основі вказаного набору даних. Результатом роботи функції є натренована нейронна мережі, яка зберігається на диску, щоб потім використати її в основній частині програми.
* Main – реалізує основний алгоритм роботи програми. Містить функцію predict\_genre що виконує класифікацію музичного жанру. Приймає на вхід шлях до композиції яку потрібно класифікувати, шлях до файлу з натренованою нейронною мережею. Результатом роботи даної функції є текстова стрічка зі списком музичних жанрів до яких була віднесена дана на вхід композиція.

3.4 Керівництво користувача

Перш за все, для зручності використання даної інформаційної системи було створено веб інтерфейс для користувача. Оскільки виконання програми у командному рядку було б дуже незручним для підтримки різних платформ, розповсюдження та користування.

Тому для початку користувачу потрібно бути підключеним до мережі інтернет. Користувач має використовувати будь-який веб браузер, що підтримує останні версії мови JavaScript, HTML5, CSS3. Після чого користувач повинен перейти за посиланням https://audio-genre-recognition.herokuapp.com/. Головна сторінка веб сайту показана на рис. 3.2.

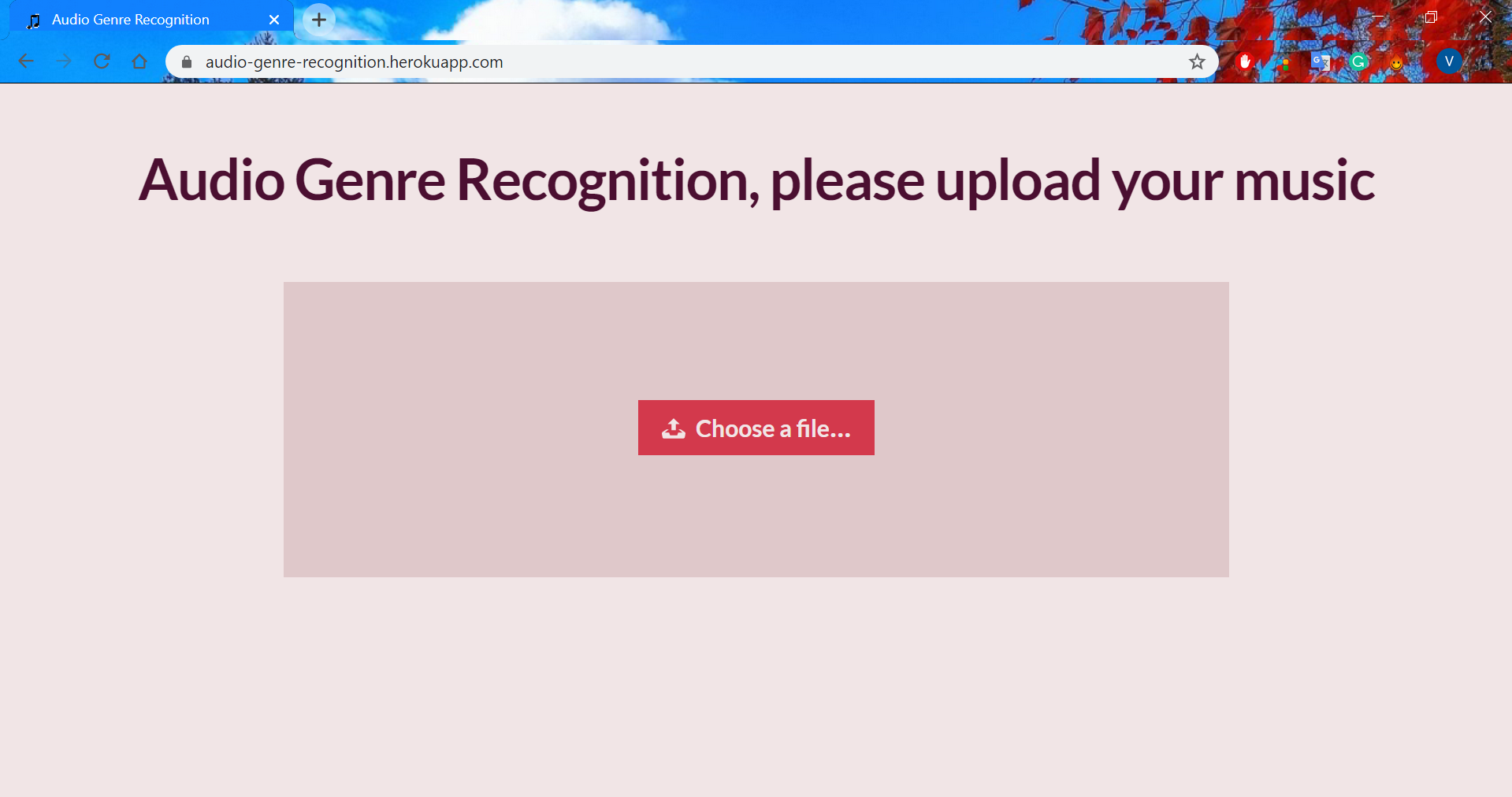


Рис. 3.2 Головна сторінка веб сайту

Інтерфейс виконаний у мінімалістичному стилі та є максимально простим для використання оскільки містить лише одну кнопку. Після натискання на дану кнопку відкривається вікно провідника у якому користувач повинен вибрати певну композицію. Також на рис. 3.3 ми можемо помітити, що здійснюється обмеження на формати файлів, які завантажує користувач. Це необхідно для того, щоб не виникало помилок при завантаженні некоректних файлів, які можуть заплутати користувача та змусити його завершити користування даним веб інтерфейсом надавши перевагу конкурентам зі схожою функціональністю.

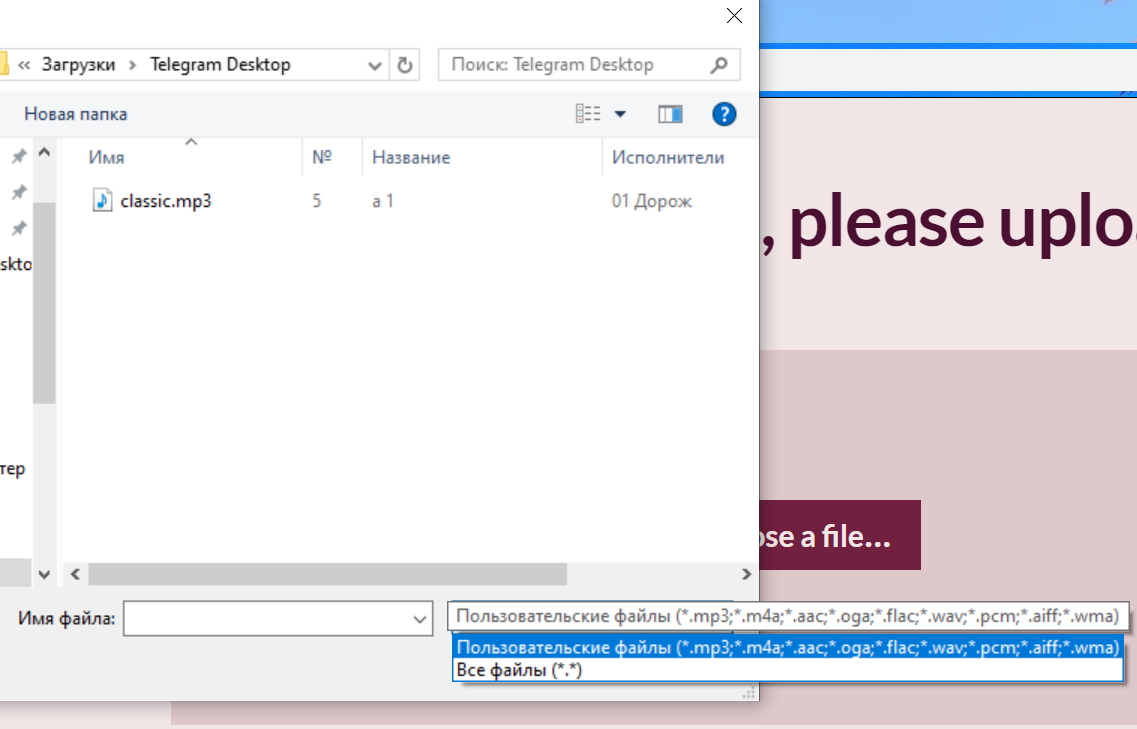


Рис. 3.3 Приклад обмеження на формат файлу, що завантажується

Як ми можемо бачити, проста перевірка на тип файлу у провіднику нас не врятує, оскільки у користувача завжди буде можливість обрати опцію всі файли. Для вирішення цієї виключної ситуація було вирішено здійснити перевірку розширення файлу за допомогою засобів JavaScript. В результаті у випадку завантаження користувач файлу неправильного формату для нього буде показано повідомлення про помилку та список дозволених форматів, що ми і можемо побачити на рис. 3.4.

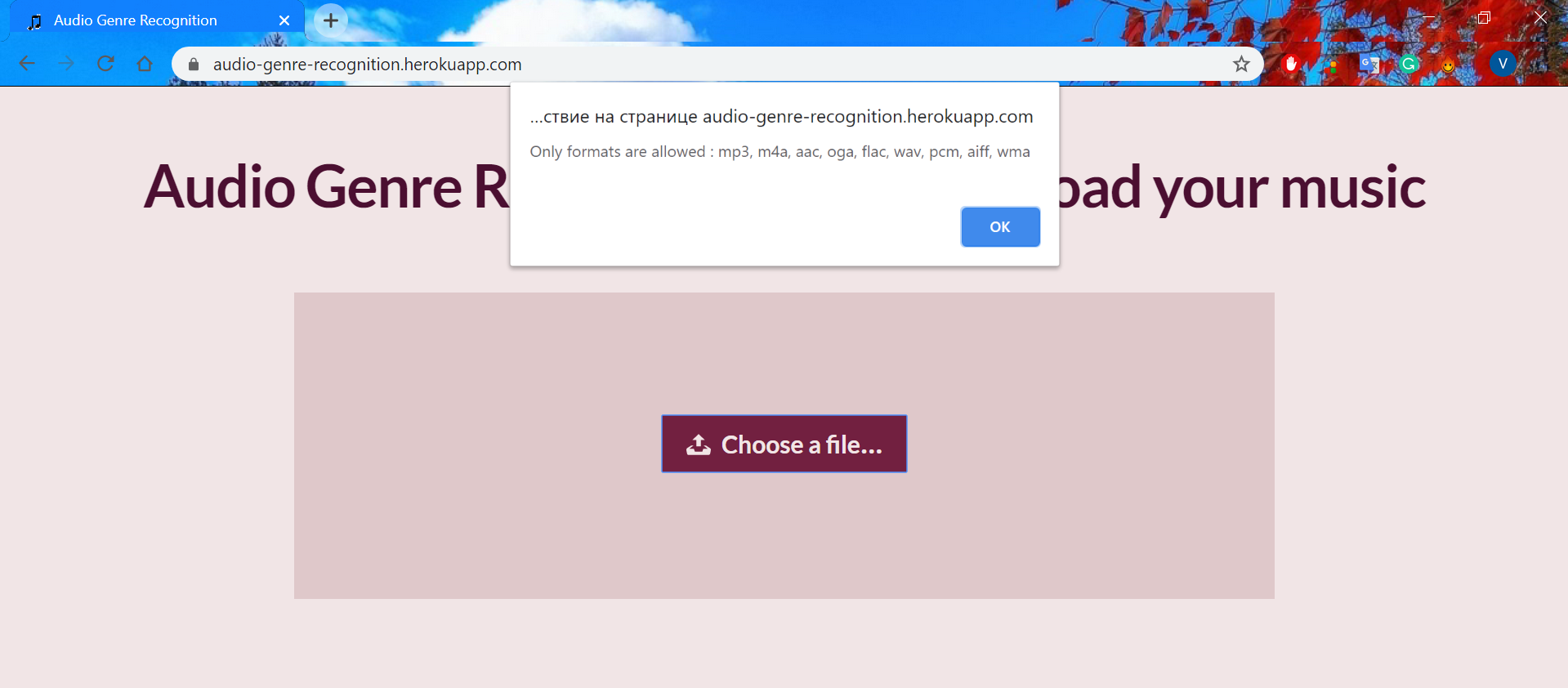


Рис. 3.4 Перевірка формату файлу засобами JavaScript

Після чого у разі успішного проходження всіх перевірок файл передається на сервер де і в подальшому аналізується нейронної мережею, що супроводжується візуальними ефектами, оскільки цей процес аналізу та автоматичного визначення жанру не є швидким. Тому візуальні ефекти необхідні, щоб показати користувачу, що файл успішно обробляється в даний момент і потрібно лише трішки почекати. На рис. 3.5 продемонстровано використані мною візуальні ефекти – у вигляді кола, що анімується і показує про те, що триває певна операція.

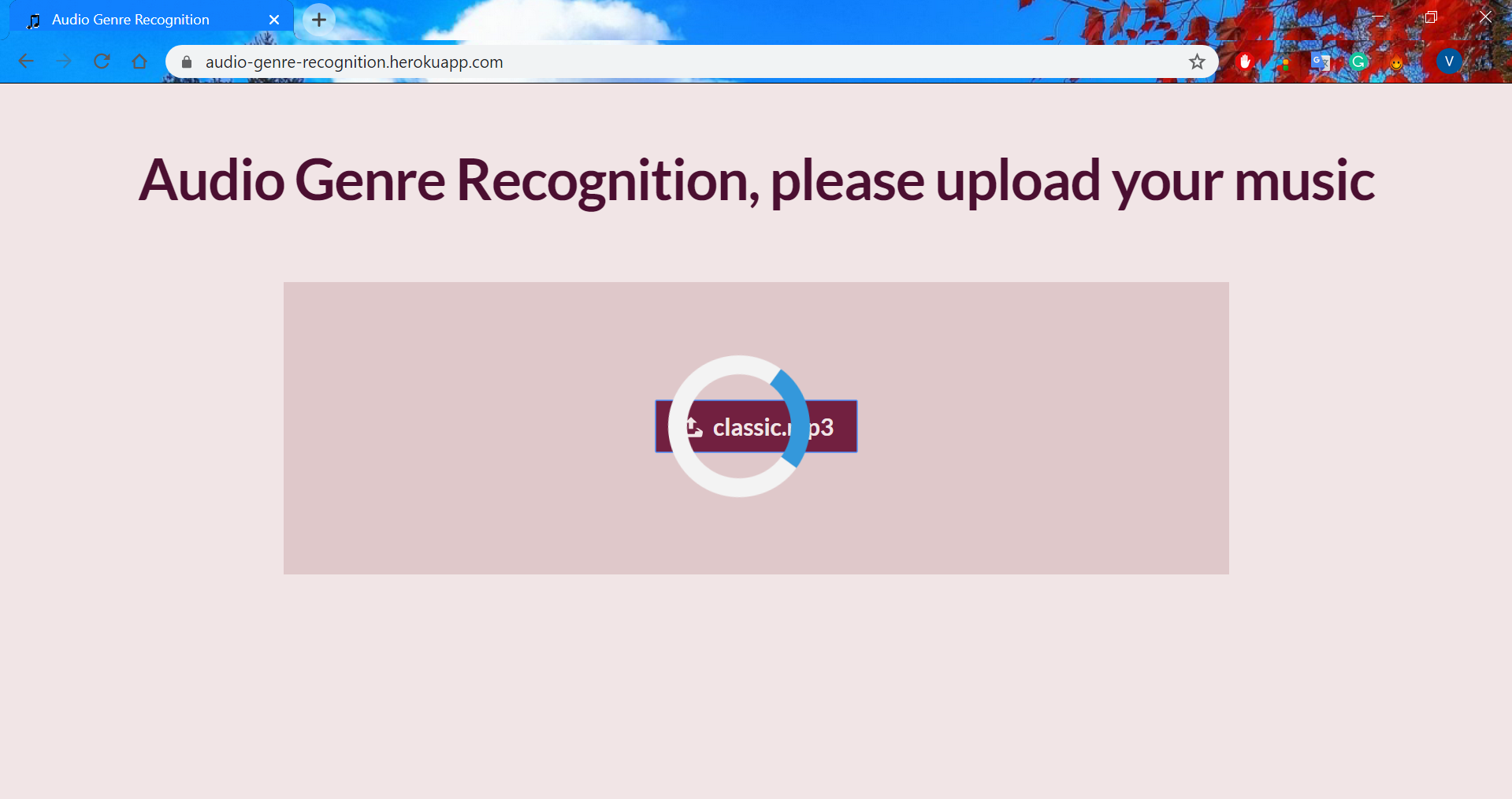


Рис. 3.5 Візуальний ефект завантаження

Надалі у разі виникнення якоїсь помилки на сервері вона буде відправлена користувачу. Якщо файл був успішно опрацьований та визначення жанру також пройшло успішно користувач побачить спливаюче вікно в якому буде жанр завантаженої раніше композиції, що показано на рис. 3.6.

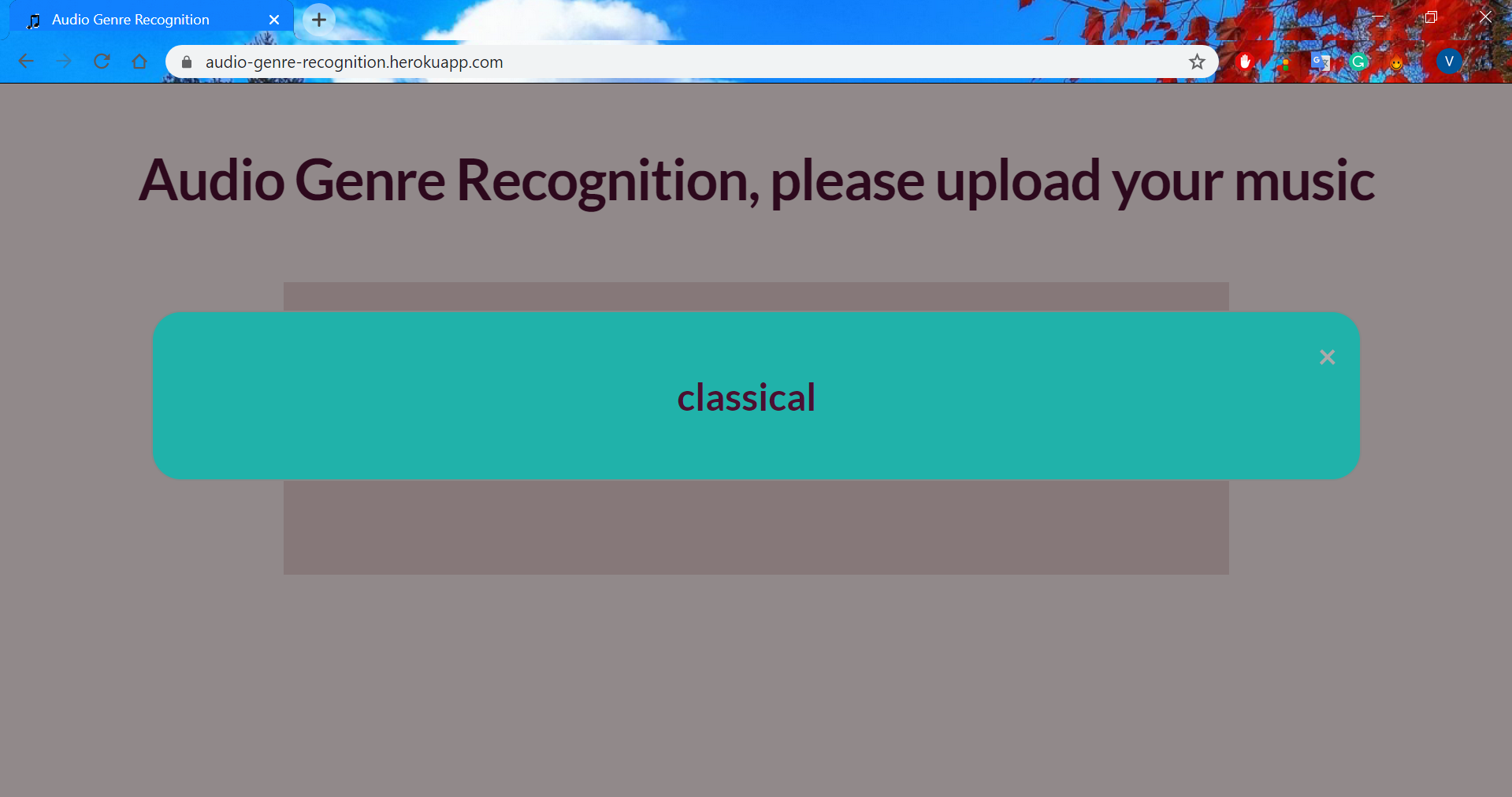


Рис. 3.6 Результат роботи системи

ВИСНОВКИ ДО РОЗДІЛУ

У третьому розділі даної дипломної роботи було здійснено обґрунтування вибору засобів розробки, сформовані вимоги до технічного та програмного забезпечення, була виконана та описана програмна реалізація та керівництво користувача. На основі проведеного дослідження ми можемо зробити такі висновки:

* Серед найважливіших вимог до програмного та технічного забезпечення варто виділити: популярність серед розробників, велика база знань, простота та висока швидкість розробки, висока стабільність та надійність. Основною вимогою до технічного забезпечення була стабільність, безперебійність роботи, висока надійність, швидкість обробки даних, оскільки для роботи нейронної мережі необхідні значно більші технічні ресурси.
* Підготовка даних є надзвичайно важливою у разі роботи зі штучним інтелектом, та складається з основних етапів: виділення ознак, вибірка ознак, нормалізація даних.
* Головні ознаки, що використовувались мною при автоматичному розпізнаванні музичного жанру: спектр частот, частотні спектральні коефіцієнти, середня кількість перетинів нуля, ритм, такт, тембр.
* Для користувача є дуже важливою швидкість обробки інформації, тому завжди варто знаходити компроміс між точністю розпізнавання та швидкістю. Варто надавати перевагу більш простим моделям, якщо вони дають задовільний результат.
* Необхідно реалізовувати зручний інтерфейс для користувачів, оскільки користувач сприймає всю програмну реалізацію загалом не тільки на основі якості її роботи, а також на основі її візуальних даних.

РОЗДІЛ 4

ОЦІНЮВАННЯ РІВНЯ ГОТОВНОСТІ НАУКОВО-ТЕХНІЧНОЇ РОЗРОБКИ ДО КОМЕРЦІАЛІЗАЦІЇ

Метою роботи є розробка програмного продукту, що дозволяє швидко визначити музичний жанр аудіо файлу.

Об’єктом дослідження є інформаційна система автоматичного визначення музичного жанру аудіо файлу.

Предметом дослідження є способи визначення жанру аудіо файлу за допомогою систем штучного інтелекту.

Оцінювання проводять за двома етапами:

1. Попереднє оцінювання НТР
2. Оцінювання готовності НТР до комерціалізації.

Для виконання першого етапу - попереднього оцінювання необхідно дати відповіді на такі питання.

1. Назва НТР:

Інформаційна система розпізнавання музичного жанру аудіо файлу.

1. Галузь промисловості:

Галузь інформаційних технологій (інформаційні послуги).

1. Загальне призначення НТР (зокрема потреби, яку вона повинна задовольнити).

Визначення музичного жанру аудіо файлу.

1. Чи існують аналоги (прототипи) Вашої НТР (якщо так, то їх зазначити):

YouTube Music, Apple Music, Sound Cloud, Spotify, Shazam.

1. Опишіть конкурентні переваги Вашої НТР, завдяки яким вона може бути привабливою для комерціалізації.

визначення жанру, швидкість роботи, низька вартість розгортання та підтримки.

1. Вкажіть можливі сфери використання Вашої НТР (вид економічної діяльності, наприклад, охорона здоров’я, кондитерська промисловість, будівництво, оборонна промисловість, туризм тощо):

Сфера послуг, сфера розваг.

1. Вкажіть прогноз масштабів можливої реалізації результатів вашої НТР (можливі кілька варіантів відповідей):

* для галузі на рівні країни; для конкретного підприємства;

1. Зазначте тип/типи ринків, на якому можливе використання результатів Вашої НТР (можливі кілька варіантів відповідей):

* ринок кінцевих споживачів;
* ринок підприємств та організацій;
* ринок посередників.

1. Вкажіть рівень новизни ідеї, яка покладена в основу Вашої НТР:

* нова для розробника (організації);

1. Вкажіть рівень глибини дослідження предмету Вашої НТР:

* ідея модифікується на підставі попередніх досліджень.

1. Виберіть один варіант відповіді щодо технологічної готовності (TRL) Вашої НТР:

* TRL5 – Перевірка прототипу в робочому середовищі користувача

2. Виберіть один варіант відповіді щодо аналітичної готовності (ARL) Вашої НТР:

* ARL4 – Уточнення якісних показників розробки та очікуваних результатів використання

3. Виберіть один варіант відповіді щодо патентної готовності (PRL) Вашої НТР:

* PRL0 – Патентування не планується

4. Виберіть один варіант відповіді щодо готовності ринку (ринкового попиту) (DRL) для Вашої НТР:

* DRL 2. Оцінювання існування попиту на розробку на внутрішньому ринку

5. Виберіть один варіант відповіді щодо впливу на суспільство Вашої НТР (SIL):

* SIL 2. Встановлення впливу розробки на особистісному рівні (нанорівень)

ЗАГАЛЬНІ ВИСНОВКИ

При вирішенні завдання аналізу предметної області в роботі показано, що системи штучного інтелекту є невід’ємною частиною сучасної музичної сфери, а визначення жанру є одним з найважливішим завдань в області класифікації аудіо сигналів та може використовуватись як окремо, та і як частина більш складної функціональності. Нажаль в наш час механізми обробки та сприйняття звуку людиною все ще не є до кінця вивченими та кожна людина все таки сприймає звук по різному, що і не дозволяє нам створювати дійсно точні системи.

При вирішенні завдання проектування інформаційної системи в роботі показано спосіб побудови діаграм потоків даних за методологією Йордана, показана декомпозиція першого рівня та здійснена деталізація усіх процесів системи. Визначено, що для ефективного порівняння таких складних даних як звук, нам потрібно спочатку перетворити їх. Для чого і було детально розібрано та досліджено алгоритм перетворення Фур’є, що використовується для виділення ознак у системах, що працюють з аудіо. На основі діаграм потоків даних побудовано загальний алгоритм роботи програми, на основі якого і буде працювати програмна реалізація.

При вирішенні завдання реалізації програмного продукту та вибору інструментарію визначено основні вимоги до програмного та технічного забезпечення серед яких основними є: популярність, велика база знань, простота, висока швидкість розробки, висока стабільність, надійність. Показано, що для будь-яких завдань пов’язаних зі штучним інтелектом найважливішим завданням є підготовка даних, оскільки від неї фактично залежить весь успіх системи. Експериментально визначено, що найважливішими ознаками в завданнях розпізнавання аудіо є: спектр частот, частотні спектральні коефіцієнти, середня кількість перетинів нуля, ритм, такт, тембр. Показано, що при побудові систем штучного інтелекту першочерговим завданням є знаходження компромісу між швидкістю та точністю її роботи.

Таким чином, завдання вирішені в повному обсязі, мета досягнута – розроблений програмний продукт, що дозволяє швидко визначити музичний жанр аудіо файлу. Практична значущість нашого дослідження полягає в тому, що отримані результати допоможуть зрозуміти як системи штучного інтелекту сприймають звук та можуть використовуватись для покращення класифікації аудіо файлів у інших завданнях.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

Wikipedia – перетворення Фур’є [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Fourier_transform>

Wikipedia – швидке перетворення Фур’є [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Fast_Fourier_transform>

Wikipedia – дискретне перетворення Фур’є [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Discrete_Fourier_transform>

Розпізнавання музичного жанру використовуючи Python [Електронний ресурс] // Режим доступу: <https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8>

Розпізнавання музичного жанру використовуючи спектрограми [Електронний ресурс] // Режим доступу: <https://www.researchgate.net/publication/224251903_Music_genre_recognition_using_spectrograms>

GTZAN genre collection [Електронний ресурс] // Режим доступу: <http://marsyas.info/downloads/datasets.html>

Wikipedia – музичний жанр [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Music_genre>

Wikipedia – виділення аудіо ознак [Електронний ресурс] // Режим доступу: <https://en.wikipedia.org/wiki/Music_information_retrieval>

Приклад виділення аудіо ознак використовуючи Python [Електронний ресурс] // Режим доступу: <https://musicinformationretrieval.com/>

Визначення музичного жанру використовуючи нейронні мережі [Електронний ресурс] // Режим доступу: <https://medium.com/@navdeepsingh_2336/identifying-the-genre-of-a-song-with-neural-networks-851db89c42f0>

Класифікація музичних жанрів використовуючи техніки машинного навчання [Електронний ресурс] // Режим доступу: <https://www.researchgate.net/publication/324218667_Music_Genre_Classification_using_Machine_Learning_Techniques>

Приклад використання Keras [Електронний ресурс] // Режим доступу: <https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/>

Методи побудови DFD [Електронний ресурс] // Режим доступу: <https://habr.com/ru/company/trinion/blog/340064/>

Як створити DFD [Електронний ресурс] // Режим доступу: <https://www.visual-paradigm.com/tutorials/data-flow-diagram-dfd.jsp>

Документація бібліотеки Librosa [Електронний ресурс] // Режим доступу: <https://librosa.github.io/librosa/>

ДОДАТОК А