

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО”  
КАФЕДРА ПРИКЛАДНОЇ МАТЕМАТИКИ

КУРСОВА РОБОТА  
з навчальної дисципліни “Бази даних та інформаційні системи-2”  
на тему  
ВИБІР МЕЛОДІЙ НА РАДІО КПІ НА ОСНОВІ ПОБАЖАНЬ СТУДЕНТІВ

Виконала студентка групи КМ-62  
ХОДОС З.О.

Керівник  
ТЕРЕЩЕНКО І.О.

Оцінка:  
Кількість балів:

## ЗАВДАННЯ НА ВИКОНАННЯ КУРСОВОЇ РОБОТИ

Створити інформаційну систему, що дозволить на основі побажань студентів формувати список пісень, які будуть програвати на радіо. Для цього виконати передпроектне дослідження (проаналізувати бізнес-процеси, бізнес-правила та елементарні події), визначити сутності, що міститимуться у базі даних, проаналізувати операції, що можуть виконуватись над кожною із сутностей, забезпечити цілісність системи. Додатково виконати кореляційний та кластерний аналіз. Розгорнути завдання на віддаленому сервері.

## АНОТАЦІЯ

В пояснювальній записці до курсової роботи описано основні етапи розробки інформаційної системи. Перший розділ містить інформацію про виконання перед проектного дослідження. У другому розділі надано детальну постановку задачі. В третьому розділі виконано моделювання бізнес-процесів. У четвертому розділі визначено модель даних, описано сутності інформаційної системи, а в розділі 5 надано графічне представлення моделі.

## РЕФЕРАТ

В ході виконання курсового проекту було створено інформаційну систему на базі веб-фреймворку Flask. База даних, на якій виконано завдання – PostgreSQL 10. Власне модель створюється на базі Python, SQLAlchemy, а зі сторони PostgreSQL налаштовано обмеження на унікальність, які виникли після оптимізації бази даних та введення штучних ключів. Також з боку PostgreSQL створені послідовності.

Веб-застосунок дозволяє переглядати дані, наявні в інформаційній системі, змінювати та видаляти дані, де це допустимо та додавати дані у всі сутності. Додавання даних валідується на WTForms, на яких також, за допомогою налаштування випадаючих списків та обмеженого вибору підтримано цілісність даних. З Python до HTML-сторінок передається інформація та відображається з допомогою шаблонізатора Jinja, а веб-сторінки зверстані з використанням фреймворку Bootstrap-4.

Також можна виконувати пошук за факультетом для виявлення психотипів студентів, що навчаються на ньому – кластерний аналіз. Машинне навчання та кореляційний аналіз (залежність між виходом пісні виконавця та частотою на радіо) виконані на базі відкритих даних із Kaggle.com з використанням Random Forest Regressor, а віддалений сервер для розгортання виконаного завдання міститься на Heroku.

## ЗМІСТ

ВСТУП.....	6
1 АНАЛІЗ ПІДПРИЄМСТВА АВТОМАТИЗАЦІЇ.....	7
2 ПОСТАНОВКА ЗАДАЧІ.....	8
2.1 Категорії користувачів.....	8
2.2 Класи даних.....	9
2.3 Бізнес-правила.....	12
2.4 Матриця елементарних подій (сценарії).....	12
3 МОДЕЛЮВАННЯ БІЗНЕС-ПРОЦЕСІВ.....	16
4 ІНФОЛОГІЧНЕ ПРОЕКТУВАННЯ.....	17
5 ДАТАЛОГІЧНЕ ПРОЕКТУВАННЯ.....	18
ВИСНОВКИ.....	20
СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ.....	22

## ВСТУП

Станом на сьогодні вже існує сервіс, що приймає замовлення від студентів на програвання мелодій з Радіо КПП. Це телеграм-бот. Він містить цілком прийнятний інтерфейс, дозволяє переглядати, яка мелодія грає зараз, які гратимуть наступні. Студенти можуть замовляти пісні, проте ті, які наявні у базі. Власне замовлення ще й цензуються. Проте невідомо, чи використовують творці телеграм-бота отриману інформацію, чи роблять дослідження і чи якимось оптимізують створення плейлістів.

Метою курсової роботи є створити підґрунтя для мікросервісу, що зможе доповнити розважальний сегмент додатку інструментом для проведення досліджень. Згодом на базі виконаних досліджень можна буде прогнозувати замовлення, або пропонувати мелодію користувачеві, що спростить йому користування. Недоліком такого рішення є те, що не всі зможуть погодитись надавати персональні дані: такі люди виберуть додаток, який цього не потребує.

Об'єктом дослідження у курсовій роботі є функціонал наявних інформаційних систем, існуючі рішення, які піддаються інтеграції та психотипи особистостей, їх співставлення із жанрами музичних композицій.

Метод дослідження – пошук та систематизація наявних рішень для створення інформаційних систем та мікросервісів.

## 1 АНАЛІЗ ПІДПРИЄМСТВА АВТОМАТИЗАЦІЇ

Відповідно до work-flow із постановки задачі, маємо процеси збору побажань, додавання мелодії у плейліст та програвання. Задача кластеризації студентів по психотипам привела до процесів пошуку психотипу за факультетом та перегляду стовпчикової діаграми розподілу. У короткому представленні ієрархія процесів наведена на рисунку 1.1.

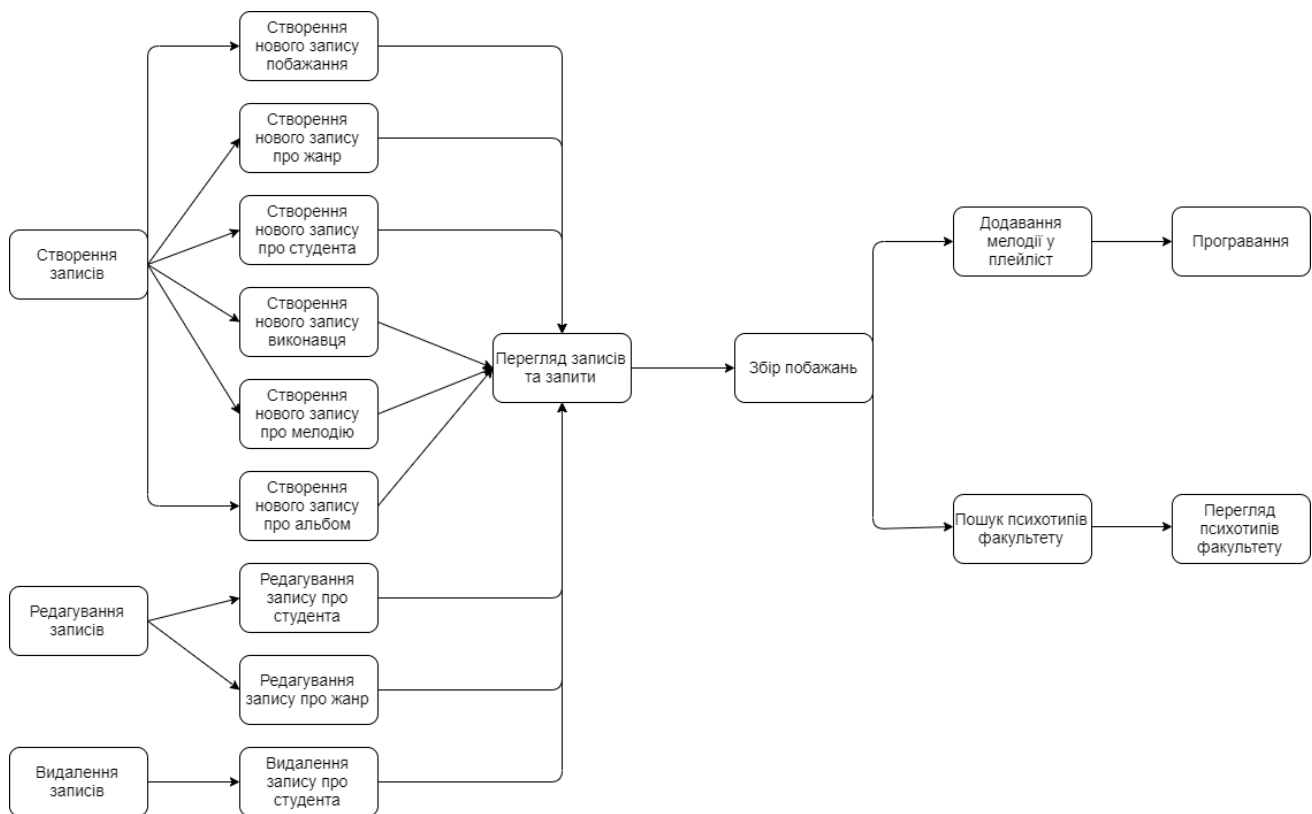


Рисунок 1.1 – Ієрархія процесів

## 2 ПОСТАНОВКА ЗАДАЧІ

Основна відмінність простої автоматизації робочого процесу від пропонованого мікросервісу в тому, що будемо здобувати корисну інформацію та робити дослідження. Наприклад, психотипу людини від жанрів мелодій, які замовляє користувач.

Основна функція мікросервісу: враховуючи побажання, надавати аудіо-файли.

Вхідні дані мікросервісу: критерій відбору (жанр, виконавець).

Приклад структури вхідних даних мікросервісу: меню або текстове повідомлення (стосується вже інтерфейсу безпосередньо у телеграм-боті). Словник: користувач – замовлення.

Результат роботи мікросервісу: дата – файл.

Задача кластеризації: залежність між психотипами та жанрами використовувати для аналізу студентів факультету.

Мета задачі кластеризації для поставленої задачі: для того, щоб будувати правила залежності між психотипами людей та мелодією.

Задача кореляційного аналізу: пов'язувати піки виходу нових треків виконавця та частоту замовлень на радіо.

Визначити мету кореляційного аналізу та для яких атрибутів: щоб на основі психотипу та правил залежності прогнозувати мелодію.

Далі розглянемо основні категорії користувачів сервісу, класи даних та елементарні події, що відбуваються у ході користування системою.

### 2.1 Категорії користувачів

У сервісі, що реалізовано в курсовому проекті є тільки одна категорія користувачів. Звичайно, сервіс можна розширити з додаванням авторизації та різних ролей. Тоді до різних сутностей та дій будуть різні права доступу.

Наприклад, адміністратор зможе робити будь-які CRUD операції прямо на сервісі, не дістаючись при цьому до інтерфейсу PostgreSQL, все буде повністю автоматизовано. Також можна розробити інтерфейс для спілкування із базою.



Наприклад, при натисканні на певну кнопку буде проходити аналіз таблиці «Побажання», вибрати серед них ті, на основі яких може формуватись плейліст та пропонуватиметься композиція. Зараз же повної цілісності для цієї таблиці немає, не можна залишити пусті критерії та побажання можуть бути фіксованими, не відповідати дійсності.

Для користувача тоді ж буде логічно реєструватися, додавати нові записи до таблиці побажання.

## 2.2 Класи даних

Опишемо наявні у системі класи даних у таблицях 2.1 –2.6.

Таблиця 2.1 – Клас даних «Студент»

<b>Сутність</b>	<b>Студент</b>	
<b>Опис сутності</b>	Студент, що є основним користувачем системи	
<b>Атрибути сутності</b>	<b>Опис атрибуту</b>	<b>Пов’язана сутність із атрибутом</b>
ID	Універсальний ідентифікатор	Побажання
Факультет	Факультет, на якому навчається студент	—
Група	Група, в якій навчається студент	—
Прізвище	Прізвище студента	—
Ім’я	Ім’я студента	—
Нікнейм	Псевдонім у Телеграм	—

Таблиця 2.2 – Клас даних «Жанр»

<b>Сутність</b>	<b>Жанр</b>	
<b>Опис сутності</b>	Жанр мелодії	
<b>Атрибути сутності</b>	<b>Опис атрибуту</b>	<b>Пов'язана сутність із атрибутом</b>
ID	Універсальний ідентифікатор	Побажання
Назва	Як називається жанр	Побажання
Психотип	До якого психотипу відноситься людина, що слухає мелодію цього жанру	Студент

Таблиця 2.3 – Клас даних «Виконавець»

<b>Сутність</b>	<b>Виконавець</b>	
<b>Опис сутності</b>	Співаки та гурти	
<b>Атрибути сутності</b>	<b>Опис атрибуту</b>	<b>Пов'язана сутність із атрибутом</b>
ID	Універсальний ідентифікатор	Альбом, мелодія
Псевдонім	Як називається жанр	Мелодія

Таблиця 2.4 – Клас даних «Мелодія»

<b>Сутність</b>	<b>Мелодія</b>	
<b>Опис сутності</b>	Пісні, що належать виконавцям та містяться в альбомах	
<b>Атрибути сутності (1)</b>	<b>Опис атрибуту (2)</b>	<b>Пов'язана сутність із атрибутом (3)</b>

Продовження таблиці 2.4

1	2	3
ID	Універсальний ідентифікатор	Альбом, побажання
Назва	Назва мелодії	Альбом
Жанр	До якого жанру належить мелодія	Жанр
Дата релізу	Коли мелодія була випущена	—
Виконавець	Який співак/гурт виконує	Виконавець
Альбом	В якому альбомі міститься	Альбом

Таблиця 2.5 – Клас даних «Альбом»

<b>Сутність</b>	Альбом	
<b>Опис сутності</b>	Альбом, що об'єднує пісні	
<b>Атрибути сутності</b>	<b>Опис атрибуту</b>	<b>Пов'язана сутність із атрибутом</b>
ID	Універсальний ідентифікатор	Мелодія
Назва	Як називається альбом	—
Виконавець	Який співак/гурт випустив	Виконавець

Таблиця 2.6 – Клас даних «Побажання»

<b>Сутність</b>	Побажання	
<b>Опис сутності</b>	Замовлення, яке робить студент	
<b>Атрибути сутності (1)</b>	<b>Опис атрибуту (2)</b>	<b>Пов'язана сутність із атрибутом (3)</b>

Продовження таблиці 2.6

1	2	3
ID побажання	Універсальний ідентифікатор	—
ID студента	Ідентифікатор студента, що зробив замовлення	Студент
Дата побажання	Дата утворення замовлення	—
Жанр побажання	Жанр, мелодії в якому бажає отримати користувач	Жанр
Мелодія побажання	Назва мелодії, яку бажає почути користувач	Мелодія
Бажаний виконавець	Виконавець, пісні якого хоче побачити користувач	Виконавець

### 2.3 Бізнес-правила

Далі слід визначити бізнес-правила для інформаційної системи:

- а) студент повинен мати факультет, групу, де він навчається;
- б) студент повинен мати прізвище, ім'я та нікнейм у Телеграм;
- в) користувач не може замовити мелодію, якої немає в базі даних;
- г) мелодія може бути визначена лише жанром, що міститься у базі;
- д) мелодія може належати лише альбомам, що у базі;
- е) дивитися можна статистику психотипів лише на тих факультетах, що наявні у базі.

### 2.4 Матриця елементарних події (сценарії)

Перший рівень містить багато однотипних процесів, тому в їх рамках будемо описувати лише один, а інші – аналогічні.

Таблиця 2.7 – Процес створення нового запису в сутності

Назва процесу	Додавання запису до сутності
Сутності	Відповідно кожна із сутностей, зазначена у таблицях 2.1–2.6 цього розділу
Вхідні атрибути сутності	В процесі додавання вхідні атрибути пусті, бо запис ще не вставлено
Опис функціоналу	Дані з форми на веб-сторінці записуються у поля класу, що відповідають сутності та відправляються на сервер
Змінені атрибути сутності	Рядок вставлених даних

Таблиця 2.8 – Процес редагування запису в сутності

Назва процесу	Редагування запису у сутності
Сутності	Студент/Жанр
Вхідні атрибути сутності	Рядок, отриманий за ID поля, яке збираємося відредагувати
Опис функціоналу	Дані приходять із серверу та пишуться у поля форми. Після цього у формі користувач змінює значення атрибутів. Дані зчитуються та відправляються на сервер
Змінені атрибути сутності	Рядок відредагованих даних

Таблиця 2.9 – Процес видалення запису із сутності

Назва процесу	Видалення запису із сутності
Сутності	Студент

Продовження таблиці 2.9

Вхідні атрибути сутності	Унікальний ідентифікатор
Опис функціоналу	Отримуємо запис сутності за ID, що надійшов після натискання на відповідну кнопку, та видаляємо запис за цим ID
Змінені атрибути сутності	Тепер рядок за цим ідентифікатором відсутній

Таблиця 2.10 – Процес збору побажань

Назва процесу	Збір побажань
Сутності	Побаження
Вхідні атрибути сутності	Всі атрибути сутності побажання
Опис функціоналу	Обробка побажань, які з них відповідають дійсності (немає конфліктів виконавець-жанр, жанр-мелодія, мелодія-виконавець)
Змінені атрибути сутності	Процес не змінює атрибути

Таблиця 2.11 – Процес пошуку психотипів факультету

Назва процесу	Пошук психотипів факультету
Сутності	Студент, Побаження, Мелодія, Жанр
Вхідні атрибути сутності	Факультет
Опис функціоналу	Користувач обирає факультет з випадającego списку та натискає кнопку пошуку. Відбувається операція SELECT із об'єднання сутностей по ключам
Змінені атрибути сутності	Сутність не змінює атрибутів

Таблиця 2.12 – Процес перегляду психотипів факультету

Назва процесу	Перегляд психотипів факультету
Сутності	Студент, Жанр
Вхідні атрибути сутності	Факультет, Психотип
Опис функціоналу	У вигляді графіку відображаються кількості студентів факультету за психотипами
Змінені атрибути сутності	Сутність не змінює атрибутів

Таблиця 2.13 – Процес додавання мелодії у плейліст

Назва процесу	Додавання мелодії у плейліст
Сутності	Побажання
Вхідні атрибути сутності	Назва мелодії/Жанр/Виконавець
Опис функціоналу	Оброблені в процесі з таблиці 2.10 побажання формують плейлист на основі VIEW із SELECT
Змінені атрибути сутності	Сутність не змінює атрибутів

### 3 МОДЕЛЮВАННЯ БІЗНЕС-ПРОЦЕСІВ

Змодельований Use-Case для процесів, що можна виконувати наразі у веб-застосунку, зображено на рисунку 3.1.

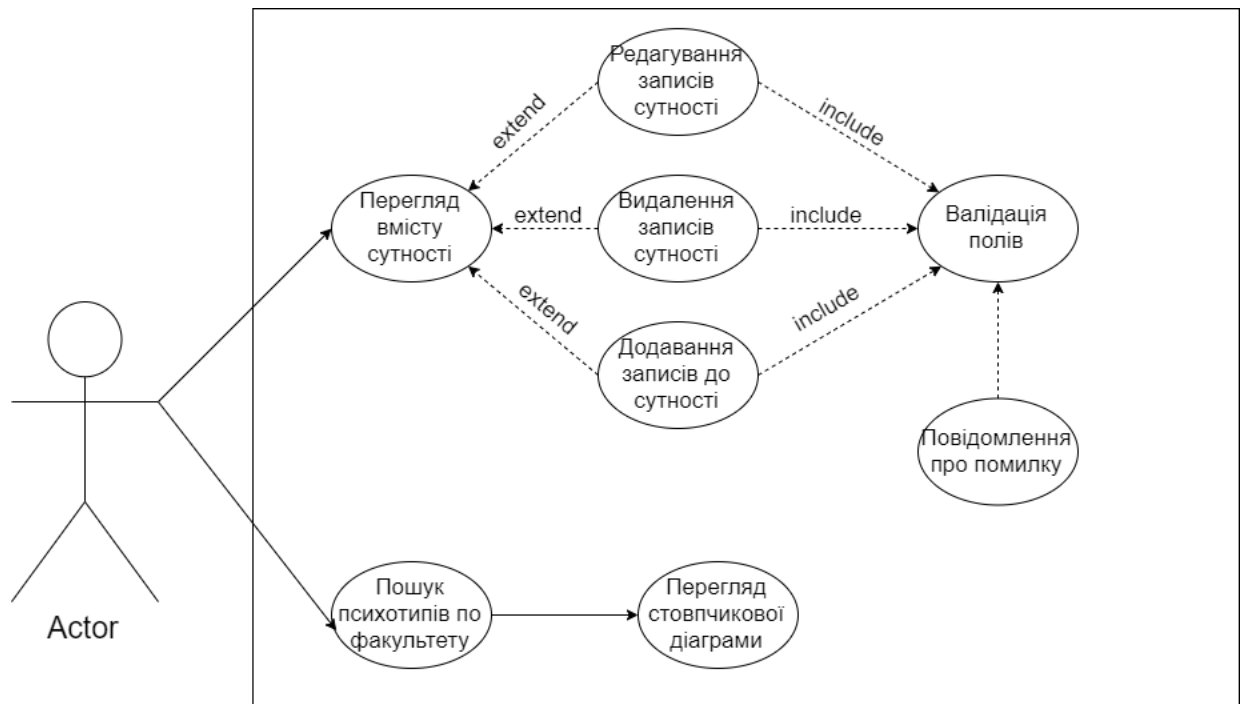


Рисунок 3.1 – Use-case diagram для веб-застосунку



## 4 ІНФОЛОГІЧНЕ ПРОЕКТУВАННЯ

Змоделюємо сутності та зв'язки звичною мовою.

а) Студент робить замовлення мелодії на радіо. Студент може робити багато побажань.

б) Мелодія має назву, виконавця та жанр, належить (майже завжди) до альбому. Мелодія входить до одного альбому та має один жанр. Мелодія виконується одним виконавцем, а виконавець має багато мелодій.

в) Альбом належить виконавцеві та містить більше однієї пісні.

г) Виконавець має багато альбомів.

д) Побажання належить одному студенту.

е) Жанр може належати різним мелодіям.

Таким чином, студента ідентифікують прізвище, ім'я, нікнейм, факультет та група. Виконавця можна визначити за псевдонімом. Мелодію та альбом – за назвою та виконавцем. Жанр визначається назвою. Побажання – датою та нікнеймом. Таким чином вводиться обмеження на одне замовлення на добу.

Концептуальна діаграма зображена на рисунку 4.1.

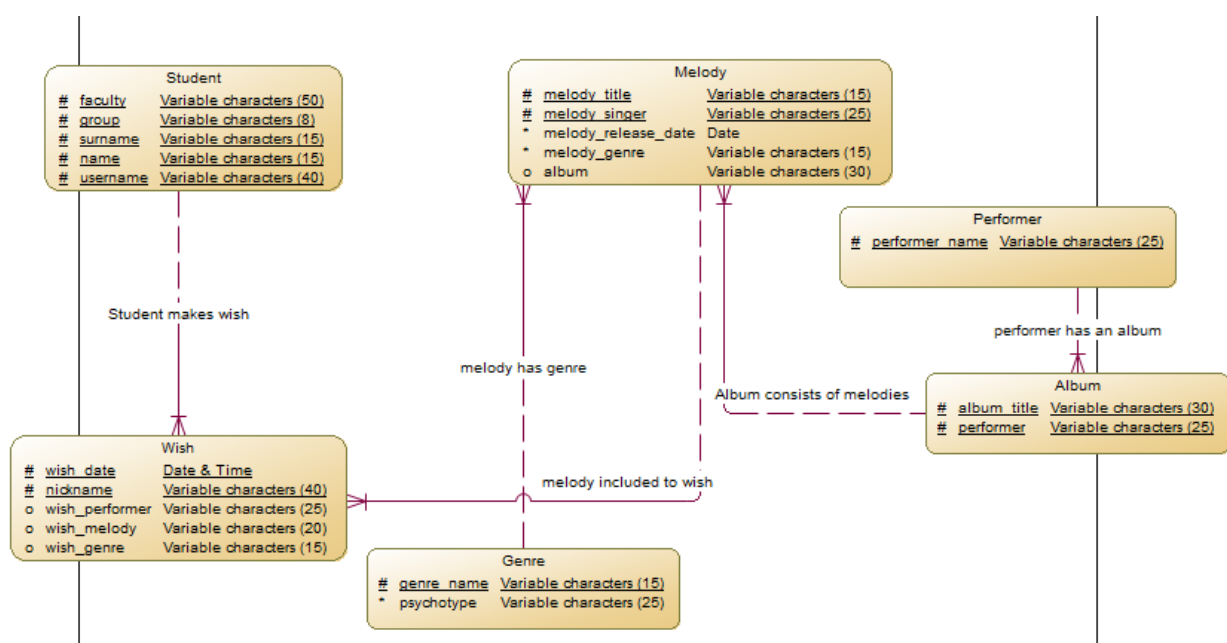


Рисунок 4.1 – Концептуальна діаграма бази даних

## 5 ДАТАЛОГІЧНЕ ПРОЕКТУВАННЯ

Наступний крок – перетворити концептуальну модель на логічну та оптимізувати її. При такому перетворенні вводять штучні ключі, а на атрибути, що були ключовими у концептуальній моделі накладається обмеження унікальності. Оптимізована логічна модель зображена на рисунку 5.1.

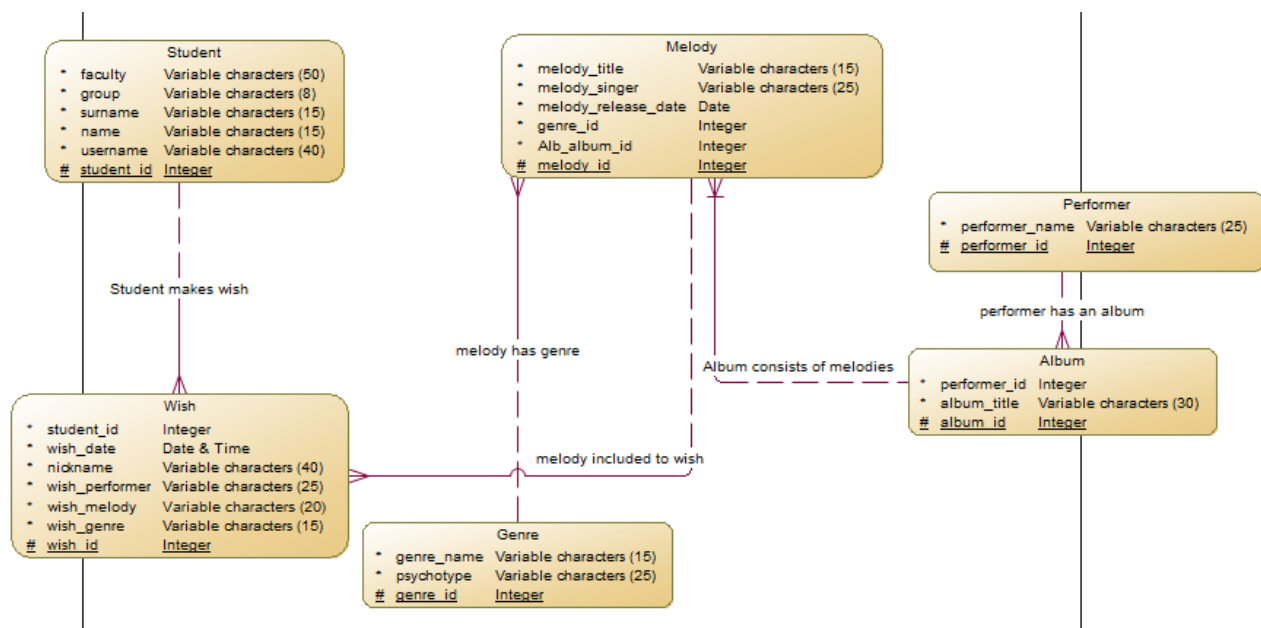


Рисунок 5.1 – Оптимізована логічна модель даних

Далі логічна модель перетворюється на фізичну. Для цього в інтерфейсі Power Designer обираємо PostgreSQL (в цій версії програми наявна версія 8). Формується модель, що зображена на рисунку 5.2. Разом із моделлю Power Designer надає код, в який залишається додати обмеження унікальності, що виникли на етапі оптимізації логічної моделі.

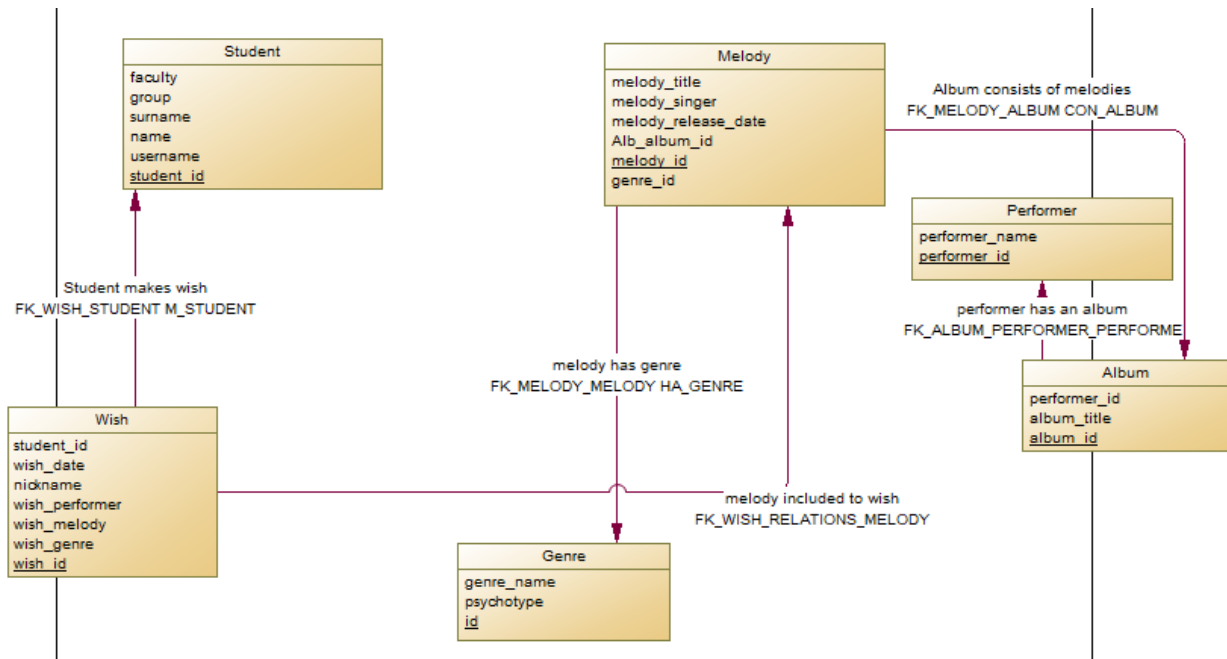


Рисунок 5.2 – Фізична модель даних

Таким чином, отримали модель даних, де кожна сутність містить штучні ключі. Штучні ключі в отриманому сервісі редагуватись користувачем не будуть: відбуватиметься автоінкремент у влаштованих процедурах. Тепер завдяки тому, що маємо штучні ключі, WTF-форми на сайті працюватимуть коректно, будуть наявні HiddenField().

Кожну таблицю моделі можна буде пов'язати з будь-якою іншою, виконавши перехресні записи із JOIN. Цілісність за посиланнями підтримуватиметься перехресними запитами з кожної форми, якщо у відповідній формі сутності є батьківська. Якщо батьківських сутностей немає – не обмежимо користувача у вводі інформації, окрім валідації полів.

## ВИСНОВКИ

В ході виконання курсової роботи було визначено наступні етапи впровадження проекту: побудова ієрархії процесів, обґрунтування постановки задачі, визначення сценаріїв перебігу елементарних процесів, визначення класів даних, сутностей та атрибутів, побудова моделей даних, інтеграція відомих рішень та їх покращення новим функціоналом.

Як результат виконання проекту досягнуто: інформаційна система, що дозволяє переглядати, додавати, іноді змінювати та видаляти зміст сутностей. База даних, що взята за основу – PostgreSQL, а на Python – SQLAlchemy. Веб-застосунок збудований на базі фреймворків Flask та Bootstrap-4. Ресурс розгорнутий на хостингу Heroku.

Окрема вкладка, зроблена на сайті, дозволяє переглядати розподіл студентів певного факультету за кластерами, визначаються вони психотипом, що поставлено у відповідність жанру мелодії.

У частині, що відповідає за машинне навчання застосовано Random Forest Regressor для визначення кореляції між виходом пісні у виконавця та його популярністю на радіо – дані взято не з власної бази даних, а з відкритих датасетів на ресурсі kaggle.com. Точність моделі – 97%. В залежності від кількості дерев прийняття рішень варіюється. При точності моделі 97% (200 дерев, 70% даних навчальні, 30% – тестові) середньоквадратичне відхилення становить 1.7. При зменшенні кількості навчальних даних точність може збільшуватись. Це обумовлене тим, що кореляція нерівномірна в рамках датасету.

Розроблена система безумовно потребує допрацювання. Для того, щоб далі з нею могло працювати більше людей, необхідно розробити авторизацію та різні ролі, а веб-інтерфейс для користувачів системи перенести у систему таку, як Telegram, зробити зручні кнопки та повну інтерактивність. А для режиму розробника залишити можливість виконання усіх CRUD-операції із веб-інтерфейсу. До збудованої бази даних можна робити запити складнішої структури, щоб діставати

коректні дані для побудови плейліста – поки що список побажань не може забезпечити повну відповідність мелодія-жанр-виконавець та не дозволяє залишати пустих полів. Також можна навчити модель пропонувати користувачам мелодію залежно від їх психотипу, незначною мірою при цьому змінивши задачу кластеризації.

## СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Create An API To Deploy Machine Learning Models Using Flask and Heroku [Електронний ресурс] – Режим доступу до ресурсу: <https://towardsdatascience.com/create-an-api-to-deploy-machine-learning-models-using-flask-and-heroku-67a011800c50>.
2. PostgreSQL 10.11 Documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://www.postgresql.org/docs/10/index.html>.
3. SQLAlchemy 1.3 Documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://docs.sqlalchemy.org/en/13/>.
4. WTForms Documentation [Електронний ресурс] – Режим доступу до ресурсу: <https://wtforms.readthedocs.io/en/stable/>.