DBS Data Scientist Mini-Project

You will be developing and serving a classification model for musical genres. This mini project will be used to gauge your aptitude of the core technical skills required to be a data scientist at DBS Bank.

Deliverables:

1) Git Repo link with all the codes and configurations required to deploy a containerized ML web service - Important to design for maintainability and extendibility.
    a. Detailed README.md containing instructions to run the service and screenshots of the key functionalities.
    b. Dockerfile for the web service and the database
    c. Tests for the data processing and application
    d. Basic documentation for the web application
    e. OpenAPI specification
2) Jupyter Notebook containing all your data exploration and experimentation steps taken to develop your final model
    a. While performance of your final prediction is important, more weight will be placed on your ability to explore the data, develop a structured approach to the problem and provide sound justification for choices made.
3) Prediction csv file containing predicted labels to the given unlabelled test.csv file.

Git Repo Details:

You may choose any public Git repo for this project.

The classification service should be in python (preferably 3.7) with type annotations. You can use any web framework (FastAPI is recommended) and minimally provide the following API:

- Classify input data and persist the results and titles into an embedded database (sqlite).
- Return a list of classified genres in the database.
- Return a list of titles to a provided genre.

Note: the classification and database service can be put into the same docker container.