Project Report: Music Genre Classification Task

1. The Task I Was Assigned

I was given the task of building a machine learning model to classify music genres from audio files using the GTZAN dataset. The core requirements involved processing audio data to extract features, training a multi-class classification model, and evaluating its performance. A key bonus objective was to explore both a traditional tabular approach and a more advanced image-based approach using a Convolutional Neural Network (CNN) and transfer learning.

2. The Problem and Its Solution

The Problem: Automatically classifying music genres is a fundamental challenge in audio analysis, with applications in music recommendation, library organization, and digital music services. The difficulty lies in extracting meaningful features from raw audio signals—complex, high-dimensional data—and building a model that can accurately distinguish between subtle differences in genres. This project addressed this challenge by exploring two distinct methodologies: one based on handcrafted audio features and another based on a deep learning model.

The Solution: This project solved the problem by developing and comparing two different machine learning pipelines:

- **Tabular Feature-Based Approach:** I extracted a set of well-established audio features (MFCCs, spectral centroid, etc.) to create a structured, tabular dataset. I then trained a **Random Forest Classifier** on this data. This method relies on the effectiveness of human-engineered features to capture the essence of a music genre.
- **Image-Based (Transfer Learning) Approach:** I converted the audio signals into Mel spectrograms, which are essentially images. I then used a powerful pre-trained **VGG16 model** as a base for a new classifier, leveraging the knowledge from a large image dataset to classify music genres.

By comparing the performance of these two approaches, the project identified the most effective strategy for music genre classification with this dataset.

3. How I Approached the Task

I followed a structured, multi-stage process to complete the assignment.

Tabular Approach: From Audio to Classification

- **Data Preparation:** I first extracted 20 MFCCs along with other key audio features (spectral centroid, chroma, etc.) from all 1,000 audio files in the GTZAN dataset. This process created a single, comprehensive extracted_features.csv file. The data was then split into training and testing sets, with the numerical features scaled using StandardScaler and the genre labels encoded.

- **Model Training & Evaluation:** I trained a **Random Forest Classifier** on the prepared tabular data. The model achieved a **67.00% accuracy** and a detailed classification report showed it performed well on genres like classical and pop but struggled with rock and hiphop.

Image-Based Approach: Spectrograms to CNN

- **Data Preparation:** I converted each 30-second audio file into a Mel spectrogram image. I then split these images into training and validation sets, preparing them for a CNN model.
- **Model Training & Evaluation:** I implemented a **Transfer Learning model** using **VGG16**, a pre-trained CNN. The model was trained on the spectrogram images and achieved a **64.50% validation accuracy**. This was a significant improvement over a simple CNN from scratch, but it was still slightly below the performance of my tabular model.

Final Comparison and Insights

- **My Goal:** To compare the best models from both approaches and select a final best model.
- **What I Did:** I compared the final accuracy of the **Tabular (Random Forest)** model (67.00%) with the **Transfer Learning (VGG16)** model (64.50%). I also considered the performance of the CNN from scratch (47.00% accuracy).
- **The Result:** The Tabular model performed best, with a slight edge over the Transfer Learning model. The CNN from scratch was the worst performer.
- **My Conclusion:** The Tabular model, using well-engineered features, proved to be the most effective for this dataset. While transfer learning showed great promise and was a significant improvement over a simple CNN, the handcrafted features provided a more robust representation of the audio for this specific classification task.

4. Final Result of My Work

After completing all assigned tasks and bonus objectives, my final conclusion is that the **Tabular model, powered by a Random Forest Classifier and handcrafted audio features, is the most effective solution** for this Music Genre Classification task. It achieved a final accuracy of **67.00%**, outperforming the more complex image-based approach in my evaluation.

This project successfully demonstrated the effectiveness of both feature engineering and transfer learning for audio data. The results highlight that the choice of approach is highly dependent on the problem and dataset characteristics, and sometimes, a simpler, well-designed model can outperform a more complex one. The project provides a solid foundation for building practical and accurate audio classification systems.