

How I Annotate Music & Metadata Datasets (Semi-Automatic + Manual)

This document summarizes my workflow for **Automatic & Semi-Automatic Music Data Labeling**, combining Python-based feature extraction with expert manual annotation on **AI-generated audio** and **MAESTRO dataset (MIDI+WAV)**.

1) Tools & Environment

- **IDE:** Visual Studio Code (VS Code)
- **Python:** 3.11 with **venv**
- **Libraries:** librosa, music21, numpy, pandas, soundfile
- **Dependencies:** pip freeze > requirements.txt

2) Folder Structure

```
week06_auto_labeling/
  src/
    data/                                # AI-generated audio
    data_external/maestro_subset/ # MAESTRO MIDI+WAV subset + CREDITS.md
    csv/
    reports/
    guides/
    venv/
    CREDITS.md
    README.md
```

3) Workflow Overview

1. **Automatic extraction (audio):**
 - **BPM** (librosa.beat)
 - **Key** (librosa + chroma) → **confidence auto-computed** : the confidence field refers exclusively to the reliability of automatic key estimation and does not represent a global confidence score for the track. The automatic confidence score ranges from 0 to 1.
 - **MFCC** (13-D, mean per file) : MFCC features are extracted exclusively from audio files (WAV/MP3).
2. **Automatic extraction (MIDI):**
 - **Key + Instruments** via **music21** (MAESTRO MIDI tracks)
3. **Source auto-labeling:**
 - source inferred from folder: data/ → “AI_generated”;
data_external/maestro_subset/ → “MAESTRO”
4. **Manual labeling:** title, source/model, duration, audio_link, bpm_manual, key_manual, instruments_manual, genre, mood, usage_context, notes.
5. **Merge & QA:** unify columns (manual/auto), compare auto vs manual, log edge cases in /reports/

4) CSV Schema (“labels_core.csv”)

```
title, source, source/model, duration, audio_link, bpm_manual, bpm_est,  
key_manual, key_est, instruments_manual, instruments_auto, instruments_merged,  
genre, mood, confidence, error, usage_context, notes
```

- **instruments_merged** = merged column combining **manual (AI audio)** and **automatic (MAESTRO MIDI)**
- **source** = **auto-generated** from directory (AI_generated / MAESTRO)
- **confidence** = **auto-computed** by key estimation script. It does not express overall musical quality, correctness of BPM detection, or artistic evaluation. The automatic confidence score ranges from 0 to 1.
- **error** = only filled when an extraction error occurred; otherwise blank/“None”

5) Other CSV files (e.g., “ai_tracks_auto_bpm.csv”, etc.)

Automatic BPM, key, and MFCC features were extracted programmatically and stored in separate CSV files (MFCC not included in the core labels schema). Instrument information was extracted automatically only for MIDI files; for audio tracks, instrumentation was labeled manually.

. Summary:

- **Manual fields** (filled by listening / reasoning):

```
title, source/model, duration, audio_link, bpm_manual, key_manual,  
instruments_manual, genre, mood, usage_context, notes.
```

- **Automatic fields** (filled by scripts):

```
source, bpm_est, key_est, instruments_auto, instruments (merged),  
confidence, error, mfcc_1 ... mfcc_13.
```

6) Method Details

- **BPM:** librosa.beat.beat_track; manual cross-check for rubato/double-time; notes saved in /reports/
- **Key (audio):** librosa key estimation + **confidence**; manual verification in ambiguous cases (relative modes for example)
- **Key/Instr (MIDI):** music21 for MAESTRO midi tracks → reliable **Piano** detection and symbolic key
- **MFCC:** librosa.feature.mfcc(y, sr, n_mfcc=13) → mean across frames (timbre fingerprint). MFCC features are extracted exclusively from audio files (WAV/MP3). MIDI files do not contain spectral audio information, therefore MFCC fields are intentionally not applicable for MIDI rows.
- **Merge:** prefer **manual** instruments for AI audio; use **auto** instruments for MAESTRO MIDI

7) Results (concise)

- **BPM:** mostly accurate but with several **double-time or some pulse-misaligned** readings (e.g., **100→198.8; 50→99.4; 40→120 BPM**). **Expressive rubato** and **non-percussive textures** reduced beat stability.
- **Key:** mostly accurate; **Track 6** overall **G_b major** with **E_b minor** sections (auto **E_b minor** coherent); **Track 7** overall **G major** with auto **E minor** (relative).
- **Instruments:** MAESTRO MIDI → **Piano** correct in 2/3; one auto mislabel (“Denisova11-15.MID”) **corrected manually to Piano**.
- **Source & Confidence:** source auto-labeled from folder; **key confidence** auto-computed and recorded.

8) Outcome

A clean, employer-ready pipeline that fuses **symbolic (MIDI)** and **audio (WAV/MP3)** data with **auto features + manual QA**, aligned with roles like **AI Music Annotator, Music Data Specialist, Prompt Engineer (Music)**.