To Bach or not to Bach - classification of music pieces based on the MIDI files

Team_1:

Emilia Wiśniewska,
Hanna Laszkiewicz,
Kalina Białek,
Mikołaj Fornal

Supervisor:

Małgorzata Biernacka

Classification of music

Motivation:

In music school we (well, at least some of us) had to learn to distinguish between composers by listening to their pieces. We want to teach a machine to do something similar – classify the music to composers based on MIDI file.

Data:

- MIDI files with labels including the composer of the music
- From MIDI files we extracted chosen features

Additionally:

- There are some papers on that, but we didn't find any musicians-models comparison.

What is a MIDI file?

- MIDI (Musical Instrument Digital Interface) is a digital format used to connect musical instruments, computers, and other devices.
- It doesn't record sound directly like a microphone; instead, it saves instructions for playing music.
- These instructions include notes, their duration, volume, and the type of instrument to use (like piano, guitar, or drums).

MIDI editor vs. Text editor

Fighting with the dataset

- A bunch of smaller datasets of MIDI files.
- 50 composers from 2 to 1000+ files each.
- We chose 10 popular composers to make it more reasonable.
- Mission:
  - delete duplicates (manually check what we have and figure out what 07456b_.mid means),
  - replace some broken files,
  - repair some more files (fix tempo, delete empty measures etc.),
  - download missing files (for every composer to have around 100 files),
  - choose a reasonable subset from more than 1000 files for the one and only - Bach.
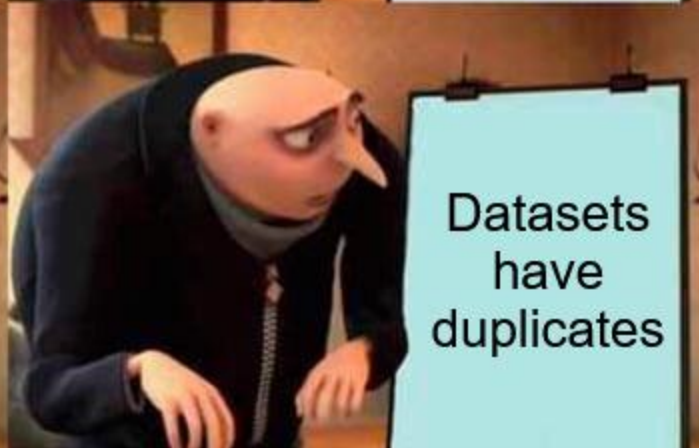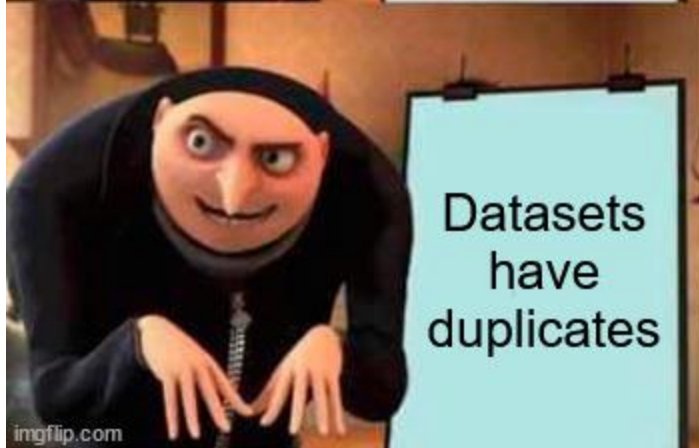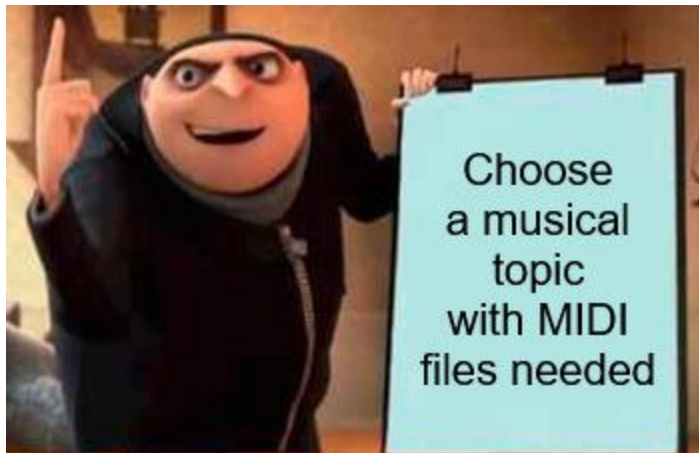
Fighting with the dataset pt. 2

We did some feature extraction:

- pitches
- velocity
- key signature
- tempo, rubato
- instruments (well, about that…)

Instruments

1st try: names of instruments.Instrument:

- irregular names :(

2nd try: programs of instrument.Instrument

- programs 0 - 127 :)
- instruments without programs :(
- no instruments in the music piece :(

Last version: getInstruments()

- .Instrument **and** .Piano, …
- not perfect :(

'Track 1: '

'Strings1'

':'

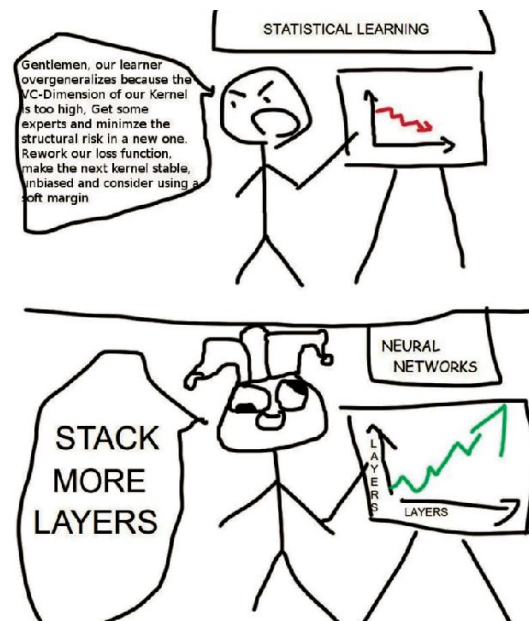'Bach Toccata & Fugue: '

'' 'L.H.: '

Methods and the results

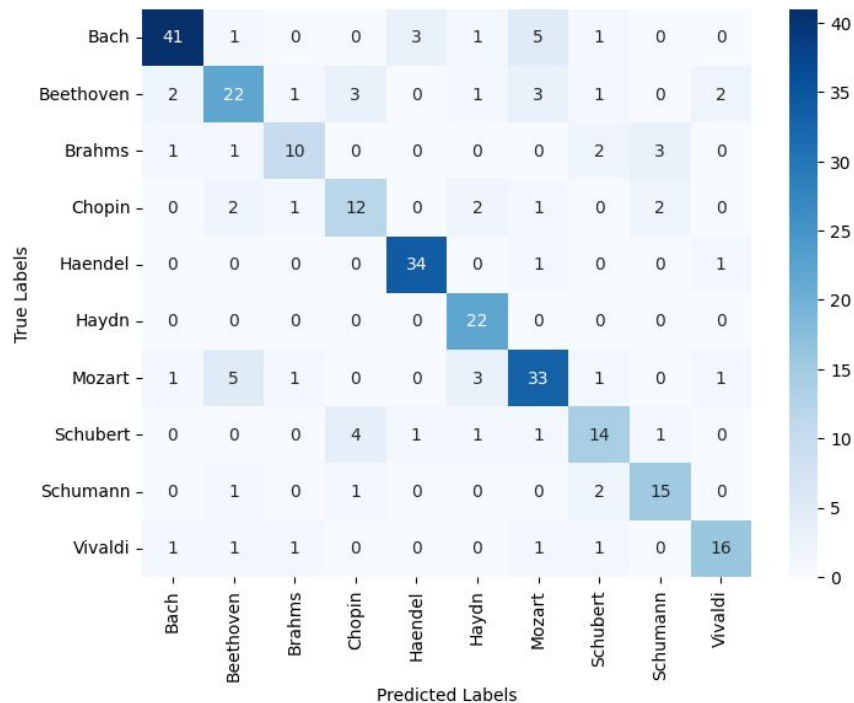| Method	Accuracy	Weighted F1-score
KNN	60%	60%
Logistic Regression (1 vs 1)	69%	69%
Logistic Regression (1 vs rest)	65%	66%
Decision Trees	57%	56%
Random Forests	71%	71%
Gradient Boosting (sklearn)	60%	58%
Gradient Boosting (xgboost)	76%	77%
Support Vector Machines	59%	57%
Neural Networks	76%	76%

Confusion matrices for 2 best methods

eXtreme Gradient Boosting
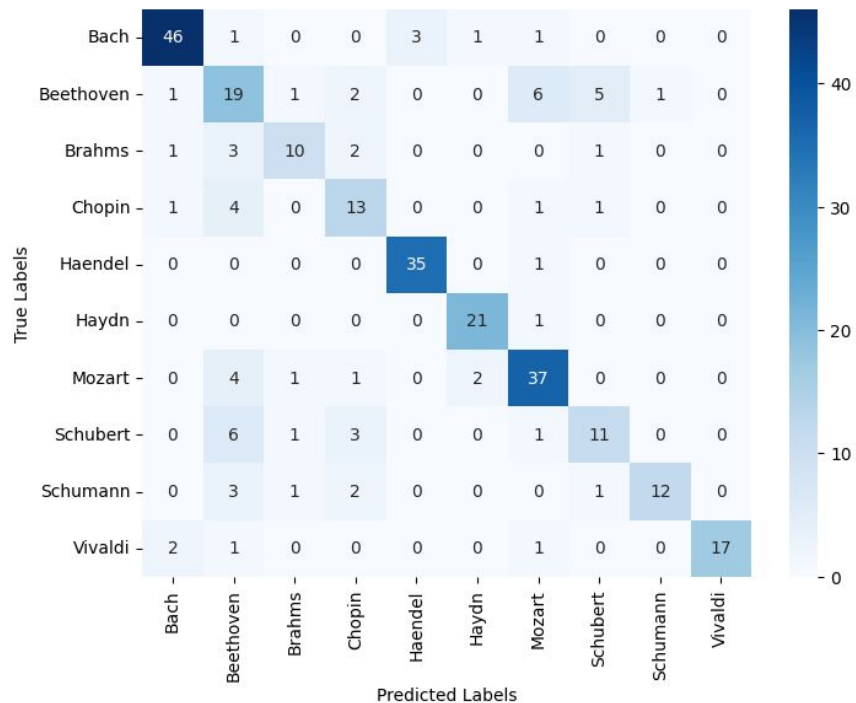
Neural Network

What about musicians?

1. **Ludwig van Beethoven** – *Rage over a Lost Penny, Op. 129*
2. **Wolfgang Amadeus Mozart** – *Piano Concerto No. 25, K503, mvt. 2*
3. **Robert Schumann** – *Arabesque, Op. 18*
4. **Fryderyk Chopin** – *Tarantella, Op. 43*
5. **Franz Schubert** – *Piano Sonata in A Major, D959, mvt. 1, Allegro*
6. **Fryderyk Chopin** – *Nocturne in G Minor, Op. 37, No. 1*
7. **Johannes Brahms** – *Academic Festival Overture, Op. 80*
8. **George Frideric Handel** – *Messiah, No. 14: Behold the Lamb of God*
9. **Ludwig van Beethoven** – *Piano Sonata No. 29 in B-flat Major, "Hammerklavier," mvt. 1*
10. **Antonio Vivaldi** – *Concerto for 2 Violins in A Minor, Op. 3, No. 8, RV578, mvt. 1*
11. **Robert Schumann** – *Fugue op. 126, No. 5*
12. **Johann Sebastian Bach** – *Herzliebster Jesu, from the St. John Passion, BWV 245*
13. **Joseph Haydn** – *Symphony No. 9 in C Major, Hob.I:9*
14. **Franz Schubert** – *Frühlingsglaube, D686*
15. **Ludwig van Beethoven** – *Lieder Op. 99: "Der Mann von Wort"*
16. **Johann Sebastian Bach** – *Fugue in G Minor, BWV 006/3*
17. **Johannes Brahms** – *Scherzo in E-flat Minor, Op. 4*

The results

Musicians tend to say that:

- Brahms is Beethoven
- Beethoven is Mozart
- Händel is Vivaldi
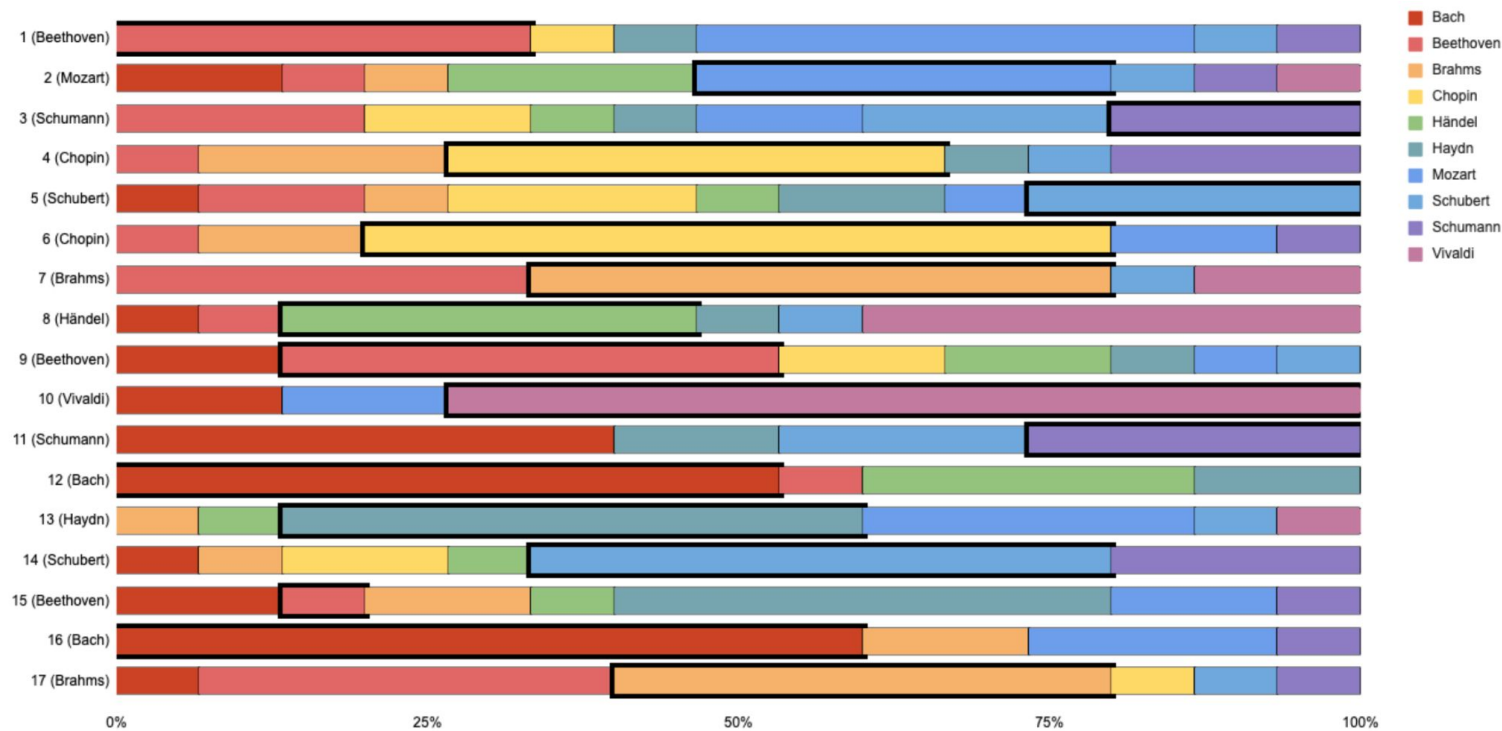- Bach is Händel
- Vivaldi is Bach

Models tend to say that:

- Brahms is Beethoven
- Beethoven is Mozart
- Bach is Händel

As expected, the composers that musicians confuse the most were composing in the same or similar time periods.

Our models tend to confuse composers similarly, but they are a bit better :)

Musicians and their results

Some observations on the quiz

	Musicians	Models
Most problematic composer	Schumann (23,33%)	Brahms (33,33%)
Least problematic composer	Vivaldi (73,33%)	Vivaldi **(100%)**
Most problematic pieces	Beethoven's *Lieder*	Brahms' *Scherzo* and *Overture*
Favourite composer	Bach	Schumann
Best result	76,47%	88% (NN)
Average accuracy	40,39%	68,29%

The end