Music Genre Classification with Machine Learning

W207 Spring 2024

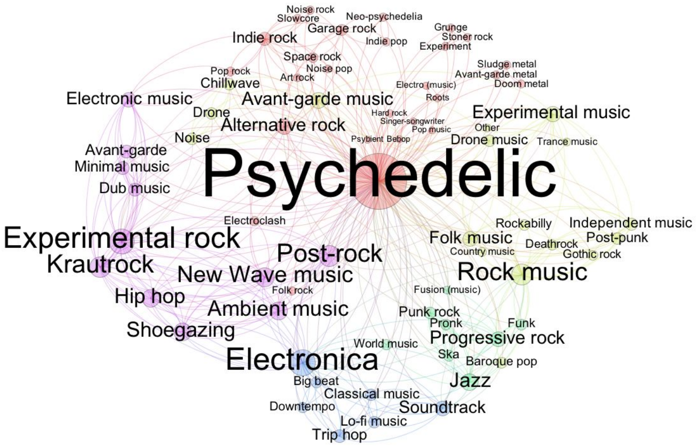
Nicholas Lin, Rex Gao, Yoni Nackash

Berkeley
UNIVERSITY OF CALIFORNIA

Problem Statement & Motivation

Our objective is to develop a machine-learning based classifier for classifying musical genres using audio characteristics.

Can be used in both track recommendation system for users and automated tagging in the track database



Berkeley
UNIVERSITY OF CALIFORNIA

Dataset & Variables

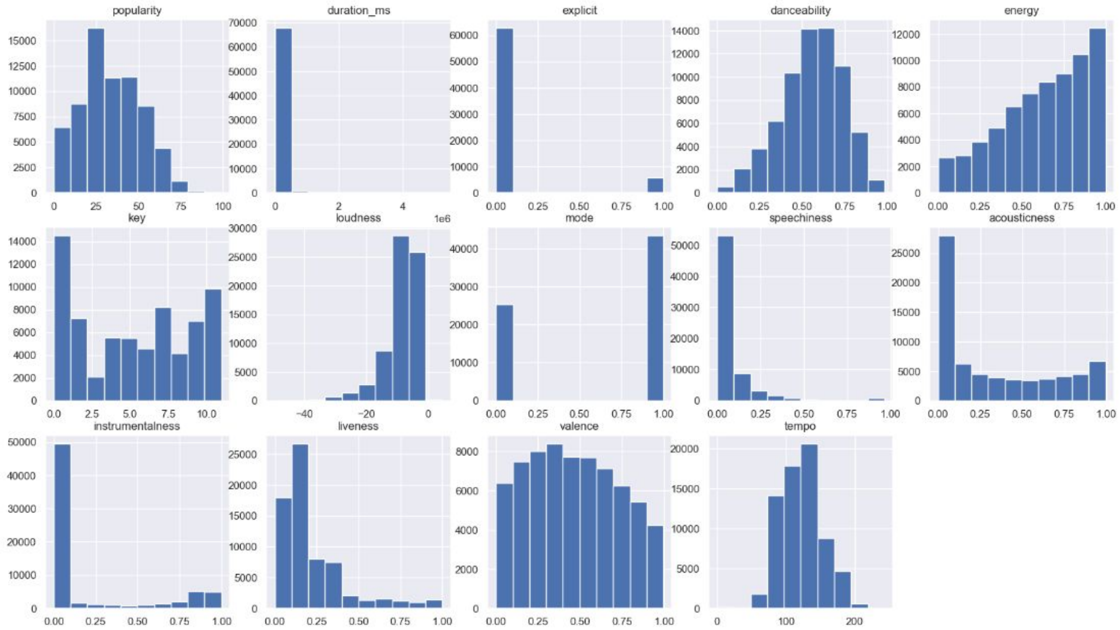
Kaggle dataset about music tracks from Spotify's web API

Contains 114 genres with 1000 tracks each for 114,000 total tracks

- Categorical variables: track_id, key, time_signature, track_genre
  - Text variables: artists, album_name, track_name
- Quantitative variables: popularity, duration_ms, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo
- Binary variables: mode and explicit

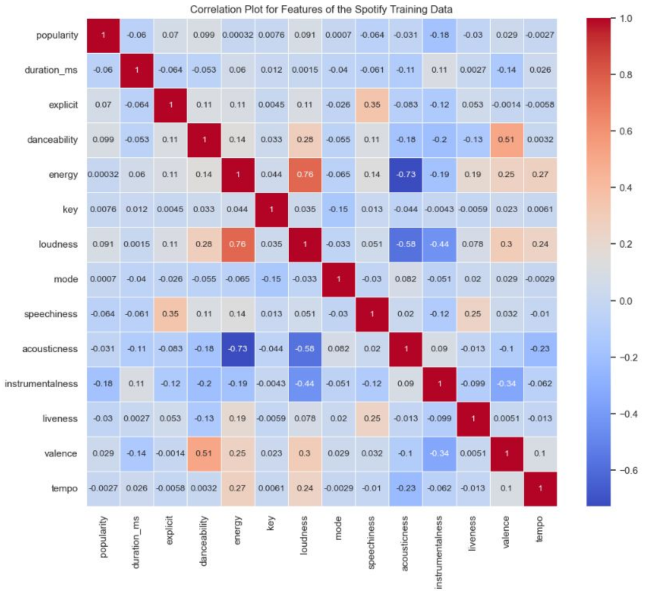
EDA & Initial Feature Engineering

- Most features are on varying scales
  - Normalization needed
- Binarized explicit and mode
- One hot encoded key
- Small sample of songs longer than the average song
- Binned Popularity & Duration
- Processed data: 112 genres



EDA & Initial Feature Engineering

- No redundant features extremely correlated with each other (~1.0)
- Notable Correlations
  - Acousticness & Energy (-0.73)
  - Acousticness & Loudness (-0.58)
  - Loudness & Energy (0.76)
  - Valence & Danceability (0.51)
  - Speechiness & Explicit features have a weak positive correlation - maybe a good indicator for “Rap”

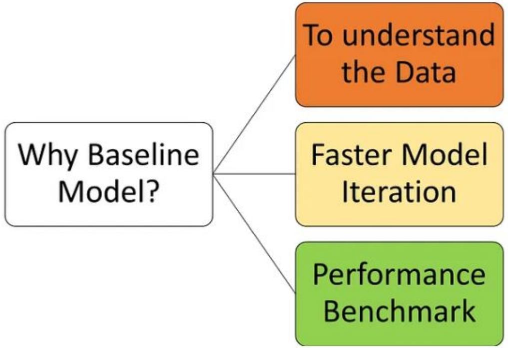


Berkeley
UNIVERSITY OF CALIFORNIA

Baseline Model

- Uneven class distribution
- Using a DummyClassifier
  - Stratified Strategy
  - Accuracy = ~1%
  - Highest precision for a genre = ~3%

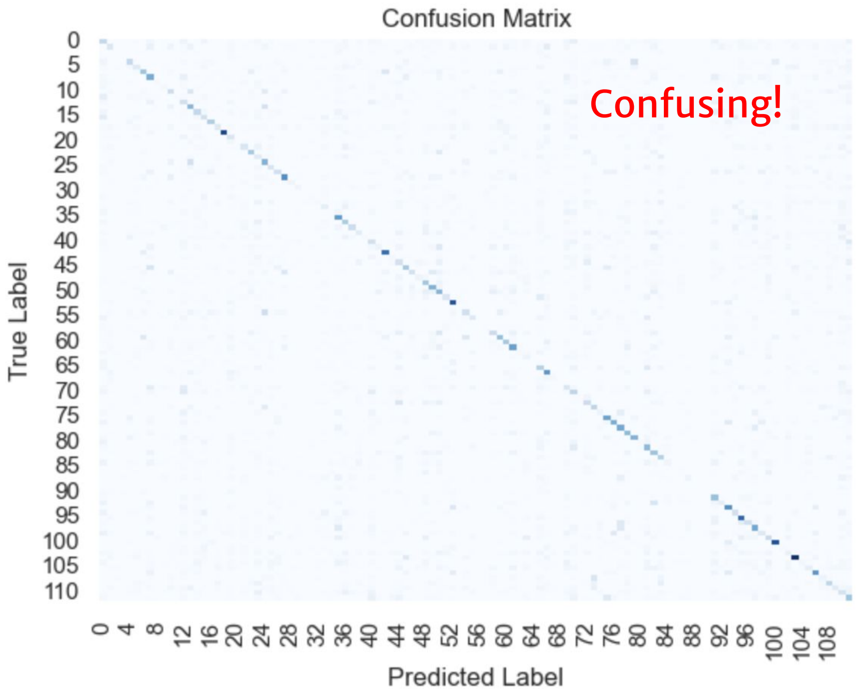
- Highlights the room for improvement



KNN

31%
Testing Accuracy

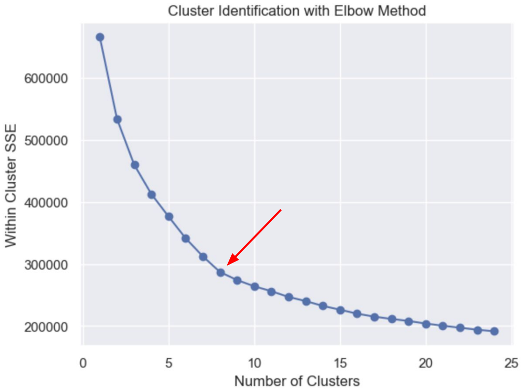
- One of the simplest multi-class classifiers
- No training step
- 31% accuracy with 32 nearest neighbors weighted using inverted Manhattan distance

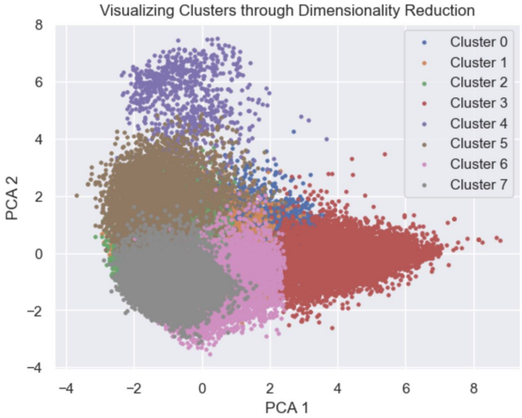


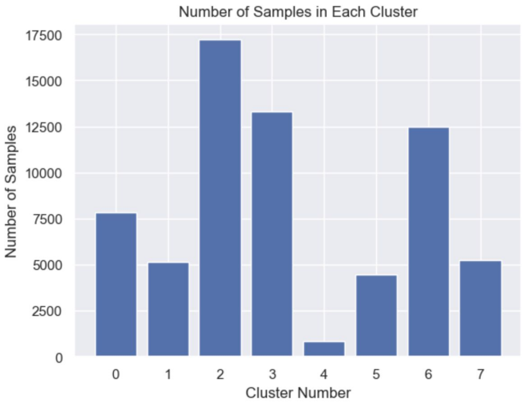
Berkeley
UNIVERSITY OF CALIFORNIA

K-Means Clustering

- Subjectivity of music labeling – songs can have many influences and styles
- Genre overlap that our model would not be able to identify
    - Pop vs K-Pop vs J-Pop vs Mandopop
- Possible solution → **higher-level genre categorization**
- First categorization experiment: K-Means Clustering. Generate mutually exclusive clusters of similar genres



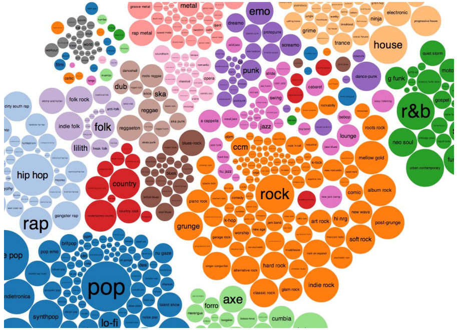




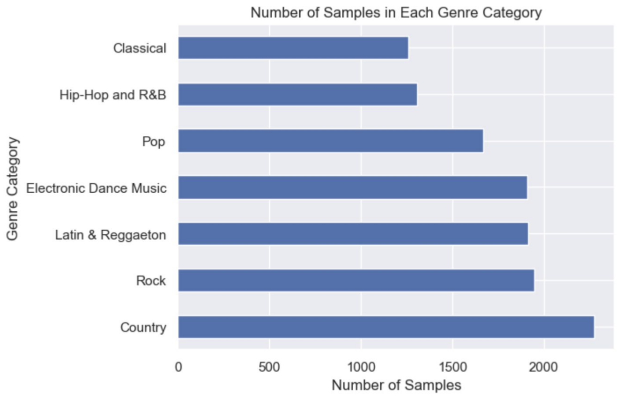
Berkeley
UNIVERSITY OF CALIFORNIA

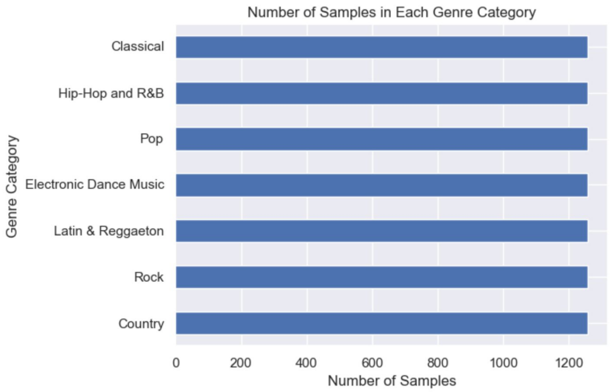
Manual Genre Categorization

- Second Experiment: Manual genre categorization
- Through investigation and research, most prolific genres:
  - Rock, Pop, Electronic Dance Music, Hip-Hop/R&B, Country, Classical, and Latin & Reggaeton
- Mapped a subset of genres from our data to these categories
- Rebalanced so genre categories contained equal number of samples



https://musicmachinery.com/2013/09/22/5025/





Berkeley
UNIVERSITY OF CALIFORNIA

KNN Revisited

72%
Testing Accuracy

- 72% accuracy with 12 nearest neighbors weighted on inverted Manhattan distance

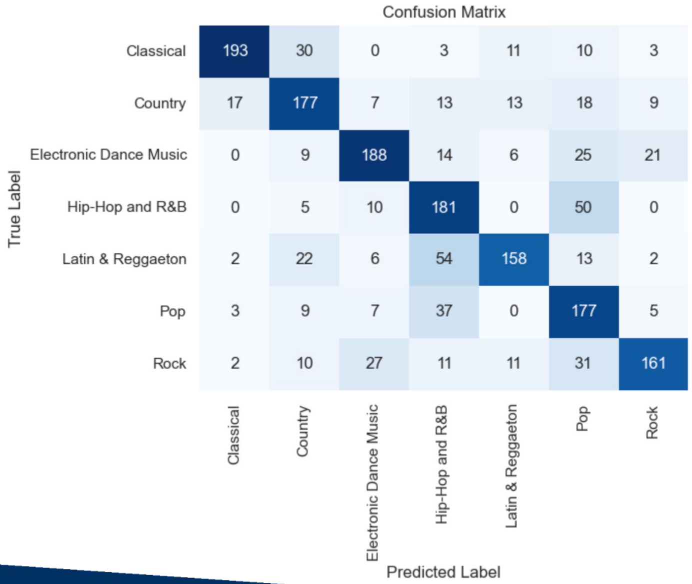
Confusion Matrix

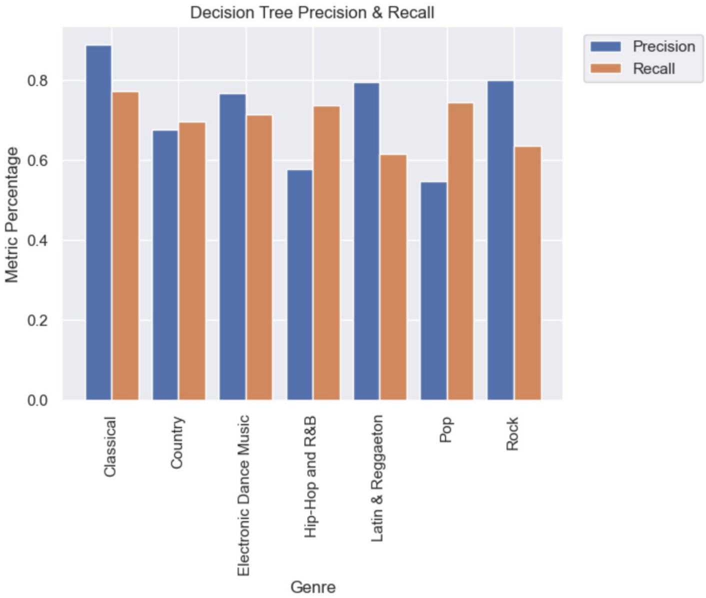
True Label \ Predicted Label	Classical	Country	Electronic Dance Music	Hip-Hop and R&B	Latin & Reggaeton	Pop	Rock
Classical	217	21	1	2	11	12	4
Country	20	182	6	2	14	7	11
Electronic Dance Music	0	8	165	16	4	16	20
Hip-Hop and R&B	0	3	8	182	7	55	9
Latin & Reggaeton	1	18	5	15	180	11	8
Pop	6	14	16	19	7	186	10
Rock	2	17	32	14	16	31	151

Decision Trees

70%
Validation Accuracy

Validation accuracy of ~70% through decision forest of 17 estimators and max depth of 8.



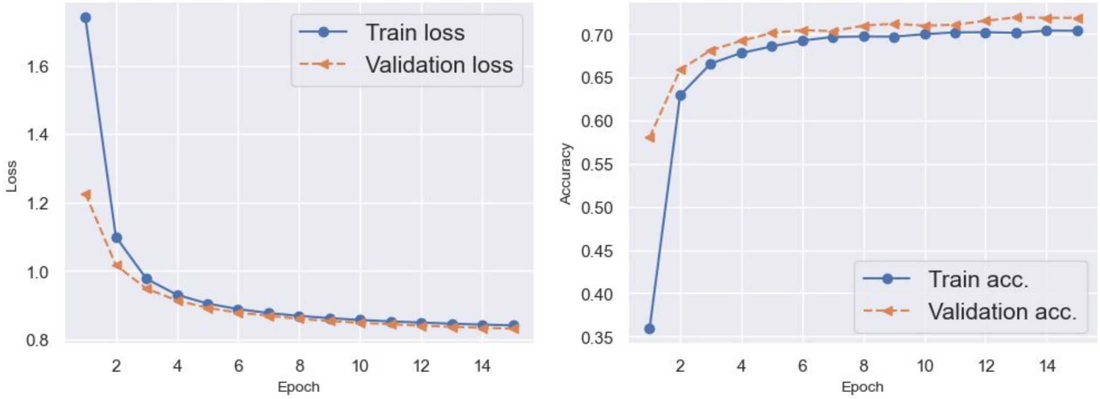


Berkeley
UNIVERSITY OF CALIFORNIA

Neural Network – Single Layer

72%
Validation Accuracy

- Baseline Neural Network Model
- Input Layer - Output layer
- Room for hyperparameter tuning

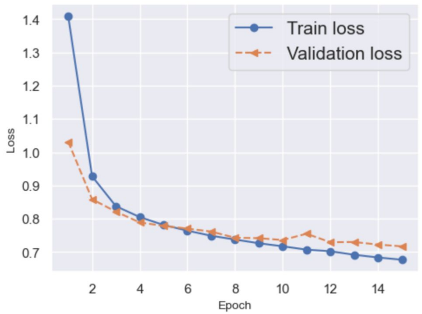


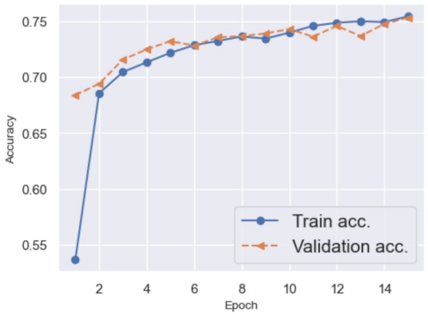
Berkeley
UNIVERSITY OF CALIFORNIA

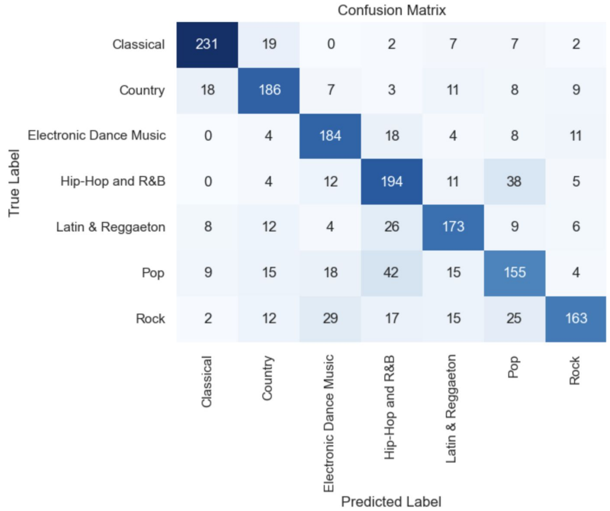
Neural Network – Multiple Layers

75%
Validation Accuracy

- Experimented with additional hidden layers → best performance with two hidden layers: 256 + 128 nodes
- Hyperparameter tuning → adjusted learning rate (0.1) and number of epochs (15)
- Next steps → continue to explore confusion matrix







Conclusion

- With accuracies of 70-75% in all of our approaches, we suspect we reached an upper limit on the insight/information found in the data set.
    - Musical genres can overlap and influence one another in the real world
- Future exploration:
    - Language tagging
    - More in-depth investigation on genre categories
    - Potential still available for raw genres

Berkeley
UNIVERSITY OF CALIFORNIA

Thank You

Contributions

- **Data Preprocessing**: Nick, Rex, Yoni
- **EDA**: Nick, Rex, Yoni
- **Feature Engineering**: Nick, Rex, Yoni
- **Baseline Model**: Nick
- **K-nearest Neighbors**: Rex
- **K-means Clustering**: Yoni, Rex
- **Research & Categorization**: Nick, Rex, Yoni
- **Decision Trees**: Yoni
- **Neural Networks**: Nick, Yoni

References

- Dataset: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset
- https://rateyourmusic.com/genres/
- https://blog.novecore.com/top-10-most-popular-music-genres/
- https://www.statista.com/chart/30575/share-of-us-respondents-that-listens-to-different-music-genres/
- https://mastersofmedia.hum.uva.nl/blog/2011/04/26/visualising-music-the-problems-with-genre-classification/
- https://musicmachinery.com/2013/09/22/5025/

Berkeley
UNIVERSITY OF CALIFORNIA

NeurIPS Checklist

1. **Author Questions**
   a. Yes – the introduction of our notebook, as well as the introduction slide of this presentation, mention the scope and contributions.
   b. Yes – we have read the ethics review guidelines, and ensure our work conforms to them.
   c. Yes – our team discussed the issue of potential cultural appropriation - we wanted to avoid accidentally grouping certain genres that we may not be fully familiar with into higher level categories that may not accurately represent the underlying music. For the purposes of this project, we chose to include a subset of the existing genres that we are knowledgeable on.
   d. Yes – our notebook, as well as this presentation, includes mentions of limitations.
2. **Theoretical Result Questions** - N/A for all questions
3. **Experimental Questions**
   a. Yes – all code, data, and instructions are linked in: (Github - notebook)
   b. Yes – training details are mentioned in both slides and our notebook
   c. N/A – testing statistical significance not included
   d. Yes – experiments were run using our local machines: One 14-core CPU, 40GB memory, 1 GPU
4. **Asset Questions**
   a. Yes – citations included in notebook as well as References slide
   b. Yes – license: Open Data Commons, mentioned in our notebook
   c. No – Open Data Commons license is open for public use
   d. Yes - License: Open Data Commons, mentioned in our notebook
   e. Yes - Data doesn't include personally identifiable information. It is possible that certain track / artist / album names contain explicit content, however we dropped all textual fields from our dataset, except genre.
5. **Research with Human Subject Questions:** N/A for all questions

Code Submission

Link to [GitHub repo]