ROBACH -
MUSIC CLASSIFICATION

David Ristau. Kieran Berton, Natali Ojeda Meneses, Ziling Huang

01 MOTIVATION

RESEARCH QUESTION

Can we accurately classify instrument ensembles from audio?

EVALUATION CRITERION

- F1 Score
- Incorporates prediction error in imbalanced datasets

Research Importance

- Challenging representation of features
- Proof of Concept Testing for Spectral Features on Classical Music ensemble prediction

01 MOTIVATION

BUSINESS VALUE

Audio Classification with Spectral Features

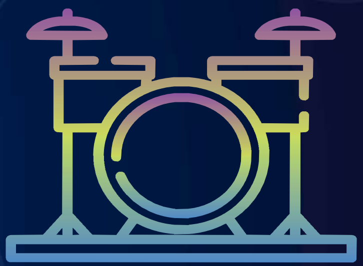
- Identify instrument ensembles from Audio Spectral Features (Shazam)
- Recommending similar songs to listeners based on preference to enhance user experience (Spotify, Amazon, Apple)

Audio Classification with MIDI Features

- Helpful for academic learning
- Labelled data might not be available in practice.

02 LITERATURE RELATED TO INSTRUMENT IDENTIFICATION

- **Commonly used features:**
  - Spectral (relating to frequencies in a given time period of music)
  - Non-spectral
- **Instruments classified:**
  - String, brass, vocals, keyboard, winds, organ etc...
- **Algorithms used:**
  - Naive Bayes, Logistic regression, Neural Nets, SVM, Random Forest
- **Successes:**
  - Some papers identify a 100% accuracy, with limited scope
  - Reasonable accuracy scores ~ 55% - 65%
- **Challenges:**
  - Percussive sounds are challenging to categorize and are essentially left out of all research in this field.
  - Only harmonic instruments are incorporated in the research of instrument classification



EXTRACTION, SEPARATION & CLASSIFICATION

03 OUR CONTRIBUTION

SIZE

Large dataset of 330 classical movements spanning 34 hours

REPLICATION STUDY

Replicating previous studies using new and different types of data

MULTI-INSTRUMENT

Instrument ensembles are used

04 OUR DATA

MIDI FILE DATA

Sheet Music

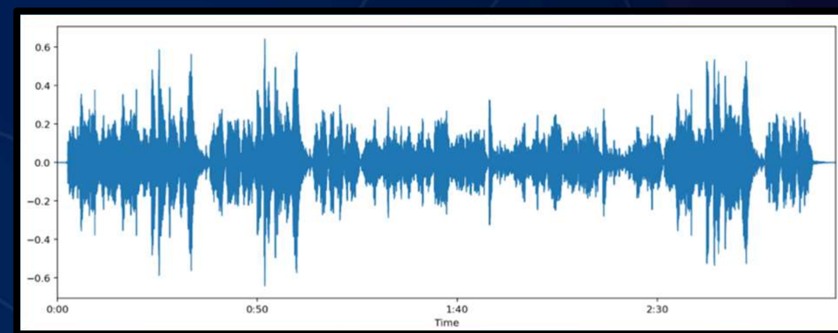
Piano Roll

Tabular Format

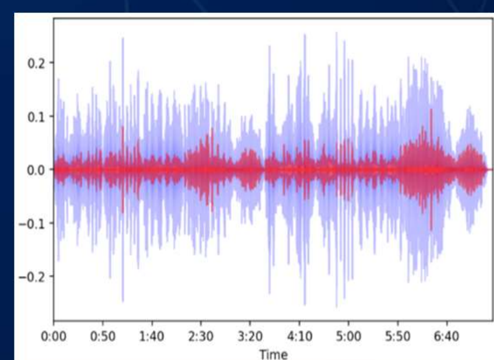
Start Time	End Time	Instrument	Note	Note Value	Piece ID
4062	20446	41	71	Quarter	2191
4062	20446	41	64	Quarter	2191
20446	35294	41	73	Quarter	2191
20446	35294	41	69	Quarter	2191
35294	45534	41	69	Eighth	2191

SPECTRAL DATA

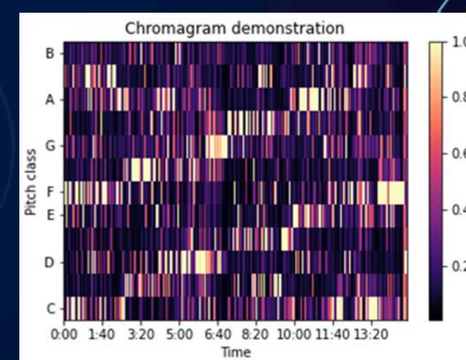
Amplitude over Time



Harmonic vs. Percussive



Pitch over Time



Source: MusicNet Data

04 OUR DATA - SOURCES AND PREPROCESSING

MIDI DATA

Number of Unique Instruments	Number of Unique Notes	Number of Notes Total	Minimum Note	Second Quntile Note	Median Note	Fourth Quintile Note	Maximum Note	Average Number of Notes per Instrument	Seconds
5	75	413,845	24	57	63	70	100	82,769	447
5	63	378,694	31	62	69	78	98	75,739	251
5	71	500,871	26	57	66	76	100	100,174	444
1	70	145,296	30	51	67	78	100	145,296	546

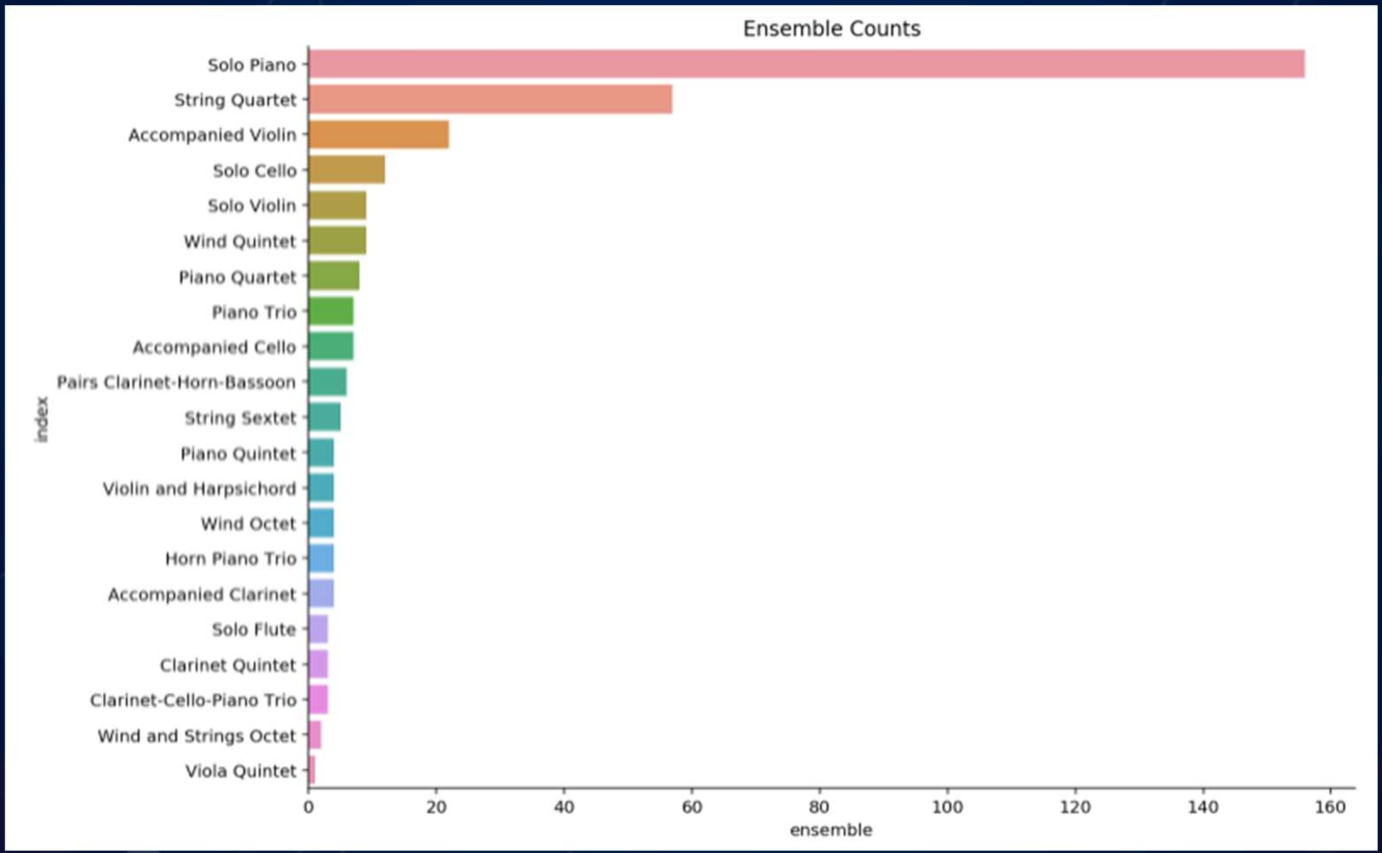
SPECTRAL DATA

Harmonic or Percussive	Mel Frequency Cepstral Coefficient 0	Mel Frequency Cepstral Coefficient 1	Mel Frequency Cepstral Coefficient 2	Spectral Feature 0	Spectral Feature 1	Spectral Feature 2	Chroma Feature 0	Chroma Feature 1	Chroma Feature 2	Contrast Chroma Feature 0	Contrast Chroma Feature 1	Seconds
1	-251.72	196.22	-10.74	0.0019	0.0102	0.0516	0.1286	0.1355	0.3570	0.1286	0.1355	124
1	-391.63	185.80	-1.77	0.0926	0.0286	0.1323	0.1889	0.1000	0.2731	0.1889	0.1000	313
1	-403.81	160.52	16.72	0.0182	0.0363	0.1495	0.1616	0.3098	0.1259	0.1616	0.3098	121
0	-350.29	157.60	-2.90	0.1298	0.0340	0.0206	0.2532	0.0866	0.2677	0.2532	0.0866	460
0	-395.80	189.02	-11.18	0.0340	0.0206	0.2810	0.1498	0.1604	0.1884	0.1498	0.1604	339
1	-256.14	166.47	-8.96	0.2965	0.2635	1.9783	0.1882	0.1940	0.3887	0.1882	0.1940	533

04 OUR DATA

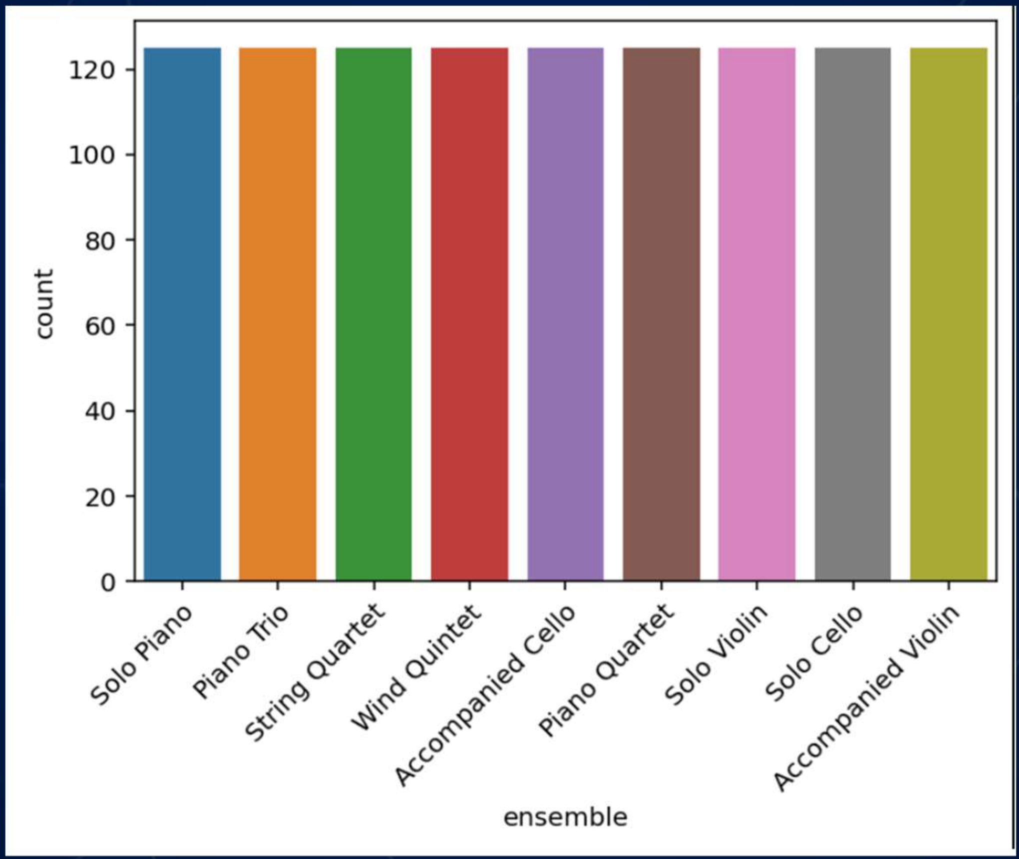
PREDICTED CLASSES : ENSEMBLE TYPES

Observation : Imbalanced Data.



05 MODEL
THIS IS WHAT SMOTE DOES!

Observation : Balanced Data after SMOTE algorithm is applied.

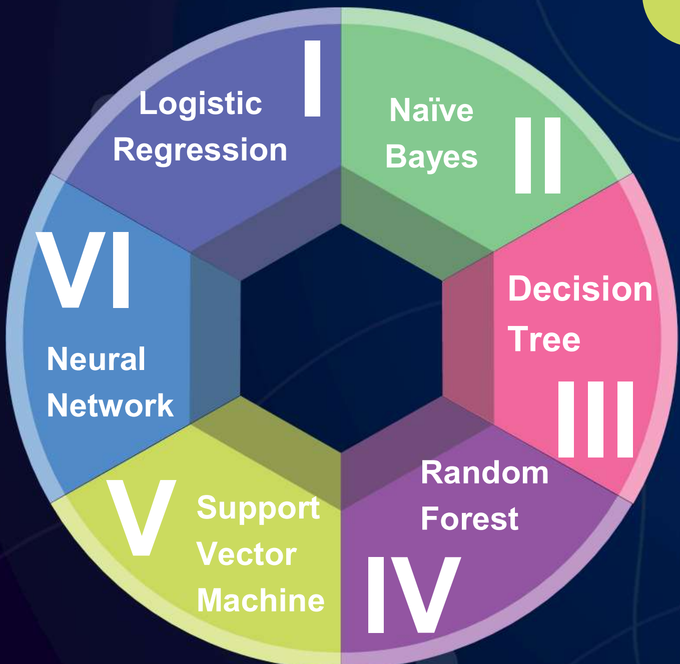


05 MODELING METHODOLOGY

Objective: To predict the instrument(s) used to perform a specific melody in a multi-class classification problem.

INDEPENDENT VARIABLES → MODEL → PREDICTION

Type of model
MULTI-CLASS MODELS



1 Type of data

Every type of model was ran in 3 datasets:

- *Spectral data*
- *MIDI data*
- *Combination of Spectral and MIDI data*

2 Complexity

We ran our most Naïve Model and depending on the situation, we continued with:

- *PCA*
- *SMOTE*
- *Hyperparameter tuning*

3 Cross-validation

Training sample was split in 5 pieces, we ran the analysis on each partition and then averaged the overall model performance metric

~30 models → **Test data**

05 MODEL DEV-SET RESULTS

F1-Score

Training Data		***Feature Set: Spectral***			***Feature Set: MIDI***		***Feature Set: Combined***	
Model Type	**Model Variant**	**Highest CV Score**	**Mean CV Score**	**Rank**	**Highest CV Score**	**Mean CV Score**	**Highest CV Score**	**Mean CV Score**
Logistic	Baseline	97.6%	94.4%	5	69.6%	65.0%	91.4%	82.4%
Logistic	PCA	63.6%	59.2%	12	31.2%	27.4%	31.2%	29.9%
Naive Bayes	Baseline	70.9%	66.7%	11	75.8%	58.9%	76.8%	74.3%
Decision Trees	Baseline	75.7%	71.4%	9	84.8%	82.5%	91.8%	81.4%
Random Forest	Baseline	83.9%	78.6%	8	92.7%	87.5%	91.3%	82.9%
Random Forest	w/SMOTE	100.0%	99.6%	**2**	99.6%	99.4%	100.0%	99.3%
SVM	Baseline	57.6%	54.2%	13	65.6%	64.1%	65.6%	64.1%
SVM	w/Hyperparameter Tuning	94.6%	90.4%	6				
SVM	w/Hyperparamater Tuning + PCA	91.5%	89.1%	7				
SVM	w/Hyperparameter Tuning + PCA + SMOTE	100.0%	99.4%	**3**				
SVM	w/Hyperparameter Tuning + PCA + SMOTE + MinMaxScaler	75.5%	70.2%	10				
SVM	w/Hyperparameter Tuning + PCA + SMOTE + StandardScaler	97.3%	96.3%	4				
Neural Network	Baseline (ADAM, 5 Hidden Layers)		100.0%	**1**		80.8%		81.2%

- Applying SMOTE resulted in the highest performance models since this creates a more balanced dataset for model training
- NN is the most predictive model in the Training Data, followed by the Random Forest w/SMOTE and SVM

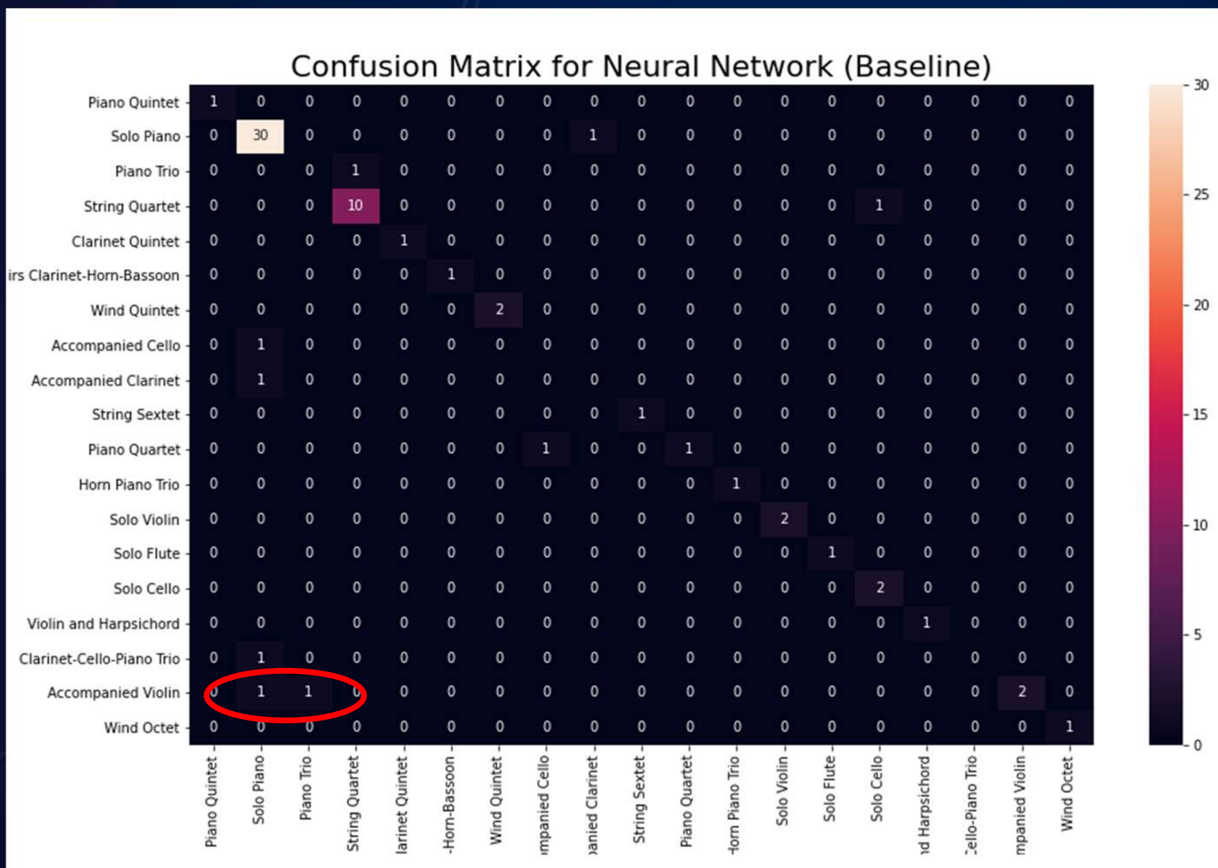
05 MODEL TEST-SET RESULTS

Training Data vs. Test Data			*Feature Set: Spectral*		
			Training	***Test***	
Model Type	**Model Variant**	**Rank @ Training**	**Mean CV Score**	**Mean CV Score**	**Rank @ Test**
Logistic Regression	Baseline	5	94.4%	72.8%	3
Logistic Regression	PCA	12	59.2%	54.8%	11
Naive Bayes	Baseline	11	66.7%	60.6%	10
Decision Trees	Baseline	9	71.4%	68.3%	7
Random Forest	Baseline	8	78.6%	64.0%	9
Random Forest	w/ SMOTE	2	99.6%	69.1%	5
SVM	Baseline	13	54.2%	47.1%	12
SVM	w/ Hyperparameter Tuning	6	90.4%	70.2%	4
SVM	w/ Hyperparamater Tuning + PCA	7	89.1%	66.3%	8
SVM	w/ Hyperparameter Tuning + PCA + SMOTE	3	99.4%	72.8%	2
SVM	w/ Hyperparameter Tuning + PCA + SMOTE + MinMaxScaler	10	70.2%	34.9%	13
SVM	w/ Hyperparameter Tuning + PCA + SMOTE + StandardScaler	4	96.3%	68.4%	6
Neural Network	Baseline (ADAM, 5 Hidden Layers)	1	100.0%	73.1%	1

- High performance models in the training dataset ssuccessfully generalized to test set.
- SMOTE causes ensembles with counts below 6 to be dropped in training but not in test.

05 CONFUSION MATRIX
BEST MODEL : NEURAL NETWORK

True Labels



Predicted Labels

Common Misclassification:

- Accompanied Violin as Solo Piano/ Piano Trio

SMOTE

- 30% of test errors Removes low sample counts in train only

06 CONCLUSION

Our Performance

Score	Metric	Model
73.1%	F1 Score	Neural Network

* For dataset containing 20 instrument ensembles (combinations of instruments)

Previous Research Performance

Score	Metric	Model
99%	Accuracy	Multi-layer perceptron
100%	Accuracy	SVM
85%	Accuracy	Random Forest

* Note: These methods used single instrument soundbytes, with between 4 and 10 possible outcomes.

THANK YOU FOR LISTENING!

QUESTIONS?

APPENDIX

APPENDIX : REFERENCES

- Brownlee, J. (2021, March 16). *SMOTE for Imbalanced Classification with Python*. Machine Learning Mastery. https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/.
- Chandwadkar, D. M., & Sutaone, M. S. (2013). Selecting Proper Features and Classifiers for Accurate Identification of Musical Instruments. *International Journal of Machine Learning and Computing*, 172–175. https://doi.org/10.7763/ijmlc.2013.v3.296
- Das, O. (2019, June). *Musical Instrument Identification with Supervised Learning*. Orchisama Das. https://ccrma.stanford.edu/~orchi/ccrma.html.
- Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., & Sutskever, I. (2020, April 30). *Jukebox: A Generative Model for Music*. arXiv.org. https://arxiv.org/abs/2005.00341.
- Engdahl, S. (2008). *How to run an AI powered musical challenge: "AWS DeepComposer Got Talent."* AWS Machine Learning Blog. https://aws.amazon.com/blogs/machine-learning/how-to-run-an-ai-powered-musical-challenge-aws-deepcomposer-got-talent/.
- Gavrilova, Y. (2020, August 5). *Machine Learning: Types of Classification Algorithms*. Serokell Software Development Company. https://serokell.io/blog/classification-algorithms
- Kawwa, N. (2019, April 29). *Can We Guess Musical Instruments With Machine Learning?* Medium. https://medium.com/@nadimkawwa/can-we-guess-musical-instruments-with-machine-learning-afc8790590b8.

APPENDIX : REFERENCES

- Mor, N., Wolf, L., Polyak, A., & Taigman, Y. (2018, May 23). *A Universal Music Translation Network*. arXiv.org. https://arxiv.org/abs/1805.07848.
- MusicNet. (n.d.). https://homes.cs.washington.edu/~thickstn/musicnet.html.
- Nabi, J. (2018, December 23). *Machine Learning-Multiclass Classification with Imbalanced Data-set*. Medium. https://towardsdatascience.com/machine-learning-multiclass-classification-with-imbalanced-data-set-29f6a177c1a.
- Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., Dinculescu, M., Radebaugh, C., Hawthorne, C., & Eck, D. (1970, January 1). *Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live*. Google Research. https://research.google/pubs/pub48280/.
- Thickstun, J., Harchaoui, Z., & Kakade, S. (2017). Learning features of music from scratch. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.

AUDIO ANALYSIS DOMAINS

SPEECH TRANSCRIPTION

ANATOMY / ZOOLOGY

MUSIC RECOMMENDATION

AUTOMATIC TRANSCRIPTION

MUSIC GENERATION

EXTRACTION, SEPARATION & CLASSIFICATION

MUSIC GENERATION AND TRANSLATION IN VOGUE

OPEN AI

JUKEBOX

GOOGLE AI

MAGENTA

FACEBOOK AI

“A Universal Music Translation Network”

AMAZON

AWS DeepComposer

OUR DATA

SOURCE : MUSICNET (CLASSICAL MUSIC DATASET)

34 hours

10 Composers

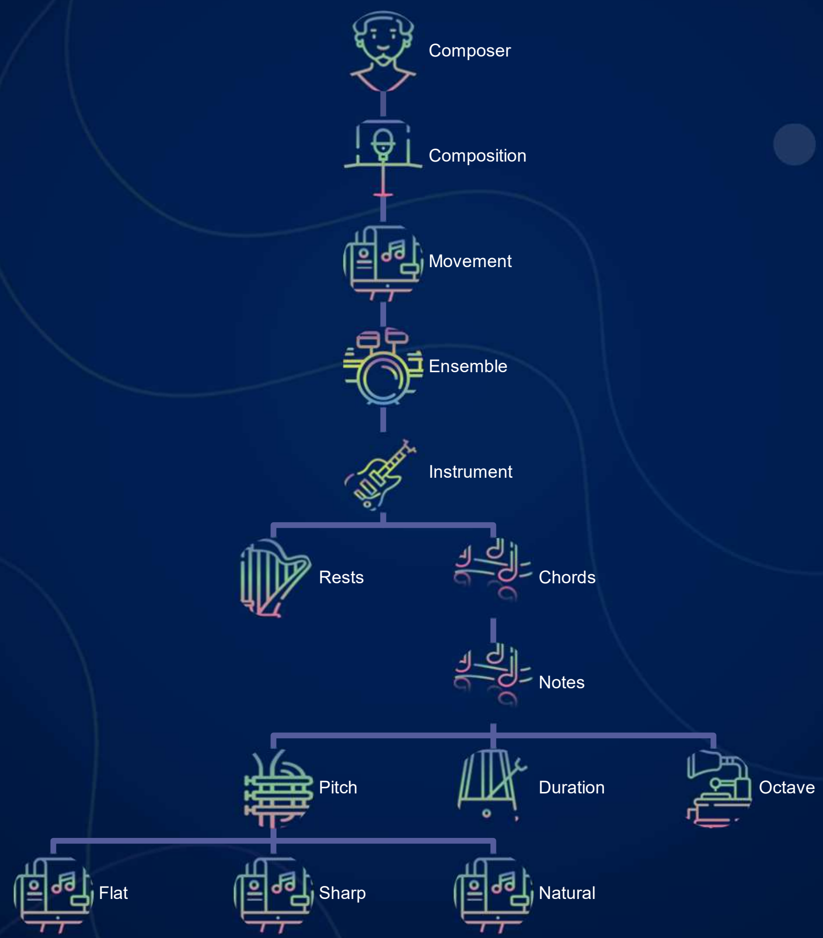
330 movements

21 Ensembles

11 Instruments

OUR DATA

DATA HIERARCHY



DATA AUGMENTATION

EFFECTS OF SMOTE

Type	Train Wav

Count of ensemble name		
ensemble label	ensemble name	Total
1	Solo Piano	125
4	String Quartet	46
19	Accompanied Violin	18
16	Solo Cello	10
14	Solo Violin	7
7	Wind Quintet	7
2	Piano Trio	6
12	Piano Quartet	6
8	Accompanied Cello	6
6	Pairs Clarinet-Horn-Bassoon	5
11	String Sextet	4
20	Wind Octet	3
17	Violin and Harpsichord	3
13	Horn Piano Trio	3
9	Accompanied Clarinet	3
0	Piano Quintet	3
15	Solo Flute	2
5	Clarinet Quintet	2
18	Clarinet-Cello-Piano Trio	2
10	Wind and Strings Octet	2
Grand Total		263

SMOTE
Removed

Type	Test Wav

Count of ensemble name			
ensem	ensemble name	Total	
1	Solo Piano	31	In Training
4	String Quartet	11	In Training
19	Accompanied Violin	4	In Training
12	Piano Quartet	2	In Training
14	Solo Violin	2	In Training
16	Solo Cello	2	In Training
7	Wind Quintet	2	In Training
17	Violin and Harpsichord	1	Not Found in Training
2	Piano Trio	1	In Training
13	Horn Piano Trio	1	Not Found in Training
5	Clarinet Quintet	1	Not Found in Training
15	Solo Flute	1	Not Found in Training
0	Piano Quintet	1	Not Found in Training
6	Pairs Clarinet-Horn-Bassoon	1	Not Found in Training
18	Clarinet-Cello-Piano Trio	1	Not Found in Training
8	Accompanied Cello	1	In Training
20	Wind Octet	1	Not Found in Training
9	Accompanied Clarinet	1	Not Found in Training
11	String Sextet	1	Not Found in Training
Grand Total		66	

	Number	Percentage
In Training	56	84.8%
Not Found in Training	10	15.2%
Total	66	100%

Type	Excluded

Count of ensemble name		
ensem	ensemble name	Total
3	Viola Quintet	1
Grand Total		1

04 OUR DATA

SPECTRAL DATA CORRELATIONS

Spectro Correlations are high