Financial Audio Analysis:

Leveraging Speech for Sentiment Classification and Predictability

Springboard DSC - Capstone Project 3

John Arancio

Can financial sentiment analysis be improved using speech audio?

- Sentiment good indicator of market momentum
  - Price action
- NLP sentiment improvement with audio
  - More accurate?
  - NLP sentiment for annotation/labeling
- Audio provides a different way to interpret sentiment
  - Can increase confidence in prospective decisions
  - Signal confluence

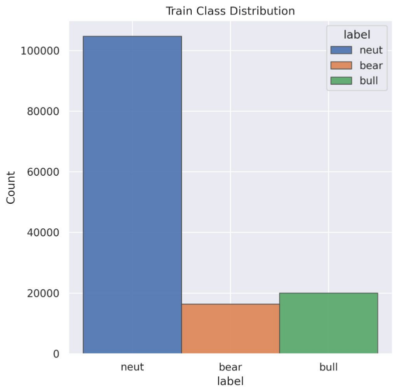
Approach

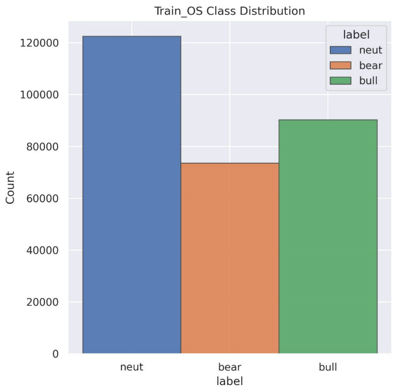
Ensemblement + Wrangling

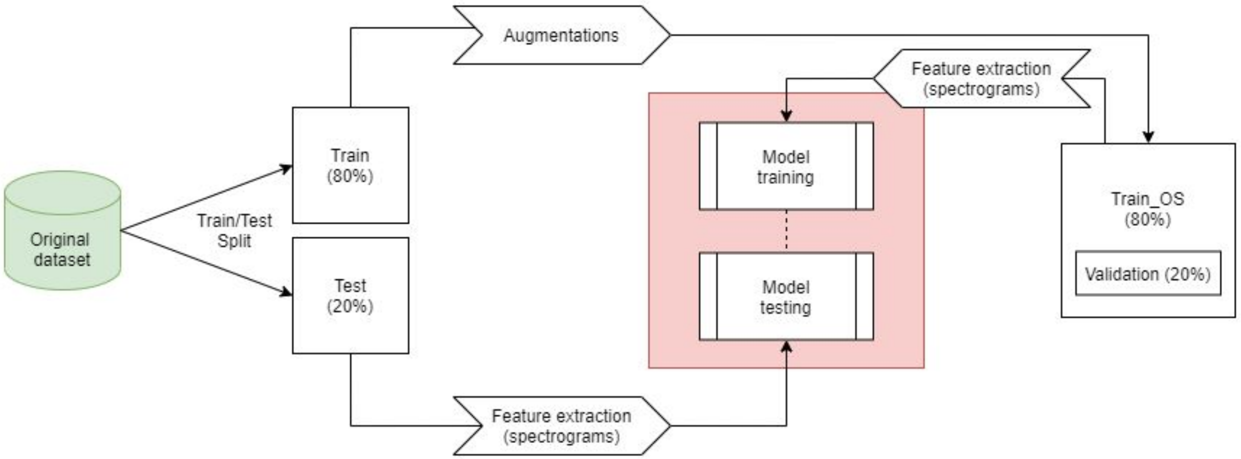
- Web scraped YouTube videos
  - Extract audios
- Transcribe audios
- Segmentation
  - By sentence uttered
- Labeling segments
  - NLP transformer
  - FinBERT
- Padding and reformatting
  - 16-bit, mono, 16000 Hz
  - 8 seconds length

Exploratory Data Analysis

- Class distribution
- Audio properties count (duration, sample rates etc.)
- View waveplots, spectrograms







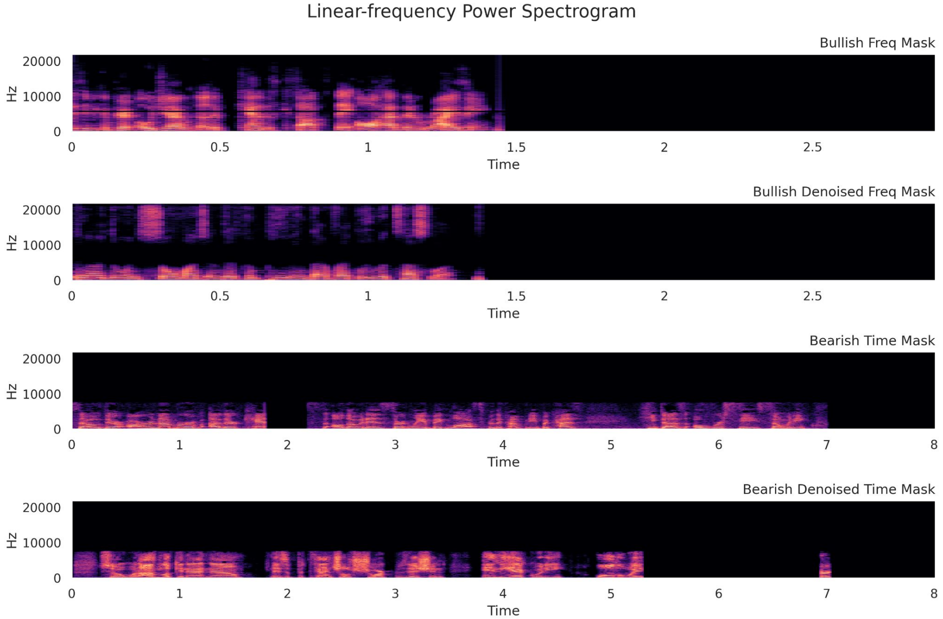
- Split original dataset
- Augmented train set
    - Oversample

- Oversampled dataset training/validation
- Evaluate original test set

Approach - Train/test split

Augmentations

- Sample denoising
- Sample shifting
- Sample frequency mask
- Sample time mask

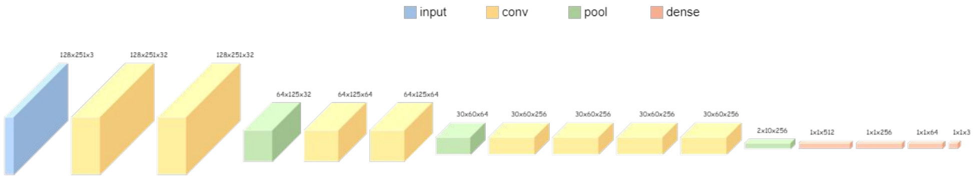


Baseline Modeling

- Logistic Regression
- Support Vector Classifier (SVC)
- XGBoost

		0 (BEARISH)	1 (BULLISH)	2 (NEUTRAL)	W.AVG
LOG. REGRESSION	PRECISION	0.40	0.38	0.48	0.43
	RECALL	0.27	0.45	0.51	0.43
	F1	0.32	0.41	0.50	0.42
		0	1	2	
XGBOOST	PRECISION	0.70	0.56	0.44	0.55
	RECALL	0.16	0.20	0.89	0.47
	F1	0.26	0.29	0.58	0.40
		0	1	2	
SVC (SKLEARN)	PRECISION	0.92	0.90	0.92	0.91
	RECALL	0.95	0.93	0.88	0.91
	F1	0.93	0.92	0.90	0.91
		0	1	2	
SVC (RAPIDS)	PRECISION	0.94	0.90	0.92	0.92
	RECALL	0.95	0.93	0.89	0.92
	F1	0.94	0.91	0.90	0.92

Extended Modeling



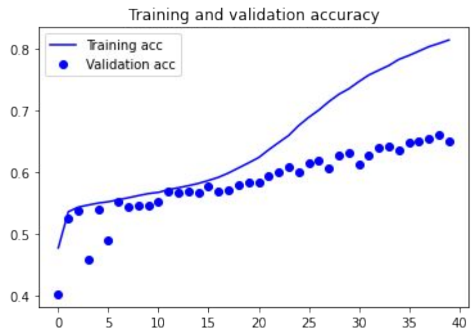
CNN VERSION	TEST LOSS	TEST ACCURACY
v1 (Shallow + tuned)	0.9341	0.5358
v2 (Larger filters + depth)	0.8658	0.5871
v3 (Depth + dropout)	0.8323	0.6192
v4 (Depth + dropout - FC dropout)	0.8134	0.6345

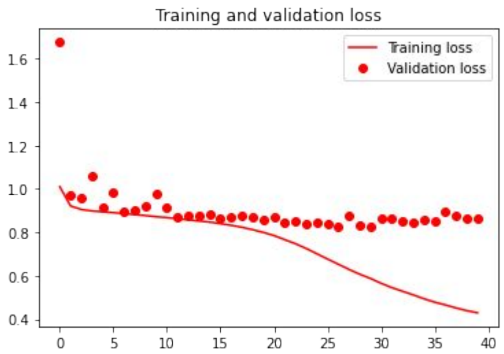
		0 (BEARISH)	1 (BULLISH)	2 (NEUTRAL)	W.AVG
CNN v1	PRECISION	0.60	0.33	0.43	0.45
	RECALL	0.35	0.42	0.58	0.45
	F1	0.50	0.42	0.49	0.47
		0	1	2	
CNN v2	PRECISION	0.44	0.55	0.61	0.53
	RECALL	0.50	0.52	0.58	0.53
	F1	0.51	0.52	0.54	0.52
		0	1	2	
CNN v3	PRECISION	0.63	0.57	0.59	0.60
	RECALL	0.67	0.50	0.58	0.58
	F1	0.67	0.53	0.60	0.60
		0	1	2	
CNN v4	PRECISION	0.65	0.60	0.65	0.63
	RECALL	0.64	0.61	0.63	0.63
	F1	0.66	0.52	0.65	0.61

- 4 CNN models
    - Depth added with each version
- V4
    - Max depth
    - Dropout regularization
    - Removed FC dropout

Findings

- Inconsistent results across models
  - SVC performed the best
    - F1(weighted avg): 0.92
    - Precision(weighted avg): 0.92
    - Recall(weighted avg): 0.92
  - Logistic Regression performed the worse
    - F1(weighted avg): 0.42
    - Precision(weighted avg): 0.43
    - Recall(weighted avg): 0.43
- CNN v4 has potential
  - Test loss: 0.8134
  - Test accuracy: 0.6345
    - Overfitting (epochs 10-15)
    - Learning rate, no change in training





Why are our results inconsistent?

- Possibilities:
  - To many augmentations
  - Unbalanced dataset
  - Mel-spectrograms not intricate enough for models to properly generalize

Conclusions

- Mel-spectrograms useful for SVC
- There are useful properties within each class
- Automated annotation process successful
    - Each class was represented by audio properly
- Sentiment can be classified using audio
- Re-evaluate augmentations
- Re-evaluate dataset balancing

Recommendations

1. Based off of SVC recall score, it can be used to positively predict bearish audios, being right 95% of the time.
2. Since both SVC and XGBoost have a 89% recall rate in predicting neutral audios, either one can be utilized.
3. SVC will be the best option in predicting bullish audios (93% recall), so if you chose one model only the best option is our support vector classifier (SVC).