Audio Classification

https://github.com/ItaiShchorry/AudioClassifier

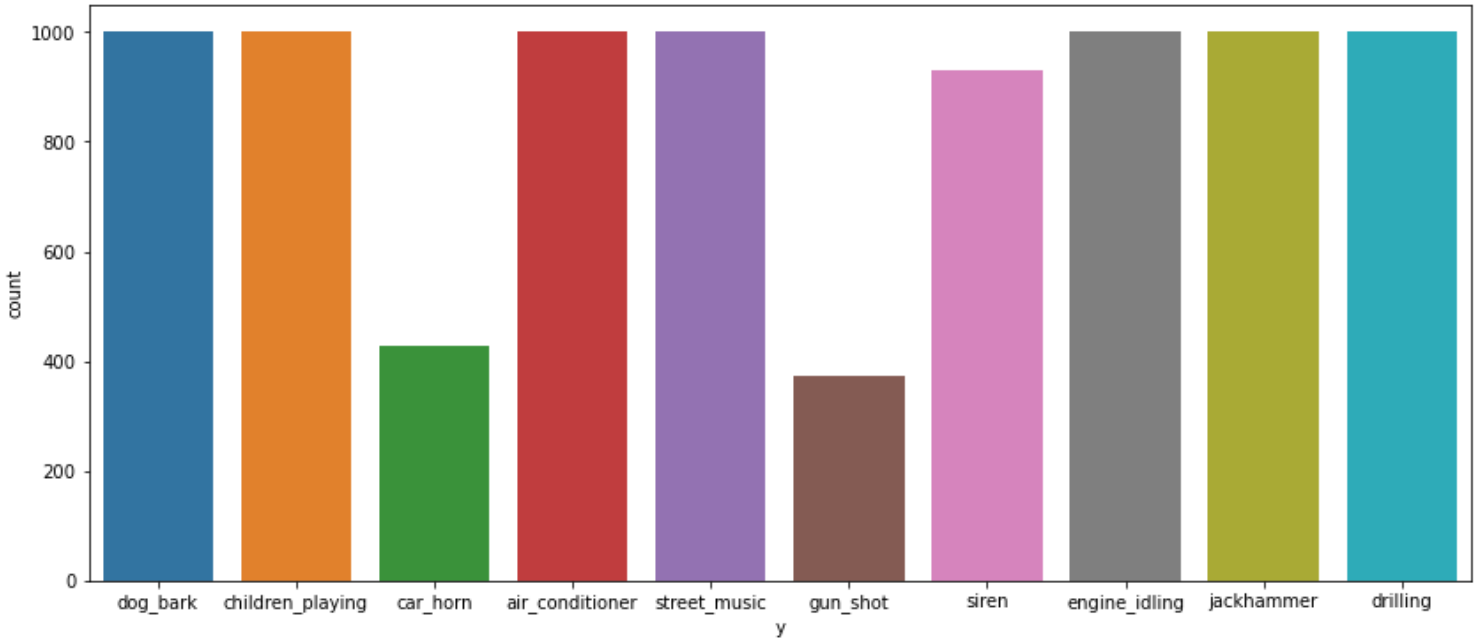
Itai Shchorry

Overview

- Project Goal – build a template methodology for developing & deploying ML applications
  - Modular & reusable as possible
- Approach – Build an audio classifying tool
  - Assumption – audio can be viewed as tabular, sequential and window-based, hence developing a utility for audio potentially extends to supporting all these datatypes

Dataset

- Urban Sounds Dataset
  - 10 classes
  - 9K examples



Processing pipeline

traceable preprocessing, managed through configuration

Create Summary View & Profiler

We'll summarize some meta-level information + aggregative measures for each column

[8]:
```
summary_table = create_col_level_summary_table(df)
disp(summary_table,max_rows=41)
```

	col_name	current_dtype	suggested_type	module	is_knob	num_unique_values	max	min	mean	median	std	missing_total	missing_percent	should_drop	reason_to_drop
0	mfcc_0	float32	numeric	None	False	8731	131.924	-767.043	-230.579	-217.399	129.832	0	0.0	False	
1	mfcc_1	float32	numeric	None	False	8730	276.346	-66.2804	115.278	121.24	48.3054	0	0.0	False	
2	mfcc_2	float32	numeric	None	False	8730	83.8572	-165.823	-28.5451	-25.467	31.4154	0	0.0	False	
3	mfcc_3	float32	numeric	None	False	8729	114.07	-79.6807	21.1783	23.8427	21.8529	0	0.0	False	
4	mfcc_4	float32	numeric	None	False	8730	44.8626	-100.634	-13.6552	-13.2264	15.7291	0	0.0	False	
5	mfcc_5	float32	numeric	None	False	8730	69.2653	-58.8799	15.0063	16.5854	14.742	0	0.0	False	
6	mfcc_6	float32	numeric	None	False	8731	54.5647	-65.2727	-8.26108	-8.19522	13.4207	0	0.0	False	
7	mfcc_7	float32	numeric	None	False	8728	50.4174	-34.8839	11.7982	12.3849	11.17	0	0.0	False	
8	mfcc_8	float32	numeric	None	False	8731	29.7796	-59.183	-9.6155	-9.53855	10.4692	0	0.0	False	
9	mfcc_9	float32	numeric	None	False	8731	57.3629	-57.3119	9.39924	10.3084	8.62147	0	0.0	False	

```
GLOBAL_CONFIG
```


```
{'BASE_FOLDER': '/dsp/dsp_portal/personal/itai.shchorry/AudioClassificationProject/AudioClassifier/Data/preprocessed/mfcc',
 'CONFIG_PATH': Path('/dsp/dsp_portal/personal/itai.shchorry/AudioClassificationProject/AudioClassifier/Data/preprocessed/mfcc/config.p'),
 'FILE_PATH': '/dsp/dsp_portal/personal/itai.shchorry/AudioClassificationProject/AudioClassifier/Data/preprocessed/mfcc/data.p',
 'TARGET': 'y',
 'ANALYSIS_RESULT_FOLDER': '/dsp/dsp_portal/personal/itai.shchorry/AudioClassificationProject/AudioClassifier/Data/preprocessed/mfcc/data',
 'REMOVE_OUTLIERS': False,
 'OUTLIER_SIGMA_CUTOFF': 5,
 'PERFORM_NORMALIZATION': True,
 'NORMALIZATION_METHOD': 'minmax',
 'NORMALIZATION_MAPPING': {1: {'groups': [], 'columns': []}},
 'RANDOM_STATE': 42,
 'NORMALIZATION_STATS': {'mfcc_0': {'X_MIN': -767.0429077148438,
   'X_MAX': 131.9243927001953},
  'mfcc_1': {'X_MIN': -66.28038787841797, 'X_MAX': 276.3455505371094},
  'mfcc_2': {'X_MIN': -165.82318115234375, 'X_MAX': 83.85719299316406},
  'mfcc_3': {'X_MIN': -79.68070220947266, 'X_MAX': 114.06986236572266},
  'mfcc_4': {'X_MIN': -100.63379669189453, 'X_MAX': 44.862606048583984},
  'mfcc_5': {'X_MIN': -58.87994384765625, 'X_MAX': 69.26529693603516},
  'mfcc_6': {'X_MIN': -65.2727279663086, 'X_MAX': 54.56472396850586},
  'mfcc_7': {'X_MIN': -34.883914947509766, 'X_MAX': 50.41743850708008},
  'mfcc_8': {'X_MIN': -59.1829833984375, 'X_MAX': 29.779634475708008},
  'mfcc_9': {'X_MIN': -57.311866760253906, 'X_MAX': 57.36292266845703},
```

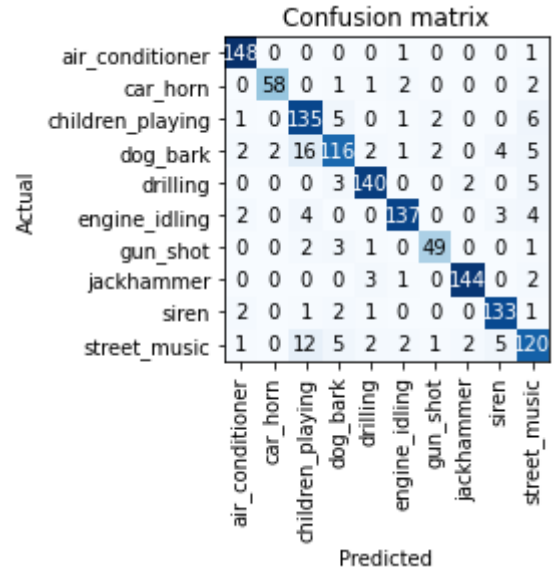
Baselines

Classical Tabular Models

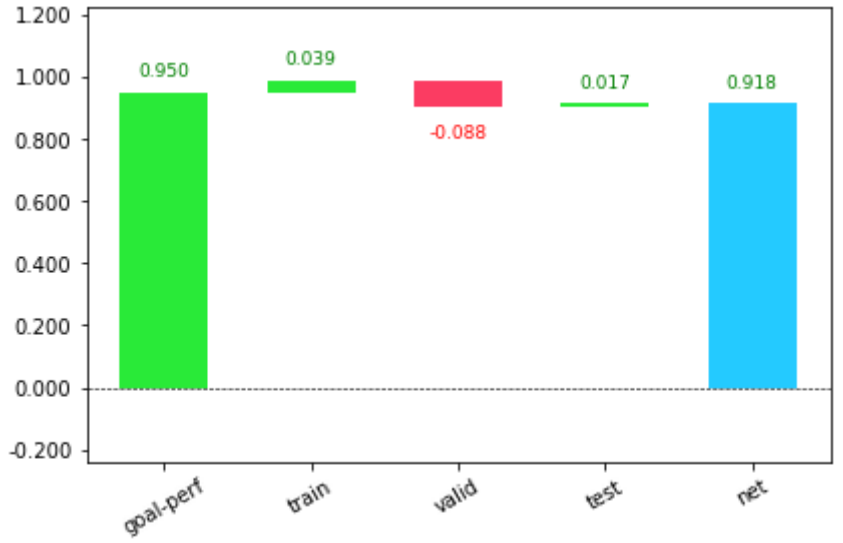
	Model	Description	Data Version	Accuracy	Kappa	Run Time
1	Baseline Classification	Random Baseline	Edition 1	0.117939	0.011735	0:00:00
2	Decision Tree	Default settings	Edition 1	0.672901	0.633752	0:00:00
3	Random Forest 1	Default settings	Edition 1	0.891221	0.877943	0:00:00
4	Random Forest 2	Greedy Feature Selection, n_estimators=40, max_features = 1	Edition 1	0.861832	0.844998	0:00:00
5	Random Forest 3	Very Shallow, min_samples_leaf=0.1	Edition 1	0.612595	0.564400	0:00:00
6	Random Forest 4	Greedy Feature Selection, min_samples_leaf=0.01, max_features=1	Edition 1	0.612595	0.564400	0:00:00
7	Random Forest 5	Less Shallow, Greedy, min_samples_leaf=0.001, max_features=0.8	Edition 1	0.810687	0.787507	0:00:00
8	Random Forest 6	Less Greedy FS, min_samples_leaf=0.001, max_features=0.8	Edition 1	0.811832	0.788772	0:00:00
9	Random Forest 7	Regularized, min_samples_leaf=0.001, max_features=sqrt	Edition 1	0.828244	0.807191	0:00:00
10	Gradient Boosting	Default settings	Edition 1	0.815267	0.792760	0:01:16
11	XGBoost	Default settings	Edition 1	0.891221	0.877997	0:00:30
12	LightGBM	Default settings	Edition 1	0.901527	0.889529	0:00:07

Fast.ai's Tabular Learner (FC-based NN)

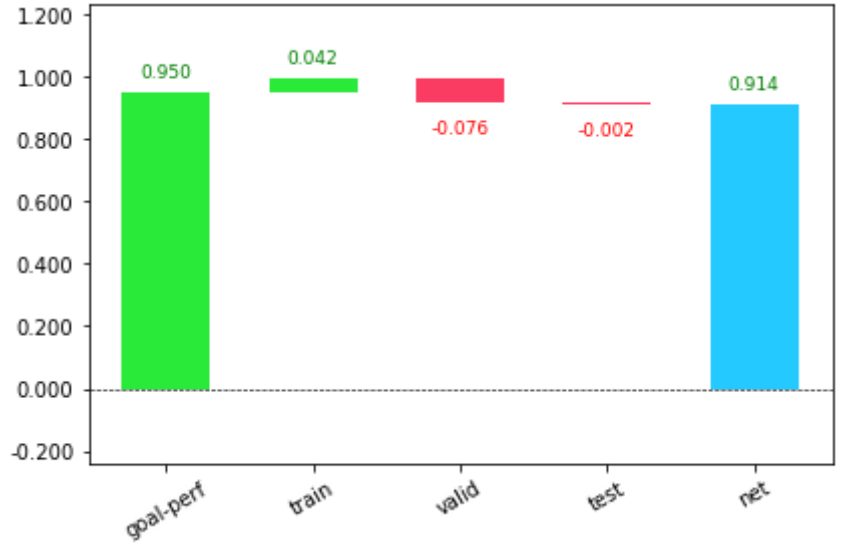
- Overfit a single batch
- Train with defaults. Score – 90.07% (accuracy)
- Bias-variance decomp. In next slide



Bias-Variance Decomposition



Observing Overfitting -> Adding regularizations (weight decay)



Less overfitting but worse test accuracy. We'll try an HP search.

Experiment Tracking & Hyperparameter Optimization

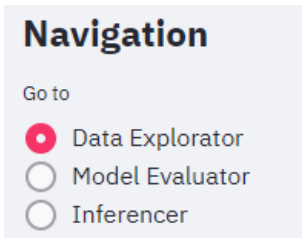
- Experiment Tracking
  - W&B
- HP Optimization
  - Data aspects as hyperparameters (outlier-dropping thresholds)
  - Grid-search based (focus on lr, layer groups momentums & weight decay)

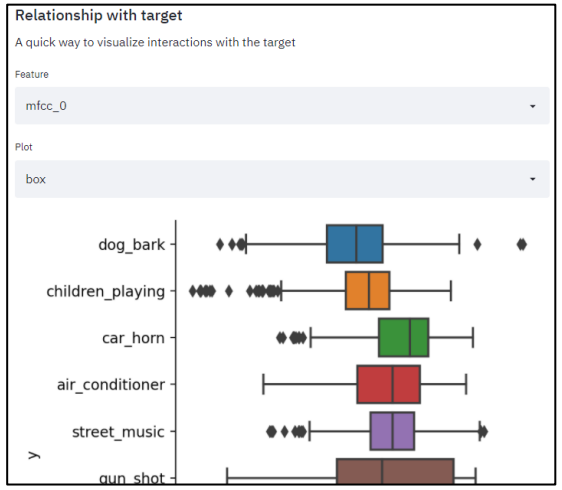
Name (144 visualized)	Runtime	Notes	Tags	Learner.lr	Learner.mo	Learner.wd	accuracy	lr_0	mom_0	raw_loss	train_loss	valid_loss	wd_0
tuning_cutoff_0_16	4m 24s	Add notes		0.005	[0.85,0.9,0	0.1	0.9443	5.152e-8	0.95	0.01312	0.01518	0.2361	0.1
tuning_cutoff_0_22	4m 28s	Add notes		0.005	[0.9,0.9,0.!	0.1	0.9427	5.152e-8	0.9	0.03256	0.01595	0.2344	0.1
tuning_cutoff_0_20	4m 39s	Add notes		0.005	[0.95,0.9,0	0.01	0.9427	5.152e-8	0.85	0.01897	0.01454	0.2747	0.01
tuning_cutoff_0_13	4m 54s	Add notes		0.005	[0.95,0.85,	0.1	0.9412	5.152e-8	0.95	0.01395	0.01623	0.2311	0.1
tuning_cutoff_0_21	4m 45s	Add notes		0.005	[0.9,0.9,0.!	-	0.9397	5.152e-8	0.9	0.004743	0.01195	0.2679	0.01
tuning_cutoff_0_12	4m 55s	Add notes		0.005	[0.95,0.85,	-	0.9397	5.152e-8	0.95	0.009253	0.01102	0.2999	0.01
tuning_cutoff_0_15	4m 22s	Add notes		0.005	[0.85,0.9,0	-	0.9397	5.152e-8	0.95	0.003691	0.01325	0.2466	0.01
tuning_cutoff_0_17	4m 22s	Add notes		0.005	[0.85,0.9,0	0.01	0.9389	5.152e-8	0.95	0.01403	0.0165	0.3068	0.01
tuning_cutoff_0_19	4m 58s	Add notes		0.005	[0.95,0.9,0	0.1	0.9389	5.152e-8	0.85	0.02006	0.01769	0.2346	0.1
tuning_cutoff_0_18	4m 37s	Add notes		0.005	[0.95,0.9,0	-	0.9374	5.152e-8	0.85	0.02184	0.01274	0.3014	0.01
tuning_cutoff_0_23	4m 38s	Add notes		0.005	[0.9,0.9,0.!	0.01	0.9366	5.152e-8	0.9	0.03111	0.01313	0.3094	0.01
tuning_cutoff_5_20	4m 16s	Add notes		0.005	[0.95,0.9,0	0.01	0.9359	5.152e-8	0.85	0.006952	0.01182	0.3244	0.01
tuning_cutoff_5_19	4m 29s	Add notes		0.005	[0.95,0.9,0	0.1	0.9351	5.152e-8	0.85	0.01159	0.01572	0.2809	0.1
tuning_cutoff_5_16	4m 29s	Add notes		0.005	[0.85,0.9,0	0.1	0.9351	5.152e-8	0.95	0.01948	0.01507	0.2584	0.1

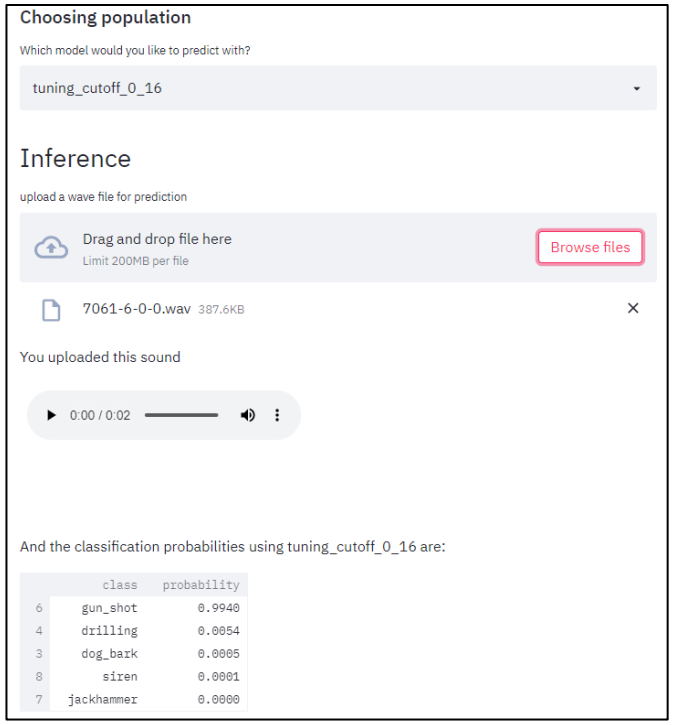
Managed to produce ~3% improvement

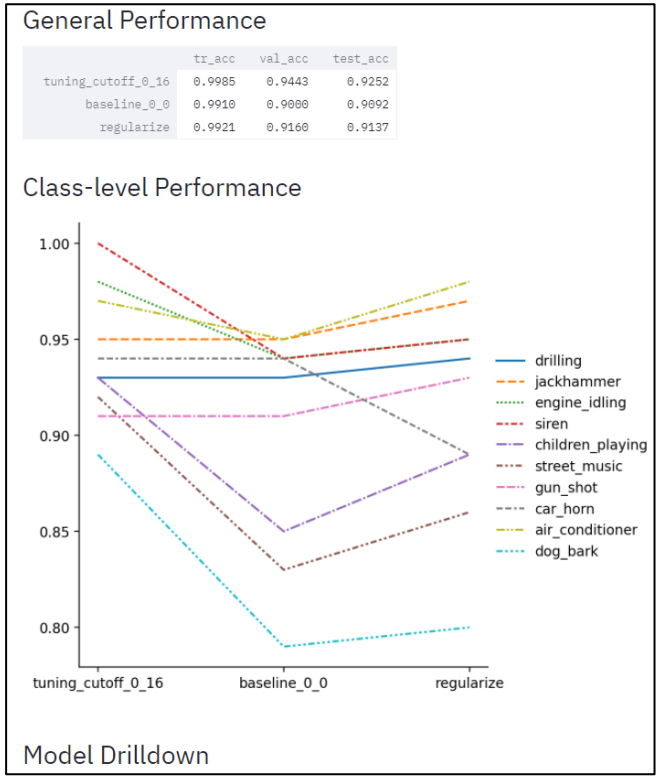
Deployment

- Based on Streamlit
- Supports EDA, Evaluations Drilldown & Inferencing









Next Steps

- Application
  - Training API
  - Labeling Section
- Training
  - Transformer-based Tabular (e.g. TabNet)
  - Extend to images, sequences
- Stability & Scalability
  - Monitoring
  - Additional testing
  - Dockerization
  - Better data versioning (currently hard-copies)
  - DAG Orchestration

Thank You
for a great course =)