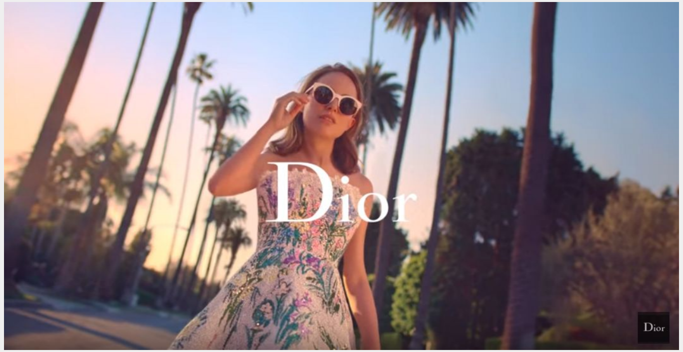
Can we tell what an advert is about based on its audio?

(yes)

The Plan

- Select distinct advertising classes
- Scrape videos from youtube for each class
- Extract and analyse audio from video
- Use supervised machine learning to train a classifier to recognise ad class based on audio only



Scraping the data

Download videos from YouTube using Pytube python library

Classes:

- Alcohol
- Cars
- Perfume
- Banks
- Supermarkets

Filters:

- Upload age
- Duration
- Keyword in title

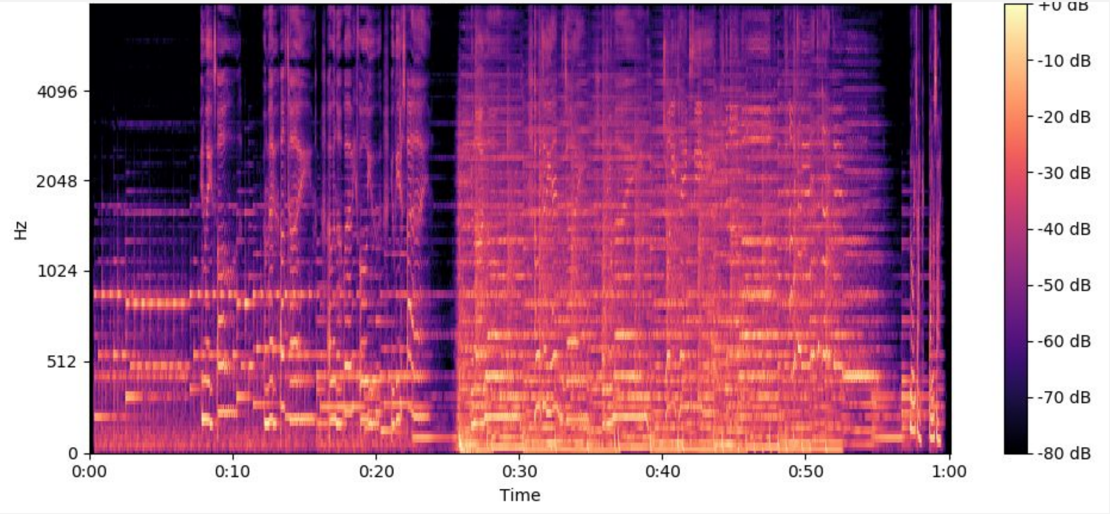
Analysing the audio spectrogram

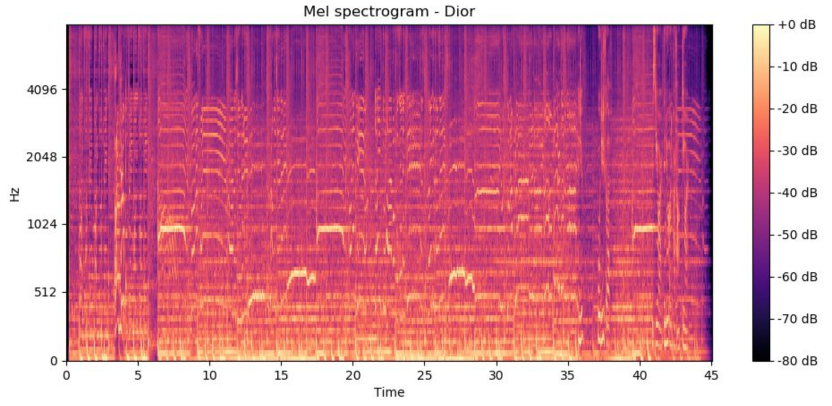
- Extract features in time and frequency domain using librosa

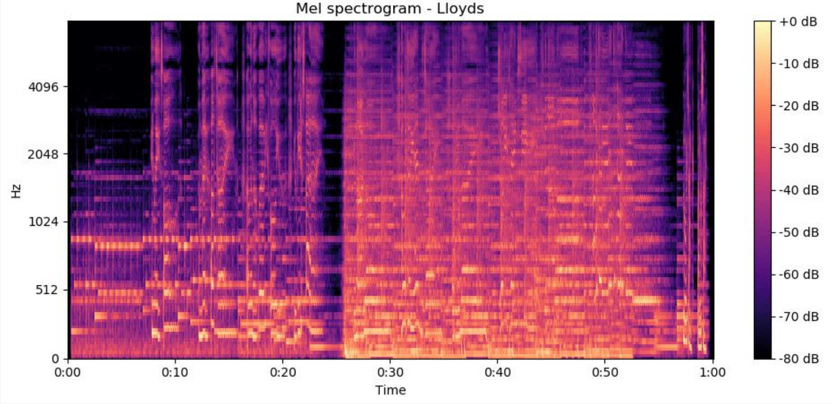
Slice song into many frequency bands and time chunks and calculate:

- RMS energy (mean, standard deviation, skew, kurtosis)
- Zero-crossing rate
- Spectral centroid
- 1st order coefficient of a polynomial fit to the power spectrum
- Harmonic & percussive separation
- Chromatic feature

Also capture time varying properties via a running RMS window of varying width



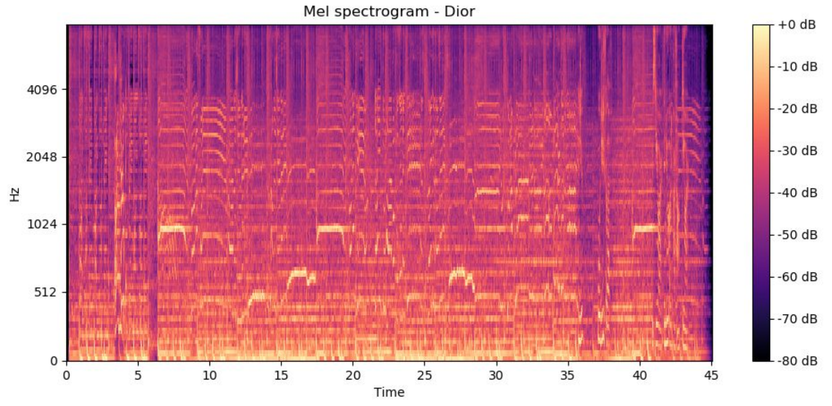


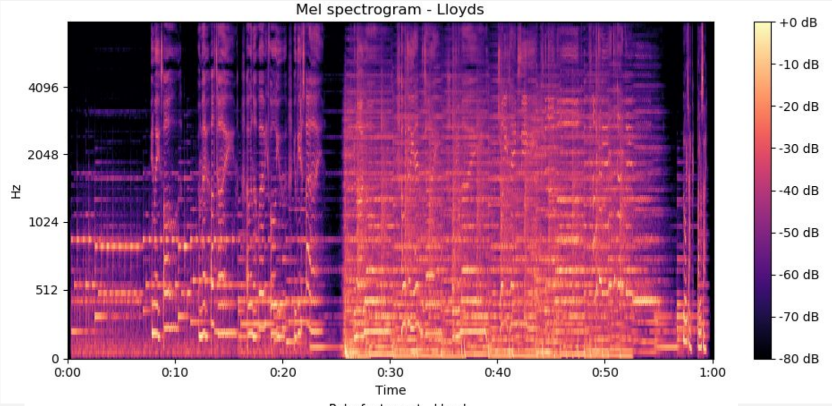


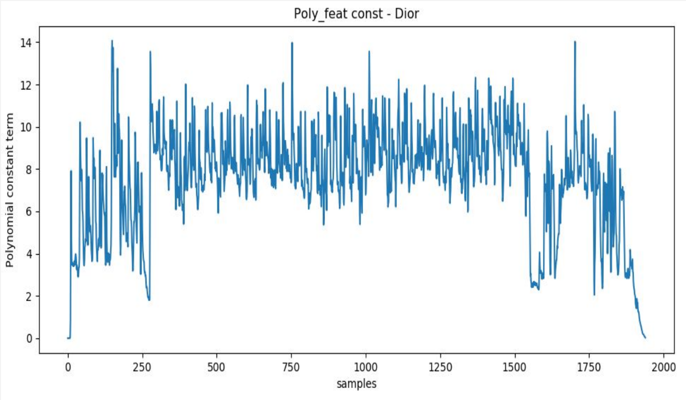
Fit an n^{th} order polynomial to the columns of a spectrogram.

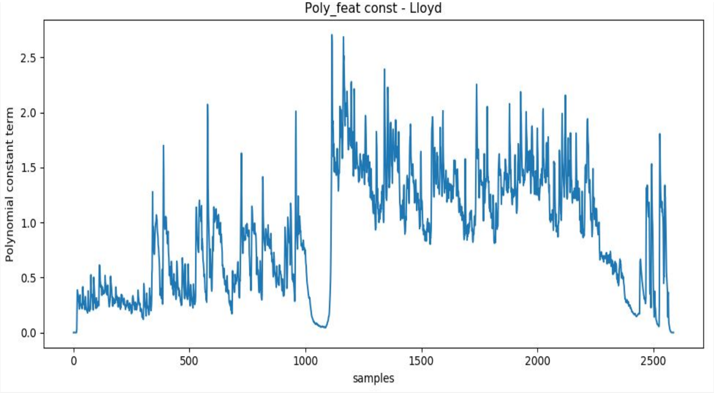
Take the RMS of windows of varying width


Dior


Mel spectrogram - Dior
Hz
4096
2048
1024
512
0
0 5 10 15 20 25 30 35 40 45
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB


Mel spectrogram - Lloyds
Hz
4096
2048
1024
512
0
0:00 0:10 0:20 0:30 0:40 0:50 1:00
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB


Poly_feat const - Dior
Polynomial constant term
0 2 4 6 8 10 12 14
0 250 500 750 1000 1250 1500 1750 2000
samples


Poly_feat const - Lloyd
Polynomial constant term
0.0 0.5 1.0 1.5 2.0 2.5
0 500 1000 1500 2000 2500
samples

Now have 419 features describing the audio

2500 videos (500 per class)

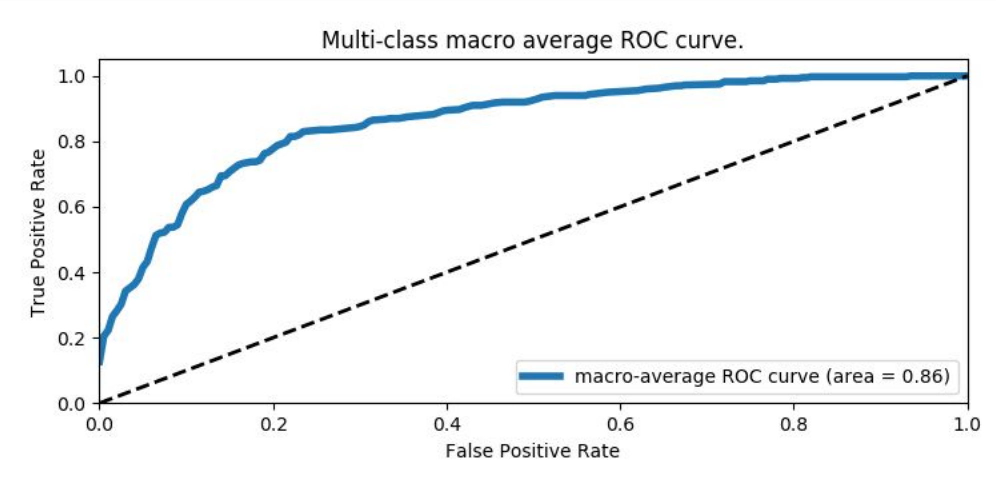
- Alcohol
- Cars
- Perfume
- Banks
- Supermarkets

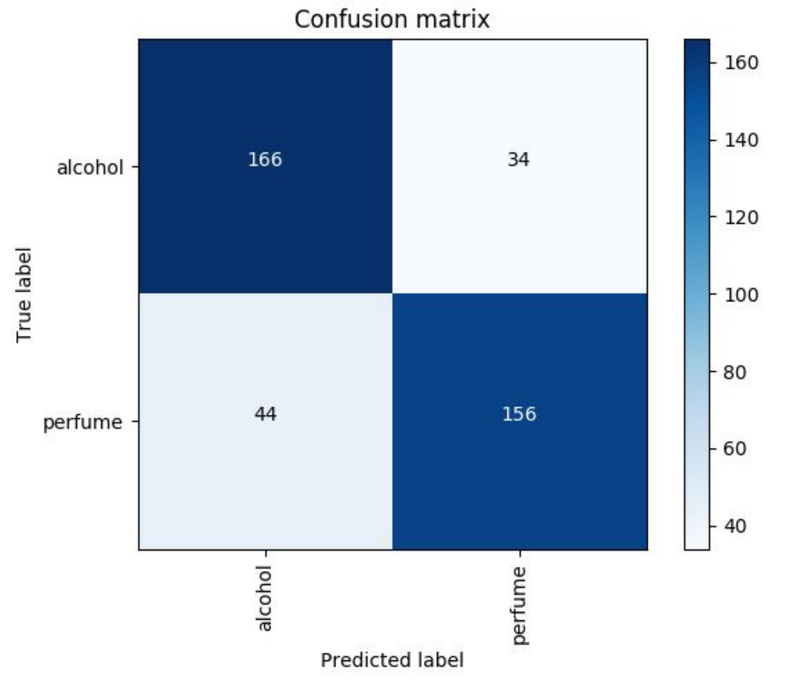
Train a Random Forest on 60% (300 per class) of them, test on remaining 40% (200 per class)

Random Forest

- Binary class: Perfume and Alcohol

```
--------------------------------------------------------------
Random forest before pruning:
             precision    recall  f1-score   support

    alcohol       0.79      0.83      0.81       200
    perfume       0.82      0.78      0.80       200

avg / total       0.81      0.81      0.80       400

--------------------------------------------------------------
```





Random Forest - multiclass

- Explore selecting most important features maintains similar performance
- Seems to handle fewer features better than an excess of features.

44/115 features:

accuracy before pruning features: 0.53
accuracy after pruning features: 0.54

Random Forest report after feature pruning:

	precision	recall	f1-score	support
alcohol	0.53	0.39	0.45	200
banks	0.59	0.58	0.58	200
cars	0.45	0.50	0.48	200
perfume	0.56	0.65	0.60	200
supermarkets	0.55	0.57	0.56	200
avg / total	0.54	0.54	0.53	1000

178/419 features:

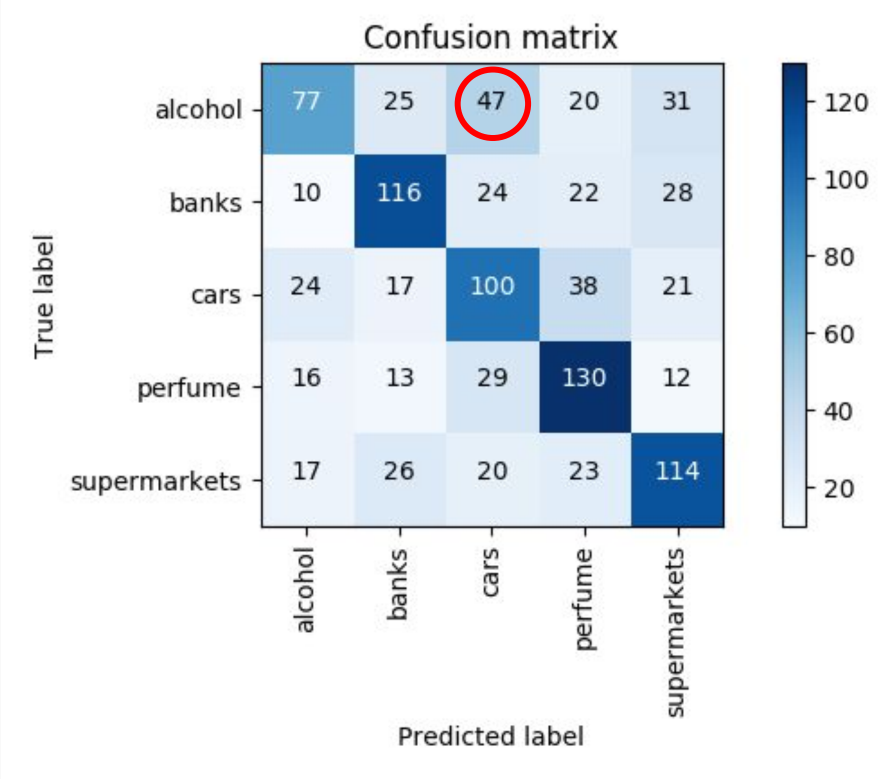
accuracy before pruning features: 0.49
accuracy after pruning features: 0.49

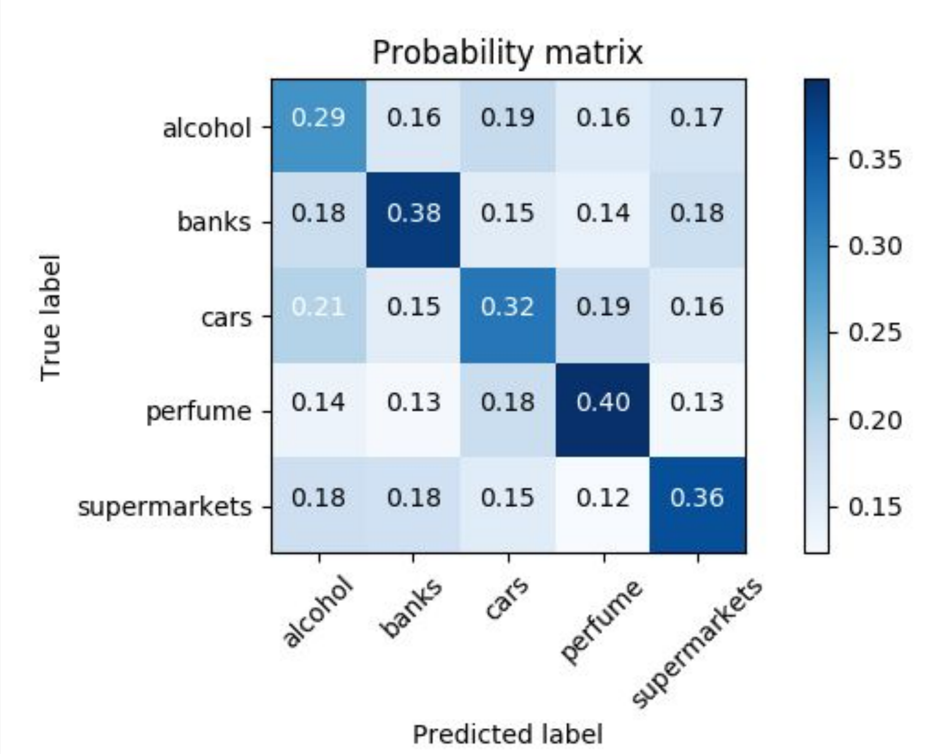
Random Forest report after feature pruning:

	precision	recall	f1-score	support
alcohol	0.51	0.35	0.42	200
banks	0.50	0.57	0.53	200
cars	0.42	0.51	0.46	200
perfume	0.60	0.57	0.58	200
supermarkets	0.48	0.47	0.47	200
avg / total	0.50	0.49	0.49	1000

Random Forest

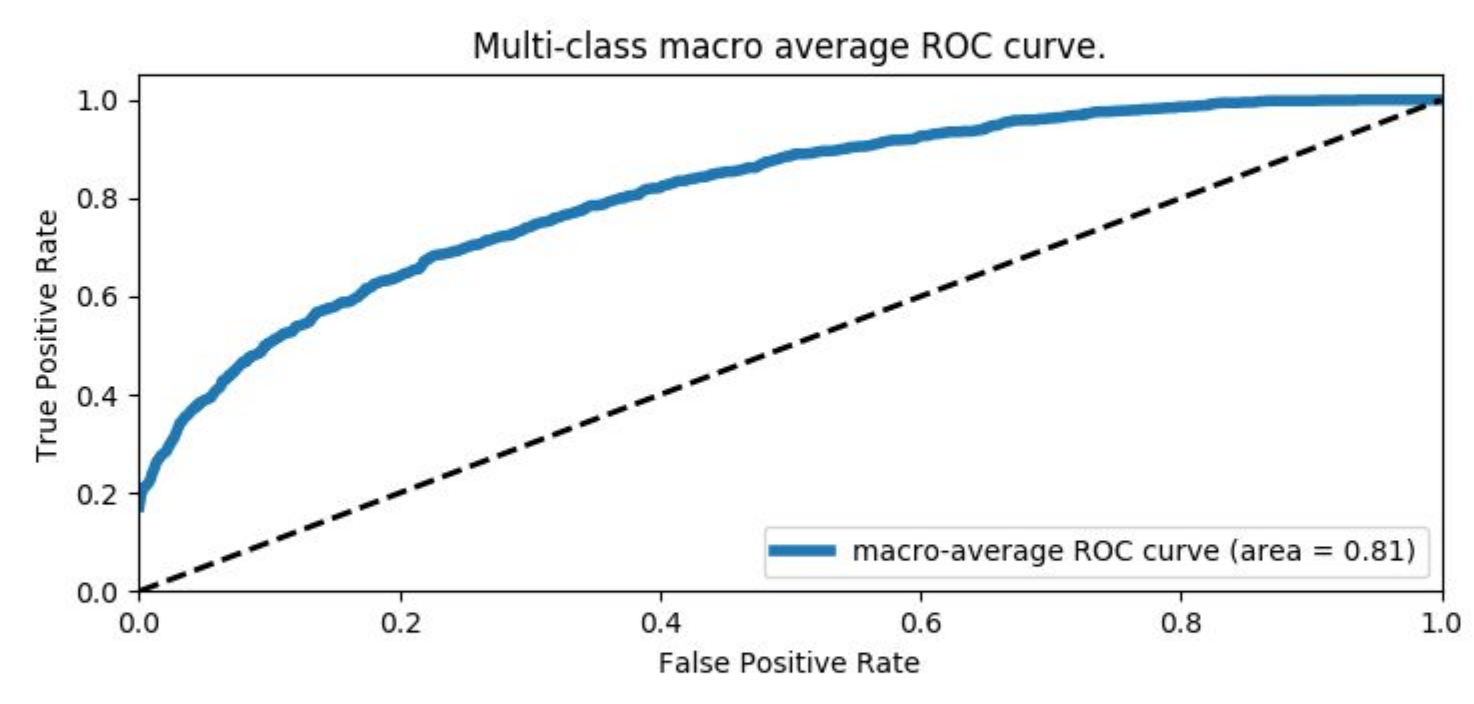
- Confusion matrix





Random Forest

- Receiver Operating Characteristic



Neural Network

- Performance increased with more features

115 features:

AUC = 0.7398625000000001
number of features : 115
accuracy = 0.452

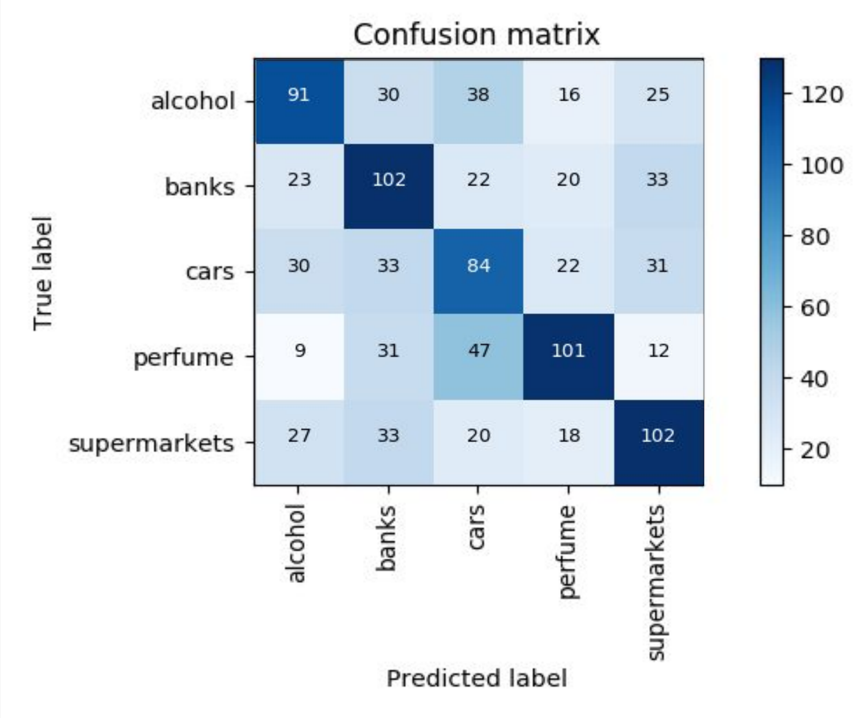
	precision	recall	f1-score	support
alcohol	0.40	0.38	0.39	200
banks	0.46	0.46	0.46	200
cars	0.39	0.46	0.42	200
perfume	0.53	0.55	0.54	200
supermarkets	0.49	0.43	0.46	200
avg / total	0.45	0.45	0.45	1000

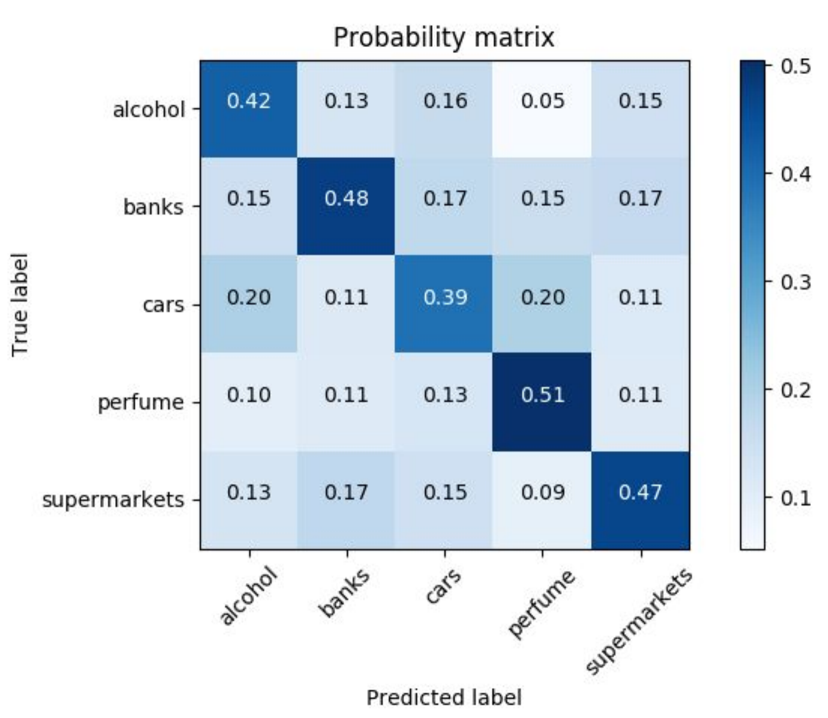
419 features:

AUC = 0.759625
number of features : 419
accuracy = 0.48

	precision	recall	f1-score	support
alcohol	0.51	0.46	0.48	200
banks	0.45	0.51	0.48	200
cars	0.40	0.42	0.41	200
perfume	0.57	0.51	0.54	200
supermarkets	0.50	0.51	0.51	200
avg / total	0.48	0.48	0.48	1000

Neural Network - metrics





Which is best?

- Very similar performance

Neural Network

```
AUC = 0.759625
number of features : 419
accuracy = 0.48
----------------------------------------
             precision    recall  f1-score   support

     alcohol       0.51      0.46      0.48       200
       banks       0.45      0.51      0.48       200
        cars       0.40      0.42      0.41       200
     perfume       0.57      0.51      0.54       200
supermarkets       0.50      0.51      0.51       200

 avg / total       0.48      0.48      0.48      1000
```

Random Forest

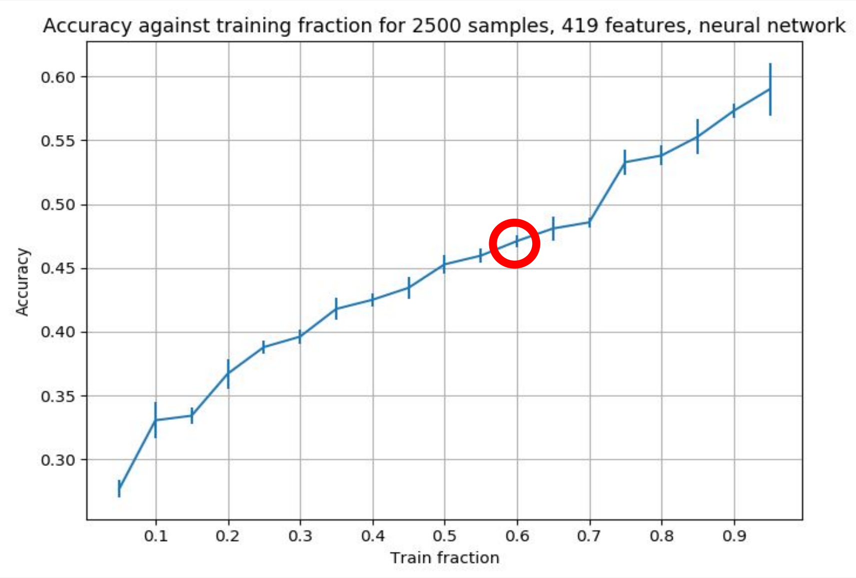
```
accuracy before pruning features: 0.49
accuracy after pruning features: 0.49
-----------------------------------------------------
Random Forest report after feature pruning:
             precision    recall  f1-score   support

     alcohol       0.51      0.35      0.42       200
       banks       0.50      0.57      0.53       200
        cars       0.42      0.51      0.46       200
     perfume       0.60      0.57      0.58       200
supermarkets       0.48      0.47      0.47       200

 avg / total       0.50      0.49      0.49      1000
```

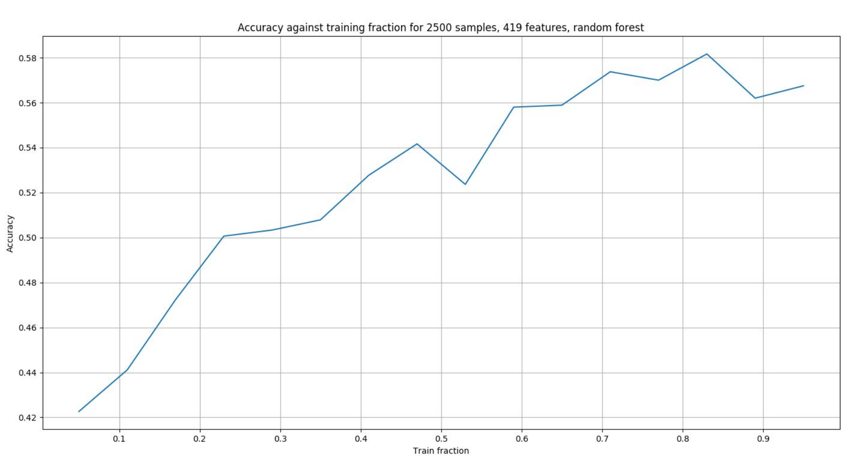
Which is best?

It’s hard to tell…

Neural Network may need more tuning

Try with more data





Thoughts...

Both classifiers are successful - as shown with various metrics.

Why are they not perfect? Because of features or data?

Is audio in adverts too similar? What degree of accuracy should we expect?

Can we use ML to inform companies about how to make their ads more distinctive?

Conclusions

Understanding the problem, the data, and the questions one wants to ask is essential.

Looking ahead: more complex features, such as vocal separation/analysis.

Ask other questions: A “not an advert” class, e.g. podcasts (just speech). Can we incorporate analysis of the video too?