Predicting Song Popularity from Spotify Attributes

Presenter: Steve Feng

Team members: Anna Feng, Linhao Chen, Alex Chiu, Alejandro Estrada-Esparza, Emmy Huynh, Ryan Kim, Joyce Lu, Jose Ordaz, Suryakiran Santhosh, Zhetan Zhang

Problem Description

We wondered…

- Recipe to create a popular song?
- Subconscious preferences?
- Patterns?

Goals & Application

- Craft more popular songs
- No more “flops”

Spotify

Related Work

- Spotify quantify music for computational studies
- Database of the calculated attributes of each song.
- Literature: “Predicting Popularity on Spotify - When Data Needs Culture More than Culture Needs Data” by Philip Peker.


```
{
  "acousticness": 0.00242,
  "analysis_url": "https://api.spotify.com/v1/audio-analysis/2takcwOaAZWiXQijPHIx7B\n",
  "danceability": 0.585,
  "duration_ms": 237040,
  "energy": 0.842,
  "id": "2takcwOaAZWiXQijPHIx7B",
  "instrumentalness": 0.00686,
  "key": 9,
  "liveness": 0.0866,
  "loudness": -5.883,
  "mode": 0,
  "speechiness": 0.0556,
  "tempo": 118.211,
  "time_signature": 4,
  "track_href": "https://api.spotify.com/v1/tracks/2takcwOaAZWiXQijPHIx7B\n",
  "type": "audio_features",
  "uri": "spotify:track:2takcwOaAZWiXQijPHIx7B",
  "valence": 0.428
}
```


Spotify

Proposed Methodology

- Dataset Description and Data Analysis

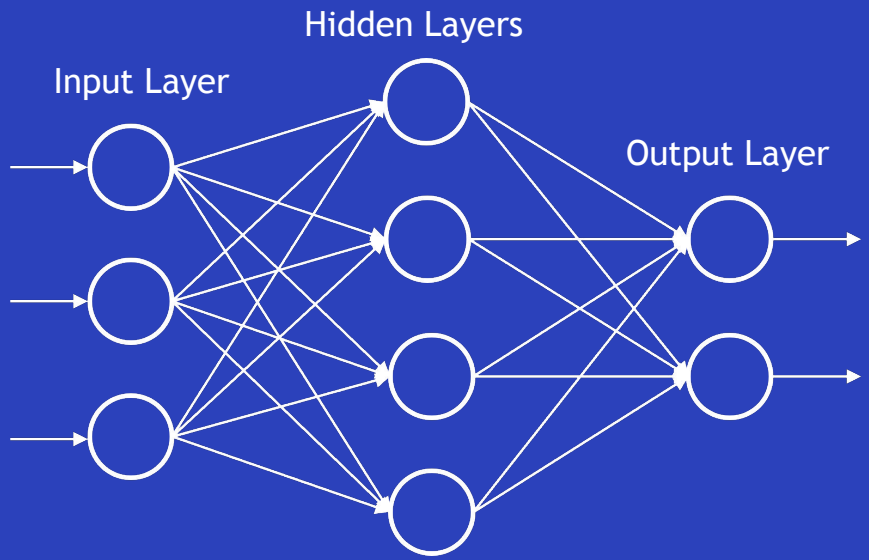
	Popularity	Danceability	Instrumentalness	Liveness	Loudness	Speechiness	Valence
count	16984	16984	16984	16984	16984	16984	16984
mean	37.65	0.54	0.17	0.23	-10.01	0.13	0.45
std	16.24	0.19	0.32	0.21	6.35	0.21	0.27
min	1	0.06	0.00	0.01	-52.46	0.02	0.01
50%	38	0.56	0.00	0.13	-8.06	0.05	0.44
max	100	0.99	0.99	1.00	3.74	0.97	1.00

- Preprocessing

Prediction Set	1	2	3	4	5
Popularity	1-20	21-40	41-60	61-80	81-100
Data Points	23769	67075	63926	15037	1001

NN & Evaluations

Neural Network

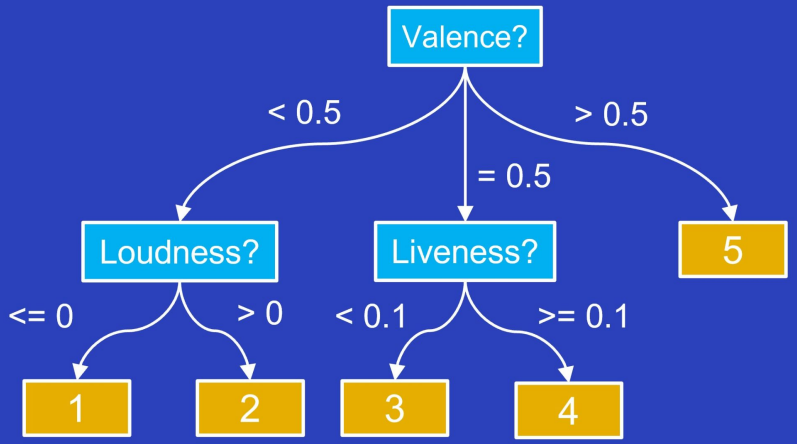


Classes	precision	recall	f1-score	instances	Confusion Matrix
1-20	0.64	0.27	0.38	5149	[27997 791] [3762 1387]
21-40	0.51	0.63	0.56	13755	[11839 8343] [5110 8645]
41-60	0.50	0.60	0.55	12322	[14281 7334] [4885 7437]
61-80	0	0	0	2614	[31323 0] [2614 0]
81-100	0	0	0	97	[33840 0] [97 0]
Accuracy			0.5147	33937	
Weighted Average	0.49	0.51	0.48	33937	
MSE			0.6448	33937	

Spotify

Random Forest & Evaluations

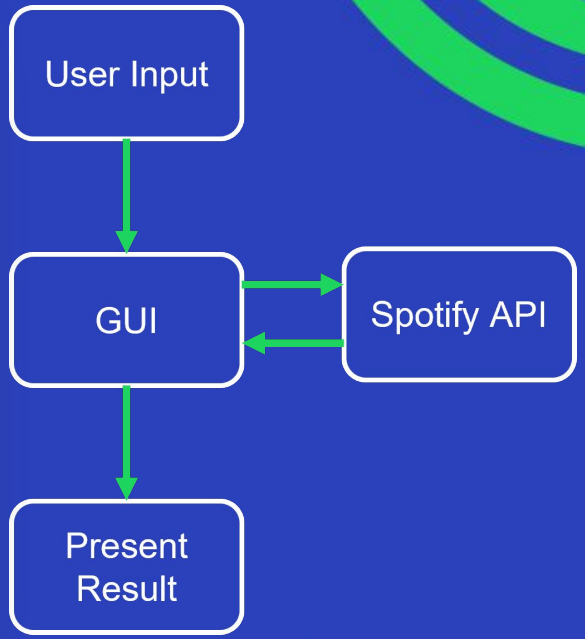
Decision Tree



Classes	precision	recall	f1-score	instances	Confusion Matrix
1-20	0.60	0.35	0.44	7716	[[41821 1819] [5018 2698]
21-40	0.51	0.59	0.55	20602	[19267 11487] [8546 12056]
41-60	0.49	0.61	0.55	18655	[21136 11565] [7323 11332]
61-80	0.33	0.03	0.05	4146	[46967 243] [4028 118]
81-100	0.66	0.11	0.18	237	[51106 13] [212 25]
Accuracy			0.51	51356	
Weighted Average	0.50	0.51	0.49	51356	
MSE			0.6795	51356	

Spotify

GUI & Spotify API



Spotify

Conclusion

Wrapping up with conclusion and what is yet remains to be addressed before the final report is due.

Problems

- Too much data
- Not enough information

Solutions

- Narrow scope to a time frame
- Create new ways to analyze music