Predicting Music Popularity Using the Spotify Dataset

By: Xinyang Zhou, Lim Jia Ying Jermaine, Hay Man Hnin Lwin, and Vania Rohmetra

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Table of contents

Mars Is a Cold Place
The 15th Planet
2:54
3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

01

Business Problem

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Did You Know?

>90%

of all music artists are undiscovered.

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Why do so many music producers fail?

Giving up too easily

Lacking social networks

Misinformation

Poor work ethics

Lacking direction

Mars Is a Cold Place
The 15th Planet
2:54
3:49

The Million Dollar Question

So… how can music producers create hit songs?

Mars Is a Cold Place
The 15th Planet
2:54
3:49

What combination of song elements will lead to the most popularity?

- We can generate revenue for records company
- The prediction model can also be useful to singers, too

Mars Is a Cold Place
The 15th Planet

2:54 3:49

We conducted a short survey...

What element(s) of a song do you think are the most important for making a hit song?

Mars Is a Cold Place
The 15th Planet
2:54
3:49

We want to know what YOU think!

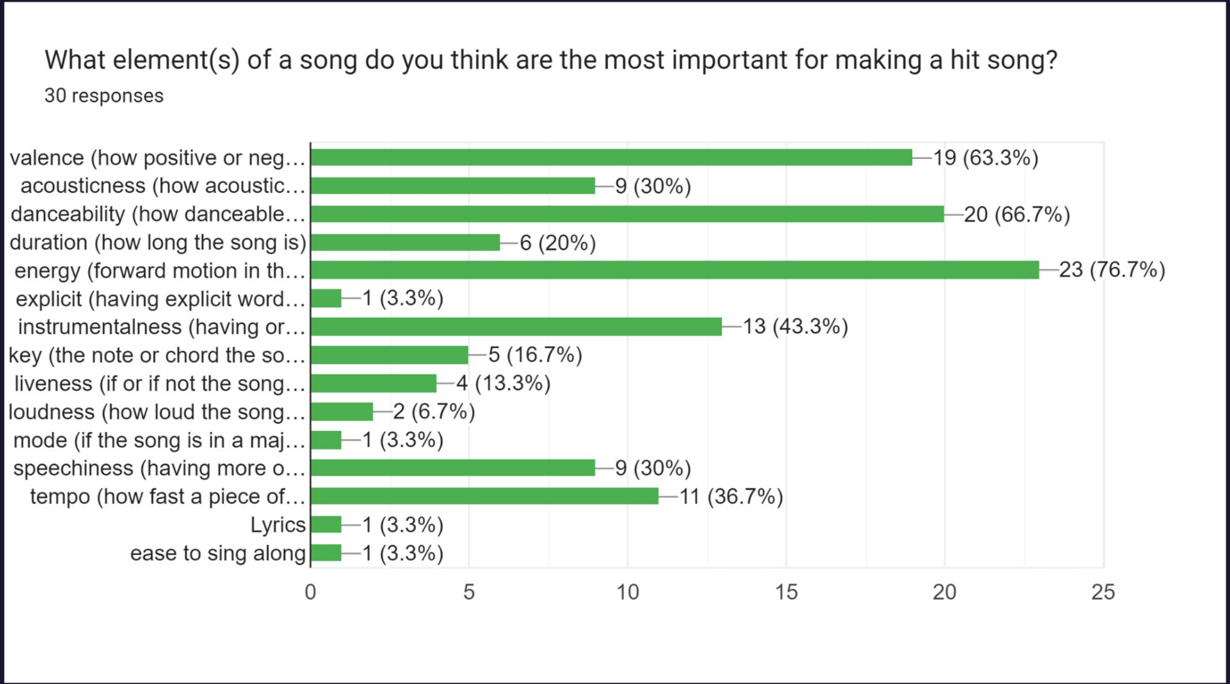
Scan the QR code and enter your answer!

Tick 5 boxes for your answers.

Closest/ correct answers will get a small reward!

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Survey Results



Top 5 Elements

1. Energy
2. Danceability
3. Valence
4. Instrumentalness
5. Tempo

Mars Is a Cold Place
The 15th Planet
2:54
3:49

02

Background

Mars Is a Cold Place

The 15th Planet

2:54

3:49

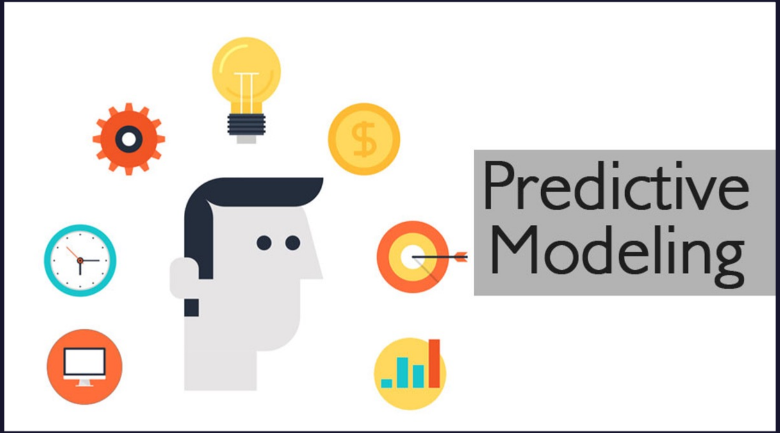
Dataset

- We obtained the dataset from Kaggle.
- An overview:

valence	year	acousticness	artists	danceability	duration_ms	energy	explicit	id	instrumentalness	key	liveness	loudness	mode	name	popularity	release_date	speechiness	tempo
0.0594	1921	0.982	['Sergei Rachmaninoff', 'James Levine', 'Berliner Philharmoniker']	0.279	831667	0.211	0	4BJqT0PrAfrxzMOxytFOIz	0.878	10	0.665	-20.096	1	Piano Concerto No. 3 in D Minor, Op. 30: III. Finale. Alla breve	4	1921	0.0366	80.954
0.963	1921	0.732	['Dennis Day']	0.819	180533	0.341	0	7xPhfUan2yNtyFG0cUWkt8	0	7	0.16	-12.441	1	Clancy Lowered the Boom	5	1921	0.415	60.936
0.0394	1921	0.961	['KHP Kridhamardawa Karaton Ngayogyakarta Hadiningrat']	0.328	500062	0.166	0	1o6I8BglA6ylDMrIELygv1	0.913	3	0.101	-14.85	1	Gati Bali	5	1921	0.0339	110.339
0.165	1921	0.967	['Frank Parker']	0.275	210000	0.309	0	3ftBPsC5vPBKxYSee08FDH	2.77E-05	5	0.381	-9.316	1	Danny Boy	3	1921	0.0354	100.109
0.253	1921	0.957	['Phil Regan']	0.418	166693	0.193	0	4d6HGyGT8e121BsdKmw9v6	1.68E-06	3	0.229	-10.096	1	When Irish Eyes Are Smiling	2	1921	0.038	101.665
0.196	1921	0.579	['KHP Kridhamardawa Karaton Ngayogyakarta Hadiningrat']	0.697	395076	0.346	0	4pyw9DVHGStUre4J6hPngr	0.168	2	0.13	-12.506	1	Gati Mardika	6	1921	0.07	119.824
0.406	1921	0.996	['John McCormack']	0.518	159507	0.203	0	5uNZnElqOS3W4fRmRYPk4T	0	0	0.115	-10.589	1	The Wearing of the Green	4	1921	0.0615	66.221
0.0731	1921	0.993	['Sergei Rachmaninoff']	0.389	218773	0.088	0	02GDntOXexBFUvSgaXLPkd	0.527	1	0.363	-21.091	0	Morceaux de fantaisie, Op. 3: No. 2, Prélude in C-Sharp Minor. Lento	2	1921	0.0456	92.867
0.721	1921	0.996	['Ignacio Corsini']	0.485	161520	0.13	0	05xDjWH9ub67nJJk82yfGf	0.151	5	0.104	-21.508	0	La Mañanita - Remasterizado	0	20/3/1921	0.0483	64.678
0.771	1921	0.982	['Fortugé']	0.684	196560	0.257	0	08zfJvRLp7pjAb94MA9JmF	0	8	0.504	-16.415	1	Il Etait Syndiqué	0	1921	0.399	109.378
0.826	1921	0.995	['Maurice Chevalier']	0.463	147133	0.26	0	0BMkRpQtDoKjcgzCpnqLNa	0	9	0.258	-16.894	1	Dans La Vie Faut Pas S'en Faire	0	1921	0.0557	85.146
0.578	1921	0.994	['Ignacio Corsini']	0.378	155413	0.115	0	0F30WM8qRpO8kdolepZqdM	0.906	10	0.11	-27.039	0	Por Que Me Dejaste - Remasterizado	0	20/3/1921	0.0414	70.37
0.493	1921	0.99	['Georgel']	0.315	190800	0.363	0	0H3k2CvJvHULnWChlbeFgx	0	5	0.292	-12.562	0	La Vipère	0	1921	0.0546	174.532
0.212	1921	0.912	['Mehmet Kemiksiz']	0.415	184973	0.42	0	0LcXzABeA84EgudqpNUN1I	0.89	8	0.108	-10.766	0	Ud Taksimi	0	1921	0.114	70.758
0.493	1921	0.0175	['Zay Gatsby']	0.527	205072	0.691	1	0MJZ4hh60zwsYleWWxT5yW	0.384	7	0.358	-7.298	1	Power Is Power	0	27/3/1921	0.0326	159.935
0.282	1921	0.989	['Sergei Rachmaninoff', 'Ruth Laredo']	0.384	221013	0.171	0	0NFeJgmTAV1kDfzSQNK41Z	0.82	7	0.116	-20.476	0	10 Préludes, Op. 23: No. 5 in G Minor. Alla marcia	4	1921	0.0319	107.698
0.218	1921	0.957	['Phil Regan']	0.259	186467	0.212	0	0Nk5f07H3JaEunGrYfbqHM	0.000222	2	0.236	-13.3	1	Come Back To Erin	1	1921	0.0358	85.726
0.664	1921	0.996	['Hector Berlioz', 'Arturo Toscanini']	0.541	250747	0.283	0	0POO8XaUgW5Qjbv9sbDy8W	0.898	9	0.393	-14.808	1	Rákóczy March	0	1921	0.0477	108.986
0.0778	1921	0.148	['THE GUY']	0.604	204957	0.418	1	0QQmUf4aPFaN9U2yRko595	0.0382	4	0.102	-11.566	0	When We Die	0	11/9/21	0.0417	80.073

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Predictors



- We 've removed "release_date", "name", "id", and "artists" columns.
- There are 14 predictors left for our model.
- It's a lot! So we expect to have some reasonable outcomes.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Response

- Our response variable, hence the predictive value, is **continuous**, instead of binary.

- We will utilize functions with modifications.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

03

Business Models

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Linear Regression

- Initial Linear Regression

  - We includes all predictors.
  - RMSE = 10.85, Rsquared = 0.75, MAE = 8.02

- Improved Linear Regression

  - We select some statistically significant predictors
  - RMSE = 10.81, Rsquared = 0.75, MAE = 8.01

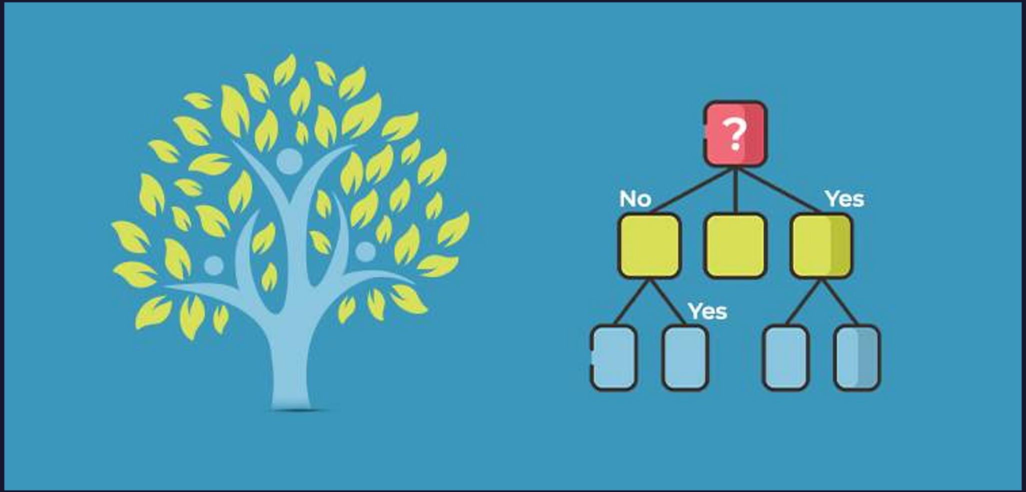
Mars Is a Cold Place
The 15th Planet

2:54

3:49

Regression Tree

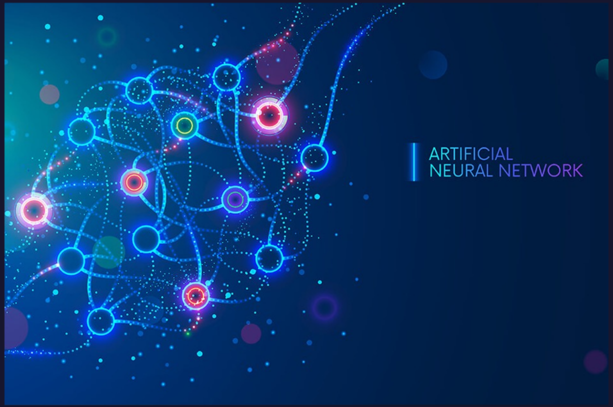
- This is similar to Decision Tree, but we will have a continuous output here.



Mars Is a Cold Place
The 15th Planet
2:54
3:49

Artificial Neural Network (ANN)

- Normalizing data
- Build model and try different number of layers
- Prediction



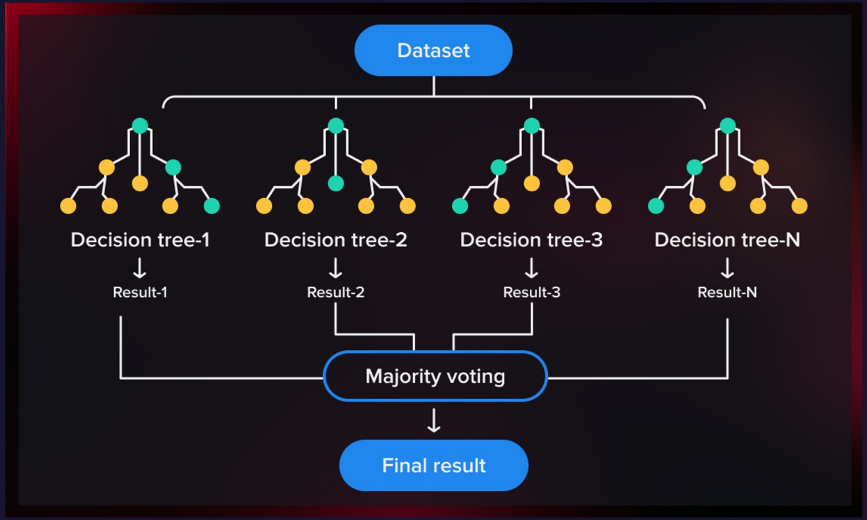
Mars Is a Cold Place
The 15th Planet
2:54
3:49

k-Nearest Neighbour Regression(KNNREG)

- Instead of KNN, we need to use KNNREG.
- Initial KNNREG: with all predictors included.
- Improved KNNREG: with selective predictors.

Random Forest

- A combination of Decision Trees.
- Initial Random Forest
- Improved Random Forest: increase the number of trees to 1200



Mars Is a Cold Place
The 15th Planet
2:54
3:49

Stacked Model

- **Stacked Regression Tree**
- Data are the predictions of previous 5 models
- We divide it to train/test and build a stacked model
- Make predictions on the test dataset and check against to test labels for performance

Mars Is a Cold Place
The 15th Planet
2:54
3:49

HOME

TABLE OF CONTENTS

PLAYLIST

THANKS!

04

Conclusions

Mars Is a Cold Place
The 15th Planet

2:54

3:49

Our parameters

- Instead of using those parameters for a binary case, we need to use other parameters to assess our model performance

- RMSE : average error between predicted values and actual values
- R-Squared : the proportion of the variance in the dependent variable that can be explained by the independent variables in the model
- MAE : the average absolute difference between predicted values and actual values.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Model Comparison

	Linear Regression	Regression Tree	ANN	KNNREG	RandomForest	Stacked Tree
RMSE	10.81	10.71	0.11	37.89	37.86	13.88
R-Squared	0.75	0.76	0.76	0.78	1.00	0.75
MAE	8.01	7.62	0.08	31.12	31.11	10.97

Mars Is a Cold Place
The 15th Planet
2:54
3:49

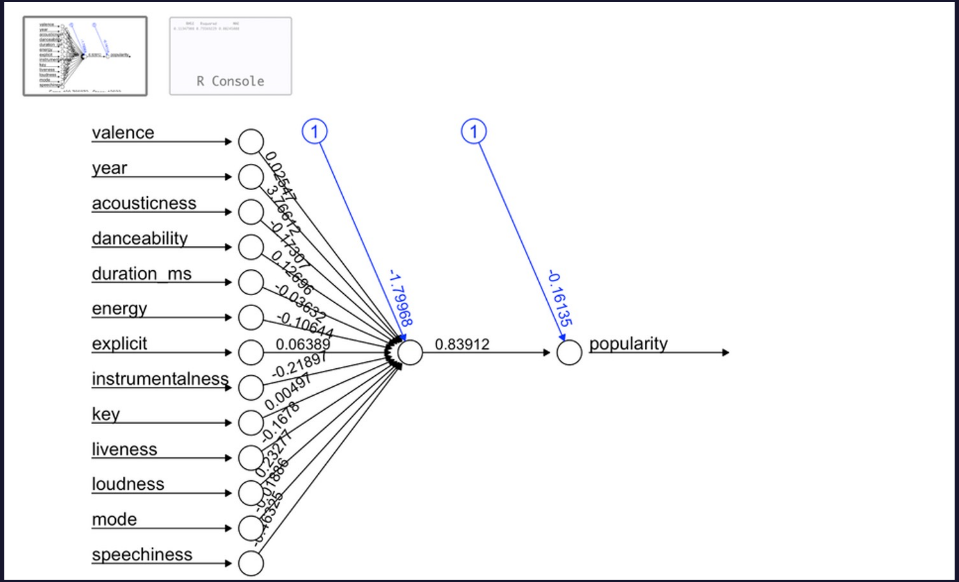
Model Interpretation

Top Individual Model: ANN (lowest RMSE and MAE)

Combined model:

- Both RMSE and MAE looks somehow like an average of previous 5 models
- R-Squared, however, it actually the lowest among all models

Model Interpretation



- We are using a one layer model.
- Predictors that affect the prediction of popularity are: year, loudness, danceability, explicit, and valence.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Model Interpretation

- **Loudness**: measures audio level based on the way humans perceive sound
- **Danceability** : measured through beat strength, overall stability, etc..
- **Explicit**: measures how explicit the lyrics are
- **Valence**: measures the musical positiveness
- **Year**: the importance of year may due to the growth of population.

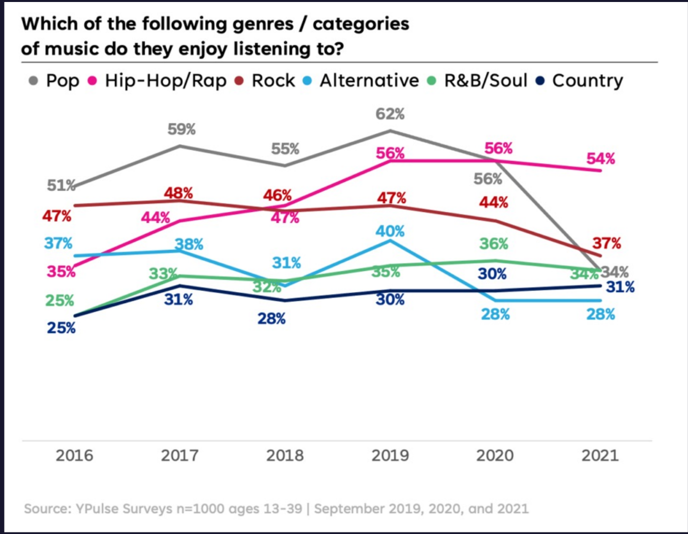
\- Two of the features (danceability and valence) matched our survey!

Mars Is a Cold Place
The 15th Planet
2:54
3:49

Further Discussions

- Our dataset contains information from 1921 until recent. It is highly possible that people's taste changed over time.

- If a question asked us targeting modern consumers, we could try using only the later half of the data, as they can be considered "modern", and we have enough rows to support it.



Mars Is a Cold Place
The 15th Planet
2:54
3:49

Further Discussion

- Musicians and record companies can use our model to predict the popularity of their new releases.

Mars Is a Cold Place
The 15th Planet
2:54
3:49

WINNERS (CONGRATULATIONS!)

(Jermaine to collate responses during presentation and fill in the winner(s) names)

Mars Is a Cold Place
The 15th Planet
2:54
3:49

THANK YOU FOR YOUR KIND ATTENTION!

Mars Is a Cold Place
The 15th Planet
2:54
3:49