AI choose celebrity song list

6/27 – 8/12

Use all my data science technique to predict which songs should be in a concert

* What songs do the audience like

Why these songs should be in the concert or not?

* <https://www.ericsson.com/en/blog/2022/4/differences-between-machine-learning-machine-reasoning>
* <https://towardsdatascience.com/what-is-reasoning-526103fd217>

Challenging Questions:

1. how does data difference affect model
2. understand the difference between model1 vs model2
3. Core fence index
4. How would you design a core fence

Question:

1. Singers will ask what songs should I add in the concert?

Baseline:

Rank these song and get top 10

Binary Classification

1. (true positive + true negative) / total
2. True positive / (top 20)

Need to have same performance metrics

I have a song list but I need to choose top 20

Metircs, baseline

What ZeroR measures?

* Measuring the percentage of songs that aren’t in concert

What log regression measures?

* Sing or not song

What random forest measures?

* Measuring if the model correctly predicts if the song is in the concert or not
  + We can separate songs that are in and not in the concert
* Accuracy = TP + TN

What we want our model to measure?

* Need a ranking
* Can’t sing all songs in a concert

What my baseline fails to measure?

Why balancing data is important?

* avoid distribution bias in predictive ability

kNN

1. explain the difference in logic of different regression
2. understand the main 3 classifier
3. problem
4. a (explainable) model everyone can understand