# Explaining Music Popularity Using Audio Features and Machine Learning

By: Deema Hazim and Ameera Alrahmah

# What audio features matters the most for popularity?

- Features that are positively correlated with being popular (or tend to be more popular) include danceability, loudness, and tempo.
- On the other hand, instrumentalness and valence (cheerfulness) show the strongest negative correlation with popularity in this model.
- Computation of the R-squared showed that these audio features explain only 2% of what makes songs successful

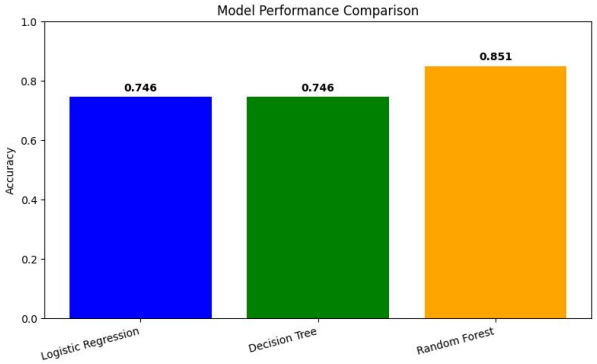
| | Feature | Coefficient |
|---|---|---|
| 0 | danceability | 1.593170 |
| 2 | loudness | 0.580323 |
| 8 | tempo | 0.436624 |
| 6 | liveness | 0.261692 |
| 4 | acousticness | -0.466508 |
| 1 | energy | -0.769545 |
| 3 | speechiness | -1.291957 |
| 7 | valence | -2.537125 |
| 5 | instrumentalness | -2.577483 |

# Can we classify songs as “hit” vs “non-hit”?

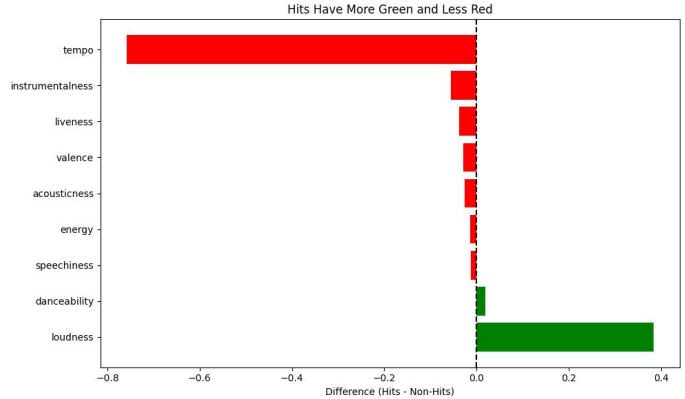
Best Model: Random Forest

Higher thresholds = better accuracy

Loudness+Danceability = Hit

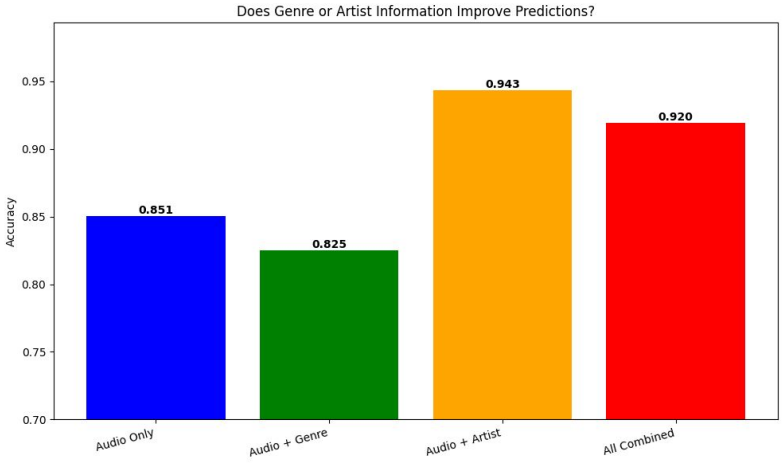



```
=== Does the hit definition matter? ===

Threshold 40: 42.6% of songs are hits
  Accuracy: 0.775

Threshold 50: 25.8% of songs are hits
  Accuracy: 0.851

Threshold 60: 13.0% of songs are hits
  Accuracy: 0.933

Threshold 70: 4.8% of songs are hits
  Accuracy: 0.978
```

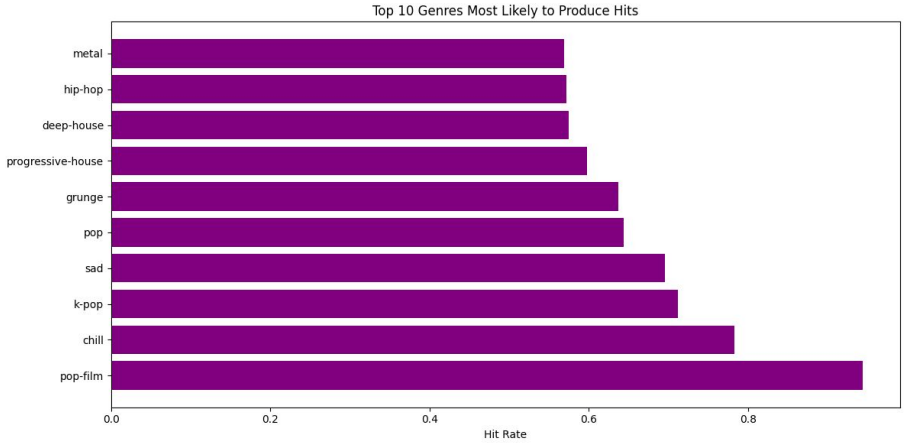


# Do genre or artist popularity improve prediction accuracy?

Audio + Artist had highest accuracy




Genre Hit Rate Analysis



# PCA Results

```
Component 1 - Top features:
loudness         0.516955
energy           0.509169
acousticness     0.439340
Name: Comp 1, dtype: float64

Component 2 - Top features:
danceability         0.574753
valence              0.522034
instrumentalness     0.305671
Name: Comp 2, dtype: float64

Component 3 - Top features:
liveness         0.678620
speechiness      0.641286
acousticness     0.227287
```

- Using PCA, we discovered that the nine audio features cluster into three underlying dimensions:

    1. Energy versus Acoustic (32%)
    2. Danceability and Mood (16%)
    3. Live Performance (14%)

- These explain 61% of how songs differ from each other.

- However, none of these dimensions predict popularity well (hits span all styles)