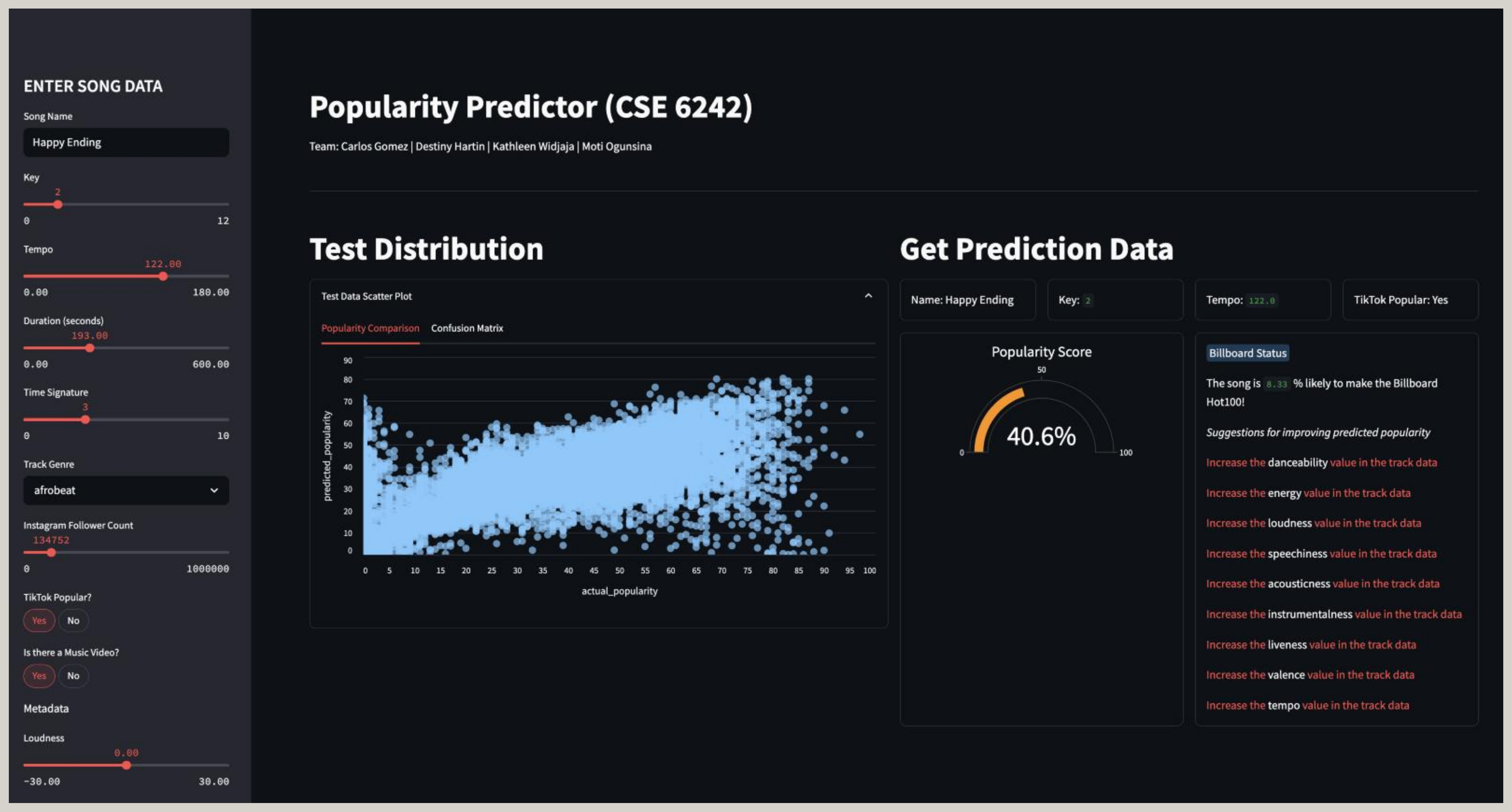
The Hit Song Science Problem

Kathleen Widjaja
Destiny Hartin
Carlos Gomez
Motiloluwa Ogunsina

INTRODUCTION

What makes a song a hit? Can we predict musical success before it happens?

In our project, we tackle the Hit Song Science Problem by asking whether it's possible to predict a song's popularity based not just on its musical features (like tempo or loudness) but also on external signals, such as the artist's social media following or whether the song has trended on TikTok. While past research tends to focus on only one of these factors, our work combines both musical and social signals to build a more complete model of what drives success. We also go a step further by comparing different definitions of what it means for a song to be "popular" and testing how well our model predicts each one. By identifying which features matter most, our results can help artists and labels focus their creative and promotional strategies toward maximizing a song's potential for success.



Above shows a screenshot of our interactive visualization which predicts song success by comparing features which are important in a song's success. The scatterplot compares actual vs predicted values and users will be able to hover over points to see information about the song.
The user will also be able to input variables and gain recommendations on increasing popularity of the song, setting it apart from other current methods.

DATA

We utilized publicly available Spotify dataset on Kaggle, which includes both musical features and popularity scores.
To enhance this base, we merged in additional data sources: APIs for artist information and social media stats (e.g., The Audio DB, Instagram Follower Scraper), a dataset of TikTok trending songs, and a Billboard Hot 100 archive. We expanded song entries to the artist-song level, helping us match across datasets.
The final dataset contains over 158,000 records and 55 features, covering audio characteristics, social metrics, and metadata like genre and artist gender. However, not all features were used in the model due to null values from missing data in the chosen APIs. After cleaning, the dataset totals around 73 MB and was processed using Python and R.

OUR APPROACH

To predict whether a song becomes a hit, we developed a suite of machine learning models, including Random Forest, Gradient Boosting, Neural Networks, KNN, and AdaBoost. These models were trained using a combination of audio features (such as tempo and loudness) and social media signals (like TikTok virality, YouTube music video presence, and Instagram follower count).
Our models tackle the problem from two angles: predicting a song's Spotify popularity score (0–100), and determining the probability that a song appears on the Billboard Hot 100.
What sets our approach apart is its integration of diverse data sources and definitions of success. While most prior studies rely heavily on musical attributes alone, we expand the feature space to include artist-level and platform-driven variables, reflecting how modern-day hits are influenced by visibility as much as sound. This hybrid perspective not only improves predictive accuracy but also offers insights into how different factors contribute to a song's breakout potential.

EXPERIMENTS

To test our models, we structured the problem in two ways: predicting a song's Spotify popularity score and classifying whether it appeared on the Billboard Hot 100. We split our dataset into training and testing sets and evaluated each model using appropriate metrics: Mean Squared Error for regression, and accuracy, precision and recall for classification.
Each model was trained twice, once for each target variable, using the same set of features. These included both traditional audio characteristics (like energy, danceability, and tempo) and social media-based variables (such as TikTok trending status, Instagram followers, and music video presence). This allowed us to see how well each model could generalize across different definitions of success.

RESULTS

The random forest was most successful in both. We observed improvements to similar studies in our model, although certain biases must be taken into consideration such as a vast majority of samples not making it onto the Billboard Hot 100. Overall, only certain external factors increased predictive power.

	Popularity Score (0 – 100)	
Algorithm:	Test Mean Square Error	R^2
Random Forest	**205**	**0.51**
Gradient Boosting	212	0.50
Neural Network	423	0
KNN Regression	285	0.32
AdaBoost Regressor	394	0.07

Table 1 – Test metrics for popularity score

	Billboard 100			
Algorithm:	Accuracy	Precision	Recall	F1 Score
Random Forest	**0.96**	**0.69**	**0.58**	**0.63**
Gradient Boosting	0.88	0.26	0.67	0.37
Neural Network	0.05	0.05	1	0.09
KNN	0.91	0.33	0.59	0.42
AdaBoost	0.90	0.30	0.64	0.41

Table 2 – Test metrics for billboard classification