

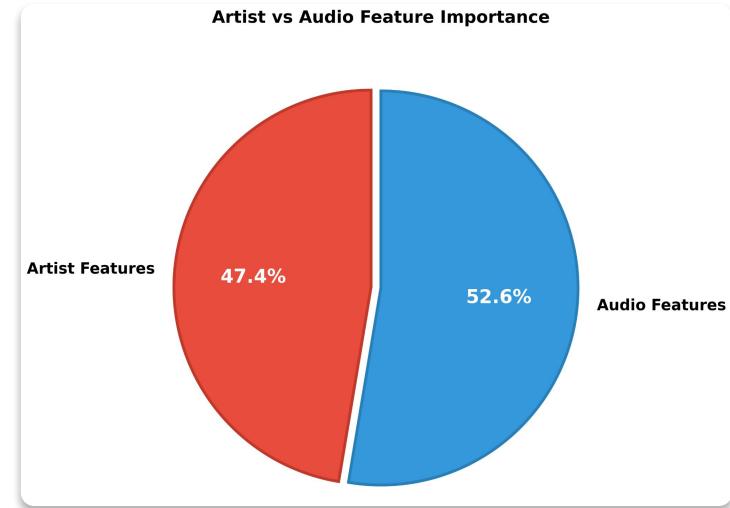
# Predicting Billboard Hits from Audio Features

Kevin Rutledge

Knowledge Discovery and Data Mining Portfolio Project

# What Drives Hit Prediction?

- Artist identity rivals audio quality in hit prediction.<sup>1,2</sup>
- Artist features account for 47% of prediction power.



<sup>1</sup> Stanford CS229, 2018; <sup>2</sup> Nature, Fraiberger et al., 2018

# The \$7.1 Billion Question

- Music industry spent \$7.1B on A&R and marketing in 2023.<sup>1</sup>
- Labels need to predict hits before signing artists.
- Current gut-instinct methods are ineffective.



<sup>1</sup> IFPI Global Music Report, 2024

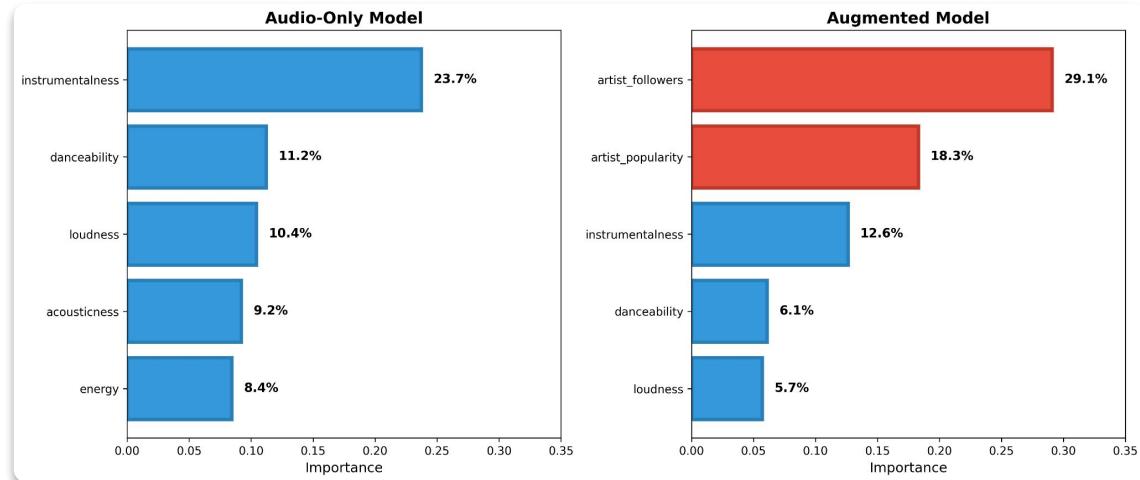
# Methodology: Three Progressions



What made a hit in 1965 differs from 2015.

# What Features Drive Predictions?

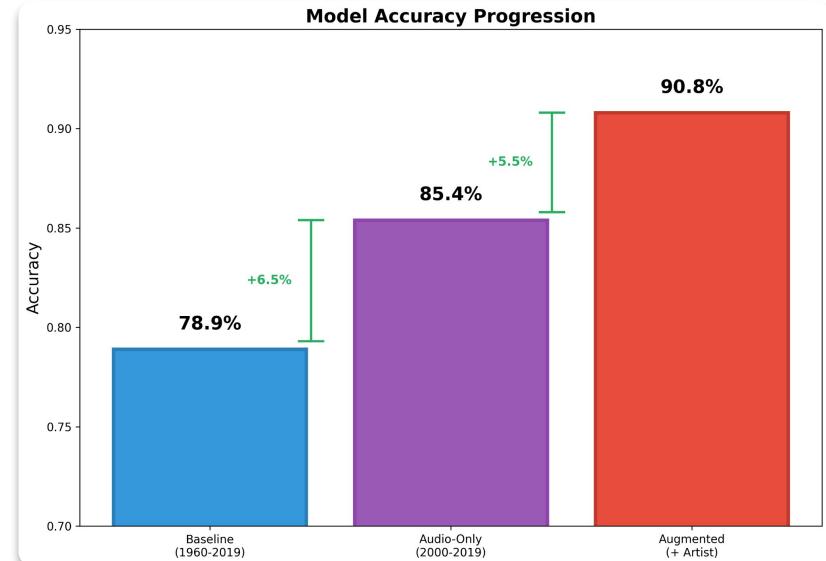
- **Audio-only:** Instrumentalness dominates (24%).<sup>1,2</sup>
- **Augmented:** Artist features take over (47% combined).



<sup>1</sup> Stanford CS229, 2018; <sup>2</sup> Nature, Fraiberger et al., 2018

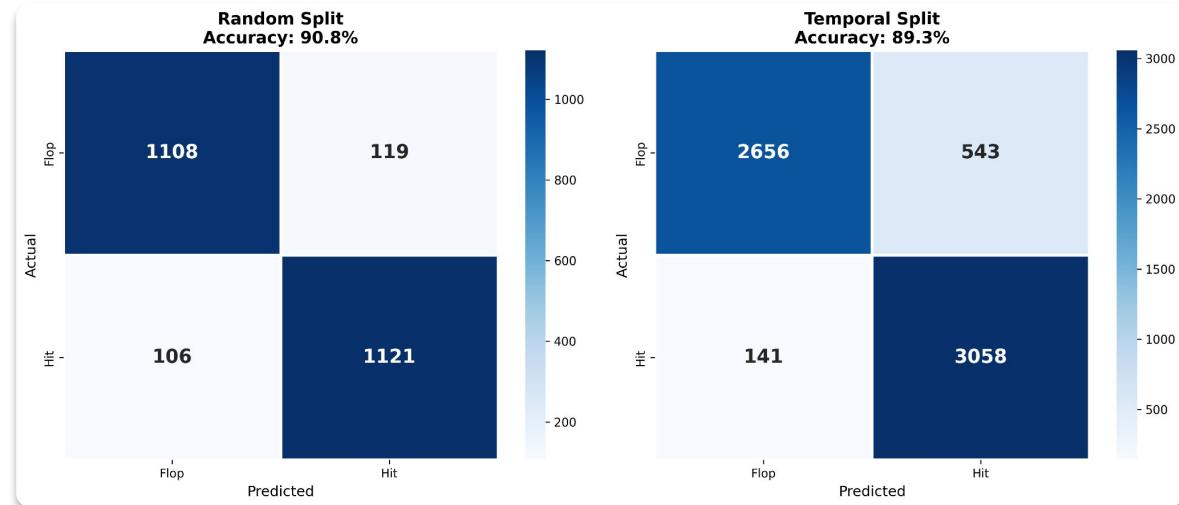
# The Path to 91% Accuracy

- Data quality beats sophisticated algorithms.
- XGBoost and hyperparameter tuning added nothing.



# Model Validates on Future Data

- Model holds up on future data (only 1.5% drop).
- Matches published benchmarks (86-90%).<sup>1-3</sup>

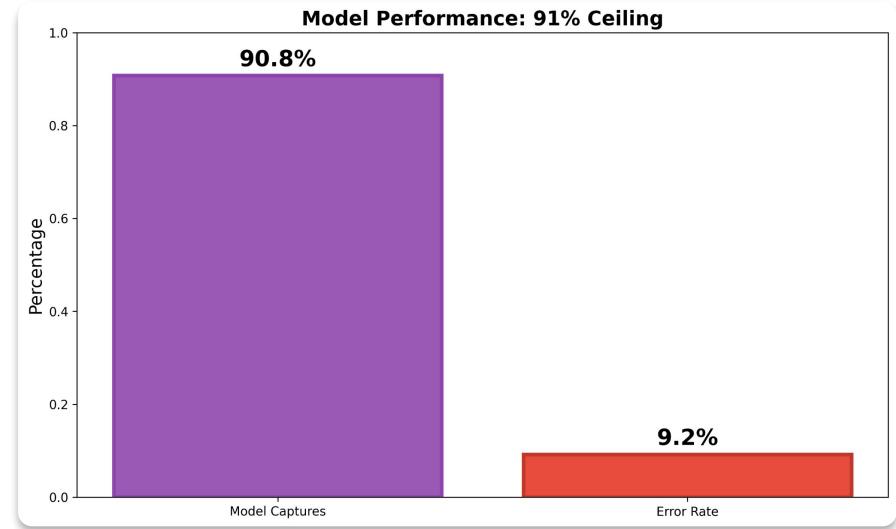


<sup>1</sup> Middlebrook, 2019; <sup>2</sup> SpotHitPy, 2023; <sup>3</sup> Mountzouris, 2025

# Why Further Gains Are Limited

## What it misses (9%):

- **TikTok viral moments (84% of 2024 hits)**<sup>1</sup>
- **TV/film sync placements**<sup>2</sup>
- **Social contagion (20-30% random variance)**<sup>3</sup>



<sup>1</sup> TikTok & Luminate, 2024; <sup>2</sup> Wikipedia - Kate Bush revival via Stranger Things; <sup>3</sup> Salganik et al., Science 2006

# Questions?

Kevin Rutledge

[github.com/kevinrutledge/](https://github.com/kevinrutledge/)