

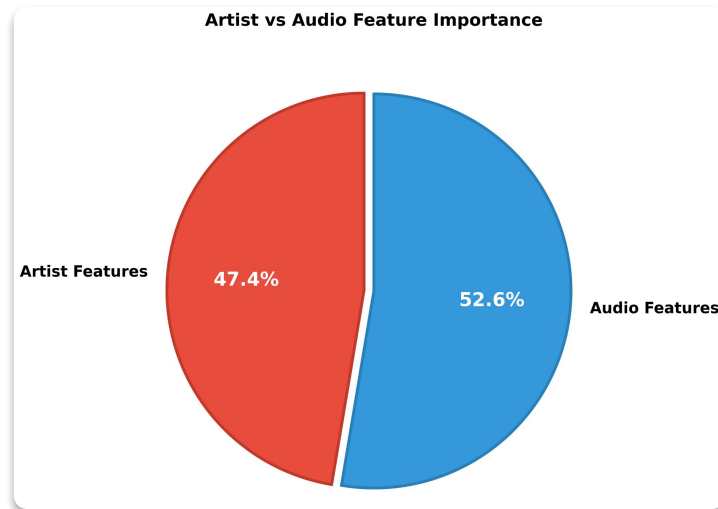
Predicting Billboard Hits from Audio Features

Kevin Rutledge

Knowledge Discovery and Data Mining Portfolio Project

What Drives Hit Prediction?

- Artist identity rivals audio quality in hit prediction.^{1 2}
- Artist features account for 47% of prediction power.



¹ Stanford CS229, 2018; ² Nature, Fraiberger et al., 2018

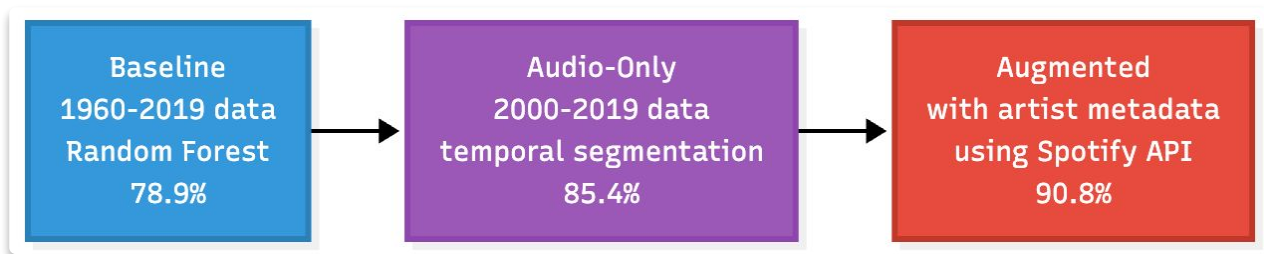
The \$7.1 Billion Question

- Music industry spent \$7.1B on A&R and marketing in 2023.¹
- Labels need to predict hits before signing artists.
- Current gut-instinct methods are ineffective.



¹ IFPI Global Music Report, 2024

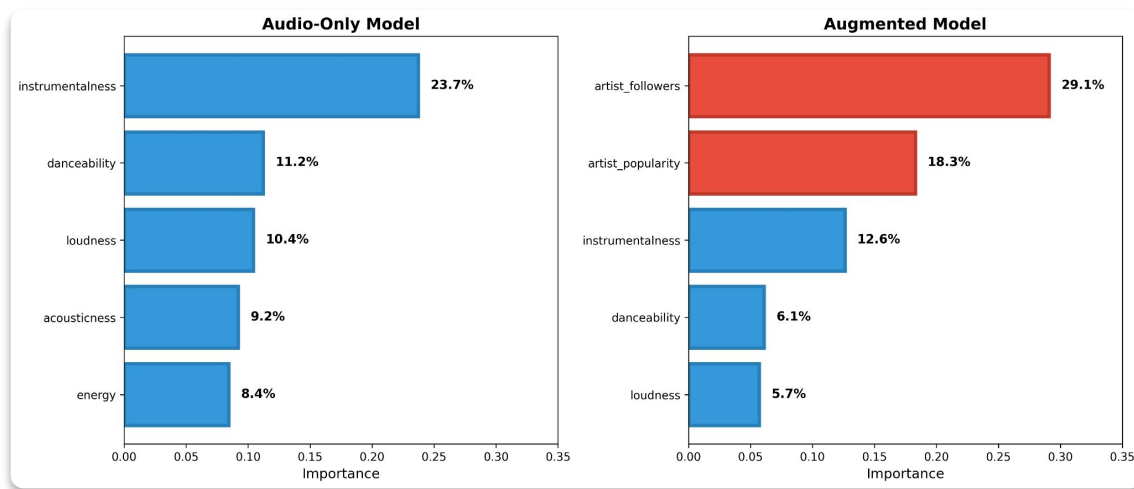
Methodology: Three Progressions



What made a hit in 1965 differs from 2015.

What Features Drive Predictions?

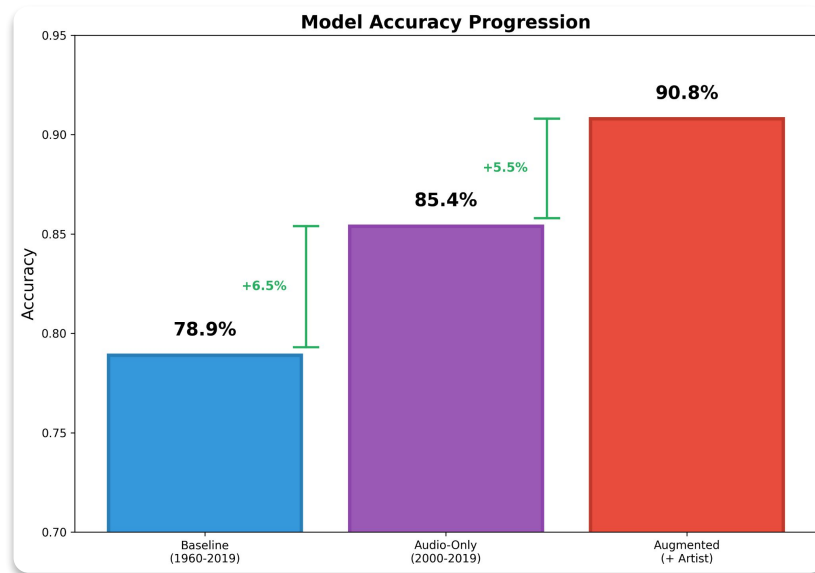
- **Audio-only:**
Instrumentalness dominates (24%).^{1 2}
- **Augmented:** Artist features take over (47% combined).



¹ Stanford CS229, 2018; ² Nature, Fraiberger et al., 2018

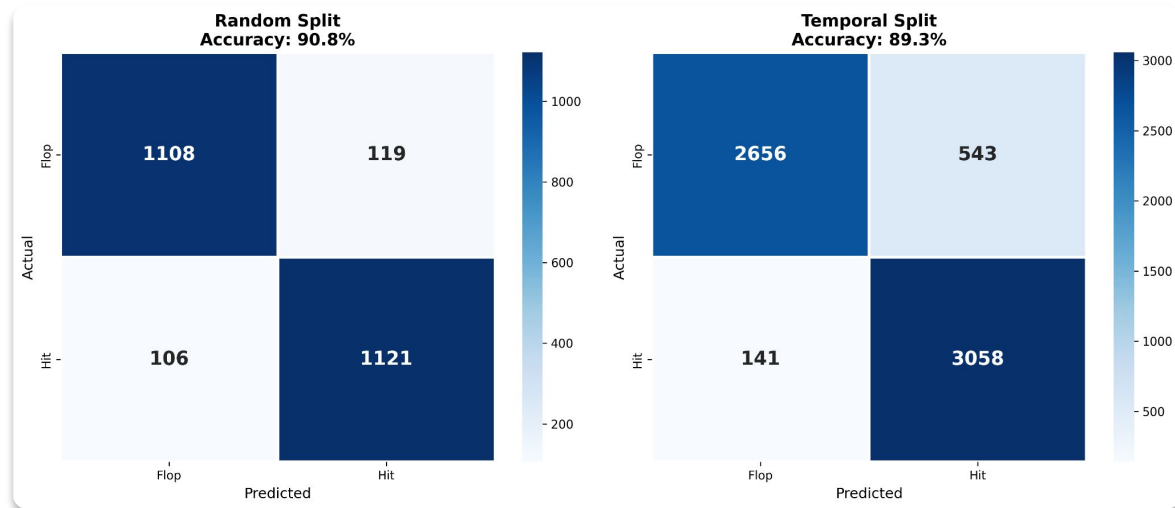
The Path to 91% Accuracy

- 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%).¹⁻³

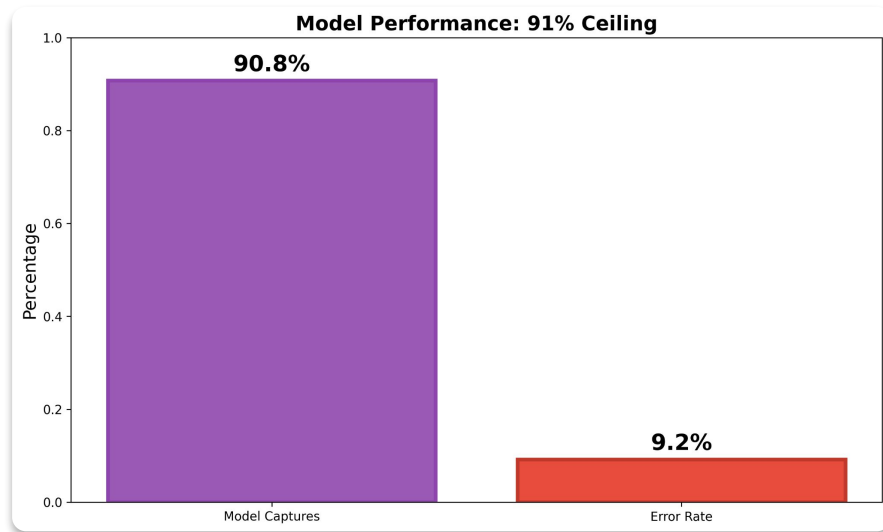


¹ Middlebrook, 2019; ² SpotHitPy, 2023; ³ Mountzouris, 2025

Why Further Gains Are Limited

What it misses (9%):

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



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

Questions?

Kevin Rutledge

github.com/kevinrutledge/