Predicting Spotify Hits: A Multimodal Classification Modeling Approach

01 Motivation & Goal

- 100K+ songs uploaded daily; <1% succeed
- Success depends on artist momentum + song content
- Need an early, scalable prediction model

02 Data Overview

Base Dataset

1.16M Spotify tracks with audio features + metadata

Added modalities

generated text description

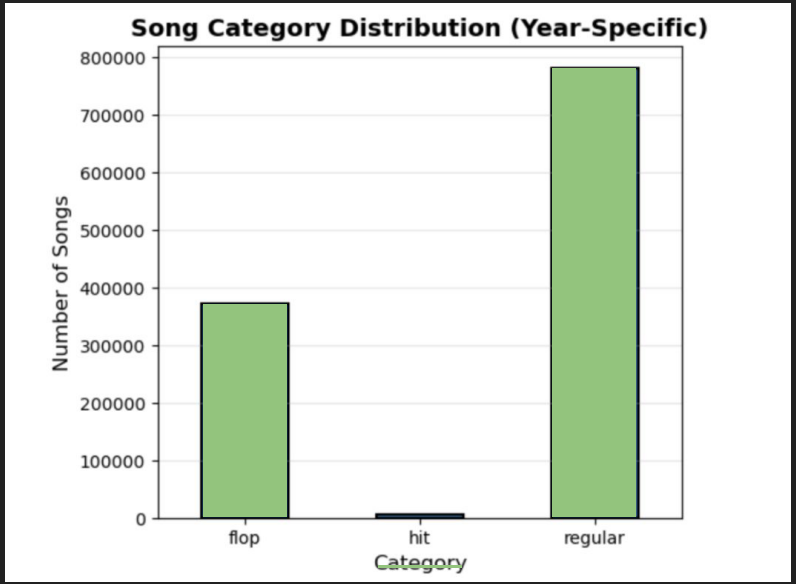
30s audio previews

lyrics

album cover

03 Key Preprocessing

- Year-normalized popularity → flop / regular / hit



04 Key Preprocessing

Artist tier

Last.fm listener percentile → S/A/B/C/F/U tier

atist_prev_pop

Average past Spotify popularity

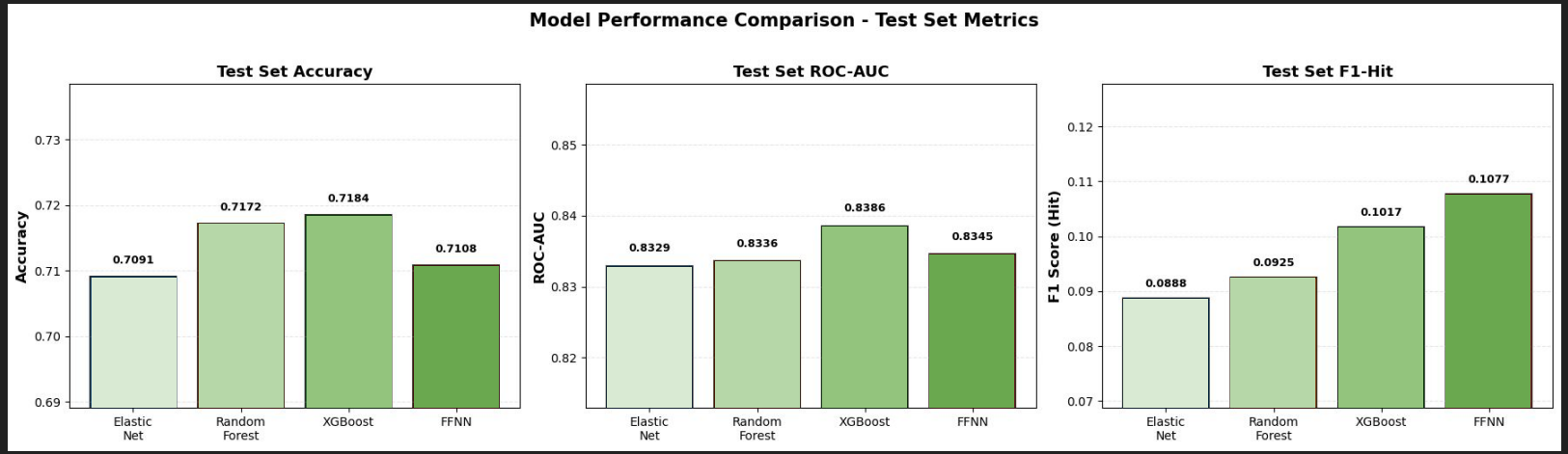
Undersample majority classes

5.5K per class to address extreme class imbalance

Time-based split

train: 2000–21, test: 2022–23

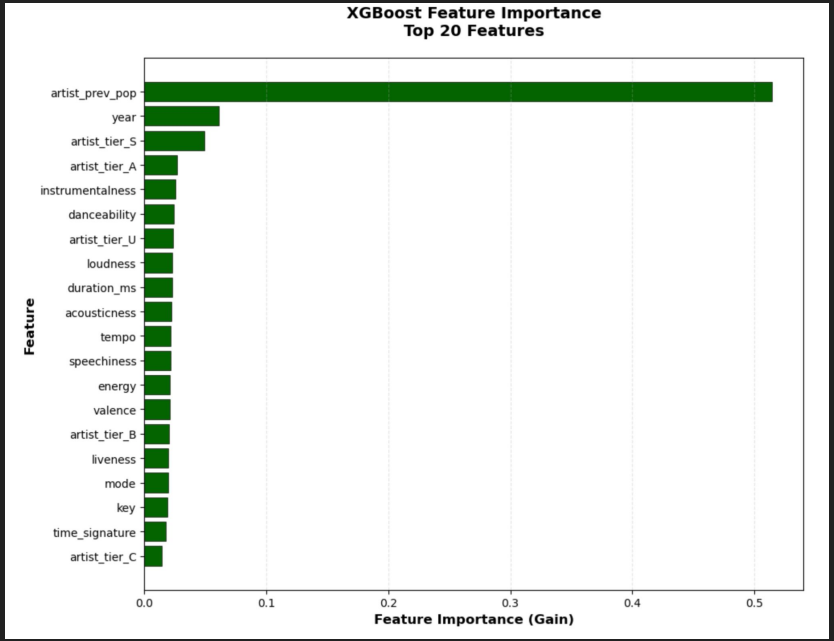
05 Basline Model Selection



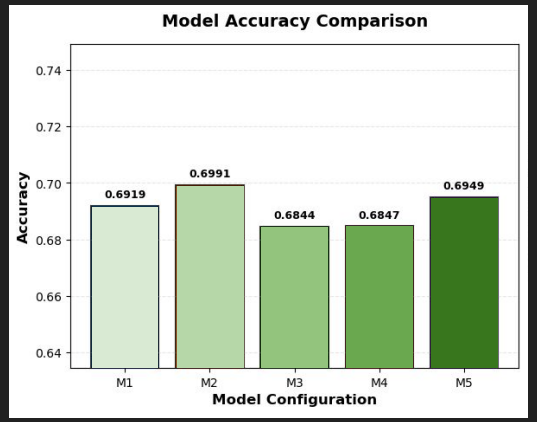
We select FFNN as the baseline

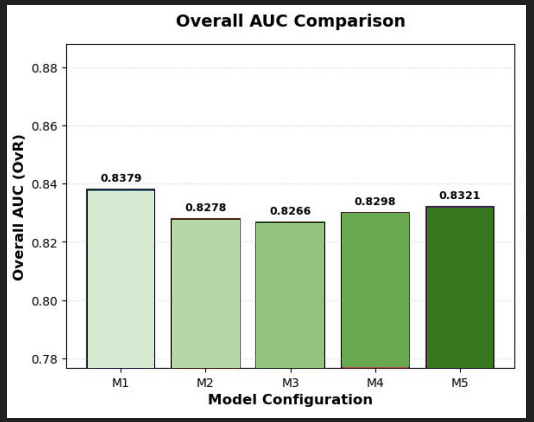
05 Basline Model Selection

- artist_prev_pop dominates all other features by a large margin

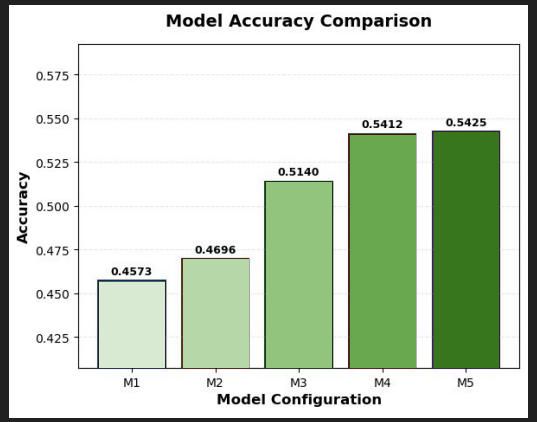


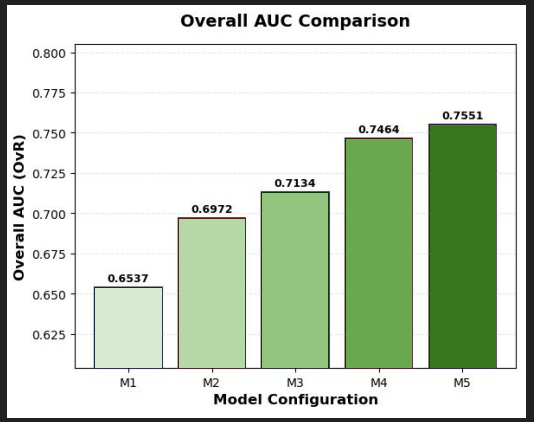
06 Multimodal Pipeline and Result: With artist_prev_pop





07 Multimodal Pipeline and Result: Without artist_prev_pop





08 Key Insights

- Artist history drives song success more than content
- Multimodal signals matter when history is weak or unavailable
- Best use cases: cold-start recommendations, new artists, early content screening

Thank you!