Audio Analysis and Intelligence Integration Using ML & LLMs

1. Instrument Detection and Proportion Estimation

This is an identification of music instruments in an audio recording and calculation of their relative percentages by machine learning models such as YAMNet. It is done through preprocessing (normalization, resampling), feature extraction (spectral, model embeddings), and classification to map audio segments into instrument classes. Results are given in JSON format, highlighting the percentage contribution of each instrument (e.g., guitar: 35%). The solution spans a broad variety of instruments, resolves overlapping sounds, and stores analysis automatically for future reuse. Methodology shown below,

Audio Input → Preprocessing → Feature Extraction → Classification → Instrument Mapping → Proportion Calculation → JSON Output

Fig.1 Methodology of Instrument Detection and Proportion Estimation

2. Predicting Virality of the Song

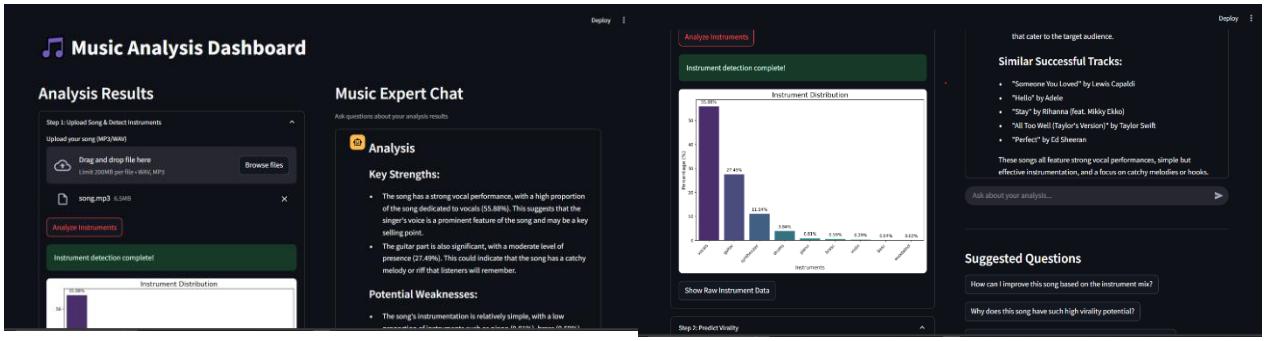
In this section, created an end-to-end song virality prediction system by resolving data quality problems, handling extreme class imbalance using strategic sampling, and feature engineering key audio features (mood score, energy-danceability interaction). Dataset using [Spottify-1-Million-Track]. In modeling, we crafted a bespoke ANN architecture with dropout and batch normalization optimized for recall on viral cases, which are few. Used SHAP/LIME for interpretability. The deployment itself entailed the development of a simple-to-use Gradio interface, feature mismatch resolution through training/inference pipeline alignment, and the incorporation of strong port handling. Simplifying the preprocessing pipeline, ensuring backward compatibility with dummy columns by being consistent in predictions while keeping explanations open using visualization tools, were some of the most important enhancements. This end-to-end solution brings data science to real-world use, offering accurate predictions as well as actionable intelligence for use in the music domain.

Data Cleaning → Feature Engineering → Imbalance Handling → Model Development → Explainability → Interface Design → Error Resolution → Deployment

Fig.2 Methodology of Predicting virality of song.

3. Future driven LLM based decision

We enhanced the music analysis dashboard by implementing a professional two-panel interface with interactive visualizations and AI-powered chat. The left panel displays instrument detection results and virality predictions through clean, annotated charts, while the right panel offers a conversational interface with the LLM expert. The system now features improved error handling, responsive design, and suggested questions to guide users through comprehensive song analysis.



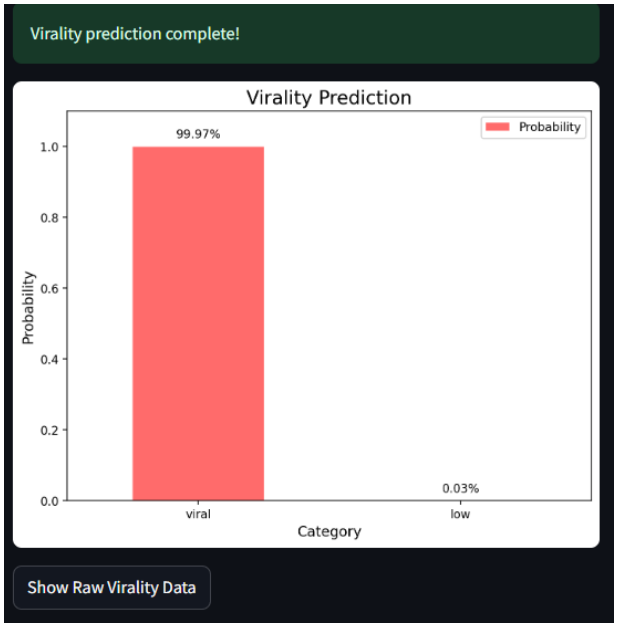


Fig.3 Glimpse of proposed solution.

Methodology:

Data Collection → Feature Extraction → Visualization → Virality Prediction → LLM Integration → Chat Interface

Fig.4 Methodology of LLM based decision.

Limitation and Potential Improvements

The current virality prediction system uses YAMNet for instrument detection, which is not optimized for musical instruments and has limited predictive power. The system's features lack depth in content and context, including lyric analysis, cultural trends, and social media signals. Real-time analysis is hindered by fixed windows and Groq API's free-tier rate restrictions. The system requires users to manually input song features, making it time-consuming and error-prone. Deployment challenges include complex containerization and the 200MB size of each model file. To improve audio analysis, the system could fine-tune on custom instrument datasets, incorporate Retrieval-Augmented Generation, and offer interactive charts with Plotly. Deployment optimization could be achieved by converting models to ONNX format, automating feature extraction, and offering artist-specific dashboards.