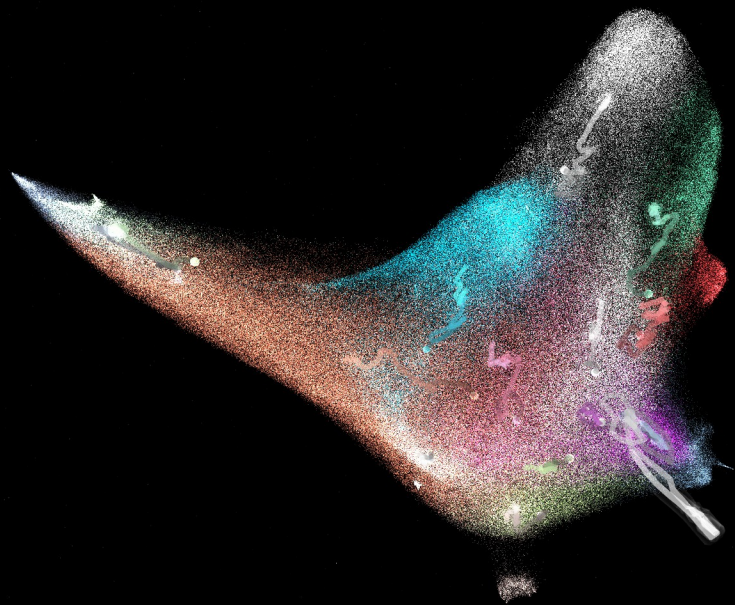


components ✖ genre classification from audio features

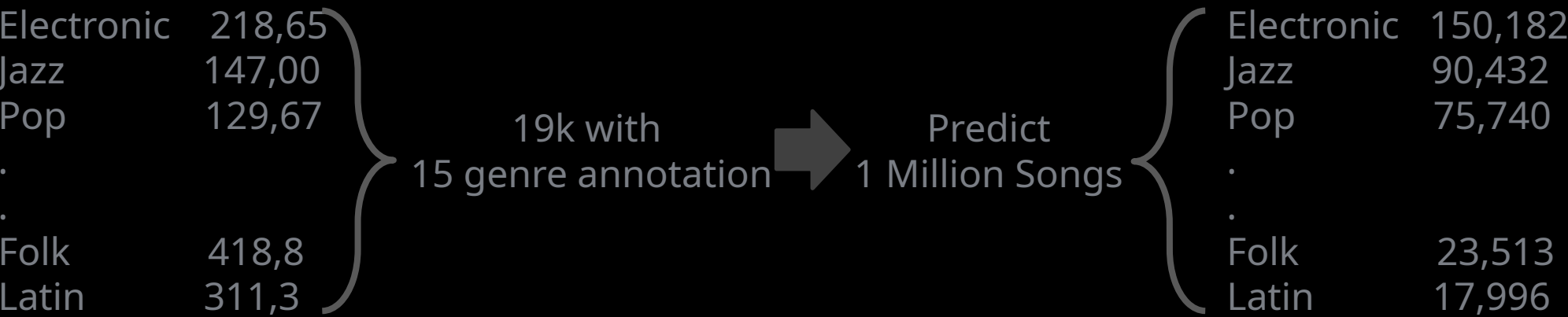
Research Objective 📊

Train a **supervised classification model** to transfer genre knowledge from 190k labeled songs to 1 million complete dataset, and achieve **spatial visualization** of music styles through **dimensionality reduction**.



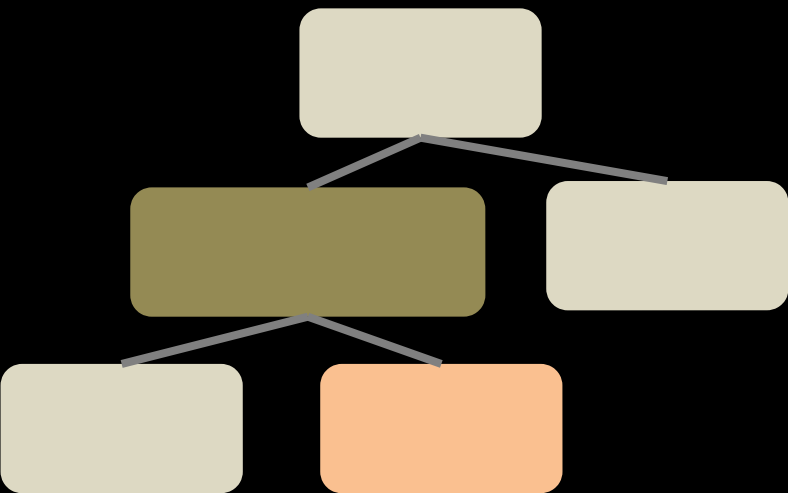
Genre Annotation Dataset 🗄️

Use the CD2C dataset provided by **tagtraum industries**, containing 191,401 songs with **15 genre annotations** for training the supervised classification model.



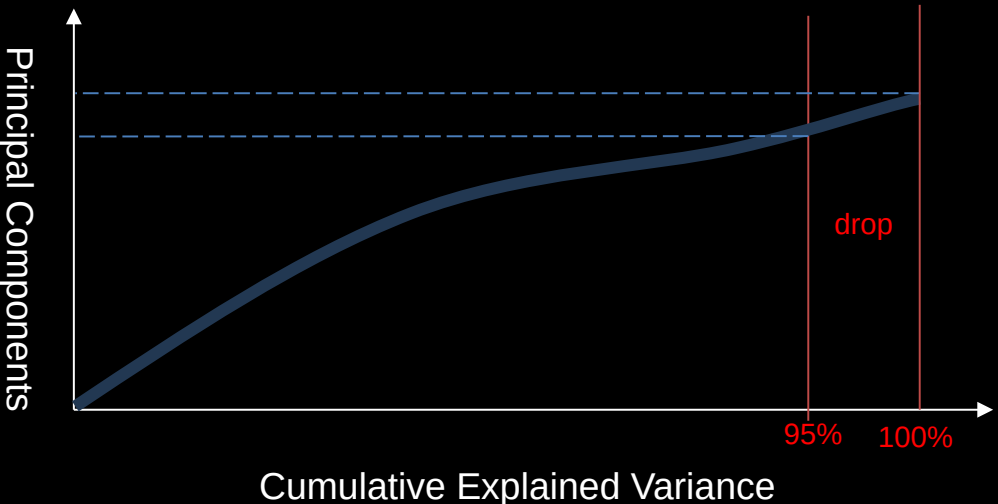
Model Selection 🏗️

Selected XGBoost for its class **imbalance handling** and **tree structure** that models hierarchical derivation relationships between genres.
(e.g., Rock → Metal → Heavy Metal)



Dimensionality Reduction 🎵

•**PCA (272 → 50):**
Keep main variance directions, reduce noise



•**UMAP (50 → 3):**
Preserve local structure for visualization

