

Jad Khater – CNET ID: jadkhater

Cross-Platform Music Streaming Quality of Experience

Music streaming services (Spotify, Apple Music, YouTube Music, Tidal etc) dynamically adapt audio quality based on network conditions. Unlike video streaming, users often multitask while listening—browsing, working, or using other applications—creating mixed traffic patterns that complicate quality prediction. This project investigates whether ML models can predict music streaming quality from network traffic features, and critically, whether models trained on one platform generalize to others. This addresses the course theme of non-representative training data: if trained on Spotify, will the model work on Apple Music in production? The project involves feature engineering for encrypted traffic, cross-platform transfer learning experiments, and comparison to video QoE from Assignment 1. Related work includes video streaming QoE prediction and any literature on adaptive bitrate (DASH, Pensieve). The goal is research: understand transferability of QoE models across platforms, identify platform-agnostic versus platform-specific network features, with practical implications for network operators encountering multiple platforms.

Data

Primary Source: MIRAGE-2019 Dataset (has spotify)

- Public mobile app traffic dataset including music streaming (Spotify)
- Available: <https://traffic.comics.unina.it/mirage/mirage-2019.html>
- Pre-labeled traffic from multiple platforms with realistic usage patterns

MAppGraph Dataset (requested from authors)

- Paper: "Mobile-App Classification on Encrypted Network Traffic using Deep Graph Convolution Neural Networks"
- Link: <https://www.researchgate.net/publication/354687255>
- Contains encrypted mobile app traffic including Spotify, SoundCloud
- Will email authors requesting access
- **SoundCloud Traffic Data** (from related research)
 - Reference: BMC Research Notes paper
(<https://bmcresearchnotes.biomedcentral.com/articles/10.1186/s13104-024-06718-7/tables/1>)
 - Will check data availability statement and request if available

Backup/Supplementary:

- Self-captured traffic from Spotify/Apple Music/YouTube Music with controlled quality settings if needed to fill gaps

Requirements: Flow-level features (5-tuple statistics) from 3-4 platforms, ground truth quality labels, sufficient samples per platform/quality level (50-100 sessions), mixed traffic scenarios. Will use NetML for feature extraction, avoiding destination IPs to prevent CDN overfitting.

Deliverables

1. **Clean Jupyter Notebook:** End-to-end reproducible analysis with data loading, feature extraction (NetML), cross-platform transfer experiments (leave-one-out, single-platform, all-platform training), model training (Random Forest, Logistic/Linear Regression), evaluation metrics (confusion matrices, precision/recall/F1, ROC curves), feature importance analysis, clear markdown documentation, tested with "Restart Kernel and Run All"

2. **Sphinx-Formatted Report:** Introduction, related work, methodology, results (within-platform and cross-platform performance with visualizations), discussion (why models transfer/fail, platform-agnostic vs. specific features, music vs. video comparison), inline code snippets, professional writing suitable for portfolio
3. **Supporting Materials:** README with run instructions, requirements.txt, processed data files with documentation, data source links/citations