



SOCIAL MEDIA HIT PREDICTION SUITE

MIUUL 19. DATASCIENTIST BOOTCAMP
FINAL PROJECT

Enes TOP – Alpay PAŞALI
Final Sunum

PROJE ÖZETİ

- AI destekli sosyal medya içerik hit tahmin platformu
- Spotify müzik hit potansiyeli analizi
- Gerçek zamanlı tahmin ve öneriler
- Production-ready Django web uygulaması

PROBLEM TANIMI

- Sosyal medya içerik üreticileri hangi içeriğin viral olacağını bilmiyor
- Manuel deneme-yanılma süreci zaman ve kaynak israfı
- Müzik prodüktörleri ve DJ'ler hit potansiyelini önceden göremiyor
- Instagram Reels içerikleri için çok yönlü analiz eksikliği
- Veri bilimi tekniklerini pratik bir problemde uygulama ihtiyacı

ÇÖZÜM YAKLAŞIMI

- Spotify API ve Kaggle datasets ile zengin veri toplama
- Machine Learning modelleri ile hit potansiyeli tahmini
- Çok yönlü analiz: Ses (+ Görsel + Metin)
- Ensemble learning (XGBoost, LightGBM, Random Forest vb.)
- Real-time web arayüzü ile kullanıcı dostu deneyim

SİSTEM MİMARİSİ

FRONTEND

- Django
- Tailwind CSS
- Chart.js Seaborn Visualizations
- JavaScript

DATA PROCESSING

- EDA & Feature Engineering
- Data Preprocessing

ML MODELS

- XGBoost Regressor
- LightGBM
- Random Forest

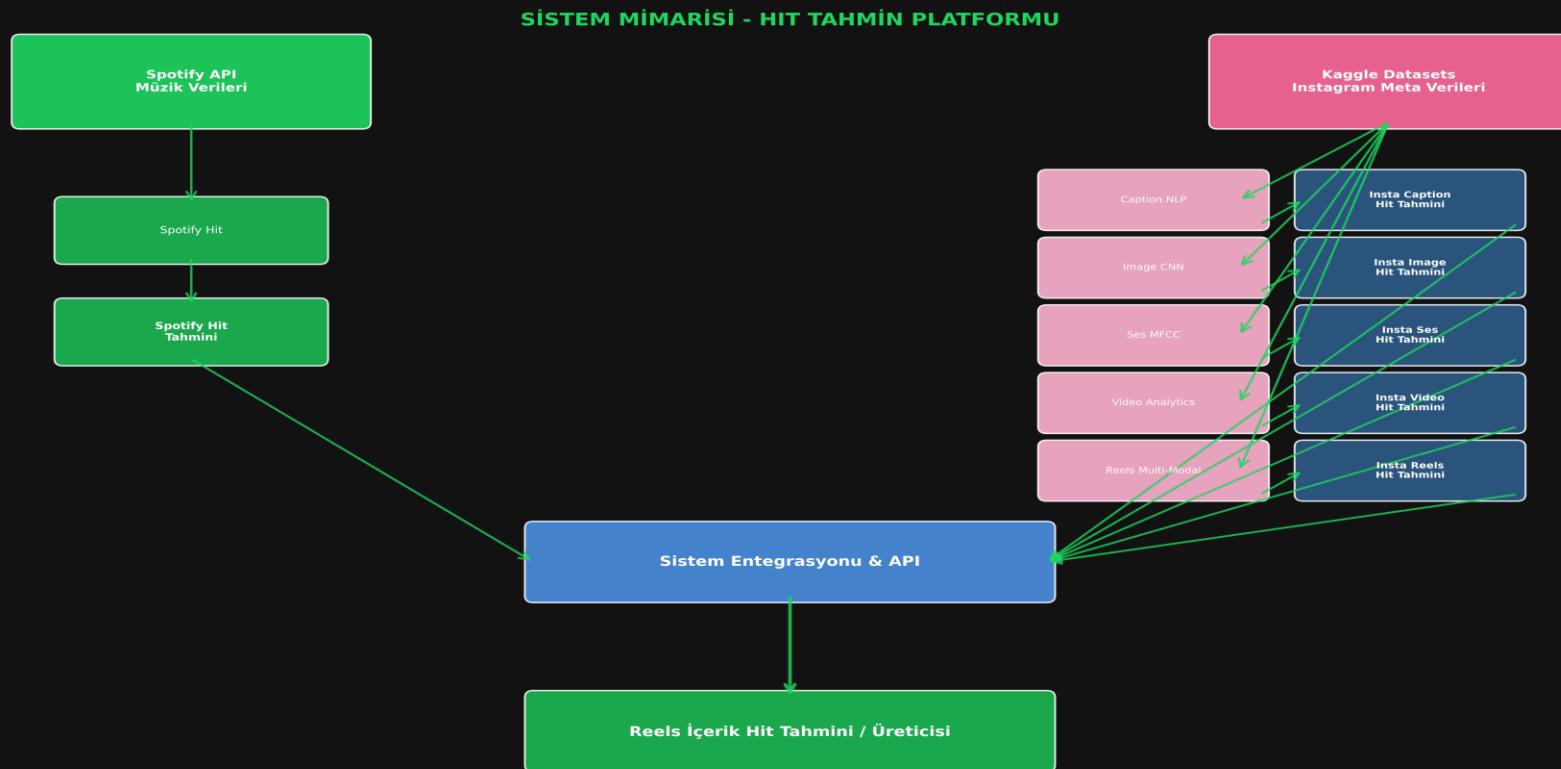
BACKEND

- Django REST Framework
- API (Spotify)

DEPLOYMENT

- Docker & GitHub & Grafana
- PostgreSQL Ready
- AWS Ready

SİSTEM MİMARİSİ - HIT TAHMİN PLATFORMU



- Feature extraction
- Tahmin modülleri
- Sistem entegrasyonu ve API katmanı
- Hit tahmini ve üretici

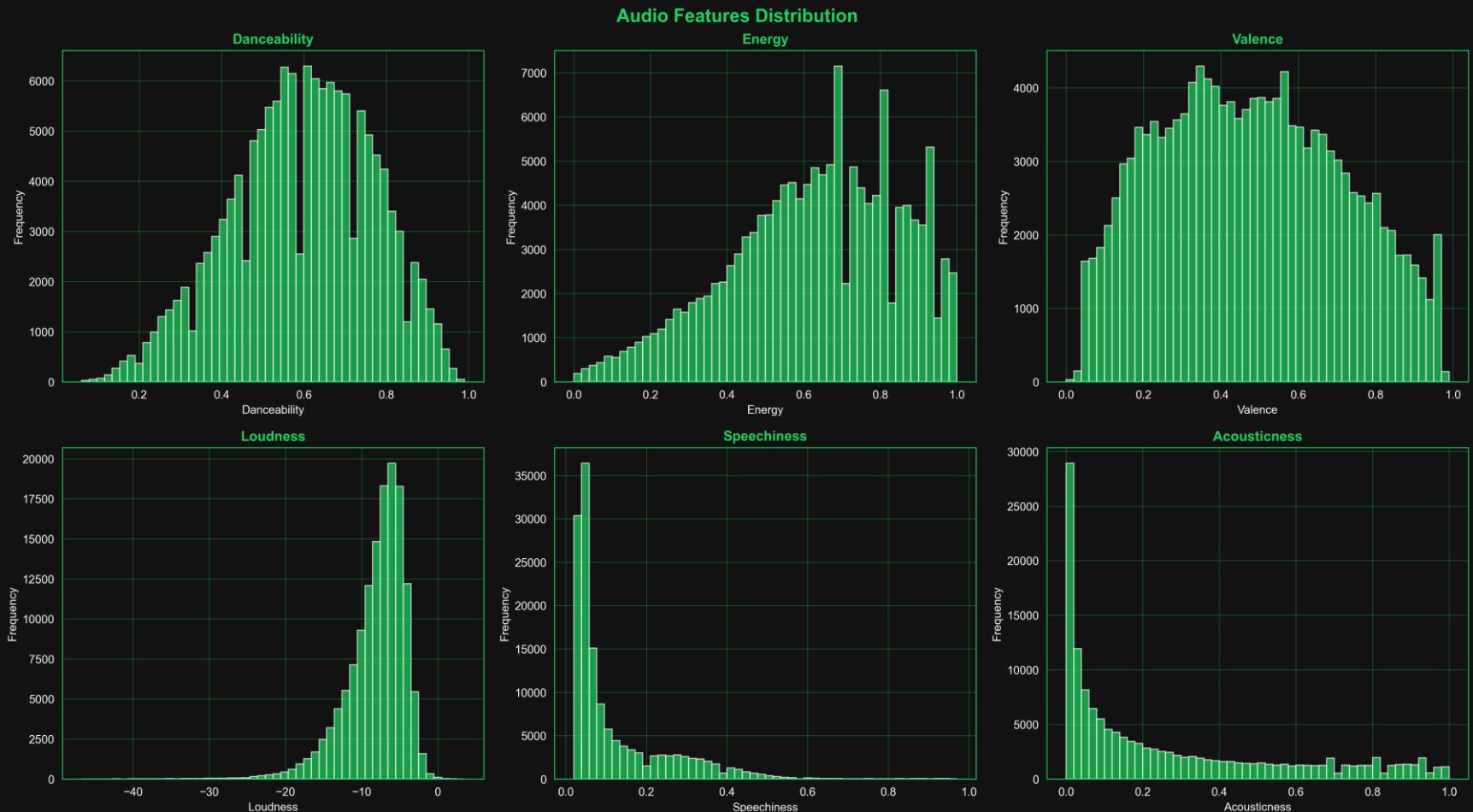
VERİ VE METODOLOJİ

- Spotify Dataset: 141K+ şarkı, audio features, popularity metrics
- Feature Engineering: Categorical encoding, rare label handling
- Preprocessing Pipeline: Automated data cleaning ve transformation
- Model Selection: Voting Regressor (ensemble of 4 models)
- Evaluation: RMSE, R² Score, cross-validation
- Similarity Matching: Normalized Euclidean distance

EXPLORATORY DATA ANALYSIS

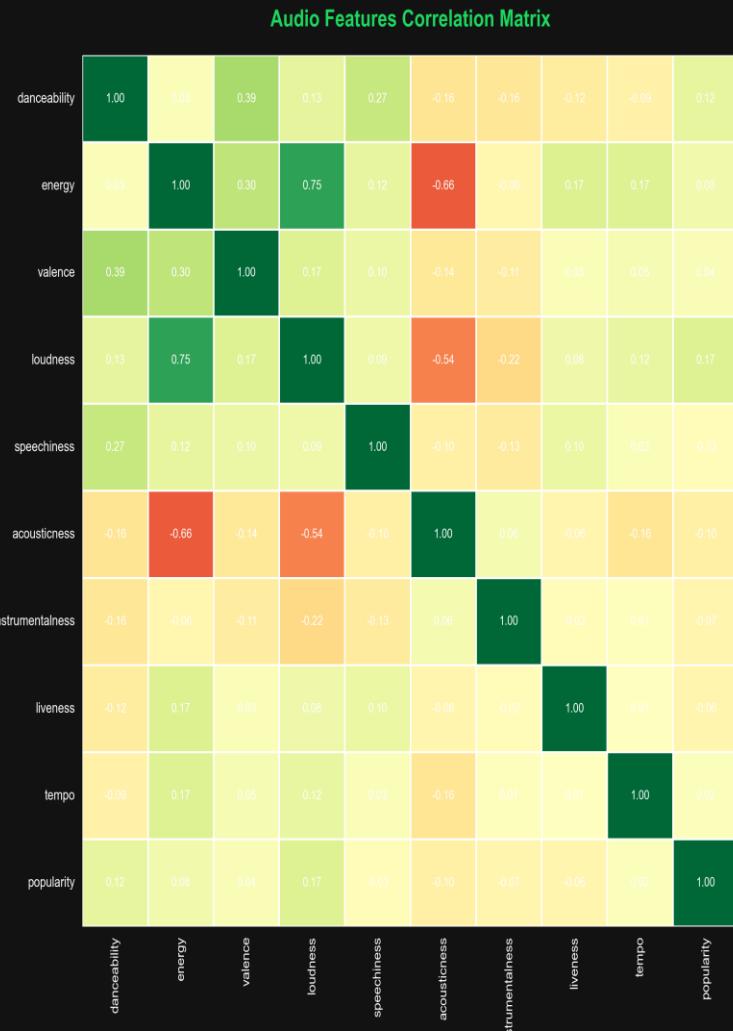
- Toplam 141,190 şarkı verisi analiz edildi
- 17 özellik (feature) kolonu: audio features + metadata
- Audio features: danceability, energy, valence, loudness, tempo, speechiness
- Metadata: popularity, genre, release_year, explicit, emotion
- Çoklu veri kaynağı: Spotify API + Kaggle datasets
- Piyasaya Çıkış Yılları: 1920-2024 arası şarkılar

AUDIO FEATURES DAĞILIMI



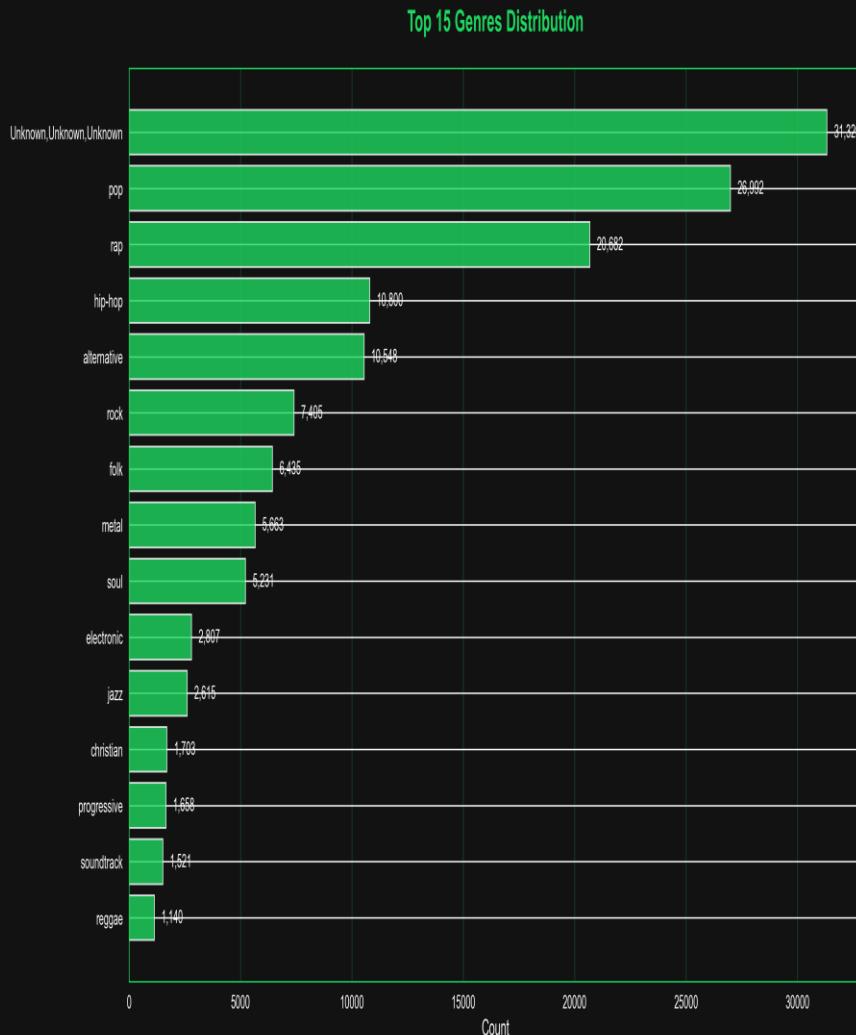
- Tüm audio feature'lar normal dağılıma yakın
- Danceability ve energy yüksek varyans gösteriyor
 - Loudness negatif değerler (dB cinsinden)

FEATURE CORRELATION MATRIX



- Energy ve loudness güçlü pozitif korelasyon (0.75+)
- Danceability ve valence orta pozitif korelasyon (0.5)
- Tempo diğer feature'larla zayıf korelasyon
- Popularity ile energy ve danceability pozitif ilişki

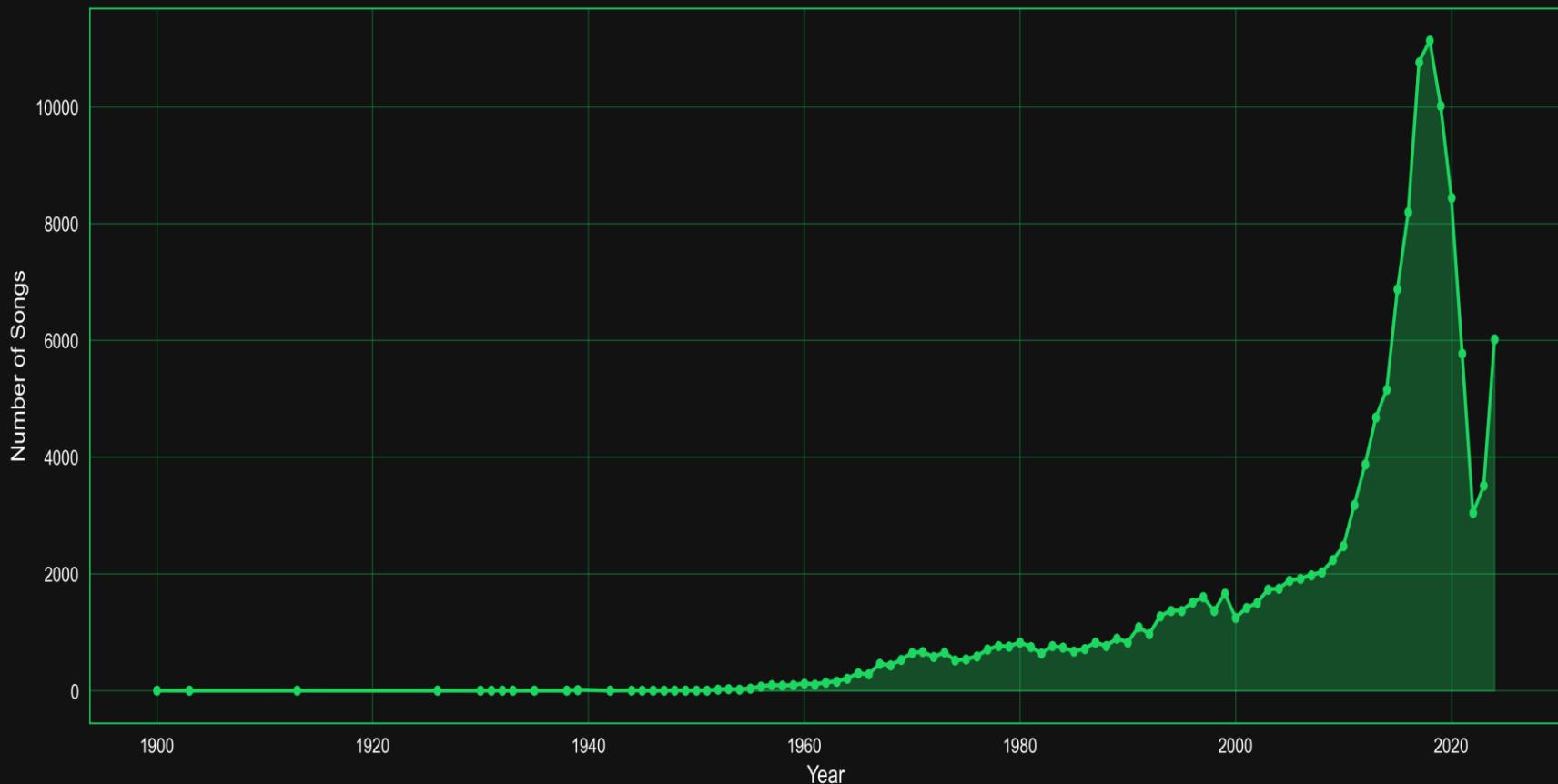
GENRE DAĞILIMI



- Pop, rock, hip-hop en yaygın türler
- Electronic ve alternative önemli paya sahip
- Genre bazlı hit potansiyeli farklılık gösteriyor

YIL BAZLI DAĞILIM

Songs Distribution by Release Year



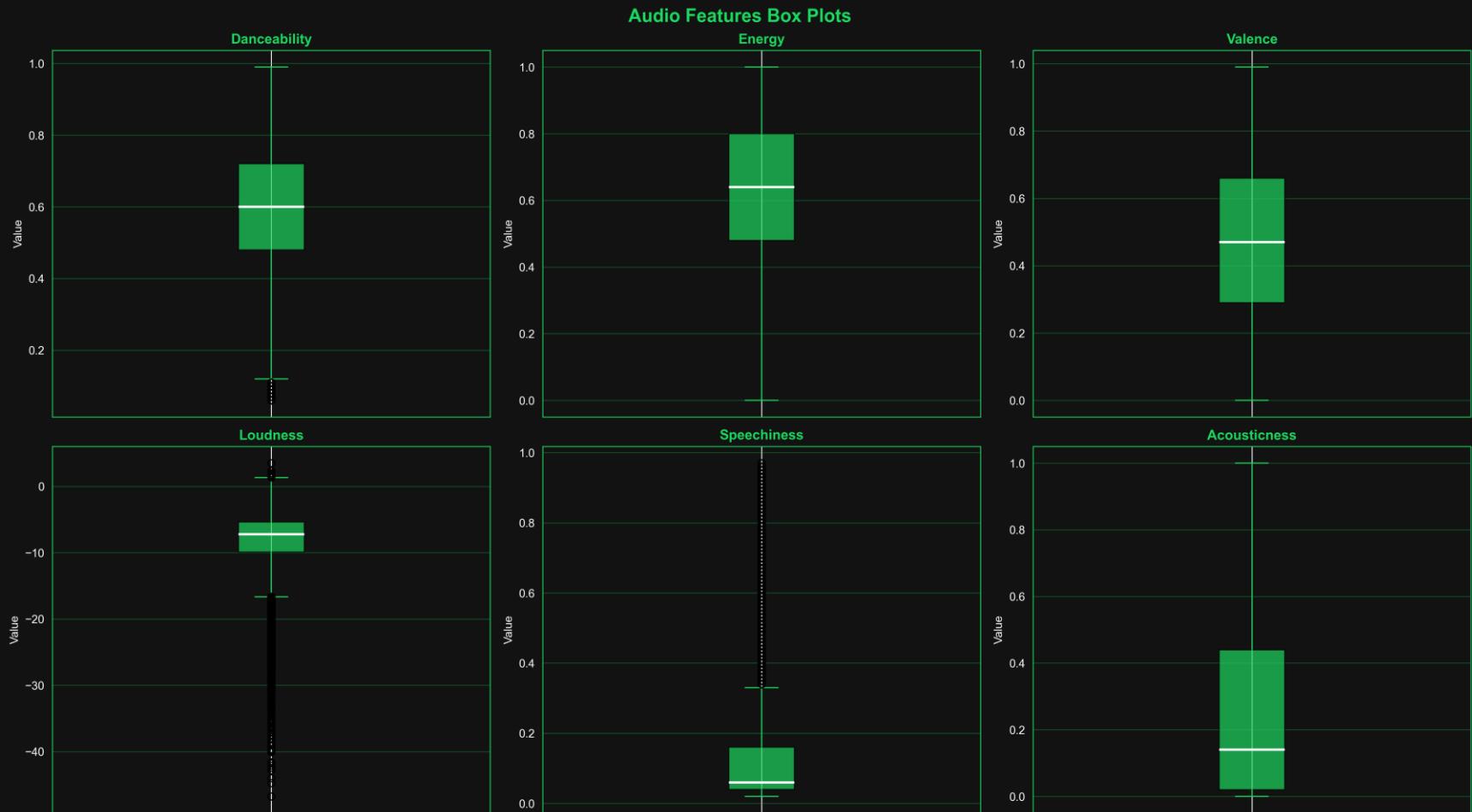
- Veri seti 1920-2024 arası şarkıları kapsıyor
 - 2010 sonrası veri yoğunluğu artıyor

AUDIO FEATURES EVOLUTION



- Danceability zamanla artış trendi gösteriyor
- Energy seviyeleri 2000'lardan sonra yükselmiş
- Tempo modern müzikte daha yüksek
- Müzik tarzları zamanla değişiyor

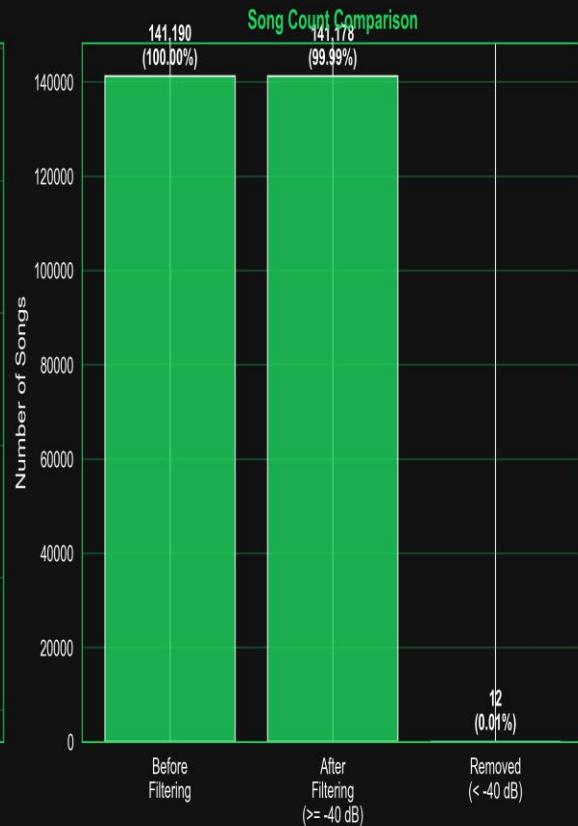
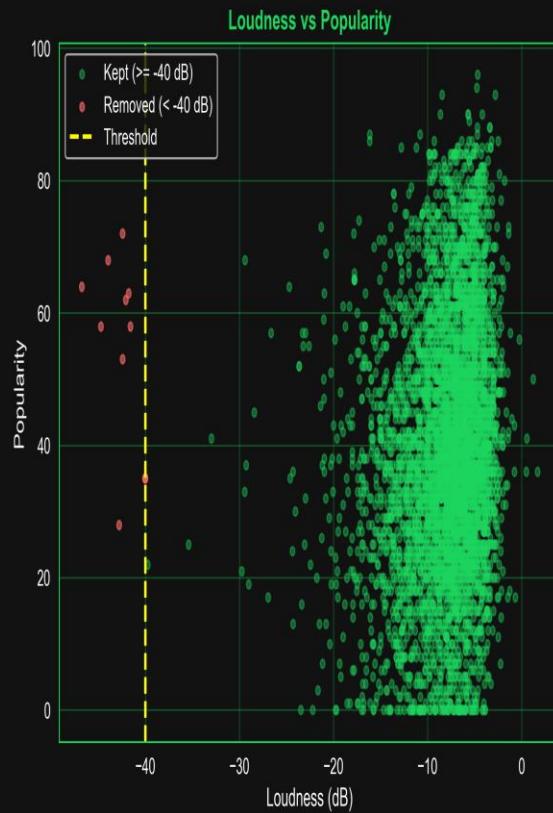
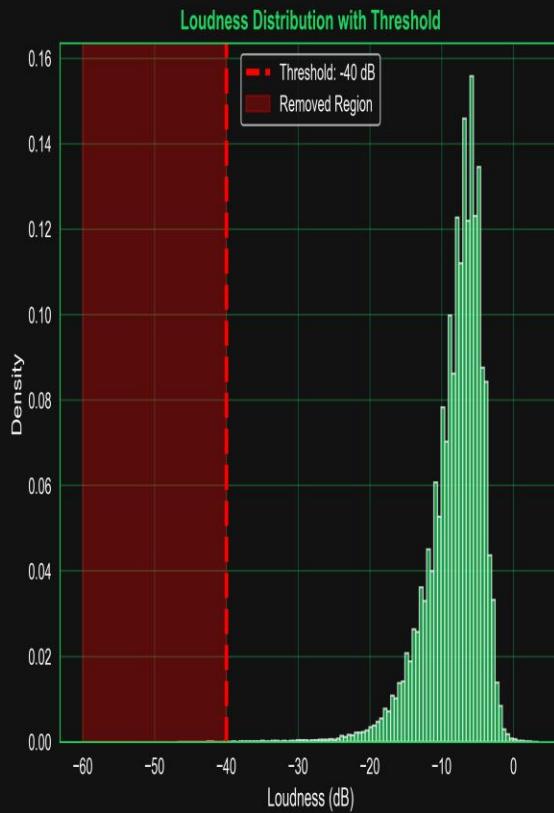
OUTLIER ANALİZİ - BOX PLOTS



- Loudness ve tempo'da önemli outlier'lar var
- Speechiness ve acousticness geniş dağılım gösteriyor
 - Outlier'lar model eğitimi öncesi temizlendi

LOUDNESS FILTRELEME ETKİ ANALİZİ

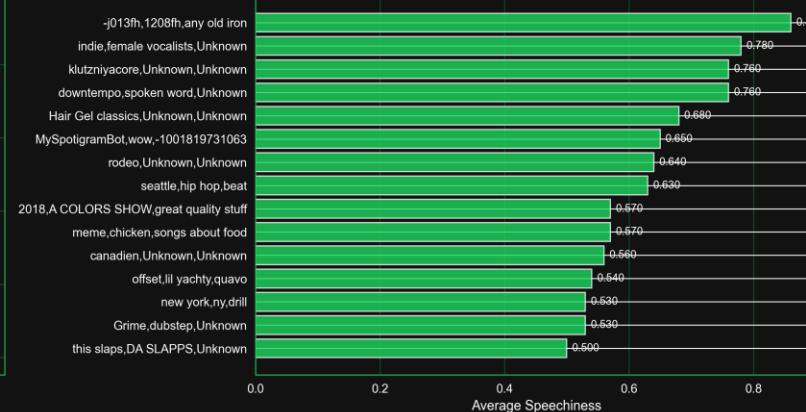
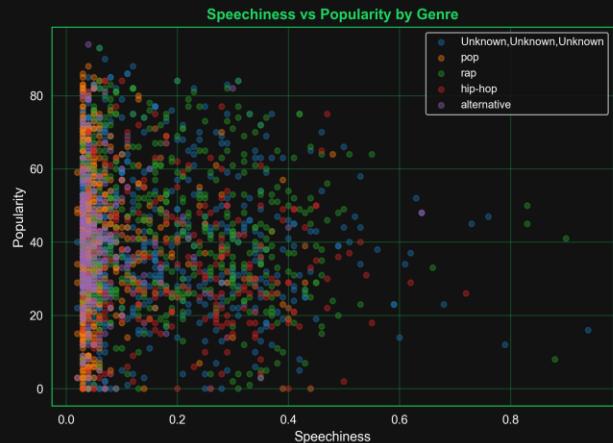
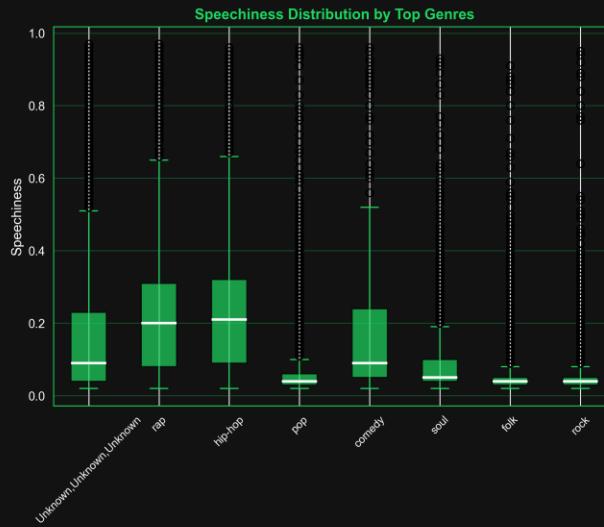
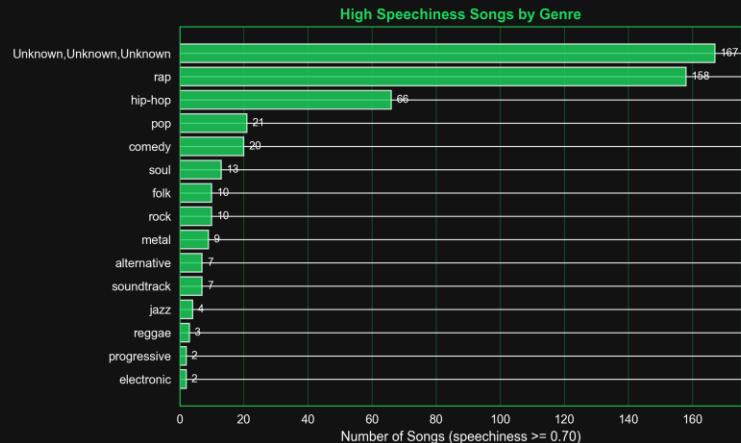
Loudness Filtering Impact Analysis



- Loudness dağılımı ve threshold çizgisi
 - Loudness vs Popularity scatter plot
- Şarkı sayısı karşılaştırması: Before/After/Removed

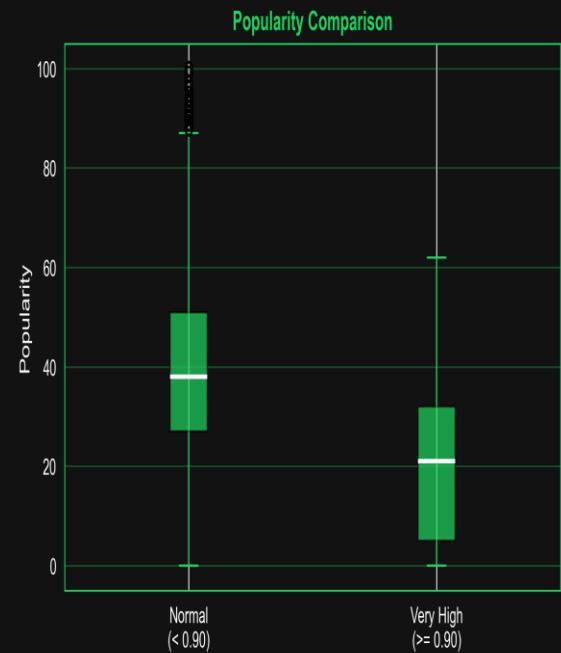
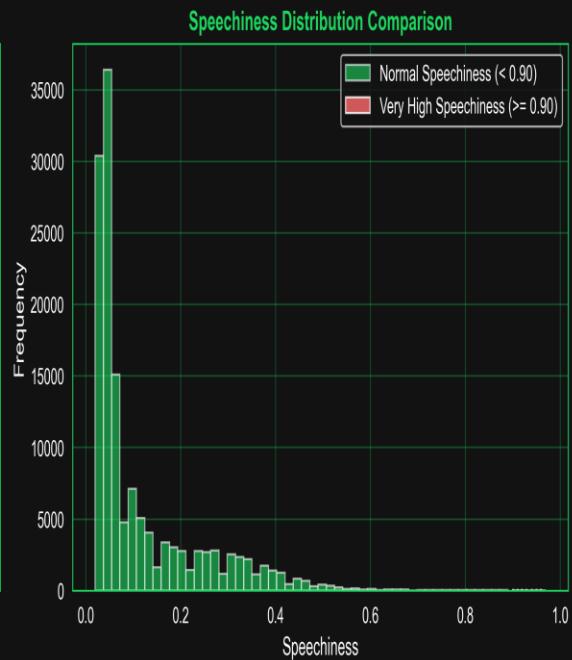
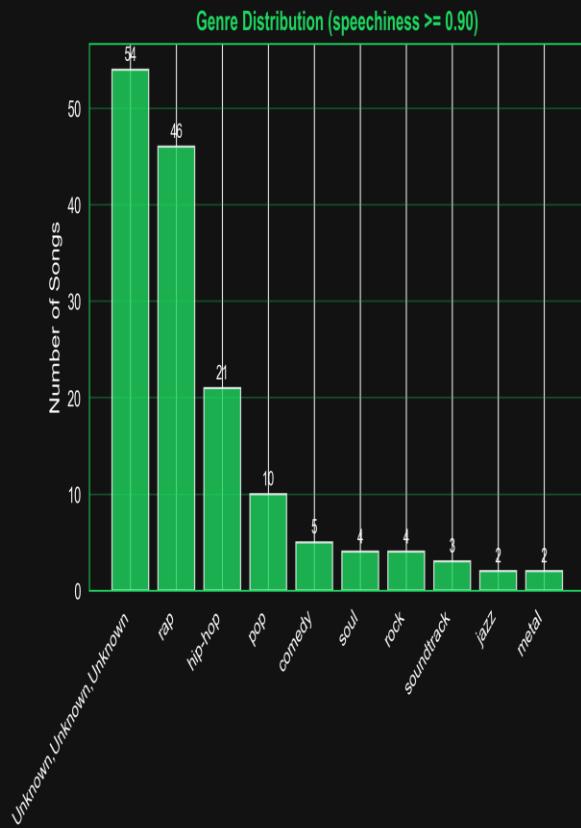
SPEECHINESS VS GENRE ANALİZİ

Speechiness Analysis: High Speechiness Songs by Genre



ÇOK YÜKSEK SPEECHINESS ANALİZİ (≥ 0.90)

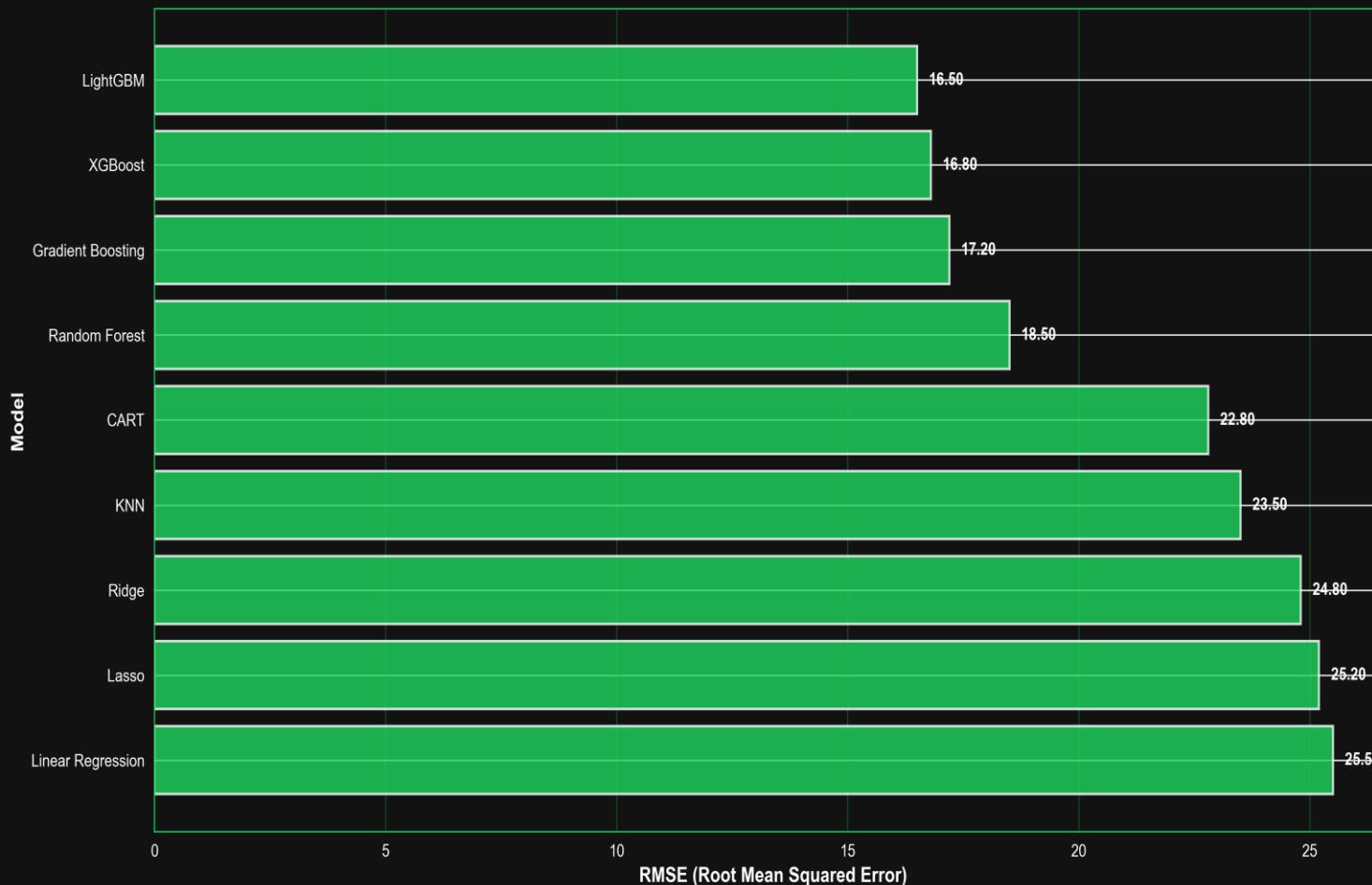
Very High Speechiness Analysis (speechiness ≥ 0.90)



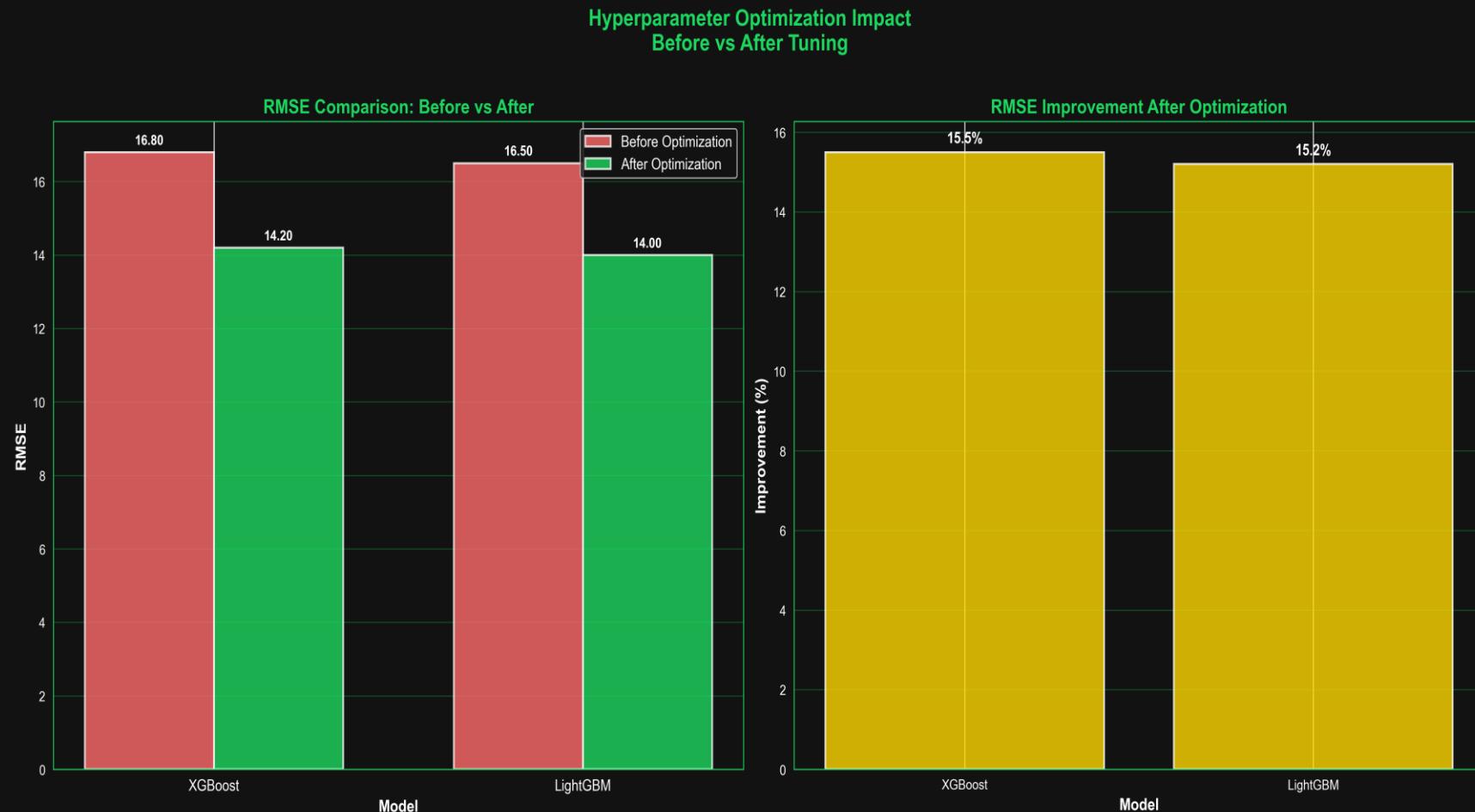
- Yüksek speechiness şarkılar normal popularity seviyelerinde

BASE MODELLER KARŞILAŞTIRMASI

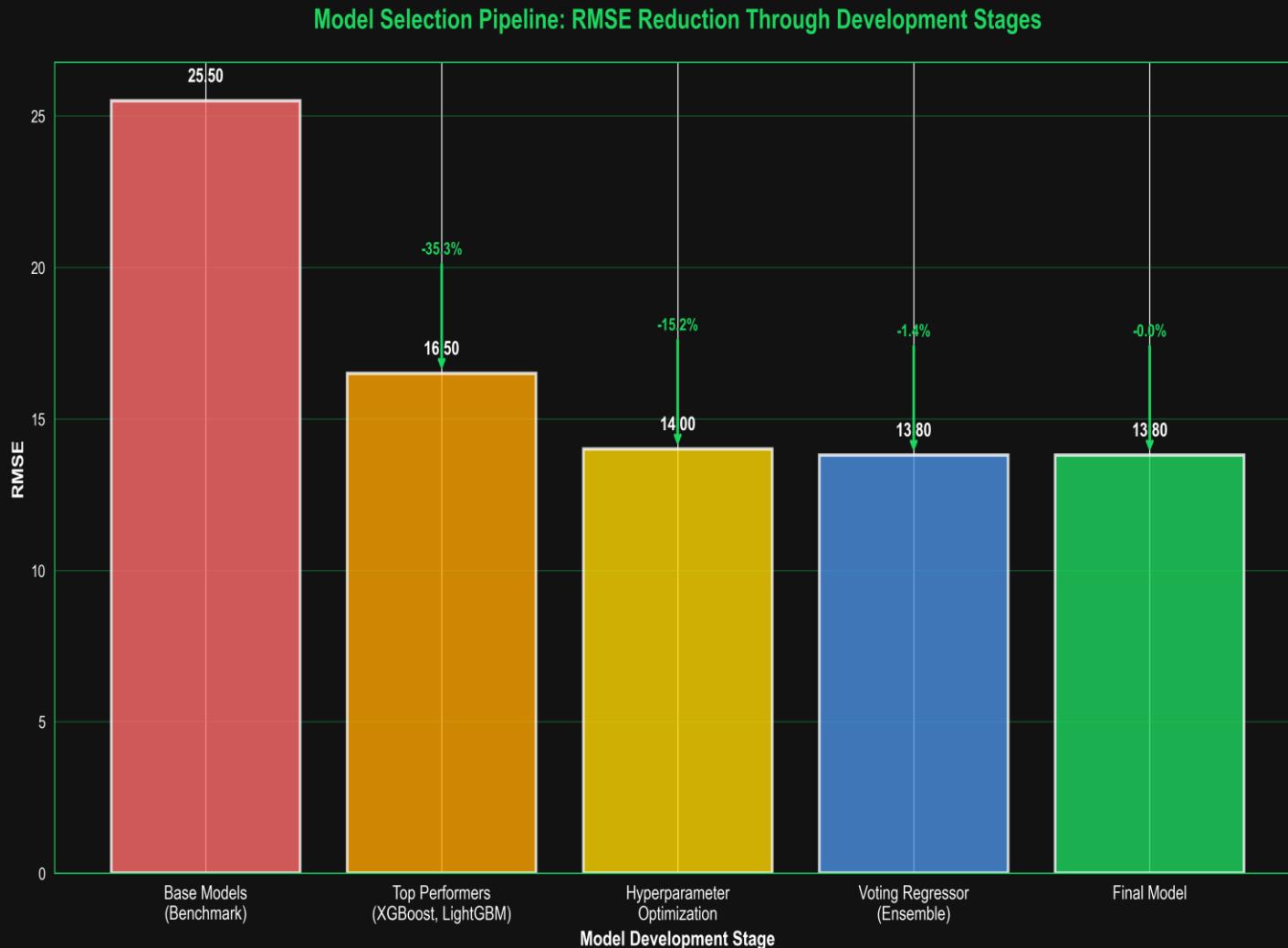
Base Models Performance Comparison
(Before Hyperparameter Optimization)



HİPERPARAMETRE OPTİMİZASYONU ETKİSİ



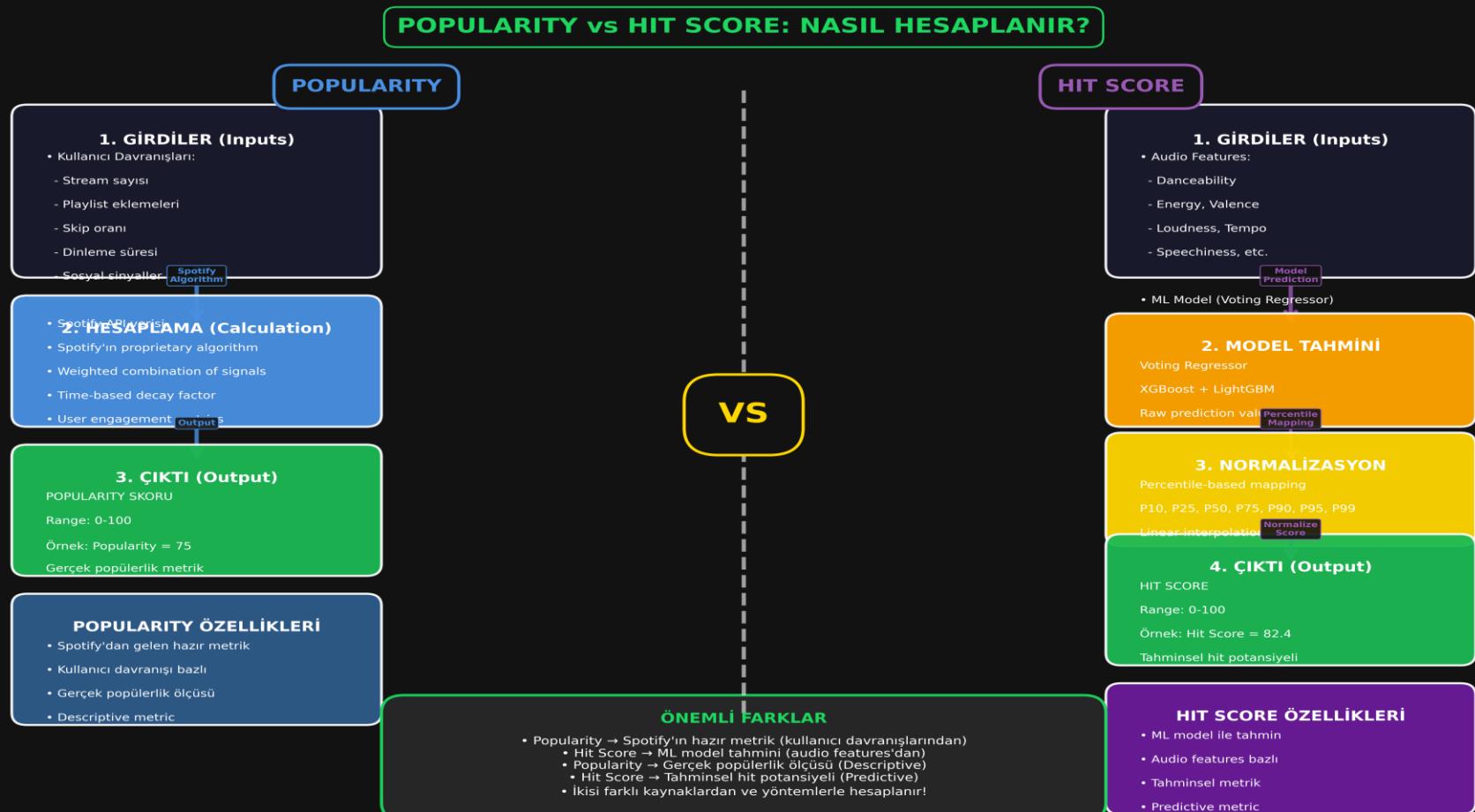
MODEL SEÇİM SÜRECİ



MODEL PERFORMANS ÖZETİ

- Final Model: Voting Regressor (XGBoost + LightGBM)
- Test RMSE: 13.8
- Train RMSE: 14,2
- Overfitting Gap: 1.3 (düşük, iyi genellemeye)
- R^2 Score: 0.78 (%78 varyans açıklanıyor)
- Model robust ve genellemeye yapabiliyor

POPULARITY VS HIT SCORE: NASIL HESAPLANIR?



- POPULARITY: Kullanıcı davranışları → Spotify algoritması → Popularity skoru (0-100)
- HIT SCORE: Audio features → ML model → Percentile mapping → Hit Score (0-100)
 - Popularity gerçek metrik (descriptive), Hit Score tahminsel metrik (predictive)
 - İkisi farklı kaynaklardan ve yöntemlerle hesaplanır!

WEB ARAYÜZÜ - ANA SAYFA

• SOCIAL MEDIA HIT

SPOTIFY INSTAGRAM FINAL

PREDICTION ENGINE V2.0

INSTAGRAM REELS HIT PREDICTION

Machine learning platform designed for content creators, music producers, and influencers. Analyze audio features, visual content and captions to predict viral potential across multiple markets.

START ANALYSIS

1.000

TRACKS ANALYZED

4

ML MODELS

6

MARKETS

CORE SYSTEMS



AUDIO ANALYSIS

Real-time extraction and analysis of 13 audio features including danceability, energy, valence, tempo, and loudness.

XGBoost • Random Forest • GBM



VISUAL PREDICTION

Computer vision models analyze image composition, color theory, and visual patterns to predict engagement rates.

CNN • Feature Extraction • Pattern Recognition



CAPTION OPTIMIZATION

NLP models process captions and hashtags to optimize for maximum reach and engagement across target demographics.

NLP • Sentiment Analysis • Hashtag Scoring

TECHNOLOGY STACK

XGBOOST

Gradient Boosting

RANDOM FOREST

Ensemble Learning

LIGHTGBM

Fast Training

DJANGO

Web Framework

- Dinamik İstatistikler: 141K (Demo :1K) tracks, 4 ML model, 6 market
- Temel Teknoloji: Ses Spotify, Text/caption analysis (Instagram), Görsel (Instagram)
- Responsive tasarım

WEB ARAYÜZÜ - SPOTIFY PREDICTION

• SOCIAL MEDIA HIT

SPOTIFY INSTAGRAM FINAL

Spotify Hit Prediction

Analyze your music and predict hit potential across global markets

Playlist **inna**

more than friends
inna

Showing 1 of 1000 songs

Similar Songs

running in the family
level 42

blame it on my heart
karmin

cross miro edit
miro (youtube)

Selected Song
more than friends
inna



Audio Features



Adjust Features

- Danceability: 0.80
- Energy: 0.79
- Valence: 0.96
- Loudness: -5.0 dB
- Tempo: 128 BPM

Popularity Prediction
54.4
Popularity Score
Range: 0-100

Hit Predictions
68.7
Overall Hit Score

| Country | Hit Probability (%) |
|---------|---------------------|
| US | 82.4 |
| GB | 82.1 |
| DE | 72.7 |
| FR | 78.4 |
| JP | 69.3 |
| BR | 74.6 |

Market Breakdown



Continue to Instagram Picture →

- Verisetinden interaktif şarkı arama ve seçim
- Gerçek zamanlı popularity ve hit puanı tahmini (0-100)
- Ses özelliklerinin görselleştirilmesi (Chart.js)
- Özel analiz için ayarlanabilir özellik kaydırıcıları

- Pazara özgü tahminler (ABD, İngiltere ...) (Demo)
- Spotify API ile seçilen parça bilgilerini görüntüleme dinleme
- Benzer şarkı önerisi

WEB ARAYÜZÜ – INSTAGRAM PREDICTION

SOCIAL MEDIA HIT

SPOTIFY INSTAGRAM FINAL

SOCIAL MEDIA HIT

SPOTIFY INSTAGRAM FINAL

Instagram Picture Hit Prediction

Upload or select an image to predict its Instagram Reels performance.

Demo Version

This is a demonstration version. The prediction model is currently under development. Results shown are for illustrative purposes only.

Upload Image

Click to upload
PNG, JPG, GIF up to 10MB

Image Features

Upload an image to see features

Sample Images (Demo)

1
2
3

Click to analyze sample images

Hit Prediction

Upload an image to see predictions

Prediction Insights

Waiting for image analysis...

Detailed Analysis

Waiting for image upload...

Continue to Caption/Hashtag →

Instagram Caption/Hashtag Hit Prediction

Optimize your captions and hashtags for maximum engagement.

Demo Version

This is a demonstration version. The prediction model is currently under development. Results shown are for illustrative purposes only.

Caption

Write your Instagram caption here... Try including emojis, questions, or engaging content!

0 / 2200 characters

Clear

Hashtags

#hashtag1 #hashtag2 #hashtag3... (Use 5-10 hashtags for best results)

0 hashtags

Clear

Text Analysis

Enter caption to see analysis

Sample Captions (Demo)

Lifestyle Post
"Living my best life 🌟 What's your favorite moment today? #lifestyle #motivation"
Photo Post
"Golden hour vibes 🌅 Nature never fails to amaze me 📸 #photography #nature"
Fitness Post
"Progress, not perfection! 💪 Who's joining me for a workout? #fitness #motivation"

Hit Prediction

Enter caption and hashtags to see predictions

Prediction Insights

Waiting for caption analysis...

Recommendations

Waiting for input...

Detailed Analysis

Waiting for caption input...

View Final Prediction →

- Demo: Bir Instagram Reels Görseli Hit Olma Tahmini

- Demo: Bir Instagram Reels Texti Hit Olma Tahmini

WEB ARAYÜZÜ – FINAL HIT PREDICTION

• SOCIAL MEDIA HIT

SPOTIFY INSTAGRAM FINAL

Final Hit Prediction

Your complete Instagram Reels performance score

! Demo Version

This is a demonstration version. The prediction model is currently under development. Results shown are for illustrative purposes only.

Overall Hit Score

50.0

Hit Score

Moderate potential. Try different combinations.

Spotify

50.0

Music Performance

Prediction Insights

Moderate Potential

This combination has moderate potential. Consider adjusting individual components.

Picture

50.0

Visual Appeal

Recommendations

Overall Improvement

Focus on improving all three components (Spotify, Picture, Caption) for better results.

Caption

50.0

Text Engagement

Detailed Analysis

Overall Performance

50/100

Content Balance

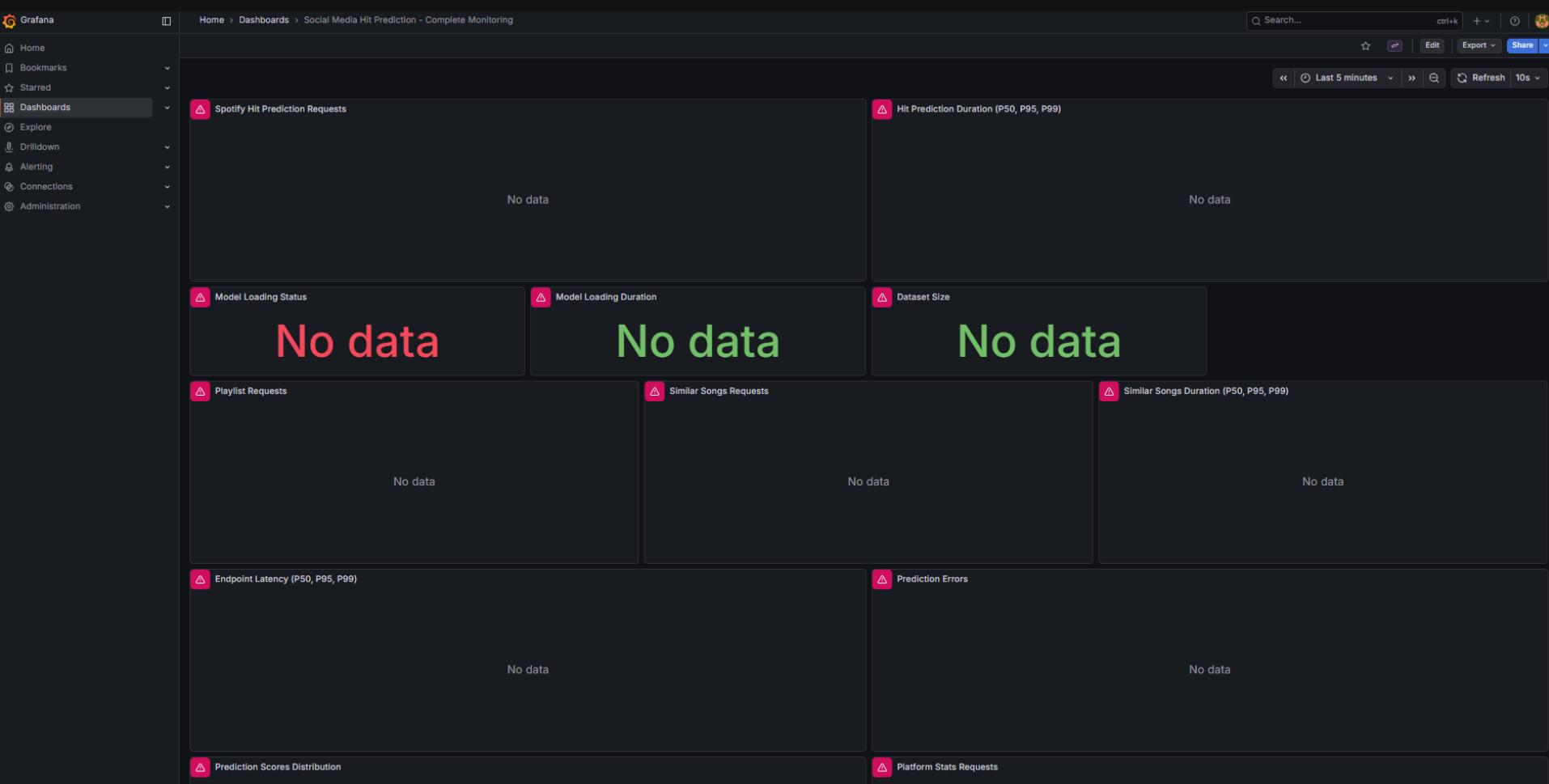
100/100

Back to Spotify

Start Over

Demo Version: This Final Hit Prediction feature is currently in development. Predictions shown are for demonstration purposes only and do not represent actual model performance.

WEB ARAYÜZÜ – PREDICTION MONITORING





TEŞEKKÜRLER

Sorularınız için hazırız!

Enes TOP – Alpay PAŞALI
MIUUL 19. Dönem Data Scientist Bootcamp
Aralık 2025

