FALL 2023 - DEC. 19

Music Listening Behavior: Forecasting Track Skipping

Harnessing LSTM for Enhanced User Experience

Ghina Al Shdaifat, Ansh Jhaveri, and Kendall Brogle

Link to GitHub Repo

EXECUTIVE SUMMARY

1 PROBLEM STATEMENT

The need for **improved Music Recommendation Systems (MRS)** that can predict and adapt to user skip behavior

2 GOAL

Predict the likelihood of track skipping based on the **duration** for which a track has been played (skip_1, skip_2, skip_3, not_skipped)

3 TECHNICAL CHALLENGES

Balancing model **accuracy** with computational efficiency

Mitigating overfitting risks

Decipher & **understand** the meanings and implications of **each variable** within our dataset

4 SOLUTION APPROACH

DEEP LEARNING

LSTM MODELS

HYPERPARAMETER TUNING

SHAP

MACHINE LEARNING

DECISION TREE

GRADIENT BOOSTING

RANDOM FOREST

5 VALUE & BENEFIT

Enhanced predictive accuracy that can drive user engagement and refine MRS personalization.

Motivation / Related Works

1. Gap in Existing Research

- Existing studies in music streaming, like Spotify's sequential skip prediction challenge, primarily focus on machine learning
- Noticeable gap in the application of deep learning techniques, particularly LSTM networks

2. Limitations of Previous Studies

- Spotify’s study, despite its large dataset, faced challenges in computational efficiency and overfitting
- The MusicRecNt study's deep learning for genre classification was limited by its dataset's scale and diversity

3. Our Drive

- We aim to explore the uncharted territory of using LSTM networks for a deeper understanding of user skip behavior in music streaming
- This approach seeks to uncover complex temporal patterns overlooked by traditional models

4. Anticipated Impact

- By leveraging LSTMs, our project aspires to refine music recommendation systems, achieving a more personalized user experience by aligning with individual listening habits and preferences

METHOD & APPROACH

ITERATIVE CYCLE

1 DATA PREPARATION

One hot encoding

2 MODEL SELECTION

1 Recurrent Neural Networks (RNNs) with LSTM

Simple LSTMs for skip_1, skip_2, skip_3, not_skipped

2 Hyperparameter Optimization

- Keras Tuner's Hyperband
- Adjusting parameters
- # of LSTM units
- Dropout rates
- Batch Normalization

3 Traditional Machine Learning Models

- Random Forest
- Gradient Boosting
- Logistic Regression

3 FEATURE ENGINEERING AND SELECTION

- SHAP (SHapley Additive exPlanations)
- Feature Importance

4 MODEL TRAINING & VALIDATION

- 70-30 split of the dataset
- Early stopping
- Model checkpointing

5 PERFORMANCE EVALUATION

- Precision
- Accuracy
- Recall
- F1-Score
- AUC-ROC scores & curves

Implementation / Experimentation

DATA PROCESSING

- Merged and preprocessed dataset involving user **interaction logs and track features**
- **One-hot encoding** for categorical variables; **StandardScaler** applied for numerical data **normalization**

LSTM MODEL SPECIFICS

- **Batch Size:** 64
- **Initial Learning Rate:** 0.001
- **Data Split:** 70% Training, 30% Evaluation
- **Architecture:** Configured with specific layers for **skip_1** and **skip_3**

HYPERPARAMETER TUNING VIA KERAS TUNER

- **Algorithm**: Hyperband, adaptive resource allocation, early stopping, batch normalization
- **Tuned Parameters:** LSTM Units (**32 to 256**), Dropout Rates (**0.0 to 0.5**), Optimizers (**'adam', 'sgd'**)

BASELINE ML MODELS CONFIGURATION

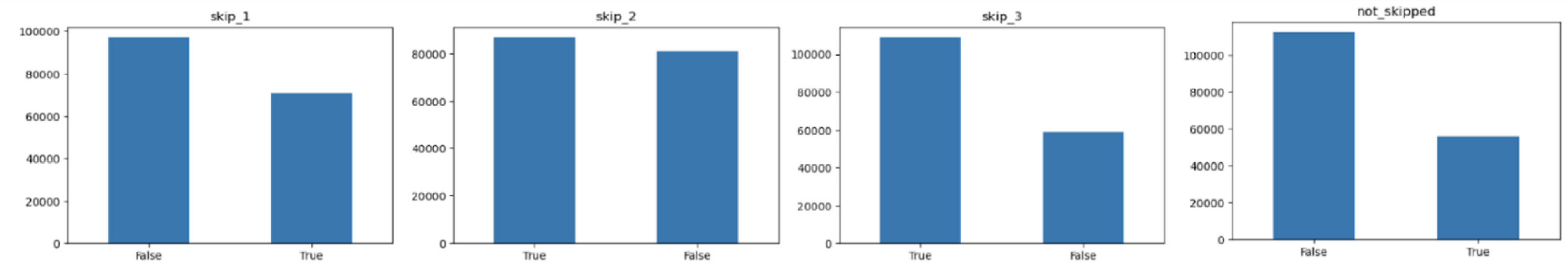
- **Random Forest:** 100 Trees, Depth: 5
- **Gradient Boosting:** 100 Stages, Learning Rate: 0.1, Max Depth: 3
- **Logistic Regression:** Max Iterations → 1,000

EVALUATION METRICS AND VALIDATION TECHNIQUES

- Metrics: Accuracy, Precision, Recall, F1 Score, AUC-ROC
- Early Stopping: Patience set at 10 epochs; ModelCheckpoint based on validation loss

Results and Experimental Evaluation

of times users skipped



Performance Metrics on LSTM Models

Targets	Train Accuracy	Test Accuracy	Val Accuracy	Test Loss
skip_1	84.91% – 88.44%	88.35%	~88.69%	28.62%
skip_2	84.07% – 87.54%	87.61%	~87.42%	26.94%
skip_3	95.75% – 98.15%	98.18%	~98.32%	7.50%
not_skipped	95.89% – 98.82%	98.96%	~98.77%	4.24%

- High predictive performance
- Strong correlation between user engagement and track completion

Could also indicate a sign of overfitting?

Performance metrics of the Random Forest

Targets	Accuracy	Precision (F/T)	Recall (F/T)	F1-Score (F/T)
skip_1	87.80%	0.91/0.84	0.88/0.88	0.89/0.86
skip_2	87.33%	0.90/0.85	0.83/0.91	0.86/0.88
skip_3	98.10%	0.99/0.98	0.96/0.99	0.97/0.99

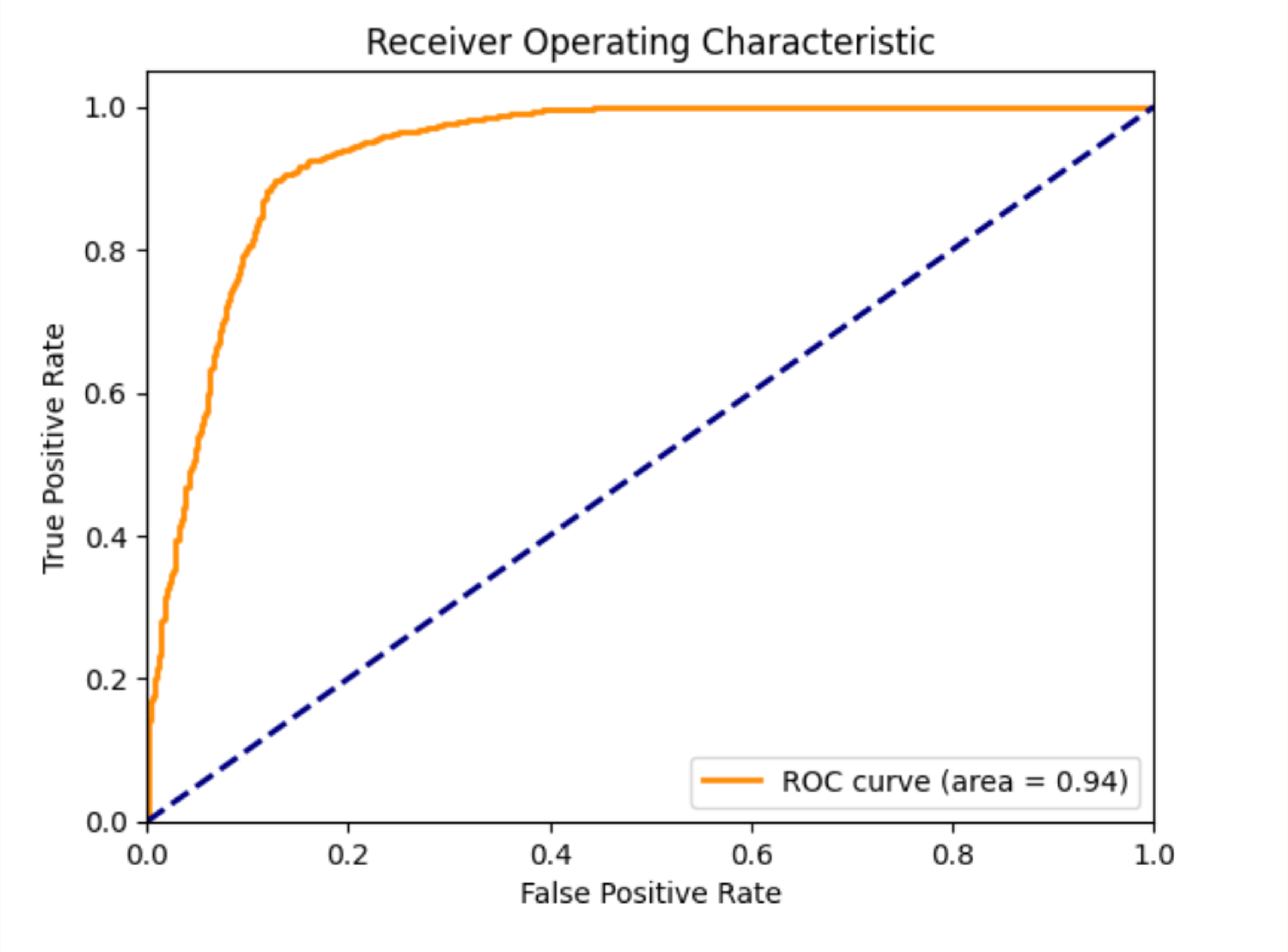
Targets	AUC-ROC Score	Accuracy
skip_1	93.88%	87.80%
skip_2	94.46%	87.33%
skip_3	98.88%	98.10%

Performance metrics of the Gradient Boosting

Targets	Accuracy	Precision (F/T)	Recall (F/T)	F1-Score (F/T)
skip_1	88.12%	0.92/0.84	0.87/0.89	0.89/0.86
skip_2	87.53%	0.90/0.86	0.84/0.91	0.87/0.88
skip_3	98.21%	0.99/0.98	0.95/1.00	0.97/0.99

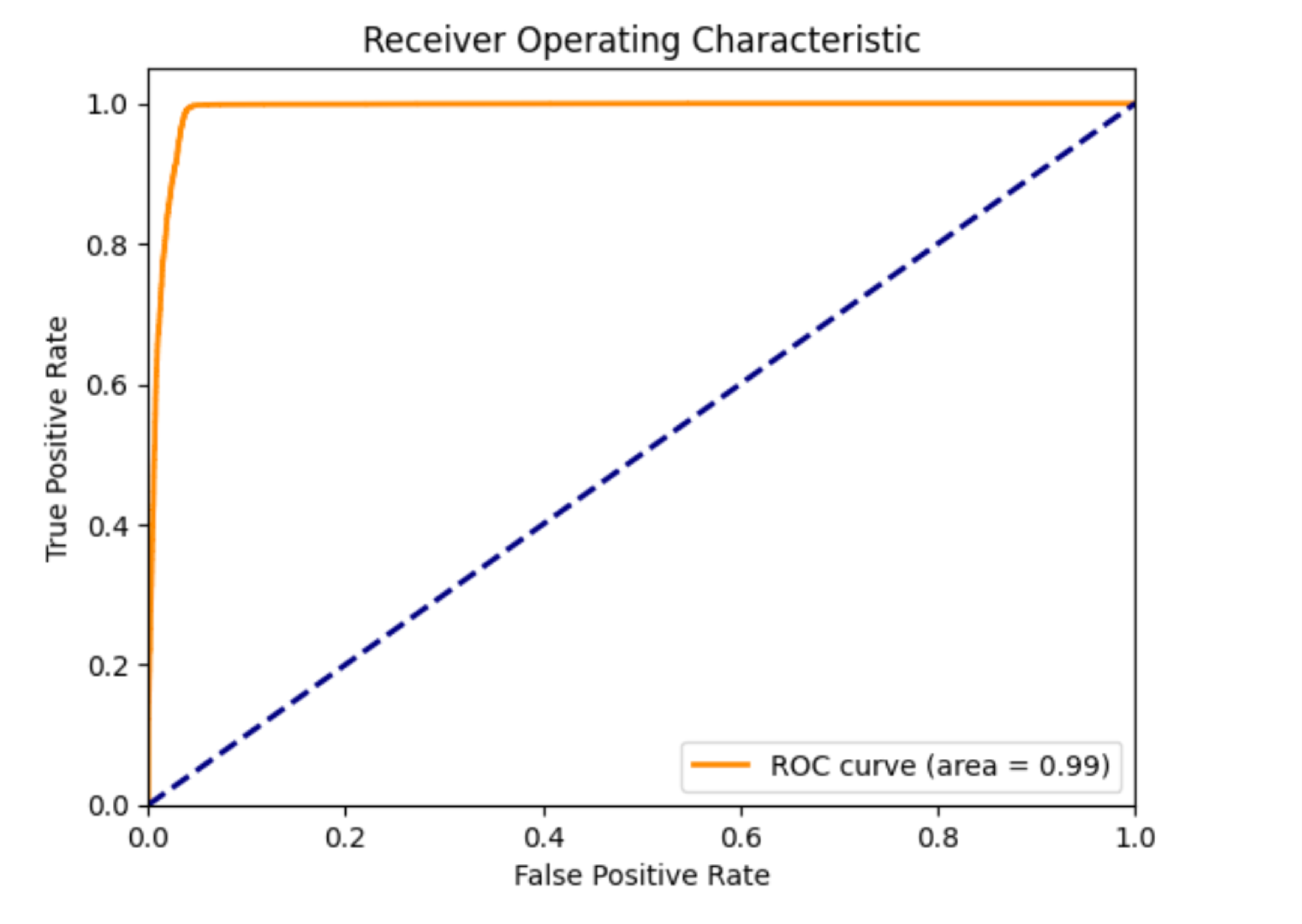
COMPLEX LSTM MODELS

Skip_1



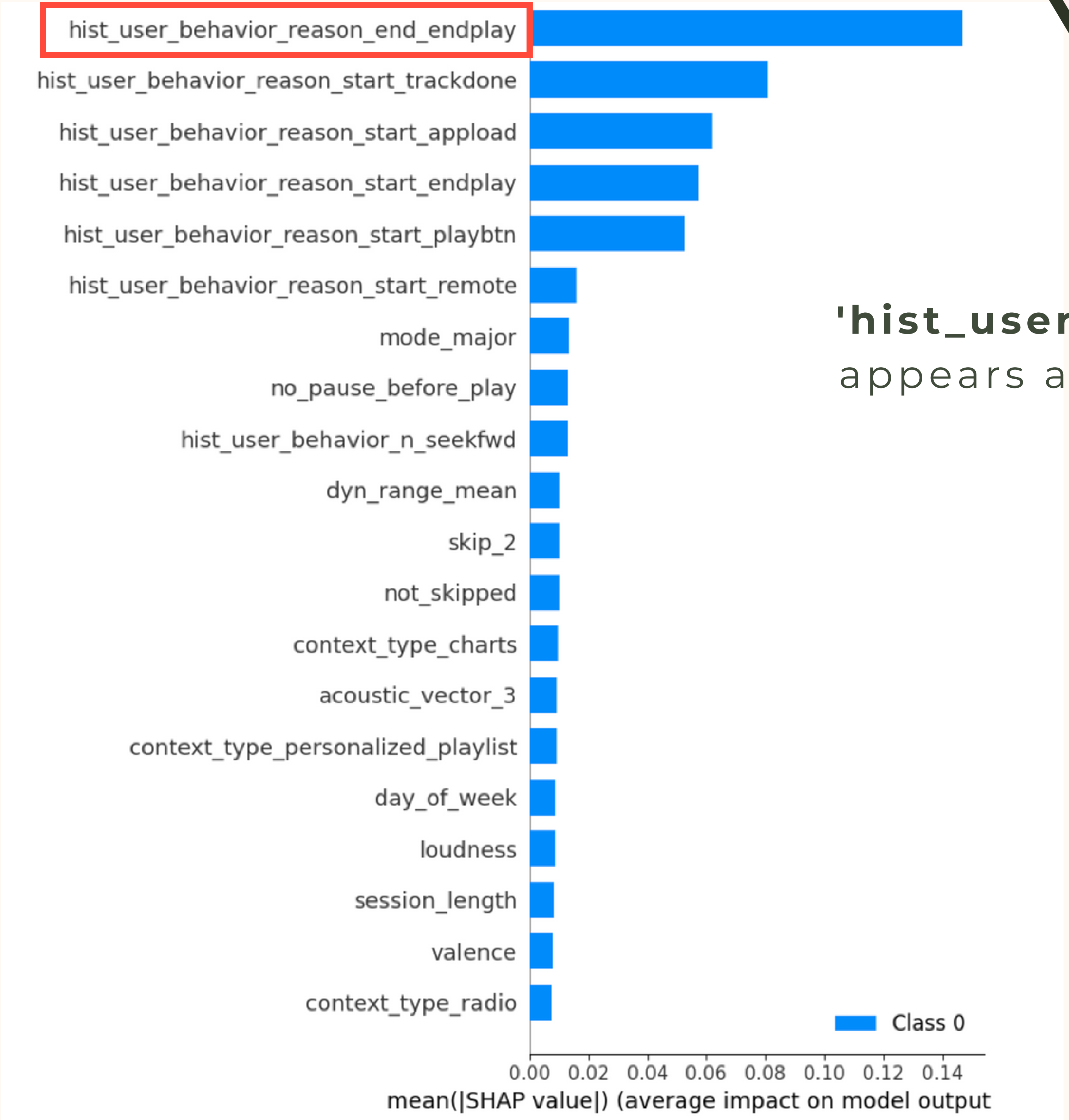
Validation Accuracy: 88.73%
Test Accuracy: 88.20%
Test Loss: 31.55%
AUC/ROC: 0.94

Skip_3

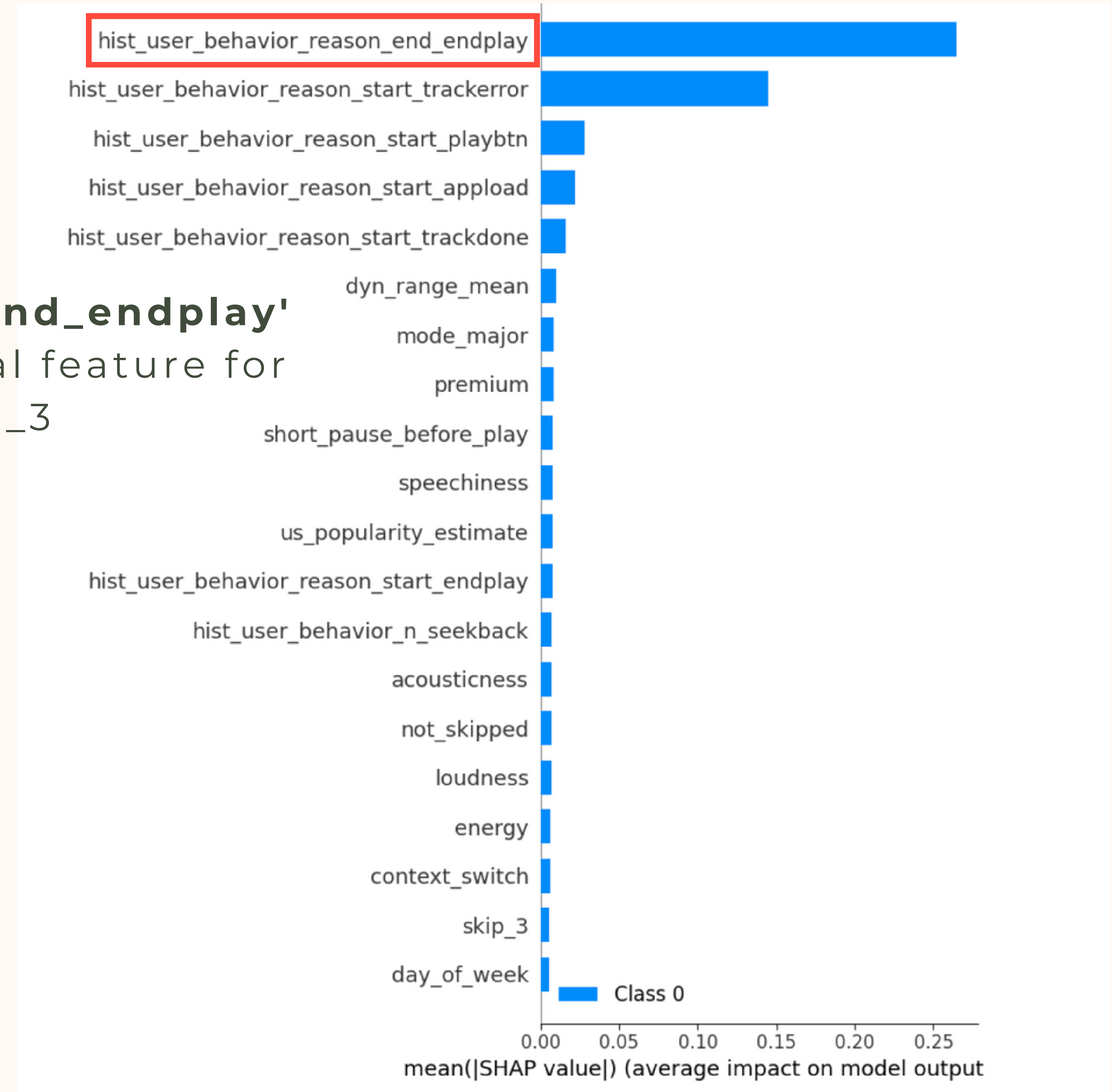


Validation Accuracy: 98.17%
Test Accuracy: 98.17%
Test Loss: 7.46%
AUC/ROC: 0.99

Skip_1



Skip_3



'hist_user_behavior_reason_end_endplay' appears as the most influential feature for both skip_1 and skip_3

Feature	Random Forest	Gradient Boosting	Logistic Regression
End Trackdone	0.186/0.286/0.474	0.595/0.828/0.975	-2.09/-2.08/-2.58
Start Trackdone	0.115/0.084/0.071	0.214/0.086/NA	-0.86/-0.56/NA
End Fwdbtn	0.106/0.114/0.186	0.074/0.113/0.013	1.16/1.12/1.45

Conclusion

Project Recap

Developed LSTM models to forecast user track skipping behavior in a music streaming context

Key Insights

- **skip_3 behavior predicted with high accuracy (AUC/ROC of 0.99)**, with **hist_user_behavior_reason_end_endplay as a key feature**
- Marginal accuracy difference between complex and simple LSTM models, suggesting **similar predictive capabilities**

Complexity VS Efficiency

- Despite the introduction of complex LSTM models with Hyperparameter Tuning, **simple LSTM models yielded comparable accuracy**
- Given the **trade-off** between performance and computational efficiency, simpler LSTMs are recommended for operational use

Directions for Future Research

- Employ larger datasets to validate findings and expand model applicability
- Explore other deep learning techniques and their impact on prediction accuracy and interpretability
- Examine external factors influencing user engagement beyond the provided features

Ghina Al Shdaifat
Kendall Brogle
Ansh Jhaveri

Thank You!

Any questions?