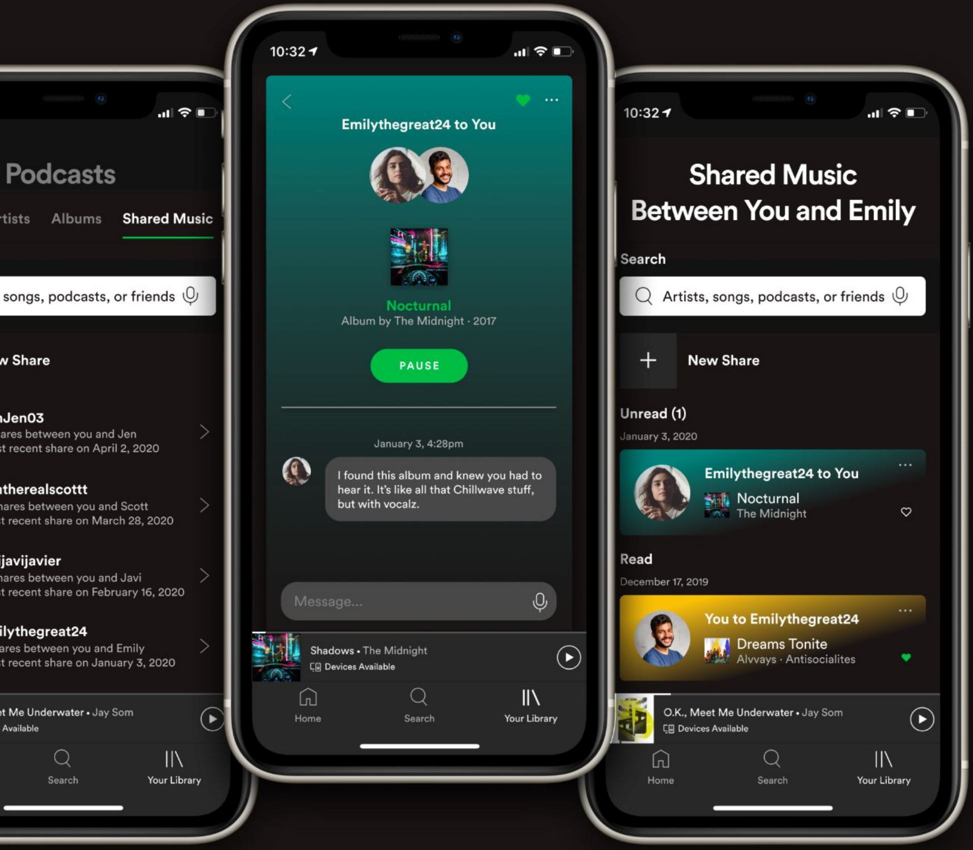


Spotify Skip Prediction

Team 9: Varshitha, Winnie, Claudia

Data Science
CS-GY 6053

Background & Data

With over 430 million active listeners, how can Spotify personalize recommendations that caters to users' varying music tastes?

To help recommend the right songs and keep users streaming for longer hours on the app, we built a user behavior prediction model using methods like **Decision Tree, SVM, and RNN** to predict whether a song will be **skipped**.

User Interaction

Over 160k listening sessions with associated user interactions on the Spotify service.

skip_1,2,3,non-skip	binary	0,1
Hist_user_behavior_reason_start	categorical	n/a
Hist_user_behavior_reason_end	categorical	n/a

Audio Features

Acoustic features and metadata for 50k different tracks played during listening sessions.

danceability	Continuous ordinal	{0,1}
mechanism	Continuous ordinal	{0,1}
acousticness	Continuous ordinal	{0,1}

Prediction Models

- **Data preprocessing:** Integrate session with corresponding tracks; Target variable created using skip_1,2,3,non-skip; one-hot encoding
- **Feature selection methods:** Pairwise correlation; Decision Tree feature importance
- **Hyperparameter tuning:** K-fold cross validation

Logistic Regression

Regularization hyperparameter (C=0.1)
Evaluation metric: accuracy_score, ROC-AUC

Support Vector Machine

Regularization hyperparameters (C= 0.01), kernel functions
Evaluation metric: accuracy_score, ROC-AUC

Neural Network

Simple NN with 4 Dense Layers with 20,12,8,1 units.
Evaluation metric: accuracy_score

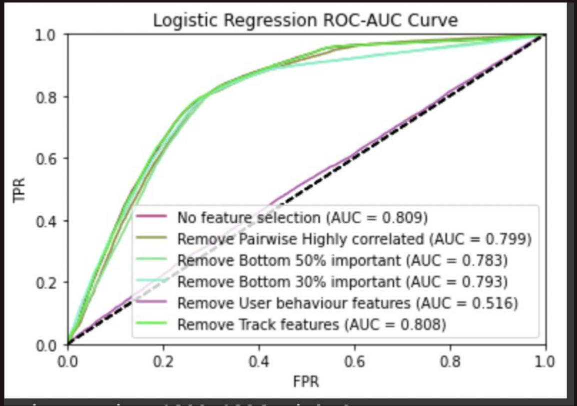
Recurrent Neural Network

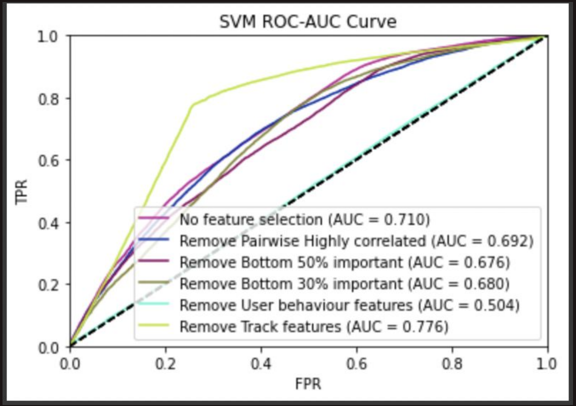
Reconstruction of data - Input: Entire session (20 tracks), **Output:** Skipability (20)
Simple RNN with 20 units + same dense layers as Simple NN.

Evaluation

Removed Features	**NN**	**LR**	**SVM**	**RNN**
No Feature Selection	76.042	75.991	64.668	**77.719**
Pairwise Highly Correlated	75.208	75.488	64.864	**78.552**
Bottom 50% important	53.517	75.762	62.115	**75.920**
Bottom 30% important	76.012	75.899	63.870	**78.219**
User Behavior features	53.317	53.469	50.515	**62.610**
Track features	**76.459**	**76.033**	**75.759**	**79.790**

- No Feature Selection doesn't include removal of unnecessary features - ID, Date, History End Variables.
- User Behavior is more important than track features.





Spotify skip prediction
Team 9

Assumptions, Limitations, Next Steps

Assumptions & Limitations

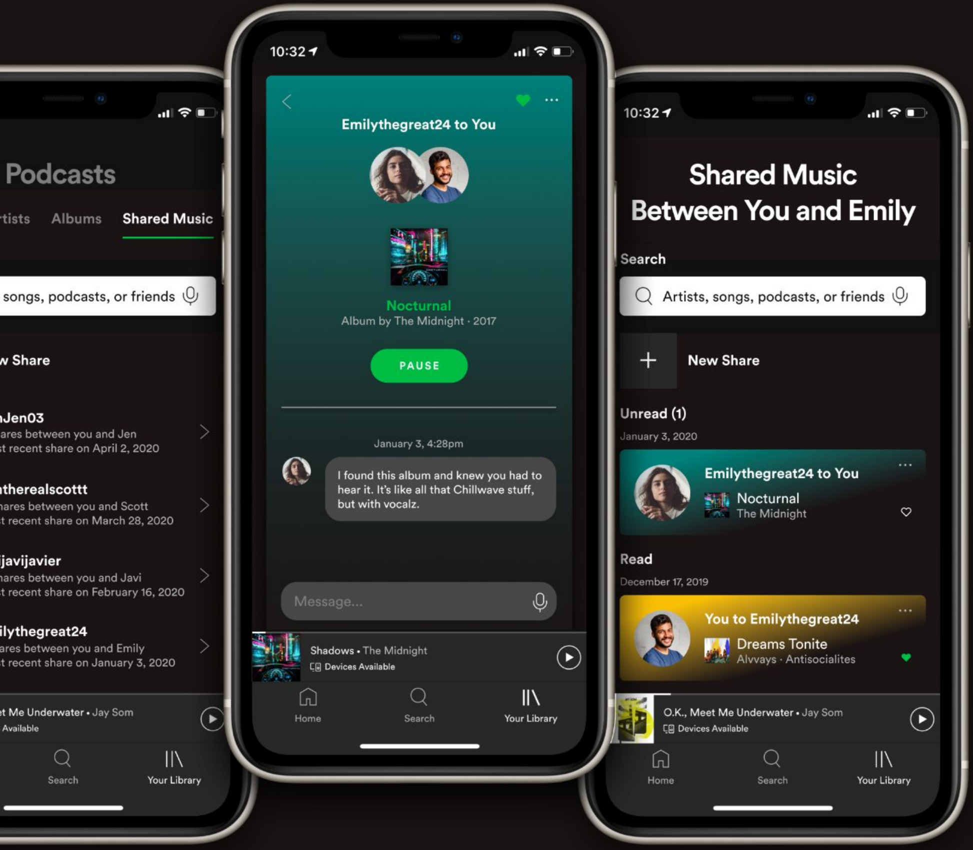
- Integrating original features into one custom variable **will not lead to significant information loss.**
- **Premium users** will nonchalantly skip more than **non-premium users**
- **Confounding variables**: genre of the track, and if the session is of a user's playlist, an album, or Spotify's generated discover playlist

Conclusion

- Skipability prediction accuracy of **~80% with RNN**
- **Track features did not play as important of a role in user skipability;** this suggests further research on immediate user interaction

Next Steps

- **Clustering** to create pseudo genres
- Investigate the **possible correlation** between similar tracks in one session



Thank you! Questions?

Team 9: Varshitha, Winnie, Claudia

Data Science
CS-GY 6053