Machine Learning Classification Project

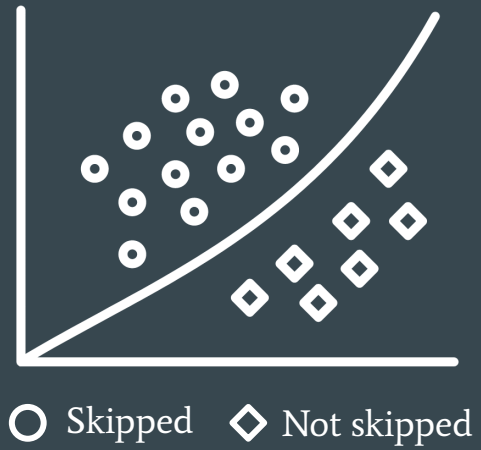
July 12, 2019

Chelsea Power

Project objective:
Determine if a song will be skipped or fully played in order to guide the business on how to negotiate upcoming licensing renewal fees.

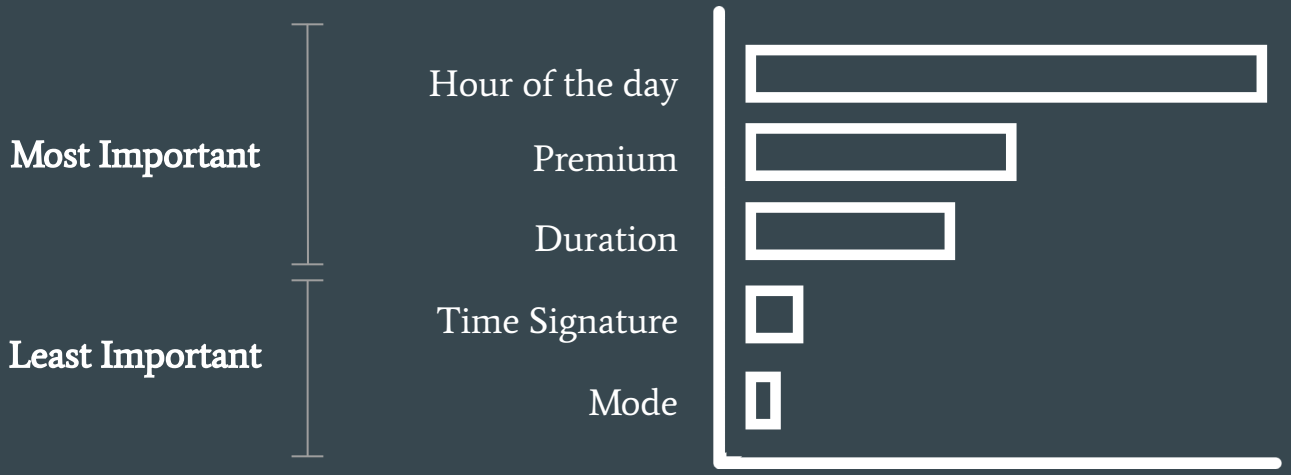
Overview

Based on Spotify 2018 streaming session data across 167,880 tracks, this machine learning classification model determines if a track will be skipped or fully played (not skipped).



Song Attributes

The streaming session data started with 50 attributes (features) and was narrowed down to 33. Of the 33 attributes, the chart below shows the three most important and two least important attributes.



Conclusion

With this model, you can now determine with 91% accuracy whether a track will be skipped or fully played that will help you to determine if the licensing model per track should be adjusted.

91% Accuracy

Next Steps

Since first Random Forest classifier model produced **91% accuracy,** it would be interesting to know if the model can be further improved by running **Principal Component Analysis (PCA)**.

Another consideration would be to re-run this project using a different **predictor variable** to see if the accuracy increases further.

Thank You