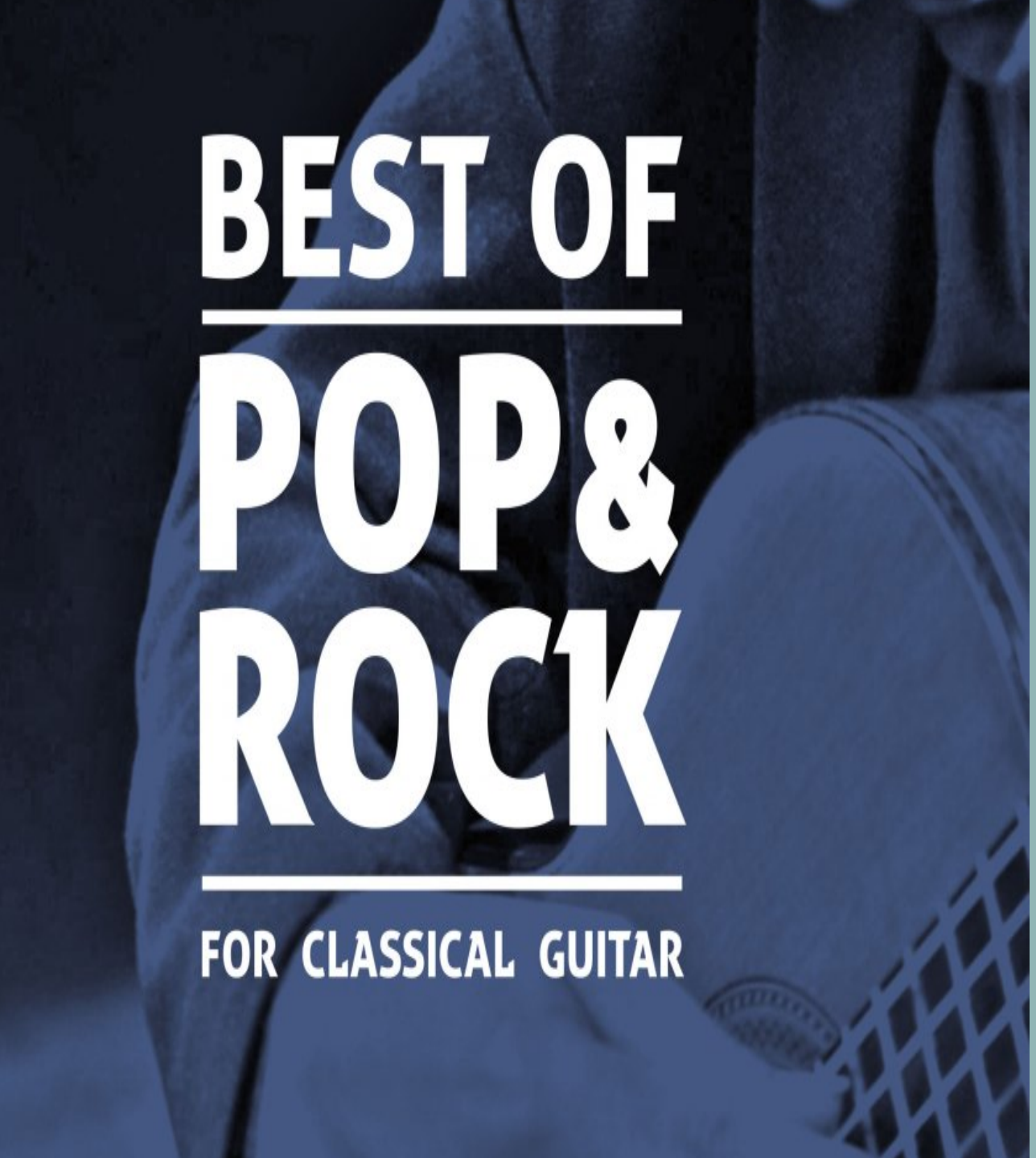
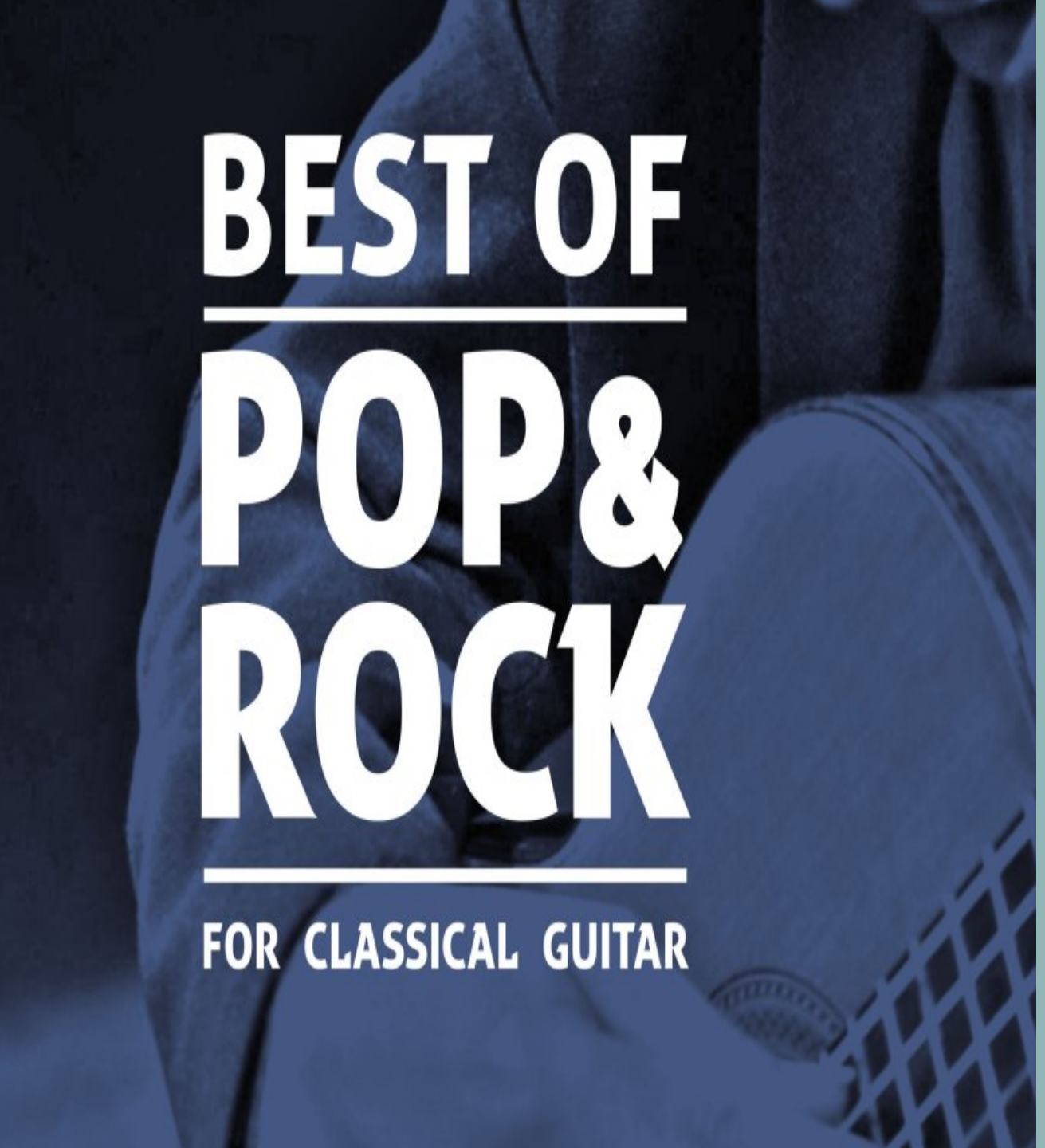


What is “Old Town Road” ?

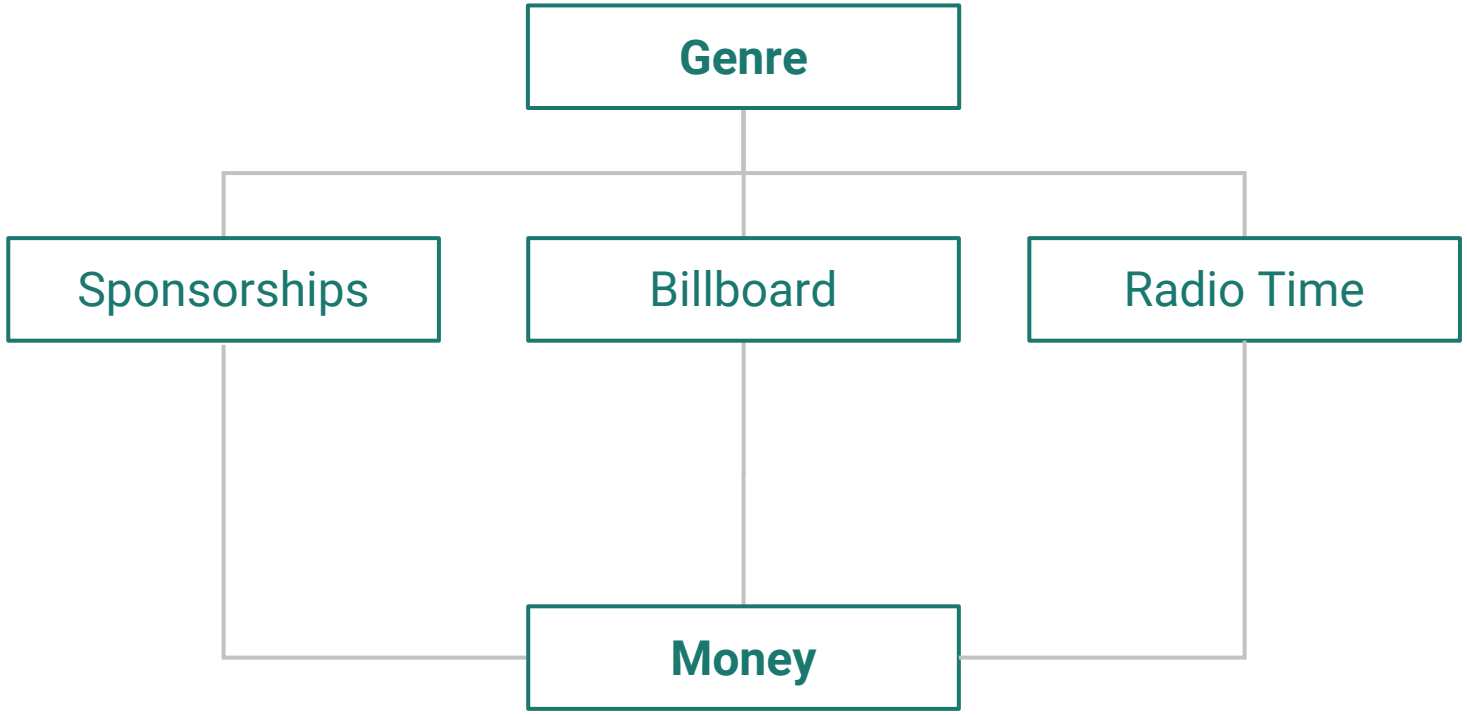
Arnav Deshwal | Pedro Ivo Rivas | Michael Sparkman
He Wang | Rui Ying



Content

1. Main Problem
2. Data Transformation
3. Descriptive Statistics
4. Model Selection
5. Results
6. Conclusion & Recommendations

Main Problem



Artist's Intent ≠ **Audience's Interpretation**

Data Transformation

135,000 Songs

Dataset was found containing 135,000 songs pulled from Spotify's API

Cut Short Songs

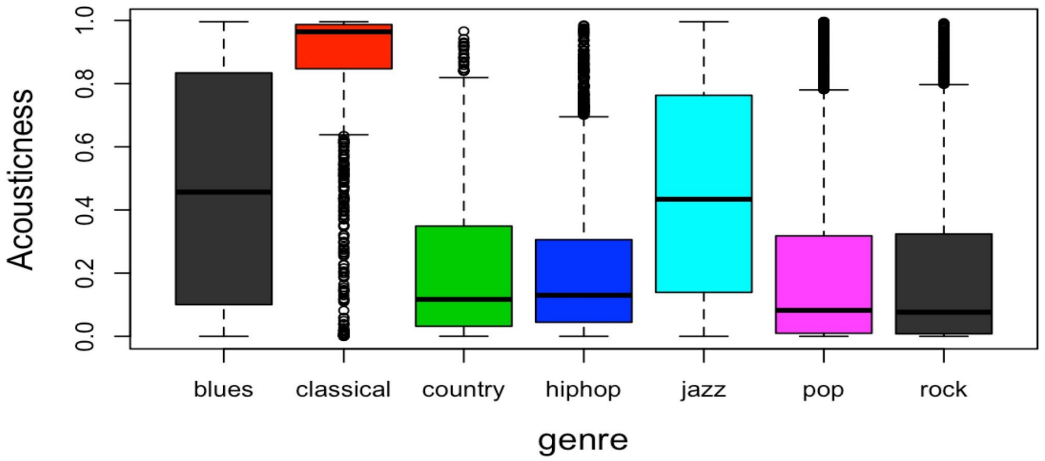
Cut songs less than 2 minutes to eliminate potential "dirty" data(ie songs that lasted one second)

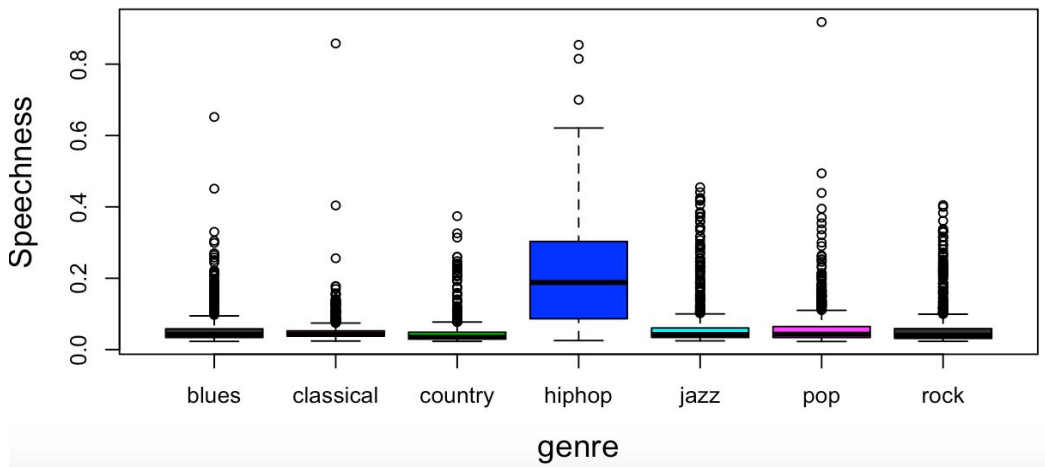
Filtered Genres

Chose 7 main genres and dropped sub-genres that crossed between them (ie **blues rock** or **pop rock**)

Final data set was **~35,000** songs

Descriptive Statistics





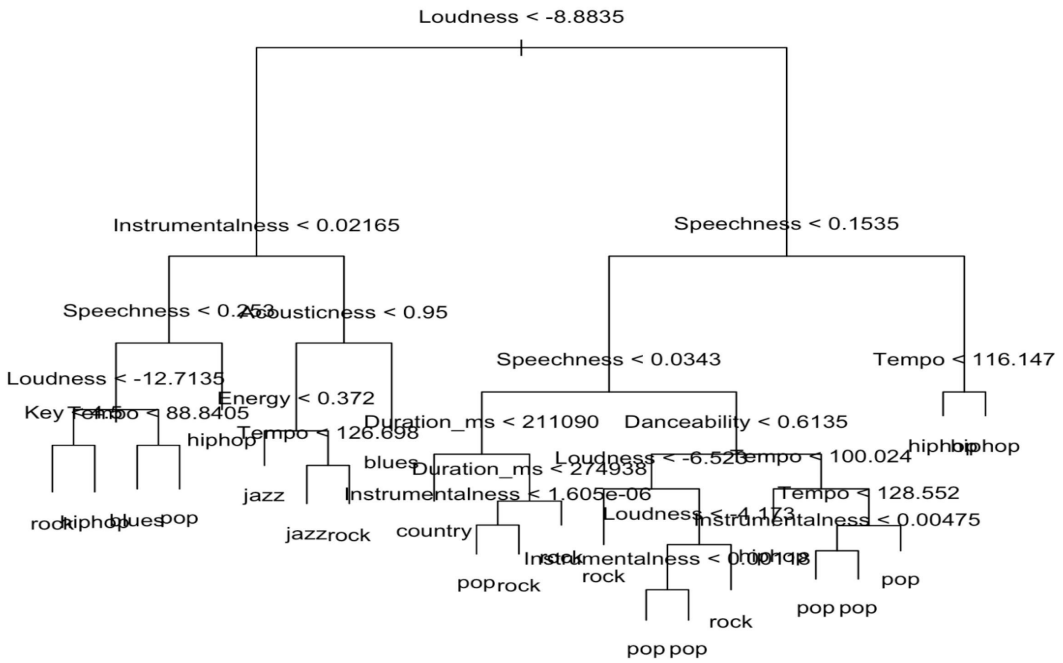
01

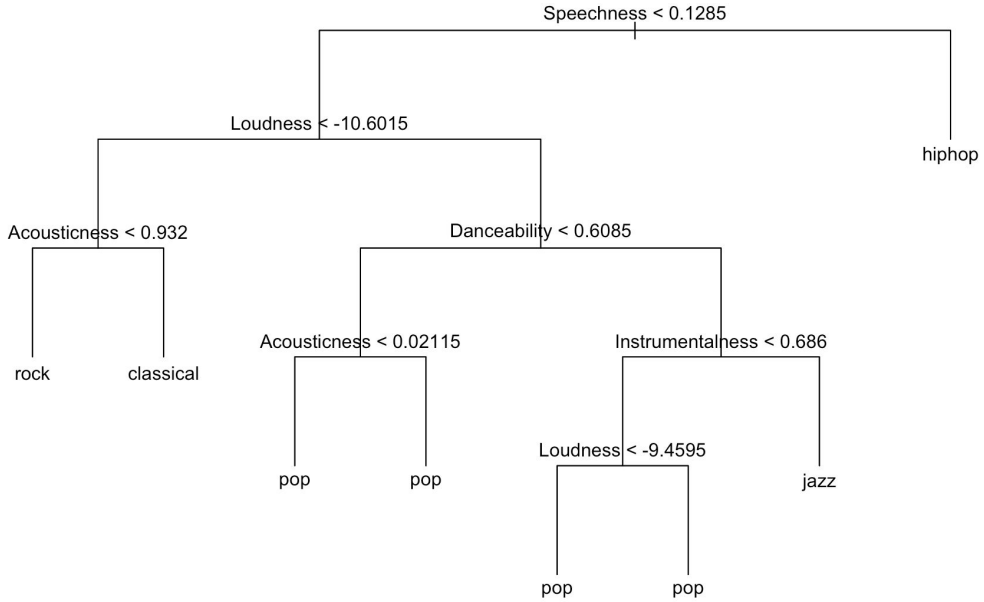
12 independent variables (e.g. loudness, key, energy, valence) ~ dependent variable: genre

02

Pop and rock are similar to each other; Classical, blues and jazz are very different from other categories.

Classification Tree





A Big Tree

Predict using test dataset:
Accuracy_bigtree = **45.57%**

After pruning it

Set the number of tree nodes = 8
Predict using test dataset:
Accuracy_prunetree = **48.06%**

K-Nearest Neighbor

Method 1

Using all the 12 predictors

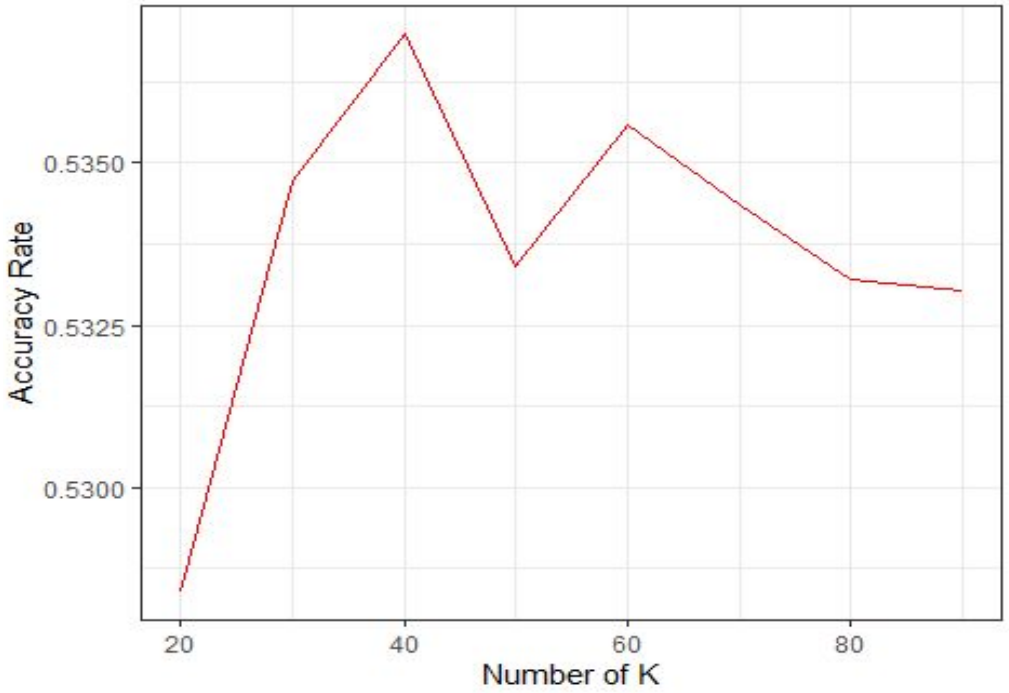
Cross validation:
K = 63

Accuracy:
53.80%

Method 2

Variable Selection: ended up with 9 variables

- Danceability
- Energy
- Loudness
- Speechness
- Acouticness
- Instrumentalness
- Valence
- Liveness
- Duration_ms

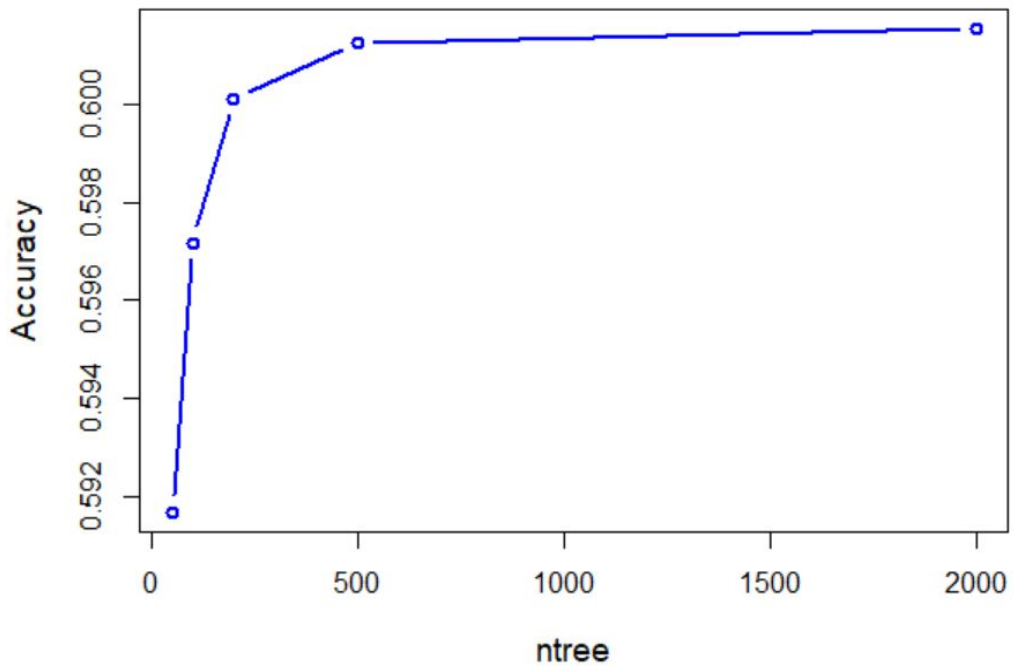


Accuracy:
53.88%

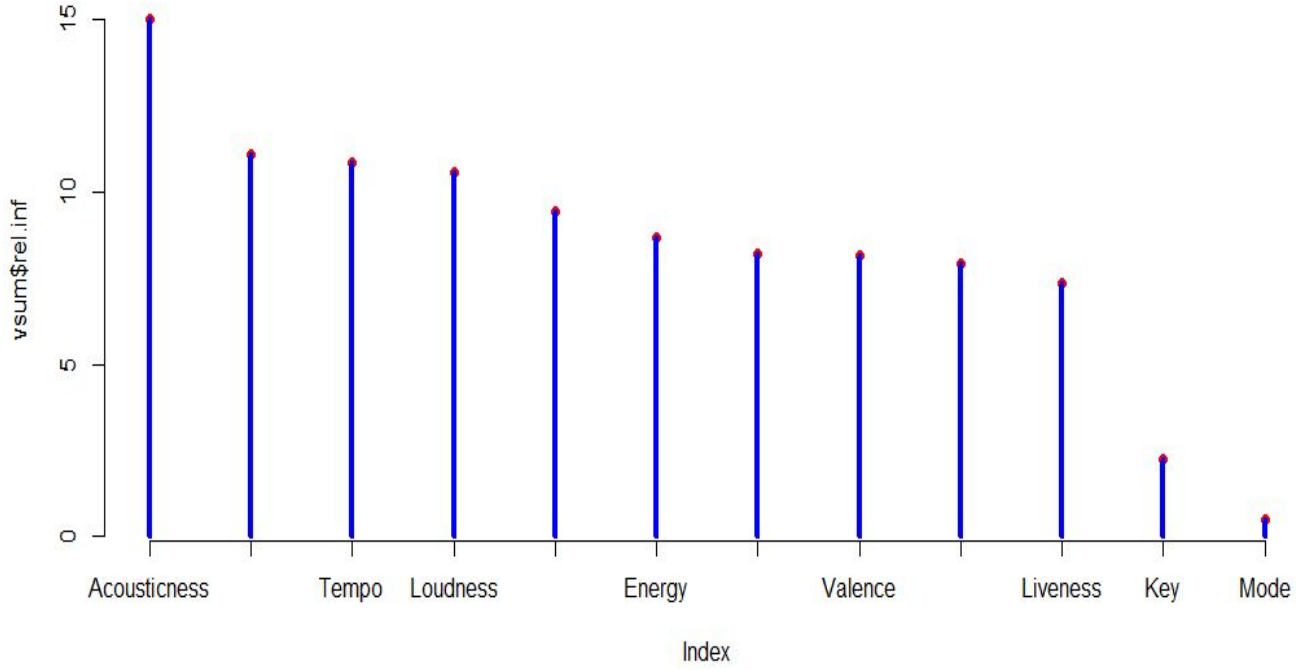
K=43

Bagging and Boosting

Number of Trees vs. Accuracy



Variable Importance



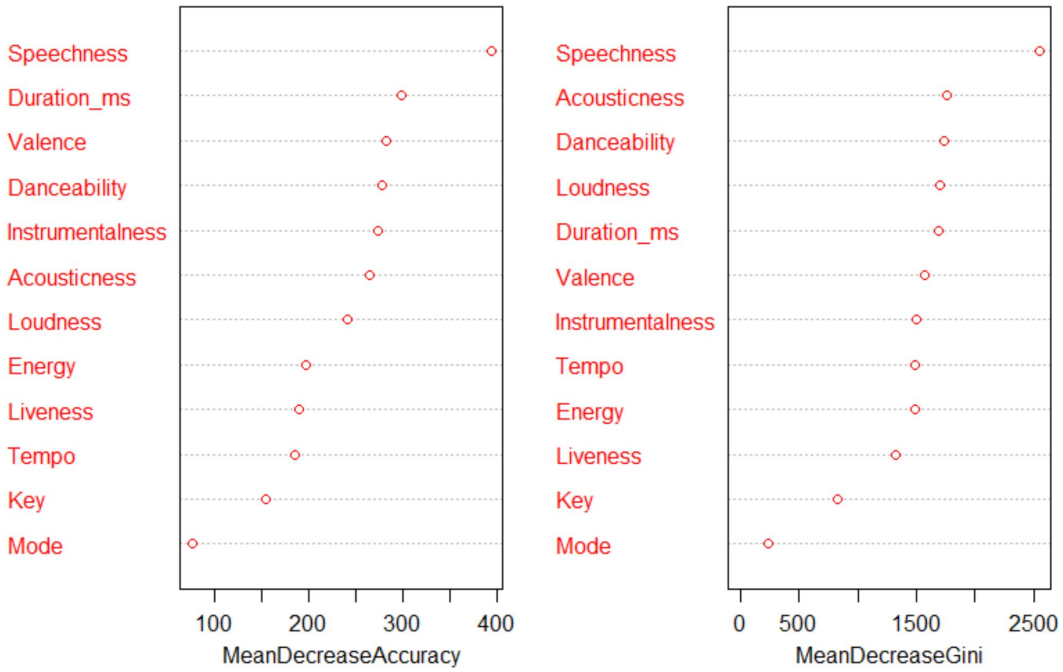
Bagging:
Accuracy: **60.15%** with **2000** trees

Boosting:
Accuracy: **56.52%** with 1000 trees

Acoustiness and Speechness shown to be most important; **Key and Mode** by far least important

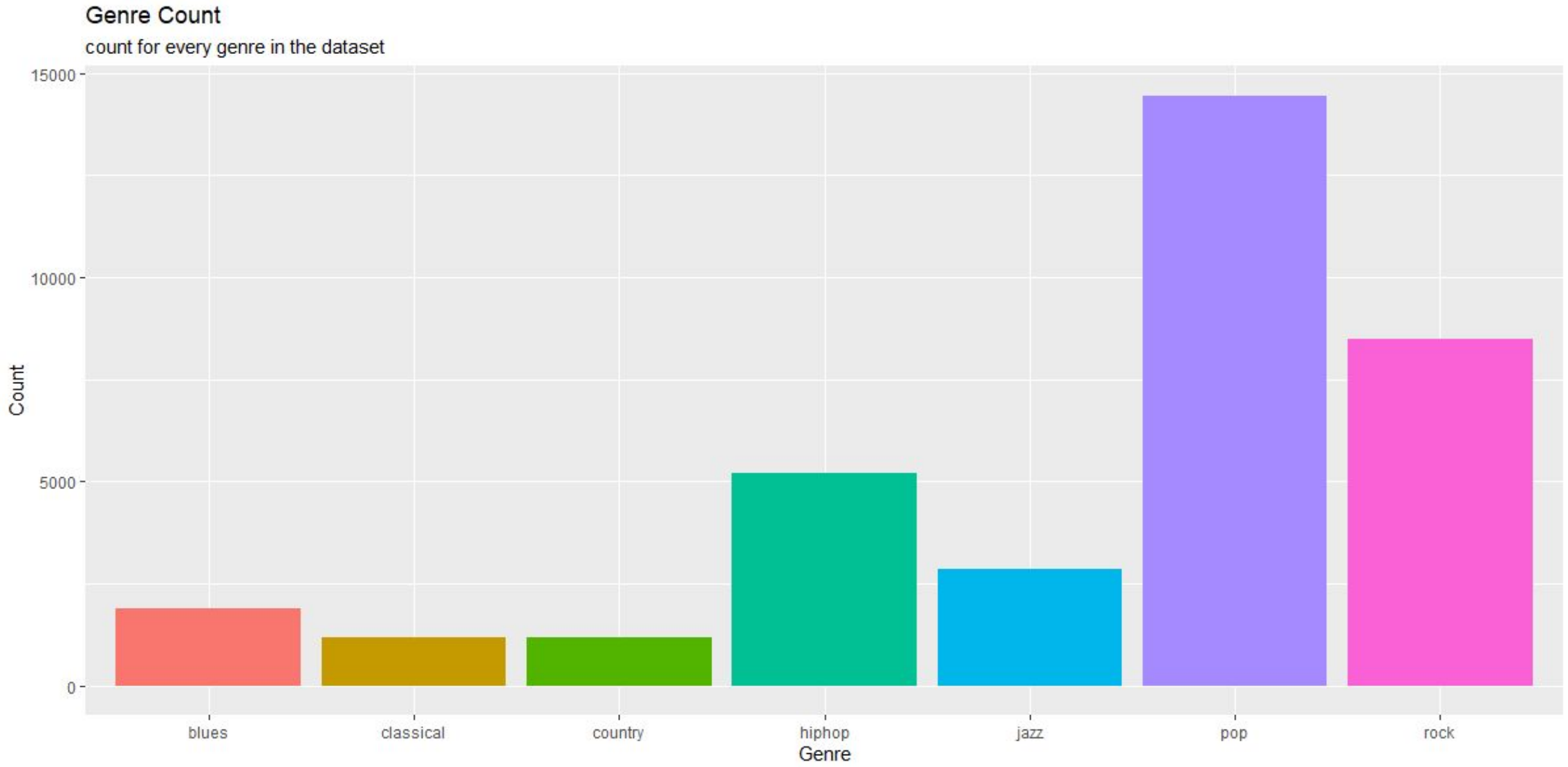
Random Forest

Variable Importance

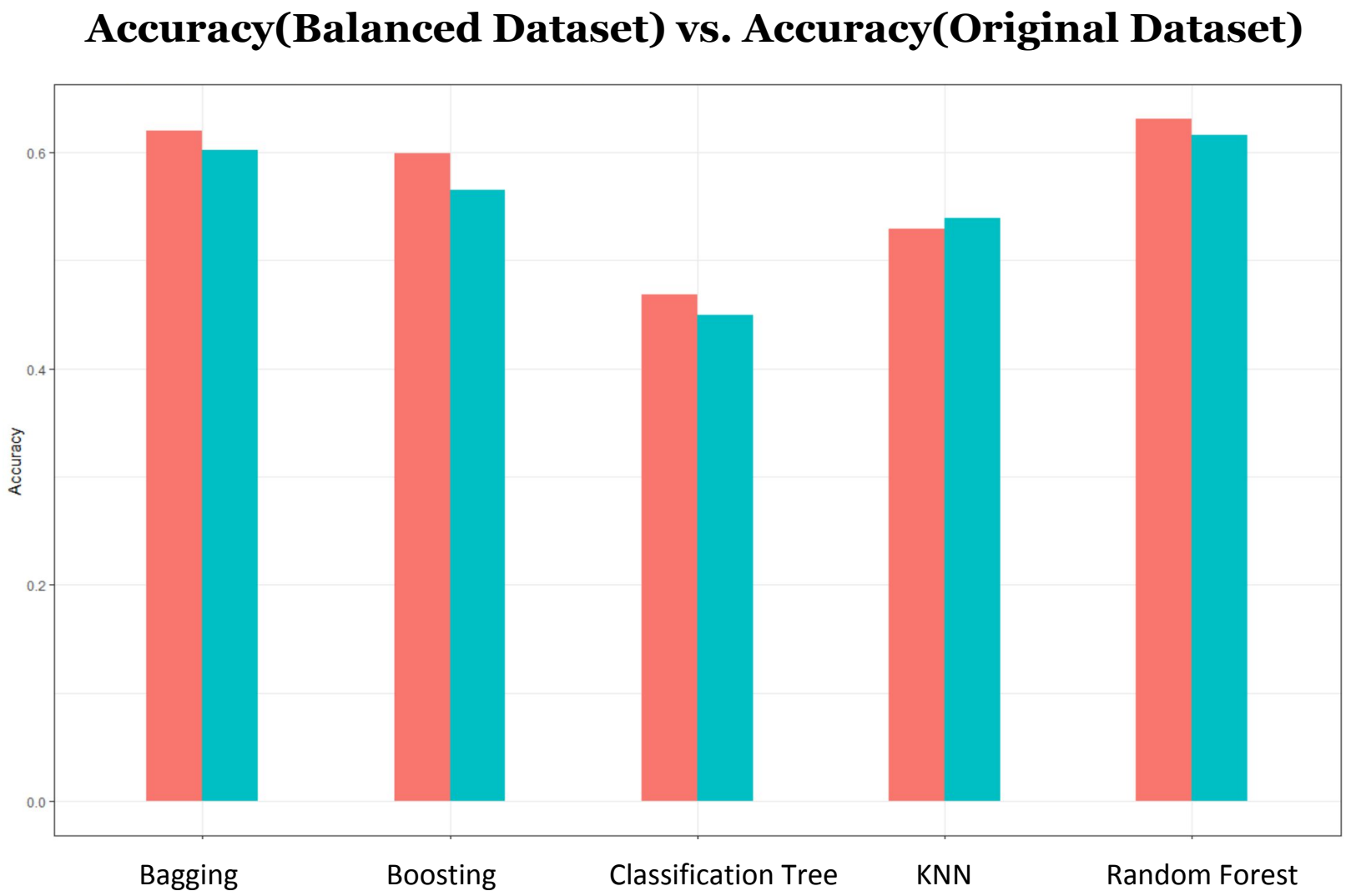


Number of Trees	m(Max Accuracy)	Accuracy
50	6	60.97%
200	6	61.30%
500	3	61.46%
2000	3	61.56%

Balancing Dataset



Results



63.10%

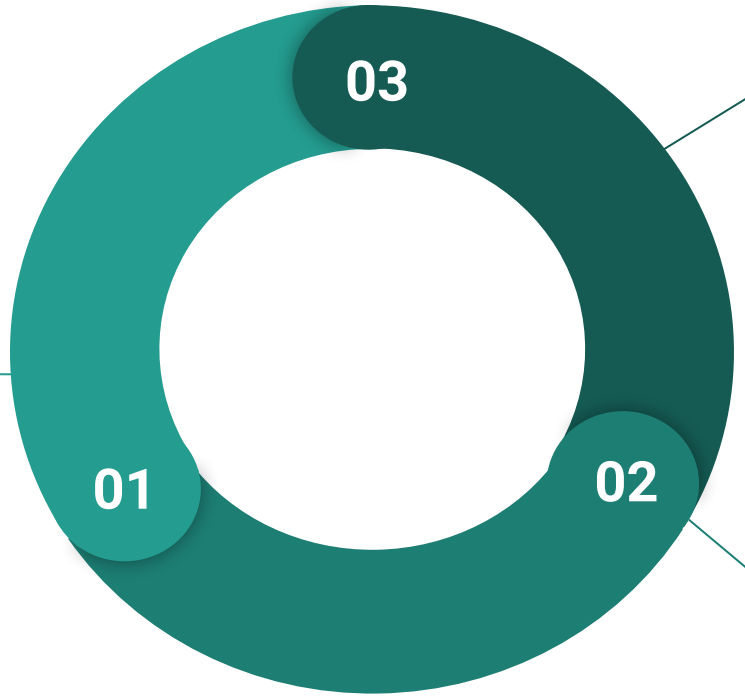
Overall Prediction Accuracy

Prediction

Song	Actual Genre	Predicted Genre
Old Town Road	**Hiphop**	**Hiphop**
Old Town Road (Ft. Billy)	**??**	**Country**
I want it that way	Pop	Pop
Metallica -Fade to black	Rock	Blues
Iron Maiden - Trooper	Rock	Pop

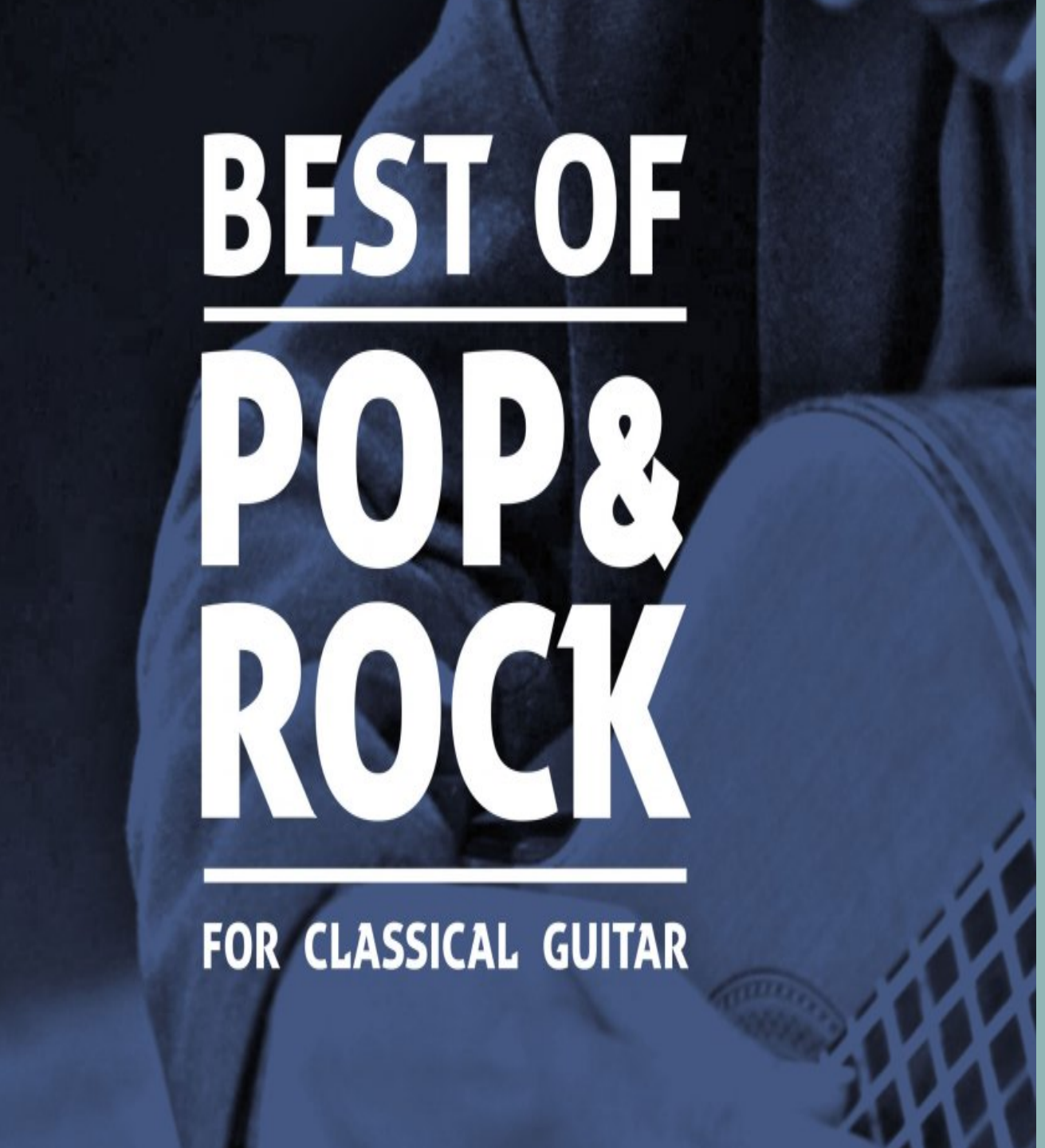
Conclusion and Future Recommendations

Old Town Road(Ft. Billy) is Country



Rebalancing data improves accuracy
Random forest performs best even after balancing data

Future Recommendations
Neural net
Add more genres including cross-genre (e.g. pop-rock)
Find a better way to rebalance data



Thank You !

Q&A