

The background of the slide is a solid black color, overlaid with a dense, repeating pattern of the Spotify logo. The logos are a vibrant green color and are slightly transparent, allowing them to overlap and create a textured effect. They are scattered across the entire slide, with some appearing larger and more prominent than others.

# SPOTIFY MUSIC GENRE CLASSIFICATION

*ATHARVA AGRAWAL, DHANUSH GOWDA, HARSHAD BANDI,  
KARTHIK M K, YESHPAL SINGH*

*25 - NOVEMBER - 2022*

# INDEX



- PROBLEM STATEMENT
- FEATURES
- METHODOLOGY
- RESULTS

The background of the slide is a dark green/black color with a pattern of Spotify logos. The logos are in a lighter green color and are of varying sizes and opacities, creating a subtle, repeating pattern. Some logos are solid, while others are semi-transparent, allowing the background to show through.

# PROBLEM STATEMENT

We have selected this data as it is used regularly in daily life by most of the people, which makes this project a real time use case. The tracks are classified with the genre of the playlist, posted by Spotify or relevant agencies from the music industry.

This data set can be used for both classification and clustering problems.

The aim of the project is to classify the songs based on its features to particular genre.

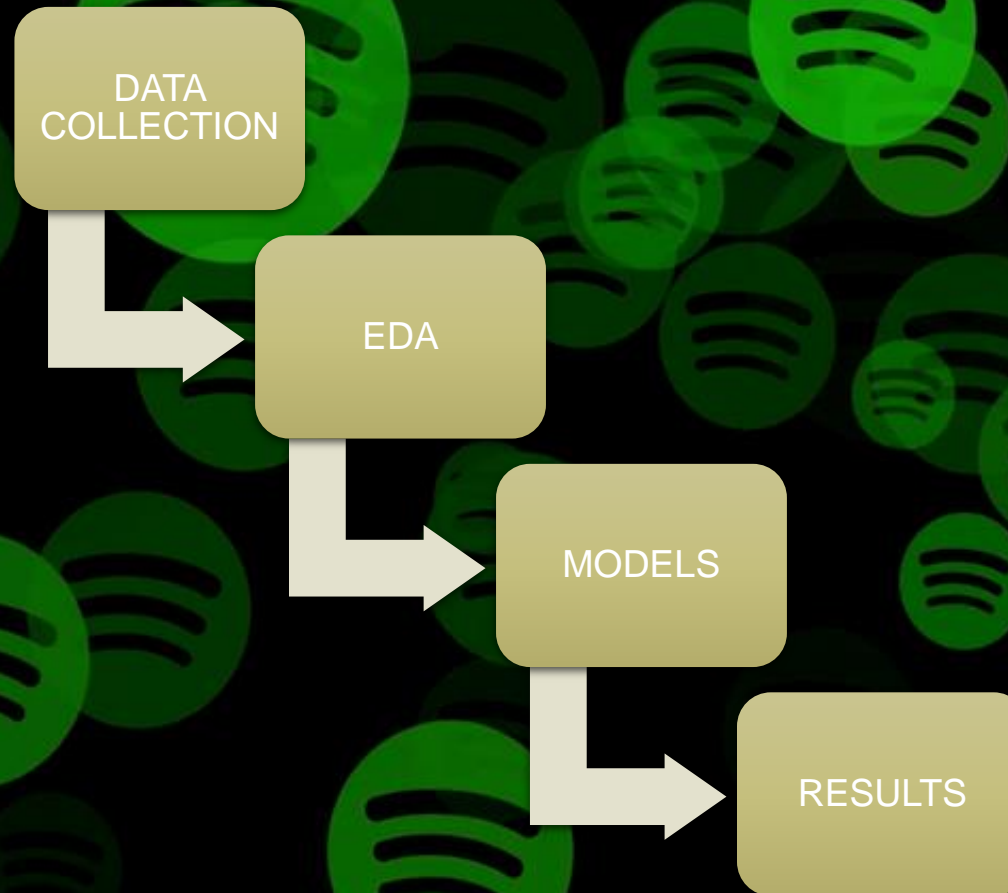


# FEATURES

- Track ID
- Playlist
- Playlist name
- Track name
- Track popularity
- Artist name
- Album
- Album cover
- Artist genres
- Artist popularity
- Danceability
- Energy

- Key
- Loudness
- Mode
- Speechiness
- Acousticness
- Instrumentalness
- Liveness
- Valence
- Tempo
- Duration ms
- Time signature
- Genre

# METHODOLOGY



The background of the slide is a repeating pattern of Spotify logos. The logos are green circles with three curved lines inside, representing the Spotify 'S' logo. They are arranged in a grid-like pattern, with some logos appearing slightly larger or more prominent than others, creating a textured effect. The logos are set against a dark background, and the overall color scheme is green and black.

# DATA COLLECTION

This data is downloaded from Kaggle where every record had been taken from Spotify API. The records in this data set contains both categorical and numerical columns. Data contains 9,198 attributes with 24 features.



# EXPLORATORY DATA ANALYSIS

- Dropping unwanted columns
- Missing value treatment
- Data conversion
- Hypothetical analysis
- Linearity
- Correlation
- Outlier treatment
- Value counts of target column
- Label encoding

# MODELS

- Naïve bayes
- Decision tree classifier
- Support vector machine
- Random forest classifier
- XGBoost classifier
- KNeighborsClassifier
- Naïve bayes
- Gradient Boost Classifier
- CatBoost Classifier



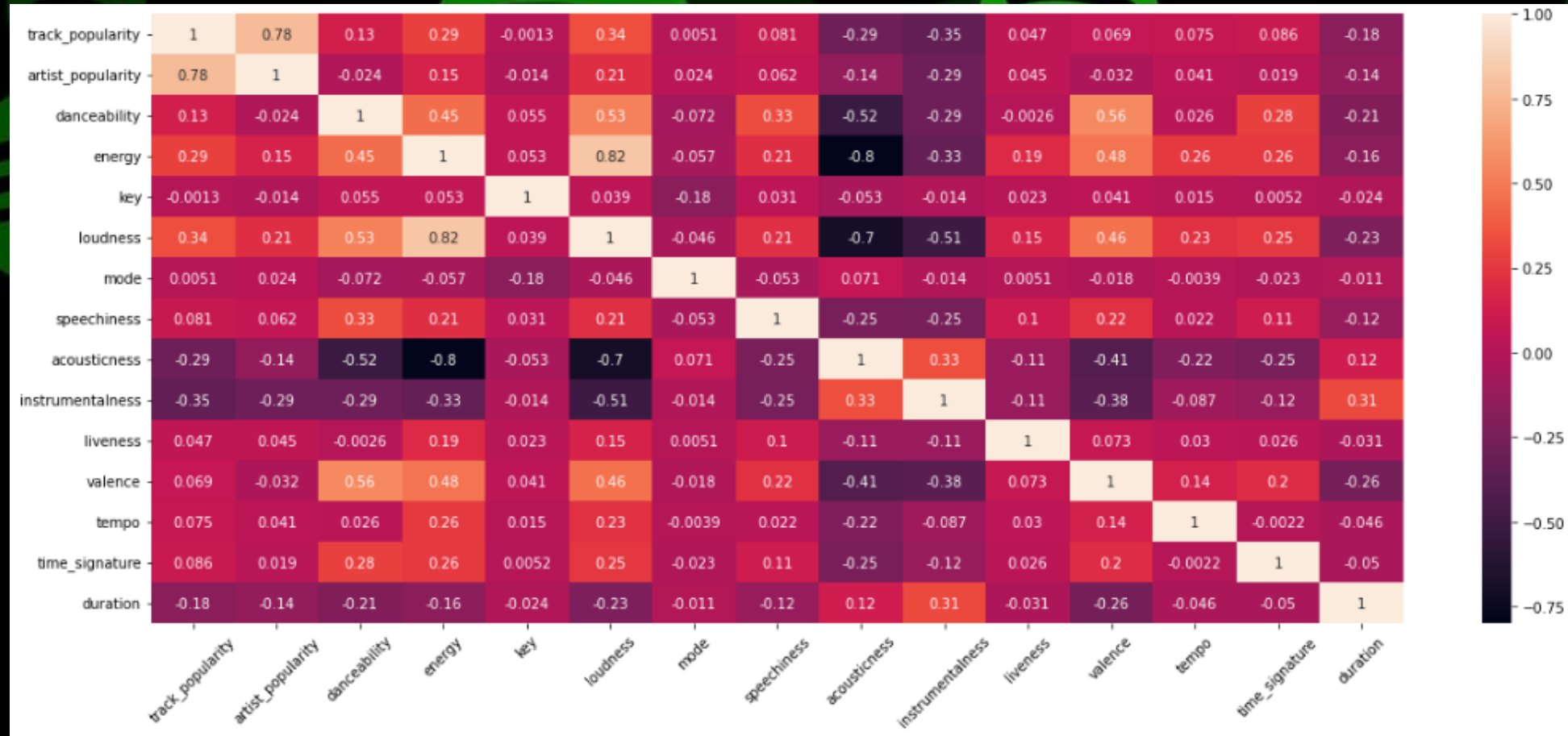
The background of the slide is a repeating pattern of Spotify logos. The logos are arranged in a way that they appear to be floating or scattered across the frame. The top and bottom sections of the slide have a dark background with bright green logos. The middle section, where the text is located, has a light pink background with semi-transparent, lighter green logos.

# RESULTS

# EDA and Visualization

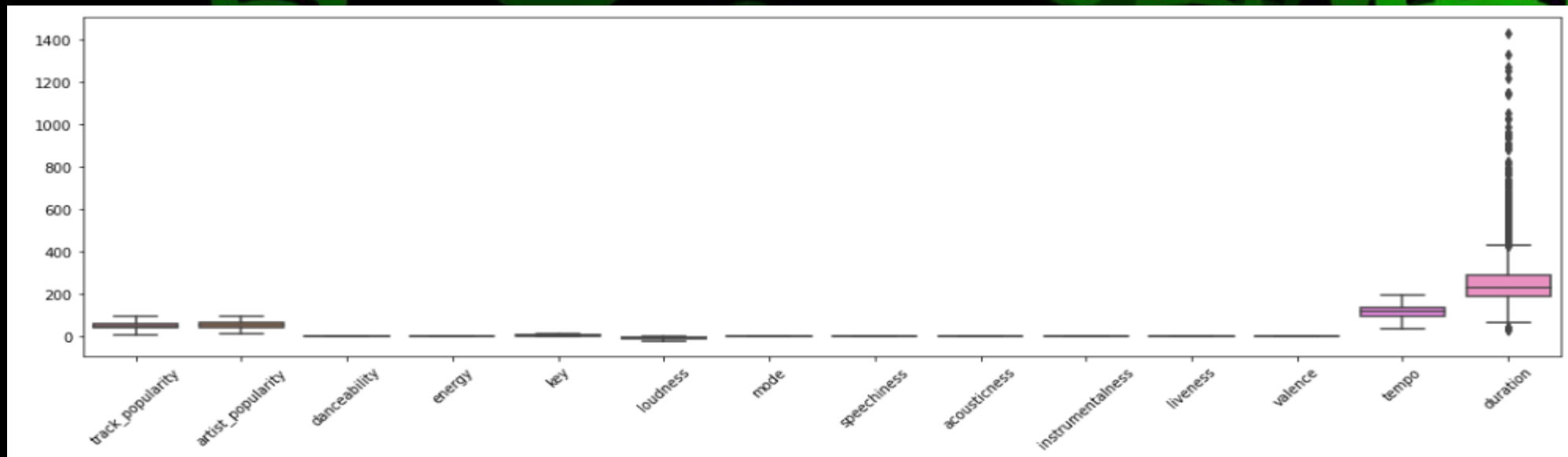
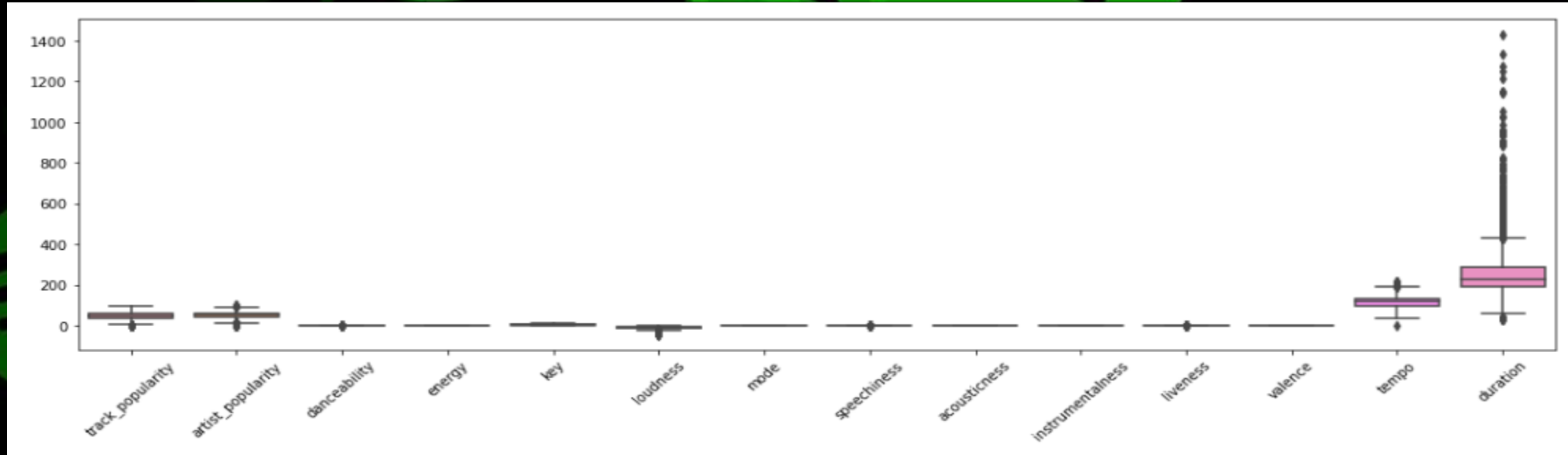
1. We dropped Track\_id column as it was 9198 data with unique values
2. There are no missing values in the data
3. Duration from milliseconds to seconds
4. From checking linearity, we observe that most of the columns are linearly related various ways
5. Hypothetical analysis suggests that the data is not scaled therefore we scale the data.

## 6. No columns have high correlation between each other

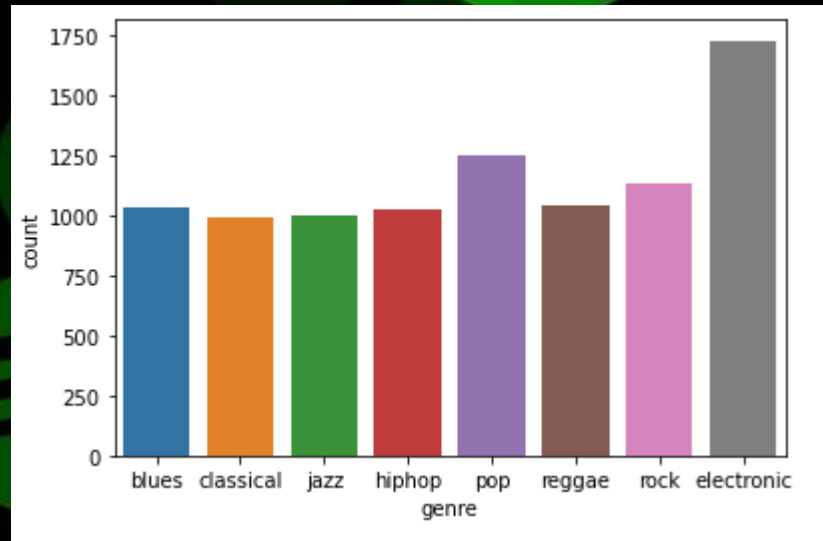




## 7. Treatment of outliers based on the thresholds and understanding



## . Value counts for target column



## 9. Conversion of categorical columns to numerical columns

# MODELS

- Naïve bayes gives 87% accuracy in classifying genres.
- Decision tree classifier model has 98.2609% accuracy
- Support vector machine gives 92.3370% accuracy
- Random forest classifier model provides 99.5109% accuracy
- Naïve bayes – 87% accuracy
- KNeighborsClassifier – 79.946% Accuracy

We can observe that from above models and all other models KNeighborsClassifier gives low accuracy.





# CONCLUSION

1. **People choose electronic music over other music genres.**
2. **XGBoost, Gradient Boost, and CatBoost Classifier with 99+% accuracy gives best model for classification.**

Code source : [SPOTIFY\\_GENRE\\_CLASSIFICATION](#)