



Exploring Music Classification with Spotify API

Predicting Song Categories

Team

Pınar Yıldız

Halil Kolatan

Muhammed Maral

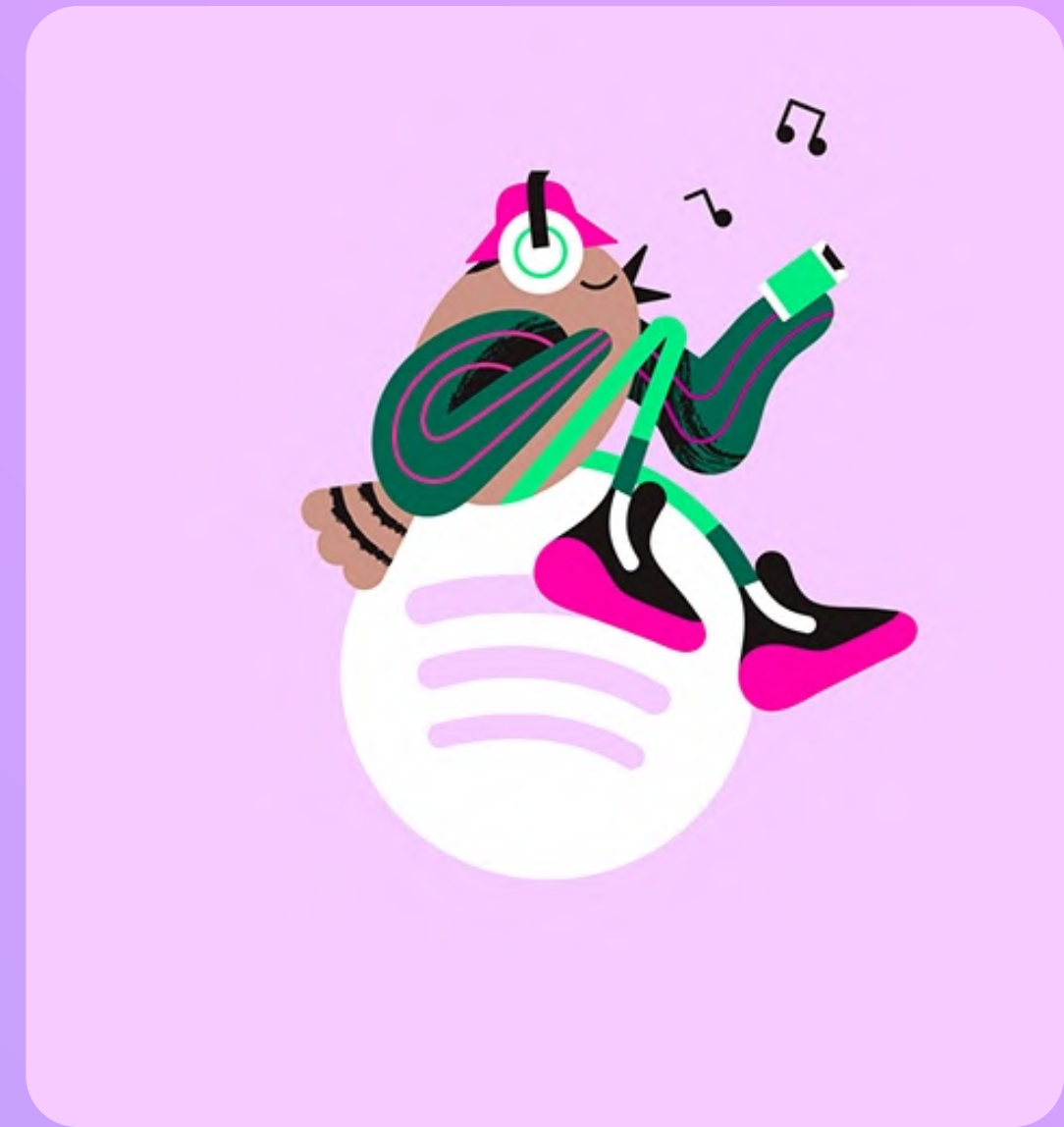
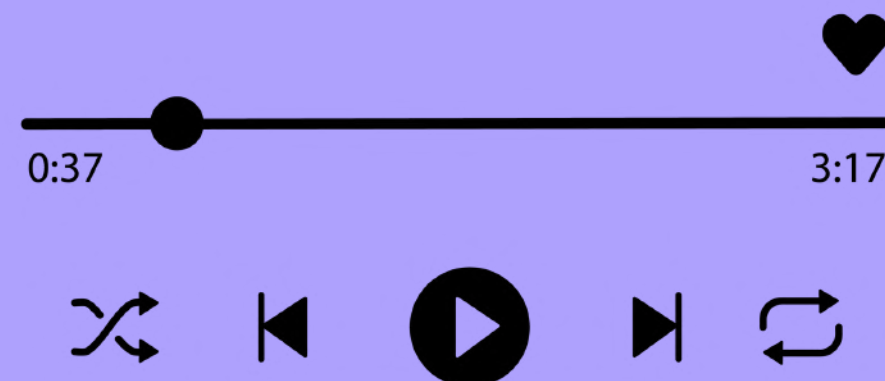


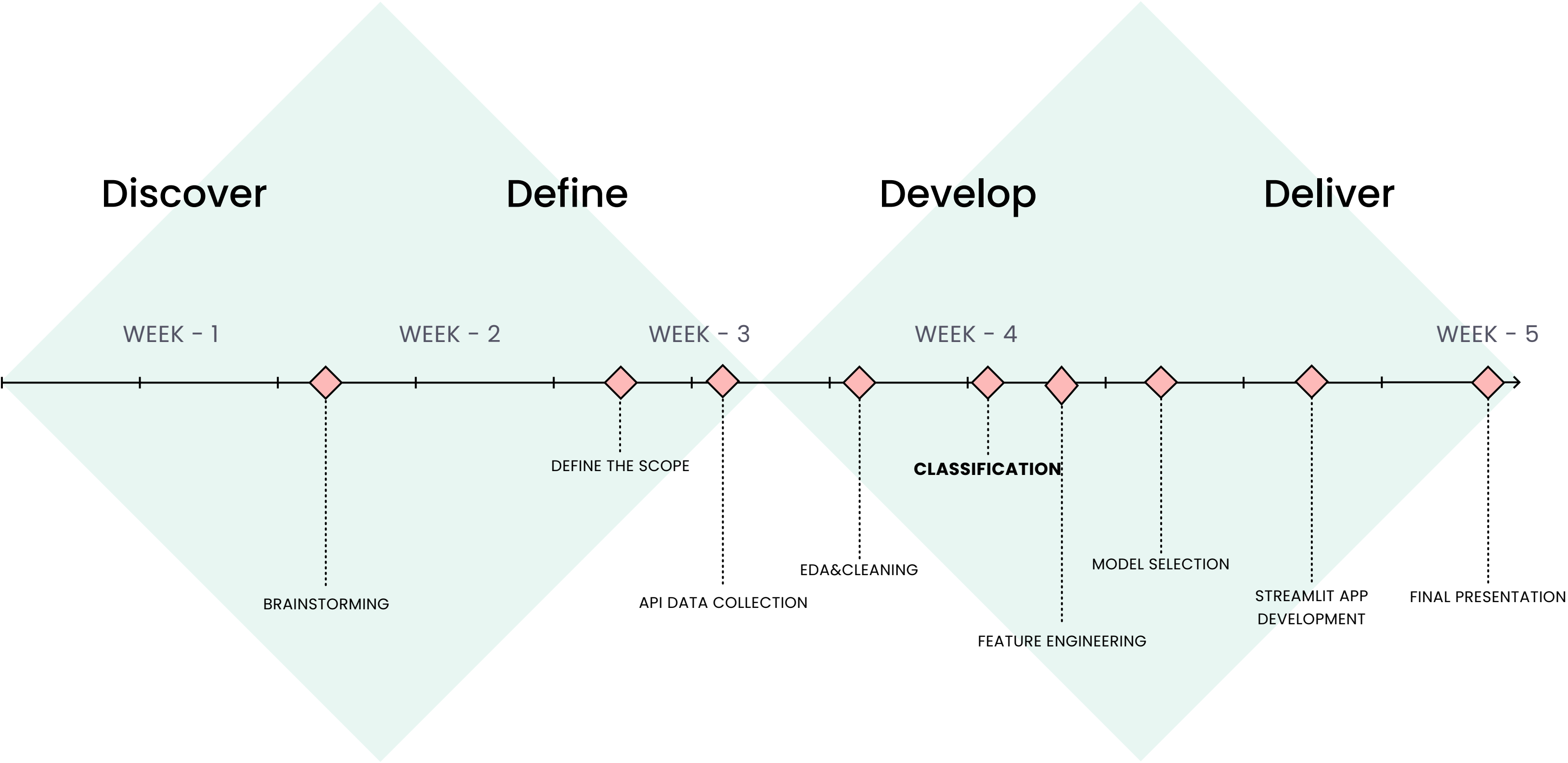
Table of Contents




1. Introduction
2. Data Collection
3. Exploratory Data Analysis
4. Data Cleaning & Classification
5. Feature Engineering
6. Model Selection
7. KNN
8. Random Forest Classifier
9. Gradient Boosting Classifier
10. Streamlit Live-Demo
11. Conclusion & Future Work

1. INTRODUCTION

PROJECT OVERVIEW



MOTIVATION

You are  *What you listen to*

PERSONA



Data Engineer and CO-Owner of DS Academy

ZEKERIYA

AGE: 48

EDUCATION: COMPUTER PROGRAMMING

LOCATION: ISTANBUL

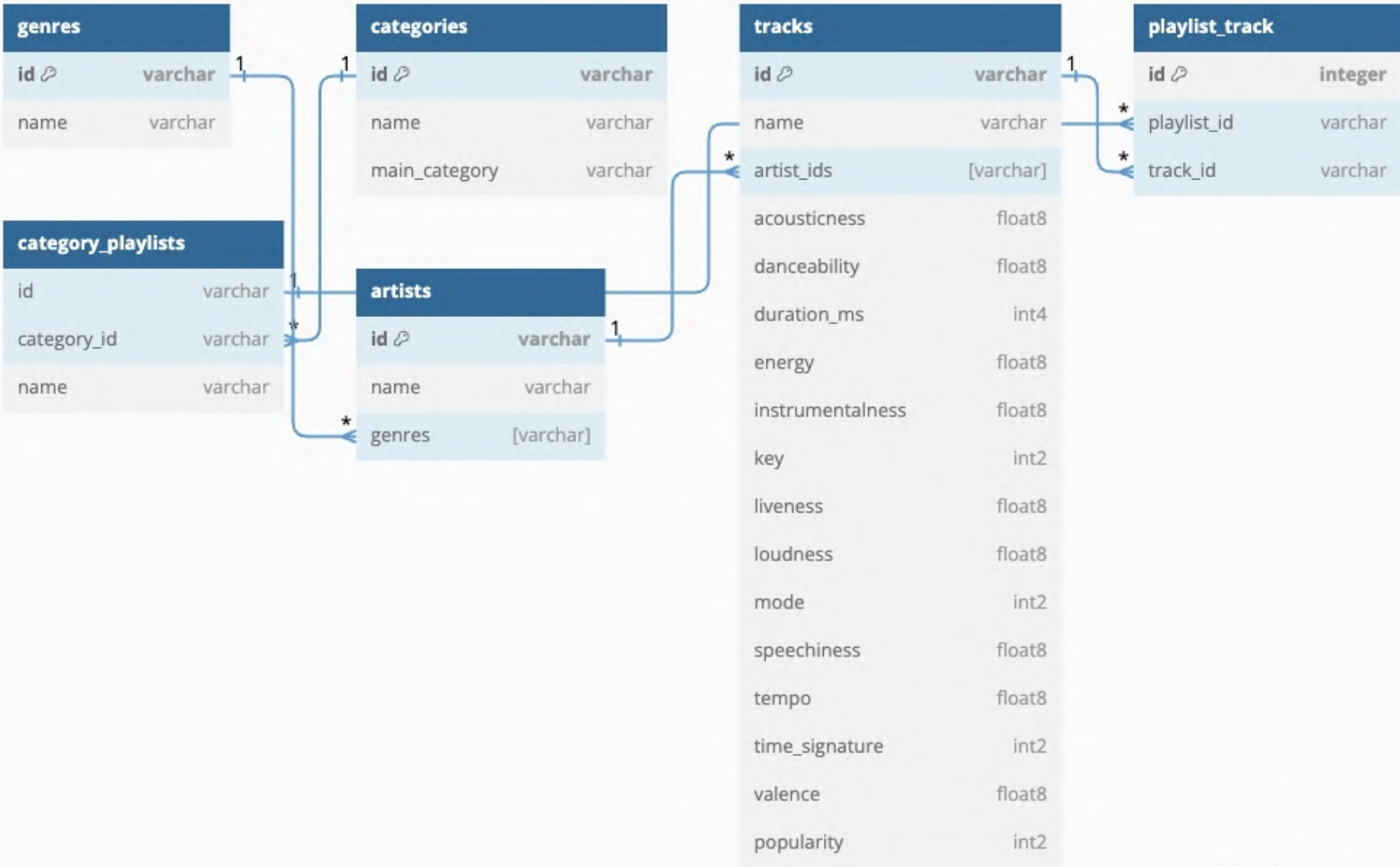
Meet Zekeriya Besiroglu, a passionate Big Data Architect and Google Cloud Certified Engineer. When it comes to music, Zekeriya's heart beats for rock. With our Spotify genre classification project, he explores music that matches his taste and energy level. Always in the mood for active songs, Zekeriya uses the project's personalized inputs to discover new rock tracks that keep him in high spirits.

2.DATA COLLECTION

TOOLS USED

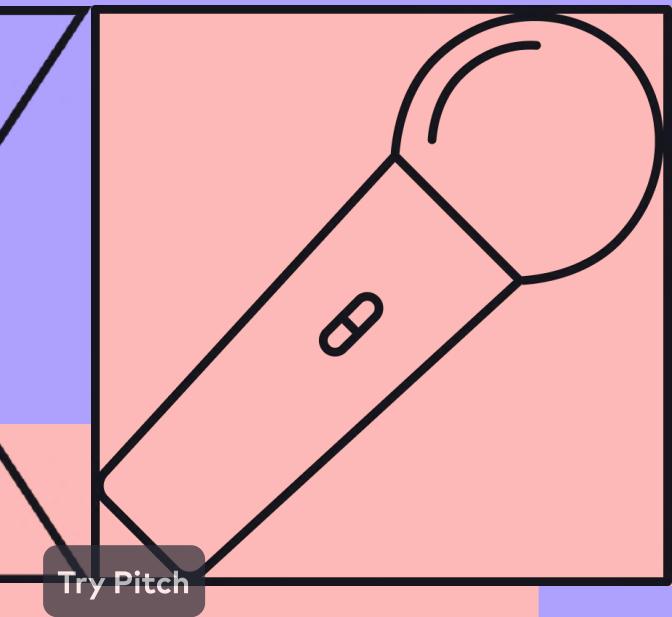


API STRUCTURE



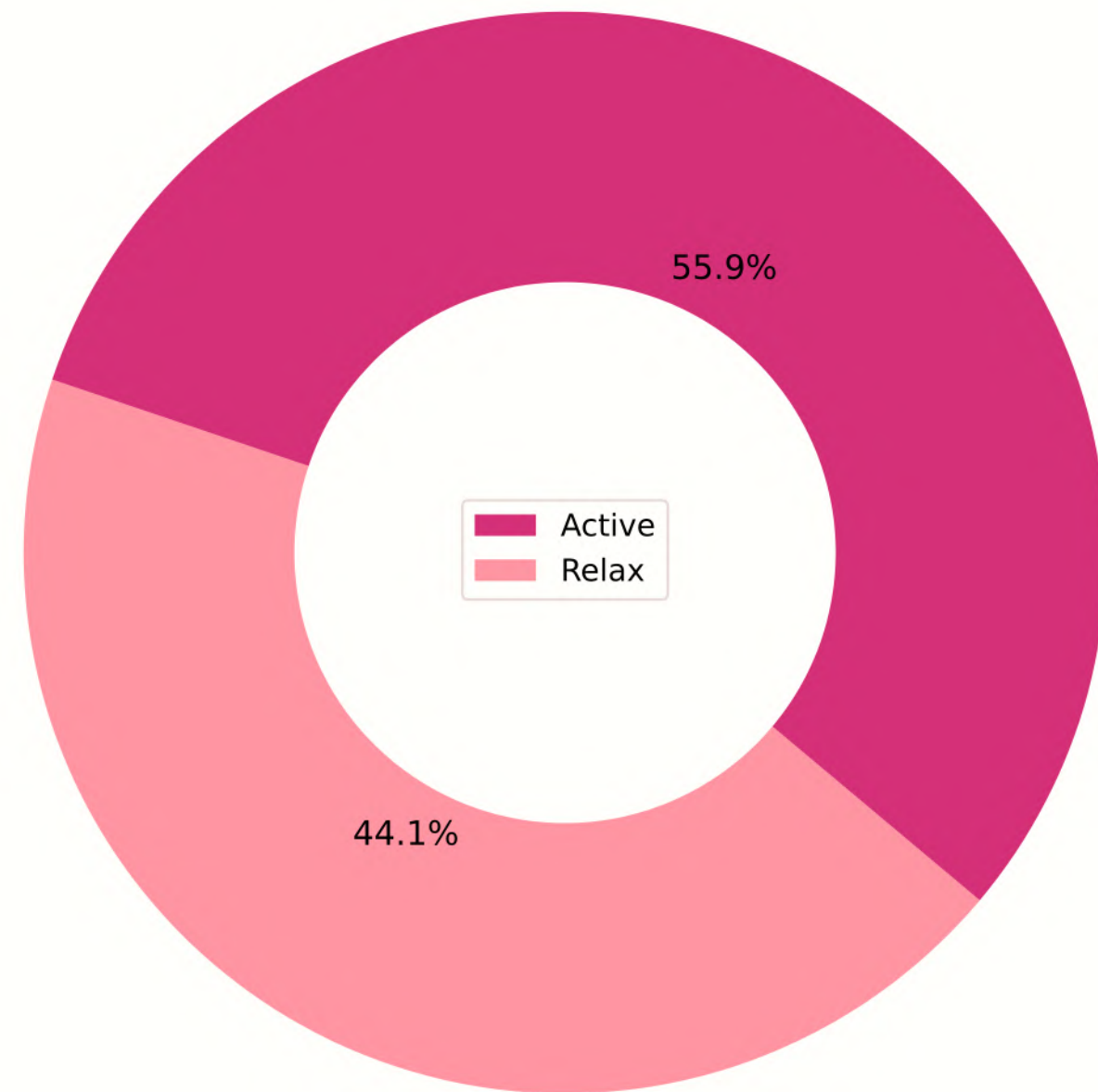
DATAFRAME

	id	name	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness	mode	speechiness	tempo	time_signature	valence	popularity	category_name
0	087fGGq27BXpxKAlbBwNM	Rhythm Is A Dancer - 7' Edit	0.27300	0.794	225882	0.7450	0.000414	9.0	0.1430	-11.367	0.0	0.0370	124.249	4.0	0.7060	32	relax
1	0CsM8VGDi38kusMv3pxyj1	True - Single Edit	0.45900	0.742	329146	0.5330	0.148000	0.0	0.0640	-9.098	1.0	0.0288	97.395	4.0	0.3910	73	relax
2	0DiWol3AO6WpXZgp0goxAV	One More Time	0.01940	0.613	320357	0.6970	0.000000	2.0	0.3320	-8.618	1.0	0.1330	122.746	4.0	0.4760	80	relax
3	0LzeivEHO16a8eBQGlpVKE	Love Shack - Edit; 2019 Remaster	0.11000	0.704	262586	0.8280	0.000000	5.0	0.7470	-7.304	0.0	0.0514	133.496	4.0	0.8660	61	relax
4	0MHXrqn909p0LRTPsNsGEi	Move on Up - Single Edit	0.41700	0.538	165789	0.9690	0.001540	11.0	0.1190	-5.626	0.0	0.0906	138.652	4.0	0.9280	73	relax
...
29040	2riw7lzJhEHDNDNQKj3ml9	Sojourner Truth	0.96800	0.264	330146	0.0783	0.251000	8.0	0.0722	-19.087	1.0	0.0457	82.021	4.0	0.0634	22	NaN
29041	6T572wD8zgOLOlB9TzxUmh	Pool Party	0.02930	0.543	238000	0.6120	0.653000	0.0	0.1070	-7.832	0.0	0.0318	172.035	4.0	0.2770	48	NaN
29042	6lrRlQDcikJ5y5DtyZPuJ	Can't Let You Go (feat. Mike Shorey & Lil' Mo)	0.23100	0.646	223973	0.6000	0.000000	9.0	0.0794	-6.569	1.0	0.4580	192.082	4.0	0.8110	68	NaN
29043	7ITCyjmVcBWKAE38V8epGU	Swaggin' At The Partment	0.00183	0.562	124667	0.9310	0.884000	9.0	0.5240	-5.377	1.0	0.0743	92.532	4.0	0.4670	11	NaN
29044	42F16X6WRCAFWPyRY4nH7O	Dou	0.98200	0.540	369626	0.1670	0.833000	1.0	0.0643	-15.765	0.0	0.0393	93.990	3.0	0.3940	41	NaN
29045 rows × 17 columns																	



3. EXPLORATORY DATA ANALYSIS

Category Distribution



- Total Data → 29045 Rows & 17 Columns
- Total 14 Features
- **Two Main Category**
 - **Active:** Pop, Hip-Hop, Top Lists, Trending, Dance/Electronic, Rock, Workout, Summer, Focus
 - **Relax:** Chill, Sleep, Classical, Student, Travel, Tastemakers

4.DATA CLEANING & CLASSIFICATION

Removed Columns

Song Name

Song ID

Removed Rows

- **Duplicated Songs** —→ **9809 Rows**
- **Rows without Category Name** —→ **3914 Rows**
- **Empty Rows** —→ **75 Rows**

Edited Columns

Category Name —→ **Mood**

Active & Relax

1 & 0

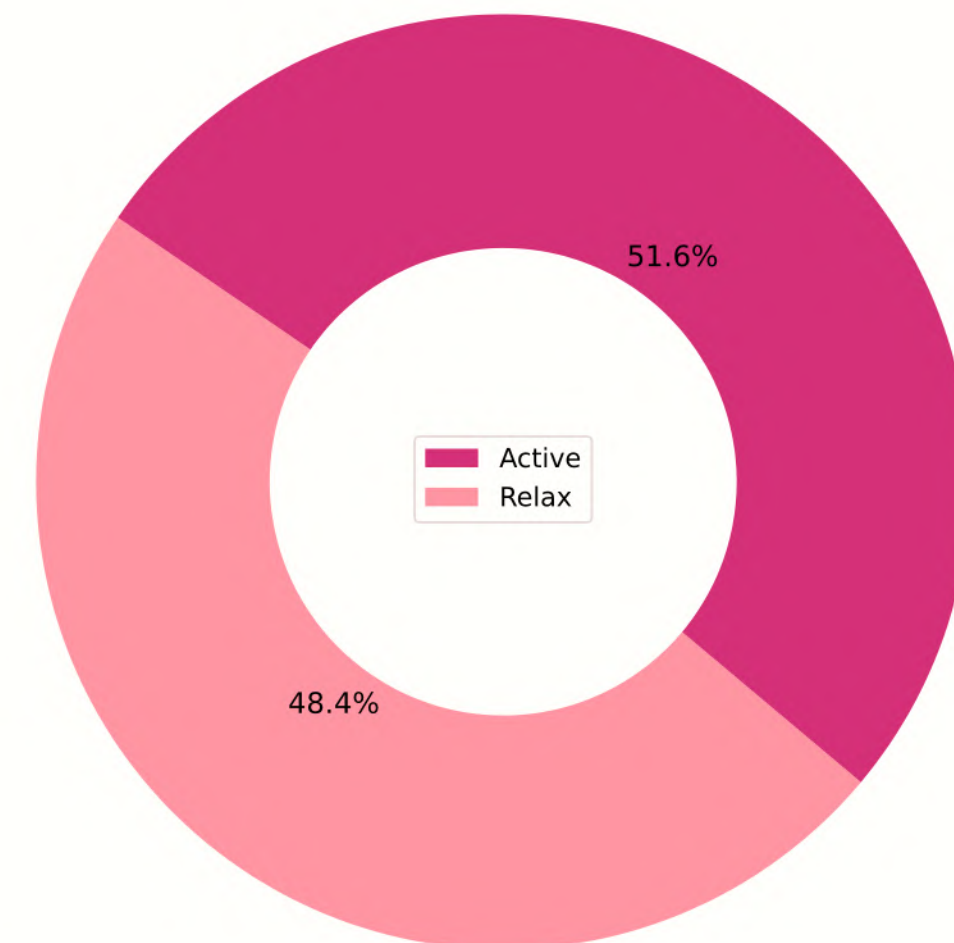
New Dataframe Rows

29045



15247

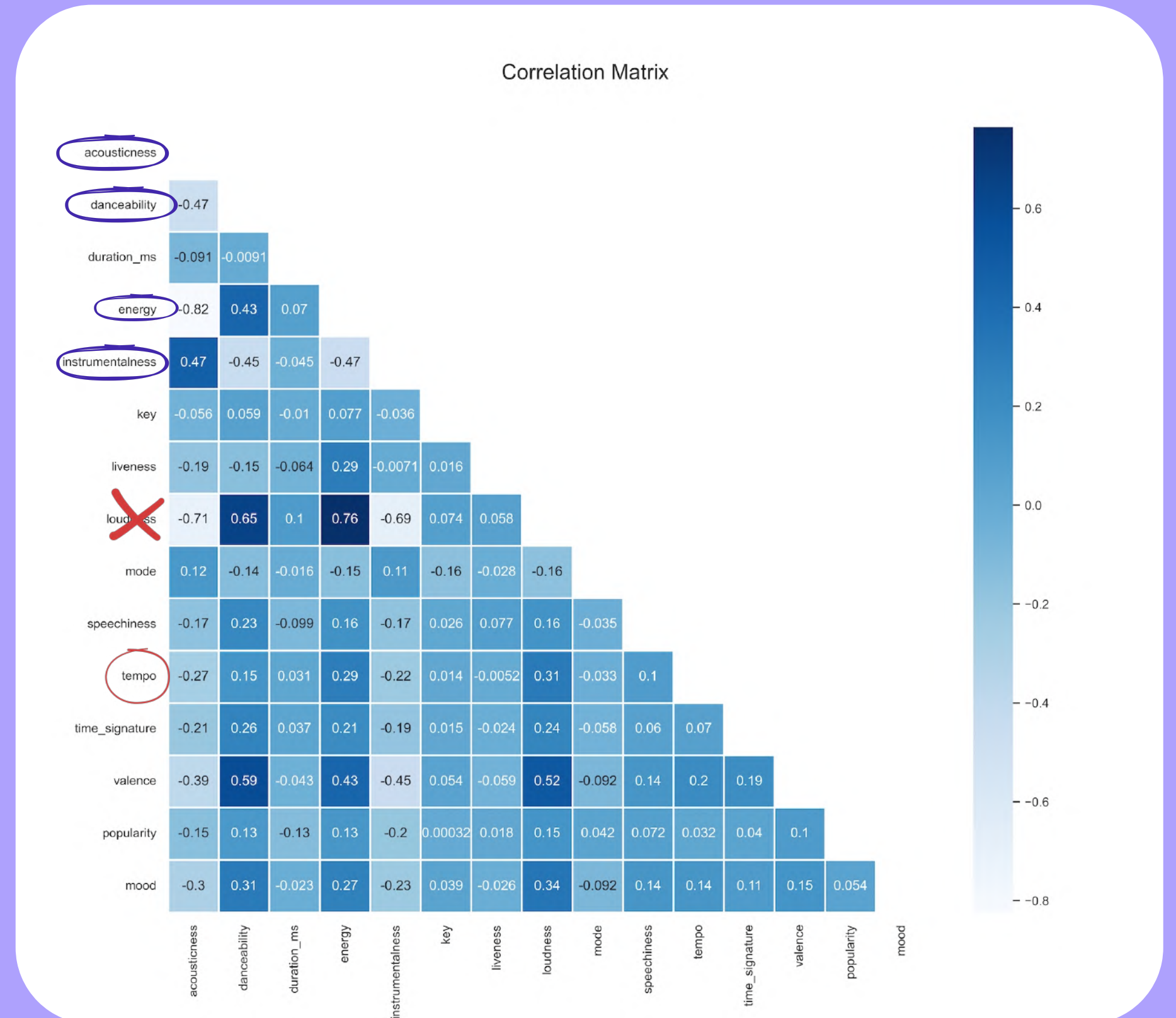
Category Distribution



5. FEATURE ENGINEERING

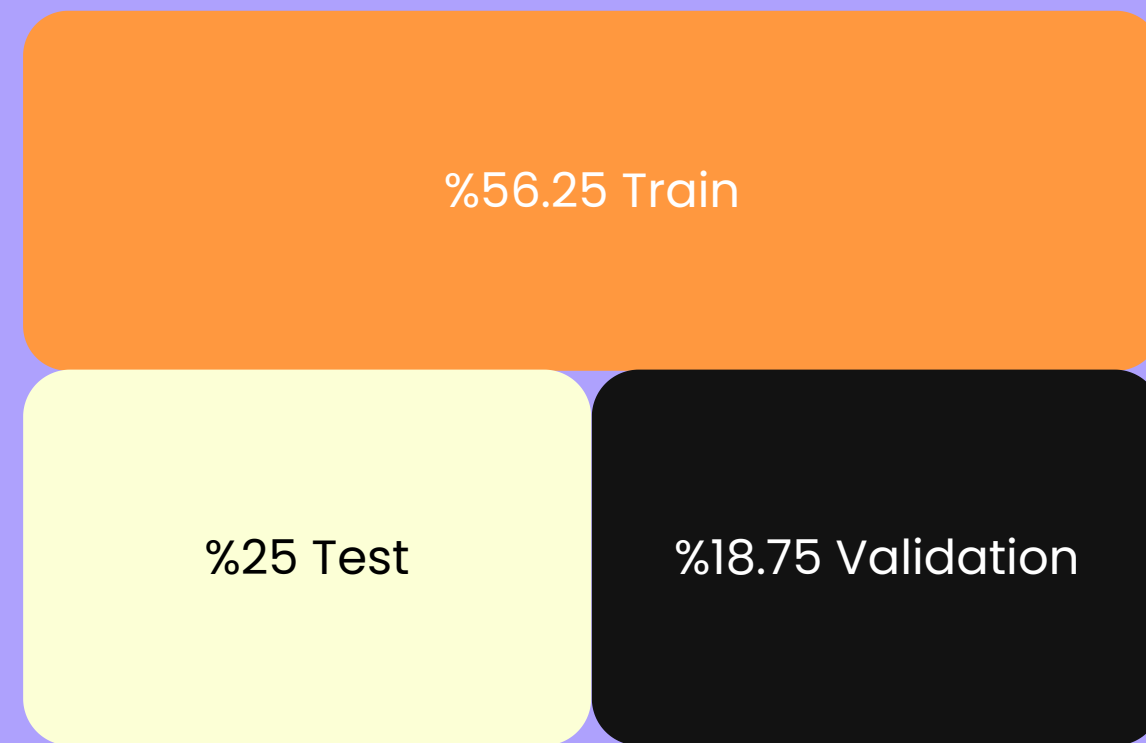
5. Feature Engineering

- The relationships between the features in our dataset are **generally weak**.
- **Important Features:** Energy, Acousticness, Danceability, Instrumentalness + Tempo
- These four features are important factors that influence our models. We will examine them in detail later.

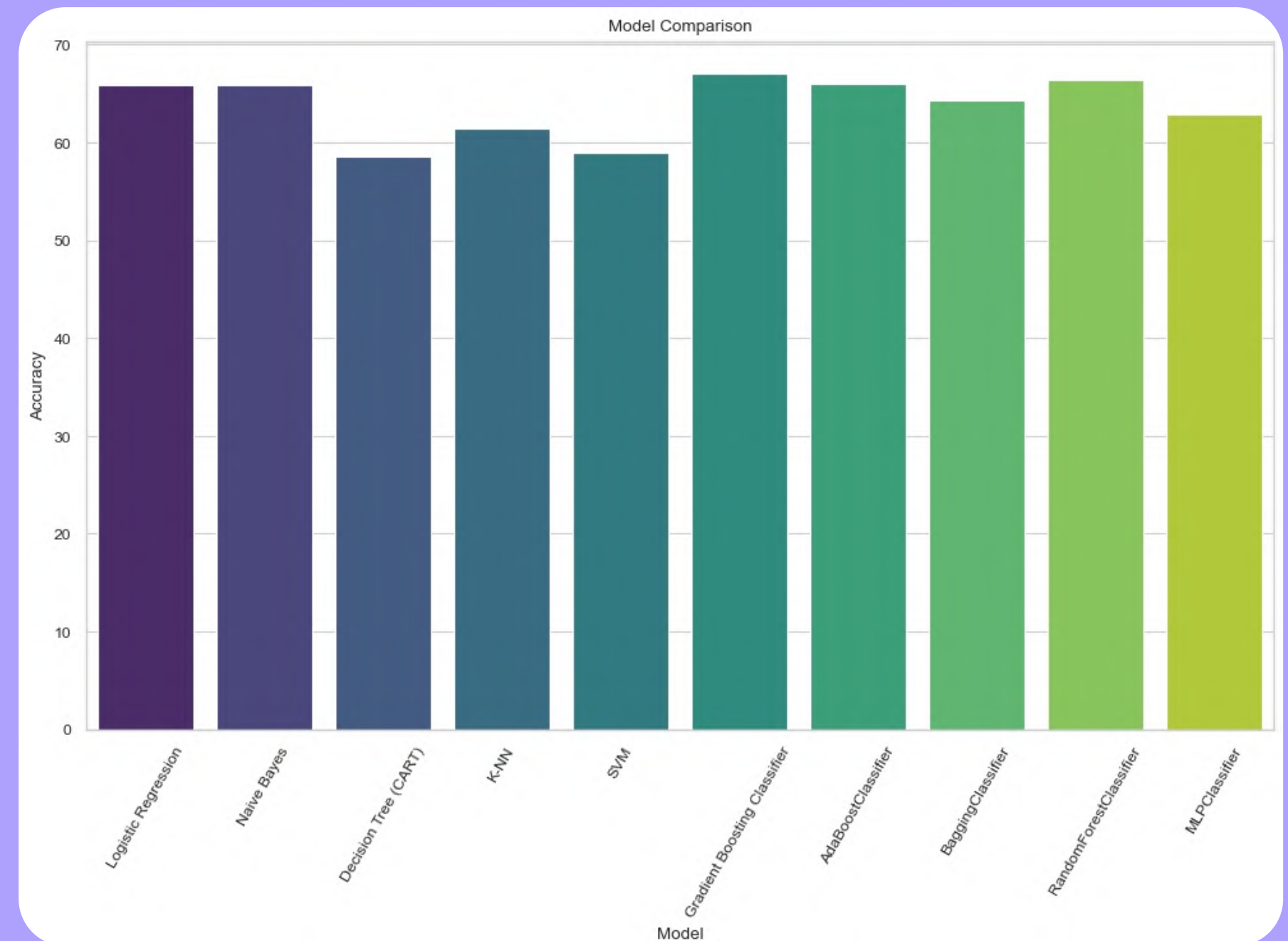


6. MODEL SELECTION

Train-Validation-Test Split

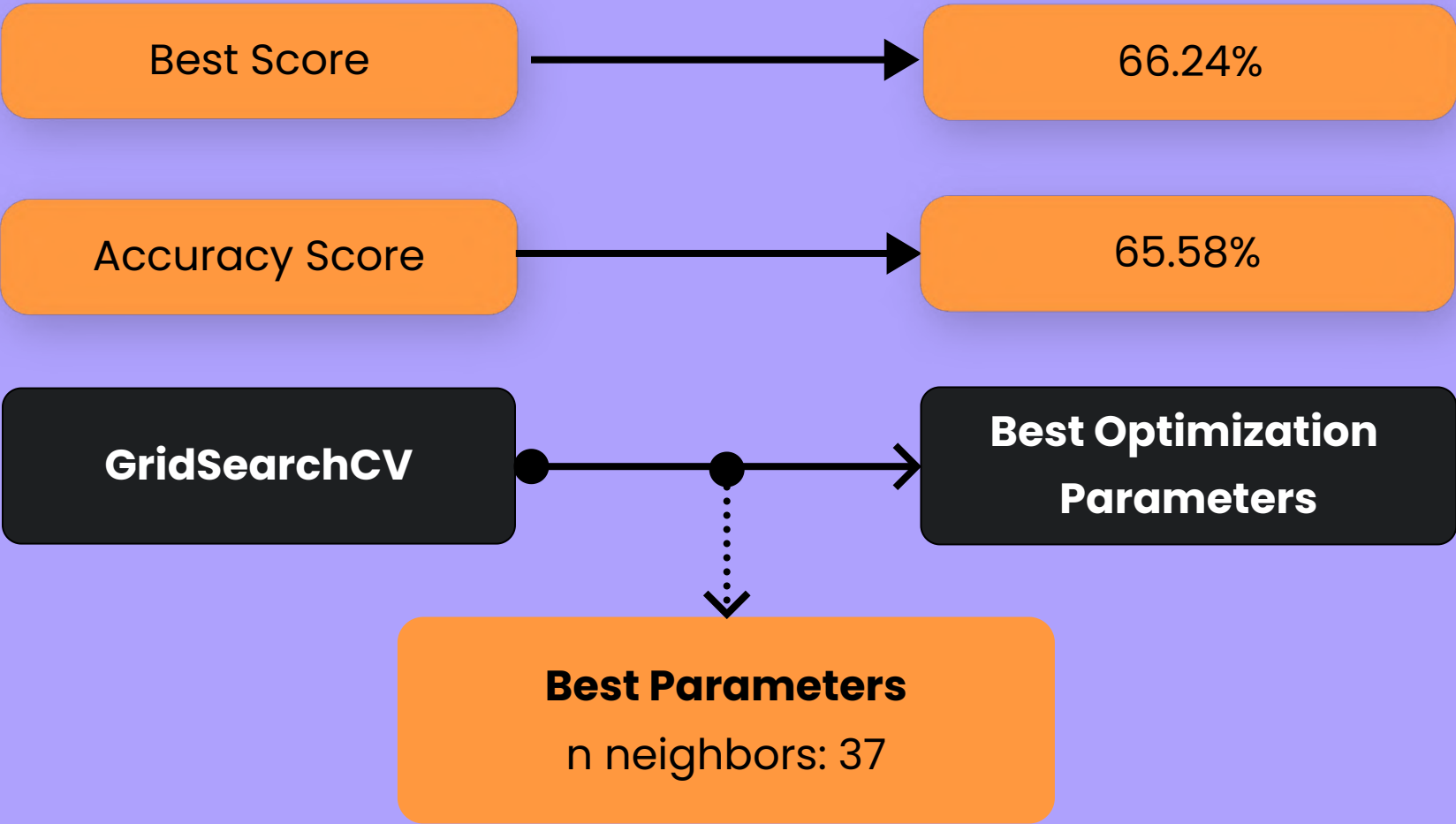


- **Model Selection:** We have selected three models, namely **KNN**, **Random Forest Classifier** and **Gradient Boosting Classifier**, for conducting detailed analysis on our dataset.
- These models will be examined in detail.



7.KNN

Model Control

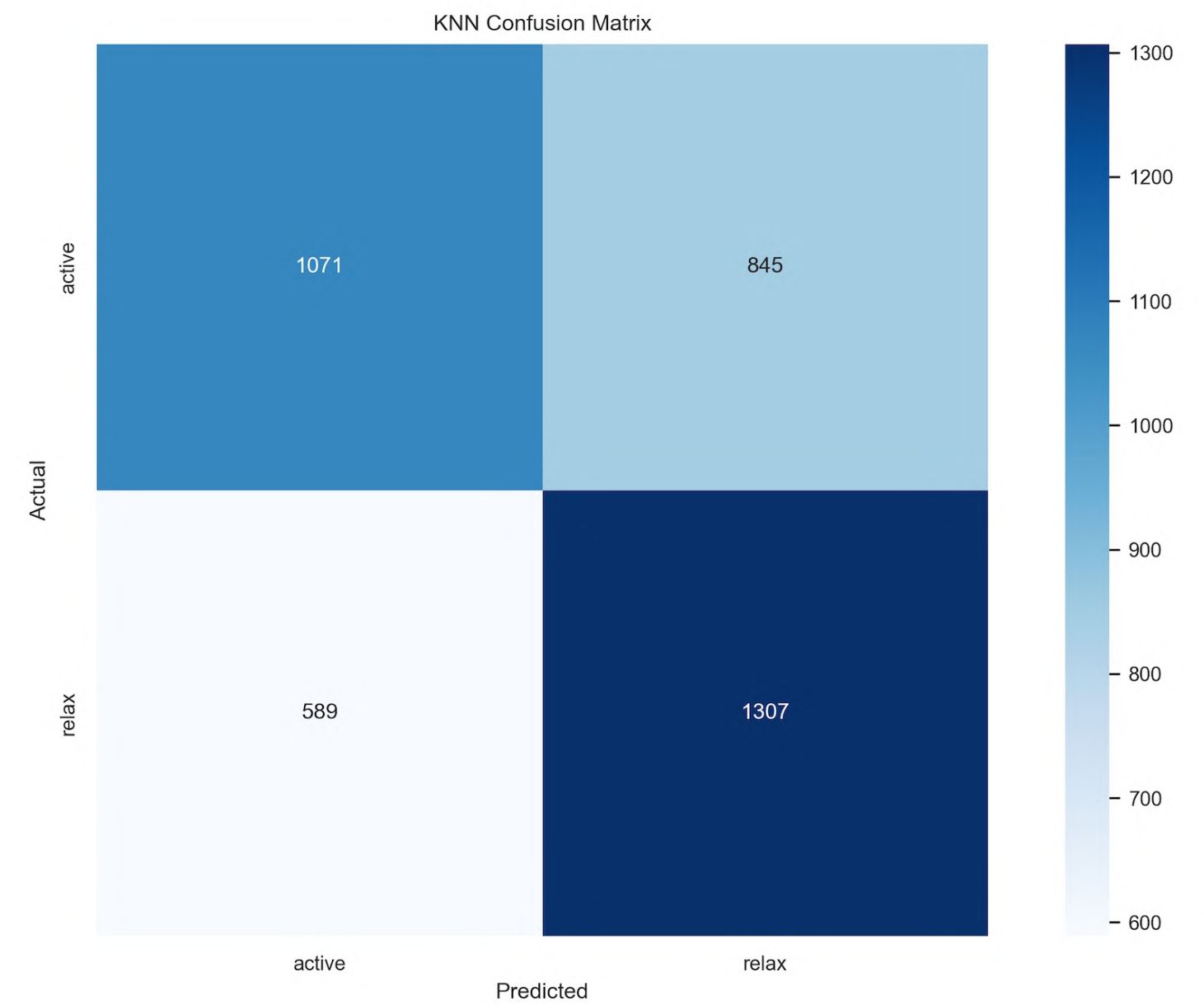
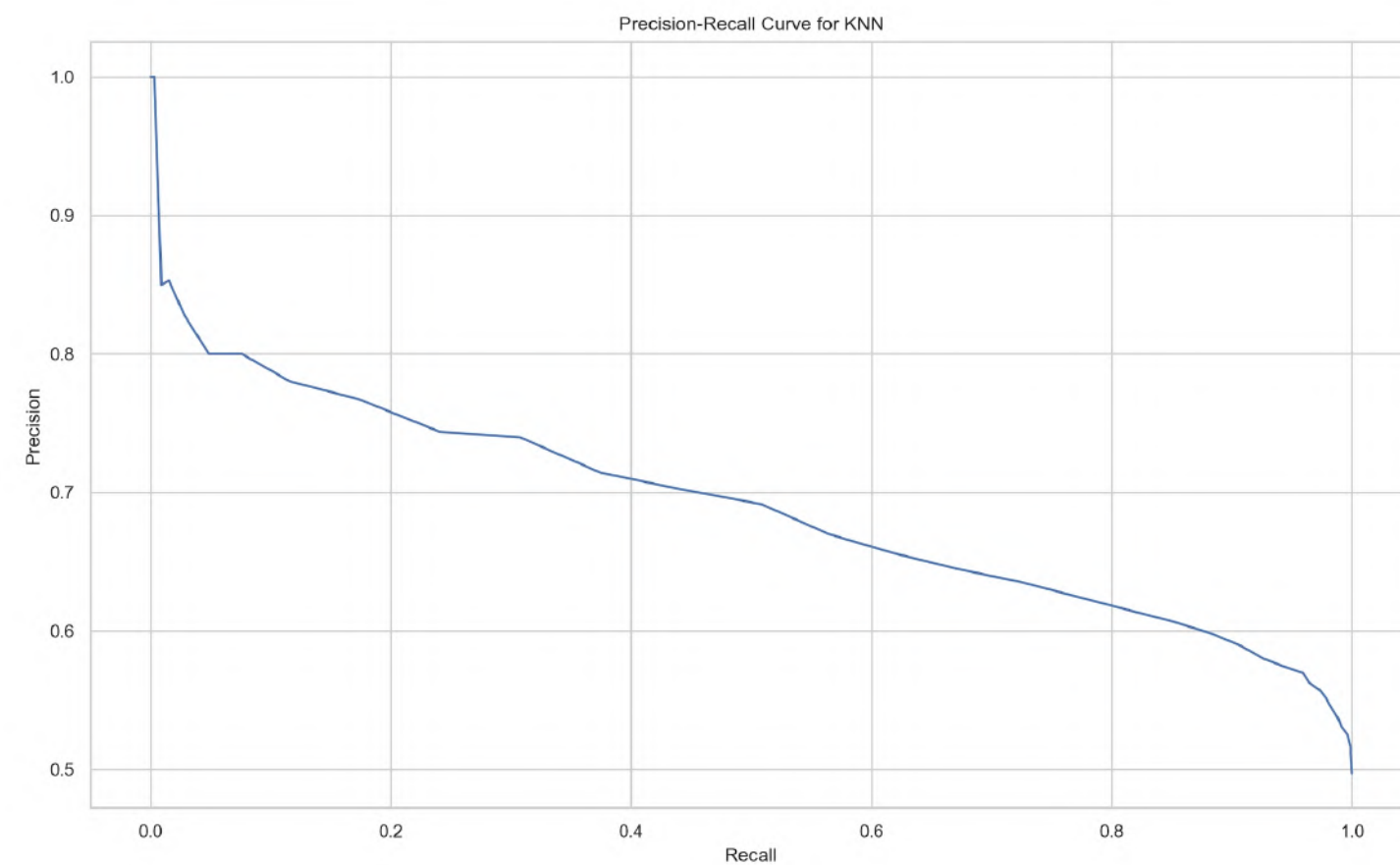


Model Evaluation

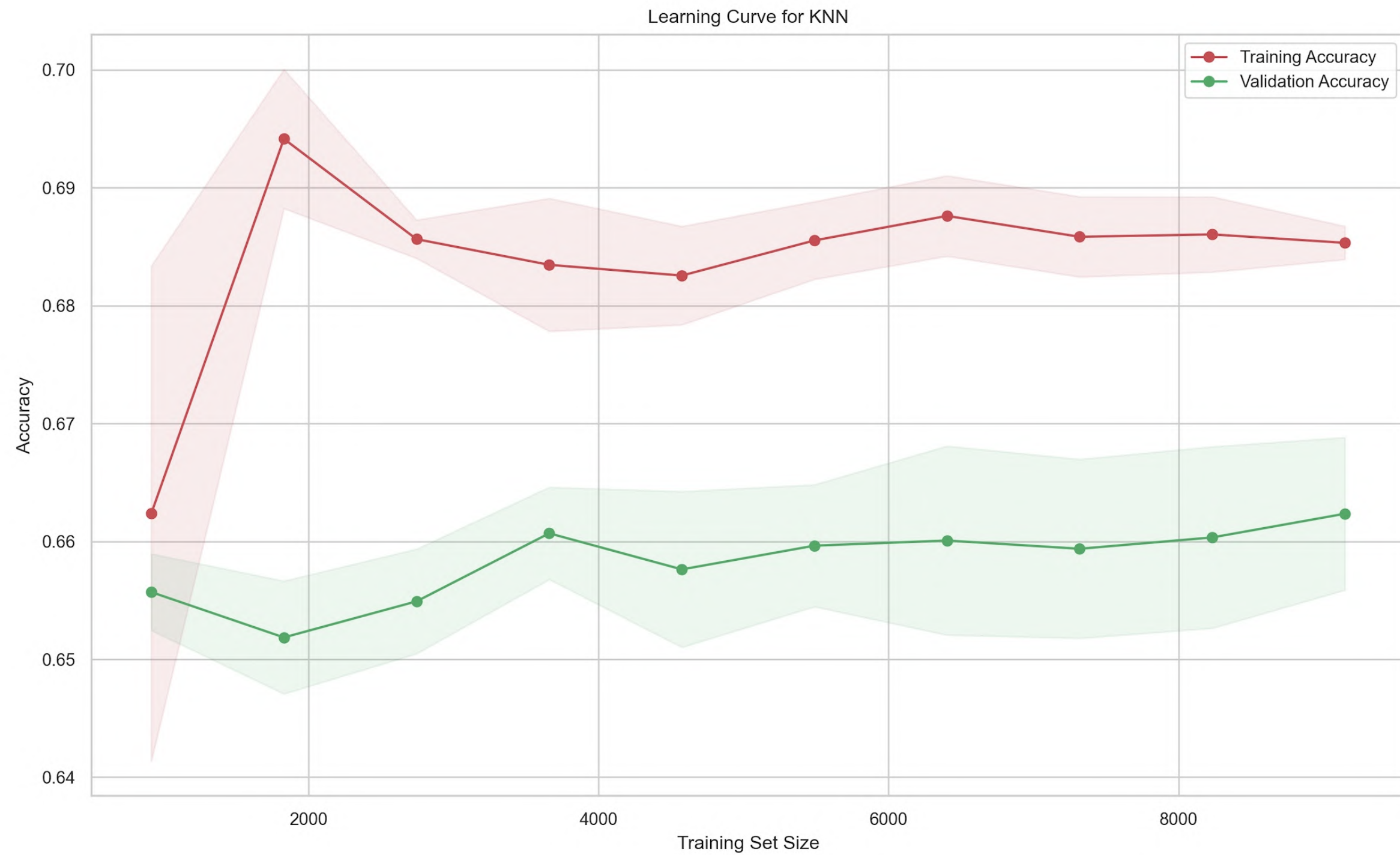
<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Class 0	0.71	0.54	0.61
Class 1	0.62	0.77	0.69
Macro Avg	0.66	0.66	0.65
Weighted Avg	0.67	0.66	0.65



Model Evaluation

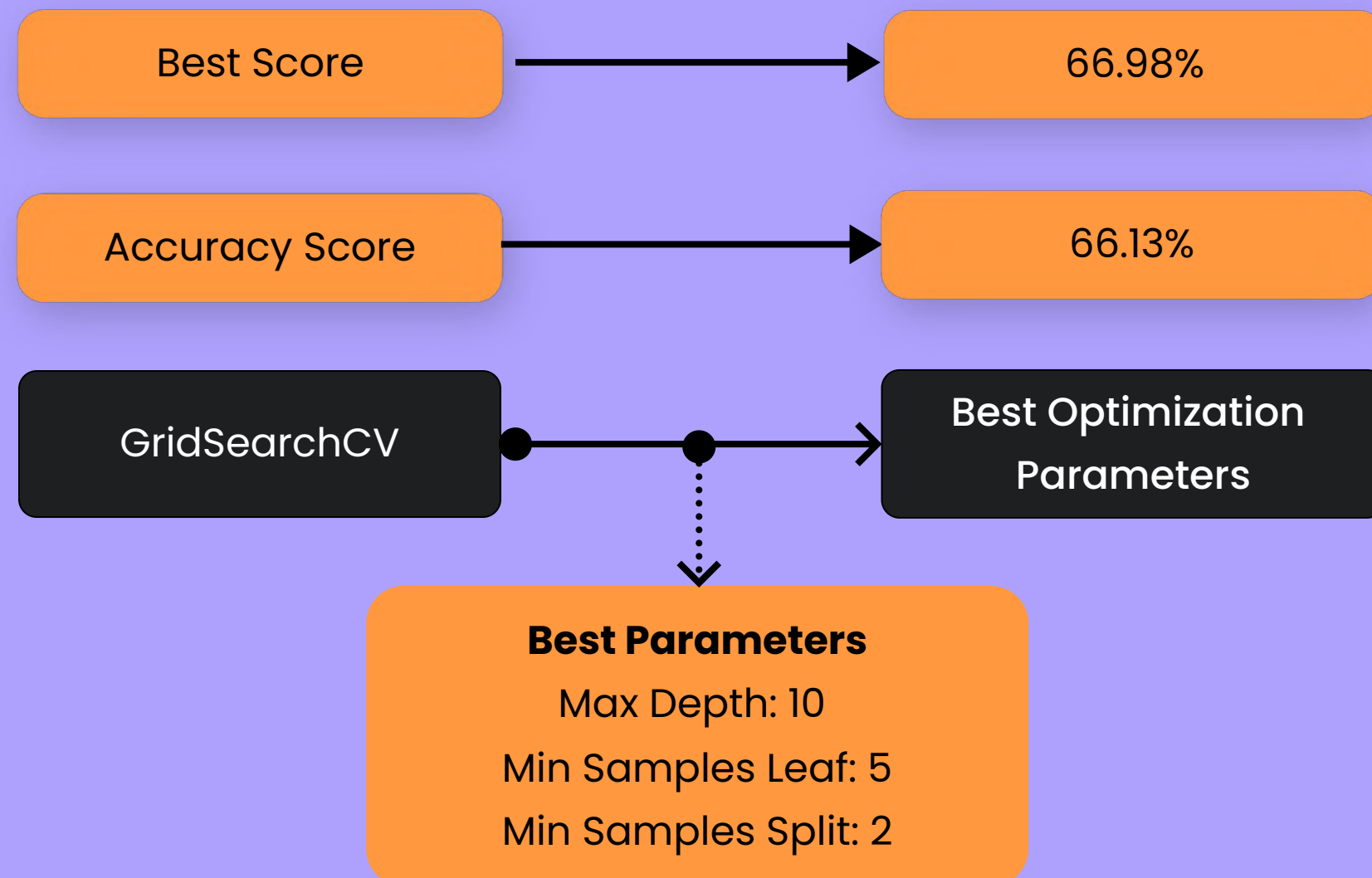


Learning Curve



8. RANDOM FOREST CLASSIFIER

Model Control



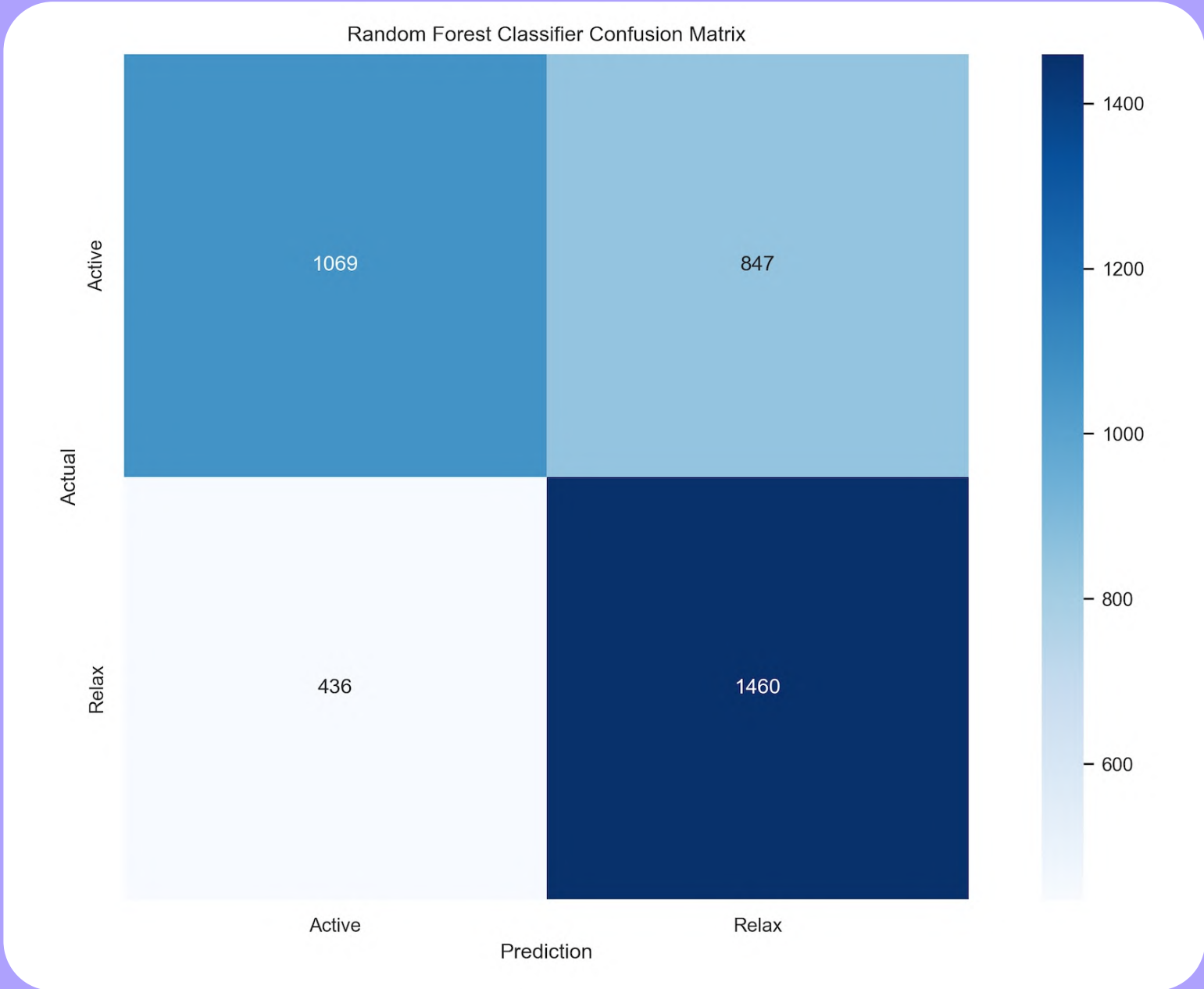
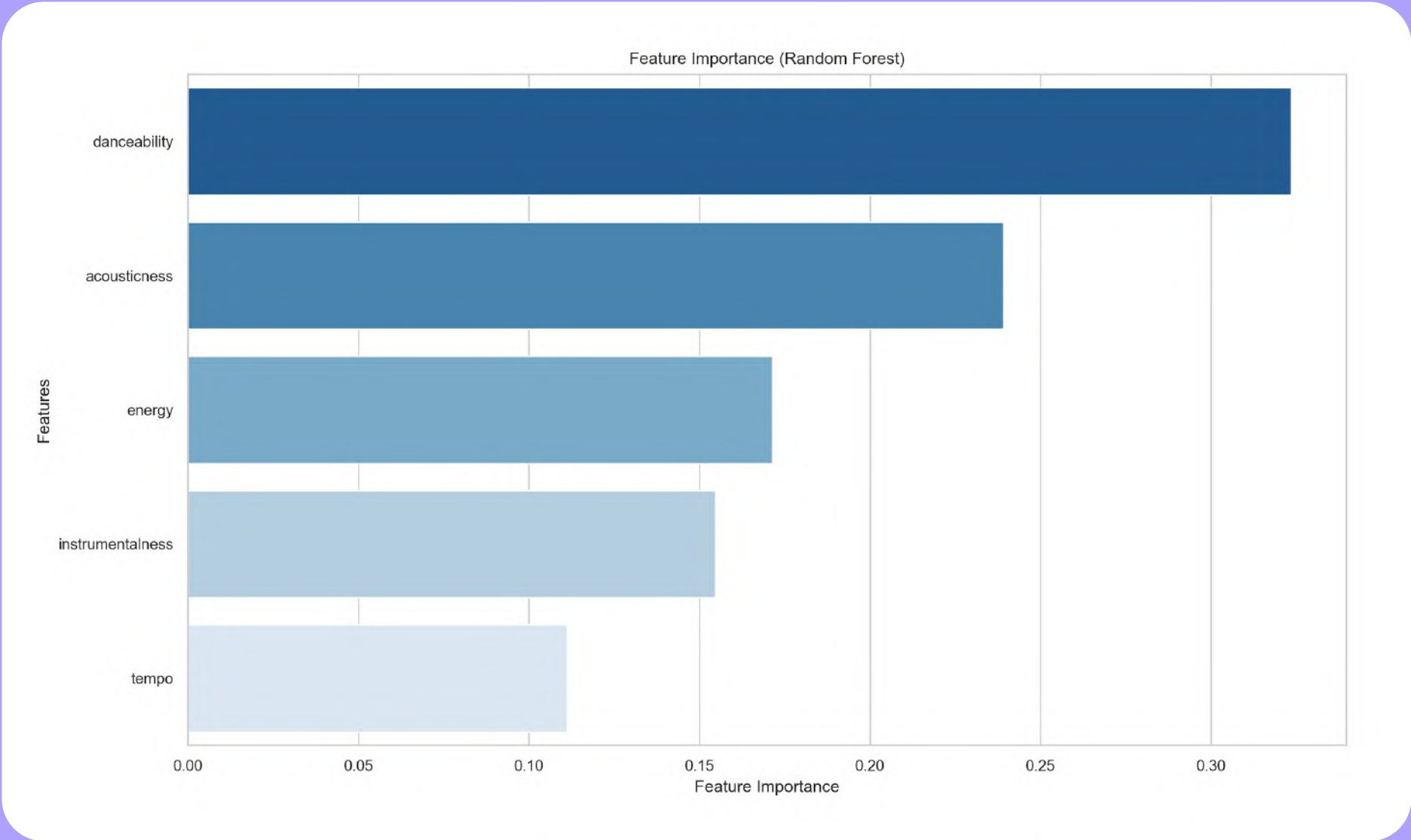
Model Evaluation

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Class 0	0.70	0.56	0.63
Class 1	0.63	0.76	0.69
Macro Avg	0.67	0.66	0.66
Weighted Avg	0.67	0.66	0.66

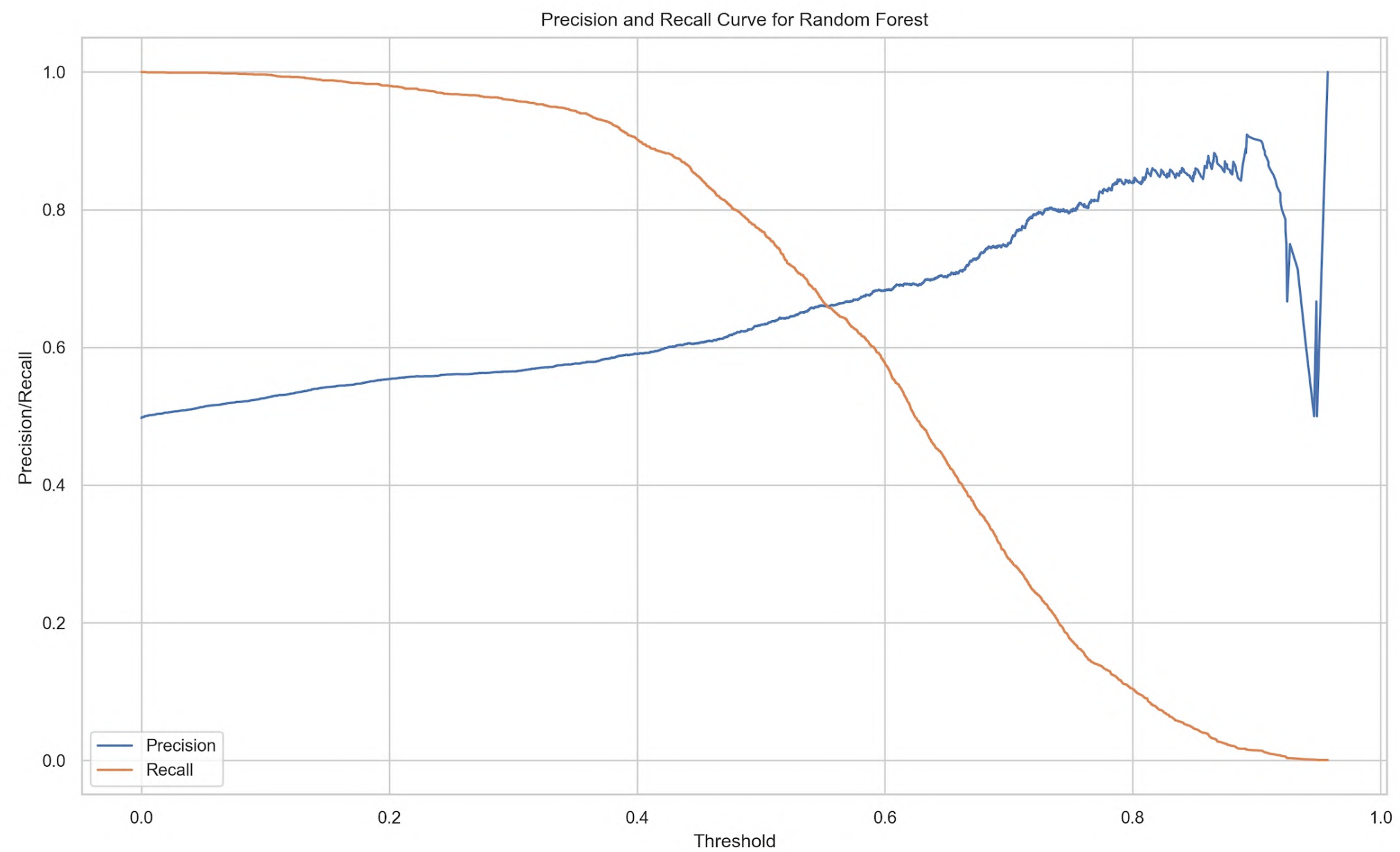
AUC Score

0.661

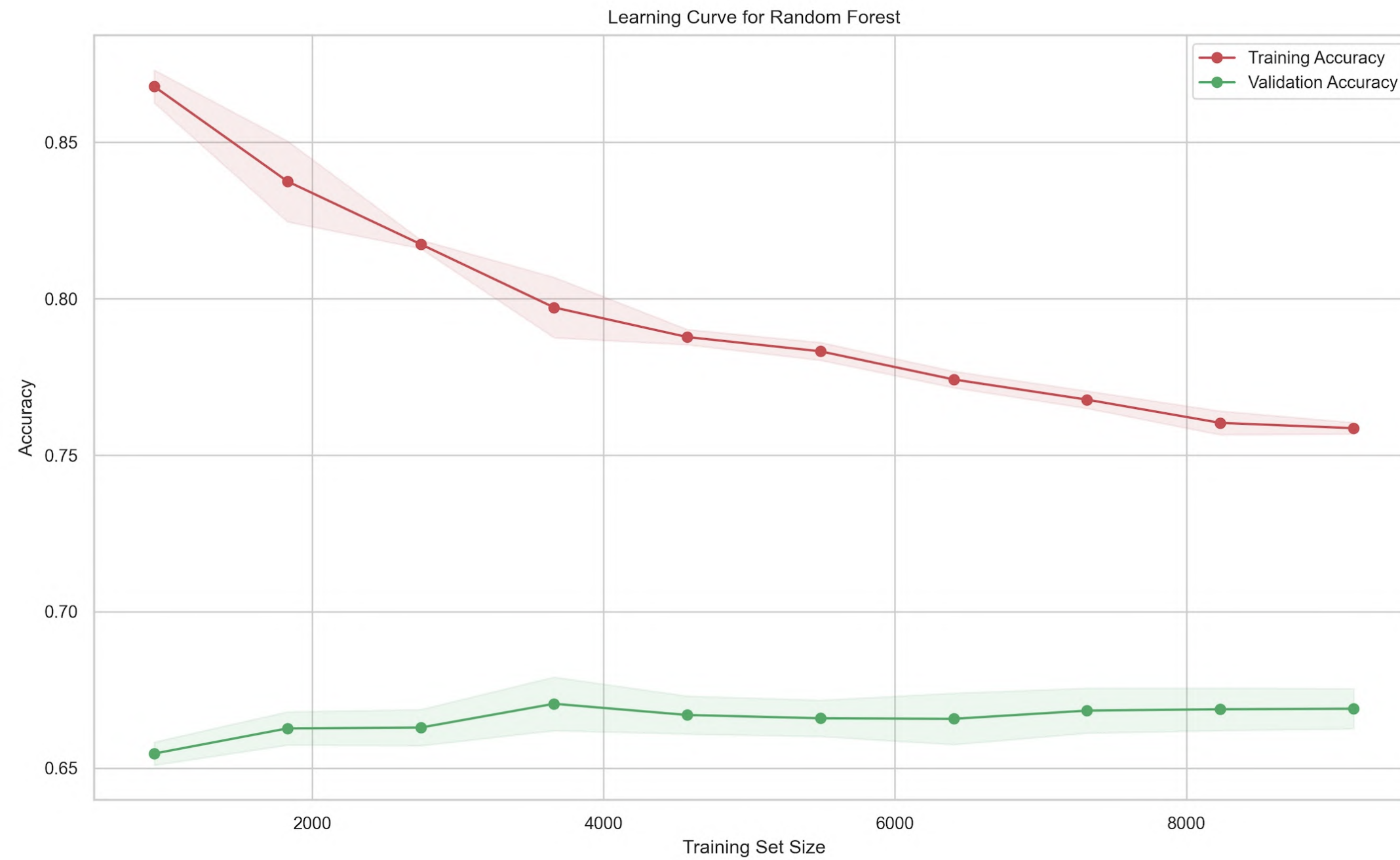
Model Evaluation



Model Evaluation

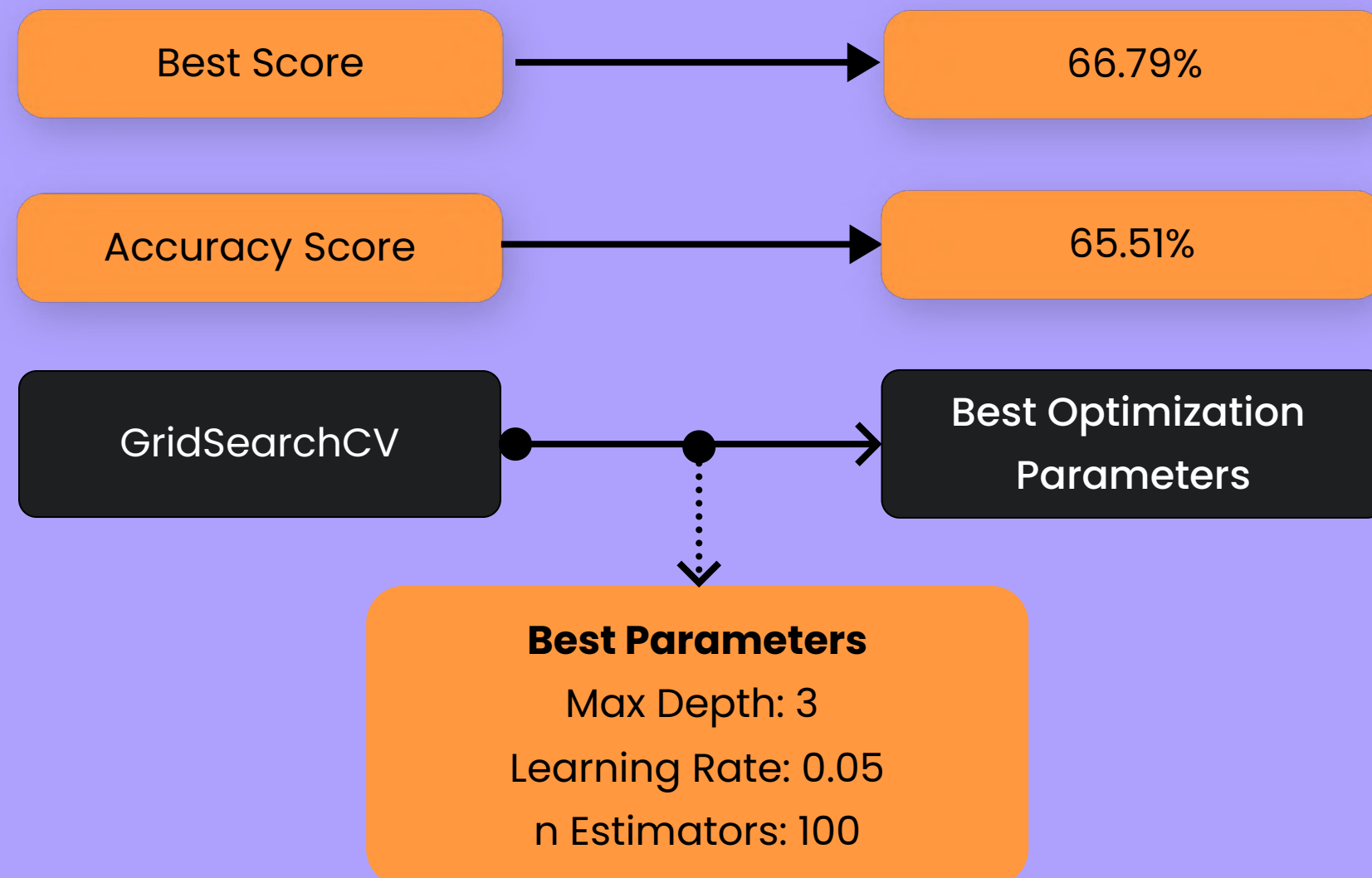


Learning Curve



9. GRADIENT BOOSTER CLASSIFIER

Model Control



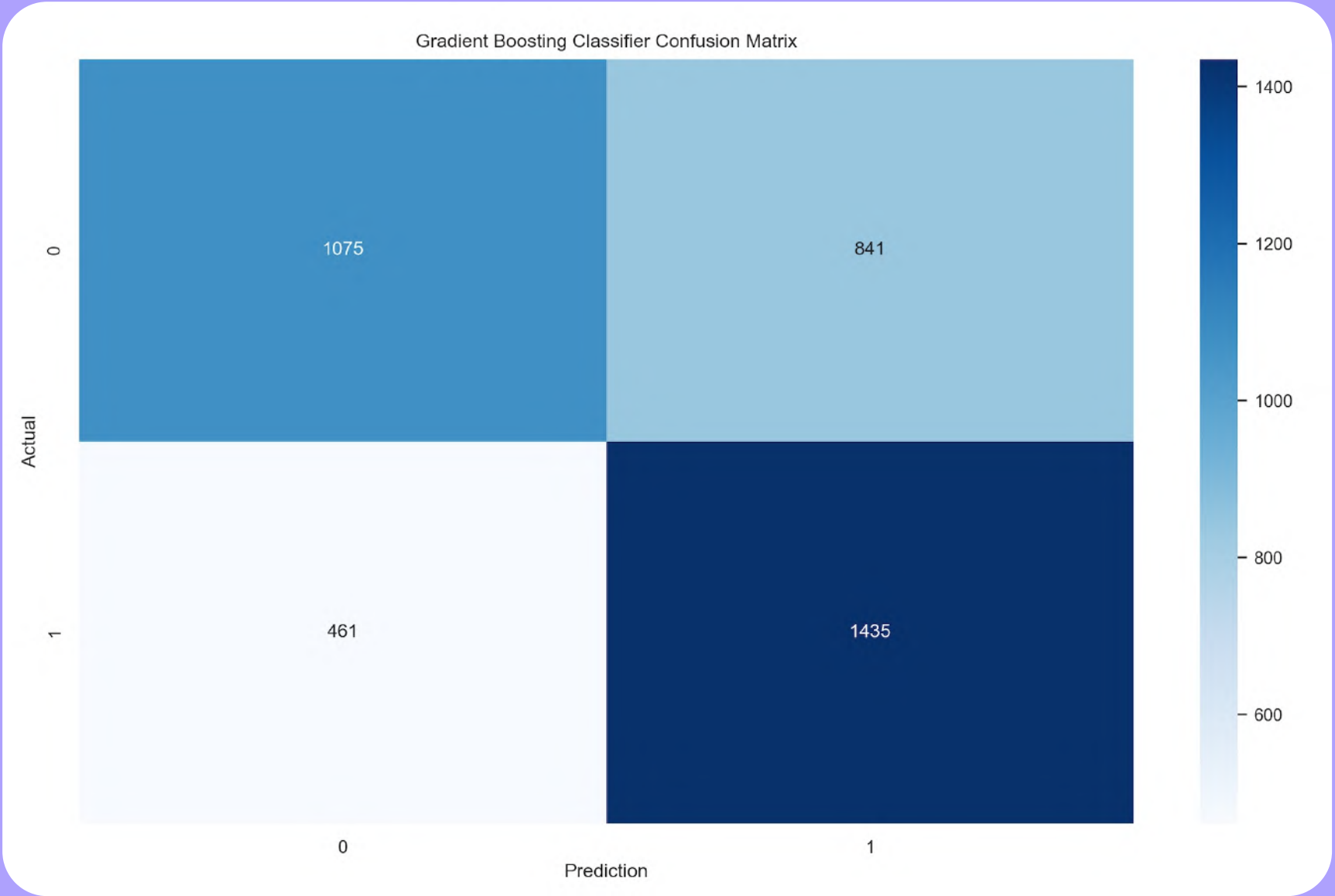
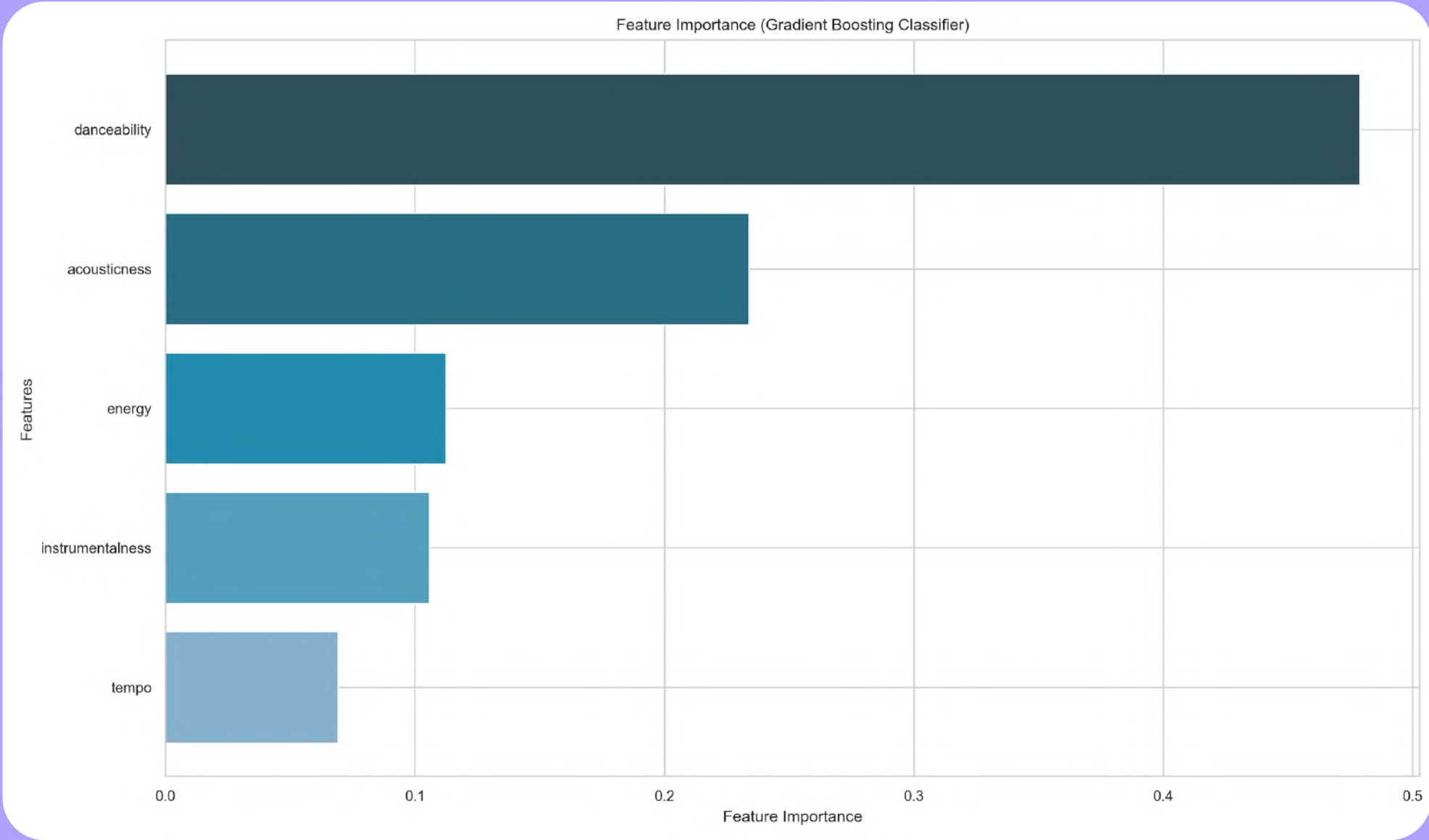
Model Evaluation

<i>Class</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Class 0	0.70	0.55	0.62
Class 1	0.63	0.76	0.69
Macro Avg	0.66	0.66	0.65
Weighted Avg	0.66	0.66	0.65

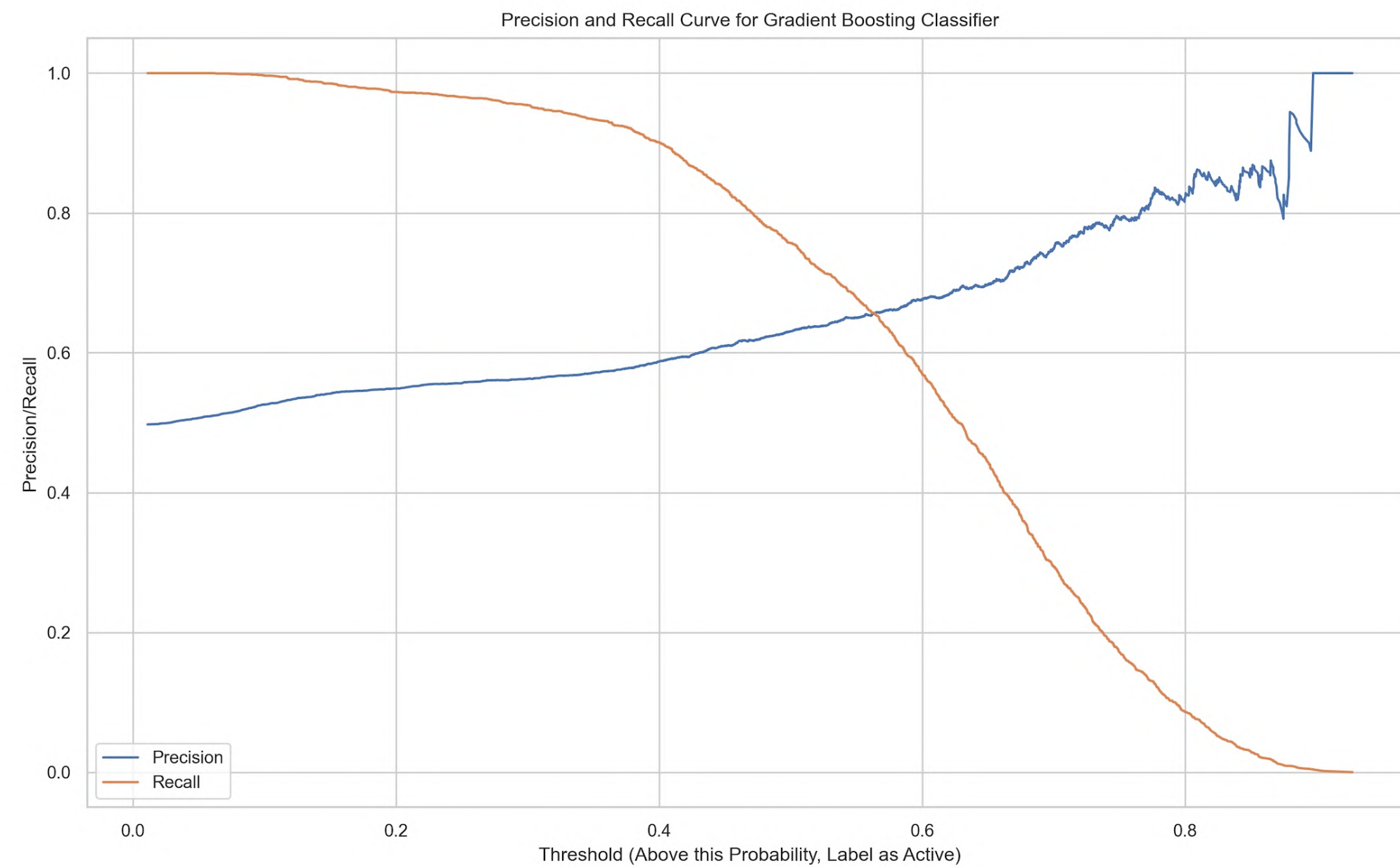
AUC Score

0.724

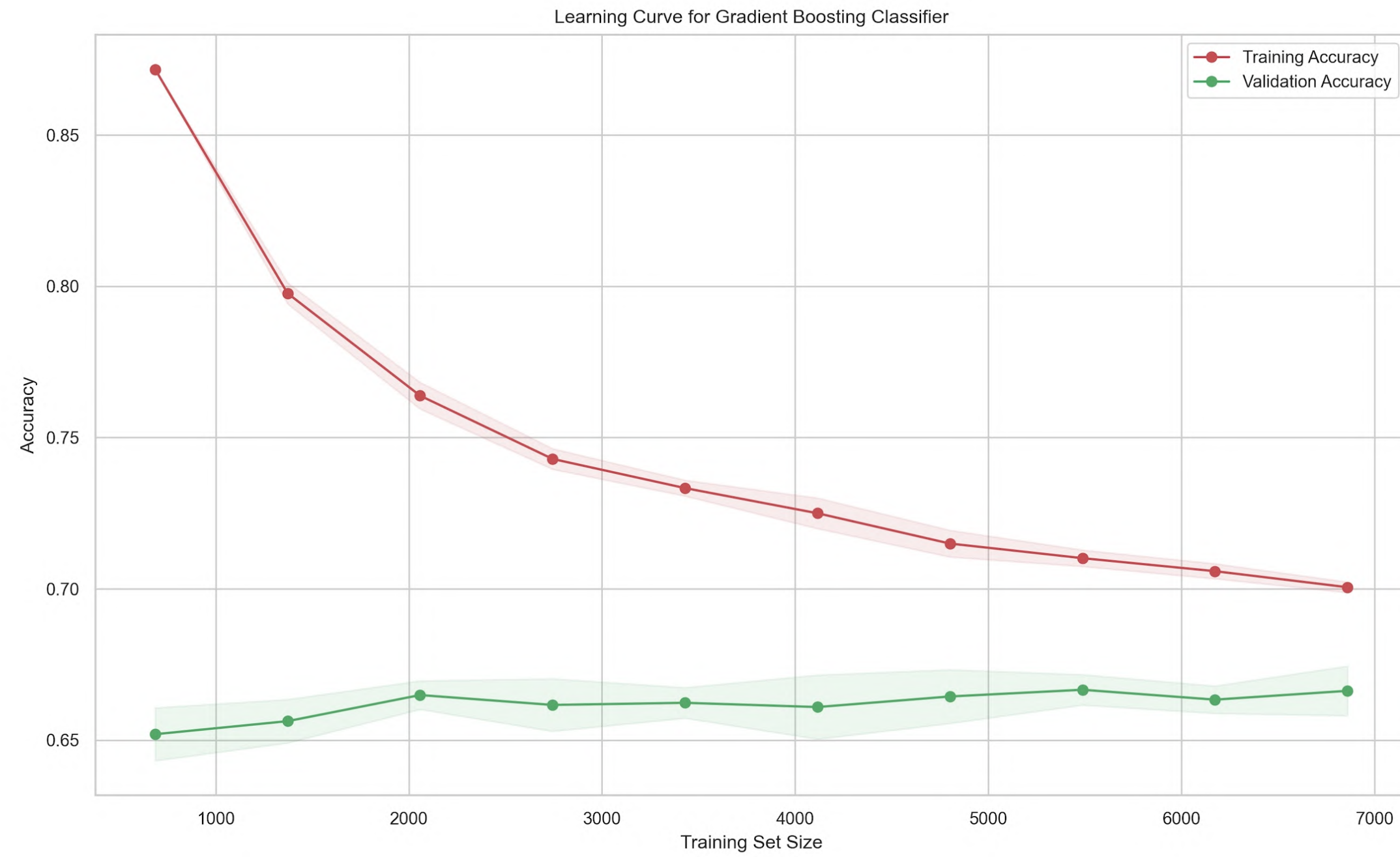
Model Evaluation



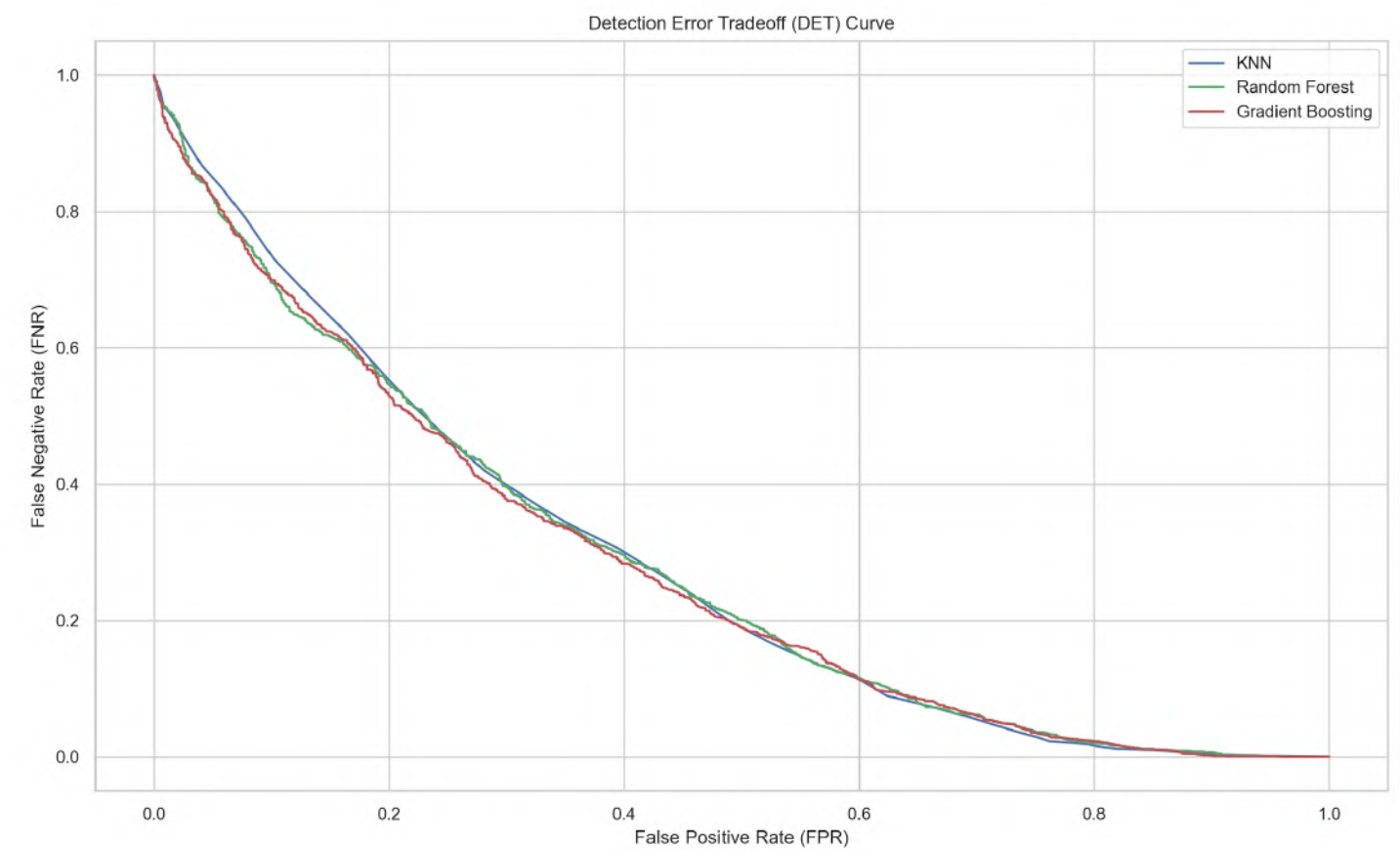
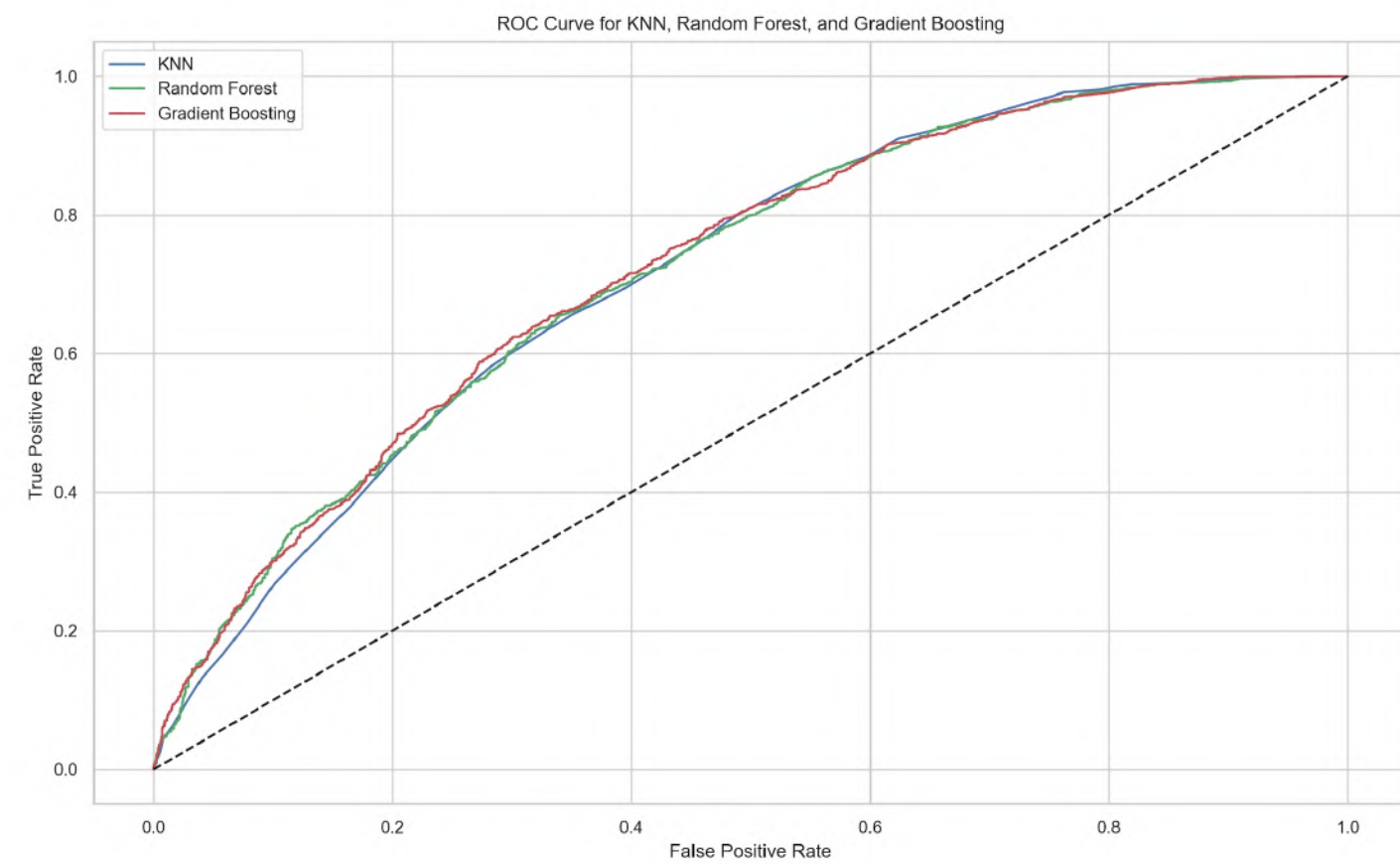
Model Evaluation



Learning Curve



ROC & DET Curve

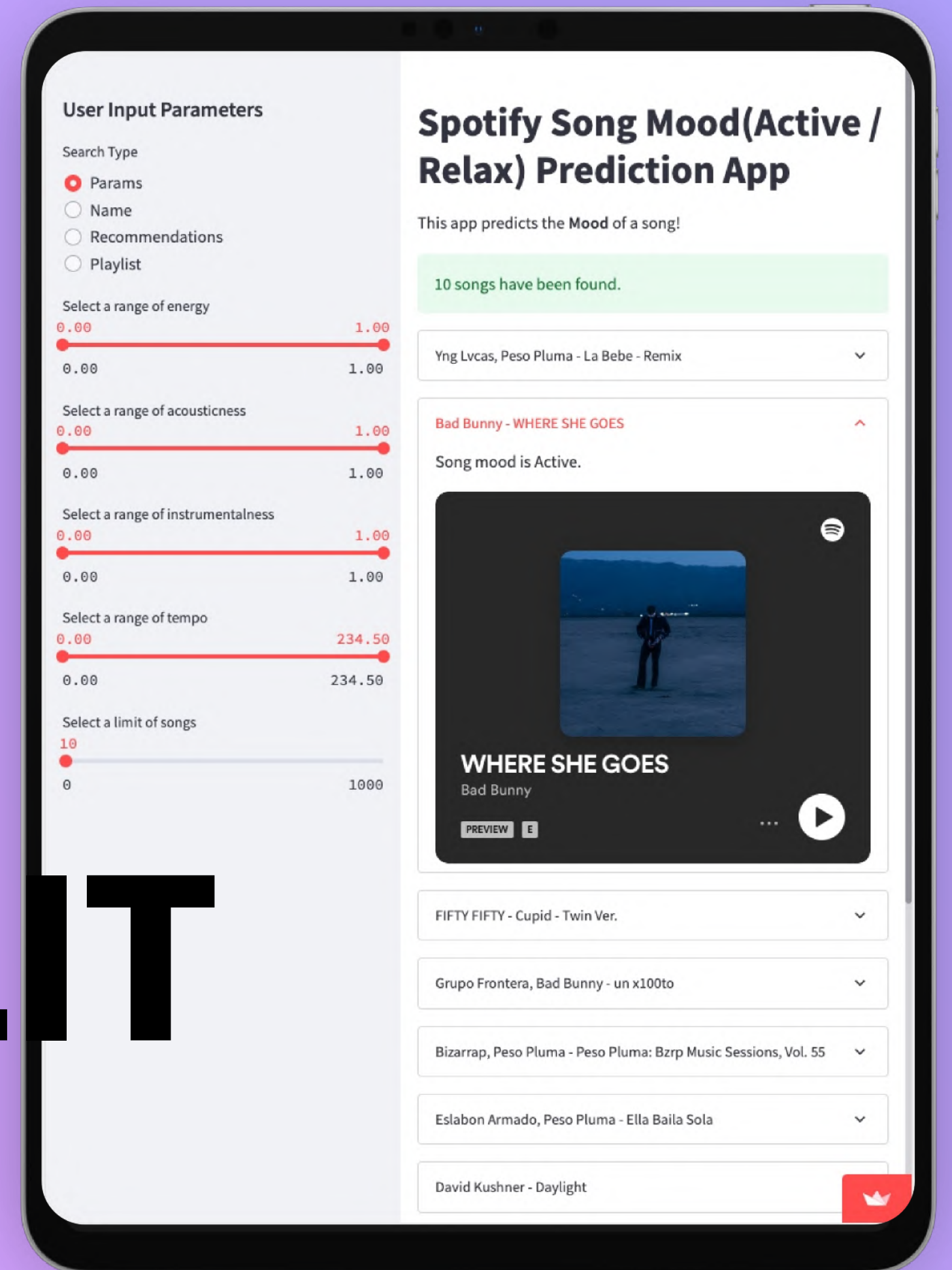


Click the link to Discover.

 Streamlit

10.STREAMLIT LIVE DEMO

Try Pitch



Song Name	Predicted	Actual
Another Love	Active	Active
Lover	Active	Active
Someone You Loved	Active	Active
Self Love	Active	Active
Lovers Rock	Active	Relax
All of the girls	Active	Relax
Girl with the tattoo	Relax	Active
When the party over	Relax	Relax
Vor Deinen Thron	Relax	Relax
Solo Piano	Active	Relax

6/10



11. CONCLUSION & FUTURE WORK

- **Spotipy** API
- **KNN** Modelling for Streamlit
- **66.24%** Prediction rate
- **Equal** distribution of target class
- **Streamlit** UI for tangible use-case.

<i>Model</i>	<i>Best Score</i>	<i>Accuracy Score</i>	<i>AUC Score</i>
KNN	66.24%	65.58%	0.656
Random Forest Classifier	66.98%	66.13%	0.661
Gradient Booster Classifier	66.79%	65.61%	0.724

Future Works:

- Spotify sign-in to sync all playlists
- Playlist classification (Active / Relax)
- Danceability Feature addition