SENTIMENT-DRIVEN SPOTIFY MUSIC RECOMMENDATION

Presented By : Md Badiuzzaman Pranto

Friedrich-Alexander-Universität

Erlangen-Nürnberg

OVERVIEW

- **Introduction**

- **Datasets**

- **Methodology**

- **ETL Pipeline**

- **Data pre-processing**

- **Recommedation engine**

- **Results and conclusion**

- **Demonstration**

INTRODUCTION

The project aims to redefine personalized music experiences by seamlessly integrating social media sentiment analysis into the realm of music curation. The approach involves training a Logistic Regression model on Twitter posts to extract sentiment, achieving a commendable 76% test accuracy. Building upon this foundation, the system identifies users' sentiments and combines them with their existing music playlists to offer truly personalized and emotionally resonant music recommendations.

Project Goals

- Sentiment Analysis Model: Train a Logistic Regression model to discern sentiment from social media posts, capturing the emotional nuances that influence music choices.

- User Playlist Integration: Leverage sentiment scores alongside existing playlist data to gain a holistic understanding of a user's musical inclinations. This project uses a dummy playlist instead of real user playlist.

- Recommendation Engine: Develop a sophisticated recommendation engine that dynamically adapts to users' changing emotions and preferences, ensuring personalized and responsive music suggestions.

MADE | Winter 2023/24

DATASETS

Sentiment 140:
- Metadata URL: https://www.kaggle.com/datasets/kazanova/sentiment140
- Data URL: https://www.kaggle.com/datasets/kazanova/sentiment140
- File Type: ZIP
- Total number of items: 1.6 millions
- Attributes used for training: target(represents the sentiment), text (represents the tweet)

Top Spotify Songs 2023:
- Metadata URL: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023
- Data URL: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023
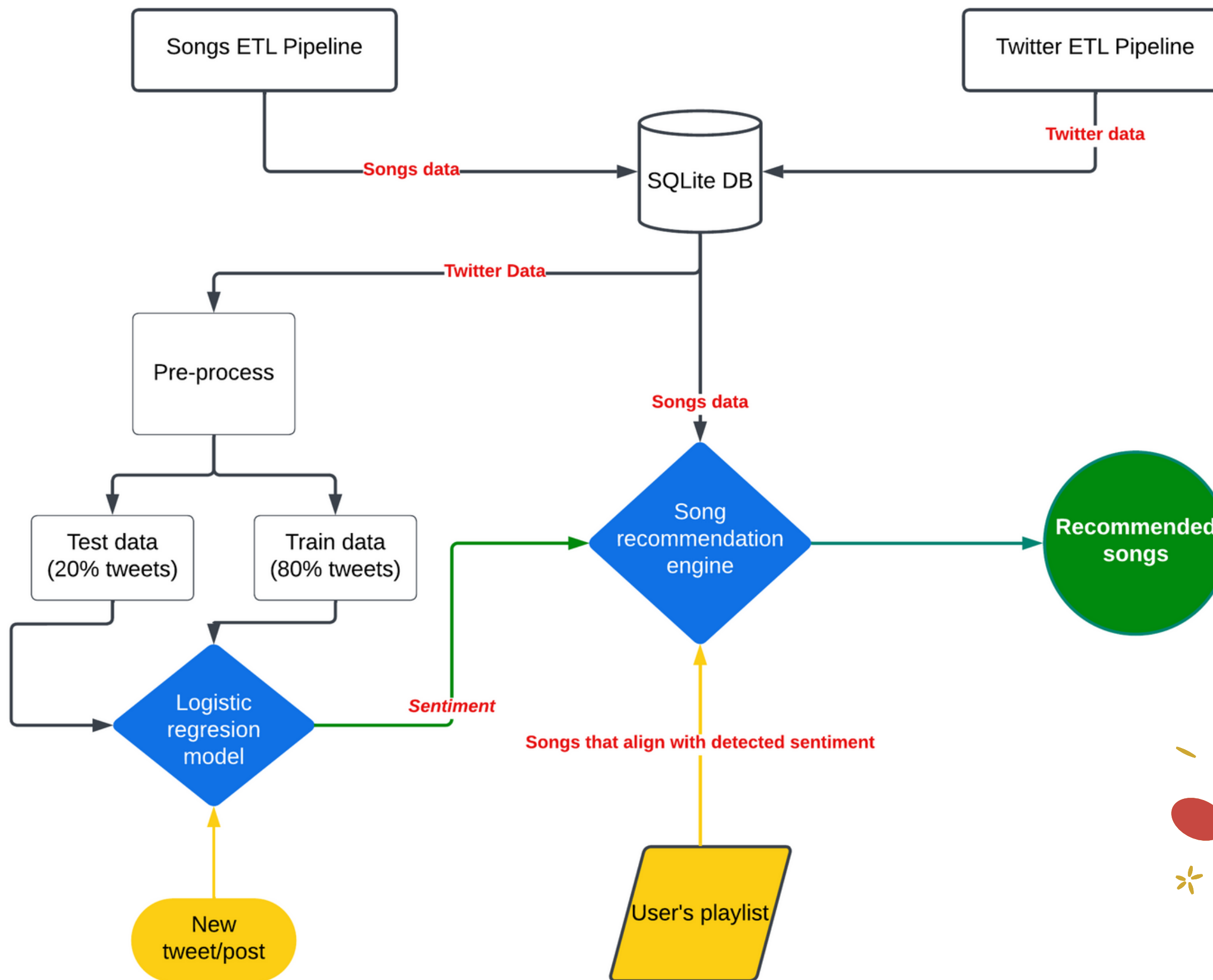- File Type: ZIP
- Total number of items: 943

User playlist:
- A user's playlist is always a subset of all available songs.
- 10% random songs of the spotify songs 2023 dataset has been used as user's playlist.
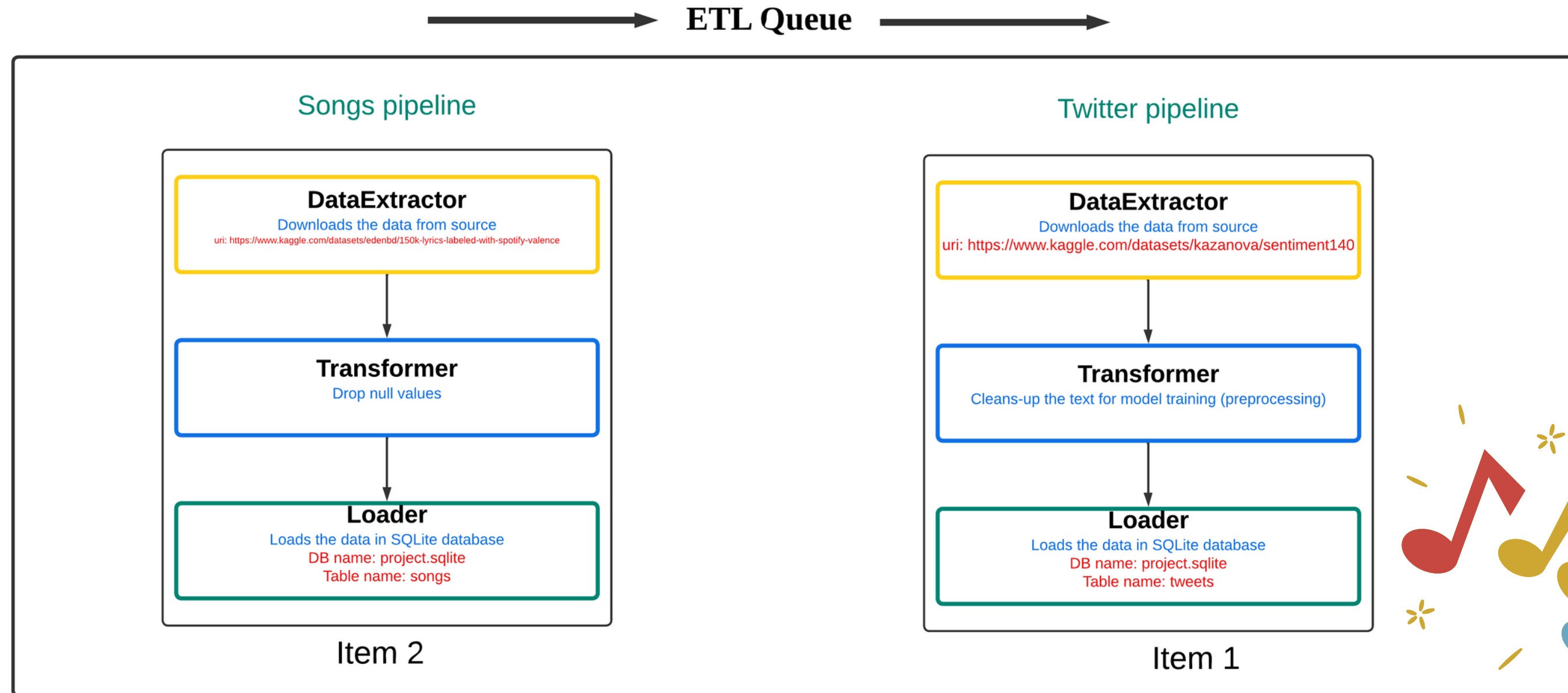- The number of songs in user's playlist is 82.
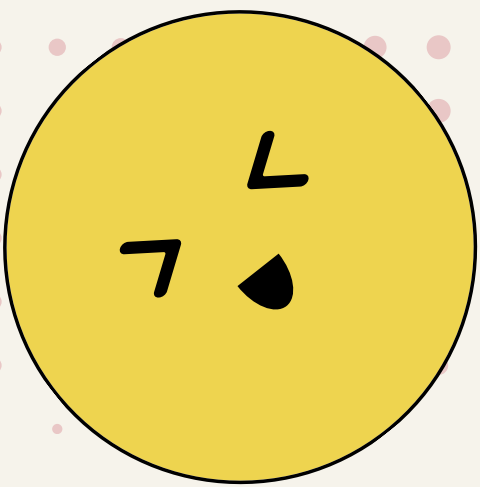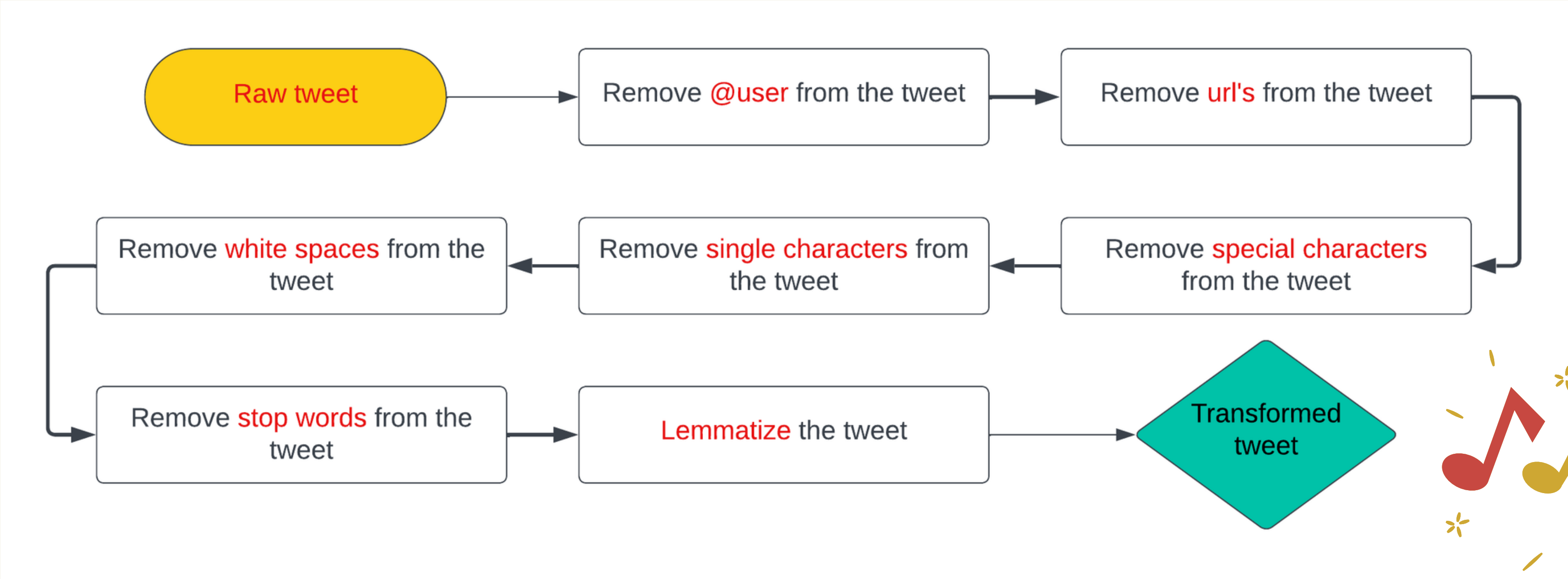
METHODOLOGY

ETL Queue

Songs pipeline

DataExtractor
Downloads the data from source
uri: https://www.kaggle.com/datasets/edenbd/150k-lyrics-labeled-with-spotify-valence

Transformer
Drop null values

Loader
Loads the data in SQLite database
DB name: project.sqlite
Table name: songs

Item 2

Twitter pipeline

DataExtractor
Downloads the data from source
uri: https://www.kaggle.com/datasets/kazanova/sentiment140

Transformer
Cleans-up the text for model training (preprocessing)

Loader
Loads the data in SQLite database
DB name: project.sqlite
Table name: tweets

Item 1

Throughout the project, a collaborative effort has been made to initiate a open-source Python package (https://github.com/prantoamt/etl-pipeline-runner) for executing ETL pipelines. Take a moment to review our contributions and share your feedback. If you like our work, please give a star to the repository.

ETL PIPELINE CONT.


```
Raw tweet  →  Remove @user from the tweet  →  Remove url's from the tweet
                                                        ↓
Remove white spaces from the  ←  Remove single characters from  ←  Remove special characters
tweet                             the tweet                          from the tweet
   ↓
Remove stop words from the  →  Lemmatize the tweet  →  Transformed tweet
tweet
```

DATA PREPROCESSING

- **The cleaned tweets underwent initial tokenization, breaking them down into individual words or phrases using the tokenization capabilities provided by nltk.**
- **Subsequently, the tokenized data was transformed into a numerical format using TF-IDF algorithm of sklearn making it suitable for input into machine learning models.**

Algorithmic workflow of song recommendation engine.

1. Get user's sentiment from the trained model.
2. Get songs from the user's playlist that align with the predicted sentiment.
3. Calculate the average values of audio features[1] based on the songs selected from the user's playlist.
4. Exclude user's playlist songs from all songs
5. Filter songs from step-4 whose valence aligns with the predicted sentiment and the average values of the audio features[1] are in between calculated values + / – 10 in step 3.
6. Order the selected songs by a random audio feature.
7. Return a random song.

[1] bpm, danceability %, energy %, acousticness %, instrumentalness %, liveness %, and speechiness % are referred as audio features.

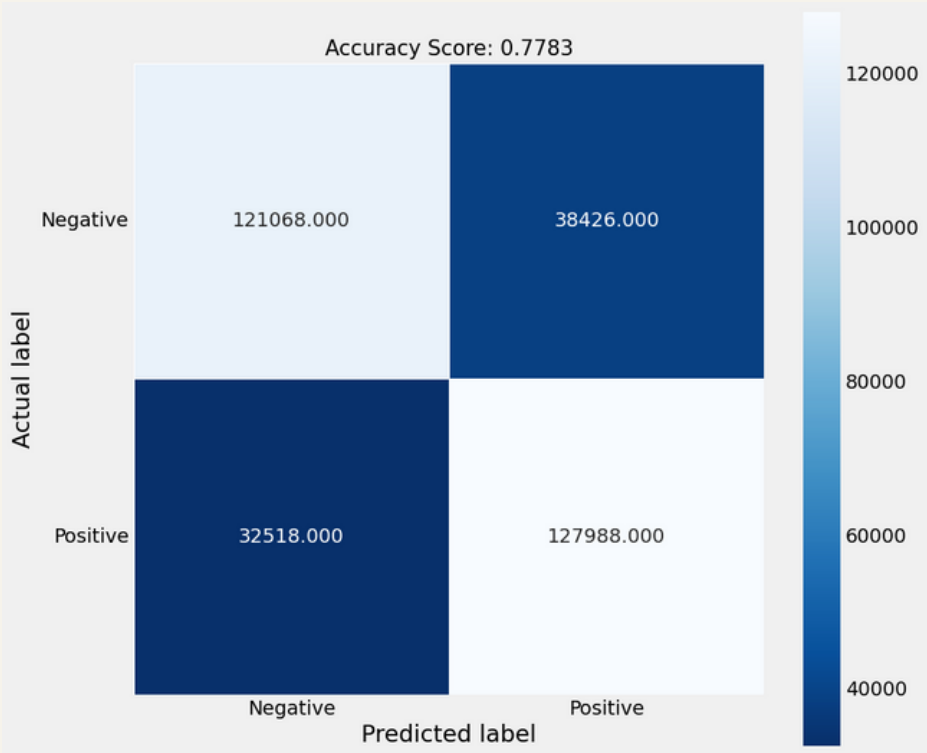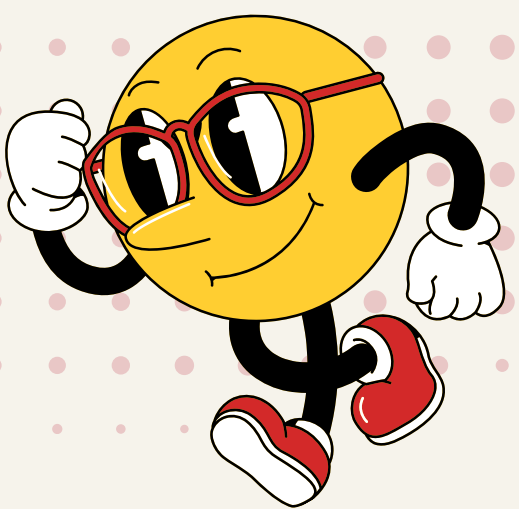A GridSearchCV search was performed to find best hyper-parameter out of:

penalty : [l1, l2, elasticnet],

solver : [lbfgs, newton−cg, sag, saga] ,

l1 ratio : [1] and,

max iter : [1000, 2000, 3000]

The best parameter found was: LogisticRegression(l1 ratio = 1, max iter = 1000, solver =′ saga′)

Sentiment	Accuracy	Precision	Recall	F1-score
Positive	0.757	0.77	0.80	0.78
Negative	0.757	0.79	0.76	0.77
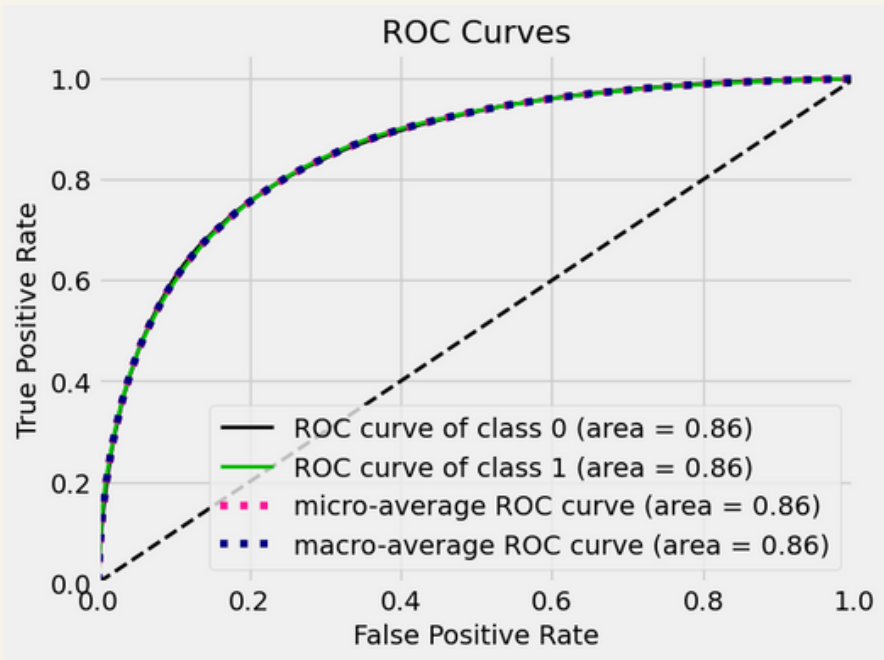
Source	Sentiment	BPM	Dancea.	Ener.	Acous.	Instru.	Live.	Speech.
Playlist	Negative	122.03	59.48	55.90	30.85	0.28	17.73	7.88
Recommendation Engine	Negative	123.51	66.79	57.58	35.62	0.09	15.33	3.98
Playlist	Positive	123.53	75.29	68.32	23.26	0.39	18.05	9.82
Recommendation Engine	Positive	123.21	73.36	67.86	30.77	0.08	15.46	5.39

Recommendation engine performance report. All the audio feature values are average values.

```
 ./project/main.sh

Make a new post in tweeter: A fresh morning with a delicious cup of coffee!

Seems like you are Happy! The song 'Don't Be Shy' by Tiï¿½ï¿½sto, Kar released in 2021 may allign with your mood.

Make a new post in tweeter: A fresh morning with a delicious cup of coffee!

Seems like you are Happy! The song 'Late Night Talking' by Harry Styles released in 2022 may allign with your mood.

Make a new post in tweeter: People are losing their right to speak

Seems like you are Sad! The song 'Still Life' by BIGBANG released in 2022 may allign with your mood.

Make a new post in tweeter: █
```

Danke schön

ধন্যবাদ

Thank You!