**Data Scientist Mini-Project**

**Manisha Pradhan**

**Genre Classification of Million Song Dataset**

**Objective:**

To provide a solution to classify tracks in different genres using dataset from Million Song Dataset.

**Dataset overview:**

Features

\* trackID: unique identifier for each song (Maps features to their labels)

\* title: title of the song. Type: text.

\* tags: A comma-separated list of tags representing the words that appeared in the lyrics of the song and are assigned by human annotators. Type: text / categorical.

\* loudness: overall loudness in dB. Type: float / continuous.

\* tempo: estimated tempo in beats per minute (BPM). Type: float / continuous.

\* time\_signature: estimated number of beats per bar. Type: integer.

\* key: key the track is in. Type: integer/ nominal.

\* mode: major or minor. Type: integer / binary.

\* duration: duration of the song in seconds. Type: float / continuous.

\* vect\_1 ... vect\_148: 148 columns containing pre-computed audio features of each song.

- These features were pre-extracted (NO TEMPORAL MEANING) from the 30 or 60 second snippets, and capture timbre, chroma, and mfcc aspects of the audio. \

- Each feature takes a continuous value. Type: float / continuous.

Labels

\* trackID: unique id for each song (Maps features to their labels)

\* genre: the genre label

1. Soul and Reggae

2. Pop

3. Punk

4. Jazz and Blues

5. Dance and Electronica

6. Folk

7. Classic Pop and Rock

8. Metal

**Steps**:

1. Data Cleaning: Handled missing values – 400 rows deleted

2. Feature Extraction: Total features – 157

- Removed highly correlated features.

- Used recursive feature elimination, and removed features.

- Final features used – 81.

3. Data Preprocessing – Used StandardScalar.

4. Feature Engineering – 2 columns – Title and Tagse.

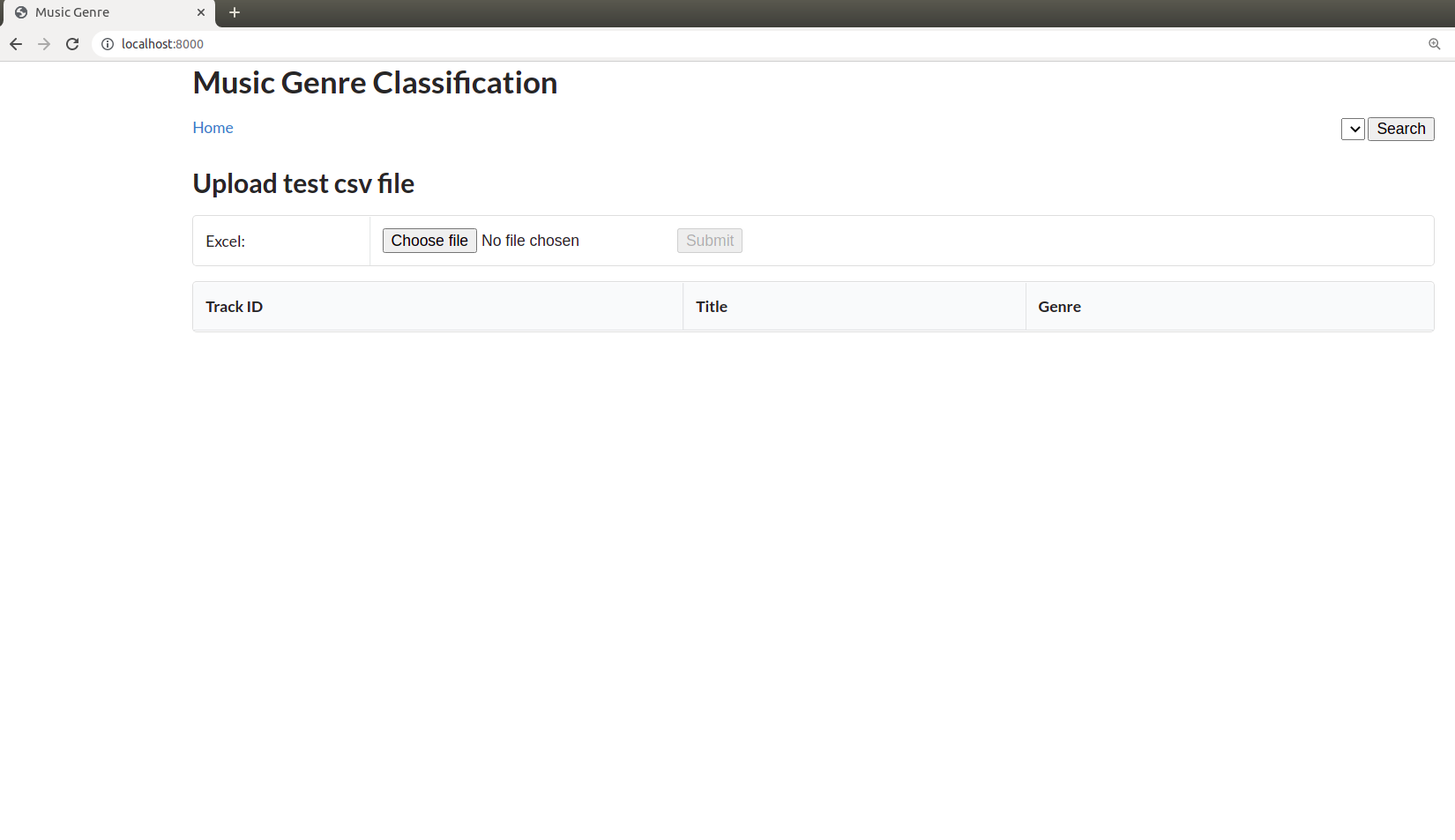
- Converted them to matrix to be used in DL

5. Tested with different CNN models.

6. Used a dropout based model to avoid overfitting.

**Page:**

**Home Page:**

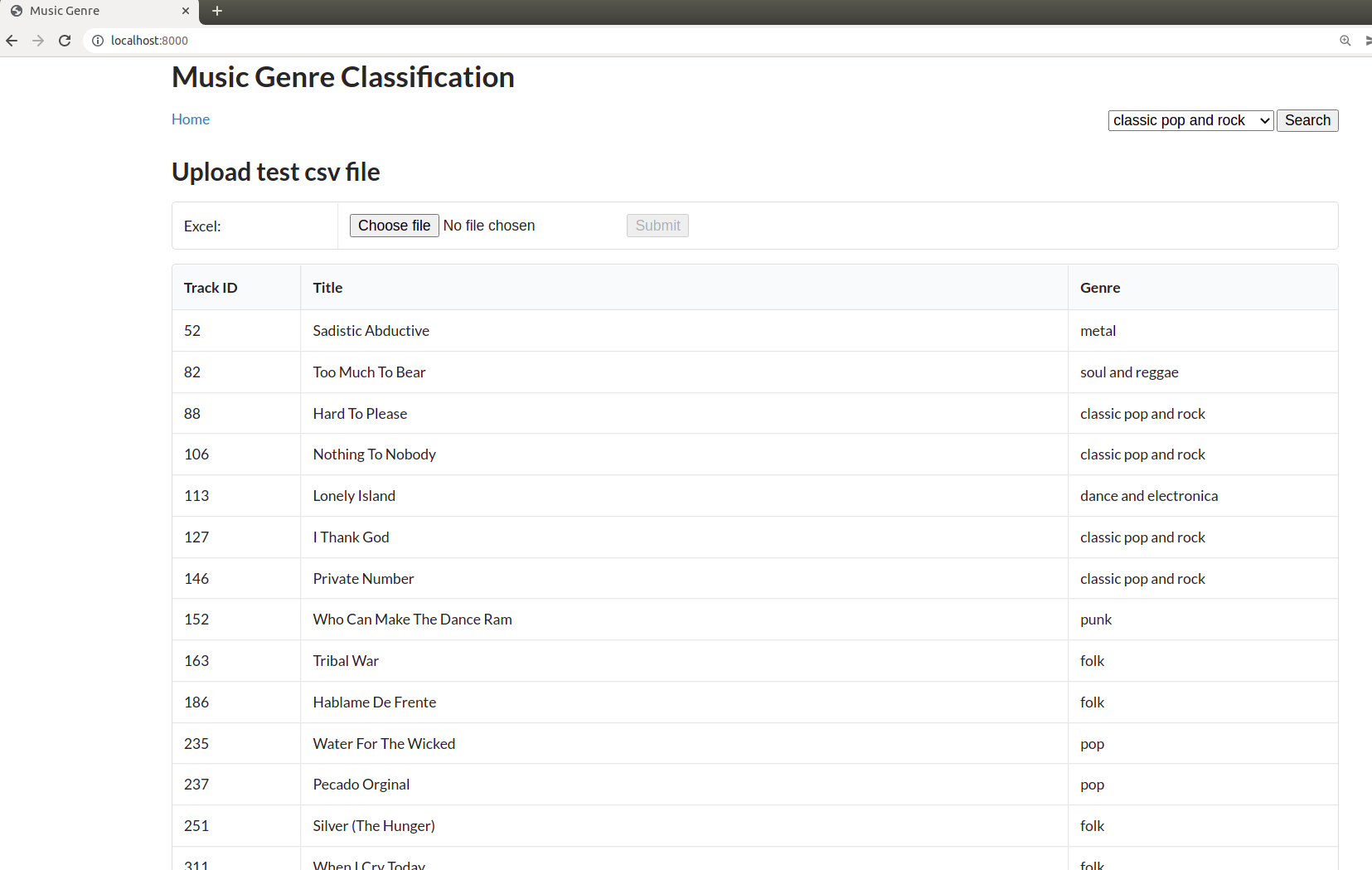


1. Upload a test.csv file for prediction and click on Submit.

2. On click of submit, the data is processed and classified.

3. Data is uploaded to database and displayed on home page.

Output: Prediction



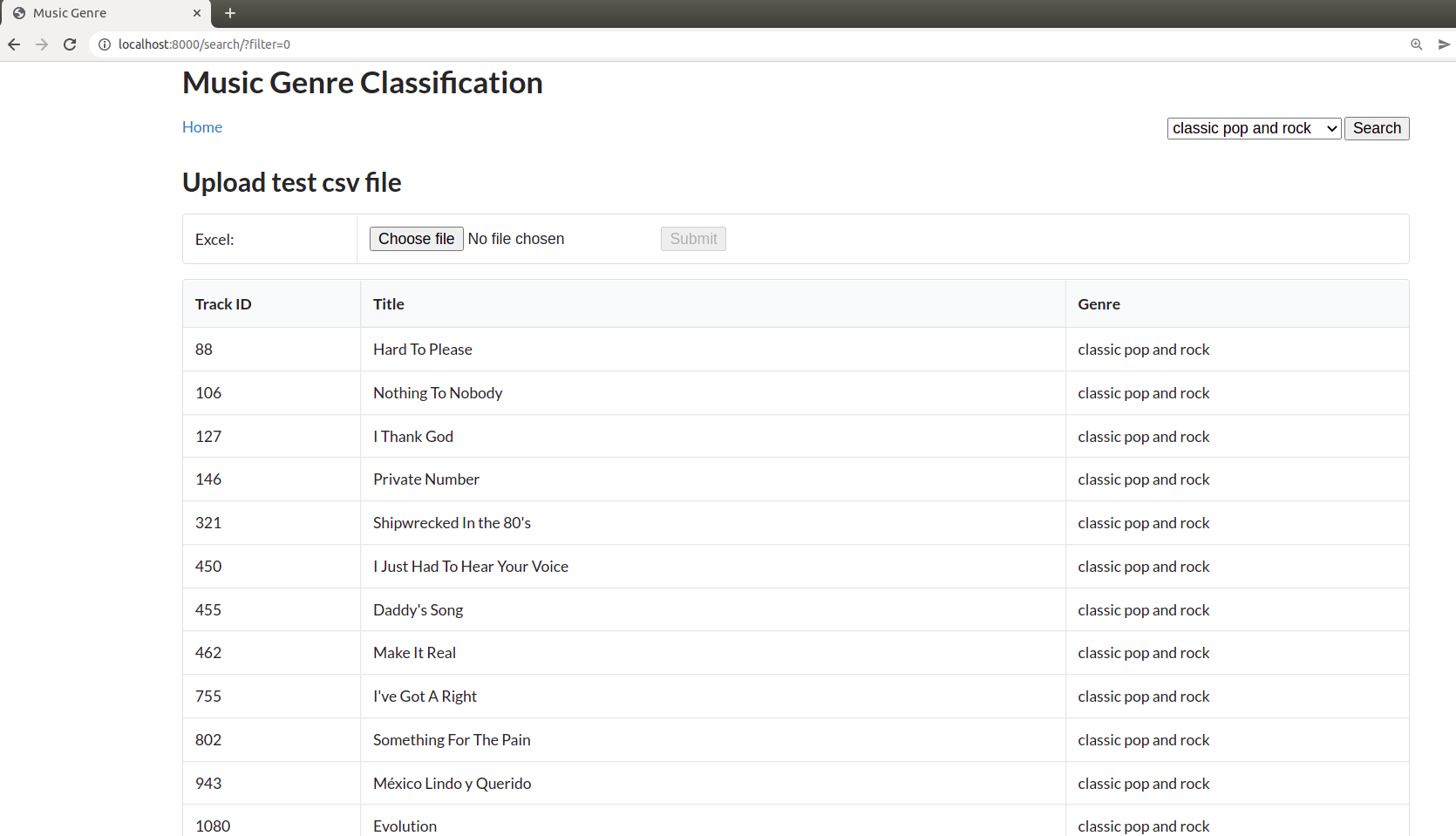
**Search**

1. Once the prediction is completed, the genres are also inserted into database.

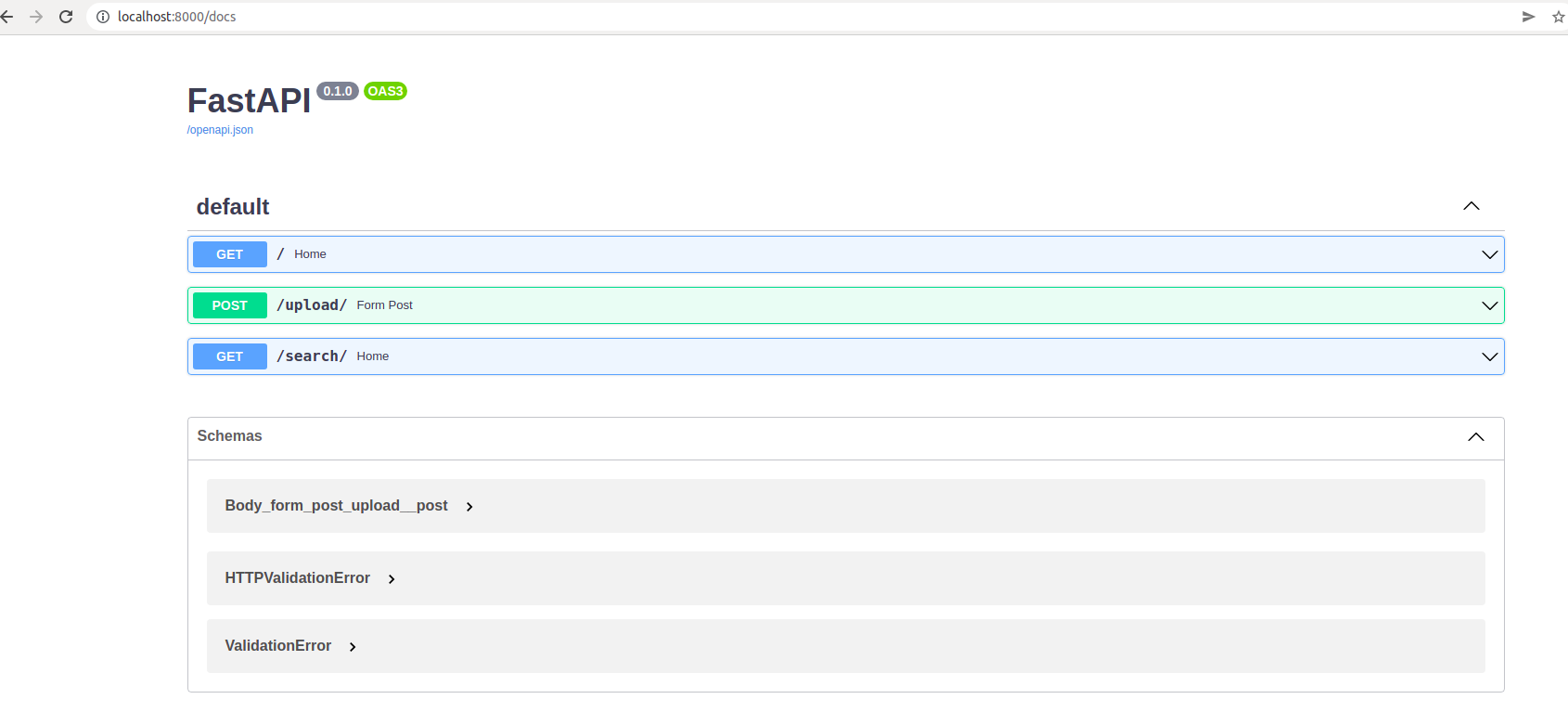
2. Genres is populated in the Search dropdown.

3. Select a genre and click on Search button.

4. All the tracks for the selected genre is displayed.



**API:**

****

**Future Works:**

1. To add few validations to the web page – Completed.

2. To add css to make UI better. For Eg: Top navigtaion bar – Completed.

3. Test with various different neural networks especially LSTM.