SUBJECT CODE : CS380 DATE: 11/03/2024 FM: 30

Question:

Design a deep learning model to develop a multi-label genre prediction system. You are given the movie's information through textual data and its multi-hot genre as a label.

Objective:
The objective is to design a multi-label genre prediction system.

Tasks:

1. Data Exploration:
The dataset is provided in the given [link]. Altogether, there are 18 distinct genres. The dataset has 1249 instances.
You may also use the already extracted embeddings using SBERT and label npy file via [link].

2. Data Preprocessing:
Divide the dataset using train_test_split in 70:10:20(train:val:test) split.

3. Model Selection:
You can develop any model and you can use pytorch or tensorflow as per your choice.

4. Model Evaluation:
The model will be evaluated with respect to classification metrics. Use sklearn to call classification_report api.

Use the below code for classification metrics calculation:

from sklearn.metrics import **classification_report**

genre_list = *["Action", "Adventure", "Animation", "Children's", "Comedy", "Crime", "Documentary", "Drama", "Fantasy", "Film-Noir", "Horror", "Musical", "Mystery", "Romance", "Sci-Fi", "Thriller", "War", "Western"]*

classification_report(y_test, y_pred, target_names=genre_list)

5. **Dashboard to put your macro_average** :

LINK TO DASHBOARD SHEET

GUIDELINES:

- Kindly observe and uphold a disciplined environment within the laboratory premises and during the exam duration.
- Utilize Google Colab for coding tasks, and feel free to leverage code snippets from your prior laboratory sessions.
- Limit your online activities to Colab only, refraining from accessing any external websites.
- Your task involves constructing a deep learning model; the utilization of conventional algorithms is not permitted.
- Thank you for your cooperation.