

Ahsanullah University of Science and Technology



Department of Computer Science and Engineering

Bachelor of Science in Computer Science and Engineering

Course No: CSE 4108

Course Title: Artificial Intelligence Lab

Dataset Documentation: Movie Recommendation System.

Date of Submission: 04/09/2022

Submitted to:

Ms. Tamanna Tabassum

Lecturer, Department of CSE, AUST.

Mr. Md. Siam Ansary

Lecturer, Department of CSE, AUST.

Submitted by,

Name: Sanjay Kumar Mandal

Student ID: 18.02.04.039

Movie Recommendation System

Documentation of Dataset

Dataset description:

There are 2 dataset used. One for movies and another one is for credits. These two datasets are from tmdb which is a database for movies. These datasets contains 5000 samples. Movies dataset contains 20 attributes and credit database contains 4 attributes. Both datasets has two attributes common which are movie_id and title. The attributes are described below:

❖ Tmdb_5000_movies:

Size:

5000 samples.

Attributes:

- Budget: Budget of movies.
- genres: Genre of movies.
- homepage: Webpage of movies.
- id: Unique id of movies.
- keywords: Word helps to search.
- original_language: Original language of movies.
- original_title: Original title of movies.
- overview: Short overview of movies.
- popularity: Popularity among fans.
- production_companies: Production companies involved.
- production_countries: Countries where filmed.
- release_date: Release date of movies.
- revenue: Revenue of of movies.
- runtime: Runtime of of movies.
- spoken_languages: Languages that are spoken in movies.
- status: released or unreleased.

- tagline: One sentence that defines the movie.
- title: Title of movies.
- vote_average: Vote of movies.
- vote_count: Vote count of movies.

❖ **Credits database:**

Size:

5000 samples.

Attributes:

- movie_id: Unique id of movies.
- Title: Title of movies.
- Cast: Cast of movies.
- Crew: Crew of movies.

Reference:

https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?select=tmdb_5000_movies.csv