



ELEMENTS OF AIML ASSIGNMENT 1

STUDENT NAME: SAKSHAM KASHYAP

SAP ID: 590014183

BATCH NO: 28 B.TECH CSE 2ND YEAR

YEAR:2024-2028

SUBMITTED TO: DR. KAMAKSHI RAUTELA

Objective:

The objective of this assignment was to create a dataset on a topic of personal interest, which would help in understanding data collection, structuring, and preparation – fundamental skills for AI/ML projects.

Topic Chosen:

Movies Dataset,

I chose the topic of movies because films are one of the most popular sources of entertainment worldwide, and analyzing movie data (ratings, genres, release years, etc.) can provide valuable insights. Movies also provide a rich dataset that can be used for AI/ML tasks like recommendation systems, sentiment analysis, and popularity prediction.

Dataset Description:

The dataset consists of multiple movies with attributes such as:

- **movie_id:** Unique identifier for each movie
- **movie_title:** Name of the movie
- **genre:** Category/genre (e.g., Action, Sci-Fi, Drama, Comedy, etc.)
- **release_year:** Year the movie was released
- **duration_minutes:** Length of the movie in minutes
- **imdb_rating:** IMDb rating (out of 10)

Data Collection Process:

- I manually curated a list of popular movies from different genres and years.
- I referred to **IMDb** and general movie knowledge to include realistic movie titles, ratings, and durations.
- I ensured a mix of **Hollywood and Bollywood** movies to keep the dataset diverse.

Challenges Faced:

- Choosing which movies to include, since the film industry is very large.

- Keeping ratings and durations realistic but not copying entire datasets directly.
- Ensuring dataset uniformity (e.g., consistent formatting of titles, years, and ratings).

Assumptions Made:

- IMDb ratings are approximated (rounded to one decimal).
- The dataset is not exhaustive but representative of different genres and years.
- Some values like duration and rating were slightly modified for uniformity.

Conclusion:

This dataset forms the foundation for applying AI/ML techniques in future labs, especially for building movie recommendation systems, trend analysis, or predictive modeling.