

# Text Classification

## (Sentimental Analysis)

### 1- Project:

Sentiment Analysis is a common NLP task that refers to extracting information from reviews. This work aims to build a movie review classifier that is able to understand the text and classify it as positive or negative review.

### 3- Company & Dataset:

Amazon.com, Inc., is an American multinational conglomerate technology company based in Seattle. It focuses on e-commerce, cloud computing, digital streaming, and artificial intelligence. It is considered one of the Big Four tech companies, along with Google, Apple, and Microsoft. Furthermore, It is the holder of the internet movie database (IMDB), which is the most interesting part in this work since it uses the IMDB dataset for the task.

IMDb is an online database of information related to films, television programs, home videos, video games, and streaming content online – including cast, production crew and personal biographies, plot summaries, trivia, ratings, and fan and critical reviews.

For this analysis we'll be using a dataset of 50,000 movie reviews taken from IMDb.

The data is split evenly with 25k reviews intended for training and 25k for testing your classifier.

Moreover, each set has 12.5k positive and 12.5k negative reviews.

IMDb lets users rate movies on a scale from 1 to 10. To label these reviews the curator of the data labeled anything with  $\leq 4$  stars as negative and anything with  $\geq 7$  stars as positive. Reviews with 5 or 6 stars were left out.

### 4- Ansatz:

The first step is to process the data, particularly loading and understanding the dataset. After that, the data must be cleaned and analysed. Then the classifier model is developed and the performance is tested.