

# Certified Specialist in Machine Learning and Artificial Intelligence

## Coding Test

### Sentiment Analysis on Movie Reviews

Movie review dataset of a popular review site is given:

The sentiment of reviews is binary, meaning the rating  $< 5$  results in a sentiment score of 0, and rating  $\geq 7$  have a sentiment score of 1. No individual movie has more than 30 reviews.

Data fields

- id - Unique ID of each review
- sentiment - Sentiment of the review; 1 for positive reviews and 0 for negative reviews
- review - Text of the review

Build a sentiment analysis model to classify movie reviews as positive or negative based on the text content. Utilize techniques such as word embeddings, word to vec, Bag of words etc.

To accomplish this, you need to follow the below steps:

- Data Preprocessing:
  - Feature Extraction:
  - Model Building and Evaluation:
  - Evaluate the model's performance
- 
- Give headings and detailed description to each step you are doing.
  - Create a repository in GitHub account as "Public".
  - Upload the notebook file (. ipynb) to the repository.
  - Please make sure that you are uploading the notebook file including the outputs as well.

- Share the link of this notebook from GitHub in the online text editor provided in Paatshala.

- [Click here](#) for the dataset.

NB: Tab separated file is given. Use sep="\t" while reading the dataset.

