

Word Embeddings and text Classification

Problem Statement:

Develop an approach to effectively train a feedforward neural network for text classification while simultaneously deriving meaningful word embeddings. The objective is to build a neural network model that not only classifies text accurately but also learns high-quality word representations, optimizing both classification performance and embedding utility.

Dataset:

MovieDataset.csv uploaded along with problem statement.

Part 1 : The goal in this part is to prepare the text data. (3 Marks)

- a. Read the text data and lowercasing and EDA
- b. Tokenization, encoding and Text vectorization
- c. Extract features and labels

Part 2 (3 marks):

Build a simple feedforward Neural Network with the below architecture:

1 Embedding layer of shape (,10)

3 Dense layers with 100 neurons and relu activation functions

1 appropriate output dense layer with appropriate activation function

Compile the model and train for 50 epochs (or accuracy score of at least 75%)

Part 3:

Extract the embedding vector from the model – 1 marks

Visualize the word embeddings in a scatter plot – 1 marks (use PCA)

Predict the class for the below given description (1 marks):

- a. "In a city of anthropomorphic animals, a rookie bunny cop and a cynical con artist fox must work together to uncover a conspiracy."
- b. "A young boy befriends a giant robot from outer space that a paranoid government agent wants to destroy."

Calculate Cosine Similarity of the below pair of words from the extracted word embeddings: (1 marks)

- a. Ancient, Historic
- b. Swift, Rapid
- c. Humble, Modest

Expected Output :

Jupyter notebook with clear mention of all the members of your team including Name and Bits ID.

There should not be any error cell, and all the cells should be executable.