

# School of Computer Science Engineering and Technology

Course- B. Tech  
Course Code- CSET346

Year- 2022  
Date: 06-09-2022

Type- Elective  
Course Name: Natural language  
processing  
Semester- odd  
Batch- ALL

## Lab Assignment 04 – Text Vectorization/Word Embedding

### CO-Mapping

Exp. No.	Name	CO1	CO2	CO3
04	Text Vectorization/Word Embedding	✓	✓	--

**Objective:** The main objective of this assignment is to know how to compare different word vectorization technique based on execution time and efficiency.

**Dataset:** Collect any data from Times of India article.

Question 1. Demonstrate the utility of N-Gram model to create word embedding for the above data.

Question 4- Use Continuous bag of word without negative sampling to create word embedding  $N=50, 100$  Where  $N$  is embedding vector dimension.

Question 5. Apply TF-IDF technique to compute the weights of words in the collected text data.