

School of Computer Science Engineering and Technology

Course- B. Tech
Course Code- CSET346

Year- 2022
Date: 12-09-2022

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

Lab Assignment 05 – Text Mining and Classification

CO-Mapping

Exp. No.	Name	CO1	CO2	CO3
05	Text Mining and Classification	✓	✓	✓

Objective: The Prime objective of this laboratory session is to make students learn to apply the concept of NLP in real life scenario objective of this assignment is to know how to convert words into numerical vectors and its application.

1. Create one hot encoding for the sentence and news category
 - a. Perform tokenization
 - b. Find number of unique tokens
 - c. Find maximum length sentence to determine number of neurons in the input node
 - d. Find number of news category to determine number of neurons in the output node
 - e. Create One hot encoding for sentence using SUM of One hot encoding on each word
2. Split your data into train, test and validation (80:10:10)
3. Create a neural network with different hyperparameters (your choice) using keras/tensorflow/pytorch
4. Train and test your network
5. Once you satisfied with the performance of the model then evaluate model with validation data.