

PG Certificate in Software Engineering for Data Science

NLP

The Objective of this assignment is to learn how to obtain and use vector representation of text using BOW and skip gram methods. Learn how to obtain Parts of Speech tagging and Named Entity Recognition in a sentence.

Programming Language: Python

Experiment 1:

CBOW and Skip Gram

Dataset: A book in the format of a txt file will be used as a dataset to obtain a bag of words and skip gram representations.

Tasks:

1. Understand how the word2vec cbow model is generated using the gensim package.
2. Train a word2vec skip gram model with gensim package and compare the performance with cbow model based on known similarity between words.

Submission: Make changes in Word2vec.ipynb file and submit the same in the portal.

Experiment 2:

Parts of Speech Tagging and Named Entity Recognition

Tasks:

1. Tag all the parts of speech for the given sentences using nltk library or library of your choice.
2. List the named entities that you want a model to predict and check if the model predicts those named entities for a sample sentence.

Submission: Make changes in POS_NER.ipynb file and submit the same in the portal.