

# Assignment-8

1. Download the file research-papers.csv from the link below:

<https://mitu.co.in/dataset>

This file contains the collection research papers received in different categories.

Perform the following operations on it:

1. Read the file
2. Separate input (summary) and output data (primary category)
3. Explore the output data
4. Prepare and vectorize the data
5. Cross validate the data
6. Build the KNN model for classification with neighbors = 6
7. Evaluate the performance of algorithm

2. Download the file sarcasm-clusters.txt from the link below:

<https://mitu.co.in/dataset>

This file contains the collection unlabeled tweets where few are sarcastic and few are not.

Perform the following operations on it:

1. Read the file
2. Clean, Prepare and vectorize the data
3. Build the K-means clustering algorithm with 2 clusters.
4. Show the tweets with 0 and 1 labels respectively.
5. Find and print sarcastic and non-sarcastic tweet's top 10 words each.