

## Project Synopsis Information Retrieval 2022

Name of Student: **Sneegdh Krishnna**

Registration No: **199302164**

1. **Project Title:** E-comm Text Classification
2. **Overview of Project (8-10 lines):**

In this project, I will classify the description of E-commerce products into 4 categories by implementing one of the NLP approaches for analysis and modelling on the dataset. The objective is to recognize whether the given description is related to Electronics (0), Households (1), Books (2), or Clothing & Accessories (3) products. My focus will be majorly on unique pre-processing techniques.

3. **Any previous work done in same field (if any, mention the detail):** The dataset was scraped from an Indian e-comm website.
4. **Dataset (it should be current dataset not 2020-2022):**
  - i. **Dataset Description:** This is the classification-based E-commerce text dataset for 4 categories - "Electronics", "Household", "Books" and "Clothing & Accessories", which almost cover 80% of any E-commerce website. The dataset is in ".csv" format with two columns - the first column is the class name and the second one is the datapoint of that class. The data point is the product and description from the e-commerce website.
  - ii. **Dataset Features:**
    - Multivariate
    - Number of Instances: 50425
    - Number of classes: 4
    - Associated Tasks: Classification
  - iii. **Dataset Link:** [E commerce text dataset | Zenodo](#)