**Experiment No.-8**

**Title:**  Mini Project / Case Study

(e.g., Breast Cancer Detection, Drug Discovery, etc.)

(A Constituent College of Somaiya Vidyavihar University)

**Batch: Roll No.: Experiment No.:4**

**Aim:** To reproduce a machine learning research work from a reputed journal article and demonstrate its practical implementation.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Resources needed:** Any programming language.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### 1. Group Formation

* Each group can have a **maximum of 2 students** and a **minimum of 1 student**. **Add group details here,** [**shared sheet**](https://docs.google.com/spreadsheets/d/1h7al9GiFlK73i9YDmEBMeXak-SpNH2N2AQ5p8PlKxBQ/edit?usp=sharing)

### 2. Paper Selection

* Select a **good-quality ML-related journal article** (2020–2025).
* Only papers from **reputed journals** (IEEE, Springer, Elsevier, Wiley, ACM, etc.) are allowed.
* The paper must be **application-oriented**, specifically focusing on **applications of ML in different domains**, such as:  
  + **Education** (student performance prediction, adaptive learning, exam analysis).
  + **Healthcare/Medical** (disease diagnosis, drug discovery, medical imaging).
  + **Agriculture** (crop yield prediction, soil analysis, pest detection).
  + **Social Good** (fake news detection, crime prediction, disaster management, poverty mapping).
  + **Other impactful real-world domains.**

### 3. Reproduction of Work

* Study the selected paper in detail.
* **Reproduce the methodology from scratch** using appropriate ML libraries/tools (Python, TensorFlow, PyTorch, Scikit-learn, etc.).

### 4. Extension to New Dataset / Domain

* After reproducing the original paper, **run the code on a different dataset**:  
  + Download a dataset from **Kaggle** (or any other open-source platform).
  + This dataset must be **different from the one used in the original paper**, or you must **apply the same implementation to a different domain**.
* This ensures you:  
  + Learn **how to preprocess datasets independently**.
  + Avoid any **copyright/ethical issues** with the original authors’ dataset.

### 5. Implementation & Documentation

* Format your code properly and structure it as per **GitHub repository standards**:  
  + Include all supporting files such as:  
    1. README.md (project description, steps to run code).
    2. Other .md files (e.g., methodology, results, references if needed).
    3. Requirements file (requirements.txt or environment.yml).
* Upload your complete implementation in the [**shared folder**](https://drive.google.com/drive/folders/1Oyv3WElzNMt-x2Tw7BhmKrJ_7u9bNJHW?usp=drive_link):  
  + Create a folder named after your **Group Number**.
  + Inside the folder, include:  
    1. All code files.
    2. Supporting .md files.
    3. Dataset link (if large, just provide URL).
    4. The selected **paper PDF** (file name format given below).

**File Naming Rule for Paper PDF:** BatchName\_GroupNumber\_TitleOfPaper.pdf

* *(Example: TYBTechML\_Group3\_BreastCancerDetection.pdf)*

### 6. Excel Sheet Update

Update the [**shared Excel sheet**](https://docs.google.com/spreadsheets/d/1h7al9GiFlK73i9YDmEBMeXak-SpNH2N2AQ5p8PlKxBQ/edit?usp=drive_link) with:

1. Group details (student names, roll numbers).
2. Title of the selected paper.
3. URL of the paper.
4. URL of the GitHub project repository (original code, if any).
5. Description of the **changes you made** (new dataset/domain, modifications in preprocessing, model adjustments, etc.).

### 

### 7. Demonstration

* Each group must **demonstrate their project before the end of the semester**.
* The demo should include:  
  + Problem statement.
  + Original dataset and new dataset/domain used.
  + Methodology and reproduction steps.
  + Results (comparison with the original paper + your new dataset results).
  + Challenges faced and learnings.

## Expected Outcomes

* Ability to reproduce and understand ML research papers.
* Skills in dataset preprocessing and adapting models to new domains.
* Experience in project structuring for GitHub repositories.
* Hands-on practice in documenting and presenting ML projects.
* Exposure to **ML applications in real-world impactful domains** like **education, healthcare, agriculture, and social good**.