**Exercise: Sentiment Analysis on Kindle Reviews**

**Objective**:  
Learn how to perform basic sentiment analysis using **TF-IDF** or **Word2Vec**.

**Instructions**

**Note:**

* **The dataset will be provided**. It contains two columns:
  + Review\_Text: Contains the review text.
  + Sentiment: Contains the sentiment label (0 for negative, 1 for positive).

**Step 1: Load the Dataset**

1. Download a dataset containing Kindle reviews with sentiment labels (0 for negative, 1 for positive).
2. Load the dataset into a pandas DataFrame.
   * The dataset should have two columns:
     + Review\_Text: Contains the review text.
     + Sentiment: Contains the sentiment label.

**Step 2: Preprocess the Text**

1. Convert the review text to **lowercase**.
2. Remove **punctuation** and **special characters**.
3. Optionally, remove **stop words**.

**Step 3: Feature Extraction**

**Option 1: TF-IDF**

1. Use TfidfVectorizer from sklearn to transform the review text into numerical features.

**Option 2: Word2Vec**

1. Train a **Word2Vec** model using gensim.

**Step 4: Train a Classifier**

1. Split the dataset into **train** and **test** sets using train\_test\_split.
2. Train a classifier using one of the following models:
   * **Logistic Regression**
   * **SVM (Support Vector Machine)**
   * **Naive Bayes**
3. Fit the classifier on the training data and predict sentiment on the test data.

**Step 5: Evaluate the Model**

1. Calculate the **accuracy**, **precision**, **recall**, and **F1-score** using appropriate functions from sklearn.metrics.
2. Compare the results of the TF-IDF model with the Word2Vec model

git init

git add -A

git commit -m "first commit"

git branch -M main

git remote add origin https://github.com/sharmistha2725/CDW-INTERN.git

git push -u origin main