BIA 5401 Group Case Study #3 (10%)

Building a Customer Feedback Intelligence System using Web Scraping & Text Mining

- This is a self-directed case study to be completed in group-base.
- Individual work or work in pairs will **not** be accepted.

Background & Objective

In the digital era, customer reviews and feedback play a vital role in shaping a company's business decisions. Businesses often struggle to analyze large volumes of textual customer feedback, especially when it is spread across various platforms.

In this group project, you are hired by a **Canadian Retail Company** to build an automated solution to:

1. **Scrape and collect customer feedback data from the web.**
2. **Organize the data into structured datasets.**
3. **Analyze and classify the customer reviews into Positive and Negative categories.**

Your solution will help the retailer improve their products, services, and customer experience using **Business Intelligence tools and techniques.**

Project Tasks

Phase 1: Data Collection & Web Scraping (50%)

1. Scrape product reviews from any e-commerce website.
2. Extract the following information:
    - Review text
    - Review rating (if available)
    - Reviewer name (optional)
3. Store the extracted data in a **NoSQL database** (e.g., MongoDB) **directly through Python code**. Manual data entry is not allowed.
4. Display all hyperlinks retrieved during the scraping process.
5. Provide a `README.pdf` explaining the scraping process, methodology, and assumptions.

Phase 2: Data Preprocessing & Feature Extraction (20%)

Once you have collected the review data:

1. Apply the following NLP techniques:
    - Sentence segmentation
    - Word tokenization
2. Store the preprocessed data in your NoSQL database **directly through Python code**. Manual data entry is not allowed.

Phase 3: Sentiment Classification using Python (20%)

1. Apply any **ready-to-use Python library or API** to classify the reviews as **Positive** or **Negative**.
   - Recommended: TextBlob, Vader, or any similar tool.
2. Store the sentiment results in the NoSQL database using Python.
3. Provide visualizations (optional) to show sentiment distribution.

Phase 4: Business Intelligence Insights & Recommendations (10%)

Prepare a **report** including:

- Percentage of positive and negative reviews.
- Most frequent words in positive and negative reviews.
- Key products with the most positive or negative feedback.
- Business recommendations based on the analysis.

Deliverables

- Python code files with clear comments
- NoSQL database file (`.json`, `.bson`, or `.csv`)
- **Report** (PDF)

Python Code connecting to MongoDB (NoSQL):

```
from pymongo import MongoClient

# Connect to MongoDB
client = MongoClient("mongodb://localhost:27017/")

# Create database and collection
db = client["customer_reviews"]
collection = db["reviews"]

# Insert one document
review = {"product": "TV", "review_text": "Excellent product", "rating": 5}
collection.insert_one(review)
```