

**TITLE: -** AI Legal Sentiment Analyzer

**NAME: -** N. Manmohan Reddy

**Roll NO: -** 23BCE9942

**College Name: -** Vellore Institute of Technology (VIT), Amaravati

**Branch: -** Computer science and Engineering

**Date of Submission: -** 23-06-2025

**Table of Content**

|  |  |  |
| --- | --- | --- |
| S.NO | Topic | Page.NO. |
| 1. | Introduction | 3 |
| 2. | Objectives | 3 |
| 3. | Tools | 3 |
| 4. | Methodology | 4 |
| 5. | Code snippets | 5-12 |
| 6. | GitHub Link | 13 |
| 7. | Conclusion | 13 |

**1. INTRODUCTION**

This project is an AI-based tool that analyzes the sentiment of legal documents.It tells whether the text is positive, negative, or neutral.The tool also gives a short summary of the document to help understand it quickly.  
It’s designed to save time for legal teams by reducing manual document review.

**2. OBJECTIVES**

To build an AI system that automatically detects sentiment in legal texts.It also provides summaries to help users quickly grasp the overall meaning of the document.

**3. TOOLS**

|  |  |
| --- | --- |
| Category | Tools/Technologies |
| Programming Language | Python |
| Libraries | pandas, sklearn, collections, countvectorizer |
| AI Model | IBM Watson FLAN-T5-XXL |
| Platform | IBM Watson Machine Learning |
| Cloud Storage | IBM Cloud Object Storage (COS) |
| IDEs | Google Colab |

**4. Methodology:**

**1. Data Collection:**  
Legal documents or case summaries are collected and stored in IBM Cloud Object Storage (COS) in CSV format.

**2. Data Preprocessing:**

* Dataset is loaded into the project.
* Sentiment labels are mapped to positive, neutral, negative.

**3. Model Setup:**  
IBM Watson’s FLAN-T5-XXL model is used as the foundation model for generating sentiment predictions.

**4. Few-Shot Learning:**  
Example sentences with known sentiments are provided to guide the model (few-shot prompting).

**5. Prediction Generation:**  
Sentiment is predicted for each sentence or document using zero-shot or few-shot prompts.

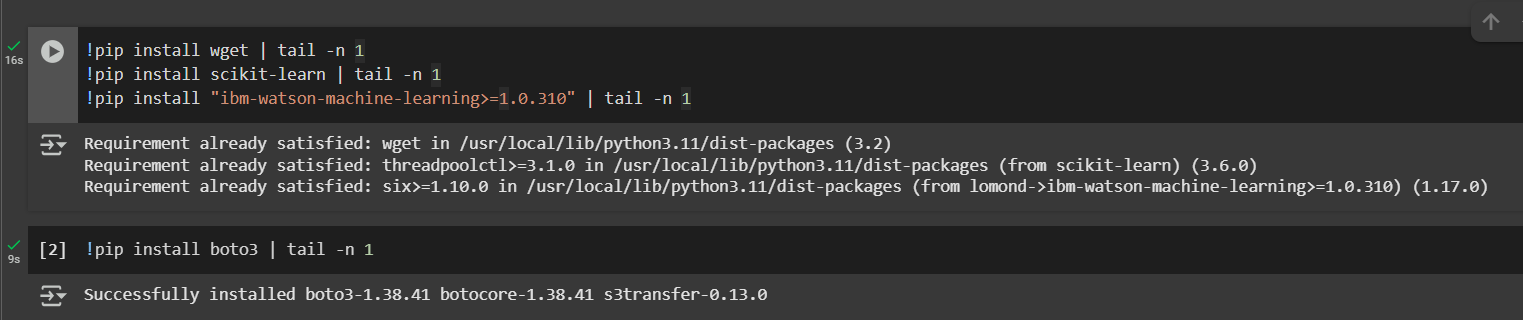
**6. Summarized Insights:**

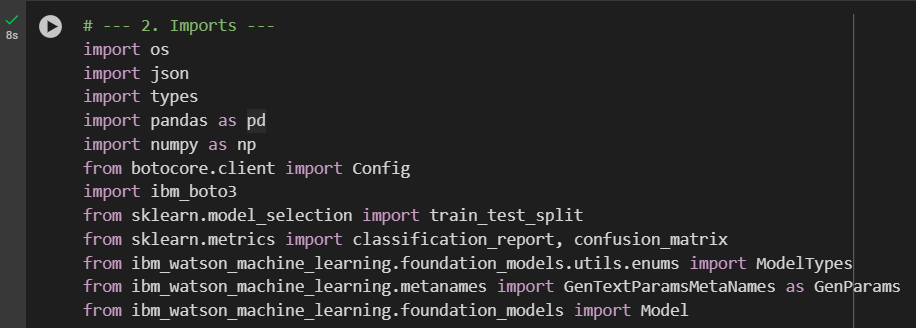
* Sentiment distribution (% of positive/neutral/negative).
* Frequent negative terms extracted using NLP techniques.

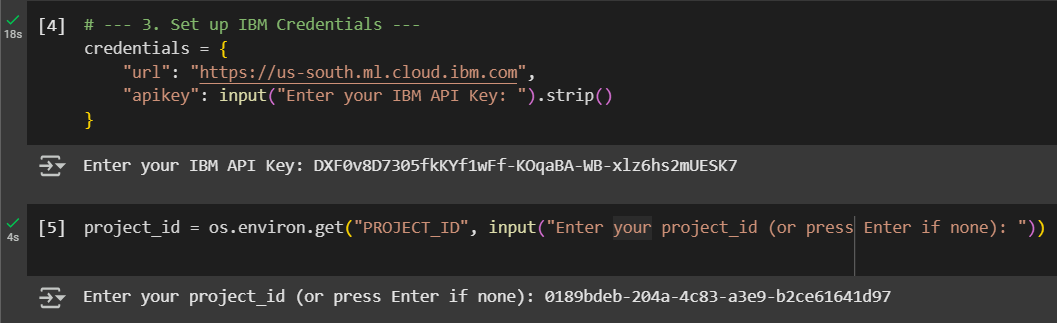
**7. Output Generation:**  
Results are saved into a CSV file for reference.  
Manual input analysis is also supported interactively via the console.

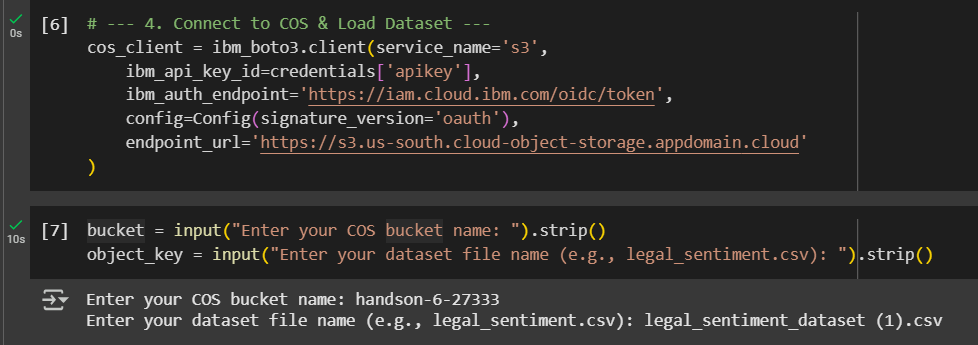
**5. Code Snippets**

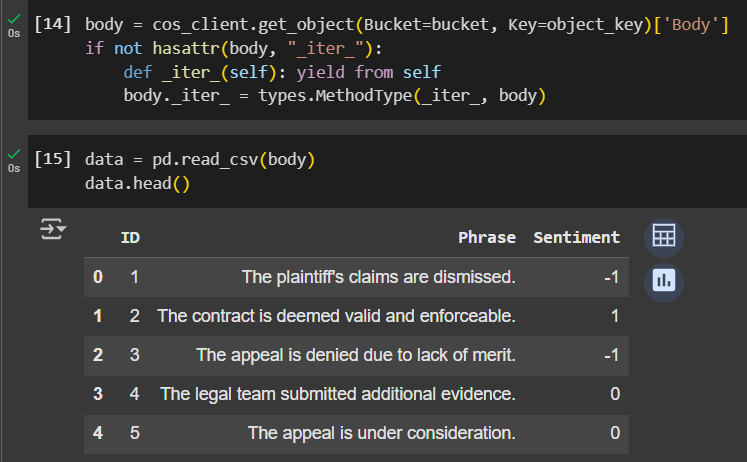
1. The required libraries for data handling, machine learning, and IBM Watson integration were successfully installed.  
These packages enable connecting to IBM Cloud, running the foundation model, and processing legal text for sentiment analysis.

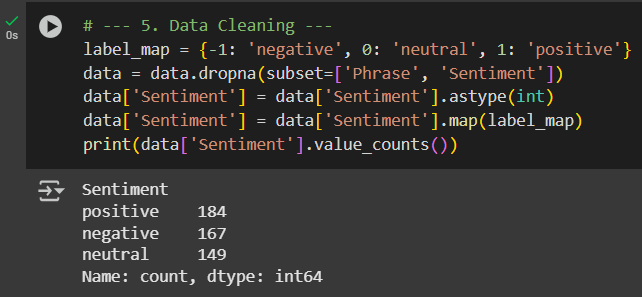
****

2. This section imports all necessary libraries for data handling, model usage, evaluation, and IBM Watson integration to build and evaluate the legal sentiment analyzer.****

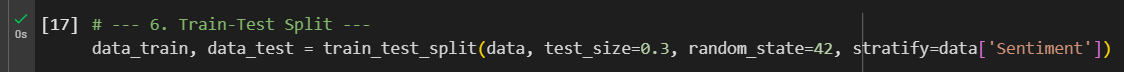
3. This section sets up the **IBM Cloud credentials** (API key and project ID) required to authenticate and connect to **IBM Watson Machine Learning** services for model execution.

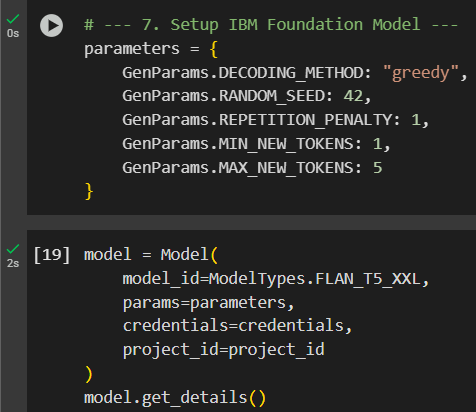
4. This section establishes a connection to **IBM Cloud Object Storage (COS)** using the provided API key. It authenticates with IBM's IAM service and specifies the endpoint for the **US-South** region. The user inputs the **COS bucket name** and **dataset file name** to locate the dataset. This enables the program to access the CSV file needed for legal sentiment analysis.

5. This code fetches the dataset file from IBM COS, reads it into a **Pandas DataFrame**, and displays the first few rows. The dataset contains **legal sentences** along with their corresponding **sentiment labels**.

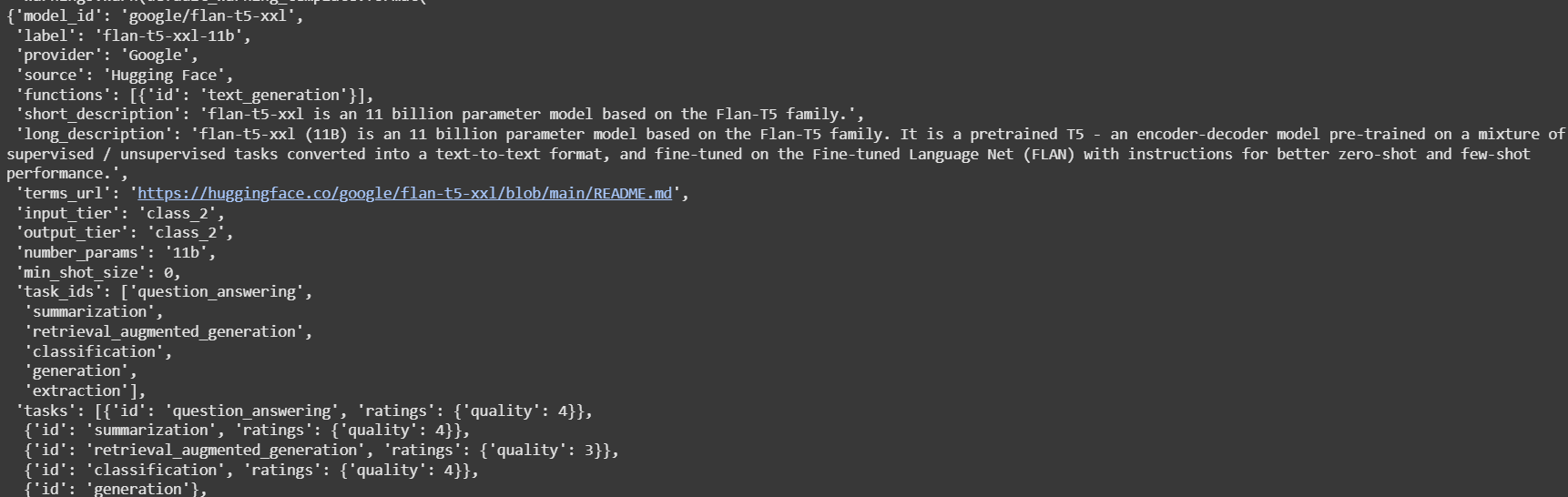
6. This section cleans the dataset by removing missing entries and mapping sentiment labels (-1, 0, 1) to their respective **negative**, **neutral**, and **positive** text labels. It then displays the count of each sentiment category in the dataset.

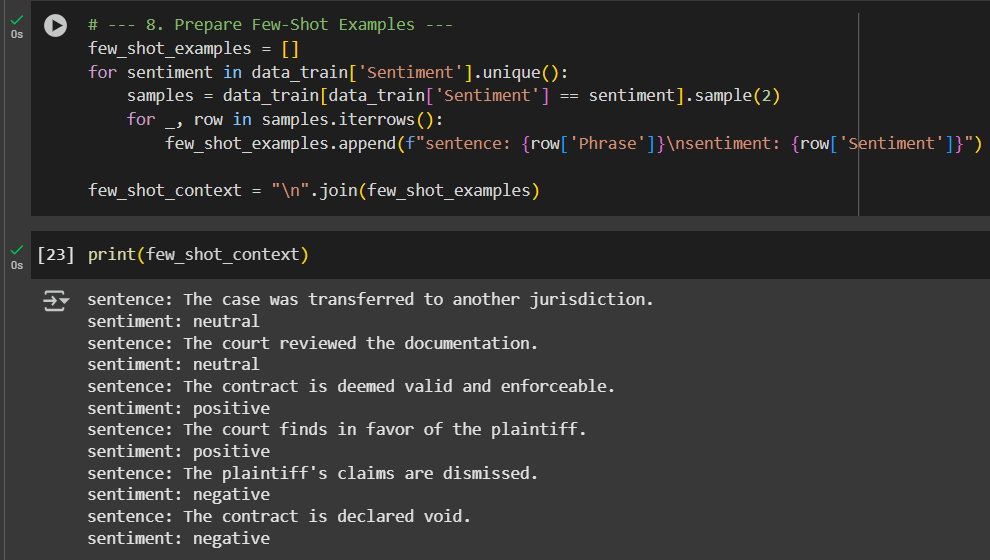
7. Splitting the Dataset for training and testing.

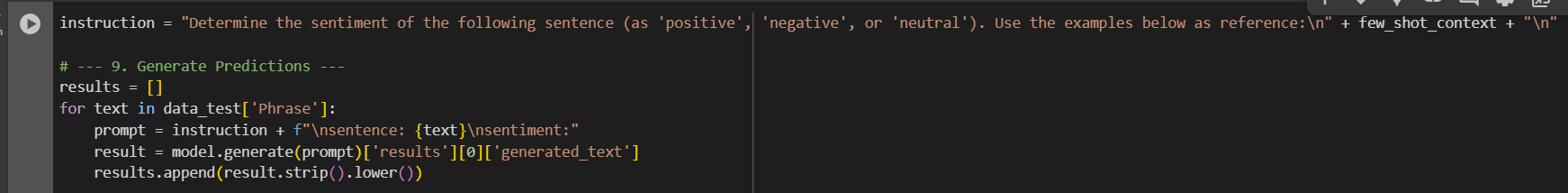


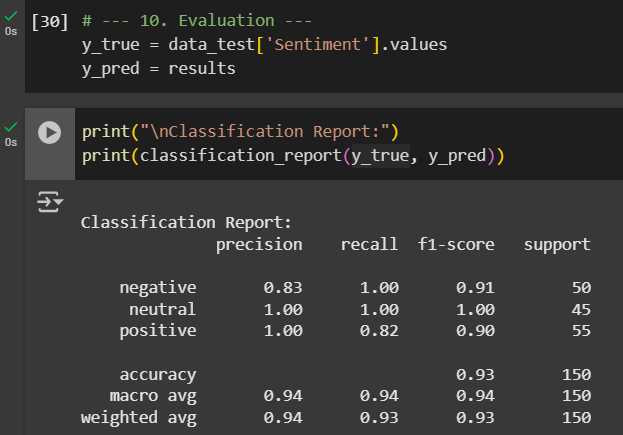
8. This section initializes the **IBM FLAN-T5-XXL** foundation model with specified parameters for text generation. It connects the model to the project using the provided credentials and project ID for performing sentiment analysis.

Showing all the details of the model FLAN\_T5\_XXL

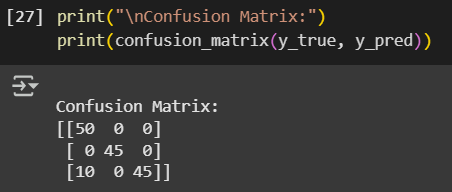


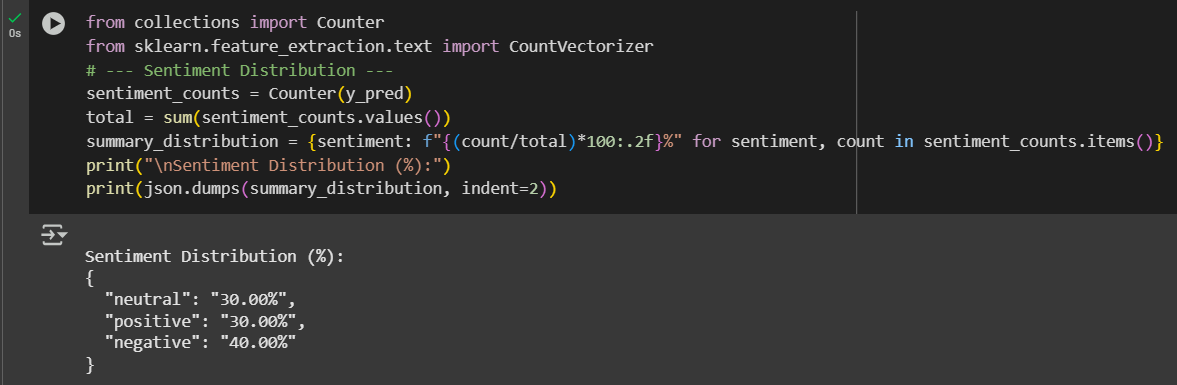
9. This part prepares **few-shot examples** by selecting sample sentences for each sentiment category from the training data. These examples help guide the foundation model during prediction for better accuracy.

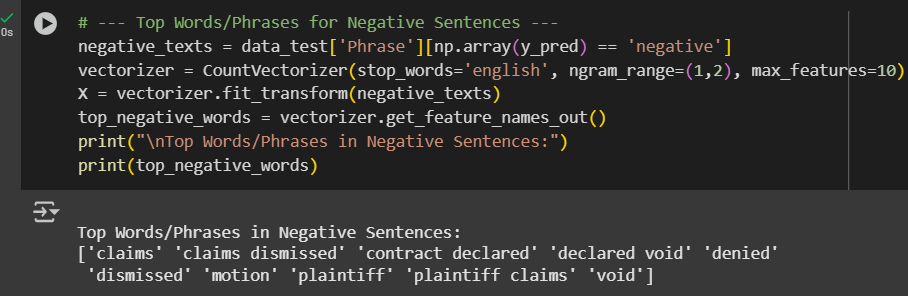
10. This block generates sentiment predictions by giving the model an instruction along with each input phrase. The model outputs 'positive', 'negative', or 'neutral', which is then cleaned and stored for evaluation.

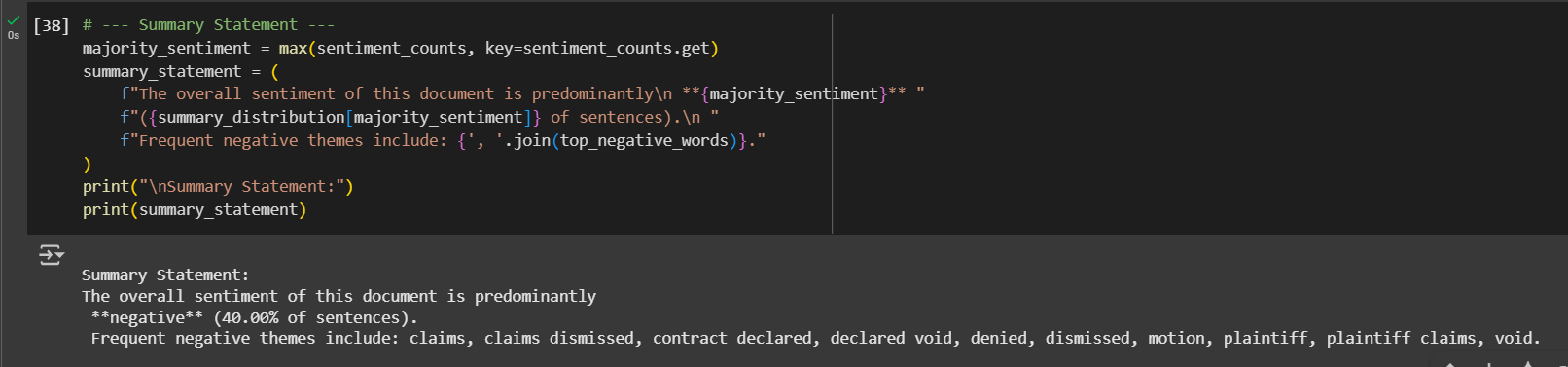
11. This code compares the predicted sentiments (y\_pred) with the actual labels (y\_true) and prints a **classification report** using classification\_report() from scikit-learn. The classification report shows **precision, recall, F1-score, and support** for each class, helping evaluate how well the model predicts each sentiment category. The model achieved **high accuracy (93%)**, perfect precision on neutral/positive, but slightly lower recall for positive class — suggesting some positive examples were misclassified.

12. This code generates a **confusion matrix** to show how many samples were correctly or incorrectly predicted for each sentiment class.From the output, the model perfectly predicted all *negative* and *neutral* samples.

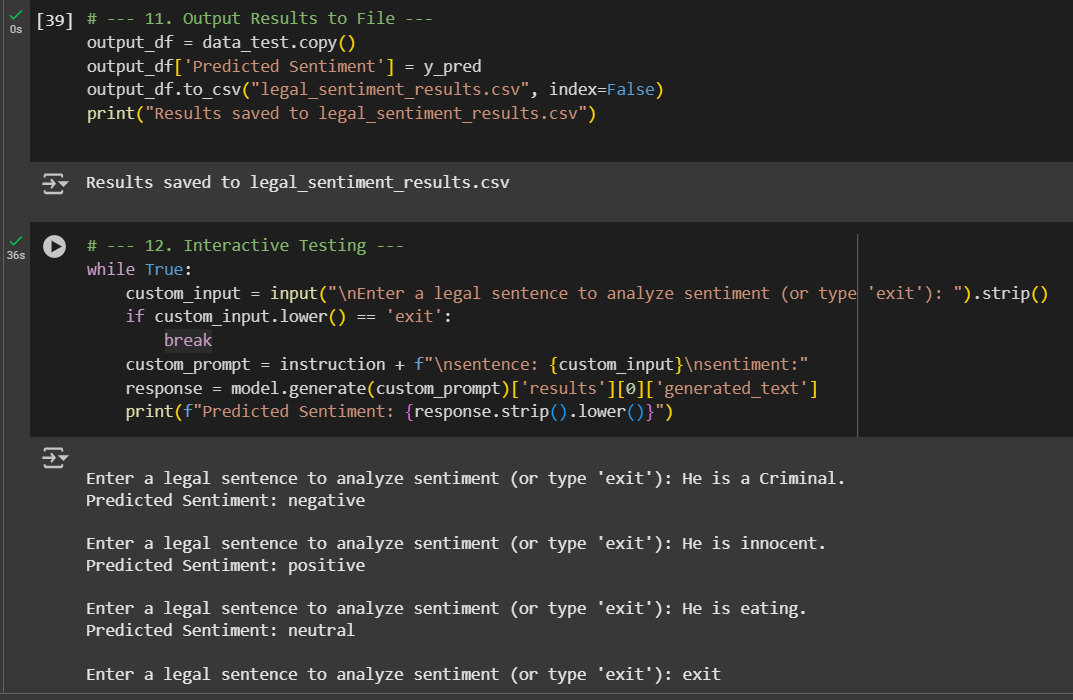


13. This code calculates the **percentage distribution** of each predicted sentiment by counting how many times each label appears in y\_pred.The output shows that 40% of the predicted sentences were negative, while neutral and positive sentiments each made up 30%

14. This code uses CountVectorizer to extract the **top 10 most frequent unigrams and bigrams** from the predicted negative sentences.The result shows common legal terms like *“claims dismissed”*, *“void”*, and *“denied”*, which frequently appear in negative sentiment phrases.

15. This code generates a **summary statement** that identifies the most common sentiment in the document and its percentage.It also lists the **frequent negative phrases**, giving a concise overview of the dominant tone and legal themes in the analyzed text.

16. Output Saved to the file and manual testing is done.



**6. GitHub link**

**7. Conclusion**

The AI Legal Sentiment Analyzer successfully automates the task of identifying sentiment in legal documents.It classifies text into positive, negative, or neutral, reducing the need for manual review by legal teams.The tool also generates summarized insights, including sentiment distribution and key negative terms.This helps users quickly understand the overall tone and focus of legal texts.By leveraging IBM Watson’s FLAN-T5-XXL model, the system ensures high-quality predictions.The integration with IBM Cloud Object Storage makes handling datasets efficient and secure.Interactive input support further enhances its usability for analyzing custom legal sentences.Overall, the project provides a smart, time-saving solution for sentiment analysis in legal workflows.