NewsBot Reflection – Individual Project by Erick Banegas

Project Overview

Working on the NewsBot Intelligence System was a very meaningful experience for me. I completed this project individually, which gave me the opportunity to fully understand every part of the NLP pipeline — from text cleaning to model training and sentiment analysis.
My main goal was to build an end-to-end system capable of processing real news articles, classifying them into categories, and extracting valuable insights using Natural Language Processing techniques. It was a challenging but rewarding process that helped me apply everything I learned from Modules 1–8.

Technical Implementation

Since I was working on my own, I had to design and integrate all components step by step. I started with **data preprocessing**, cleaning the raw BBC News dataset by removing duplicates, normalizing text, and preparing it for analysis.
Then, I used **TF-IDF vectorization** to represent the text numerically and trained three machine learning models: **Logistic Regression**, **Linear SVM**, and **Naive Bayes**. After comparing results, I found that the LinearSVC model performed best, reaching around **96% accuracy**.
Next, I expanded the system by including **sentiment analysis** with NLTK's Vader, **part-of-speech tagging**, **dependency parsing**, and **Named Entity Recognition (NER)** using spaCy. This integration gave me a complete system capable of classifying articles, analyzing emotions, and identifying entities such as people, organizations, or locations.

Challenges Faced

One of the main challenges I faced was managing all parts of the project alone. I had to plan my time carefully to ensure that each notebook worked correctly and connected smoothly with the next.
I also encountered some technical issues in Google Colab, such as package errors, missing resources, and memory limits when loading larger datasets. To solve this, I optimized the data size, limited the number of TF-IDF features, and processed the data in smaller batches.
Although it took patience to debug everything, these problems helped me strengthen my problem-solving skills and taught me how to work efficiently under time constraints.

Key Learnings

This project helped me understand how multiple NLP techniques can be combined to create a functional system that produces real insights.
I learned how important **data cleaning** and **feature representation** are for model accuracy. I also developed a deeper understanding of **linguistic features** such as syntax and entity extraction, which made me appreciate the complexity behind human language.
Working independently also helped me improve my organization, coding discipline, and confidence when explaining technical results.

Business Value

From a real-world perspective, the NewsBot system could be used by media companies, research organizations, or business analysts to monitor public opinion, categorize articles automatically, and extract key entities or trends.
For example, companies could use it to track how often their brand or competitors appear in the news, while researchers could analyze sentiment across political or economic topics.
Through this experience, I realized how NLP can provide practical business intelligence by transforming raw text into actionable data.

Future Improvements

If I had more time or resources, I would like to expand this project with **deep learning models** like BERT or DistilBERT to compare their performance against classical algorithms.
I would also like to build an **interactive web dashboard** using Streamlit or Gradio, allowing users to upload news articles and instantly see classifications, sentiments, and entities.
Another possible extension would be adding a **temporal sentiment trend analysis** to show how public tone changes over time by category or topic.

Personal and Professional Growth

Completing this project by myself helped me grow both technically and personally. I became more confident in handling complex machine learning workflows and NLP tasks.
It also helped me improve my time management and perseverance, as I had to take full responsibility for every step — from the dataset selection to the final documentation.
This experience confirmed my passion for Artificial Intelligence and motivated me to continue learning advanced NLP and automation techniques. I believe that the knowledge I gained here will be very valuable for my future professional career in AI and software engineering.