

Erwin Cheng

NewsBot\_Midterm\_Reflective Journal

ITAI 2373 Natural Language Processing

Anna Rachapudi

## NewsBot Midterm

Although this project was designed for group collaboration, I completed it independently as a solo developer. This required me to take full ownership of every phase—from dataset acquisition and cleaning to model design, testing, and analysis. Working alone helped me strengthen my project management and problem-solving abilities, as I needed to simulate a team environment by organizing my workflow, setting milestones, and managing my own deadlines. While I didn't have direct teammates to collaborate with, I engaged in self-reflection and used online resources, forums, and documentation as forms of indirect collaboration and peer learning. This independent approach enhanced my adaptability and technical confidence, teaching me how to approach AI projects from multiple perspectives including data science, development, and evaluation.

The NewsBot system was built using the BBC News Dataset obtained from Kaggle. The most challenging part of the process was ensuring that the data was properly cleaned and preprocessed before model training. The raw dataset contained inconsistencies, missing values, and category imbalances, which required data engineering techniques such as NaN value handling, text normalization, and tokenization. I implemented TF-IDF vectorization to convert textual content into numerical features and trained a Multinomial Naive Bayes classifier to categorize news articles into relevant topics. Integrating these components in Python using pandas and scikit-learn required careful testing to ensure data compatibility and performance optimization. Through trial and error, I refined the model and achieved a strong accuracy score, demonstrating effective feature selection and preprocessing strategies.

From a business perspective, NewsBot offers significant potential. Automated news classification can help organizations, media companies, and researchers efficiently organize and analyze large volumes of information. By categorizing news articles by topic, companies can track media trends, sentiment, and audience engagement more effectively. This capability can also improve recommendation systems, provide targeted content delivery, and support decision-making in marketing and public relations. In future iterations, integrating real-time news feeds and sentiment analysis would increase its value for businesses seeking timely insights into market and media behavior.

As the sole contributor, I handled every phase of the project. My contributions included dataset acquisition from Kaggle, data cleaning using pandas, exploratory data analysis, feature engineering, and model training using scikit-learn. I developed the classification logic using the Multinomial Naive Bayes algorithm and evaluated its accuracy using test data. I also prepared the documentation, reflected on performance metrics, and visualized category distributions to ensure the system's robustness. This end-to-end process gave me a deeper understanding of AI pipeline design, from concept to deployment readiness.

Future development of NewsBot would include adding a live API connection to pull real-time articles, implementing deep learning models such as BERT or DistilBERT for more accurate text classification, and expanding the system to support multilingual datasets. I also aim to integrate the model into a user-friendly web dashboard that displays categorized news and keyword trends interactively. These enhancements would make NewsBot more practical for both individual users and enterprises interested in automated news analytics.

This project was a major step in my professional development. I gained hands-on experience with natural language processing, machine learning, and Python-based data workflows. More importantly, I learned how to troubleshoot complex problems independently, adapt to challenges, and maintain consistency in project documentation and evaluation. Completing NewsBot alone gave me confidence in managing full AI project cycles—from concept to execution—and reinforced my interest in building intelligent, data-driven systems. The experience will directly inform my future work in AI model deployment and applied machine learning.