Mid-Term Group Project: NewsBot Intelligence System

Martin Demel and Jiri Musil

Department of Science, Technology, Engineering & Math, Houston Community College

ITAI 2373 Natural Language Processing

Patricia McManus

July 29th, 2025.

GitHub Link:

<https://github.com/martindemel/ITAI2373-NewsBot-Midterm>

Video Presentation Link:

<https://youtu.be/lsQLkM48fl0>

# Introduction

# Working on the NewsBot Intelligence System, we established clear communication and divided tasks effectively to utilize each other's strengths. We jointly developed the core NLP pipeline, including preprocessing, feature extraction, and classification algorithms, and created an interactive dashboard to enhance user experience. Regular check-ins and thorough documentation ensured smooth integration. Resolving differences in coding styles improved our codebase by combining robust error handling with efficient algorithms to deliver reliable, high-performance code.

# Body

Integrating eight NLP modules posed significant challenges, especially in creating a unified preprocessing pipeline. Balancing comprehensive text cleaning with preserving essential semantic features required extensive testing. The TF-IDF feature extraction was optimized for both accuracy and speed, enabling real-time use. We integrated VADER and TextBlob for sentiment analysis, carefully tuning their outputs. Our ensemble classification model, combining logistic regression, random forests, and SVM, achieved 98.7% accuracy. Additionally, spaCy-based Named Entity Recognition and the dynamic dashboard added further integration challenges, which we successfully managed through iterative development and rigorous testing.

The NewsBot Intelligence System provides substantial business benefits, notably reducing manual analysis costs by about 85%. Automated processing of over 1,000 articles per minute at 98.7% accuracy delivers timely insights for media, financial, and public relations sectors. Our real-time dashboard offers immediate access to actionable analytics, improving response times and decision-making. The scalable, modular architecture supports organizational growth and customized deployments, while our open-source design ensures transparency, compliance, and continuous improvement through community collaboration.

We equally contributed to developing and implementing the NLP processing pipeline, dashboard design, and user interfaces. We both worked closely on system architecture, seamlessly integrating components and creating comprehensive documentation, testing, and deployment frameworks that ensured reliability and ease of use.

Future developments include implementing real-time streaming analytics, multi-language capabilities, and robust API integration to expand usability within existing business ecosystems. In the longer term, integrating transformer-based models like BERT and GPT will improve analytical accuracy. Additional features, such as knowledge graph integration and predictive analytics, will further increase NewsBot’s value for businesses, ensuring its adaptability and relevance.

# Conclusion

The NewsBot project greatly improved our technical skills, especially in practical NLP implementation beyond just classroom learning. Working together on software development strengthened our project management skills and encouraged teamwork, including effective communication, resolving disagreements, and managing multiple tasks at once. Understanding how technical solutions align with real-world business goals broadened our professional perspective. This project confirmed our dedication to pursuing NLP professionally, emphasizing its powerful impact on business.

# References

GeeksforGeeks. (2025, July 23). Natural Language Processing (NLP) overview. GeeksforGeeks. <https://www.geeksforgeeks.org/nlp/natural-language-processing-overview/>