

Streamlit News Categorization Project – Summary

1. Model Choice:

- Model: Support Vector Classifier (SVC) with TF-IDF vectorization.
- Course: Cloud computing course from the previous semester, final team project.
- Why this model?
SVC is well-suited for text classification tasks due to its strong decision boundary for high-dimensional data (such as text TF-IDF vectors). It is computationally efficient and works well with relatively small datasets, making it ideal for a lightweight Streamlit prototype. Compared to deep learning models, it provides a faster response time, which is crucial for an interactive app where real-time classification is expected.

2. Prototype Utility

This prototype serves as an interactive news categorization app, allowing users to classify news articles into predefined categories using NLP techniques. Users can enter text manually, upload files, or select sample articles for classification. The app provides sentiment analysis, and word cloud visualizations to enhance user understanding.

By automating news classification, it can assist journalists, researchers, and content curators in organizing and analyzing news content efficiently. The prototype demonstrates how machine learning can be integrated into real-world applications for news filtering and recommendation systems. The use of Streamlit makes it accessible, easy to use, and deployable for non-technical users, making this project an excellent demonstration of how interactive AI-powered applications can be built quickly with minimal effort.

3. Main Difficulties Found

- User Input Handling: Ensuring seamless transitions between manual entry, file upload, and sample article selection required careful state management.
- Performance Optimization: The model and dataset were initially reloading too often, which was resolved using `@st.cache_resource` and `@st.cache_data` to enhance efficiency.

Link to video: [Recording-20250305_134638.webm](#)