Jade Sanchez

ITAI 2373

11/1/2025

News Bot Midterm Written Report Team InsightAI.

Project Overview: NewsBot Intelligence System is a system that will extract insights, categorise and process news articles automatically in a news article using advanced NLP techniques. The system combines the Modules 1-8 giving a full end-to-end pipeline of raw text input to actionable business insights. The main objectives include categorizing the articles (Politics, Sports, Technology, Business, Entertainment, Health), named entity extraction, sentiment analysis, and the main patterns are to determine the key patterns based on TF-IDF, POSTagging and syntax parsing.

Methodology:

Acquisition of Data: Data was collected by observing the study participants and their actions within the laboratory.<|human|>Preprocessing of Data: Data were obtained through observation of the members of the study and their behaviors in the laboratory.

BBC News Classification Dataset (not more than 2,000 articles)

Lemmatized, normalized, cleaned and tokenized text.

Stop word elimination and dealing with non-observed values.

Feature Extraction:

TF-IDF 1000-feature feature vectorization.

Dependency parsing and syntactic analysis through POS tagging.

Sentiment Analysis:

Sentiment scoring of the emotional tone of articles with VADER.

Classification:

SVM and Naive Bayes multi-class classifier.

Split/train test (acc, F1-score) and model analysis (accuracy, F1-score).

Named Entity Recognition:

org, person, amount, date, entity.

Cross-category mapping and frequency analysis.

Visualization:

Bar graphs of the number of articles in each category.

POS and sentiment distribution charts.

Results & Insights:

SVM was very much accurate in classifying the news articles into six categories.

TF-IDF top terms prioritized category specific words (e.g., "government" in Politics, "stock" in Business).

Sentiment analysis showed the skew in Sports articles has been towards positive and Politics towards more neutral/negative.

NER gave common entities, which made it possible to trace major organizations and individuals.

Business Value NewsBot can be of great use to businesses in terms of media monitoring, competitor tracking, and market researching, which is a fast and automated means to understand

a huge news dataset.

Future Work:

Add BERT-based classifier to enhance performance.

Enlarge data set of real-time surveillance.

Implement trend detection and topic modeling.

Maximize computational performance with big data.