Analysis of Text Representation Methods for Romanian Fake News Detection

Supervisor
Lect. PhD. Lupea Mihaiela-Ana

Author
Pop Răzvan-Gabriel

Babeș-Bolyai University, Cluj-Napoca, Romania

Faculty of Mathematics and Computer Science

July 3, 2025

Problem & Motivation

Problem

- Fake news spreads rapidly on digital platforms.

- There are very few automated detection tools specifically for the Romanian language.

Motivation

- Disinformation has influenced public opinion on topics like **elections, public health, and the economy.**

- There is a clear need for automated tools that can **analyze and understand Romanian text.**

Related Work and Main Objectives

Related Work

- Previous studies show good results for Romanian fake news classification using models like **CNN, SVM, and BERT.**

- Recent research explored emotional processing and how it affects belief in fake news.

Main Objectives

- **Work** with a dataset in the Romanian Language.

- **Analyze** the data to find the key linguistic and emotional patterns that can be turned into features.

- **Compare** different text representations to see which performs best with an SVM classifier.

- **Build** a working web application to apply the findings in practice.

Text Representation Techniques

Corpus-Based

- **Word2Vec**: dense, static, word-level.

- **Doc2Vec**: dense, static, document-level.

- **TF-IDF**: sparse, high-dimensional, keyword-based

Dynamic and Contextualized

- **RoBERT (CLS Token):** dynamic, context-aware representation.

Classification and Clustering Algorithms

Supervised

Support Vector Machine(SVM):

- The main algorithm used for classification in all experiments.

- Works by finding the best possible boundary (hyperplane) to separate the different news categories.

- Chosen to ensure a fair and direct comparison between all text representation methods.

Unsupervised

K-Means Clustering:

- Used to automatically group similar documents into clusters based on their Doc2Vec vectors.

- Its role was for **feature engineering**: the ID of the cluster a document belonged to was used as a single feature in the hybrid model.
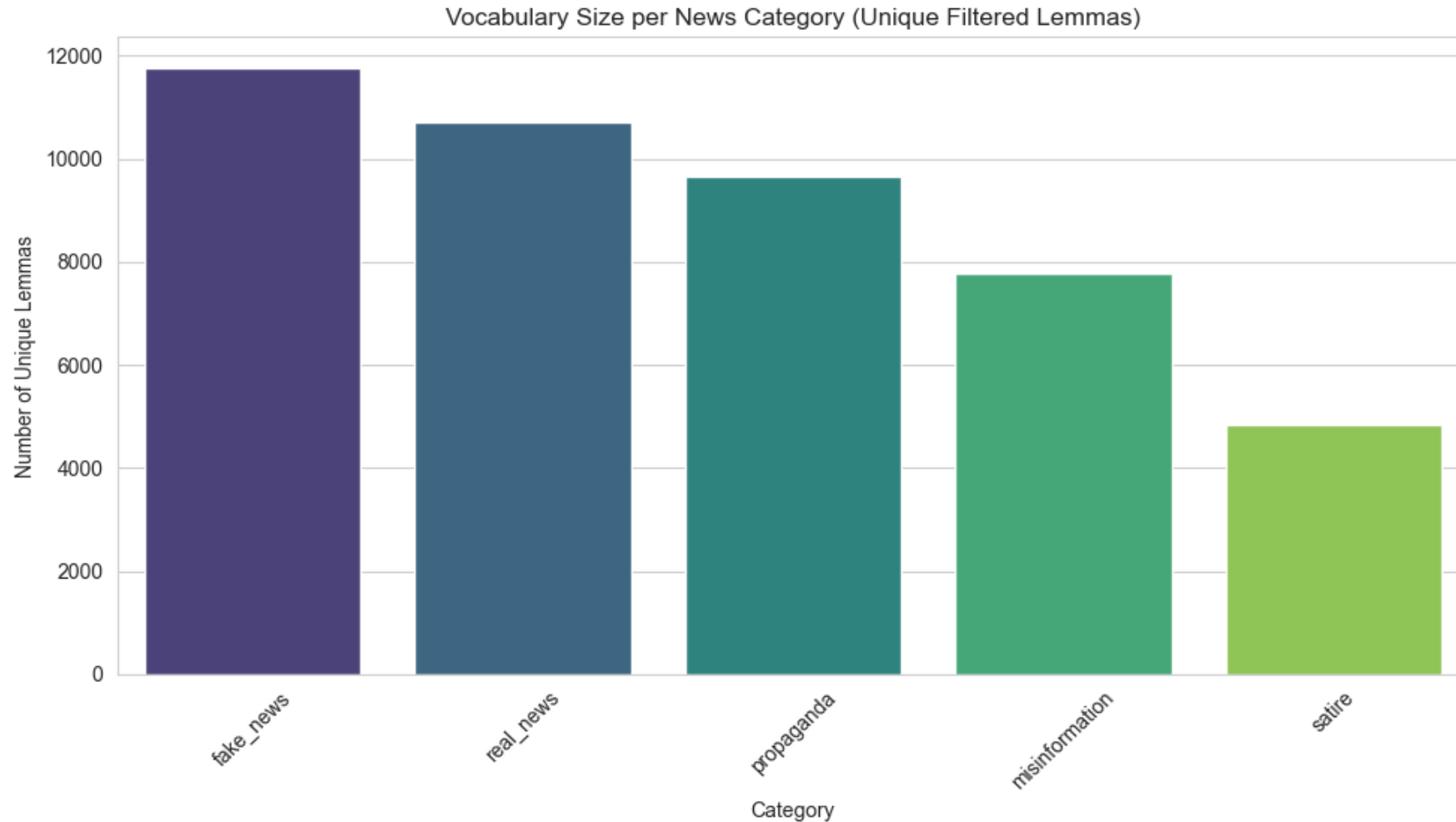
Dataset Analysis Perspective
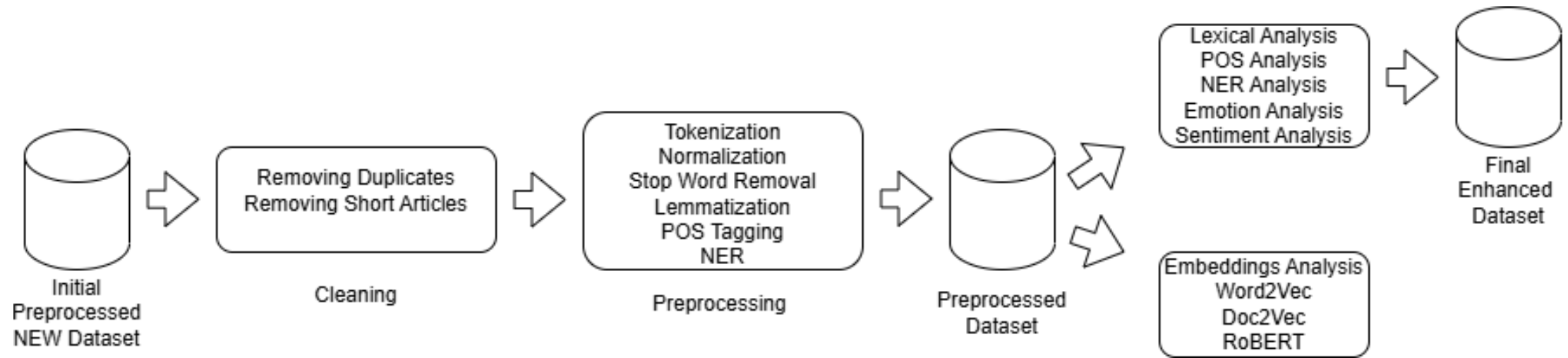
- **Preprocessing**: Tokenization, Normalization, Stop Word Removal, Lemmatization, POS Tagging, Named Entity Recognition (NER).

NEW Dataset for Romanian Fake News Detection

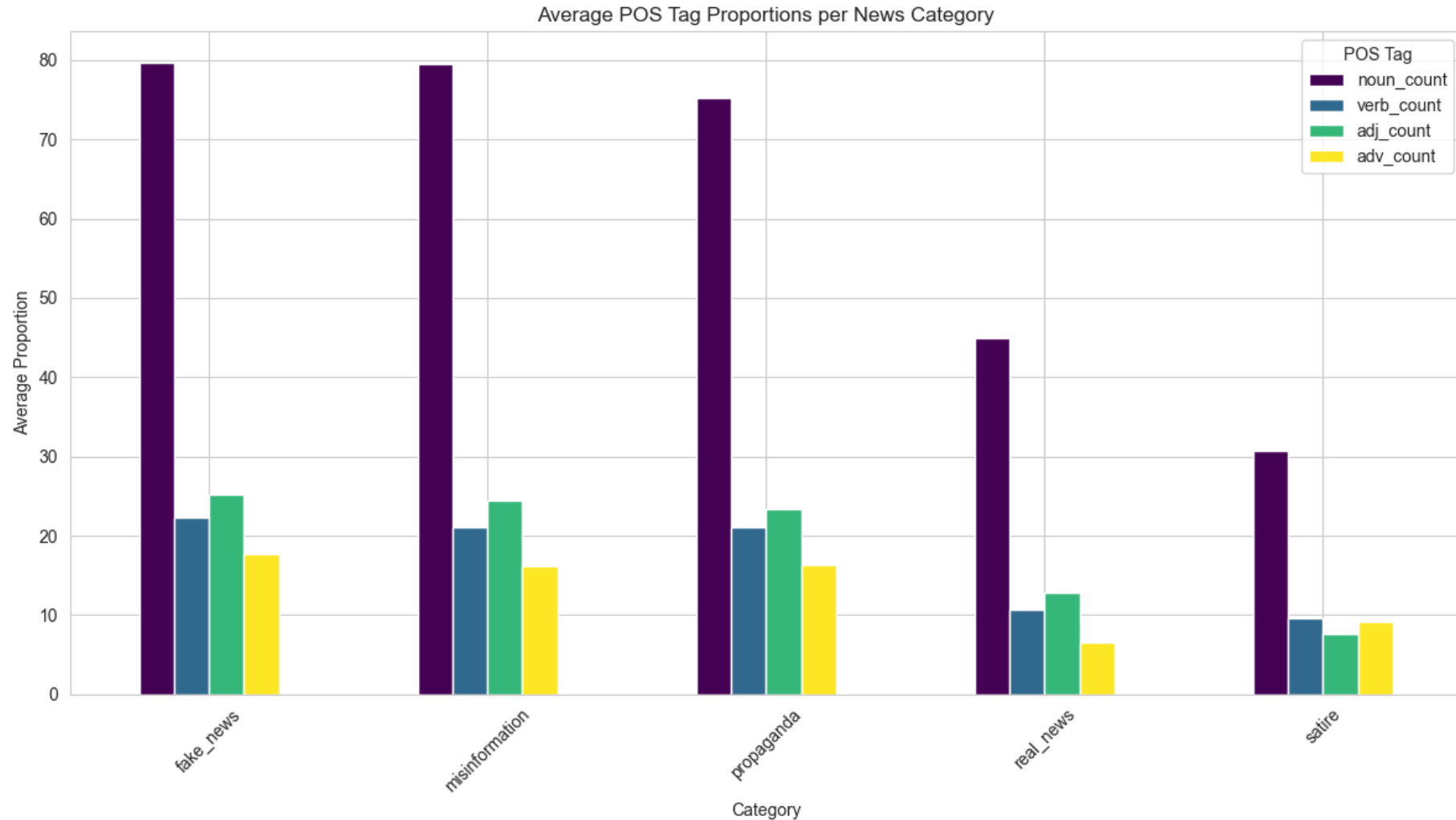| Category	Count	Percentage
Real	1,032	23.8%
Misinformation	916	21.1%
Satire	870	20.1%
Fake	812	18.7%
Propaganda	704	16.2%

- **Real News:** Factual, verifiable information that aims to inform objectively.
- **Fake News:** Intentionally false content designed to deceive an audience, often for financial or political gain.
- **Misinformation:** False information that is spread, but not necessarily with the intent to deceive.
- **Propaganda:** Biased or misleading information used to promote a specific political cause or point of view.
- **Satire:** Content that uses humor, irony, and exaggeration to criticize or offer commentary on issues.

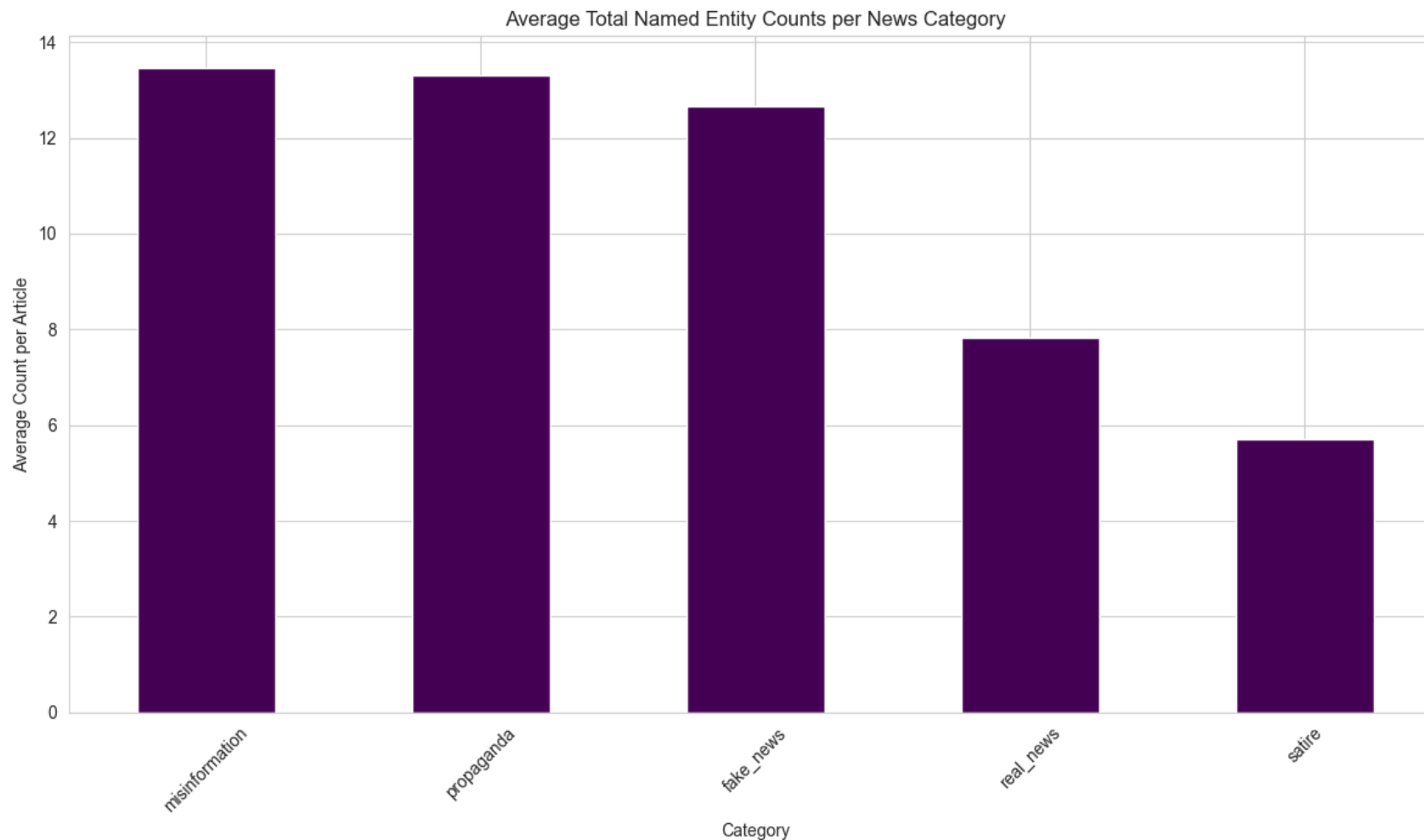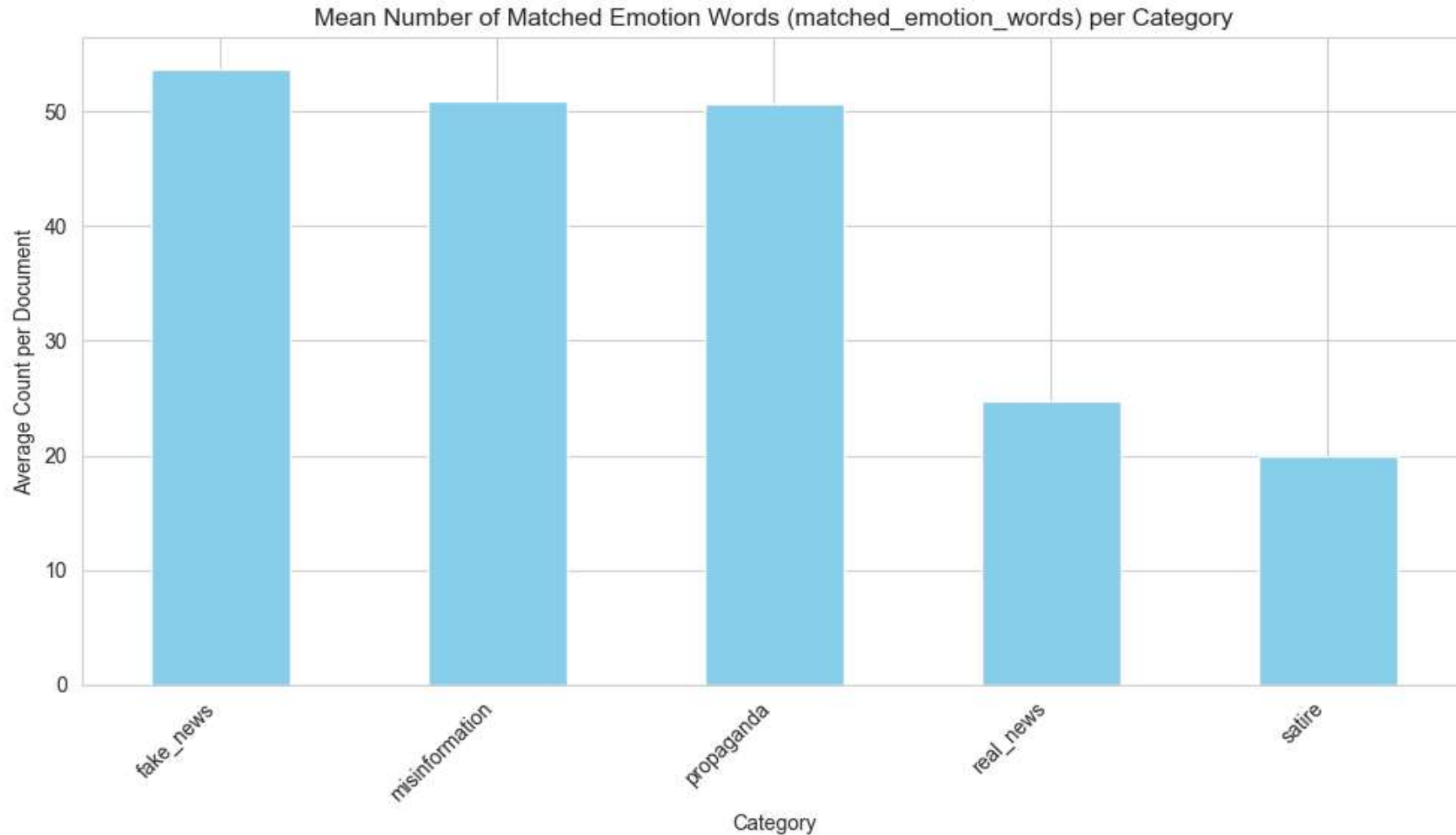# Dataset Analysis: Lexical and Syntactic Analysis I

Vocabulary Size per News Category (Unique Filtered Lemmas)

Dataset Analysis: Lexical and Syntactic Analysis II

Average POS Tag Proportions per News Category

Dataset Analysis: Lexical and Syntactic Analysis III

Average Total Named Entity Counts per News Category

Dataset Analysis: Emotion and Sentiment Analysis I

Mean Number of Matched Emotion Words (matched_emotion_words) per Category

Dataset Analysis: Emotion and Sentiment Analysis II

Mean Emotion & Polarity Scores per News Category

Dataset Analysis: Emotion and Sentiment Analysis III

Average Proportions of Sentiment-Specific Words
(as % of Total RoEmoLex Matched Words per Article) by Category
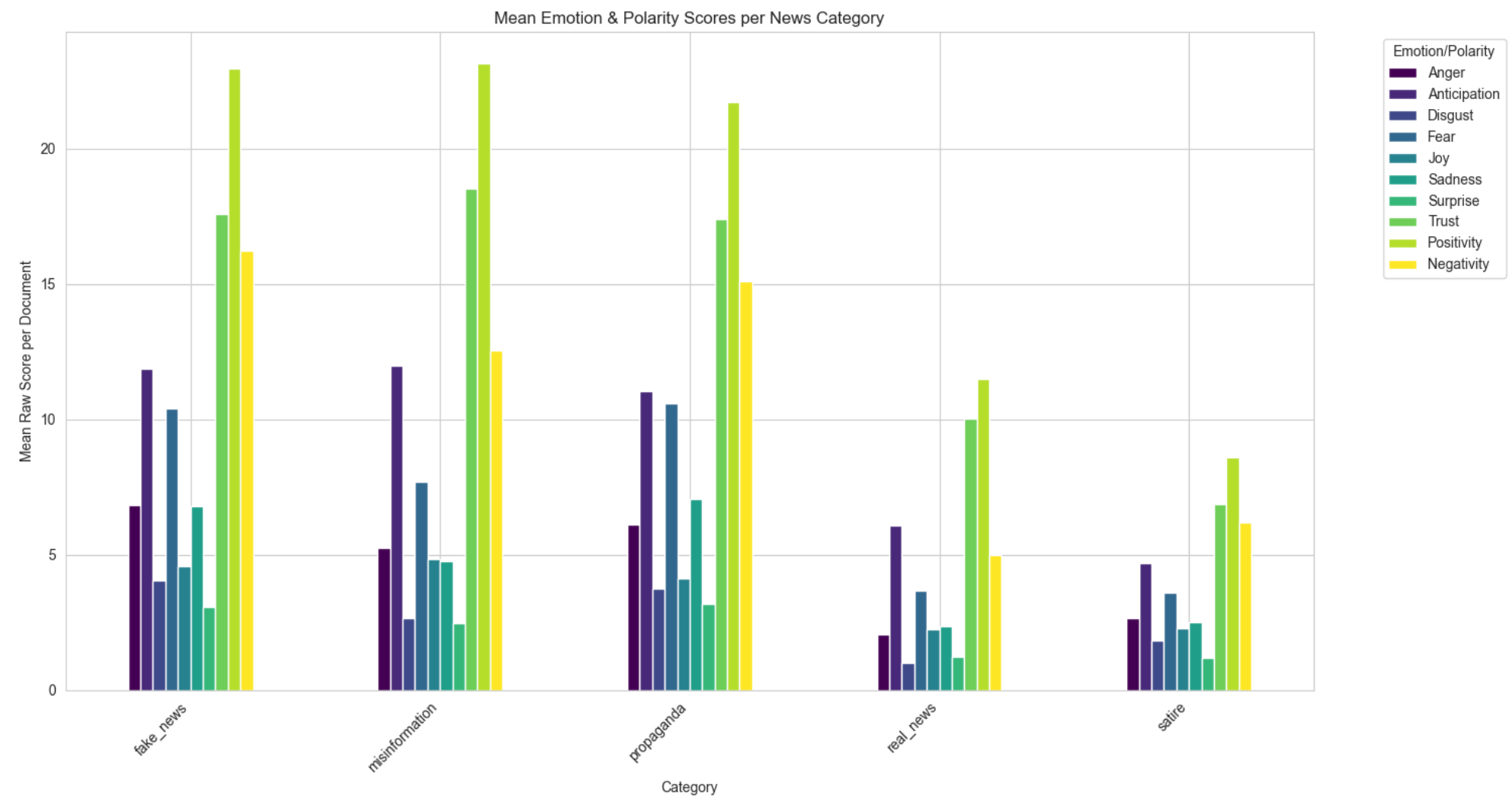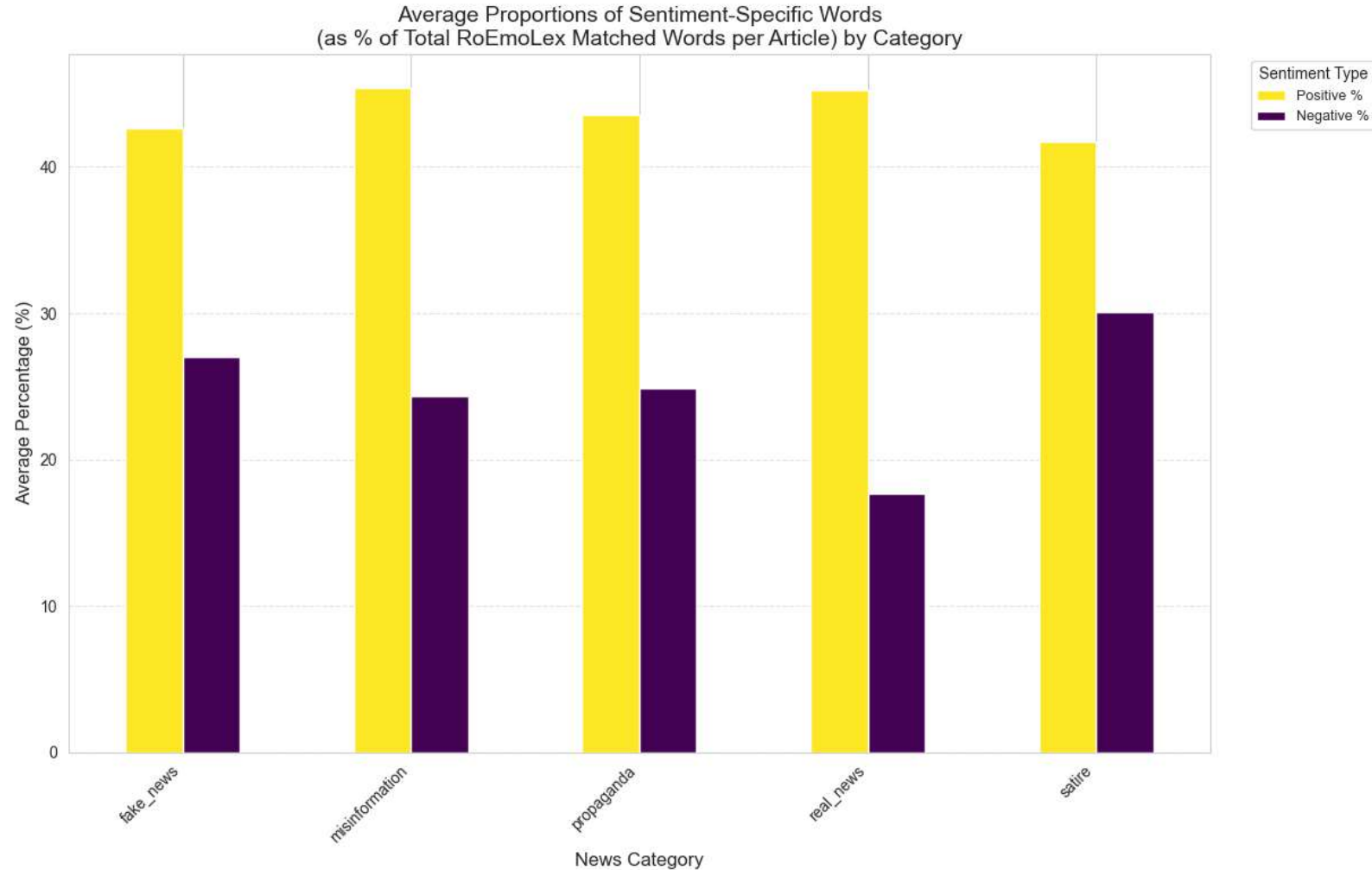
Dataset Analysis: Embeddings Analysis
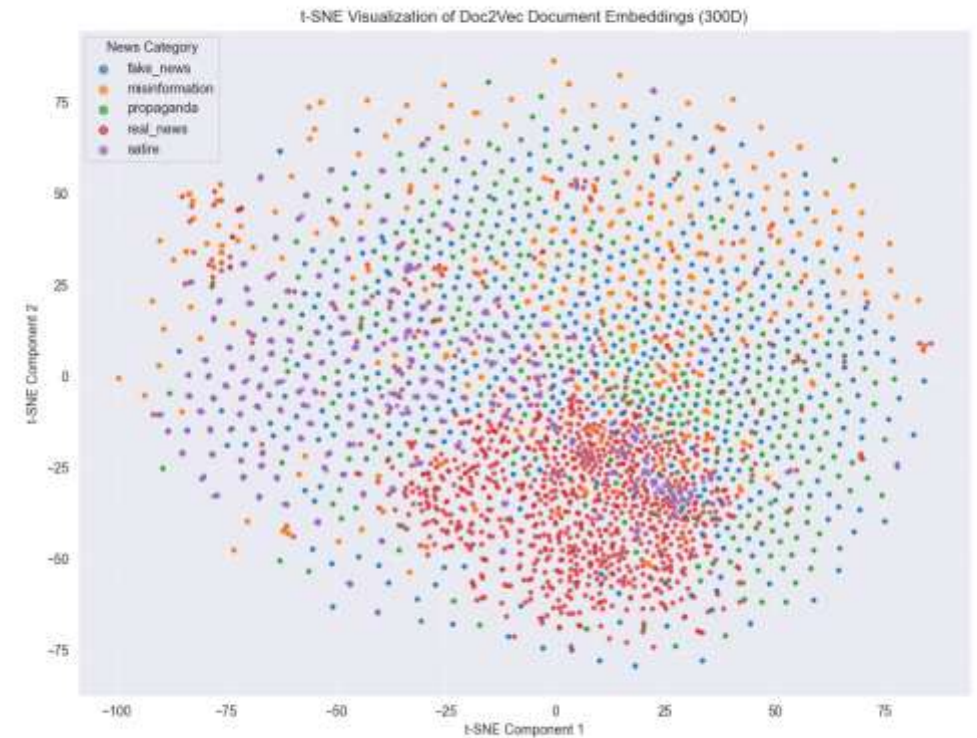
Word2Vec(300 dimensions)

- The custom Word2Vec model successfully learned the specific, biased meanings of words within the news dataset.

- It clearly separated the semantic neighborhoods of opposing concepts such as **știință (science)** and **conspirație (conspiracy)**, showing it understood their usage in this context.

t-SNE Comparison: 'știință' vs 'conspirație' Neighborhoods (300d Word2Vec Model)
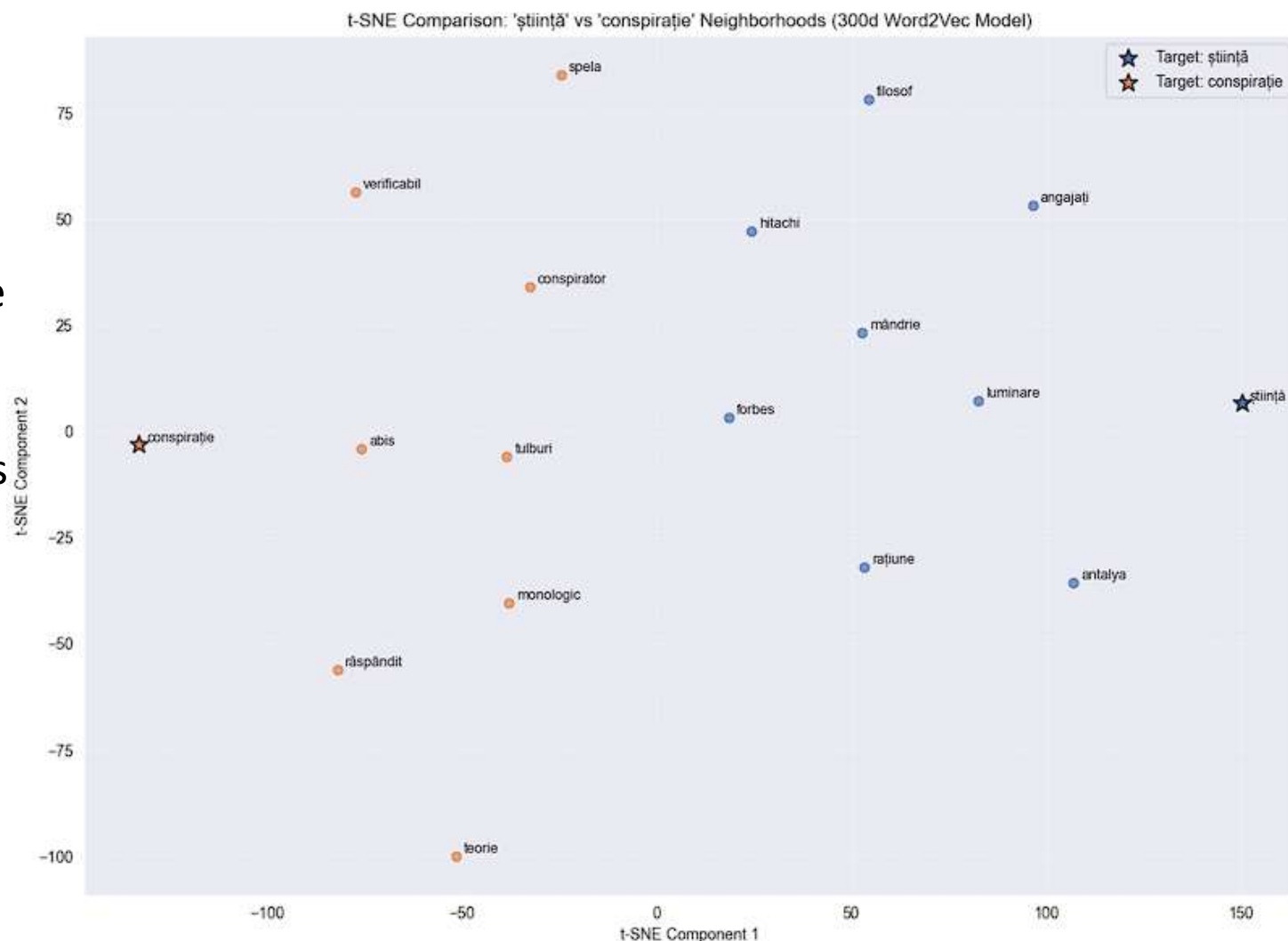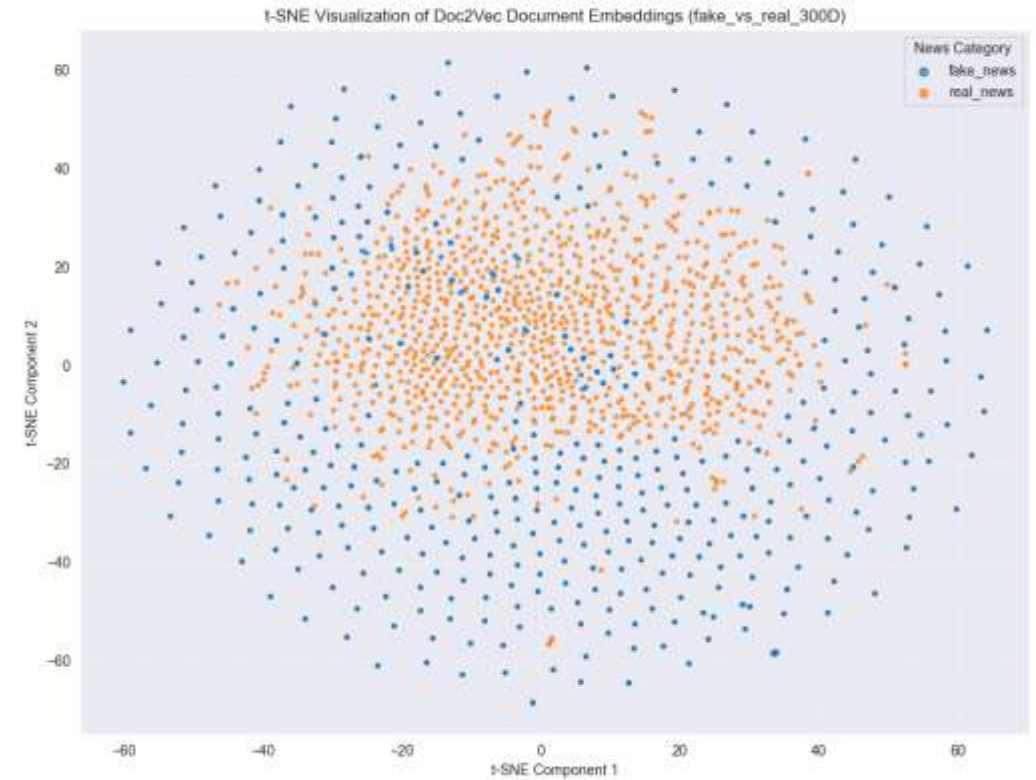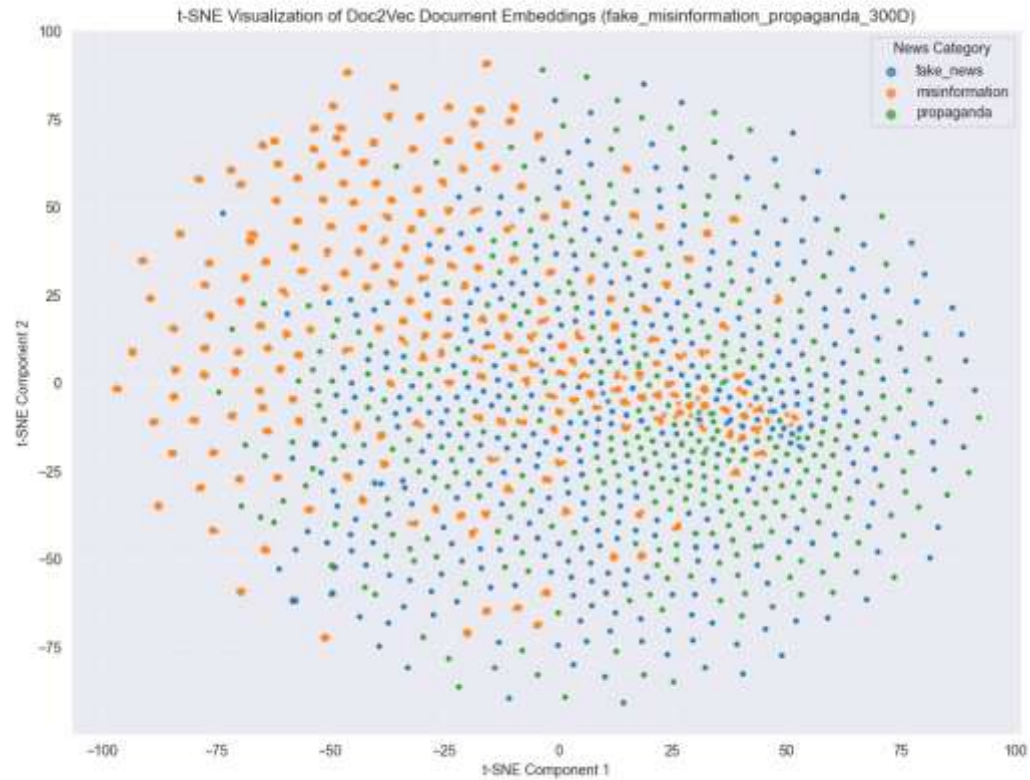
Dataset Analysis: Embeddings Analysis

Doc2Vec(300 dimensions)

- The Doc2Vec model was used to create a single vector representing the overall meaning of each article.

- **significant overlap** between all five categories, meaning they often discuss the same topics.

- **Observations: Real news** articles (in red) form a **dense semantic core**. The other categories, especially **fake news** (blue) and **satire** (purple), are more scattered around this central core.

- **Conclusion**: This suggests that real news is more thematically consistent, while disinformation is more varied in its style and content.

t-SNE Visualization of Doc2Vec Document Embeddings (300D)

Dataset Analysis: Embeddings Analysis

Doc2Vec(300 dimensions)

t-SNE Visualization of Doc2Vec Document Embeddings (fake_misinformation_propaganda_300D)

t-SNE Visualization of Doc2Vec Document Embeddings (fake_vs_real_300D)

Classification: Methodology

Text Representations Tested

- **TF-IDF** (Keyword-based vectors)

- **Word2Vec** (Averaged word vectors)

- **Doc2Vec** (Document vectors)

- **RoBERT [CLS]** (non-fine-tuned)

- **Hybrid Model** (Doc2Vec + Engineered Features)

Evaluation

- To ensure a fair comparison, a **Support Vector Machine (SVM)** was used as the classifier for all experiments.

- Performance was measured using **5-fold stratified cross-validation**.

Engineered Features

- **Word count**: Total filtered lemmas
- **Vocabulary richness**: Unique filtered lemmas / total filtered lemmas ration
- **Part-of-speech proportions**: Nouns, verbs, adjectives and adverbs
- **Total entities**: people, nationalities, religious, political groups, organizations, facilities, events, money, locations
- **Entity density**: Entities / total lemmas ratio
- **Emotion word count**: Lemmas matching RoEmoLex lexicon
- **Emotions proportions**: anger, joy, sadness, trust, anticipation, disgust, fear, surprise
- **Polarity counts (Positivity & Negativity)**: Positive and negative lemma frequencies
- **Polarity percentages**: Sentiment proportions as percentages
- **Cluster ID**: Document cluster assignment

Classification: Results

Experimental Method	Mean F1-Score
TF-IDF	0.9570 ± 0.0067
Hybrid Model	0.9169 ± 0.0141
Doc2Vec	0.8046 ± 0.0117
Word2Vec (Averaged)	0.7686 ± 0.0113
RoBERT [CLS] (no fine-tune)	0.7376 ± 0.0072

Web Application

- **Technology:** The application was developed using the **Django** web framework.

- **Core Feature:** It classifies text using the best-performing method from the experiments (**TF-IDF + SVM**), which was then re-trained on the full dataset.

- **Unique Logic:** The results are displayed differently based on the model's confidence score (>80%, 60-80%, <60%).

- **Analysis Tool:** It includes a separate page for exploring the custom **Word2Vec models** and allowing users to download them.

Conclusion & Main Contributions

Main Conclusions

- The linguistic and emotional analysis revealed clear, measurable differences between real news and disinformation.

- For classification, a **TF-IDF model performed best (~96%)**, showing that specific keywords are a very powerful signal.

- A **Hybrid Model (~92%)**, combining document vectors with engineered features, is also a highly effective strategy.

Key Contributions

- A detailed analysis of a Romanian news dataset, creating a "fingerprint" for disinformation.

- Two custom, lemmatized **Word2Vec models** for the Romanian news domain.

- A functional **web application** that demonstrates the practical use of the research.

Future Work

- Create a more diverse dataset.

- Fine-tune a RoBERT model.

- Deploy the application to the cloud.

Thank You

Questions?