

# **Sentiment Analysis of News Entities in Armenia**

Generated by GPT-3.5, edited by Mher Movsisyan

## **1. Introduction**

The rapid proliferation of online news sources has led to a wealth of information that can be analyzed to gain insights into public sentiment and opinions about various entities and topics. This project aims to harness the capabilities of GPT-3.5, a state-of-the-art language model, to scrape news articles from three distinct online sources, extract entities, and perform sentiment analysis on each entity mentioned. By doing so, we will create a time-series dataset and investigate if there are significant differences in the sentiment expressed toward these entities across different news sources.

## **2. Objectives**

The primary objectives of this research project are as follows:

- Collect news articles from three online sources.
- Utilize GPT-3.5 to extract entities mentioned in the news articles.
- Conduct sentiment analysis on the extracted entities.
- Aggregate sentiment data over time to create a time-series dataset.
- Statistically analyze the data to determine if there are significant differences in the sentiments expressed toward the same entities across different sources.

## 3. Methods

### 3.1 Data Collection

We will gather news articles from three diverse online sources to ensure a broad spectrum of news perspectives. Sources will be selected to encompass various domains, such as politics, finance, and technology. Web scraping will be employed to collect a substantial number of articles for analysis.

### 3.2 Entity Extraction

GPT-3.5 will be used to identify and extract entities, including people, organizations, and locations, from the news articles. Named Entity Recognition (NER) capabilities of the model will be utilized to ensure accuracy and comprehensiveness.

### 3.3 Sentiment Analysis

Sentiment analysis will be conducted using GPT-3.5's default, pre-trained capabilities for each extracted entity. GPT-3.5 will be responsible for determining the sentiment of each entity and classifying it as positive, negative, or neutral without any specific training or fine-tuning for this particular task.

This approach will leverage GPT-3.5's existing language understanding capabilities to provide sentiment scores for the identified entities in the news articles. The sentiment analysis will be carried out in an out-of-the-box manner without any modifications to the model.

The remaining methods and tests, as outlined in the proposal, will remain consistent with this

approach, acknowledging that GPT-3.5 will not undergo additional training or fine-tuning specifically for the sentiment analysis task.

### 3.4 Time-Series Dataset

To analyze the evolution of sentiment over time, we will organize the sentiment scores for each entity into a time-series dataset. The data will be aggregated at daily or weekly intervals to capture trends and fluctuations in public sentiment.

### 3.5 Statistical Analysis

Statistical tests will be used to determine if there are significant differences in sentiment toward the same entities across the different news sources. We will employ the following methods:

- **ANOVA (Analysis of Variance):** To test for significant differences in sentiment scores across multiple sources.
- **Post-hoc Tests:** If significant differences are found, post-hoc tests (e.g., Tukey's HSD) will be used to identify which sources differ significantly.
- **Time-Series Analysis:** Trends and patterns in sentiment data will be analyzed using time-series techniques, including Autoregressive Integrated Moving Average (ARIMA) or Exponential Smoothing.
- **Hypothesis Testing:** Hypothesis tests will be conducted to confirm or reject the presence of statistically significant differences.

#### **4. Ethical Considerations**

Ethical considerations will be taken into account throughout the research process, including only publicly available data, ensuring that the analysis is unbiased, and being transparent about the methodology.

#### **5. Expected Outcomes**

The project aims to provide insights into how different online news sources influence the sentiment expressed toward various entities over time. The outcomes will contribute to a better understanding of the impact of media on public perception and potentially have applications in media studies, data journalism, and sentiment analysis.

#### **6. Conclusion**

This research project seeks to uncover valuable insights into how different online news sources portray entities in terms of sentiment and to assess the significance of these differences. By employing web scraping, entity extraction, sentiment analysis, and statistical methods, we aim to provide a comprehensive analysis of sentiment trends across multiple sources.