Sentiment Analysis of Western News about India

Aryaman Pathak

June 30, 2024

1 Introduction

This document presents the outcomes and results of a sentiment analysis project focused on Western news articles about India. The analysis includes overall sentiment distribution, company-wise sentiment analysis, topic modeling using LDA, and clustering visualization. The methods used include ensemble sentiment analysis with VADER, TextBlob, and Flair.

2 Overall Sentiment Distribution

The overall sentiment distribution in the dataset is visualized in the bar chart and pie chart below.

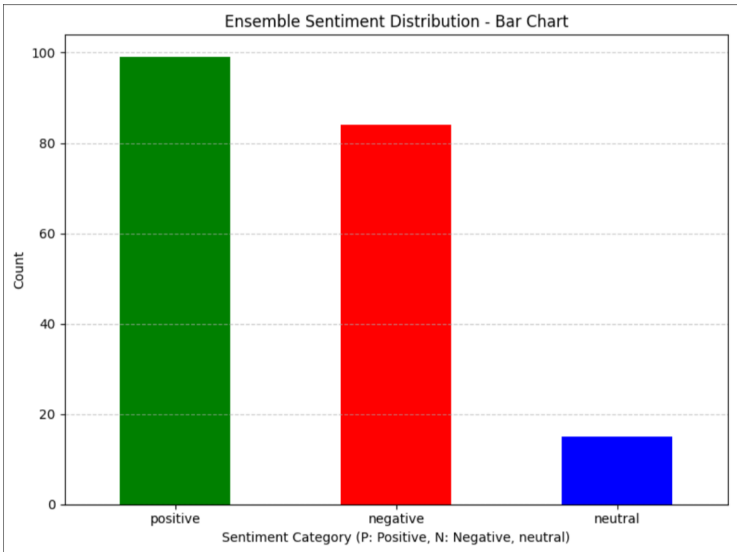

Figure 1: Overall Sentiment Distribution - Bar Chart

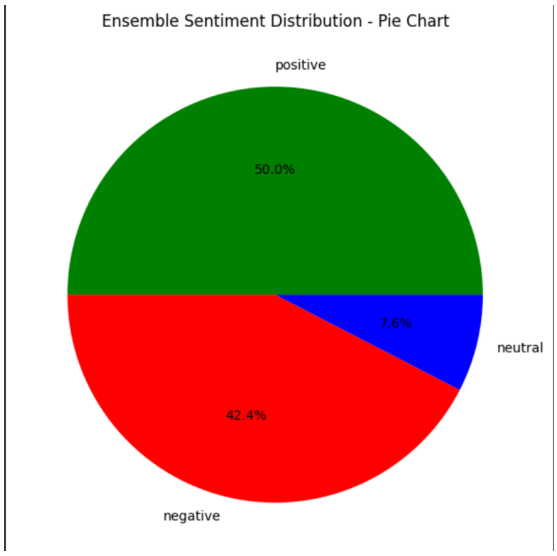

Figure 2: Overall Sentiment Distribution - Pie Chart

3 Company-wise Sentiment Analysis

The sentiment distribution for each news company is analyzed and visualized in the following bar charts and pie charts.

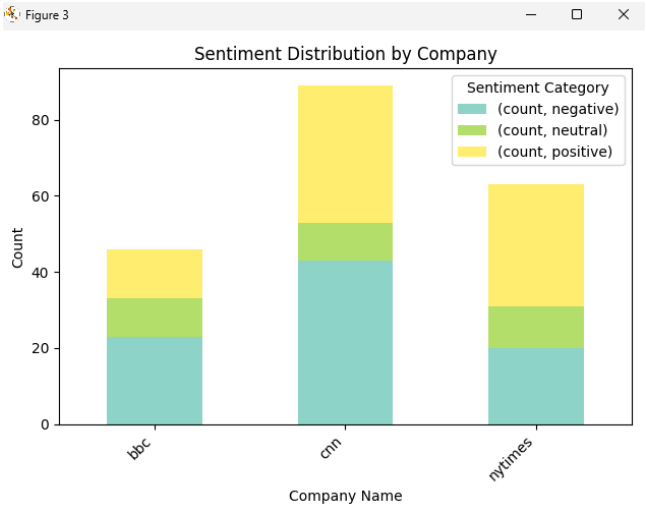

Figure 3: Sentiment Distribution for Company A - Bar Chart

4 Main Topics Using LDA

The main topics in the dataset were identified using Latent Dirichlet Allocation (LDA). The word clouds below illustrate the key topics.

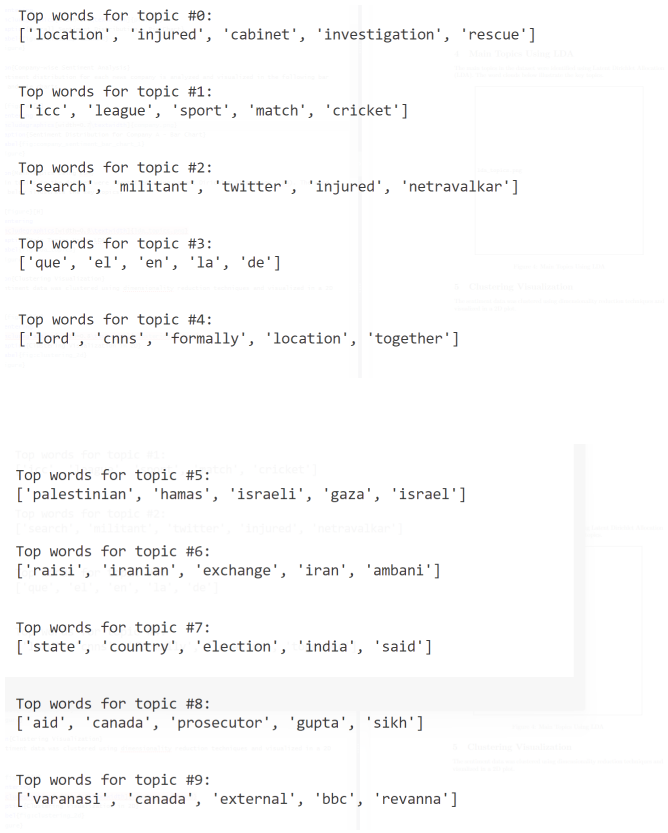

Figure 4: Main Topics Using LDA

5 Clustering Visualization

The sentiment data was clustered using dimensionality reduction techniques and visualized in a 2D plot using Kmeans

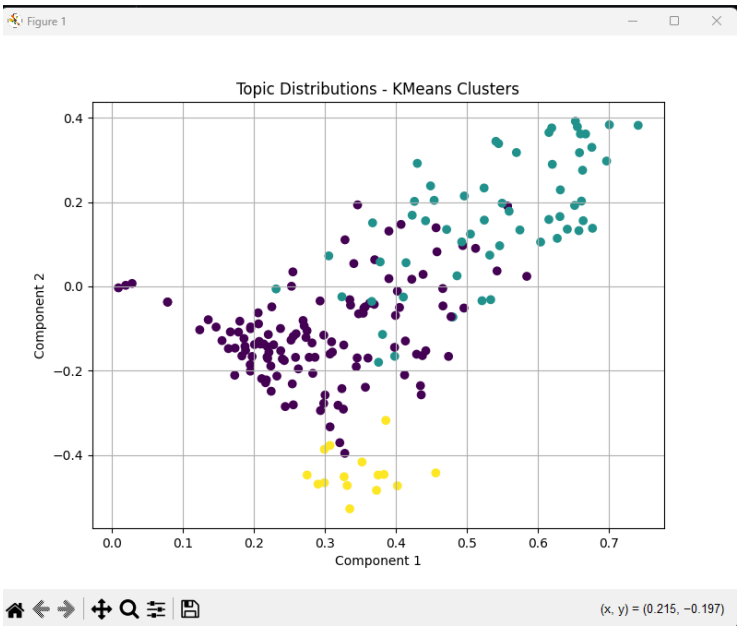

Figure 5: Clustering Visualization in 2D

6 Methods and Tools Used

This analysis utilized an ensemble method combining the sentiment analysis tools VADER, TextBlob, and Flair to achieve robust sentiment classification.

6.1 VADER

The VADER (Valence Aware Dictionary and sEntiment Reasoner) tool is a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media.

6.2 TextBlob

TextBlob is a simple Python library for processing textual data, providing a consistent API for diving into common natural language processing (NLP) tasks.

6.3 Flair

Flair is a powerful NLP library developed by Zalando Research, capable of performing a wide range of NLP tasks, including named entity recognition (NER), part-of-speech tagging (POS), and sentiment analysis.

7 Data Source

The dataset used in this analysis is available for download here(Link). It consists of news articles from various Western media outlets, categorized by company