Stevens Center for Sustainability

525 River Street (Babbio 317A)
Hoboken, NJ 07030
stevens.edu

NAME: _____Tina Nosrati_____

Project Report for PSEG Summer Undergraduate Research Program in Environmental Justice

Date: 8/1/2024

Project Title:
An Analytical Study of the Challenges Faced by Underprivileged Communities in New Jersey: Insights from Social Media Expressions

1. Introduction:

This research aims to understand the challenges faced by underprivileged communities in New Jersey by analyzing social media and news article content from general online sources using Bing search results. Our Dataset contains news articles for more than 350 areas in New Jersey in the past year. This research has used text mining techniques and large language models to process the data and perform topic modelling to identify the challenges communities face in New Jersey. Sentiment analysis has also been used to detect negative-toned articles and investigate them more. The result of the Analysis has been summarized in 3 main indicators, including Weighted Social Negative Index (WSNI), Crime Index and Social Negative Articles Ratio for each area. The final results were presented using two interactive maps and a dashboard summarizing the findings.

1.1 Project overview:

This study holds great significance as it takes a data-driven approach to understand the challenges faced by underprivileged communities in New Jersey. By identifying and analyzing these issues, the research provides valuable insights that can help guide policymakers, community leaders, and social organizations in making informed decisions.

1.2 Research questions:

1. Based on news and social media analysis, what are the main challenges faced by underprivileged communities in New Jersey?
2. How do negative news articles relate to the challenges in these areas?

3. What do the WSNI, Crime Index, and Social Negative Articles Ratio reveal about the conditions in New Jersey communities?

2. Literature Review:

The literature review covers the design science methodology. It also looks at studies using machine learning to analyze social media data, showing how these methods can identify community challenges. Previous research proves these techniques give useful insights into socio-economic issues and help guide interventions and policies.

3. Methodology

3.1 Research Design

The project starts by initiating an automated data-gathering process using Python that searches for more than 350 areas in New Jersey on Bing.com and then crawls a number of articles related to each of these areas. The text content of each article will be saved in a Jason file, and all the articles related to an area are gathered in a folder.

3.2 Procedures

Each article has been broken down into sentences and each sentence has been passed through 2 different labeling processes. First, the sentence has been passed to the Fine-BERT model to be labelled as Environmental, Social, Governance or None. Next, sentiment analysis will be applied to decide whether the sentence has a positive, negative or neutral tone. The labels will be saved next to each sentence in the Jason files.

4. Results

4.1 Quantitative Data Analysis

The data analysis process includes calculating different facts and approaching the data from different perspectives. Firstly, the ratios of Environmental, Social, and Governance articles have been calculated for each area. Secondly, the ratio has been calculated for all the combinations of labels, such as Social Positive, Social Negative, Environmental Positive, and Environmental Negative….
Finally, three main indicators have been defined to represent the ongoing patterns in the data:

a. **Social Negative Articles Ratio:** The ratio of negative social articles in a specific area to the total number of such articles.

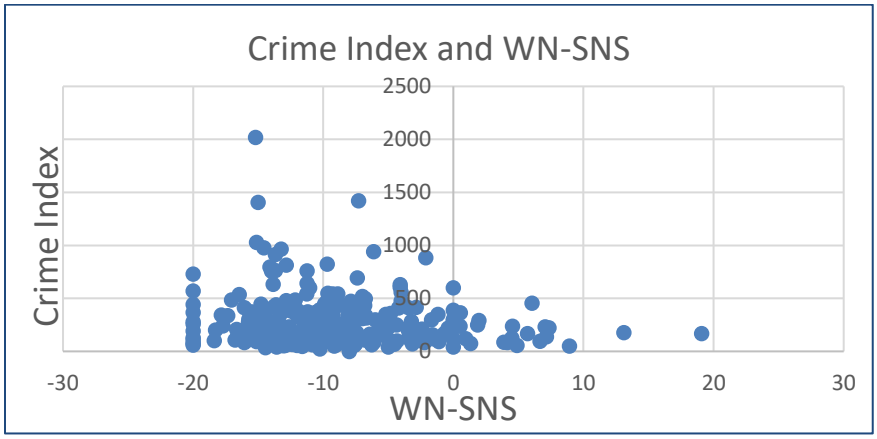
b. **Crime Index:** The total number of reported crimes per capita, multiplied by 10,000. The Crime information has been obtained from https://ucr.fbi.gov/.
c. **Weighted Social Negative Index (WSNI):** Measures the proportion of negative social sentiment relative to total interactions, scaled by 100. A higher WSNI indicates more negative interactions. Positive WSNI (up to 60) suggests dominance of negative interactions; negative WSNI (down to -40) indicates positive and neutral interactions are more prevalent; a score near zero shows a balance.

4.2 Topic modeling

We applied Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) topic modelling techniques on articles labelled as socially negative. The analysis revealed three main categories: Local Areas and Communities, Municipal Services and Emergency Management, and Health and Demographics. Significant words for each category have been identified and reported in a table within the dashboard, providing a detailed overview of the key themes and terms associated with these socially negative articles.

4.3 Interpretation

The presented dashboard summarizes our research findings and provides interactive maps for viewers to explore and draw their own conclusions. While the distribution of socially negative articles is an important indicator, it does not necessarily imply a direct correlation with higher crime rates in those areas.



5. Discussion

5.1 Summary of Findings

In addition to a comprehensive dataset that can be used for various types of analytics, this project provides a dashboard to summarize the result of its analysis. One of the most significant findings of this research is that Local Areas and Communities, Municipal Services and Emergency Management, and Health and Demographics are the three main concerns, according to social articles in New Jersey areas. Another important finding is that although the distribution of socially negative articles is a significant indicator, it does not automatically indicate a direct correlation with higher crime rates in those areas. Finally, each of the indicators in this research can be used for planning and investing to address problems in different areas.

5.2 Comparison with Literature

Compared to other literature that can draw a more specific conclusion, our findings should be used with caution. They highlight the complexity of the relationship between social negativity and crime rates, indicating that socially negative articles reflect community concerns without directly predicting crime. Like other literature, the design science methodology used in this research is helpful for understanding and addressing social issues.

5.3 Limitations

The limitations of this project include a lack of articles or reliable crime indexes for some of the areas in New Jersey. Also, the data-gathering process has some limitations because of its speed and the nature of web page articles.

5.4 Future Directions

We are considering expanding this research and the dashboard to represent news articles in New Jersey in real time by designing a more efficient data gathering process and more comprehensive indicators, according to the feedback from this phase of research.

6. Conclusion

This research provides a detailed analysis and interactive dashboard summarizing key concerns in New Jersey areas, particularly Local Areas and Communities, Municipal Services and Emergency Management, and Health and Demographics. While the distribution of socially negative articles is a significant indicator, it does not directly correlate with higher crime rates, reflecting the complexity highlighted in existing literature. Future research will focus on enhancing real-time data gathering and developing comprehensive indicators to further improve understanding and planning for social issues.

7. Appendices

Dashboard: https://tinrt.github.io/PSEG/Dashboard.html