An Analytical Study of the Challenges Faced by Underprivileged Communities in New Jersey: Insights from Social Media Expressions
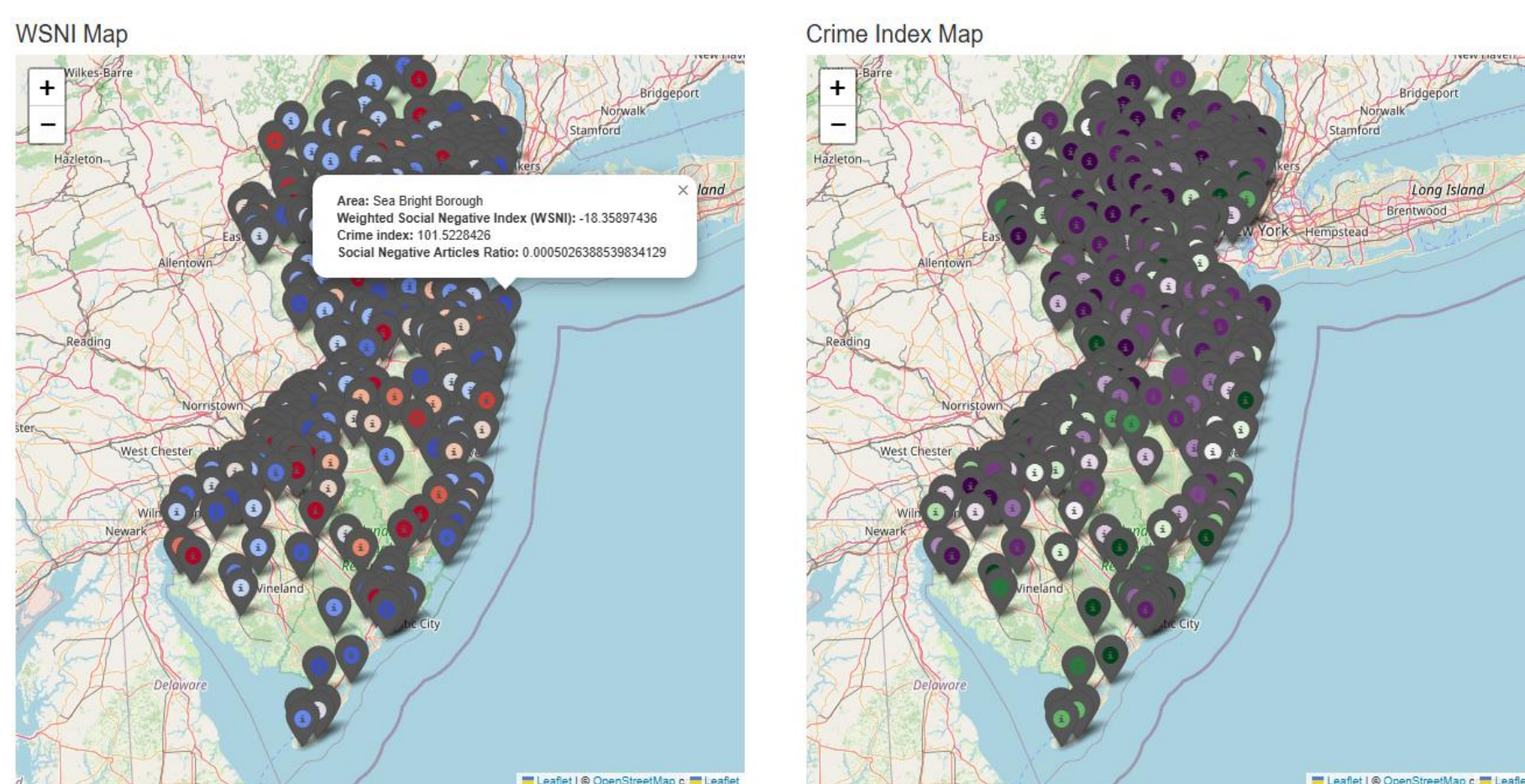
Tina Nosrati, Ramapo College of New Jersey,
Professor Arion Cheong, Stevens Institute of Technology

Introduction

This research analyzes challenges faced by underprivileged communities in New Jersey using social media and news articles from Bing search results covering over 332 areas in the past year. It employs text mining, large language models, topic modeling, and sentiment analysis to identify and investigate issues. Results are summarized through three indicators: Weighted Social Negative Index (WSNI), Crime Index, and Social Negative Articles Ratio, presented via interactive maps and a dashboard.

Research questions:

1. What are the main challenges faced by underprivileged communities in New Jersey based on news and social media analysis?
2. How do negative news articles reflect these challenges?
3. What do WSNI, Crime Index, and Social Negative Articles Ratio reveal about these communities?

Methodology

Research Design: The project employs an automated Python-based data-gathering process to search over 350 areas in New Jersey using Bing.com. Articles related to each area are crawled, with their text content stored in JSON files, organized by area in separate folders.

Procedures: Each article is divided into sentences, and each sentence undergoes two labeling processes:

Fine-BERT Model: Labels sentences as Environmental, Social, Governance, or None.

Sentiment Analysis: Determines whether the sentence has a positive, negative, or neutral tone. The labels are saved alongside the sentences in the JSON files.

Results

Quantitative Data Analysis:
The analysis calculates ratios of Environmental, Social, and Governance articles and combinations like Social Positive/Negative. Three key indicators summarize patterns:

Social Negative Articles Ratio: Proportion of negative social articles in an area.

Crime Index: Crimes per capita scaled by 10,000 (from FBI data).

WSNI: Measures negative social sentiment, ranging from -40 (positive/neutral dominance) to 60 (negative dominance).

Topic Modeling:
LDA and NMF were applied to socially negative articles, identifying three categories: Local Communities, Municipal Services, and Health. Key terms for each category are displayed on the dashboard.

Topic Modeling Result Important Words		
Local Areas and Communities	Municipal Services and Emergency Management	Health and Demographics
township	fire	health
wildwood	office	people
school	municipal	average
county	community	age
city	clerk	medical
site	emergency	sex
beach	services	number
residents	department	behavior
new	planning	sexual
pittsgrove	court	population

Discussion

This project provides a comprehensive dataset and a dashboard summarizing its analysis. Key findings include Local Areas and Communities, Municipal Services, and Health being the main concerns in New Jersey. While socially negative articles are a significant indicator, they do not directly correlate with higher crime rates. The indicators can guide planning and investments to address area-specific issues.

Limitations: The project's limitations include the absence of articles or reliable crime indexes for certain areas in New Jersey and constraints in the data-gathering process due to its speed and the nature of web articles

Conclusion

This research benefits New Jersey communities by providing an interactive dashboard and highlighting key concerns like Local Communities, Municipal Services, and Health from online content. It helps address challenges using socially negative articles as indicators without oversimplifying them. Future efforts will improve real-time data collection and tools to support informed community planning.

Center for Sustainability

STEVENS
INSTITUTE OF TECHNOLOGY

PSEG

PSEG
FOUNDATION