AI-Based Modeling for Textual Data on Solar Policies in Smart Energy Applications

Sarahana Shrestha[1,4], Isabele Bittencourt[2,3], Aparna S. Varde[1,3,4], Pankal Lal[1,4]
1. Earth and Environmental Studies, 2. Computational Linguistics, 3. School of Computing, 4. CESAC
Montclair State University, Montclair, NJ 07043, USA
(shresthas1 | bittencourti1 | vardea| lalp)@montclair.edu ORCID ID: 0000-0002-9829-9607 (Shrestha), 0000-0002-3170-2510 (Varde)

2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA) | 979-8-3503-6883-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISA62523.2024.10786713

Abstract— **This study aims to examine sentiments and themes in deploying solar energy, after the recent enactment of the Clean Energy Act. It uses text on solar energy from public dockets and open access social media platforms to fathom mass opinion. It is crucial for policymakers / stakeholders in smart and renewable energy to address challenges of energy efficiency and equity. While we deal with solar energy, many claims here are valid for other sources such as offshore wind. We propose to harness the AI-based models of VADER, SentiWordNet and TextBlob for sentiment analysis, and Latent Dirichlet Allocation (LDA) for topic modeling, entailing NLP (natural language processing), to discover knowledge from the given text on prevalent attitudes concerns and endorsements. The approach dissects textual data, from assessing nuanced sentiment in social media to extracting major discussion themes. Numerical analysis on the modeled data with Pearson's correlation coefficient offers more insights. Findings reveal public reaction to solar energy policy, levels of awareness, mass acceptance, equity issues etc. This novel study underscores an interplay of solar policy, its mass perception, and renewable energy usage. It applies AI with a multi-modal holistic approach, mining textual data to help overcome the challenges of policy acceptance. It also aims to use AI to balance energy efficiency and equity, enhancing public engagement for transition to a fair, sustainable energy future for a smart planet.**

Keywords—AI in smart cities, energy equity, renewable sources, opinion mining, policy, smart energy, solar panels, topic modeling

I. INTRODUCTION

Background: The world today is at the forefront of transition to renewable energy policies and international agreements such as the Paris Agreement, which aim to reduce greenhouse gas emissions and enhance energy sustainability [1]. In alignment with these goals, the International Energy Agency projects that solar energy capacity will see an increase of over 125% in the next five years, indicating a significant shift towards this clean energy source [2]. This worldwide momentum is significantly highlighted in the United States through regional efforts, such as the enactment of the New Jersey's Clean Energy Act [3], outcomes of which appear in Fig. 1 here. This legislative framework sets ambitious goals to increase the use of clean energy, pivoting notably towards solar energy. The state has made substantial progress in this area, as evidenced by the installation of more than 5,276 megawatts (MW) of solar capacity, positioning it 9th nationally in solar installations to power 904,948 homes. 7.44% of the state's total electricity production comes from solar energy, which is almost twice as much as the national average [4]. As we intensify efforts to meet renewable energy targets including solar, offshore wind etc. the dynamics of public sentiment and societal acceptance are crucial [5]. Public sentiment and awareness play critical roles in the successful adoption of technologies, making stakeholder engagement essential in shaping effective energy policies [6]. It is vital to ensure that this transition promotes energy equity to guarantee that solutions are equitable and affordable, addressing disparities in access to such resources. Advanced analytical tools and modeling methodologies offer nuanced insights into public attitudes towards solar energy, aiding policymakers and industry stakeholders in decision-making processes [7]. Fathoming these aspects is essential to navigate potential challenges and to leverage opportunities that arise in the implementation of transformative energy policies.

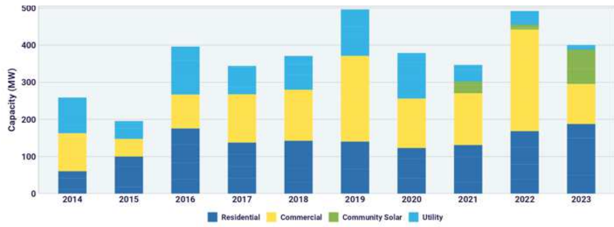

Fig 1. Solar installations in New Jersey from 2014-2023 [3].

Motivation: The transition to smart and renewable energy, involves complex and multi-faceted stakeholder engagement and public perception challenges. Fig 2. shows a series of solar panels installed at NJ's Atlantic County Utilities Authority (ACUA), captured during our own recent site visit there, showcasing the tangible application of solar technologies as a part of a global commitment to expanding renewable energy infrastructure. Previous studies have shown that public sentiment can highly influence policy effectiveness and adoption rates [5-7]. Given recent aggressive trends on smart and renewable energy goals, there is a pressing need to systematically fathom public sentiment to ensure successful policy implementation and to foster public acceptance. This aligns with progressive themes of citizen science and smart governance. Our study is motivated by the opportunity to fill gaps in understanding how the public perceives solar energy initiatives, and to explore the impact of these perceptions on future energy strategies. Additionally, a key aspect of our motivation is to address energy equity, ensuring that all communities benefit fairly from smart and renewable energy advances, not only to enhance the energy systems but also contribute to social justice and economic inclusivity.

Fig 2. Solar installations in ACUA's Solar Array Project [8].

979-8-3503-6883-3/24/$31.00 ©2024 IEEE

Problem Definition: While legislative strides have been made in promoting solar energy, there is limited empirical data on how these policies are received by the public and the broader implications for policy adherence and success. This study Our study in this paper thus has the following goals:

1. Analyze mass sentiment towards solar energy policy.
2. Identify core themes on solar energy deployment.

Approach and Contributions: We propose an approach with a holistic multi-modal analysis framework. We collate data from a variety of sources: stakeholder comments on statewide dockets, social media platforms, and press releases. Multiple AI-infused models of VADER, SentiWordNet, TextBlob, and Latent Dirichlet Allocation (LDA) are deployed for nuanced analysis on the collated data, followed by quantitative analysis via Pearson's correlation coefficient (PCC). Our work offers application-oriented AI research. The main contributions are:

- Provide AI-based sentiment analysis to highlight mass opinions and trends on solar energy policies
- Perform topic modeling with AI-guided methods to identify main themes of interest across the sources
- Compare the analyses on formal and informal data sources to gauge context-specific nature of results
- Integrate qualitative & quantitative analyses to guide stakeholders on energy efficiency, equity and more

To the best of our knowledge, ours is the 1st novel study in textual data on solar energy usage with a holistic multi-modal AI-based approach, aiming to address the key challenges of maintaining energy efficiency and equity. Our findings can be applicable to other smart energy sources, e.g. offshore wind.

II. Literature Review

Research on renewable energy technologies shows that public acceptance is significantly influenced by perceived fairness and the effectiveness of government policies [6-9]. Real-time sentiment analysis can enable more responsive governance by aligning policy implementation with public sentiment and expectations. This insight is particularly vital to understand responses to globally interesting policies, e.g. NJ's Clean Energy Act, which aims to significantly increase the use of clean energy, especially solar power. Additionally, a recent survey assessing public willingness to pay for community solar projects further illustrates the vital role of economic considerations in public acceptance of renewable energy initiatives [10]. Studies further demonstrate that state-level policies with strong incentives and regulatory support correlate positively with higher rates of solar energy adoption and favorable public opinion [11].

A recent study on sentiment analysis to gauge public opinion on offshore wind projects revealed key concerns related to environmental impacts and aesthetics [12], which are crucial insights for addressing potential barriers in implementation. In the general context of environmental management, opinion mining helps in understanding public reactions to urban projects, assisting policymakers to prioritize initiatives and enhance community engagement strategies [13]. Likewise, application of sentiment analysis extends to related avenues, as demonstrated in the air travel sector where it serves as an important tool for understanding changes in public priorities and behaviors in response to global events like the COVID-19 pandemic. Analysis of sentiment from traveler feedback before and after the pandemic identified significant shifts in public concerns towards health safety [14]. This information is important for adjusting services and policies to better align with current public expectations and concerns.

Insights into public sentiment and the multifaceted impact of smart energy policies highlight the need for further research. Our study in this paper provides valuable insights into the interplay of public opinion, policy effectiveness, and renewable energy implementation through sentiment analysis techniques combined with topic at the community and state levels. Hence, our work contributes to a growing body of literature that leverages AI-based modeling alongside policy analysis to explore the dynamics of the smart governance in the realm of equitable and accessible smart energy.

III. Data Harvesting

A. Sources of Data

In order to investigate public sentiments and thematic trends regarding solar energy adaptation, we consider the state of New Jersey as the major source of data, especially due to its high emphasis on clean energy. Comments on New Jersey's Board of Public Utilities (NJBPU) dockets [15], and comments from related Reddit [16] posts are collected. The latter ensures capturing mass sentiment from an open access social media platform where anyone can post their views (without payment), thus helping to target multiple groups including underrepresented ones and their spokespersons (who may have wider access to the Internet). Sample positive, negative and neural comments are depicted in Fig. 3 & 4 here.

NJBPU Comments: A total of 759 documents with comments are systematically retrieved from specific dockets provided by the NJBPU focusing on the period following the introduction of the state's Clean Energy Act in 2018. The dockets used in this dataset are:

1. QO18060646: In the Matter of the Community Solar Energy Pilot Program
2. QO20020184-: In the Matter of a Successor Solar Incentive Program Pursuant to P.L.2021
3. QO22030153-: In the Matter of Community Solar Energy Program
4. QO22080540-: In the Matter of New Jersey Energy Storage Incentive Program
5. QO21101186 - In the Matter of the Competitive Solar Incentive Program
6. QO19010068-: In the Matter of a New Jersey Solar Transition Pursuant to P.L. 2018, C.17

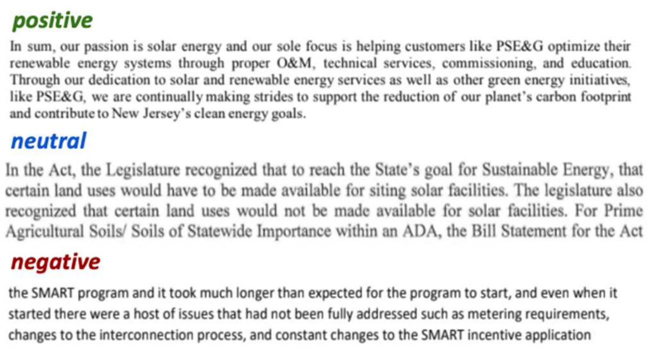

Fig 3. Sample of comments with positive, negative, and neutral sentiment from NJBPU Comments dataset.

Reddit Data: Public discussions related to solar energy are extracted from Reddit using the Python Reddit API Wrapper (PRAW), which facilitated the scraping of relevant subreddit threads. This platform is chosen because of its active user engagement and diverse range of opinions. A total of 959 comments are scrapped from January 2018 to March 2024.

These comments are extracted from various subreddit threads pertinent to solar energy discussions.

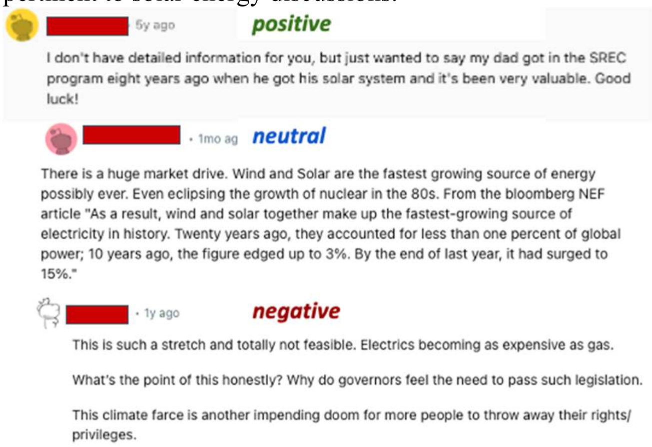

Fig 4. Sample of comments with positive, negative, and neutral sentiment from Reddit Comments dataset.

B. Data Preprocessing with NLP

Preprocessing of the data is executed in all the datasets using the Natural Language Toolkit (NLTK) [17], a comprehensive Python library designed for analyzing human language data via AI techniques by incorporating automated methods in Natural Language Processing (NLP). The preprocessing in our work involves the removal of stopwords, stemming, and lemmatization to reduce words to their base or root form. Additionally, regular expressions are employed to eliminate irrelevant characters and symbols, streamlining the dataset to consist solely of meaningful text. This is done to ensure consistency and accuracy in subsequent analyses.

IV. Methods and Models

We propose a methodology of combining qualitative and quantitative AI-based modeling on solar energy data. The qualitative methods entail sentiment analysis and topic modeling while the quantitative analysis applies the well-known measure of Pearson's Correlation Coefficient (PCC). These methods are integrated to study mass opinions about policies on solar energy, fathom the major topics of interest in the data sources, and assess differences in scores obtained from various sources. We discuss these methods next.

A. Sentiment Analysis on Solar Energy Posts

In order to systematically identify, extract, and quantify affective states and subjective information embedded within the textual data, we propose to deploy sentiment analysis in this work. A tripartite model, encompassing TextBlob [18], VADER [19], and SentiWordNet [20] is employed ensuring a comprehensive and nuanced approach. This approach serves as a validation mechanism, enhancing the reliability of the sentiment assessment by cross-verifying results across different algorithms and lexicons. The justification for each of these models is provided here.

1) TextBlob: The TextBlob tool for sentiment analysis evaluates text by aggregating sentiment scores of individual words. This provides a straightforward API for common natural language processing (NLP) tasks, including sentiment analysis. It generates polarity scores classifying comments as positive, negative, or neutral based on the numerical score, and subjectivity scores ranging from 0 to 1 [18]. We aim to use TextBlob to gauge the extent to which users express sentiments in a purely opinionated versus a more factual manner, as gathered from subjectivity scores. Additionally, the polarity scores of TextBlob are useful to assess the nature of the users' sentiments.

2) VADER: The VADER technique i.e. Valence Aware Dictionary and sEntiment Reasoner specializes in calculating normalized compound scores that reflect the overall sentiment of a text. These scores are derived from the sum of positive, negative, and neutral word evaluations and are normalized between -1 (most negative) and 1 (most positive). VADER is particularly suited for sentiment analysis on social media text, with its built-in lexicon and rule-based analysis framework adept at interpreting the nuances of text-based sentiment [19]. This explains our need for VADER; we aim to obtain compound scores that highlight the overall nature of the sentiment in a given text, going beyond the individual polarity of its terms.

3) SentiWordNet: The SentiWordNet approach leverages WordNet's lexical database, assessing words within the context of their use, utilizing 'synsets', or sets of semantically equivalent synonyms. It is able to contextually analyze semantic content allows for a context-sensitive sentiment evaluation of the collated data [20]. This is crucial in our work as we need to fathom the sentiments in given textual sources with reference to context, especially as we deal with a variety of sources ranging from highly formal to very informal across a diverse range of population demographics.

B. Topic Modeling for Themes in Solar Energy

We propose to harness topic modeling to identify themes and main keywords in large text sets. GENSIM's LDA (Latent Dirichlet Allocation) [21] model is adapted to discover latent thematic structures in text. LDA is a generative probabilistic model which assumes that documents are mixtures of topics; a topic is defined as a distribution on a fixed vocabulary. We justify the use of this model because it helps us highlight, rank and connect the major themes of interest within each dataset analyzed here. This can provide at-a-glance visualization of the interesting results in a user-friendly manner, particularly given that they are extracted from plain text including natural language with its subtleties. Human judgment is automatically embedded in this AI-infused model, thus facilitating the task of identifying key themes and topics in various data sources.

Algorithm 1: Sentiment Analysis & Topic Modeling on Solar Energy

```
INPUT: Preprocessed solar energy data δ
FOR comment α in δ DO:
    IF polarity score φ > 0 DO:
        Add positive label
    IF φ = 0 DO:
        Add neutral label
    ELSE DO:
        Add negative label
    END IF
END FOR
OUTPUT: polarity scores φ for each comment α

COMPUTE topic model with LDA
SET value of k
OUTPUT: top k values (topics) from data δ
```

Algorithm 1 has our pseudocode to systematically outline the methods and models harnessed in our proposed approach for sentiment analysis and topic modeling on solar energy data in this study. This algorithm operates on the preprocessed solar energy textual data, denoted here as δ. It integrates NLP (natural language processing) techniques to assess public opinion quantitatively and qualitatively on solar energy using polarity scores to categorize comments into positive, neutral, or negative sentiments. The LDA component for topic modeling identifies prevailing topics, enabling a deeper understanding of the thematic structures within the discourse. The k value in LDA alters granularity and comprehensiveness of the topics extracted from the data. This structured approach helps in deriving meaningful information from the textual data in the evolving landscape of smart energy systems. It intuitively captures human judgment through the AI-based modeling in the NLP techniques for preprocessing as well as the sentiment analysis and topic modeling techniques.

C. *Pearson's Correlation Coefficient for Quantification*

Pearson's Correlation Coefficient is a statistical measure of the strength and direction of a linear relationship between two continuous variables [22]. Its correlation coefficient, denoted as *"r"*, ranges from -1 to +1. The *p-value* associated with this correlation coefficient assesses whether the observed relationship is statistically significant. A small p-value (typically <0.05) suggests the correlation is likely significant and not due to chance, enhancing confidence in the results, whereas a large p-value indicates that the correlation may not reliably reflect a true relationship and could be coincidental. We justify employing PCC in this study because it can help to quantify outcomes with statistical significance. Displaying numbers as outputs is convincing to policymakers and other stakeholders who are interested in quantitative analysis; they may need numerical values to assist them in making good decisions. This also helps researchers to conduct further studies via numerical modeling.

The steps taken to calculate the PCC are as follows.

a) Sentiment Classification: Use the sentiment scores calculated using all three models [18-20] for both datasets.

b) Binning: Define bins for sentiment scores ranging from -1 to 1, with intervals of 0.1, and categorized sentiment scores into these bins for both datasets.

c) Normalization: Count the number of comments in each sentiment bin for both datasets; normalize the counts to account for different dataset sizes; convert the counts to proportions of the total counts in each dataset.

d) Data Alignment: Combine the datasets to create a new DataFrame where each row corresponds to a bin of sentiment scores, with columns for the normalized counts from both the datasets.

e) Correlation Analysis: Compute the PCC between the two sets of normalized sentiment data using the Pearson r function from SciPy's stats module [23]. Then, calculate the p-value in order to assess the statistical significance of the observed correlation.

Note that the methods proposed in this analysis with these steps can be seamlessly adapted to assess the mass opinions on other smart energy sources such as offshore wind. Based on the above methods and models in our proposed approach for knowledge discovery on solar energy, we now present a summary the results along with suitable discussion.

V. RESULTS AND DISCUSSION

A. *Sentiment Analysis Results*

The analysis of sentiment towards solar energy policies reveals that the overall sentiment is fairly positive in all three models [Fig 5.]. In the NJBPU dataset, almost all of the comments are positive, whereas the Reddit dataset shows some variation with about 59-65% positive, 17-21% negative and 13-20% neutral comments. Reddit's sentiment diversity reflects its broad, anonymous user base, while the NJBPU dataset likely consists of comments submitted in a more formal, possibly curated environment where individuals or organizations may choose to present a more positive or constructive stance, perhaps due to the official nature of the submissions. These comments may be from stakeholders who have a vested interest in promoting positive aspects of policies and to provide constructive feedback. However, Reddit is an informal platform where users may feel more comfortable expressing diverse opinions and criticisms, leading to a broader range of sentiments coming from various demographic groups, possibly including negative and neutral positions. Fig 3 & 4 show the difference in the datasets, and how NJBPU comments are more formal, whereas the Reddit ones are less formal. This is indeed a surprising and very interesting finding! The fact that the mass sentiments across a wide range of the population differ much from those of the policymakers can call for further research on a deeper level.

On another note, the consistent sentiment analysis results across three different tools, i.e., VADER, SentiWordNet, and TextBlob (see Fig. 4) reinforces the validity of the findings, suggesting reliable depiction of opinion in the analyzed data.

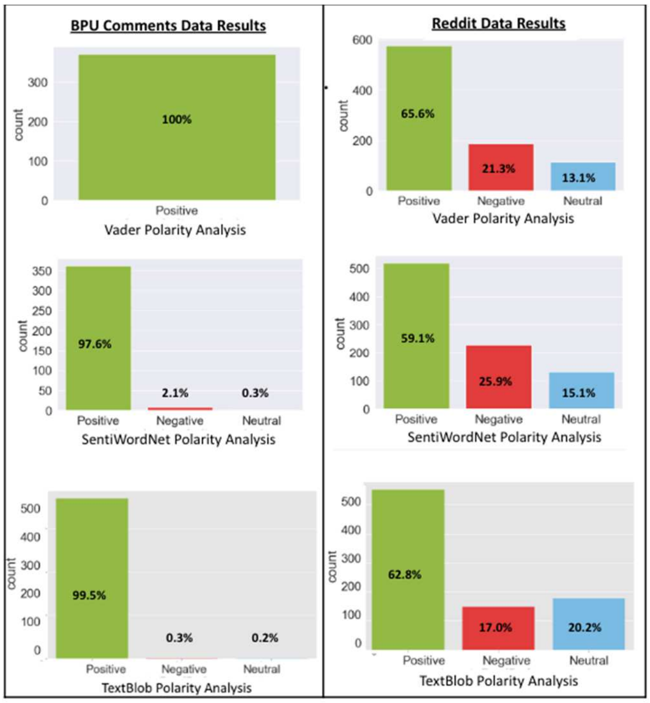

Fig 5. Comparison of Different Sentiment Analysis Models on NJBPU and Reddit Comments (Note that BPU comments are highly positive)

Moreover, Word cloud visualization of sentiment data from both NJBPU and Reddit platforms show an overlap on the major topics and keywords being used in both the datasets [Fig 6, 7]. The NJBPU dataset appears to use more formal and technical language. Some words such as "project," "community," "storage," and "incentive" hint at discussions related to the technical and policy aspects of solar energy implementation. The presence of terms like "clean energy," "support," "development," and "customer" could suggest a focus on the development and support of solar energy initiatives, possibly with a tone of professionalism and formality. On the other hand, the Reddit dataset appears to contain a wide array of terms that suggest diverse and possibly more personal opinions on solar energy. Terms like "bill," "company," "pay," and "price" suggest discussions around the cost and financial aspects of solar energy highlighting a key aspect of energy equity. This raises important questions about whether solar energy policies are adequately addressing the financial burdens on different socioeconomic groups. Words like "system," "energy," and "solar" are expected as central to the topic, but the presence of emotionally charged words like "scam" or "deal" could indicate skepticism or concerns regarding solar energy offerings. Once more, we observe a difference between the comments from the policymakers versus those by the masses.

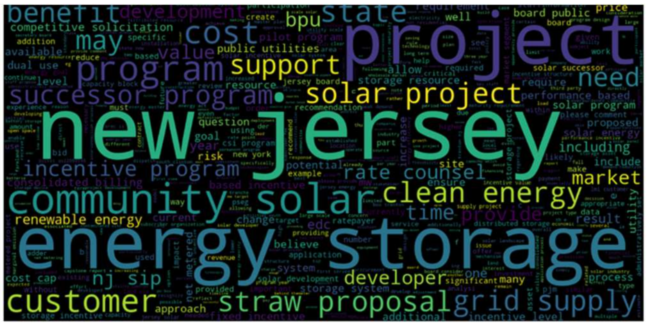

Fig 6. Word Cloud Visualization of Key Topics from NJBPU Comments on New Jersey's Solar Energy Policies.

Fig 7. Word Cloud Visualization of Key Topics Reddit Comments on New Jersey's Solar Energy Policies.

The trend from 2018 to 2024 indicates a predominance of positive sentiment through the period, suggesting a generally optimistic public view towards solar energy [Fig 8]. Positive sentiments maintain a majority, which could correlate with major policy implementations or public awareness campaigns. Notably, there is a presence of negative sentiment that, while significantly less than positive, persists across the years, hinting at consistent concerns or areas of dissatisfaction that may require attention. Neutral sentiment remains the least common, suggesting that most respondents have a definitive opinion on the matter. This is another interesting finding.

Subjectivity scores for NJBPU data and Reddit data vary as well [Fig 9 & 10]. For the NJBPU data, there is a narrower range of subjectivity scores, peaking in a central range and lacking scores close to 0 or 1. This suggests that the comments from the BPU dataset tend to be moderately subjective. They may express some personal opinions or experiences but generally stay within a more balanced or factual tone. However, Reddit data shows wider distribution of subjectivity scores with significant frequency at both extremes, indicating a mix of highly objective (scores close to 0) and highly subjective (scores close to 1) comments. This range suggests that Reddit comments vary widely, with some expressing strong opinions and emotions and others conveying facts and neutral information. This is yet another avenue of significant variation between policymakers and the common masses, which is noteworthy.

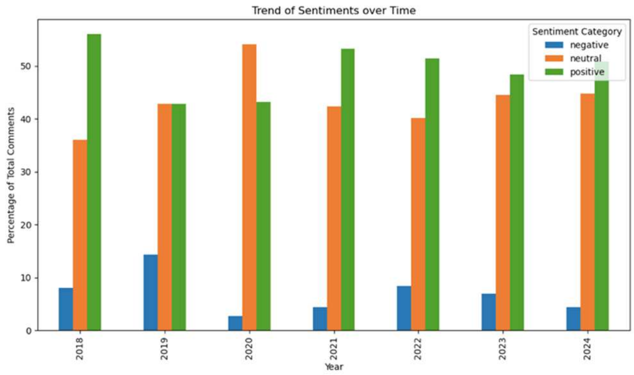

Fig 8. Trend of sentiments over time on Reddit dataset.

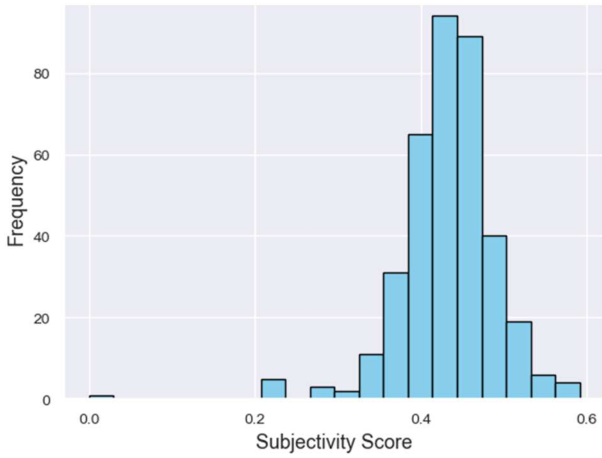

Fig 9. Distribution of Subjectivity Scores for NJBPU dataset.

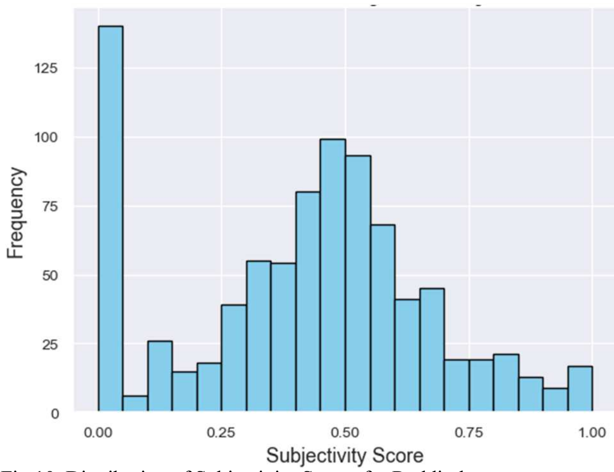

Fig 10. Distribution of Subjectivity Scores for Reddit dataset.

B. Topic Modeling Results

The LDA topic model as visualized in the inter-topic distance maps presents a structured overview of the main themes [see

Figs. 11 & 12]. Larger circles in the distance map for certain topics suggest that a higher proportion of the discourse is dedicated to those themes. A lower k value represents a "zoomed-out" view, where the topics are broader and more general. A higher k value represents a "zoomed-in" view, offering a detailed look at the range of discussions within text.

In the BPU data, topics related to "energy," "solar," "storage," and "incentive" seem particularly prominent, indicating that discussions often revolve around the technical aspects of solar energy systems, as well as the incentives provided for their implementation [Fig 11]. The frequent mention of "program," "project," "board," and "Jersey" aligns with discussions specific to New Jersey's energy programs and the governance body overseeing them. Words like "cost," "development," "capacity," and "market" suggest economic considerations and market dynamics are crucial parts of the documents.

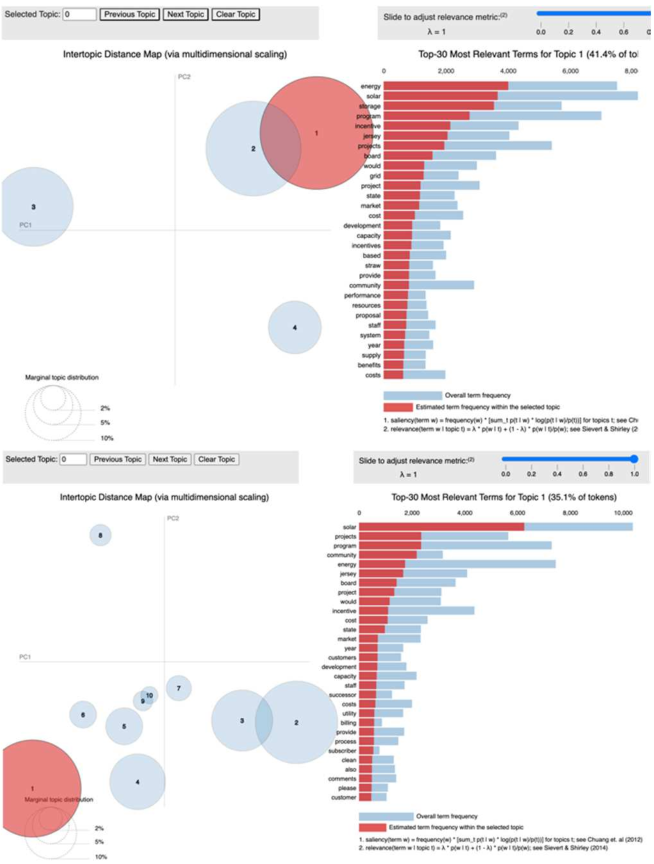

Fig.11. LDA visualization on NJBPU data, k = 4 (above) and 10 (below)

However, the results of topic modeling from Reddit data reveals distinct conversational clusters, each representing a different set of topics discussed by the users [Fig 12]. Discussions seem to center around the cost of energy, the logistics of solar panel installation, and interactions with energy companies. The presence of words such as "solar," "energy," "bill," "electric," "company," and "panels" indicates a strong focus on the economic and affordability aspects of solar energy from a consumer perspective. This highlights the broader community's concerns about the financial impact of adopting solar technology, pointing to a critical area for energy equity considerations. These and other discussions offer insights to policymakers, suggesting a need to address affordability and accessibility in the development of solar energy policies to ensure that they meet the needs of a diverse population. This is in line with energy equity.

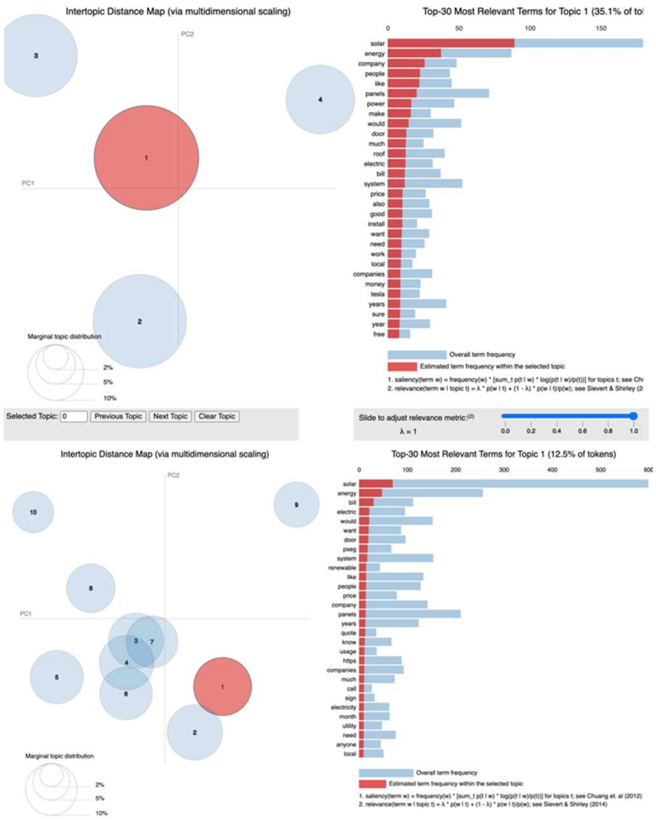

Fig12. LDA visualization on Reddit data with k = 4 (above) and 10 (below)

C. Pearson's Correlation Coefficient Results

Based on the results from PCC, it was found that there is a strong and statistically significant positive correlation between the sentiment scores from the NJBPU and Reddit datasets. This means that as sentiment scores in one dataset increase, they also tend to increase in the other dataset. This suggests that public opinions or sentiments expressed on the more formal NJBPU platform align closely with those expressed on the more informal Reddit platform, at least in the context of the topics analyzed. It is an interesting finding despite the fact that there are highly more positive sentiments in the NJBPU data than the Reddit data.

Table 1 presents outcomes of Pearson's correlation analyses conducted on sentiment scores derived from three sentiment analysis models: TextBlob, VADER, and SentiWordNet. Results show strong positive correlation between TextBlob sentiment scores from the two datasets. The statistical significance of this correlation *($p < 0.001$)* underscores a reliable alignment in sentiment trends across the platforms, suggesting that TextBlob effectively captures consistent sentiment patterns. Results from SentiWordNet are also statistically significant, indicating a consistent sentiment correlation across all the datasets. VADER sentiment scores display weak positive correlation, not statistically significant *($p > 0.05$)*. This outcome suggests that VADER may respond differently to the linguistic nuances of each platform.

Analysis with such AI-based models is applicable to offshore wind energy [12] and other sources, aiding decision-making.

TABLE I. PEARSON'S CORRELATION COEFFICIENT RESULTS FOR SENTIMENT ANALYSIS ACROSS ALL THE MODELS

Model	*PCC (r)*	*p-value*
TextBlob	0.77	6.30 e-05
VADER	0.20	0.40
SentiWordNet	0.97	1.12 e-12

VI. CONCLUSIONS AND FUTURE WORK

Our comprehensive study with AI-based models on textual data in solar energy reveals how public sentiment and policy effectiveness shape the landscape of implementation. Major findings of this applied-AI study are listed below.

1) Discourse around solar energy on a formal platform (NJBPU) seems to be centered more on themes about policy specifics, program implementation, and technical details.

2) Discussions on the open access social media platform (Reddit) reflect more consumer-centric themes with practical concerns on affordability, reliability, and customer service.

3) Variation in sentiment between formal and informal platforms (in this case highly positive in NJBPU versus mixed in Reddit) reinforces the fact that comprehensive analysis should consider multiple varied sources to capture the full spectrum of diverse public opinion among the masses.

4) Analysis of models revealing topics on "cost," "access," and "affordability" can reveal vital issues on energy equity, highlighting the need for solar energy policies to advance sustainability as well as promote fairness in energy access.

5) Fluctuations in sentiment discourses year after year can reflect evolving public response to temporal advances in energy policy, or to broader socioeconomic events.

6) Numerical analysis of the modeled data with Pearson's correlation coefficient shows a considerably good & reliable alignment of opinion trends with high statistical significance.

Based on these findings, key takeaways of this research from an application standpoint are as follows. In order to address a broad range of opinions on energy, policymakers should consider inclusive approaches. Community forums, online surveys, and social media engagement can be helpful [24, 25]. There should be a focus on encouraging open dialogue and providing spaces for dissent and discussion, especially in communities disproportionately affected by the energy burden. This is in line with other studies on equity as well [26-28]. Continuous monitoring of trends should be done to track shifts in public discourse and inform adaptive policy strategies that respond to public needs and sustain the momentum of positive perception. Sentiment analysis can go hand-in-hand with policy implementation in smart energy initiatives, as in related work, e.g. on offshore wind [12] where equity is important as well. Analysis via AI-based methods and models in smart energy initiatives such as solar energy and offshore wind is not only insightful but also indispensable today, as we gear towards smart governance and maintain a smart environment. For instance, positive opinions on such initiatives can propel the growth of energy incentives for the concerned residents. Negative opinions or concerns, e.g. timeliness of programs, feasibility of solar panels, wildlife impacts due to offshore wind sources etc. can help the governmental agencies conduct improvements and make better policy decisions for the future. Our study is application-oriented research focusing mainly on a region in the Northeastern United States. It can well exemplify citizen science and emphasize the role of AI in global applications for smart energy systems, smart cities, and a smart planet. Future work emerging from this research can potentially aim to answer the following questions.

1) Does the mass sentiment along policy developments in solar technology undergo a significant shift over time and place, i.e. are temporal / spatial aspects highly crucial?

2) In what manner can the insights from this study be leveraged to promote equitable access to solar energy across different communities, as per environmental justice (EJ)?

3) How do solar energy posts (from formal / informal sources) compare with those on other smart energy sources (nuclear, offshore wind [12]), though analyses are similar?

4) Is the content of the posted text genuine with respect to data veracity issues; in other words, is there a tendency to spread fake news on any platforms?

5) Can other AI models and methods, e.g. LLMs (large language models) including BERT (Bidirectional Encoder Representations from Transformers) and GPTs (generative pretrained transformers) [29] do better?

6) Would commonsense knowledge in AI, e.g. cultural commonsense [30], enhance such research?

The comprehensive analysis in this paper, focusing on smart energy, also supports the development of smart governance and smart environment practices which can thrive much on some of our earlier research in this area [9, 31], including that on smart manufacturing for energy efficiency [32, 33]. Such work promotes adaptive policies in dynamically responding to rapid technological advances as well as shifts in public attitudes. Hence, it gears towards ensuring that urban development is sustainable as well as responsive to citizens' needs. It highlights the need for energy equity, related to EJ, i.e. environmental justice [34]. This aims for energy-related technological advances and policy implementations that do not exacerbate existing disparities. It is ethical as well as beneficial to discover knowledge about the needs of multiple sectors in the overall population demographics to head more towards energy equity. A crucial step for this knowledge discovery can be the analysis of mass opinions from residents across the locale, including those from underrepresented groups and their spokespersons. Much of this data is available on open access social media platforms (e.g. Reddit). We have utilized it in our study, hence emphasizing the broader impacts of this work. In the future, we aim to address larger data sets including more diverse sources of public opinion providing a broader perspective on the underlying matter. Ultimately this research with its multifaceted approach helps understand the public sentiment and guide more informed and effective policy-making in the future to promote inclusive growth as part of the smart energy transition. It examines the role of AI-based methods in this venture, paving the way for more multidisciplinary research. This is analogous to other work on data mining and information retrieval including some of our own group's research [35 – 37]. It is also in line with AI for smart energy / smart cities [38 – 42]. *The novelty of our study in this paper is that it is among the 1st to integrate multiple AI-based models, leveraging NLP with qualitative* *and quantitative analysis for analyzing text on solar energy, addressing energy efficiency and equity in policy acceptance.* Hence, the work in this paper can make modest impacts on AI for smart energy and sustainability.

ACKNOWLEDGMENT

Sarahana Shrestha is a Doctoral Assistant in the Environmental Science & Mgmt. PhD program at Montclair State University. Isabele Bittencourt is supported by the NJ Wind Tech Fellowship of NJEDA. Dr. Aparna Varde acknowledges NSF grants 2018575 and 2117308. She is an Associate Director of the Clean Energy & Sustainability Analytics Center (CESAC). Dr. Pankaj Lal is the Founding Director of CESAC at Montclair State University.

REFERENCES

[1] United Nations Framework Convention on Climate Change (UNFCCC), "Paris Agreement," https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement, 2015.

[2] International Energy Agency (IEA), "Tracking Clean Energy Progress 2023," IEA, Paris https://www.iea.org/reports/tracking-clean-energy-progress-2023, Licence: CC BY 4.0, 2023.

[3] New Jersey's Energy Master Plan. 2019. Retrieved from https://nj.gov/emp/.

[4] New Jersey Solar. Solar Energy Industries Association. 2024. https://www.seia.org/state-solar-policy/new-jersey-solar.

[5] D. Kocakusak, J. Senick, and C.J. Andrews, "Implementing the energy transition: lessons from New Jersey's residential solar industry," Climate policy, 2024, 24(5), 646-659.

[6] J.E. Carlisle, S.L. Kane, D. Solan, M. Bowman and J.C. Joe, "Public attitudes regarding large-scale solar energy development in the US," Renewable and Sustainable Energy Reviews, 2025, 48:835-847.

[7] M. Zhang, H. Su, and J. Wen, "Analysis and mining of internet public opinion based on LDA subject classification," Journal of Web Engineering, 2021, 20(8): 2457-2472.

[8] Atlantic County Utilities Authority, "Solar Array Project Fact Sheet," https://www.acua.com/ACUA/media/Acua/uploadedfiles/Solar-Array-Project-Fact-Sheet-2024.pdf, 2024.

[9] M. Puri, A. Varde, X. Du, and G. De Melo, "Smart governance through opinion mining of public reactions on ordinances," IEEE Intl. Conference on Tools with Artificial Intelligence, 2018, pp. 838-845.

[10] S. Oluoch, P. Lal, A. Bevacqua, and B. Wolde, "Consumer willingness to pay for community solar in New Jersey," The Electricity Journal, 2021, 34(8): 107006.

[11] W. Kimberly S., P.C. Stern, and T. Dietz, "Explaining interest in adopting residential solar photovoltaic systems in the United States: Toward an integration of behavioral theories." Energy research & social science, 2017, 25: 134-151.

[12] I. Bittencourt, A. S. Varde, and P. Lal, "Opinion Mining on Offshore Wind Energy for Environmental Engineering," IEMTRONICS Conference, Springer Nature, 2024, pp. 252-264.

[13] X. Du, M. Kowalski, A. S. Varde, G. de Melo, and R. W. Taylor, "Public opinion matters: Mining social media text for environmental management," ACM Sigweb 2019, Autumn (2020): 1-15.

[14] F. Avery, A. Varde, and P. Lal, "Sentiment Analysis and Topic Modeling for Public Perceptions of Air Travel: COVID Issues and Policy Amendments," 13th Language Resources and Evaluation Conference (LREC), Workshops, 2022.

[15] New Jersey Board of Public Utilities (NJBPU), https://publicaccess.bpu.state.nj.us/, 2024.

[16] Reddit, Home Page, https://www.reddit.com/, 2024.

[17] S. Bird, E. Klein, and E. Loper, "Natural language processing with Python: analyzing text with the natural language toolkit," O'Reilly Media, Inc., 2009.

[18] S. Loria, textblob Documentation. Release 0.15, 2018, 2(8): 269.

[19] H. Clayton J, and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Eighth International AAAI Conference on Weblogs and Social Media, ojs.aaai.org/index.php/ICWSM/article/view/14550, 2014.

[20] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," LREC 2010, Vol. 10.

[21] Ř. Radim, and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," LREC Conference, Workshop on New Challenges for NLP Frameworks, 2010.

[22] P. Sedgwick, "Pearson's correlation coefficient," BMJ, 45(jul041):e4483–e4483, https://doi.org/10.1136/bmj.e4483, 2012.

[23] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, D., … and P. Van Mulbregt, "SciPy 1.0 Contributors," SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 2020, 17(3):261-272.

[24] O.B. Driss, S. Mellouli, and Z. Trabelsi, " From citizens to government policy-makers: Social media data analysis," Government Information Quarterly, 2019, 36(3):560-570.

[25] N.J. Adams-Cohen, "Policy change and public opinion: Measuring shifting political sentiment with social media data," American Politics Research, 2020, 48(5):612-621.

[26] F. Fang Yu, M. Alharthi, Z. Bhatti, L. Sun, F. Rasul, I. Hanif, and W. Iqbal,"The dynamic role of energy security, energy equity and environmental sustainability in the dilemma of emission reduction and economic growth," Journal of Env. Mgmt. 280: 111828, 2021.

[27] S. Cong, D. Nock, Y.L. Qiu, and B. Xing, "Unveiling hidden energy poverty using the energy equity gap," Nature communications, 2022, 13(1):2456.

[28] M.A. Brown, A. Soni, M.V. Lapsa, K. Southworth, and M. Cox, "High energy burden and low-income energy affordability: Conclusions from a literature review," Progress in Energy, 2020, 2(4):042003.

[29] Yang, B., Luo, X., Sun, K., & Luo, M. Y. "Recent progress on text summarisation based on BERT and GPT", Intl. Conf. on Knowledge Science, Engineering & Mgmt. 2023, pp. 225-241.

[30] T. P. Nguyen, S. Razniewski, A. S. Varde, and G. Weikum, G. "Extracting cultural commonsense knowledge at scale". In Proceedings of the ACM Web Conference, WWW 2023, pp. 1907-1917.

[31] M. J. Pawlish, and A. S. Varde, A.S., "A decision support system for green data centers", ACM Workshop on Ph. D. students in information and knowledge management, ACM CIKM 2010, pp. 47-56.

[32] C.J. Conti, A.S. Varde, and W. Wang, "Human-robot collaboration with commonsense reasoning in smart manufacturing contexts", IEEE Transactions on Automation Science and Engineering (TASE), 2022, 19(3), pp.1784-1797.

[33] Conti, C. J., Varde, A. S., & Wang, W. "Robot action planning by commonsense knowledge in human-robot collaborative tasks", IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1-7.

[34] US EPA, https://www.epa.gov/environmentaljustice

[35] Kaluarachchi, A., Roychoudhury, D., Varde, A. S., & Weikum, G. "SITAC: discovering semantically identical temporally altering concepts in text archives", ACM Intl. Conf. on Extending Database Technology (EDBT), 2011, pp. 566-569.

[36] Varde, A., Rundensteiner, E., Javidi, G., Sheybani, E., & Liang, J. "Learning the relative importance of features in image data", IEEE Intl. Conf. on Data Engineering (ICDE) workshop, 2007, pp. 237-244.

[37] Varde, A. S. "Challenging research issues in data mining, databases and information retrieval", ACM SIGKDD Explorations Newsletter, 2009, 11(1):49-52.

[38] A. I. Arvanitidis, L. A. Valdez, and M. Alamaniotis, "A Quantum Machine Learning Methodology for Precise Appliance Identification in Smart Grids", 14th Intl. Conf. on Information, Intelligence, Systems & Applications (IISA), Volos, Greece, 2023, pp. 1-6.

[39] E. Sarmas, V. Marinakis, and H. Doukas, "Correction: A data-driven multicriteria decision making tool for assessing investments in energy efficiency". Oper Res Int J, 2023, 23:21.

[40] Pawlish, M. J., Varde, A. S., and Robila, S. A. (2015). The greening of data centers with cloud technology. International Journal of Cloud Applications and Computing (IJCAC), 5(4), 1-23.

[41] N. G. Bourbakis, I. P. Ktistakis, T. Seleem, "Smart Cities - Detecting Humans in Regions of Disasters: Synergy of Drones, Micro-robots in Underground Tunnels". Int. J. Artif. Intell. Tools, 2020, 29(5):1-22.

[42] K. Sorri, K. Yrjönkoski, and M. Seppänen, M. "Smart cities, smarter values: Unpacking the ecosystem of urban innovation", Technology in Society, 2024, 77:102499.