Review of: "A Comparative Analysis of Advertising in the 2020 Presidential Elections & Phoenix Mayoral Elections using Natural Language Processing"

Alaa Fareed Abdulateef[1]

1 Universiti Utara Malaysia

Potential competing interests: No potential competing interests to declare.

1. Abstract :

   While the abstract provides a succinct overview of the study's objectives and methods, it lacks specific details about the findings and implications of the research. To enhance the abstract, it should include a brief summary of the key results and their significance, offering readers a clearer understanding of the paper's contributions.
2. The introduction adequately introduces the research question and provides context for the study. However, it lacks a comprehensive literature review that synthesizes existing research on political advertising, microtargeting, and related topics. Incorporating relevant theoretical frameworks and empirical studies would strengthen the introduction and provide a more robust foundation for the research.
3. The section on related work briefly mentions a prior study on political microtargeting but fails to engage with a broader range of literature on the topic. To improve this section, the authors should conduct a more comprehensive review of relevant research, discussing key findings, methodologies, and theoretical perspectives. This will demonstrate a deeper understanding of the research context and contribute to the paper's academic rigor.

Suggestions for Improvement:

- Critically evaluate the strengths and weaknesses of previous studies to highlight gaps or inconsistencies in the literature. For example, the literature review lacks discussion on recent advancements in microtargeting techniques and their implications for political advertising. Case studies highlighting the impact of microtargeting strategies in recent elections, such as the 2020 US presidential campaign or the Brexit referendum, could provide valuable context (Bovet & Makse, 2019).
- Explore the historical evolution of political advertising and the role of social media platforms like Facebook in shaping electoral campaigns.
- Discuss theoretical perspectives from communication studies, political science, and computational linguistics to provide a multidisciplinary framework for understanding political microtargeting.

- To better integrate existing research, the authors could incorporate recent studies by Kruikemeier et al. (2020) and Stroud (2019), providing insights into the role of social media in political communication.
- Drawing from theoretical frameworks such as the elaboration likelihood model (Petty & Cacioppo, 1986) and the spiral

of silence theory (Noelle-Neumann, 1984) would offer a deeper understanding of how microtargeting shapes voter attitudes and behaviors.

4. The description of data acquisition methods is sufficient, but the authors should provide more details about the dataset obtained from the Facebook Ad Library API. Specifically, they should clarify how the dataset was processed and any limitations associated with the data collection process.
5. While the authors outline their data cleaning process, they do not thoroughly address potential sources of bias or error in the dataset. To address biases, the authors could consider implementing propensity score matching techniques to balance the distribution of covariates across different demographic groups (Austin, 2011).

- Utilizing Bayesian statistical models, such as hierarchical Bayesian models, can help account for uncertainty in estimates and improve the robustness of the analysis, especially in the presence of sparse data or small sample sizes (Gelman et al., 2013). A more comprehensive discussion of data quality issues, such as missing values, outliers, and sampling biases, is needed to ensure the validity and reliability of the findings. Additionally, the authors should explain how they addressed these issues in their analysis to mitigate their impact on the results. The authors should conduct a thorough analysis to identify and mitigate potential biases in the dataset. This could involve examining demographic distributions and comparing them to census data to ensure representativeness.
- Implementing techniques like data augmentation or oversampling can help mitigate biases, especially if certain demographic groups are underrepresented in the dataset (Chawla et al., 2002).
- Employing adversarial training approaches, such as adversarial debiasing (Zhang et al., 2018), can also help mitigate bias by explicitly training the model to disregard sensitive features that could lead to biased predictions.

4. The methodology lacks justification for the choice of specific NLP techniques and machine learning algorithms. The authors should articulate the rationale behind their methodological decisions and discuss alternative approaches considered. Additionally, they should provide more information about model validation and performance evaluation to assess the robustness of the findings. The article lacks detailed descriptions of the data preprocessing steps. For instance, it's unclear how missing values were handled or how text data was transformed into numerical format. Providing step-by-step explanations or code snippets in an appendix would enhance clarity.

- Additionally, the description of model training techniques, especially for SVM and RNN, is somewhat vague. The authors could clarify which specific hyperparameters were tuned during training and how cross-validation was performed to ensure model generalization.

Suggestions for Improvement:

- Provide a clear and concise overview of each stage of the methodology, including data acquisition, preprocessing, and model training.
- Include specific examples of data preprocessing techniques and rationale behind their selection.
- Specify hyperparameters used in model training and the process of model evaluation to enhance reproducibility.

- Employ machine learning algorithms like Random Forest and Gradient Boosting to enhance the robustness of the

analysis (Chen & Guestrin, 2016).

- Implement advanced NLP techniques such as Transformer models (Vaswani et al., 2017) and pre-trained language models like GPT (Radford et al., 2018) to extract nuanced sentiment and thematic patterns from political advertisements.

7. The authors briefly mention the issue of overfitting in their analysis but do not provide a comprehensive discussion of how they addressed this challenge. To enhance the paper, they should describe techniques used to prevent overfitting, such as regularization methods or cross-validation. Additionally, they should provide evidence of model generalization by testing the performance of the models on independent datasets.
8. While the authors present the results of their analysis, they do not provide sufficient interpretation or discussion of the findings. A more detailed analysis of the results, including comparisons with previous research and theoretical implications, is needed to contextualize the findings within the broader literature. Additionally, the authors should acknowledge any limitations or uncertainties in the data and discuss their implications for the validity of the conclusions.
9. overall comments there are several significant weaknesses and limitations:

   1. Lack of Clarity: The paper lacks clarity and coherence in several sections, making it difficult to follow the flow of ideas and understand the methodology clearly.

2. Inadequate Data Analysis: The data analysis seems superficial and lacks depth. There is no detailed explanation of the methods used for data analysis, making it challenging to assess the validity of the findings.

3. Limited Scope: The paper focuses primarily on describing the process and outcomes of using NLP techniques for analyzing political advertisements. However, it lacks a comprehensive literature review, theoretical framework, or discussion of broader implications.

4. Methodological Issues: The paper fails to address important methodological issues such as data cleaning, feature selection, model validation, and generalizability. There is also no mention of ethical considerations or potential biases in the data analysis process.

5. Weaknesses in Model Training: The paper briefly mentions using SVM and RNN models but provides limited details on model architecture, hyperparameter tuning, or performance evaluation metrics. Additionally, the reported accuracy scores for both models seem implausible and raise questions about the validity of the results.
Several revisions are necessary:

1. Clarify the research question and objectives.
2. Provide a comprehensive literature review to contextualize the study within existing research.
3. Clearly articulate the methodology, including data collection, preprocessing, modeling techniques, and evaluation metrics.
4. Conduct a rigorous data analysis with detailed explanations of findings and interpretation.
5. Address methodological limitations and ethical considerations.

6. Improve the quality of writing, including grammar, structure, and coherence.
7. Provide additional details on model training, validation, and performance evaluation.
8. Ensure transparency and reproducibility of the study by sharing code and data.
9. And finally, the structure of the paper needs to be more organized to keep the flow of the paper neat and clear.