Spring 2021

The Role of Social Media in Policy Formulation Improvement in California

Khoi Minh Nguyen
San Jose State University

The Role of Social Media in Policy Formulation Improvement in California

by

Khoi Minh Nguyen

A Thesis Quality Research Project
Submitted in Partial Fulfillment of the
Requirements for the
Master's Degree
in

PUBLIC ADMINISTRATION

Professor Frances Edwards. Ph.D.
Adviser

The Graduate School
San Jose State

May 2021

Table of Contents

INTRODUCTION

The Problem

The volume of public opinion presented by social media can inform policy formulation and evaluation to a great extent, with a lot of political, social and economic activity happening over the internet. Data on social networks is being used by a lot of people to provide insights into various aspects of public policy, while the private sector has increasingly used social media to understand consumer behavior and use this information to make business decisions. Social media presents the government with an opportunity to overhaul its processes and enhance its understanding of the people and the impact of policies on the society in general, yet this has not always been the case (Leavey, 2013).

The tremendous growth in the use of social media makes it a necessary component of the policy making and evaluation process, yet most studies have focused on the effects of social media and not its relationship with policy formulation and evaluation (Lant, 2014). Big data is currently changing the way information is collected, stored, organized and processed, and this technology presents governments and government agencies with great opportunities for policy making on matters concerning innovation, productivity and economic growth. Governments could embrace the use of big data in policy making with regards to education, healthcare and security, yet this has not always been the case (Williams, 2019).

Past research confirms that the use of research-based evidence in policy formulation has often failed due to disparities between the state decision making process and the differing priorities of policy and science, coupled with barriers in the political process used by states in formulating policies. These barriers include political priorities, constraints of time, information overload, institutional features, budgetary constraints and insufficient evaluation skills (Payán &

Lewis, 2019). The emergence of technologies such as machine learning and artificial intelligence make it possible to obtain critical insight regarding the perceptions and attitudes of individual citizens and inform the policy making process, yet this has not been the case in California (Nemes & Kiss, 2020). This study assesses how staff members of the California State Assembly could use social media to improve policy formulation in the state.

Purpose of the Research Question

How could staff members of California Assembly members use social media to improve policy formulation? The purpose of this study is to assess the role that social media platforms such as Twitter and Facebook could play in the formulation and evaluation of policies in the state of California, given the increasing application of big data in decision making in the private sector. This study further evaluated how emerging technologies such as machine learning and artificial intelligence could be used to determine the attitudes and perceptions of the citizens of California, specifically on policy issues, and analyze how these technologies could be used by California as a means of gaining useful insights, gauging sentiments and collecting data, prior to formulating and evaluating laws and policies.

Currently, the state legislature, state departments, departmental committees and public sector agencies are the leading policy making bodies in California, alongside organized groups made up of citizens, electronic media, and public figures such as political consultants (California State Senate, 2019). The state's constitution requires that the authority be shared with voters by using their sentiments throughout policy initiation, recall and referenda. Ideas come from a variety of sources, among them scientific research on different aspects, and state assembly staff analyze these issues, meet with lobbying groups, who therefore influence the process (Cahn & Schockman, 1996).

The legislative process starts with an idea from either an individual or group of individuals who want certain socioeconomic or political aspects legislated, and they approach a member of the legislature to craft a bill. The bill and the language to be used are then submitted by the interested member to the legislative counsel's office, where attorneys draft it into an actual bill. The legislative counsel returns the crafted bill to the legislator who had received its idea, who may give it back to the originators for review before taking it to the policy committee. Bills can only be heard by the committee some thirty days after being received and being on the Daily File for four straight days. Bills that have financial implications are then taken to the fiscal committee. After discussions by committees, the bill is returned to the floor of the Assembly, where if it passes, it goes to the State Senate. Bills that successfully pass through the State Senate are then taken to the governor for signing into law (California State Assembly, 2020).

Sources reveal that sponsored bills are becoming more common in the California Assembly, with lobbyists, advocacy groups and private industries advocating for policy change. They craft bills and introducing them to the Assembly through a legislator who is willing to support their positions (Sá, 2010). The use of big data technologies can improve the efficiency of this process, make policy making more citizen-centric, enhancing the speed of data collection, increasing the volume and quality of data collected, answering questions on what the citizens want in a more efficient way. The use of public opinion that has been expressed through social media could lead to higher confidence about the public acceptance of the policies being formulated, empowering all sectors of a community to contribute to policy development, and creating data-driven governance (Williams, 2019).

BACKGROUND

There is dramatic change regarding the environment within which public policies are formulated and evaluated all over the world. Digital technologies, such as social media platforms and the internet in general, are providing traces of information from the public that can be harnessed through big data technologies. These technologies provide a chance for the process of policy making, implementation and evaluation to be more citizen focused, while taking into account the needs of the citizens, their experience with public service and their preferences when it comes to policy making, through information recorded on social media networks such as Facebook and Twitter (Margetts, 2013).

California can use big data science in its policy formulation and evaluation process. The state's Assembly is comprised of eighty members and forms the lower house of the state's legislature. Each of the eighty members currently represents close to half a million people, making the state have the largest ratio of population per every representative. The Assembly sits in Sacramento and is presided over by the Speaker of the Assembly, who is nominated by the majority party and elected by the full state Assembly. Assembly members serve for 2-year terms with a maximum of 12 years of service. All the eighty members are therefore subject to an election every two years. Apart from the elected members, the Assembly also employs the chief clerk, who is responsible for administrative duties and serves as the custodian of all bills, and publishes the minutes of all the Assembly's sessions (California State Assembly, 2020).

According to the National Collaborating Center for Healthy Public Policy (2013), various assembly committees and subcommittees that deal with different socioeconomic issues oversee policy issues in terms of formulation, implementation and evaluation. The process of formulating policy begins with agenda setting, where issues are first acknowledged to be of public interest,

become the subject of discussion, with different groups identifying the problem, proposing solutions, and engaging in activities that are likely to pressure the government into intervening. The public agency concerned with the problem then evaluates different options considered as possible solutions and prioritizes a specific policy direction before adopting one of the proposed solutions (National Collaborating Centre for Healthy Public Policy, 2013).

California's Assembly therefore works with different organizations to develop policies. For the most part, the decisions by policy makers regarding policies are informed by community attitudes and behaviors that are often determined through quantitative research that involves labor-intensive surveys or interviews. Research results in the form of factsheets and in-depth reports are then used to solve the many complex issues that the state faces. The process of gathering information continues during policy implementation, as authorities look for ways of improving what has been formulated, and what is reflected in public sentiments through statewide surveys conducted by groups like the Public Policy Institute of California (PPIC, 2020).

Recent studies reveal that social media platforms have increased the capacity of different players to shape the policy agenda of their countries, with the social media handles of politicians and political parties playing a role in what the mainstream media reports, especially on political matters. An analysis of tweets, comments, trends and hashtags has been used in the past to effectively forecast what matters to individuals and groups at any one given time (Gilardi, Gessler, Kubli & Muller, 2020). Data mined from these platforms and their administrative operation can help state and government agencies to monitor and improve their own performance by getting insights into what people are discussing about governance, and the policies that are

attracting negative publicity and complaints, by soliciting data from their own sites or those sites operated by social enterprises (Margetts, 2013).

Social media is increasingly being used to shape public policy (Rehr, 2017). The author cites the increased usage of social media platforms for advocacy, as well as the extensive use of Twitter by President Donald Trump to further underline this importance, while noting that little objective research has been carried out to gauge the role of social media in advocacy and policy evaluation. Social media data scraping tools such as Octoparse, Datahut, Mozenda, Apify, Import.io and ParseHub can be used to predict trends, keep track of what California residents think, and efficiently automate data collection from social media platforms, with a view to improving the formulation and evaluation process, as well as government performance (Agrawal, 2019). Studies have shown that web scraping methods are used to search for information, combine such information from different data sources and present based on user preferences (Dewi, Meiliana & Chandra, 2019).

Research reveals that epistemic practices such as web scraping and analysis of big data have in the recent past been used by the elite to represent citizen opinion and legitimize the process of making and evaluating policies (Ulbricht, 2019). The author refers to this process of collecting information about citizen sentiments on public policy as demo scraping and claims that the big data analysis and web scraping have the potential to reduce the gap between the citizens and the political or bureaucratic elite, while addressing the paradox in democratic societies of increasing the expectations of the citizens while avoiding a legitimacy crisis in policy formulation, implementation and evaluation. Ulbricht's views are consistent with the views of Rehr (2017) as well as Gilardi, Gessler, Kubli and Muller (2020).

There is need to move away from traditional research methods and place extra emphasis on technological infrastructure and data analytic tools in the monitoring and evaluation of data by policy institutes around the world. Evidence based public policy would enhance the understanding of societal issues, and lead to the design of interventions that are better targeted, while avoiding haphazard decision making, overcoming individual biases and overcoming a political agenda that is counterproductive (Bhattarai and Adhikari, 2020). The authors also explain that using evidence from data analytic tools introduces transparency and accountability into the policy decisions of the government, provided that this data is collected in an open, trustworthy, comprehensive and transparent manner.

Experts suggest that the use of social media in public policy formulation would also enable more responsive public service delivery, and early warning about possible failures in transport, security, healthcare, law enforcement, as well as social care, while saving on cost and improving accuracy (Leavey, 2013). Some authors feel that public servants involved in governance, policy making, implementation and evaluation do not fully appreciate the role that data science can play in their work as their counterparts in the private sector do, yet data science offers a framework that is powerful enough to improve service delivery by expanding current understanding of policy choices (Engler, 2020). The author explains that the analysis of networks, texts and images as well as predictive analytics, big data methods and clustering can be used to glean insights into policy studies, while cloud computing and programming languages can be used in data collection and information storage.

Organizations can either pay a developer to line up an in-house web scraper at a fee, or use already existing web scraping companies that are available worldwide (Deysarkar, 2017). According to the author, the entire one-off cost of developing an in-house web scraper is

$12,500, and a recurrent expenditure of $1,000 monthly. This cost is prohibitive for little business organizations, making the utilization of already existing web scrapers a better alternative. Some freelancers offer the service for as low as $99 initial set up fee and $79 monthly fee for maintenance (Garcia & Perez, 2019).

Companies like Zyte, formerly referred to as Scrapinghub, scrape an estimated 8 billion pages monthly, from which they collect information for the businessmen, especially e-commerce stores that require price intelligence (Zyte, 2018). Web scraping has become easier and cheaper, with open-source programs available to be used by individuals and corporations that need the technology. According to Parsehub, there are free web scraping tools online, with freelancers offering the service for as little as $10 per website scraped (Garcia & Perez, 2019). According to the authors, these freelancers use automated bots or tools that collect information from websites and put in a simple to interpret format like xml, excel or JSON.

Garcia and Perez (2019) explain that web scraping software from companies like Parsehub are often employed by clients who have no technical ability, allow maintenance support from the developers of the software, are often maintained by the clients without necessarily contacting the developers, are cheaper, and can scrape through multiple sites, with the downside being that some ready-made software cannot scrape through complex websites. It also requires time, and should need help from technicians if advanced features are to be used. The authors further explain that the demand for web scraping technology is currently high, revealing its popularity in research and market intelligence.

The duration of web scraping is additionally extremely rapid, with sources revealing that some web scrapers can handle more than 40,000 requests a day, averaging one complete request every two seconds (Zyte, 2018). According to the article, this allows organizations to collect

information from many people within a short time, provided the data is publicly available, with organizations like Facebook having issues with collection of private information (Hutchinson, 2020).

LITERATURE REVIEW

Social Media and Policy Making

Social networks offer policy makers the ability to access a wide range of information, such as opinions, direct messages, debates and raw material for policy making that is wider in scope than mainstream media, which indicates how useful they can be in the policy making process (Lant, 2014). This broad use of social media, regardless of whether requested or spontaneous, offers the residents of a state, for example California, an enormous open door for associating with policy making (Bekkers and Edwards, 2017). The authors point out that social media networks enable many more people to participate in policy making, a greater opportunity for ideas to be exchanged, as well as an opportunity for knowledge and arguments to be advanced.

The magnitude of social media usage, with some networks like Facebook registering almost a billion daily users, with 20% of these in USA and Canada, is too huge to ignore (Lant, 2014). Lant (2014) further points out that the size and diversity of the population on social media platforms cannot be underestimated and calls for an evaluation of the role of these platforms on policy making. With recent advances in computing and modelling, data collection from social media networks has become easier, making it easier for these networks to inform policy formulation and evaluation (Battaglini and Patacchini, 2019). Kouam (2019) calls on policy makers to take a more technocratic approach to formulation, implementation and evaluation of policies, by making social media an optional source of public opinion (Kouam, 2019).

The innovativeness that social media platforms provide has made possible the tendency to scrutinize all the steps of the policy making process in public organizations, and the necessity of policy makers engaging these platforms in deliberation can no longer be ignored (Lant, 2014). The rapid rise of social media platforms has made it possible for citizens to raise their voices and

get the attention they deserve by using social media to frame their policy demands, mobilize or lobby fellow citizens, and explore alternative sources of knowledge on policy issues, as well as new ways of citizens participating in policy formulation and review (Bekkers and Edwards, 2017). Some authors have also claimed that advances made recently in theoretical modelling, computing and data collection enable the study of the role played by social networks in policy making (Battaglini and Patacchini, 2019).

Three characteristics make social media an important part of policy making (June, 2011).The three characteristics are the ability to access a wide range of citizen opinion and reveal what can easily be ignored or overlooked by traditional means of gathering opinion, the ability to connect users and offer them the opportunity to discuss or share opinions either in real-time or within a short time, and lastly, the ability to reduce the anonymity of users, making them much more careful with what they post online. This makes it easy for social media to be used in crowdsourcing of ideas for policy formulation, implementation and evaluation (June, 2011). These views on the importance of social media are consistent with Lant (2014).

The Potential of Social Media

Social media is a collective term used to refer to websites and applications which enable users online to create online communities, share information, ideas, personal messages and other content, such as videos (Merriam-Webster, 2004). According to studies, recent advances in technology, such as social media analytics, can greatly benefit policy interventions by complimenting traditional methods of data collection used by social policy formulators (Yeung, 2018). According to Yeung (2018), social media platforms remain largely untapped by policy makers as a source of data on citizen sentiments that can be used in policy actions aimed at bringing about social changes. The author further claims that social media platforms capture

systemic relationships and upstream factors that may not be captured by traditional methods of data collection. These findings are consistent with those of Nemes and Kiss (2020).

Traditional methods of data collection are prone to barriers such as political priorities, constraints of time, information overload, institutional features, budgetary constraints and insufficient evaluation skills (Payán & Lewis, 2019). This can be solved by network analysis of connections on social media, which can serve as a useful data source when seeking to understand societal attitudes with regard to policies, provided limitations such as privacy risks, data unreliability and validity of data are surmounted (Yeung, 2018). The availability of large quantities of data on networks such as Twitter and Facebook have attracted many researchers, leading to fears of digital trespass (Chiauzzi & Wicks, 2019).

Mendoza-Herrera et al. (2020) has in the recent past assessed the potential of social media platforms being used to address public healthcare issues such as nutrition, and notes that Twitter and Facebook are inexpensive platforms that can be used to encourage citizen participation in the promotion, development and evaluation of policy actions or the trends in opinion analyzed using neural networks and other data mining techniques. Yeung (2018) explains that rapidly refreshed and constantly changing social media data can help track attitudes and behaviors of individuals, which can provide insight into different aspects of healthcare and wellbeing, and being used to improve policies related to healthcare issues and the well-being of the citizens. This view is consistent with past research in the field of public health and nutrition (Mendoza-Herrera et al., 2020).

Studies have shown social media to be a useful tool for determining public sentiment. In a recent study by Nemes and Kiss (2020), sentiments in ambiguous posts, tweets, hashtags and comments can be classified using recurrent neural networks and by means of methods such as

natural language processing, which can assess the polarity of emotions. The authors used natural language processing and recurrent neural networks to develop a model based on the keywords ‘covid’ and ‘coronavirus’. and trained the model to work much more accurately so as to minimize the margin of error for use in determination of emotional manifestation on any topic at any given time. Based on their analysis, social media can be useful in determining public sentiment, provided models are trained to eliminate ambiguity and minimize errors.

According to Mendoza-Herrera et al. (2020), social media platforms are relevant sources of information for policy formulators that have gained recognition among academics, since they are updated continuously, provide up-to-date indicators that are useful to policy analysts as well as implementers, and can also be used to optimize interventions, assist with surveillance, and identify the socially vulnerable, based on geographical location, and inform resource allocation. The authors established that at least two-thirds of all adults in some forty countries use internet, of which 76% are regular users of social media platforms, making these networks hard to ignore.

The study by Yeung (2018) hypothesized that social media can be a catalyst for policy action, in particular in the realm of healthcare policies. While the study focused on healthcare, an exploration of its relevance in policy. and whether or not it can provide useful data that would provide insight into the process of policy making, can be used to provide a way of improving the policy making process and in particular, data collection. This has been suggested in part by Gilardi, Gessler, Kubli & Muller (2020), who point out that some social issues are yet to benefit from the vast information on social media the way politicians and political parties have benefitted.

Current Social Media Usage

Social media has transformed the way corporations collect and use information, with social networks increasingly being used as a critical source of data that can be used to understand consumer perceptions. Specifically, it has been used to understand their preferences, which can then be used to show them ads, while at the same time overcoming implied lack of attention from mainstream media (Tonello, 2016). According to the author, the use of social media, such as Twitter and Facebook, has changed the US economy in terms of communication and transformed the way information flows on matters concerning military intelligence, court decisions, law enforcement, and the two networks were even credited with encouraging the development of the Arab Spring movements in the Middle East (Tonello, 2016).

Twitter, Facebook and YouTube are currently being used to not only advertise and promote brands, but have also found usage in dispute resolution, customer service, making some important announcements, as well as general content sharing with the public (Petersen, 2019). Firms now employ social media teams to update clients and the general public, and handle different aspects of customer service, while collecting information that helps improve their goods and services. Blog posts, announcement of new product offerings, demonstration videos on how products work, tips to users or potential clients, and even recruitment of staff for vacant positions, are being accomplished through social media (Petersen, 2019).

In research, social media as an ever-present part of modern life has made it possible for individuals and companies to reach a massive audience at a fraction of the budget that one would be required to spend o reach respondents through other means (Rogers, 2019). Social media can also be used to promote research findings, apart from collecting views and providing a platform for engagement. In public health and nutrition, the World Health Organization has for a long

time used social media platforms such as Facebook, Twitter and others to help improve population health by influencing behavior change, changing lifestyles, collecting information on metabolic indicators and strengthening of patient counseling in resource constrained locations around the world (Mendoza-Herrera et al., 2020).

In a little more than a decade, social media has grown from a form of entertainment to an integral part of many people's lives. It is currently affecting almost every aspect of society and, in some ways, policymaking. For example, in the past two years, several police officers have been terminated due to racist comments that they made online. Employees from different companies and public figures have also been admonished, and, in some cases, lost their jobs, due to their activity online, and this shows how important social media platforms have become (University of Pennsylvania, 2019). This means that social media and the internet are fast becoming part of every aspect of people's lives and playing a central role in informing the public about what is happening in the world (Happer & Philo, 2013).

Research indicates that a majority of young people get their breaking news from social media platforms, with more than 2.4 billion people, or close to 65% of internet users, depending on social media platforms like Twitter, Facebook, YouTube, Instagram, and Snapchat for knowledge of what is happening in the world around them, and equally joining discussions, tweeting, retweeting or sharing such news content, instead of depending on mainstream media (Martin, 2018). The author notes that with the advent of big data and artificial intelligence, the danger of fake news diminishes, making social media platforms an important medium that can easily affect people's beliefs and attitudes, especially concerning the changes in society; and be used not only in lobbying, but also in shaping public opinion that may in turn affect policy

making. This research suggests that social media platforms be considered as a data source that informs policy making.

Social Media in Evidence-Based Policy Making

Evidence-based policy making uses decisions based on the rational analysis of the evidence available and its incorporation into policy (Sutcliffe and Court, 2005). Information obtained from surveys, for example, is used to influence policy making and reform, even though some barriers, such as differing political priorities, constraints of time, too much information, differences in institutional features, budget deficits and insufficient evaluation skills, have often interfered with this process (Payán & Lewis, 2019). Social media provides an opportunity for agencies involved in policy making to achieve collection of data for such evidence-based policy making at a fraction of the cost spent on surveys (Rogers, 2019).

According to Paquette and Redaelli (2014), evidence-based policymaking is most successful when policymakers refer to diversified information portfolios, and social media can offer this diversification. Despite the push by agencies towards evidence-based policy making, there still exists a huge disconnect between researchers, who produce new evidence, and the public, who are supposed to consume this information (Bogenschneider & Corbett, 2011).The huge investment in scientific research by the government and other agencies has not shaped policies to the extent expected, with the real beneficiaries not being reached either by the surveys or the findings, hence the need to try other avenues through which the mass market that should give and get this information can be reached. The challenges of transmitting research findings to policy makers and knowledge consumers have also been documented in financial policy making (Leuz, 2018).

Studies have in the past explored the possibility of using social media to inform public policy formulation. According to Happer and Philo (2013), use of social media led to better informed debate by the public on issues related to climate change by stimulating discussion of findings from research, which shows how important these platforms can be in advocacy, as well as disseminating the findings of empirical research. Other studies have explored the same issue and come to a similar conclusion, but call for further research to determine the suitability of social media technologies in policy formulation and implementation (Giustini et al., 2018).

However, while many of these studies established a relationship between social media trends and lobbying and other events that followed, which drive policymaking, they failed to explore how social media can be used to inform policymaking. Based on studies that explored factors such as data mining and machine learning, different scholars have determined that a wealth of information can be obtained from social media and other online platforms (Yeung, 2018). This data has been known to offer businesses information on their users' preferences, attitudes, and beliefs (Giustini et al., 2018). As such, the data is often used to determine advertisement content, which in turn drives profit. However, these studies have only explored the usage of data collected on their platforms to drive profits.

This study aimed to explore how staff members of the California Assembly used social media as a source of additional information for inform policymaking, especially with regard to improving the efficiency of the data collection and analysis phase of policymaking, and to assess the role played by social media platforms such as Twitter and Facebook in the formulation and evaluation of policies. This study also hypothesized that insight can be drawn from engagement, broad sentiment analysis, and a deeper granular analysis of the conversations on social media about issues that would then serve to build a robust enough data source that can be used to

inform public policy decisions related to these issues. These methods, coupled with the data the social sites have already collected about each user, may paint a better picture about who the users are, whether or not they are constituents in California (using their location data and registration information) and what elements about their behavior regarding societal issues can be used to gain an understanding about how they perceive such issues.

Extracting Social Media Data

Big data has changed the way data is collected, analyzed, stored and retrieved, and this can be of great use in policy making (Williams, 2019). Using big data techniques to collect data is not only fast, allows for diversity, and generates huge quantities of information, but also enhances the quality of information gathered, and can therefore be used to improve policy formulation and evaluation. Scientists, however, have identified various threats, such as technological challenges for governments, cultural barriers in the use of social media, and moral as well as ethical dilemmas for public officials and agencies involved in policy making and evaluation (Margetts, 2013).

The augmentation of opinion captured through social media posts may enhance the policymaking process. For instance, traditional surveys may target only a few members of the public who are in support of a specific policy, whereas the majority of the members of the public might oppose the issue (Yeung, 2018). Data mined from social media platforms allows policy makers to focus more on citizen opinion in terms of preferences and improve their own performance (Margetts, 2013). Additionally, the surveys may be influenced by researcher and respondents' biases, while monitoring of unsolicited comments on social media might be more representative of general public opinion. Social media-based data analysis can come from far more members of the public, and may give a more accurate depiction of their values and wishes

(Yeung, 2018). This can then be used to improve policymaking by making it easier to collect and develop an understanding of the public's opinions.

This study hypothesizes that social media could equally provide an opportunity through which state agencies tasked with rule making authority can understand values and expectations about different societal issues and help improve policy making in California. Currently, California has about 200 state agencies that can use social media platforms to analyze how residents perceive current policies, and their expectations for these policies in future (Micheli, 2018). Web scraping and big data analysis techniques can be used to access citizen opinion and represent their views, thereby legitimizing the whole policy making process (Ulbricht, 2019). This paper will assess how social media could be used by policymakers in California to inform their decisions on policy formulation, and how the process can be improved using big data and social media analytics.

METHODOLOGY

Policy making agencies in California largely depend on evidence-based approaches to policy formulation, where surveys are conducted among residents before findings are used to inform new policies or review of existing ones. Currently, this appears to be the only method of collecting information from the residents, despite barriers to valid surveys that have been identified in the past, such as political priorities, constraints of time, information overload, institutional features, budgetary constraints and insufficient evaluation skills, as well as the disconnect that exists between researchers as originators of new knowledge, and residents and policy formulators as the consumers of their research findings. This study hypothesizes that social media networks can help inform policy formulation by supplying data on community attitudes and perceptions on societal issues in the state.

Type of Analysis

The survey method, based on a Qualtrics anonymous survey, was used to assess the methods currently being used by policymakers in the California Assembly to inform their decisions and assess how these methods can be improved using social media analytics. A directory of Assembly chiefs of staff was used to get contacts and administer the survey to the staff members of the Members of the California Assembly who are involved in the policymaking process. These surveys were used to gather information on whether the Assembly could use social media platforms and big data technologies to inform policy formulation and evaluation.

The researcher aimed to reach staff in at least 25% of the Assembly members' offices. Survey questions included the current timelines involved in the policy making processes, and the costs involved in soliciting public opinion to inform policy formulation.

Finally, respondents were questioned on the legal implications of using social media analytics to inform policy formulation, and how such implications might be mitigated to allow for greater use of social media in policy formulation and evaluation. Using the data collected from this survey, a better understanding of the data collection aspect of the policymaking process was developed, and ways of improving the overall efficiency and speed of the process were suggested.

Data Collection

A directory of the chiefs of staff was used to administer the anonymous online survey containing questions, with the aim of getting a response from at least 25% of the legislative policy makers in the California assembly. Each of the respondents was mailed a survey with questions regarding the current policy formulation process, the timeline of the whole process, and the approximate cost of the process. This survey also sought to understand how these respondents could use social media and apply social media networks in the collection of data that informs the policy making process. The staff at the Assembly were also asked how web scraping and data mining from social media platforms could be used to improve policy making in the state, as well as the legal implications of such data mining, including how they might overcome these legal obstacles. The survey is in Appendix A.

Data Analysis

The information given by the respondents was transferred from the Qualtrics anonymous survey instrument to a spreadsheet. Content analysis was used to create summaries of the surveys and process the data obtained into codes. Data codes applied to each of the answers were reviewed and answers combined based on the themes derived from the research questions. The questions on the surveys were used to categorize the answers given by the members of staff into

themes, and the information checked for similarities and differences based on the experience of the members of staff in California's policy formulation process. The respondents did not have to identify themselves or display their personal identifiable information on the surveys.

IRB Exclusion

This study qualified for exclusion from IRB review, since the respondents who filled in the Qualtrics anonymous surveys are employees of public agencies involved in policy formulation, and personally identifiable information was not collected from any of the respondents. The questions on the surveys were about factual information based on what is currently being done in the Assembly, and not their personal experiences or opinions. All the surveys were designed to ensure that they comply with the anonymity needs of an excluded research process. All information collected was for the purposes of research only, and its design ensured that the identity of all respondents who choose to take part in the study was not revealed. The researcher did not collect personally identifiable information from respondents.

Data Selection

Only respondents directly involved in policy making and information communication technology were selected to form part of the study sample. Respondents were free to choose not to take part in the anonymous survey or to withdraw at any time during the process.

Study Limitations

The Covid-19 pandemic, which has made it necessary for most employees to keep social distance and minimize contact with the public, played a role in the relatively low response rate, with 61 respondents completing the interview schedules and mailing them back. On-line Qualtrics surveys were used to collect the qualitative data, which was then processed and presented in the form of pie charts and tables.

FINDINGS

The purpose of this study was to assess the role that social media platforms, such as Twitter and Facebook, could play in the formulation and evaluation of policies in the state of California, given the increasing application of big data in decision making in the private sector. This study evaluated how emerging technologies, such as machine learning and artificial intelligence, could be used to determine the attitudes and perceptions of the citizens of California, specifically on policy issues, and analyze how these technologies could be used by California as a means of gaining useful insights, gauging sentiments and collecting data, prior to formulating and evaluating laws and policies.

Response Rate

Out of 80 potential respondents, 61 agreed to take the survey, giving a 76% response rate. Males represented 39% of respondents, 61% of respondents were female, and one preferred no to state gender. There was missing data on a few of the questions, with some respondents skipping some of the questions and electing to leave blank spaces in the interview schedule. No reason was given for the blank spaces.

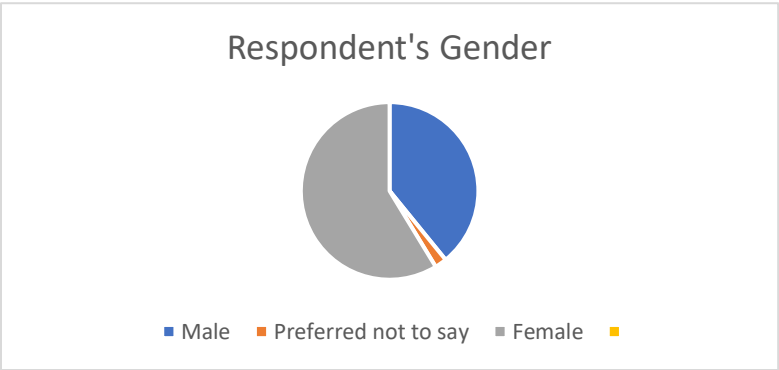

Figure 1: Respondent's Gender

Descriptive Analysis

The respondents were first asked about their sources of information whenever they developed new legislation, with 37% of them claiming that they depended on special interest groups to research and develop the bills before submitting them to a member of the Assembly to introduce, while an estimated 35% confirmed that the policy formulators did not depend on public opinion or citizen sentiments when researching bills and crafting new policies. An estimated 17% explained that bills were crafted by committee members, who then asked members of the California assembly for support, while 7% states that the number of bills created in a year were very few to warrant research, even though research was necessary if some members disagreed.

Table 1: Sources of Information for Policy Formulation

Source of Information	Percentage of respondents
Research by special interest groups	37%
No public opinion research	35%
Committee members	17%
Number of bills too few to warrant research	7%
Others	4%

About 2% of the staff stated that public opinion was only sought on bills sponsored by a member of the Assembly, while the remaining 2% stated that they only did public opinion research on bills opposed by some members of the legislative Assembly. These responses revealed that public opinion research was less likely to be taken into consideration by policy formulators, with public sentiments only becoming useful whenever members of the Assembly either sponsored a bill, or one of them opposed it, making public opinion research necessary. The

responses also revealed that most bills did not originate from the Assembly, with policy formulators depending more on special interest groups rather than members of the Assembly or committee members, to introduce a policy issue for crafting into a bill.

On how much time it takes policy formulators to gather information into a bill, staff members interviewed gave varying responses, with some of the respondents unsure how long such a process would take. A significant 37% of the respondents stated that the process takes less than six months, 28% of the respondents did not know how long it takes or had no idea, or were not sure, 20% of the respondents stated that it takes more than one year, even as long as three years, or simply too long, 15% of the respondents stated that data collection takes 6 to 12 months, while a respondent stated: 'Hopefully enough time to get an actual representation of public opinion and not just what opinion is being pushed at the moment. Some people try to change public opinion with half-truths and exaggerated information…'

Table 2: Duration of Data Collection for Policy Formulation

Duration of data collection for policy formulation	Percentage of Respondents
Less than 6 months	37%
Six months to one year	15%
More than a year	20%
Do not know how long it takes	28%

This indicated that the time taken was often too long, with respondents appearing to have no interest in how long it took. One of the respondents replied with a laughter emoji and provided no answer, while another claimed it took only two sessions for them to gather information on the proposed legislation, without substantiating. The length of time taken by policy formulators is therefore not fixed and can last for several years or even longer. Answers

such as ‘sometimes weeks, sometimes months’ given by one of the respondents show that members deal with bills originating from different sources differently, which is consistent with the responses received on the question of seeking public opinion on policies before formulating bills, in which respondents admitted to only doing public opinion research on those sponsored by members or opposed by one of them.

The third question dealt with the cost of the whole policy formulation process. Answers ranged from $0 to $1,000,000, with a significant 26% of them admitting that they do not know how much it costs. A respondent boldly claimed, “$200-$1,000 depending on how many pages there are to be mailed…” while another stated “$255,000 dollars.” Such exact figures were common among the answers, even though most seemed unsure of the exact figure, either replying with “Around $2.200”, “Approximately $5,000”, “haven’t asked” or “we don’t collect”. These answers were unsatisfactory, making it hard for the researcher to estimate the amount of money it costs the state to formulate or review policies under development

Table 3: Amount Spent on Data Collection

Cost of data collection for policy formulation	Amount in US Dollars
Do not know how much	26%
Varying amounts (0 – 1,000,000)	74%

On whether or not the use of social media to collect public opinion and research into public sentiments could reduce time and cost spent on the process, 61% of the respondents agreed that social media was bound to help in easing the cost and time burden, with 26% of them saying that it could not work, while the remaining used “maybe” or “probably” to indicate that they were not sure about whether or not it can work. One of the respondents said, “Yes if they keep it on track and not get into tangential issues…”, another one asked, “Does it matter?” while

another claimed, “Social media isn’t real” to register his or her disapproval of its use in collecting information on public opinion to help in the policy formulation process. The overwhelming response in the affirmative shows how popular social media networks have become with a majority of the citizenry.

Table 4: Opinion of Social Media Use in Formulating Pubic Policy

Opinion on Social media	Percentage of Respondents
Social media can be useful in reducing cost and time	61%
Social media cannot work	26%
Others	13%

While disputing the usefulness of social media, a respondent said, “No, not everyone tells the truth, and some people have multiple media accounts, and can be swayed to believe what is the most popular public opinion of the people they are interacting with...”, which reveals one of the complexities of social media usage. Individuals with multiple accounts can easily sway public opinion through multiple posts, while the risk of some individuals not giving accurate information would render the data gleaned from such accounts not useful to opinion researchers. Another one said, “No, social media isn't real, only what people want others to believe is real…” perhaps an indicator that what people post may not always be their opinion or conviction on an issue, but rather what they want the world to believe. However, one of the respondents appeared convinced on the usefulness of such networks, stating, “Yes. It's easier and a faster method to reach more people and get lots of opinions…”, which emphasized the speed with which opinions could be collected, and the huge number of people that could be potentially reached if social media was to be used. However, another said, “Yes, but the data would be garbage…”, emphasizing how much social media is mistrusted as a serious source of information.

Fifth, the respondents were interviewed on the methods of data mining that they were aware of. Interestingly, nearly half, 46%, flatly revealed that they did not know of any data mining methods that could be used to extract information from social media for purposes of policy formulation and or evaluation, while 26% of the respondents mentioned social media networks such as twitter, Facebook, LinkedIn, Tiktok, Instagram, and snapchat as data mining methods.

Table 5: Awareness of Data Mining Methods

Awareness of data mining methods	**Percentage of respondents**
Do not know any	46%
Facebook, twitter, Instagram etc.	26%
Other	28%

This reveals the lack of knowledge about data mining techniques among policy makers' staff members in the California Assembly. Two of the respondents even mentioned 'newspapers' as data mining techniques, while another two only mentioned that they were aware of several but did not substantiate on the specific techniques. Two strange responses appeared in form of "Web page visits, click thru rates, time spent on sites, network contacts, influencer posts…" and "Door to door and canvasing in the mall. also 'damn' phone calls…", further emphasizing the lack of knowledge on the subject.

Respondents were also asked about the current legal limitations on the use of social media to extract data for policy formulation and evaluation purposes, with 78% of the respondents stating that they did not know of any legal limitations to data mining and use of social media information. A respondent claimed that there were no limitations, while the rest claimed that the limitations available were not strict and policy makers were only supposed to

“hide personal information” “abide by privacy settings and drop people who ask to be unsubscribed”, “screen out personal information” “net neutrality”, “not many, if any” and “yes” without substantiation. This indicates that this is an issue that policy makers are not well versed with and could do with help from legal experts and or data mining professionals, if data mining from social media is to be of use in the policy formulation and evaluation processes.

Table 6: Knowledge of Legal Limitations on Data Mining from Social Media

Legal limitations to use of social media	Percentage of respondents
Do not know	78%
Other	22%

On the statutes that the California assembly would have to change in order to make data from social media useful to the policy formulation and evaluation process in the state, respondents gave varying answers, with 63% of the respondents claiming that they did not know of any statutes that required change, further indicating how new the idea is among policy makers’ staff members. One respondent stated there was need for HIPAA to be made more lenient, while another claimed that there was too much to be changed, without mentioning any specific statutes. Respondents appeared less informed on the laws guiding data mining as a process, whether from social media or elsewhere, and did not offer useful information on what they felt should be changed in the existing California statutes.

Answers such as: “See how many residences actually use social media to see if it would even be a good platform to use”, “Verify answers and mesh results with other market research efforts” “Allow for more information to be shared”, “Make sure the detection of spammers is faster so useless, false information is not passed to far”, “Making the companies to ask for permission to use people's personal information “signify lack of understanding of social media

data mining, while another respondent answered that the use of social media was in itself a bad idea, which meant he or she did not see the need of changing the law to make it possible for policy makers to mine data from social media platforms. Still another said, “I don't believe legislation should be created to divulge person's ID with their political opinions” to further express dissatisfaction with social media usage.

On how social media platforms can contribute to policy formulation and evaluation, many of the respondents felt that social media is a powerful tool that can be used to gauge public opinion by asking for citizens views and recording their answers, educating the public on the contents of the policy, gathering opinions from as many people in the state as possible, conducting surveys and gathering data, using the platforms to collect information on how to revise or review policy, or asking social media users questions and listening to their answers on policy. However, a significant number of respondents still did not know how social media could be used, with 35% of them admitting they did not know how.

Table 7: Social Media Used to Contribute to Policy Formulation

Contribution of social media to policy	Percentage of respondents
Useful in gauging public opinion, educating the public	65%
Do not know how	35%

According to some respondents, policy makers must make sure that debate in social media remains focused if it is to be used to inform the policy making process, while others felt it should not be used at all. “You should not use these platforms at all to formulate policies”, said one, while another stated, “I think it is a terrible idea”, indicating the fear that opinion leaders have of social media. A respondent felt it would be better to hire a policy firm to use social media instead, perhaps afraid of the intricacies the whole process would entail. Even though the

respondent did not substantiate, policy firms are expected to shoulder the liability and circumvent some of the legal hurdles expected if social media is to be used at any stage of the policy formulation process (Soskolne, 2004).

Respondents were also interviewed on the social media platforms they felt could potentially be of use to the policy making process, with 52% of them mentioning Facebook as the social media platform with the potential to inform policy making. Another 35% of the respondents mentioned twitter, 20% of them mentioned Instagram, 7% mentioned LinkedIn, while one each mentioned snapchat, myspace and Tiktok. However, a number still felt the use of social media in policy formulation was a bad idea, while others admitted they did not know. Some 11% of the respondents did not know which platform would be useful, while one stated, "This is a bad idea" to suggest how difficult it would be to use social media, without substantiating, while another pointed out that social media platforms are private and business oriented, making it hard for policy makers to use them.

Table 8: Preferred Social Media Platforms (multiple responses accepted)

Social media platforms that could be useful	**Percentage of Respondents**
Facebook	52%
Twitter	35%
Instagram	20%
LinkedIn	7%
Do not know	11%
Bad Idea	2%

On the type of useful information that can be gleaned from social media platforms for use in policy making and evaluation, 15% of the respondents thought that the platforms could be

useful in opinion research, where they could be used to give different views on policy aspects. One of the respondents stated, “Opinions and which options are preferred”, while another stated that they could be used to gauge “People’s opinions, wants, and observations”, while yet another claimed that these platforms could be useful in getting “Personal interest, ideas from chats, open conversations and from open surveys”, which indicated that there is a lot that can be mined from social media to inform the policy making process.

Others thought that social media platforms could be useful in capturing demographic and socioeconomic information such as “Age, income, interests, marital status, voting preferences”, “Household demographics” percentage of the populace who are willing to respond and their personal demographics”, “All sorts of things if I actually worked in this field”, even though 28% of the respondents still said they did not know how it could be used, while another said it could be used for health issues shared by the users of the platforms. The large number of policy makers’ staff members who do not know what useful information can be mined from social media is, however, significant, which indicates the amount of work that should be done before these platforms become useful to policy making and evaluation.

Respondents were asked about the web scrapers that could be used to mine data from social media platforms, with 39% of the respondents admitting that they did not know what web scrapers are, while 11% of the respondents gave the wrong answers, including Google and Yahoo, which clearly are not web scrapers for data mining. Two respondents gave the right answer, with one stating: “The only ones I have heard of is Scrapehub and Parsehub, from articles on LinkedIn,” while another was aware of data scrapers but not their actual names. This further indicated how knowledge of data scrapers is scarce among policy makers’ staff members.

This suggests that social media data craping might be harnessed in the future provided staff members get the necessary information on the available options.

ANALYSIS

The Importance of Social Media

The idea of using social media in policy formulation, implementation and evaluation is relatively new, with many public agencies depending on the mainstream media such as radio, television and newspapers to set the agenda and lead the audience in discussing what is perceived to be of public importance (Feezell, 2017). Some of the respondents interviewed mentioned newspapers as the only method they use to research issues of public interest, with much of the research work being left to special interest groups that sponsor the policy formulation process.

Feezell (2017), however, notes that this is about to change, with social media platforms such as Facebook having taken on extra importance, by raising the perceived importance of some issues among the audience, especially on political issues. The author explains that the audience for mainstream media has been shrinking, with the audience becoming increasingly fragmented. Social media is taking on extra importance and impacting the issues that affect a huge percentage of the population, with individuals more likely to consider issues shared through social media as important, even when they had no interest in political issues.

Some of the respondents from California who were interviewed acknowledged the increasing popularity of social media platforms such as Facebook, Twitter, LinkedIn and Instagram. They acknowledge that social media holds large amounts of data from demographics to opinions that can be used in policy formulation, but they did not understand how this information could be mined, or who could do the mining. These platforms carry information that individuals consider important and keep discussing for weeks or even months and can be useful in policy making. Studies have shown that one does not need mainstream media forms such as

radio, television and newspapers, especially if they live in urban areas (Salman, Mustaffa, Salleh & Ali, 2016).

Consistent with Feezell (2017), studies have shown the impact that media, whether social or mainstream, can have on policy making, with an integrative systemic review concluding that media interventions had the potential to influence the policy making process, even though many researchers had shied away from establishing how. Interventions based on social media are scarce, despite its growing significance in agenda setting, particularly in political and economic discourses (Bou-Karroum et al., 2017). The authors argue that the use of the media in a purposive and organized manner to motivate, inform and persuade populations can greatly inform the making of policies. The lack of knowledge among the survey respondents on how this can be done does not mean that it cannot be done, and better organized ways of gleaning the massive information from social media have to be used to make it useful in policy making.

Social media has been effectively used to inform political campaigns in the United States and beyond, and its growing significance in agenda setting cannot be ignored, with authors noting that Facebook and Twitter influenced Donald Trump's election against Hillary Clinton in 2016, through supervised machine learning, and mining of texts to analyze user posts and comments (Rossini, Stromer-Galley & Zhang, 2020). Policy makers in California and beyond have a responsibility to find ways of capturing citizen sentiments from social media platforms and using them to inform the policy making and evaluation process in a faster and more cost-effective way. The rapid rise of social media, and the spread of its use has made it possible for millions of citizens to make their voices heard and get the attention they deserve (Bekkers and Edwards, 2017).

How Social Media Can be Used

Many of the respondents seemed unsure about how social media platforms can be used to inform the policy making process. For this reason, social media has remained largely untapped, with many scholars arguing that social media platforms can be a source of data on citizen sentiments that can be used in policy actions aimed at bringing about social changes (Yeung, 2018). The author further claims that social media platforms capture systemic relationships and upstream factors that may not be captured by traditional methods of data collection, which can be used to inform the policy making process in the state of California and make the process faster and cheaper for the state government. Other studies have also documented how important social media can be as a source of information and an agent of behavior change communication, in a world where the use of these platforms has been expanding to levels that cannot be ignored (Mendoza-Herrera et al., 2020).

Some respondents explained that social media had great potential and contained a lot of information that could be useful but lamented the complexity of getting this information, with some talking of there being a lot of garbage on platforms like Facebook and Twitter, while others pointed out the privacy settings on these privately owned networks, which would make it difficult for public entities to access and use the information in policy formulation, implementation, and evaluation. Scholars however feel that this potential can be harnessed if there is the will by state agencies and point to the effectiveness with which Facebook and twitter have been used to influence political and economic discourses as reason enough to believe social media can be useful to policy making (Bekkers and Edwards, 2017).

Social media platforms are relevant sources of information for policy formulators that have gained recognition among academics, since they are updated continuously, provide up to

date indicators that are useful to policy analysts as well as implementers, and can also be used to optimize interventions, assist with surveillance and identify the socially vulnerable based on geographical location, and inform resource allocation (Mendoza-Herrera et al., 2020). Based on this, it is possible for policy makers in California to use Facebook, twitter and LinkedIn, for example, to do opinion research that will most likely enhance the quality of policies being made, given that most respondents admitted that they do not do public opinion research unless some members of the state assembly are opposed to a bill.

Yeung (2018) hypothesized that social media can be a catalyst for policy action, in particular in the realm of healthcare policies. While the study focused on healthcare, an exploration of its relevance in broader policy, and whether or not it can provide useful data that would provide insight into the process of policy making, can be used to provide a way of improving the policy making process and in particular, data collection.

Cost and Duration of Data Collection

Estimates by respondents put the cost of data collection at between $0 and $1,000,000, depending on how much data is to be collected, with some of the respondents seemingly not aware of the exact figure that is spent on this phase of policy formulation. The expensive nature of the data collection process, and the time taken by the process, are likely the reasons that respondents admitted that opinion research is not carried out prior to policy formulation and crafting of bills, unless some members raise an objection to the policy or bill. Some sources put the amount of data generated daily by online users at a staggering 44 billion gigabytes, costing business organizations $800 billion a year (Kaufman, 2018). This amount can greatly be reduced through use of social media platforms, while the amount of time required to collect the data can also be reduced significantly.

California's state assembly can either pay a developer to set up an in-house web scraper at a fee, or use already existing web scraping companies that are available worldwide. Building their own web scraper would involve paying a web development company such as datafiniti to set up one (Deysarkar, 2017). According to the author, building each web scraper can cost an estimated $500, while the server cost of running the scraper and storing the scraped data is also estimated at $500. Maintenance of the scrapers would cost $500 a month, while the one-off fee for processing the data scraped would cost $10,000, bringing the total one-off cost to $12,500, and a recurrent expenditure of $1,000 every month. The developer explains that this cost could be prohibitive for small business organizations, making the use of already existing web scrapers the better alternative. Some freelancers offer the service for as low as $99 initial set up fee and $79 monthly fee for maintenance (Garcia & Perez, 2019).

Companies such as Zyte, formerly known as Scrapinghub, have been scraping web pages since 2010 and scrape more than 8 billion pages every month, from which they collect information for the business community, especially e-commerce stores that need price intelligence (Zyte, 2018). According to the company, web scraping has become easier and cheaper, with open-source programs available for use by individuals and companies that require the technology. The authors however note that the quality and speed of these open-source programs may be a challenge, particularly for large amounts of data, making it necessary for organizations to pay for the service. Other challenges are listed as the sloppy and ever-changing web formats, the large number of scalable requests, maintenance of throughput performance, and the anti-bot performance measures employed by some websites.

According to Parsehub, one of the most popular web scrapers, there are free web scraping tools online, with freelancers offering the service for as little as $10 per website scraped (Garcia

& Perez, 2019). These freelancers use automated bots or tools that collect information from websites on the internet and put it in an easy to interpret format such as xml, excel or JSON. Scraping can be done for market research, to get useful information such as contacts or addresses, for research purposes, or even for aggregating data from blog posts. Reviews and comments can also be analysed by web scrapers. Scraping can be done through custom made web scrapers developed by programmers using languages such as Python, or by means of ready-made web scraping software.

Garcia and Perez (2019) explain that web scraping software from companies such as Parsehub can be used by clients who have no technical ability, allows maintenance support from the developers of the software, can be maintained by the clients without necessarily contacting the developers, pricing is lower than using custom made web scrapers, and the same software can be used to scrape through multiple sites, with the downside being that some ready-made software cannot scrape through complex websites. It also requires time, and may require help from technicians if advanced features are to be used. The authors further explained that the demand for web scraping technology is currently high, revealing its popularity in research and market intelligence.

The duration of web scraping is also extremely rapid, with sources revealing that some web scrapers can handle more than 40,000 requests every day, averaging one complete request every two seconds (Zyte, 2018). This is faster than analysis of surveys can ever occur, and would allow organization to get information from millions of people within days. Scraped data is stored in the cloud, and for this reason, the California Assembly may not require any special hardware to process the information, or access the processed information. While there have been cases involving Facebook against some data scrapers in the past, such lawsuits have been levelled

against companies that accessed private information from YouTube, Twitter, Facebook, LinkedIn, Instagram, and Amazon to sell as market intelligence (Hutchinson, 2020). Ethically using publicly available information, however, has no legal hurdles.

According to some studies, social media provides an opportunity for agencies involved in policy making to achieve collection of data for such evidence-based policy making at a fraction of the cost spent on surveys (Rogers, 2019). Researchers have also claimed that use of social media led to better informed debate by the public on issues related to climate change by stimulating discussion of findings from research, which shows how important these platforms can be in advocacy, as well as disseminating the findings of empirical research (Happer & Philo, 2013). Another study by Giustini et al. (2018) acknowledged the significant role that social media can play in policy, but called on further research to determine the suitability of social media technologies in policy formulation and implementation. The present research shows that there is a long way to go, since most respondents were not aware of what role platforms such as Facebook and twitter could play in this process.

Mining Data from Social Media Platforms

Most respondents did not know of any data mining techniques that could be used to glean useful data from social media platforms, yet studies reveal that data mined from social media platforms has the potential to allow policy makers to focus more on citizen opinion, including preferences, and improve their own performance (Margetts, 2013). Social media-based data analysis can come from far more members of the public and may give a more accurate depiction of their values and wishes, compared to other evidence-based methods. Most respondents did not know of any web scraping and big data analysis techniques, yet these techniques can be used to

access citizen opinion and represent their views, thereby legitimizing the whole policy making process (Ulbricht, 2019).

Data mining techniques have the capacity to determine value from structured and unstructured data, provided the huge amounts of data collected from these platforms is cleaned, and converted to actionable information. These techniques employ statistics, data management, and machine learning to process and draw useful conclusions from large quantities of information. Techniques allow organizations to track patterns within sets of data, classify these sets of data, link sets of data that are associated or correlated, identify anomalies or outliers, and predict what might happen in future based on the collected data (Alton, 2017).

Technology can enhance the use of data collected from social media networks. While respondents felt much of this data would not be useful, the data collected can be cleaned of garbage, classified and converted into useful information that can influence policy making, implementation and evaluation. According to studies, using big data techniques to collect data is not only fast, allows for diversity, and generates huge quantities of information, but also enhances the quality of information gathered, and can therefore be used to improve policy formulation and evaluation. Scientists, however, have identified various threats such as technological challenges for governments, cultural barriers in the use of social media, and moral as well as ethical dilemmas for public officials and agencies involved in policy making and evaluation (Margetts, 2013).

CONCLUSION AND RECOMMENDATIONS

There has been tremendous growth in the use of social media, with more than 70% of all adults in the United States being regular users of social media platforms. This ubiquitous use of these platforms makes them a necessary component of the policy making and evaluation process, yet studies have not focused on their relationship with policy formulation and evaluation. The increasing usage of big data is currently changing the way information is collected, stored, organized and processed, and this technology presents governments and government agencies, such as the California State Assembly, with great opportunities of using social media for policy making on matters concerning the public.

Social media has become indispensable as an agenda setting platform, and the emergence of technologies such as machine learning and artificial intelligence make it possible to obtain critical insight regarding the perceptions and attitudes of individual citizens, and inform the policy making process. Yet this has not been the case in California, since most of the policy makers' staff members are unaware of the ways in which these networks can be used to provide information that can be used in the policy making process. Using social media to collect information will not only speed up the data collection process, but also reduce the cost of data collection, which according to respondents can top a million dollars and last for more than three years.

This study recommends that the California Assembly incorporates the use of popular social media platforms such as Facebook, Twitter and LinkedIn in its policy making process. Currently, policy makers do not carry out opinion research on policies before drafting bills and enacting legislation, unless some of the assembly members object to the bills. It is necessary that the public participate more in the data collection process, with social media being one of the

ways in which their sentiments can be taken into account before proposed policies are formulated, implemented, revised or evaluated. While a lot of information contained within social networks may not be useful as some of the respondents pointed out, web scrapers and data mining techniques can help glean this data, clean it to determine what is useful to the policy making process, and incorporate the findings into the process.

Respondents admitted that social media platforms can easily shorten the data collection process and lessen the budget for the state government, yet for this to happen, there is need for policy makers' staff members to be educated on web scrapping and data mining. Most of them did not have a grasp of these technologies, or how they can inform the policy making process. Educating the policy makers on data mining will help them acknowledge how indispensable social media could become to the policy making process. The assembly could change its legislative formulation processes to pave the way for opinion research on bills, or simply use a policy research firms to collect information on proposed policies from the social media, clean the data, identify what is useful and use it to shorten the data collection process and make it less costly.

Alternatively, this study further recommends that the Assembly's information technology department invests in web scraping technology to help with its in-house data collection process. Tools such as Apify SDK, Octoparse, Scrapingbee, Luminati, Scraper API, scraping bot, and Parsehub, among others, can make data collection from social media platforms easy. These platforms are here to stay and have changed the whole agenda setting process, with political and economic discourses being shaped by social media platforms. Policy making can greatly benefit from web scraping technologies, which make mining data from the internet not only faster, but cheaper, with a company confirming that it can develop a data scraping tool for a client for only

$12,500 , with a recurrent expenditure of $1,000 monthly, or offer the same service at a cheaper cost. The cost quoted by datafiniti includes developer time to develop a web scrapper for the California Assembly, the cost of the server, data storage costs, maintenance cost, as well as normalizing, merging and processing of data scraped from five different sites.

References

Agrawal, H. (2019, February 18). *7 Social Media Scraping tools for 2019*. Datahut: https://blog.datahut.co/social-media-scraping-tools/

Alton, L. (2017, December 22). *The 7 Most Important Data Mining Techniques*. Retrieved from Data Science Central: https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques

Battaglini, M. and Patacchini, E. (2019). Social Networks in Policy Making. *Annual Review of Economics* , 473-494 .

Bekkers, V., and Edwards, A. (2017). The role of social media in policy processes. *Handbook of the Policy Process*, University of Rotterdam.

Bhattarai, P., and Adhikari, P. (2020, July 29). Data for public policy. My Republica My City: https://myrepublica.nagariknetwork.com/news/data-for-public-policy/

Bogenschneider, K., & Corbett, T. J. (2011). *Evidence-Based Policymaking: Insights from Policy-Minded Researchers and Research-Minded Policymakers.* New York: Routledge.

Bou-Karroum et al. (2017). Using media to impact health policy-making: an integrative systematic review. *Implementation Science, 12*(52), 1-14.

Cahn, M. A., & Schockman, H. E. (1996). *California: An Owner's Manual.* Pearson.

California State Assembly. (2020, September 27). *Assembly*. California State Assembly: https://www.assembly.ca.gov/

California State Senate. (2019, April 1). The Legislative Process: A Citizen's Guide to Participation. California State Senate: https://www.senate.ca.gov/sites/senate.ca.gov/files/legislative_process.pdf

Chiauzzi, E., & Wicks, P. (2019). Digital Trespass: Ethical and Terms-of-Use Violations by Researchers Accessing Data From an Online Patient Community. *Journal of Medical Internet Research, 21*(2), e11985.

Dewi, L. C., Meiliana & Chandra, A. (2019). Social Media Web Scraping using Social Media Developers API and Regex. *Procedia Computer Science*, 444-449.

Deysarkar, S. (2017, August 10). *The Money Pit: Why You Shouldn't Build Your Own Web Scrapers*. Retrieved from Datafiniti: https://blog.datafiniti.co/the-money-pit-why-you-shouldnt-build-your-own-web-scrapers-1d2b3c7a899

Engler, A. (2020). What all policy analysts need to know about data science. Brookings. https://www.brookings.edu/research/what-all-policy-analysts-need-to-know-about-data-science/

Feezell, J. T. (2017). Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era. *Political Research Quarterly, 71*(2), 482-494 . doi:10.1177/1065912917744895

Garcia & Perez. (2019, August 1). *A Complete Guide on Offering Web Scraping Services*. Retrieved from Parsehub: https://www.parsehub.com/blog/a-complete-guide-on-offering-web-scraping-services/

Gilardi, F., Gessler, T., Kubli, M., & Muller, S. (2020). *Social Media and Political Agenda Setting.* University College Dublin.

Giustini, D., Ali, S. M., Fraser, M., & Kamel Boulos, M. N. (2018). Effective uses of social media in public health and medicine: a systematic review of systematic reviews. *Online Journal of Public Health Informatics, 10*(2), e215.

Happer, C., & Philo, G. (2013). The role of the media in the construction of public belief and social change. *Journal of Social and Political Psychology,* 321-336.

Hutchinson, A. (2020, October 1). *Facebook Takes Legal Action Against Data Scrapers in Latest Enforcement Effort* . Retrieved from Social Media Today: https://www.socialmediatoday.com/news/facebook-takes-legal-action-against-data-scrapers-in-latest-enforcement-eff/586288/

June, P. (2011, October). Social Media Impact on Policy Making. *SERI Quarterly,* pp. 125-129.

Kouam, H. (2019, April 22). *Policy-making in the era of social media*. Medium: https://medium.com/@henrikouam/social-media-for-solutions-oriented-policy-6588f1b62a95

Lant, M. (2014). *Social Media and the Policy-Making Process: A Traditional Novel Interaction.* A research project submitted to the Ai War College, Air University.

Leavey, J. (2013). *Social Media and Public Policy: What is the Evidence?* Economic and Social Research Council.

Leuz, C. (2018). Evidence-based policymaking: promise, challenges and opportunities for accounting and financial markets research. *Accounting and Business Research, 48*(5), 582-608.

Margetts, H. (2013, October 28). The promises and threats of big data for public policy-making. The Policy and Internet Blog: https://blogs.oii.ox.ac.uk/policy/promises-threats-big-data-for-public-policy-making/

Martin, N. (2018, November 30). How social media has changed how we consume news. *Forbes*. https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/

Mendoza-Herrera, K., Valero-Morales, I., Ocampo-Granados, M. E., Reyes-Morales, H., Arce-Amaré, F., & Barquera, S. (2020). An overview of social media use in the field of public health nutrition: benefits, scope, limitations, and a Latin American experience. *Prevdenting Chronic Disease, 17*(1), e76.

Micheli, C. (2018, June 25). The Role of State Agencies in Policy Making. Cap Impact:, McGeorge School of Law. https://www.capimpactca.com/2018/06/role-state-agencies-policy-making /

National Collaborating Centre for Healthy Public Policy. (2013). *Public Policy Models and Their Usefulness in Public Health: The Stages Model.* Quebec: NCCHPP.

Nemes, L., & Kiss, A. (2020). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication* , 1-16.

Payán, D. D., & Lewis, L. B. (2019). Use of research evidence in state health policymaking: Menu labeling policy in California. *Preventive Medicine Reports*, 16(1), 1-5.

Petersen, L. (2019, March 24). Role of Social Media in Organizations. *Houston Chronicle*: https://smallbusiness.chron.com/role-social-media-organizations-30678.html

PPIC. (2020). *PPIC Statewide Survey: Californians and Their Government*. Public Policy Institute of California: https://www.ppic.org/publication/ppic-statewide-survey-californians-and-their-government-september-2020/

Rehr, D. K. (2017, July 2). How Is Social Media Being Used In Advocacy*? Huffpost*: https://www.huffpost.com/entry/how-is-social-media-being-used-in-advocacy_b_589a7b12e4b0985224db5bac

Rogers, J. (2019). The use of social media and its impact for research. *BioResources*. https://bioresources.cnr.ncsu.edu/resources/the-use-of-social-media-and-its-impact-for-research/

Sá, K. d. (2010, July 6). How our laws in California are really made. *The Mercury News:* https://www.mercurynews.com/2010/07/06/how-our-laws-in-california-are-really-made/

Salman, Mustaffa, Salleh & Ali. (2016). Social Media and Agenda Setting: Implications on Political Agenda. *Malaysian Journal of Communication, 32*(1), 607-623.

Sutcliffe, S., & Court, J. (2005). *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?* London: Overseas Development Institute.

Tonello, M. (2016, May 17). *Corporate Use of Social Media*. Retrieved from Harvard Law School Forum on Corporate Governance: https://corpgov.law.harvard.edu/2016/05/17/corporate-use-of-social-media-2/

Ulbricht, L. (2019). Scraping the demos. Digitalization, web scraping and the democratic project. *Democratization*, 426-442.

University of Pennsylvania. (2019, July 26). *The Impact of Social Media: Is it Replaceable?* Knowledge@Wharton: https://knowledge.wharton.upenn.edu/article/impact-of-social-media/

Williams, J. (2019, October 5). How Government Policy Making will be Impacted by Big Data? – *PromptCloud*. Prompt Cloud: https://www.promptcloud.com/blog/big-data-influence-on-government-policy-making/

Yeung, D. (2018). Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint. *Journal of Medical Internet Research, 20*(3), e94.

Zyte. (2018, July 2). *Data for price intelligence: Lessons learned scraping 100 billion products pages*. Retrieved from Zyte: https://www.zyte.com/blog/price-intelligence-web-scraping-at-scale-100-billion-products/

Appendix A

Appendix 1: Survey on the Use of Social Media for Public Policy Formulation by the California Assembly

Social media users express opinions on a variety of current events through their posts. Technology now exists that allows this information to be subject to data mining, which could be a way to collect public opinion on a public policy issue efficiently. The purpose of this survey is to understand whether you currently use social media to understand public opinion on public policy issues, or what techniques you currently use to gauge public opinion when formulating public policy.

Categories of Information Included in the Survey	Number of Questions
Current policy formulation practice	7
Social media for public policy formulation	4
Total	11

Current Policy Formulation Practice

1. When you are researching the creation of a new law, what sources of information do you use to develop it? [Check all that apply]
    a. Most bills are created by interest groups and submitted to a member to carry
    b. Most bills are created by an Assembly committee and members are asked to support them
    c. We only create a few bills a year in our office, and we do public opinion research on them

d. We only do public opinion research on bills that the Assembly member wants to support
e. We only do public opinion research on bills that the Assembly member wants to oppose
f. We do not do public opinion research.

2. How long does the current process of public opinion data collection take during policy formulation?

_____

_____

_____

3. Approximately how much does it cost your office to collect public opinion data from the residents of your district using the current data collection method?

_____

_____

_____

4. Do you think that the increased use of social media to collect residents' opinions on policy matters can reduce the time and cost spent by the Assembly on data collection for policy making purposes?

_____

_____

_____

5. What methods of social media data mining are you familiar with now?

_____

_____

_____

6. What are the current legal limitations on social media data mining?

_____

_____

_____

7. What changes in statutes could the Assembly sponsor in order to use social media platforms more for policy making purposes?

_____

_____

_____

Please answer the following questions if you are familiar with using social media for public policy formulation'

8. Which social media platforms could you potentially use to inform policy making?

_____

_____

_____

9. How could you use these platforms in formulating policies?

_____

_____

_____

10. What type of useful information could be mined from these social media platforms for use in policy making?

_____

_____

_____

11. Which web scrapers could be used to mine data from social media platforms?

_____

_____

_____