Social Media Analysis and Opinion Mining on Politics

1Mr. Rohit Pal, 2Dr. A.P. Shukla

1Research scholar, Department of Computer Science and Engineering, Dr. K. N. Modi, University, Rajasthan.
2Department of Computer Science, Government Polytechnic, Bijnor.

Abstract—The majority of computational methods for opinion mining have concentrated on sentiment categorization for blogs or reviews, including product reviews. In order to ascertain sentiment or stance toward a certain target of interest, this paper analyzes socio-political opinion mining, with a focus on social media platforms like Facebook and Twitter. We investigate the problem of automatically inferring from political opinion in social media whether the author of a given text is in favor of or against a target, a problem known as stance detection, and whether the author is positive or negative towards a target and aspects of the target, a problem known as fine-grained opinion mining. The suggested models have been tested using political Twitter and Facebook datasets.

Keyword: opinion, Politics, social media, etc.

Introduction
Industry and the scientific community have traditionally used surveys or polls of the general public, given to a properly selected sample of the population, as the most trustworthy ways to find out "what people think?" Public opinion surveys are helpful for obtaining a somewhat accurate picture of how different people feel about a particular subject within a population, but they are expensive to conduct, biased, and have poor response rates. Public opinion research is currently entering a new era by offering options to alter the way conventional public opinion polls function, thanks to the expansion of internet infrastructure and internet access techniques.

These days, a lot of people are willing and able to publicly express their ideas on well-known websites like Facebook and Twitter. Millions of reviews and answers are accessible online as a result of these online activities. Due to the abundance of data available, there are now excellent opportunities to comprehend public policy, government decisions, social/cultural events, vacation spots, political issues, and new product opinions, among other topics. For example, policymakers may use posts in forums, Facebook, and Twitter relating to their target policy or legislation to gather feedback from the public on how they feel about new policy decisions or legislative amendments.

As a result, traditional surveys and polls are no longer the only option for businesses, governments, or private citizens to get public opinion. However, it takes a lot of time and money to manually analyze this massive volume of textual data.

The goal of opinion mining research is to provide automated methods for reliably analyzing such opinion data. Opinion mining is a computer study that identifies subjective information, like opinions, sentiments, or emotions revealed in texts, by utilizing computational linguistics, text analysis, natural language processing, and machine learning. Finding the explicit or implied opinions about a target entity, such as a semantic orientation or polarity like positive/negative, like/dislike, for/against, etc., is the main goal of opinion mining research.

Political Opinion Mining
Opinions matter a great deal in politics. Nowadays social media plays a vital role in enabling the general public to view political opinions and engage in political discussions related to different political events. Elections are one event where social media provides an opportunity for both voters and candidates to publish their opinions online. Similarly, social media offers a platform for social/political activists (left-wing and right-wing activists), non-government organizations (NGOs), and civic organizations to set up a virtual space by forming social media groups such as Facebook groups/pages to publish opinions, engage with society and organize events including protests. A significant portion of the first research on opinion mining was devoted to business and e-commerce uses, such as product and movie reviews. One motivation for concentrating on these domains was the accessibility of labeled data sources, such as star ratings, which may be utilized as quantitative measures of opinion. With Twitter and Facebook's recent growth, a wide range of opinion-mining research topics have become possible for various new applications. A few examples include using these platforms to predict election outcomes, track the spread of diseases, predict movie box office
success, and monitor the effects of disasters. Views regarding electrical products, lodging, dining establishments, etc., have been the subject of a lot of early research. Research on comprehending opinions in a political environment is scarcer. The majority of these studies focus on analyzing various aspects of political tweets to identify contentious issues and political opinions, measure the degree of polarization in the electorate, forecast voting intentions or election outcomes, and identify sentiment, emotion, and purpose in electoral tweets.

**Social Media Opinion Mining in Politics**

This section reviews work on opinion mining in the domains of natural language processing and machine learning, with a focus on political opinion mining for social media. Fine-grained opinion mining and stance detection are the two opinion mining research challenges that we are interested in.

The study of people's opinions, feelings, assessments, appraisals, attitudes, and emotions regarding various entities—such as goods, services, companies, people, problems, occasions, subjects, and their characteristics—is known as opinion mining. Organizations and individuals are using opinions found on social media and other online sources more frequently for decision-making due to the rapid development of user-driven social media and internet infrastructure. However, because of their physical and mental limitations, people find it challenging to produce consistent results due to the vast volume of available data. Opinion mining has become a particularly active study subject during the last two decades due to the necessity for automated opinion-mining methodologies.

Opinion mining is a field of study in computer science that spans computational linguistics, natural language processing, and machine learning. It uses text analytics to extract subjective information from text. Furthermore, industry interest is focused on a number of practical applications of opinion mining, including risk identification in banking systems, nation-state relationship analysis, market prediction, competitive intelligence, and election predictions.

Opinion mining applications have been applied to a variety of data sources, including blogs, review sites, microblogs, and, more recently, social networks. In recent years, opinion mining research has made extensive use of microblogs, like Twitter, and a wide range of applications, including tracking the spread of disease, observing the consequences of disasters, and predicting elections, have been developed. Facebook is a social network that is less frequently utilized for opinion-mining research than Twitter because most of its data is restricted.

That being said, there are intriguing study options using the publicly available Facebook data. In this thesis, we assess models for fine-grained opinion mining and stance identification using data from Facebook and Twitter.

In general, three levels of granularity have been used in past studies to study opinion polarity classification. Analysis at the document and phrase levels was the focus of several early research. Opinion mining at the document level is the most extensively researched issue. This mostly focuses on categorizing documents according to the overall sentiment orientation of the document—that is, whether it is positive, negative, or neutral. Finding each sentence's opinion polarity is the goal of sentence-level opinion mining.

Opinion mining at the phrase and document levels is helpful in many situations, but there is still a lot that needs to be explored. For example, labeling a document as positive does not necessarily mean that the person who wrote it thinks well of all entities and their aspects. Therefore, a closer look at the aspect level is needed to get a more detailed opinion analysis. Opinion mining techniques based on machine learning define the opinion polarity classification problem as a conventional text classification problem that frequently depends on syntactic and linguistic elements, machine learning approaches rely on learning algorithms. Based on the set of characteristics in the underlying text, the classification model is trained on a training dataset, and each instance is assigned a class label. The model can then be used to predict a class label for a given new observation. There are four types of machine learning approaches: unsupervised, semi-supervised, weakly supervised, and supervised.

### Machine Learning Approaches

- Supervised.
  - (All data is labeled)
For opinion polarity categorization, supervised learning's main benefit is its ability to automatically learn from a wide range of features through optimization. On the other hand, the availability of labeled training data is a prerequisite for supervised machine learning techniques. Data labeling is a laborious and time-consuming operation, especially when dealing with huge datasets. A further drawback of the method is that supervised classifiers that are trained using labeled data in one domain are frequently ineffective when applied to another.

The two primary kinds of supervised opinion polarity classification techniques are probabilistic and non-probabilistic classification. Opinion polarity was classified using probabilistic algorithms like Maximum Entropy, Naive Bayes (NB), and Bayesian networks in the early days of opinion mining research. Neural networks, support vector machines (SVM), closest neighbors, decision trees, and rule-based techniques are some of the non-probabilistic classifiers that are frequently employed in opinion mining research. However, in order to determine which algorithm performs best for the chosen characteristics and dataset, the majority of work tested both probabilistic and non-probabilistic techniques for evaluation.

Pang et al. set the groundwork for the field of document-level sentiment classification by using machine learning techniques including SVM, NB, and Maximum Entropy to identify if a movie review indicated a "thumbs up" or "thumbs down" mood. Whether or not a unigram was present, they found that employing binary features coding to classify data produced the best classification accuracy using SVM. SVM is a cutting-edge learning method that is utilized in numerous different systems for sentiment classification because it has been shown to be reliable and efficient in sentiment classification on big feature spaces.

For polarity classification, Wilson et al. employed a boosting classifier called AdaBoost, a rule learning method, and a memory-based learning k-nearest neighbor (KNN). Ortigosa-Hernandez et al. employed Bayesian network classifiers based on three aspects (opinion polarity, subjectivity, and the will to influence) to ascertain a customer's opinion from customer-written evaluations. Using multi-dimensional Bayesian network classifiers, they demonstrated how their suggested method outperforms other widely used sentiment analysis techniques.

As a result, there are numerous study directions for analyzing this kind of data. On the other hand, opinion mining on Facebook datasets has received relatively little study attention. Ortigosa et al. used a Facebook application they designed to gather user comments, messages, and likes. They categorized each user's polarity based on the information they were able to obtain. Esther and colleagues employed political party pages on Facebook to forecast users' political inclinations.

There's been a lot more effort done on sentiment analysis on Twitter datasets than on Facebook. Prior to 2017, tweets could only contain 140 characters; however, in 2017, that number was raised to 280 characters. Therefore, in contrast to other review kinds like product reviews, which are typically longer and more verbose, old tweets are shorter and may contain severe spelling errors. Reviewers typically focus on a single product or entity and don't include a lot of unrelated information. But tweets are typically far more varied, particularly when it comes to sociopolitical topics. Researchers have recently become interested in a number of Twitter opinion-mining tasks, including opinion retrieval, sentiment monitoring over time, irony/sarcasm recognition, emotion detection, and sentiment quantification. By utilizing social and opinion attendees' information, Luo et al. suggested a learning-to-rank model to handle the challenge of retrieving relevant and opinionated tweets toward a user's query. O'Connor et al. computed sentiment scores for each day over a period of time in order to examine the relationship between opinions stated in tweets and public opinion as determined by polls. One of the earliest research projects on irony detection in tweets was given by Reyes et al. A semi-supervised learning technique for tweet emotion identification was introduced by Sintsova et al.

In the past ten years, the modeling stance has advanced significantly. The bulk of research, however, concentrated on three distinct debate environments: (1) corporate internal discussion forums; (2) legislative debates; and (3) online social and political public forums. Using such domains has the benefit of author-provided gold standard labeling.
However, there is a dearth of research on attitude detection for other types of user-generated content on social media platforms, such as Facebook. However, until the SemEval2016 competition1 released a Twitter stance detection dataset, Twitter was likewise not utilized for stance detection. Textual content-based methods see the issue as a binary text classification problem and extract posture information from the text by analyzing textual features like sentiment inferred from lexicons and syntactic patterns. The relationship between postings (disagreement, conflict) or users is used in collective models. Most of the methods for stance detection that are now in use employ supervised learning. Conditional Random Fields and Linear Integer Programming were utilized in addition to SVM and logistic regression in order to capture agreement and disagreement in user interactions. Somasundaran and Wiebe's work involved the creation of a vocabulary that could be used to identify arguments by identifying argument trigger expressions. These taken-out arguments were used as features for stance classification in a supervised learner, along with sentiment expressions and their targets. A rule-based classifier with many features, including punctuation marks, bigrams, unigrams, grammatical dependencies, and the dialogic structure of the postings, was constructed by Anand et al. Walker et al. made use of the dialogic relationships between posts' disagreements and agreements. They suggested a graph-based method in which every node in the graph represents a post and every edge denotes a relationship—either agreement or disagreement—built by the postings' and authors' relationships.

The concept of agenda setting, which is related to the topics that are discussed, and framing, which is related to the perspectives of the authors on those topics, are identified by Nguyen through the use of supervised hierarchical LDA (SHLDA), which simultaneously captures the multi-level topic structure of documents and their polar response variables. Chen and Ku proposed a joint user-post strategy for Facebook post stance tagging, which is one of the few published papers on stance detection on Facebook. The suggested method took into account user-involved posting and liking behavior in addition to post content. After applying sentiment analysis to postings to determine their initial stance, correlated posting-related activities were used to iteratively update both the post and user stance. In order to depict the relationship between users and postings, they created an engagement graph, which they combined with the content of each post to identify the post's viewpoint.

Mining Political Opinions
A significant portion of the initial research on opinion mining was devoted to business and e-commerce uses, like product and movie evaluations. One motivation for concentrating on these domains was the accessibility of labeled data sources, such as star ratings, which may be used as numerical representations of opinions. Eventually, annotated datasets from a variety of sources, including news stories, microblogs, RSS feeds, and more, were accessible. Few studies have attempted to comprehend beliefs within a political framework. Analysis of social and political ideas is more difficult than analysis of product assessments because of the nature of these viewpoints, which includes the use of idioms, irony, and sarcasm.

Maynard and Funk identified similar challenges when examining the outcomes of their opinion-mining method for microblogging posts about politics. On the other hand, a lot of scholars think that ideas shared on social media can be transformed into valuable information sources that the public can use. Consequently, there are considerable prospects for interdisciplinary study between the social science and computer science fields. Among the political domain-focused works, there has been a great deal of interest in the analysis of political tweets to identify sentiment, emotion, and purpose in electoral tweets [80], as well as to ascertain the political alignment of Twitter users, identify divisive issues and political opinions, gauge the degree of polarization among the electorate, and forecast voting intentions or election results.

Through crowdsourcing, Mohammad et al. automatically annotated a set of electoral tweets from the 2012 US Presidential election for information such as sentiment (positive or negative), emotion (joy, sadness, anger, etc.), and the purpose of the tweet (to support, ridicule, point out a mistake, etc.). They used information from cutting-edge sentiment analysis systems to create a supervised classifier that predicted purpose and emotion labels.

The topic of forecasting users' political inclinations on the Twitter network was examined by Makazhanov et al. They proved that one could infer a user's political inclination from the way they behaved toward political parties on Twitter. Every party has a language model that it learns from the content of the tweets sent by its candidates. A user's choice is determined by how closely their tweets match the language models of the parties.

Conover et al. approached the task of forecasting users' political affinities in two stages. In the first stage, they extracted over 250,000 relevant tweets by expanding a list of two popular political hashtags to 66 target keywords using a co-occurrence-based ranking. After that, each user's tweets were sorted, and SVM models based on unigram features were constructed. The accuracy, according to the authors, was roughly 79%; however, if the features were limited to hashtags alone, the accuracy could increase to nearly 91%.

ISSN: 2455-2631  April 2024 IJSDR | Volume 9 Issue 4

IJSDR2404008 | International Journal of Scientific Development and Research (IJSDR) www.ijsdr.org | 66
Maynard and Funk presented an example application they designed to uncover political leanings from a set of pre-election tweets, as well as a number of concerns connected to opinion mining from micro-blogs and the difficulties, they present for a Natural Language Processing (NLP) system. The GATE natural language processing toolbox was utilized in the development of the program to extract the triples of Person, Opinion, and Political Party from every tweet. The individual and political parties were identified using GATE's built-in named entity identification technology. Opinion mining was aided by a lexicon-based technique.

**Conclusion**

Social media provides a forum for civic organizations, non-governmental organizations, and social activists to establish their online presence, share their opinions on various social and political topics, interact with the public, and plan events. Even with the recent achievements in opinion mining, there is still much to learn about this relatively young sector in the political sphere. In this paper, we have concentrated on two sub-aspects of opinion mining: social media attitude detection in the political sphere and fine-grained opinion mining. The majority of current methods for tackling these two problems rely on supervised or semi-supervised models that have been trained using labeled data, which may not always be readily available in practical settings. The identification of the target entity or features of the opinion, as well as the relationships between sentiment/stance, entity, and aspect, are also absent from the majority of earlier studies.

**REFERENCES:**