Quantifying the impact of Twitter activity in political battlegrounds

by

Manmeet Kaur Baxi

A thesis submitted in partial fulfillment of the

requirements for the degree of

Master of Science

in

Computer Science

in the

Faculty of Science and Environmental Studies

of

Lakehead University, Thunder Bay

Committee in charge:

Dr. Vijay Mago (Principal Supervisor)
Dr. Yimin Yang (Co-Supervisor)

Dr. Gurjit Randhawa (External Examiner)
Dr. Thiago E. Alves de Oliveira (Internal Examiner)

Spring/Summer 2022

The thesis of Manmeet Kaur Baxi, titled Quantifying the impact of Twitter activity in political battlegrounds, is approved:

Chair ______________________________ Date ____________

______________________________ Date ____________

______________________________ Date ____________

Lakehead University, Thunder Bay

Declaration

I certify that,

- The work contained in this thesis is original and has been done by myself and under the general supervision of my supervisor(s).
- The work reported herein has not been submitted to any other Institute for any degree or diploma.
- Whenever I have used materials (concepts, ideas, text, expressions, data, graphs, diagrams, theoretical analysis, results, etc.) from other sources, I have given due credit by citing them in the text of the thesis and giving their details in the references. Elaborate sentences used verbatim from published work have been clearly identified and quoted.
- I also affirm that no part of this thesis can be considered plagiarism to the best of my knowledge and understanding and take complete responsibility if any complaint arises.
- I am fully aware that my thesis supervisor(s) are not in a position to check for any possible instance of plagiarism within this submitted work.

Quantifying the impact of Twitter activity in political battlegrounds

Copyright 2022
by
Manmeet Kaur Baxi

Abstract

QUANTIFYING THE IMPACT OF TWITTER ACTIVITY IN POLITICAL BATTLEGROUNDS

It may be challenging to determine the reach of the information, how well it corresponds with the domain design, and how to utilize it as a communication medium when utilizing social media platforms, notably Twitter, to engage the public in advocating a parliament act, or during a global health emergency. Chapter 3 offers a broad overview of how candidates running in the 2020 US Elections used Twitter as a communication tool to interact with voters. More precisely, it seeks to identify components related to internal collaboration and public participation (in terms of content and stance similarity among the candidates from the same political front and to the official Twitter accounts of their political parties). The 2020 US Presidential and Vice Presidential candidates from the two main political parties, the Republicans and Democrats, are our main subjects. Along with the content similarity, their tweets were assessed for social reach and stance similarity on 22 topics. This study complements previous research on efficiently using social media platforms for election campaigns. Chapter 4 empirically examines the online social associations of the top-10 COVID-19 resilient nations' leaders and healthcare institutions based on the Bloomberg COVID-19 Resilience Ranking. In order to measure the strength of the online social association in terms of public engagement, sentiment strength, inclusivity and diversity, we used the attributes provided by Twitter Academic Research API, coupled with the tweets of leaders and healthcare organizations from these nations. Understanding how leaders and healthcare organizations may utilize Twitter to establish digital connections with the public during health emergencies is made more accessible by this study. The thesis has proposed methods for efficiently using Twitter in various domains, utilizing the implementations of various Language Models and several data mining and analytics techniques.

Dedication

This is dedicated to my family, friends, and everyone who believed in me and supported me along the way throughout my academic journey. To my grandfather, Late Mr. Narinder Singh Baxi, for inspiring me to dream big and work persistently to attain that ambition. And finally, my supervisors for their mentorship.

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Description**
AM	Attention Mechanisms
API	Application Programming Interface
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CV	Computer Vision
DL	Deep Learning
G2C	Government to Citizen
GEC	Grammatical Error Correction
GNN	Graph Neural Networks
HDP	Hierarchical Dirichlet Process
LDA	Latent Dirichlet Allocation
LM	Language Model
LSI	Latent Semantic Indexing
LSTM	Long Short Term Memory
MAP	Mean Average Precision
METEOR	Metric for Evaluation of Translation with Explicit ORdering
ML	Machine Learning
MRR	Mean Reciprocal Rank
MT	Machine Translation
NER	Named Entity Recognition
NLP	Natural Language Processing
nltk	Natural Language Tool Kit
NMF	Non-negative Matrix Factorization
NMT	Neural Machine Translation
PHAC	Public Health Agency of Canada
PTLM	Pre-trained Language Model
QA	Question Answering

RNN	Recurrent Neural Networks
ROUGE	Recall-Oriented Understudy for Gisting Assessment
SMOTE	Synthetic Minority Oversampling Technique
SMPs	Social Media Platforms
STS	Semantic Textual Similarity
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverse Document Frequency
U.A.E.	United Arab Emirates
U.S.	United States
VADER	Valence Aware Dictionary for Sentiment Reasoning

Acknowledgments

First and foremost, I want to praise and thank the Almighty for His bountiful Grace and Strength, which allowed me to complete my research successfully. I express my sincere gratitude to my supervisors Dr. Vijay Mago and Dr. Yimin Yang, for their academic, financial and moral support, without which this research would not have been possible. I am grateful for their encouragement and confidence in me, which enabled me to overcome numerous obstacles efficiently. I also extend my gratitude to Dr. Rajesh Sharma (Associate Professor, Software Engineering and Information Systems Group, Institute of Computer Science, University of Tartu) and all of my colleagues at DaTALab in the CASES building at Lakehead University.

I am grateful for all the resources made available to me at DaTALab, the Vector Institute's Artificial Intelligence Award and the NSERC Discovery Grant held by my supervisors, for supporting me throughout my degree.

Last but not least, I want to thank my parents and my sister for believing in my aspirations and for their unwavering love and affection.

Chapter 1

Introduction

This thesis is primarily made up of articles written throughout the degree. The primary goal of these works was to use data mining, analytics, and the implementations of Language Models (LMs) to propose novel approaches to measure the societal impact, engagement, and synergy among candidates using the tweets from various entities in two different real-world scenarios. In Chapter 2, the history, applications, and assessment criteria of Language Models have been condensed since they are widely utilized to assess and comprehend the enormous quantity of information available through different social media platforms (SMPs).

A thorough examination of the candidates competing in the 2020 US Elections is provided in Chapter 3, along with qualitative and quantitative insights regarding their online behaviour. The public metrics of candidates' tweets—the number of likes, replies, retweets, and quotes—are examined in our study to assess their social reach. It is qualitative as we deduce the topics discussed by candidates on Twitter and compare the similarity of the candidate's stance across the two major political fronts, Republicans and Democrats. Therefore, this study aims to identify the topics focussed by candidates in both offline (Presidential debates) and online (Twitter) forums and the level of citizen participation on these issues during various election-related phases. We also look into how the candidates from the same political front differ in their tweet content and how they take a stance on certain topics. Additionally, we look for any connections between internal collaboration and public involvement (similarities in content and stance) that would have assisted the candidates in the running for the 2020 US Presidential election.

People express their thoughts and attitudes—generally referred to as "sentiment"—through various SMPs, including political leaders and healthcare institutions[56]. However, it is unclear how sentiment and allusions to distinct groups could impact information dynamics in a social-media scenario. In order to assess the impact of leaders and healthcare organizations on society, it is crucial to consider sentiment, inclusiveness, diversity as a whole, and public participation. The study in chapter 4 considers all the factors mentioned earlier to gauge the societal association of leaders and healthcare institutions with the citizens of their respective countries during COVID-19 through SMPs, like Twitter.

Chapter 2

Background

All of this chapter will be submitted as the following peer-reviewed journal article:

- Baxi, MK., Mago, V. (2022). How social media has complemented the growth of Language Models? A survey.

Throughout my degree programme, I conducted research on themes related to the confluence between social media use and politics, in general, to broaden my knowledge in the field. Since Language Models are extensively used to evaluate and decipher the humongous amount of information accessible through various social media platforms, I have distilled their history, applications, and evaluation standards in this chapter. On the basis of the information provided here and forthcoming research, I plan to submit a comprehensive survey paper.

2.1 Introduction

The development and use of computational models and procedures to address real-world issues in understanding human languages is known as natural language processing (NLP) [122]. NLP is predominantly a data-driven discipline that incorporates statistical and probabilistic calculations, machine learning (ML) and deep learning (DL) models to achieve a human-like comprehension of [122]. Research in NLP helps address fundamental issues, including language modelling, structural analysis, sentential processing, parsing [140, 238], and semantic analysis [150, 204]. In addition, NLP also focuses on heuristic subjects, like the automatic extraction of pertinent information from enormous volumes of unstructured and low-quality text available across multiple social media platforms, translation across languages, document summarization, automatic question-answering [6, 24], document categorization and clustering [235, 252], and many more.

With the advancement of deep learning (DL), several neural networks, such as convolutional neural networks (CNNs) [125, 49, 109, 87], recurrent neural networks (RNNs) [212, 62, 136, 240], graph neural networks (GNNs) [234, 220, 201], and attention mechanisms (AMs) [101, 228], have been widely employed to tackle NLP tasks. One of the numerous advantages of these neural models is their capability to overcome the feature engineering barrier. Although neural NLP approaches frequently use low-dimensional and dense vectors (also known as distributed representations), the issue of disappearing or exploding gradients [19] has been a stumbling block for employing them compared to the non-neural NLP methods, which heavily rely on discrete handmade features. Specific NLP tasks need to acquire these features. Thus, neural approaches facilitate the development of diverse NLP systems.

Despite the effectiveness of neural models for NLP tasks, the performance gain may not be as substantial as in the computer vision (CV) research community. The fundamental causes for this are the lack of large annotated datasets (except for machine translation), and overfitting the limited training data, resulting in poor generalization performance for the models [16, 242]. As a result, incipient neural models for many NLP tasks, such as Word2Vec [146, 147, 149] and GloVe [167], were comparatively shallow and often only included very few neural layers. While these pre-trained word embeddings are crucial for many NLP applications, they restrict the representation of polysemous words in varied contexts because a single dense vector represents each word. For instance, the word "*bank*" has distinct meanings in the sentences "The *bank* raised it's interest rates yesterday" vs "Mary walked along the *bank* of the river". This stimulates pre-training RNNs to provide contextualised word embeddings [98, 143, 170], but their performance is still constrained by model size and depth.

With the advancement of deep neural networks in the NLP community, the advent of Transformers [219] and pre-trained language models (PTLMs) [177] makes it feasible to train very deep neural models for NLP applications. The NLP community has hence adopted self-supervised learning [137] to create PTLMs to reap the maximum benefits of diverse linguistic information that can be yielded from large unlabeled corpora for NLP tasks. Utilizing inherent correlations in the text as supervision signals with minimal labor rather than human supervision is the driving force behind self-supervised learning. In essence, this self-supervised environment adheres to the well-known language model learning [18].

In terms of inferring from large-scale PTLMs like GPT [181], BERT [59] and GPT-3 [37], increasing model size and parameters allows machines to interpret the language better to capture polysemous disambiguation, lexical and syntactic structures, and factual knowledge from text. Calibrating these PTLMs on previously neglected but readily available social media data as is the case of ArabicTransformer [6], TweetBERT [180], HateBERT [43], BanglaBERT [24], and many more [154, 13, 85], helps widen the linguistic understanding and application of PTLMs with outstanding performance.

The performance of the models on numerous NLP tasks (such as, text classification, named entity recognition, part-of-speech tagging, semantic and syntactic analysis) has been enhanced by the current large-scale PTLMs, which have even challenged our preconceived notions about how well deep learning models function. We still do not fully understand the nature of the enormous number of model parameters, and the high computing cost of training these giants also prohibits us from going further in our research. However, training them with publicly available, unlabeled and humongous corpora from various social media platforms (like Twitter, Reddit, Instagram, and Facebook) might justify the prudent use of the available intuitive models. Most recent survey studies provide extensive details about language models relating to a specific downstream NLP task [245, 22, 172], their societal influence and release techniques [209], their applicability in a particular area [111, 42], or the potential threats to these large-scale models [83]. This research aims to present an empirical overview of how social media data has complemented the growth of PTLMs, including the most recent developments based on PTLMs utilizing these datasets. This study also examines how PTLMs have changed throughout the years, differentiating them according to the underlying models, datasets, and parameters.

2.2 Language Models

PTLM's serve as the foundation for many NLP tasks. Language models (LMs) are essentially probability distributions over words or phrases. For instance, in tasks involving machine translation, PTLMs are used to assess the likelihood of the model's output to increase translation fluency in the target language. Bengio, Ducharme, and Vincent suggested a feed-forward neural network that can anticipate the next word in a series as a conventional neural language model [18]. Their model is a preliminary prototype that has been refined steadily over time via various subsequent modifications. The following sections discuss the evolution of Language Models alongwith their applications in various domains.

2.3 Brief History of Language Models

The word "*language model*" emerges from the probabilistic language synthesis models created for automatic voice recognition systems in the early 1980s [4]. A language model is used to supplement the results of an acoustic model, which simulates the relationship between words (or phrases, called phonemes) and the auditory signal in speech recognition applications. However, language models have a long history dating back to Andrei Markov, who utilized language models (also known as Markov models) to simulate character sequences in Russian literature in the early 1900s [9]. Claude Shannon's models of words and phrases, which he used to emphasize the significance and inferences from coding and information theory, are another well-known application of language models [14]. In 1982, John Hopfield proposed the Recurrent Neural Network (RNN), which was utilised for text- or voice-based computations on sequential data [97]. The earliest concepts for expressing words as vectors appeared in 1986 when Geoffrey Hinton, carried out these investigations [141]. Language models were used as a generic tool in numerous NLP applications in the 1990s, including machine translation, part-of-speech tagging, and speech recognition. Several research groups employed language models for information retrieval in the late 1990s [20, 27, 92]. They gained popularity in information retrieval studies swiftly. In 1997, Hochreiter and Schmidhuber first proposed the concept of Long Short Term Memory networks (LSTM) [93]. However, there was still a dearth of computational capacity during this time period to properly employ the neural language models to their full potential. By 2001, there were two different sessions on language models at the ACM SIGIR conference, with a total of five articles [95]. A research roadmap entitled "Challenges in information retrieval and language modelling" [103] was released in 2003 by a group of eminent information retrieval experts, stating that the futures of information retrieval and language modelling cannot be understood in isolation.

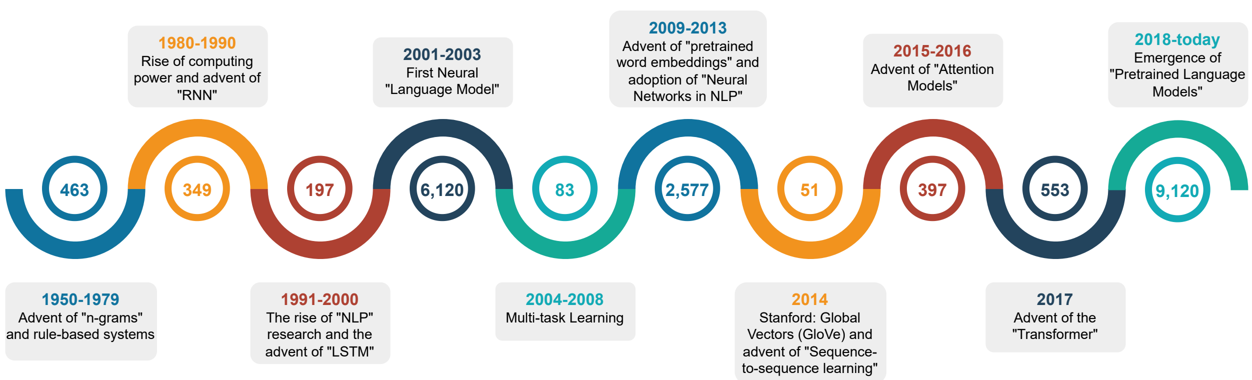


Figure 2.1: Key milestones in the evolution of Language Models (The numerals in the circles represent the number of research articles listed on *Google Scholar* as a search result for a particular term during that time-period.)

Bengio, Ducharme, and Vincent proposed the world's first feed-forward neural network language model in 2003 [18]. In order to anticipate the subsequent word in a sequence, their model uses a single hidden layer feed-forward network. Although feature vectors were already in use at the time, Bengio, Ducharme, and Vincent were the first who popularized the idea. Although conventional feed-forward neural networks have been gradually displaced by RNNs [148] and LSTMs [76] for language modelling, the latter are still competitive in some contexts due to "*catastrophic forgetting*", i.e., the propensity of a neural network to entirely and suddenly loose previously learned information, when it learns new information [58]. Additionally, most of the contemporary neural language and word embedding models still use the basic building block of Bengio, Ducharme, and Vincent's network. Afterwards, Collobert and Weston implemented multi-task learning to neural networks for NLP in 2008, a branch of ML where many learning tasks are handled concurrently [51, 52]. They employed a single CNN, capable of producing a variety of language processing predictions, including part-of-speech tags, named entity tags, and semantic roles, when given a text. Multi-task learning has grown in significance and is currently employed widely across NLP as models are being assessed on several tasks to determine their generalization capacity. It pioneered concepts like pre-training word embeddings and employing CNNs for text that have recently gained widespread acceptance. Figure 2.1 provides a summary of the significant milestones in the evolution of language models over time. The numerals in the circles represent the number of articles listed on *Google Scholar* as a search result for a particular term during that time-period. For example, during 1991-2000, 197 papers were listed as a search result for the term "NLP and Long Short Term Memory (LSTM)".

The year 2013 saw the acceptance of three well-defined forms of neural networks in NLP: *vanilla RNN* [65], *CNNs* and *RNNs* [208]. RNNs gained popularity for handling the dynamic input sentences, common in NLP because of their structure. However, the traditional LSTMs rapidly substituted vanilla RNNs since they were more immune to the disappearing and expanding gradient issue [93].As CNNs were extensively used by researchers in the CV community, they were now also being used to analyze natural language [109, 123]. Since the state at each time step just depends on the local context (by the convolution operation) instead of all the previous states as in the case of RNNs; CNNs have the benefit of being more parallelizable than RNNs when dealing with text sequences. Finally, RNNs were motivated by the idea that human language is essentially hierarchical, i.e., words are assembled into higher-order sentences, which may then be recursively concatenated as per a set of production rules. This linguistic viewpoint led RNNs to consider sentences as trees instead of as a series of events.

Sequence-to-sequence learning, a broad end-to-end strategy for mapping one sequence to another using a neural network, was introduced in 2014 [212]. According to this technique, a phrase is parsed word-by-word and compressed into a vector form using an encoder neural network. The output sequence is then predicted symbol by symbol by a decoder neural network using three attributes – the encoder state, the symbols that were previously predicted and input at each step. Although alternative designs have also emerged, RNNs often provided the foundation of sequence encoders and decoders. Deep-LSTMs [240], convolutional encoders [110, 73], the Transformer [219], and a hybrid of an LSTM and a Transformer [48] are examples of recent models. Sequence-to-sequence learning turned out to be the ideal application for machine translation, and the advancements were so substantial that Google made the formal announcement in 2016 that it was switching *Google Translate* on to a neural sequence-to-sequence model for eight language pairs, officially replacing its rigid phrase-based machine translation algorithm [217].

The notion of attention mechanisms, one of the fundamental advancements in neural machine translation (NMT), was first proposed in 2015 [7]. This concept is crucial to understanding how NMT models outperform traditional sentence-based machine translation (MT) systems. It essentially removes the primary obstacle of sequence-to-sequence learning, i.e., the need to compress the full source sequence's information into a fixed-size vector. In fact, attention mechanism enables the decoder to review the source sequence concealed states, which are then merged by a weighted average and delivered as extra input to the decoder. Attention mechanism has the potential to be effective for any activity that needs making judgments based on specific sections of the information. Until now, it has been used in syntactic constituency parsing [223], reading comprehension [91], and one-shot learning [224]. Self-attention, which is at the centre of the Transformer architecture, is a new type of attention mechanism that has lately emerged. In order to get better contextually aware word representations, it is utilised to look at the words that are immediately adjacent in a phrase or paragraph.

Large pre-trained language models are undoubtedly the most significant development in the field of NLP recently. They were initially proposed in 2015 [55], but it wasn't until recently that it was demonstrated that they outperformed state-of-the-art approaches across a wide range of tasks. In order to enable efficient learning with substantially less data, pre-trained language model embeddings can be employed as features in a target model [170] or a pre-trained language model can be fine-tuned using target task data [60, 99, 182, 253]. These pre-trained language models' key benefit is their capacity to learn word representations from sizable unannotated text corpora, which is especially helpful for low-resource languages or where labelled data is rare, as in the case of social media. Thus, language models have evolved from probability models (such as, TF-IDF [105]), through n-grams [36] subsequently, and finally to the standard models like, CNNs, RNNs, LSTM, GNNs, and many more, to tackle NLP applications. Recent attempts have explored NLP tasks in a generic model with the increasing volumes of social media data available. Kalyan, Rajasekharan, and Sangeetha and Qiu et al. presented a detailed review of the PTLMs applicability in the field of natural language processing in the past, present and how they could be leveraged in the future [112, 177].

2.4 Applications of Language Models

Language models are the foundation of various NLP tasks and many models have used social media data. For instance, models like HateBERT [43] and BERT-SentiX [251] have used data from Reddit(Rale-E) [43] and product reviews from Amazon [251] as a pre-training corpus for abusive language detection and cross-domain sentiment analysis, respectively. The following list of language modelling-based NLP tasks includes an explanation of each task's purpose as well as examples of its applications:

Natural Language Inference (NLI)

NLI is an essential NLP activity that necessitates the knowledge of sentence-level semantics. It is structured as a three-way classification problem of sentence pairs. NLI determines how a pair of statements relate to one another, such as if the second sentence implies, contradicts, or is coherent with the first. When a model is trained using NLI datasets, it gains knowledge of sentence-level semantics, which is helpful for various tasks, including question answering [6, 24], normalizing concepts in specific domains [114, 218], information retrieval [39, 184], and paraphrasing [140, 238]. Models like BERT [59] combine the representation of the two phrases to predict the connection between the supplied sentence pair using a three-way task-specific softmax classifier. For instance, Kecht et al. utilized Twitter interactions between the consumers and support representatives at prominent companies (like, Apple, Spotify and

Amazon) to create conversation logs using NLI and subsequently deduce topics and process subsequent actions [117].

Entity extraction

Entity extraction is the first step in obtaining relevant information from unstructured text data. It is helpful for various applications such as entity linking, relation extraction, knowledge graph generation, etc. Both the clinical field and recently low resource languages like Hindi-English [207], Telugu-English [211], and Arabic [102] have made extensive use of entity extraction. In the field of biomedical sciences, entity extraction has been used to extract data from scholarly publications about proteins, chemicals, and drugs [90]. Additionally, a number of code-mixed Twitter datasets for low resource languages have been made available for the Named Entity Recognition (NER) and extraction task using LSTMs, Bidirectional LSTMs (Bi-LSTMs) and BERT [102, 207, 211].

Semantic textual similarity

Semantic textual similarity (STS) quantifies the degree of semantic resemblance between two phrases or sentences. In contrast to NLI, which assigns the provided sentence pair to one of three groups (i.e., if the sentence pair implies, contradicts or is coherent), STS yields a numerical value of the degree of similarity for the sentence pair. Sentence-level semantics are required for both NLI and STS. Concept relatedness [113], concept normalization [114, 218], duplicate text identification [151], question answering [6, 24] and text summarizing [213] are a few scenarios where STS is helpful. Reimers and Gurevych have also demonstrated that training transformer-based PTLMs on STS datasets enables the model to acquire sentence-level semantics and subsequently improve the representation of variable-length texts like phrases or sentences [184]. Models such as BERT learn the combined representation of a particular phrase pair, and a task-specific sigmoid layer provides the similarity value. CORD19STS [84], MedSTS [233], and the Arabic [150] and Bengali [204] twitter datasets by Mohammad et al. and Shajalal and Aono are a few examples of social media data used for training LMs for the STS task.

Text categorization/ classification

Text categorization/ classification requires tagging variable-length texts, such as phrases, sentences, paragraphs, or pages, with one of the predetermined labels. The process of text classification requires a task-specific softmax classifier, an encoder (often a transformer-based PTLM), and text. The weighted sum of the final hidden state vectors, or "[CLS]" vector, represents the given text as a whole. The classifier's fully connected dense layer projects the text representation vector into n-dimensional space, where n is the number of predetermined labels. Then, the softmax function is used to determine the probability of all the labels. Zia et al. fine-tuned an XLM-R (Large) based model for zero-shot cross-lingual hate speech detection from English to six different target languages (Spanish, Italian, German, Arabic, Greek and Turkish) using Twitter and Reddit datasets (AMIEvalita2018 [67], GermEval2018 [236], HatEval 2019 [15], OffensEval2020 [246]) [252]. Whitehouse et al. proposed a knowledge-enhanced LM [235] based on ERNIE [249], KnowBERT [169], KEPLER [232] and K-ADAPTER [230] to evaluate fake news detection over two social media based datasets – LIAR [231] and COVID-19 [45].

Question Answering

Question answering (QA) is the process of eliciting answers to the prompt questions. QA assists in swiftly extracting knowledge from notes or books, saving a significant amount of time. Qi et al. published a Chinese dataset for open-domain visual question answering [176]. Xiong et al. also released a Twitter-based QA dataset targeted toward the Twitter profiles of the journalists who write articles for major news channels (CNN, NBC) in four sections – World, Politics, Money and Tech [241]. Oniani and Wang leveraged the COVID-19 Open Research dataset (CORD-19) [229] with GPT-2 [182] and transfer learning to evaluate LMs on automatic QA [158].

Grammatical Error Correction

Grammar Error Correction (GEC) is the process of rectifying textual flaws such as spelling, punctuation, grammatical issues, and word choice. A GEC system accepts a possibly incorrect statement as input and rectifies its grammatical errors. ESC [178], TMTC [126], T5 [187], GECTOR [157] and UA-GEC [214] are a few examples of the models that have helped in multi-lingual GEC for languages like English, Arabic, Czech, German and Russian.

Relation Extraction

Identifying the semantic relationships between text items is a vital step in information extraction known as relation extraction. Many tasks, such as knowledge network creation, text summarization, and question answering, benefit from extracting relations between entities. Unstructured text may be transformed into structured data using entity and relation extraction. He et al. presented a PTLM enhanced on knowledge graphs, KLMo to extract relationships between entities [88].

2.5 Performance and Evaluation Metrics

During pre-training, a LM acquires information about the syntactic, semantic, factual and common-sense knowledge encoded in the pre-training corpus. Both intrinsic and extrinsic methods can be used to assess the efficacy of a PTLM. Intrinsic evaluation examines the information encoded during pre-training in PTLMs, whereas extrinsic evaluation assesses how well the PTLMs perform on the downstream tasks in practice. The intrinsic evaluation of PTLMs enables the model developers to create more effective pre-training tasks and increase the knowledge the model acquires during pre-training.

Intrinsic Evaluation

Intrinsic evaluation concentrates on intermediate goals (i.e., how well an NLP component performs on a specific task). For instance, when completing the word vector analogies using cosine similarity, the word vector with the maximum cosine similarity is given preference over others. . During model building, testing, and deployment, various ML metrics typically satisfy the performance evaluation requirements for basic LM tasks. However, additional probing-based and quality-based methods help gauge the performance of PTLMs as per different applications.

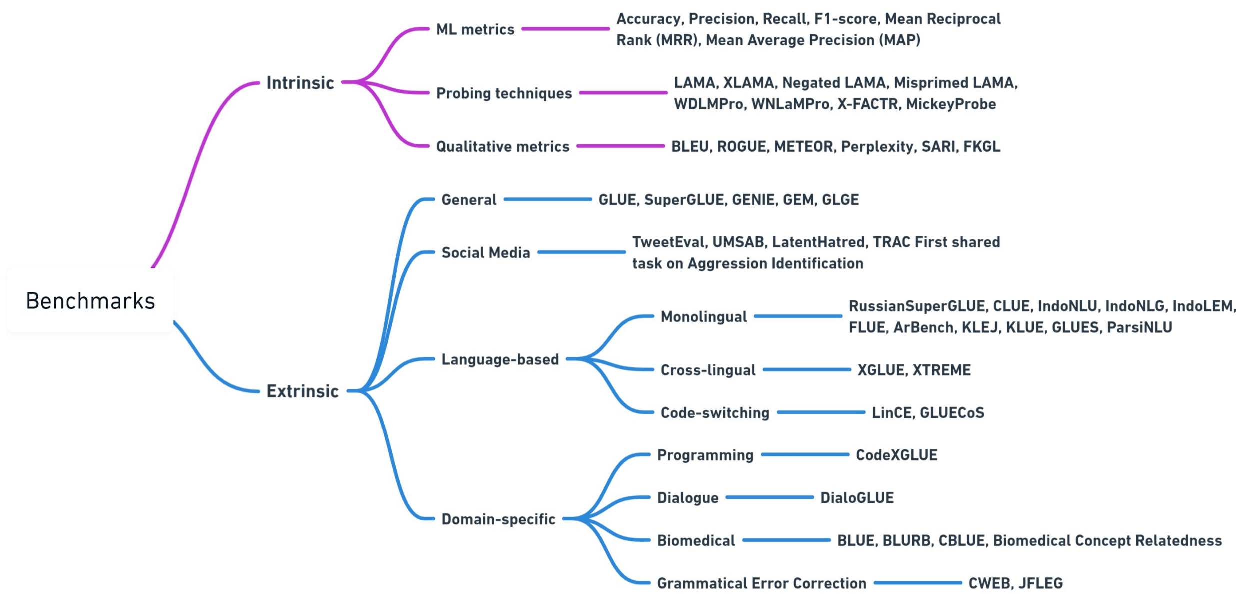


Figure 2.2: Benchmarks for evaluating the development of PTLMs (adapted from [112])

ML metrics like Accuracy, Precision, Recall and F1-score (macro, micro, and weighted) help the PTLM understand the closeness of an observed value to a known value. Therefore, these metrics are frequently applied when the output variable is categorical or discrete, such as in classification or categorization tasks. The Mean Reciprocal Rank (MRR) and Mean Average Precision (MAP) are frequently used for information retrieval tasks. MRR ranks the responses concerning a question based on their likelihood of correctness, and MAP determines the mean precision across all the results retrieved.

To examine the knowledge attained by models, intrinsic assessment also entails probes like LAMA [171], XLAMA [115], X-FACTR [104], MickeyProbe [132], Negated LAMA [116], Misprimed LAMA [116], WDLMPro [202], or WNLaM-Pro [200], to name a few. LAMA was one of the first probes to assess factual and common-sense information in PLTMs in zero-shot conditions. It comprises a corpus of facts where each fact might be either a question-answer pair or a triplet of relations retrieved from SQUAD. LAMA is predicated on the supposition that a model with adequate factual information accurately predicts the blank tokens, i.e., the ground truth tokens are predicted with the highest probability compared to other tokens in the model vocabulary. The Negated LAMA and Misprimed LAMA probes reveal that the language models cannot consider negated or misprimed terms in the templates. For instance, regardless of whether the template is negated or not, the model predicts the same token. Poerner, Waltinger, and Schütze proposed LAMA-UHN as a collection of difficult-to-guess triples from the LAMA probing benchmark [173]. The LAMA probing technique is enhanced by XLAMA to 53 more languages and multi-token entities [115]. The model must predict over the full model vocabulary in LAMA, but in XLAMA, the model must predict over a defined collection of candidates appropriate to each relation type. This form of querying is known as TypedQuery (TQ), while UnTypedQuery (UnTQ) is used in LAMA [171]. X-FACTR, like XLAMA, is a multi-lingual probe that supports 23 languages. Additionally, to forecast multi-token entities, the developers of X-FACTR have proposed several decoding algorithms. A zero-shot common-sense probe called MickeyProbe [132] employs sentence-level ranking based on pseudo-likelihood [193]. Thus, the model rates a collection of declarative sentences with related terms and grammatical elements.

Assessing inter-entity relationships is the main emphasis of probes like LAMA, XLAMA, and X-FACTR. In contrast to these probes, WDLMPro and WNLaMPro concentrate on comprehending how the pre-trained models interpret the words. WDLMPro comprises synset groups, each containing a word, its taxonomic sibling term from WordNet, and their meanings. The WDLMPro probe is predicated on the idea that a model can only accurately match a word with its meaning when the model comprehends the term. While WDLMPro analyses the model by matching a word with its description, WNLamPro employs templates that require filling in the blanks.

The iterative nature of machine translation (MT) system development needs frequent assessments to provide immediate feedback on the effectiveness of continuously evolving development strategies for LMs. Therefore, it becomes essential to use automated quality measurement metrics like BLEU, ROUGE, and METEOR to speed up the feedback process for the researchers and developers of data-driven MT systems. Bilingual Evaluation Understudy (BLEU) measures the quality of MT systems that aim to assess the degree of agreement between a machine translation's output and a human translation [159]. BLEU is based on the core principle that a machine translation is better when it is more similar to a qualified human translation. BLEU ratings indicate how an MT model performs on the specific subset of source texts and translations chosen for the test. Metric for Evaluation of Translation with Explicit ORdering (METEOR) addresses some of the shortcomings of the BLEU score, such as the need for precise word matching while computing the accuracy of MT models for a given subset of data [10]. The METEOR score also enables matching synonyms and stemmed words with a reference term. Unlike the BLEU score, the Recall-Oriented Understudy for Gisting Assessment (ROUGE) evaluation metric examines recall. It provides methods (such as ROUGE-N, ROUGE-L, and ROUGE-S) for automatically determining a summary's quality by contrasting it with other (ideal) summaries generated by people [133]. The metrics tally the quantity of n-grams (ROUGE-N), word sequences (ROUGE-L), and word pairings (ROUGE-S) included in the ideal human-written summaries and the computer-generated ones being assessed.

Extrinsic Evaluation

Extrinsic assessment aids in determining a model's performance in subsequent tasks. To get the most out of a model, it should perform efficiently across various activities rather than just one or two. Furthermore, a benchmark is required to examine the differences in the performance across models on a specific task and better understand the underlying problems in existing models. A benchmark offers a consistent method of assessing how well the model generalizes across tasks by evaluating the overall performance of the models using a single metric. It typically consists of a collection of datasets, a leader board, and a single statistic [225]. The datasets chosen for the benchmark are demanding and representative of various activities. A leaderboard is a database that allows users to compare and rate models. To perform well in a benchmark, a model must communicate information, i.e., parameters across tasks, with one or two layers specialized to each task [225]. Without a benchmark, it is challenging to assess models consistently and monitor the advancement of PTLMs.

GLUE [225] and SUPERGLUE [226] benchmarks are frequently used to measure how well the PTLMs can comprehend natural language. The GLUE benchmark consists of nine problem tasks, including sentence pair and single sentence challenges. With the rapid advancement in model development, the models obtained good performance in the GLUE benchmark, leaving minimal room for additional improvement [226]. Inspired by the success of the GLUE and SuperGLUE benchmarks in general English, benchmarks like as GENIE [120], GEM [74], and GLGE [134] have been introduced to evaluate NLG models in general English. The XGLUE [130] and XTREME [100] benchmarks have been established to assess cross-lingual models. In contrast to the XGLUE benchmark, which contains both XNLU and XNLG workloads, XTREME benchmark only covers XNLU jobs. The XGLUE benchmark is also more demanding and valuable since it uses various datasets relevant to the search, advertisements, and news contexts.

For evaluating social media-based datasets, and PTLMs, benchmarks like TweetEval [13], UMSAB [12], and LatentHatred [66] have been proposed. While UMSAB has datasets from eight languages, including English, TweetEval exclusively contains datasets from the English language. Both benchmarks' tasks are presented as tweet categorization/classification. Besides the presence of XGLUE and XTREME to evaluate cross-lingual models, we have different benchmarks in each language to assess monolingual language models, such as Russian (Russian SuperGLUE [205]), Indian (IndicGLUE [108]), Chinese (CLUE [243]), Indonesian (IndoNLU [237], IndoNLG [40], IndoLEM [124]), French (FLUE [128]), Arabic (ArUE [1]), Polish (KLEJ [191]), Korean (KLUE [160]), Spanish (GLUES [41]), and Persian (ParsiN [121]). Furthermore, we have benchmarks such as GLUECoS [119] and LinCE [2] for evaluating CodeSwitching; BLUE [166], BLURB [82], and Chinese-BLUE [248] for the Biomedical models; CodeXGLUE [138] for the code intelligence domain; and DialogGLUE [142] to evaluate the dialogue-based models. Furthermore, to assess T-PTLMs in the few shot scenarios, benchmarks like FewCLUE [244], FewFLEX [35], FewGLUE [199], and RAFT [3] have also been proposed recently.

The study in the following chapters aims to carefully examine the publicly accessible data on Twitter, analyze, evaluate and propose novel approaches to understand the engagement, synergy, and influence of diverse individuals in two very different real-world circumstances, an election and a healthcare emergency. The articles presented in the following chapters, investigate two scenarios – first, the 2020 US Elections, and second, the resilience of political leaders and healthcare institutions during COVID-19. The articles will use Twitter datasets and applications of LM embeddings and LMs for tasks like machine translation, stance detection, and stance classification.

Chapter 3

Studying topic engagement and synergy among candidates for 2020 US Elections

All of this chapter is under revision at a reputed journal as:
- Baxi, MK., Sharma, R., & Mago, V. (2022). Studying topic engagement and synergy among candidates for 2020 US Elections.

This chapter gives a comprehensive overview of how candidates contesting for 2020 US Elections employed Twitter as a communication tool to interact with potential voters. It seeks to identify the factors related to public engagement and internal cooperation using the attributes available through the Twitter Academic Research API and the tweets from the Presidential and Vice-Presidential candidates contesting the 2020 US Elections. The internal cooperation is measured in terms of content and stance similarity among the candidates from the same political front and to the official Twitter accounts of their political parties. This study adds to the existing work on using social media platforms for electoral campaigns and can be effectively utilized by contesting candidates.

Keywords: social media, electoral campaigns, public engagement, content similarity, 2020 US Elections

3.1 Introduction

Social media platforms (SMPs) like Facebook, Twitter and Instagram have become the conventional modes of online campaigns for elections after being first used by Barack Obama while contesting for his 2008 candidacy [64]. Researchers from around the world have expressed a strong interest in analyzing and evaluating social media data in the context of elections across these different platforms, as discussed in the works of [34], [189], and [222].

Various politicians have widely used SMPs to express their views on current topics, share the latest developments in their constituencies, and communicate with their potential voters strategically. From the statistics in [44], out of the top 10% adult Twitter users in the US, 92% of them are politicians, with Democrats or Democratic-leaning independents being hyperactive and capturing 69% positions in the top Twitter users; while Republicans or Republic-leaning independents occupying the remaining 26%. Additionally, recent studies have shown that maintaining an active presence on SMPs has helped politicians address social concerns and build a stronger relationship with the audience [29], [80], [192]. Furthermore, the cooperation among the candidates from the same political front has helped them communicate their policy initiatives clearly and organize support from the related interest groups [78]. The authors of [239] emphasize the benefits of cooperation among the political parties and interest groups in European Union policy-making by examining their information networks. Hence, there is a requirement to investigate the influence of citizens' engagement and the internal cooperation among the politicians on the election results. The authors of [33, 155] have stressed the importance of a future qualitative study to quantify the genuine impact of social media on Government to Citizen (G2C) interactions. As a result, the goal of this research is to measure the utility of Twitter for politicians during the 2020 Presidential elections in the United States. In particular, we investigate the following two research questions:

Research Question 1 – What topics did the candidates discuss through online (Twitter) and offline (Presidential debates) mediums? How engaging were these topics and to what extent during the different phases of the electoral campaign? Understanding what information and topics appeal to the audience the most, may be an effective method for gaining attention and increasing involvement [31]. The authors of [28] found that certain types of contents are more engaging than others. Therefore, identifying such materials and developing thorough plans for maintaining a continuous dialogue with the citizens, responding to their grievances, recommendations and wants, would help improve governance quality. Following the identification of objectives and goals, norms could be produced to facilitate the contesting candidates with an effective tool for communication on SMPs. If implemented properly, it is indeed a win-win situation for both the candidates and the citizens. On that account, we calculate the impact of candidates and the engagement received by them on various topics, followed by classifying them according to the topic stickiness in different phases of the election campaign.

Findings: Joe Biden was the most impactful candidate among all, and Democrats tweeted more about topics of public interest during the electoral campaign as compared to Republicans. The detailed observations can be found in Section 3.4.

Research Question 2 – Did the candidates from the same political front have similarities in their tweets and the stance for the topics with respect to thier political front ?

Politicians collaborate to share resources and coordinate political support. Wonka and Haunss highlight the various types of cooperation networks formed inside a political front during the European Union policy-making [239]. Furthermore, the smaller networks inside a political party or interest group may reconfigure themselves based on the reputation (impact of the candidates) and the internal reciprocity (similarity in thoughts/actions) [8], [70], and [77]. Thus, to analyze the synergy among the candidates during the electoral campaign of 2020 US Elections, we employ two methods, i.e., content similarity-based on the tweets, and stance similarity-the standpoint of candidates with respect to different topics. Understanding these aspects would help us identify which political front was more cooperative among themselves and echoed similar thoughts on Twitter.

Findings: Kamala Harris depicted a higher amount of cooperation with both Joe Biden and the official account of Democrats in both - content and stance. On the other hand, Republicans portrayed comparatively lower synergy in their stance with respect to different topics. Refer to Section 3.5 for more details.

This research provides a comprehensive analysis of the contesting candidates, a combination of both qualitative and quantitative insights into their online behaviour. Previously, this form of hybrid research has proven to be beneficial [194]. Our study uses statistical methods to examine the public metrics of candidates' tweets (the number of likes, replies, retweets, and quotes) and determine their social reach. It is qualitative as we infer the topics the candidates tweeted about and the similarity in the content and stance of the candidates from the same political front. Therefore, the objectives of this study are to discover the topics discussed by the candidates through online (Twitter) and offline (Presidential debates) channels, as well as the civic engagement on these topics during the different phases of the electoral campaign, and throughout the whole election campaign. Additionally, we also investigate the similarities in the tweets of the candidates from the same political front with respect to the tweet content and their stance on the topics. Furthermore, we try to uncover any relationships between public engagement and internal cooperation (content and stance similarities) that might have aided the candidates in contesting the 2020 US Presidential elections. Figure 3.1 presents an overview of the research framework.

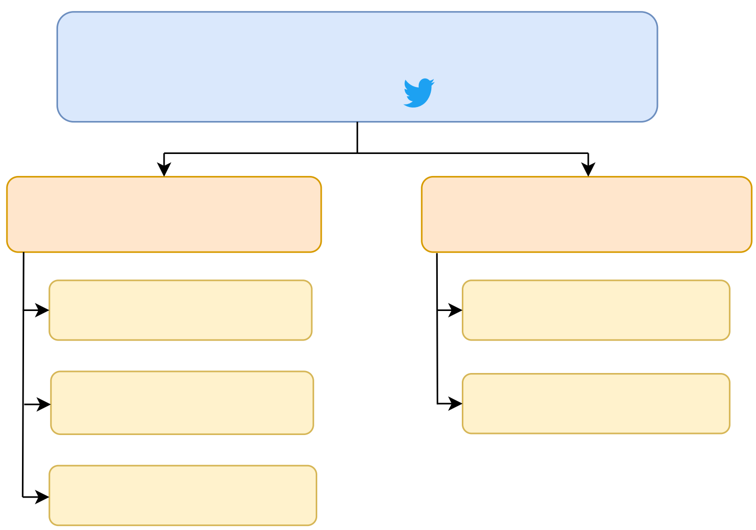


Figure 3.1: Overall research framework

The remaining part of the article is arranged as follows: Section 3.2 gives a summary of prior work linked to the use of SMPs in the electoral campaigns. Section 4.3 explains the data utilized in this study. The techniques used and insights of public engagement during the electoral campaign are presented in Section 3.4. Section 3.5 discusses the approach followed and observations for the identifying the collaboration among the candidates from the same political front. Finally, the key inferences and further research directions are presented in Section 3.6.

3.2 Related Work

The analysis on elections has been widespread across different SMPs, like authors of [38] compared the candidate and audience activity on Twitter for the 2013 and 2016 federal elections in Australia. Additionally, Dzisah studied the role of SMPs in enhancing democratic participation during the 2012 and 2016 Ghana elections [63]. Praznik et al. took a different approach by analyzing the strength of networks based on the usage of hashtags on Twitter during electoral campaigns [174]. The researchers of [68] tried to contrast different opinion groups using network representations for replies and retweets in the context of Saxon state elections and violent riots in the city of Leipzig, Germany in 2019 and, Bilal et al.'s work surveyed the current state-of-the-art approaches to analyze the election prediction mechanisms in use [26]. As far as SMPs are concerned, researchers of [34] investigated the use of Facebook for campaign strategies used in 2008 and 2012 US Presidential elections, authors of [189] analyzed Instagram for Swedish elections, and the authors of [222] studied the use of YouTube as an advertising tool during the campaign of European Parliament elections.

According to previous studies ([21], [33], [50]), SMPs can help enhance the transparency, involvement, and correspondence in governance. Also, researchers have revealed the influence on different features of public interaction through various instances ([81], [96], [185]). Several authors ([29], [30], [32], [80], [192], [206]) have highlighted the relevance of SMPs as a vital instrument for amplifying social reach, and to help understand the audience better. However, earlier studies have also found that the sentiment, emotion, stance of the tweets, the promotion of tweets by bots; and collaboration between interest groups (polarization) may mitigate the impact of different aspects of public involvement and steer the change in public opinions ([23], [69], [79], [195], [196]).

Internal cooperation and collaboration play an essential role for a political party to convey their policy initiatives during the electoral campaigns and gather support. The survey by Khanam et al. explores various methodologies proposed till date on the usage of the *Homophily* principle (likelihood of similar-minded people to engage with one another in communities) across different domains [118]. Another survey by Chandrasekaran and Mago lists the different methods available to evaluate the semantic similarity between texts, ranging from traditional Natural Language Processing (NLP) techniques to deep neural-network-based hybrid methods [46], out of which, we alter one method as per the objectives of this study, to measure the internal cooperation among candidates based on their tweets. However, several factors affect the conflicts and synergy within a group as highlighted in the studies ([77], [127], [139], [168]). It is important to identify them for effective operations and governance. Our study evaluates the cooperation among the candidates by comparing the content of their tweets and collating the similarities in their stance on different topics. We also uncover the relationship between internal synergy and electoral campaigns.

Formerly, researchers of [11], [106], [156], and [216] have performed sentiment analysis on tweets using machine learning techniques, like lexicon-based models–VADER, and decision trees to predict the election results by focusing on a single aspect. Additionally, [47] have released a dataset for analyzing the 2020 US Elections. However, there has not been an empirical analysis to understand and examine the utility of Twitter (with emphasis on public participation and internal cooperation) as a communication tool during the 2020 US electoral campaign, considering different factors like social reach and internal cooperation (stance and content similarity), which is hence the focus of this work.

3.3 Dataset

We collected a total of 99,784 tweets authored from the accounts of Presidential and Vice-Presidential candidates of two major political fronts—Republicans and Democrats using the Twitter API v2[1], during the time frame of January 21, 2019, to January 27, 2021. Additionally, the tweets created by the official Twitter handles of both the political fronts were also collected. The candidates selected for our analysis and the number of tweets scraped from their accounts are discussed in Table 3.1. For the scope of our research,

[1]https://developer.twitter.com/en/docs/twitter-api

the political fronts (Democrats/Republicans) confine to the Presidential candidate, Vice-Presidential candidate, and the official Twitter handles of the political fronts.

Political Party	**Candidates (Twitter Handle)**	**Number of Tweets**
Democrats	*Presidential Candidate*: Joe Biden (@JoeBiden)	5,486
	Vice Presidential Candidate: Kamala Harris (@KamalaHarris)	5,835
	Official Twitter Handle (@TheDemocrats)	30,465
	Total	*41,786*
Republicans	*Presidential Candidate*: Donald J. Trump (@realDonaldTrump, @POTUS)	21,007
	Vice Presidential Candidate: Mike Pence (@Mike_Pence, @VP, @VP45)	12,003
	Official Twitter Handle (@HouseGOP, @GOP)	24,988
	Total	*57,998*

Table 3.1: Tweet distribution of the candidates selected from both the political fronts.

3.4 Engagement and Stickiness of Topics

Identifying topics

We analyze the most and least discussed topics by the candidates from the political fronts through two sources – *offline* and *online* as defined below:

1. The *offline* source is the topics that were discussed in the Presidential debates by both the candidates as given by 'The Commission on Presidential Debates'[2] and the events synchronous with the US Elections (Current/Snapshot events),

2. The *online* source of topics is topic modelling on the tweets authored by the candidates. We first preprocess, and then cluster the tweets using various clustering algorithms and leverage the topics yielded by the best-performing topic model.

[2]Topics for first presidential debate,Topics for second presidential debate

Topic Source	Topic Category	Topic	Abstract Category
Online	Modelled Topics (Topics generated from NMF)	Legalization of Medical Marijuana	Social Issues
		Equality rights for LGBTQ	Social Issues
		Weapon Ban	Social Issues
		Build Back Express Tour	Social Issues
		Affordable Health Care Act	Healthcare
Offline	Presidential Debate (1st)	The Economy	Social Issues
		The Supreme Court Appointments	National Security
		COVID-19	Healthcare
		Race & Violence in our cities	Social Issues
		The Integrity of Elections	Elections
		The Trump Biden Records	Elections
		Trump Healthcare Plan	Healthcare
	Presidential Debate (2nd)	Fighting COVID-19	Healthcare
		American Families & The Economy	Healthcare
		Race in America	Social Issues
		Climate Change	Social Issues
		National Security	National Security
		Leadership	National Security
	Snapshot Events	Black Lives Matter	Social Issues
		Capitol Hill Incident	National Security
		US Elections	Elections
		Inauguration Ceremony	Elections

Table 3.2: Topics selected for analysis.

Clustering Algorithm	c_v	c_umass	CPU Time (min:sec)
LDA	0.70	-2.26168	52:52
Parallel LDA	0.5921	-2.41955	12:12
NMF	**0.773022**	**-1.61094**	07:37
LSI	0.585223	-2.59355	00:27
HDP	0.640714	-17.3223	01:38

Table 3.3: Mean coherence scores and CPU time for different clustering algorithms with TF-IDF embeddings over five runs with varying random states.

Clustering Algorithm	**Epochs**	**Chunk Size**	**Workers (Number of CPU cores)**	**Evaluation Period (seconds)**	**Alpha (A-priori belief on document-topic distribution)**	**Eta (A-priori belief on topic-word distribution, also known as beta)**	**Kappa (Gradient descent step-size)**	**Minimum normalizing probability**
LDA	205	1000	NA	10	0.01	0.9	NA	NA
Parallel LDA	205	1000	7	10	0.01	0.9	NA	NA
LSI	NA	1000	NA	NA	NA	NA	NA	NA
NMF	205	1000	NA	10	NA	NA	1	0
HDP	NA	1000	NA	NA	0.01	NA	1	NA

Table 3.4: Model parameters for topic clustering with TF-IDF document embeddings

Preprocessing: Firstly, all the non-alphabets (numbers, punctuation, new-line characters and extra spaces) were removed from the text using the regular expression module (*re 2.2.1*). Then, the text was tokenized using *nltk 3.2.5*, followed by the removal of stopwords. Also, tiny words (i.e., words with a length of fewer than three characters) were removed from the text. This was followed by stemming the text using *PorterStemmer* and lemmatizing it using the *WordNetLemmatizer* from *nltk*.

Topic Modelling: Researchers have relied on Term Frequency-Inverse Document Frequency (TF-IDF) for generating document embeddings for short-text ([131], [197]). Tweets are categorized as short texts[3], and after preprocessing them, we generate document embeddings using TF-IDF and then pass them to five different clustering algorithms, namely – Latent Dirichlet Allocation (LDA), Parallel LDA, Non-negative matrix factorization (NMF), Latent Semantic Indexing (LSI), and Hierarchical Dirichlet Process (HDP) to generate topic clusters. Due to the short and noisy nature of the data, we ran these models five times over the data with varying random seeds. We check the coherence scores of topic models based on words, the 'c_umass' [153] and 'c_v' [186] measure, to confirm the performance consistency over multiple runs and finally use the best model to extract the top five topics.

We used Latent Dirichlet Allocation (LDA)[4] and LDA multi-core[5](Parallel LDA) pro-

[3]https://developer.twitter.com/en/docs/counting-characters
[4]https://radimrehurek.com/gensim/models/ldamodel.html
[5]https://radimrehurek.com/gensim/models/ldamulticore.html

vided by Gensim. Non-negative matrix factorization (NMF) model[6] uses the online NMF proposed in [250] for large corpora. Latent Semantic Indexing (LSI) model[7] implements fast truncated SVD (Singular Value Decomposition). And, for HDP[8], we use the improved online variational inference model proposed in [227]. The details of parameters used for each of the models have been listed in Table 3.4, and performance for each clustering algorithm in terms of their coherence scores ('c_v' and 'c_umass') and the amount of CPU time taken are mentioned in Table 3.3. From Table 3.3, we can see that NMF had the highest coherence scores ('c_v' and 'c_umass'), followed by LDA and HDP. Hence, we selected the top five topics yielded by NMF to search across the first page of Google search results. The content from the first page of Google search results was then retrieved to make sense of the extracted topic keywords to suggest a good topic name. For example, for the set of keywords yielded by the topic model: ['Paris', 'climate', 'green', 'change', 'science', 'reforms', 'environment', 'sustainable', 'urgency'], we did a Google search with these keywords and looked up for content and connections between them to deduce a suitable topic-phrase, i.e. *((Paris) Climate Agreement)*.

Hence, for each of the 22 selected topics (combined from both the sources, i.e., *online and offline*), we assign them an abstract category out of 'Social Issues', 'Healthcare', 'Elections' and 'National Security'. The abstract categories chosen serve as the foundation for political campaigns [135], and they are the most significant categories to consider when trying to persuade the public to vote for a particular political front. We utilize all these topics to analyze the engagement, stickiness, and to predict the candidate's stance. Table 3.2 presents the details of the selected topics as per their source and the abstract category they fall into. Each topic category consists of topics from at least two abstract categories. There are nine topics in 'Social Issues', five in 'Healthcare', and four in both 'Elections' and 'National Security'. The distribution of topics as per their abstract categories can be seen in Figure 3.2. Furthermore, we analyze the most and least talked about topics and abstract categories for each candidate and political party based on Engagement and Stickiness.

[6]https://radimrehurek.com/gensim/models/nmf.html
[7]https://radimrehurek.com/gensim/models/lsimodel.html
[8]https://radimrehurek.com/gensim/models/hdpmodel.html

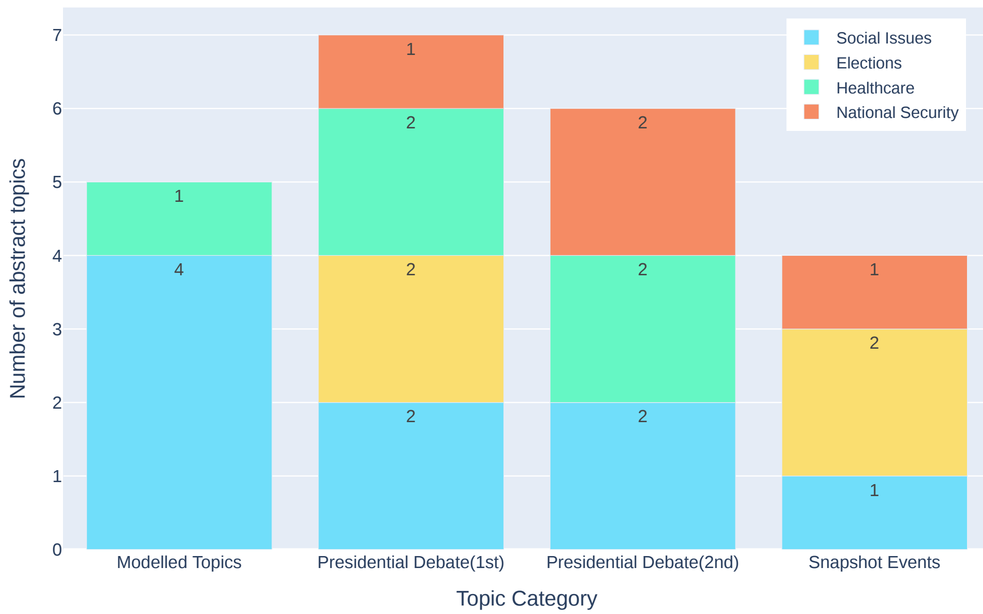


Figure 3.2: Distribution of topics as per their Abstract categories.

Engagement on topics

The amount of engagement received on a particular topic helps us quantify how popular the topic was among the general public. To quantify a topic's the engagement on Twitter, we first define each of the selected candidate's (user's) engagement on Twitter and then aggregate the tweets published by them, as well as their engagement as per the topic categories defined. The engagement for each user is defined as the product of average engagement per day and their impact.

User engagement was formerly quantified in terms of community features (the number of communities a user is a member of), author features (number of followers/ following, author influence) and content features (the number of retweets, mentions, URLs, hashtags, keywords, comments, and sentiment subjectivity) [89, 175]. Similarly, we aim to include all the accessible features through the Twitter Academic API in this work, and the average engagement per day for a user ($engagementPerDay_{user}$) is computed as the product of average engagement for a tweet each day ($avgEngagement\_day$) and the user impact ($userImpact$).

Algorithm 1: Pseudocode for engagement of a user per day

```
Input  : likesCount, repliesCount, retweetsCount, quotesCount,
         tweetCreationDate, tweet, user_df, followers, following, listedCount,
         totalTweets, profileCreationDate
Output: engagementPerDay_user (Engagement of a user per day)
Function avgEngagementPerDay(likesCount, repliesCount, retweetsCount,
 quotesCount, tweetCreationDate, tweet, user_df):
    // Count the number of tweets in a day
    tweetsPerDay ← user_df.groupby(['tweetCreationDate'])['tweet'].count()
    // Calculate total engagement for the day
    user_df['engagement_rate']
     ← likesCount + repliesCount + retweetsCount + quotesCount
    engagement_day ← user_df.groupby(['tweetCreationDate'])
    // Calculate average engagement per day
    avgEngagement_day ← engagement_day/(4 * tweetsPerDay)
    // Weighted exponential moving average of avgEngagement_day
    avgEngagement_day
     ←avgEngagement_day.ewm(span = 20, adjust =FALSE).mean()
    z_score ← stats.zscore(avgEngagement_day)
    // Remove outliers using z-score
    avgEngagement_day ←avgEngagement_day[-3 ≤ z_score ≤ 3]
    // Smoothen the average engagement per day to 8th degree using
       Savitzky-Golay filter
    avgEngagement_day ← savgol_filter(avgEngagement_day, polyorder = 8)
return avgEngagement_day
Function
 userImpact(followers,following,listedCount,totalTweets,profileCreationDate):
    profileAge ← days(January 27, 2021 − profileCreationDate)
    // To quantify whether a user is active producer/consumer of
       content
    FtFRatio ← log10(followers/following + 1)
    userImpact
     ← (followers * listedCount * FtFRatio)/(totalTweets * profileAge)
return userImpact
engagementPerDay_user ← avgEngagement_day * userImpact
```

The average engagement for a tweet is aggregated by measuring the reactions received on tweets (such as the number of replies, retweets, likes, and quotes) over the course of the day. The data obtained from the Twitter API for the reactions to each tweet have been aggregated from January 21, 2019, to January 27, 2021. We propose average engagement per day for a tweet by taking inspiration from the *Engagement rate* defined by Twitter[9]. For a given user, Twitter defines *Engagement Rate* as:

$$Engagement\ rate = \frac{Engagement}{Impressions} * 100 \tag{3.1}$$

where *Engagement* is the summation over the number of likes, replies, retweets, media views, tweet expansion, profile/ hashtag/ URL clicks, and new followers gained for every tweet, and *Impressions* is the total number of times a tweet has been seen on Twitter, such as through a follower's timeline, Twitter search, or as a result of someone liking your tweet9[0].

Due to limitations with the API, we only have access to the public metrics, i.e., number of likes, retweets, replies, and quotes. Therefore, to calculate the average engagement rate for a user per day, we use the function *avgEngagementPerDay* as proposed in Algorithm 1 (line number: 1). To normalize the fluctuating values of Average Engagement, we calculate its Exponential Moving Average (EMA) with a window span of 20 days for every candidate[10] and remove the outliers using z-score, followed by smoothening the average engagement per day to the 8$^{\text{th}}$ degree using Savitzky Golay filter4[0].

Every user has a different number of followers, following and they receive varied responses from the users on Twitter (which may or may not be their followers); hence, it is essential to consider their impact (popularity) on Twitter to calculate the number of users they reach through their tweets. Researchers have tried to analyze the impact of users by proposing heuristic and neural-network-based models ([57], [183], [210]). We define the impact of a user ($impact_u$) inspired from the previous work done in [183] and define it as a function of followers, following, the total number of tweets, and the profile age, as in Algorithm 1 (line number: 11), where $followers$ is the total number of followers a user has, $listedCount$ is the number of public lists a user is a part of, $following$ is the number of people that the user follows, $log_{10}\left(\frac{followers}{following} + 1\right)$ is the ratio of followers to following (FtF ratio) to check whether a user is an active user (with more followers, producing content) or a passive user (with more following, consuming content). To avoid outliers, we take log base 10 and add one to prevent the metric from being zero when the value of followers equals to following. $tweetCount$ is the

[9]https://help.twitter.com/en/managing-your-account/using-the-tweet-activity-dashboard

[10]A grid-search analysis was performed to find the best value.

total number of tweets produced by the user for the scope of our analysis, and $profileAge$ is the difference between the profile creation date reported by Twitter and January 27, 2021, i.e., the last day for our data collection, quantified as the number of days. Our algorithm overcomes the shortcomings of [183] by incorporating the $listedCount$ factor and changing the placement of $tweetCount$. The $tweetCount$ has been deemed inversely related to the user impact, because a user tweeting sporadically but obtaining high interaction is more significant than a person tweeting recurrently but receiving low engagement.

The engagement for a user is the product of Average engagement per day and the user's impact. The engagement value helps us in quantifying the user's social reach.

Findings: Joe Biden had the highest impact, followed by Donald Trump, Mike Pence, Kamala Harris, The Democrats and HouseGOP. We normalize the user impact between the range 0 and 1 to calculate the engagement on tweets for each topic, where 0 is the lowest user-impact and 1 is the highest.

For Joe Biden, the top three topics receiving the maximum engagement during the scope of our analysis were – *The Integrity of Elections, Weapon Ban, and US Elections*. As for Kamala Harris, they were *US Elections, Fighting COVID-19 and The Integrity of Elections*, however, for the official Twitter handle of the political party (@TheDemocrats), the most engaging topics differed from these, and they were *American Families & The Economy, National Security, and Inauguration Ceremony*. Overall, for Democrats, the top three most engaging topics were *The Integrity of Elections, US Elections, and Weapon Ban*.

In the case of Donald Trump, the top three topics receiving the maximum engagement were *US Elections, The Integrity of Elections, and Affordable Healthcare Act*. For Mike Pence, they were *Inauguration Ceremony, US Elections and Fighting COVID-19*. However, for the official Twitter handle of the political party (@HouseGOP, @GOP), as they had the lowest impact, the engagement received on the tweets was too low to be quantified. Overall, for Republicans, the top three most engaging topics were *US Elections, Inauguration Ceremony and Affordable Healthcare Act*.

Stickiness of topics

Stickiness helps us to identify the favourite topics for each candidate within the scope of our analysis. We quantify stickiness based on the repetitiveness of the topics spanning across different election phases. The Presidential candidate's timeline is divided into three phases, and the Vice-Presidential candidate's timeline is divided into four phases. See Figure 3.3 for more detailed information about the timelines. The topic stickiness is checked for three candidates only, because of the unavailability of exact campaigning dates for Mike Pence. We segregate the topics into three classes based on stickiness:

1. ***Very Sticky:*** Topics are *Very Sticky* when they have been tweeted about in every election phase,
2. ***Sticky:*** If the topics were tweeted in $(n-1)$ election phases, then they were classified as *Sticky*. Here, n is the total number of election phases (i.e., $n = 3$ for Presidential, and $n = 4$ for Vice Presidential candidates).
3. ***Loose:*** If the topics were tweeted only once for Presidential candidates, and twice for Vice Presidential candidates across all the election phases, then they were classified as

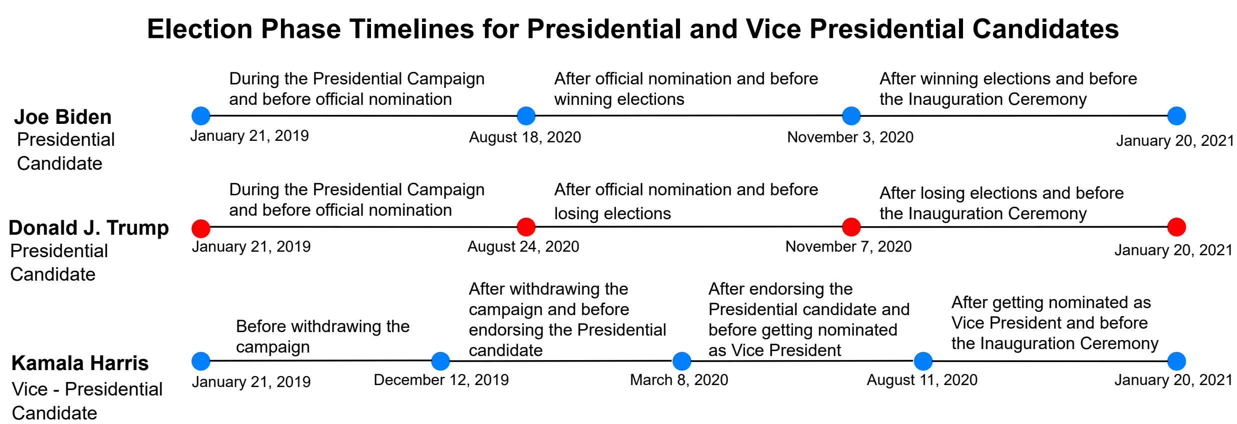


Figure 3.3: Electoral campaign timelines for Presidential and Vice Presidential candidates. The timeline is divided as per the general election phases and the ranks each candidate was contesting for. The details of campaigning for each candidate, have been taken from the news reports of the campaigns on CNBC and Politico.

Loose.

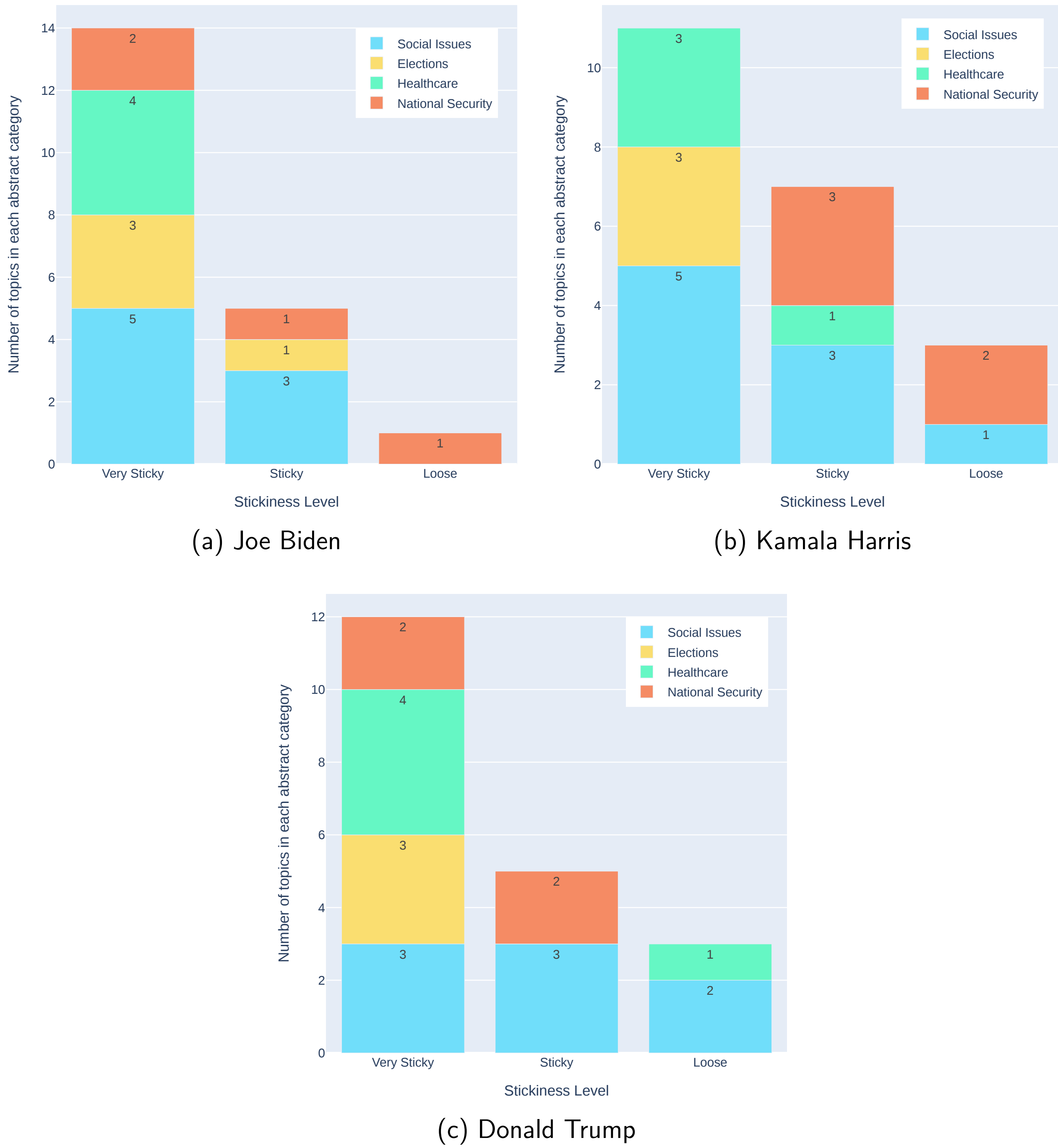
Findings: Joe Biden and Kamala Harris tweeted mostly about *Social Issues and Healthcare* as we can see these categories dominate both the *Very Sticky* and *Sticky* levels from Figure 3.4a and Figure 3.4b. However, *National Security* topics were loose in nature. Donald Trump tweeted differently from the Democrats and tweeted the most about *Elections and Healthcare*, followed by *Social Issues*.

Doing a micro-analysis, we found that Joe Biden tweeted about twenty topics out of twenty-two, with fourteen of them being *Very Sticky*, five being *Sticky*, and one being *Loose*. *Legalization of Medical Marijuana* and *Trump Healthcare Plan* were the topics that Joe Biden had not tweeted about, even once. However, Kamala Harris and Donald Trump tweeted about 21 topics. Kamala Harris had eleven topics in the *Very-Sticky* category, seven in *Sticky* and three in *Loose*. For Donald Trump, the distribution of topics per their Stickiness levels was slightly different, with thirteen being *Very Sticky*, five being *Sticky*, and three being *Loose*. Kamala Harris did not tweet about *Trump Healthcare Plan*, and Donald Trump did not tweet about the *Legalization of Medical Marijuana*.

Regardless of their political fronts, Joe Biden, Kamala Harris, and Donald Trump had nine *Very Sticky* topics in common (i.e., *The Integrity of Elections, Affordable Healthcare Act, Equality Rights for LGBTQ, Weapon Ban, Inauguration Ceremony, US Elections, American Families & The Economy, COVID-19, Race & Violence in our cities*). Joe Biden and Donald Trump stuck to *The Integrity of Elections*; however, Kamala Harris stuck to *Affordable Healthcare Act*. Comparing the candidates from the same political front, Joe Biden and Kamala Harris had eleven common topics in the *Very-Sticky* category. For the Presidential

Candidate	**Election Phase 1**	**Election Phase 2**	**Election Phase 3**	**Election Phase 4**
Joe Biden	No loose topics	The Supreme Court Appointments	No loose topics	Not Applicable (NA)
Kamala Harris	The Economy, The Trump & Biden Records	National Security	No loose topics	No loose topics
Donald Trump	The Economy, Trump Healthcare Plan, Build Back Express Tour	No loose topics	No loose topics	Not Applicable (NA)

Table 3.5: Appearance of *Loose* topics in different election phases.



(a) Joe Biden

(b) Kamala Harris

(c) Donald Trump

Figure 3.4: Abstract categories of topics segregated according to stickiness levels for all three candidates (a) Joe Biden, (b) Kamala Harris, and (c) Donald Trump.

candidates, in addition to the nine common topics, they also had *Fighting COVID-19* and *Capitol Hill Incident* repeating in the *Very-Sticky* category.

Furthermore, the topics classified as *Loose* from all three candidates appeared in only one of the election phases. For example, in Joe Biden's case, topic *Supreme Court Appointments* appeared only in the second election phase as per his timeline, and similar behavior can be seen for the other two candidates (refer Table 3.5 for details). Also, the *Loose* topics, i.e., *Trump Healthcare Plan*, *Supreme Court Appointments*, *Build Back Express Tour*, *The Economy*, and *The Trump & Biden Records*, are among the rarely tweeted topics, and corresponding behavior can be seen for the *Very Sticky* topics, i.e., the top three *Very Sticky* topics for each candidate are the most frequently tweeted and highly engaging topics.

3.5 Synergy among candidates

To quantify the cooperation among the candidates from the same political front, we highlight the content similarity and the congruities and contrasts in the stance of various topics they tweeted about, as discussed below.

Content Similarity

We check the alignment of the Presidential and Vice-Presidential candidates with the political party by comparing the similarity of their tweets as per our proposed Algorithm 2. When comparing the similarity of two users, there is a high probability that the topic of tweets from one user may be repeated by the second user a couple of days before or after the first user's tweet during the election campaign. So, to address this, we compare each tweet of $user_1$ with all the tweets of $user_2$ and store the maximum similarity between the tweet text. We repeat this process for all the tweets, then average the results to determine how similar two candidates' content is. The tweets of Vice-Presidential candidates are compared with both the Presidential candidate and the political party, however, the tweets of the Presidential candidate are compared with the political party only. We compute the content similarity (cosine similarity) between the candidates by using the top-5 models from HuggingFace[11]

[11]https://huggingface.co/models

(grouped by the sentence-similarity task, sorted by the number of downloads) to generate text embeddings. Table 3.6 elaborates the performance details for each of them.

Algorithm 2: Content Similarity of two users

Input : $tweets_{user1}$, $tweets_{user2}$
Output: Average cosine similarity of users' content ($avgCosineSim$)

embeddings$_{user1}$ $\leftarrow model.encode(tweets_{user1})$
embeddings$_{user2}$ $\leftarrow model.encode(tweets_{user2})$
$cosineSimilarity = [\,]$
`/* Compare each tweet embedding of user 1 with all tweet embeddings of user 2 and store the max similarity in list */`
for $embedding_{user1} \leftarrow 0$ **to** $len(embeddings_{user1})$ **do**
  $cosineScoresList$ $\leftarrow$cosineSimilarity(embedding$_{user1}$, $embeddings_{user2}$[:])
  $maxScore$ $\leftarrow$max(cosineScoresList)
  $cosineSimilarity.append(maxScore)$
end for
$avgCosineSim = mean(cosineSimilarity)$

Findings: From all the models tested, 'bert-base-mean-nli-tokens' performs the best in computing the content similarity between the candidates, followed by 'paraphrase-multilingual-MiniLM-L12-v2'. The common trend noticed while computing the content similarity is that the tweets by Kamala Harris are more aligned with the Presidential candidate than the political party; however, for Mike Pence, it's the opposite, i.e., Mike Pence aligns more with the

Political Party	Candidate	all-MiniLM-L6-v2		paraphrase-xlm-r-multilingual-v1		paraphrase-mpnet-base-v2		bert-base-nli-mean-tokens		paraphrase-multilingual-MiniLM-L12-v2	
		w.r.t Presidential Candidate	w.r.t Political Party	w.r.t Presidential Candidate	w.r.t Political Party	w.r.t Presidential Candidate	w.r.t Political Party	w.r.t Presidential Candidate	w.r.t Political Party	w.r.t Presidential Candidate	w.r.t Political Party
Democrats	Kamala Harris	0.6537	0.6473	0.6465	0.6256	0.6959	0.7016	**0.8333**	**0.8387**	0.6914	0.6794
	Joe Biden	NA	0.6548	NA	0.6457	NA	0.7143	NA	**0.8496**	NA	0.7192
Republicans	Mike Pence	0.6829	0.6974	0.6452	0.6680	0.6669	0.7262	**0.8439**	**0.8561**	0.7177	0.7318
	Donald Trump	NA	0.6453	NA	0.6262	NA	0.6986	NA	**0.8485**	NA	0.7187

Table 3.6: Content similarity between candidates using different BERT based embeddings.

political party instead of the Presidential candidate. Also, the Presidential candidates have a high similarity rate with the political party. Therefore, the results portray coordination in the candidates' tweets and the tweets from their political parties.

Stance Similarity

Stance Similarity is an important technique to analyze textual data and is frequently used in NLP to analyze the standpoint of a person towards a topic or an event. We test different models that classify the candidates' stance for the selected topics in three categories: favour, against, and neutral. Although there is no universal number for how many tweets should be sampled, for testing a model, we observe a range across studies from under 2,000 labelled tweets ([5], [165], [198]) to several thousand ([75], [152], [190], [247]). For this study, we sample 3,015 tweets evenly distributed across the timeframe of our data collection and from all the candidates. Three different annotators then labelled these tweets, and the stance category having the majority among the three annotators was chosen as the overall response. The annotators had no known prior political biases and they annotated the tweets solely on the basis of the tweet content. The Fleiss' Kappa statistical test was performed to determine the inter-annotator agreement in labelling, and the *kappa* score is '*0.7*'. We divide the data into an 80:20 ratio for training and testing multiple classification methods, and we annotate the dataset using standard procedures, as defined above.

We try various traditional and modern algorithms to estimate the performance for stance classification on the labelled tweets. We use TF-IDF and Hashing Vectorizer for the conventional algorithms to generate embeddings as inputs to Support Vector Machine (SVM), Linear SVM, and Logistic Regression to compute the performance. Synthetic Minority Oversampling Technique (SMOTE) is used to oversample the tweets' vectorized features. However, we don't notice a rise in classification performance after oversampling. We also report the classification performances on modern algorithms, like the Deep Neural Network (DNN) based classifier from the Spark NLP pipeline, which takes Universal Sentence Encod-

[12] *rbf: Radial basis function,*https://en.wikipedia.org/wiki/Radial$_{b}asis_{f}unction_{n}etwork$

[13] *rbf: Radial basis function,*https://en.wikipedia.org/wiki/Radial$_{b}asis_{f}unction_{n}etwork$

Algorithm used	**Oversampled (Yes/No)**	**Classification Performance**
Hashing Vectorizer and Linear SVM	No	0.70
	Yes	0.54
Hashing Vectorizer and SVM(kernel=‘rbf’[12], degree=3, Cubic)	No	*0.73*
	Yes	0.66
Hashing Vectorizer and Logistic Regression	No	*0.73*
	Yes	0.57
TF-IDF and Linear SVM	No	0.72
	Yes	0.54
TF-IDF and SVM (*kernel=‘rbf’[13], degree=3, Cubic*)	No	0.70
	Yes	0.66
TF-IDF and Logistic Regression	No	*0.73*
	Yes	0.56
Spark NLP (Universal Sentence Encoder and Deep Learning Classifier)	No	0.72
BERT-base (uncased)	No	0.69
XLNet (base, epochs=10)	No	0.71
XLNet (large, epochs=10)	No	0.71
facebook/bart-large-mnli (fine tuned)	No	**0.75**

Table 3.7: Stance classification performance on the testing set (i.e., 20% of the sampled dataset) using different algorithms.

ings of tweets as inputs[14], BERT-base-uncased[15], XLNet (base-cased[16], large-cased[17]) and fine-tuned Facebook's Zero-shot learning-based, bart-large-mnli[18]. From the traditional algorithms, TF-IDF combined with Logistic Regression, Hashing Vectorizer with SVM, and Logistic Regression perform equally well with their stance classification accuracies on the testing set of the sampled data as 73%. From the modern ones, Facebook's ‘bart-large-mnli’ performs at par with 75% classification accuracy on the test set. We then use Facebook's bart-large-mnli to predict the stance on the remaining tweets.

Findings: From the predicted stance, we notice that Democrats favoured most of the topics,

[14]https://nlp.johnsnowlabs.com/api/com/johnsnowlabs/nlp/annotators/classifier/dl
[15]https://huggingface.co/bert-base-uncased
[16]https://huggingface.co/xlnet-base-cased
[17]https://huggingface.co/xlnet-large-cased
[18]https://huggingface.co/facebook/bart-large-mnli

apart from *Trump Healthcare Plan* and *Build Back Express Tour*, with no tweets for *Trump Healthcare Plan* and neutral stance for *Build Back Express Tour*. However, Republicans had a favourable outlook for 14 topics, an unbiased view for six (*The Trump & Biden Records, Leadership, Black Lives Matter, Capitol Hill Incident, Inauguration Ceremony, and Build Back Express Tour*), and they were against one topic (*Trump Healthcare Plan*). Additionally, the predicted stance illustrates symmetry between the candidates' tweets and the tweets from their political parties.

3.6 Conclusion & Future Work

Social media platforms (SMPs) have evolved into strategic spaces critical to modern political campaigns. Because of the interactivity of social media, candidates can establish direct relations with their audience without an intermediary (e.g., newspapers, news channels). Not surprisingly, campaign strategists prioritize social media as the primary channel for delivering and persuading messages. Therefore, understanding how the internal co-operation of a political front helps in contesting for elections is just as important as understanding the external factors, like the candidate's impact and the audience's engagement.

In terms of the candidate's impact and engagement on topics, it was found that the candidate with the highest impact did not have a higher number of tweets as compared to the candidate with a lower impact. User impact ultimately depends on the topic referred to in the tweets. Also, Joe Biden and Kamala Harris understood their audience well. They tweeted about the topics receiving higher engagements than Donald Trump and Mike Pence, who were unable to identify the topics of social interest. Additionally, the Democratic candidates displayed higher internal cooperation through their tweets and stance on different topics than the Republican candidates.

This study extends political campaign research by investigating two broad aspects — engagement and stickiness of topics, the candidate synergy (i.e., the content and stance similarity) and their social reach on Twitter during the 2020 US Presidential elections. The results clearly indicate that the tweet's topic, and the candidate's influence, both have an impact on the amount of public engagement it receives. Internal cooperation (i.e., similarities in the content and stance for specific topics) also helps the candidates in creating a stronger hold on the thoughts of the public during the election campaign. Furthermore, this study employed an empirical approach by examining how internal cooperation between candidates and their engagement influenced the election results and it may be utilized to create a Twitter communication model for the candidates and government to assist in effective campaigning and governance.

Several restrictions, as well as recommendations, must be noted for further research. As the data for this study is focused on the tweets from Presidential and Vice-Presidential candidates for the 2020 US Elections, it was not feasible to distinguish between organic and synthetically generated (i.e., through bots, masked profiles) engagements. The impact of moving window for the content similarity between two candidates needs to be investigated further. Also, we do not intend to generalize our results for every political campaign. Moreover, our research is based on text features; we could not account for the influence of image features and knowledge graphs related to a particular tweet. Investigating the separation and impact of organic and sponsored audience involvement may benefit future studies. Additionally, factors that might influence public engagement and internal cooperation, such as emotions, timing impacts, tweet formats, should be examined precisely. Furthermore, investigating the extent of interactions between users and political candidates through their social media managers will contribute significantly to a better understanding of the democratic ability of SMPs.

Acknowledgments

We are grateful to the research team at Lakehead University's DATALab for their help - Andrew Fisher and Aditya Singhal for proof-reading this paper, along with Aditya Singhal, Amir Shahbandegan and Subhodip Panda for annotating the sampled tweets, to evaluate the stance similarity among candidates from the same political front. We would also like to thank the CASES building, Lakehead University for providing the infrastructure; the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant and Vector Institute since this study would not have been feasible without their support.

Chapter 4

Resilience of Political leaders and Healthcare organizations during COVID-19

All of this chapter is under major revision at a reputable journal as follows:
• Baxi, MK., Philip, J., Mago, V. (2022). Resilience of political leaders and healthcare organizations during COVID-19.

According to the Bloomberg COVID-19 Resilience Ranking, this chapter empirically analyzes the online societal associations of the top ten COVID-19 resilient nations' leaders and healthcare institutions. The attributes supplied by Twitter Academic Research API, along with the tweets of leaders and healthcare organizations, are used to quantify the strength of the online social association in terms of public involvement, sentiment strength, and inclusivity and diversity. This study aids in comprehending how leaders and healthcare organizations may effectively use social media platforms to foster digital affiliations with the public amid health emergencies.

Keywords : social media, Twitter, COVID-19, user engagement, content analysis, sentiment strength, inclusivity and diversity strength, crisis communication

4.1 Introduction

The exponential spread of the 2019 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has flooded various social media platforms (SMPs) with a plethora of information about the disease, pandemic trajectory, influence on human fatalities, and global and regional consequences for the governments and health organizations ([72]). SMPs such as Twitter, Facebook and Instagram have become the norm for broadcasting and acquiring pandemic-related information by the leaders and healthcare organizations. Researchers have demonstrated an interest in examining and interpreting the social media data of leaders and healthcare organizations across various SMPs, by evaluating their Twitter usage through content analysis and the change in their number of followers ([86, 188]).

Political leaders are followed on Twitter for a number of reasons, including convenience, expressiveness, knowledge and sociability ([161]). According to research, 70% of healthcare institutions in the United States also utilize SMPs, and their social media presence influences 57% of clients' decisions about where to seek medical care ([164]). Thus, it is necessary to investigate the societal impact of leaders and healthcare organizations on the citizens during a crisis situation. The importance of studying the qualitative factors to determine the societal impact of social media on Government to Citizen interactions has been highlighted previously ([33, 155]). Therefore, the objective of this study is to quantify the societal impact of leaders and healthcare organizations from the top-10 COVID-19 resilient nations by analyzing the following factors – user engagement, sentiment strength, and the inclusivity and diversity of various communities in the tweets authored by them.

Understanding which information (content, type) appeals to the audience the most might be an effective way of amplifying the involvement ([31, 28]), and hence can help in perpetuating a regular conversation with the public, addressing their concerns, recommendations, and desires, and thus assist in pacifying them, building trust, and fighting through crisis situations together. On this rationale, this paper calculates the influence of leaders and healthcare organizations, and the engagement they receive during COVID-19. Furthermore, people, especially political leaders and healthcare organizations, communicate their opinions and attitudes – which are generally termed as '*sentiment*' through several SMPs ([56]). However, it is ambiguous how sentiment and the references to different communities might influence information dynamics in a social-media setting. Therefore, it is critical to evaluate the collective influence of sentiment, inclusion and diversity along with the public engagement on the leaders' and healthcare organizations' societal impact.

This study offers a thorough assessment of the leaders and healthcare organizations, including both qualitative and quantitative insights regarding their online behaviour. This type of hybrid study has yielded promising findings ([194]). Our research examines the public metrics of their tweets (likes, replies, retweets, and quotes) as well as utilizes statistical methodologies to compute sentiment strength, inclusiveness, and diversity strength. It is qualitative in nature as we analyze the tweet content to see whether there are any parallels to real-life events, as well as higher levels of engagement and societal influence. As a result, the purpose of this study is to explore how audience engagement, sentiment, inclusion and diversity strength may assist leaders and healthcare organizations develop a trustworthy relationship with the public, gathering support for policies that limit the spread of COVID-19, and overcoming the crisis situations together. The key findings of the research are:

1. Amongst politicians, United Arab Emirates' Prime Minister had the highest societal impact,

2. The Canadian public health agency demonstrated a prominent level of social impact amid the healthcare organizations, and,

3. Individual aspects of the societal impact were better understood by leaders than by healthcare organisations.

The remainder of the paper is structured as follows: Related work section discusses the summary of the past research relevant to the usage of SMPs by leaders and healthcare organizations. Dataset and the statistical methods used to compute the societal impact are discussed in Methodology section. The results and implications of this study are discussed in Results, followed by Discussion & Conclusion section, where the key inferences and further research directions are discussed.

4.2 Related Work

Leaders and healthcare organizations utilize social media platforms (SMPs) including Facebook, Twitter, Instagram, and Reddit for election campaigns, broadcasting public health information, announcing significant events, and improving public relations ([25, 21, 33, 50]).

In lieu of the use of SMPs for election campaigns, the candidate and audience engagement on Twitter for the 2013 and 2016 Australian federal elections was compared ([38]), while the use of Facebook for strategic campaigning in the 2008 and 2012 US Presidential elections was analyzed ([34]). Additionally, the use of Instagram as an advertising tool for Swedish elections and YouTube for the electoral campaign of European Parliamentary elections was examined ([189, 222]). SMPs, according to earlier studies ([21, 33, 50]) may assist in improving governance transparency, engagement, and accountability. Furthermore, using a variety of examples, researchers have highlighted the impact on several aspects of public interaction through SMPs ([81, 96, 185]). These platforms have been cited by several academicians ([29, 30, 32, 80, 192, 206]) as an important tool for expanding the social reach and better understanding the audience. However, previous research has demonstrated that the sentiment, emotion, and diversity of tweets, can mitigate public engagement and persuade the audience ([23, 196, 107, 179]).

As seen recently, SMPs have also been used by healthcare professionals and organizations as a communication tool to promote healthy habits, share announcements, disseminate awareness, motivate the patients on the way to their recovery, support emergency response, and eventually boost readiness during exceptional circumstances ([17, 54, 129, 221, 145, 203, 163, 162]). The idea of SMPs having a positive influence on public awareness and behavioural changes by disseminating succinct information to specified audiences was explored ([61]). Further, to evaluate public reactions during the epidemic, topic identification and sentiment analysis was utilized to examine the change of public attitude over time in relation to the published news, and reddit posts ([71, 144]). Also, a previous study has identified factors associated with the levels and duration of engagement, for the Facebook accounts of U.S. Federal health agencies ([25]). Furthermore, researchers have investigated the influence of world leaders during the COVID-19 pandemic and how Twitter was used to swiftly transmit information to the public ([86, 188]). However, there has not been an extensive analysis to understand and examine the societal impact of leaders and healthcare organizations during COVID-19, considering different factors like user engagement, sentiment strength, and, inclusivity and diversity strength, which is hence the focus of this work.

4.3 Methodology

Dataset

Twitter Academic Research API v2[1] was utilized to retrieve the information of the political leaders' and health organizations' tweets. 173,071 tweets were collected and analyzed from December 1, 2019, to December 31, 2021. The dataset was curated based on the Bloomberg COVID-19 Resilience Ranking[2], as of January 8, 2022, at 5 p.m. EST, selecting the health organizations and leaders of the top-10 COVID-19 resilient countries. The COVID-19 Resilience Ranking is a monthly impression of the countries handling the virus most effectively, with the least social and economic disruption. The ranking is calculated based on the factors of virus containment, quality of healthcare, vaccination coverage, overall mortality and progress towards restarting travel. The timeline was chosen to include the outbreak COVID-19 to the vaccination period of the pandemic. Official health organizations of the respective countries and personal accounts of the political leaders were analyzed in this specific study. This provides an opportunity to get a sense of the contrasting dynamics between the accounts; to truly encapsulate the societal impact on the particular country.

The collected tweets spanned across 19 different languages and were translated to English using the Neural Machine Translation (NMT) models from the Tatoeba Translation Challenge, which consists of NMT models trained on a compressed dataset of over 500GB, encompassing 2,961 language pairings, and 555 languages ([215]). For this study, each Twitter account is referred to as a *user* and the type of account (i.e., leaders, health organizations) is referred to as a *user group*. The details of the tweets authored by each of the selected users in the order of their COVID-19 Resilience ranking (i.e., from the best country to live in during COVID-19, like Chile, to the good ones, like United Kingdom) can be found in

[1]https://developer.twitter.com/en/products/twitter-api/academic-research
[2]https://www.bloomberg.com/graphics/covid-resilience-ranking/

Account Type	**Name (Twitter Handle)**	**Country**	**Number of Tweets**
Leader (President or Prime Minister)	Sebastián Piñera (@sebastianpinera) (President)	Chile	622
	Micheál Martin (@PresidentIRL, @MichealMartinTD)	Ireland	3,641
	Mohammed bin Rashid Al Maktoum (@HHShkMohd)	U.A.E	839
	Sanna Marin (@MarinSanna)	Finland	2,007
	Justin Trudeau (@JustinTrudeau, @CanadianPM)	Canada	13,778
	Iván Duque (@IvanDuque) (President)	Colombia	7,059
	Recep Tayyip Erdoğan (@RTErdogan) (President)	Turkey	1,943
	Pedro Sánchez (@sanchezcastejon)	Spain	4,290
	Magdalena Andersson (@SwedishPM)	Sweden	282
	Boris Johnson (@BorisJohnson)	United Kingdom	2,335
	Total		36,796
Health Organization/ Health Minister	Ministerio de Salud (@ministeriosalud)	Chile	39,401
	HSELive (@HSELive), Department of Health (@roinnslainte)	Ireland	18,332
	Ministry of Health and Prevention, U.A.E. (@mohapuae)	U.A.E	8,424
	Ministry of Social Affairs and Health (@MSAH_News)	Finland	1,009
	Health Canada and PHAC (@GovCanHealth)	Canada	38,715
	Ministry of Health and Social Protection of Colombia (@MinSaludCol)	Colombia	11,346
	Ministry of Health of the Republic of Turkey (@saglikbakanligi)	Turkey	4,119
	Ministry of Health (@sanidadgob)	Spain	8,595
	The Public Health Agency of Sweden (@Folkhalsomynd)	Sweden	701
	UK Health Security Agency (@UKHSA)	United Kingdom	5,633
	Total		136,725

Table 4.1: Distribution of tweets for the selected user accounts.

Table 4.1.

Societal Impact

The societal impact (denoted by, $societalImpact$) is defined as the product of engagement per day with user impact ($dailyAvgEng_{user}$), sentiment strength ($sentiStrength$), and inclusivity and diversity strength ($iDStrength$) in the user's tweets as in Equation (4.1). Further details of the variables can be found in the following subsections.

$$societalImpact = dailyAvgEng_{user} * sentiStrength * iDStrength \tag{4.1}$$

Engagement with impact

The engagement per day is the measure of the social interaction of the post, including the likes, replies, retweets and quotes. The engagement per day represents the relationship between the followers and the user, and the resonation of their tweets. Twitter defines engagement rate, as the ratio of engagements to impressions: $\frac{Engagement}{Impressions} \times 100$. The engagements are defined as an aggregate of interactions of a tweet – retweets, replies, follows, likes, links, cards, hashtags, embedded media, profile photo, username or tweet expansion. The impressions account for times a user has observed a particular tweet in their search results or timeline (Twitter Account Activity Analytics – Engagement, Impressions)[3]. This study analyzes only public metrics such as the count of likes, replies, retweets and quotes-as a result of the limitations of Twitter API.

To evaluate a user's engagement ($dailyAvgEng_{user}$, Equation (4.2)); firstly, their tweet-wise engagement ($dailyAvgEng_{(tweet,user)}$, Equation (4.3)) is calculated by multiplying the user impact ($impact_{user}$) and average engagement per day for a tweet, ($dailyAvgEng_{tweet}$), followed by taking an average of the tweet-wise engagement ($dailyAvgEng_{(tweet,user)}$) for the user.

$$dailyAvgEng_{user} = \frac{\sum dailyAvgEng_{(tweet,user)}}{totalTweets_{user}} \tag{4.2}$$

$$dailyAvgEng_{(tweet,user)} = dailyAvgEng_{tweet} \times impact_{user} \tag{4.3}$$

[3]https://help.twitter.com/en/managing-your-account/using-the-tweet- activity-dashboard

The impact of a user ($impact_{user}$, Equation (4.4)) is quantified based upon the hyperbolic tangent function (tanh) of followers, the total number of tweets, following, public lists and profile age. The $listedCount$ is the total amount of public lists of a user. Lists indicate popularity - generally revolving around the concept that other users are engaged with one's content. Furthermore, $log_{10}\left(\frac{\sqrt{followers}}{following}\right)$ is the followers to following ratio, indicating the general nature of the account. The ratio is within a base-10 log to elude outlier values. The $totalTweetCount$ is the number of tweets from the account during our data collection timeframe. The $profileAge$ represents the number of days between the profile creation date to December 31, 2021; the last analysis day. Because a freshly joined user with more followers would be more influential than a previously joined user with fewer followers, the square of a user's profile age has been deemed inversely proportional to the user's impact.

$$impact_{user} = \frac{\tanh\left(\log_{10}\left(\frac{\sqrt{followers}}{following}\right) \times listedCount \times tweetCount\right)}{(profileAge)^2} \tag{4.4}$$

To quantify the average engagement per day ($dailyAvgEng_{tweet}$, Equation (4.5)), the collated number of likes, replies, retweets, quotes and tweets per day, from December 1, 2019, to December 31, 2021, is used. Furthermore, the $dailyTweetCount$ are multiplied by 4 (equal to the number of variables in the numerator).

$$dailyAvgEng_{tweet} = \frac{likes + replies + retweets + quotes}{4 * dailyTweetCount} \tag{4.5}$$

To standardize the shifting values of average engagement, we calculate the Exponential Moving Average with a 151-day window span[4], eliminate outliers using $z - score$ and smoothen the average engagement per day to the 8$^{\text{th}}$ degree using the Savitzky Golay filter[5].

Sentiment Strength

To quantify the strength of sentiment for every user, we first calculate the sentiments of all the tweets collected for our analysis using CardiffNLP's '*twitter-roberta-base-sentiment*' model, which is trained on a 60 million Twitter corpus, and then calculate the sentiment

[4]A grid-search analysis was performed to find the best value.
[5]A grid-search analysis was performed to find the best value.

strength for every user as mentioned in Equation (6), i.e., based on the sentiment category with the maximum number of tweets for that day, followed by assigning the sentiment score based on the sentiment: 10^{-6} for neutral, the ratio of the count of positive tweets to total tweets for positive, and the negation of the ratio of the count of negative tweets to the total tweets for negative sentiment.

$$sentiStrength = \begin{cases} 10^{-6} & ; maxSentiment(tweets) = neutral \\ \dfrac{count(positiveTweets)}{totalTweets} & ; maxSentiment(tweets) = positive \\ -\dfrac{count(negativeTweets)}{totalTweets} & ; maxSentiment(tweets) = negative \end{cases} \tag{4.6}$$

Inclusivity and Diversity Strength

We assess the inclusivity and diversity in the tweets of the users (denoted by, $iDStrength$) by computing the usage of different keywords pertaining to various communities from the countries selected for our analysis. The keywords are selected based on gender, age, cultural inferences, ethnicity, and employment sectors of each of these countries. The detailed list of keywords can be found in GitHub Repository[6]. The usage frequency for each of these keywords is calculated for all users with respect to the total number of tweets from that user, as given in Equation (7). If there exists a user who has not referred to any community in his tweets, a default value of 10^{-6} is assigned.

$$iDStrength = \begin{cases} 10^{-6} & ; count(communityMentionTweets) = 0 \\ \dfrac{count(communityMentionTweets)}{totalTweets} & ; otherwise \end{cases} \tag{7}$$

Content analysis

The tweets of all the users are analyzed for the most-frequent topics and the most-referred users by assessing the usage of hashtags and mentions. The tweets are examined by extracting the top-10 hashtags and mentions using the '*advertools 0.13.0*' module[7]. We compare the

[6]https://github.com/manmeetkaurbaxi/Societal-Impact-on-Twitter
[7]https://pypi.org/project/advertools/

similarities and differences in the tweeting habits of health organizations and leaders of the top-10 COVID-19 resilient countries.

Computational Resources and GitHub

The analysis was done using Compute Canada's service[8]. The computational resources provided by the '*graham*' cluster of the Compute Canada were as listed below:

1. **CPU**: 2x Intel E5-2683 v4 Broadwell @ 2.1GHz
2. **Memory (RAM)**: 30 GB

The supplementary material for this study – data, code, and results are available on the GitHub repository[6].

4.4 Results

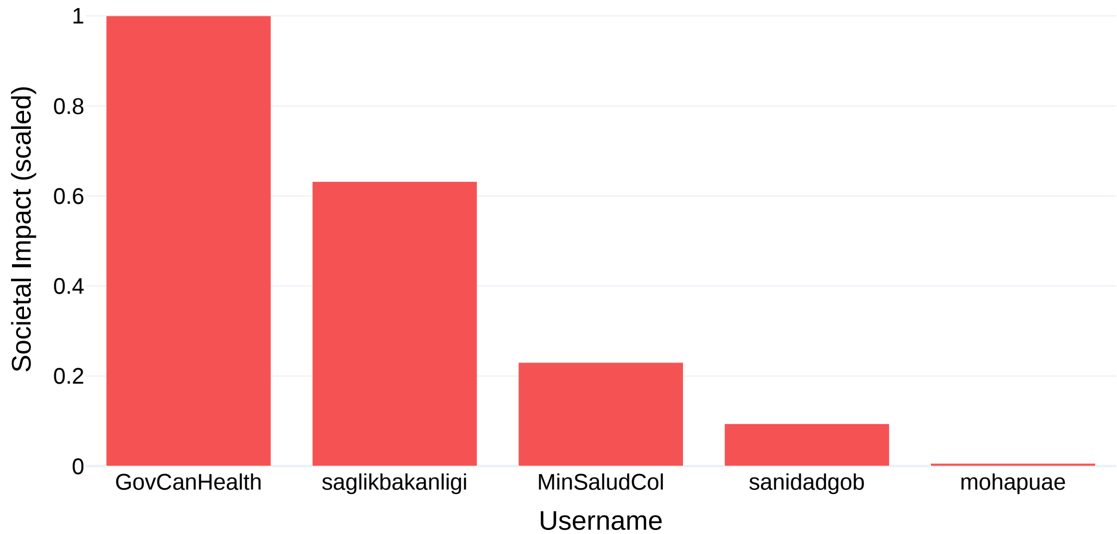
Societal Impact

Prime Minister of U.A.E, Mohammed bin Rashid Al Maktoum (societal impact: 1.000), had the highest societal impact overall, followed by Canadian Prime Minister Justin Trudeau (societal impact: 0.068) and Turkish President Recep Tayyip Erdoğan (societal impact: 0.033), among the leaders. Out of the health organizations, the Health Canada and Public Health Agency of Canada (PHAC, societal impact: 1.000) had the highest societal impact, followed by the Ministry of Health of the Republic of Turkey (societal impact: 0.632) and the Ministry of Health of Spain (societal impact: 0.094). The results for each of the factors affecting the societal impact, are individually explained in the following subsections.

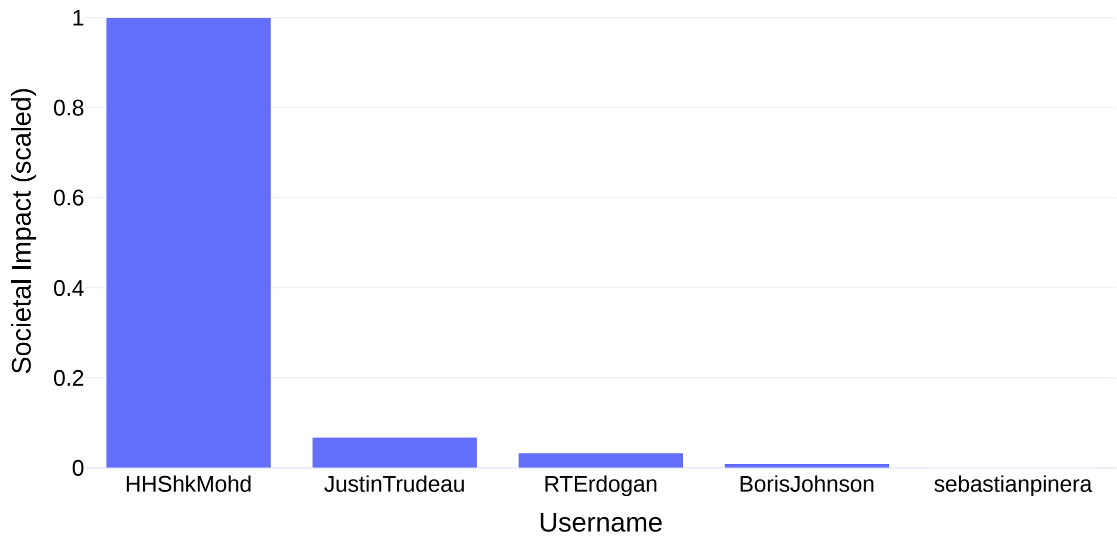
Engagement with impact

The user impact was scaled between the range 0 and 1 (1 denotes high user impact, and 0 denotes low user impact). The results indicate that the Turkish President (Recep Tayyip Erdoğan) had the greatest user impact (1.000), followed by U.K. Prime Minister (Boris

[8] https://www.computecanada.ca/research-portal/accessing-resources/available-resources/



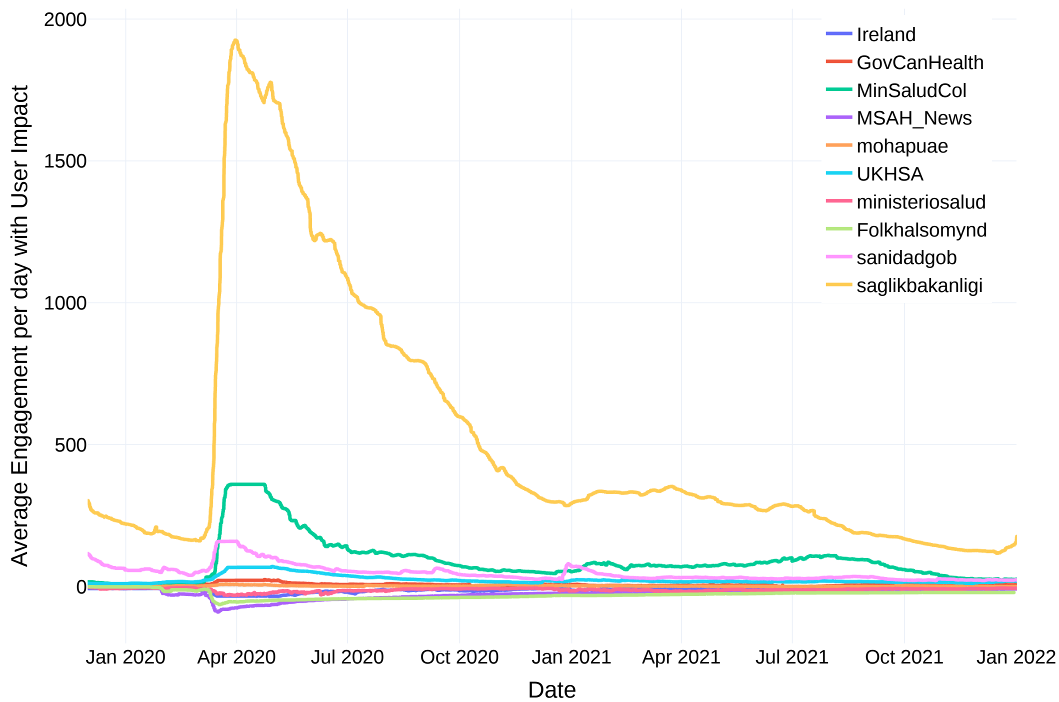
(a) Health organizations



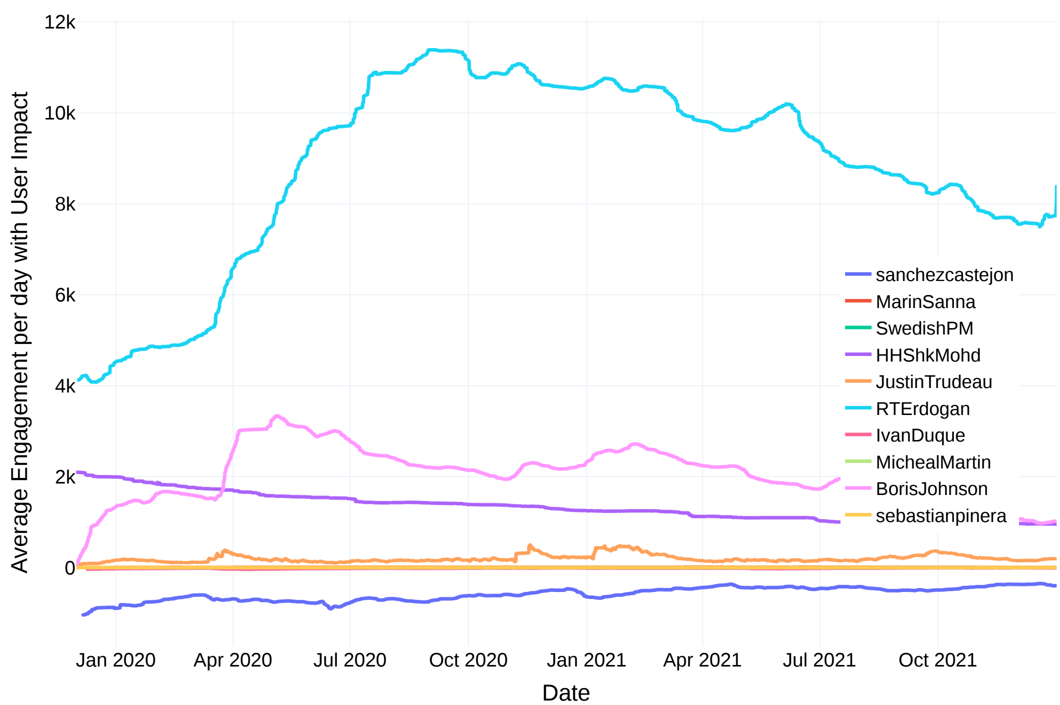
(b) Leaders

Figure 4.1: Societal impact (scaled) of the top-5 most impactful users by their user group, i.e., (a) health organizations, and (b) leaders of the top-10 COVID-19 resilient countries.

Johnson, user impact: 0.978), and the Prime Minister of U.A.E (Mohammed bin Rashid Al Maktoum, user impact: 0.663) among the leaders. Among the health organizations, the



(a) Health organizations



(b) Leaders

Figure 4.2: Average Engagement per day with user impact for (a) health organizations, and (b) leaders of the top-10 COVID-19 resilient countries.

Ministry of Health of the Republic of Turkey had the highest user impact (1.000), followed by the Ministry of Health and Social Protection of Colombia (user impact: 0.887) and the UK Health Security Agency (user impact: 0.778).

Among the health organizations, The Ministry of Health of the Republic of Turkey's user engagement is considerably higher than the other organizations (Figure 4.2a). The highest engagement was observed during April, 2020. This can be attributed to the impacts of COVID-19, specifically, the curfew mandate imposed by the Turkish government during this time. The user engagement gradually decreased, as the COVID-19 measures lifted, and the normalization process continued. Similar to the health organizations, Turkish President Recep Tayyip Erdoğan's user engagement (as shown in Figure 4.2b) is considerably higher than the other leaders, with the highest engagement recorded during August-October, 2020. The initial rise in engagement came in response to the finding of 320 billion cubic metres of natural gas in the Black Sea, which was made possible by drilling in the Danube-1 well, which began on July 20, 2020, as part of their goal of being a massive energy exporter [9].The subsequent spike in engagement occurred in the aftermath of the 6.6 magnitude earthquake that struck Izmir, Turkey on October 30, 2020, with the government agencies rallying to save people who were trapped[10].

Sentiment Strength

After computing the Sentiment Strength for all the users, it was found that most of the users had a neutral outlook on Twitter, except for the UK Health Security Agency (UKHSA), who had a highly negative opinion (sentiment strength: -0.999). Only six user accounts out of 20 reflected positive sentiment through their tweets; these were the leaders of Chile, U.A.E., Canada, Colombia, Sweden and the official account of the Public Health Agency of Canada (PHAC) (sentiment strength: 0.411). Among the leaders, U.A.E.'s Prime Minister Mohammed bin Rashid Al Maktoum had the highest positive sentiment strength (i.e., 0.746), followed by the Swedish Prime Minister, Magdalena Andersson (sentiment strength: 0.706)

[9]https://www.forbes.com/sites/arielcohen/2020/09/18/turkeys-new-natural-gas-find-in-the-black-sea-exciting-but-tricky-process-ahead/?sh=3c7697d15a86

[10]https://reliefweb.int/report/turkey/zmirturkey-earthquake-situation-report-no-01-30-october-2020

and Canadian Prime Minister, Justin Trudeau (sentiment strength: 0.512). Figure 4 depicts the sentiment strength of the top-5 leaders.

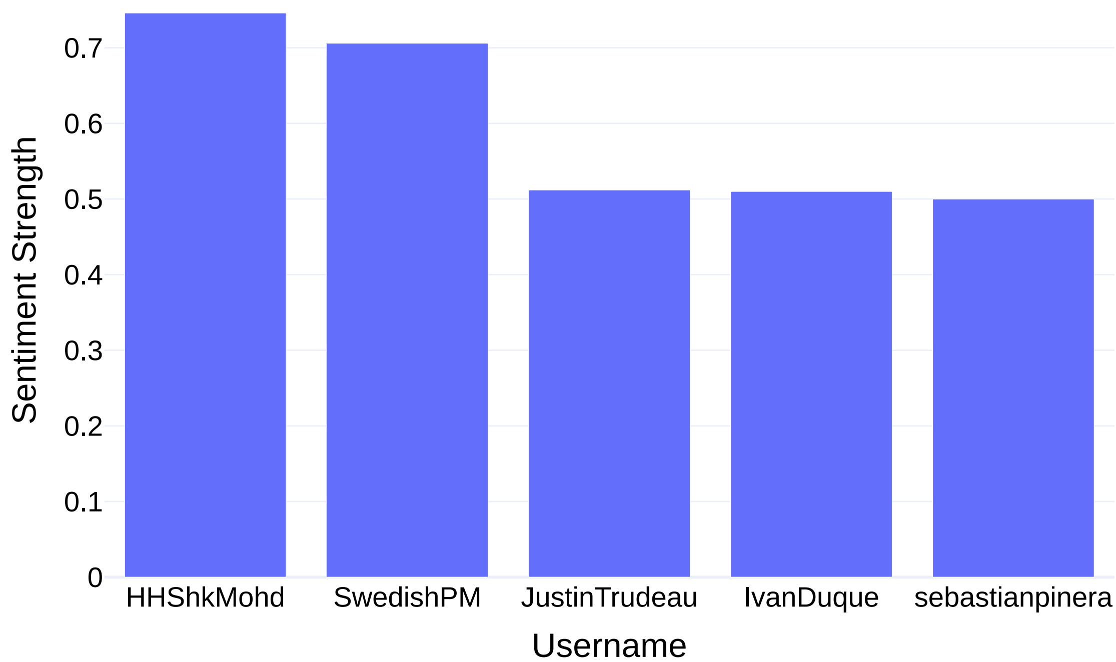
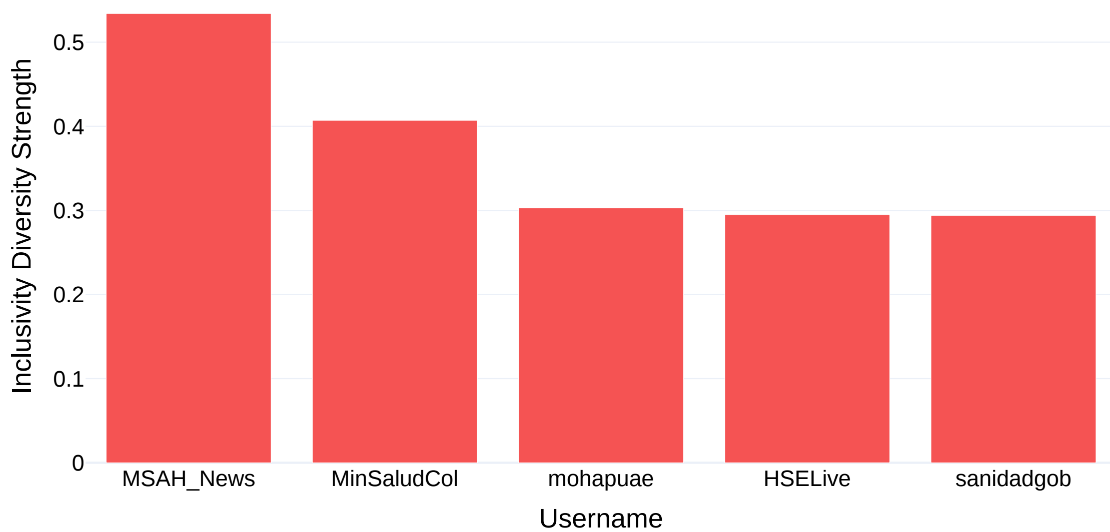


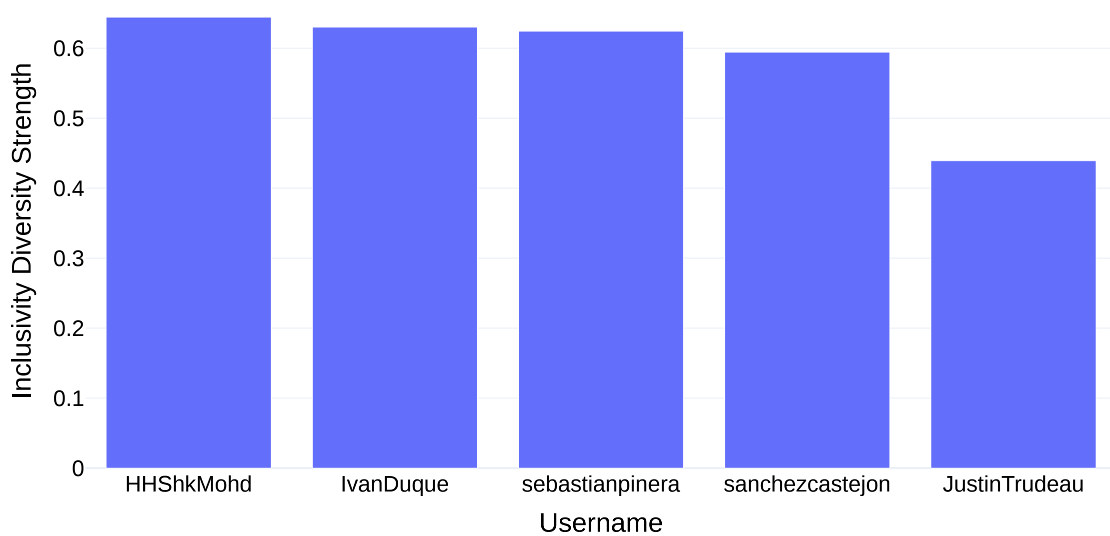
Figure 4.3: Sentiment Strength of top-5 leaders.

Inclusivity and Diversity Strength

Leaders of the top-10 COVID-19 resilient countries were more inclusive and diverse while tweeting compared to the health organizations of these countries. Among the leaders, U.A.E's Prime Minister had the highest inclusivity and diversity strength (i.e., 0.644), followed by the Colombian President, Iván Duque (inclusivity and diversity strength: 0.63), and the Chilean President, Sebastián Piñera (inclusivity and diversity strength: 0.624). For the health organizations, the results were slightly different, with Finland's Ministry of Social Affairs and Health having the highest inclusivity and diversity strength (i.e., 0.534), followed by the Colombian Ministry of Health and Social Protection (inclusivity and diversity strength: 0.407) and U.A.E.'s health organization, MOHAP (inclusivity and diversity strength: 0.303). Figure 4.4a and 4.4b illustrates the *Inclusivity and Diversity Strength* of the top-5 health organizations and leaders respectively.

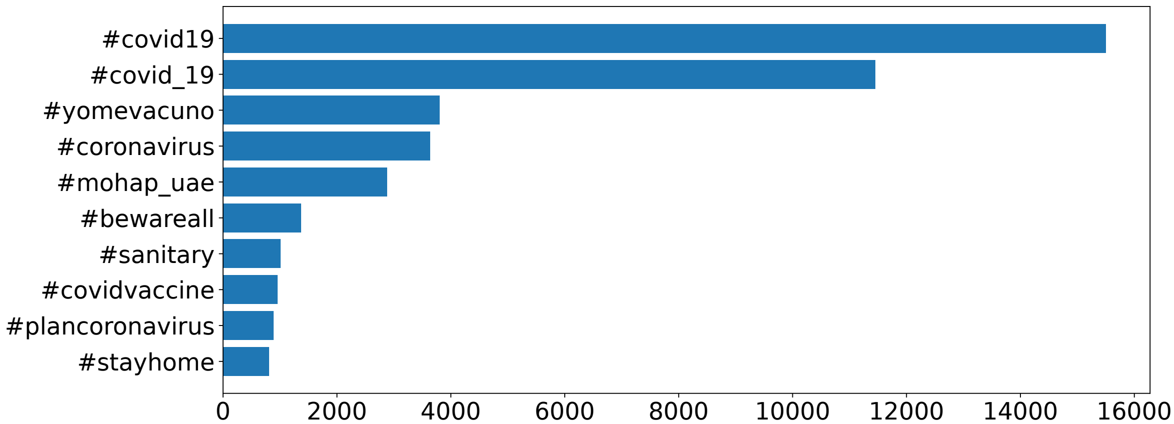


(a) Health organizations

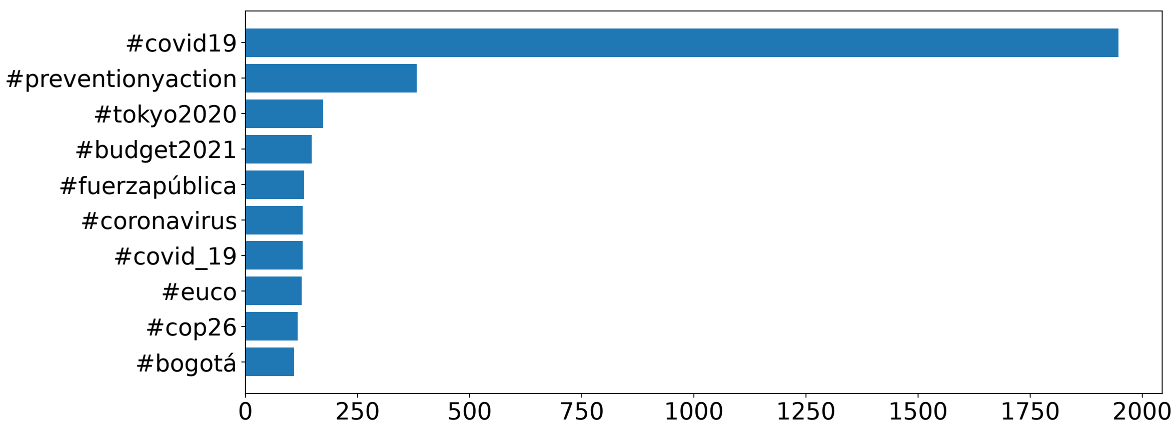


(b) Leaders

Figure 4.4: Inclusivity and Diversity Strength for top-5 (a) health organizations, and (b) leaders.



(a) Top #tags for health organizations



(b) Top #tags for leaders

Figure 4.5: Top #tags for (a) health organizations, and (b) leaders of the top-10 COVID-19 resilient countries.

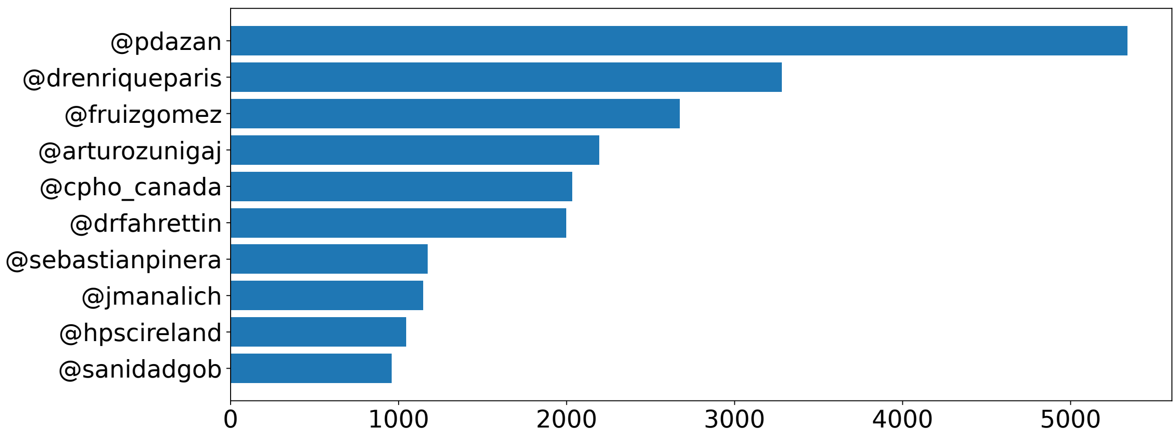
Content Analysis

The analysis of the hashtags served as a representation of the content, for each of the users. '*COVID-19*' was the most discussed topic among the accounts of the health organizations- as shown in Figure 4.5a which displays the high frequency of '*#covid19*', '*#covid_19*', '*#yomevacano*' (referring to the vaccination plans and status of Chile)[11], and '*#coronavirus*' hashtags in user's tweets. The data indicates that the health organizations communicated

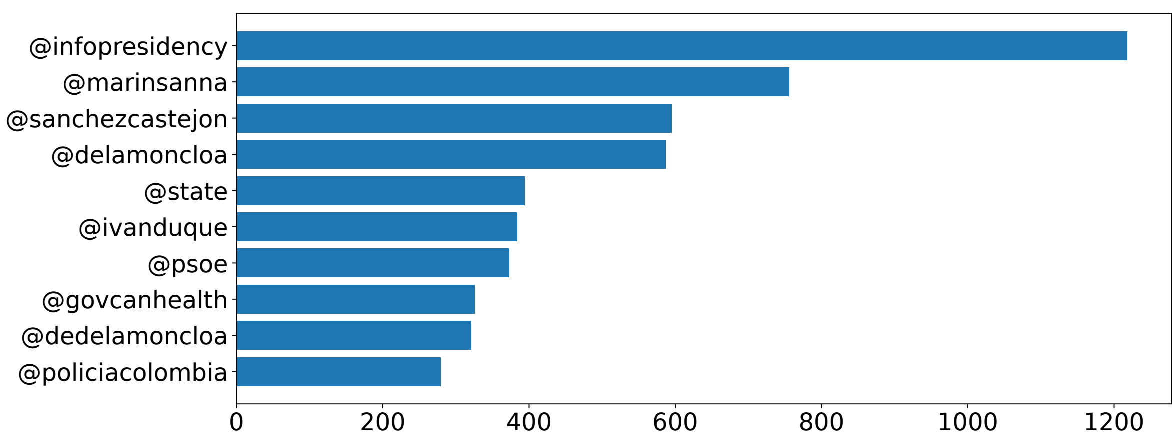
[11]https://www.gob.cl/yomevacuno/

about the COVID-19 pandemic with the same, or similar words. However, the results of the political leaders indicated contrasting content discussion. This is due to the fact that the political leaders would discuss relevant political issues, respective to their country-attributing to the diversity of hashtags. This concept is supported by Figure 4.5b which shows the variety of hashtags related to different topics; '*#covid19*', '*#cop25*' (referencing the 25th United Nations Climate Change Conference, held from December 2 to 13, 2019[12]), '*#tokyo2020*',

[12] https://unfccc.int/cop25



(a) Top mentions for health organizations



(b) Top mentions for leaders

Figure 4.6: Top mentions for (a) health organizations, and (b) leaders of the top-10 COVID-19 resilient countries.

'*#budget2021*', '*#euco*' (regarding the European Council[13]), '*#Bogotá*' (referring to the capital of Columbia), '*#FuerzaPública*' (referencing the public force of Columbia[14]), and '*#machnamh100*' (regarding an initiative of Ireland's President, Michael D. Higgins, which explores influential events during Ireland's Decade of Commemorations[15]). The frequency of mentions were measured to understand the interactions of each account. Figure 4.6a indicates the frequency of mentions among the health organizations. The mentioned accounts are current or former health ministers, or politicians, respective to each of the top-10 COVID-19 resilient countries selected for our analysis. Figure 4.6b represents the most recurrent mentions for the political leaders. The mentioned accounts were relevant political figures or organizations to the country.

4.5 Discussion & Conclusion

In this study, we present an approach for evaluating the societal impact of leaders and health organizations from the top-10 COVID-19 resilient countries using NLP-based text mining algorithms. We analyzed fairly significant volumes of textual data for assessing the societal impact by evaluating their public engagement, sentiment strength and, inclusivity and diversity strength.Our findings indicate that being the most active user on social media does not necessarily imply a higher level of societal impact. The Prime Ministers of the United Arab Emirates and Canada had significantly more societal impact than the leaders of Colombia and Spain, despite the latter's having a higher number of tweets. A similar observation is made for the health organizations, where the Canadian and Turkish health agencies created a substantially greater societal impact than those of Colombia and Ireland. People are also more inclined to engage with neutral tweets, which normally contain some sort of public notification, rather than entirely positive or negative tweets, according to our findings. Using specific hashtags, undoubtedly aids in driving engagement, as we have seen that the majority of public engagement is highly slanted towards tweets containing hashtags related to '*COVID-19*'. Furthermore, we note that user engagement for both the user groups,

[13] https://www.consilium.europa.eu/en/european-council/
[14] https://www.constitucioncolombia.com/titulo-7/capitulo-7
[15] https://president.ie/en/news/article/machnamh-100-president-of-irelands-centenary-reflections

i.e., health organizations and leaders, follows a predictable pattern, with peaks emerging around events of emergency or public welfare announcements. Additionally, leaders of the top-10 COVID-19 resilient nations targeted wider audience than their health organizations when it came to inclusion and diversity. As a result, each of the individual characteristics, i.e., public involvement, sentiment strength, inclusivity and diversity strength, played an equally important role in determining a user's societal influence. It is observed that the societal impact would be affected if any of the three were neglected. Thus, the tweets that cover more current events, are neutral and target a wider audience may have a greater societal impact. Leaders and health organizations may incorporate this NLP-based social media research to develop content that has a greater societal impact. Overall, we believe that quantifying the societal impact and analyzing the tweet content provides a better understanding of how posting the appropriate tweet at the right time may make all the difference in communal impact.

The results of this study are confined to the COVID-19 timeline selected, i.e., between December 1, 2019, to December 31, 2021. To further comprehend the societal impact of leaders and health organizations in different timeframes, the researchers might use alternative approaches to organize their data. Moreover, our research focuses on leaders' and healthcare organizations' Twitter data, which is often clean and requires little pre-processing. Because our research was confined to textual data, we could not account for the influence of image characteristics or knowledge graphs related to individual tweets. However, it would be intriguing to see how this methodology behaves on the tweets of other cabinet members and decision-makers of these countries, as well as investigate the organic and paid audiences, if any exist. Another area that future research might look at is the demographics of the individuals who are interacting with these contents.

Chapter 5

Conclusion

The first chapter of this thesis provides background information on LMs by outlining the historical development, the tasks for which they are applied, and the evaluation criteria that have been employed to assess their efficacy. Additionally, it aids in comprehending how social media datasets have been applied throughout the time to enhance the application of LMs in various domains and what new benchmarks have been established to date for diverse tasks and domains.

In chapter 3, a novel approach is put forward for assessing the candidate engagement, topics' stickiness, candidate synergy (i.e., content and stance similarity), and candidate's social reach on Twitter during the 2020 US Presidential Elections. An effective Twitter communication strategy for the candidates and government might be developed using the findings of this study, which used an empirical technique to examine how internal collaboration and participation among candidates affected the election outcomes.

The last chapter discusses the approach suggested to assess the societal association/ impact of healthcare organizations and political leaders from the top-10 COVID-19 resilient nations, according to the Bloomberg rating. The social association includes computing the public engagement, sentiment strength, and inclusivity and diversity strength using the tweets the leaders and organizations created. According to research, having a high degree of societal effect may not always have the most significant social association. A user's social association may be affected if one of the three metrics—public engagement, sentiment strength, or inclusivity and diversity strength—falls below par. Thus, understanding how publishing the correct tweet at the right moment may significantly influence society can be better understood by quantifying the societal impact.

The research presented in this thesis was intended to inspire further, more in-depth investigation required to create models for analyzing social media data for various use-cases. Users on social media platforms tend to compose texts informally and without considering the content's audience, sentiment, and impact. Additionally, because of the limitations on tweet length, the content may be condensed and may not express the intended idea. Therefore, mining and analyzing tweets for situations like elections and healthcare emergencies, can help eliminate the challenges caused by undirected dialogues and help the government and citizens foster a better digital environment.

Bibliography

[1] Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. "ARBERT & MARBERT: deep bidirectional transformers for Arabic". In: *arXiv preprint arXiv:2101.01785* (2020).

[2] Gustavo Aguilar, Sudipta Kar, and Thamar Solorio. "Lince: A centralized benchmark for linguistic code-switching evaluation". In: *arXiv preprint arXiv:2005.04322* (2020).

[3] Neel Alex et al. "RAFT: A real-world few-shot text classification benchmark". In: *arXiv preprint arXiv:2109.14076* (2021).

[4] James Allan et al. "Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, university of massachusetts amherst, september 2002". In: *ACM SIGIR Forum*. Vol. 37. 1. ACM New York, NY, USA. 2003, pp. 31–47.

[5] Khaled Mohammad Alomari, Hatem M ElSherif, and Khaled Shaalan. "Arabic tweets sentimental analysis using machine learning". In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. 2017, pp. 602–610.

[6] Sultan Alrowili and K Vijay-Shanker. "ArabicTransformer: Efficient Large Arabic Language Model with Funnel Transformer and ELECTRA Objective". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021, pp. 1255–1261.

[7] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473* (2014).

[8] Daniel Balliet, Junhui Wu, and Carsten KW De Dreu. "Ingroup favoritism in cooperation: a meta-analysis." In: *Psychological bulletin* 140.6 (2014), p. 1556.

[9] Krisztian Balog, Leif Azzopardi, and Maarten De Rijke. "Formal models for expert finding in enterprise corpora". In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 2006, pp. 43–50.

[10] Satanjeev Banerjee and Alon Lavie. "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments". In: *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*. 2005, pp. 65–72.

[11] Barkha Bansal and Sangeet Srivastava. "On predicting elections with hybrid topic based sentiment analysis of tweets". In: *Procedia Computer Science* 135 (2018), pp. 346–353.

[12] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. "Xlm-t: A multilingual language model toolkit for twitter". In: *arXiv preprint arXiv:2104.12250* (2021).

[13] Francesco Barbieri et al. "TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1644–1650. DOI: `10.18653/v1/2020.findings-emnlp.148`. URL: `https://aclanthology.org/2020.findings-emnlp.148`.

[14] Gely P Basharin, Amy N Langville, and Valeriy A Naumov. "The life and work of AA Markov". In: *Linear algebra and its applications* 386 (2004), pp. 3–26.

[15] Valerio Basile et al. "Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter". In: *Proceedings of the 13th international workshop on semantic evaluation*. 2019, pp. 54–63.

[16] Mikhail Belkin et al. "Reconciling modern machine-learning practice and the classical bias–variance trade-off". In: *Proceedings of the National Academy of Sciences* 116.32 (2019), pp. 15849–15854.

[17] A Benetoli, TF Chen, and P Aslani. "How patients' use of social media impacts their interactions with healthcare professionals". In: *Patient Education and Counseling* 101.3 (2018), pp. 439–444.

[18] Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. "A neural probabilistic language model". In: *Advances in neural information processing systems* 13 (2000).

[19] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks* 5.2 (1994), pp. 157–166.

[20] Adam Berger and John Lafferty. "Information Retrieval as Statistical Translation In proceedings of the 22nd Conference on Research and Development in Information Retrieval,(SIGIR)". In: *CA August* (1999), pp. 15–19.

[21] John Carlo Bertot et al. "Social media technology and government transparency". In: *Computer* 43.11 (2010), pp. 53–59.

[22] Prajjwal Bhargava and Vincent Ng. "Commonsense Knowledge Reasoning and Generation with Pre-trained Language Models: A Survey". In: *arXiv preprint arXiv:2201.12438* (2022).

[23] Vaibhav Bhat et al. "AdCOFE: Advanced Contextual Feature Extraction in Conversations for emotion classification". In: *arXiv preprint arXiv:2104.04517* (2021).

[24] Abhik Bhattacharjee et al. "Banglabert: Lagnuage model pretraining and benchmarks for low-resource language understanding evaluation in bangla". In: *Findings of the North American Chapter of the Association for Computational Linguistics: NAACL* (2022).

[25] Sanmitra Bhattacharya, Padmini Srinivasan, and Philip Polgreen. "Social media engagement analysis of US Federal health agencies on Facebook". In: *BMC Medical Informatics and Decision Making* 17.1 (2017), pp. 1–12.

[26] Muhammad Bilal et al. "Predicting elections: Social media data and techniques". In: *2019 international conference on engineering and emerging technologies (ICEET)*. IEEE. 2019, pp. 1–6.

[27] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan (2003), pp. 993–1022.

[28] Enrique Bonsón and Michaela Bednárová. "The use of YouTube in western European municipalities". In: *Government information quarterly* 35.2 (2018), pp. 223–232.

[29] Enrique Bonsón, David Perea, and Michaela Bednárová. "Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities". In: *Government Information Quarterly* 36.3 (2019), pp. 480–489.

[30] Enrique Bonsón and Melinda Ratkai. "A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page". In: *Online Information Review* (2013).

[31] Enrique Bonsón, Sonia Royo, and Melinda Ratkai. "Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe". In: *Government information quarterly* 32.1 (2015), pp. 52–62.

[32] Enrique Bonsón, Sonia Royo, and Melinda Ratkai. "Facebook practices in Western European municipalities: An empirical analysis of activity and citizens' engagement". In: *Administration & Society* 49.3 (2017), pp. 320–347.

[33] Enrique Bonsón et al. "Local e-government 2.0: Social media and corporate transparency in municipalities". In: *Government information quarterly* 29.2 (2012), pp. 123–132.

[34] Porismita Borah. "Political Facebook use: Campaign strategies used in 2008 and 2012 presidential elections". In: *Journal of Information Technology & Politics* 13.4 (2016), pp. 326–338.

[35] Jonathan Bragg et al. "Flex: Unifying evaluation for few-shot nlp". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 15787–15800.

[36] Peter F Brown et al. "Class-based n-gram models of natural language". In: *Computational linguistics* 18.4 (1992), pp. 467–480.

[37] Tom Brown et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[38] Axel Bruns and Brenda Moon. “Social media in Australian Federal elections: comparing the 2013 and 2016 campaigns”. In: *Journalism & Mass Communication Quarterly* 95.2 (2018), pp. 425–448.

[39] Guillaume Cabanac, Ingo Frommholz, and Philipp Mayr. “Scholarly literature mining with information retrieval and natural language processing: Preface”. In: *Scientometrics* 125.3 (2020), pp. 2835–2840.

[40] Samuel Cahyawijaya et al. “IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation”. In: *arXiv preprint arXiv:2104.08200* (2021).

[41] José Canete et al. “Spanish pre-trained bert model and evaluation data”. In: *Pml4dc at iclr* 2020 (2020), pp. 1–10.

[42] Casimiro Pio Carrino et al. “Biomedical and Clinical Language Models for Spanish: On the Benefits of Domain-Specific Pretraining in a Mid-Resource Scenario”. In: *arXiv preprint arXiv:2109.03570* (2021).

[43] Tommaso Caselli et al. “Hatebert: Retraining bert for abusive language detection in english”. In: *arXiv preprint arXiv:2010.12472* (2020).

[44] Pew Research Centre. *Differences in How Democrats and Republicans Behave on Twitter*. `https://www.pewresearch.org/politics/2020/10/15/differences-in-how-democrats-and-republicans-behave-on-twitter/`. Oct. 2020.

[45] Tanmoy Chakraborty et al. *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers.* Vol. 1402. Springer Nature, 2021.

[46] Dhivya Chandrasekaran and Vijay Mago. “Evolution of Semantic Similarity—A Survey”. In: *ACM Computing Surveys (CSUR)* 54.2 (2021), pp. 1–37.

[47] Emily Chen, Ashok Deb, and Emilio Ferrara. “# Election2020: the first public Twitter dataset on the 2020 US Presidential election”. In: *Journal of Computational Social Science* (2021), pp. 1–18.

[48] Mia Xu Chen et al. “The best of both worlds: Combining recent advances in neural machine translation”. In: *arXiv preprint arXiv:1804.09849* (2018).

[49] Yahui Chen. "Convolutional neural network for sentence classification". MA thesis. University of Waterloo, 2015.

[50] Soon Chun et al. "Government 2.0: Making connections between citizens, data and government". In: *Information Polity* 15.1, 2 (2010), pp. 1–9.

[51] Ronan Collobert and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning". In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 160–167.

[52] Ronan Collobert et al. "Natural language processing (almost) from scratch". In: *Journal of machine learning research* 12.ARTICLE (2011), pp. 2493–2537.

[53] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". In: *Machine learning* 20.3 (1995), pp. 273–297.

[54] K Courtney et al. "The use of social media in healthcare: organizational, clinical, and patient perspectives". In: *Enabling health and healthcare through ICT: available, tailored and closer* 183 (2013), p. 244.

[55] Andrew M Dai and Quoc V Le. "Semi-supervised sequence learning". In: *Advances in neural information processing systems* 28 (2015).

[56] Linh Dang-Xuan et al. "An investigation of influentials and the role of sentiment in political communication on Twitter during election periods". In: *Information, Communication & Society* 16.5 (2013), pp. 795–825.

[57] Michal Daniluk et al. "Synerise at RecSys 2021: Twitter user engagement prediction with a fast neural model". In: *RecSysChallenge'21: Proceedings of the Recommender Systems Challenge 2021*. 2021, pp. 15–21.

[58] Michał Daniluk et al. "Frustratingly short attention spans in neural language modeling". In: *arXiv preprint arXiv:1702.04521* (2017).

[59] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association

for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: https://aclanthology.org/N19-1423.

[60] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[61] Hani Al-Dmour et al. "Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: integrated model". In: *Journal of Medical Internet Research* 22.8 (2020), e19996.

[62] Jeffrey Donahue et al. "Long-term recurrent convolutional networks for visual recognition and description". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2625–2634.

[63] Wilberforce S Dzisah. "Social media and elections in Ghana: Enhancing democratic participation". In: *African Journalism Studies* 39.1 (2018), pp. 27–47.

[64] Robin Effing, Jos Van Hillegersberg, and Theo Huibers. "Social media and political participation: are Facebook, Twitter and YouTube democratizing our political systems?" In: *International conference on electronic participation*. Springer. 2011, pp. 25–35.

[65] Jeffrey L Elman. "Finding structure in time". In: *Cognitive science* 14.2 (1990), pp. 179–211.

[66] Mai ElSherief et al. "Latent Hatred: A Benchmark for Understanding Implicit Hate Speech". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 345–363. DOI: 10.18653/v1/2021.emnlp-main.29. URL: https://aclanthology.org/2021.emnlp-main.29.

[67] Elisabetta Fersini, Debora Nozza, and Paolo Rosso. "Overview of the evalita 2018 task on automatic misogyny identification (ami)". In: *EVALITA Evaluation of NLP and Speech Tools for Italian* 12 (2018), p. 59.

[68] Felix Gaisbauer et al. "Ideological differences in engagement in public debate on Twitter". In: *Plos one* 16.3 (2021), e0249241.

[69] Michael C Galgoczy et al. "(Re) shaping online narratives: when bots promote the message of President Trump during his first impeachment". In: *PeerJ Computer Science* 8 (2022), e947.

[70] Edoardo Gallo and Chang Yan. "The effects of reputational and social knowledge on cooperation". In: *Proceedings of the National Academy of Sciences* 112.12 (2015), pp. 3647–3652.

[71] Klaifer Garcia and Lilian Berton. "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA". In: *Applied Soft Computing* 101 (2021), p. 107057.

[72] Bill Gates. "Responding to Covid-19—a once-in-a-century pandemic?" In: *New England Journal of Medicine* 382.18 (2020), pp. 1677–1679.

[73] Jonas Gehring et al. "Convolutional sequence to sequence learning". In: *International conference on machine learning*. PMLR. 2017, pp. 1243–1252.

[74] Sebastian Gehrmann et al. "The gem benchmark: Natural language generation, its evaluation and metrics". In: *arXiv preprint arXiv:2102.01672* (2021).

[75] Anton Golubev and Natalia Loukachevitch. "Improving results on Russian sentiment datasets". In: *Conference on Artificial Intelligence and Natural Language*. Springer. 2020, pp. 109–121.

[76] Alex Graves. "Generating sequences with recurrent neural networks". In: *arXiv preprint arXiv:1308.0850* (2013).

[77] Jörg Gross and Carsten KW De Dreu. "The rise and fall of cooperation through reputation and group polarization". In: *Nature communications* 10.1 (2019), pp. 1–10.

[78] Matthew Grossmann. *Artists of the possible: Governing networks and American policy change since 1945*. Oxford University Press, 2014.

[79] Purva Grover et al. "Polarization and acculturation in US Election 2016 outcomes–Can twitter analytics predict changes in voting preferences". In: *Technological Forecasting and Social Change* 145 (2019), pp. 438–460.

[80] Anatoliy Gruzd, James Lannigan, and Kevin Quigley. "Examining government cross-platform engagement in social media: Instagram vs Twitter and the big lift project". In: *Government Information Quarterly* 35.4 (2018), pp. 579–587.

[81] Anatoliy Gruzd and Jeffrey Roy. "Social media and local government in canada: an examination of presence and purpose". In: *Social Media and Local Governments*. Springer, 2016, pp. 79–94.

[82] Yu Gu et al. "Domain-specific language model pretraining for biomedical natural language processing". In: *ACM Transactions on Computing for Healthcare (HEALTH)* 3.1 (2021), pp. 1–23.

[83] Shangwei Guo et al. "Threats to Pre-trained Language Models: Survey and Taxonomy". In: *arXiv preprint arXiv:2202.06862* (2022).

[84] Xiao Guo et al. "Cord19sts: Covid-19 semantic textual similarity dataset". In: *arXiv preprint arXiv:2007.02461* (2020).

[85] Suchin Gururangan et al. "Don't stop pretraining: adapt language models to domains and tasks". In: *arXiv preprint arXiv:2004.10964* (2020).

[86] Michael Haman. "The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic". In: *Heliyon* 6.11 (2020), e05540.

[87] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[88] Lei He et al. "KLMo: Knowledge Graph Enhanced Pretrained Language Model with Fine-Grained Relationships". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 4536–4542. DOI: `10.18653/v1/2021.findings-emnlp.384`. URL: `https://aclanthology.org/2021.findings-emnlp.384`.

[89] Lihong He et al. "On the dynamics of user engagement in news comment media". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.1 (2020), e1342.

[90] Sam Henry et al. "2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records". In: *Journal of the American Medical Informatics Association* 27.1 (2020), pp. 3–12.

[91] Karl Moritz Hermann et al. "Teaching machines to read and comprehend". In: *Advances in neural information processing systems* 28 (2015).

[92] Djoerd Hiemstra and Wessel Kraaij. "Twenty-One at TREC-7: Ad-hoc and cross-language track". In: (1999).

[93] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[94] Matthew Hoffman, Francis Bach, and David Blei. "Online learning for latent dirichlet allocation". In: *advances in neural information processing systems* 23 (2010).

[95] Thomas Hofmann. "Probabilistic latent semantic indexing". In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. 1999, pp. 50–57.

[96] Linda D Hollebeek, Mark S Glynn, and Roderick J Brodie. "Consumer brand engagement in social media: Conceptualization, scale development and validation". In: *Journal of interactive marketing* 28.2 (2014), pp. 149–165.

[97] John J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the national academy of sciences* 79.8 (1982), pp. 2554–2558.

[98] Jeremy Howard and Sebastian Ruder. "Universal Language Model Fine-tuning for Text Classification". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, July 2018, pp. 328–339. DOI: `10.18653/v1/P18-1031`. URL: `https://aclanthology.org/P18-1031`.

[99] Jeremy Howard and Sebastian Ruder. "Universal language model fine-tuning for text classification". In: *arXiv preprint arXiv:1801.06146* (2018).

[100] Junjie Hu et al. "Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 4411–4421.

[101] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. "Spatial transformer networks". In: *Advances in neural information processing systems* 28 (2015).

[102] Mustafa Jarrar, Mohammed Khalilia, and Sana Ghanem. "Wojood: Nested arabic named entity corpus and recognition using bert". In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France, June*. 2022.

[103] Frederick Jelinek. *Statistical methods for speech recognition*. MIT press, 1998.

[104] Zhengbao Jiang et al. "X-FACTR: Multilingual factual knowledge retrieval from pre-trained language models". In: *arXiv preprint arXiv:2010.06189* (2020).

[105] Karen Sparck Jones. "A statistical interpretation of term specificity and its application in retrieval". In: *Journal of documentation* (1972).

[106] Ferdin Joe John Joseph. "Twitter based outcome predictions of 2019 indian general elections using decision tree". In: *2019 4th International Conference on Information Technology (InCIT)*. IEEE. 2019, pp. 50–53.

[107] Jakob Jünger and Birte Fähnrich. "Does really no one care? Analyzing the public engagement of communication scientists on Twitter". In: *New Media & Society* 22.3 (2020), pp. 387–408.

[108] Divyanshu Kakwani et al. "IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages". In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. 2020, pp. 4948–4961.

[109] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences". In: *arXiv preprint arXiv:1404.2188* (2014).

[110] Nal Kalchbrenner et al. "Neural machine translation in linear time". In: *arXiv preprint arXiv:1610.10099* (2016).

[111] Katikapalli Subramanyam Kalyan, Ajit Rajasekharan, and Sivanesan Sangeetha. "AMMU: a survey of transformer-based biomedical pretrained language models". In: *Journal of biomedical informatics* (2021), p. 103982.

[112] Katikapalli Subramanyam Kalyan, Ajit Rajasekharan, and Sivanesan Sangeetha. "Ammus: A survey of transformer-based pretrained models in natural language processing". In: *arXiv preprint arXiv:2108.05542* (2021).

[113] Katikapalli Subramanyam Kalyan and Sivanesan Sangeetha. "A hybrid approach to measure semantic relatedness in biomedical concepts". In: *arXiv preprint arXiv:2101.10196* (2021).

[114] Katikapalli Subramanyam Kalyan and Sivanesan Sangeetha. "Target concept guided medical concept normalization in noisy user-generated texts". In: *Proceedings of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*. 2020, pp. 64–73.

[115] Nora Kassner, Philipp Dufter, and Hinrich Schütze. "Multilingual LAMA: Investigating knowledge in multilingual pretrained language models". In: *arXiv preprint arXiv:2102.00894* (2021).

[116] Nora Kassner and Hinrich Schütze. "Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly". In: *arXiv preprint arXiv:1911.03343* (2019).

[117] Christoph Kecht et al. "Event Log Construction from Customer Service Conversations Using Natural Language Inference". In: *2021 3rd International Conference on Process Mining (ICPM)*. IEEE. 2021, pp. 144–151.

[118] Kazi Zainab Khanam, Gautam Srivastava, and Vijay Mago. "The homophily principle in social network analysis: A survey". In: *Multimedia Tools and Applications* (2022), pp. 1–44.

[119] Simran Khanuja et al. "GLUECoS: An evaluation benchmark for code-switched NLP". In: *arXiv preprint arXiv:2004.12376* (2020).

[120] Daniel Khashabi et al. "Genie: A leaderboard for human-in-the-loop evaluation of text generation". In: *arXiv preprint arXiv:2101.06561* (2021).

[121] Daniel Khashabi et al. “Parsinlu: a suite of language understanding challenges for persian”. In: *Transactions of the Association for Computational Linguistics* 9 (2021), pp. 1163–1178.

[122] Diksha Khurana et al. “Natural language processing: State of the art, current trends and challenges”. In: *Multimedia Tools and Applications* (2022), pp. 1–32.

[123] Y Kim et al. “Convolutional neural networks for sentence classification. arxXiv, doi”. In: *arXiv preprint arXiv:1408.5882* (2014).

[124] Fajri Koto et al. “IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP”. In: *arXiv preprint arXiv:2011.00677* (2020).

[125] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “Imagenet classification with deep convolutional neural networks”. In: *Advances in neural information processing systems* 25 (2012).

[126] Shaopeng Lai et al. “Type-Driven Multi-Turn Corrections for Grammatical Error Correction”. In: *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 3225–3236. DOI: 10.18653/v1/2022.findings-acl.254. URL: https://aclanthology.org/2022.findings-acl.254.

[127] Jennifer M Larson. “Networks of Conflict and Cooperation”. In: *Annual Review of Political Science* 24 (2021), pp. 89–107.

[128] Hang Le et al. “Flaubert: Unsupervised language model pre-training for french”. In: *arXiv preprint arXiv:1912.05372* (2019).

[129] Huaye Li and Yasuaki Sakamoto. “Social impacts in social media: An examination of perceived truthfulness and sharing of information”. In: *Computers in Human Behavior* 41 (2014), pp. 278–287.

[130] Yaobo Liang et al. “XGLUE: A new benchmark datasetfor cross-lingual pre-training, understanding and generation”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 6008–6018.

[131] Joseph Lilleberg, Yun Zhu, and Yanqing Zhang. "Support vector machines and word2vec for text classification with semantic features". In: *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*. IEEE. 2015, pp. 136–140.

[132] Bill Yuchen Lin et al. "Common sense beyond english: Evaluating and improving multilingual language models for commonsense reasoning". In: *arXiv preprint arXiv:2106.06937* (2021).

[133] Chin-Yew Lin. "Rouge: A package for automatic evaluation of summaries". In: *Text summarization branches out*. 2004, pp. 74–81.

[134] Dayiheng Liu et al. "Glge: A new general language generation evaluation benchmark". In: *arXiv preprint arXiv:2011.11928* (2020).

[135] Dilin Liu and Lei Lei. "The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election". In: *Discourse, Context & Media* 25 (2018), pp. 143–152.

[136] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. "Recurrent neural network for text classification with multi-task learning". In: *arXiv preprint arXiv:1605.05101* (2016).

[137] Xiao Liu et al. "Self-supervised learning: Generative or contrastive". In: *IEEE Transactions on Knowledge and Data Engineering* (2021).

[138] Shuai Lu et al. "Codexglue: A machine learning benchmark dataset for code understanding and generation". In: *arXiv preprint arXiv:2102.04664* (2021).

[139] Dario Madeo and Chiara Mocenni. "Self-regulation versus social influence for promoting cooperation on networks". In: *Scientific reports* 10.1 (2020), pp. 1–10.

[140] Jonathan Mallinson, Rico Sennrich, and Mirella Lapata. "Paraphrasing revisited with neural machine translation". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2017, pp. 881–893.

[141] James L McClelland, David E Rumelhart, and Geoffrey E Hinton. "The appeal of parallel distributed processing". In: *MIT Press, Cambridge MA* 3 (1986), p. 44.

[142] Shikib Mehri, Mihail Eric, and Dilek Hakkani-Tur. “Dialoglue: A natural language understanding benchmark for task-oriented dialogue”. In: *arXiv preprint arXiv:2009.13570* (2020).

[143] Oren Melamud, Jacob Goldberger, and Ido Dagan. “context2vec: Learning Generic Context Embedding with Bidirectional LSTM”. In: *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 51–61. DOI: 10.18653/v1/K16-1006. URL: https://aclanthology.org/K16-1006.

[144] Chad A Melton et al. “Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence”. In: *Journal of Infection and Public Health* 14.10 (2021), pp. 1505–1512.

[145] Raina M Merchant and Nicole Lurie. “Social media and emergency preparedness in response to novel coronavirus”. In: *JAMA* 323.20 (2020), pp. 2011–2012.

[146] Tomas Mikolov et al. “Distributed representations of words and phrases and their compositionality”. In: *Advances in neural information processing systems* 26 (2013).

[147] Tomas Mikolov et al. “Efficient estimation of word representations in vector space”. In: *arXiv preprint arXiv:1301.3781* (2013).

[148] Tomas Mikolov et al. “Recurrent neural network based language model.” In: *Interspeech*. Vol. 2. 3. Makuhari. 2010, pp. 1045–1048.

[149] Tomáš Mikolov, Wen-tau Yih, and Geoffrey Zweig. “Linguistic regularities in continuous space word representations”. In: *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*. 2013, pp. 746–751.

[150] AL-Smadi Mohammad et al. “Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features”. In: *Information Processing & Management* 53.3 (2017), pp. 640–652.

[151] Faith W Mutinda et al. “Detecting redundancy in electronic medical records using clinical bert”. In: *26* (2020).

[152] Mahmoud Nabil, Mohamed Aly, and Amir Atiya. "Astd: Arabic sentiment tweets dataset". In: *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2015, pp. 2515–2519.

[153] David Newman et al. "Automatic evaluation of topic coherence". In: *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. 2010, pp. 100–108.

[154] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. "BERTweet: A pre-trained language model for English Tweets". In: *arXiv preprint arXiv:2005.10200* (2020).

[155] Donald F Norris and Christopher G Reddick. "Local e-government in the United States: Transformation or incremental change?" In: *Public Administration Review* 73.1 (2013), pp. 165–175.

[156] Deni Kumianto Nugroho. "US presidential election 2020 prediction based on Twitter data using lexicon-based sentiment analysis". In: *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE. 2021, pp. 136–141.

[157] Kostiantyn Omelianchuk et al. "GECToR – Grammatical Error Correction: Tag, Not Rewrite". In: *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*. Seattle, WA, USA → Online: Association for Computational Linguistics, July 2020, pp. 163–170. DOI: `10.18653/v1/2020.bea-1.16`. URL: `https://aclanthology.org/2020.bea-1.16`.

[158] David Oniani and Yanshan Wang. "A qualitative evaluation of language models on automatic question-answering for covid-19". In: *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. 2020, pp. 1–9.

[159] Kishore Papineni et al. "Bleu: a method for automatic evaluation of machine translation". In: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002, pp. 311–318.

[160] Sungjoon Park et al. "Klue: Korean language understanding evaluation". In: *arXiv preprint arXiv:2105.09680* (2021).

[161] John H Parmelee and Shannon L Bichard. *Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public*. Lexington books, 2011.

[162] Krunal Dhiraj Patel et al. "Analyzing use of Twitter by diabetes online community". In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2019, pp. 937–944.

[163] Krunal Dhiraj Patel et al. "Using Twitter for diabetes community analysis". In: *Network Modeling Analysis in Health Informatics and Bioinformatics* 9.1 (2020), pp. 1–16.

[164] Jessica L Peck. "Social media in nursing education: responsible integration for meaningful use". In: *Journal of Nursing Education* 53.3 (2014), pp. 164–169.

[165] Jānis Peisenieks and Raivis Skadiņš. "Uses of machine translation in the sentiment analysis of tweets". In: *Human Language Technologies–The Baltic Perspective*. IOS Press, 2014, pp. 126–131.

[166] Yifan Peng, Shankai Yan, and Zhiyong Lu. "Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets". In: *arXiv preprint arXiv:1906.05474* (2019).

[167] Jeffrey Pennington, Richard Socher, and Christopher Manning. "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. URL: https://aclanthology.org/D14-1162.

[168] Evan Perkoski. "Internal politics and the fragmentation of armed groups". In: *International Studies Quarterly* 63.4 (2019), pp. 876–889.

[169] Matthew E Peters et al. "Knowledge enhanced contextual word representations". In: *arXiv preprint arXiv:1909.04164* (2019).

[170] Matthew E. Peters et al. "Deep Contextualized Word Representations". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 2227–2237. DOI: `10.18653/v1/N18-1202`. URL: `https://aclanthology.org/N18-1202`.

[171] Fabio Petroni et al. "Language models as knowledge bases?" In: *arXiv preprint arXiv:1909.01066* (2019).

[172] Flor Miriam Plaza-del-Arco et al. "Comparing pre-trained language models for Spanish hate speech detection". In: *Expert Systems with Applications* 166 (2021), p. 114120.

[173] Nina Poerner, Ulli Waltinger, and Hinrich Schütze. "E-BERT: Efficient-yet-effective entity embeddings for BERT". In: *arXiv preprint arXiv:1911.03681* (2019).

[174] Logan Praznik et al. "Analysis of Link Prediction Algorithms in Hashtag Graphs". In: *Big Data and Social Media Analytics*. Springer, 2021, pp. 221–245.

[175] Hemant Purohit et al. "Understanding user-community engagement by multi-faceted features: A case study on twitter". In: *WWW 2011 Workshop on Social Media Engagement (SoME)*. 2011.

[176] Le Qi et al. "DuReadervis: A: A Chinese Dataset for Open-domain Document Visual Question Answering". In: *Findings of the Association for Computational Linguistics: ACL 2022*. 2022, pp. 1338–1351.

[177] Xipeng Qiu et al. "Pre-trained models for natural language processing: A survey". In: *Science China Technological Sciences* 63.10 (2020), pp. 1872–1897.

[178] Muhammad Qorib, Seung-Hoon Na, and Hwee Tou Ng. "Frustratingly Easy System Combination for Grammatical Error Correction". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics, July 2022, pp. 1964–1974. URL: `https://aclanthology.org/2022.naacl-main.143`.

[179] Mohiuddin Md Abdul Qudar, Palak Bhatia, and Vijay Mago. "ONSET: Opinion and Aspect Extraction System from Unlabelled Data". In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2021, pp. 733–738.

[180] Mohiuddin Md Abdul Qudar and Vijay Mago. "Tweetbert: a pretrained language representation model for twitter text analysis". In: *arXiv preprint arXiv:2010.11091* (2020).

[181] Alec Radford et al. "Improving language understanding by generative pre-training". In: (2018).

[182] Alec Radford et al. "Language models are unsupervised multitask learners". In: *OpenAI blog* 1.8 (2019), p. 9.

[183] Gerasimos Razis and Ioannis Anagnostopoulos. "InfluenceTracker: Rating the impact of a Twitter account". In: *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer. 2014, pp. 184–195.

[184] Nils Reimers and Iryna Gurevych. "Sentence-bert: Sentence embeddings using siamese bert-networks". In: *arXiv preprint arXiv:1908.10084* (2019).

[185] Ana-Marıa Rıos, Bernardino Benito, and Francisco Bastida. "Factors explaining public participation in the central government budget process". In: *Australian Journal of Public Administration* 76.1 (2017), pp. 48–64.

[186] Michael Röder, Andreas Both, and Alexander Hinneburg. "Exploring the space of topic coherence measures". In: *Proceedings of the eighth ACM international conference on Web search and data mining*. 2015, pp. 399–408.

[187] Sascha Rothe et al. "A simple recipe for multilingual grammatical error correction". In: *arXiv preprint arXiv:2106.03830* (2021).

[188] Sohaib R Rufai and Catey Bunce. "World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis". In: *Journal of Public Health* 42.3 (2020), pp. 510–516.

[189] Uta Russmann and Jakob Svensson. "Interaction on Instagram?: Glimpses from the 2014 Swedish Elections". In: *International Journal of E-Politics (IJEP)* 8.1 (2017), pp. 50–66.

[190] Furqan Rustam et al. "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis". In: *Plos one* 16.2 (2021), e0245909.

[191] Piotr Rybak et al. "KLEJ: comprehensive benchmark for polish language understanding". In: *arXiv preprint arXiv:2005.00630* (2020).

[192] Abdulsamad Sahly, Chun Shao, and K Hazel Kwon. "Social media for political campaigns: An examination of Trump's and Clinton's frame building and its effect on audience engagement". In: *Social Media+ Society* 5.2 (2019), p. 2056305119855141.

[193] Julian Salazar et al. "Masked Language Model Scoring". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, pp. 2699–2712.

[194] Roberto Hernández Sampieri. *Metodologıa de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw Hill México, 2018.

[195] Mannila Sandhu et al. "From associations to sarcasm: Mining the shift of opinions regarding the supreme court on Twitter". In: *Online Social Networks and Media* 14 (2019), p. 100054.

[196] Rodrigo Sandoval-Almazan and David Valle-Cruz. "Facebook impact and sentiment analysis on political campaigns". In: *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. 2018, pp. 1–7.

[197] Yunita Sari and Mark Stevenson. "Exploring Word Embeddings and Character N-Grams for Author Clustering." In: *CLEF (Working Notes)*. Citeseer. 2016, pp. 984–991.

[198] Emre Şaşmaz and F Boray Tek. "Tweet Sentiment Analysis for Cryptocurrencies". In: *2021 6th International Conference on Computer Science and Engineering (UBMK)*. IEEE. 2021, pp. 613–618.

[199] Timo Schick and Hinrich Schütze. "It's not just size that matters: Small language models are also few-shot learners". In: *arXiv preprint arXiv:2009.07118* (2020).

[200] Timo Schick and Hinrich Schütze. "Rare words: A major problem for contextualized embeddings and how to fix it by attentive mimicking". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 05. 2020, pp. 8766–8774.

[201] Michael Schlichtkrull et al. "Modeling relational data with graph convolutional networks". In: *European semantic web conference*. Springer. 2018, pp. 593–607.

[202] Lutfi Kerem Senel and Hinrich Schütze. "Does he wink or does he nod? a challenging benchmark for evaluating word understanding of language models". In: *arXiv preprint arXiv:2102.03596* (2021).

[203] Neel Shah et al. "Assessing Canadians health activity and nutritional habits through social media". In: *Frontiers in public health* 7 (2020), p. 400.

[204] Md Shajalal and Masaki Aono. "Semantic textual similarity in bengali text". In: *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE. 2018, pp. 1–5.

[205] Tatiana Shavrina et al. "RussianSuperGLUE: A Russian language understanding evaluation benchmark". In: *arXiv preprint arXiv:2010.15925* (2020).

[206] Vinitha Siebers, Raymond Gradus, and René Grotens. "Citizen engagement and trust: A study among citizen panel members in three Dutch municipalities". In: *The Social Science Journal* 56.4 (2019), pp. 545–554.

[207] Vinay Singh et al. "Named entity recognition for hindi-english code-mixed social media text". In: *Proceedings of the seventh named entities workshop*. 2018, pp. 27–35.

[208] Richard Socher et al. "Recursive deep models for semantic compositionality over a sentiment treebank". In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. 2013, pp. 1631–1642.

[209] Irene Solaiman et al. "Release strategies and the social impacts of language models". In: *arXiv preprint arXiv:1908.09203* (2019).

[210] Jaebong Son et al. "Using a Heuristic-Systematic Model to assess the Twitter user profile's impact on disaster tweet credibility". In: *International Journal of Information Management* 54 (2020), p. 102176.

[211] Vamshi Krishna Srirangam et al. "Corpus creation and analysis for named entity recognition in Telugu-English code-mixed social media data". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 2019, pp. 183–189.

[212] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks". In: *Advances in neural information processing systems* 27 (2014).

[213] Ayesha Ayub Syed, Ford Lumban Gaol, and Tokuro Matsuo. "A survey of the state-of-the-art models in neural abstractive text summarization". In: *IEEE Access* 9 (2021), pp. 13248–13265.

[214] Oleksiy Syvokon and Olena Nahorna. "UA-GEC: Grammatical error correction and fluency corpus for the ukrainian language". In: *arXiv preprint arXiv:2103.16997* (2021).

[215] Jörg Tiedemann. "The Tatoeba Translation Challenge–Realistic Data Sets for Low Resource and Multilingual MT". In: *arXiv preprint arXiv:2010.06354* (2020).

[216] Meng-Hsiu Tsai et al. "A machine learning based strategy for election result prediction". In: *2019 international conference on computational science and computational intelligence (CSCI)*. IEEE. 2019, pp. 1408–1410.

[217] Barak Turovsky. *Found in translation: More accurate, fluent sentences in Google Translate*. Nov. 2016. URL: `https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/`.

[218] Elena Tutubalina et al. "Medical concept normalization in social media posts with recurrent neural networks". In: *Journal of biomedical informatics* 84 (2018), pp. 93–102.

[219] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[220] Petar Veličković et al. "Graph attention networks". In: *arXiv preprint arXiv:1710.10903* (2017).

[221] C Lee Ventola. "Social media and health care professionals: benefits, risks, and best practices". In: *Pharmacy and Therapeutics* 39.7 (2014), p. 491.

[222] Lucia Vesnic-Alujevic and Sofie Van Bauwel. “YouTube: A political advertising tool? A case study of the use of YouTube in the campaign for the European Parliament elections”. In: *Journal of Political Marketing* 13.3 (2014), pp. 195–212.

[223] Oriol Vinyals et al. “Grammar as a foreign language”. In: *Advances in neural information processing systems* 28 (2015).

[224] Oriol Vinyals et al. “Matching networks for one shot learning”. In: *Advances in neural information processing systems* 29 (2016).

[225] Alex Wang et al. “GLUE: A multi-task benchmark and analysis platform for natural language understanding”. In: *arXiv preprint arXiv:1804.07461* (2018).

[226] Alex Wang et al. “Superglue: A stickier benchmark for general-purpose language understanding systems”. In: *Advances in neural information processing systems* 32 (2019).

[227] Chong Wang, John Paisley, and David Blei. “Online variational inference for the hierarchical Dirichlet process”. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 752–760.

[228] Fei Wang et al. “Residual attention network for image classification”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 3156–3164.

[229] Lucy Lu Wang et al. “Cord-19: The covid-19 open research dataset”. In: *ArXiv* (2020).

[230] Ruize Wang et al. “K-adapter: Infusing knowledge into pre-trained models with adapters”. In: *arXiv preprint arXiv:2002.01808* (2020).

[231] William Yang Wang. “” liar, liar pants on fire”: A new benchmark dataset for fake news detection”. In: *arXiv preprint arXiv:1705.00648* (2017).

[232] Xiaozhi Wang et al. “KEPLER: A unified model for knowledge embedding and pretrained language representation”. In: *Transactions of the Association for Computational Linguistics* 9 (2021), pp. 176–194.

[233] Yanshan Wang et al. "MedSTS: a resource for clinical semantic textual similarity". In: *Language Resources and Evaluation* 54.1 (2020), pp. 57–72.

[234] Max Welling and Thomas N Kipf. "Semi-supervised classification with graph convolutional networks". In: *J. International Conference on Learning Representations (ICLR 2017)*. 2016.

[235] Chenxi Whitehouse et al. "Evaluation of Fake News Detection with Knowledge-Enhanced Language Models". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16. 2022, pp. 1425–1429.

[236] Michael Wiegand, Melanie Siegel, and Josef Ruppenhofer. "Overview of the germeval 2018 shared task on the identification of offensive language". In: (2018).

[237] Bryan Wilie et al. "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding". In: *arXiv preprint arXiv:2009.05387* (2020).

[238] Sam Witteveen and Martin Andrews. "Paraphrasing with large language models". In: *arXiv preprint arXiv:1911.09661* (2019).

[239] Arndt Wonka and Sebastian Haunss. "Cooperation in networks: Political parties and interest groups in EU policy-making in Germany". In: *European Union Politics* 21.1 (2020), pp. 130–151.

[240] Yonghui Wu et al. "Google's neural machine translation system: Bridging the gap between human and machine translation". In: *arXiv preprint arXiv:1609.08144* (2016).

[241] Wenhan Xiong et al. "TWEETQA: A social media focused question answering dataset". In: *arXiv preprint arXiv:1907.06292* (2019).

[242] Keyulu Xu et al. "How neural networks extrapolate: From feedforward to graph neural networks". In: *arXiv preprint arXiv:2009.11848* (2020).

[243] Liang Xu et al. "CLUE: A Chinese language understanding evaluation benchmark". In: *arXiv preprint arXiv:2004.05986* (2020).

[244] Liang Xu et al. "Fewclue: A chinese few-shot learning evaluation benchmark". In: *arXiv preprint arXiv:2107.07498* (2021).

[245] Lin Yue et al. “A survey of sentiment analysis in social media”. In: *Knowledge and Information Systems* 60.2 (2019), pp. 617–663.

[246] Marcos Zampieri et al. “SemEval-2020 task 12: Multilingual offensive language identification in social media (OffensEval 2020)”. In: *arXiv preprint arXiv:2006.07235* (2020).

[247] Li Zhang et al. “Sentiment analysis methods for hpv vaccines related tweets based on transfer learning”. In: *Healthcare*. Vol. 8. 3. Multidisciplinary Digital Publishing Institute. 2020, p. 307.

[248] Ningyu Zhang et al. “Conceptualized representation learning for chinese biomedical text mining”. In: *arXiv preprint arXiv:2008.10813* (2020).

[249] Zhengyan Zhang et al. “ERNIE: Enhanced language representation with informative entities”. In: *arXiv preprint arXiv:1905.07129* (2019).

[250] Renbo Zhao and Vincent YF Tan. “Online nonnegative matrix factorization with outliers”. In: *IEEE Transactions on Signal Processing* 65.3 (2016), pp. 555–570.

[251] Jie Zhou et al. “Sentix: A sentiment-aware pre-trained model for cross-domain sentiment analysis”. In: *Proceedings of the 28th international conference on computational linguistics*. 2020, pp. 568–579.

[252] Haris Bin Zia et al. “Improving Zero-Shot Cross-Lingual Hate Speech Detection with Pseudo-Label Fine-Tuning of Transformer Language Models”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16. 2022, pp. 1435–1439.

[253] Daniel M Ziegler et al. “Fine-tuning language models from human preferences”. In: *arXiv preprint arXiv:1909.08593* (2019).

Chapter 6

Appendix

6.1 Resources

Description	**Link**
A repository of the codebase used in Chapter 3	`https://github.com/manmeetkaurbaxi/2020-US-Elections`
A repository of the codebase used in Chapter 4	`https://github.com/manmeetkaurbaxi/Societal-Impact-on-Twitter`

6.2 Definitions and formulae

1. **TF-IDF**: The acronym TF-IDF, which stands for the term frequency-inverse document frequency, is a numerical statistic used in information retrieval that aims to capture the significance of a word to a corpus of documents. It is frequently applied as a weighting factor in information retrieval, text mining, and user modelling searches. TF-IDF is determined by multiplying two metrics: the number of times a word appears in a document and the word's inverse document frequency across a group of documents. In more formal mathematical terms, the TF-IDF score for the word t in the document d from the document set D is calculated as follows:

$$tf\_idf(t, d, D) = tf(t, d).idf(t, D) \tag{6.1}$$

where,

$$tf(t, d) = log(1 + freq(t, d)) \tag{6.2}$$

$$idf(t, D) = log\left(\frac{N}{count(d \in D : t \in d)}\right) \tag{6.3}$$

2. **LDA**: Latent Dirichlet Allocation (LDA) is a prominent statistical topic modelling technique. Documents in LDA are represented as a collection of topics, and a topic is a collection of words. In LDA, "latent" refers to the topics in the data that we wish to extract, and are concealed or not fully formed yet. Following the Dirichlet distribution model comes the Dirichlet allocation. The Dirichlet model depicts the pattern of words that regularly appear together, recur and are related to one another.

3. **Parallel LDA**: Parallel LDA refers to the Online LDA, utilizing all the CPU cores to parallelize and speed up the model training. Online LDA [94] is based on online stochastic optimization with a natural gradient step and can easily handle massive document collections.

4. **NMF**: Non-Negative Matrix Factorization (NMF) is an unsupervised statistical technique that decomposes the input corpora's high-dimensional vectors into a lower-

dimensional representation. Internally, it uses the factor analysis method to give comparatively less weightage to words with less coherence.

5. **LSI**: Latent semantic indexing (LSI) is a method of indexing and retrieval that uses a mathematical approach: singular value decomposition (SVD) to find patterns in the relationships between terms and concepts in an unstructured collection of text. The foundation of LSI is that words employed in related circumstances have a propensity to have similar meanings. LSI can extract the intellectual content of a text document by creating linkages between terms that appear in related situations. It was first used on the text at Bellcore[1] in the late 1980s and is known as "latent semantic indexing" because of its capacity to correlate latently present semantically linked terms in a text corpus. The technique, also known as latent semantic analysis (LSA), identifies the underlying latent semantic structure in the word usage in a body of text and demonstrates how it might be applied to deduce the text's meaning in response to user inquiries, also known as concept searches.

6. **HDP**: Hierarchical Dirichlet Process (HDP) is a non-parametric Bayesian model that may be used to model mixed-membership data with unlimited components. It has been extensively used in probabilistic topic modelling, where the components are distributions of words representing recurrent patterns (or themes) in the collection. The number of topics required and their distributions are determined using posterior inference and a document collection[227].

7. **Cosine Similarity**: Cosine similarity calculates the cosine of the angle between two vectors projected in a multi-dimensional space, and is used to determine the similarity between two documents regardless of their size. It is independent of the magnitudes of the vectors, and depends only on their angle. The cosine similarity is always in the range [-1,1]. Mathematically, the cosine similarity of two vectors is given as:

$$cosine\ similarity = S_C(A,B) := cos(\theta) = \frac{A.B}{||A||\ ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{6.4}$$

where, A_i and B_i are components of vectors A and B respectively.

[1] https://iconectiv.com/

8. **Hashing Vectorizer**: A vectorizer that employs the hashing method to determine the token string name to feature name index mapping is known as hashing vectorizer. This vectorizer converts text documents into matrices, creating a sparse matrix that contains the token occurrence counts for the collection of documents. It has very low memory scalability for substantial data sets since the vocabulary dictionary does not need to be stored in memory. It may be utilized in a parallel or streaming pipeline because there is not any state during the fit.

9. **SVM**: Support-vector machines (SVMs, also known as support-vector networks [53]) evaluate data for classification and regression analysis using supervised learning models and associated learning methods. An SVM training algorithm creates a model that categorizes fresh samples into one of two categories when given a series of training examples, making it a non-probabilistic binary linear classifier. SVM assigns training samples to spatial coordinates to maximize the distance between the two categories. Then, based on which side of the gap they fall, new samples are projected into that region and predicted to belong to a category. Using a technique known as the kernel trick, SVMs may effectively conduct non-linear classification in addition to linear classification by implicitly translating their inputs into high-dimensional feature spaces.

10. **Linear SVM**: When a dataset can be divided into two groups using just one straight line, it is said to be linearly separable, and a classifier known as a Linear SVM classifier is used to separate the data into these two categories.