Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Chapter 5
Applying Machine Learning to Online Data?
Beware! Computational Social Science Requires Care

Ulya Bayram
https://orcid.org/0000-0002-8150-4053
Çanakkale Onsekiz Mart University, Turkey

ABSTRACT

The immense impact of social media on contemporary cultural evolution is undeniable, consequently declaring them an essential data source for computational social science studies. Alongside the advancements in natural language processing and machine learning disciplines, computational social science researchers continuously adapt new techniques to the data collected from social media. Although these developments are imperative for studying the sociological transformations in many communities, there are some inconspicuous problems on the horizon. This chapter addresses issues that may arise from the use of social media data, like biased models. It also discusses various obstacles associated with machine learning methods while also providing possible solutions and recommendations to overcome these struggles from an interdisciplinary perspective. In the long term, this chapter will guide computational social science researchers in their future studies, from things to be aware of with data collection to assembling an accurate experimental design.

DOI: 10.4018/978-1-7998-8553-5.ch005

Copyright © 2022, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

INTRODUCTION

In the current digital age, online data availability is exponentially growing. Consequently, the interest in computational social science (CSS) research is rising, which spans many disciplines (social psychology, anthropology, economics, political science, sociology, and various levels of analysis) (Oboler, Welsh, & Cruz, 2012). Studies conducted on past and recent Congressional records of many countries including the United States (Bayram, Pestian, Santel, & Minai, 2019; Diermeier, Godbout, Yu, & Kaufmann, 2012; Gentzkow, Shapiro, & Taddy, 2016; Iyyer, Enns, Boyd-Graber, & Resnik, 2014; Jensen et al., 2012; Lauderdale & Herzog, 2016; Thomas, Pang, & Lee, 2006; Yu, Kaufmann, & Diermeier, 2008), political records from the British Parliament (Peterson & Spirling, 2018), and the Irish Dail debates (Lauderdale & Herzog, 2016), are among the plethora of studies that became possible thanks to the online availability of these collections. Similarly, old and new newspaper articles (Burley et al., 2020; Neresini, 2017), speeches of political figures including those from the past centuries (Jackson, Watts, List, Drabble, & Lindquist, 2021; Savoy, 2010), digital books (Brooke, Hammond, & Hirst, 2015), and many other textual online data have been the main focus of CSS research. While all these online data sources bring rich contributions to the broad range of CSS research areas, there is one domain that requires special attention: social media data.

Data collected from social media platforms facilitate many possibilities such as answering serious social science-related questions, finding insights into both individual-level and anthropological phenomena (Harford, 2014; Lazer et al., 2009; Olteanu, Castillo, Diaz, & Kiciman, 2019). There is also a growing consensus that data collected from these domains can provide more than simple observations (Olteanu et al., 2019); social media domains are amongst the most valuable data sources for CSS research areas. The wide-range public usage of social media domains, the absence of data use restrictions, the simplicity of data acquisition through application programming interfaces (APIs), and the valuable content of the data made them attractive for researchers. These social media and social network platforms (e.g. Twitter, Facebook, Reddit, Wikipedia, other forums) can easily capture the evolution of sociological norms and rapid changes culturally and globally. For example, recent movements such as “Me Too” and “Black Lives Matter” could not expand globally without these platforms. This fact makes these platforms one of the principal sources of information for CSS research on such events. Recent studies utilize social media data to evaluate the effects of crises like the mass killings in the United States (Burley et al., 2020).

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

The large availability of data drove CSS researchers towards using natural language processing (NLP), machine learning, and deep learning methods. They follow the NLP and machine learning literature to determine which techniques would be beneficial for their studies. Additionally, their openness to collaborate with professionals from different domains prompted the CSS studies to become more interdisciplinary. Studies they conduct utilizing NLP, machine learning, or deep learning techniques report new findings that were previously unattainable due to lack of proper automated tools. While all these developments on computational methods and online data availability are very advantageous for the future of CSS research, certain aspects of utilizing them demand more care and attention because there are significant problems inconspicuously haunting the studies. Data collected from social media may contain racist, sexist, or otherwise abusive language. These collections may also include samples with misinformation about the subject of interest and may create biased models. Also, machine learning methods are susceptible to predicaments like overfitting, underfitting, unsuitable method selection, and costly parameter tuning. Likewise, improper experiment designs could affect the generalizability of the findings, which would invalidate the reported results. Meanwhile, the black-box nature of machine learning and deep learning models causes a lack of interpretability regarding how or on which basis the results were obtained from these methods, making it difficult to justify the findings.

This chapter focuses on problems related to using online data and machine learning methods in CSS research from an interdisciplinary perspective. Each of the four main sections explains a range of problems related to a subject through examples to raise awareness towards constructing better studies. The following sections also propose recommendations and resolutions on how to overcome some of these obstacles. This chapter aims to inform a wide range of audiences that aspire to conduct CSS research utilizing online data and machine learning methods; while also intends to serve as a guide for future studies.

BACKGROUND

Since social media domains become a part of everyday life and the common subject of many CSS studies, researchers began to raise concerns. Among these concerns is how the use of social media in CSS should be constrained legally and ethically to protect society (Oboler et al., 2012). CSS studies on areas like social anthropology and political science can have significant

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

negative impacts, such as the subversion of democracy, profiling of individuals or groups for the development of better manipulation or even brainwashing techniques (Oboler et al., 2012). It is known that during the last few presidential elections in the United States, social media data was actively utilized to shape the course of elections (Bessi & Ferrara, 2016; Issenberg, 2013; Sen, Flock, &Wagner, 2020). In that perspective, the power of conducting CSS studies on online data is unequivocal. Therefore not committing any measurement errors is of utmost importance for a functioning democracy as it is vital for scholars (Sen et al., 2020). There is a great responsibility on the part of CSS researchers for making sure their models and results are correct.

In addition to the ethical concerns, there are several pitfalls related to using data acquired from online platforms such as social media (Olteanu et al., 2019). For instance, the existence of big data may not always be a blessing. In some cases, problems can be solved with a subsection of the big data, and using more than enough can make the estimations worse (Harford, 2014). Besides the data size, there is a notorious phenomenon called "bias," which is a word very generously uttered by studies without a clear definition (Blodgett, Barocas, Daume III, & Wallach, 2020). Algorithms are not known for being ethically neutral regardless of the circumstances (Tsamados et al., 2021). They are not magical; they are prone to learning noise or bad samples from the data. Also, there are inherent problems related to the algorithms/machine learning methods as well. All these issues can produce different types of biases. For example, various studies show that algorithmic advertisements regarding opportunities for higher-paying jobs in science and technology fields are advertised to men more often than those advertised to women (Datta, Tschantz, & Datta, 2015; Lambrecht & Tucker, 2019). It is only one class of studies demonstrating that machine learning methods can become biased towards gender. Another example shows that the prediction algorithms used in the United States health care systems enable white patients to receive measurably better care than black patients (Obermeyer, Powers, Vogeli, & Mullainathan, 2019). It is another demonstration of how serious the real-life consequences of a biased algorithm can be. The existence of bias in machine learning methods in CSS is dangerous and has to be eliminated. Olteanu et al. (2019) provide a detailed insight on different types of biases (population bias, behavioral bias, content bias, linking bias, temporal bias, and redundancy) with corresponding definitions and possible methods to overcome them.

Although studies acknowledge the existence and importance of various kinds of biases, only a few studies propose techniques for measuring or alleviating them (Blodgett et al., 2020). Nonetheless, using these techniques

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

may not be enough to guarantee a problem-free CSS research. The study of Jacobs and Wallach (2021) shows that even when the systems are designed to be unbiased, the fairness measures used can also be harmful due to the assumptions made during the design. They state that making assumptions can introduce potential mismatches between the theoretical understanding of the measured subject and the actual model constructed to perform the measurement. There are many other challenges regarding the social media data, such as labeling, feature extraction, feature elimination, class imbalance, among others. However, CSS research is not only threatened by data-related problems. Besides these issues, there are also other difficulties associated with using machine learning algorithms. Studies like L. Zhou, Pan, Wang, and Vasilakos (2017) and Domingos (2012) summarize some of these difficulties. There are challenges like overfitting, underfitting, scalability, incorrect feature selection implementations, selecting the wrong objective functions, among many others. Meantime, the popular deep learning methods also have specific drawbacks (Miotto, Wang, Wang, Jiang, & Dudley, 2018). In addition to the computational complexity, they can suffer from issues like a lack of interpretability, i.e. their black-box nature. This causes a requirement for a level of expertise in machine learning as not every method is suitable for every data type, size, or subject under analysis (L. Zhou et al., 2017). CSS researchers are from various backgrounds and the field is interdisciplinary, but not all researchers can be experts at machine learning.

As these background studies show, conducting CSS studies using online data and machine learning methods is a sensitive quest. Using algorithmic methods and interpreting the findings demand attention, knowledge and caution, all proportional to the subject-sensitivity. A computer scientist conducting CSS research without proper data sensitivity knowledge on the research subject may fail to be careful or fail to interpret the results accurately. Similarly, a political scientist conducting a CSS research study using machine learning might select methods based on their popularity rather than their suitability to the research subject, or might fail to build an optimal model and reach the wrong conclusions. This chapter is to guide these researchers in realizing some of these issues briefly and effectively. Those who seek more information regarding specific obstacles related to data or machine learning can refer to the Additional Reading section.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

WORKING WITH ONLINE DATA

The wide-range availability of online data has increased the interest in NLP research. Consequently, various CSS research areas became attractive for researchers from distinct fields. Especially the emergence of social media platforms where individuals can freely express their thoughts and opinions shaped the current state of CSS research interest. Researchers collect data from these platforms for purposes like studying the recent social movements like "Me Too" and "Black Lives Matter," for modeling the information flow concerning the evolution and the effects of conspiracy theories like "QAnon," or the aftereffects of some critical events like "US Capitol riots of 2021." They employ standard NLP techniques on these rich data collections and implement the fundamental pre-processing methods before training the algorithmic models. To label the data, they commonly look for pre-selected keywords related to the subject or the hashtags to avoid manual labeling costs (Grace, Halse, Aurite, Montarnal, & Tapia, 2019). Alternatively, some studies use crowd-sourcing to label the data (Chang, Amershi, & Kamar, 2017). Next, they train selected machine learning models with an assortment of these labeled data and conduct their analyses, classifications, or predictions based on their hypotheses and goals. However, all these operations have several associated costs. Among them, the data alone can vitiate a CSS study.

The challenge of using social media data begins with the data collection step. Properly collecting samples that include data from various demographics and groups uniformly and fairly is very important for CSS research (Olteanu et al., 2019). Also, data collected from social media domains contain many typos and errors, which may induce problematic results (Olteanu et al., 2019). Further, the existence of sarcasm or spam within the samples can go unrecognized by the data collection algorithms, and their presence would harm the machine learning results. The vast volumes of collected tweets, for example, will contain samples that are sarcastically mocking the movements or ideologies under analysis, as these mockers could use similar keywords or hashtags as the supporters. CSS researchers who are unaware of these possibilities will include measurements from these bad samples and report incorrect results. Also, there could be data in the collections intentionally containing misinformation to subvert or taint the movements or ideologies. Studies show how "fake news" emerges, especially during the elections (Allcott & Gentzkow, 2017; Wu, Morstatter, Carley, & Liu, 2019). Likewise, the existence of abusive language or curse words in data might result in incorrect measurements about the analyzed problem. Machine learning methods that

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

are trained to perform a CSS study, unless specifically fine-tuned to do otherwise, cannot detect the existence of misinformation or sarcasm. Thus, all these "bad samples" present within the data collection procured from the social media domains might bias the studies (Bender, Gebru, McMillan-Major, & Shmitchell, 2021).

As mentioned in the previous section, the word bias is commonly encountered in the literature. There are many different types of biases. The study of Olteanu et al. (2019) provides a detailed survey of possible biases with proper definitions and explanations. Because of the high number of bias types, how a model becomes biased is a complex problem to detect. However, creating a model containing bias is very easy. A data collection may appear to be balanced, but the samples might induce a biased algorithmic outcome. A study demonstrates this possibility through a real-life example: machine learning models proposed for use in the judicial system in the United States return racially inequitable decisions because the training data has latent patterns correlated with race (Rice, Rhodes, & Nteta, 2019). The examples in the previous section related to gender and healthcare are also critical examples regarding the biased models. These examples are the types of biases caused by the data. Meantime, there are many other possible ways to produce bias in models. Regardless of the type of bias present, none of the classification outcomes proclaimed by these biased models would be reliable (Bender et al., 2021). However, detecting and eliminating the possible sources of bias is not an easy task. While it is crucial to manually inspect a small subset of the data collection to grasp the contents, it is unattainable to manually investigate every sample in a dataset containing thousands or even millions of samples. A feasible approach for detection and elimination of bias is through analysis and data elimination.

Researchers have focused on possible ways to detect and reduce the likely roots of problems in data collections. These ways can be divided into two based on their complexity. Some of these methods used in text related tasks can be summarized as follows:

- Some simple methods
  - Removing typos
  - Checking language
  - Detecting semantics
  - Using ontologies
  - Plotting distributions
  - Using experts during labeling

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

- Some complex methods
  - Using power-law distributions
  - Following auto-text generation literature

Simple methods include tools available in NLP literature for detecting typos, language, and semantics within the data. Removing typos would eliminate many uninformative samples that could harm the models. Similarly, identifying which language do the texts contain is essential for social media datasets. Because domains like Twitter comprise posts in many languages and the tweets containing multiple languages can have the same hashtags or keywords used for labeling, and their existence may cause failure during the experiments. Meanwhile, checking the semantics can help eliminate texts containing indirectly expressed opinions such as sarcasm. Also, ontologies like Linguistic Inquiry and Word Count (Pennebaker, Francis, & Booth, 2001) can be used for removing the texts falling under undesired categories. Alternatively, the study of Olteanu et al. (2019) proposes comparing the samples in the data collection with reference samples from other sources. To handle these comparisons, techniques developed for textual data matching operations could be of help. Meantime, researchers can plot distributions over statistics such as demographics (if available) for creating balanced datasets. They also can reach out to experts to eliminate labeling-related biases.

Compared to the simple methods, there are alternative methods for detecting and eliminating data related problems. However, employing them would require adding another layer of a computational process to the CSS research. For example, to remove samples containing hateful comments, a researcher would need to implement one of the methods in the hatefulness-detection literature (Fortuna & Nunes, 2018). Similarly, studies like X. Zhou and Zafarani (2020) that automatically detect the existence of "fake news" and misinformation can be used to eliminate bad samples from the data collections. Yet, there are some challenges regarding bad sample detection that are waiting to be resolved in the future (Blodgett et al., 2020). Also, the addition of another layer of computation to detect and eliminate unwanted samples comes with the cost of increasing the computational complexity of research. Therefore, social scientists should consider the sensitivity of their research subject before deciding whether increasing the level of complexity is worth the effort.

Problems in a CSS research that may emerge due to online data are not limited to these. Another potential problem is the assumption that the texts are created by humans. There are many language models like Long short-term

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

memory (LSTM) (Sherstinsky, 2020), Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2019), generative pre-training (GPT) (Radford, Narasimhan, Salimans, & Sutskever, 2018), among others. They are analyzed and improved frequently to auto-generate better and more believable texts (Otter, Medina, & Kalita, 2020). In the past, evaluating whether the data follows a power-law distribution would be beneficial to determine whether the data was human-generated (Clauset, Shalizi, & Newman, 2009). However, the recent developments in NLP and machine learning disciplines are changing the level of difficulty. A model called GPT-3 shows strong capabilities in convincing readers as to whether the generated texts were actually from a human (Hao, 2020). These models are trained over big data and learn which words are more likely to follow each other through complex mathematical methods and then generate texts automatically on a given subject (Devlin et al., 2019). They are likely to affect the future of NLP research on online data. Hence, the CSS researchers will also have to deal with these computer-generated texts besides other problems. Researchers need to be aware of such developments in the literature.

WORKING WITH MACHINE LEARNING METHODS

Machine learning is part of the artificial intelligence field where methods learn patterns present within the given labeled or unlabeled dataset, based on the algorithmic structure of the selected model. These methods can roughly be divided into three based on the learning process they contain: supervised, unsupervised, and semi-supervised. Supervised methods require labeled training sets for learning the optimal ways of separating the data into classes based on the labels. Meanwhile, the unsupervised techniques do not require any labeled sets. Instead, they utilize different processes to find the optimal ways to separate the data without label-related knowledge. Alternatively, the semi-supervised learning methods use a mixture of a few labeled and many unlabeled data. These learning processes are a collaboration of representation, evaluation, and optimization (Domingos, 2012). A sub-field of machine learning research, known as deep learning, can also be explained through these general descriptions. However, deep learning also has its specific differences. These methods contain multiple layers of neural networks with complex mathematical processes. This section focuses on some problems the CSS researchers who are not experts in machine learning may encounter and need to be aware of, followed by related recommendations and solutions:

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

- The importance of feature elimination/selection
- Assumptions made by machine learning algorithms
- Selecting a method based on biased opinion
- Deep learning burdens
- Experience and knowledge required for properly tuning the methods
- High expectations from machine learning methods

Most machine learning methods require data to be transformed into a feature space through certain mathematical operations. In working with textual data, one of the choices is using n-grams. They are the word frequencies for unigrams (n=1), concurrent word pair frequencies for bigrams (n=2). These unigram and bigram features together showed success in the past (Bayram et al., 2019). Thanks to the developments in NLP literature, there are other options like the word embeddings (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). For a successful machine learning application, it is necessary to use features that can capture the patterns relevant to the data and to the main aim. The most appropriate solution for this is to search the literature to determine the most preferred/recommended for the studied subject. However, extracting the correct set of features is not enough to promise a successful machine learning application for a CSS study. Before using them for learning processes, it is also necessary to perform elimination after feature extraction (Domingos, 2012). The presence of a large number of uninformative features can introduce a burden over the learning process. It can also cause problems like learning noise and damaging the generalizability of the model. For avoiding these, it is possible to eliminate features using statistical methods such as the Kolmogorov-Smirnov test or Principal Component Analysis, where the feature variances are computed. However, creating a golden feature set cannot promise a problem-free machine learning application.

The use of machine learning methods possesses various possible inherent issues independent from the feature elimination. One of these issues is how every method makes certain assumptions about the data beyond the existing knowledge, which is the central goal (Domingos, 2012). For example, the naive Bayes classifier that uses the Bayes theorem makes an assumption of independence between the features given the class. On that note, whether this assumption holds for the actual data used to train this classifier is not considered. Similarly, linear classifiers always have the assumption that the data is linearly separable, which may not be true. Some data collections might have complex patterns that would thrive with non-linear models instead of linear ones. Nevertheless, without experimenting with both linear and non-linear models, it is nearly impossible to determine. Thus there is no

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

guarantee that any of these inherent assumptions of the selected algorithms will be accurate for every dataset and every classification problem. A way to handle any predicaments that may emerge from these assumptions is to experiment with different types of methods. Another concern regarding the use of machine learning is method selection. It is a known phenomenon that the method selection typically reflects the past experience of a researcher, which is justifiable since it is easier than learning and experimenting with a new model. However, this method selection bias can create more problems like using methods that are not suitable for the study. There could be methods more fitting for the type of data, size of the collection, or the subject of interest, and relying on the knowledge a researcher already possesses might result in incorrect results. Thus, researchers need to be deliberate and reasonable when making their choices to overcome this issue. Perhaps the most straightforward solution could be to conduct a literature search over possible methods used for the studied problem and stating the arguments for method selections before initiating the implementations.

The mentioned problems are also valid for the subfield of machine learning called deep learning. There is an undeniable bias towards preferring deep learning methods over more manageable methods because of their popularity. While deep learning methods have proven to be successful on various difficult data problems, thus rightfully deserving this popularity, a drawback is that they expect plenty of training for returning acceptable results (Otter et al., 2020), which may not be possible in some problem domains. Yet, they cannot guarantee success in every problem and data domain. They come with the burden of computational complexity. In other words, they require additional computational resources for providing adequate performance. However, not every CSS researcher can access expensive servers to run these complicated learning algorithms over large data collections. Researchers should consider using deep learning methods when it is necessary to bring the computational complexity to the studied subject. They should initially experiment with simple, well-tuned machine learning methods, and pursue deep learning methods only when the simple ones do not return satisfactory results. In fact, any study from any research filed that uses a deep learning method should also report results from the well-tuned simple machine learning methods. That is a good way of rationalizing the necessity for complexity.

Machine learning and deep learning methods require some level of knowledge from the users regarding parameter tuning and objective function selection. An objective function is a proxy to the true goal, which is to generalize the learned model over previously unseen data (Domingos, 2012). Methods can return completely conflicting results if different objective functions are

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

selected (Olteanu et al., 2019). Therefore, it is of utmost importance to perform the tunings and function selections cautiously. The needed knowledge is available if the researchers are willing to seek them. For example, to handle the parameter tuning properly, literature shows that sometimes finding a local optimum by a randomized search can be more efficient than finding a global optimum through a greedy search (Bergstra & Bengio, 2012). Thus, researchers need to refer to the literature when facing such choices during their implementations instead of willfully selecting parameters or objective functions. Finally, the most effective recommendation is to be aware that machine learning models are not magical; they cannot produce something out of anything (Domingos, 2012). These methods are limited by the information and patterns present in the data. Consequently, researchers should have realistic expectations when working with machine learning methods and should be prepared to handle failures.

EVALUATION AND GENERALIZATION

The fundamental goal of machine learning is to generalize the evaluations beyond the examples present in the labeled set (Domingos, 2012). Any CSS research that uses machine intelligence algorithms also has the same goal: to make general claims about the world based on what is observed from the data evaluations. However, obtaining results from one collection and accepting the observations as the golden truth about the studied social science phenomenon can induce various problems. Because the researcher could be using the incorrect metrics for the evaluations, and the observations could be proportionally problematic. Similarly, the dataset could be problematical, thus causing faulty generalizations. A few of such problems are explained in this section:

- Selecting the correct evaluation metrics
- Relying on the generalization results

Following the implementation of a machine learning method, the next step is to evaluate the results, which is not easy. Selecting the correct evaluation metric might affect a research study takeaways (Olteanu et al., 2019). There are many metrics to evaluate the success of a machine learning model, including accuracy, precision, recall, the area under the ROC curve, among others. Before selecting metrics to evaluate the results of a CSS study, researchers need to consider a few factors. One of these factors is the class imbalance

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

in the data collection. An example of imbalanced data collection could be forum posts containing many positive views about a political party, where the research goal is to detect the few negative opinions. There are methods to handle training set class imbalance problems at the data level (e.g. reducing the number of positive view posts from the collection), algorithm level (e.g. using machine learning methods that can handle learning imbalanced data), or hybrids that combine the both (Krawczyk, 2016). When the test set has class imbalance, the evaluation metric selection becomes important. Meanwhile, the existence of class imbalance on test sets also requires attention. For example, the accuracy metric is not appropriate for imbalanced test sets since its equation (the number of correct detections divided by all) is not fit well to represent the classes with small number of samples compared to the other classes. In such cases, a researcher can use balanced accuracy metrics (Brodersen et al., 2010). In addition to class imbalance, the meaning of the evaluation metric alone can be important depending on the subject of interest. For instance, the recall metric might be more suitable when detecting the positive class data in health-care related domains.

Another error a CSS researcher may face towards the end of a study is to have too much faith in the results and to make global generalizations about the world based on the observed outcomes. However, machine intelligence results are only as good as the data and the algorithm. Therefore, the trained models that return the evaluation scores or the analyzed outcomes should also be studied to ensure whether the results are interpretable. The study of Ribeiro, Singh, and Guestrin (2016) shows an example where a deep learning model that returns high classification scores actually learned to detect husky images based on the existence of snow in the background instead of learning any features related to specifying the shape or properties of the animal. This problem would not be discovered if the researchers did not analyze the learned features. In that case the model would be applied to real-world classification tasks and could return incorrect results. Another study demonstrates that pre-trained deep learning methods that return top scores on ImageNet competition actually have been overfitting (Recht et al., 2019). They demonstrate that when slightly harder generalization experiments are introduced using the same or similar data, the reported high scores drop by 3-15%. Problems like overfitting can cause observing high results, which is due to memorizing the training data to the point of eventually damaging the ability to generalize to other data (Domingos, 2012). It is fundamental not to rely on within-dataset scores alone. Studies like Sen et al. (2020) are good examples regarding the importance of performing generalization experiments. The generalization

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

evaluation is also known as measuring the external validity of the outcomes found through within-dataset evaluations (Olteanu et al., 2019). One of the goals of the NLP studies is constructing and training methods that can be re-used (Sen et al., 2020). When the CSS research areas are considered, the value of this generalizability principle becomes more significant because the found outcomes should hold for other data collections obtained from the same or different sources. Otherwise, the complete foundation of the CSS research would fall apart, and the validity of claims and the credibility of researchers would become dubious. Researchers should not prefer some methods over others based on their reported results in the literature as the study of Recht et al. (2019) demonstrates how the generalization experiments are useful to detect the faulty models.

In addition to the generalization experiments, comparing the evaluation and generalization results of different methods is necessary for the credibility of the findings. For example, some experiments show simple methods can over-perform the complex ones on various data collections (Bayram et al., 2019; Bayram & Benhiba, 2021). In a similar case where data favors simple methods over the complex ones, if the researchers only experiment with the popular complex methods and report their results, they might come to wrong conclusions about the subject of interest due to low performance they observe. Thus including a range of methods with different complexities is also essential for a machine learning study. However, it is important to note that the method comparisons should be conducted fairly. For example, a researcher who favors a method over the others might tune the favored method well, while not conducting any necessary parameter search or fine tuning on the other compared methods. Such an experiment would return results favoring the favored method, but the reliability of the study and its claims would be questionable. Conducting a good CSS research requires a lot of caution and attention so the reported evaluation and generalization results would be correct, trustworthy, and can aid the future research studies.

INTERPRETING THE RESULTS

The vitality of interpreting the evaluation results in CSS earns it a separate section in this chapter. Correctly interpreting the results is at the core of social science. Therefore it is essential for a researcher to be aware of some potential issues when interpreting the findings obtained from the trained models. Some of the elements to consider when interpreting the results of a CSS study:

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

- Not mixing correlations with causations
- The black-box property of machine learning methods
- Ethics/need for an expert opinion

One of the most common pitfalls in research is mixing the correlation with causation (Domingos, 2012). Especially in CSS studies, it is easy to make this mistake when interpreting the findings. Machine learning methods are powerful. Especially the power of deep learning technology enables researchers to automatically detect complex patterns present within the large data collections. However, the expectation should be limited to finding the patterns present within the data, and not to lean towards causative interpretations. Because, in principle, machine learning only learns patterns algorithmically found to be correlated to the goal, but it does not or cannot learn all the variables that exist in the real-world (Leetaru, 2019). For example, a hypothetical study could be detecting a positive correlation between the male gender and a high number of research publications after analyzing the author genders in a journal. Thus, interpreting this correlation with the causation of "male authors are better at publishing than the authors of other genders," or even as "male academics are more successful than the others" would be profoundly problematic and false. There could be other factors behind the correlation results such as the training data or the used model having a gender bias. In another example, a CSS researcher conducting a machine-learning study to detect anti-vaxxers from online posts might find a correlation between anti-vaxxers and where they live, e.g. majority of the anti-vaxxers are found to be living in the coastal areas. Therefore, when interpreting the results, the researcher should refrain from making remarks such as "living in the coastal areas causes individuals to oppose to vaccines." Instead, the CSS researcher should look deeper into the data, which might show that the majority of the posts are collected from coastal area residents, or there could be other latent problems in the data.

Another matter to be aware of when interpreting the machine learning results is the black-box property of the models. Some machine learning libraries provide tools to analyze the trained models by ranking the features based on their effectiveness during the training process. Simple, linear classifiers commonly provide such techniques to detect the top features. However, in complex machine learning and deep learning methods, interpreting the decision-making mechanism behind a result is no longer straightforward. Especially those researchers that lack proper education or training on machine learning theory might find it arduous to understand the underlying structure that returns the obtained results. Thanks to the growing research interest in

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

deep learning, new methods are being developed to overcome the back-box property of these complex methods (Montavon, Samek, & Muller, 2018). A study conducted by Google researchers attempts to go through the deep learning process backwards and returns visually appealing images they call "how machines dream" (Mordvintsev, Olah, & Tyka, 2015). While the researchers unveil the structures of these trained complex models, they are still difficult to interpret for those who lack theoretical training in the field. Researchers should consider the importance of interpreting the results by taking the underlying structure of their models into account because the models are responsible for the obtained results. Otherwise, their interpretations might lack proper reasoning and may mislead other researchers.

Perhaps one of the most important issues related to the use of machine learning methods in CSS problems is the possibility of obtaining biased models, where bias in this sentence refers to ethical issues. In the previous sections, how the use of machine learning models in the justice system could cause a bias towards race was mentioned. A study by Angwin et al. (2016) expands the racial bias in judicial system algorithms in great detail with real-life examples. They find that the scores returned by the system were remarkably unreliable in forecasting future violent crime. Additionally, the formula was found to be falsely flagging black defendants as future criminals very frequently, almost twice the rate as the white defendants. Meantime, white defendants were mislabeled with low future crime risk scores more often than the black defendants. Two individuals with different race who commit the same crime receive different future crime prediction scores, which is highly unfair. Another study by Buolamwini and Gebru (2018) studies algorithms that classify gender from the face images and report that three commercial gender classification systems perform significantly worse on the images of individuals with darker skin tones. These studies show that the machine learning algorithms used in CSS problems are far from being perfect for many reasons. While there is a research focus on finding ways to improve the fairness of these algorithms, CSS researchers should be aware of the possibility of obtaining unfair results. The ethical side of research is important. Any interpretation of the results obtained from an unfair model could be harmful in many aspects. A solution could be to consult social scientists (Olteanu et al., 2019). A social scientist could perform an independent statistical analysis over the data and the machine learning results to validate their fairness, could provide different interpretations, and could eliminate the ethical concerns. Also, consulting the experts would be useful when the studies are conducted by interdisciplinary researchers, because the lack of social science knowledge may cause incorrect interpretations of the findings as well.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

FUTURE RESEARCH DIRECTIONS

The growth of social media evokes more attention to CSS research, thus causing the field to become progressively more interdisciplinary. Thanks to all this attention CSS receives, the latest developments in NLP and machine learning research domains are getting executed by the interdisciplinary researchers over various CSS problems. They inform the public with their new and significant findings, causing conspicuous changes in everyday life through studies covering critical domains such as political science and health care. With more CSS studies conducted, the society becomes more conscious of the ongoing problems reported by these studies, thus hopefully enforcing a change towards the positive in everyday life. For example, those who read the previously mentioned CSS studies reporting how machine learning is actively operated on social media during the elections might become more aware of the content they see, thus might become more resistant to potential manipulations. Meanwhile, studies like Bender et al. (2021) inform the researchers and the public about problems like race or gender bias in machine learning. Thanks to the increasing awareness, the future may behold better tools to detect and eliminate these biases.

There are also negative consequences to the advancements in NLP and machine learning disciplines regarding the future of CSS research. While these advancements are making it possible to develop better automated language models (e.g. GTP-3), these language models can be used for malicious purposes. For example, the language auto-generator models can be used on social media to change public opinion regarding politics, alter election outcomes, easily manipulate large crowds, and ultimately cause adverse events in the future like the US Capitol riots of 2021, which was also due to the spread of manipulative information. These generative models can even be utilized to further ignite the ongoing riots through trolling activities and generate more chaos. Thus these language models can become an immense threat to the future of society. Moreover, texts created by these models can escape detection tools due to their complex nature. While these models are very recent and are yet to become widely available, CSS researchers need to be aware of this potential threat.

With the wide-range availability of machine learning courses, CSS researchers who do not have a deep computational background are learning more information than the basics of the models. Also, the developers are working on constructing open-source tools to help researchers efficiently access and use the new techniques that are better at handling the problems explained in the previous sections. Meanwhile, studies such as this chapter

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

and those present in the Additional Reading section are helpful in further apprising the researchers regarding the value of awareness. Overall, in the light of these developments and problems, the future of CSS research has a great potential to unveil more unknowns regarding society as a whole. The central core of future CSS research should be to endure all these obstacles and persevere through them. The significance of the CSS research problems makes them indispensable.

CONCLUSION

The impact of social media on contemporary cultural evolution is undeniable. This impact led many computational social scientists to dedicate their efforts to capturing these evolutions over data collected from social media using natural language processing and machine learning methods. However, in these efforts, they struggle with various problems. Among these problems, there are many issues related to working with social media data. Data from online sources are prone to containing typos, possessing words from different languages, comprising texts with irony and sarcasm, and may include texts generated by machines. There are also problems inherent to the machine learning methods. These methods can learn bias from data or the extracted features, while researchers can select incorrect methods or parameters without proper knowledge. Also, researchers can employ complex learning methods in their computational social science studies like the deep learning methods solely based on their popularity in the recent literature. Choosing these complex methods over the simpler alternatives without a justifiable decision-making process may increase the computational burden immensely while failing to return correct results. Furthermore, researchers can select incorrect evaluation metrics, which may disturb the accuracy of the reported results. The computational social science researchers may also neglect to perform generalization experiments on different collections and might be reporting results from methods that are suffering from overfitting. This chapter addresses a range of these issues. Under each section, it proposes recommendations on how to avoid, detect, and handle these issues. Yet, the problems computational social science researchers might encounter are uncountable. When starting a new study, the researchers are taking a great responsibility in terms of awareness towards these issues and readiness for handling them. The future computational social science research will become a lot different than the current version. Methods will become more advanced, while newer and more troubling problems emerge. Until then, this chapter

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

will guide the computational social science researchers in constructing better studies and it will prepare them for the inevitable future difficulties.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *The Journal of Economic Perspectives*, *31*(2), 211–236. doi:10.1257/jep.31.2.211

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. Propublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bayram, U., & Benhiba, L. (2021). Determining a Person's Suicide Risk by Voting on the Short-Term History of Tweets for the CLPsych 2021 Shared Task. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access* (pp. 81-86). 10.18653/v1/2021.clpsych-1.8

Bayram, U., Pestian, J., Santel, D., & Minai, A. A. (2019). What's in a Word? Detecting Partisan Affiliation From Word Use In Congressional Speeches. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). 10.1109/IJCNN.2019.8851739

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). 10.1145/3442188.3445922

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*(2).

Bessi, A., & Ferrara, E. (2016). Social Bots Distort the 2016 US Presidential Election Online Discussion. *First Monday*, *21*, 11–17.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Blodgett, S. L., Barocas, S., Daume, H. III, & Wallach, H. (2020). Language (Technology) Is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). 10.18653/v1/2020.acl-main.485

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. In *Proceedings of the IEEE 20th International Conference on Pattern Recognition* (pp. 3121-3124). 10.1109/ICPR.2010.764

Brooke, J., Hammond, A., & Hirst, G. (2015). Gutentag: An NLP-Driven Tool for Digital Humanities Research in the Project Gutenberg Corpus. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature* (pp. 42-47). 10.3115/v1/W15-0705

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 2018 ACM Conference on Fairness, Accountability and Transparency* (pp. 77-91). ACM.

Burley, T., Humble, L., Sleeper, C., Sticha, A., Chesler, A., Regan, P., Verdeja, E., & Brenner, P. (2020). NLP Work Flows for Computational Social Science: Understanding Triggers of State-Led Mass Killings. In Practice and Experience in Advanced Research Computing (pp. 152-159). Academic Press.

Chang, J. C., Amershi, S., & Kamar, E. (2017). Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2334-2346). 10.1145/3025453.3026044

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, *51*(4), 661–703. doi:10.1137/070710111

Datta, A., Tschantz, M. C., & Datta, A. (2015). *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies*. doi:10.1515/popets-2015-0007

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). 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* (vol. 1, pp. 4171-4186). Academic Press.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Diermeier, D., Godbout, J.-F., Yu, B., & Kaufmann, S. (2012). Language and Ideology in Congress. *British Journal of Political Science*, *42*(1), 31–55. doi:10.1017/S0007123411000160

Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. *Communications of the ACM*, *55*(10), 78–87. doi:10.1145/2347736.2347755

Fortuna, P., & Nunes, S. (2018). A Survey on Automatic Detection of Hate Speech in Text. *ACM Computing Surveys*, *51*(4), 1–30. doi:10.1145/3232676

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2016). *Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech*. National Bureau of Economic Research.

Grace, R., Halse, S., Aurite, W., Montarnal, A., & Tapia, A. (2019). Expanding Awareness: Comparing Location, Keyword, and Network Filtering Methods to Collect Hyperlocal Social Media Data. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 2699-2708). 10.24251/HICSS.2019.325

Hao, K. (2020). *A College Kid's Fake, AI-Generated Blog Fooled Tens of Thousands. This Is How He Made It*. https://www.technologyreview.com/2020/08/14/1006780/ai-gpt-3-fake-blog-reached-top-of-hacker-news/

Harford, T. (2014). Big Data: A Big Mistake? *Significance*, *11*(5), 14–19. doi:10.1111/j.1740-9713.2014.00778.x

Issenberg, S. (2013). How President Obama's Campaign Used Big Data to Rally Individual Voters. *Technology Review*, *116*(1), 38–49.

Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P. (2014). Political Ideology Detection Using Recursive Neural Networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (vol. 1, pp. 1113-1122). 10.3115/v1/P14-1105

Jackson, J., Watts, J., List, J.-M., Drabble, R., & Lindquist, K. (2021). From Text to Thought: How Analyzing Language Can Advance Psychological Science. *Perspectives on Psychological Science*. Advance online publication. doi:10.1177/17456916211004899 PMID:34606730

Jacobs, A. Z., & Wallach, H. (2021). Measurement and Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 375-385). 10.1145/3442188.3445901

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Jensen, J., Naidu, S., Kaplan, E., Wilse-Samson, L., Gergen, D., Zuckerman, M., & Spirling, A. (2012). Political Polarization and the Dynamics of Political Language: Evidence From 130 Years of Partisan Speech. *Brookings Papers on Economic Activity*, *2012*(1), 1–81. doi:10.1353/eca.2012.0017

Krawczyk, B. (2016). Learning From Imbalanced Data: Open Challenges and Future Directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. doi:10.100713748-016-0094-0

Lambrecht, A., & Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*, *65*(7), 2966–2981. doi:10.1287/mnsc.2018.3093

Lauderdale, B. E., & Herzog, A. (2016). Measuring Political Positions From Legislative Speech. *Political Analysis*, *24*(3), 374–394. doi:10.1093/pan/mpw017

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. doi:10.1126cience.1167742 PMID:19197046

Leetaru, K. (2019). A Reminder That Machine Learning Is About Correlations Not Causation. *Forbes*. https://www.forbes.com/sites/kalevleetaru/2019/01/15/a-reminder-that-machine-learning-is-about-correlations-not-causation/?sh=2f44af576161

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. In Advances in Neural Information Processing Systems (pp. 3111-3119). Academic Press.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep Learning for Healthcare: Review, Opportunities and Challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246. doi:10.1093/bib/bbx044 PMID:28481991

Montavon, G., Samek, W., & Muller, K.-R. (2018). Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing*, *73*, 1–15. doi:10.1016/j.dsp.2017.10.011

Mordvintsev, A., Olah, C., & Tyka, M. (2015). *Inceptionism: Going Deeper Into Neural Networks*. https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Ulya Bayram (ulya.bayram@comu.edu.tr) Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Neresini, F. (2017). Old media and new opportunities for a computational social science on PCST. *Journal of Science Communication*, *16*(2), C03. doi:10.22323/2.16020303

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science*, *366*(6464), 447–453. doi:10.1126cience.aax2342 PMID:31649194

Oboler, A., Welsh, K., & Cruz, L. (2012). The Danger of Big Data: Social Media as Computational Social Science. *First Monday*, *17*(7).

Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*, 13. doi:10.3389/fdata.2019.00013 PMID:33693336

Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(2), 604–624. doi:10.1109/TNNLS.2020.2979670 PMID:32324570

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic Inquiry and Word Count: LIWC 2001. Lawrence Erlbaum Associates.

Peterson, A., & Spirling, A. (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring polarization in Westminster systems. *Political Analysis*, *26*(1), 120–128. doi:10.1017/pan.2017.39

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training. Technical Report*. OpenAI.

Recht, B., Roelofs, R., Schmidt, L., & Shankar, V. (2019). Do ImageNet Classifiers Generalize to ImageNet? In *International Conference on Machine Learning* (pp. 5389-5400). Academic Press.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining The Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). 10.1145/2939672.2939778

Rice, D., Rhodes, J. H., & Nteta, T. (2019). Racial Bias in Legal Language. *Research & Politics*, *6*(2), 2053168019848930. doi:10.1177/2053168019848930

Savoy, J. (2010). Lexical Analysis of Us Political Speeches. *Journal of Quantitative Linguistics*, *17*(2), 123–141. doi:10.1080/09296171003643205

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Sen, I., Flock, F., & Wagner, C. (2020). On the Reliability and Validity of Detecting Approval of Political Actors in Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1413-1426). 10.18653/v1/2020.emnlp-main.110

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D. Nonlinear Phenomena*, *404*, 132306. doi:10.1016/j.physd.2019.132306

Thomas, M., Pang, B., & Lee, L. (2006). Get Out the Vote: Determining Support or Opposition From Congressional Floor-Debate Transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 327-335). 10.3115/1610075.1610122

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The Ethics of Algorithms: Key Problems and Solutions. *AI & Society*, 1–16.

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in Social Media: Definition, Manipulation, and Detection. *SIGKDD Explorations*, *21*(2), 80–90. doi:10.1145/3373464.3373475

Yu, B., Kaufmann, S., & Diermeier, D. (2008). Classifying Party Affiliation From Political Speech. *Journal of Information Technology & Politics*, *5*(1), 33–48. doi:10.1080/19331680802149608

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine Learning on Big Data: Opportunities and Challenges. *Neurocomputing*, *237*, 350–361. doi:10.1016/j.neucom.2017.01.026

Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, *53*(5), 1–40. doi:10.1145/3395046

ADDITIONAL READING

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91).

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Dressel, J., & Farid, H. (2018). The Accuracy, Fairness, and Limits of Predicting Recidivism. *Science Advances*, *4*(1), eaao5580. doi:10.1126ciadv.aao5580 PMID:29376122

Molina, M., & Garip, F. (2019). Machine Learning for Sociology. *Annual Review of Sociology*, *45*(1), 27–45. doi:10.1146/annurev-soc-073117-041106

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2020). Deep Double Descent: *Where Bigger Models and More Data Hurt. In International Conference on Learning Representations (ICLR).*

Stilgoe, J. (2018). Machine Learning, Social Learning and the Governance of Self-Driving Cars. *Social Studies of Science*, *48*(1), 25–56. doi:10.1177/0306312717741687 PMID:29160165

KEY TERMS AND DEFINITIONS

Bias: Prejudice towards or against a person, a group, or a class. In a machine learning context, there are various types of biases. Each bias can affect a model differently.

Deep Learning: A subfield of machine learning that works with artificial neural networks containing many hidden layers and complex structures.

Interdisciplinary Research: It is a type of research that employs knowledge, data, techniques, and theories from multiple disciplines. The main goal is to analyze or solve a specific problem of interest using the strengths of these different disciplines.

Language Generation: It is the process of automatically generating texts from models containing artificial intelligence. The goal of this process is to auto-generate texts that appear to be created by humans.

Machine Learning: A subfield of artificial intelligence where models can learn patterns from the data in a supervised or unsupervised fashion and tune themselves during the learning process.

Noise: In the context of machine learning, noise corresponds to the type of data or features that do not contain meaningful patterns related to the problem of interest and have a possibility of disrupting and harming the learning process.

Overfitting: Corresponds to the event when a machine learning model memorizes the training set data instead of learning the patterns present within it for accurate generalization. When a model overfits data, the model can return high within-dataset performance while it fails to generalize to other data.

Ulya Bayram (ulya.bayram@comu.edu.tr)
Downloaded: 2/9/2022 3:17:39 AM
IP Address: 176.217.109.201

Tuning: The set of operations within machine learning algorithms to improve the performance of classification and prediction. Some of the tuning procedures happen during the learning process, while it is also possible to tune a model after the initial run of the training process.

Underfitting: Corresponds to the event when a machine learning model does not learn the patterns present within the training set properly for reasons such as incorrect parameter selection or a small number of epochs for the case of neural networks. A machine learning that suffers from underfitting would fail to return acceptable results from the within-dataset experiments and the generalization experiments.