Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.023914
Article

Tech Science Press

Political Ideology Detection of News Articles Using Deep Neural Networks

Khudran M. Alzhrani*

Department of Information Systems, Al-Qunfudhah Computing College, Umm Al-Qura University, Al-Qunfudhah, Kingdom of Saudi Arabia
*Corresponding Author: Khudran M. Alzhrani. Email: kmzhrani@uqu.edu.sa
Received: 26 August 2021; Accepted: 15 November 2021

Abstract: Individuals inadvertently allow emotions to drive their rational thoughts to predetermined conclusions regarding political partiality issues. Being well-informed about the subject in question mitigates emotions' influence on humans' cognitive reasoning, but it does not eliminate bias. By nature, humans tend to pick a side based on their beliefs, personal interests, and principles. Hence, journalists' political leaning is defining factor in the rise of the polarity of political news coverage. Political bias studies usually align subjects or controversial topics of the news coverage to a particular ideology. However, politicians as private citizens or public officials are also consistently in the media spotlight throughout their careers. Detecting political polarity in the news coverage of politicians rather than topics adds a new perspective. Determining the best approach for detecting political polarity in the news relies on the news delivery method. Data types such as videos, audio, or text could summarize the news delivery methods. Text is one of the most prominent news delivery methods. Text pattern recognition and text classification are well-established research areas with applications in many multidisciplinary domains. We propose to use deep neural networks to detect ideology in news media articles that cover news related to political officials, namely, President Obama and Trump. Deep network models were able to identify the political ideology of articles with over 0.9 F1-Score. An evaluation and analysis of deep neural network performance in detecting political ideology of news articles, articles' authors, and news sources are presented in the paper. Furthermore, this paper experiments on and provides a detailed analysis of newly reconstructed datasets.

Keywords: Deep neural networks; data analysis; text classification; natural language processing; ideology detection; political science; media bias; political ideology

1 Introduction

Fortunately, facts are still absolute, but fact interpretations are subjective. As humans, we yield to confirmation bias [1], where we present evidence selected obliviously to support a claim or an argument. We recall, seek, or interpret information that agrees with our personal beliefs. We use the words to describe an event or argue an issue carefully picked to fit a narrative. In this kind of situation,

fact-checking news articles is not a suitable solution. Journalists, like any other human, are subject to confirmation bias. The inability of a journalist to separate his\her political ideology from logically analyzing events and reporting unfiltered news is a critical flaw in journalism. One would like to know to which degree confirmation bias affects journalists' work. Would journalists' political beliefs influence their way of reporting or writing about an event? Commonly, news media sources describe the same event in opposing views. Exposure to multiple points of view in the news media is beneficial for attentive readers. In an ideal scenario, an alert reader would go through various news articles and then follow the most persuasive argument, which is not often the case. Hence, the questions we are attempting to answer are: can deep neural networks identify news articles political ideology and how accurate are the models? And can the models predict authors and news sources political ideology based on articles contents?

Currently, information is being circulated to more information outlets than ever before, yet in 2017, only 16% of Americans have high confidence in news media [2]. At first glance, the increasing number of news sources is advantageous since it enables the viewers to be exposed to various points of view. However, several studies [3,4] asserted that people are selectively exposed to media sources aligned with their political beliefs. Other studies point out that social networks foster people's selective exposure patterns by befriending or following like-minded people, which heavily influences the information types shared among them [5]. Limited exposure to diverse media outlets can shape people's conception of issues, which will extend the impact of politically aligned news from opinions to actions such as election participation and voting [6,7].

New internet technologies enabled traditional news media to reinforce its presence and reach more audiences. However, the same technologies contributed to creating new political news sources that fit an intended audience rather than aiming for neutrality [8]. Inattentive readers might not know the news sources' ideology found in search engines or any platform other than the original publisher. In social network platforms, labeling debatable user content with explanatory or informative content to avoid misconceptions is rising. Applying the same concept by labeling news across platforms based on their political ideology will improve readers' attention to these issues. One way to find the political ideology of the news media source is through self-proclaimed neutral third parties that undertake scientific methods to categorize media sources. Other news media sources provide information regarding their political ideology on their websites or social media accounts.

Political news attracts more readers during significant events such as presidential elections, though the newcomers will find it challenging to read all published articles thoroughly. Hence, it is logical to assume that the attention given to each political news article during such events is limited [9]. It takes attentiveness and interest in political matters to find detailed information regarding politics or certain politicians. The Presidential dataset consists of articles written about President Obama and Trump, extracted from news sources with explicit political ideologies. Unlike celebrities' news, politicians' news coverage could easily be politicized. The study of politicians' media coverage in political science is dispersed and inconclusive. Political ideology is often associated with controversial topics, sources ideology, and authors' political leanings, more details in Section 2. Due to the nature of political news coverage, we believe that ideology detection has latent attributes such as news framing, agenda settings, and political stances.

Text is one medium for rapid information sharing on the internet through social media, blogs, or news websites. Since texts are unstructured, distinctions between different categories are not easily recognizable by humans or require a knowledgeable person. Machine learning and deep neural networks can automatically categorize sensitive texts [10], medical records [11], financial news [12], and political sentiments [13]. We experiment with the most prominent deep neural networks in the text classification literature to detect news ideology. We collected the data from the self-identified far-left and far-right news sources that specifically targeted politicians in their articles. The data is left unbalanced to simulate a real-world scenario where

media outlets do not balance their news coverage of a person, ideology, or topic. The results have proven the existence of political ideology in the news. The contributions of this paper are summarized as follows:

- We experimented and analyzed newly reconstrued datasets consisting of articles, authors, sources, and political ideologies.
- We implemented four well-received deep neural networks in text classification literature to detect articles' ideologies.
- We measured the political ideology detection models' performance on articles, news outlets, and authors.

This paper's remaining part is structured as follows: Section 2 briefly reviews news ideology detection literature. We elaborate on technical aspects of the data construction procedures in Section 3 and provide a statistical description and analytical insights on the data itself. Details on the deep neural networks implemented in this paper are in Section 4, with illustrative figures to better understand neural network architectures. Section 5 articulates the experimental setup and analyzes the deep neural network models' detection results. Finally, the paper is concluded in Section 6.

2 Related Work

In general, political ideology is highly associated with political parties. Take as an example; in the U.S., the two dominating parties are republican and democratic. A republican voter most likely has conservative or right-wing views on the economy, education, health systems, and other policies, whereas a Democratic voter has liberal or left-wing views. Conservative or liberal voters do not have identical views on all policies; therefore, some could be positioned on the center-right or left of the political spectrum. On that account, researchers consider one of two methods to label political ideologies of a textual news article, namely, the top-to-bottom method or bottom-to-top method. In the top-to-bottom approach [14–19], textual articles are labeled with the ideology corresponding to its author's partisanship or explicit political ideology. As an example of top-to-bottom approach Preoțiuc-Pietro et al. [14] collected tweets from users participated in questioner to determine their political ideology. The tweets are labeled based on the users' political ideology. Similar to this paper Kulkarni et al. [15] extracted news articles and labeled them based news sources political affiliation. News sources political ideology rating is derived from the ALLSIDES.COM. Others Rahat et al. [16], Li et al. [17] and Bayram et al. [19] relied on publicly available resources of parliamentary and Congressional debates marked with party affiliations.

On the other hand, the bottom-to-top marks textual news articles with the ideology that matches its political stance regardless of its author political leanings [20,21]. The pros and cons of the two labeling approaches are briefly summarized: The top-to-bottom approach avoids confusion and takes self-identified authors' ideologies for granted, yet it risks mislabeling political ideology texts with uncommon political stances within the ideology. In the bottom-to-top, researchers examine each text content to associate it with a specific ideology. However, it requires an expert or annotators with an adequate understanding of the targeted country's political system [20,21], and its policies.

This paper follows the first approach for two reasons: First, we aim to identify two broad political ideologies of the articles extracted from media sources placed on the far-left or far-right of the political spectrum. Political ideology generalization from the extreme left and right minimizes the number of stance outliers because stances are most likely crossed between ideologies in the middle right or left of the political map. Second, bottom-to-top approach labeling is more suited for sentences or short texts [20,21], not long articles, where some portions of the articles do not necessarily exhibit any political alignments. On that note, a single textual news article may include paragraphs or even sentences with different political stances.

In the literature of political ideology detection, researchers obtained ideologized textual content through multiple sources. One can find ideologized text in social media [14], campaign speeches [21], congressional transcripts, debates [16,19,20], news websites or blogs [15,20]. While we are addressing the problem of political alignments in relatively long textual news articles, ideology detection models could be applied to sentences [20], short texts [14,17,21], and documents [15,16,18–20]. We collected textual political posts tagged with politicians' names from various news websites and blogs; however, others collected politically ideologized texts within selected sets of topics [21] or unbounded by any topics [14–16,18–20]. Authors of the ideologized content is an additional variable to the problem; authors could be categorized as ordinary people [14], politicians, [14,16,19,20], and journalists [15,18].

The data collection and construction processes of the political ideologies datasets vary. For instance, a questioner can help to identify news sources' ideology through their readers [14], others employed annotators to go through the news articles to label them [20,21], or by taking advantage of Natural Language Processing (NLP) techniques [20] to label the examples in the dataset automatically. The number of political ideology labels in the dataset is also a crucial aspect of the problem. It is possible to categorize the political ideologies into coarse polarized ideologies [14–16,18–20], or fine-grained ideologies [14,21]. The ideology detection method can incorporate textual content only [14,16,18,20,21], or by its content and context [15]. Examples of ideology detection methods are statistical inference [21], machine learning [14,18–20], deep neural networks [14–16,20,22].

3 Presidential Dataset

3.1 Dataset Construction

Exploring and analyzing the dataset is critical to finding suitable solutions for the ideology detection problems and understanding the experiments' results. This section provides information about the construction process of the newly made dataset known as the Presidential dataset. An earlier version of the Presidential dataset briefly introduced in [23]. The following summarizes the differences between the older and current dataset versions: the older version only focuses on the articles' ideology detection regardless of the sources, authors, and dates. The Presidential dataset consists a collection of articles written about Trump and Obama, the dataset can be downloaded from this link (https://bit.ly/3jgKzmw). The older version did not sort articles based on the publication date. We also filtered out articles with missing values, short content, or no mentions of the targeted politicians. In the reconstruction process, we stratified Trump and Obama's corpora into separate training and testing set based on the attributes such as sources, authors, and ideologies. The reconstructed dataset ensures that the researcher will access the data once to explore and experiment on it. For instance, one can easily incorporate authors or any other attributes into the ideology detection algorithm. Another possible use is to analyze the similarities between articles covering a specific event; the articles originated from different sources, authors, or ideologies. Therefore, we believe that the reconstructed Presidential dataset, which includes multiple perspective datasets and more features, is more suitable for experimental purposes. Also, the difference in experimental settings and research approaches between [23] and this paper makes the two incomparable.

3.2 Web Crawlers

The web is rich in the unstructured text but obtaining meaningful labeled information from the internet is not straightforward. Hence, many researchers rely on agencies that specialize in constructing labeled datasets. However, other methods enable researchers to automate the collection process of the desired data, such as by web crawlers. Web crawlers are programs that automate the process of web browsing, data selection, and extraction. Every website has a different Document Object Model (DOM) structure,

HTML element names, classes, or I.D. Therefore, we developed several web crawlers for each media website to go through web pages and single out needed information.

The search scope is determined by examining the website structure and local URLs. We did the same for every media source website to program the algorithm that formulates the crawler browsing steps. Since we are interested in personalized news articles rather than articles covering a specific event or topic, the selection process is limited to articles about or associated with a predetermined politician. Conventional search techniques such as searching for the exact word in the content might retrieve articles that barely mention the intended person. Often news website utilizes tags to link articles with locations, events, objects, or persons. Therefore, we only picked the articles that were tagged with the politician's name.

Although political ideology alignments in the news media are universal, we constructed and experimented on data published in United States' media. We decided to examine the political polarity of the U.S. news media for a couple of reasons. First, the U.S. media heavenly engages in politics, which means many political news articles are published in the U.S. media. Second, the extent of political polarization in the U.S. is continuing to grow. Third, the U.S. presidents' popularity and position as leaders of the country with dominant power are rational reasons that attract researchers worldwide to U.S. politics.

Our crawler should be able to automate the browsing, clicking, reading, and writing actions. The objective is to select articles tagged with a predetermined politician and extract needed information to build the dataset. Tags cannot contain empty spaces; thus, we used one or more tags to search for articles related to two U.S. presidents, Barack Obama, and Donald Trump. However, one should be aware that some websites do not allow explicit looping through their pages and require virtual browsing. The browsing in these websites simulates user interaction, which is achieved by virtual WebDrivers.

Once we implement the virtual user who can browse and click through all selected articles, the next step would be to inspect the page's structure manually. Web page inspection is required to decide which HTML elements or CSS selectors will be crawled. Nevertheless, websites place different information in the HTML elements; for instance, the HTML element for the article's title is not the same as the article itself or its author. Some other times websites group several HTML elements with a unique I.D. There are numerous styles for building websites, highlighting the need for different web crawlers for each website. Each web crawler is designed for automatic HTML element selection and content extraction. After obtaining the needed information from the web crawler, the program sorts the collected data based on some criteria and then stores it into a flat-file. We opted for flat files as a storage mechanism to store the data due to its ease of accessibility, relatively small storage size, and portability. Lastly, we developed the web crawlers in Python with several external packages such as Beautiful Soup, Selenium, and Google Chrome WebDriver.

3.3 Data Sources

Articles' political ideology labels are inherited from their publisher; hence, data source selection is critical. This research paper purposely avoided news media outlets known as mainstream media that claim to be bias-free. All the news media sources included in this research are self-claimed to align with one side of the political spectrum. We assume that Liberal or Conservative media will attract bloggers, journalists, and editors with the same political affiliation. Meaning, each article will be labeled with a political ideology that corresponds to its source. A brief description of each media source and its political ideology based on Allsides Media Bias Ratings and Media Bias Factcheck will follow.

- **DailyKos** (https://www.dailykos.com) is a well-established liberal collaborative blog and news media platform. Since it was founded in 2002, the DailyKos has covered multiple presidential election cycles, and its contents have grown over time. Tags are one of the prominent search and navigation techniques used on their website. It is placed on the far-left.

- **DailyWire** (https://www.dailywire.com/) is a conservative news website founded in 2015; its focus is on U.S. politics. This website utilizes tagged keywords as a search technique, but not as explicit as DailyKos. It is placed on the far-right of the political map.
- **National Review Online** (https://www.nationalreview.com/) is a right-wing news media; it is considered moderate compared to the DailyWire, and ILoveMyFreedom. National Review Online does not shy away from criticizing Donald Trump. The magazine version of this media outlet was founded in 1955. The National Review Online is placed to the far right of the political spectrum.
- **WorldSocialist** (https://www.wsws.org/) is a far-left news media site a critic of capitalism and the Democratic Party. It was founded in 1998.
- **TheBlaze** (https://www.theblaze.com/) is a conservative news website founded in 2011.
- **ILoveMyFreedom** (https://ilovemyfreedom.org/) is a conservative news and opinion website. Not much information is available about this website other than its explicit support for U.S. President Donald Trump.

3.4 Dataset Description

As mentioned earlier, the dataset is stored in flat files for easier accessibility and a limited number of retrieving calls. Each article and other extracted information fit in a single line divided by a delimiter. A single line in the file consists of eight fields, namely, I.D., Date, Title, Content, Author, Source, and Political Ideology. The I.D. field is a unique identifier that captures the source, person, and number representing the article's extraction order. The Date field states the day, month, and year of the article's publication. Articles' titles are stored in the Title field. Articles' author attribute contains authors' actual name, the anonymous user I.D., or an external site name. The origin or the website that published the article is the value stored for each article in the Source field. Ideology classes, which are limited to Conservative or Liberal, marked the articles in the Ideology field. In addition, to ease data accessibility and minimize machine memory loading, structuring the dataset in this way allows for flexibility in experimenting. To avoid overfitting and replication near real-world scenarios, we partition the Presidential dataset into two sets: training and testing. We only use the training set in the learning process in all experiments, not the testing set. We also follow the tradition of keeping the more significant portion of the dataset in training set to build a better detection model. However, 15% of the training set is utilized to develop validating procedures to fit the model on unseen data before testing. Although we report some validation set results, we only rely on the testing set detection results in model performance assessments and comparison.

This section includes a descriptive analysis of the presidential dataset articles' ideology, sources, and authors. The Presidential dataset is a combination of two corpora known as Trump and Obama. The training and testing set of the Presidential dataset is also divided between the two separate datasets. We first will explore the statistics of the articles' ideology of the Presidential dataset. Then we will go through a much more detailed analysis. As we have stated in previous sections, Conservative and Liberal are political ideology classes addressed in this paper. The difference between Conservative and Liberal ideologies lies in their stand on issues, such as economic, social, governance policies. It is worth noting that more sophisticated taxonomy systems can divide the political spectrum into more than two classes. We believe an inclusive political ideology labeling system is more suitable for the Presidential dataset since the news sources included in this study are often placed on the extreme right or left. Unlike center-right and center-left, far-left and far-right are less likely to agree on controversial issues. The total number of articles in the presidential dataset is 178572. 72% of the Presidential dataset is Liberal, and the remaining 28% of the articles are Conservative, See Tab. 1. The exact reason behind this is not apparent, but some of the conservative news media are relatively recent compared to their counterpart. We had no

control over the ideology class size since the data collection process was automated and might reflect the representation size of the far left and right ideology in the U.S. media. There are some interesting insights into the presidential dataset that we believe should be highlighted. Out of the entire data, 70% and 30% are allocated for training and testing sets. As illustrated in Tab. 1, the ideology articles sizes in the training and testing set are consistent with the Presidential Dataset, 28% for Conservative and 72% for Liberal. Some other key factors that might be played a role in the size difference between conservative and liberal classes are the number of contributors, news website age, or crawling scope.

Table 1: Presidential dataset partitioned into two sets, train and test. The table shows the number of conservative and liberal articles in each one of the sets

Set	Conservative	Liberal
Train	35035 (28%)	90016 (72%)
Test	15017 (28%)	38504 (72%)

The number of unique authors in this dataset is 17388, meaning that a single author writes 10.2 articles on average. Considering that some media sources are also blogging platforms, it is reasonable to assume that authors are employed, owners, or volunteers. It explains why there are more articles written by specific authors than the others do. Articles written by the same author can improve ideology detection models' performance by identifying writers' styles. Tab. 2 lists the top 10 authors ranked based on their number of contributions. The top authors in both training and testing sets are pretty similar but not identical. Like other platforms that rely on the collaboration of contributors to build their content [24], news websites approximate a power-law distribution. It is quite interesting that just ten authors out of the 17388 are responsible for 17% of all training and testing sets articles. Few people with substantial contributions can direct the news media outlet towards one side of the political spectrum or the other.

Table 2: Top ten authors with the highest number of written articles in train and test sets

Order	Top authors in train	Top authors in test
1	Jim Geraghty (7273, 5.8%)	Jim Geraghty (3266, 6.1%)
2	Joan McCarter (2551, 2%)	Joan McCarter (1136, 2.1%)
3	Poopdogcomedy (2174, 1.7%)	Poopdogcomedy (921, 1.7%)
4	Laura Clawson (1838, 1.4%)	Laura Clawson (752, 1.4%)
5	Hank Berrien (1498, 1.2%)	Mark Sumner (588, 1.09%)
6	Mark Sumner (1405, 1.1%)	Hank Berrien (581, 1.08%)
7	Clayton Keirns (1297, 1%)	Clayton Keirns (507, 0.94%)
8	James Barrett (1210, 0.96%)	Ben Shapiro (489, 0.91%)
9	Ben Shapiro (1155, 0.92%)	James Barrett (481, 0.89%)
10	KOS (1019, 0.81%)	Martin (458, 0.85%)
Total	(21384, 16.9%)	(9179, 17.06%)

4 Research Models

Although the political ideology detection of articles shares some characteristics with sentiment classification where the views of two opposing sides are either with or against a politician, we did not denote it as a negative and positive problem. We expect news articles to be lengthy and cover topics from a political perspective; For this kind of articles' sentiment words are not necessarily directed towards a single politician. Deep neural networks do not require feature engineering algorithms that extract features to improve the prediction outcome. Since they do not require engineering algorithms, this is an advantage for deep neural networks over traditional sentiment analysis techniques. Furthermore, it is logical to assume that the news articles dataset size would increase overtime to keep up with the latest events; deep neural networks are better scalable with more data than other classical methods. Therefore, the use of DNN for sentiment applications has been increasingly popular [25].

4.1 FastText

FastText model, since its introduction in [26], has been proven to be one of the effective models for text classification tasks [27,28]. FastText is not considered a deep learning model, and it consists of a text embedding layer. The model also includes a single average pooling layer and a fully connected layer with Softmax as an activation function. As described in the paper, FastText text representation overcomes traditional linear classifiers' limitations by linking the model's parameters with text features and classes. Fig. 1 shows that the articles to fixed-length sequences conversion is achieved by padding short articles and truncating the longer ones. The word embedding is a matrix with vocabulary counts as rows and a fixed number of columns. Each sequence is represented with a numeric vector corresponding to its words identification number in the text representation matrix. The word weights in the word embeddings and other network parameters are continuously updated during the learning process. The outputs of embedding layers are pooled with a non-overlapping average pooling layer to return a vector derived from averaging sequence dimension values. The last layer is a fully connected layer that takes averaged pooling vectors as an input to compute the dot product of input vectors and Dense layer weights, then feeds it to the Softmax function. The Softmax output will determine the input sequence label by normalizing values to a probability distribution between 0 and 1.

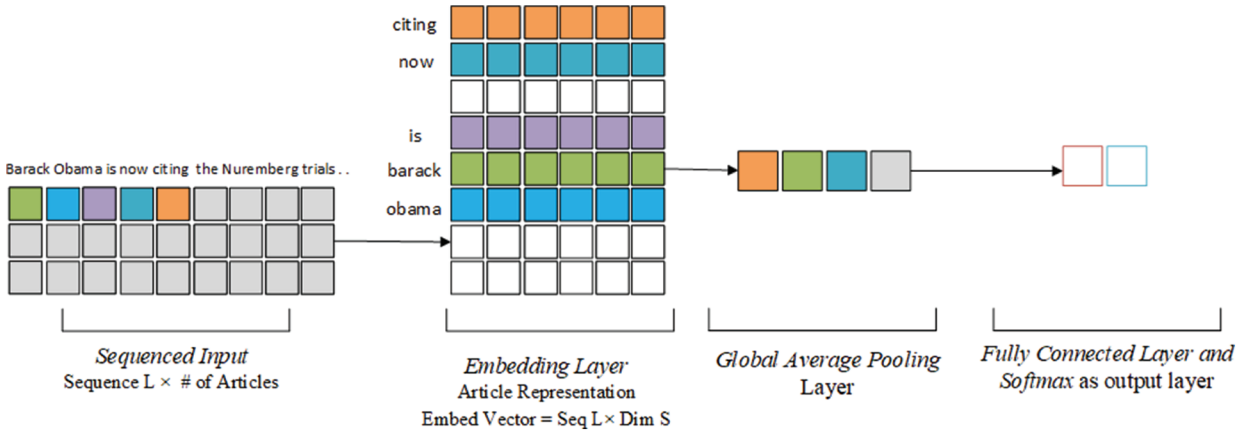

Figure 1: FastText network architecture

4.2 TextCNN

Fig. 2 illustrates the design of the Convolutional Neural Networks for the text classification. The original paper proposed four variations of the same model; in this paper, we are only interested the variation known as

CNN-rand, which have been tested on multiple text classification task [29]. TextCNN is similar to FastText in terms of sequencing and word embeddings procedures. In other words, all articles are sequenced into a fixed-length, and word embedding is randomly initialized. The main difference between the FastText and TextCNN lies in the use of multiple filter convolution layers. This paper uses the standard window sizes 3, 4, and 5 to extract different feature maps from each convolution layer in a non-linear way. Each output in the convolution layers is fed to a separate layer that takes maximum values from the feature map. This layer is commonly known as the max-pooling layer. The original paper argues that features with maximum values are the most crucial features within the features map; hence, it advances the performance of the TextCNN model. The three different feature maps are merged in the concatenation layer. Finally, the concatenation layer's results are delivered to the last layer in the TexCNN model, a fully connected layer with the Softmax activation function. The fully connected in TextCNN operates similarly to the one in the FastText network.

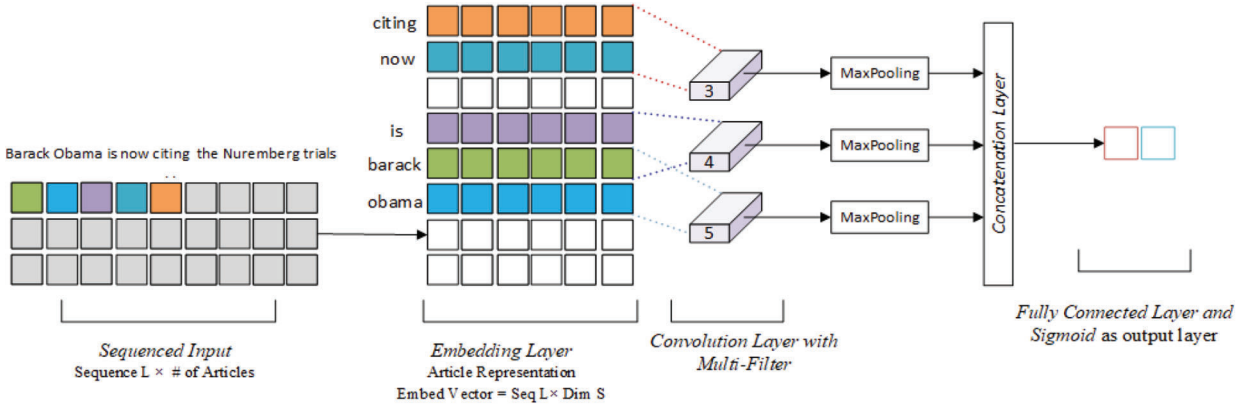

Figure 2: TextCNN network architecture

4.3 RCNN

The previous two models emphasized the importance of the model architecture; the following two models, including Recurrent Convolutional Neural Networks (RCNN) [30], added another dimension to the text classification problem using deep neural networks. RCNN is also an influential text classification model proven its effectiveness in text classification. RCNN, as displayed in Fig. 3, computes three text representations stemmed from words' position in the sequence. The sequence is portioned into two, left and right. The shared word embeddings layer learns from words in the left, right, and entire sequences. The embedding layer is shared, but the output is not. Different embedding vectors exist, including the output of the embedding layer, depending on the words' context. The left and right embedding vectors are fed to the Forward RNN and Backward RNN layers, respectively. Both RNN layers are fully connected, operate over the sequence's timesteps, are initialized uniformly, and use Tanh as an activation function. The two layers have critical differences, whereby in the backward RNN process, the sequence is reversed, yet not shown in the model's design. The output of the backward RNN is reversed back. The current text representation is for all previously mentioned contexts. The forward RNN, word embedding, and reversed backward RNN outputs are joined together in the concatenation layer. Then a convolution layer takes the merged values to convolve on with a window size set to 1. According to the original paper, the convolution layer represents the words' contextual information more meaningfully than traditional convolution layers. The max-pooling layer picks the maximum values from the convolution layer's output to get the most informative features. At last, a fully connected layer with Softmax activation function for predicting documents label based on its probability.

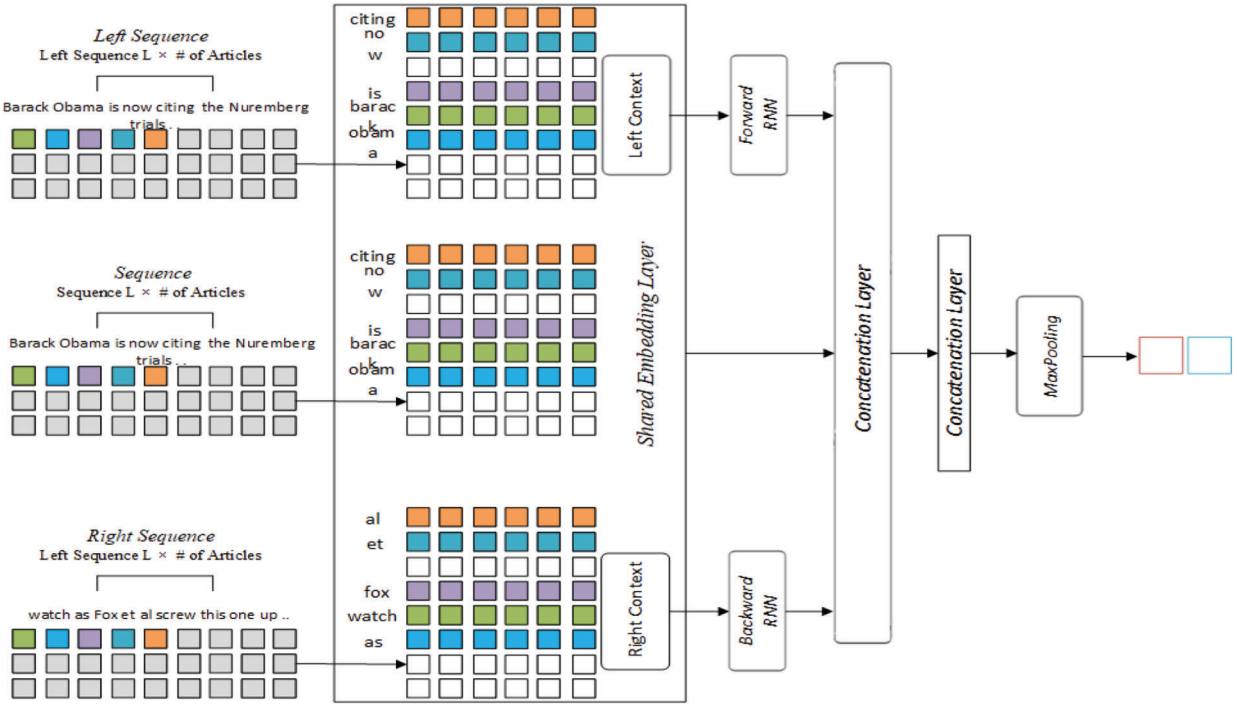

Figure 3: RCNN network architecture

4.4 HAN

In Hierarchical Attention Networks (HAN) [31], the network structure mirrors words, sentences, and document components in documents; thus, it is most suited for document classification. As shown in Fig. 4, the HAN model design expects documents to be reconstructed into sentences and words for each sentence. Hence, input sequences are reshaped to produce multidimensional input where each sequence has a fixed number of sentences, and sentences have a fixed number of words. In the first level of the HAN hierarchy, words are encoded by feeding each word embedding value to the forward RNN variation known as Long-short-term-memory (LSTM) layer. LSTM is wrapped in the Bidirectional RNN layer that computes backward RNN on the layer input. It also concatenates LSTM and Bidirectional RNN outputs. An attention mechanism is used directly to the word level to form a sentence after selecting the most informative words. Word encoder and attention are wrapped in a time Distributed layer to ensure that all the previous operations are correctly executed on all the words across the two layers. Formed sentences are encoded, and attention is applied to encoded sentences to assemble the most informative sentences in a document—finally, a fully connected layer with Softmax activation to predict the document class.

5 Results

5.1 Experiments Setup & Data Pre-Processing

All our experiments have the same settings, regardless of the model. The experimental settings can fit into two categories, model settings and text representation. As for the model settings, the models' optimizer is an extension of stochastic gradient descent named Adaptive Moment Estimation (Adam). While there might be better optimizers to update the model weights and minimize the loss function, Adam optimization is well-known and has proven to be effective. We set Adam's learning rate to 0.0001 and its decay parameters beta_1 and beta_2 to 0.9, 0.999, respectively. Ideology detection is a binary classification problem; hence, we utilize Binary Cross-entropy as a loss function. Model parameters are

updated to find the best fit for a model on the data; the updating process usually occurs during the training phase. It can be fitted on the validation set based on its accuracy as a measurement metric. We will use different metrics to evaluate models' performance on the testing set, detailed later in this section. We follow the standard batch size 32 with 50 training epochs maximum. 15% of the training set is set aside for validation after each epoch with an early-stop scheduler to terminate training if the validation accuracy did not increase for three consecutive epochs. The maximum number of extracted features is 35000 unigrams.

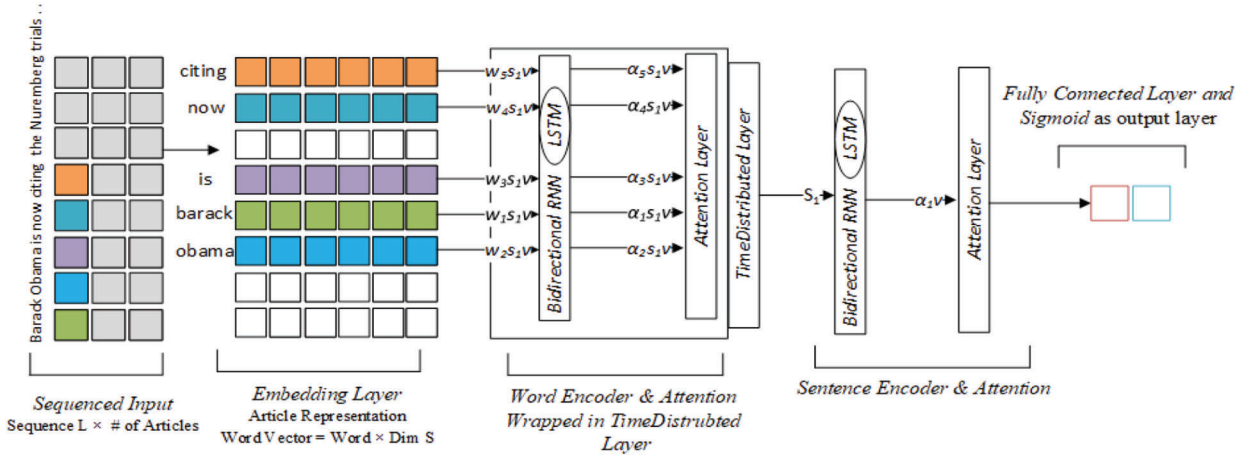

Figure 4: HAN network architecture

Text tokenization occurs before converting the unstructured texts to structured. We used white spaces as tokenization delimiters. Then, texts are cleaned by removing all punctuations and lowering the remaining letter cases to avoid duplication. Article strings are converted to a padded sequence of unique integers with 400 as a fixed sequence length. The same features that were found in the training set will be looked up in the testing set. As required by deep neural networks, padded sequences are represented by continuous values vectors or word embeddings. Word embeddings size was set to 50, and its weights were initiated randomly and updated by training on the underlining data, meaning no pre-trained embeddings were used in this experiment. Finally, we converted articles' labels from text to a binary representation to train and match with model predictions. We built the networks in the Keras platform with TensorFlow as the backend in all the experiments.

5.2 Models' Validation Accuracy

Analyzing models' performance on the validation fold is one way to find the most appropriate parameters for each model. The models are trained on the Presidential training set, a mixture of two corpora containing Obama and Trump articles. Observing the neural network performance during the training phase helps researchers assess detection models accurately and open new possibilities for future adjustments. Fig. 5 illustrates the accuracy of the ideology detection models on the validation set, which is 15% of the training set. Each model behaved differently; for instance, the RCNN model terminated its training in less than ten epochs, yet RCNN takes a longer time to train. Also, the decrease in prediction accuracy over epochs is more notable than other deep learning models. Hence, the margin between training accuracy and validation accuracy widens after the sixth epoch. Although FastText is much faster than any other model we experimented on in this paper, it required a more considerable number of epochs before reaching the point of termination. In contrary to the RCNN model, the decline in validation accuracy is not rapid and more stable. RCNN and FastText got around 0.85 of accuracy at some point in

the training process; the same applies to TextCNN and HAN models; see Fig. 5. However, the HAN model exhibits distinct behavior where validation accuracy oscillates before training ends at epoch 16. Finally, TextCNN is second to FastText in the number of training epochs needed to build the model where training aborted right before epoch 25. Furthermore, the accuracy results are similar to and as stable as FastText.

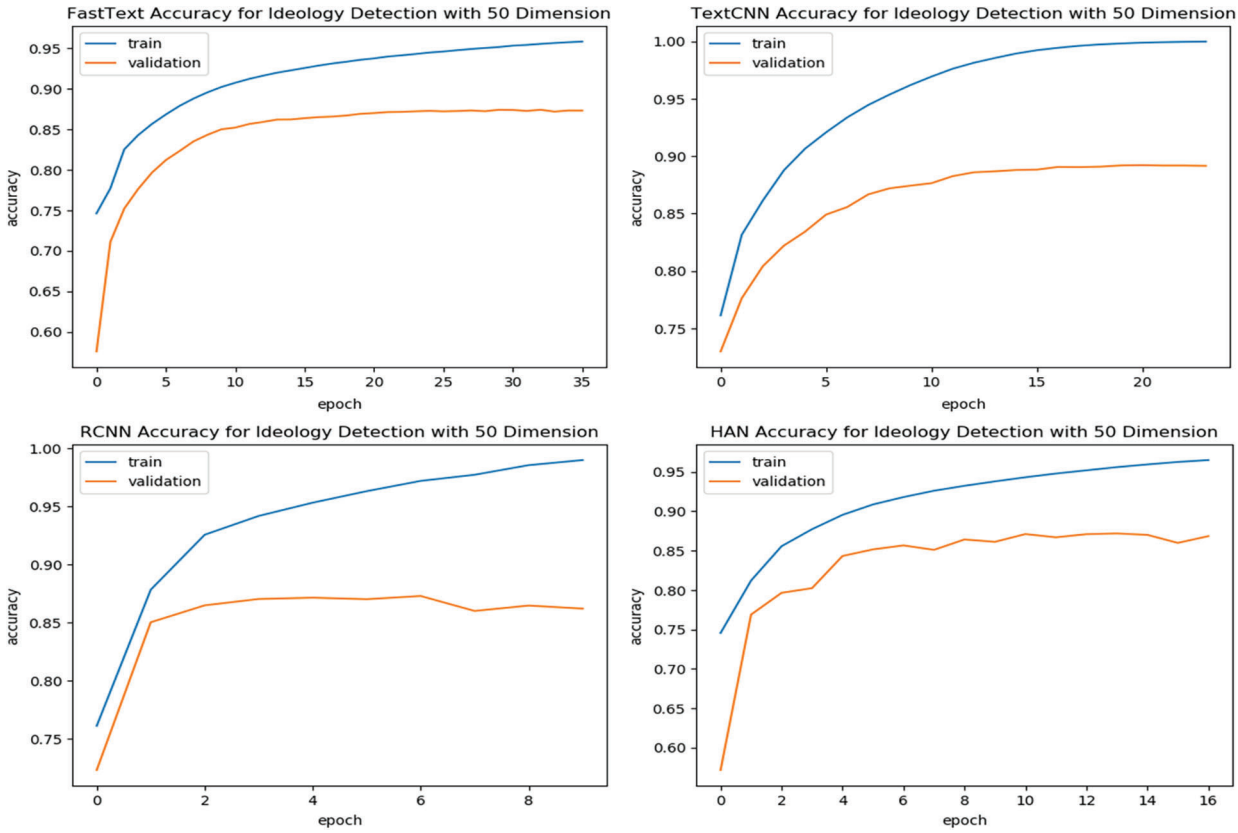

Figure 5: Ideology detection models' performance on the validation set

5.3 Ideology Detection Results

Aside from prediction accuracy, several metrics could be used to measure the performance of detection models. Although some can argue that the accuracy is sufficient for binary classification problems, simple accuracy is often misinterpreted for imbalanced datasets as the Presential dataset. Therefore, more suited metrics are employed, namely, Recall, Precision, and F1-Score, which are mathematically expressed as $Precision = \frac{True\ Positive}{True\ Positive+False\ Positive}$, $Recall = \frac{True\ Positive}{True\ Positive+False\ Negative}$, and $F1 = 2\ \times \frac{Precision \times Recall}{Precision+Recall}$. In Fig. 6, measures are provided in two types of graphical representations. The first graph, Fig. 6, shows higher values in lighter colors with numerical values in the center of each box in the heatmap. While the other diagram, Fig. 7, presents the four ideology detection models' findings side by side in a single vertical bar chart, both graphical representations show that all models achieved better Recall, Precision, and F1-score for Liberal class than Conservative class. We expected this outcome considering that the dataset is skewed towards the Liberal class. However, all models reported at least 0.8 Recall on Conservative class, meaning 80% of conservative articles are detected, which is reasonable considering that Conservative articles comprise only 28% of the Presidential dataset. We argue that the presence of heavy ideological

polarity in articles minimized the impact of data imbalanced, yet it did not eliminate it. The same reasoning could be applied to the Precision results that scored 0.84 and higher for the Conservative class.

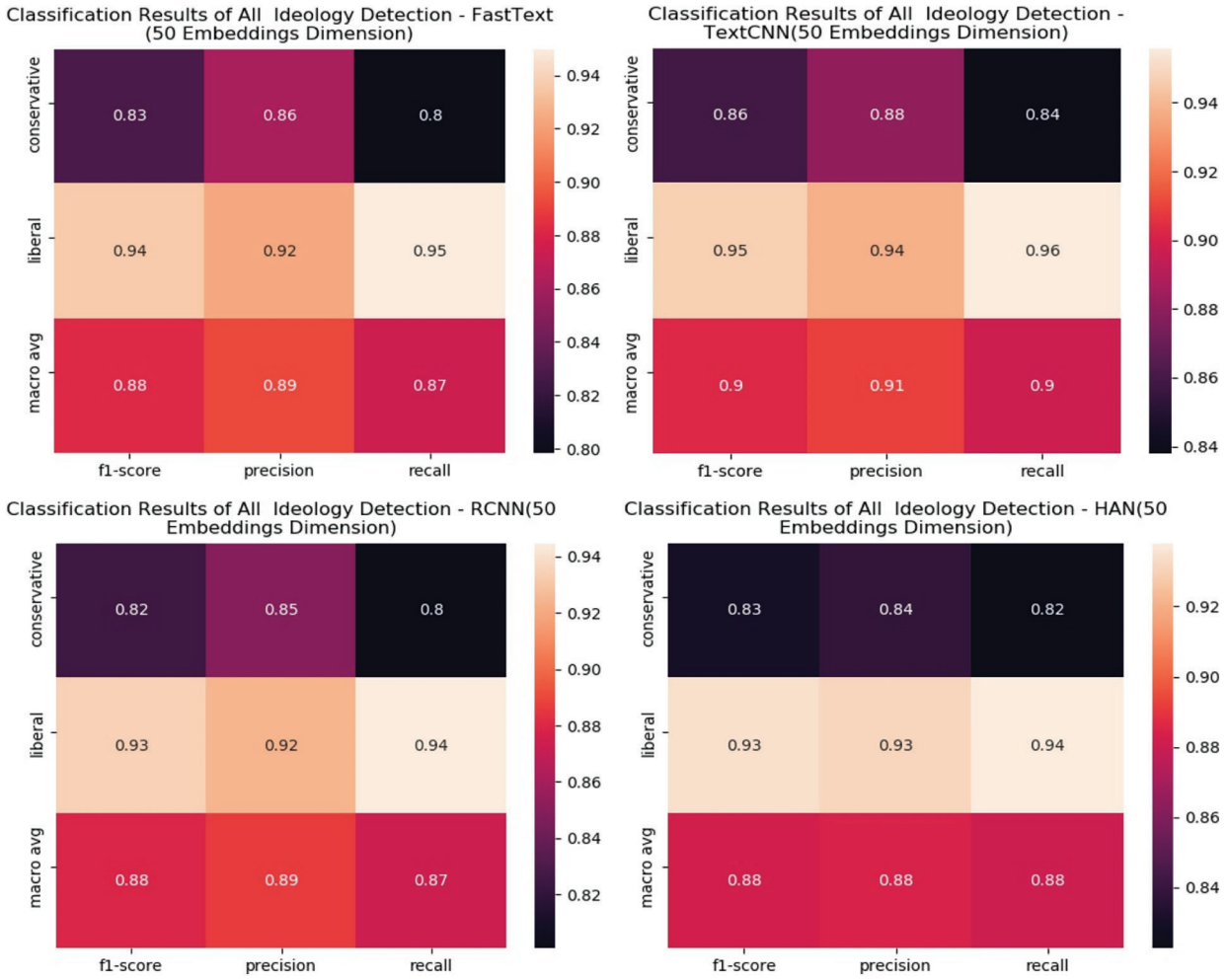

Figure 6: Heatmap of FastText, TextCNN, RCNN, and HAN ideology detection models' performance

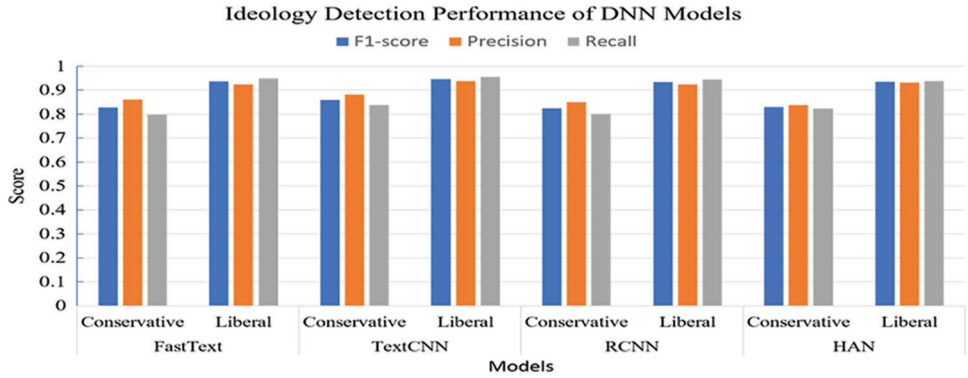

Figure 7: Comparison of ideology detection models' performance

Nonetheless, it is more likely that a conservative article to be mislabeled as a Liberal article. Overall models result, except for TextCNN, are similar. TextCNN performance exceeded other models on all metrics for Liberal and, more notably, for Conservative classes. Therefore, we can assume that TextCNN is less susceptible to data imbalance. Compared to the other three, it is the best model to detect articles' ideology with a 0.9 Macro averaged F1-Score. The difference is observable in the bar chart, where all models' results are plotted side by side in a single chart.

5.4 Sources

Articles' source is not included in the ideology detection training process as an attribute, yet we examine models' prediction accuracy for each source. DailyKos has the most significant number of articles in both training and testing. From first glance at Fig. 8, we can see that all models detected over 90% of articles originated from Liberal sources, namely DailyKos and WorldSocialist. The detection accuracy of articles posted in liberal sources is anticipated since the four deep networks recorded a 0.9375 F1-Score on an average for the Liberal class. Although DailyKos articles outnumber the WorldSocialist ones, some of the ideology models detected 100% of the articles published in WorldSocialist. This result might be attributed to the WorldSocialist's position on the political spectrum, which is further to the left than DailyKos, making it easier to distinguish from other conservative articles. Another factor is that DailyKos's size increased the number of authors, topics, and writing styles, making it harder to achieve better results than the ones reported for WorldSocialist. Despite all the shortcomings we listed earlier, TextCNN, Fig. 8, can detect over 95% of the DailyKos articles. Moving on to the other side, out of the five conservative media sources, TextCNN accurately predicted the 90% and over of articles' ideologies published by the Blaze, DailyWire, and ILoveMyFreedom. The same model correctly labeled around 87% of NewsBusters Articles.

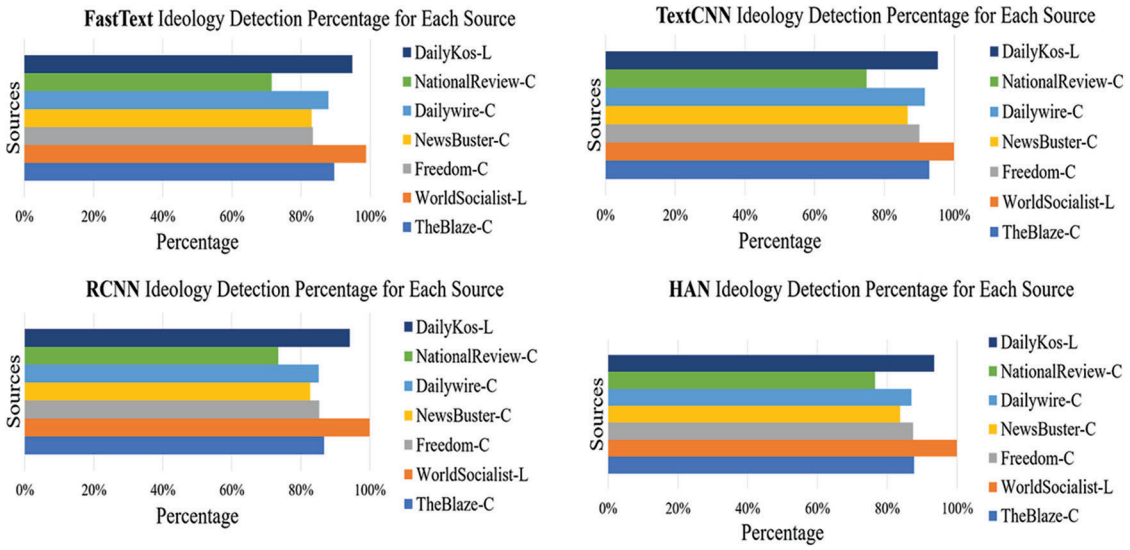

Figure 8: Political ideology detection percentage of Presidential Dataset's media sources

Interestingly, all these conservative sources are considered further to the right of the political spectrum compared to National Review Online, closer to the center-right. National Review Online comprises 40% of conservative articles, though TextCNN accurately predicted only 75% of the articles. In comparison, the conservative data consist of 26% of Dailywire articles, and 92% of them were successfully detected. Despite the lower detection performance of the other three models, the rank of articles' sources based on

their detection accuracy is similar to TextCNN. National Review has criticized President Donald Trump on several occasions, contributing to its relatively low accuracy detection rate.

5.5 Authors

Leveling down from data sources to articles' authors provides new insights into the data and the predictive models. As discussed before, we assume that writers or journalists often contribute to media platforms that fit their political beliefs. Similar to news sources, the authors' data are not included as attributes in the training process. Evaluating the ideology detection models' performance on authors with the highest publications number might enable us to estimate authors' political alignments. Meaning that if we can accurately predict 100% of the authors' articles' ideology, it is more likely that the author has a political bias or at least have more common grounds with the political ideology in question. Fig. 9 shows the ideology detection accuracy of the top 25 authors with the highest number of articles in the testing set, the threshold for the minimum number of written articles set to above 250 articles per author. It is worth noting that all detection models recorded 100% accuracy for TheBradBlog, News Corpse, Martin, and Poopdogcomedy. TheBradBlog is a website founded by Brad Friedman that re-posted articles on DailyKos. BRAD FRIEDMAN interests lie in covering election and voting integrity from a progressive standpoint. News Corpse is another independent website published by Mark Howard, who also writes for the DailyKos. According to the News Corpse mission, Mark Howard established the website as a response to corporate-dominated institutions. Martin is a writer for ILOVEMYFREEDOM, and we were not able to find any information about him. Finally, our ideology models detected 100% of Samuel Sero's articles, a progressive democrat writer for the DailyKos under the username Poopdogcomedy.

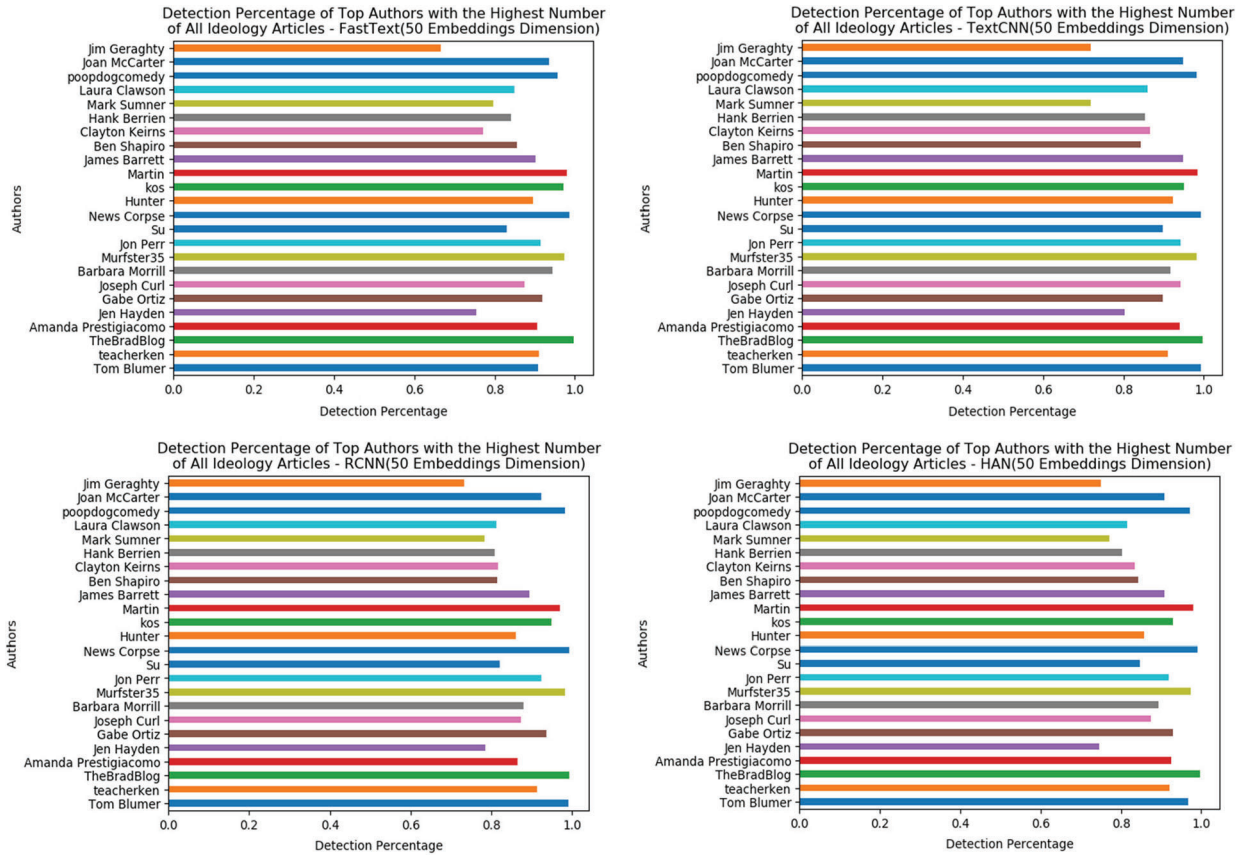

Figure 9: Political ideology detection rate of the top authors in the Presidential Dataset

On the other hand, the models predicted the ideology of articles written by Jim Geraghty, a political correspondent on National Review, and Jen Hayden, a Trending News Manager and a writer at DailyKos, with lower than 80% accuracy. It is feasible to assume that news media outlets are inclusive of authors with somewhat dispersed political stances. We can assume that conservative news outlets would include writers with far-right views or center-right, which applies to the liberal media. However, it is easier to predict the ideology of articles written by authors who identify as progressive democrats or far-right than those on the center-left or right. We can deduce that two criteria assist in placing news outlets on the political spectrum, its authors' political ideology and the number of articles written by each one of the authors.

6 Conclusion

This research paper solely focused on the articles' content to predict articles' political ideology. The results of the ideology detection models are promising, with over 0.9 F1-score. Detection results of the ideology of articles authors and sources are reasonable with accuracy as high as, or around, 100% for some news media sources and authors. Therefore, deep neural networks have repeatedly proven their effectiveness, and there are still windows for improvements. For instance, one might argue that the robustness of the models is better evaluated when the testing set includes news sources that do not exist in the training set. However, we believe that having a large number of unique authors in the dataset improved models' robustness. Moreover, deep neural networks' performance could improve by applying data augmentation, sampling, or pre-trained embeddings techniques. Other attributes in the Presidential datasets never used in the models' training process, such as articles' sources, authors, and date, could enhance the performance of the detection models. Furthermore, the Presidential dataset allows for additional unexplored tasks such as multi-label and multiclass detection problems by combining two or more of the following articles' labels: ideology, personalization, authors, or news sources.

Funding Statement: The author received no specific funding for this study.

Conflicts of Interest: The author declares that they have no conflicts of interest to report regarding the present study.

References

[1] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[2] T. T. Lee, "Virtual theme collection: "Trust and credibility in news media," *Journalism & Mass Communication Quarterly*, vol. 95, no. 1, pp. 23–27, 2018.

[3] N. J. Stroud, "Media use and political predispositions: Revisiting the concept of selective exposure," *Political Behavior*, vol. 30, no. 3, pp. 341–366, 2008.

[4] J. A. Frimer, L. J. Skitka and M. Motyl, "Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions," *Journal of Experimental Social Psychology*, vol. 72, pp. 1–12, 2017.

[5] E. Bakshy, S. Messing and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[6] D. V. Dimitrova, A. Shehata, J. Strömbäck and L. W. Nord, "The effects of digital media on political knowledge and participation in election campaigns: Evidence from panel data," *Communication Research*, vol. 41, no. 1, pp. 95–118, 2014.

[7] K. Holt, A. Shehata, J. Strömbäck and E. Ljungberg, "Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveler?," *European Journal of Communication*, vol. 28, no. 1, pp. 19–34, 2013.

[8] H. K. Meyer, D. Marchionni and E. Thorson, "The journalist behind the news: Credibility of straight, collaborative, opinionated, and blogged news," *American Behavioral Scientist*, vol. 54, no. 2, pp. 100–119, 2010.

[9] S. Schäfer, "Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions," *Computers in Human Behavior*, vol. 103, pp. 1–12, 2020.

[10] K. Alzhrani, F. S. Alrasheedi and F. A. Kateb, "CNN with paragraph to multi-sequence learning for sensitive text detection," in *Proc. 2nd Int. Conf. on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, IEEE, pp. 1–6, 2019.

[11] M. Hughes, I. Li, S. Kotoulas and T. Suzumura, "Medical text classification using convolutional neural networks," *Studies in Health Technology and Informatics*, vol. 235, pp. 50–246, 2017.

[12] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega and A. L. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Systems with Applications*, vol. 55, pp. 194–211, 2016.

[13] A. Tumasjan, T. O. Sprenger, P. G. Sandner and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proc. Fourth Int. AAAI Conf. on Weblogs and Social Media*, Washington, DC, USA, 2010.

[14] D. Preoțiuc-Pietro, Y. Liu, D. Hopkins and L. Ungar, "Beyond binary labels: political ideology prediction of twitter users," in *Proc. The 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, pp. 729–740, 2017.

[15] V. Kulkarni, J. Ye, S. Skiena and W. Y. Wang, "Multi-view models for political ideology detection of news articles," ArXiv Preprint ArXiv:1809.03485, 2018.

[16] G. Rahat and T. Sheafer, "Word embeddings for the analysis of ideological placement in parliamentary corpora," *Political Analysis*, vol. 28, no. 1, pp. 112–133, 2020.

[17] X. Li, W. Chen, T. Wang and W. Huang, "Target-specific convolutional bi-directional LSTM neural network for political ideology analysis," in *Proc. Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conf. on Web and Big Data*, Beijing, China, pp. 64–72, 2017.

[18] V. Fiore, "Automatic identification of political ideology in online news articles," *RAIS Journal for Social Sciences*, vol. 3, no. 2, pp. 50–54, 2019.

[19] U. Bayram, J. Pestian, D. Santel and A. A. Minai, "What's in a word? detecting partisan affiliation from word use in congressional speeches," in *Proc. 2019 Int. Joint Conf. on Neural Networks (IJCNN)*, Budapest, Hungary, IEEE, pp. 1–8, 2019.

[20] M. Iyyer, P. Enns, J. Boyd-Graber and P. Resnik, "Political ideology detection using recursive neural networks," in *Proc. The 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, USA, pp. 1113–1122, 2014.

[21] Y. Sim, B. D. Acree, J. H. Gross and N. A. Smith, "Measuring ideological proportions in political speeches," in *Proc. Conf. on Empirical Methods in Natural Language Processing*, Seattle, USA, pp. 91–101, 2013.

[22] C. Li and D. Goldwasser, "MEAN: Multi-head entity aware attention network for political perspective detection in news media," in *Proc. Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, Online, 2021, pp. 66–75, 2021.

[23] K. Alzhrani, "Ideology detection of personalized political news coverage: A new dataset," in *Proc. The 4th Int. Conf. on Compute and Data Analysis*, Silicon Valley, CA, USA, pp. 10–15, 2020.

[24] T. Chełkowski, P. Gloor and D. Jemielniak, "Inequalities in open source software development: Analysis of contributor's commits in apache software foundation projects," *PLoS One*, vol. 11, no. 4, pp. e0152976, 2016.

[25] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 335–4385, 2020.

[26] A. Joulin, E. Grave, P. Bojanowski and T. Mikolov, "Bag of tricks for efficient text classification," ArXiv Preprint ArXiv:1607.01759, 2016.

[27] J. Liu, W. C. Chang, Y. Wu and Y. Yang, "Deep learning for extreme multi-label text classification," in *Proc. The 40th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, Tokyo, Japan, pp. 115–124, 2017.

[28] H. T. Le, C. Cerisara and A. Denis, "Do convolutional networks need to be deep for text classification?," in *Proc. Workshops at the Thirty-Second AAAI Conf. on Artificial Intelligence*, New Orleans, Louisiana USA, pp. 29–36, 2018.

[29] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang *et al.,* "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[30] S. Lai, L. Xu, K. Liu and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. Twenty-ninth AAAI Conf. on Artificial Intelligence*, Austin, Texas, USA, pp. 2267–2273, 2015.

[31] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola *et al.,* "Hierarchical attention networks for document classification," in *Proc. Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, California, USA, pp. 1480–1489, 2016.