

Article

Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic

Abdullah Bin Shams ^{1,*}, Ehsanul Hoque Apu ^{2,3,*}, Ashiqur Rahman ^{4,†}, Md. Mohsin Sarker Raihan ^{5,‡}, Nazeeba Siddika ⁶, Rahat Bin Preo ⁷, Molla Rashied Hussein ⁸, Shabnam Mostari ⁹ and Russell Kabir ¹⁰

- ¹ The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada
 - ² Institute of Quantitative Health Science, Department of Biomedical Engineering, Michigan State University, East Lansing, MI 48824, USA
 - ³ The Intervention Centre, Oslo University Hospital, 0372 Oslo, Norway
 - ⁴ Department of Computer Science, Northern Illinois University, DeKalb, IL 60115-2828, USA; ashiqur@niu.edu
 - ⁵ Department of Biomedical Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh; msr.raihan@gmail.com
 - ⁶ Center for Environmental and Respiratory Health Research (CERH), Faculty of Medicine, University of Oulu, 90014 Oulu, Finland; nazeeba.siddika@oulu.fi
 - ⁷ Department of Electrical and Electronic Engineering, Bangladesh University of Business and Technology, Dhaka 1216, Bangladesh; rahatpreo1004@gmail.com
 - ⁸ Department of Computer Science and Engineering, University of Asia Pacific, Dhaka 1205, Bangladesh; mrh.cse@uap-bd.edu
 - ⁹ Aspire to Innovate (a2i) Programme, ICT Division, Dhaka 1207, Bangladesh; shabnam.mostari@a2i.gov.bd
 - ¹⁰ School of Allied Health, Faculty of Health, Education, Medicine and Social Care, Anglia Ruskin University, Chelmsford CM1 1SQ, UK; russell.kabir@anglia.ac.uk
- * Correspondence: abdullahbin.shams@mail.utoronto.ca or abdullahbinshams@gmail.com (A.B.S.); hoqueapu@msu.edu (E.H.A.)
† Equal contribution and joint first author.
‡ Equal contribution and joint second author.



Citation: Shams, A.B.; Hoque Apu, E.; Rahman, A.; Sarker Raihan, M.M.; Siddika, N.; Preo, R.B.; Hussein, M.R.; Mostari, S.; Kabir, R. Web Search Engine Misinformation Notifier Extension (SEMiNExt): A Machine Learning Based Approach during COVID-19 Pandemic. *Healthcare* **2021**, *9*, 156. <https://doi.org/10.3390/healthcare9020156>

Academic Editor: Yu-Jun Zheng
Received: 25 December 2020
Accepted: 28 January 2021
Published: 3 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Misinformation such as on coronavirus disease 2019 (COVID-19) drugs, vaccination or presentation of its treatment from untrusted sources have shown dramatic consequences on public health. Authorities have deployed several surveillance tools to detect and slow down the rapid misinformation spread online. Large quantities of unverified information are available online and at present there is no real-time tool available to alert a user about false information during online health inquiries over a web search engine. To bridge this gap, we propose a web search engine misinformation notifier extension (SEMiNExt). Natural language processing (NLP) and machine learning algorithm have been successfully integrated into the extension. This enables SEMiNExt to read the user query from the search bar, classify the veracity of the query and notify the authenticity of the query to the user, all in real-time to prevent the spread of misinformation. Our results show that SEMiNExt under artificial neural network (ANN) works best with an accuracy of 93%, F1-score of 92%, precision of 92% and a recall of 93% when 80% of the data is trained. Moreover, ANN is able to predict with a very high accuracy even for a small training data size. This is very important for an early detection of new misinformation from a small data sample available online that can significantly reduce the spread of misinformation and maximize public health safety. The SEMiNExt approach has introduced the possibility to improve online health management system by showing misinformation notifications in real-time, enabling safer web-based searching on health-related issues.

Keywords: COVID-19; public health misinformation; web search engine; notifier extension; natural language processing; machine learning; artificial neural network

1. Introduction

The current internet-based smart technological advancements, such as wearable technology and mobile phone applications (apps), have enabled constant real-time health monitoring of people in various diseases, and users have become more conscious regarding their health and about that of their loved ones [1]. Even long before the COVID-19 pandemic, in 2012, researchers reported that almost 59% of US adults searched for health information online. However, the number increased to 75% more recently, with more than a billion health-related searches occurring on Google search engines daily. There is no doubt that individuals are relying more on internet search engines regarding their health-related queries [2].

Misinformation is fake, unreliable, or not scientifically validated written material regardless of intentional authorship. Most misinformation-oriented discussions have been focused on venomous acts to taint the social media platforms with harmful and inaccurate information. False or fake news, misinterpretation of a drug protocol, or the presentation of unrealistic claims have dramatic effects on public health [3]. It draws attention away from scientific fact and real public health challenges and misleads healthcare professionals [3]. The flow of biased and inaccurate information can quickly dilute the seriousness of the actual issue and hamper the functioning of healthcare policy and disaster management [4].

Misinformation was widespread during the primary stages of the human immunodeficiency viruses (HIV) epidemic. It was also plagued by conspiracy theories, rumors, and misinformation for many years, with the effects still visible in regions to this day [5]. At the time of the avian influenza H5N1 outbreak in 2004, the World Health Organization's (WHO) Western Pacific Regional Office identified around forty rumors. Among them, nine were verified to be accurate [6]. During the West African Ebola virus epidemic in 2014, there was widespread fear and attention among the United States-based users, followers of Western media, and social media platforms such as Twitter [7].

The current ongoing coronavirus disease 2019 (COVID-19) pandemic situation has potentially affected the capacity of health facilities, even in developed countries where there are proven and robust healthcare systems [8,9]. The pandemic has brought unexpected, sudden, and unparalleled damages, changes to global health and socio-economic frameworks. For minimizing COVID-19 spread, most countries have enforced a societal-level lockdown, and citizens have resumed offices remotely and online activities by staying in residences as much as possible [10]. During the lockdown, people are using internet search engines and social media platforms to gain information about COVID-19. The nature of social media impact varies depending on an individual's gender, age, and level of education. Social media has played a vital role in spreading anxiety about COVID-19 in many territories. The COVID-19 pandemic has been termed as the first social media infodemic [11]. At the Italian lockdown period, even near bedtime, people increased the usage of digital media and internet usage [12]. In an Italian CoMuNe laboratory, Gallotti et al. have set up a COVID-19 "infodemic observatory," where they used artificial intelligence (AI)-integrated automated software to follow 4.7 million tweets on COVID-19 streaming past every day [13]. On the other hand, Cinelli et al. reported about 1.3 million posts and 7.5 million comments on COVID-19 from several social media platforms [12].

Pandemic fear among the population can promote online searches for unproven and unprescribed therapies. If fake news and misinformation spread in an uncontrolled manner, it may become fatal [14]. Moreover, the poisoning of over two thousand Iranians by swallowing methanol took place since a misleading social media message urged the mass people in Iran to prevent the SARS-CoV-2 infection by drinking alcohol. Around nine hundred illicit alcohol poisoned patients had to get admitted to intensive care unit (ICU) with a fatality of almost three hundred unfortunate deaths [15]. Global coverage about panic-buying in the online media and social networks only served to promote the same behavior, which caused stockpiling of drugs and vaccines as a method of preparation in the COVID-19 pandemic [16].

False news generally travels faster than reliable and authentic reports on social media platforms such as Twitter [17,18]. Much misinformation and rumor spread around online search engines and social platforms regarding COVID-19. Technology entrepreneurs and investors shared a document on Twitter, promoting the malaria drug chloroquine for treating COVID-19. Many mentioned successful therapeutic outcomes in China and South Korea. Despite that, when high profile individuals, such as entrepreneur Elon Musk, promoted chloroquine, it attracted the attention of the general population, which could lead to personal decision-making about treatment options [17,19]. Misinformation circulated rapidly before the results of a small, non-randomized French trial of the related drug hydroxychloroquine; at that time the article was in press [20]. Hospitals have reported poisoning cases, where individuals suffered from toxicity from chloroquine containing pills, which they intended to take for COVID-19 [19]. During early July 2020, the WHO released a press note about discontinuing hydroxychloroquine. In the most recently published original article, it has been shown that hydroxychloroquine did not improve clinical status at two weeks, compared with standard care [21].

To control and track the COVID-19 trajectory, different governments have implemented various digital health surveillance tools, such as smartphone-based apps for COVID-19 contact tracing [22]. Researchers have suggested that the US Food and Drug Administration (FDA) must warn the general population against collecting unapproved, unprescribed therapies and medicines [17]. Google integrated an educational website into search results related to the COVID-19 outbreak, and they mentioned that it could be extended for unapproved COVID-19 medication-related searches [17]. The rise in the use of social media spreads misinformation and unproven treatments about infectious diseases such as COVID-19 at lightning speed. An infinite amount of misinformation is available over the internet and it keeps rising day by day. Therefore, machine learning (ML) is one of the potential candidates as a counterstrategy to handle and classify the authenticity of a large volume of information automatically and in real-time. To identify fake news, scientists have developed ML-based false news and misinformation credibility inference models, which form a deep diffusive network model to memorize news articles, writers, and topics [23,24].

To combat the propagation of misinformation, some of the biggest internet and social media companies have introduced technical gatekeepers to control what information can be sent over their networks. WhatsApp launched a chatbot to connect its millions of users with various fact-checking organizations across the globe [25]. This allows a user to double-check the information. The platform also introduced the WHO's alert notification. This service responds to public queries about Coronavirus and provides official information 24 h a day, worldwide [26]. Facebook deployed machine learning algorithms to detect the advertisement of false claims such as homeopathic remedies that can prevent, cure, or protect against COVID-19 [27]. It also bans the sale of commercial safety products during the outbreak, such as medical grade face masks and hand sanitizers, to ease the panic buying. Google has introduced "Fact Check Explorer," which verifies the information available online through authenticated third-party fact-checkers [28]. The results of a search query indicate whether the claim from each source is valid, false, partly right or partly wrong. The fact check service is available in multiple languages and is also applicable to validate an image.

As we have mentioned above, based on the scientific literature, to date, scientists have been utilizing surveillance systems for infodemia tracking and analyzing disease outbreak-related fake news, misinformation, and rumor spread in online media and social platforms [19]. However, there is an immediate need for a tool for notifying any user about the scandal, fake news, and misinformation, while browsing in the web search engines. A web browser extension for a web search engine can be illustrated as a computer programming code package, which could be installed into a web browser that the user primarily uses for searching on the world wide web (WWW). The extension has the capability of adding a new feature to a web browser's search engine, or augment a current

functionality, update a visual theme [29], or in this case, screen a search result's content and show warning as per severity.

This study's main objective was to evaluate the potential ML-based approach integrated with search engine extension for notifying any public health misinformation during this COVID-19 pandemic. This paper proposes a novel approach to prevent the spread of misinformation through a web search engine extension (SEMiNExt). Once a user inquires for specific information on the web search engine bar, SEMiNExt is activated. The extension reads the search bar's user query and converts the texts into a number using NLP. This numerical data is then inputted into a trained ML algorithm that classifies the query's integrity and notifies its authenticity to the user using a message box on the screen. The overall procedure transpires in real-time to alert the user of potential misinformation pitfalls before clicking on any search results. This may significantly help to prevent the spread of misinformation.

2. Materials and Methods

2.1. Search Engine Misinformation Notifier Extension (SEMiNExt)

The comprehensive data flow diagram of our proposed methodology is illustrated in Figure 1. Google Chrome is chosen as the benchmark web browser, as 65.47% of the netizen uses this web browser as their primary one [30]. When a user enters a text in the web search engine bar, SEMiNExt is activated and the textual content is copied and then screened for sensitive keywords related to public health. The trigger keywords for this proof-of-concept extension are—'covid', 'COVID', 'covid19', 'COVID19', 'hydroxychloroquine', 'alcohol', 'corona', and 'coronavirus'. If any keyword exists, it is pre-processed further using NLP. This step is crucial to prepare the data suitable for an ML algorithm.

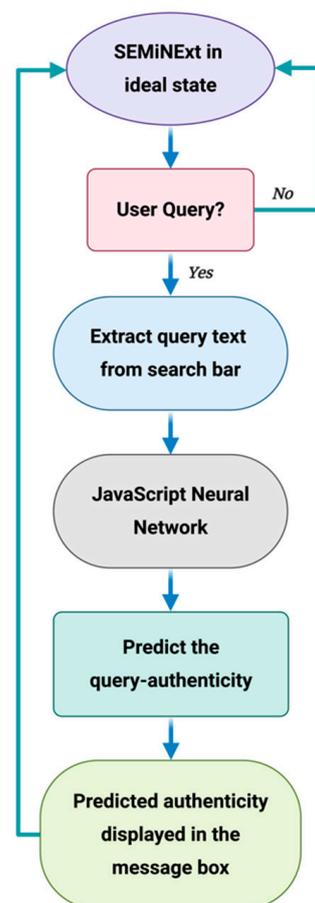


Figure 1. Data flow diagram of SEMiNExt. The illustration is created with [Biorender.com](https://www.biorender.com/).

After pre-processing, the data is then input into a trained ML algorithm which then predicts the authenticity of the original query. The validity of the user query is then displayed in a message box with additional recommendations to create awareness e.g., a hyperlink to visit the World Health Organization (WHO) for updated and authentic public health related information. The ML algorithm is implemented in a JavaScript Object Notation (JSON) and integrated within the extension.

There are several machine learning steps behind the process of determining the authenticity of the query. At first, we test and compare the performances of the mentioned machine learning algorithms using python. The results show the highest accuracy for the Neural Network (NN) algorithm. Based on this, we have integrated Artificial Neural Network (ANN) into our browser extension using the built-in JavaScript focused machine learning library called Brain.js. This is a graphics processing unit (GPU) accelerated library of Neural Networks written in JavaScript for browsers. Brain.js allows training of a NN in JavaScript and then predicts based on the training outcome. The data used for training (see Section 2.2.3) has been saved in a separate JavaScript Object Notation (.json) file.

The NN in Brain.js is defined by three parameters: activation function, number of hidden layers with their respective sizes, and the learning rate. The activation function is a mathematical operation attached to each node in the network that determines the output of the NN. The function activates its corresponding node based on whether each node's input is relevant for the prediction. The output of each node is normalized between -1 and 1 . We employed the sigmoid activation function and the backpropagation method to train the nodes of our network. The learning rate determines the network tuning in response to the estimated error each time the model weights are updated. Once a user searches for anything online, a URL is generated automatically. The user query is an integral part of the URL. The extension uses the URL extractor to copy the generated URL. Then by using the content script, we can extract the query string of the user from the generated URL. When the query string has any of the trigger words, the extension Application Programming Interface (API) calls the .json file where the NN is implemented. The overall procedure transpires in real-time. The extension we developed will also work on Bing, Duckduckgo, and Yahoo! web search engines. In addition, a privacy notice is integrated with the extension (Figure 2). This informs the user about the exchange of data with the machine learning model and ensures that no personal identifiable information will be collected.

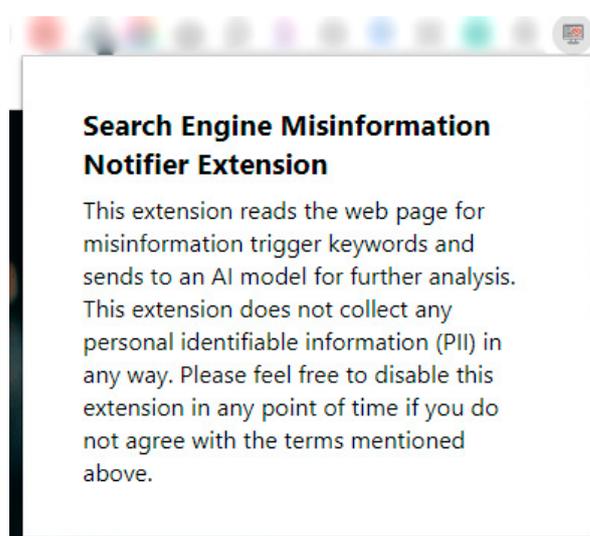


Figure 2. Privacy notice.

2.2. Natural Language Processing

2.2.1. Text Pre-Processing:

At first, we clean all the text samples by removing any punctuation marks e.g., full stop, comma, question mark, etc. to get samples composed of only words. In the second step, all the upper-case letters are converted to their respective smaller cases. The third step involves the removal of words that are not relevant for prediction, such as “the”, “that”, “in”, preposition, grammatical articles, stop words (e.g., “is” and “are”), etc. Now only the words remain that are relevant for prediction, but the inflection remains which may exist in a variety of forms either as prefixes or suffixes. To eliminate the inflections, we first split each sample sentence into its constituent words to remove the unnecessary characters, a process known as stemming. Finally, the words are stitched back to form a sentence.

2.2.2. Bag-of-Words (BoW) Model

ML algorithms only work with numbers and cannot recognize raw text. To circumvent this issue, the text must be translated into a vector of numbers that reflect various linguistic properties of the text. This is where the BoW model steps in. The idea is that texts with similar content are alike and from the content alone, we can learn about the meaning of the text. The BoW model is a text classification method where the occurrence of a word in a sample and its multiplicity are used as a feature to train a classifier. Other information such as grammar, the structure of the words, and position in a text are disregarded [31]. For an example if we have a sample: “John likes statistical tools. He likes neural network too.” it will be converted into a set of distinct words: [“John”, “likes”, “statistical”, “tools”, “He”, “neural”, “network”, “too”] and based on the frequency of the words the corresponding vector is generated: [1 2 1 1 1 1 1]. Once all the samples in the data set are transformed, we have a matrix of numbers that is now suitable to use in any ML algorithm.

2.2.3. Machine Learning Algorithms:

To train various ML algorithms, we first created a data set. Each sentence of a search query is considered as a single sample and our data set holds about in total 130 samples. The samples were inspired by basic questions found in popular social media platforms such as Twitter, Facebook, and various news articles.

In our study, we have implemented the following ML algorithms on our dataset: <https://github.com/ashiqur-rony/search-engine-misinformation-notifier-extension/tree/v1.0.0>.

- Logistic Regression: This classifier computes the relationship between the target class (which we want to predict) and other independent variables. It is then used to estimate the probability of an outcome using a logistic function also known as sigmoid function. Therefore, the output of the function is limited between 0 and 1.
- K Nearest Neighbors (KNN): This method uses the Euclidean distance as the similarity measure. It stores all the input samples then computes the Euclidean distances of a new case relative to all the input samples. Next, K closest neighbors are used for prediction.
- Support Vector Machine (SVM): This is mostly used as a binary classifier i.e., categorize the input samples into true and false classes. At first, a mathematical hyperplane/line is introduced into the variable space. Then, the best possible coefficients are found for which the hyperplane establishes a maximum separation between the classes. The prediction is then based on the location relative to this hyperplane.
- Naive Bayes: This method computes the likelihood of every words in the sentence to appear in the true and false classes. These probabilities are used to predict the authenticity of a given sentence.
- Decision Tree: In this algorithm, a decision tree is constructed. Then based on the top-down approach, the entire input samples are searched to test every attribute at each node. Next, entropy and information gain are calculated to identify which attribute to test at each node. This information is then used for prediction.

- **Random Forest:** This is an ensemble ML algorithm. Random forest combines bootstrap aggregation and random feature selection to construct a collection of decision trees to predict the final output.
- **Artificial Neural Network (ANN):** In this method, there is an input and an output layer. In between, there can be one or multiple layers known as hidden layers. Every layer contains at least one node and all the nodes of layer y is interconnected to every node at layer $y - 1$ and layer $y + 1$. ANN learns by adjusting the weights of every nodes iteratively via backpropagation to predict an outcome.

2.2.4. K-Fold Cross-Validation

To use an ML algorithm, we must split the input data into training and testing sets. The size of the training dataset must be maximized to achieve the best learning outcome and test data size must be optimized for an accurate prediction. More test samples mean fewer data available for the learning process and vice versa. To overcome this tradeoff, a cross-validation approach is generally employed. Cross-Validation is a statistical method of evaluating learning algorithms to estimate the error in the prediction of a model by dividing the input data into two sections: one part is used to train a model and the other is used to validate the model. In this study, we have used the K-fold cross-validation approach [32]. Here, the input dataset is randomly partitioned into K disjoint subsets (referred to as folds) of approximately equal size with no overlap between two subsets. Kohavi recommended 10-fold cross-validation for real-world data sets [33]. So, if the dataset has a total of 130 samples, then for 10-fold cross-validation, the dataset is divided into 10 equal subsets each with 13 samples. Then, 10 separate (K , in general) learning experiments are carried out. In each run, one subset is selected as the testing set, and the remaining $K-1$ subset is used to train the classifier. The testing set is then used to validate the trained ML algorithm. This procedure is repeated 10 times (K times in general) with 10 different testing and training sets and the results of all the experiments are averaged. In this process, the assessment of the classifier is more accurate, and all the samples are in turn used for both training and testing.

2.2.5. Classification Parameters

To measure and compare the performance of various algorithms, we need to compute several parameters, namely precision, recall, accuracy, and $F1$ score. All these classification parameters can be calculated from the confusion matrix [34]. Confusion matrix is a table that visually illustrates the performance of an algorithm and provides an insight into the detail diagnosis of the model outcomes. The rows of the matrix represent the predicted classes and the columns represent the actual classes. It depicts the number and types of errors made by an algorithm. A typical confusion matrix is shown by Table 1. True Positive (TP)/True Negative (TN) are the samples that are correctly identified by the ML algorithm, and False Positive (FP)/False Negative (FN) are the samples that are incorrectly identified. These values are used to calculate the classification parameters.

Table 1. Example of a confusion matrix.

Predicted \ Actual	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Accuracy can be defined as the ratio of the correctly predicted outcomes over the total number of predictions. A drawback of this parameter is that it does not consider the subtleties of class imbalances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Recall/Sensitivity is the portion of the actual positives the classifier picks up correctly among the total test data set.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

Precision is the portion of correct positive prediction over the total positive predictions made by the classifier.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

F1-score measures the accuracy of a model on a given dataset and is calculated as the harmonic mean of the precision and recall. It is a better measure parameter, especially in the presence of an uneven class distribution e.g., many actual negatives. An F1-score of 1 indicated a perfect precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

2.2.6. Parameters of the Implemented Machine Learning (ML) Algorithms

For the proper outcome of the ML algorithm, we followed the parameters listed below:

- Logistic Regression.
- KNN: K = 10 nearest neighbors are used for prediction. Minkowski distance metric is used with Euclidian distance as the power parameter.
- SVM: Linear kernel is used.
- Naive Bayes: Gaussian classifier is implemented.
- Decision Tree: Entropy criterion is used in the decision tree classifier.
- Random Forest: A total number of 300 trees is used with the entropy criterion.
- ANN: Total 3 layers i.e., input, hidden and an output layers with 44, 22, 1 number of nodes respectively.

3. Results

3.1. Machine Learning (ML) Predicts Real-Time Misinformation

Figure 3a–d illustrates the accuracy, F1 score, precision and recall values of the ML algorithms w.r.t the size of the input data size used for training. ANN demonstrates the highest accuracy, F1 score and recall. Its performance decreases slightly but remains stable with a change in the data size. In general, the other statistical methods experience a slight increase in the accuracy and F1 score with the increase in the data size. This may be because a larger training data (pilot data set) size makes these algorithms learn and predict better. At all cases, Naive bayes obtained a precision of 100% but it suffers from a relatively lower recall value that reduces its prediction accuracy and F1 score.

The overall performances between the statistical methods and ANN is shown in Figure 4. A 10-fold cross validation was implemented to generate the results for the statistical algorithms. Since cross validation approach is not applicable for ANN algorithm, in this case, 80% of data was split for training while the remaining 20% was used for the validation. This process was repeated several times with random splitting of the input sample and the results are then averaged for a better estimation of its performance. Figure 4 illustrates that ANN achieves comparatively the highest accuracy of 93% and a F1 score of 92%. A high F1 number is important when there is an imbalanced dataset available e.g., when there are more fake news than true news. Since F1 score is the harmonic mean of the precision and recall, it is less affected by the extreme values. Therefore, it is a better performance indicator for an imbalanced dataset. So, ANN with the highest F1 score suggests that it can handle the data imbalances much better giving a more accurate prediction in classifying a given misinformation.

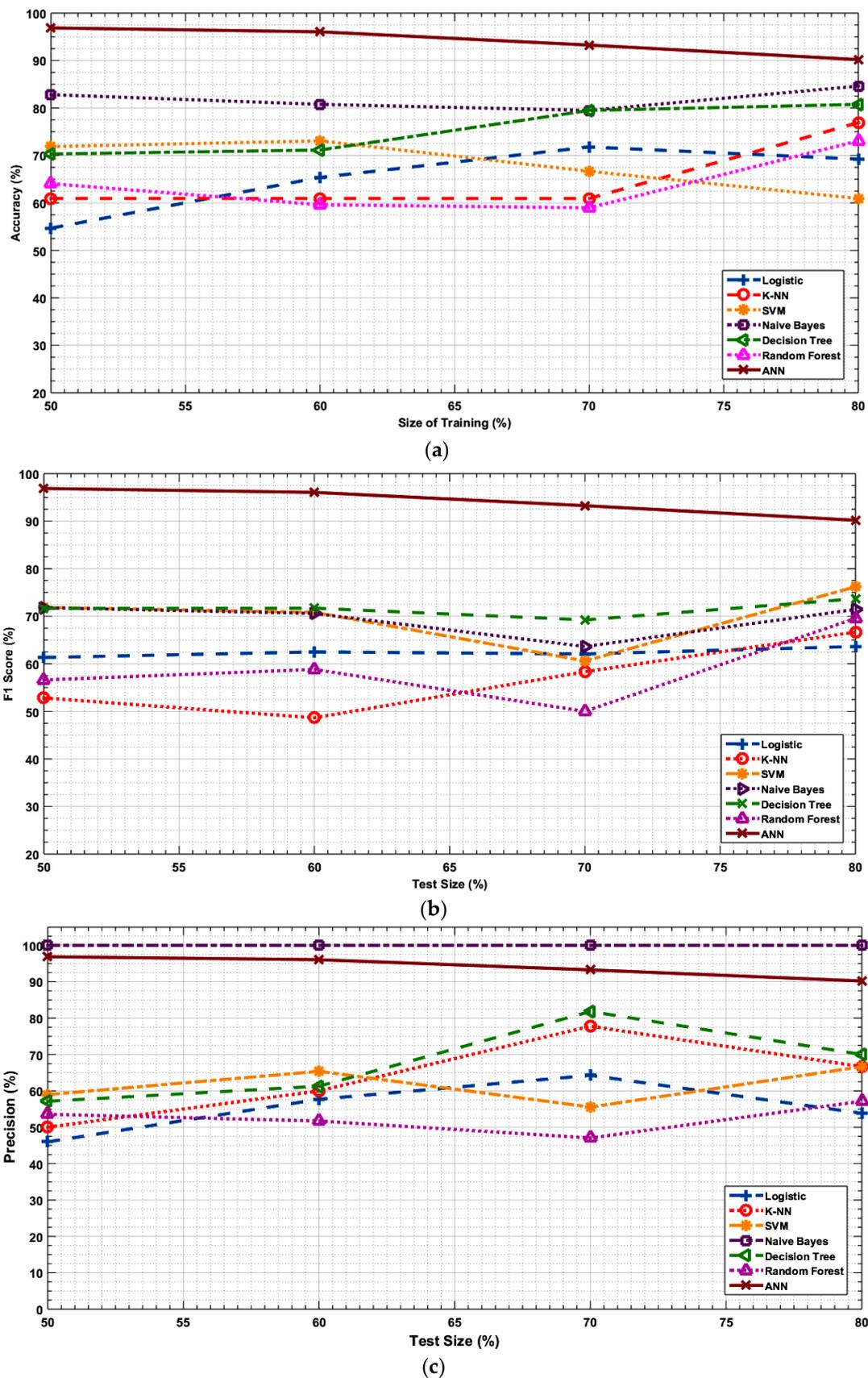


Figure 3. Cont.

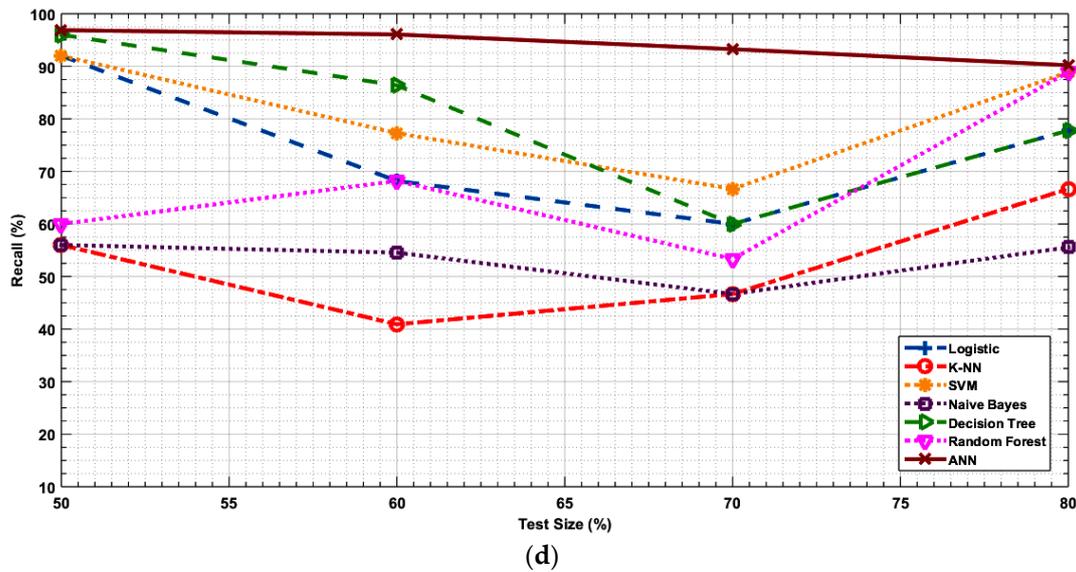


Figure 3. (a). Accuracy vs. size of the training set as a percentage of the total input data set used for training. (b). F1 score vs. size of the training set as a percentage of the total input data set used for training. (c). Precision vs. size of the training set as a percentage of the total input data set used for training. (d). Recall vs. size of the training set as a percentage of the total input data set used for training.

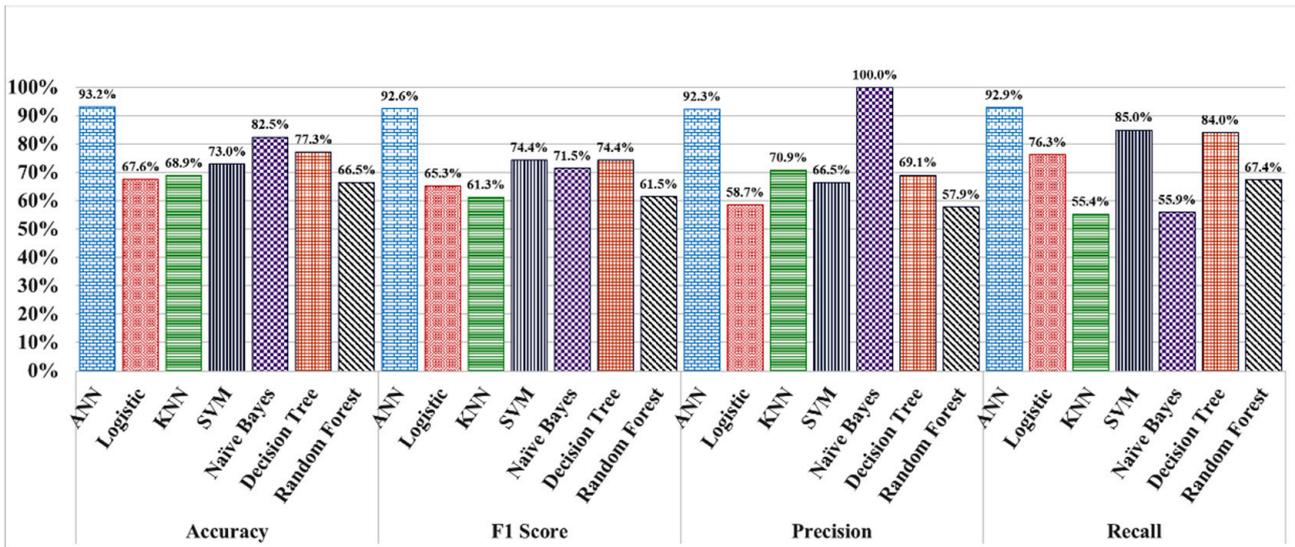


Figure 4. Performance comparison between the statistical methods and the artificial neural network (ANN).

3.2. SEMiNExt Notifies Real-Time Misinformation and False Health-Related News

The SEMiNExt works as a browser extension that silently observes the search queries on different search engines. With the help of a machine learning algorithm, then it generates a message for the user to indicate the query’s authenticity. The extension is activated only when any of the predefined keywords are present in the search query.

The trigger keywords for SEMiNExt are—‘covid’, ‘COVID’, ‘covid19’, ‘COVID19’, ‘hydroxychloroquine’, ‘alcohol’, ‘corona’, and ‘coronavirus’. When any trigger keywords are present in the user query, the extension shows a small popup at the right side of the browser with query authenticity. Additionally, the message box suggests the user to visit the WHO website through the provided hyperlink (Orange boxes in Figure 5). The proposed novel SEMiNExt works with high accuracy in the Google Chrome browser. As

displayed in Figure 5, the screenshot below, a popup with the query authenticity and trigger keywords is displayed along with a hyperlink to the WHO website.



Figure 5. In this portion, the overall performances between the statistical methods and ANN is shown.

4. Discussion

We have proposed and demonstrated a real-time web search engine misinformation notifier extension (SEMiNExt) that uses an integrated ML algorithm to predict the accuracy of the user query. When a user inquires for a piece of health-related information in the online search engine, SEMiNExt identifies the sensitive keywords, predicts the query accuracy using ML algorithm, and triggers a message box to the user regarding the authenticity

of the searched query. It also suggests that the user visit the official website of the World Health Organization via a hyperlink for raising more awareness. Our results show that SEMiNExt under Artificial Neural Network works best with an accuracy of 93%, *F1*-score of 92%, the precision of 92%, and a recall of 93% when 80% of the data is trained. Moreover, ANN can predict with very high accuracy even for a small training data size. This is very important for the early detection of new misinformation from a small data sample available online that can significantly reduce the spread of misinformation in its infancy and maximize public health safety.

The SEMiNExt utilization will be vital for the ongoing second wave and mass panic about different COVID-19 vaccines' updates. On 20 December 2020, the American Medical Association (AMA) has requested that social media companies combat vaccine misinformation [35]. Already the social media platform Instagram's leading management said months ago that they would take measures to slow down the spread of any anti-COVID vaccine-related misinformation and fake news. However, several vaccine-related contents were reported by different news agencies [36]. Surprisingly, it has increased to spread virally in the Instagram platform, which created a lot of alarm for the public health officials and government authorities [37].

Therefore, our proposed novel protocol can be utilized as a useful tool for obstructing the spread of new circulating false information. We can observe from Figure 4 that unlike the statistical methods, ANN can predict with relatively very high accuracy even for the availability of smaller data size. This is very important, especially in predicting new misinformation that started making its rounds online very recently and has minimal related data available online. Therefore, using ANN, the early detection of potentially new rumors from a small data sample can be possible with high accuracy. In this regard, SEMiNExt can intervene and significantly reduce the spread of misinformation and maximize public health safety before the fake news escalates into a significant panic amongst the population.

The current pandemic has caused hospital ICUs and emergency departments to become overwhelmed by untenable patient volumes and care requirements [38]. Any novel viruses cause intense anxiety among people searching for a rapid cure. In this circumstance, false treatments may appear online, encouraging someone to try it against scientific validations. Such extra patients can quickly saturate hospital beds and other medical equipment during an ongoing pandemic. When searching for cures online against coronavirus, SEMiNExt validates the user query in real-time to warn the user against the pitfalls of such false treatments. This can prevent the overload of hospitalizations from non-COVID patients. It would also alert the healthcare authorities about the potential regions of heavy misinformation related online searches and alarm the local authorities to take measures, even advising the hospitals about any rush in the emergency.

5. Conclusions

We have mentioned earlier that during any disease outbreak, the health care system faces massive challenges for online misinformation and fake news. SEMiNExt can successfully prevent the depletion of healthcare resources: in novel disease outbreaks, resource management in healthcare can be overwhelming because the demand is dynamic. Our proposed method can be extended to real-time healthcare resource management. At the onset of the COVID-19 pandemic, a strict lockdown was implemented to slow down the virus's widespread. This caused anxiety amongst the general population about food availability, medical resources, and other necessities over the lockdown period. As a result, a private accumulation of common medical resources such as medical masks, oxygen cylinders, etc., was observed. Later, a shortage of resources was experienced in many hospitals when the number of critical patients started rising exponentially. This was a familiar scenario in most of the developing countries. When inquiring online about the essential medicines or medical equipment that are more critical to the hospitals, SEMiNExt can inform and discourage the user against a private stockpile. This can ensure that the medical inventories are never understocked with healthcare resources.

The SEMiNExt approach has introduced the possibility of improving the online health communication system by showing misinformation notifications in real-time, enabling safer web-based searching while inquiring on health-related issues. SEMiNExt is designed to function properly in multiple web search engines platforms, such as Google, Bing, Duckduckgo, and Yahoo. Additionally, in this study, only English has been chosen as the natural language; additional natural languages, such as (by the descending order, according to the speaking population) Spanish, Bengali, Russian, German, French, and Italian, will be integrated to the extension in the future.

Author Contributions: E.H.A. and A.B.S. initiated and conceived the study. A.B.S., E.H.A., A.R. and M.M.S.R. planned, and designed the study. A.B.S., M.M.S.R. and R.B.P. conducted the machine learning experiments and acquired the data. A.R. developed the extension and exchanged data with the machine learning classifier. A.B.S., E.H.A., N.S. and R.K. interpreted the results. A.B.S., E.H.A., A.R., M.M.S.R., N.S., R.B.P., M.R.H. and S.M. drafted the first version of the manuscript. All authors critically reviewed and edited the manuscript. A.B.S. and E.H.A. share equal authorship as first author. A.R. and M.M.S.R. share equal authorship as second author. A.B.S., E.H.A. and R.K. supervised the project. A.B.S. and E.H.A. takes full responsibility for the contents of the publication as the corresponding author. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The novel search engine misinformation notifier extension (SEMiNExt) introduced and presented in this original paper have been stored in the GitHub repository: <https://github.com/ashiqur-rony/search-engine-misinformation-notifier-extension/tree/v1.0.0>.

Conflicts of Interest: The authors declare no conflict of interest.

References

- García-Magariño, I.; Sarkar, D.; Lacuesta, R. Wearable Technology and Mobile Applications for Healthcare. Available online: <https://www.hindawi.com/journals/misy/2019/6247094/> (accessed on 23 December 2020).
- Rodgers, K.; Massac, N. Misinformation. *J. Public Heal. Manag. Pract.* **2020**, *26*, 294–296. [[CrossRef](#)]
- Southwell, B.G.; Niederdeppe, J.; Cappella, J.N.; Gaysynsky, A.; Kelley, D.E.; Oh, A.; Peterson, E.B.; Chou, W.-Y.S. Misinformation as a Misunderstood Challenge to Public Health. *Am. J. Prev. Med.* **2019**, *57*, 282–285. [[CrossRef](#)]
- Na, K.; Garrett, R.K.; Slater, M.D. Rumor Acceptance during Public Health Crises: Testing the Emotional Congruence Hypothesis. *J. Heal. Commun.* **2018**, *23*, 791–799. [[CrossRef](#)] [[PubMed](#)]
- Mian, A.; Khan, S. Coronavirus: The spread of misinformation. *BMC Med.* **2020**, *18*, 1–2. [[CrossRef](#)]
- Samaan, G.; Patel, M.; Olowokure, B.; Roces, M.C.; Oshitani, H.; The World Health Organization Outbreak Response Team. Rumor Surveillance and Avian Influenza H5N1. *Emerg. Infect. Dis.* **2005**, *11*, 463–466. [[CrossRef](#)] [[PubMed](#)]
- Sell, T.K.; Hosangadi, D.; Trotochaud, M. Misinformation and the US Ebola communication crisis: Analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. *BMC Public Heal.* **2020**, *20*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Pillai, S.; Siddika, N.; Apu, E.H.; Kabir, R. COVID-19: Situation of European Countries so Far. *Arch. Med. Res.* **2020**, *51*, 723–725. [[CrossRef](#)] [[PubMed](#)]
- Isha, S.N.; Ahmad, A.; Kabir, R.; Apu, E.H. Dental Clinic Architecture Prevents COVID-19-Like Infectious Diseases. *HERD Heal. Environ. Res. Des. J.* **2020**, *13*, 240–241. [[CrossRef](#)]
- Ding, D.; Cruz, B.D.P.; Green, M.A.; Bauman, A.E. Is the COVID-19 lockdown nudging people to be more active: A big data analysis. *Br. J. Sports Med.* **2020**, *54*, 1183–1184. [[CrossRef](#)]
- Ahmad, A.R.; Murad, H.R. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. *J. Med. Internet Res.* **2020**, *22*, e19556. [[CrossRef](#)]
- Cinelli, M.; Quattrocchi, W.; Galeazzi, A.; Valensise, C.M.; Brugnoli, E.; Schmidt, A.L.; Zola, P.; Zollo, F.; Scala, A. The COVID-19 social media infodemic. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Gallotti, R.; Valle, F.; Castaldo, N.; Sacco, P.L.; de Domenico, M. Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics. *Nat. Hum. Behav.* **2020**, *4*, 1285–1293. [[CrossRef](#)] [[PubMed](#)]
- Bin-Naeem, S.; Bhatti, R.; Khan, A. An exploration of how fake news is taking over social media and putting public health at risk. *Heal. Inf. Libr. J.* **2020**. [[CrossRef](#)]

15. Soltaninejad, K. Methanol Mass Poisoning Outbreak, a Consequence of COVID-19 Pandemic and Misleading Messages on Social Media. *Int. J. Occup. Environ. Med.* **2020**, *11*, 148–150. [CrossRef] [PubMed]
16. Arafat, S.M.Y.; Kar, S.K.; Marthoenis, M.; Sharma, P.; Apu, E.H.; Kabir, R. Psychological underpinning of panic buying during pandemic (COVID-19). *Psychiatry Res.* **2020**, *289*, 113061. [CrossRef] [PubMed]
17. Liu, M.; Caputi, T.L.; Dredze, M.; Kesselheim, A.S.; Ayers, J.W. Internet Searches for Unproven COVID-19 Therapies in the United States. *JAMA Intern. Med.* **2020**, *180*, 1116–1118. [CrossRef] [PubMed]
18. Vosoughi, S.; Roy, D.; Aral, S. The spread of true and false news online. *Science* **2018**, *359*, 1146–1151. [CrossRef]
19. Ball, P.; Maxman, A. The Epic Battle against Coronavirus Misinformation and Conspiracy Theories. Available online: <https://www.nature.com/articles/d41586-020-01452-z> (accessed on 23 December 2020).
20. Gautret, P.; Lagier, J.C.; Parola, P.; Hoang, V.T.; Meddeb, L.; Mailhe, M.; Doudier, B.; Courjon, J.; Giordanengo, V.; Vieira, V.E.; et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: Results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* **2020**, *56*, 105949. [CrossRef]
21. Cavalcanti, A.B.; Zampieri, F.G.; Rosa, R.G.; Azevedo, L.C.; Veiga, V.C.; Avezum, A.; Damiani, L.P.; Marcadenti, A.; Kawano-Dourado, L.; Lisboa, T.; et al. Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *N. Engl. J. Med.* **2020**, *383*, 2041–2052. [CrossRef]
22. Jalabneh, R.; Zehra, S.H.; Pillai, S.; Hoque, A.E.; Hussein, M.R.; Kabir, R.; Arafat, S.M.Y.; Azim, M.M.A. *Use of Mobile Phone Apps for Contact Tracing to Control the COVID-19 Pandemic: A Literature Review*; Social Science Research Network: Rochester, NY, USA, 2020.
23. Jain, A.; Shakya, A.; Khatter, H.; Gupta, A.K. A smart System for Fake News Detection Using Machine Learning. In Proceedings of the 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 27–28 September 2019; pp. 1–4.
24. Zhang, J.; Dong, B.; Yu, P.S. Fakedetector: Effective Fake News Detection with Deep Diffusive Neural Network. In Proceedings of the 36th IEEE International Conference on Data Engineering (ICDE 2020), Dallas, TX, USA, 20–24 April 2020; pp. 1826–1829.
25. WhatsApp IFCN Fact Checking Organizations on WhatsApp. Available online: <https://faq.whatsapp.com/general/ifcn-fact-checking-organizations-on-whatsapp/?lang=fb> (accessed on 23 December 2020).
26. WhatsApp WHO Health Alert Brings COVID-19 Facts to Billions via WhatsApp. Available online: <https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp> (accessed on 23 December 2020).
27. Facebook Working to Stop Misinformation and False News. Available online: <https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news> (accessed on 23 December 2020).
28. Google Find Fact Checks in Search Results—Google Search Help. Available online: <https://support.google.com/websearch/answer/7315336?hl=en> (accessed on 23 December 2020).
29. Selig, R.A. Toolbar/Sidebar Browser Extension 2012. U.S. Patent 8,255,824, 28 August 2012.
30. StatCounter Browser Market Share Worldwide. Available online: <https://gs.statcounter.com/browser-market-share> (accessed on 23 December 2020).
31. Goldberg, Y. Neural Network Methods for Natural Language Processing. *Synth. Lect. Hum. Lang. Technol.* **2017**, *10*, 1–309. [CrossRef]
32. Berrar, D. Cross-Validation. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, UK, 2019; pp. 542–545. ISBN 978-0-12-811432-2.
33. Kohavi, R. *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*; Morgan Kaufmann: Burlington, MA, USA, 1995; pp. 1137–1143.
34. Benfenati, E. *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*; Elsevier: Amsterdam, The Netherlands, 2007; ISBN 978-0-444-52710-3.
35. AMA Urges Social Media Companies to Combat Vaccine Misinformation. Available online: <https://www.ama-assn.org/press-center/press-releases/ama-urges-social-media-companies-combat-vaccine-misinformation> (accessed on 24 December 2020).
36. Harvard University Instagram Remains a Hotbed of Vaccine Misinformation. Available online: <https://www.hsph.harvard.edu/news/hsph-in-the-news/instagram-vaccine-misinformation> (accessed on 21 December 2020).
37. Cook, J. Instagram’s Search Results for Vaccines Are A Public Health Nightmare. Available online: https://www.huffpost.com/entry/instagram-promoting-anti-vax-anti-vaccine_n_5e347c50c5b69a19a4aede0c (accessed on 21 December 2020).
38. Mareiniss, D.P. The impending storm: COVID-19, pandemics and our overwhelmed emergency departments. *Am. J. Emerg. Med.* **2020**, *38*, 1293–1294. [CrossRef] [PubMed]