The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

ANALYSIS OF SOCIO-ECONOMIC IMPACT OF COVID-19 THROUGH PUBLIC OPINION MINING USING THE SEMANTIC WEIGHING DEEP NEURAL NETWORK ARCHITECTURE AND ALGORITHM

SHIVANI YADAO[1, a*], Dr. A. VINAYA BABU[2, b], Dr. MIDHUNCHAKKARAVARTHY JANARTHANAN[3, C] and Dr. AMIYA BHAUMIK[4, d]

[1] Department of Computer Science and Engineering, Lincoln University College, Selangor, D.E., Malaysia.
[2] Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women Hyderabad, India.
[3] Department of Computer Science and Multimedia, Lincoln University College, Selangor, D.E., Malaysia.
[4] Lincoln University College, Selangor, D.E., Malaysia.
E-mail: [a,*]shivaniyadao123@gmail.com (Corresponding author), [b]Avb1222@gmail.com

Abstract:

Web mining has made it easier to extract application-oriented intelligible data from the vast amount of data available on the internet. There is a new field called web mining that was born out of data mining. Web mining, as opposed to traditional data mining, aims to find patterns in data that has been made available to the public. Deep learning makes it much easier to recognize patterns. Deep learning operates in the same manner as the human brain does when it comes to anticipating outcomes from a large amount of data. A key focus is on developing mathematical models that can detect patterns fast. WCM, WSM, and WUM (web usage mining) are only a few of the web-mining approaches that have been developed in the past several years. To better understand how the epidemic affects society's economic standing, this initiative uses a reputation system to gather information from researchers and economists around the world. We deal with the covid- 19 data collection from twitter because it is a hub of varied public viewpoints. The socio-economic status of tweets about the Covid-19 dataset can be determined using a reputation system. For the reputation system, this research proposes a framework in which web mining is accomplished using a semantic augmented deep neural network technique.

Keywords: Web mining, data mining, deep learning, socio-economic, semantic, neural networks, and reputation system.

1. INTRODUCTION

Understanding a website or an application's global popularity is known as a Reputation System, and the phrase refers to this understanding. For any business to be successful, especially one that is online, it is critical to look into the trustworthiness of the company's customer reviews. Many organizations, from e-commerce to e-learning to data analysis to socioeconomic research all rely on internet reputation systems to ensure that their services are reliable and effective. The lack of mining optimization efficiency has made online reputation systems a target for

ISSN 1533-9211

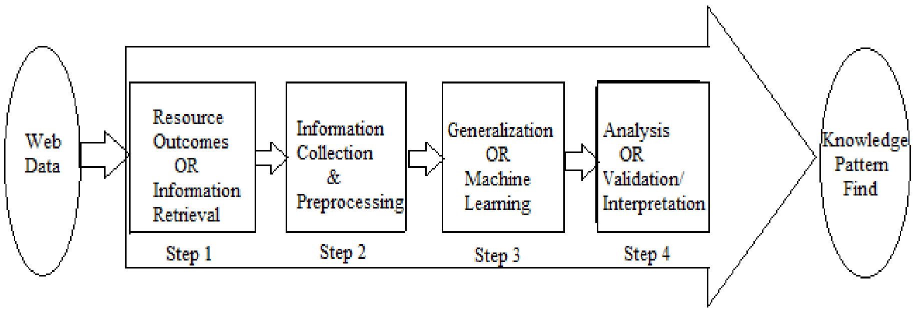
DOI 10.5281/zenodo.7025053

many critics. In the event that a user is still dissatisfied with the system's outcome, they are more likely to leave. Web data models have become increasingly inaccurate, resulting in client attrition and threatening online firms with substantial revenue losses, according to numerous writers and articles. Traditional machine learning algorithms can be used to classify opinion mining data. However, the usage of deep neural networks in this article distinguishes it from others of its kind because the computer learns on its own with the least amount of human intervention. Web mining and deep learning are used together to produce a new model distinct from the previous one. For the covid-19 twitter data sets, this research proposes a model in which the reputation system is found.

There are a lot of things that people do today that don't involve going to the store or banking: they buy things online, and they use the internet for almost every part of their daily lives. Most businesses, medical fields, educational institutions, banks, and even office gatherings are relying on internet resources. The web makes an enormous amount of knowledge readily available to everyone with an internet connection. Web mining, as a result, is a significant area of study. For many years, data mining was a subset of web mining. The process of uncovering patterns in data that are relevant to a specific client's needs is known as data mining. Web mining, as opposed to data mining, focuses on the patterns that can be found in publicly available information on the web [1, 2, and 3]. Each of these types of mining is called a "web mine." They are all called "web content mining," "web structure mining," and "web use mining". Web Content Mining [7, 8] is a procedure that allows people to search websites for text, images, and videos. Web Content Mining is the term for this procedure. A web or application server hosts a large number of web pages. Web Structure Mining looks for the links between web pages in order to learn more about how the pages are put together. Other than that, Online Usage Mining is concerned with web users or clients' utility. This makes it easier to identify the most common searches and areas of interest for a user.

All three web mining methods follow a similar set of mining procedures. Subtasks of mining techniques are depicted in Figure 1.

Figure1: Subtasks regarding Web Mining



Step 1: It's a preliminary step in which data is retrieved or web guidance is retrieved for this preliminary subtask.

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

Step 2: In the second step, data is preprocessed before it can be used. Preprocessing is nothing more than the removal of extraneous data that could cause problems later on.

Step 3: Generalization - This step focuses on getting data into a form that is easy to understand and modify.

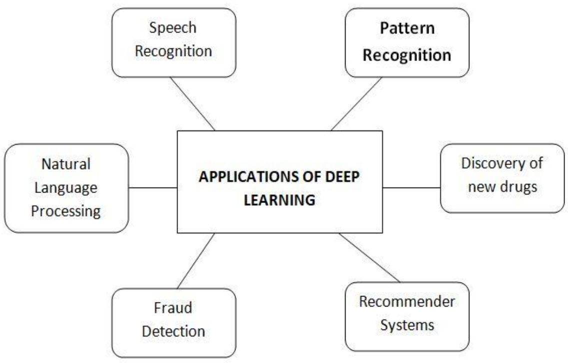
Step 4: This is the most critical phase in the process since it provides the logic for how data should be transformed into the patterns requested by the client. Eventually, a pattern of knowledge will emerge.

For each of these three separate types of web mining, a diverse set of tools and techniques are employed. Among the various web structure mining methods available are PageRank, weighted PageRank, topic-sensitive PageRank (Hits), and distance rank (SimRank).PageRank and SimRank are the most widely used structure mining algorithms. In comparison to PageRank, SimRank is a modern algorithm that utilizes a variety of strategies to provide better results. Other web mining algorithms include Page Gather, CDL4, Leader, Cobweb, and Iterate. Some of the algorithms used in online content mining include correlation algorithms, weighted page content ranks, cluster hierarchies, fuzzy c-means algorithms and building algorithms. [11].

Deep Learning with web mining

When it comes to deep learning, artificial neural networks are employed in order to examine massive amounts of data in an advanced manner. In other words, it's a form of artificial intelligence based on the way the human brain works [13]. Deep learning algorithms that learn from examples are used to train the machines. Healthcare and e-commerce are only two examples of fields where deep learning is being used. Figure 2 [14] shows an example of a deep learning app.

Figure 2: Applications in Deep Learning



When it comes to web mining, pattern detection is the key goal, and deep learning can be used for this. Where I got the idea to combine "Web mining" and "deep learning." Deep learning for reputation systems is discussed in this study, which includes sentiment analysis as a key component [15]. The term "sentiment analysis" refers to the study of how people's thoughts,

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

feelings, and attitudes are expressed in writing. For an opinionated text, sentiment analysis provides a digital representation of it. In order to derive an individual's viewpoint from a body of text, a technique known as sentiment classification must be used. The goal of sentiment analysis is to glean information about things, events, or traits from people's thoughts, judgments, and feelings about them. The polarity of a statement can be either positively or negatively polarized, or it might be neutrally polarized. Sentiment analysis is frequently used in contact centres and social media monitoring. There are two kinds of values returned by this function, one representing "polarity" and the other "subjectivity," both of which are floating-point values with ranges between [-1.0, 1] and [0.0, 1.0]. Whereas [0.0, 1.0] is the range for subjectivity, [0.0] symbolizes the very objective, and [1.0] the very subjective, [-1.0, 1] represents the very negative, and the [-1.0, 1] the very positive in polarity.

2. EXISTING SYSTEM

Over the Support Vector Machine (SVM) approach, deep neural network frameworks are compared and created [12]. Classification algorithms that use supervised learning to determine whether data is positive or negative are commonly used in these techniques. In order to use this strategy, you'll need data that has been labelled. With this method, it becomes clear that both positive and negative features of a word's local context must be considered (e.g. Very beautiful). An important starting point for developing a feature Vector is this paradigm.

1. The first step is to tag each tweet with a part of speech tagger.
2. Gather all the adjectives you can think of in the context of the full tweet.
3. Using the top N adjectives, create a group of words that are commonly used.
4. During the experimental set, go through each tweet one by one to select the one you desire. Make a list of these things:
    - Number of positive terms
    - Number of negative words
    - Presence, absence, or frequency of each phrase

As a major advantage, this system can generate and alter trained data models to meet specific needs or criteria. However, this system has a research gap when it comes to labelled data. Using only data that has been tagged is both time consuming and inflexible, as it won't work with all platforms.

The experiment and analysis for this research were conducted utilizing the Twitter data set as a whole. As a result of the lack of a semantic bag of word model module that references several meanings of a word, the system's accuracy has been diminished.

3. Semantic enhanced weighting deep neural network frame work for reputation systems

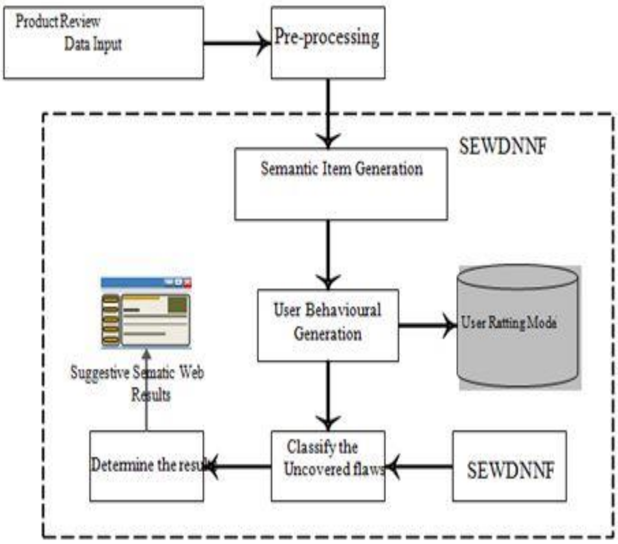
In order to create a successful website, a huge amount of data must be described and comprehended. On the basis of Covid-19's socioeconomic impact, a Twitter data ontology

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

learning model had to be used in a Deep Neural Network (DNN).Web usability is combined with semantic web expressiveness and flexibility in this meta-model [16,17,18]. Figure 3 depicts the fundamental structure of a deep neural network with semantically enhanced weighting. Deep learning is used to collect and process the words with multiple but similar meanings that exist in the paper's Semantic Item Generation module, and the technique for SWEDNNF is provided in the other module.

Figure 3: SEWDNNF Model



Making Web resources more machine-accessible and manipulable through the use of metadata that specifies Web material in a data format that is tailored to the needs of certain users would improve machine accessibility and manipulation. On the basis of the Covid-19 Twitter data collection, we'll study a number of semantic web modules.

One hundred and eighty percent of the covid-19 twitter data comes from three separate sources: Kaggle (80%), direct social media (15%), and web crawling (5%). A total of 179,000 tweets from the year 2020, when the epidemic was at its height, are included in the data collection. Figure 4 shows that there are a total of 13 columns. Among the data are: "user name," "user location," "user description," "user created," "user followers," "user friends," "user favourite," "user verified," "date, text, hash tags, source," and "is retweeted."

The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

Figure 4: Covid-19 Dataset entries

J16 | Rajasthan Government today started a Plasma Bank at Sawai Man Singh Hospital in Jaipur for treatment of COVID-19 paâ€¦ https://t.co/cwfCcWyaDA

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourite	user_verified	date	text	hashtags	source	is_retwee
2	á‰áŽ¥â˜»Č	astroworld	wednesday adda	26-05-2017 05:46	624	950	18775	FALSE	25-07-2020 12:27	If I smelled the scent of h		Twitter fo	FALSE
3	Tom Basile ć	New York, NY	Husband, Father,	16-04-2009 20:06	2253	1677	24	TRUE	25-07-2020 12:27	Hey @Yankees	@Yankee	Twitter fo	FALSE
4	Time4fisticu	Pewee Valley	#Christian #Cath	28-02-2009 18:57	9275	9525	7254	FALSE	25-07-2020 12:27	@diane3443 @	['COVID19	Twitter fo	FALSE
5	ethel mertz	Stuck in the N	#Browns #Indian	07-03-2019 01:45	197	987	1488	FALSE	25-07-2020 12:27	@brookbanktv	['COVID19	Twitter fo	FALSE
6	DIPR-J&K	Jammu and K	ðŸ–Šï¸Official Twi	12-02-2017 06:45	101009	168	101	FALSE	25-07-2020 12:27	25 July :	['CoronaV	Twitter fo	FALSE
7	ðŸŽ¹ Franz S	ÐÐ¾Ð²Ð¾Ñ€€	ðŸŽ¼ #ÐÐ¾Ð²Ð¾	19-03-2018 16:29	1180	1071	1287	FALSE	25-07-2020 12:27	#coronavirus #	['coronavi	Twitter W	FALSE
8	hr bartende	Gainesville, F	Workplace tips a	12-08-2008 18:19	79956	54810	3801	FALSE	25-07-2020 12:27	How #COVID1	['COVID19	Buffer	FALSE
9	Derbyshire LPC			03-02-2012 18:08	608	355	95	FALSE	25-07-2020 12:27	You now have to wear fa		TweetDec	FALSE
10	Prathamesh	Bendre	A poet, reiki prac	25-04-2015 08:15	25	29	18	FALSE	25-07-2020 12:26	Praying for	['covid19',	Twitter fo	FALSE
11	Member of	ðŸ‘‡ðŸ»locati	Just as the body	17-08-2014 04:53	55201	34239	29802	FALSE	25-07-2020 12:26	POPE AS GOD -	['Hurrican	Twitter fo	FALSE
12	Voice Of CBSE Students			14-07-2020 17:50	8	10	7	FALSE	25-07-2020 12:26	49K+ Covid19		Twitter W	FALSE
13	Creativegms	Dhaka,Bangla	I'm Motalib Mia,	12-01-2020 09:03	241	1694	8443	FALSE	25-07-2020 12:26	Order here:	['logo', 'gr	Twitter W	FALSE
14	SEXXYLYPPS	Hotel living -	My ink "My	25-03-2010 21:16	0	8	32	FALSE	25-07-2020 12:26	ðŸ‘‹ðŸ»@Patty	['COVID19	Twitter W	FALSE
15	Africa Youth	Africa	Official account c	13-05-2019 06:27	830	254	3692	FALSE	25-07-2020 12:26	Let's all	['COVID19	Twitter W	FALSE
16	DailyaddaaN	New Delhi	Breaking news al	22-10-2016 09:18	546	29	88	FALSE	25-07-2020 12:26	Rajasthan Government t		Twitter W	FALSE
17	Dimapur 24/	Nagaland, Inc	strive to	11-11-2019 12:02	274	32	378	FALSE	25-07-2020 12:26	Nagaland	['Covid19'	Twitter fo	FALSE
18	ChennaiCity	Now	Individual tweet	26-04-2009 09:38	3987	53	749	FALSE	25-07-2020 12:26	July 25	['COVID19	Twitter fo	FALSE
19	marc goovae	Brussels	Progressive minc	13-06-2009 13:48	283	1432	1546	FALSE	25-07-2020 12:26	Second wave c	['COVID19	Twitter fo	FALSE
20	Dorian Aur			30-01-2011 18:40	46	108	453	FALSE	25-07-2020 12:26	It is during	['light']	Twitter W	FALSE
21	Coronavirus	Florida, USA	COVID-19 Practic	03-12-2019 19:00	14	24	74	FALSE	25-07-2020 12:26	COVID Update: The infec		Twitter fo	FALSE
22	The Voice of	GenX		28-05-2011 15:28	292	1037	58	FALSE	25-07-2020 12:26	@EvanAKilgo		Twitter fo	FALSE
23	APO Group	#AFRICA #MEI	Latest #Africa & #	22-02-2011 09:09	10661	6	2037	TRUE	25-07-2020 12:26	Coronavirus - South Afric		Africa Nev	FALSE
24	Micah Pollal	Northwest In	Associate Profes	22-07-2011 13:41	751	183	1308	FALSE	25-07-2020 12:26	@JimBnntt Your image d		Twitter W	FALSE
25	CAWST	100+ countrie	Providing trainin	06-08-2009 02:43	3038	2713	7445	FALSE	25-07-2020 12:26	The first	['WASH', '	Twitter fo	FALSE

4. EXPERIMENTAL MODEL

The deep neural network framework with semantically enhanced weighting was developed by the modules listed below.

4.1. Pre-processing

During the first stage, the tweets are fetched after mining the suitable dataset, in this case, the Covid-19 twitter data. During this step, data is pre-processed, which includes deleting unnecessary or irrelevant information like as timestamps and embedded links, as well as any data that is unreadable or impossible to identify. In most cases, these outputs are of little significance and can lead to inaccurate outcomes.

4.2. Semantic Item Generation

The term "semantic" refers to the fact that a single word can have a variety of logical or linguistic connotations. To test an adjective's polarity, we establish a semantic pool containing these semantics [22, 23, 24, and 25]. If "destination" sounds like "last halt," it's because it is. Different shades of language are used based on the context in which they appear. This level also includes text classification and deep learning. This is done so that certain words and phrases may be extracted from the data set and their recurrence can be determined. We can determine the statement's polarity by looking up the word's semantic meaning in a database containing all possible semantic meanings.

In order to classify text, deep neural networks use convolution layers to construct a 1D structure from text data. Using a Deep Neural Network to categorize text requires a vector representation of data that preserves the locations of the individual words. If each word were to be represented by one pixel, then each document would be represented by an image of |D|*1 pixel in the [V] channels. In each convolution layer, an area vector that contains pixels (words) is used to build up the convolution matrix. The variable output size is what the convolution layer produces as output. Using a pooling layer, a variable-sized output was transformed into a fixed-size output.

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

The most common type of data pooling used in text is max-pooling throughout the entire dataset.

Using a Deep neural network to categorize some text is an example.. Pixels are used to symbolize each word in each document D. Vocabulary |V| was built from the papers that were provided. Each document was viewed as a single pixel in one of the |D|*1 channels. In this example, vocabulary V has two documents, D1 and D2. Now we'll develop an area vector that basically checks to see if the word from the selected document is contained inside the vocabulary or not. Write 1 if the word appears; otherwise, write 0. The following are the region vectors based on input documents D1 and D2:

Input text regarding D1 = {this is good}

Input text regarding D2 = {Doing good work}

Vocabulary: {This is doing good work}

	1	This
Region vector1:	1	is
	0	doing
	1	good
	0	work

	0	This
Region vector 2:	0	is
	1	doing
	1	good
	1	work

With this vector, we can assess how frequently a term is used in a given tweet, and then check its semantics in order to design an opinion mining system that has a high degree of accuracy.

4.3. User Behavioural Generation / User Rating Mode

From Semantic item generation, this step retrieves user information for use in this stage. User rating mode maintains a database from which this information can be accessed and retrieved.

Logistic regression is used to train the system during the creation of user behaviour. For binary input variables, logistic regression is a good choice because the polarity is usually either positive or negative. Table 3 shows that logistic regression outperforms the other two classifiers. The content in a document or a sentence can be divided into two or three distinct groups. A linear logistic regression model is frequently employed when the response variable is binary. There are two possible values for the response variable, which are zero and one. As an example, let's say we have a dataset called x that has n records. Predictor variables and

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

attributes are included in each record along with the binary answer variable. No more than 0 or 1 values are allowed for this binary outcome variable, meaning it can only have one of two potential outcomes. The goal of logistic regression is to develop a model that can predict an outcome variable based on the data that has been provided. σ(t)is a logistic function that can be defined as:

$$y = f(x) \qquad 4.1$$

$$y = \sigma(t) = \left(\frac{1}{1+e^{-t}} \right)$$

t– Linear function regarding a single explanatory variable x

$$P(y \mid x, \theta) = \sigma(\theta^T x) = \left(\frac{1}{1+e^{-\theta^T x}} \right)$$

$$y \in \{0, 1\}$$

x – Feature Vector

$x \in \theta \in R^N$ - N dimensional feature vector

θ - Parameters of the logistic regression

4.4. Classify uncovered flaws

Blanks or null values frequently appear in the statements or reviews generated since the preceding stage. A positive or negative polarity can be generated by SEWDNNF logic in this stage, which fixes the defects and adds logic related SEWDNNF to provide semantic web results.

4.5. SEWDNNF

The reader will be introduced to a method for semantically improving the weighting of deep neural networks at this point. Step two of our strategy entails determining the text's polarity, which we aim to do in this phase. Emoticon resolution is used to determine the polarity of a statement when emoticons are employed in a phrase. For example, the polarity detection may not be accurate, or the claimed sentiment may not be expressed accurately. In addition, we attempt to separate opinion phrases in a sentence in relation to the notions of the supplied sentence.

a) As a result, we educate the system to recognise the relationship between words in a variety of different situations. As a way to establish if the remark conveys a favourable or negative feeling, we count how many times the word appears in the statement.

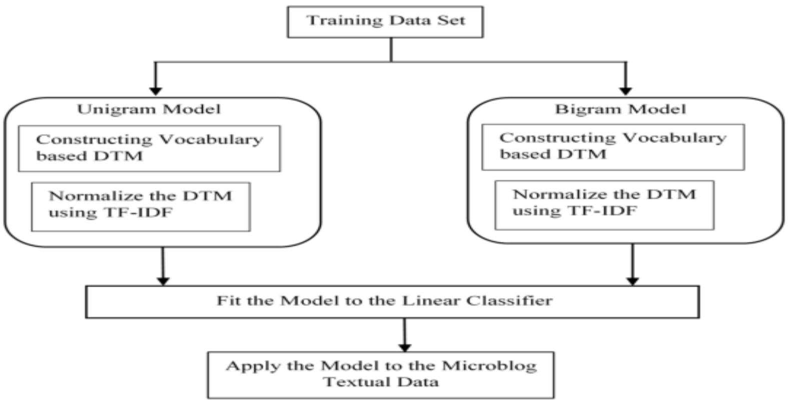
The Seybold Report
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

b) Words with polarity can be identified once the opinion words have been linked to context. Once opinion words have been identified using Context, it is possible to establish the polarity associated with those terms.

c) Our approach is trained on a huge dataset that contains a wide range of subtle and ambiguous emotions in order to aid in the detection of the concepts included in the study. This information is given in an unsupervised manner by the system.

It is possible to receive either positive or negative feedback on the default training dataset.Following the identification of the positive and negative parameters; we also consider the neutral parameter. The training dataset was thus classified into different levels using qualitative prediction during the course of this work. Among the most generally used classifiers for predicting qualitative responses, logistic regression is the most commonly encountered. For decades, the logistic regression model has been the go-to model for predicting a categorical answer. Predicted probability for the training dataset range from 0 to 1. A binary distribution is also employed in this scenario. Unigram and bigram models can be accurately trained on trained data using the TF-IDF transformation approach, as shown in figure 5, by normalizing the document-term matrix (DTM) transformation technique.

Figure 5: Process to Perform Semantic Analysis



An algorithm called TF–IDF (frequency–inverse document frequency) is used to determine the importance of a word across a collection of texts or a corpus of documents. When considering data recovery, text mining, and client demonstrating, it is commonly used as a weighing component in the decision-making process. The value of TF-IDF is directly related to the number of times a word appears in a document, according to the frequency of occurrence of that term. It's one of the most often used methods nowadays, and it's closely tied to the Term-Weighting strategy.

Recurrence of terms in a collection of reports or documents can be shown using a document-term matrix. For this matrix, the rows and columns are the same as in any other, with each row being a report or paper and each column representing a term. In the event of a transposition or term document matrix, the rows and columns are rearranged.

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

With terms in the columns and documents in rows, it generates an index of terms in a corpus regarding archives. An xy cell shows how many times a particular term appears in a report. A vector of word counts for each line addresses the content in respect to the record being compared using that column, as a result. Think on the next two (short) articles:

- R1 = "I am research scholar"
- R2 = "I am research guide",

Then the document-term matrix would be:

Table1: DTM

	I	am	research	scholar	guide
R1	1	1	1	1	0
R2	1	1	1	0	1

Even though a word's raw count is often its cell value, there are a variety of approaches to balance raw counts, including row normalization (i.e. "relative frequency/proportions") and tf-idf. The use of single words as whitespace or punctuation on each side of the term (a.k.a. unigrams). Individual word counts are kept, but the order of the words in the document is lost, leading to the term "bag of words" to describe this type of representation.

Algorithm 1: To Perform Semantic Analysis

Input: Extracted Micro-blogging data (M)

The blogged data (B) that was extracted.

Output: semantically enhanced Micro-blogged data of text = {0,1}

Begin

1. Extract the data from the data set whose characteristics are either positive or negative.
2. Generate a document term matrix $T_{D,}$ from the training dataset.
3. Create T_{DN}, a Normalized Document term matrix, by implementing normalization technique TF-IDF transformation on T_D
4. Apply Fit Function Regression R on T_{DN} for obtaining RT_{DN}
5. Design the text data of the Micro-blog B and insert it inRT_{DN} to obtain RB_{DN}
6. Return

   Semantic

   End

Algorithm 2: Dummy Variable Approach

Input: Semantic Rate of Micro-blogged Textual data that was Predicted in previous step S (RB_{DN}) = [0, 1]

The Seybold Report

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

Output: The weights in terms of polarity of semantically enhanced Micro-blogged data of text = {0, 1}

Begin

1. Import predicted Semantic rate regarding Micro-blogging Textual data in the RB_{DN} results $S(RB_{DN})$
2. If $S(RB_{DN}) > 0.5$ then
   a. S = 1
3. Else
   a. S = 0
4. Endif
5. Return S
6. End

When using logistic regression for qualitative data, the expected answer should fall into one of two categories: 0 or 1. However, the anticipated probability for micro blogging textual unseen data sits somewhere between 0 and 1 in the range. the probability are expected to occur in the following order (0-0.35), (0-0.45)- (0-0.65), and (0.65)-1 (positive-semantic-rate-ordered). Positive and neutral gaps are frequently found to occur together, as is the reverse for negative and neutral gaps.

The response to a training dataset should have two possible outcomes in order to find accuracy. The following code answer could be used to obtain a binary qualitative response using the dummy variable strategy.

Response = 0; if Predicted micro-blogging text Semantic rate is<0.5

Response = 1; if Predicted micro-blogging text Semantic rate is> 0.5.

It is necessary to create a dummy variable with a response value of 1 when the predicted semantic rate for micro blogging text is greater than 0.5, and a response value of 0 when the predicted semantic rate for micro blogging text is less than 0.5 in order to use a linear classifier model based on unseen data.

5. PERFORMANCE METRICS

The classification difficulty arises when public opinion is sorted into positive, negative, and neutral categories. [26] The accuracy of multiple classifiers, including logistic regression, SVM, and decision trees, is compared using a dataset of tweets from Covid-19. For the purpose of calculating various performance measures of the classifier, such as accuracy, recall and F measure, we calculate the confusion matrix.

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.7025053

$$\text{Accuracy} = \frac{\text{Count of accurately classified field}}{\text{Total Count in the record}}$$

The confusion matrix is utilized to update the formula below because the accuracy in the preceding formula was not adequate in terms of correctness.

Table2: Confusion matrix

Actuals	*Predicted class*	
	Positive Forecasts	Negative Forecasts
Actual Positive	*TP (Real Positives)*	*FN (Wrong Negatives)*
Actual Negative	*FP (Wrong Positives)*	*TN (Real Negatives)*

Accuracy = (TP + TN) / (TP + TN + FP +FN)

Precision (measuring exactness) = TP / (TP + FP)

Recall (measuring completeness) = TP / (TP + FN)

F1 = (2* Precision* Recall) / (Precision + Recall)

Table3: Comparison between different classifiers in terms of performance metrics

Classifier	Precision	Recall	F Score
Logistic Regression	0.94	0.99	0.96
SVM	0.92	0.98	0.92
Decision Tree	0.90	0.91	0.93

Covid-19 Twitter data was best classified using Logistic regression, as shown in table 3.

Table 4: Comparison between existing and proposed system in terms of accuracy

Research Model	Accuracy
SEWDNNF	96.80
SVM	95.70

Table 4 further shows that the suggested model, SEWDNNF, is more accurate than the standard Support Vector Machine method for Twitter Sentiment Recognition [12].

6. CONCLUSION AND FUTURE SCOPE

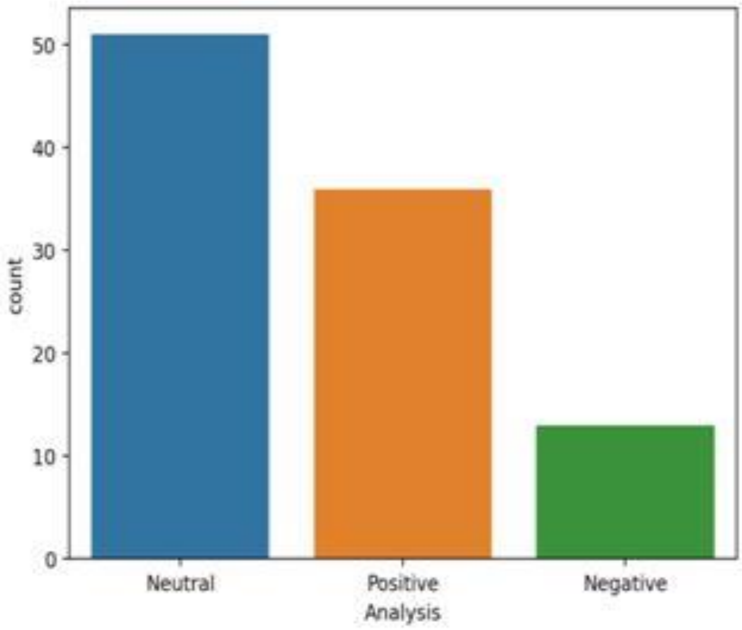
According to Twitter data analysis figure 6, most people's opinions on pandemic covid-19 were either neutral or positive by the middle of the year. People's views about combating the

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

epidemic have improved as a consequence of this study's findings, which suggest a rise in public support for the effort. Figure 7 shows the most commonly recurring words to be panic, crisis, and scam, as can be seen there. This suggests that the general public's perception of the global economic situation isn't very positive. As depicted in Figure 6, this model combines Web Mining with Deep Neural Networks to identify the strengths and weaknesses of covid-19 tweets or reviews. This model also uses online reputation networks and a semantic approach. The covid-19 dataset was used in this study to examine the influence of socioeconomic factors on public sentiment [27].

Figure 6: Analysis of covid-19 tweets in terms of polarity



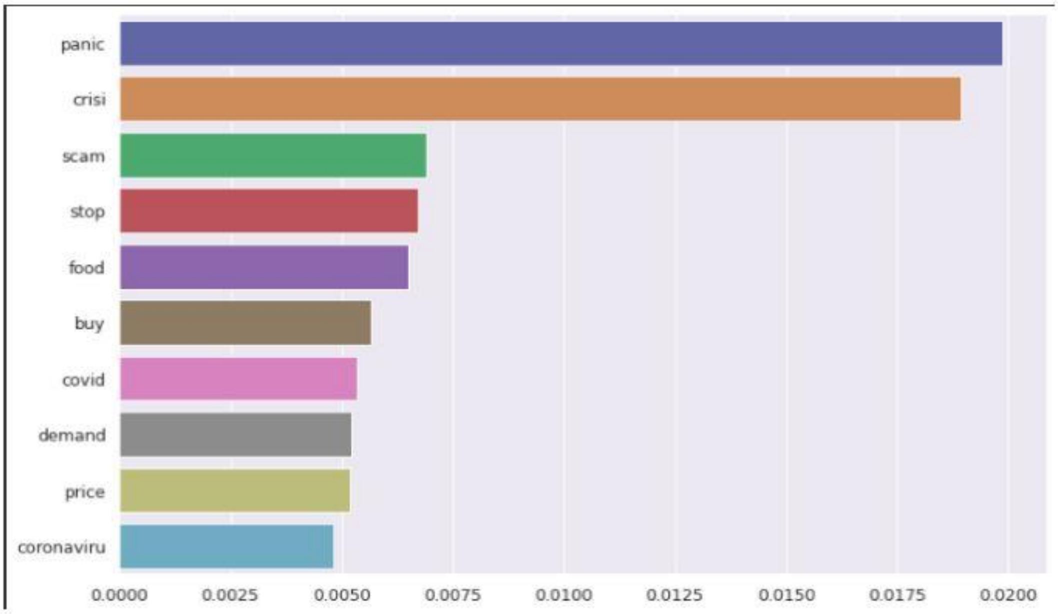
Ultimately, the goal of this research is to improve the security of computer networks. A variety of strategies exist for obtaining sensitive data from the web, both directly and indirectly. As a result, web mining security is a crucial consideration for future study.

The Seybold REPORT

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

Figure 7: Most frequently used words



References

1. V. Medvedev, O. Kurasova, J. Bernatavičienė, P. Treigys, V. Marcinkevičius,andG. Dzemyda, "A new web-based solution considering modelling data mining processes," Simulation Modelling PracticeandTheory, vol. 76, pp. 34-46, 2017.
2. S. M. Huded, S. Balutagi,andA. Ranjan, "Mapping regarding literature on data mining at j-gate database," 2019.
3. S. Taha Ahmed, R. Al-hamdani,andM. Crook, "Studying regarding Educational Data Mining Techniques," International Journal regarding Advanced Research during Science, EngineeringandTechnology, vol. 5, pp. 5742-5750, 2018
4. T. A. Al-asadi, A. J. Obaid, R. Hidayat,andA. A. Ramli, "A survey on web mining techniquesandapplications," International Journal on Advanced Science Engineeringand guidance Technology, vol. 7, pp. 1178-1184, 2017.
5. M. J. H. Mughal, "Data Mining: Web Data Mining Techniques, ToolsandAlgorithms: An Overview," guidance Retrieval, vol. 9, 2018.
6. C. Li, "Research on an Enhanced Web guidance Processing Technology based on AIS Text Mining," Recent Advances during Electrical and Electronic Engineering (Formerly Recent Patents on Electrical and Electronic Engineering), vol. 14, pp. 29-36, 2021.
7. R. H. Salman, M. Zaki,andN. A. Shiltag, "A Studying regarding Web Content Mining Tools," Al-Qadisiyah Journal regarding Pure Science, vol. 25, pp. 1-16, 2020.
8. E. T. John, B. Skaria,andP. Shajan, "An Overview regarding Web Content Mining Tools," Bonfring International Journal regarding Data Mining, vol. 6, pp. 01-03, 2016.

ISSN 1533-9211

DOI 10.5281/zenodo.7025053

9. N. PradhanandV. Dhaka, "Comparison-based study regarding pagerank algorithm using web structure miningandweb content mining," during Smart SystemsandIoT: Innovations during Computing, ed: Springer, 2020, pp. 719-729.
10. S. Shanthi, "Survey on web usage mining using association rule mining," International Journal regarding Innovative Computer Science and Engineering, vol. 4, 2017.
11. Mohd Shoaib and Ashish K. Maurya, “Comparative Study of Different Web Mining Algorithms to Discover Knowledge on the Web”, Proceedings of Elsevier Second International Conference on Emerging Research in Computing, Information, Communication and Application (ERCICA-2014), Volume-3, August 2014
12. V Uday Kumar, CMAK Zeelan Basha, M Vikas Chandra, D Sai Mahesh ,K.Anish (2019) “Twitter Sentiment Recognition using Support Vector Machine”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019
13. Biran, O.,andCotton, C. (2017). “Explanationandjustification during machine learning: a survey,” during IJCAI-17 Workshop on Explainable AI (XAI). Vol. 8, 1.
14. Cao, S., Lu, W.,andXu, Q. (2016). “Deep neural networks considering learning graph representations,” during Thirtieth AAAI Conference on Artificial Intelligence.
15. Dai, J., Wang, Y., Qiu, X., Ding, D., Zhang, Y., Wang, Y., et al. (2018). BigDL: a distributed deep learning framework considering big data. arXiv:1804.05839.
16. Guo, Y., Liu, Y., Georgiou, T. et al. A review regarding semantic segmentation using deep neural networks. Int J Multimed Info Retr 7, 87–93 (2018). https://doi.org/10.1007/s13735-017-0141-z
17. Shimoda W, Yanai K (2016) Distinct class-specific saliency maps considering weakly supervised semantic segmentation. In: ECCV
18. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning considering image recognition. In: CVPR
19. Jin B, Ortiz-Segovia MV, Süsstrunk S (2017) Webly supervised semantic segmentation. In: CVPR
20. Khoreva A, Benenson R, Hosang J, Hein M, Schiele B (2017) Simple does it: weakly supervised instanceandsemantic segmentation. In: CVPR
21. Lin G, Shen C, Van Den Hengel A, Reid I (2017) Exploring context among deep structured models considering semantic segmentation. In: IEEE transactions on pattern analysisandmachine intelligence
22. Liu Y, Guo Y, Lew MS (2017) Onexploration regarding convolutional fusion networks considering visual recognition. In: MMM
23. Liu S, Qi X, Shi J, Zhang H, Jia J (2016) Multi-scale patch aggregation (mpa) considering simultaneous detectionandsegmentation. In: CVPR
24. Lin G, Shen C, van den Hengel A, Reid I (2016) Efficient piecewise training regarding deep structured models considering semantic segmentation. In: CVPR
25. Saleh F, Akbarian MSA, Salzmann M, Petersson L, Gould S, Alvarez JM (2016) Built-in foreground/ background prior considering weakly-supervised semantic segmentation. In: ECCV
26. Sherin Mariam John, K. Kartheeban, “Sentiment Scoring and Performance Metrics Examination of Various Supervised Classifiers”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2S2, December 2019
27. Adeena Nasir, Munam Ali Shah, Ummarah Ashraf, Abid Khan, Gwanggil Jeon,An intelligent framework to predict socioeconomic impacts of COVID-19 and public sentiments,Computers and Electrical Engineering,2021,ISSN 0045-7906