

## Content Analysis of Covid-19 and Agriculture News in Bangladesh Using Topic Modeling Algorithm

M Moriom Khatun<sup>1\*</sup>, Md Saeed Siddik<sup>2</sup>, Md Abdur Rahman<sup>3</sup> and Shah Khaled<sup>4</sup>

<sup>1</sup>Department of Agribusiness and Marketing, Sher-e-Bangla Agricultural University, Bangladesh

<sup>2</sup>Institute of Information Technology, University of Dhaka, Bangladesh

<sup>3</sup>Center for Advanced Research in Science, University of Dhaka, Bangladesh

<sup>4</sup>Waterloo Management of Integrated Manufacturing Systems Lab, University of Waterloo, Canada

Received: 24 June 2020, Revised: 22 September 2020, Accepted: 5 November 2020

### Abstract

Covid-19 pandemic is an ongoing global crisis which affects all the economic sectors including the most fundamental sector, agriculture. Bangladesh, a developing country and is home to 2.11% of the world's population, is facing a massive shock in the agricultural sector which is the backbone of the country's economy. In the era of technology and having lockdown in recent pandemic, every person willingly collects information on agriculture and rural economics from publicly available online news articles. To satisfy the readers' demand, not only the online blogs and newspapers, but also printed newspapers have taken a serious look at open access articles. This research analyzed entire online articles related to Covid-19 and Bangladesh agriculture news published during this country wide lockdown after the first case was addressed. Topic modeling algorithm has been used to determine the underlying topics over these textual contents for mitigating the gap between media coverage and actual scenario. This research considered the entire set of English online articles indexed by the most prominent Google search engine to point out diverse topics which are covered by various news sources. The dataset and experimental results were statistically analyzed to define the relationship and significance among different independent variables. Numerous interesting findings have been reported in this research including the fearsome number of female agricultural reporters, high frequency topic of agricultural economics and government spokesman, significance between underlying topics and article sources, etc. It concludes with recommendations to bridge the gap between government policy and actual scenario presented by news sources on Bangladeshi agricultural sector during Covid-19.

**Keywords:** Covid-19; agriculture news; content analysis; topic modeling; clustering algorithm  
DOI.....

---

\*Corresponding author: Email: moriom@sau.edu.bd  
Tel: +8801520083662

## 1. Introduction

Covid-19 is an infectious viral disease that has created a global health emergency and a crisis in the agriculture and food sector [1, 2]. It has already grabbed the eyeballs of people all over the planet and changed communication and economic trends throughout the virtual and online systems [3]. Health experts have encouraged paperless communication and transaction during this pandemic to ensure social distancing and sanitization. That is why publicly available online contents become more and more demandable to broadcast precise information to the national and international communities. Furthermore, several newspapers have already stopped their printing operation and started fully online versions because of public demand for them to adapt to the current situation [4]. Since major business operations have been partially or fully closed down for safety reasons, locally grown and based industries e.g., agriculture, crafts, and so on, have become crucial supports for the economies of various countries. News reporters have increasingly become real pillars of democracy as they have tried to focus on the real issues and bridge the gap between policymakers and end-users. By analyzing the media contents, one may gather the real point of view about the current situation.

Bangladesh is a South Asian country. It has the 8<sup>th</sup> highest number of the global population and has been severely affected by the Covid-19 pandemic. People in Bangladesh have faced severe restrictions on their daily activities during safety lockdown. However, the majority of the country's population is directly or indirectly related to agriculture, and they are continuing with their farming, harvesting, and supplying of food to support demand and thus push along the economic cycle accordingly. However, the actual scenario of agriculture and rural economics is not positively reflected in the news in comparison with other issues, e.g., politics, media, sports, etc. [5]. Hence, rural affairs are often skipped over by policymakers, international aid groups and charities. There are some current works available on agricultural news analysis. However, these are predefined subtopics and not dynamically generated, which cannot ensure the actual underlying topics of those contents to predict the optimal results.

Newspaper contents can be a valuable source and was analyzed for wolf recolonization in France using structural topic modeling [6]. Chandelier *et al.* [6] studied media dissemination of information to detect wolf recovery in France. Since they regarded newspapers as a good source of content, they developed a statistical method for structural topic modeling that allows them to generate topics from a large number of texts about the targeted wolf recovery. They conducted two regional newspaper surveys and found a gap in representations of wolf management between citizens who had directly interacted with the wolf and who favored detailed information content. Another topic modeling application was presented by Gosh and Guha [7] for Twitter data analysis of US healthcare-related issues. The survey data, which was extracted from Twitter.com, came from obesity-related queries about childhood obesity, and intake of desserts and fast food. The data were then analyzed using statistical model and the comparative study between rural and urban areas, northern and southern states, and between coasts and inland states were determined. Bhatia *et al.* [8] presented the role of representations of the human-leopard conflict in Mumbai through media-content analysis. They conducted a content analysis of print media articles on human-leopard conflict in Mumbai, India. That research found that English-language and non-English-language print media differed significantly in their framing of human-wildlife conflicts. Another study demonstrated the coverage of the three most popular Bangladeshi domestic newspapers of the climate change issue from May 2006 to June 2009 [9]. This paper considered newspaper content as the most valuable communication research items and analyzed them to give an overview of Bangladeshi climate change concerns. They collected data in 15 predefined categories and analyzed their significance of data in those categories as observed in three domestic newspapers. It turned out that agricultural issues only featured at 8.3% of content in comparison with other climate issues.

Recently, the effects of Covid-19 on the agriculture sector have been addressed by some researches, where a Canadian Prof. Ellen Goddard presented the impact of Covid-19 on food retail and foodservice in Canada [1]. She put forward an economic model that described the Canadian food sector before the pandemic and the losses that were being incurred due to the impact of the disease. Her paper concluded with some recommended actions that should be taken by the foodservice, food retail, and government in response to remarkably uncertain projections in the future. The effect of the Covid-19 pandemic in agriculture for international trade relations was elaborately presented by Kerr [2]. The author described the short and long-run implications for the food supply chain of this pandemic. He suggested strengthening institutions that govern international trade for mitigating the post effect of this pandemic. The impact of Covid-19 on Iowa State's corn, soybean, ethanol, pork, and beef sectors was depicted by Hart *et al.* [10]. They presented the comparative study of these agricultural products with the estimated price and Covid-19 price damage. This paper was one of the first analyses that showed the slowdown in production and economic damages caused by the pandemic in the United States.

Analysis of agriculture related news only was rarely addressed by some researches, where Narayana and Kumar [5] presented a content analysis of agricultural news coverage in the daily newspapers of India. It was a study of five local daily newspapers of India and found that agriculture news was just 4.61%, where as political news was 27.3%. Ogessa and Sife [11] presented content analysis of the coverage of agricultural information in Tanzanian newspapers. They found that the prominence of agricultural information was as low as 4.9% in all the newspapers, which was too low to have any impact on and appeal to readers. They recommended that the government formulate policies to increase the coverage of development and agriculture-related news in the newspapers. Newspaper reportage and its effect on promoting agricultural development in Nigeria were presented by Okorie and Oyedepo [12], who analyzed three different domestic newspapers. They focused on 12 predefined agricultural sub-sectors and searched for related news in the newspapers. Since those sub-sectors were predefined and fixed, some relevant content may have been incorrectly assigned or missed out. In contrast, dynamic sub-sectors may increase the acceptance of relevant news information. Local and global brand food and beverages in Jamaican newspaper were analyzed to determine standardized strategies [13]. Newspaper contents were automatically analyzed to demonstrate the current views of climate change in India [14]. The researchers used a topic modeling algorithm to classify the contents into four overarching themes. Another newspaper analysis was used to determine the impact of Covid-19 on the tourism sector in China [15], which is an excellent work but its scope was limited to the tourism sector only.

Based on the analysis of the above literature, this paper presents dynamically generated underlying discussed keywords on Covid-19 and Bangladesh Agriculture using a topic modeling algorithm to predict a better view of the current situation on the agriculture sector of Bangladesh. This research also helps to find out the topics that need to be emphasized for better communication. We have collected a dataset from publicly available online news sources, including newspapers (printed and online), organizational portal, blogs, and forums. Then, the dataset is filtered and cleaned for observation using a topic modeling algorithm, which is widely used for textual data analysis. Finally, identified underlying topics, keywords, and contents are statistically analyzed with independent variables. This paper selects six research questions to determine the scenario presented by open access news sources about agriculture in Bangladesh during Covid-19, which will be helpful to bridge the gap between policymakers and the agriculture sector.

**RQ1:** What are the underlying topics that can be identified in articles related to the impact of Covid-19 on the agriculture sector in Bangladesh?

**RQ2:** What is the distribution of article sources and writers?

**RQ3:** What is the relationship between identified topics across the source types and content?

**RQ4:** Does the proportion of author type differ across the online news source types?

**RQ5:** Does the content length significantly relate to the type of content writer?

**RQ6:** Does the proportion of author types differ with identified topics and news source types?

By addressing those research questions, this paper visualizes the current point of views on the Covid-19 and Bangladesh Agriculture sector that occur in newspapers or online portals. Some important findings were also addressed in this paper including the low number of female writers, high proportion of government-related content, and the correlation between content size and source type. This paper concluded with some appropriate guidelines and suggestions for communication concerned with the agricultural sector during Covid-19 pandemic, and it is hoped that these will assist policy makers and international aid groups.

## 2. Materials and Methods

This research presents an analysis of open access online articles on the Covid-19 and Bangladesh agriculture using topic modeling algorithms in order to uncover the underlying discussed issues. A detailed description of methods, dataset collection and other materials are presented in this section.

### 2.1 Dataset

Data were collected from open access publicly available online English articles that were concerned with the effects of Covid-19 on Bangladesh Agriculture. The articles were published in local and international newspapers, organizational portals, blogs or forums. These articles selected were sampled from the date of the first Covid-19 patient confirmed in Bangladesh (March, 2020) until the date that the country reported 75,000 cumulative patients (June 10, 2020). All articles that were available online and indexed by the Google search engine were considered. The Google search engine was selected as it had the highest search engine market share worldwide (92.06%) and its market share in Bangladesh was 98.34% [16]. We used four search queries (Q) in Google with specific settings that are given below.

**Q1:** “Bangladesh agriculture covid-19”

**Q2:** “Bangladesh agriculture pandemic”

**Q3:** “Bangladesh agriculture coronavirus”

**Q4:** “Bangladesh farmer covid-19”

**Tool:** Country: Bangladesh; Time: Any time; Result: All result;

**Setting:** Region Settings: Bangladesh; Private results: Use private results.

The collected and filtered dataset was publicly available at online for further usages (<https://sites.google.com/view/moriomrma/research>). Four source types were categorized from the dataset: Blog & Forum (5%), Organizational Portal (18%), Online Newspaper (20%) and Printed Newspaper online version (57%). An overview of the observation data is listed in Table 1. It should be mentioned that the observation sample size is not that large because most of the news sources did not cover agriculture issues frequently. On the other hand, a large number of printed newspapers did not have an online version, which was a major barrier to information sharing in the recent digital world. However, people in Covid-19 lockdown have been increasing their online activities in order to maintain social distancing and safety issues. So, to cover maximum reachable articles we exclusively consider online version content for data analysis.

**Table 1.** Overview of survey data

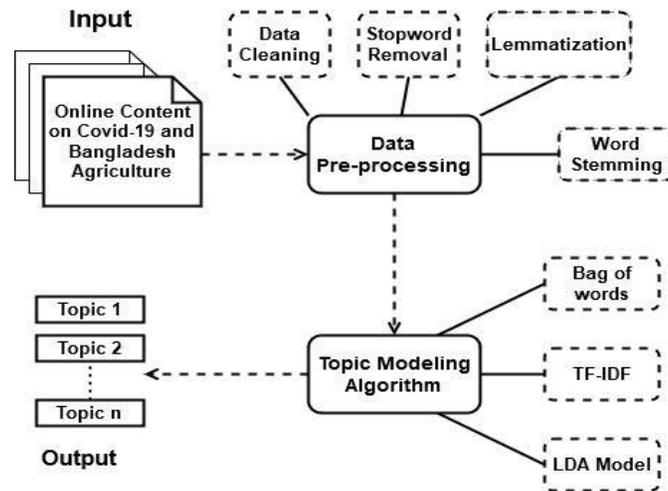
Source Type	Source Name	Observation	(%)	Cumulative
Blog and Forum	City Farmer	1	5%	5%
	Fish Tank	1		
	Techlife	1		
	East Asia Forum	1		
Organization Portal	Aesa Network	1	18%	23%
	Anadolu Agency	1		
	Atlanta Council	1		
	BAEN BD	1		
	Bangladesh Awami League	2		
	BRAC	1		
	Databd	1		
	Fish Tank	1		
	Green Watch BD	1		
	Int. Finance Corporation	1		
	KMPG	1		
	Light Castle	1		
	Relief Web	1		
	US Embassy	1		
USAID	1			
Online Newspaper	Bangladesh Monitor	1	20%	43%
	bd24live	1		
	bdnews24	6		
	Channel News Asia	1		
	Daily Industry	1		
	En Prothom Alo	1		
	IPS News	1		
	Somoy News	2		
	UCA News	1		
	United News Bangladesh	2		
	World Economic Forum	1		
Printed Newspaper	Arab News	4	57%	100%
	Bangladesh Post	5		
	Business Standard	3		
	Daily Asian Age	2		
	Daily Star	16		
	Dhaka Tribune	5		
	Financial Express	10		
	Independent	1		
New Age	4			
Total		88	100%	

## 2.2 Topic modeling algorithm for textual data classification

Topic modeling is a classic solution to information retrieval problems from textual documents [17]. Such problems are closely related to the problem presented in this paper named textual article

classification problems on Covid-19 and Bangladesh Agriculture. Topic modeling imagines a fixed set of topics, where each topic represents a set of words. One of the popular topic modeling techniques, known as Latent Dirichlet Allocation (LDA) [18], has been used for selecting the topics from the dataset of Covid-19 in Bangladesh Agriculture articles. The goal of LDA is to map all the articles to the topics in a way such that the words in each article are mostly captured by those pre-determined topics.

The textual data was scrolled from various sources with title, author, contents, date and source name. Then stop words were removed from the contents, and data were finally processed by lemmatization and word stemming. Then, the filtered words were transformed into bags of words and a probabilistic model was executed to distribute the words into different prominent topics as output. The overall process of this experiment is depicted in Figure 1 and the steps are described below.



**Figure 1.** Overview of topic modeling algorithm

**Data cleaning:**

Only Covid-19 and Bangladesh Agriculture-related online content are allowed for the next phase of data pre-processing. Others irrelevant data are erased from the dataset in the process of data cleaning.

**Stop word removal:**

Stop words, which are words that have no specific meaning in the sentence, including prepositions, conjunctions and others are removed for better data modeling.

**Lemmatization:**

Lemmatization is the process of grouping together the various forms of a word so all of them can be analyzed as a single item [19]. Filtered words are lemmatized when words having the third person form in the sentence are changed to first person and verbs in past and future tenses are changed into present tense format.

**Word stemming:**

The lemmatized words which are basically verbs need to be stemmed, and this refers to the conversion to their root form. Finally, a set of purified and filtered words are ready for processing in the topic modeling algorithms.

**Bag of words:**

A ‘Bag of words’ is a special type of dictionary containing the words related to respective articles in ‘Covid-19 and Bangladesh Agriculture’. After stemming, the purified words are counted and stored in a dictionary containing the number of times a word appears in a single article. Those frequencies will be used to determine the influential words for modeling the topics.

**TF-IDF:**

TF-IDF is a numerical statistic which is used to reflect how important a word is to an article in the collection dataset on ‘Covid-19 and Bangladesh Agriculture’. TF-IDF is a very popular approach in information retrieval systems. It is the product of two statistics, Term Frequency and Inverse Document Frequency [20]. The processes for calculating TF-IDF are presented in equations (1), (2), and (3) as follows:

$$tf(t, a) = \frac{f_{t,a}}{\sum_{t' \in a} f_{t',a}} \tag{1}$$

In the above equation (1), term frequency  $tf(t,a) = 1$  if  $t$  occurs in  $a$  and 0 otherwise, and  $\sum_{t' \in a} f_{t',a}$  is the summation of the number of words in article  $a$ .

$$idf(t, A) = \log \log \frac{N}{|a \in A: t \in a|} \tag{2}$$

In the above equation (2), the inverse document frequency is depicted as  $idf(t,A)$ . Where  $N$  is the total number of articles in the dataset  $N = |A|$  and  $|a \in A: t \in a|$  is the number of articles where the term  $t$  appears. If the term is not in the database, this will lead to a division-by-zero, hence it is therefore common to adjust the denominator to  $1 + |a \in A: t \in a|$ .

$$tfidf(t, a, A) = tf(t, a) \times idf(t, A) \tag{3}$$

Then TF-IDF is calculated by the mathematical product of TF and IDF, which is depicted in equation (3). A high weight of TF-IDF is reached by a high term frequency (in the given article); the weights hence tend to filter out common terms from the dataset on Covid-19 and Bangladesh Agriculture.

**LDA model:**

Latent Dirichlet Allocation (LDA) is a probabilistic model for analyzing the underlying probable topics based on the collection of textual articles [18]. Blei *et al.* [18] presented LDA as random mixtures over latent topics, where each topic is characterized by a distribution over words in textual dataset. It has been assumed that a generative process for each document  $w$  in a dataset  $D$  for LDA is used to implement topic modeling algorithm. The details of that are listed below.

- i. Choose  $N \sim \text{Poisson}(\xi)$ .
- ii. Choose  $\theta \sim \text{Dir}(\alpha)$ .
- iii. For each of the  $N$  words  $w_n$ :
  - a. Choose a topic  $z_n \sim \text{Multinomial}(\theta)$ .
  - b. Choose a word  $w_n$  from  $p(w_n / z_n, \beta)$ , a multinomial probability conditioned on topic  $z_n$

**2.3 Implementation**

One of the world’s most popular data analytics based programming languages named Python version 3.8 (<https://www.python.org/>) was used for data interpretation in this research. Several programming library modules were also used in the experiment, and included were pandas (<https://pandas.pydata.org/>) for data handle, Python nltk (<https://www.nltk.org/>) for data preprocessing, genism (<https://pypi.org/project/gensim/>) for LDA modeling, and matplotlib (<https://matplotlib.org/>) for result visualization. The IBM SPSS 26 and Microsoft Excel 2013 were also used for statistical data analysis.

### 3. Results and Discussion

Textual newspaper content was collected and analyzed to pull up actual scenarios in the Bangladeshi agriculture sector during Covid-19 pandemic. The findings with discussion are presented in this section.

#### 3.1 Research questions

This paper addresses six different research questions to present its findings. The results and their discussion with feedback are described below.

##### **RQ1: What are those topics to visualize the contents of Covid-19 and Agriculture in Bangladesh?**

The topic modeling algorithm dynamically created a total of 6 of the most prominent agricultural topics for presenting Covid-19 and Bangladesh Agriculture. The overviews of those topics with their relevant information are listed in Table 2, in which topic-wise article distribution, frequent terms, keywords, number of words and probable meanings are presented accordingly. The terms of the entire dataset are visualized in Figure 2(a), where several prominent terms are located with different color and size. The descriptions of six individual topics are listed below.

**Topic 1** contains 20.5% of the total articles including the keywords related to supply chain management. The probable title of this topic is the economic impact on food and agricultural products supply chain during Covid-19. It has only 485(8.25%) times social media share which indicates readers' low level of interest. The term visualization of Topic 1 is presented in Figure 2(b).

**Topic 2** contains only 5.7% of the total article and features keywords in the area of Agribusiness and Product Harvesting. The tentative title of this topic is Crisis in harvesting seasonal fruits and boro paddy by farmers during Covid-19. The automatically tagged terms under this topic are visualized at Figure 2(c). However, the social media share is 507 (8.62%) times which is also low.

**Topic 3** contains the second highest percentage of the total articles (27.3%), with keywords of Agricultural Economics and Government. The potential title of this topic is Million dollar Government support for the country's SME, agriculture and fish sector. It must be mentioned that Bangladesh Government announced a stimulus package known as Taka 5000 Crore to offer financial assistance to the farmers. A large number of news portals presented this topic positively and its social media share is 942 (16.02%) time, which is high. Term visualization of topic 3 is presented at Figure 2(d).

**Topic 4** contains the highest number of articles which is 34.1% of total. The keywords of this topic are Production Loss, Natural Disaster, and Government. This topic is related to the catastrophe of super cyclone Ampan and Covid-19 pandemic. The probable title of this topic is Government initiatives to mitigate district-wise production loss in Covid19 lockdown and Ampan cyclone. Since it has a direct effect on the country's agriculture sector, the social media share is also the highest at 2816(47.89%) time, which indicates that people are conscious of the combine natural disaster and pandemic issue. The underlying terms of Topic 4 are represented at Figure 2(e).

**Topic 5** contains only 6.8% of the total articles with the keywords of Farmers' safety and Rural Economic. This topic is closely related to the pandemic health issue and the agricultural sector. However, as part of this topic is already covered by topic 3, its coverage and level of peoples' interest are relatively low, which is indicated by the social media share value (1.52%). The underlying terms covered by this topic is presented at Figure 2(f).

**Table 2.** Topic wise term distribution and probable meaning with article frequency

Topic	Article (%)	# of terms	Keywords	Probable description of topic	Top 20 words for this topic	Social media share
1	20.50	8003	Supply Chain Management	Economic impact of food and agricultural products supply chain during Covid-19	['market', 'inputs', 'prices', 'demand', 'country', 'food', 'vulnerable', 'income', 'impact', 'sector', 'supply', 'production', 'crisis', 'government', 'economic', 'people', 'farmers', 'Bangladesh', 'covid19', 'agricultural',]	8.25%
2	5.70	5009	Agribusiness, Product Harvest	Crisis for harvesting seasonal fruits and boro paddy by farmers during Covid-19.	['farmer', 'budget', 'mango', 'paddy', 'pandemic', 'covid19', 'agricultural', 'food', 'vegetables', 'rice', 'seasonal', 'meeting', 'government', 'ministry', 'harvest', 'relief', 'crop', 'production', 'fruits', 'agriculture']	8.62%
3	27.30	15914	Agricultural Economics, Government	Million dollar Govt. support for country's SME, agriculture and fish sector	['farmer', 'food', 'sme', 'year', 'stimulus', 'lockdown', 'poor', 'fish', 'minister', 'banks', 'bank', 'coronavirus', 'Bangladesh', 'package', 'people', 'farmer', 'million', 'workers', 'government', 'April']	16.02%
4	34.10	12523	Production loss, Natural Disaster, Government	Government initiatives to mitigate district wise production loss in covid19 lockdown and Ampan cyclone	['income', 'support', 'agriculture', 'covid19', 'hospitals', 'package', 'damaged', 'amphan', 'people', 'district', 'prime', 'farmers', 'district', 'health', 'country', 'minister', 'cyclone', 'government', 'hectares', 'taka']	47.89%
5	6.80	1086	Farmers safety, Rural Economy	Prime minister announced motivational packages for farmer, health workers and victims.	['assistance', 'coronavirus', 'minister', 'production', 'world', 'need', 'economy', 'people', 'pandemic', 'food', 'crisis', 'Bangladesh', 'agriculture', 'country', 'prime', 'health', 'sector', 'farmers', 'fish', 'government']	1.52%
6	5.70	2585	Product Purchase, post-harvest storage, Government	Ministry of Bangladesh government said the collection and harvesting target of agricultural food and boro paddy.	['ministry', 'crop', 'food', '\$589m', 'price', 'haor', 'percent', 'harvesting', 'farmers', 'country', 'taka', 'harvest', 'rice', 'Bangladesh', 'government', 'boro', 'agricultural', 'paddy', 'loan', 'impact']	17.7%



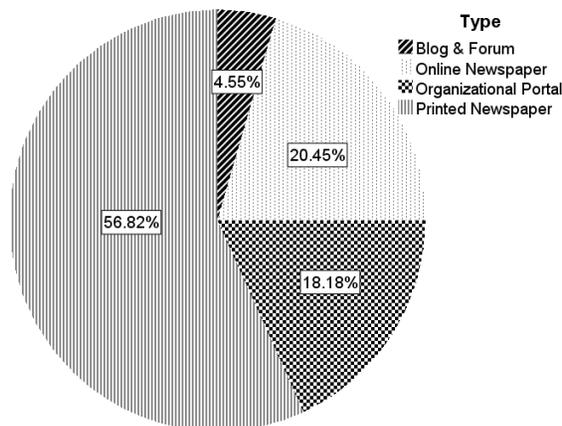
rural economics, its social media share is the second highest at 1041(17.70%) time. The underlying terms of topic 6 are visualized in Figure 2(g).

**RQ2: What is the distribution of article sources and writers for the content of Covid-19 and Agriculture in Bangladesh?**

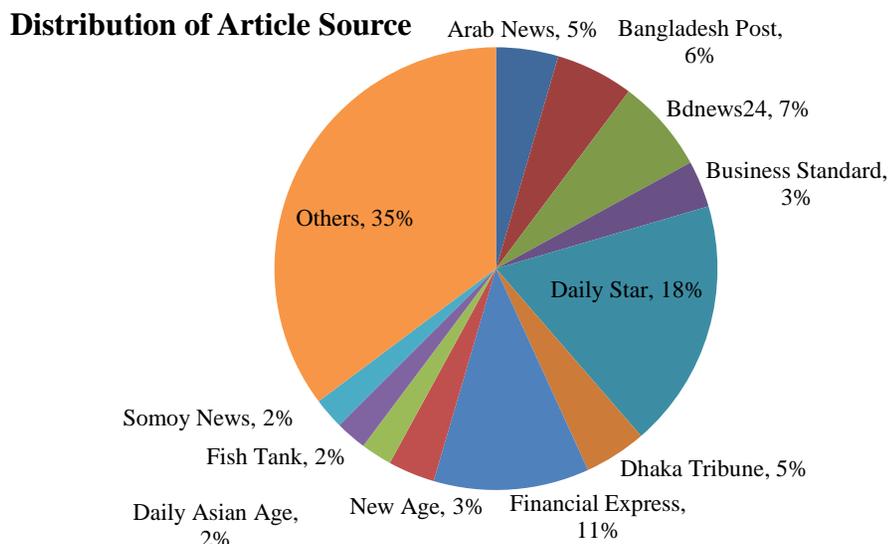
The articles concerned with Covid-19 and Agriculture in Bangladesh were collected from various sources that were categorized under four types, named as Blog & Forum, Online Newspaper, Organizational Portal and Printed Newspaper, and these had distributions of 4.5%, 20.5%, 18.2% and 56.8%, respectively. The results are presented in Table 3 and Figure 3. It can be seen that online and printed newspapers cover two-third of the dataset. It can be predicted that people are not willing to discuss about agricultural issues on their own in blogs or forums. Individual source-based content distribution is depicted as a pie chart in Figure 4.

**Table 3.** Statistical analysis of article source type

Source Type	Frequency	%	Valid %	Cumulative %
Blog & Forum	4	4.5	4.5%	4.5
Online Newspaper	18	20.5	20.5	25.0
Organizational Portal	16	18.2	18.2	43.2
Printed Newspaper	50	56.8	56.8	100.0
Total	88	100	100.0	



**Figure 3.** A pie chart showing statistical analysis of article source types



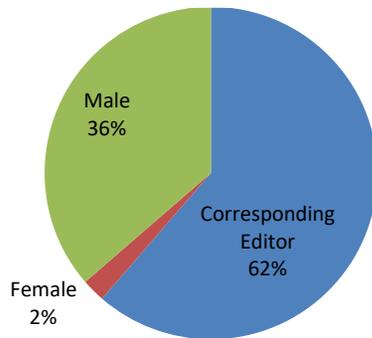
**Figure 4.** Visualization of article source names

Moreover, more than 60% of the content writers are corresponding editors, which means the ordinary agricultural events are covered by responsible person. A notable observation was that concerned with female content writers in agriculture. Only 2.3% of the articles were written by female agricultural reporters, which should be increased. The other 36% of individual reporters are male in gender. The details of article writer statistics are presented at Table 4 and Figure 5.

Table 4 shows that corresponding editors wrote the highest number of agricultural articles, and these writers may have been of either gender. However, the publicly gender-specific reporters are not equal to males and females in the agriculture sector during this Covid-19 pandemic.

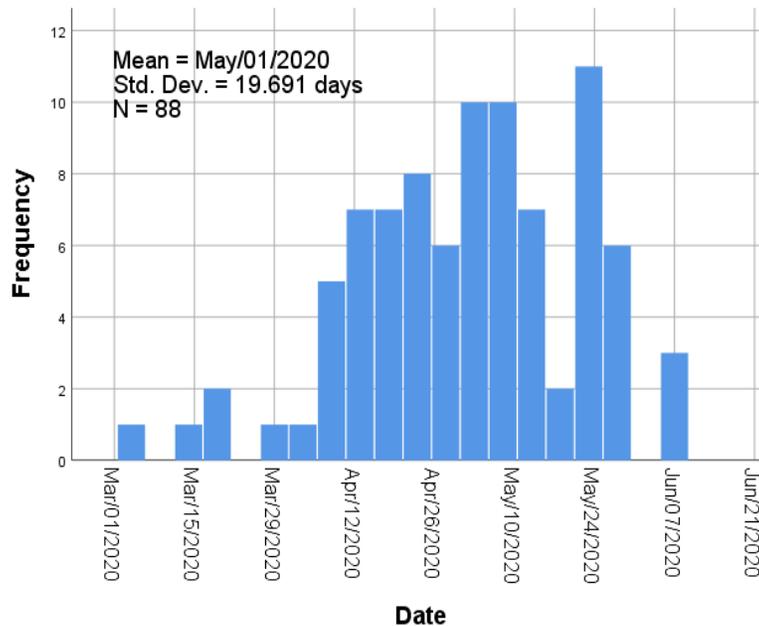
**Table 4.** Statistical analysis of content writers

Writer	Frequency	%	Valid %
Corresponding Editor	54	61.4	61.4
Female	2	2.3	2.3
Male	32	36.4	36.4
Total	88	100	100.0



**Figure 5.** A pie chart showing statistical analysis of content writers

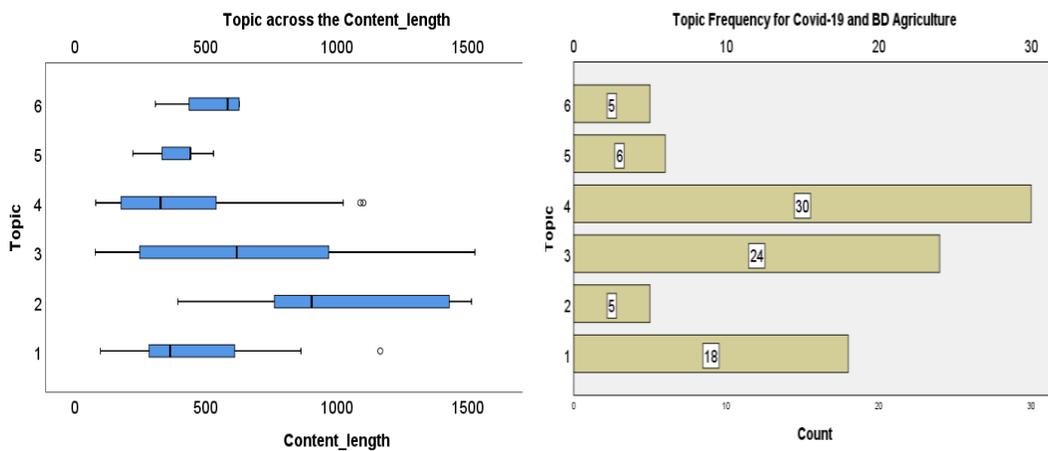
A publishing date-wise article frequency histogram can be seen in Figure 6. It shows that the date range for article consideration was March 01, 2020, to June 10, 2020. It can be seen that May 24 2020 was the day on which the maximum number of articles (11 articles) was published, and the mean publishing date was May 01, 2020. The majority of the Covid-19 and Bangladeshi agriculture-related articles were published within April and May 2020 with some outliers on June 07 and March 02, 2020. Not only were those two months peak times for the outbreak of Covid-19 in Bangladesh, they were also significant because the Government subsidiary package announcements, super cyclone Ampan stroke, and other agriculture-related issues occurred then.



**Figure 6.** Histogram of published articles based on date

**RQ3: What are the relations with identified topics across the source types and content?**

Research question 3 can be split in two different sub questions. The first one is about the relationship between identified topics and content. Figure 7 presents the relationship between identified topics and content variables, where content variables have two forms named as content length and content frequency. Our analysis found that topic 4 has the maximum number of articles but their lengths are below average. It means this type was related to typical media reports only. The maximum number of article lengths is in topic 2, but its frequency is the lowest. Our observation found that these types of articles are basically detailed observations from the reporter. It was also found that topic 3 has the optimal coverage based on frequency and length, which covers the country's agricultural economics and Government issues.



**Figure 7.** Relationship between identified topics and content

The second part of research question 3 is about the relationship between identified topic and source type, which confirms the topic coverage source. Figure 8 presents the relationship between article source and topics. It shows that printed newspapers covered the highest number of articles and obviously topic 4 has the highest frequency count. An interesting finding that can be reported from the graph is that the organizational portal covers the highest number of the country's agricultural economic issues, which is a type of investigation and R&D report. It was also found that newspapers (both online and printed) cover the maximum number of typical media issues e.g. super cyclone Ampan, product pricing, etc.

**RQ4: Does the proportion of author type differ across the online news source types?**

Research question 4 is related to the relationship between content writer and source type. We selected the hypotheses shown below for this experiment.

H0 = Author types are independent across the online news source categories.

H1 = Author types are dependent across the online news source categories.

Table 5 presents the statistical analysis of content writer and source type, where the significance of Pearson chi-square and likelihood ratio is also shown. The fact that the *p* value of chi-square is 0.862, which is far larger than 0.05, means that the null hypothesis is accepted and it can be concluded that author types are independent across the article source categories

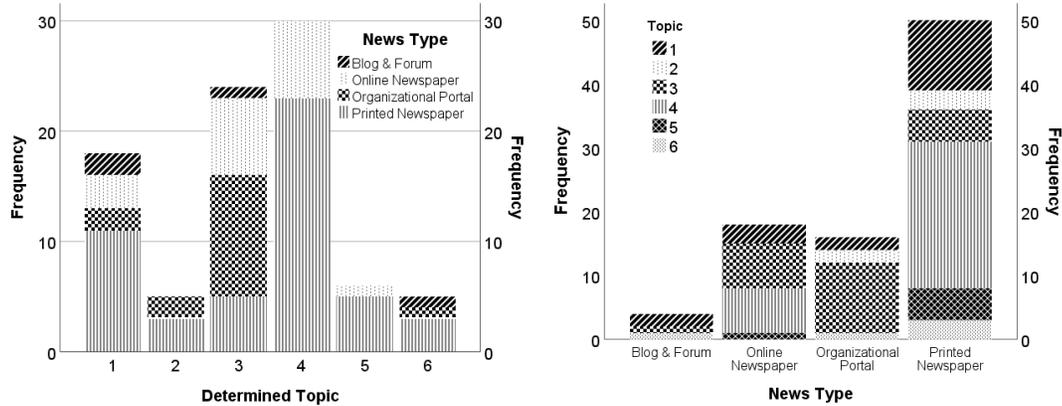


Figure 8. Identified topic and type of article relation

Table 5. Relationship between content writer and source type

		Source Type				Total
		Blog & Forum	Online Newspaper	Organizational Portal	Printed Newspaper	
Author	Corresponding Editor	2	10	10	32	54
	Female	0	0	0	2	2
	Male	2	8	6	16	32
	Total	4	18	16	50	88
		Chi-Square Tests				
		Value	df	Asymptotic Significance (2-sided)		
<b>Pearson Chi-Square</b>		2.558 <sup>a</sup>	6	.862		
<b>Likelihood Ratio</b>		3.276	6	.774		
<b>N of Valid Cases</b>		88				

<sup>a</sup>6 cells (50.0%) have expected count less than 5. The minimum expected count is .09.

**RQ5: Does content length significantly relate to the type of content writer?**

Research question 5 deals with the relationship between content writer and content length. We selected various hypotheses for this experiment. It can be clearly seen that female writer frequency is ferociously low. To measure the significance of article length and writer type, the following hypotheses were used.

H0: There is no relationship between the two variables, content length and content writer.

H1: There is a significant relationship between content length and content writer type.

Table 6 presents the statistical results of the above hypotheses via Pearson chi-square and likelihood ratio. Since the *p* value 6 is 0.011, which is less than level of significance, the null hypothesis is rejected. It can therefore be determined that content length is significantly related to content writer type.

**Table 6.** Content length and content writer cross tabulation

		Content Writer Type			Total
		Corresponding Editor	Female	Male	
Content Length	<300	25	0	5	30
	300-599	17	0	10	27
	600-899	6	2	10	18
	900-1199	3	0	5	8
	>1200	3	0	2	5
Total		54	2	32	88
Chi-Square Tests					
		Value	df	Asymptotic Significance (2-sided)	
<b>Pearson Chi-Square</b>		19.746 <sup>a</sup>	8	0.011	
<b>Likelihood Ratio</b>		18.950	8	0.015	
<b>Linear-by-Linear Association</b>		7.912	1	0.005	
<b>N of Valid Cases</b>		88			

<sup>a</sup>9 cells (60.0%) have expected counts less than 5. The minimum expected count is .11.

**RQ6: Does the proportion of author type differ with identified topics and news source types?**

Research question 6 focuses on the difference between identified topics and news source types. To evaluate their difference, the hypotheses below were selected.

H0: There is no relationship between variables news source category and newly identified topics.

H1: There is a significant relationship between the categorical variable topic and source type.

Table 7 presents the statistical analysis of the above hypotheses to measure the differences. Based on this Table, the *p* value is .002 which is less than 0.05, and this means the null hypothesis is rejected and source types and identified topics are not independent. It can be concluded that there is a significant correlation between content source type and determined topic.

**Table 7.** Identified topics and source type cross tabulation

		Topic						Total
		1	2	3	4	5	6	
Source Type	Blog & Forum	2	0	1	0	0	1	4
	Online Newspaper	3	0	7	7	1	0	18
	Organizational Portal	2	2	11	0	0	1	16
	Printed Newspaper	11	3	5	23	5	3	50
Total		18	5	24	30	6	5	88
Chi-Square Tests								
		Value	df	Asymptotic Significance (2-sided)				
<b>Pearson Chi-Square</b>		36.251 <sup>a</sup>	15	.002				
<b>Likelihood Ratio</b>		43.088	15	.000				
<b>Linear-by-Linear Association</b>		2.018	1	0.155				
<b>N of Valid Cases</b>		88						

<sup>a</sup>19 cells (79.2%) have expected count less than 5. The minimum expected count is .23.

### **3.2 Performance evaluation**

A pool of experts blindly evaluated the results found from the topic modeling algorithms. That pool consisted of five members, including two agricultural academicians, two agriculture service providers, and one agriculture expert policymaker. The review process was blind, and no one shares their scores with others. The reviews were collected on three categories, named as accepted, neutral, and rejected. Then, we use the majority voting on those topics generated by our work. It was found that the experts dismissed none of the items and the keywords. In contrast, more than eighty percent of outcomes were accepted, which confirmed the model's accuracy.

### **3.3 Recommendations**

This paper presents some recommendations based on the experiment and data analysis which depicted a proper view of online news articles on Covid-19 and Bangladesh Agriculture. These recommendations can also be used for creating agriculture policy and decision support. The recommendations for the agriculture section in Bangladesh during Covid-19 are listed below.

- 1) Too much government-related news in the agriculture sector was found. It needs to be amalgamated with detailed investigation and creative news related to actual farmers' issues during this pandemic.
- 2) National newspapers and publicly available news portals did not put agriculture at the top of their news priorities. However, Bangladesh is predominantly an agricultural country, and agricultural news should be prioritized to grab the eyeballs of policymakers.
- 3) Personal experience or opinions inside blogs related to agriculture were alarmingly low and should be increased to publicize more actual views of the current problems.
- 4) The numbers of female authors that reported on agriculture news articles was chronically low. This imbalance needs to be corrected by government subsidiary or other relevant actions.

## **4. Conclusions**

This research dealt with English articles on Covid-19 and agriculture in Bangladesh to figure out which sections were communicated enough and which needed more emphasis. This analysis was conducted using topic modeling algorithms to determine the underlying topics of the articles. Then, several statistical analyses were performed to answer six different research questions. It was found that there is a significant correlation between article source type and determined topics. Furthermore, the correlation between content length and writer is significant. It is also reported that newspapers did not cover enough detailed investigation of the current problems of agriculture in the Covid-19 issue. On the other hand, several organizations did the opposite, because most of the articles in organizational portals focused on the agricultural problematic issues.

Finally, this paper recommends that newspapers should focus more on problems with an internal investigation and engage more female writers to cover agricultural matters in Bangladesh. This research can be extended to other prominent topics such as education, labor, and social service. Comparing the articles among other countries' newspapers and publicly available online news sources may also be a future direction of this work to determine a broader view of the agriculture sector during this pandemic.

## References

- [1] Goddard, E., 2020. The impact of COVID-19 on food retail and food service in Canada: Preliminary assessment. *Canadian Journal of Agricultural Economics*, <https://dx.doi.org/10.1111/cjag.12243>
- [2] Kerr, W.A., 2020. The COVID-19 pandemic and agriculture: Short- and long-run implications for international trade relations. *Canadian Journal of Agricultural Economics*, <https://dx.doi.org/10.1111/cjag.12230>
- [3] Irvine, M. 2020. *How COVID-19 Is Shaping Google Search Trends & Patterns*. [Online] Available at: <https://www.wordstream.com/blog/ws/2020/03/30/covid-19-google-search-trends>
- [4] Bangkok Post, 2020. *Sixty Australian Newspapers to Stop Printing*. [Online] Available at: <https://www.bangkokpost.com/world/1890915/sixty-australian-newspapers-to-stop-printing>
- [5] Narayana, U. and Kumar, S. 2009. Content analysis of agricultural news coverage in leading language dailies of India-A study of Kannada language dailies of Karnataka State. *University Journal of Communication*, 3(1), 1-14.
- [6] Chandelier, M., Steuckardt, A., Mathevet, R., Diwersy, S. and Gimenez, O., 2018. Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. *Biological Conservation*, 220(1), 254-261.
- [7] Ghosh, D. and Guha, R., 2013. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartography and Geographic Information Science*, 40(2), 90-102.
- [8] Bhatia, S., Athreya, V., Grenyer, R. and Macdonald, D.W., 2013. Understanding the role of representations of human-leopard conflict in Mumbai through media-content analysis. *Conservation Biology*, 27(3), 588-594.
- [9] Miah, M.D., Kabir, M.H., Koike, M. and Akther, S., 2011. Major climate-change issues covered by the daily newspapers of Bangladesh. *The Environmentalist*, 31(1), 67-73.
- [10] Hart, C.E., Hayes, D.J., Jacobs, K.L., Schulz, L.L. and Crespi, J.M., 2020. The Impact of COVID-19 on Iowa's Corn, Soybean, Ethanol, Pork, and Beef Sectors. [online] Available at: <https://www.card.iastate.edu/products/policy-briefs/display/?n=1301>
- [11] Ogessa, C.M. and Sife, A.S., 2017. Newspaper coverage of agricultural information in Tanzania. *University of Dar es Salaam Library Journal*, 12(1), 12-26.
- [12] Okorie, N. and Oyedepo, T., 2011. Newspaper reportage and its effect towards promoting agricultural development in Nigeria. *Journal of Media and Communication Studies*, 3(2), 27-32.
- [13] Ahn, R.J., Nelson, M.R. and Ferguson, G.M., 2020. Local and standardized strategies: A content analysis of newspaper food and beverage advertising in Jamaica. *Newspaper Research Journal*, 41(2), 179-203.
- [14] Keller, T.R., Hase, V., Thaker, J., Mahl, D. and Schäfer, M.S., 2020. News media coverage of climate change in India 1997-2016: Using automated content analysis to assess themes and topics. *Environmental Communication*, 14(2), 219-235.
- [15] Chen, H., Huang, X. and Li, Z., 2020. A content analysis of Chinese news coverage on COVID-19 and tourism. *Current Issues in Tourism*, <https://doi.org/10.1080/13683500.2020.1763269>
- [16] Stat Counter Global Stats, 2020. *Search Engine Market Share Worldwide*. [Online] Available at: <https://gs.statcounter.com/search-engine-market-share>
- [17] Wallach, H.M., 2006. Topic modeling: beyond bag-of-words, *Proceedings of the 23rd International Conference on Machine Learning*, New York, United States, 2006, pp. 977-984.

- [18] Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(1), 993-1022.
- [19] Manning, C.D., Raghavan, P. and Schütze, H., 2009. An Introduction to Information Retrieval. Cambridge:Cambridge University Press.
- [20] Jones, K.S., 1972. A statistical interpretation of term specificity and its application in retrieval, *Journal of Documentation*, 28(1), 11-21.