



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

## Social Sciences &amp; Humanities Open

journal homepage: [www.elsevier.com/locate/ssaho](http://www.elsevier.com/locate/ssaho)

# Recreational and philanthropic sectors are the worst-hit US industries in the COVID-19 aftermath



Satyaki Roy<sup>a,\*</sup>, Ronojoy Dutta<sup>b</sup>, Preetam Ghosh<sup>c</sup>

<sup>a</sup> University of North Carolina, Chapel Hill, USA

<sup>b</sup> Deep Run High School, Glen Allen, VA, USA

<sup>c</sup> Virginia Commonwealth University, Richmond, USA

## ARTICLE INFO

## Keywords:

COVID-19  
Economy  
Topic model  
Policymaking  
Job losses

## ABSTRACT

Lockdown measures to curb the spread of COVID-19 has brought the world economy on the brink of a recession. It is imperative that nations formulate administrative policies based on the changing economic landscape. In this work, we apply a statistical approach, called topic modeling, on text documents of job loss notices of 26 US states to identify the specific states and industrial sectors affected economically by this ongoing public health crisis. Our analysis reveals that there is a considerable incongruity in job loss patterns between the pre- and during-COVID timelines in several states and the recreational and philanthropic sectors register high job losses. It further shows that the interplay among several possible socioeconomic factors would lead to job losses in many sectors, while also creating new job opportunities in other sectors such as public service, pharmaceuticals and media, making the job loss trends a key indicator of the world economy. Finally, we compare the low income job loss rates against overall job losses due to COVID-19 in the US counties, and discuss the implications of press reports on reopening businesses and the unemployed workforce being absorbed by other sectors.

## 1. Introduction

Human history is scarred by plague, flu and Ebola that have globally claimed millions of lives (Coronavirus: what have be, 2020). Infectious disease is the leading cause of human deaths and will continue to affect many more in the years to come (Walsh, 2020). COVID-19, the ongoing pandemic caused by acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to 1.28 million deaths since its inception in Wuhan, China in December 2019 (Coronavirus world map: wh, 2020). Despite the strictest lockdown measures (Kimball et al., 2020) to contain infection spread, efforts by the frontline health workers and clinical trials of vaccines (Smith, 2020), there is no end in sight to this global health crisis.

Epidemiologists, clinicians and computer scientists are applying their expertise to seek out factors and their implications on contagion as well as economic downturn (Adhikari et al., 2020). First, there are attempts to apply machine learning (ML) to build prediction models on epidemiological and clinical data. Given existing clinical data, prediction models (Wynantset al., 2020) and therapeutic approaches can help identify vulnerable groups (Alimadadi et al., 2020; Randhawa et al., 2020). Epidemiologists are trying to identify spread dynamics of COVID-19. Inga Holmdahl et al. (Holmdahl & Buckee, 2020) analyze the pros and cons of

forecasting models that make predictions through curve fitting or mechanistic models, while supervised and unsupervised ML is helping trace the trends in infection dynamics (Wang et al., 2020). Khan et al. used regression tree analysis, cluster analysis and principal component analysis on Worldometer infection count data to gauge the variability and effect of testing in prediction of confirmed cases (Khan et al., 2020). Roy et al. perform regression analysis to identify pre-lockdown factors that affect the post-lockdown pandemic numbers (Roy & Ghosh, 2020).

The other aspect of the ensuing lockdown is the looming financial crisis worldwide. The economists are expecting an unprecedented decline in industrial output and stock exchange percentage, increase in the price of goods (Khan et al., 2020) as well as a potential for contraction in US GDP (Baker et al., 2020). While national governments try to offset this slump in commodity prices as well as the households, firms, and financial markets by providing economic assistance to the affected groups (Gopinath, 2020), it is clear that such an arrangement is not sustainable for the developing nations. It is imperative to study the implications of COVID-19 on the global fiscal ecosystem to design effective administrative policies.

*Contributions.* In this work, we study the industrial impacts of COVID-19 in 26 US states. We apply *topic model* – a widely used statistical and

\* Corresponding author.

E-mail addresses: [satyakir@unc.edu](mailto:satyakir@unc.edu) (S. Roy), [hcps-duttar@henricostudents.org](mailto:hcps-duttar@henricostudents.org) (R. Dutta), [pghosh@vcu.edu](mailto:pghosh@vcu.edu) (P. Ghosh).

<https://doi.org/10.1016/j.ssaho.2020.100098>

Received 10 August 2020; Received in revised form 11 November 2020; Accepted 2 December 2020

Available online 9 December 2020

2590-2911/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

natural language processing approach to identify latent topics in documents – to recognize the affected job sectors. We generate *industry type* (such as hospitality, health care, etc.) repository for a state as follows: we process an open-access dataset of layoff notices from the Worker Adjustment and Retraining Notification (WARN) Act (from July 1998 to present) and connect them with the industry types from a Kaggle database of 7 million companies.

Our topic analysis reveals the following interesting details about the US states and industry types: (1) Arkansas, Colorado, Connecticut, Georgia, Kentucky, North Carolina and Virginia are the states that have the highest disparity in job loss patterns in the pre- and during-COVID timelines and (2) the *recreational* and *philanthropic* sectors are the worst affected industries, while the *blue collar jobs*, *public service* (including pharmaceuticals) and *print media* industries undergo fewer job losses in the wake of COVID-19; *food distribution*, *politics*, *electronic media*, *research* are unaffected. We also study the job loss data from the US counties to gauge the impact of COVID-19 on the low income job loss rates. Finally, we discuss the implications of media reports on businesses reopening during and after the pandemic as well as the unemployed workforce being absorbed by other sectors.

## 2. Materials and methods

### 2.1. Data preprocessing

We consider three datasets: (1) of 7 million companies from 237 countries sourced from standard sources like LinkedIn (and available at <https://www.kaggle.com/peopledatalabssf/>). It includes the following fields: *company name*, *year founded*, *industry type*, *country*, *number of employees*, etc. (2) Worker Adjustment and Retraining Notification (WARN) Act (and the U.S. Department of Labor) that helps ensure advance notice in cases of qualified plant closings and mass layoffs. For each of the 26 states (for which all the fields could be parsed properly), there are the following fields: *company name*, *layoff*, *total workforce*, *local area* and *WARN Date*. (3) Urban Institute data science team used data from US Bureau of Labor Statistics to estimate the number of low income jobs (less than \$40,000 salary) lost due to COVID-19 (<https://datacatalog.urban.org/dataset/estimated-low-income-jobs-lost-covid-19>). We make them available on <https://github.com/satunr/COVID-19/tree/master/JobDataset>.

We process the Kaggle dataset to generate a hash table of the following format  $H$ : *company name* → *industry type*. For a company-name entry in the WARN data and state  $i$ , we record the industry type (i.e.,  $H$  (*company name*)) in a new document  $d_i$ . Thus,  $d_i$  is a space-delimited text file of industry type tokens processed further to filter out stopwords (such as prepositions, punctuation, articles, etc.). We then utilize the date field in the WARN data to break down  $d_i$  into pre-COVID (up to December 31st, 2019) and during-COVID (January 1st, 2020 onwards) datasets named  $d_i^{pre}$  and  $d_i^{post}$ , respectively.

### 2.2. Topic modeling

A document  $d \in D$  can be represented as a distribution of latent topics, while a topic  $k \in K$  is a distribution over a vocabulary set  $V$ .

#### 2.2.1. Latent semantic analysis

It captures the association between documents and words using a  $|D| \times |V|$  matrix  $M$ ;  $M_{d,w}$  is the importance of word  $w$  in document  $d$ , measured as term frequency-inverse document frequency (*tf-idf*), given by:

$$M_{d,w} = tf_{d,w} \times \log \frac{D}{df_w}$$

Here  $tf_{d,w}$  is the number of occurrences of  $w$  in  $d$  and  $df_w$  is the number of documents containing  $w$ . In other words, *tf-idf* score is high

when  $w$  occurs in few documents but has a high presence in the current document  $d$  (Ramoset al.).

#### 2.2.2. Latent Dirichlet Analysis

It is a variant of latent semantic analysis that uses a Bayesian probabilistic model (Hoffman et al., 2010). Given that the distribution of topics in a document  $d_i$  is given by a Dirichlet distribution  $\theta_{d_i} \sim Dir(\alpha)$  and the distribution of words in a topic is given by a Dirichlet distribution  $\beta_k \sim Dir(\eta)$ , latent Dirichlet Analysis (LDA) models the generation of a document  $d_i$  using words  $w \in V$  in an iterative, two-step generative approach.

- Choose a topic  $k$  from  $K$  based on a multinomial distribution  $Multi(d_i)$ .
- Choose a word  $w$  from  $V$  based on a multinomial distribution  $Multi(k)$ .

*Learning the topic and word distributions.* One approach to learn the  $\alpha$  and  $\eta$  is to minimize KL divergence between the predicted and true posterior. Markov Chain Monte Carlo approach presents a Gibbs sampling method that repeatedly samples the topic assignment of the word conditioned on the data and all other topic assignments (Darling, 2011). The probability that a word  $w_j$  belongs to topic  $k$  in document  $d_i \in D$  is given by:

$$S_{w_j \in k | D} = \frac{tf_{d_i,k}}{n_{d_i}} \times \frac{n_{w_j \in k}}{\sum_{w_j \in V} n_{w_j \in k}}$$

In the above equation, the first term represents the proportion of words in document  $d_i$  that are assigned to topic  $k$  as the ratio between the number of words assigned to  $k$  and total words in  $d_i$  ( $n_{d_i}$ ); the second term is the proportion of assignments to topic  $k$  across all documents that come from  $w_j$  calculated as the ratio between the number of times  $w_j$  is assigned to topic  $k$  ( $n_{d_i \in k}$ ) to the total number of word assignments to  $k$ .

#### 2.2.3. Similarity

A document  $d_i$  is a vector of topics and a topic  $k$  is a vector of words. These metrics are used to gauge the inter-document or inter-topic distance.

- *Hellinger distance.* Given two distributions  $p$  and  $q$ , it is calculated as  $\frac{1}{\sqrt{2}} \times \sum_j (\sqrt{p_j} - \sqrt{q_j})^2$ . It always takes a value between 0 and 1 (Hellinger, 1909).
- *Kullback-Leibler divergence.* Given two distributions  $p$  and  $q$ , it is calculated as  $\sum_j p_j \times \log \frac{p_j}{q_j}$  (Kullback & Leibler, 1951).

#### 2.2.4. Implementation

We implement topic modeling with the Python Gensim library (Rehurek & Sojka, 2011) used for document indexing and similarity retrieval in text documents. Given a set of states  $\Omega$ , we have a set of three documents per state  $i (i \in \Omega)$   $d_i$ ,  $d_i^{pre}$  and  $d_i^{post}$ . Gensim processes the corpora  $L = \{d_i | i \in \Omega\}$  and represents each  $d_i$  as a list of *industry type* tokens (in a manner similar to the approach described in Sec. 2.2.1). It then applies LDA (see Sec. 2.2.2) to identify the latent topics in  $L$  and the distribution of words in each topic. We use the Hellinger distance and KL divergence (see Sec. 2.2.3) libraries to gauge inter-document distance between pre- and during-COVID documents  $d_i^{pre}$  and  $d_i^{post}$ .

#### 2.2.5. One sample proportion Z-test

It is a standard hypothesis testing approach. Given the number of trials and successor trials, one can test a null hypothesis ( $H_0$ ) whether the proportion  $\pi$  (i.e., fraction of successful trials) of the data equals a pre-specified value.

### 3. Results

The results are classified into the following two subsections: (a) distribution of industry types (or words) in topics and (b) identification of the most affected US states and industry types.

#### 3.1. Distribution of words in topics

For our experiments we consider ( $|K| = 10$ ) topics. Each topic  $k$  is a distribution of words  $w$  belonging to the vocabulary set  $V$ . Thus,  $k$  can be represented as some distribution  $[\omega_0^k, \omega_1^k, \omega_2^k, \dots, \omega_{|V|-1}^k]$ , where  $\omega_i^k$  is the weight of  $w_i$  in  $k$  such as  $\sum_{i=0}^{|V|-1} \omega_i^k = 1$ . Below are the top 5 words in each topic in the decreasing order of the weights.

- Topic 0. hospitality, health care, retail, oil energy, restaurants
- Topic 1. hospitality, health care, retail, finance, food and beverages
- Topic 2. retail, food production, oil energy, finance, outsourcing
- Topic 3. retail, hospitality, automotive, health care, mining
- Topic 4. hospitality, health care, retail, oil energy, automotive
- Topic 5. retail, hospitality, health care, restaurants, IT
- Topic 6. hospitality, retail, restaurants, advertising, health care
- Topic 7. retail, health care, hospitality, finance, automotive
- Topic 8. hospitality, restaurants, retail, food and beverage, health care
- Topic 9. finance, aviation and aerospace, food and beverages, retail, IT

For each topic  $k$ , we estimate the total number of statistically significant words, defined as words  $w_i \in V$  such as  $\omega_i^k$  is greater than a pre-defined cut-off. Fig. 1a depicts that, given a cut-off equal to 0.005, the statistically significant words are fairly distributed across topics, showing very little deviation from the mean (represented by a black dotted line).

Similar to topics, a word  $w_i$  too can be represented as a vector comprising the weights it has in each topic, i.e.,  $[\omega_0^k, \omega_1^k, \omega_2^k, \dots, \omega_{|V|-1}^k]$ , where  $\omega_i^k$  is the weight of  $w_i$  in topic  $k$ . We calculate the mean pairwise Hellinger and KL distance (see Sec. 2.2.3) between the statistically significant words in each topic. Fig. 1b shows that the Hellinger distance (equal to  $1 - \text{similarity}$ ) scores range between 0.15 and 0.2, suggesting the words in each topic are quite similar. KL divergence (shown in black line), which does not necessarily lie in the range 0 and 1, correlates with and corroborates the Hellinger distance scores.

#### 3.2. Identification of significant states and industry types

A document  $d_i$  (representing the job losses for a state  $i$ ) is broken down into pre- and during-COVID timelines  $d_i^{pre}$  and  $d_i^{post}$  (as stated in Sec. 2.1). We record the Hellinger and KL distances between the  $d_i^{pre}$  and

$d_i^{post}$  for each state  $i$ ,  $d_i^{pre}$  and  $d_i^{post}$  can be written as a vector  $[\omega_0^k, \omega_1^k, \omega_2^k, \dots, \omega_{|V|-1}^k]$  (once again  $\sum_{i=0}^{|V|-1} \omega_i^k = 1$ ). Fig. 2a shows that Arkansas, Colorado, Connecticut, Georgia, Kentucky, North Carolina and Virginia (marked in red bars) exhibit Hellinger distance higher than cut-off 0.9. These states have the highest disparity in the job loss patterns in the pre- and during-COVID timelines.

In order to pinpoint the topics showing the highest variation between the pre- and during-COVID timelines, we plot the mean topic distribution of all the states. Fig. 2b shows that the mean score for topics 3 and 9 have increased and decreased, respectively, in the during-COVID timeline. Both these topics have retail as a common word (see Sec. 3.1), suggesting the following: one or more of the industry types (1) *hospitality, automotive, health care, mining* have suffered high job losses and (2) *finance, aviation and aerospace, food and beverages, and IT* have fired fewer professionals.

**Significant industry types.** We dig deeper to recognize the specific sectors that show an increase and decrease in job loss trends in the during-COVID word. To achieve this, we calculate the cumulative contribution of each word towards the pre- and during-COVID mean scores (reported in Fig. 2b). For a word  $w_i$ , its pre-COVID score (and analogously during-COVID score) is  $S_i^{pre} = \sum_k \omega_i^k \times m_i^k$ , where  $m_i^k$  is the mean pre-COVID weight of topic  $k$  in Fig. 2b. The final score for  $w_i$  is  $\frac{S_i^{post}}{S_i^{pre}}$ . The interpretation of  $S_i$  is as follows.

1.  $S_i \gg 1$ : proportion of job losses for industry type  $w_i$  increased in the during-COVID timeline.
2.  $S_i \sim 1$ : proportion of job losses for industry type  $w_i$  remained roughly the same.
3.  $S_i \ll 1$ : proportion of job losses for industry type  $w_i$  decreased in the during-COVID timeline.

Our analysis reveals that (1) the *recreational and philanthropic* sectors show a spike in job losses (i.e.,  $S_i \gg 1$ ). In the increasing order, religious institutions, recreational facilities and services, museums, arts and crafts, luxury goods and jewelry, computer gaming, hospitality, non-profit organization, performing arts, gambling have the highest  $S_i$ ; (2) *food distribution, politics, electronic media, research* sectors are largely unaffected (i.e.,  $S_i \sim 1$ ), as supermarkets, scientific research, political organizations, broadcast media and marketing match the pre-COVID weights; and (3) *the blue collar jobs, public service and print media* sectors exhibit a drop in job losses (i.e.,  $S_i \ll 1$ ). In the increasing order, market research, telecommunication, pharmaceuticals, aviation, food and beverage production, newspapers, textiles, consumer goods, printing show the least  $S_i$ .

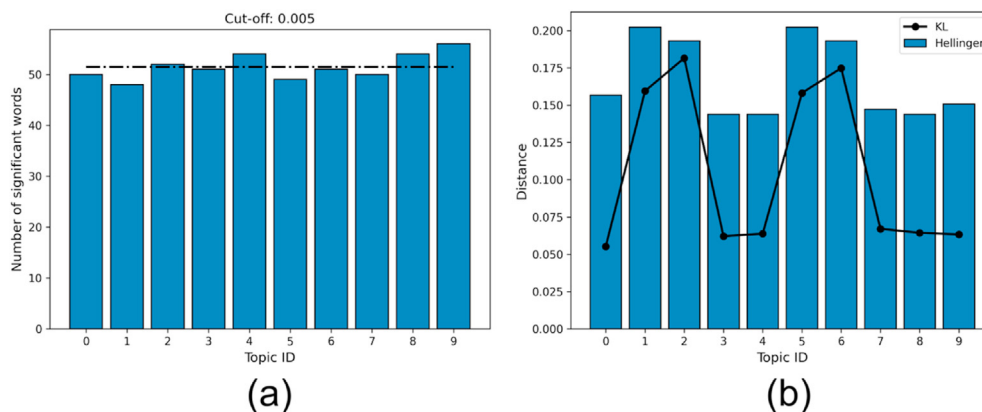


Fig. 1. Distribution of words in topics. (a) the statistically significant words (with cut-off 0.005) are fairly distributed across topics; (b) mean pairwise Hellinger and KL distance scores for words in each topic.

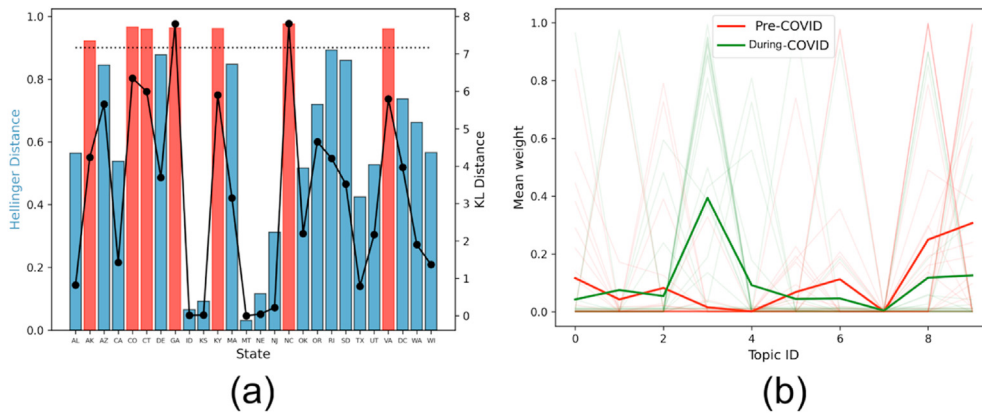


Fig. 2. Identification of significant states and industry types. (a) Hellinger and KL distances between the pre- and during-COVID timelines for each state; (b) mean weight of topics across states in pre- and during-COVID timelines.

3.3. Economic strata of unemployed workers

We utilize the job loss estimates due to COVID-19 in the US counties (see Sec. 2.1). Let the overall job loss rate and low income job loss rate of each US county  $c \in C$  be denoted by  $l_o(c)$  and  $l_{li}(c)$ . We create a vector  $V = \{I_{l_o(c)-l_{li}(c) \geq 0} | c \in C\}$ , where  $I_{l_o(c)-l_{li}(c) \geq 0}$  is an indicator variable set to 1 if  $l_o(c) - l_{li}(c) \geq 0$ , and 0 otherwise. We then apply the one-sample proportion Z-test (see Sec. 2.2.5) based on the proportion of counties where the overall job loss rates exceed low income job loss rates, i.e.,  $\pi = \frac{|\{v:v \in V \text{ and } v=1\}|}{|\{v:v \in V\}|}$ . We perform the one sample Z-test with the following null hypothesis –

**H0.** 70% US counties have higher overall job loss rates than low income job loss rates

We estimate the Z-scores for proportion = 0.60, 0.65, 0.7 and 0.75, Fig. 3 illustrates that the Z-scores and p values follow decreasing trends, as the proportions ( $\pi$ ) are statistically significant (i.e., p value =  $4.7 \times 10^{-9}$ , Z-score = 5.7) for up to 0.7 (or 70%). Thus, there is statistical evidence to reject the null hypothesis and state that the overall job loss rates are higher than low income job loss rates in over 70% of the US counties, or that the higher income job losses are more adversely affected by COVID-19.

3.4. Absorption into other sectors and reopening

There is little evidence of large-scale absorption (i.e., rehiring) of unemployed workers from one sector into another, or even into the same sector. There is no available dataset that enlists such migrations of workers to new sectors. It is worth mentioning, several artificial intelligence-based firms are attempting to minimize human contact by replacing human labor with digitization. This suggests that there is a possibility that unemployed workers will be “reskilled” in technology and inducted into the warehousing, manufacturing, and retail sectors in the near future (The Hill Vilas Dhar, 2020). As far as reopening is concerned, media reports suggest that most of the affected sectors have either stalled plans for reopening or are struggling to find consumers in the COVID “new normal” era. The U.S. Army Corps of Engineers, Jacksonville District, FL announced a reopening of Corps-managed recreational areas starting October 9, 2020, while ensuring necessary precautionary measures (U.S. ARMY CORPS OF ENGINEERS JACKSONVILLE DISTRICT, 2020). The administration of the City of San Jose, California decided to reopen parks and recreational neighborhoods on October 23, 2020 (City of San Jose, 2020). Museums across the US were originally contemplating reopening by May 2020 (Elassar, 2020), but at present most of them are looking at an indefinite shutdown and a few run the risk of closing forever (Vankin, 2020). Major jewelry outlets in the US

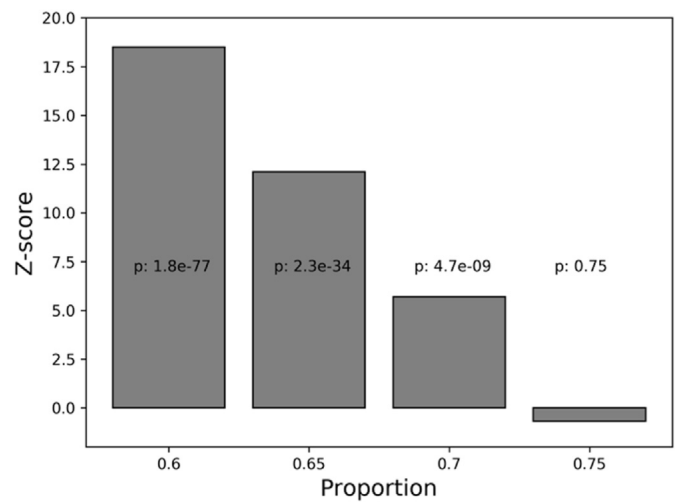


Fig. 3. Z-score and p values (in black) for the proportion of the US counties where the overall job loss rate exceed low-income job loss rates (<\$40,000 annual salary).

have reopened but are finding it difficult to garner sales (Doiron, 2020), and this trend is projected to continue for a long time (CNBC, 2020). The lodging real estate investment trusts of USA continue to under-perform as people feel wary to step out (Krishnan et al., 2020). Although, several hospitality businesses have resumed operation, they are struggling to find takers (Loh et al., 2020). Also, the performing arts sectors are considering pushing back reopening to next year (Cooper, 2020; Limbong, 2020).

4. Discussions

In this work we study the effect of COVID-19 on the American economy with respect to job losses. We apply a statistical approach, called topic modeling, on a combination of exhaustive datasets from Worker Adjustment and Retraining Notification (WARN) Act and repository of 7 million companies to unravel crucial findings on the worst affected US states and job sectors. While Arkansas, Colorado, Connecticut, Georgia, Kentucky, North Carolina and Virginia show a high incongruity in job loss patterns between the pre-and during-COVID timelines, recreational and philanthropic sectors record the highest job losses. At present, majority of the affected sectors are contemplating extended periods of shutdown or finding fewer consumers despite reopening.

Few of the observations made in course of this study stood out. First, it

is crucial to understand that the economic landscape will change immensely in the during-COVID world, i.e., some industries will take a hit, while others will get a boost. Second, our findings suggest that some industries may in fact continue to stay relevant due to the interplay of several socioeconomic factors. For instance, contrary to expectation, aviation and retail (with score  $S = 0.77$  and  $0.85$ ) show a lower proportion of job losses in the during-COVID timeline. Third, our analysis on the job loss datasets suggest that the higher income jobs are more adversely affected than the low income counterparts. This could be because some low income jobs, such as grocery stores, maintenance, food joint, security, etc., must stay operational despite the lockdown. This study can have wide-ranging implications in public policymaking to bolster economy and government subsidization of endangered sectors.

#### CRedit authorship contribution statement

**Satyaki Roy:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing - original draft. **Ronojoy Dutta:** Data curation, Formal analysis, Methodology, Software. **Preetam Ghosh:** Conceptualization, Formal analysis, Methodology, Writing - review & editing.

#### Declaration of competing interest

We have no conflict of interest.

#### Acknowledgement

This work is partially supported by NSF CBET-1802588.

#### References

- Adhikari, S., Meng, S., Wu, Y., Mao, Y., Ye, R., Wang, Q., Sun, C., Sylvia, S., Rozelle, S., Raat, H., et al. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (covid-19) during the early outbreak period: A scoping review. *Infectious diseases of poverty*, 9(1), 1–12.
- Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P., Joe, B., & Cheng, X. (2020). *Artificial intelligence and machine learning to fight covid-19*.
- Baker, S., Bloom, N., Davis, S. J., & Terry, S. (2020). *Covid-induced economic uncertainty*. National Bureau of Economic Research. Technical report.
- City of San Jose. (2020). *Parks, recreation neighborhood services news*. <https://www.sanjoseca.gov/Home/Components/News/News/1207/5103>.
- Cnbc. (2020). *Chanel warns virus impact will linger on luxury sector*. <https://www.cnbc.com/2020/06/18/chanel-warns-virus-impact-will-linger-on-luxury-sector.html>.
- Cooper, M. (2020). *The metropolitan opera won't reopen for another year*. <https://www.nytimes.com/2020/09/23/arts/music/metropolitan-opera-cancels-coronavirus.html>.
- Coronavirus world map: Which countries have the most cases and deaths?(2020). <https://www.theguardian.com/world/2020/may/06/>.
- Coronavirus: What have been the worst pandemics and epidemics in history?(2020). [https://en.as.com/en/2020/04/18/other\\_sports/1587167182\\_422066.html](https://en.as.com/en/2020/04/18/other_sports/1587167182_422066.html).
- Darling, W. (2011). A theoretical and practical implementation tutorial on topic modeling and gibbs sampling. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 642–647).
- Doiron, A. (2020). *Despite Virginia retail reopening, jewelry stores continue to struggle selling luxury items*. <https://wydaily.com/localnews/2020/05/29/despite-virginia-retail-reopening-jewelry-storescontinue-to-struggle-selling-luxury-items/>.
- Elassar, A. (2020). *This is where each state is during its phased reopening*. CNN. <https://www.cnn.com/interactive/2020/us/states-reopen-coronavirus-trnd/>.
- Gopinath, G. (2020). *The great lockdown: Worst economic downturn since the great depression*. Technical report <https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/>.
- Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1909(136), 210–271.
- Hoffman, M., Bach, F., & Blei, D. (2010). Online learning for latent Dirichlet allocation. In *Advances in neural information processing systems* (pp. 856–864).
- Holmdahl, I., & Buckee, C. (2020). Wrong but useful – what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*, 383(4), 303–305. <https://doi.org/10.1056/NEJMp2016822>
- Khan, N., Naushad, M., Fahad, S., Faisal, S., & Muhammad, A. (2020). *Covid- 2019 and world economy*. COVID-2019 AND WORLD ECONOMY.
- Kimball, K., Palder, D., & Mackinnon, A. (2020). *Tales from the lockdown: How covid-19 has changed lives around the world*. <https://foreignpolicy.com/2020/05/07/lockdown-covid-19-changed-lives-around-the-world/>.
- Krishnan, K., Mann, R., Seitzman, N., & Wittkamp, N. (2020). *Hospitality and covid-19: How long until 'no vacancy' for us hotels?*. <https://www.mckinsey.com/~media/McKinsey/Industries/Travel%20Transport%20and%20Logistics/Our%20Insights/Hospitality%20and%20COVID%2019%20How%20long%20until%20no%20vacancy%20for%20US%20hotels/Hospitality-and-COVID-19-How-long-until-no-vacancy-for-US-hotels-vF.pdf>.
- Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Limbong, A. (2020). *Broadway's reopening is pushed back again*. <https://www.npr.org/sections/coronavirus-live-updates/2020/10/09/922158893/broadwaysreopening-pushed-back-again>.
- Loh, T., Goger, A., & Liu, S. (2020). *'back to work in the flames': The hospitality sector in a pandemic*. <https://www.brookings.edu/blog/the-avenue/2020/08/20/back-to-work-in-the-flames-the-hospitality-sector-in-a-pandemic/>.
- Ramos, J., et al. (2003c). Using tf-idf to determine word relevance in document queries. In *Proceedings of the worst instructional conference on machine learning* (Vol. 242, pp. 133–142). New Jersey, USA.
- Randhawa, G., Soltysiak, M., El Roz, H., de Souza, C., Hill, K., & Kari, L. (2020). Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study. *PLoS One*, 15(4), Article e0232391.
- Rehurek, R., & Sojka, P. (2011). *Gensim – statistical semantics in python*. Retrieved from [gensim.org](http://gensim.org).
- Roy, S., & Ghosh, P. (2020). Factors affecting covid-19 infected and death rates inform lockdown-related policymaking. *PLoS One*, 15(10), Article e0241165.
- Smith, D. (2020). *When will covid end? Update on the race for a coronavirus vaccine*. <https://www.cnet.com/how-to/coronavirus-vaccine-pfizer-moderna-and-how-many-vaccine-doses-are-coming-in-2020/>.
- The Hill Vilas Dhar. (2020). *Covid-19 has destroyed millions of jobs | it's time to reskill America*. <https://thehill.com/opinion/technology/521559-covid-19-has-destroyed-millions-of-jobs-its-time-to-reskill-america>.
- U.S. Army Corps Of Engineers Jacksonville District. (2020). *Corps announces the second phase of reopenings at recreation facilities*. <https://www.usace.army.mil/Media/News-Releases/News-Release-Article-View/Article/2376972/corps-announces-the-second-phase-of-reopenings-at-recreation-facilities/>.
- Vankin, D. (2020). *16% of u.s. museums say they risk closing forever in a prolonged pandemic*. <https://www.latimes.com/entertainment-arts/story/2020-07-22/museums-risk-closing-permanently-covid-pandemic>.
- Walsh, B. (2020). *Covid-19: The history of pandemics* (BBC) <https://www.bbc.com/future/article/20200325-covid-19-the-history-of-pandemics>.
- Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 139, 110058.
- Wynants, L., et al. (2020). Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *Bmj*, 369.