



Towards an AI-Driven Marketplace for Small Businesses During COVID-19

Erik Coltey¹ · Daniela Alonso¹ · Shahin Vassigh² · Shu-Ching Chen¹

Received: 20 January 2022 / Accepted: 26 July 2022 / Published online: 11 August 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

Abstract

With the introduction of new COVID-19 variants such as Delta and Omicron, small businesses have been tasked with navigating a constantly changing business environment. Furthermore, due to supply chain issues, shortages of various critical products negatively affect businesses of all sizes and industries. However, continued innovation in Computer Science, specifically in sub-fields of Artificial Intelligence (AI), such as natural language processing (NLP), has created significant value for businesses through helpful data-driven features. To this end, we propose a platform utilizing AI-driven tools to help build an effective business-to-business (B2B) platform. The proposed platform aims to automate much of the market research which goes into selecting products and platform users during times of distress while still providing an intuitive e-commerce interface. There are three primary novel components to this platform. The first of these components is the Buyer's Club (BC), which allows customers to pool resources to purchase bulk orders at a reduced cost. The second component is an automated system utilizing Natural Language Processing (NLP) to detect trending disaster news topics. Disaster topic detection can be applied to inform buyers and suppliers on adapting to changing market conditions and has been shown to match closely with Google Trends data. The third component is a regulation matching system, using a custom data set to help inform customers when purchasing products. Such guidance is necessary to comply with a regulatory environment that will be irregular for the foreseeable future.

Keywords Natural language processing · Artificial intelligence · COVID-19 · Data mining · Data retrieval · Mobile · User interface

Introduction

The COVID-19 Pandemic could be considered the defining event of the past 2 years. Through the multiple virus waves and variants, public safety measures have taken their toll on everyone, including small businesses. From a survey completed by the US Census Bureau, it is estimated that 89.9 % of small businesses experienced a negative effect on operations due to COVID-19 [1]. Specific sectors, such as the service and hospitality sectors, were disproportionately affected. Many larger firms in these sectors rely on small businesses as suppliers and customers for B2B services [2]. Small businesses often form a significant percentage of industries that have high employment multipliers, so job losses in these industries could cause job losses in the broader economy [2]. Widespread closure of small businesses could obstruct larger firms that rely on them.

One of the primary concerns for a business of any size is procurement. For example, before vaccines were widely

This article is part of the topical collection “Innovative AI in Medical Applications” guest edited by Lydia Bouzar-Benlabiod, Stuart H. Rubin and Edwige Pissaloux.

✉ Erik Coltey
ecolt003@cs.fiu.edu

Daniela Alonso
dalon045@fiu.edu

Shahin Vassigh
svassigh@fiu.edu

Shu-Ching Chen
chens@cs.fiu.edu

¹ Knight Foundation School of Computing and Information Sciences, Florida International University, 11200 SW 8th Street, Miami 33199, FL, USA

² College of Communication, Architecture and the Arts, Florida International University, 11200 SW 8th Street, Miami 33199, FL, USA

available, shortages of Personal Protective Equipment (PPE) were extremely difficult to deal with. Many brick-and-mortar businesses had to close until they could adequately protect staff [3, 4]. Over time, shortages in other commodities have also begun to appear, with one present example being the current chip shortage. The lack of computer chips has caused uncertainty even in massive industries, such as automotive, which typically uses Just-In-Time (JIT) inventory management techniques [5]. Procurement issues can be quite dependent on location as well. For example, each US state has different regulations regarding COVID-19 response, which further complicates procurement if businesses have multiple locations.

To summarize, three primary goals need to be achieved to help aid small businesses during times of distress, such as COVID-19.

- Allowing small businesses to effectively procure products from one another which are difficult to find elsewhere.
- Ensuring that products and suppliers are relevant to buyers' needs.
- Informing businesses of the regulatory concerns when procuring key products.

To help meet these goals, we have extended our previous work in [6] on a platform for Business to Business (B2B) transactions specifically built with AI-driven features. Section “[Related Work](#)” describes related work on COVID-19 mobile applications, small business commerce, and NLP. In Sect. “[Digital Marketplace Implementation](#)”, the application architecture is discussed in detail to provide more thorough explanations for each component. These explanations include information on features, such as the Buyer's Club and Company Sourcing functions. Section “[Unsupervised Disaster Topic Detection](#)” proposes Unsupervised Disaster Topic Detection, a system using news information to help understand what buyers and suppliers may need from the platform at a given moment. Section “[Product Regulation and Guidance Matching](#)” provides additional detail and updates on Product Regulation and Guidance matching, showcasing regulations for products that are likely to be relevant. Finally, Sect. “[Conclusions](#)” summarizes the main takeaways and concludes this work.

Related Work

COVID-19 Response: Mobile Applications

Organizations had developed mobile applications to aid in the fallout of the COVID-19 Pandemic since early 2020 when COVID-19 first began to spread. Many of these applications come in the form of assessment tools, scientific

research applications, information dissemination tools, and tracking tools [7]. The vast majority of these applications deal with displaying or collecting information that directly relates to the status of the Pandemic in a given area.

There are relatively few mobile applications dedicated to addressing the downstream effects of the COVID-19 Pandemic across other factors, such as mental health and economic impact. Longyear et al. [8] argues for the importance of these sectors for helping improve a given population's well-being throughout the Pandemic. Rakshit et al. [9] Details the importance of mobile applications developed for businesses as a solution to help struggling small and medium-sized businesses, especially in developing countries. Figueroa and Aguilera [10] Proposes using mobile applications for services, such as telehealth to help vulnerable populations cope with the Pandemic.

Much research into creating mobile applications for COVID-19 response has dealt with ethics, and privacy concerns [11]. Many consumers are concerned that many applications designed to help the Pandemic collect what amounts to Personally Identifiable Information (PII) through location tracking or microphone/video capabilities. For example, many contact tracing applications require information on the movements of entire communities, making many uncomfortable as they feel monitored by the application administrators.

Small Business-to-Business Commerce

Small Businesses were hit especially hard at the beginning of the COVID-19 Pandemic, with surveys predicting that 43% of small firms will close by December 2020 [12]. As a result, many firms have been pivoting towards a Virtual Enterprise (VE) model during the Pandemic. In a VE, multiple complementary organizations form a temporary partnership to take advantage of economic opportunity or to ensure survival [13].

Business-to-Business (B2B) Commerce has become increasingly digitized since the advent of the internet. These trends have intensified since the COVID-19 Pandemic, leading to most businesses moving their operations online. [14] explains that digital sales interactions are 2–3 times more important than in-person sales interactions. In addition, many US B2B firms have increased spending throughout the Pandemic.

[15] Describes how AI systems can aid in different areas of the B2B sales funnel, especially when acquiring new businesses to join their service. For example, AI systems can generate leads, collect information on business needs and habits, and present the B2B firm's products. While AI is poised to enhance many parts of the funnel, it is necessary to have such systems be easy to use. In addition, providing intuitive user interfaces allows non-technical employees to leverage AI tools to enhance sales operations.

Natural Language Processing

NLP is one of the fastest-growing subfields in Machine Learning (ML), a subfield of AI. NLP focuses specifically on learning algorithms that can operate on natural text. One crucial task for NLP algorithms is embedding natural text into a fixed-length numerical representation. Embeddings are necessary to operate on text data further, either using supervised or unsupervised learning algorithms.

Some more traditional embedding methods used statistics across all documents in a corpus. One of the most naive embedding methods is to convert each word into an integer representing how many times it appears in a corpus. TF-IDF is a method of refining a count matrix described above by reducing the weight of tokens that appear across many documents. Therefore, TF-IDF helps avoid issues, such as words like “a” or “the” having outsized influence in an embedding.

The introduction of Deep Neural Networks (DNNs) has entirely changed perceptions on information processing [16]. Specifically, transformer architectures [17] have allowed document embeddings to take into account the context and position of tokens in the document. Multiple training methodologies for these transformer architectures have been proposed, using methods, such as masked language modeling (MLM) in BERT [18], permuted language modeling (PLM) in XLNet [19], and a mixture of the strengths of both in MPNet [20]. These models train on massive amounts of data concatenated from sources across the internet. Pretrained weights for these large-scale language models have enabled practitioners to solve many domain-specific NLP problems without training large models. Instead, all that is necessary is either finetuning existing weights for a specific task or using the resultant embeddings in more traditional ML systems.

Digital Marketplace Implementation

Application Architecture

Section 1 describes three distinct goals that had to be met to aid small businesses. The first goal relates to the idea that small businesses may need to procure products from one another. To this end, creating some form of marketplace software for small businesses to purchase products from one another is necessary. As initially described in [6], the base of this application is an e-commerce marketplace with functionality for buying products, selling products, and fulfilling orders, all through a mobile device. As demonstrated in Fig. 1, this marketplace consists of a mixture of typical back-end databases, REST APIs, AI-based systems, and interfaces to expose unique platform features to end-users.

Starting from the far left in Fig. 1, we have various data sets created manually or on-demand at specified time intervals using Selenium. Selenium is a library that can scrape website information using a web browser. Therefore, Selenium proved ideal for collecting data from sources, such as the Small Business Administration (SBA), which lack accessible APIs.

We then leverage this data with the help of various AI-driven systems. The first of these systems is Unsupervised Disaster Topic Detection, which collects data from news headlines to help inform users on the most relevant trending topics which may influence what products should be on the platform. The second system is the Regulation Matching System, which aims to populate product listings with relevant regulations. Sections [Unsupervised Disaster Topic](#)

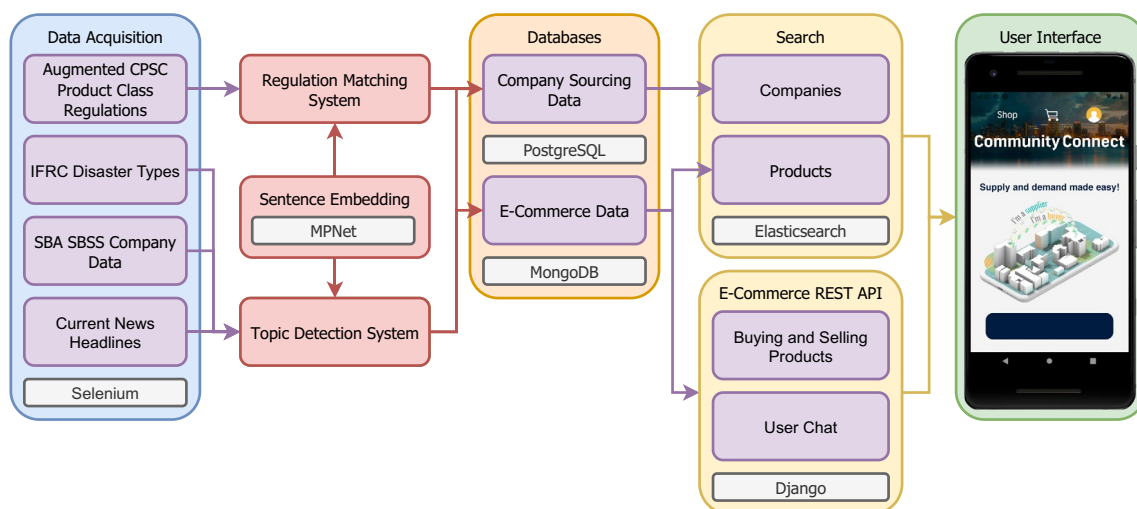


Fig. 1 Block diagram of the current software architecture of the AI-Driven Marketplace

Detection and **Product Regulation and Guidance Matching** provide more detail on how these systems are implemented.

The output information provided by these systems is then fed into two distinct databases. The first database allows administrators to search through collected company data. The second database drives the data management for the marketplace, holding information on user accounts, products, orders, and more. PostgreSQL was the first choice for the primary marketplace database due to its reliability, stability, and scalability. Next, a MongoDB data store was chosen for storing company sourcing data, as it provides better support for heterogeneous data, such as from the SBA and LinkedIn. MongoDB also made it easier to search documents with Elasticsearch, which typically expects document-based data.

Django provides the primary business-logic layer between the user interface and data stores mentioned above. Django allows for the development of business logic and REST APIs using Python. Python libraries for data science, such as Pytorch [21], and Scikit-Learn [22] could be imported directly to implement test AI systems without needing intermediate code in a different programming language.

ElasticSearch is primarily used to search through company sourcing data, to allow administrators to quickly identify leads to bring onto the platform based on keywords. ElasticSearch is a separate back-end service used to search through document-based data meaningfully. ElasticSearch is highly scalable, built to handle searching across large volumes of data.

Finally, React Native powers the front-end interface, which the end-user will access to utilize the marketplace. The main draw of React Native is the ability to create mobile

interfaces with a shared codebase for both iOS and Android. Having a shared codebase is essential in quickly iterating and more efficiently deploying the application to meet the needs of small businesses.

Buyer's Club

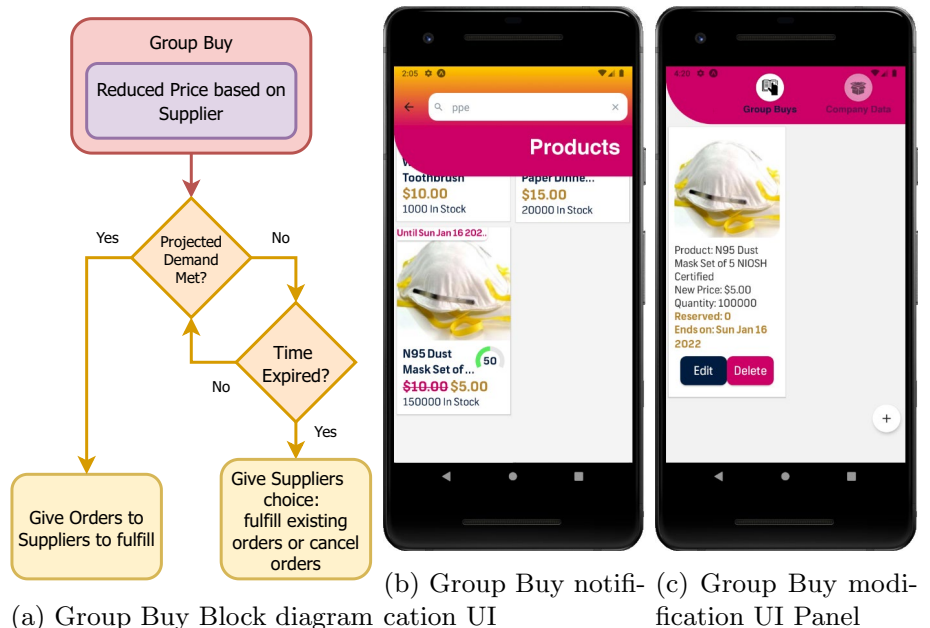
When building a new marketplace, there must be benefits for both buyers and suppliers. One way to create a win-win situation for small businesses is the addition of a Buyer's Club, as explained in [6]. For buyers, a Buyer's Club can take advantage of the collective buying power of individual members to buy large numbers of an item for a lower price. Suppliers benefit by obtaining large bulk orders, with the idea that pooling individual purchases can take advantage of economies of scale.

Historically Buyer's Clubs were created to gain access to difficult-to-acquire medical treatments, such as Hepatitis C [23], and could, in theory, be used for any product. For the implementation of a Buyer's Club, certain products are offered at a reduced price during an agreed-upon period between administrators and suppliers, similar to a sale.

As shown in Fig. 2a, these sales would be contingent upon the collection of a wide enough pool of different buyers who wish to reserve the product. If we meet the projected demand before time has expired, we will fulfill all the required orders. Otherwise, the group buy will wait until expiry, and suppliers will decide whether to fulfill orders or not.

Figure 2b, c shows how group buy information is displayed to users. Figure 2b provides a screenshot of a small banner over a product indicating a currently open group

Fig. 2 Block diagram and UI showcasing how Group Buys are created, modified, and executed on the proposed marketplace



buy. Figure 2c showcases an interface to modify group buys reserved for suppliers and administrators.

One of the most significant concerns with the proposed marketplace may involve what happens if suppliers decide to cancel group buys entirely. While this option may leave buyers with little recourse, negotiations between administrators and suppliers should minimize these unwanted outcomes. Many suppliers likely need some minimum inventory to be purchased before shipping makes economic sense. Therefore, group buy listings should ideally use this minimum or a slightly higher amount. Otherwise, creating group buys with excessively high minimum reservations will increase the likelihood of a cancelled group buy due to a lack of buyer interest.

Company Sourcing

Allowing administrators to collect information on relevant companies is essential to the early success of the proposed marketplace. Administrators could use company information to reach out to organizations that may help fix broken supply networks in the wake of events, such as COVID-19.

One way to collect company data is through directly sourcing companies in bulk using specific topical keywords. Mining data directly from places such as the SBA allows for the extraction of large amounts of data about a business. Figure 3a shows the different attributes that can be collected from the SBA, as well as an example of expected information. Some companies may have every attribute filled out, while others may be incomplete.

Figure 3b, c demonstrates how this system currently functions, with fields to enter certain keywords to look for businesses. For example, keywords, such as “PPE” can search for relevant businesses that supply Personal Protective

Equipment (PPE). The company sourcing system is built into the marketplace and is available to administrators.

Unsupervised Disaster Topic Detection

A major goal of the proposed marketplace is to help understand and predict the needs of buyers and suppliers during disaster situations, such as COVID-19. Some of how this goal presents itself in potential features of the marketplace include:

- Collecting relevant regulatory information
- Recommending products based on predicted need
- Automatically sourcing companies to become buyers or suppliers based on relevant keywords

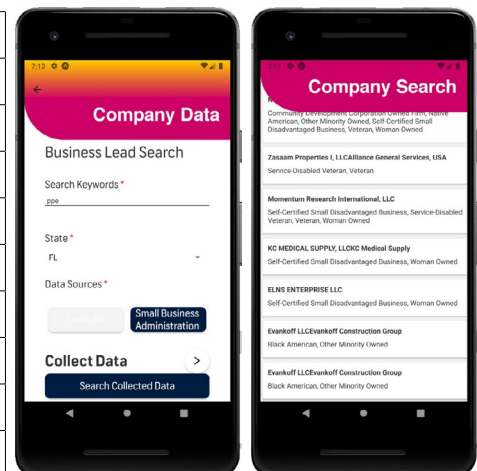
Trending topic detection is one way to help predict the needs of buyers and suppliers using data relating to current events. Given the COVID-19 Pandemic, we focus on disaster topic detection and downstream keyword extraction. Figure 4 provides a view of the entire end-to-end system for generating keywords from different topics detected in news headlines. Each component of this architecture diagram is explained in detail in the following subsections.

Data Sources

News headlines can be leveraged as a proxy for more traditional demand/trend data. Demand/trend data can typically only be acquired when a marketplace already has a relatively large user base. News content can help determine the currently trending topics in a particular category. For this task, we consider the number of headlines related to a specific

Fig. 3 Data attributes and UI showcasing how company data is collected and visualized for administrators through the Company Sourcing system

Attribute	Example
Name	Company I
Description	We sustainably manufacture PPE!
Website	http://CompanyI.com
Phone	777-777-7777
Address	1234 Main St.
Demographics	Small Business, Veteran Owned
Skills	PPE, Masks, Gowns
Industry	Manufacturing
Founding	2021



(a) Attributes which can be collected from the SBA (b) Company data collection UI (c) Collected company list UI

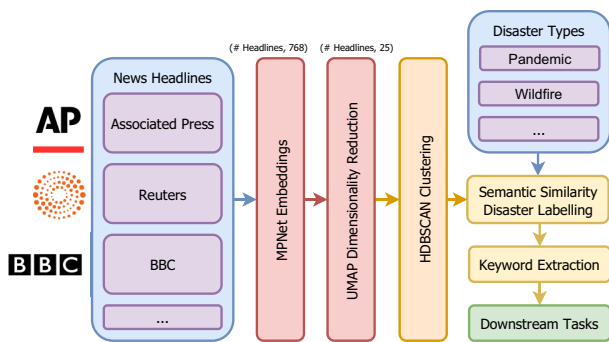


Fig. 4 Block diagram demonstrating how news article headlines are converted into trending topics and relevant keywords using the proposed Disaster Topic Detection system

topic as a coarse metric representing a topic's popularity and use only the titles of news articles. Titles typically already contain the keywords necessary for downstream tasks, such as keyword extraction. The A Million News Headlines data set is used to help validate the proposed topic detection system. This data set provides headlines from the Australian Broadcasting Corporation (ABC) from early 2003 to the end of 2020.

Headline Clustering

Sentence Embedding

News headlines are typically single sentences or phrases meant to garner public attention to an article. Therefore, the first step in processing these headlines is embedding them as a vector representing the content of that sentence for further processing.

To create meaningful clusters from headlines, we generate embeddings using TF-IDF on sentences filtered for punctuation and numerical values. Typically, news headlines are incomplete sentences and contain many keywords, so TF-IDF embeddings are favorable due to heavily weighing words unique to a topic.

For semantic similarity between sentences, headline embeddings are generated using a transformer model, pre-trained using Microsoft's MPNet [20]. MPNet aims to deal with the discrepancies of the masked (MLM) and permuted language modeling (PLM) methods used for the BERT and XLNet transformer models, and provide a unified objective that incorporates their strengths. The output of these embeddings is a 768-length vector representing the content of a given news headline.

Dimensionality Reduction

While the TF-IDF and MPNet embeddings can already be fed into various downstream NLP tasks, they are too high-dimensional for clustering. In addition, typical distance metrics, such as Euclidean Distance become much less helpful at very high dimensions. Therefore, we apply dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) [24].

UMAP works by constructing a high-dimensional graph of the input data and uses variable-sized radii to form connections between each data point. This high-dimensional graph is then weighted based on the distance between points and reprojected into the target lower dimensional space, ensuring the preservation of topological structure. UMAP's strength is the ability to better preserve the local and global data structure compared to other methods, such as t-SNE. Each headline embedding is reduced to 15 dimensions and then clustered.

Density-Based Clustering

News can be highly dynamic, with significant changes in subject matter over short timescales. Therefore, headline data is usually unsuitable for more traditional clustering methods, such as K-means, which require the number of clusters. In contrast, density-based clustering algorithms, such as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [25], create clusters based on the separation or concentration of data points. Density-based clustering can also label points as noise, which helps filter out news stories in outlier topics. Finally, HDBSCAN avoids the need to select an optimal ϵ value from DBSCAN by taking advantage of hierarchical trees.

HDBSCAN evaluation typically uses the Density-Based Clustering Validation (DBCv) technique [26]. DBCv assesses clustering quality based on the relative density between points and works with arbitrarily shaped clusters. For the marketplace implementation, DBCv is maximized using a grid-search across the minimum possible samples, minimum cluster size, and cluster selection ϵ metrics for HDBSCAN. The grid search is confined to a space between 1 and 10 for both minimum samples and cluster size. This relatively small search space is used, because if extended further, HDBSCAN may merge many smaller clusters into huge ones. In addition, these enormous clusters may have too much headline variation to meaningfully label the data as relating to any specific disaster type.

Algorithm 1 Disaster Headline Topic Detection

Require: Disaster topics ($D_1, \dots, D_m$)
Require: Array of sets for each cluster ($O_1, \dots, O_L$)
Require: 2D Array ($X_1, \dots, X_L$), where X_i are the headlines for cluster i

```

1: for  $i = 1, \dots, L$  do
2:    $\bar{X}_i \leftarrow \frac{1}{n} \sum_{j=1}^n X_{ij}$ 
3:    $O_i \leftarrow \{\}$ 
4:   for  $k = 1, \dots, m$  do
5:      $S \leftarrow \frac{\bar{X}_i \cdot D_k}{\|\bar{X}_i\| \times \|D_k\|}$ 
6:     if  $S > T$  then
7:        $O_i \leftarrow O_i \cup \{D_k\}$ 
8:     end if
9:   end for
10: end for
11: return  $O$ 

```

Topic Labelling

Once clustered, the mean embedding is calculated for all valid cluster labels $i = 1, \dots, L$, and is represented by $\bar{X}_i$. As shown in Algorithm 1, the semantic similarity is taken between the mean embedding and the embedding of each disaster topic. Semantic similarity results for each topic are then filtered using a threshold T , which defaults to 0.5 for this analysis. All disaster topic labels with a semantic similarity greater than T for the mean embedding of a given cluster i are assigned as a valid topic for that cluster. To this end, headline clusters may have multiple relevant disaster topics. 0.5 is an arbitrary selection based on the qualitative assessment that semantic scores greater than 0.5 will typically be pretty relevant. Administrators could modify T to help filter the topic detection results.

Disaster topic names are taken from The International Federation of Red Cross and Red Crescent Societies (IFRC) website. These include Earthquakes, Volcanic Eruptions, Epidemics and Pandemics, Wildfires, Cyclones, and others.

Keyword Extraction

With each cluster labeled with relevant disaster topics, it is now possible to extract related keywords. The reasoning behind including keyword extraction is that search typically requires keyword arguments. One such example is augmenting the Company Sourcing system proposed in Sect. 3.3, which uses the SBA database to collect company data. The proposed augmentation to this system would automatically

use keyword extraction to automate searching for businesses. The KeyBERT library [27] is used for this task, with slight modifications. Rather than using the eponymous BERT model to generate embeddings, the previously cached MPNet embeddings without dimensionality reduction are used. KeyBERT works by comparing the embedding of each token in a sentence with the sentence embedding itself using semantic similarity. These semantic similarity scores are then sorted, with the ability to return the top N keywords. A platform administrator can tune N for their ideal use case.

Model Validation

Table 1 demonstrates the popularity of different disaster types using the proposed topic detection method on the A Million Headlines data set from October 2019 to December 2020. This specific range of data shows a significant increase in epidemic/pandemic headlines when COVID-19 first began appearing in the news. Near the end of 2019, Australia was also experiencing a bushfire season, explaining the high number of Wildfire related headlines in December 2019 and January 2020. In November 2019, the number of clusters chosen was 2, which led to 0 values for every disaster type. Therefore, it was likely difficult for HDBSCAN to effectively cluster the headlines for that month. In this case, most headlines were allocated to one cluster, leading the topic detection to fail due to no disaster topic representing the mean embedding well.

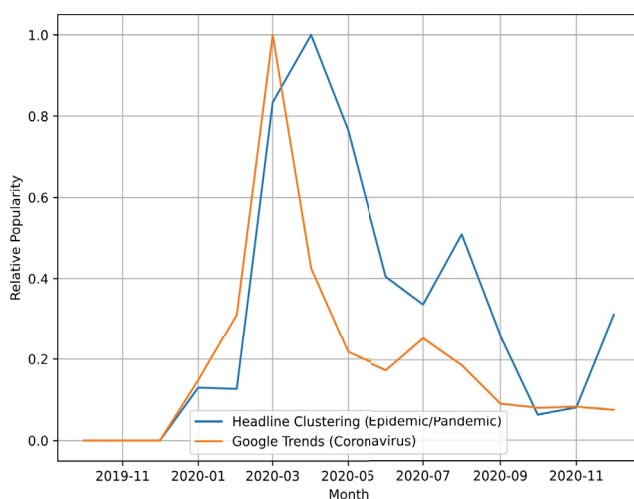
To further validate the model, Fig. 5 displays a comparison of the relative popularity between the Topic detection

Table 1 Unsupervised headline topic detection results for different disasters

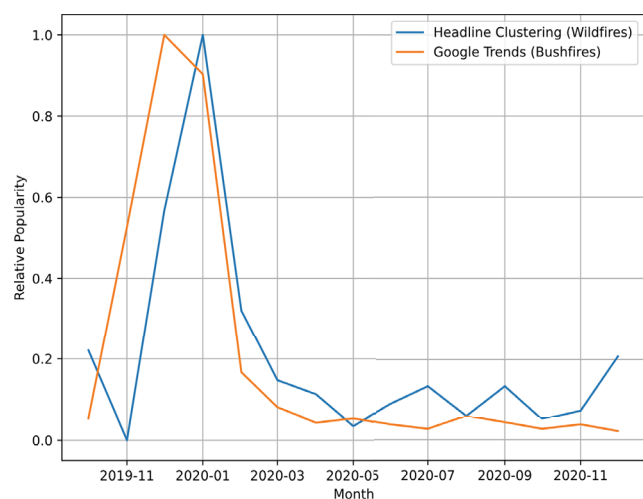
Month	DBC	Number of headlines per disaster type								
		Epidemic/ pandemic	Wildfire	Drought	Flood	Hailstorm	Volcano	Coldwave	Landslide	Cyclone
Oct 2019	0.49	0	51	0	0	20	0	0	0	0
Nov 2019	0.89	0	0	0	0	0	0	0	0	0
Dec 2019	0.53	0	131	3	0	0	41	0	0	0
Jan 2020	0.52	36	231	30	31	8	0	0	0	0
Feb 2020	0.52	35	74	9	39	0	0	0	0	21
Mar 2020	0.52	231	34	3	0	0	3	0	0	10
Apr 2020	0.53	277	26	8	6	0	0	3	0	0
May 2020	0.54	212	8	4	0	0	0	0	0	8
Jun 2020	0.50	112	21	6	0	0	0	9	0	0
Jul 2020	0.49	93	31	0	13	0	0	0	0	0
Aug 2020	0.50	141	14	21	17	3	0	0	0	0
Sep 2020	0.53	72	31	10	0	0	0	0	0	0
Oct 2020	0.50	18	12	19	12	33	0	0	0	0
Nov 2020	0.52	23	17	14	0	10	9	4	0	0
Dec 2020	0.50	86	48	0	8	0	7	0	5	0

Table 2 Selected keywords for each disaster topic analyzed

Disaster topic	Selected keywords
Epidemic/Pandemic	Coronavirus, outbreak, pandemic, recession, crisis, contagion, lockdown
Wildfire	Bushfire, farmland, grassfires, smoke, drought, sydney, mountain
Drought	Bushfires, drought, farmers, agriculture, outback, climate, summer
Flood	Flooding, rain, pier, docklands, downpour, victoria, sulawesi
Hailstorm	Shoalhaven, hailstorm, flood, hail, forecast, snow, brisbane
Volcano	Zealand, volcano, nz, eruption, erupts, indonesia, island
Coldwave	Covid, coronavirus, minks, rainfall, coldest, melbourne, victoria
Landslide	Bail, queensland, houses
Cyclone	Wa, cyclones, nsw, rain, queensland, cattle, perth



(a) Coronavirus and Epidemic/Pandemic



(b) Bushfires and Wildfires

Fig. 5 Normalized popularity between Google Trends data and headline results for COVID-19 and the Australian Bushfires

output and Google Trends for the Coronavirus Pandemic and the Bushfires in Australia. Relative popularity for topic detection is calculated by simply normalizing the number of headlines for a given disaster topic using the minimum and maximum values across the relevant period. Google trends give a popularity score for trends between 0 and 100, which is also normalized. The terms used for the topic detection and Google trends data are different as certain words are much more likely to be searched on Google. Therefore, these alternate words give a better idea of the popularity of a topic as opposed to less relevant search terms.

When comparing trends to headline popularity, the general shape is quite similar for both graphs, especially when looking at the bushfires. As the bushfire season ended after the peaks in January 2020, the topic was less relevant for users or news organizations soon after. On the other hand, the Epidemic/Pandemic topic loses popularity much more slowly than Google Trends. According to the survey in [28], many are fatigued by constant news, with about two-thirds feeling worn out. Such phenomena may help explain why news popularity may remain high even as user search popularity sharply declines.

As for keywords, some examples of detected keywords are shown from October 2019 to December 2020 for each disaster type in Table 2. We chose the top 3 keywords for each topic each month. Some disaster types have much fewer keywords than others, such as Landslides, due to that type being uncommon in the data. However, most of these keywords are related to the given disaster topic, including severity and location. Location keywords can help localize, where disasters are happening, which can be leveraged by the marketplace to curate users' products and services based on their location.

Product Regulation and Guidance Matching

When businesses need to procure products at scale to help meet critical obligations, it is essential that what they buy is what is needed for operations to run smoothly. For example, with developing situations, such as COVID-19, emergency regulations are drafted and applied, meaning businesses need to keep aware of legal changes. Therefore, allowing small businesses to see references to different regulations and guidance documents directly can be helpful for buyers in making purchase decisions.

Data Sources

To power the Regulation and Guidance Matching, it is first necessary to have an initial database of regulations that can be updated to provide relevant information for businesses. The Consumer Product Safety Commission (CPSC) provides a large amount of information on safety guidelines. The CPSC data set also provides relevant Electronic Code of Federal Regulations (E-CFR) references based on a product's type. As described in [6], these product classes and relevant E-CFR references are on the CPSC website. Some data was also manually generated relating to categories of standard COVID-19 equipment, such as PPE, from sources, such as FDA guidance documents from the early stages of the Pandemic. The final data set contains 266 distinct kinds of products, of which there are safety regulations in the E-CFR.

Algorithm 2 Product Regulation and Guidance Matching

Require: Product Name P

Require: CPSC class and regulation pairs $C = \{(c_i, r_i) : i \in \{1 \dots n\}\}$

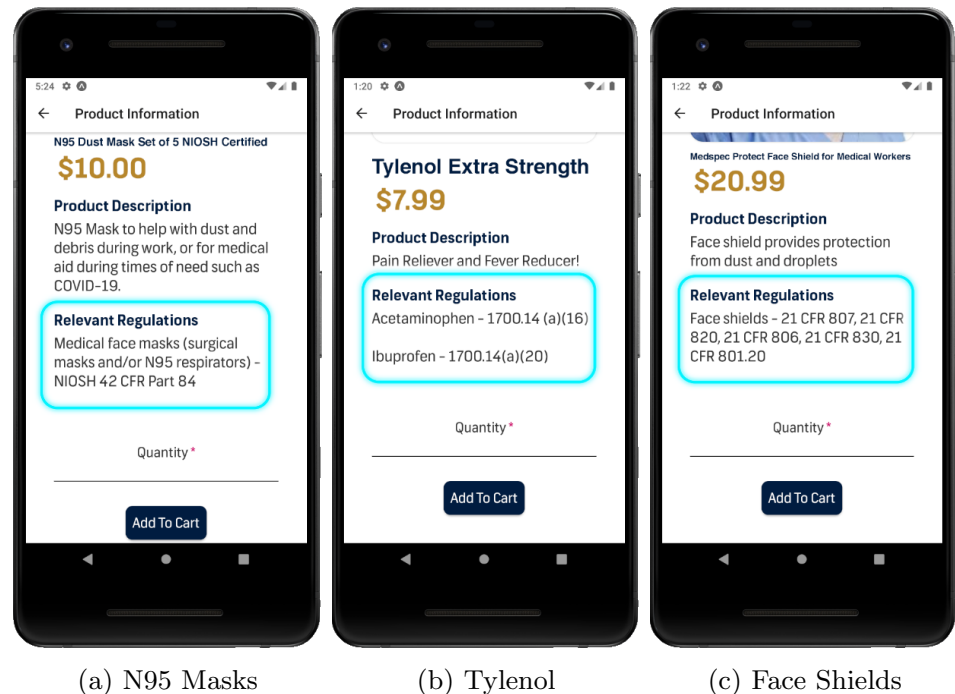
Require: List of output regulations O

```

1: for  $i = 1, \dots, n$  do
2:    $S \leftarrow \frac{P \cdot C_{ci}}{\|P\| \times \|C_{ci}\|}$ 
3:   if  $S > T$  then
4:      $O \leftarrow O \cup \{C_{ri}\}$ 
5:   end if
6: end for
7: return  $O$ 

```

Fig. 6 Relevant regulations provided for different types of COVID-19 related products based on their name using the proposed Regulation Matching system



Matching System

To perform matching, CPSC Product classes can be embedded and compared with a product's name, similar to the comparison between headlines and topics described in Sect. 4.2.4. One distinct difference from the previous version is the change in the embedding model. Previously in [6], the Universal Sentence Encoder (USE) was used to generate embeddings. However, for the same reasons described in Sect. 4.2.1, MPNet was used.

As shown in Algorithm 2, a product's name is compared against every product class in the CPSC data set. Product names and classes are not pre-processed before sending the data to MPNet for semantic similarity calculations. The lack of processing is because, typically, every character in a product name tends to be descriptive. For example, filtering out numeric characters can be misleading for N95 masks, as they may have different regulatory requirements than other masks. Removing numeric characters such as 95 would make it much more difficult to understand the nuances of the selected product. For each product class that meets a semantic similarity threshold T of 0.5, all the regulations for that product class are added to a final output array. $T = 0.5$ is selected as an arbitrary value. T can be tuned by an administrator or chosen based on user feedback, similar to the thresholding in Sect. 4.2.4.

To validate some of the results generated by the updated regulation matching system, Fig. 6 provides a qualitative view demonstrating some of the outputs for medical products used during COVID-19. Of particular interest is Fig. 6b,

which shows that even using the brand name Tylenol, the system can still see that it is similar to Acetaminophen, which is indeed accurate.

Conclusions

As the COVID-19 Pandemic continues to disrupt normal business operations for the foreseeable future, small businesses must be equipped with the tools to handle such change. Some of the goals which potential tools need to meet include:

- Small businesses need to be able to effectively procure products from suppliers or other small firms, which may be challenging to find due to shortages.
- Products need to be relevant to the needs of many buyers on the platform.
- Businesses should be aware of any regulatory concerns for products.

Given the advances in AI and, more specifically, NLP, tools driven by text data hold tremendous promise in helping meet these goals for small businesses. In this paper, we extended our proposed E-Commerce Marketplace with new data-driven features and more thorough explanations and updates to previously proposed features.

To help meet the first goal, the Buyer's Club and Company Sourcing features are presented together. The Buyer's Club aims to help businesses still procure products for lower

prices, given that many will buy products together to create bulk orders. On the other hand, Company Sourcing works by helping invite suppliers to put up products in the first place.

Unsupervised Disaster Topic Detection is an integral component of downstream systems that could meet the first two goals. In addition, results show that analyzing headline data can provide a proxy for large-scale trends, such as interest in COVID-19. Finally, keyword extraction can help further augment the platform by providing the marketplace with information on general disaster information and specific locations.

Tackling the third goal requires a system to communicate regulatory information to users. To this end, further explanations and updates are provided for the Product Regulation and Guidance Matching system, aiming to provide a solution to help businesses deal with changing regulations.

Acknowledgements This research is partially supported by NSF OIA-1937019, NSF OIA-2029557, and NSF CNS-1952089.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest to disclose.

References

1. Buffington C, Dennis C, Dinlersoz E, Foster L, Klimek S, et al. Measuring the effect of covid-19 on us small businesses: The small business pulse survey. US Census: Technical report; 2020.
2. Dua A, Ellingrud K, Mahajan D, Silberg J. Which small businesses are most vulnerable to covid-19-and when. McKinsey & Company ; 2020.
3. Cohen J, van der Meulen Rodgers Y. Contributing factors to personal protective equipment shortages during the covid-19 pandemic. *Preventive medicine*, 106263; 2020.
4. Ritter T, Pedersen CL. Analyzing the impact of the coronavirus crisis on business models. *Ind Marketing Manag*. 2020;88:214–24.
5. Cankurtaran P, Beverland MB. Using design thinking to respond to crises: B2b lessons from the 2020 covid-19 pandemic. *Ind Marketing Manag*. 2020;88:255–60. <https://doi.org/10.1016/j.indma.2020.05.030>.
6. Coltey E, Vassigh S, Chen S.-C. Community-connect: Covid-19 small business marketplace with automated regulation matching and company lead retrieval. In: 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), pp. 57–60;2021. IEEE
7. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, Votis K, Tzovaras D. Covid-19 mobile apps: a systematic review of the literature. *J Med Internet Res*. 2020;22(12):23170. <https://doi.org/10.2196/23170>.
8. Longyear R.L, Kushlev K. Can mental health apps be effective for depression, anxiety, and stress during a pandemic? *Practice Innovations* ;2021.
9. Rakshit S, Islam N, Mondal S, Paul T. Mobile apps for sme business sustainability during covid-19 and onwards. *J Business Res*. 2021;135:28–39.
10. Figueroa CA, Aguilera A. The need for a mental health technology revolution in the covid-19 pandemic. *Front Psychiatry*. 2020;11:523.
11. Cho H, Ippolito D, Yu Y.W. Contact tracing mobile apps for covid-19: Privacy considerations and related trade-offs. *arXiv preprint arXiv:2003.11511* 2020.
12. Bartik AW, Bertrand M, Cullen Z, Glaeser EL, Luca M, Stanton C. The impact of covid-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences*. 2020;117(30):17656–66. <https://www.pnas.org/content/117/30/17656.full.pdf>. <https://doi.org/10.1073/pnas.2006991117>
13. Jnr BA, Petersen SA. Examining the digitalisation of virtual enterprises amidst the covid-19 pandemic: a systematic and meta-analysis. *Enterprise Information Systems*. 2021;15(5):617–50.<https://doi.org/10.1080/17517575.2020.1829075>. <https://doi.org/10.1080/17517575.2020.1829075>
14. Bages-Amat A, Harrison L, Spillecke D, Stanley J. These eight charts show how covid-19 has changed b2b sales forever. McKinsey & Company. 2020;14
15. Paschen J, Wilson M, Ferreira JJ. Collaborative intelligence: How human and artificial intelligence create value along the b2b sales funnel. *Business Horizons*. 2020;63(3):403–14. <https://doi.org/10.1016/j.bushor.2020.01.003>.
16. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu M-L, Chen S-C, Iyengar SS. A survey on deep learning: algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*. 2018;51(5):1–36.
17. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A.N, Kaiser Ł, Polosukhin I. Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008;2017
18. Devlin J, Chang M.-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 2019.
19. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*. 2019;32
20. Song K, Tan X, Qin T, Lu J, Liu T.-Y. Mpnnet: Masked and permuted pre-training for language understanding. *arXiv preprint arXiv:2004.09297* 2020.
21. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inform Processing Syst*. 2019;32:8026–37.
22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. *J Mach Learning Res*. 2011;12:2825–30.
23. Vernaz-Hegi N, Calmy A, Hurst S, Jackson Y-LJ, Negro F, Perrier A, Wolff H. A buyers' club to improve access to hepatitis c treatment for vulnerable populations. *Swiss Med Weekly*. 2018;148:14649.
24. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* 2018.
25. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. *J Open Source Softw*. 2017;2(11):205.
26. Moulavi D, Jaskowiak P.A, Campello R.J, Zimek A, Sander J. Density-based clustering validation. In: *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 839–847 2014. SIAM
27. Grootendorst M. KeyBERT: Minimal keyword extraction with BERT. Zenodo. 2020. <https://doi.org/10.5281/zenodo.4461265>.

28. Gottfried J. Americans' news fatigue isn't going away—about two-thirds still feel worn out. Pew Research Center 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.