This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution is published in Advances in Intelligent Data Analysis XXI. IDA 2023. Lecture Notes in Computer Science, vol 13876, pp 195–207, and is available online at https://doi.org/10.1007/978-3-031-30047-9_16

AID4HAI: Automatic Idea Detection for Healthcare-Associated Infections from Twitter, A Framework based on Active Learning and Transfer Learning

Zahra Kharazian[1,3], Mahmoud Rahat[1], Fábio Gama[2], Peyman Sheikholharam Mashhadi[1], Sławomir Nowaczyk[1], Tony Lindgren[3], and Sindri Magnússon[3]

[1] Center for Applied Intelligent Systems Research (CAISR), Halmstad University, Sweden
{mahmoud.rahat, peyman.mashhadi, slawomir.nowaczyk}@hh.se

[2] Department of Innovation Management, Sweden, Halmstad University, Sweden
fabio.gama@hh.se

[3] Department of Computer and System Science (DSV), Stockholm University, Sweden
{zahra.kharazian, tony, sindri.magnusson}@dsv.su.se

Abstract. This research is an interdisciplinary work between data scientists, innovation management researchers, and experts from a Swedish hygiene and health company. Based on this collaboration, we have developed a novel package for automatic idea detection with the motivation of controlling and preventing healthcare-associated infections (HAI). The principal idea of this study is to use machine learning methods to extract informative ideas from social media to assist healthcare professionals in reducing the rate of HAI. Therefore, the proposed package offers a corpus of data collected from Twitter, associated expert-created labels, and software implementation of an annotation framework based on the Active Learning paradigm. We employed Transfer Learning and built a two-step deep neural network model that incrementally extracts the semantic representation of the collected text data using the BERTweet language model in the first step and classifies these representations as informative or non-informative using a multi-layer perception (MLP) in the second step. The package is called AID4HAI (Automatic Idea Detection for controlling and preventing Healthcare-Associated Infections) and is made fully available (software code and the collected data) through a public GitHub repository[4]. We believe that sharing our ideas and releasing these ready-to-use tools contributes to the development of the field and inspires future research.

Keywords: automatic idea detection · healthcare-associated infections · human-in-the-loop · active learning · feedback loops · supervised machine learning · natural language processing

[4] https://github.com/XaraKar/AID4HAI

1 Introduction

Healthcare-Associated Infections (HAIs) are among the most prevalent contamination events in healthcare settings, posing significant challenges to patient care. Multimodal interventions have been advocated, implemented, and studied to control and prevent HAIs, and cross-transmission of multidrug-resistant organisms worldwide [14]. Despite numerous efforts, the compliance of interventions among healthcare professionals remains below the WHO recommendations, hampering patient safety [8, 10]. Consequently, healthcare professionals, firms, and policymakers seek innovative ideas to increase interventions compliances and improve patient care. They now ask questions such as "what can healthcare professionals do in the real world to reduce HAIs?" and "how to prevent HAIs in different settings effectively?" Moreover, new knowledge is needed to promote behavior changes and education, monitor performance feedback, and create a safe climate.

Notably, the literature in idea identification has primarily investigated traditional methods involving interviews, ethnographic market research, repertory grid technique, and lead user workshops to identify novel ideas in healthcare settings [21]. Although prior research has highlighted important aspects to identify ideas and established some principles for best practices, it needs more details regarding how to scale the identification methods across different continents, reduce the operationalization costs, and ultimately seek local ideas discussed in real-time [22]. For example, Kesselheim et al. [11] conducted interviews to investigate the idea generation processes and clinical doctors' involvement in coronary artery stents but has been restricted to local settings. Likewise, Smith et al. [20] used text-matching algorithms on patents to investigate premarket approval applications in four medical device firms (Medtronic, Johnson & Johnson, Boston Scientific, and Guidant). Nevertheless, the study is restricted to secondary data, which undermines the possibility of identifying unpublished and potentially disruptive ideas.

A new and efficient way of identifying ideas is to use classification algorithms that can screen large amount of text and identify those bits of information that are more likely to contain ideas [4]. One way to access such a large pool of information is to use social media platforms such as Twitter and analyze the human-generated text using Natural Language Processing (NLP) [7]. The literature on NLP emphasizes using Transfer Learning to extract semantic representation in social media [15]. The *BERTweet* language model has been primarily investigated and provided valuable insights for various downstream tasks [17]. However, the potential of transfer learning and domain adaptation is not limited to text processing. We have explored this perspective in our previous studies in other domains and received outstanding results [12], [18].

Patients, healthcare professionals, scholars, and industry representatives are constantly using social media to communicate their needs and promote new healthcare practices [16]. This study conducts a retrospective observational analysis of Twitter user's posting related to HAIs. Ideas are identified using supervised machine learning, demonstrating how technologies such as artificial intelligence can advance HAI interventions. The idea is to analyze a set of tweets and rank them based on their probability of conveying an idea or a problem (aka informative tweets). The informative tweets form the minority class, which is also referred to as positive samples. On the other hand, the majority class is non-informative tweets and corresponds to the negative class. The proposed framework (AID4HAI) analyzes the collected HAIs ideas and validates the theoretical and practical implications of the approach with the help of a Swedish hygiene and health company. We employ Active Learning (AL) at the core of our framework to incrementally improve a discriminative model for finding as many potential ideas as possible.

In the Active Learning setup, we usually have a small labeled and a large unlabeled data pool. The goal is to pick samples from the pool of unlabeled data that produce the most significant improvement in the model's performance and then present the selected samples to the annotators for labeling, eventually adding them to the set of labeled data. This can be done in different ways. A popular approach is the *least confidence* [13] query strategy which chooses samples to query by considering the uncertainty of the classifier prediction. According to Chen et al. [2], uncertainty sampling methods are not a perfect solution for imbalanced scenarios since the majority class size is much larger than the minority one, and they will presumably query too many samples from the majority class. A more recent query strategy for handling imbalanced classification aims to find samples that are under-represented in the labeled data distribution. To this end, they either train an auxiliary binary classifier [6] to distinguish between labeled and unlabeled data or train an outlier detection algorithm [1] on the labeled data to score samples of the unlabeled pool.

However, these approaches do not promote (enforce) the selection of samples from the minority class in a manner that would be sufficient for our case. Despite improving the decision boundary, the ratio of the majority to the minority class is still preserved, and there is no mechanism to make the proportion more balanced. This drawback can potentially lead to the shortage of samples from the minority class in highly imbalanced datasets and ultimately degrades the performance of discriminative models. The proposed framework prioritizes the minority class by selecting samples that are predicted as more informative according to the feedback from the trained model. This helps to make the training dataset more and more balanced over the iterations.

The rest of this paper is structured as follows: Section 2 discusses how the dataset is collected, pre-processed, and labeled in detail. Section 3 brings up the proposed iterative method. Experiments and results are demonstrated in Section 4. Finally, the study's conclusion and future works are mentioned in Section 5.

2 Data Collection and Labeling

2.1 Data Collection

The Twitter platform has been chosen as a data source in this study. For extracting data, we selected a list of 78 HAI-related keywords and accounts with the help of experts in business and specialists in the healthcare domain. This list contains 21 personal accounts from famous Infection Prevention (IP) specialists with a high number of followers on the Twitter platform, 6 HAI-related journals, 15 public health organizations, 11 health and hygiene companies, and 25 HAI-related keywords. The following keywords and accounts were used:

Infection preventionists: Tom Frieden, Jason Gallagher, Debbie Goff, Marc Mendelson, Jon Otter, Eli Perencevich, Kevin Pho, Laura Piddock, Didier Pittet, Daniel Uslan, Marion Koopmans, Debbie Xuereb, Carole Hallam, Heather Loveday, Pat Cattini, Ermira Tartari, Karen Wares, Hannah, Evonne T Curran, Martin Kiernan, and Helen Dunn.

Journals: Infection Control & Hospital Epidemiology, Lancet Infectious Diseases, Journal American Medical Association, New England Journal Medicine, Journal of Infection and Prevention, and Journal of Hospital Infection

Organizations: CDCFlu, CDCGov, CDC, WHO NIAIDNews, ECDC_EU, SHEA_Epi, IDSAInfo, APIC, HIS_infection, IPS_Infection, IPSRnD, NHInfectPrevent, ips_epdc, IFH_HomeHygiene, and ESCMID

Companies: Purell, Clean hands safe hands, DEB, GWA, Hygiene, EcoLab, Georgia-Pacific, Ophardt, SaniNudge, Essity, and Tork.

Keywords: Cross Infection, Health acquired infection, Hospitalacquired infection, #hospitalacquiredinfection, Healthcare acquired infection, #Health acquired infection, Cross contamination, Nosocomial Infection, Healthcare-Associated Infection, Healthcare Associated Infection, Hand hygiene, Hospital Infection, Hand disinfection, Hand washing, Hand sanitizer, Infection control, Disinfection, Infection prevention, Decontaminate hands, Surgical site infection, Central line-associated bloodstream infections, Catheter associated urinary tract infections, Ventilator associated pneumonia, #HAI, and #HCAI

We searched Twitter by each of these queries and collected about 4.5 million tweets using the Twitter API v2. It resulted in a dataset containing selected HAI-related tweets posted from 2019 till the beginning of 2022. The dataset encompasses the tweet id, text, user id, time, and key metadata (i.e., number of likes, replies, and retweets) for each tweet. The collected tweets are in English, and each tweet has up to 280 characters.

2.2 Data Pre-processing

All the tweets were collected based on HAI-related keywords and also from the personal accounts belonging to infection prevention specialists whose goal is to teach and inform people about infection prevention. Although these accounts are all related to our goal, there is still no guarantee that all the tweets are on topic for our study. Therefore, we filtered out the collected tweets and kept only those that have at least one of the following terms: infection, health, contamination, nosocomial, healthcare, hand, hygiene, disinfection, prevention, decontaminate, surgical, bloodstream, catheter, urinary, ventilator, pneumonia, sanitizer, rub, hospital, disease, wash, control

Furthermore, inspired by Christensen et al. [3], we filtered the collected tweets once more with some ideation and problem terms to increase the chance of finding more ideas and problems in HAI. The terms are: need, problem, been, still, difficult, puzzle, can't, would, headache, would be, they would, i think, idea, and could be. This filtering process narrowed down the dataset to 692616 tweets (the unlabeled data pool).

2.3 Data Labeling

The collected data is unlabeled, and determining whether each tweet belongs to the "informative" or "non-informative" classes requires data annotation. For this purpose, three annotators with healthcare background were recruited and instructed to label the data independently. They were given instructions through some educational sessions by experts from a hygiene company and physicians from a hospital in Sweden. Also, some examples of informative and non-informative tweets are shown to annotators to familiarize them with the concept of controlling HAI. Then they were asked to read each tweet and label it based on this question: "Does the text below contain any information that can improve or create products or services related to HAI? (Mark 1 for Yes (informative) and 0 for No (non-informative))"

The annotation has been done in four iterations. In each iteration, a new batch of data is labeled by three annotators individually. This corpus of the labeled dataset is then published on a public GitHub repository and can be used for further studies.

After the annotation, we used our annotator's labels to identify each tweet's importance and label them. We assigned label 1 to those tweets that a majority of the annotators (2 out of 3, or unanimously) perceived as belonging to the "informative" class (referred to as 2- and 3-stars). Samples with no vote of informativeness (0-star) belonged to the "non-informative" class. Samples with only one vote (1-star) are ambiguous; in a sense, their label is unclear, making it difficult to judge to which group they belong. On the one hand, they are not "informative" since they failed to receive the majority; on the other hand, they should not be considered "non-informative" since one of the annotators voted them as informative. These ambiguous samples are removed from the data and neither used for training nor evaluating the models.

3 Methodology

3.1 Model

In this project, we used Transfer Learning and built a two-step deep neural network model that identifies the HAI-related informative tweets from non-informative ones. Transfer Learning is an ML technique for transferring the knowledge learned from one domain to another domain. More specifically, we took advantage of the transformer layers of the *BERTweet* language model [17], which has been previously trained on 850 million general English tweets, as a first step to extract the semantic representation of our HAI-related English tweets. Then, we use these representations as input for the next layer of the model, which is a multi-layer perceptron (MLP) chosen for classifying tweets. The portion of data used for training, validation, and testing is 60%, 20%, and 20%, respectively. The training process stops when the validation loss does not improve after ten epochs using the early stopping method. The structure of the model can be seen in Fig. 1.

BERTweet has the same architecture as the BERT (Bidirectional Encoder Representations from Transformers) [5] and consists of transformer layers and self-attention heads. On the other hand, the MLP we use as the classification head is structured from max-pooling, batch normalization, a dense layer with a ReLU activation function, a dropout layer, and another dense layer with a softmax activation function. Overall, this large model contains approximately 135 million parameters.

3.2 Active Learning in Data Labeling

One of the challenges of this project is how to train an ML model with unlabelled and highly imbalanced data. In this case, since we are working with a highly skewed dataset, most ML models will fail to distinguish samples from the minority class, and their decision will be biased toward the majority class.

Moreover, manually labeling a large-scale set of tweets (by humans) is laborious, time-consuming, and costly. This challenge, together with dealing with highly imbalanced data, motivated us to employ an iterative method based on Active Learning and feedback loops to boost the chance of finding more positive samples among the pool of unlabeled and imbalanced datasets. According to a survey on active learning [19], the main advantage of this learning approach is that it enables the model to achieve good performance even by training on relatively small number of labeled samples.

Active learning is a method that can be used for optimizing the labeling process by prioritizing samples to query an expert/oracle for labeling or correcting the labels. In this research, based on our need to find more informative samples within the pool of imbalanced and unlabeled data, we favor a query strategy that allows the active learner to select those samples with a higher probability of having ideas in the minority class. We called it the *"Richest Minority"* query strategy. These probabilities are the predictions produced by the trained model in each iteration.

Our work uses a two-step deep neural network model including *BERTweet* language model and a multi-layer perception (MLP) to classify and evaluate the tweets. The output of this model is a vector of probabilities produced by the softmax activation function that indicates how much each tweet is assumed to belong to the informative class. By sorting these probability scores from high to low and following the *Richest Minority* query strategy, we can select the most informative set of tweets to query the oracle. We employed this query strategy in our iterative algorithm and managed to detect a number of informative ideas and problems for controlling and preventing healthcare-associated infections.

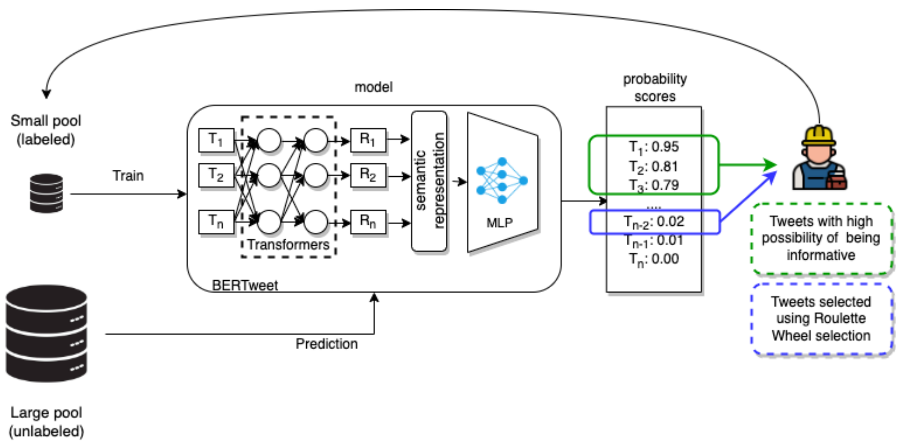


Fig. 1: Proposed algorithm based on Active Learning and Transfer Learning

Fig. 1 and Algortihm 1 demonstrate the proposed iterative method. The algorithm follows the below steps:

1. A small subset of data is selected based on each tweet's metadata score, including the number of likes, replies, retweets, and quotes. To do so, we normalized each of these features using the MinMax scaler method, summed up these four normalized values, and considered it a new feature called "rate." Afterward, we sorted the rate values for each set of tweets grouped by their keywords/accounts and selected the first three tweets plus a random tweet among the rest. This procedure yielded a diversified "small pool," which contains 586 tweets. The rest of the unlabeled data is stored in the "large pool."
2. The small pool is labeled by human annotators.
3. Labeled samples (if any) from the previous iteration(s) are added to the small labeled pool.
4. The two-step model is used to extract the semantic representation of each tweet and train a classifier on them.
5. The trained model is used to predict the probability class score of the remaining unlabeled data.
6. To find more positive samples, using the *richest minority* query strategy, the 700 most informative samples (i.e., ones with the highest probability scores), as well as 300 random samples (using the *roulette wheel selection* method) are selected as the next small batch of data to query for annotation. The reason for adding random samples selected using the roulette wheel method

is to avoid the Echo Chamber phenomenon [9], where the same or similar ideas are repeatedly discovered in the dataset.

7. Go back to step 2

Algorithm 1 Proposed algorithm

$large_pool \leftarrow$All filtered Tweets ▷ size:692616
$small\_pool \leftarrow$[]
for $i \leftarrow 1$ *to* 4 **do**
  if $i = 1$ **then**
    $selected\_tweets \leftarrow Initial\ Seeds.\ Tweets\ with\ high\ rates$ ▷ size:586
  else
    $selected\_tweets \leftarrow richest\_minority\_query(large\_pool, \hat{y})$ ▷ size:1000
    large_pool $-=$ $selected\_tweets$ ▷ remove the selected tweets from large pool
  end if
  Annotate(selected_tweets)
  $small\_pool$ += $selected\_tweets$ ▷ add selected tweets to small pool
  $model.fit(small\_pool)$
  $\hat{y}$ ←model.predict(large_pool)
end for

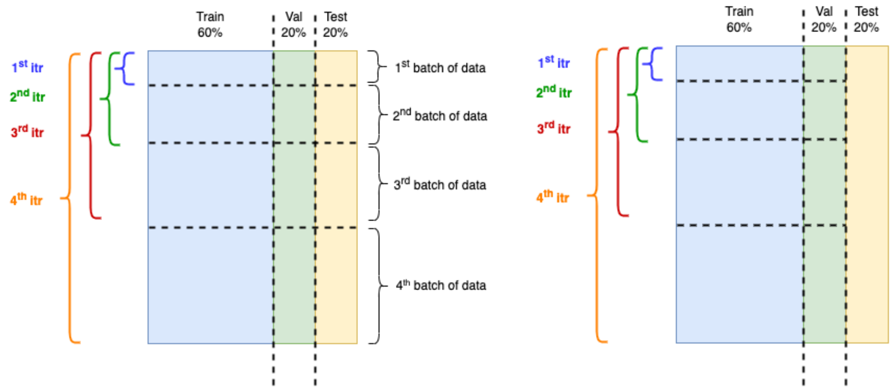
Following the proposed iterative algorithm, Table 1 demonstrates the number of informative and non-informative samples in different iterations. The "3-star", "2-star", "1-star", and "0-star" columns show the score of tweets. The "Informative" column shows the summation of 3 and 2-star tweets, while the "non-informative" column shows the number of 0-star tweets in each iteration. "Aggregated informative" is the number of accumulated informative samples from the current and previous iterations. This growing data is used to retrain the model in further iterations.

4 Experiments and Results

We have designed an experiment to evaluate the performance of the proposed framework based on two data split configurations. Fig. 2a shows the portions of data used to train models in each iteration. For instance, the first model is

Table 1: Statistics of the labeled data in all iterations

iteration	3-star	2-star	1-star	0-star	total	informative	non-info	agg informative	agg non-info
1st	15	42	122	407	586	57	407	57	407
2nd	19	85	152	731	987	104	731	161	1138
3rd	26	90	197	664	974	116	664	277	1802
4th	17	81	196	676	970	98	676	375	2478
total	77	298	667	2478	3517	375	2478		



(a) first configuration (b) second configuration

Fig. 2: The data split used for the experiment. Note that the second configuration uses the same test set for all iterations

trained, validated, and tested on 60, 20, and 20 percent of the first batch of data. This proportion is kept for all other iterations but with the difference that the data for the rest of the iterations are being aggregated (from that iteration and the previous ones). The second configuration uses a similar setting (Fig. 2b). The only difference is that the test set here is the collection of test sets from all iterations. The purpose is to evaluate the model's performance without the influence of various test sets.

In each iteration, the performance of the trained model is evaluated and reported using the f1-score measure for each class and the macro average of both classes. It is known that the f1-score evaluation metric is less sensitive to class imbalance. Furthermore, another evaluation metric, the Area under the Precision-Recall Curve (PR-AUC), is calculated to assess the performance of the models.

One common practice for training models with imbalanced datasets is to add weight to the samples from the minority class. To evaluate the effect of adding sample weights, we trained the same model twice, once in the presence and once in the absence of sample weights. We assign weights equal to one to the samples from the majority class, while the samples from the minority class receive weights of 2 or 3 according to their respective number of votes. In other words, positive samples with two votes get weight 2, and positive samples with three votes get weight 3. The samples with only one vote are being ignored, as mentioned earlier. Tables 2 and 3, respectively, represent the results for the first and second data splits.

Comparing the values from "f1-score macro avg" of "non-weighted" and "weighted samples" in Table 2, one can conclude that assigning weight to the samples, improves the performance of the model, as expected. On average, the "f1-score macro avg" has increased by 7 percent over four iterations. The amount

Table 2: Comparison of the performance of the trained model on normal and weighted samples in each iteration

iteration	non-weighted samples				weighted samples			
	f1-score informative	f1-score non-info	f1-score macro avg	PR-AUC	f1-score informative	f1-score non-info	f1-score macro avg	PR-AUC
1st	0.00	0.9333	0.4667	0.5625	0.4000	0.9492	0.6746	0.6718
2nd	0.9459	0.9909	0.9684	0.9565	1.00	1.00	1.00	0.9782
3rd	0.8872	0.9785	0.9329	0.8644	0.9552	0.9914	0.9733	0.9055
4th	0.7627	0.9729	0.8678	0.7960	0.7692	0.9673	0.8683	0.7791

Table 3: Comparison of the performance of the trained model over iteration both for normal and weighted samples

iteration	non-weighted samples				weighted samples			
	f1-score informative	f1-score non-info	f1-score macro avg	PR-AUC	f1-score informative	f1-score non-info	f1-score macro avg	PR-AUC
1st	0.00	0.9392	0.4696	0.5572	0.2655	0.9201	0.5928	0.3174
2nd	0.3974	0.9091	0.6532	0.4349	0.3638	0.8108	0.5872	0.5004
3rd	0.4444	0.9529	0.6987	0.6081	0.6032	0.9513	0.7772	0.6288
4th	0.6435	0.9605	0.8020	0.6830	0.6719	0.9596	0.8154	0.6924

of improvement decreases gradually from the first to the last iteration. In the first iteration, sample weighting helped the model by 20.79 percent. This value decreased to 0.05 percent in the last iteration. Our hypothesis to explain this phenomenon is that as we go through the iterations, the training set's size increases, reducing the need for adding sample weights. Also, by comparing the "PR-AUC" of "non-weighted" versus "weighted samples" in each iteration, we can see this number has increased chiefly when the model has trained on weighted samples. A similar pattern for the effect of sample weight can be concluded from the results of Table 3.

By comparing the macro average f1-score of the trained models in consecutive iterations (see Table 3), we can see that the ability of the model to distinguish between informative and non-informative tweets gradually increases both for weighted and non-weighted scenarios. The performance starts from 0.46 and increases all the way to 0.80 for the non-weighted samples. The corresponding numbers for the weighted samples show an increase from 0.59 to 0.81.

Moreover, by subtracting the value of the f1-score of the informative class from the f1-score of the non-informative class in consecutive iterations, we can see this value is relatively high in the first iteration (0.93 for non-weighted and 0.65 for weighted). This number gradually decreased to the last iteration (0.31 for non-weighted and 0.28 for weighted). Also, by comparing the "PR-AUC" of the trained model on "non-weighted" versus "weighted samples" in consecutive iterations in Table 3, we can see this value has increased from 0.55 to 0.68 for "non-weighted" and has increased from 0.31 to 0.69 for "weighted samples." These patterns show the performance improvement of the model on classifying the imbalance dataset over iterations.

5 Conclusions and Future Works

The main contribution of this paper is to introduce a full framework capable of discovering ideas and problems to control and prevent Healthcare-Associated Infections (HAI). This framework contains a corpus of 4.5 million HAI-related tweets posted from 2019 till the beginning of 2022 using the Twitter API v2 from the Twitter platform. Moreover, our work introduces an iterative machine learning method based on active learning and feedback from the model's decision. It selects the informative tweets based on the novel *richest minority* query strategy. The collected and labeled dataset, as well as the algorithm's code, are published in a GitHub repository called AID4HAI.

In our experiments, the proposed framework managed to discover 375 informative HAI-related ideas and problems, within the four iterations. The ideas and problems concern a number of various topics and directions. Figure 3 plots a handful of automatically extracted ideas/problems from Twitter. Our innovation team helped us visualize the core idea of each tweet across a two-dimensional chart. The x-axis represents the spectrum of ideas suggesting products to services. The y-axis spreads ideas based on behavior or technological-driven.

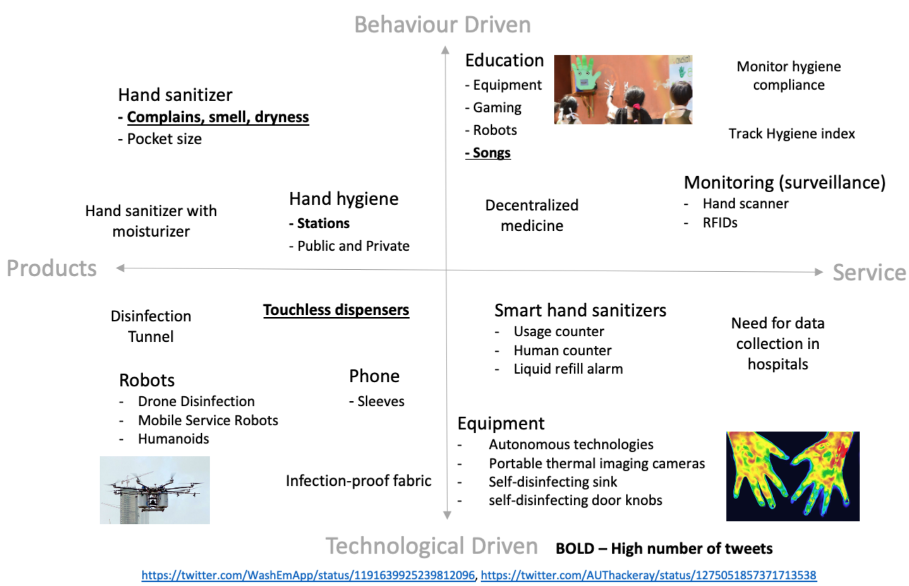


Fig. 3: Examples of extracted ideas plotted on a chart with an x-axis (service/product spectrum) and a y-axis (behavior/technology-driven).

The deep neural network model used in this study categorizes a tweet based on its informativeness. As a future work, it would be interesting to evaluate the tweets across additional dimensions of interest. Moreover, one could visualize the vectors from the model's attention layers and validate if they are focusing on the sensible tokens of the text.

Acknowledgment: The work was supported by research grants from KK-Foundation, Scania CV AB and the Vinnova program for Strategic Vehicle Research and Innovation (FFI).

We would like to express our utmost gratitude to Håkan Lindström and Peter Blomström and their team for their valuable help and professional support during this project. Also, we would like to thank our annotation team for helping us to provide this labeled data.

References

1. Barata, R., Leite, M., Pacheco, R., Sampaio, M.O., Ascensão, J.T., Bizarro, P.: Active learning for imbalanced data under cold start. In: Proceedings of the Second ACM International Conference on AI in Finance. pp. 1–9 (2021)
2. Chen, Y., Mani, S.: Active learning for unbalanced data in the challenge with multiple models and biasing. In: Active Learning and Experimental Design workshop In conjunction with AISTATS 2010. pp. 113–126. JMLR Workshop and Conference Proceedings (2011)
3. Christensen, K., Liland, K.H., Kvaal, K., Risvik, E., Biancolillo, A., Scholderer, J., Nørskov, S., Næs, T.: Mining online community data: The nature of ideas in online communities. Food Quality and Preference **62**, 246–256 (2017)
4. Christensen, K., Scholderer, J., Hersleth, S.A., Næs, T., Kvaal, K., Mollestad, T., Veflen, N., Risvik, E.: How good are ideas identified by an automatic idea detection system? Creativity and Innovation Management **27**(1), 23–31 (2018)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Gissin, D., Shalev-Shwartz, S.: Discriminative active learning. arXiv preprint arXiv:1907.06347 (2019)
7. Goff, D.A., Kullar, R., Newland, J.G.: Review of twitter for infectious diseases clinicians: useful or a waste of time? Clinical Infectious Diseases **60**(10), 1533–1540 (2015)
8. Gould, D., Lindström, H., Purssell, E., Wigglesworth, N.: Electronic hand hygiene monitoring: accuracy, impact on the hawthorne effect and efficiency. Journal of Infection Prevention **21**(4), 136–143 (2020)
9. Guo, L., A. Rohde, J., Wu, H.D.: Who is responsible for twitter's echo chamber problem? evidence from 2016 us election networks. Information, Communication & Society **23**(2), 234–251 (2020)
10. Irgang, L., Holmén, M., Gama, F., Svedberg, P.: Facilitation activities for change response: a qualitative study on infection prevention and control professionals during a pandemic in brazil. Journal of Health Organization and Management (2021)
11. Kesselheim, A.S., Xu, S., Avorn, J.: Clinicians' contributions to the development of coronary artery stents: a qualitative study of transformative device innovation. PLoS One **9**(2), e88664 (2014)
12. Kharazian, Z., Rahat, M., Fatemizadeh, E., Nasrabadi, A.M.: Increasing safety at smart elderly homes by human fall detection from video using transfer learning approaches. In: 30th European Safety and Reliability Conference (ESREL2020)

& 15th Probabilistic Safety Assessment and Management Conference (PSAM15), Venice, Italy, 1-5 November, 2020 (2020)
13. Lewis, D.D.: A sequential algorithm for training text classifiers: Corrigendum and additional data. In: Acm Sigir Forum. vol. 29, pp. 13–19. ACM New York, NY, USA (1995)
14. Lotfinejad, N., Peters, A., Tartari, E., Fankhauser-Rodriguez, C., Pires, D., Pittet, D.: Hand hygiene in health care: 20 years of ongoing advances and perspectives. The Lancet infectious diseases **21**(8), e209–e221 (2021)
15. Malte, A., Ratadiya, P.: Evolution of transfer learning in natural language processing. arXiv preprint arXiv:1910.07370 (2019)
16. Martischang, R., Tartari, E., Kilpatrick, C., Mackenzie, G., Carter, V., Castro-Sánchez, E., Márquez-Villarreal, H., Otter, J.A., Perencevich, E., Silber, D., et al.: Enhancing engagement beyond the conference walls: analysis of twitter use at# icpic2019 infection prevention and control conference. Antimicrobial Resistance & Infection Control **10**(1), 1–10 (2021)
17. Nguyen, D.Q., Vu, T., Nguyen, A.T.: Bertweet: A pre-trained language model for english tweets. arXiv preprint arXiv:2005.10200 (2020)
18. Rahat, M., Mashhadi, P.S., Nowaczyk, S., Rognvaldsson, T., Taheri, A., Abbasi, A.: Domain adaptation in predicting turbocharger failures using vehicle's sensor measurements. In: PHM Society European Conference. vol. 7, pp. 432–439 (2022)
19. Settles, B.: Active learning literature survey (2009)
20. Smith, S.W., Sfekas, A.: How much do physician-entrepreneurs contribute to new medical devices? Medical care pp. 461–467 (2013)
21. Thune, T., Mina, A.: Hospitals as innovators in the health-care system: A literature review and research agenda. Research Policy **45**(8), 1545–1557 (2016)
22. Weigel, T., Goffin, K.: Creating innovation capabilities: mölnlycke health care's journey. Research-Technology Management **58**(4), 28–35 (2015)