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], Fabio Gama[2], Peyman Sheikholharam Mashhadi[1], Slawomir Nowaczyk[1], Tony Lindgren[3], Sindri Magnusson[3], and Håkan Lindström[4]

***Abstract*— This study is a collaboration between data scientists, innovation management researchers from academia, and experts from a hygiene and health company. The study aims to develop an automatic idea detection package to control and prevent healthcare-associated infections (HAI) by extracting informative ideas from social media using Active Learning and Transfer Learning. The proposed package includes a dataset collected from Twitter, expert-created labels, and an annotation framework. Transfer Learning has been used to build a two-step deep neural network model that gradually extracts the semantic representation of the text data using the BERTweet language model in the first step. In the second step, the model classifies the extracted representations as informative or non-informative using a multi-layer perception (MLP). The package is named AID4HAI (Automatic Idea Detection for controlling and preventing Healthcare-Associated Infections) and is publicly available on GitHub.**

I. INTRODUCTION

Healthcare-Associated Infections (HAIs) are a significant problem in healthcare settings. Various interventions have been implemented to prevent and control HAIs, including multimodal approaches, but healthcare professionals' compliance with these interventions remains below the recommended level by the World Health Organization (WHO) [2], [3]. Therefore, there is a growing need for innovative ideas to increase compliance and improve patient care. Healthcare professionals and firms are asking questions about how to effectively prevent HAIs and create a safe climate. New knowledge is required to address these challenges. A more effective approach to identify ideas is to use classification algorithms that can quickly scan large amounts of text and identify information that is likely to contain ideas [1]. Social media platforms such as Twitter provide access to a vast pool of information that can be analyzed using Natural Language Processing (NLP) techniques. The literature on NLP recommends using Transfer Learning to extract semantic representations from social media data. The BERTweet language model [5] has been extensively studied and has provided valuable insights for various downstream tasks. The goal is to analyze a set of tweets and prioritize them based on the likelihood of conveying an idea or problem, which are referred to as informative tweets. Informative tweets form the minority class, while non-informative tweets form the majority class. The proposed approach employs Active Learning (AL) to gradually enhance a discriminative model for identifying as many potential ideas as possible. The framework's theoretical and practical implications are validated with the help of domain experts.

II. DATA COLLECTION AND ANNOTATION

A. Data collection: In this study, Twitter is used as the source of data. About 4.5 million HAI-related tweets were collected using the Twitter API v2 by searching Twitter with each of the 78 HAI-related queries. The dataset contains selected HAI-related tweets posted from 2019 till the beginning of 2022. The selection of HAI-related queries was made with the help of experts in business and specialists in the healthcare domain, and the 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.

B. Data annotation: The data collected from Twitter for the study is not labeled and needs to be annotated to determine whether each tweet is informative. Three healthcare experts were recruited as annotators and trained with educational sessions led by experts from a hygiene company and physicians from a hospital in Sweden. After annotation, tweets that received a majority of two or three votes (3-star and 2-star) were labeled as informative and assigned a label 1. Non-informative tweets with no vote (0-star) received a label of 0. Tweets that received only one vote (1-star) were considered ambiguous and removed from the dataset since their label was unclear. These tweets were neither used for training nor evaluating the models.

III. METHODOLOGY

Active learning is a technique that can enhance the labeling process by selectively choosing which samples to query an expert for labeling. In this study, a query strategy called "Richest Minority" was developed to select informative samples from imbalanced and unlabeled data. This strategy prioritizes samples with a higher likelihood of belonging to the minority class, based on predictions made by the trained model in each iteration. A two-step deep

This work was supported by research grants from KK-Foundation, Scania and Vinnova program for Strategic Vehicle Research and Innovation (FFI)

[1] Center for Applied Intelligent Systems Research, Halmstad University, Sweden

[2] Department of Innovation Management, Halmstad University, Sweden

[3] Department of Computer and System Science, Stockholm University, Sweden

[4] Essity Hygiene and Health AB, Gothenburg, Sweden

neural network model was created using Transfer Learning to identify informative tweets related to HAI. This project used the BERTweet language model's transformer layers to extract semantic representation from HAI-related English tweets. Then, a multi-layer perceptron (MLP) was used to classify the tweets. The model's structure can be seen in Fig. 1. The algorithm follows the bellow steps.

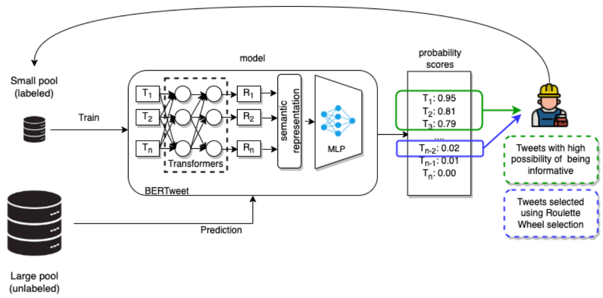

Fig. 1: Proposed algorithm based on Active Learning and Transfer Learning

1) Select and label a small subset of data.
2) Add to the small labeled pool the labeled samples from previous iteration(s).
3) Extract the semantic representation of each tweet and train a classifier on them.
4) Predict probability class score of remaining unlabeled data.
5) Selecting 700 samples using the "richest minority" query strategy and 300 random samples.
6) Go back to step 2

Table I displays the results of the proposed iterative algorithm, which shows the number of informative and non-informative samples in each iteration. The scores of tweets are categorized into 3-star, 2-star, 1-star, and 0-star columns. The "Informative" column shows the total number of 3-star and 2-star tweets, while the "non-informative" column displays the number of 0-star tweets in each iteration. The "Aggregated informative" column indicates the cumulative number of informative samples from the current and previous iterations, which is used to retrain the model in subsequent iterations.

IV. Experiments and Results

The study conducted an experiment to evaluate the performance of a proposed framework using two data split configurations. The first configuration involved training, validating, and testing the model on 60%, 20%, and 20% of the first batch of data, respectively, with subsequent iterations aggregating data from previous iterations. The second configuration was similar, but the test set comprised all test sets from all iterations. To address imbalanced datasets in model training, adding weight to samples from the minority class is a common practice. To evaluate the impact of sample weights, the study trained the same model twice, once with and once without sample weights. Weights of 1 were assigned to samples from the majority class, while samples from the minority class were assigned weights of 2 or 3, depending on their number of votes. Tables II and III show the results for the first and second data splits, respectively.

TABLE I: 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		

TABLE II: 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 III: 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

From Table II, the values of both "f1-score macro avg" and PR-AUC for "weighted samples" were higher than those for "non-weighted" samples, indicating that assigning weights to samples improves the model's performance, as expected. On average, the "f1-score macro avg" increased by 7 percent over four iterations. Similar findings were observed in Table III. In Table III, comparison of the macro average f1-score of the trained models in consecutive iterations demonstrated that the model's ability to differentiate between informative and non-informative tweets gradually improved, regardless of whether the samples were weighted or non-weighted.

V. Conclusions

The primary contribution of this paper is the introduction of a comprehensive framework for identifying ideas and issues related to controlling and preventing HAI, utilizing a corpus of 4.5 million HAI-related tweets collected via Twitter API v2. The paper also proposes an iterative machine learning approach based on active learning and feedback from the model's decision, which selects informative tweets using the innovative richest minority query strategy. The labeled dataset and algorithm code are shared in a GitHub repository named AID4HAI. To read the full paper please refer to [4].

References

[1] 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)

[2] 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)

[3] 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)

[4] Kharazian, Z., Rahat, M., Gama, F., Mashhadi, P.S., Nowaczyk, S., Lindgren, T., Magnússon, S.: Aid4hai: Automatic idea detection for healthcare-associated infections from twitter, a framework based on active learning and transfer learning. In: Advances in Intelligent Data Analysis XXI: 21st International Symposium on Intelligent Data Analysis, IDA 2023, Louvain-la-Neuve, Belgium, April 12–14, 2023, Proceedings. pp. 195–207. Springer (2023)

[5] Nguyen, D.Q., Vu, T., Nguyen, A.T.: Bertweet: A pre-trained language model for english tweets. arXiv preprint arXiv:2005.10200 (2020)