Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231005

The Role of Data Stream Mining in Improving Self-Medication Practices in the Syrian Arab Republic

Labeb Abood [1]
Belgorod State University, Belgorod 308007, Russian Federation

Abstract. Self-medication is a widespread practice throughout the world, especially in the Syrian Arab Republic. However, if done incorrectly, this method might be harmful. Data stream mining, which involves analyzing vast amounts of data from multiple sources, has proven to be an effective technique for enhancing self-medication practices. The role of data stream mining, and how it may evaluate information from various sources like social media, electronic health records, pharmacy sales data, sensor-based medical devices, etc. have been explored for the research. Healthcare professionals in Syria can improve patient outcomes and safety related to self medication by employing data stream mining to uncover important information about patient behaviors, drug effectiveness, and adverse events. In this paper, a strategy for generating evidence-based medical data utilizing data stream mining techniques is suggested. Here, methods including association rule mining, categorization, and data clustering have been described. For future progression of this research, the information gap and other data collection-related problems need to be resolved. The implementation demonstrates that healthcare professionals can use a variety of data stream mining techniques to better understand drug usage patterns and spot opportunities while also learning about the prevalence of self-medication in different parts of the Syrian Arab Republic.

Keywords. Self-medication, data stream mining, Syria, data mining in healthcare, data clustering.

1. Introduction

By providing real-time analysis of health data that can spot changes and patterns in a person's health, mining data streams can enhance self-medication. The term "self-medication" is a common term in the medical field. People engage in daily self-care of their health, and practice self-medication in the form of personal care. [1] Self-medication has traditionally been defined as "the taking of drugs, herbs or home remedies on one's own initiative, or on the advice of another person, without consulting a doctor." [2]

There are no accurate statistics about self-medication specifically for those countries where the medical sector lacks many basic facilities to fulfill the demand of the people. The Syrian Arab Republic is one of them. It's difficult to identify how many people

[1] Corresponding Author, Labeb Abood, Belgorod State University, Belgorod 308007, Russian Federation; E-mail: marcabood028@gmail.com.

exercise self-medication and the consequences related to this practice. The lack of access to the information necessary to improve "self-medication" in Syria, make it more difficult to implement the proposed approach. For this reason, a unique method has been created to gather the data in a continuous flow. Memory errors can be decreased and dynamic elements of health phenomena can be better understood by combining self-reported data with other reality mining data streams [3]. In the areas of disease risk assessment and clinical decision support, the use of big data mining in clinical care has been examined [4].

The use of data stream mining in self-medication practices has several benefits. Firstly, it can help identify the most commonly used drugs and their indications, which can aid in the development of evidence-based treatment guidelines. Secondly, it can help healthcare providers identify potential drug interactions and side effects, which can reduce the risk of adverse events. It also can assist healthcare professionals in identifying individuals who are more susceptible to prescription misuse or abuse, allowing them to take early action and minimize harm. Taking into account all of these, an effort has been made through this research to improve the practice of self-medication in Syria.

Overall, data stream mining has the potential to revolutionize self-medication practices in Syria and other parts of the world. By leveraging this technique, healthcare providers can improve patient outcomes, reduce healthcare costs, and enhance the overall quality of care.

2. Problem Statement

In Syria, the need for improved self-medication practices is particularly acute due to the ongoing conflict and resulting healthcare system challenges. According to a study conducted at Damascus Hospital in Syria, 67.3% of 453 adult inpatients practiced self-medication, with analgesics, antipyretics, and antibiotics being the most commonly used drugs [5].

Another study found that the 7-day prevalence of medicine use among Syrian adults was 34.9%, indicating a significant percentage of individuals practicing self-medication [6]. However, it is essential to note that self-medication with antibiotics as prophylactics against COVID-19 has also been observed, with an estimated 19.5% of participants engaging in this practice [7]. These findings suggest that self-medication is widespread in Syria and highlight the need for further investigation to control its risks, such as antibiotic resistance and misdiagnosis.

Therefore, the use of data stream mining could be particularly beneficial in this context. By analyzing data from multiple sources, healthcare providers can gain a more comprehensive understanding of drug usage patterns and identify opportunities for intervention.

3. Research Methodology

The information from numerous medical devices and search engines was used to construct the dataset for this research. While different search engines are used to gather data on self-medication activities, medical devices are utilized to collect data on vital signs like heart rate, blood pressure, and others. Using data stream mining techniques,

the data is gathered in real-time. The process of dataset creation and stream mining is defined in figure 1.

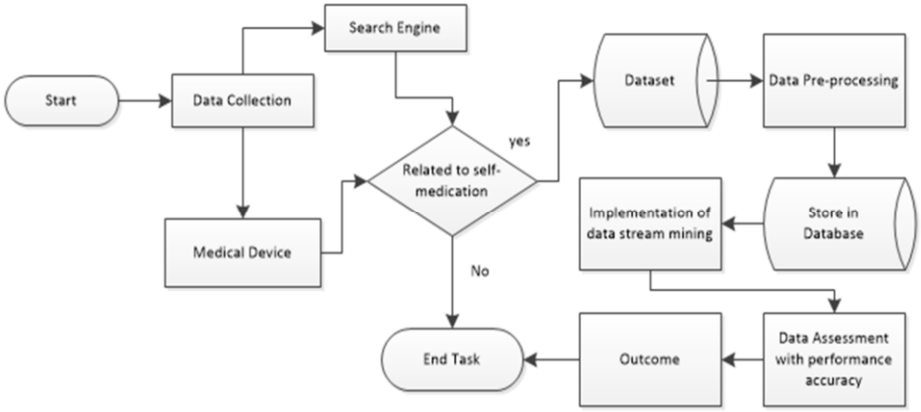

Figure 1. The process of data collection and implementation of stream mining techniques.

The collected data is preprocessed to remove any noise and outliers. This involves data cleaning, data transformation, and data reduction techniques. We then stored the cleaned data in a database for further analysis.

The cleaned data is analyzed using three data stream mining techniques:

- Clustering,
- Classification, and
- Association rule mining.

While classification is used to forecast the likelihood of self-medication activities based on specific criteria, clustering is used to group similar data points together. To find patterns and connections between self-medication practices and other factors, association rule mining is performed.

The analysis's findings are assessed using measures for performance accuracy. By contrasting the outcomes of the data stream mining techniques with the source data, the evaluation will be carried out.

4. Implementation

Collected data related to self medication from various regions of Syria has been prepared for implementing different data steaming techniques. Various search engines, such as Google, Yandex, Yahoo, Bing, and YouTube, have been utilized for data collection. Health-related trending topics in the Syrian Arab Republic have been examined. Then, the topics that had the highest percentage of exploration were chosen using data stream mining techniques described below -

4.1 Data stream Clustering

The pre-processed data is used for applying stream clustering. It is one of the most suitable methods for real-time data stream processing, because it can be applied with less prior information about the data and it does not need labeled instances. [8]

A Python implementation of the data stream clustering algorithm named "DenStream" is applied for the collected data. [9] Where micro clusters have been used. Micro-clusters are a popular technique in stream clustering, which maintain the compact representation of the clustering. [10]

For clustering text documents, a spherical k-means clustering algorithm is generally used. [11] This multivariate numerical data clustering is done using a traditional clustering method. Documents d_i, are represented by feature vectors x_i, clustering this data into k groups is to minimize the function:

$$\sum_i d(x_i, p_{c(i)}) \quad (1)$$

Where, $p_{c(i)}$ is the function of centroid representing an assignment of c of objects i to cluster ids $c(i) \in \{1,\ldots, k\}$ for a suitable similarity measured. [11]

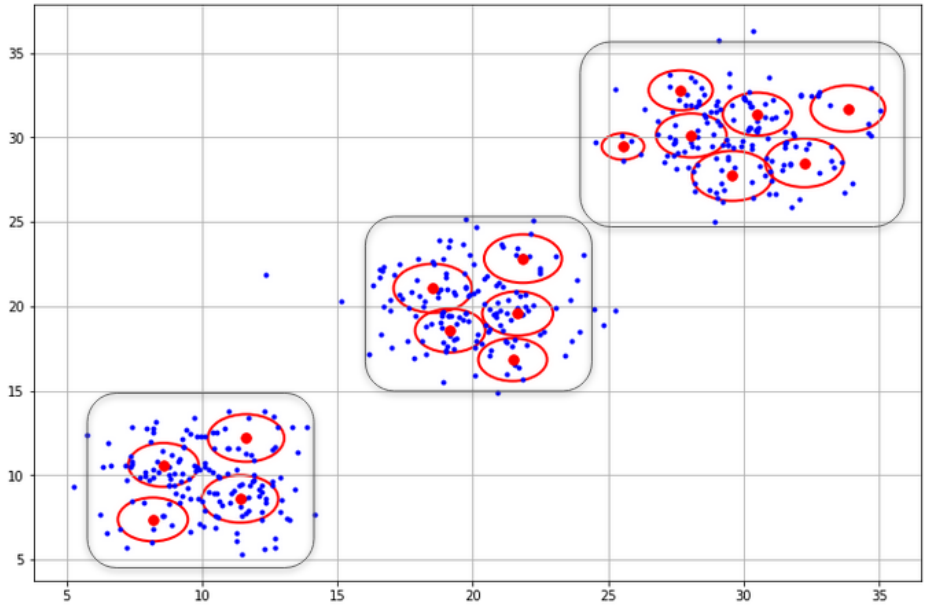

Figure 2. Random sample of potential micro-cluster using DenStream. [9]

4.2 Data classification

Data classification is the process of classifying data into appropriate groupings (or "classes") based on shared traits, such as their degree of sensitivity, the dangers they pose, and the compliance laws that protect them. [12]

4.3 Association rule mining

Association rule mining (ARM) is used to define the relation between a large number of data objects. [13]

5. Results and Analysis

The result shows Cholera has been a health related most discussed topic for the last 12 months in the Syrian Arab Republic.

Table 1. Keywords categorization using search queries in search engines for the Damascus region in the Syrian Arab Republic.

Keywords	Total Search (Damascus)
cholera	4384
cholera symptoms	662
cholera treatments	41
cholera home remedies	33
how to treat cholera at home	67
cholera medicine	987
cholera complications	598
cholera side effects	521

Having this in mind, these topics and questions were selected for the categorization of the keywords of the dataset. The information is gathered for various locations of Syria using different search engines, and then it is integrated to use data stream techniques.

The data reveals that searches for self-medication are more prevalent in the Damascus area. Given that medical equipment was only examined from there, a variety of data mining approaches were used. The percentage of keyword searches in figure 3 is significantly consistent across all of Syria.

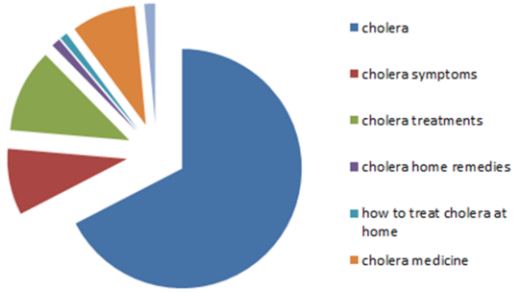

Figure 3. An average percentage for various cholera-related keywords in the Syrian Arab Republic.

While other strategies also had an average accuracy of over 80%, the data classification technique demonstrated an accuracy of over 85%.

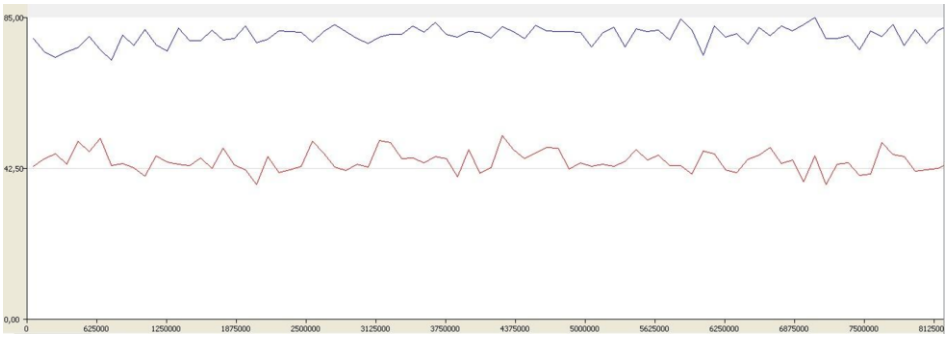

Figure 4. Average accuracy, and loss of the training and test set.

MOA (Massive online Analysis) framework has been used to evaluate methods for mining tasks on evolving data streams over the full space only. [14] With the assistance of around 4,000 participants over the course of 30 days, a second individual study was conducted in the Damascus region, and the sample data was gathered while taking the findings of the earlier studies into consideration. To assess the situation relating to cholera self-medication, data has been gathered from various medical devices and surveys.

The analysis of the result shows the following findings -

- Women look for self-medication more than men.
- Patients have a true sign of eagerness in order to improve the practice.
- Proper information, guidelines, and awareness can improve this practice rapidly among both men and women.
- It is necessary to look for the data which are not only written in English but also in Arabic.
- A high-quality dataset with an approach for collecting data can make the result more accurate.

6. Conclusion

Data stream mining has become a potent tool for enhancing self-medication practices and unearthing insightful information from a variety of healthcare data sources. The study has shown the method's potential for detecting self-medication-related patient behaviors, pharmacological efficacy, and adverse events in the Syrian Arab Republic. Healthcare practitioners can create evidence-based medical data to improve patient outcomes and safety by using techniques including association rule mining, classification, and data clustering. To assure the efficacy and dependability of data stream mining in healthcare, more research is necessary to address the knowledge gap and data collection-related issues. Overall, the application of data stream mining techniques can give healthcare practitioners a plethora of knowledge and enable them to make informed decisions.

References

[1] Vizhi SK, Senapathi R. Evaluation of the perception, attitude and practice of self-medication among business students in 3 select Cities, South India. International Journal of Enterprise and Innovation Management Studies (IJEIMS) July-December. 2010;1(3):40–4.

[2] Hernandez-Juyol M, Job-Quesada JR. Dentistry and self-medication: A current challenge. Med Oral. 2002;7:344–7.

[3] Pentland A, Lazer D, Brewer D, Heibeck T. Using reality mining to improve public health and medicine. Stud Health Technol Inform. 2009;149:93-102. PMID: 19745474.

[4] Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in Clinical Medicine. Chin Med J (Engl). 2016 Mar 20;129(6):731-8. doi: 10.4103/0366-6999.178019. PMID: 26960378; PMCID: PMC4804421.

[5] Abdelwahed RNK, Jassem M, Alyousbashi A. Self-Medication Practices, Prevalence, and Associated Factors among Syrian Adult Patients: A Cross-Sectional Study. J Environ Public Health. 2022 Jun 28;2022:9274610. doi: 10.1155/2022/9274610. PMID: 35800339; PMCID: PMC9256391.

[6] Aljadeeah S, Wirtz VJ, Nagel Cross-sectional survey to describe medicine use among Syrian asylum seekers and refugees in a German federal state: looking beyond infectious diseases BMJ Open 2021;11:e053044. doi: 10.1136/bmjopen-2021-053044

[7] Marcel Alied, Antimicrobial resistance, COVID-19 and self-medication in Syria: a potential boost for an already escalating problem, JAC-Antimicrobial Resistance, Volume 3, Issue 3, September 2021, dlab100, https://doi.org/10.1093/jacamr/dlab100

[8] Alaettin Z., Volkan A., Data Stream Clustering: A Review, CoRR, 28 Jul 2020, https://doi.org/10.48550/arXiv.2007.10781.

[9] Feng Cao, Martin Ester, Weining Qian, Aoying Zhou, Density-Based Clustering over an Evolving Data Stream with Noise, Conference: Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD, USA, 2006, DOI: 10.1137/1.9781611972764.29.

[10] Amineh Amini, Teh Ying Wah, Density Micro-Clustering Algorithms on Data Streams: A Review, Proceedings of the international MultiConference of Engineer and Computer Scientists 2011, Vol I, IMECS 2011, March 16-18, 2011, Hong Kong.

[11] Kokate, U.; Deshpande, A.; Mahalle, P.; Patil, P. Data Stream Clustering Techniques, Applications, and Models: Comparative Analysis and Discussion. Big Data Cogn. Comput. 2018, 2, 32. https://doi.org/10.3390/bdcc2040032

[12] Data Classification Guide, URL: https://www.spirion.com/data-classification

[13] Akash S., Vikram R., A Comparative Analysis of Association Rule Mining Algorithms, 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012032, DOI 10.1088/1757-899X/1099/1/012032.

[14] Hassani M., Kim Y., Seidl T. (2013). Subspace MOA: Subspace Stream Clustering Evaluation Using the MOA Framework. 7826. 10.1007/978-3-642-37450-0_33.