Nat. Volatiles & Essent. Oils, 2021; 8(5): 1164 - 1173

NVEO
Natural Volatiles &
Essential Oils

Application of Data Mining for Risk mining in Medicine

[1]Rajat Puri, [2]Deepali Giri, [3]Shraavya R.Srinivasarao, [4]Ujwalla Gawande, [5]Dattu Hawale

[1]Dr. Vishwanath Karad MIT World Peace University, Pune, India, Email address: purirajat.rp@gmail.com
[2]Datta Meghe Ayurvedic Medical College Hospital and Research Centre, Wanadongri, Nagpur, India.
[3]Dr. Vishwanath Karad MIT World Peace University, Pune, India.
[4]Associate Professor Dept. of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur.
[5]Tutor, Dept. of Biochemistry Jawaharlal Nehru Medical College, Datta Meghe Institute of Medical Sciences, Sawangi (Meghe), Wardha

Abstract

Introduction: As the demands of society grow larger, the organizations grow larger and becomes complex to provide services. The value of organisational risk management has been recognised in recent years, according to the current scenario. With the introduction of information technology to organisations, a vast volume of data is being processed automatically. However, this data is not utilized efficiently. It is expected that this data can be used effectively to contribute to risk-management for such organizations.

Methodology: Literature study was done using various search engines like google scholar, pubmed. Various publications related to data mining and risk mining in health sector were collected and studied thoroughly.

Discussion: This research proposes risk mining, in which data mining techniques are used to identify and analyse potential risks in organisations, as well as to use risk data for better organisational management. This was applied to the dataset originating from hospital information systems. This is applied to the following two cases: infection control and risk aversion of nurse incidents.

Result: The results of this study show that data mining methods are effective for detection of risk factors in health sector.

Keywords: Risk mining in medicine, data mining, risk aversion, decision trees, risk mining in infection control, risk mining in medical incidents.

1. Introduction:

For those working in the medical profession, decision making is of utmost importance and is constantly required. Given the severity of the situation, it is critical to achieve a proper understanding of the decision-making process, as well as a high degree of consistency in clinical decision-making, in order to reduce risk and error in real-time for the sake of patient safety.[1]

However, to achieve safe medication, medical practise should prevent as many errors as possible. Therefore, preventing close misses and achieve patient protection is a crucial problem in the clinical world. Errors can be classified into 3 types. The first, systematic (or determinate) errors are instrumental, methodological, or mistakes caused due to workflow causing "lopsided" data, which is consistently deviated in one direction from the true value. Second, is Random error that are caused by uncontrollable fluctuations in variables. And finally, Personal errors, that can happen due to an inexperienced member of staff [2]. It is crucial to detect these systematic and personal error which can be avoided with necessary steps.

This study proposes risk mining, in which data containing risk information is analysed using data mining techniques, and the findings are used to prevent risk. Risk mining consists ofthree major processes: risk detection, risk clarification and risk utilization[3].

2. Background:

Health, financial, and organisational data that evolves with the practise must be managed and integrated by hospitals. Previously, this data was organised by hand, which took a long time and did not achieve the desired degree of productivity. Hospital information systems (HIS) are now used by the majority of professionally run hospitals and clinics to help them handle all of their medical and administrative data. [4]. Preoperative treatment, procedure documents, post-operative care, and regular notes of a patient's progress and medications are all recorded in medical records[5]. Incident or accident reports are no exception to this; they are also stored in HIS as clinical data. As a result of this stored data, it's predicted that data mining these documents would reveal new information about medical accidents.

3. Methodology:

This study was conducted in collaboration with DMAMCHRC, Nagpur. Concept designing and literature survey was done by R.P. The implementation of the algorithm was done by S.R.The supervisor, D.G looked over any discrepancies and guided us though the entire research and analysis. The study included Decision making, Data mining, risk mining, data storage in hospitals.

4. Literature review:

Data Mining:

The term "data mining" refers to the process of using analytical techniques from statistics, machine learning, and pattern recognition to understand and interpret data to predict other variables or recognise associations within the data.[6] Data mining is the process of scanning vast amounts of data automatically for trends and patterns that go beyond basic analysis. To segment the data and calculate the probability of future events, data mining employs sophisticated mathematical algorithms. Information Discovery in Data Mining is another name for data mining (KDD)[7]. Data mining can open the door to infinite possibilities for the healthcare industries. The healthcare systems can systematically use the data and analytics to recognise the inefficiencies and the best practices that allow better care and lower costs[8].

Over the years, numerous decision-making models and theories have been introduced and studied[1]. According to Paley and Bjørk and Hamilton, decision making can be classified as having two kinds of validity: one, where decision making is analytical and logical, also called rational systematic-positivist and second, which is more interpretiveand intuitive approach, orphenomenological [9, 10].

Decision theory, in combination with probabilistic principles, has been used to diagnose renal disease since 1970. Several approaches to supporting the diagnosis of lymph node disease were developed in the 1990s.[11]

The 20th century was beginning to be the era of technology. More and more data is being stored and has been employed to improve the quality of life everywhere. However, many diseases have a high level of uncertainty when it comes to diagnosis. Physicians diagnose patients by identifying patterns and creating provisional theories based on the patient's symptoms, medical history, physical examination, and a few tests[12].

Even then, there’s room for error. Nevertheless, these errors should be reduced to the minimum when it comes to medical practice. As a result, data mining has been developed as a method of reducing decision-making errors.

Risk Mining:

To make use of risk data derived from the hospital information systems, this study proposes risk mining, which combines three key processes: risk detection, risk clarification, and risk utilisation.

a. Risk detection:

Patterns or knowledge that domain experts are unaware of can be crucial in detecting the risk of large-scale incidents. So, one of the essential processes in risk mining is mining trends or other forms of knowledge that are unexpected to domain experts. This process is called risk detection, where acquired knowledge is referred to as detected risk information[13].

b. Risk Clarification:

Domain experts and data miners can concentrate on clarification of modelling the hidden risk mechanism by focusing on detected risk information. If domain experts need more information at a finer level of granularity more data can be collected with more comprehensive information and apply data mining to the newly collected data. This process is called risk clarification, where acquired knowledge is referred to as clarified risk information[14].

c. Risk utilization:

To avoid risk incidents, clarified risk data needs to be tested in a real-world setting. If risk knowledge alone is not enough to keep you safe, you will need to do some more research. As a result, more data is collected in preparation for a new risk mining cycle.This process is called risk utilizationwhere acquired knowledge is referred to as clarified risk information[14].

d. Mining unbalanced data:

Large-scale events have a very low chance of happening. Typically, these are deviations of small-scale mishaps, also known as incidents. Since these events are so rare, the chance of a major accident is nearly nil. However, most data mining techniques are based on frequency, and mining such unbalanced data is a difficult problem in data mining. As a result, techniques for mining unbalanced data are critical for detecting risk information in risk mining.[15]

e. Interestingness:

In traditional data mining, mining pattern indices are based on frequency. However, measures of unexpectedness or interest can be used to extract trends from data to extract unexpected or interesting information, and such studies have been published in the data mining literature.

f. Granular Computing:

Human actions are included in the reports and these data are described subjectively with uncertainty. This creates a need to deal with the granularity, orcoarseness and fineness of information. Fuzzy sets and rough sets help us deal with this. dealing with this[16].

g. Visualization

Visualizations of such incidents help domain experts to detect risk information, to clarify or utilize risk information. Visualizations help a better representation of data.

h. Graph Mining:

Only relationships between many items in a broad network system can be used to detect or clarify danger. As a result, extracting partial structure from a network buried in data is a critical technique, with a focus on risk information centred on item relationships. [17].

i. Clustering:

Similarity can reveal connections between seemingly unrelated objects. Or events which are seemingly independent can be grouped into several “similar” events which enable us to find dependencies that normally would be missed. Thus clustering is very essential[18].

j. Evaluating Risk Probability:

Since probability is formally defined as a Lebesgue measure on a fixed sample space, its output is highly volatile when the sample space description is unstable. Instability is especially common when collecting data in a dynamic manner. As a result, careful consideration of risk likelihood is critical.

k. The computer-human interaction:

This is one the most important steps for risk mining for the following reasons. Firstly deep, detailed discussions can lead to better risk mining, as mining details may only show a part of a whole. As domain experts have a deep knowledge, they can compensate for incomplete mining. Second, mining results can lead to a deep understanding of workflow by domain experts. Finally, human-computer interaction adds a new dimension to risk management. Domain experts may look for other possibilities from the rules that seem to be less relevant, in addition to performing risk clarification results. [19]

Risk mining can be applied in various domains. Some of them are discussed bellow.

Case Study 1: Prevention of Medication Errors

Problem statement:

The Risk mining process was applied to the analysis of nurses’ incident data. The study collected data for 6 months and was thoroughly analysed by rule induction method, which helps in detection of several important factors for incidents (risk detection). Domain experts analysed all the data and discovered several significant workflow flaws (risk clarification). Based on these discussions, the nurses changed their workflow and since incidents have been reduced to one-tenth (risk utilization).

Dataset:

The conventional sheet of incident reports was used in this study to collect Nurses' incident data.This was done during the 6 months from April 2001 to September 2001 at the emergency room in Osaka Prefectural General Hospital. The dataset contains factors such as the forms of near-misses, the patients' factors, the medical staff's factors, and the shifts (early-night, late-night, and daytime), with total incidents amounting to 245.

To this dataset, Decision Tree induction and rule induction was applied [20]. The following was the result.

Rule Induction. The following decision tree and rule was found.

(medication error):

If lack of checking and late night shift,

Probability of medication error: (53.3%, 8/15).

(injection error):

If lack of checking and daytime,

Probability of injection incidents: (53.6%, 15/28).

(injection error):

If lack of checking, early-night, and error of injection rate,

Probability of injection incidents(50%, 2/4)

These findings show us that lack of checking and time of shift can be considered as primary risk factors for this.

After a period of risk clarification, it was found that mental concentration of the nurses was an important factor too. Further study found the following induction rules.

(medication error):

If the nurses are disturbed by one or more patients,

Probability of medication error: (90%, 18/20).

(medication error):

If interruption in nurses work,

Probability of medication error:(80%, 4/5).

According to the study, by using rule interpretation, the medication check system and its workflow was discussed and reformed.

The prescription was prepared at the emergency room by the shift's nurses. When the liaison conference between shifts took time, the time spent preparing prior to the start of the shift would generally be less than half an hour. In these situations, medicine sorting cannot be performed ahead of time and must be completed during the shift.

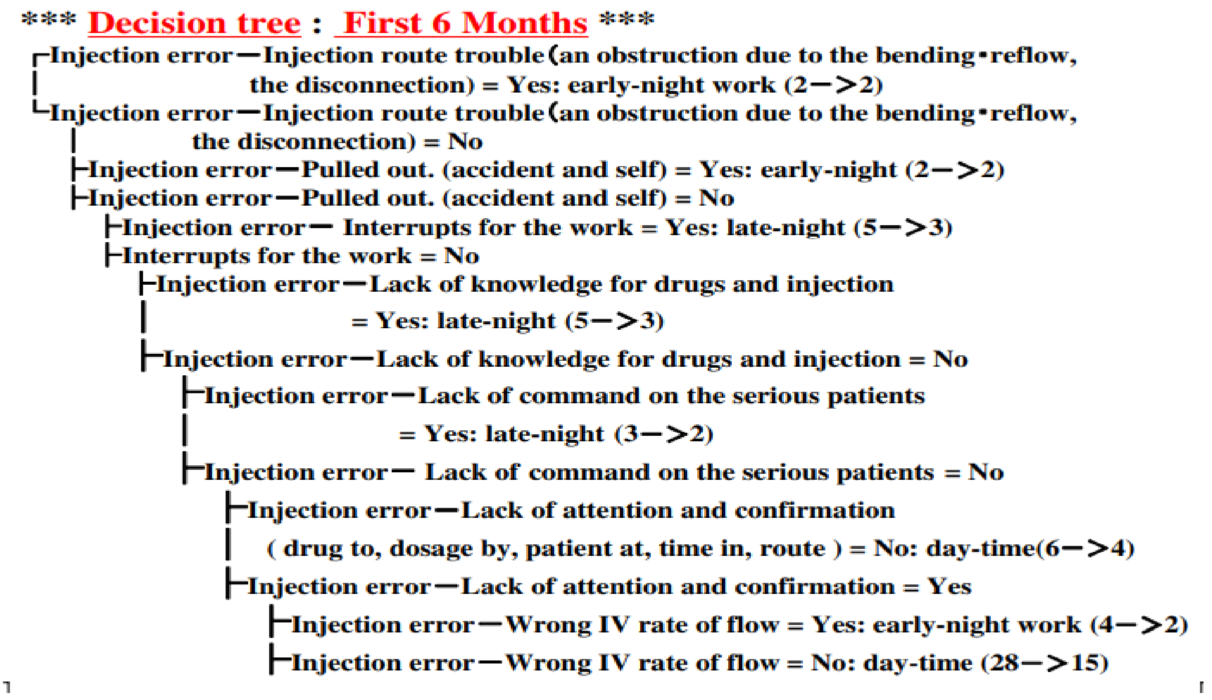
In cases where the restless patients disturbed the nurses' concentration, double checking and preparation of medicine could not be done, leading to medication errors.

Through **Risk utilization**, it was determined that two nurses who had completed their shifts would manage the preparation for the next shift, and that the one in charge of the medication would review the dosage and identification of medicines by herself. This triple check ensured that the incidences were reduced to just 24 in 3 months.

Case study 2: Infection Control

Intrahospital infection is a severe medical complication in which microorganisms are spread by physicians or a patient and nurses admitted to the hospital without infection becomes infected. Methicillin-resistant Staphylococcus aureus (MRSA) is one of the most dangerous bacteria since it is resistant to almost all antibiotics and can kill immunocompromised patients.As a result, extracting risk factors for intrahospital infection from data as proof is critical.

To avoid MRSA infection, the background risk factors for MRSA detection were collected from a clinical database and extracted from a HIS, including microbial examination data and laboratory examination results. The study included 236 patients, including 118 patients with microorganisms other than MRSA and 118 patients with MRSA, who were treated between 1995 and 1998.



```
*** Decision tree :  First 6 Months ***
┌Injection error—Injection route trouble(an obstruction due to the bending•reflow,
|                        the disconnection) = Yes: early-night work (2—>2)
└Injection error—Injection route trouble(an obstruction due to the bending•reflow,
    |            the disconnection) = No
    ├Injection error—Pulled out. (accident and self) = Yes: early-night (2—>2)
    ├Injection error—Pulled out. (accident and self) = No
        ├Injection error— Interrupts for the work = Yes: late-night (5—>3)
        ├Interrupts for the work = No
            ├Injection error—Lack of knowledge for drugs and injection
            |                        = Yes: late-night (5—>3)
            ├Injection error—Lack of knowledge for drugs and injection = No
                ├Injection error—Lack of command on the serious patients
                |                        = Yes: late-night (3—>2)
                ├Injection error— Lack of command on the serious patients = No
                    ├Injection error—Lack of attention and confirmation
                    |  ( drug to, dosage by, patient at, time in, route ) = No: day-time(6—>4)
                    ├Injection error—Lack of attention and confirmation = Yes
                        ├Injection error—Wrong IV rate of flow = Yes: early-night work (4—>2)
                        ├Injection error—Wrong IV rate of flow = No: day-time (28—>15)
```

For analysis, decision tree and rule induction had been used. The global structure of the important features was caught using Decision tree. A set of "If-then" rules were then extracted from the decision trees.

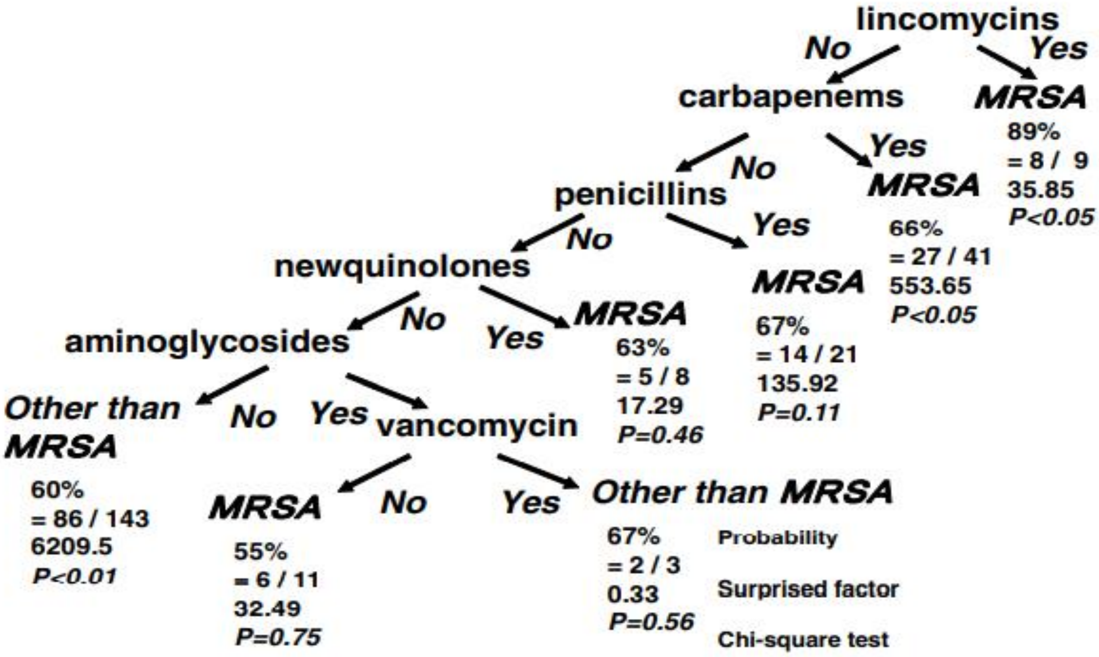
The following rules were obtained.

Rule-1:

If no transfusion catheter and no CVP and a catheter,

Probability of MRSA detected: 0.88 (n=7 / 8).

Rule-2:



If no transfusion catheterand an intra-arterial catheter and no urinary tract catheter,

Probability of MRSA detected: 1.00 (n=3 / 3).

As a relationship between MRSA detection and antimicrobial agent treatment, the following decision tree was developed.

In general, MRSA identification and cephem medication may have a close relationship. However, this decision tree revealed that Cephem medication has a poor relationship with MRSA detection: MRSA detection and Cephem medication may not be related.

The first node in the decision tree is Lincomycin drug. Since Lincomycin medication is the first node's factor, the following rule was discovered:

Rule-3:

If Lincomycin medication = YES, then MRSA detection = YES.

(probability: 89%= 8/9).

Lincomycin is the least effective antibiotic for treating MRSA, according to this rule. Carbapenem, Penicillin, and new Quinolon drugs are all inadequate for MRSA treatment, according to the decision tree.

5. Discussion:

After extensive review of literature, it was found that there was a problem in the workflow. It was concluded that a change in the system was required. After careful deliberations, it was determined that nurses who oversee medication should be held accountable for errors rather than the ones who made arrangements, and that nurses from the previous shift should make necessary arrangements for the following shift. This method also demonstrates how important information granularity is for risk clarity. Items like "lack of checking, lack of attention, etc." in a traditional report form are too wide for risk clarity. Detailed descriptions of environmental factors, on the other hand, are much more critical in eliciting domain experts' discussion and risk utilisation.

For the second case study, using data mining techniques, the study was able to extract context risk factors for MRSA infection. To assess the relationship with MRSA identification, decision trees and if-then rules were extracted. The identification of MRSA and the use of cephem medication are said to have a close relationship. However, it was discovered in this study that cephem medication is not closely related to MRSA. This was further supported by the mining results.

In addition, the results of the mining could indicate that Lincomycin causes MRSA infection. Clindamycin Since Lincosamides: Macrolide-resistant isolates of S. Staphylococcus aureus and Staphylococcus coagulasenegative spp. [methylation of the 23S rRNA encoded by the erm gene, also known as MLSB (macrolide, lincosomide, and type B streptogramin) resistance] may be resistant to clindamycin either permanently or inducibly, or may only be resistant to macrolides (effluxmechanism encoded by the msrA gene).[21]

Moreover, inducible clindamycin resistance was found in 7 and 12 percent of methicillin-resistant Staphylococcus aureus (MRSA), 20 and 19 percent of methicillin-susceptible Staphylococcus aureus (MRSE), and 14 and 35 percent of coagulase-negative staphylococci at two hospitals, respectively. The variability of inducible clindamycin resistance was discovered in those two hospitals because of their investigation. It was concluded that the disk diffusion test (D-test) should be included with susceptibility testing of staphylococci.

As a result, rule-3 might indicate the above inconsistency in inducible MRSA-Lincomycin resistance. The unpredictabilityof inducible MRSA-resistance to Lincomycine in hospital infection control will be critical. As a result, staphylococci susceptibility testing should provide a test for inducible MRSA tolerance to Lincomycin.

Conclusion:

As all the medical information has been recorded electrically as a hospital information system (HIS), mining such combined data is expected to provide new insight into medical injuries. To exploit information about risk mined from information systems, the study proposes risk mining which uses the three important procedures: risk detection, risk clarification and risk utilization.

Risk Detection uncovers trends or knowledge that domain experts are unaware of, which can be interpreted as a warning sign of large-scale incidents. In risk clarification, domain experts and data miners create a model of the secret mechanism of based on detected risk information. If domain experts need more information with finer granularity, more data with comprehensive information should be collected and data mining applied to newly collected data.

To avoid risk incidents, risk utilisation analysed explained risk information in a real-world setting. More study is needed if risk knowledge is insufficient for prevention. As a result, more data is gathered in preparation for a new risk mining period.

The risk mining process was investigated in the following domains as examples.

The analysis begins withanalysing incident data from nurses using the whole procedure. The data was collected over a six-month period and evaluated using rule induction techniques, which identified many key factors that contribute to accidents (risk detection). Since data does not include exact information about these factors, the incident data was recollected for 6 months. Rule induction is then applied to the new data.

Domain experts analysed all the data and discovered a number of significant workflow flaws (risk clarification). Finally, nurses changed their workflow to avoid incidents, and data was collected for a period of six months. Surprisingly, medication errors had dropped to one-tenth of their previous level (risk utilization).

Then the risk detection process was applied to MRSA infection control and discovered a number of useful rules for history MRSA infection risk factors. Cephems are a low-risk antimicrobial agent for MRSA infections, while lincomycins may have the variability of inducible MRSA resistance.

These examples show us how data mining and risk mining can help us avoid incidents and create a safer work environment.

References:

[1] "Chen and Fawcett - 2016 - Using Data Mining Strategies in Clinical Decision .pdf." Accessed: Mar. 03, 2021. [Online]. Available: https://nursing.ceconnection.com/ovidfiles/00024665-201610000-00009.pdf.

[2] "Error Analysis.pdf." .

[3] "ResearchGate Link." Accessed: Mar. 03, 2021. [Online]. Available: https://www.researchgate.net/publication/300580196_Risk_Mining.

[4] "Hospital Information Systems | Electronic Medical Record, EMR Software, Electronic Health Record, EHR Software Customized Recommendations," Jun. 13, 2010. https://web.archive.org/web/20100613022621/http://www.emrconsultant.com/education/hospital-information-systems (accessed Mar. 03, 2021).

[5] J. Thomas, "Medical records and issues in negligence," *Indian J. Urol. IJU J. Urol. Soc. India*, vol. 25, no. 3, pp. 384–388, 2009, doi: 10.4103/0970-1591.56208.

[6] S. Sumathi and S. N. Sivanandam, Eds., "Introduction to Data Mining Principles," in *Introduction to Data Mining and its Applications*, Berlin, Heidelberg: Springer, 2006, pp. 1–20.

[7] "Snapshot." Accessed: Mar. 03, 2021. [Online]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#CHDFGCIJ.

[8] "What is Data Mining in Healthcare?," *Health Catalyst*, May 28, 2014. https://www.healthcatalyst.com/data-mining-in-healthcare/ (accessed Mar. 03, 2021).

[9] J. Paley, H. Cheyne, L. Dalgleish, E. A. S. Duncan, and C. A. Niven, "Nursing's ways of knowing and dual process theories of cognition," *J. Adv. Nurs.*, vol. 60, no. 6, pp. 692–701, Dec. 2007, doi: 10.1111/j.1365-2648.2007.04478.x.

[10] I. T. Bjørk and G. A. Hamilton, "Clinical Decision Making of Nurses Working in Hospital Settings," *Nursing Research and Practice*, Sep. 28, 2011. https://www.hindawi.com/journals/nrp/2011/524918/ (accessed Mar. 03, 2021).

[11] K. B. Wagholikar, V. Sundararajan, and A. W. Deshpande, "Modeling paradigms for medical diagnostic decision support: a survey and future directions," *J. Med. Syst.*, vol. 36, no. 5, pp. 3029–3049, Oct. 2012, doi: 10.1007/s10916-011-9780-4.

[12] P. Yildirim, L. Majnarić, O. Ekmekci, and A. Holzinger, "Knowledge discovery of drug data on the example of adverse reaction prediction," *BMC Bioinformatics*, vol. 15 Suppl 6, p. S7, 2014, doi: 10.1186/1471-2105-15-S6-S7.

[13] "Rahman - 2009 - Data Mining Applications for Empowering Knowledge .pdf." .

[14] H. Kargupta, J. Han, P. S. Yu, R. Motwani, and V. Kumar, *Next Generation of Data Mining*. CRC Press, 2008.

[15] "A comparison of two approaches to data.pdf." .

[16] J. Yang, T. Xu, and F. Zhao, "Modified Uncertainty Measure of Rough Fuzzy Sets from the Perspective of Fuzzy Distance," *Mathematical Problems in Engineering*, Aug. 06, 2018. https://www.hindawi.com/journals/mpe/2018/4160905/ (accessed Mar. 07, 2021).

[17] S. Iwata, Y. Ohsawa, S. Tsumoto, Y. Shi, N. Zhong, and L. Magnani, *Communications and Discoveries from Multidisciplinary Data*. Springer Science & Business Media, 2008.

[18] R. Xu and D. Wunsch, *Clustering*. John Wiley & Sons, 2008.

[19] S. Tsumoto, Y. Tsumoto, K. Matsuoka, and S. Yokoyama, "Risk Mining in Medicine: Application of Data Mining to Medical Risk Management," in *Web Intelligence Meets Brain Informatics*, Berlin, Heidelberg, 2007, pp. 471–493, doi: 10.1007/978-3-540-77028-2_28.

[20] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/BF00993309.

[21] J. B. Patel and Clinical and Laboratory Standards Institute, *Performance standards for antimicrobial susceptibility testing*. 2017.