Universidad Politécnica Salesiana

SMART TECHNOLOGIES, SYSTEMS AND APPLICATIONS

SMARTTECH-IC 2022

Third International Conferenceon Smart Technologies, Systems and Applications

Fabián R. Narváez / Fernando Urgilés / Teodiano Freire Bastos-Filho / Juan Pablo Salgado-Guerrero (Eds.)

Carrera de Biomedicina
Grupo de Investigación en Ingeniería Biomédica - GIIB
Grupo de Investigación en Bioingeniería
y Biomecatrónica - GIByB

This book contains the papers presented on poster sessions at the Third International Conference on Smart Technologies, Systems and Applications (SmartTech-IC 2022) held on November 16-18, 2022 in Cuenca, Ecuador. The SmartTech-IC conference aims to attract researchers, scientists and technologists from some of the top companies, universities, research groups, and government agencies from Latin America and around of the world to communicate their research results, inventions and innovative applications in the area of smart science and the most recent smart technological trends. SmartTech-IC 2022 was organized by the Universidad Politécnica Salesiana, a private institution of higher education with social purposes, nonprofit and co-financed from the Ecuadorian State.

The SmartTech-IC conference has been conceived as an academic platform to promote the creation of technical and scientific collaboration networks. The goal of the conference was addressed some relevant topics related to smart technologies, smart systems, smart trends and applications in different domains in the field of computer science and information systems that represent innovation in current society.

Universidad Politécnica Salesiana Ecuador

Abya Yala UPS

Salesianos Don Bosco Ecuador

SmartTech-IC 2022

ISBN 978-9978-10-844-4
9 789978 108444

Fabián R. Narváez / Fernando Urgilés /
Teodiano Freire Bastos-Filho / Juan Pablo Salgado-Guerrero (Eds.)

Smart Technologies, Systems and Applications

SmartTech-IC 2022: Third International Conference on Smart Technologies, Systems and Applications

ABYA YALA | UPS

SmartTech-IC 2022

2023

SMART TECHNOLOGIES, SYSTEMS AND APPLICATIONS
SmartTech-IC 2022: Third International Conference on Smart Technologies, Systems and Applications

Autores: *Rafael Wilmer Contreras Urgilés, Bryan J. Jiménez, Carlos X. Alvarez, Julio E. Guillén, Araceli Olmos-Vallejo, Lisbeth Rodríguez-Mazahua, José Antonio Palet-Guzmán, Isaac Machorro-Cano, Giner Alor-Hernández, José Luis Sánchez-Cervantes, Fernando Chávez-Castrillón, Massimo Coltorti, Roberta Ivaldi, Guido Sciavicco, Diana C. Burbano G., Jaime A. Álvarez, Jose López, Claudia Isaza, David Luna-Naranjo, Ángela Sucerquia, Camilo Sánchez, Juan Daza, Diego Naula Valladarez, Pablo Aguirre, Gabriel Bermeo, Sebastián Luna, Jennifer Vidal-Sánchez, Fabián Guaraca-Defas y Ana Cecilia Villa-Parra.*

1.ª edición: Universidad Politécnica Salesiana
Av. Turuhuayco 3-69 y Calle Vieja
Cuenca-Ecuador
Casilla: 2074
PBX: (+593 7) 205 0000
Fax: (+593 7) 408 8958
e-mail: publicaciones@ups.edu.ec
www.ups.edu.ec

CARRERA DE BIOMEDICINA
Grupo de Investigación en Ingeniería Biomédica GIIB
Grupo de Investigación en Bioingeniería y Biomecatrónica-GIByB

ISBN digital: 978-9978-10-844-4

Edición, diseño, diagramación e impresión: Editorial Universitaria Abya-Yala
Quito-Ecuador

Tiraje: 300 ejemplares

DOI: https://doi.org/10.17163/abyaups.23

Impreso en Quito-Ecuador, noviembre de 2023

Publicación arbitrada de la Universidad Politécnica Salesiana

Table of Contents

Application of Supervised Descriptive Rule Discovery Methods: Review and Architecture

Araceli Olmos-Vallejo
Tecnológico Nacional de México/I.T. Orizaba
Orizaba, Ver., México
aolmos@ito-depi.edu.mx

Lisbeth Rodríguez-Mazahua
Tecnológico Nacional de México/I.T. Orizaba
Orizaba, Ver., México
lrodriguezm@ito-depi.edu.mx

José Antonio Palet-Guzmán
Hospital Regional de Río Blanco
Río Blanco, Veracruz, México
joseantoniopaletguzman@gmail.com

Isaac Machorro-Cano
Universidad del Papaloapan
Tuxtepec, Oaxaca, México
imachorro@unpa.edu.mx

Giner Alor-Hernández
Tecnológico Nacional de México/I.T. Orizaba
Orizaba, Ver., México
giner.ah@orizaba.tecnm.mx

José Luis Sánchez-Cervantes
CONACYT-Tecnológico Nacional de México/I.T. Orizaba
Orizaba, Ver., México
jose.sc@orizaba.tecnm.mx

Abstract

An autopsy is a recognized procedure for achieving continuous improvement in the quality of medical work despite its worldwide decline, while data mining is the process of finding patterns, anomalies, or correlations between data in a data set. Supervised descriptive rule discovery (SDRD) brings together the two types of tasks that exist in data mining, i.e., both descriptive and predictive areas. The objective is to describe data with respect to a property of interest. In this paper, we provide the results of the analysis of 39 articles related to SDRD to select the adequate techniques and technologies to design the architecture of a module for the comparison of medical opinions related to the decrease ot autopsies in Mexican hospitals. In this way, we seek to address the current problem of the decrease in the number of autopsies in the country, which resulted in the interest of the director of the Pathology area of the Hospital Regional of Río Blanco (H.R.R.B.) in designing and applying a survey to physicians to know the causes of this autopsy decline to take actions that increase this study. The SDRD works were analyzed following a methodology wtih three stages: search, selection and analysis, to conclude that the technique that provides the ability to perform all the features proposed in this work is subgroup discovery (SD), therefore, the proposed module architecture was designed considering SD to compare the results obtained in the H.R.R.B. with those of other hospitals.

Keywords

Supervised Descriptive Rule Discovery, Visualization, Contrast Sets, Subgroup Discovery.

Introduction

An autopsy is a recognized procedure to achieve continuous improvement in the quality of medical work. It is where the pathology specialist who performs it examines a corpse to determine the causes of death, it is an orderly and thorough process. The practice of such a procedure makes an important contribution to the medical knowledge and learning of physicians in training [1].

Due to the decrease in the number of autopsies performed at the Hospital Regional of Río Blanco (H.R.R.B.), the director of the hospital's Pathology area prepared a survey and applied it to the medical staff of his hospital to find out the causes, reasons, and circumstances of this decrease. A system for analyzing these surveys has been developed in earlier work [2]. The system uses descriptive (association) and predictive (classification) data mining tasks. The system revealed that the main reasons for not performing autopsies at the H.R.R.B. are known underlying disease, refusal of family members, disinterest, social factors, fear of demand, cultural issues, religious issues, lack of indication, lack of human resources, and lack of material resources.

Subsequently, a module for comparing medical opinions from various hospitals was added to the system [1]. The module uses Emerging Pattern Mining, a Supervised Descriptive Rule Discovery technique. The main problem of the system is that the generated models are presented in the form of rules, which have to be evaluated by the specialist, when the number of rules obtained increases, the complexity of the evaluation also augments.

The SDRD concept is a combination of the two types of tasks that exist in data mining, i.e. descriptive and predictive tasks. The purpose is to describe data related to a target property. The techniques found in the SDRD are subgroup discovery, emerging patterns, and contrast sets [3].

In this context, the contribution of this paper is two-fold: 1) A review of 39 articles on SDRD techniques was carried out, and 2) A module architecture was proposed to facilitate the comparison of the results of the surveys applied to physicians in Mexican hospitals.

The main benefit expected from the development of this module is to use other SDRD techniques, specifically SD, to compare the reasons why autopsies are not performed in other hospitals, to determine if they are the same as at H.R.R.B. so that the Pathology Department staff can take measures to increase the number of autopsies performed at the hospital. In addition, the selected techniques are intended to allow better data visualization to easily distinguish differences in the medical opinions of the different hospitals.

This article is organized as follows: section 2 presents the analysis methodology applied, section 3 describes the architecture and the solution proposed for this development, and finally, section 4 describes the conclusions obtained based on the analysis and future work.

Methodology of analysis

The study and analysis of advanced SDRD techniques were based on 39 articles that were downloaded from the main digital Computer

Science libraries (IEEE Xplore, ACM digital library, SpringerLink, ScienceDirect, among others). Article selection was based on the technique used: Subgroup Discovery or Contrast Sets. Only articles published in English in journals or conference proceedings during the past five years (2018-2022) were eligible.

Subsequently, the articles were analyzed and compared against certain parameters, for example, visualization, comparison of opinions, focusing on the medical field, to name a few. Most of these opt for the use of visualization, and the use of subgroup discovery, however, very few of them are based on the comparison of opinions or focus on the medical field. Fig. 1 shows the analysis methodology used for the study of the works.

Figure 1
Methodology of analysis

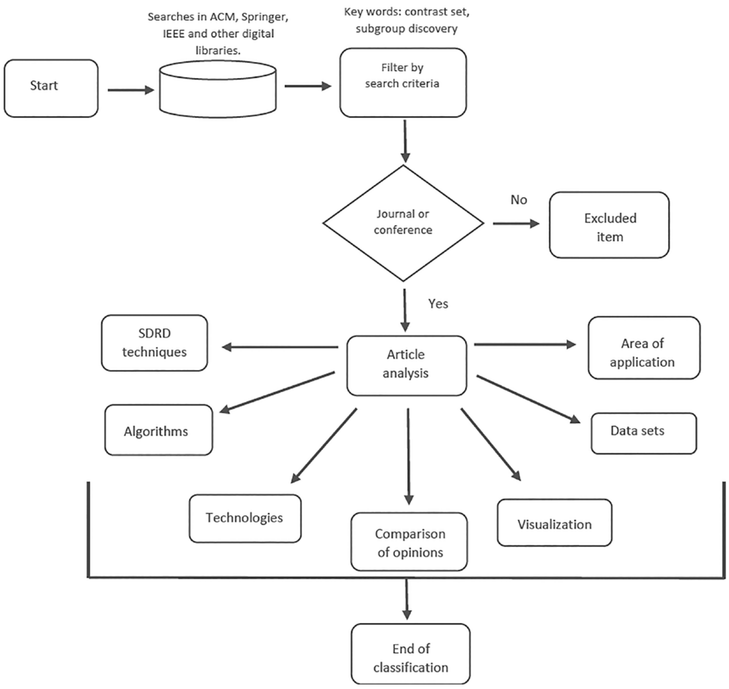


Table 1 shows the comparative table of the articles analyzed for which the following criteria were taken into account: A) Comparison of opinions, B) Use of visualization, C) Contrast sets, D) Subgroup discovery, and E) Focus on the medical area.

Table 1
Comparative analysis of the SDRD works

Work	A	B	C	D	E
Carmona et al. [3]		x	x	x	
Centeio et al. [4]				x	
Lopez-Martinez-Carrasco et al. [5]		x		x	x
Lucas et al. [6]		x		x	
Costa et al. [7]				x	
Wang and Rudin [8]		x		x	
Atzmueller [9]				x	
Al-Taie et al. [10]		x		x	x
Luna et al. [11]				x	
Mattos et al. [12]		x		x	x
Park et al. [13]				x	x
Foppa and Ghiringhelli [14]		x		x	
Hendrickson et al. [15]		x		x	
Helal et al. [16]				x	
de Leeuw et al. [17]	x			x	
Esnault et al. [18]				x	x
Hammal et al. [19]		x		x	
Proença et al. [20]				x	
Luna et al. [21]		x		x	
Kiefer et al. [22]		x		x	
Cervantes et al. [23]		x	x		
Valdivia et al. [24]				x	

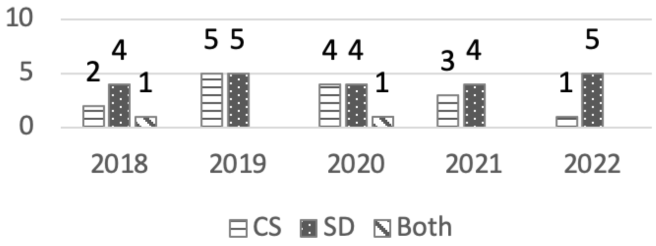
Remil et al. [25]		x		x	
Liu et al. [26]		x	x	x	x
Loyola-Gonzalez et al. [27]		x	x		
Loyola-González et al. [28]		x	x		
Varlamis [29]		x	x		x
Kong et al. [30]		x	x		
Cañete-Sifuentes et al. [31]		x	x		x
Tallon et al. [32]			x		x
Loyola-González et al. [33]		x	x		
Qian et al. [34]		x	x		
Li et al. [35]		x	x		
Al-Taie et al. [36]		x	x		x
Duan et al. [37]	x	x	x		
Alipourchavary et al. [38]			x		
Luna et al. [39]				x	
Loyola-González et al. [40]		x	x		
Jahan et al. [41]	x	x	x		x

After studying the 39 articles, it was concluded that only 28% (11) of them focused on the medical area and none of these eleven works considered the comparison of opinions. In addition, most of the SDRD works (62%) opted for the use of subgroup discovery as a data mining technique. For these reasons, this paper proposes an architecture of a module that uses subgroup discovery to compare medical opinions on the decrease of autopsies performed in Mexican hospitals.

Classification of analyzed works

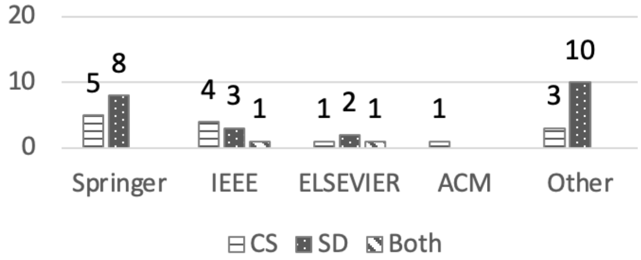
Year of publication. Figure 2 shows the number of articles published annually, separated by the technique: Discovery of Subgroups (SD) and Contrast Sets (CS). Of these, 15 articles are related to CS, 22 focus on SD, and 2 consider both techniques. The graph indicates that there is a higher number of SD publications in more years (2018, 2021, and 2022), while there is the same number of CS works between 2019 to 2020.

Figure 2
Publications by year



Publisher. Fig. 3 depicts that the publishers with the most publications of the subgroup discovery were Springer and others; whereas for contrast sets, Springer and IEEE.

Figure 3
Papers by publisher.



Data Mining Tool. Fig. 4 exhibits the tools used in the DSRD works, where it can be seen that the CS approaches utilized WEKA (e.g., [28], [31], [40]), while most of the SD techniques used VIKAMINE ([9], [15], [16]).

Algorithms. Fig. 5 shows the algorithms used in the analyzed papers, the most frequent algorithms for DSRD applications were: SD-Map* [9], Top-k, SD-Map, Single Shot Detector (SSD), Pattern-based Classifier for Class Imbalance Problem (PBC4cip), and Others, of which the most frequent are PBC4cip and SD-MAP.

Figure 4
Technologies used in the analyzed works.

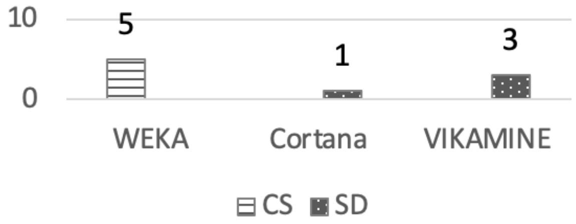
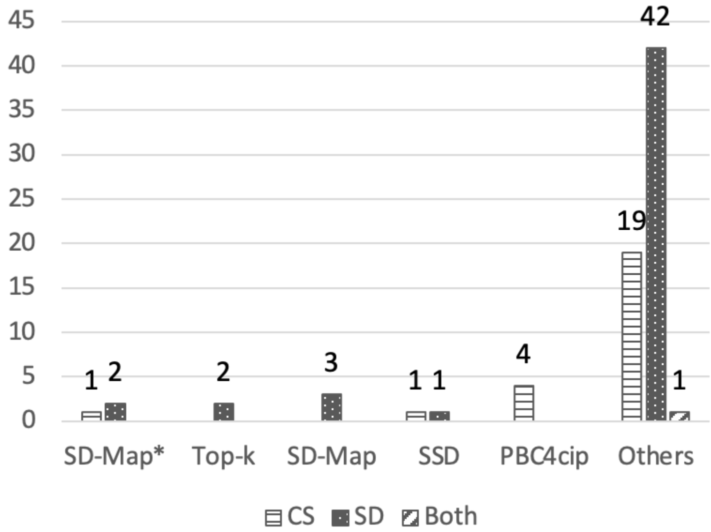
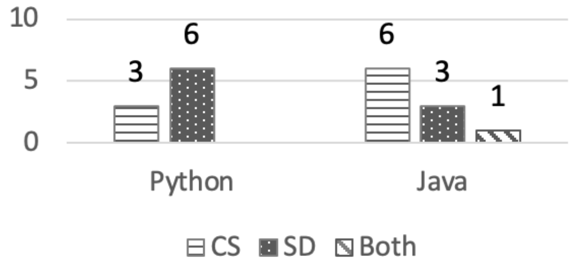


Figure 5
Algorithms used in the DSRD approaches.



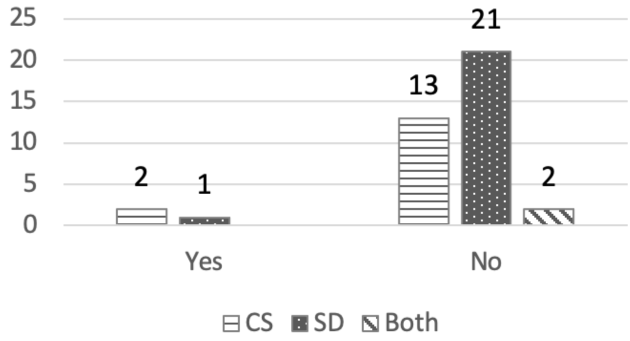
Programming languages. Fig. 6 shows the programming languages most commonly used in the reviewed papers. The most popular language for contrast sets was Java, while for subgroup discovery it was Python.

Figure 6
Programming languages used in the analyzed papers.



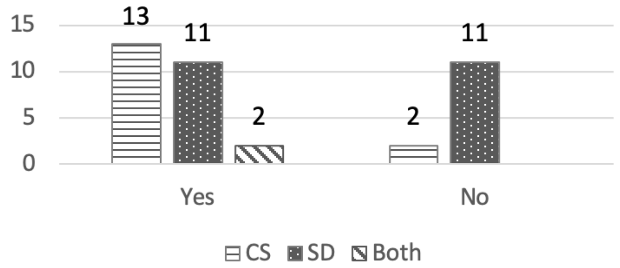
Comparison of opinions. Works that base their research and/or development on opinion comparison are shown in Fig. 7. As can be seen, the use of opinion comparisons is infrequent in the articles analyzed, as only 3 [17, 37, 41] out of 39 articles consider them.

Figure 7
Comparison of opinions in SDRD works.



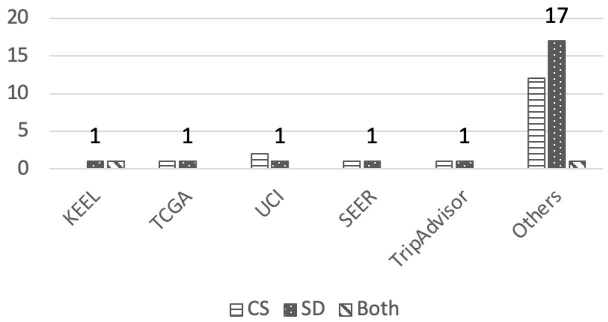
Visualization. As can be seen in Fig. 8, the graph shows that 26 of the 39 papers analyzed make use of some visualization method to improve the interpretation and analysis of the results obtained.

Figure 8
Use of visualization in the SDRD applications.



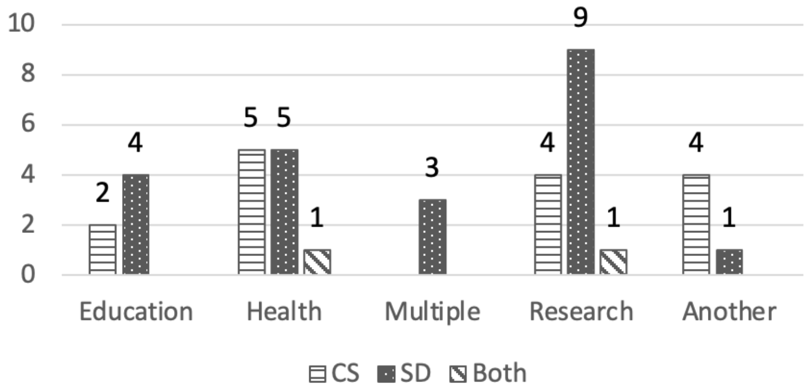
Data sets. Fig. 9 shows the places where the data were obtained and their frequency in the cited works, among which repositories such as Knowledge Extraction Based On Evolutionary Learning (KEEL), The Cancer Genome Atlas (TCGA), UCI Machine Learning Repository, Surveillance Epidemiology, and End Results Program (SEER), TripAdvisor, and Others, were found.

Figure 9
Data collection in SDRD works.



Area of application. Fig. 10 shows the application areas of the 39 analyzed papers, obtaining as a result that the areas of greatest application were: Medical area and Research, e.g., [12, 26].

Figure 10
Research areas of SDRD applications.

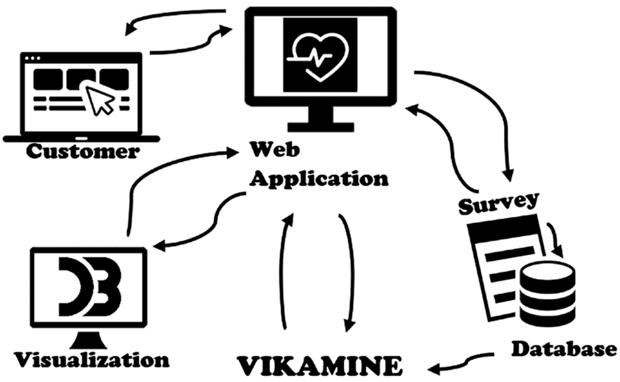


In the following section, the description of the proposed architecture is presented.

Architecture

In this section, we present the architecture of a Subgroup Discovery module for the comparison of medical opinions about the decrease of autopsies in Mexican hospitals. Figure 11 shows the proposed solution scheme for this architecture.

Figure 11
Solution outline.



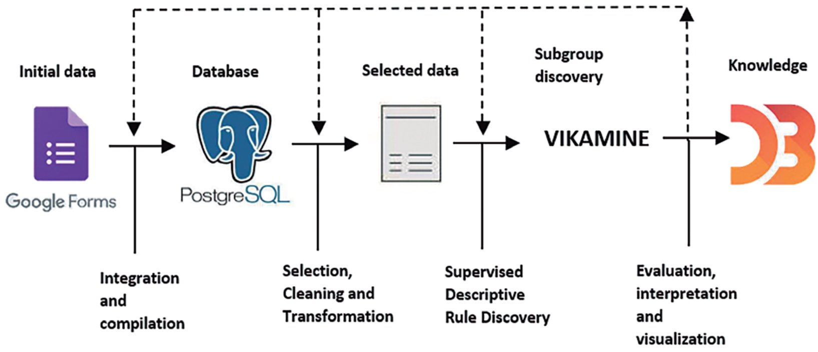
As Figure 11 shows the user is expected to add and visualize the results of the surveys to the database through a Web application, where the algorithms provided by VIKAMINE are expected to generate the models based on such data and present them to the user in a graphical way for a better and quick understanding for its analysis. The new system module for survey analysis is intended to be developed according to the UWE (*UML Web Engineering*) methodology since it adapts to the characteristics of Web systems, has views graphically represented through UML (*Unified Modeling Language*) diagrams, is easy to learn, and allows for quick correction of bugs and problems encountered throughout the Web system development life cycle. In addition, for the programming language, we intend to use Java, because it is a fast language (compared to other alternatives for development like Python), therefore, the IDE (*Integrated Development Environment*) analyzed with the most optimal features for this procedure is NetBeans because, unlike Eclipse (for example), the project structure in NetBeans provides a much clearer view of large applications, from the DBMS (Database Management System) we intend to use PostgreSQL since it has support for Java, high reliability and robustness, extensive documentation, free installation and use, it is multiplatform, to mention some of its most important characteristics.

JavaServer Faces (JSF) is intended to be used as a development framework because it provides a clear separation between behavior and presentation, facilitates the construction and maintenance of Web applications, and provides a clear separation between behavior and presentation. As a data mining tool, VIKAMINE allows the use of multiple data types, features a generic subgroup discovery algorithm that can be configured in many ways to implement various forms of local pattern discovery, is Java-based, and is freely licensed, among other features.

As a data mining methodology, it is intended to use KDD (*Knowledge Discovery in Databases*) as this is an iterative and interactive process (Fig. 12 shows this process in greater detail), and is in rapid development, Finally, we will use D3.js data visualization library because it is a JavaScript library suitable for manipulating documents without the need to be coupled to a proprietary framework, thus facilitating the interpretation of the results obtained by the pathologists.

Fig. 12 shows the KDD process that is intended to be used as a methodology for data mining adapted to this work:

Figure 12
KDD Process.

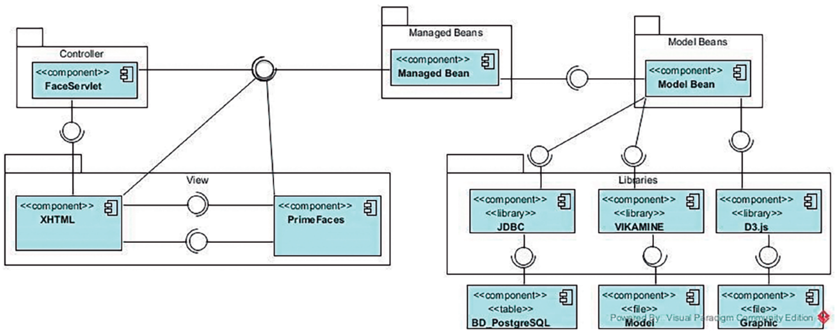


KDD is a methodology proposed in 1996, it is an interactive process consisting of five stages [41], each stage is described below considering the analysis of the medical opinions proposed in this paper:

1. Selection: the algorithms that could be used to compare medical opinions will be analyzed and the results of the surveys will be collected utilizing Google Forms.
2. Preprocessing: the target data set (selected data) will be obtained/created.
3. Transformation: basic operations will be applied, including noise elimination, and empty fields, to mention a few, the main objective of this stage is to reduce dimensionality.
4. Data Mining: Subgroup discovery algorithms will be applied using the VIKAMINE data mining tool and looking for models of interest in the dataset.
5. Implementation: an interpretation and evaluation of the results will be carried out using the D3 visualization library.

The proposed architecture for the new Web application module is shown in Fig. 13. The proposed architecture for the new system module is based on one of the most widely used models for web application development, called MVC (Model-View-Controller). The functionality of each component is briefly explained below:

Figure 13
Proposed architecture.



- **FaceServlet:** This component represents the elements provided by the FaceServlet libraries.
- **Managed Beans:** This component includes the classes necessary for the control and management of the system logic.
- **Model Beans:** This component contains the classes required for the creation of the models within the system.
- **XHTML:** The function of this component is to provide the ability to create/support XML files.
- **PrimeFaces:** This component represents the elements offered by the PrimeFaces library.
- **JDBC:** The function of this component is to provide access to the database management system, employing the necessary drivers,.
- **BD_PostgreSQL:** Consists of the database managment system to be used.
- **VIKAMINE:** This component provides the necessary plugins for the use and creation of models using the VIKAMINE data mining tool.
- **Model:** This component is responsible for generating the models of the data already processed and obtained from the data sets.
- **D3.js:** This component provides the necessary plugins for the use and creation of the models in a graphical way by using the D3 visualization library.
- **Graphic:** This component presents as a final result the graphs to visualize, interpret and evaluate the results obtained.

Conclusions and future work

After analyzing the 39 SDRD works, it was concluded that supervised descriptive rule discovery techniques are used to solve problems in different areas and it was observed that subgroup discovery is used in most of these applications. Nevertheless, it was found that SDRD approaches that focus on the medical area and opinion comparison are scarce, therefore, none of the reviewed methods addresses the five criteria selected for the study of the analyzed papers. Therefore, with the new proposed architecture, we intend to develop a module for the comparison of medical opinions on the decrease of autopsies, using the DS to address the problem that pathologists have in understanding the results obtained, thus providing an easier and faster interpretation of these results. It is expected that this will contribute to the decisión-making of the corresponding areas of the hospitals to increase the number of autopsies.

In the future, the following activities will be performed: Implementation of the survey using the Google Forms platform; design of the new system module; insert the data obtained from the surveys into a database, the database management system will be PostgreSQL; perform the coding of the new system module to compare the data sets using SD and generate the results using the D3.js data visualization library, and finally, validate the newly developed module using a case study.

Acknowledgments. The authors of this work would like to thank the Consejo Nacional de Ciencia y Tecnología (CONACyT) for its financial support and the Tecnológico Nacional de México for allowing the development of this project.

References

[1] Ingrid Aylin Ríos-Méndez, Lisbeth Rodríguez-Mazahua, José Antonio Palet Guzmán, Isaac Machorro-Cano, Silvestre Gustavo Peláez-Camarena, Celia Romero-Torres, "Discovering Emerging Patterns from Medical Opinions about the Decrease of Autopsies Performed in a Mexican Hospital," Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE), pp. 798–803, 2020.

[2] E. Rubio Delgado *et al.*, "Analysis of Medical Opinions about the Nonrealization of Autopsies in a Mexican Hospital Using Association Rules and Bayesian Networks," *Sci. Program.*, vol. 2018, 2018, doi: 10.1155/2018/4304017.

[3] C. J. Carmona, M. J. del Jesus, and F. Herrera, "A unifying analysis for the supervised descriptive rule discovery via the weighted relative accu-

racy," *Knowledge-Based Syst.*, vol. 139, pp. 89–100, 2018, doi: 10.1016/j.knosys.2017.10.015.

[4] C. Centeio Jorge, M. Atzmueller, B. M. Heravi, J. L. Gibson, R. J. F. Rossetti, and C. Rebelo de Sá, "'Want to come play with me?' Outlier subgroup discovery on spatio-temporal interactions," *Expert Syst.*, no. November 2020, pp. 1–20, 2021, doi: 10.1111/exsy.12686.

[5] A. Lopez-Martinez-Carrasco, J. M. Juarez, M. Campos, and B. Canovas-Segura, "A methodology based on Trace-based clustering for patient phenotyping," *Knowledge-Based Syst.*, vol. 232, p. 107469, 2021, doi: 10.1016/j.knosys.2021.107469.

[6] T. Lucas, J. Gomes, R. Vimieiro, R. Prudencio, and T. Ludermir, "A multivariate method for group profiling using subgroup discovery," *Proc. - 2019 Brazilian Conf. Intell. Syst. BRACIS 2019*, pp. 371–376, 2019, doi: 10.1109/BRACIS.2019.00072.

[7] A. J. Costa, M. S. Santos, C. Soares, and P. H. Abreu, "Analysis of Imbalance Strategies Recommendation using a Meta-Learning Approach," 7th ICML Workshop on Automated Machine Learning(2020), 2020.

[8] T. Wang and C. Rudin, "Causal Rule Sets for Identifying Subgroups with Enhanced Treatment Effects," *INFORMS J. Comput.*, pp. 1–40, 2022, doi: 10.1287/ijoc.2021.1143.

[9] M. Atzmueller, "Compositional Subgroup Discovery on Attributed Social Interaction Networks," *Lect. Notes Comput. Sci.*, vol. 11198 LNAI, pp. 259–275, 2018, doi: 10.1007/978-3-030-01771-2_17.

[10] Z. Al-Taie, M. Hannink, J. Mitchem, C. Papageorgiou, and C. R. Shyu, "Drug repositioning and subgroup discovery for precision medicine implementation in triple negative breast cancer," *Cancers (Basel).*, vol. 13, no. 24, pp. 1–23, 2021, doi: 10.3390/cancers13246278.

[11] J. M. Luna, M. Pechenizkiy, W. Duivesteijn, and S. Ventura, "Exceptional in so Many Ways - Discovering Descriptors That Display Exceptional Behavior on Contrasting Scenarios," *IEEE Access*, vol. 8, pp. 200982–200994, 2020, doi: 10.1109/ACCESS.2020.3034885.

[12] J. B. Mattos, E. G. Silva, P. S. G. de Mattos Neto, and R. Vimieiro, "Exceptional Survival Model Mining," in *Lecture Notes in Computer Science* , 2020, vol. 12320 LNAI, pp. 307–321. doi: 10.1007/978-3-030-61380-8_21.

[13] J. V. Park, S. J. Park, and J. S. Yoo, "Finding characteristics of exceptional breast cancer subpopulations using subgroup mining and statistical test," *Expert Syst. Appl.*, vol. 118, pp. 553–562, 2019, doi: 10.1016/j.eswa.2018.10.016.

[14] L. Foppa and L. M. Ghiringhelli, "Identifying Outstanding Transition-Metal-Alloy Heterogeneous Catalysts for the Oxygen Reduction and Evolution Reactions via Subgroup Discovery," *Top. Catal.*, vol. 65, no. 1–4, pp. 196–206, 2022, doi: 10.1007/s11244-021-01502-4.
[15] A. T. Hendrickson, J. Wang, and M. Atzmueller, "Identifying exceptional descriptions of people using topic modeling and subgroup discovery," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11177 LNAI, pp. 454–462, 2018, doi: 10.1007/978-3-030-01851-1_44.
[16] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D. J. Murray, "Identifying key factors of student academic performance by subgroup discovery," *Int. J. Data Sci. Anal.*, vol. 7, no. 3, pp. 227–245, 2019, doi: 10.1007/s41060-018-0141-y.
[17] A. W. de Leeuw, S. van der Zwaard, R. van Baar, and A. Knobbe, "Personalized machine learning approach to injury monitoring in elite volleyball players," *Eur. J. Sport Sci.*, vol. 22, no. 4, pp. 511–520, 2022, doi: 10.1080/17461391.2021.1887369.
[18] C. Esnault, M. L. Gadonna, M. Queyrel, A. Templier, and J. D. Zucker, "Q-Finder: An Algorithm for Credible Subgroup Discovery in Clinical Data Analysis — An Application to the International Diabetes Management Practice Study," *Front. Artif. Intell.*, vol. 3, no. December, pp. 1–25, 2020, doi: 10.3389/frai.2020.559927.
[19] M. A. Hammal, H. Mathian, L. Merchez, M. Plantevit, and C. Robardet, "Rank correlated subgroup discovery," *J. Intell. Inf. Syst.*, vol. 53, no. 2, pp. 305–328, 2019, doi: 10.1007/s10844-019-00555-y.
[20] H. M. Proença, T. Bäck, and M. van Leeuwen, "Robust subgroup discovery," 2021, [Online]. Available: http://arxiv.org/abs/2103.13686
[21] J. M. Luna, H. M. Fardoun, F. Padillo, C. Romero, and S. Ventura, "Subgroup discovery in MOOCs: a big data application for describing different types of learners," *Interact. Learn. Environ.*, pp. 1–19, 2019, doi: 10.1080/10494820.2019.1643742.
[22] A. M. Christoph Kiefer, Florian Lemmerich, Benedikt Langenberg, "Subgroup Discovery in Structural Equation Models," *Science*, vol. 327, no. 5962, pp. 167–70, Jan. 2010, doi: 10.1126/science.1179555.
[23] B. Cervantes, F. Gómez, R. Monroy, O. Loyola-González, M. A. Medina-Pérez, and J. Ramírez-Márquez, "Pattern-based and visual analytics for visitor analysis on websites," Appl. Sci., vol. 9, no. 18, 2019, doi: 10.3390/app9183840.

[24] A. Valdivia *et al.*, “What do people think about this monument? Understanding negative reviews via deep learning, clustering and descriptive rules,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 1, pp. 39–52, 2020, doi: 10.1007/s12652-018-1150-3.

[25] Y. Remil, A. Bendimerad, R. Mathonat, P. Chaleat, and M. Kaytoue, “‘What makes my queries slow?’: Subgroup Discovery for SQL Workload Analysis,” 2021, [Online]. Available: http://arxiv.org/abs/2108.03906

[26] D. Liu, W. Baskett, D. Beversdorf, and C. R. Shyu, “Exploratory Data Mining for Subgroup Cohort Discoveries and Prioritization,” *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 5, pp. 1456–1468, 2020, doi: 10.1109/JBHI.2019.2939149.

[27] O. Loyola-Gonzalez, M. A. Medina-Perez, R. A. C. Valdez, and K. K. R. Choo, “A contrast pattern-based scientometric study of the QS world university ranking,” *IEEE Access*, vol. 8, pp. 206088–206104, 2020, doi: 10.1109/ACCESS.2020.3037665.

[28] O. Loyola-González, M. A. Medina-Pérez, D. Hernández-Tamayo, R. Monroy, J. A. Carrasco-Ochoa, and M. García-Borroto, “A pattern-based approach for detecting pneumatic failures on temporary immersion bioreactors,” *Sensors (Switzerland)*, vol. 19, no. 2, pp. 1–19, 2019, doi: 10.3390/s19020414.

[29] I. Varlamis, “Affinity analysis for studying physicians’ prescription behavior.: The case of hypertension and hyperlipidemia,” *Data Min. Knowl. Discov.*, vol. 35, no. 4, pp. 1739–1759, 2021, doi: 10.1007/s10618-021-00758-4.

[30] J. Kong, J. Han, J. Ding, H. Xia, and X. Han, “Analysis of students’ learning and psychological features by contrast frequent patterns mining on academic performance,” *Neural Comput. Appl.*, vol. 32, no. 1, pp. 205–211, 2020, doi: 10.1007/s00521-018-3802-9.

[31] L. Cañete-Sifuentes, R. Monroy, M. A. Medina-Pérez, O. Loyola-González, and F. Vera Voronisky, “Classification Based on Multivariate Contrast Patterns,” *IEEE Access*, vol. 7, pp. 55744–55762, 2019, doi: 10.1109/ACCESS.2019.2913649.

[32] E. M. Tallon, M. J. Redondo, C. R. Shyu, D. Liu, K. Boles, and M. A. Clements, “Contrast Pattern Mining With the T1D Exchange Clinic Registry Reveals Complex Phenotypic Factors and Comorbidity Patterns Associated With Familial Versus Sporadic Type 1 Diabetes,” *Diabetes care*, vol. 45, no. 3. pp. e56–e59, 2022. doi: 10.2337/dc21-2239.

[33] O. Loyola-Gonzalez, R. Monroy, J. Rodriguez, A. Lopez-Cuevas, and J. I. Mata-Sanchez, "Contrast Pattern-Based Classification for Bot Detection on Twitter," *IEEE Access*, vol. 7, pp. 45800–45817, 2019, doi: 10.1109/ACCESS.2019.2904220.

[34] R. Qian, Y. Yu, W. Park, V. Murali, S. Fink, and S. Chandra, "Debugging crashes using continuous contrast set mining," in *Proceedings - International Conference on Software Engineering*, 2020, pp. 61–70. doi: 10.1145/3377813.3381369.

[35] L. Li, S. Erfani, C. A. Chan, and C. Leckie, "Discovery of contrast corridors from trajectory data in heterogeneous dynamic cellular networks," in *Proceedings of the International Joint Conference on Neural Networks*, 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9207060.

[36] Z. Al-Taie *et al*., "Explainable artificial intelligence in high-throughput drug repositioning for subgroup stratifications with interventionable potential," *J. Biomed. Inform*., vol. 118, p. 103792, 2021, doi: 10.1016/j.jbi.2021.103792.

[37] L. Duan *et al*., "Mining distinguishing customer focus sets from online customer reviews," *Computing*, vol. 100, no. 4, pp. 335–351, 2018, doi: 10.1007/s00607-018-0601-1.

[38] E. Alipourchavary, S. M. Erfani, and C. Leckie, "Mining Rare Recurring Events in Network Traffic using Second Order Contrast Patterns," in *Proceedings of the International Joint Conference on Neural Networks*, 2021, vol. 2021-July, no. ii. doi: 10.1109/IJCNN52387.2021.9533918.

[39] J. M. Luna, C. J. Carmona, A. M. García-Vico, M. J. Del Jesus, and S. Ventura, "Subgroup discovery on multiple instance data," *Int. J. Comput. Intell. Syst*., vol. 12, no. 2, pp. 1602–1612, 2019, doi: 10.2991/ijcis.d.191213.001.

[40] O. Loyola-González, "Understanding the Criminal Behavior in Mexico City." Lect. Notes Comput. Sci., vol. 11835 LNAI, pp. 136–149, Springer, 2019.

[41] B. H. F. y R. A. Munima Jahan, Elham Akhond Zadeh Noughabi, "Finding Influential Factors for Different Types of Cancer: A Data Mining Approach." Lect. Notes in Social Networks, pp. 147-168, Springer, 2018. doi: 10.1007/978-3-319-95810-1_11.