REME • Rev Min Enferm. 2021;25:e-1362
DOI: 10.5935/1415.2762.20210010

REPORT

KNOWLEDGE DISCOVERY IN DATABASES FOR THE SUPPORT TO DECISION-MAKING IN NURSING-RELATED ACTIVITIES

KNOWLEDGE DISCOVERY IN DATABASES PARA APOIO À DECISÃO EM ATIVIDADES VINCULADAS À ENFERMAGEM

KNOWLEDGE DISCOVERY IN DATABASES PARA APOYAR LA DECISIÓN EN ACTIVIDADES RELACIONADAS CON LA ENFERMERÍA

Ana Paula Sartorelli[1]
Carlyne Lopata[1]
Denilsen Carvalho Gomes[1]
Marcia Regina Cubas[1]
Deborah Ribeiro Carvalho[1]

[1]Pontifícia Universidade Católica - PUC, Tecnologia em Saúde. Curitiba, PR - Brazil.

Corresponding Author: Márcia Regina Cubas
E-mail: mcubas@pucpr.br

Authors' Contributions:
Data Collect: Ana P. Sartorelli, Carlyne Lopata; Conceptualization: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes, Marcia R. Cubas, Deborah R. Carvalho; Funding Acquisition: Ana P. Sartorelli, Carlyne Lopata, Marcia R. Cubas, Deborah R. Carvalho; Investigation: Ana P. Sartorelli, Carlyne Lopata; Methodology: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes, Marcia R. Cubas, Deborah R. Carvalho; Project Management: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes, Marcia R. Cubas, Deborah R. Carvalho; Resource Management: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes; Supervision: Marcia R. Cubas, Deborah R. Carvalho; Validation: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes, Marcia R. Cubas, Deborah R. Carvalho; Visualization: Ana P. Sartorelli, Carlyne Lopata, Marcia R. Cubas, Deborah R. Carvalho; Writing - Original Draft Preparation: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes, Deborah R. Carvalho; Writing – Review and Editing: Ana P. Sartorelli, Carlyne Lopata, Denilsen C. Gomes.

Funding: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES. Financing Code 001.

Submitted on: 2020/19/09
Approved on: 2021/12/02

Responsible Editors:
Kênia Lara Silva
Tânia Couto Machado Chianca

ABSTRACT

Objective: to report experiences of data mining use strategies in two Nursing practice settings. Description of the experience: in both experiences, the Apriori algorithm was used to discover association rules and to identify exception situations. The first experience used data from infant deaths in the metropolitan region of Curitiba, PR. In the second experience, medical records of patients assisted by nurses were used in the risk classification at a private hospital in Curitiba -PR. The first study identified 374general rules and the second, 108, both with their respective exception rules. Conclusion: the application of Knowledge Discovery in Databases can be demonstrated and carried out in two different settings, in order to contribute to decision-making by the manager. It is expected that the report reinforces the importance of teaching Nursing Informatics as a decision-making support tool.

Keywords: Data Mining; Databases, Bibliographic; Nursing Informatics; Education, Nursing; Decision Making.

RESUMO

Objetivo: relatar experiências de estratégias do uso da mineração de dados em dois cenários de práticas de Enfermagem. Descrição da experiência: em ambas as experiências foi utilizado o algoritmo Apriori para descoberta de regras de associação e identificado as situações de exceção. A primeira experiência utilizou dados provenientes de óbitos infantis da região metropolitana de Curitiba - PR. Na segunda experiência utilizaram-se prontuários de pacientes atendidos por enfermeiros na classificação de risco em um hospital particular de Curitiba - PR. O primeiro estudo identificou 374 regras gerais e o segundo, 108 regras gerais, ambos com suas respectivas regras de exceção. Conclusão: a aplicação do Knowledge Discovery in Databases pode ser demonstrada e efetivada em dois cenários distintos, a fim de contribuir para a tomada de decisão pelo gestor. Espera-se que o relato reforce a importância do ensino da informática em Enfermagem como ferramenta de apoio à decisão.

Palavras-chave: Mineração de Dados; Bases de Dados Bibliográficas; Informática em Enfermagem; Educação em Enfermagem; Tomada de Decisões.

RESUMEN

Objetivo: reportar experiencias de estrategias de uso de minería de datos en dos escenarios de prácticas de enfermería. Descripción del experimento: en ambos experimentos se utilizó el algoritmo Apriori para descubrir reglas de asociación e identificar situaciones excepcionales. El primer experimento utilizó datos de muertes infantiles en la región metropolitana de Curitiba-PR. En el segundo experimento, se utilizaron los registros médicos de los pacientes atendidos por enfermeros en la clasificación de riesgo en un hospital privado de Curitiba-PR. El primer estudio identificó 374 reglas generales y el segundo 108 reglas generales, ambas con sus respectivas reglas de excepción. Conclusión: la aplicación del Knowledge Discovery in Databases se puede demostrar e implementar en dos escenarios diferentes, con el fin de contribuir a la toma de decisiones por parte del gerente. Se espera que el informe refuerce la importancia de la enseñanza de la informática en enfermería como herramienta de apoyo a la toma de decisiones.

Palabras clave: Minería de Datos; Bases de Datos Bibliográficas; Informática Aplicada a la Enfermería; Educación en Enfermería; Toma de Decisiones.

How to cite this article:
Sartorelli AP, Lopata C, Gomes DC, Cubas MR, Carvalho DR. Knowledge Discovery in Databases for the support to decision-making in Nursing-related activities. REME - Rev Min Enferm. 2021[cited _____ ____ ___];25:e-1362. Available from: _____ DOI: 10.5935/1415.2762.20210010

INTRODUCTION

The storage and use of data from records made by health professionals are challenges that involve technical-organizational processes.[1] Among the health professionals, the Nursing team works in direct patient care and, consequently, produces care and management data.

However, their skills and competences to extract and use the generated data are incipient. As managers, nurses must improve their ability to analyze a large amount of data and appropriate technologies that enable their use to support decision-making.[2]

Among the alternatives for data exploration, the most frequent are based on statistics, but it is possible to adopt artificial intelligence technological resources to discover information capable of complementing the statistical results, especially data mining (KDD, Knowledge Discovery in Databases), which allows for the identification of valid, new, useful and comprehensive standards.[3] Such patterns can be represented, among other ways, through association rules. A number of studies relating data mining in the health area have presented results that can be employed in disease prevention;[4] however, there are no reports of applications directly or indirectly related to Nursing.

In this context, this article aims to report experiences of data mining use strategies in two settings of Nursing practices.

EXPERIENCE DESCRIPTION

In both experiences, the Apriori algorithm was adopted and, on the set of association rules discovered, the exception situations were identified,[5] being referred to as discovery of exception rules (DER). The association rules are represented in the following format: A → C, read as: "IF [Antecedent], THEN [Consequent]". The exception rules are represented by general rules, followed by the respective exception situations: IF [A] and [B], THEN NOT [C], read as: "IF [A] AND [B], THEN [C] DENIED",[6] where [B] represents an item combined with the [Antecedent] of the general rule.

In addition, two percentages are associated with each rule discovered (Sup, Conf). Sup refers to the likelihood for the [Antecedent] occurring. Conf refers to the conditional probability for the [Consequent] occurring, given the occurrence of the [Antecedent].

Experience1 had the original research approved by the Research Ethics Committee (Comitê de Ética em Pesquisa, CEP) of Pontifícia Universidade Católica do Paraná (PUCPR), under Opinion number 2.229.737. Experience 2 was not submitted to the CEP, for dealing with the use of a secondary database; however, use was authorized by the institution's care management, with coding of the patients' and nurses' identities. Before starting pre-processing, data coding was performed by the unit's coordinator. The nurses' names were replaced by codes from 1 to 12, in order to ensure anonymity.

Case 1

This is an example taken from the master's thesis entitled "Data mining for the identification of new relationships between the variables associated with child mortality, in a municipality from the metropolitan region of Curitiba".[7] The data refer to a sample of 358 infant death records from 2010 to 2016, recorded in the Mortality Information System (Sistema de Informação sobre Mortalidade, SIM), investigated and analyzed by the Municipal Maternal and Child Mortality Prevention Committee(Comitê de Prevenção da Mortalidade Materna e Infantil Municipal).

A total of 124 nominal variables were considered, obtained from pre-processing data from the research summary form, which is structured in the SIM, and from the research script for infant and fetal death. The data from the variables were organized in an Excel spreadsheet. Some examples of the variables used are the following: recommendations and prevention measures; reducibility criteria; referral to a reference service; and classification of the newborn. For data mining, the variable "avoidability" of infant death was used as the focus variable (outcome), for being a governability variable in public management. The possible values for the avoidability variable are as follows: avoidable, unavoidable and inconclusive death.

A total of 2,592,045 association rules were identified in data mining. Post-processing with DER resulted in 374 general rules for "avoidability", with their respective exception rules. The selected rules presented statistical significance (chi-square test).

Figure 1 shows an example of a general rule, with its respective exception rules. In general rule1, the variable "referred to the reference system" was associated with "unavoidable death". When this variable is combined with the "recommendations and preventive measures" and "reducibility criteria" variables, the association was changed to "avoidable death".

Figure 1 - General Rule 1, respective exception rules and rule reading structure

Rule	Antecedent	Consequent	Rule Reading
General 1	IF Referred to reference system = Yes	THEN Unavoidable death (29.6%; 52.8%)	29.6% of the pregnant women were referred to the reference system, among which 52.8% infant deaths were unavoidable
Exception 1.1	IF Referred to reference system = Yes AND Recommendations and prevention measures = Quality of the prenatal consultation	THEN Avoidable death (3.4%; 100.0%)	3.4% of the pregnant women were referred to the reference system and had recommendations and prevention measures in relation to the quality of the prenatal consultation, among which 100.0% of the infant deaths were avoidable
Exception1.2	IF Referred to reference system = Yes AND Reducibility criterion = Reducible by adequate care to the woman during pregnancy	THEN Avoidable death (9.5%; 97.1%)	9.5% of the pregnant women were referred to the reference system and met the reducibility criterion of "reducible by adequate care to the woman during pregnancy", among which 97.1% of the infant deaths were avoidable

Source: Research Data.

Case2

Data from a management report were used regarding 10,813 medical records of patients seen by 12nurses responsible for risk classification at a private hospital in the city of Curitiba, Paraná, in March2019. The criterion for risk classification was the Emergency Severity Index[8] protocol, which classifies risk based on the estimation of the number of resources needed for the service, considering higher resource utilization for the most severe patients.

For pre-processing, the data from the management report were categorized. The variables used were the following: nurse responsible for screening (numerical reference 1 to 12); age group of the physician who performed the service (up to 40 years old, 41 to 60 years old or 61 to 80 years old); outcome of the service provided (discharge, hospitalization or death); medical specialty(General Clinic, Orthopedics or Cardiology); agreement between screening and the outcome of medical care(yes or no); and day of the week of the service(Monday to Sunday).

For the discovery of association rules, the agreement between screening and the outcome of the medical care was adopted as a focus variable (outcome).

A total of 8,099 association rules were identified, which resulted in 108general rules, with their respective exception rules.

Figure 2 shows an example of a general rule, with its respective exception rule. As a general rule, the antecedent of the compound rule for the conditions "nurse=3" and "age of the physician=up to 40 years old" was associated with the consequent "agreement between the risk classification and the medical care outcome=yes". However, when the antecedent of the general rule is complemented with the "medical specialty" condition, the consequent changes to "agreement between the risk classification and the medical care outcome=no".

Figure 2 - General Rule 2, respective exception rule and rule reading structure

Rule	Antecedent	Consequent	Rule Reading
General 2	IF Nurse = 3 AND Age of the physician = Up to 40 years old	THEN Medical discharge and agreement between the risk classification and the outcome of medical care = yes (13.3%; 63.7%)	13.3% of the patients were evaluated by nurse3 and treated by a physician aged up to 40 years old; in 63.7% of these cases there was agreement between the risk classification and the medical care outcome
Exception rule 2.1	IF Medical specialty = Orthopedics AND Nurse = 3 AND Age of the physician = Up to 40 years old	THEN Medical discharge and agreement between the risk classification and the outcome of medical care=no(3.9%; 52.9%)	3.9% of the Orthopedics patients were evaluated by nurse 3 and treated by a physician aged up to 40 years old; in 52.9% of these cases there was no agreement between the risk classification and the medical care outcome

Source: Research Data.

DISCUSSION

In experience 1, the use of KDD allowed understanding variables associated with infant mortality consistent with the local reality of a medium-sized municipality. Although infant mortality has causes and risk factors evidenced in the literature, the discovery of exception rules using local basis allows for municipal planning with assertive decision-making, based on targeted interventions and focusing on the integrality of the actions.

In the example highlighted, the rules were related to both reference and prenatal services. By highlighting situations in which infant deaths could have been avoided, the exception rules indicate the need to improve the monitoring of the actions performed in the reference services, which can serve as a warning to the managers of the municipal services for directing resources in specialized team training. It can also contribute to the reflection by managers and health professionals to provide improvements in the care practices.[9]

On the other hand, the rules indicated that the prevention measures related to the quality of the prenatal consultation and to adequate care to the woman during pregnancy can also reduce the number of infant deaths. This reinforces the findings of the study published in 2018, in which 51.3% of the infant deaths were reduced by means of adequate care for the women during pregnancy.[10]

In this example, the application of the KDD process may reinforce the importance of the Nursing consultation in usual-risk prenatal care and, at the same time, ensure the principle of integrality of the actions. Even strengthening policies of universality, integrality and equality, the priority actions to reduce infant mortality are still insufficient.[9]

It is noteworthy that the professional categories involved in the analysis of issues referring to the reduction of infant mortality are not restricted to the nurses. However, the team coordinated by them corresponds to an expressive number of workers in the primary health care services, health promotion and surveillance departments, and reference services for women, pregnant women and newborns. Thus, when appropriating tools such as KDD, managing care or service nurses enhance quality of care by intervening in variables under their technical-scientific governability.

In experience 2, the use of KDD showed potential for knowledge discovery from data related to the risk classification performed by the nurse in the emergency service. The agreement between the risk classification and the medical care outcome can suggest an alignment of institutional conducts among the professionals, as well as better targeting of scarce resources in critical specialties and adoption of the protocols recommended by the institution. The age of the physician, also present in the exception rules, can be related to the longer experience time.

A study that evaluated the application of a risk classification protocol among nurses showed full agreement on the extreme levels of need for care – emergency and non-urgency; however, doubts still remain about the effectiveness of the professional action.[11] In the example reported, the results of the KDD process reinforce positive situations regarding the emergency care initiated by the nurse, which may justify the presence of such professional.

In turn, the divergences in risk classification obtained in the example raise the discussion that the attribution of a risk level to the patient comprises a complex decision-making process, in which it is necessary to identify the data to be collected and the instruments to be used, precluding decision-making being conducted only by the experience and subjective evaluation of each nurse.[12]

Based on the two examples reported, it can be inferred that the contribution of adopting data mining to support the decision-making process consists in complementing the results obtained from statistical tests, which, in the case of the two experiments presented, could be the chi-square test, which allows for the identification of the relationships between nominal variables, which we represent in the results as "general rule".

Data mining becomes a differential from the moment it allows identifying exceptions that, for local managers, are more relevant than the association or the establishment of a relationship from statistical tests already established in the literature. Due to the use of local databases, the result favors the understanding of the relationships, especially when discussing the reading of the rule and identifying the critical nodes of the problems presented.

It is noteworthy that statistics and data mining are complementary. The examples presented reveal that the KDD process explained relationships that would be redundant without the use of exception rules.

The reports submitted may come to extend to other situations experienced in the nurses' routine and contribute to the teaching of Nursing Informatics in the undergraduate and graduate courses as a decision-making support tool.

CONCLUSION

The use of KDD in the setting related to infant mortality helped to understand the associated variables that are consistent with the local reality, and may be the basis for planning and for assertive decision-making, with integrality of health care. Its application reinforces the identification of known causes of infant mortality, but explains in detail the relationship with local variables.

In the case of the risk classification by Nursing care, KDD may suggest an alignment of institutional conducts between physicians and nurses, as well as better targeting of resources in critical areas.

The application of KDD can be carried out in different scenarios, contributing to decision-making by the manager. And in the case of Nursing, it is expected that the report reinforces the importance of teaching Nursing Informatics in the undergraduate and graduate courses as a decision-making support tool.

The limitation of this report lies in the impossibility of generalizing the examples, due to the small sample of records used in case1 and to the use of local databases.

REFERENCES

1. Geilsa SCV, Augusta MD. A gestão da informação em atenção básica de saúde e a qualidade dos registros de Enfermagem. Investig Enferm. 2018[cited 2020 Abr 09];20(2):01-8. Available from: https://pesquisa.bvsalud.org/portal/resource/pt/biblio-995111

2. Topaz M, Pruinelli L. Big Data and Nursing: implications for the future. Stud Health Technol Inform 2017[cited 2020 June 18];232:165-71. Available from: https://www.ncbi.nlm.nih.gov/pubmed/ 28106594

3. Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. Advances in knowledge discovery and data mining. American Association for Artificial Intelligence. Menlo Park: MIT Press; 1996.

4. Li Q, Zhang Y, Kang H, Xin Y, Shi C. Mining association rules between stroke risk factors based on the Apriori algorithm. Technol Health Care. 2017[cited 2020 Apr 09];25(S1):197-205. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28582907

5. Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB Conference. Santiago, Chile; 1994. p.487-99.

6. Hussain F, Liu H, Suzuki E, Lu H. Exception Rule Mining with a Relative Interestingness Measure. Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining PAKDD, Kyoto, Japão; 2000 Apr. 18-20.

7. Sartorelli AP. Mineração de dados para identificação de novas relações entre variáveis associadas à mortalidade infantil, em um município da região metropolitana de Curitiba [dissertação]. Curitiba: Pontifícia Universidade Católica do Paraná; 2018.

8. Emergency Severity Index (ESI). A triage tool for emergency. Department care. Version 4, Implementation Handbook. Rockville: AHRQ Publication; 2012[cited 2017 Apr 16]. Available from: https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/esi/esihandbk.pdf

9. Pinheiro JMF, Tinoco LS, Rocha ASS, Rodrigues MP, Lyra CO, Ferreira MAF. Atenção à criança no período neonatal: avaliação do pacto de redução da mortalidade neonatal no Rio Grande do Norte, Brasil. Ciênc Saúde Colet. 2016[cited 2020 Dec 27];21(1):243-52. Available from: https://doi.org/10.1590/1413-81232015211.09912014

10. Oliveira CM, Bonfim CV, Guimarães MJB, Frias PG, Medeiros ZM. Mortalidade infantil: tendência temporal e contribuição da vigilância do óbito. Acta Paul Enferm. 2016[cited 2020 Dec 27];29(3):282-90. Available from: https://doi.org/10.1590/1982-0194201600040

11. Malfussi LBH, Bertoncello KCG, Nascimento ERP, Silva SG, Hermida PMV, Jung W. Concordância de um protocolo institucional de avaliação com classificação de risco. Texto Contexto Enferm. 2018[cited 2020 Apr 30];27(1):e4200016. Available from: https://www.scielo.br/ pdf/tce/v27n1/0104-0707-tce-27-01-e4200016.pdf

12. Souza CC. A tomada de decisão do enfermeiro na classificação de risco: qual o embasamento teórico para a prática neste cenário de cuidado? REME - Rev Min Enferm. 2015[cited 2020 Apr 08];19(2):09-11. Available from: http://www.dx.doi.org/10.5935/1415-2762.20150021

This is an open-access article distributed under the terms of the Creative Commons Attribution License