Discretization for Naive Bayes Taking the Specifics of Heart Data into Account

By: Bohacik, J (Bohacik, Jan)[1]; Zabovsky, M (Zabovsky, Michal)[2]

JOURNAL OF INFORMATION AND ORGANIZATIONAL SCIENCES
Volume: 43 **Issue:** 1 **Pages:** 1-14
DOI: 10.31341/jios.43.1.1
Published: 2019
Document Type: Article

Abstract

At the present time heart disease is a major cause of death. Factors such as physical inactiveness, obesity, diabetes, social isolation and aging are expected to make the situation worse. It is worsened even further with misdiagnosis of patients describing heart related issues. A probability decision support approach to diagnosis of heart disease based on Naive Bayes is discussed here as most hospitals collect patient records but these are rarely used for automatic decision support. The approach is analyzed on Statlog heart data with the focus on improving preprocessing methods. As the result, a discretization algorithm with Equal Frequency Discretization which considers the specifics of engaged heart disease patients is presented. Enhancements of achieved accuracy with the added discretization and in comparison with other machine learning algorithms are shown in experiments founded on 10-fold cross-validation.

Keywords

Author Keywords: discretization; Naive Bayes; diagnosis; heart disease
KeyWords Plus: CARDIOVASCULAR-DISEASE; BURDEN; RISK

Author Information

Reprint Address: Bohacik, J (reprint author)

Univ Zilina, Fac Management Sci & Informat, Zilina, Slovakia.
Organization-Enhanced Name(s)
University of Zilina

Addresses:

[1] Univ Zilina, Fac Management Sci & Informat, Zilina, Slovakia
Organization-Enhanced Name(s)
University of Zilina

[2] Univ Zilina, Univ Sci Pk, Zilina, Slovakia
Organization-Enhanced Name(s)
University of Zilina

E-mail Addresses: jan.bohacik@fri.uniza.sk; michal.zabovsky@fri.uniza.sk

Publisher

UNIV ZAGREB, FAC ORGANIZATION & INFORMATICS, PAVLINSKA 2, VARAZDIN, 42000, CROATIA

Categories / Classification

Research Areas: Computer Science
Web of Science Categories: Computer Science, Interdisciplinary Applications

Document Information

Language: English
Accession Number: WOS:000472601100002
ISSN: 1846-3312
eISSN: 1846-9418

Other Information

IDS Number: IE8CI

Cited References in Web of Science Core Collection: 23

Times Cited in Web of Science Core Collection: 0

See fewer data fields

1 of 50

Cited References: 23

Showing 23 of 23 View All in Cited References page *(from Web of Science Core Collection)*

1. Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey — Times Cited: 4
   By: Banu, Salma N. K.; Swamy, Suma
   2016 INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS, COMMUNICATION, COMPUTER AND OPTIMIZATION TECHNIQUES (ICEECCOT) Pages: 256-261 Published: 2016

2. The Burden of Cardiovascular Disease in Low- and Middle-Income Countries: Epidemiology and Management — Times Cited: 37
   By: Bowry, Ashna D. K.; Lewey, Jennifer; Dugani, Sagar B.; et al.
   CANADIAN JOURNAL OF CARDIOLOGY Volume: 31 Issue: 9 Pages: 1151-1159 Published: SEP 2015

3. Cardiovascular disease and hypertension in sub-Saharan Africa: burden, risk and interventions — Times Cited: 34
   By: Cappuccio, Francesco Paolo; Miller, Michelle Avril
   INTERNAL AND EMERGENCY MEDICINE Volume: 11 Issue: 3 Pages: 299-305 Published: APR 2016

4. Supervised and unsupervised discretization of continuous features — Times Cited: 232
   By: Dougherty, J.; Kohavi, R.; Sahami, M.
   Machine Learning. Proceedings of the Twelfth International Conference on Machine Learning Pages: 194-202 Published: 1995

5. Comparative Study of Heart Disease Classification — Times Cited: 2
   By: Ekiz, Simge; Erdogmus, Pakize
   2017 ELECTRIC ELECTRONICS, COMPUTER SCIENCE, BIOMEDICAL ENGINEERINGS' MEETING (EBBT) Published: 2017

6. From data mining to knowledge discovery in databases — Times Cited: 1,296
   By: Fayyad, U; PiatetskyShapiro, G; Smyth, P
   AI MAGAZINE Volume: 17 Issue: 3 Pages: 37-54 Published: FAL 1996

7. MULTI-INTERVAL DISCRETIZATION OF CONTINUOUS-VALUED ATTRIBUTES FOR CLASSIFICATION LEARNING — Times Cited: 444
   By: FAYYAD, UM; IRANI, KB
   IJCAI-93, VOLS 1 AND 2: PROCEEDINGS OF THE THIRTEENTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE Pages: 1022-1027 Published: 1993

8. Title: [not available] — Times Cited: 3
   By: McLachlan, G. J.; Do, K.-A.; Ambroise, C.
   Analyzing Microarray Gene Expression Data Published: 2004
   Available
   Publisher: Willey, San Diego, USA
   URL: http://onlinelibrary.wiley.com/book/10.1002/047172842X

9. Title: [not available] — Times Cited: 1
   Group Author(s): National Center for Chronic Disease Prevention and Health Promotion
   Know the Facts About Heart Disease Published: 2013
   Available
   Publisher: National Center for Chronic Disease Prevention and Health Promotion

10. Title: [not available] — Times Cited: 2
    Group Author(s): National Heart Foundation of Australia

Heart Information: Coronary Heart Disease Published: 2013
Available

Publisher: National Heart Foundation of Australia, Australia

11. **Cholesterol levels: What you need to know** Times Cited: 3

Group Author(s): National Institutes of Health

NIH MedlinePlus Volume: 7 Issue: 2 Pages: 6-7 Published: 2012
Available

12. **Neural Network Diagnosis of Heart Disease** Times Cited: 6

By: Olaniyi, Ebenezer O.; Oyedotun, Oyebade K.; Helwan, Abdulkader; et al.

2015 INTERNATIONAL CONFERENCE ON ADVANCES IN BIOMEDICAL ENGINEERING (ICABME) Pages: 21-24 Published: 2015

13. **Intelligent heart disease prediction system using data mining techniques** Times Cited: 78

By: Palaniappan, Sellappan; Awang, Raflah

2008 IEEE/ACS INTERNATIONAL CONFERENCE ON COMPUTER SYSTEMS AND APPLICATIONS, VOLS 1-3 Book Series: International Conference on Computer Systems and Applications Pages: 108-115 Published: 2008

14. **Screening for asymptomatic coronary artery disease: A systematic review for the U. S. Preventive Services Task Force** Times Cited: 3

By: Pignone, M.; Fowler-Brown, A.; Pletcher, M.; et al.

Systematic Evidence Review Issue: 22 Published: 2003
Available

URL: https://www.ahrq.gov/downloads/pub/prevent/pdfser/chdser.pdf

[Show additional data]

15. **Analysis of Data Mining Techniques for Heart Disease Prediction** Times Cited: 5

By: Sultana, Marjia; Haider, Afrin; Uddin, Mohammad Shorif

2016 3RD INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING AND INFORMATION & COMMUNICATION TECHNOLOGY (ICEEICT) Book Series: International Conference on Electrical Engineering and Information Communication Technology Published: 2016

16. **Improving Naive Bayes classifier using conditional probabilities** Times Cited: 5

By: Taheri, S.; Mammadov, M.; Bagirov, A. M.

P 9 AUSTR DAT MIN C Pages: 63-68 Published: 2011
Available:

URL: https://dl.acm.org/citation.cfm?id=2483637

17. Title: [not available] Times Cited: 1

Group Author(s): Victorian Government

Fainting and Collapse Published: 2011
Available:

Publisher: Victorian Government, Melbourne, Australia

18. Title: [not available] Times Cited: 213

Group Author(s): WHO

Global Atlas on cardiovascular disease prevention and control Published: 2011
Accessed July 12, 2018

Publisher: World Health Organization, Geneva

19. Title: [not available] Times Cited: 18

By: Witten, I. H.; Frank, E.; Hall, M. A.

Practical Machine Learning Tools and Techniques Published: 2011
Available:

Publisher: Morgan Kaufman Publishers, Burlington, MA, USA

URL: https://www.sciencedirect.com/science/book/9780123748560

20. Title: [not available] Times Cited: 1,423

Group Author(s): World Health Organization

Global status report on noncommunicable diseases 2010 Published: 2011
(accessed August 15, 2014), 2011

Publisher: World Health Organization

21. **The next generation of clinical decision support: linking evidence to best practice.** (View record in MEDLINE) **Times Cited: 24**
By: Wu, Robert; Peters, William; Morgan, Matthew W
Journal of healthcare information management : JHIM Volume: 16 Issue: 4 Pages: 50-5 Published: 2002

22. **Race and Sex Differences in the Incidence and Prognostic Significance of Silent Myocardial Infarction in the Atherosclerosis Risk in Communities (ARIC) Study** **Times Cited: 34**
By: Zhang, Zhu-Ming; Rautaharju, Pentti M.; Prineas, Ronald J.; et al.
CIRCULATION Volume: 133 Issue: 22 Pages: 2141-2148 Published: MAY 31 2016

23. **A comparative study for predicting heart diseases using data mining classification methods** **Times Cited: 2**
By: Zriqat, I. A.; Altamimi, A. M.; Azzeh, M.
International Journal of Computer Science and Information Security Volume: 14 Issue: 12 Pages: 868-879 Published: 2016
Available:
URL: https://arxiv.org/ftp/arxiv/papers/1704/1704.02799.pdf

Showing 23 of 23 **View All in Cited References page**