Analysis of Heart Disease Data

Project Report

presented by
Club der toten Dichten (Team 12)
Finn Hülsbuch, 1913864
Thilo Dieing, 1692328
Lasse Lemke, 1914420
Eric Jacquomé, 1903834
Timotheus Gumpp, 1913876

submitted to the
Data and Web Science Group
Prof. Dr. Heiko Paulheim
University of Mannheim

December 2022

Contents

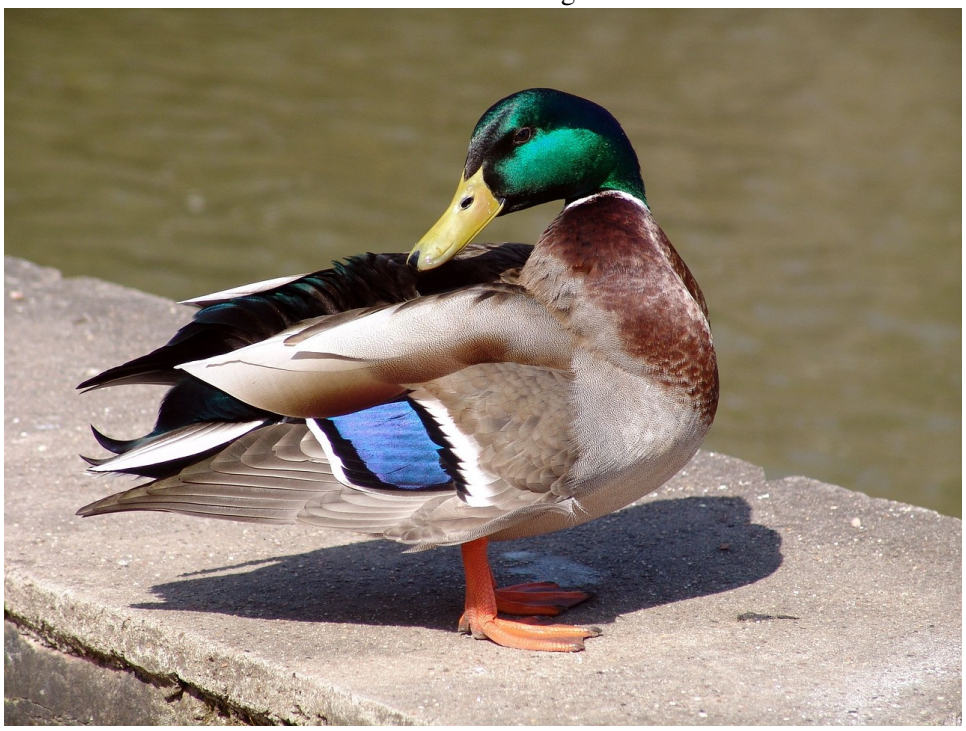
1 Application Area and Goals

Heart disease is currently still one of the highest causes of mortality on earth (Nahar et al., n.d.; Kavitha and Kannan, n.d.; Statistisches Bundesamt, n.d.). Given the successful application of data mining in other sectors e.g. banking and finance or marketing (Keleş, n.d.) possible applications in the medical industry are plentiful. Yet the healthcare sector is information rich but knowledge poor (Soni et al., n.d.). According to Soni et al. (n.d.) medical data sets provide great potential for data mining to be used in clinical diagnosis.

This aim of this project was the application of data mining methods, more specifically classification methods, to predict whether or not a patient could suffer from a heart disease. The successful application could help doctors and medical staff with diagnosing patients by automatically analysing historical test result data of the patient and give a prediction when a higher potential of heart problems arise. By doing this analysis patients flagged for potential heart disease could possibly be prioritised. Due to the immense amount of stress and long working hours medical personal are facing, having an additional instance looking at the data could be beneficial. In the past such approaches have already been tested and proven to be a good diagnostic option (Usha Rani, n.d.). Jabbar, Deekshatulu, and Chandra (n.d.) state that data mining techniques answer several important and critical questions related to healthcare and that they can improve the provision of quality services to patients.

This project report is based on the "Heart Disease Data Set" (Janosi et al., n.d.) which, despite its age is still relevant given the fact that it consists of results of medical tests. In addition to that the validity is assumed because it is frequently used in contemporary research (see Nahar et al., n.d.; Usha Rani, n.d.; Aha and Kibler, n.d.).

Hallo Lasse. So kannst du Bilder in Latex anzeigen.



Da du aber wahrscheinlich ja eher Abbildungen machen willst versuche es hiermit.

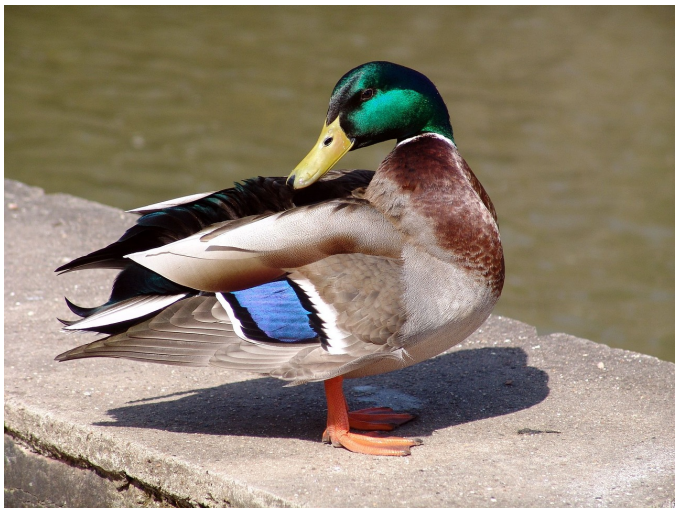


Figure 1: This is a smaller duck

References

Nahar, Jesmin et al. (n.d.). "Association Rule Mining to Detect Factors Which Contribute to Heart Disease in Males and Females". In: *Expert Systems with Applications* 40.4 (), pp. 1086–1093. ISSN: 09574174. DOI: 10.1016/j.eswa.2012.08.028. URL: https://linkinghub.elsevier.com/retrieve/pii/S095741741200989X.

Kavitha, R. and E. Kannan (n.d.). "An Efficient Framework for Heart Disease Classification Using Feature Extraction and Feature Selection Technique in Data Mining". In: *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*. 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS). Pudukkottai, India: IEEE, pp. 1–5. ISBN: 978-1-4673-6725-7. DOI: 10.1109/ICETETS.2016.7603000. URL: http://ieeexplore.ieee.org/document/7603000/.

Statistisches Bundesamt (n.d.). *Gestorbene: Deutschland, Jahre, Todesursachen, Geschlecht*. URL: https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=0&step=0&titel=Tabellenaufbau&levelid=1665392693797&acceptscookies=false#abreadcrumb.

Keleş, Mümine KAYA (n.d.). "An Overview: The Impact of Data Mining Applications on Various Sectors". In: (), p. 5.

Soni, Jyoti et al. (n.d.). "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction". In: *International Journal of Computer Applications* 17.8 (), pp. 43–48. ISSN: 09758887. DOI: 10.5120/2237-2860. URL: http://www.ijcaonline.org/volume17/number8/pxc3872860.pdf.

Usha Rani (n.d.). "Analysis Of Heart Diseases Dataset Using Neural Network Approach". In: *International Journal of Data Mining & Knowledge Management Process* 1.5 (), pp. 1–8. ISSN: 2231007X. DOI: 10.5121/ijdkp.2011.1501. URL: http://www.aircconline.com/ijdkp/V1N5/0911ijdkp01.pdf.

Jabbar, M. Akhil, B.L. Deekshatulu, and Priti Chandra (n.d.). "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm". In: *Procedia Technology* 10 (), pp. 85–94. ISSN: 22120173. DOI: 10.1016/j.protcy.2013.12.340. URL: https://linkinghub.elsevier.com/retrieve/pii/S2212017313004945.

Janosi, Andras et al. (n.d.). *UCI Machine Learning Repository: Heart Disease Data Set*. URL: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

Aha, D and Dennis Kibler (n.d.). "Instance-Based Prediction of Heart-Disease Presence with the Cleveland Database". In: *University of California* 3.1 (), pp. 3–2.