## **An Analysis of Free Data-Mining Tools with Regards to Medical field**

* 1. References:

Santos-Pereira, J., Gruenwald, L., & Bernardino, J. (2022). Top data mining tools for the healthcare industry. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 4968–4982. <https://doi.org/10.1016/j.jksuci.2021.06.002>

* 1. Purpose of paper

The primary objective of this paper is to provide an overview of the data mining techniques, methods, and various data types that contribute more towards the development of the health care industry. authors had clearly classified different types of data mining methods and different data sources that where we can get the information to mine, such as data from an image data in the form of a graph etc. and classified the data by clustering and identifying the outlines of the selected and getting only the use full information and authors suggested for the health care experts that to implement a data mining technique a tool should be used where it is readily available and can use it. authors had clearly classified different types of data mining methods and different data sources that By this point, the author has conveyed all of the necessary information to be covered by this paper.

* 1. Research Design or Strategy of paper:

We are all aware that the health care industry generates complex large data from a variety of data sources, such as images, medical reports, devices, and billing systems. These generated data are too complex to analyze using traditional methods, and there are no advanced statistical methods that can analyze large amounts of data; consequently, new data mining tools are being developed with the capacity to analyze data that is both complex and large. There are some features that can automatically clean the data when the data mining tool detects a change in the data. Knime is an end-to-end analytics tool developed in the United States that is based on the java programming language and used to integrate, transform, and analyze data. It has the capabilities of preprocessing, cleansing, and data mining tasks for knowledge discovery, among other things. One of the most useful features of this instrument is the ease with which it can be incorporated with other instruments. Following that is R, which is a programming language that was initially developed in C++ but now all programming is written in R itself. This instrument is helpful in all stages of the data mining analysis process. The fact that it can support extensions is one of its primary advantages. Scikit learning is the machine learning library for python. Its main advantages are that it can mine all data types and that it has good delivery performance. However, it is not end to end when it comes to spark fast general large scale data processing developed in scala programming. Next is Rapid miner, which is an end to end developed tool that has auto cleaning techniques and analyzing techniques. Its weakness is that the user needs to have advanced technical knowledge.

**1.4** **Results, Discussion, Conclusion of paper:**

**Results:**

It was found that there were gaps between the health care industry and data scientists. This led to the discovery of the significance of data mining, which led to the recommendation of some open-source data mining tools. Among these tools, rapid miner is the only one that is fully implemented from beginning to end. It will be more accurate in predicting results and will contribute more to the health care industry.

**Discussions:**

The rapid miner does a good job of covering the gap between the health care industry and advanced data statistics; however, there is no proper application of health care data into the tool, and analyzing the results is an area that needs additional discussion.

**Conclusions:**

From this paper, we were able to determine the gaps that exist within the health care industry, as well as how certain open-source software tools can assist in filling those gaps and achieving the best possible results among the tools that were suggested. The one that has all of the better features and is easier to implement in analyzing the statistical data is rapid miner.

* 1. Contribution from paper:

The author of this paper explained the various data sources that are available to us from the health care industry and suggested some tools that can contribute more towards the health care industry. One such tool is rapid miner, which gives the more satisfying results compared to other tools mentioned in terms of better results.

## SECTION II: CRITICAL ANALYSIS from paper

* 1. Overall Assessment of paper

In this article, the author presents a clear picture of the divide that exists between the health care industry and data scientists. The author also discusses data mining and the various data mining methods, as well as the different data types, data sources, and visualization techniques. Finally, the author provides an overview of the data mining tools that are open source and that contribute more to the health care industry, as well as a comparison of their capabilities.

* 1. Research Methodology of paper:

The author of this paper conducted a theoretical survey of the gaps that exist between data scientists and health care professionals, how those gaps should be filled in, and the various tools that are available on the internet and can be freely implemented towards research in the health care industry. At first, the authors categorized what data mining is and what the various methods of data mining are. They also explained about clustering, association, and how to detect outliers. Next, they focused on the various types of data and datatypes, data sources, cloud data, and various data sources. Finally, they explained about the data mining tools such as rapid miner knime, spark, etc., and their capabilities towards the health care industry. Finally, they compared the various tools and came to the conclusion about the tools.

* 1. Future Research

In this paper, the authors had clearly described the gap between data scientists and the health care industry and how data-driven decisions are not up to mark in the health care industry. They also suggested some data mining tools that are open sourced and the impact of the suggested tools in the health care industry, and they found rapid miner to be one of the best in the field of health care; however, there is room for improvement in that the tool has not been tested with the multiple scenarios with multiple regression models.

* 1. New Knowledge Learned:

As a result of reading this paper, I now understand that there is a knowledge gap between the scientists who work in health care and the data scientists who work in the health care industry. Since health care experts are knowledgeable about health care topics but lack the same level of expertise in data as computer experts, it is extremely challenging for health experts or data scientists to make decisions that are driven by data in the health care industry. In order to bridge this knowledge gap, several popular open-source data mining tools have been considered and compared with one another.

## Section III: Question to discussed

1) In what are the most important areas should the data mining tool focus its attention in terms of the aspects involved?

2) Does the health care industry have access to any other software besides rapid miner that is capable of more accurate prediction of the outcomes?