



# HEART FAILURE ANALYSIS AND PREDICTION

Advanced Business Analytics and  
Visualization

THINES KUMAR NADARAJA

TP063097

## Table of Contents

Introduction .....	2
Aims and Project Methodology.....	2
Project Scope .....	3
Conclusion.....	5
References .....	5

## Table of Figures

Figure 1: Three Vs of Big Data .....	3
Figure 2: Schematic Diagram of Big Data Analytics in the Healthcare Industry .....	4
Figure 3: Big Data Analytics in the Healthcare Industry .....	4

# Heart Failure Analysis and Prediction

---

## Introduction

Data published by the World Health Organization (WHO) in 2018 shows that Cardiovascular Diseases (CVD) are a contributing factor in causes of death, on a global scale (Lanzer et al., 2020). Big Data Analytics in Healthcare is pivotal in ensuring that CVD analysis and failure prediction can aid cardiologists and doctors alike (Sammani et al., 2019), with the appropriate treatment for patients (Kim, 2021). Big Data tools such as Apache Spark, is used for Predictive and Prescriptive Analytics in Healthcare (Thammasudjarit et al., 2018). Apache Spark is useful in healthcare organizations as it provides an open-source framework (Rammal & Z., 2018), which complements the Hadoop Distributed File System (HDFS) (Lanzer et al., 2020). Running HDFS along with Machine Learning (ML) algorithms, will certainly enable Big Data Analytics in Healthcare in prediction of CVD (Alexander & Wang, 2017).

## Aims and Project Methodology

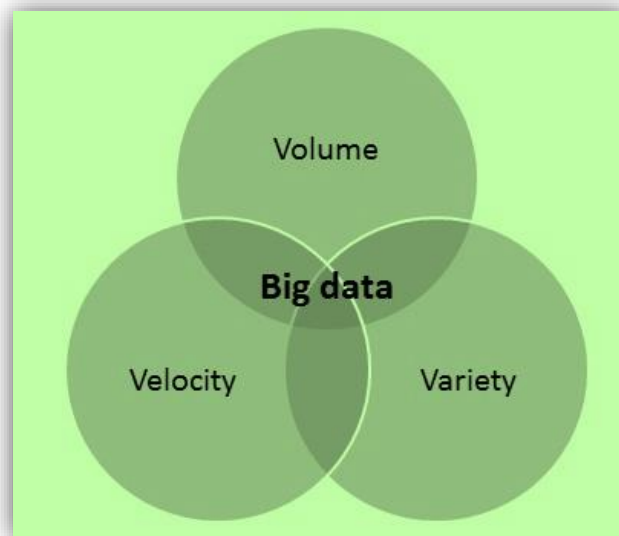
The aim of this project is to apply Big Data Analytics in Heart Failure Prediction, which will in turn, provide prevention of future diseases, based on factors that cause CVD. Procuring the dataset from Kaggle ("Heart Failure Prediction", 2021) containing 300 patients which have been previously affected with heart failure. Factors contributing to heart failure such as Anaemia, Level of CPK enzyme in blood, Diabetes, Blood Ejection Fraction, Hypertension, Blood Platelets, Creatinine Serum in Blood, Sodium Serum in Blood, Gender, Cigarette Smoking, Follow-Up Period and Death Events are analysed for this dataset. This dataset is then cleaned, using pre-processing techniques. Once dataset has been pre-processed, data analytics is then applied. Finally, using various data visualization techniques, a report that contains the analysis of the dataset is produced. Applying various Machine Learning modelling techniques on the dataset would provide a solution for Heart Failure Prediction, that would be deemed useful for the healthcare industry. The software used for this project are Tableau and SAS-Miner Studio.

## Project Scope

The scope of this project is to provide a solution for the healthcare industry using Big Data Analytics. Big Data Analytics ensure that knowledge procured from data in the healthcare industry, can be used to affect change and disrupt said industry (Roy et al., 2020). Big Data Analytics consist of 3 components, known as the 3 Vs of Big Data (Seward, 2021):

- 1) Volume – Volume of data is high, which would require analytics to be performed, to process and provide analysis of data.
- 2) Velocity – Data moves at a quick rate, and with real-time, processing such data is important for analytics.
- 3) Variety – Data comes in different forms; structured, unstructured, and semi-structured data. This would require pre-processing of data, which is a key component of Big Data Analytics.

The 3 Vs of Big Data is shown in the figure below:



*Figure 1: Three Vs of Big Data (Deka, 2014)*

Big Data Analytics in the healthcare industry, which is the domain for this project, can be applied with Predictive and Prescriptive Analytics (Roy et al., 2020). Predictive Analytics provide an indication of the future of the data, whilst Prescriptive Analytics provide a solution that considers historical data (Khan & Alotaibi, 2020). Machine Learning is an example of statistical techniques used for Predictive and Prescriptive Analytics (Deka, 2014).

Big Data Analytics can be applied in the healthcare industry, with the schematic diagram shown below:

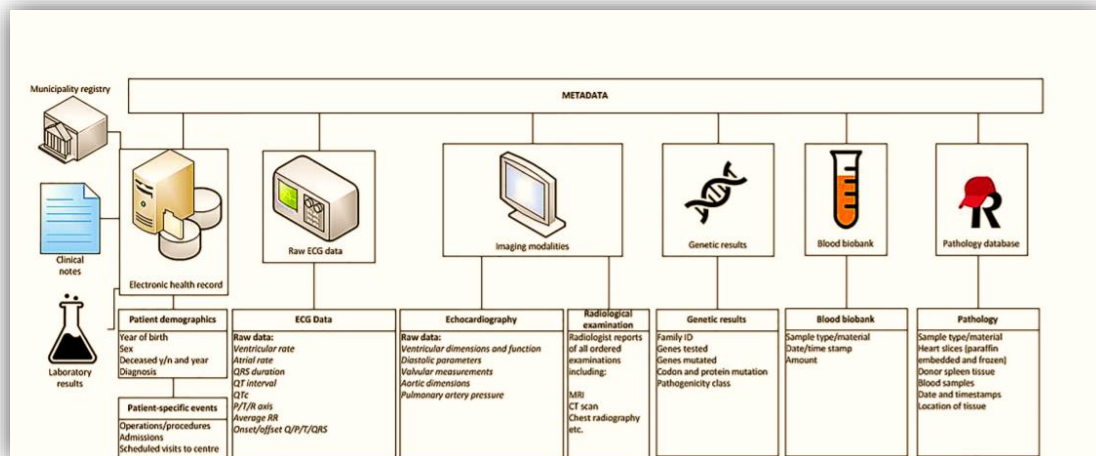


Figure 2: Schematic Diagram of Big Data Analytics in the Healthcare Industry (Sammani et al., 2019)

Big Data Analytics can be applied in the healthcare industry, by processing all forms of data, storing data in Hadoop, and finally, classifying data using Predictive Analytics. Next, Prescriptive Analytics is then applied to the data, as shown in Figure 3:

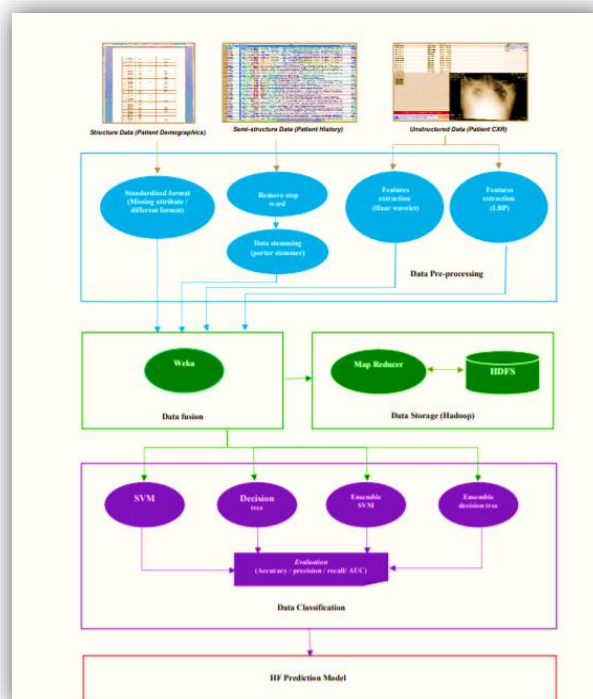


Figure 3: Application of Big Data Analytics (Sammani et al., 2019)

## Conclusion

Big Data Analytics in the Healthcare industry is a pivotal form of analytics using Data Science, and for the purpose of this project, predicting the rate of heart failure based on various dependent and independent variables, is sufficient in providing an accurate solution for prediction. This project outlines the methodology and purpose of carrying out Big Data Analytics for Heart Failure Prediction, using various techniques available in data science.

## References

1. Heart Failure Prediction. Kaggle.com. (2021). Retrieved 10 July 2021, from <https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>.
2. Alexander, C., & Wang, L. (2017). Big Data Analytics in Heart Attack Prediction. *Journal Of Nursing & Care*, 06(02). <https://doi.org/10.4172/2167-1168.1000393>
3. Deka, G. (2014). Big Data Predictive and Prescriptive Analytics. *Advances In Data Mining and Database Management*, 370-391. <https://doi.org/10.4018/978-1-4666-5864-6.ch015>
4. Khan, Z., & Alotaibi, S. (2020). Applications of Artificial Intelligence and Big Data Analytics in m-Health: A Healthcare System Perspective. *Journal Of Healthcare Engineering*, 2020, 1-15. <https://doi.org/10.1155/2020/8894694>
5. Kim, J. (2021). Big Data, Health Informatics, and the Future of Cardiovascular Medicine.
6. Lanzer, J., Leuschner, F., Kramann, R., Levinson, R., & Saez-Rodriguez, J. (2020). Big Data Approaches in Heart Failure Research. *Current Heart Failure Reports*, 17(5), 213-224. <https://doi.org/10.1007/s11897-020-00469-9>
7. Rammal, H., & Z., A. (2018). Heart Failure Prediction Models using Big Data Techniques. *International Journal Of Advanced Computer Science And Applications*, 9(5). <https://doi.org/10.14569/ijacsa.2018.090547>
8. Roy, A., Bruce, C., Schulte, P., Olson, L., & Pola, M. (2020). Failure prediction using personalized models and an application to heart failure prediction. *Big Data Analytics*, 5(1). <https://doi.org/10.1186/s41044-020-00044-2>
9. Sammani, A., Jansen, M., Linschoten, M., Bagheri, A., de Jonge, N., & Kirkels, H. et al. (2019). UNRAVEL: big data analytics research data platform to improve care of patients with cardiomyopathies using routine electronic health records and standardised biobanking. *Netherlands Heart Journal*, 27(9), 426-434. <https://doi.org/10.1007/s12471-019-1288-4>
10. Seward, J. (2021). Paradigm Shift in Medical Data Management.
11. Thammasudjarit, R., Pattanateep, A., & Pattanaprateep, O. (2018). Big Data Analytics in Healthcare. *Ramathibodi Medical Journal*, 41(2), 116-123. Retrieved 13 July 2021.