Data Science: A Novel Analytical Structure in Public Mental Health

Mohan Satvik Adusumalli

Disha Delphi Public School, Kota

DOI:10.37648/ijrst.v14i02.005

[1]*Received: 22 March 2024; Accepted: 19 June 2024; Published: 21 June 2024*

ABSTRACT

Applying data science to public mental health concerns and devising remedies based on research findings can be challenging and call for sophisticated methods. In contrast to traditional data analysis initiatives. It's critical to possess an extensive project management procedure to guarantee that Project associates are capable and knowledgeable enough to Implement the data science process. As a result, this essay offers a fresh paradigm that mental health practitioners might apply to address issues people encounter when applying data science. Even so, a sizable portion of Many studies on the mental health of the public have been published, not many have talked about data science's application to public mental health.

Data science has recently transformed how the healthcare business manages, analyses, and uses data. Because of the scientific methodology employed in data science initiatives, they differ from traditional data analysis. Motivating medical practitioners to use "Data Science" to mental health issues is one of the goals of launching this new framework. It's usually advantageous to have a strong data analysis framework and precise instructions for a thorough examination. Estimating the time and resources required early on in the process can also be helpful in gaining a clear understanding of the problem that needs to be solved.

INTRODUCTION

Any diagnosable mental illness that is typified by aberrant thinking, feeling, or behaviour is referred to as mental illness [1]. Even though mental illness is widespread and has a significant negative social and economic impact on society globally, diagnosis is still based on clinical judgment and expert opinion because there are currently no conventional biological diagnostic tests available [2]. For humans to be healthy generally, mental wellness is essential. The ability to deal with day-to-day challenges and the quality of life are both impacted by mental health. Physical and mental illnesses are closely related and can have disastrous effects on one another. Nowadays, there are numerous scientifically supported strategies for treating mental illness and promoting mental health. Both socially and health-wise, mental wellness has several advantages. The World Health Organisation (WHO) works to promote mental health among people everywhere. In addition, WHO works to safeguard human rights and supports mental health and well-being as well as the treatment of those who suffer from mental illnesses and their prevention. The World Health Organisation estimates that 800,000 individuals commit suicide annually, and many more make attempts at it [3]. Most of these documented suicide cases have some connection to common mental health issues. Most people with severe mental problems are frequently unable to access necessary mental healthcare services. Every suicide is a tragedy that has an impact on people's lives, families, and communities, and it often leaves a lasting legacy for those left behind. Suicide is the second largest cause of death worldwide for individuals aged 15 to 29 and can happen at any point in life [3]. Over the past few years, there has been a significant growth in the amount of data being collected in the mental healthcare industry. Instead of using typical statistical analyses on mental healthcare data, a novel exploratory research approach is made possible by the abundance of high-quality mental healthcare data sources and their availability. This approach helps uncover hidden patterns and new insights from the data. Hospitals' use of electronic medical records, or EMRs, has increased globally. Digital patient health records are created and kept safe and secure at hospitals using EMR software. Thus, the availability of patient data allows researchers to evaluate and

[1] How to cite the article: Adusumalli M.S.; June 2024; Data Science: A Novel Analytical Structure in Public Mental Health; *International Journal of Research in Science and Technology*, Vol 14, Issue 2, 42-49, DOI: http://doi.org/10.37648/ijrst.v14i02.005

uncover intriguing information that may help treat patients with mental health issues. Data-driven research has produced new scientific information, enhanced patient treatment quality, and decreased administrative costs in public mental health services throughout the last ten years. Recently, data science techniques have been applied to studies on mental health. Using scientific techniques like data mining, machine learning, time series analysis, or statistics, data science is the act of looking for patterns and hidden correlations in massive datasets. Data science frequently uses mathematical and statistical techniques to identify trends and determine correlations between variables.

Most data science projects in mental health now concentrate on algorithmic technical aspects rather than user involvement and project results. There is a communication gap between domain specialists and the technical team conducting the study because doctors who must deal with the results of a data science project are typically not familiar with the notion of data science [4]. Moreover, if healthcare professionals are not included in the project, they will feel inferior, preventing them from embracing technology and ultimately leading to project failure. Solving this issue is a collaborative effort that significantly reduces the difficulty of interacting with healthcare experts in the field during data science projects. Their regular participation in the study is not just necessary, but valued. Working together with subject matter specialists in this field is crucial, as it not only improves analysis but also makes it easier to put the findings into practice, thereby making them an integral part of the project.

Healthcare professionals can complete most of the analytical process using the suggested visual data mining approach. Visual data mining is the creation of visual representations that can be used in the data preparation, model derivation, and validation stages of the three stages of the data mining life cycle [5]. In this publication, we presented a novel visual data mining approach for the research of mental health.

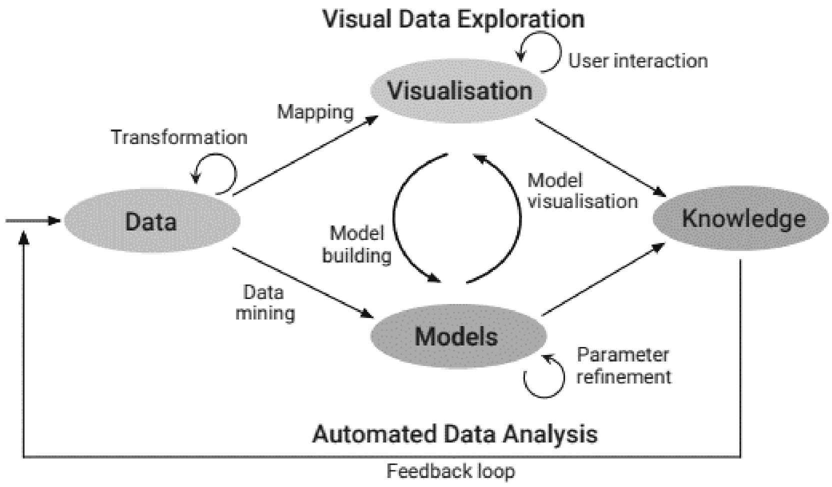


Fig 1: Visual analytics process for knowledge discovery [12]

PROPOSED FRAMEWORK

This study presents a novel framework for data science projects on public mental health for data analysis, fusion, storing, managing, processing, analysing, visualising, and modelling. The creation of this new framework was primarily driven by the need to support medical professionals in addressing public mental health challenges. In the absence of an appropriate framework for coordination and structure, there will likely be a lot of overlap and duplication between project stages, which could lead to misunderstandings about the roles of each project participant.

A. Definition of the problem

An issue definition is a summary of the topics that, including Data Science project storage, management not necessary analysis, visualisation to solve the issue in this phase. The problem can be described using the 5W1H approach. Who? What? Where? When? Why? and how are the acronyms represented by the five W's and the H. It is an effective way to systematically collect data in a difficult scenario. Keep your focus on the issue at hand and question your presumptions.

Participant: Problem definition must be defined by the healthcare professionals.

B. Acquisition of Needs

For each kind of project, gathering business needs is an essential first step. Requirement gathering is a crucial component of any data science or analytics effort. Many project failures can be attributed to inadequate requirements-gathering procedures. Obtaining insufficient requirements is the root of many design errors. Early in the project, a wide range of needs should be developed. This will enable more accurate cost estimates, shorter project durations, higher patient satisfaction, and increased accuracy of the final solution. It is advisable to wait until the project owners and participants completely understand the requirements before discussing technology or solutions. Producing a comprehensive and lucid requirements document and disseminating it to project stakeholders is crucial.

Participant: This will be a joint excise of project owners such as healthcare professional and data analysists/scientists.

C. Data Acquisition

Applications and projects involving data on mental health must get sophisticated, resource-rich data sets. Data science initiatives' most time- and money-consuming parts are data gathering and database upkeep. Data about mental health may be found in a variety of file types. Consequently, getting the correct data for the project could be challenging. Finding hidden patterns, trends, and linkages at national and local levels is easier with integrated analysis of various data sources from regional and national repositories. As such, obtaining a wide variety of linked datasets is usually beneficial.

i. Risk factors

ii. Services

iii. Protective factors

iv. Quality & outcomes

v. Prevalence & incidence

Participant: As this is a technical task, data analyst will perform this task. But we highly recommend getting help from healthcare professional who typically own the data sources.

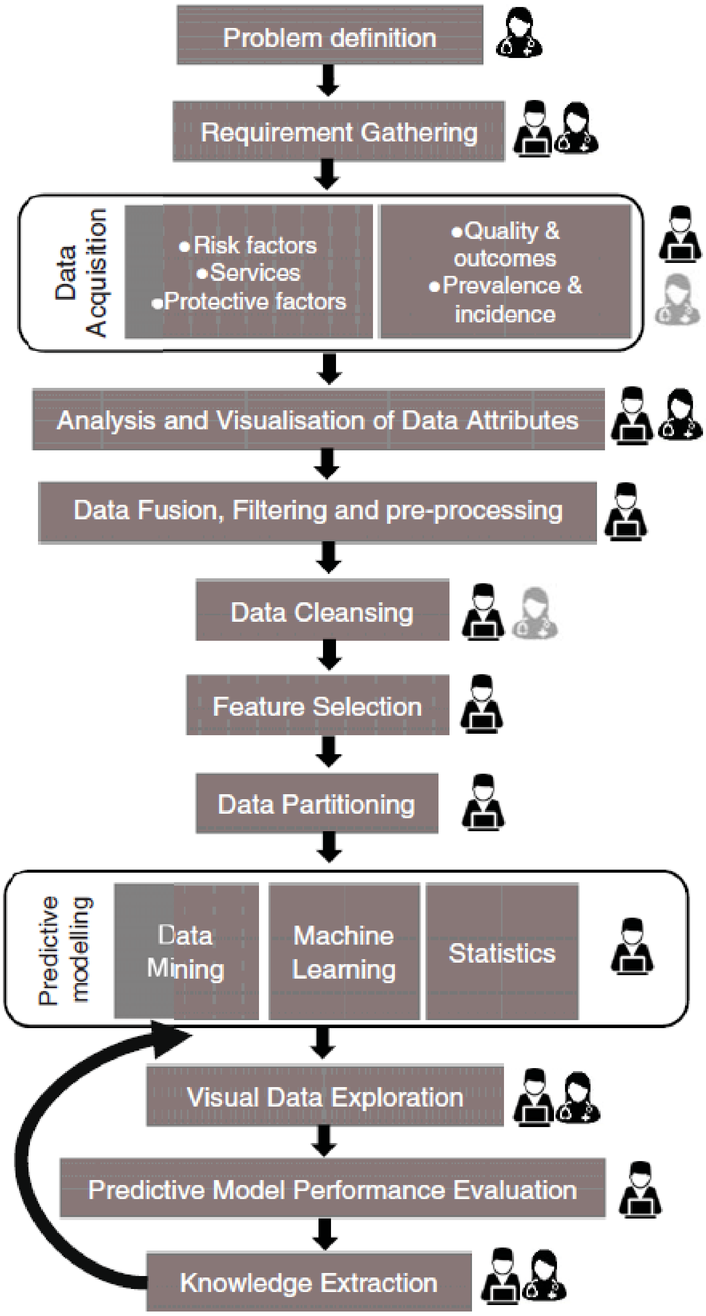


Fig 2: PROPOSED FRAMEWORK

D. Analysis and Visualisation of Data Attributes

By placing the data in a visual context, data visualisation helps project participants understand the information. If users can use the proper data visualisation techniques, correlations, patterns, trends, and relationships that might not be seen in the data can be found and identified more quickly. A data field that describes the qualities or features of a data object or dataset is known as a data attribute. Early in the project lifecycle, data attribute visualisation can prove advantageous for the overall project because it makes it easy to identify some qualities with high correlations among themselves and can identify some correlations, if any, between the attributes. In this phase, we evaluate and visualise the data attributes obtained from the earlier data-gathering stage. We construct the data in a way that allows project participants to benefit from the analysis without delving further into the data. At this point, we can employ a straightforward tool-simple pairwise scatter plots-to enhance our analysis, making the process more accessible and reassuring for everyone involved.

Participant: Project owners who work in the healthcare industry, data analysts, and scientists will collaborate on this.

E. Data Fusion, Filtering and pre-processing

To deduce pertinent circumstances and events connected to the observed environment, a data fusion (data integration) procedure seeks to maximise the valuable information content obtained by heterogeneous sources [13]. Integrating data from disparate sources enhances the potential to spot hidden trends, patterns, and connections. Data pre-processing transforms data into a format that facilitates more accessible and faster processing. One of the steps to examine, filter, and state data before going on to the modelling process is data filtering. Data processing and filtering steps could consume a significant amount of project time. For instance, to choose the optimal filter and pre-treatment strategy, one must have a solid understanding of the data content because an untreated NULL value can ruin future modelling efforts.

Participant: As this is highly technical task, data analyst will perform this task.

F. Data Cleansing

Data from the real world are sometimes noisy, inconsistent, and incomplete [18]. Data cleansing aims to enhance the quality of the data that the data analyst will use. If the data is accurate, consistent, complete, and comprehensive, the prediction model will have a high level of accuracy. Missing data can be handled using statistical techniques. Data cleansing is an expensive and time-consuming, laborious procedure, particularly when a lot of data is employed. Domain expertise is essential for tasks involving data purification.

Participant: Since it's a technical task, a data analyst will complete it. However, we strongly advise obtaining assistance from a medical professional. Understanding the mental healthcare domain is essential at this step, which might involve both data analysts and healthcare professionals.

G. Feature Selection

Feature selection is the process of choosing a subset of pertinent attributes to be utilised in model building in the discipline of data science. In addition to reducing the execution time and total dimensionality of the data with an improvement in execution performance, it offers the mechanism for identifying the useful patterns in the data [14]. The main goals of feature selection strategies are to decrease overfitting in the model, boost generalisation, and enhance model prediction performances and runtimes [15].

Participant: As this is highly technical task, data analyst will perform this task.

H. Data Partitioning

Building data mining and machine learning models requires dividing the data into training and testing sets. A model is trained or developed using the training set. Once a predictive model has been developed using the training set, a new dataset—also called the test or validation dataset—must be used to verify the predictive model's performance. Inadequate data partitioning can lead to subpar inference outcomes. Thus, data scientists and analysts should consider data partitioning techniques before creating the models.

Participant: As this is highly technical task, data analyst will perform this task.

I. Predictive modelling

The method of using data, arithmetic, and statistics to create data models for outcome prediction is known as predictive modelling [16]. By building, testing, and validating a model, predictive modelling determines the likelihood of a certain event. This is an iterative process that often entails training the model, testing it on the same set of data, and then determining which model best fits the business requirement [17]. Developing predictive models is helpful for any industry because it gives users access to hidden information about most of the difficult problems they encounter and enables them to make predictions that are comparatively accurate. Organisations must be able to anticipate future developments and opportunities that could challenge established beliefs in order to stay ahead of the competition in any given industry.

Participant: As this is highly technical task, data analyst will perform this task.

J. Visual Data Exploration

Automated analysis methods and interactive visualizations are combined in visual analytics to enable efficient comprehension, deductive reasoning, and decision-making from considerable and complicated information [18]. Visual data exploration is entirely human-guided, which sets it apart from data mining significantly. The academic and industry groups have expressed interest in the visually explored data and model findings, which are seen as a fascinating application. You can choose from a range of visual analytics techniques for your solution, from straightforward bar plots to intricate visualization plots.

Visual data exploration is a process that demands domain expertise. This expertise is crucial for bringing a wealth of information to the stage, enabling a deeper understanding and analysis of the results.

Participant: This will be a joint excise of project owners such as healthcare professional and data analysists/scientists.

K. Predictive Model Performance Evaluation

Knowing how well a predictive model fits the data is essential to determining the model's genuine worth. Performance evaluation, therefore, has a significant influence on predictive modelling technologies. Choosing the appropriate metrics allows for calculating and comparing predictive model performance. Consequently, selecting the proper measures for a particular prediction model is crucial to obtaining an accurate result. Additionally, as different kinds of data sets will be used for the same predictive model, it is essential to assess suitable prediction models. The Confusion Matrix and ROC Curve can be utilised for a given prediction model for a critical performance evaluation.

Participant: As this is highly technical task, data analyst will perform this task.

L. Knowledge Extraction

The intricate knowledge extraction process makes it possible to identify novel, potentially helpful information and hitherto unidentified structures from massive amounts of data [19]. The difficult task of extracting legitimate, unique, useful, and eventually intelligible patterns from huge data sets is known as knowledge discovery. An interdisciplinary field called knowledge extraction and discovery focuses on techniques for drawing meaningful conclusions from data.

Participant: This will be a joint excise of project owners such as healthcare professional and data analysists/scientists.

EXPERIMENTAL EVALUATION

A. Problem definition

Objective of this study is understanding the behavioural factors associated with mental health among various geographical areas.

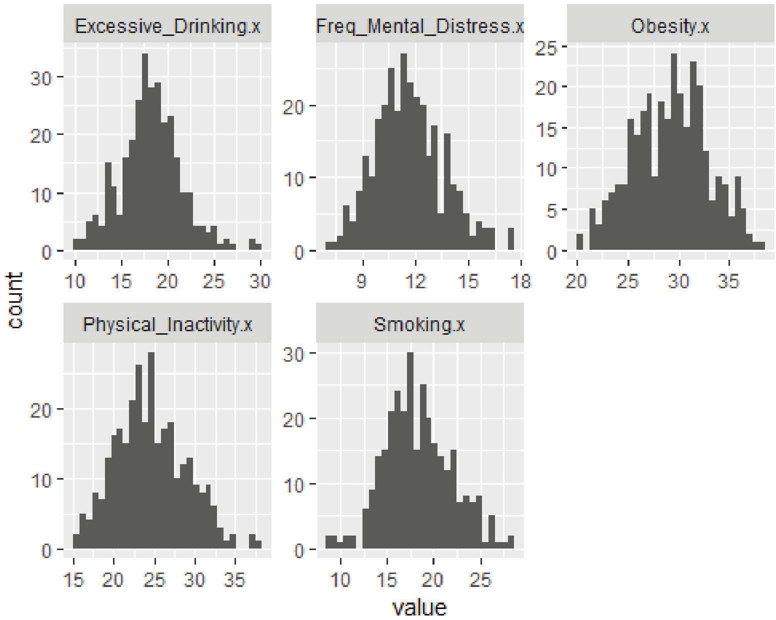
B. Requirement Gathering

In national and local level, understanding of factors associated with public health issues like mental health is paramount important. This project aims to use the decision Tree technique to improve the degree of understanding of the mental health among various geographical areas by identifying behavioural factors associated with mental health.

C. Data Acquisition

This research study has used last 6 years of America's Health Rankings Annual Report [20], which is the longest-running annual assessment of the nation's health on a state-by-state basis in US to extract data for this study. The research dataset has been constructed by the authors using five key health measures reported in the America's Health Rankings Annual Report.

D. Analysis and Visualisation of Data Attributes



E. Data Fusion, Filtering and pre-processing

The dataset consists of 306 data points each described by 6 attributes. Even there are 6 attributes in the dataset, attribute called "State" will not been used to do the analysis due to the nature of the attribute. Other five numerical attributes will be considered to do the study, but discretization has been applied to those five attributes due to the continues values. This study has used k-means clustering method to do the discretization and in this study k=3 has been used.

F. Data Cleansing

In this study US territories like Puerto Rico, Virgin Islands have been excluded from the research due to the incomplete data throughout the study period.

G. Feature Selection

Feature section is not necessary as only 4 independents variables and one dependent variable exist in the study.

CONCLUSION

Compared to traditional data analysis projects, comprehending public mental health concerns and finding solutions can be challenging and require advanced methodologies. Data science is a rapidly developing healthcare subject with enormous promise for improving our ability to predict and comprehend illness classification, particularly mental health. A data science project management procedure is necessary to guarantee that project participants are capable and knowledgeable enough to begin and carry out the project effectively. This paper's primary goal is to provide a new framework that data scientists in the mental health field can use to solve problems.

This framework is very regimented and strongly emphasizes preplanning. With this waterfall-style architecture, every project must be planned from start to finish, with each phase starting only after the completion of the preceding one. Motivating medical personnel to use "data science" to address mental health issues is one of the goals of launching this new framework. To demonstrate to readers the utility of the proposed framework, it has been adopted and validated in the experimental assessment part using an actual case study.

REFERENCES

1. Dooshima, M. P. (2018). A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining. International Journal of Immunology, 6(1), 5. doi:10.11648/j.iji.20180601.12
2. Diederich, J., Al-Ajmi, A., & Yellowlees, P. (2007). Ex-ray: Data mining and mental health. Applied Soft Computing, 7(3), 923-928. doi:10.1016/j.asoc.2006.04.007
3. Suicide. (n.d.). Retrieved April 21, 2019, from https://www.who.int/newsroom/fact-sheets/detail/suicide

4. Menger, V., Spruit, M., Hagoort, K., & Scheepers, F. (2016). Transitioning to a Data Driven Mental Health Practice: Collaborative Expert Sessions for Knowledge and Hypothesis Finding. Computational and Mathematical Methods in Medicine, 2016, 1-11. doi:10.1155/2016/9089321
5. Kopanakis, I., & Theodoulidis, B. (2003). Visual data mining modeling techniques for the visualization of mining outcomes. Journal of Visual Languages & Computing, 14(6), 543-589. doi:10.1016/j.jvlc.2003.06.002
6. Public Health. (2017, August 30). Better mental health: JSNA toolkit. Retrieved April 21, 2019, from https://www.gov.uk/government/publications/better-mental-health-jsnatoolkit
7. Saunders, M. N., Lewis, P., & Thornhill, A. (2019). Research methods for business students. New York: Pearson.
8. Exploratory Case Study. (2010). Encyclopedia of Case Study Research. doi:10.4135/9781412957397.n139
9. Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. Sustainable Cities and Society, 38, 230-253. doi:10.1016/j.scs.2017.12.034
10. Pedrycz, W. (2011). Introducing WIREs Data Mining and Knowledge Discovery. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 1-1. doi:10.1002/widm.17
11. Oliveira, M. D., & Levkowitz, H. (2003). From visual data exploration to visual data mining: A survey. IEEE Transactions on Visualization and Computer Graphics, 9(3), 378-394. doi:10.1109/tvcg.2003.1207445
12. Oliveira, M. D., & Levkowitz, H. (2003). From visual data exploration to visual data mining: A survey. IEEE Transactions on Visualization and Computer Graphics, 9(3), 378-394. doi:10.1109/tvcg.2003.1207445
13. F. Mastrogiovanni, A. Sgorbissa and R. Zaccaria. A Distributed Architecture for Symbolic Data Fusion. In IJCAI-07, pp 2153-2158. 2007.
14. Abdullah, A. S., Ramya, C., Priyadharsini, V., Reshma, C., & Selvakumar, S. (2017). A survey on evolutionary techniques for feature selection. 2017 Conference on Emerging Devices and Smart Systems (ICEDSS). doi:10.1109/icedss.2017.8073659
15. I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182 . http://dl.acm.org/citation.cfm?id= 944919.944 96 8 .
16. Predictive Modeling: The Only Guide You Need. (n.d.). Retrieved from https://www.microstrategy.com/us/resources/introductoryguides/predictive-modeling-the-only-guide-you-need
17. Saraee, M., & Silva, C. (2018). A new data science framework for analysing and mining geospatial big data. Proceedings of the International Conference on Geoinformatics and Data Analysis - ICGDA 18. doi:10.1145/3220228.3220236
18. D.A. Keim, J. Kohlhammer, G. Ellis, F. Mannsmann (Eds.), Mastering the Information Age. Solving Problems with Visual Analytics, Eurographics Association, Goslar, 2010.
19. Jebbor, F., & Benhlima, L. (2014). Overview of knowledge extraction techniques in five question-answering systems. 2014 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14). doi:10.1109/sita.2014.6847287