

Data Science in Public Mental Health: A New Analytic Framework

Charith Silva

School of Computing, Science and
Engineering
University of Salford-Manchester
h.c.e.silva@edu.salford.ac.uk

Mahsa Saraee

School of Cardiovascular Sciences
Faculty of Biology, Medicine and Health
University of Manchester
mahsa.saraee@postgrad.manchester.ac.uk

Mo Saraee

School of Computing, Science and
Engineering
University of Salford-Manchester
m.saraee@salford.ac.uk

Abstract— Understanding public mental health issues using data science and finding solutions based on the findings from the data science projects can be complex and requires advanced techniques, compared to conventional data analysis projects. It is important to have a comprehensive project management process to ensure that project associates are competent and have enough knowledge to implement the data science process. Therefore, this paper presents a new framework that mental health professionals can use to solve challenges they face using data science. Although a large number of research papers have been published on public mental health, few have addressed the use of data science in public mental health. Recently, Data Science has changed the way we manage, analyze and leverage data in healthcare industry. Data science projects differ from conventional data analysis, primarily because of the scientific approach used during data science projects. One of the motives for introducing a new framework is to motivate healthcare professionals to use "Data Science" to address the challenges of mental health. Having a good data analysis framework and clear guidelines for a comprehensive analysis is always a plus point. It also helps to predict the time and resources needed in the early in the process to get a clear idea of the problem to be solved.

Keywords- Mental Health, Public Health, Data Science, Data Mining, Visual Data Exploration

I. INTRODUCTION

Mental illness refers to all diagnosable mental disorders which are characterized by abnormalities in thinking, feelings or behaviours [1]. Mental illness is very common and causes substantial social and economic burden worldwide, yet no standardised biological diagnostic tests are available, and diagnosis is still dependent on clinical skills and expert opinion [2]. Mental health is vital for overall well-being of human. Mental well-being is important for the quality of life and the ability to cope with data to day life. Mental illness can have a dire impact on and correlation with physical illness. Currently, there are many scientifically based methods that promote mental well-being and treat mental illness. Mental well-being offers many health benefits as well as social benefits. World Health Organisation (WHO) work in all round the world to improve the mental health of individuals. Also, WHO is promoting mental well-being, care for persons with mental disorders, prevention of mental disorders and work to protection of human rights. According to the WHO statistics, every year approximately 800,000 people take their own life and there are many more people who attempt suicide [3]. The majority of these reported suicide cases are related to common mental health conditions. In many cases, required mental healthcare services remains out of reach for most people with severe

mental disorders. Every suicide is a tragedy that affects individuals, families, and communities, having long-term consequences for people who are left behind. Suicide occurs throughout the lifespan and is the second leading cause of death among 15-29-year-olds globally [3].

The amount of data that is being collected in mental healthcare sector has increased enormously over the past years. Availability of large number of mental healthcare data sources and the quality of data enables a novel exploratory research approach to finding hidden patterns and new knowledge from the data rather than conducting traditional statistical analyses on mental healthcare data. Electronic medical records (EMR) has become more widespread in hospitals in worldwide. EMR software helps create and store patient's digital health records safe and securely in the hospital. So, availability of patient data enable researcher to analyse and find interesting information which can be beneficial to treat mental health patients. Over the past 10 years, data-driven research has led to new scientific knowledge, improved quality of patient care and reduced administrative cost in public mental health services. Data Science approaches have been recently used in mental health studies. Data Science is a process of searching for previously unknown patterns, and hidden relationships in a large dataset using scientific methods such as data mining, machine learning, time series analysis, or statistics. In many cases, data science uses statistical and mathematical algorithms to detect patterns and find correlation between variables.

Most mental health data science projects currently focus on algorithm driven technical characteristics of the project, not on user participation and project outcomes. The clinicians that have to work with the outcomes of a data science project are usually unfamiliar with the concept of data science, which creates a gap between domain experts and the technical staff performing the analysis [4]. Furthermore, not involving clinical professionals in the project will make them feel surpassed, resulting in failure to adopt to technology thus leading to project failure. Tackling this problem will greatly alleviate the challenge of interaction with the local healthcare professionals during the data science project, requiring them to consistently engage in the analysis. Collaboration with domain experts in this area will enhance analysis and pave the way for easy implementation of the results that will eventually be discovered. The proposed visual data mining approach contains the ability to engage healthcare professionals in the most parts of analysis process. Visual data mining involves the invention of visual representations that could be applied in all three data-mining

life cycle stages, as partitioned to the data preparation, model derivation and validation stage [5]. In this paper, we proposed novel visual data mining framework for Mental Health study.

II. RELATED WORK

The Joint Strategic Needs Assessment (JSNA) is a process by which local governments and clinical groups assess the present and future healthcare and wellbeing needs of the local community to inform local decision-makers in UK. In the context of mental health, the main aim of this framework is to understand the determinants of mental health in an area and consider social and contextual factors that affect mental health, such as employment, crime, safety and housing [6]. The mental health of most of the people are influenced by their social settings. Good mental health is fundamental to our physical health. The JSNA toolkit (Fig: 1) contains resources designed to support local government and healthcare planners in understanding the mental health and wellbeing needs of their local population. The two main components of the toolkit are the Data Profile and the Knowledge Guide document.

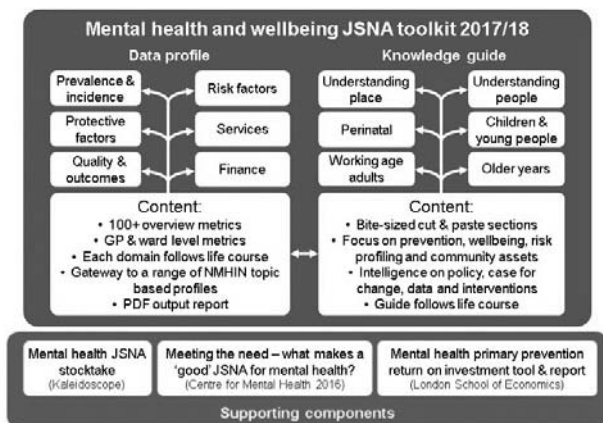


FIGURE 1: JSNA TOOLKIT STRUCTURE

The toolkit aims to support and assist the local process. Toolkit components are designed to work together to meet different localised needs. They provide an overview of the path of public mental healthcare and provide a portal for detailed information and knowledge about specific aspects of mental illness care, mental well-being, and support available and requirements by people with mental health problems. It is directed towards those seeking to understand the breadth and complexity of mental health issues in their area, particularly those working in public health. Our proposed framework is an extension of this JSNA toolkit and it allows to use JSNA recommended data indicator and carryout advance scientific experiment on public mental health.

III. METHODOLOGY

This section explains the research methodology and approach used for this study. As outlined by Saunders et al. [7] the purposes of a research could be categorised as exploratory, descriptive and explanatory. An exploratory study can be described as valuable means of finding out what is happening to seek new insights [7]. It can be particularly useful in

supporting understanding of the problem, explaining the nature of the challenge, or identifying problems that are usually included. In addition, it allows you to develop suggestions and assumptions for further research; discover new insights or better understand the problem. Authors have used exploratory case study approach for this study. The exploratory case study investigates distinct phenomena characterized by a lack of detailed preliminary research, especially formulated hypotheses that can be tested, and/or by a specific research environment that limits the choice of methodology [8]. Authors will use existing literature of Knowledge Discovery Tools, Visual Data Exploration technique and Mental Health and Wellbeing JSNA Toolkit to develop the new framework.

A. Knowledge Discovery (KDD)

Authors of this study use Knowledge Discovery (KDD) process conjunction with data mining to identify the any hidden patterns in the data. Knowledge Discovery (KDD) is an interdisciplinary area focusing upon methodologies for extracting useful knowledge from data [9]. It is a complete process of digging deep and discovering useful previously unknown knowledge from data. The knowledge discovery process consists of a set of clearly defined sub steps to be followed by practitioners when executing a knowledge discovery project [10]. The knowledge obtained from the proposed model and its findings can be analysed by coinciding the context information. The impact and relationships of knowledge can then be correlated. Thus, this knowledge can be merged through its interlinkages and relationships, creating comprehensive knowledge for public mental health. There are several challenges exist in the KDD process, including data acquisition, data extraction, data store, data cleansing and filtering and visualizing. Saree et al. [17] has introduced A New Data Science Framework for Analysing and Mining Geospatial Big Data. This framework can be used to plan, design and execute entire data science project. This framework will be used as a KDD process in this study.

B. Visual Data Exploration

Visual Data Exploration technique and data mining techniques will be used gain hidden knowledge and understand the correlation between factors associate with mental health. Both techniques require user involvement at different phases, and visualisation will be used to support the knowledge discovery and pattern recognition. Exploratory Data Analysis is the special data mining task which visualisation is playing a major role. Model visualization is the process of using visual techniques to make the discovered knowledge understandable and interpretable by humans. The degree of automation of data mining algorithms varies considerably as different levels of human guidance and interaction are usually required. However, it is the algorithm, not the user, that will look for patterns [11]. Geospatial data have become important as many visual analytics approaches require finding spatial patterns and relationships between the data points.

Below Visual Analytics process (Shown in Fig: 2) was introduced by Kohlhammer and his team [12] to describe the process of visual analysis. The process begins with data retrieval and transformation, then an automatic data analysis approach such as data mining is employed to evaluate models and extract the information. Information visualisation implies

that users collaborate with the visual interface towards exploring and analysing data to a greater extent [12].

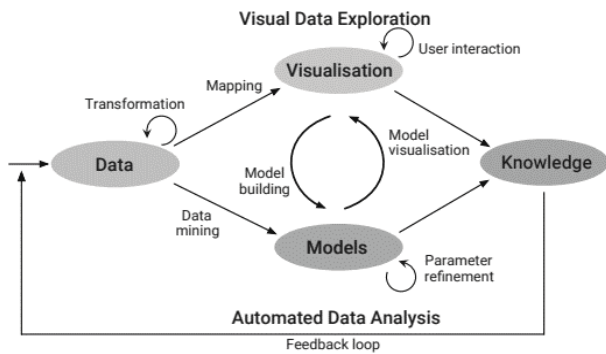


FIGURE 2: VISUAL ANALYTICS PROCESS FOR KNOWLEDGE DISCOVERY [12]

IV. PROPOSED FRAMEWORK

This research presents a new data analytic framework for data acquisition, data fusion, data storing, managing, processing, analysing, visualising and modelling for public mental health related data science projects. One of the main motivations for building this new framework is to help healthcare professionals to address public mental health issues. We believe that without a proper coordination and structuring framework there is likely to be much overlap and duplication amongst project phases, and can cause confusion around the responsibilities of each project participant.

A. Problem definition

A problem definition is a brief description of the subjects that need to be addressed by the Data Science project. In this step, it does not need to describe your approach to addressing the problem. The 5W1H technique can be used to describe the problem. The five W's and the H are acronyms for Who? What? Where? When? Why? and How? It is a good means of gathering information methodically in a challenging situation. When considering the problem, stay focused and challenge your assumptions.

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

B. Requirement Gathering

Compiling business requirements is a vital first step for any type of project. Requirement gathering is an important part of any Data Science or Data Analytic project. Poor requirements gathering techniques are the cause of many project failures. Gathering incomplete requirements are the cause of many design flaws. The development of a wide range of requirements early in the project will allow for accurate cost estimates; shorter project time periods; greater patient satisfaction, and improved final solution accuracy. It is always best to avoid talking about technology or solutions until the requirements are fully understood by the project owners and participants. It is important to create a clear, and complete requirements document to share it with project participants.

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

C. Data Acquisition

The acquisition of complex, resource-rich data set is essential

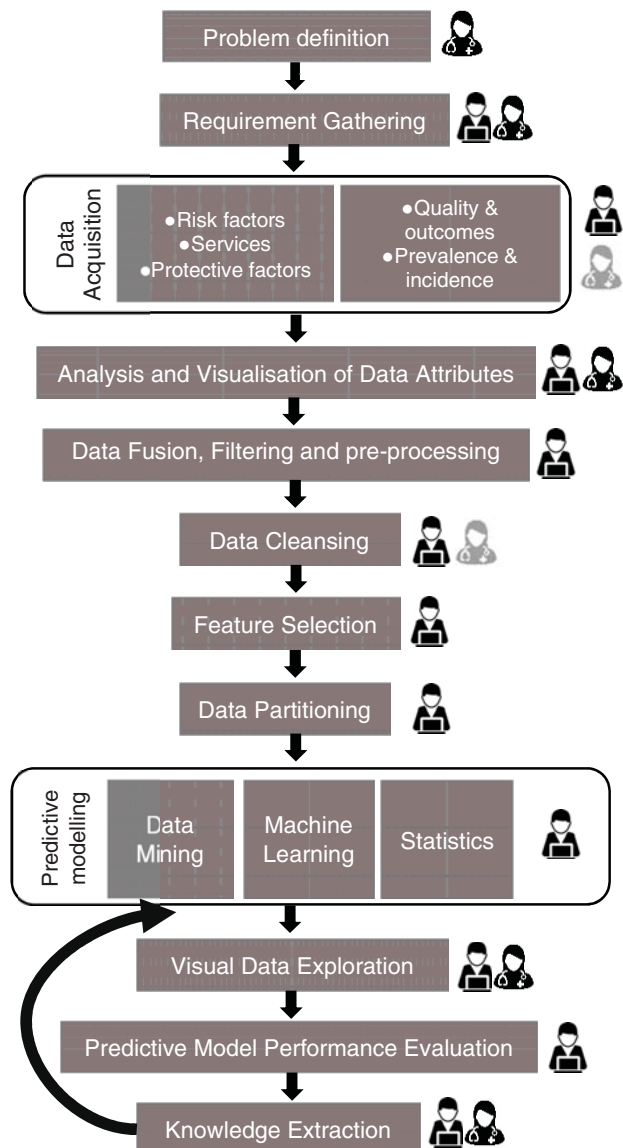


FIGURE 3: PROPOSED FRAMEWORK

for mental health data projects and applications. Data collection and database maintenance are the most cost-effective and time-consuming aspects of data science projects. Mental health related data can be available in different file formats. As a result, it may be difficult to obtain the correct type of data for the project. Integrated analysis of different types of data from the local and national repositories improves the ability to identify hidden patterns, trends and relationships at a national and local level. Therefore, it is always good to acquire many different related datasets.

- 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.

D. Analysis and Visualisation of Data Attributes

Data visualization aids project participants to understand the data by engaging it in a visual context. Correlations, Patterns, trends, and links that may not be detected in the data are detected and identified more easily if users can use appropriate data visualization techniques. Data Attribute can be describing as a data field that represents characteristics or features of a data object/ Dataset. Data attribute visualisation in the early stage of the project can be beneficial for entire project. Because, early data attribute visualisation can spot some of the correlation (if any) and you can easily spot some attributes having strong correlations amongst themselves. In this stage, we visualize the data attribute which acquire from the previous data acquisition stage which has been analysed and design it in a way that project participants can take advantage of the analysis without mining deeper into data. Basic pair-wise scatter plots can be used in this stage.

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

E. Data Fusion, Filtering and pre-processing

The aim of a data fusion (data integration) process is to maximize the useful information content acquired by heterogeneous sources in order to infer relevant situations and events related to the observed environment [13]. Data integration from heterogeneous data sources helps to improve the capability to identify hidden patterns, trends, and relationships. Data pre-processing converts data to a format that will be processed more easily and efficiently. Data filtering is one of the steps to explore, filter, and state data before shifting to the modelling process. Data filtering and processing steps may take a lot of allocated project time. A good understanding of data content is necessary to determine the best way to filter and pre-treat, for example: the untreated NULL value can destroy any modelling activities in the future.

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

F. Data Cleansing

Real world data are often incomplete, inconsistent, and noisy [18]. The purpose of data cleansing is to improve the quality of data that will be exercised by the data analyst. High accuracy of the prediction model will be achieved if the data is comprehensive, complete, consistent and accurate. Statistical methods can be used to treat missing data. Data cleansing is a tedious process that takes a lot of time and cost, especially when large amounts of data are used during the data cleansing process. Domain knowledge is vital for data cleansing activities.

Participant: As this is technical task, data analyst will perform this task. But we highly recommend getting help from healthcare professional. This is because mental healthcare domain knowledge is vital for this stage, both healthcare professional and data analyst can be involved in this stage.

G. Feature Selection

In the Data Science field, feature selection is the process of selecting a subset of relevant attributes to be used in model building. It provides the mechanism of determining the useful patterns in the data, which then decreases the execution time, and decreases the overall dimensionality of the data with an improvement in execution performance [14]. Feature selection techniques are primarily intended to improve model prediction performances and runtimes, to reduce model overfitting, and to increase generalization [15].

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

H. Data Partitioning

Dividing data into training and testing sets is an important part of building data mining and machine learning models. The training set is used to train or build a model. Once you create a predictive model using the training set, you must validate the performance of the predictive model using new dataset, the data set is known as the test dataset or validation dataset... Poor data partitioning may cause poor inference results. Thus, data analyst/ Data Scientist should consider data partitioning methods before building the models data.

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

I. Predictive modelling

Predictive modeling is a process that uses data, mathematics and statistics to predict outcomes with data models [16]. Predictive modelling build, test and validate a model to predict the probability of an outcome. This is an iterative process and often involve training the model; testing the model on the same data set and finally find best fit model based on the business requirement [17]. Building predictive models are useful in any industry because they provide hidden insight about most of the complex questions they face and allow users to create reasonably accurate predictions. To maintain a competitive advantage in any industry, it is essential that organisations have insight into future events and opportunities that challenge key assumptions.

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

J. Visual Data Exploration

Visual analytics combines automated analysis techniques with interactive visualisations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets [18]. A key difference between data mining and visual data exploration is that visual data exploration is a completely human guided process. Visual exploration of the data and the results from the models have been considered as an interesting application and attracted attention from both the academic and industry communities. Visual analytics methods can be selected for your solution in a variety of ways, ranging from simple bar plot to complex geo visualisation plots. Domain knowledge is vital for visual data exploration and it will add lot of knowledge to this step to understand and interpret the results.

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

K. Predictive Model Performance Evaluation

To get the true value of a predictive model, it is vital to know how good the model fit the data. Therefore, performance assessment plays a dominant role in predictive modeling technology. Predictive model performance is calculated and compared by selecting the correct metrics. Therefore, it is vital to choose the correct measurements for a given predictive model to get an accurate result. Also, it is very important to evaluate appropriate predictive models because several types of data sets will be used for the same predictive model. Confusion Matrix and ROC Curve can be used for basic Performance Evaluation for given prediction models.

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

L. Knowledge Extraction

Knowledge extraction is a complex process allowing the identification of previously unknown structures and potentially useful original information from large amount of data [19]. Knowledge Discovery is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from large data collections. Knowledge Discovery and Knowledge Extraction is an interdisciplinary area focusing upon methodologies for extracting useful knowledge from data.

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

V. 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.

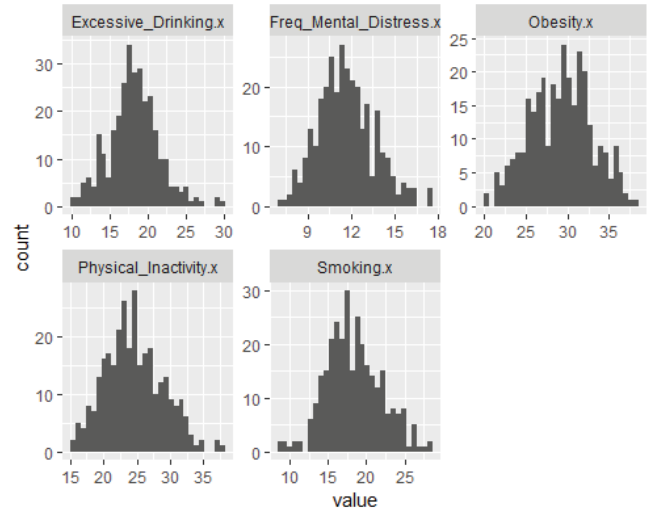
C. Data Acquisition

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.

ATTRIBUTE	TYPE
STATE	Factor
EXCESSIVE_DRINKING	Numeric
OBESITY	Numeric

PHYSICAL_INACTIVITY	Numeric
SMOKING	Numeric
FREQ_MENTAL_DISTRESS	Numeric

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 be 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 continuous values. This study has used k-means clustering method to do the discretization and in this study k=3 has been used. Table 2 shows the categorical value and the count after discretizing the continuous variables.

	Excessive Drinking	Obesity	Physical Inactivity	Smoking	Mental Distress
High	68	68	76	76	58
Medium	169	133	140	127	140
Low	69	105	90	103	108

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 independent variables and one dependent variable exist in the study.

H. Data Partitioning

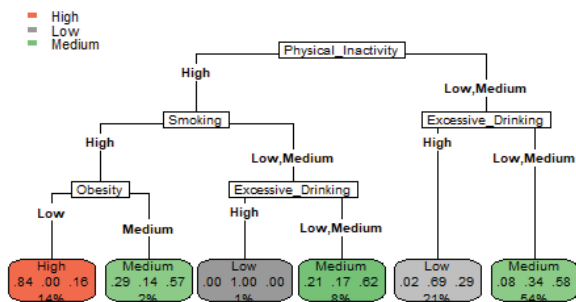
The study dataset was split into two datasets. The first subset, which contain 6 years' worth (2011- 2016) of the records, was used for training the decision tree model. The data in year 2017 was used for performance evaluation.

I. Predictive modelling

	High	Low	Medi
- Mental_Distress is High	[.84	.00	.16]
when Physical_Inactivity is High & Smoking is High & Obesity is Low			
- Mental_Distress is Low	[.02	.69	.29]
when Physical_Inactivity is Low or Medium & Excessive_Drinking is High			

- Mental_Distress is Low [.00 1.00 .00]
 when -Physical_Inactivity is High & Smoking is Low or Medium &
 Excessive_Drinking is High
 - Mental_Distress is Medium [.29 .14 .57]
 when Physical_Inactivity is High & Smoking is High & Obesity is Medium
 - Mental_Distress is Medium [.08 .34 .58]
 when Physical_Inactivity is Low or Medium & Excessive_Drinking is Low or
 Medium
 - Mental_Distress is Medium [.21 .17 .62]
 when Physical_Inactivity is High & Smoking is Low or Medium &
 Excessive_Drinking is Low or Medium

J. Visual Data Exploration



K. Predictive Model Performance Evaluation

In order to evaluate the performance of the developed model, we used a confusion matrix of the results, correctly and incorrectly predicted records for each class are illustrated in Table below:

	PREDICT: LOW	PREDICT: MEDIUM	PREDICT: HIGH
ACTUAL: LOW	7	5	0
ACTUAL: MEDIUM	4	25	4
ACTUAL: HIGH	0	2	4

Accuracy is the percentage of correctly classified instances in the validation dataset. It can be calculated as the number of all correct predictions divided by the total number of the data points. The best accuracy is 1, whereas the worst is 0.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}}$$

Accuracy = 70.58%

L. Knowledge Extraction

Knowledge extraction is a complex process and joint excise of healthcare professional and data analyst. It allows to identify previously unknown knowledge and potentially useful hidden information from predictive models. In this project, according to the previously built decision tree model, mental distress will be high when physical inactivity and smoking is high and obesity is low. But this need to be examined by healthcare professionals before drawing any conclusions.

VI. CONCLUSION

Understanding public mental health issues and find solution can be complex and requires advanced techniques, compared to conventional data analysis projects. Data Science in healthcare is an evolving field of great importance to deliver a projection and deeper understanding of disease classification, specifically in mental health. For this reason, it is essential to have a data science project management process to ensure that project participants are competent and have enough knowledge to start

and progress the project successfully. The main objective of this paper is to present a new structure that mental health professionals can use to solve challenges using data science. This framework is highly structured with a heavy emphasis on upfront planning. This waterfall style framework requires a project to be planned to begin to end, with no phase of a project beginning until the previous one has ended. One of the motives for introducing a new framework is to motivate healthcare professionals to use "data science" to address the challenges of mental health. Proposed framework has been adopted and validated in the experimental evaluation section with real case study to show the value of the proposed framework to readers.

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/news-room/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-jsna-toolkit>
- [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. Mastrogianni, 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=944919.944968> .
- [17] Predictive Modeling: The Only Guide You Need. (n.d.). Retrieved from <https://www.microstrategy.com/us/resources/introductory-guides/predictive-modeling-the-only-guide-you-need>
- [18] 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
- [19] D.A. Keim, J. Kohlhammer, G. Ellis, F. Mannsmann (Eds.), *Mastering the Information Age. Solving Problems with Visual Analytics*, Eurographics Association, Goslar, 2010.
- [20] Jebbor, F., & Benhlila, 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