Predictive Modelling in Mental Health: A Data Science Approach

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— In national and regional level, understanding of factors associated with public health issues like mental health is paramount important to improve the awareness. This study aims to use the data mining techniques such as association rule mining to improve the degree of understanding the mental health among various geographical areas by identifying various vital behavioural factors associated with mental health issues. The study will produce interesting relationships among the behavioural factors in form of Association rules. The outcomes of this research will be beneficial to organisations that work in public health sector to improve mental health among the citizens. Also, this proposed new data science approach will be beneficial to improve the degree of understanding by identifying factors associated with mental health within the different geographical areas such as city or state level. The study found that states in US which have low excessive drinking percentage and high obesity and high smoking percentage has the highest frequent of mental distress. Also, these rules have shown high confidence threshold among females rather than males. Mental health related authorities who works in local and national governments may be used these findings to improve awareness of mental health.

Keywords- Data Science, Mental Health, Public Health, Data Mining, Association rules

I. INTRODUCTION

Field of data science is a combination of statistic, data mining, machine learning and domain expertise which seeks to provide meaningful information from large amounts of complex datasets [1]. Data science may help decision making in the relevant organisation. It can be applied to many organisations such as commercial organisations, government and not-for-profit organisations. According to the Royal Statistical Society (RSS) in the UK, Data Science is still a cottage industry with small teams of artisan crafting bespoke prototypes to their own standards [2]. Data science projects differ from most traditional data research projects because they could be complex as well as in need of advanced technologies in comparison to the traditional data analysis projects [1]. Data Science is rapidly changing the way that industries operate in many sectors, including the healthcare industry. Good number of academic and industrial research works have done in data science adaptation in healthcare, only a few have addressed the field of public health. Data Science projects are different from traditional data analytics project primarily due to the complex scientific approach. One of the main motivations of this research is to introduce this new approach, which addresses global public health challenges like mental health.

Mental health is vital to public well-being. Mental well-being has a great importance to a high quality of life and the ability to cope with the stress in day-to-day life. In fact, mental illness can precipitate physical diseases, social inequalities and harmful lifestyles. There is now a large number of evidence-based approaches that promote mental well-being and prevent mental illness. Psychological well-being is a relatively recent concept in public welfare and new types of mental health related diseases are identified daily. Mental well-being offers many health benefits as well as social benefits. Since balanced healthy lifestyles and equal social benefits are both supported by optimistic psychological performance. Social well-being, which includes mental well-being, is a valuable social goal in itself. Therefore, has become a clear and explicit goal of the government in many countries of the both developing and developed world.

II. PROBLEM STATEMENT

National and local government perspective, it is important that to have comprehensive knowledge of factors associated with mental health, especially among various geographical areas. Also, there is a social need to improve mental health among the population. Thus, this study may help healthcare professionals to reduce risk factors associated with mental health problems.

The objectives of the study are:

i. To understand the behavioural factors associated with mental health among various geographical areas using Association Rule Mining.
ii. Analyse and compare the behavioural factors associated with mental health within male and female populations.

III. RELATED WORK

Dmonte et al. [3] presented a new method of data mining approach for diagnose of anxiety disorder. This method has been proposed to be used to identify and predict anxiety disorder. Huang et al. [4] has proposed new method of discovering association rules from responded questionnaires for diagnosing geriatric depression. The proposed model has helped users self diagnose their physical fitness and mental health conditions with user-friendly interfaces devised on their smart phones. This system is deployed in the waiting rooms for patients to answer the diagnosis questionnaire before they are seen in outpatient services. Experimental results verified that the proposed system is suitable for analyzing the associations among users' responded questionnaires and their negative emotions.

Halonen et al. [5] has done a study on mental health by gender-specific occupational groups and they have observed different mental health profiles for the largest occupational groups for men and women. The particular findings suggest that gendered occupational status is an underlying factor explaining distinctive mental health profiles in the working

population. This finding was supported by analysis where gender category was the main of the examined predictors for those outcomes. Within the male population, there seem to be different behavioural patterns related to emotional health the type of in the lower vs. high-skilled careers [5].

The America's Health Rankings (AHR) Annual Report 2018 [6] shows that the obesity rate in US has increased compare to the previous years. Obesity has identified by the medical experts as a leading factor to cardiovascular disease, cancer, mental health and other medical conditions. Furthermore, the AHR report found the number of people reporting frequent mental distress (FMD) is also rising. Some states have six times the concentration of mental health providers than other states [6]. In addition, the suicide rate has increased 16 percent since 2012 and is much higher among males: 22.2 deaths per 100,000 compared with females at 6.2 deaths per 100,000.

Velten et al. [12] has conducted a research to find correlation between lifestyle choices and mental health and found out that all lifestyle factors were associated with mental health outcomes. Improved mental health was associated with high frequency of physical and mental activity, moderate consumption of alcohol, non-smoking, and body mass index and regular rhythm of life.

IV. METHODOLOGY

This section describes the methodological approach adopted in the project. 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]. This study has used exploratory case study approach as the research methodology. Exploratory study can be predominantly useful in understanding of the problem, explaining the nature of the challenge, or identifying problems that are usually included. It also allows researchers to develop suggestions and assumptions for further research and discover new insights or better understand the problem. In this research, the authors are make an effort to understand most significant behavioral factors that contribute to mental health issues in public health domain, espacially using Association Rule Mining. The outcomes of this research may support to the policy design in public health organisations.

A. Data Science

Data science is the process of discovering for an unknown correlations, associations, hidden patterns and relationships in a large dataset using intelligence methods such as data mining, machine learning, time series analysis or advanced statistical methods. In most causes data science uses statistical methods and mathematical algorithms to detect patterns. In data science, algorithms are the blueprint for solving data mining problems. Usually, algorithms are recursive in design, as a set of steps are repeated iteratively until a limiting condition is met.

B. Association Rule Mining

Association analysis is an unsupervised data mining technique which measures the strength of co-occurrence between transaction items in a large transactional database. The aim of this process is not to predict the occurrence of an item like in classification and regression techniques, but to discover strong frequently occurring patterns in the itemsets called association rules. It has been mainly applied for analysing customer's shopping behaviours. This association

rules help identifying items that are likely to be bought at the same instance [9]. Most popular algorithm for discovering association rules is the Apriori algorithm. Apriori algorithm is used for discovering association rule and finding frequent item set [8]. The Apriori algorithm firstly introduced by Agrawal, Imieliński and Swami to analyse market basket data in transactional database[9].

C. K-Means Clustering using Euclidean Distances.

Cluster analysis is an exploratory analysis that tries to identify structures within the data [11]. Cluster analysis is an unsupervised data mining strategy employed in the procedure of separating data objects predicated on similarities to the other data points and dissimilarities with data points grouped into various other clusters. It includes grouping a couple of objects into different clusters in a way that things in each cluster show similarities among themselves but are dissimilar to things in other clusters. Hopkins statistic is used to assess the clustering tendency. K-Means clustering algorithm has been used for data discretisation. In this study k=3 has been used.

V. EXPERIMENTAL EVALUATION

Recent evidence indicates that smoking, drinking habits, physical activities, eating habits and mental illness are associated [12]. The objective of this experiment is to determine the relationship between frequent mental distress (FMD) and human behavioral patterns in US states. FMD is a measure of perceived poor mental health and represents the percentage of the population experiencing persistent and likely severe mental health issues [13].

A strong and healthy mental state is vital to overall positive health and well-being among the people. Poor mental health has strongly related to both suicidal ideation and attempt. Every year, suicide is among the top 20 leading causes of death globally for all ages [14]. Although occasional short periods of mental distress may be unavoidable. Since 2016, frequent mental distress increased from 11.2 percent to 12.0 percent of adults in US [6].

A. Data

This research study has used last 6 years of AHR report [6]. AHR is the longest-running annual assessment of health on a state-by-state basis in US. The research dataset has been constructed by the authors using five key health measures reported in the AHR report [6].

1. Excessive Drinking - Percentage of adults who reported either binge drinking or chronic drinking.
2. Obesity - Percentage of adults with a body mass index of 30.0 or higher based on reported height and weight.
3. Physical Inactivity - Percentage of adults who reported doing no physical activity or exercise other than their regular job in the past 30 days.
4. Smoking - Percentage of adults who are smokers (reported smoking at least 100 cigarettes in their lifetime and currently smoke every or some days).
5. Frequent mental distress (FMD) - Percentage of adults who reported their mental health was not good 14 or more days in the past 30 days.

Table 1 shows the research dataset :

7

ATTRIBUTE	TYPE
STATE	Factor
EXCESSIVE_DRINKING	Numeric
OBESITY	Numeric
PHYSICAL_INACTIVITY	Numeric
SMOKING	Numeric
FREQ_MENTAL_DISTRESS	Numeric

B. Data Cleansing & Quality Assurance

Important measure for the achievement of the data science study is usually depend on the comprehensiveness and cleanliness of the research dataset. Usually original datasets are incomplete and noisy. Therefore, data cleansing process play vital part in data science projects. During the data cleansing process, work can be done upon missing values, treat outliers and provide resolution to inconsistent data. In special cases, machine learning algorithms or data mining methods can be used to handle missing data. 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.

C. Predictive modelling using Association Rules

i. Apriori algorithm

This research utilised the Apriori algorithm for association rules mining ([9]. This algorithm was selected as it allows for extraction of rules with a high accuracy. The Apriori algorithm functions by retrieving common item sets, the algorithm then assumes that subsets of each common item set must also be frequently occurring. An example of this is if item set {XY} is frequent then both items {X} and {Y} are assumed to be commonly distributed within the data and the association rules are generated around these items. Three important measures to assess the association rules are Support, Confidence and Lift. Support represents the probability of the set of items. Confidence measure indicates how often the generated rule is found to be true. Users can set a minimum support and confidence threshold and check if the rule satisfies both minimum thresholds. Lift describes the strength of association rule. The Apriori algorithm which is employed in this study is frequently used to mine for association rules among the five key health measures reported in the AHR dataset.

ii. Implementation

Dataset 1: Entire population

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 research k has been set to three. Table 2 illustrate the categorical value and the count after discretizing the continuous variables.

TABLE 2: BREAKDOWN OF THE CATEGORICAL VALUES

	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

First experiment is to find all the factors that contributes to high mental distress among both male and female.

The first set of rules were obtained for Mental_Distress=High in RHS with the support set to 0.05 and confidence set to 0.8.

In total, these settings generated 7 rules. Table 3 demonstrate the association rules generated by the Apriori algorithm. They have multiple lengths but support, confidence and Lift threshold are the main deciding factor to be considered.

TABLE 3: GENERATED ASSOCIATION RULES

	LHS	RHS	Sup.	Con	Lift	Count
1	Excessive_Drinking =Low Obesity=High Physical_Inactivity= High Smoking=High	Mental Distress = High	0.09	0.88	4.64	29
2	Excessive_Drinking =Low Obesity=High Physical_Inactivity= High	Mental Distress = High	0.10	0.89	4.67	31
3	Excessive_Drinking =Low Obesity=High Smoking=High	Mental Distress = High	0.10	0.83	4.40	30
4	Excessive_Drinking =Low Obesity=High	Mental Distress = High	0.10	0.84	4.44	32
5	Excessive_Drinking =Low Physical_Inactivity= High Smoking=High	Mental Distress = High	0.10	0.84	4.42	31
6	Excessive_Drinking =Low Physical_Inactivity= High	Mental Distress = High	0.11	0.80	4.25	33
7	Obesity=High Physical_Inactivity= High Smoking=High	Mental Distress = High	0.12	0.84	4.42	36

A network graph (Figure 1) can be utilized to illustrate the association rules generated by the Apriori algorithm and the relationship among the attributes.
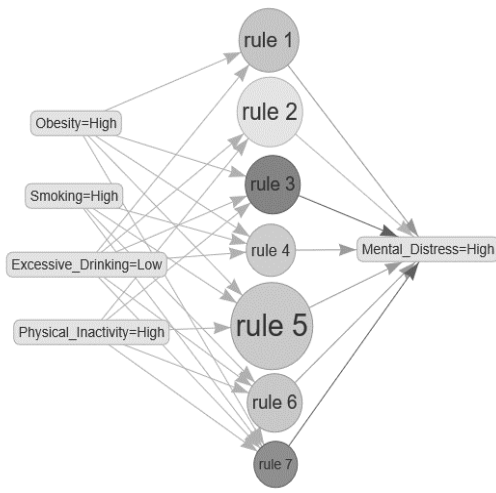
FIGURE 1: NETWORK GRAPH

The minimum threshold of the confidence measure has been reduced to identify uncommon rules. As Table 4 displays, the number of association rules and complexity of rules. According to the below figures, complexity increases when decreasing the confidence threshold.

TABLE 4: ASSOCIATION RULES WITH DIFFERENT CONFIDENCE LEVELS

CONFIDENCE	2 ITEMS	3 ITEMS	4 ITEMS	5 ITEMS
90%	0	0	0	0
85%	0	0	1	1
80%	0	2	4	1
75%	0	5	4	1
70%	0	6	4	1

Dataset 2: Male Only

The dataset consists of 306 data points each described by 6 attributes but only contained the data point form the male. Also, k-means clustering method has been used for the discretization and in this research k has been set to three. Table 5 shows the categorical value and the count after discretizing the continuous variables.

TABLE 5: BREAKDOWN OF THE CATEGORICAL VALUES

	Excessive Drinking	Obesity	Physical Inactivity	Smoking	Mental Distress
High	58	100	85	81	66
Medium	178	139	132	129	138
Low	70	67	89	96	102

This second experiment is trying to find the factors that contributes to the areas has the high mental distress only with male.

The first set of rules were obtained for Mental_Distress=High in RHS with the support set to 0.05 and confidence set to 0.75. Three rules have been generated with this settings.

Table 6 shows the association rules generated by the Apriori algorithm. They have multiple lengths but support, confidence and Lift threshold are the main deciding factor to be considered.

TABLE 6: GENERATED ASSOCIATION RULES

	LHS	RHS	Sup.	Con	Lift	Count
1	Excessive_Drinking =Low Obesity=High	Mental Distress = High	0.09	0.77	3.60	28
2	Excessive_Drinking =Low Obesity=High Smoking=High	Mental Distress = High	0.07	0.75	3.47	24
3	Excessive_Drinking =Low Obesity=High Physical_Inactivity= High	Mental Distress = High	0.07	0.76	3.55	23

A network graph (Figure 2) can be utilise to display the generated association rules and the relationship among the attributes.
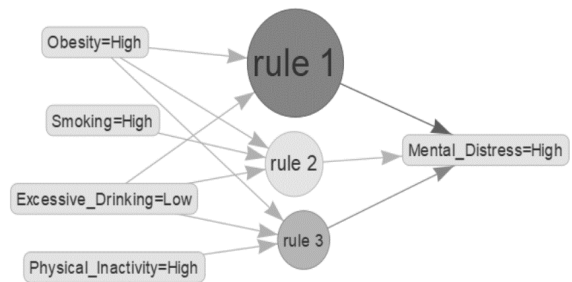

FIGURE 2: VISUALIZATION OF ASSOCIATION RULES ON NETWORK GRAPH

The minimum threshold of the confidence must be reduced to identify uncommon rules. As Table 7 illustrate, the number of association rules and complexity of the rules increases when decreasing the confidence threshold.

TABLE 7: ASSOCIATION RULES WITH DIFFERENT CONFIDENCE LEVELS

CONFIDENCE	2 ITEMS	3 ITEMS	4 ITEMS	5 ITEMS
90%	0	0	0	0
85%	0	0	0	0
80%	0	0	0	0
75%	0	1	2	0
70%	0	1	3	1

Dataset 3: Female Only

The dataset consists of 306 data points each described by 6 attributes but only contained the data point form the female. Table 8 shows the categorical value and the count after discretizing the continuous variables.

TABLE 8: BREAKDOWN OF THE ASSOCIATION RULES

	Excessive Drinking	Obesity	Physical Inactivity	Smoking	Mental Distress
High	59	78	71	59	67
Medium	163	125	124	128	133
Low	84	103	111	119	106

9

This third section of the experiment is to find the associated factors that contributes to the areas has the high mental distress only with female.

The first set of rules were obtained for Mental_Distress=High in RHS with the support set to 0.05 and confidence set to 0.8. Ten rules have been generated with this setting. Table 9 shows the association rules generated by the Apriori algorithm. They have multiple lengths but support, confidence and Lift threshold are the main deciding factor to be considered.

TABLE 9: GENERATED ASSOCIATION RULES

	LHS	RHS	Sup	Con	Lift	Coun
1	Physical_Inactivity= High Smoking=High	Mental Distress = High	0.12	0.88	4.02	37
2	Obesity=High Smoking=High	Mental Distress = High	0.11	0.85	3.89	35
3	Excessive_Drinking =Low Smoking=High	Mental Distress = High	0.10	0.83	3.82	31
4	Obesity=High Physical_Inactivity= High	Mental Distress = High	0.14	0.81	3.70	43
5	Excessive_Drinking =Low Physical_Inactivity= High	Mental Distress = High	0.12	0.80	3.67	37
6	Obesity=High Physical_Inactivity= High Smoking=High	Mental Distress = High	0.11	0.89	4.08	34
7	Excessive_Drinking =LowPhysical_Inact ivity=High Smoking=High	Mental Distress = High	0.09	0.88	4.02	30
8	Excessive_Drinking =Low Obesity=High Smoking=High	Mental Distress = High	0.08	0.87	3.97	27
9	Excessive_Drinking =Low Obesity=High Physical_Inactivity= High	Mental Distress = High	0.11	0.85	3.88	34
10	Excessive_Drinking =Low Obesity=High Physical_Inactivity= High Smoking=High	Mental Distress = High	0.08	0.90	4.11	27

In this instance also, the confidence threshold has been reduced to identify uncommon rules. As Table 10 shows, the number of rules and complexity of rules increases when decreasing the confidence threshold.

TABLE 10: ASSOCIATION RULES WITH DIFFERENT CONFIDENCE LEVELS

CONFIDENCE	2 ITEMS	3 ITEMS	4 ITEMS	5 ITEMS
90%	0	0	0	1
85%	0	2	4	1
80%	0	5	4	1
75%	0	6	4	1
70%	1	6	4	1

As each rule should be analysed and eventually confirmed by a healthcare expert.

VI. RESULTS AND DISCUSSION

Having described our approach and experiment, we now present our findings. For increased clarity, the revealed relations will be illustrated in the form of Association rules. Our experimental results suggest that, below two association rules (Table 11 and Table 12) are common in both male and female and shows high confidence level among other rules. According to the below association rules, we can argue that states which have low excessive drinking percentage, high level of obesity, high smoking percentage and people have low physical activities has the highest frequent of mental distress. Also, these rules have shown high confidence, Support and Lift thresholds among females rather than males.

TABLE 11: GENERATED ASSOCIATION RULES FROM MALE DATASET

	LHS	RHS	Sup.	Con	Lift	#
1	Excessive_Drinking=Low Obesity=High Smoking=High	Mental Distress = High	0.07	0.75	3.47	24
2	Excessive_Drinking=Low Obesity=High Physical_Inactivity=High	Mental Distress = High	0.07	0.76	3.55	23

TABLE 12: GENERATED ASSOCIATION RULES FROM FEMALE DATASET

	LHS	RHS	Sup.	Con	Lift	#
1	Excessive_Drinking=Low Obesity=High Smoking=High	Mental Distress = High	0.08	0.87	3.97	27
2	Excessive_Drinking=Low Obesity=High Physical_Inactivity=High	Mental Distress = High	0.11	0.85	3.88	34

Model validation has been done by applying rule number one mentioned in Table 11 and 12 to latest dataset from the 2018 AHR Report. Table 13 shows the results from the male dataset. It shows that 4 States out of 6 States have been correctly classified. So, accuracy of the association rule number one in male is 66%.

TABLE 13: VALIDATED RESULTS IN MALE

State	Excessive Drinking	Excessive Drinking Category	Obesity	Obesity Category	Physical Inactivity	Physical Inactivity Category	Smoking	Smoking Category	Mental Distress	Mental Distress Category
Alabama	19.1	Low	35.1	High	27.8	High	23.4	High	13.5	High
Arkansas	20.2	Low	32.8	High	29	High	22.8	High	15.4	High
Mississippi	19.1	Low	35.7	High	29.3	High	22.2	High	12.8	High
Oklahoma	18.9	Low	36.9	High	29.7	High	21.2	High	12.1	Medium
Tennessee	18.6	Low	33.3	High	28.5	High	24.3	High	10.5	Medium
West Virginia	17	Low	39.4	High	30.4	High	25.1	High	15.4	High

Table 14 shows the results from the female dataset. It shows that 5 States out of 5 States have been correctly classified. So, accuracy of the association rule number one in female is 100%.

TABLE 14: VALIDATED RESULTS IN FEMALE

State	Excessive Drinking	Excessive Drinking Category	Obesity	Obesity Category	Physical Inactivity	Physical Inactivity Category	Smoking	Smoking Category	Mental Distress	Mental Distress Category
Alabama	9.2	Low	37.4	High	35.8	High	18.7	High	17	High
Arkansas	11.6	Low	37.4	High	35.9	High	21.6	High	19.2	High
Mississippi	8.8	Low	38.8	High	36.6	High	22.1	High	18.7	High
Oklahoma	9.7	Low	36.1	High	34.8	High	19.2	High	18.9	High
West Virginia	7.8	Low	36.6	High	32.7	High	26.9	High	19.2	High

Mental health professionals may use this study findings to enhance awareness of mental health among citizens on a state level.

VII. CONCLUSION

Recent studies on mental health have shown that the strong correlation between lifestyle factors and mental health. Also, recent healthcare study found out that lifestyle factors were strongly associated with the mental health outcomes. The aim of this study is to examine the association between mental health and human behavioural factors like smoking, physical activities, eating and drinking habits among residents in city or state level using aggregated data. More specifically, the main aim of this study was to determine the relation between frequent mental distress (FMD) and human behavioural patterns among US population. This research study has used data from last 6 years of AHR report; separate analysis has been carried out on male and female datasets, then compared the behavioural factors associated with mental health with male and female. Study found that states which have low excessive drinking percentage and high obesity and high smoking percentage has the highest frequent of mental distress. Also, these rules have shown high confidence threshold among females rather than males. Furthermore, the study suggests that the association between excessive drinking, obesity, physical inactivity, smoking and frequent mental distress among residents in USA is consistent enough to assume concretely a plausible and significant association. Also, this new proposed data science based approach will help healthcare authorities to expand the degree of understanding of mental wellbeing within different geographical areas like cites or states. From this results and analysis, the study shows that there are interesting association rules which suggest some lifestyle behaviours to be responsible for mental health issues. These findings provides evidence to help local authorities in policy-making, sensitisation of the population, and providing targeted services to help reverse negative correlations.

VIII. FUTURE WORKS

Experimental results showed that the proposed method could identify significant causal relationships among behavioural factors and mental health. So, future research will focus on adding more behavioural factors like sleeping, walking, cycling and other social demographic factors so that better association rules can be found. Also, future works would assess current health policies and investigate how policies have played a role to improve or worsen the mental health in local authorities, and the possible socio-economic and socio-demographic factors that might be influencing the impact of health policies in the various local authorities.

REFERENCES

[1] 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

[2] Powell, H. (n.d.). The Industrialisation and Professionalisation of Data Science: 12 Questions. Retrieved April 10, 2019, from https://www.linkedin.com/pulse/industrialisation-professionalisation-data-science-12-harry-powell/

[3] Dmonte, S., & Panchal, D. (2015). Data mining approach for diagnose of anxiety disorder. International Conference on Computing, Communication & Automation. doi:10.1109/ccaa.2015.7148357

[4] Huang, Y., Huang, C., Chen, S., Liu, S., & Huang, H. (2012). Discovering association rules from responded questionnaire for diagnosing geriatric depression. 2012 ICME International Conference on Complex Medical Engineering (CME). doi:10.1109/iccme.2012.6275662

[5] Halonen, J. I., Koskinen, A., Varje, P., Kouvonen, A., Hakanen, J. J., & Väänänen, A. (2018). Mental health by gender-specific occupational groups: Profiles, risks and dominance of predictors. Journal of Affective Disorders, 238, 311-316. doi:10.1016/j.jad.2018.06.007

[6] 2018 America's Health Rankings Annual Report (Rep.). (2018). Retrieved April 10, 2019, from United Health Foundation website: https://assets.americashealthrankings.org/app/uploads/ahrannual-2018.pdf

[7] Saunders, M. N., Lewis, P., & Thornhill, A. (2019). Research methods for business students. New York: Pearson.

[8] Turban, E., Delen, D., & Sharda, R. (2018). Business intelligence, analytics, and data science: A managerial perspective. Harlow ; Munich: Pearson Prentice Hall.

[9] Olson, d. L. (2019). Descriptive data mining. S.l.: springer verlag, singapor.

[10] Saunders, M. N., Lewis, P., & Thornhill, A. (2019). Research methods for business students. New York: Pearson.

[11] Dubes, R., & Jain, A. (1980). Clustering Methodologies in Exploratory Data Analysis. Advances in Computers Advances in Computers Volume 19, 113-228. doi:10.1016/s0065-2458(08)60034-0

[12] Velten, J., Lavallee, K. L., Scholten, S., Meyer, A. H., Zhang, X., Schneider, S., & Margraf, J. (2016). Erratum to: Lifestyle choices and mental health: A representative population survey. BMC Psychology, 4(1). doi:10.1186/s40359-016-0106-7

[13] Rohrer, J. E., Borders, T. F., & Blanton, J. (2005). Rural residence is not a risk factor for frequent mental distress: A behavioral risk factor surveillance survey. BMC Public Health, 5(1). doi:10.1186/1471-2458-5-46

[14] Schwartz-Lifshitz, M., Zalsman, G., Giner, L., & Oquendo, M. A. (2012). Can We Really Prevent Suicide? Current Psychiatry Reports, 14(6), 624-633. doi:10.1007/s11920-012-0318-3

[15] Han, Jiawei, and Micheline Kamber. Data Mining: Concepts and Techniques. Elsevier, 2012.

[16] Sheng-yi Jiang, Xia Li, Qi Zheng and Lian-xi Wang, "An Approximate Equal Frequency Discretization method", WRI Global Congress Intelligent System, vol. 3, no. 4, pp. 514-518, 2009.