Extracting the Hidden Patterns Affecting Mental Health through Data Mining Techniques

Maryam Jahanbakhshi1, Asal Aghadavodian Jolfaee2, Roya Kelishadi3, Mohammad Sattari4

1. Health Information Technology Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
2. Dept. of Management and Health Information Technology, Isfahan University of Medical Sciences, Isfahan, Iran
3. Child Growth and Development Research Center, Research Institute for Primordial Prevention of Non-communicable Disease, Isfahan University of Medical Sciences, Isfahan, Iran

ABSTRACT

Background & Objective: This study was conducted to shed light on the hidden relationships, trends, and patterns of the teenagers’ mental health dataset based on data mining techniques.

Materials & Methods: The proposed method has four parts as follows: data preprocessing, data cleaning, target class selection, and extracting rules. The classes included inappropriate, moderate, and acceptable. The rules were extracted separately by implementing ID3, CHAID, and rule induction on the Caspian 5 dataset.

Results: It was found that the teenagers who rarely drink carbonated soda and have dinner seven days a week, have acceptable status of mental health. Besides, watching TV and playing computer games for 4 hours or more per week, drinking tea and packaged juices, eating cakes, cookies, pastries, biscuits, and chocolate weekly could lead to inappropriate status of mental health.

Conclusions: An attempt to improve health especially in youth is one of the important concerns of every country. The rules express the negative impact of soda on mental health. Besides, it can be concluded that there is a direct relationship between having breakfast and mental health.

Keywords: Mental Health, Data Mining, Hidden Pattern, Iterative Dichotomiser 3(ID3)

Introduction

Improving health in societies is one of the most important concerns of every country [1]. The World Health Organization (WHO) defines health as a set of physical, mental, and social well-being states [2]. It considers mental health as the main part of the human health. Mental health plays a central role in ensuring the dynamism and efficiency of the society [3].

Contrary to popular belief, children and adolescents are more vulnerable to mental health disorders than the other age groups [4]. Besides, many psychological disorders in adulthood are actually due to adolescence problems [5]. Half of all mental disorders begin at the age of 14 [6]. It can be stated that adolescence is one of the most valuable periods of an individual’s life [7]. If the adolescent moves through this period without any problems, he/she will become a healthy adult. Yet, if there are difficulties along the way and the adolescent’s mental health deviates from the main road, he/she will be misled [4].

Moreover, studies show that the majority of deaths among adolescents are due to suicide, conflict, and violence. These problems are more or less related to adolescents’ mental health [8]. The high prevalence of mental health problems in adolescents is estimated at 10% to 20% [9]. In Iran, 39% of students suffer from behavioral-emotional problems during adolescence [10].

In this regard, Caspian V studies have been conducted. Badihian studied the relationship between psychological distress and blood pressure in children and adolescents. The findings show that high psychological distress increases the chance of high blood pressure [11]. Riahi assessed the relationship between weight misperception and psychological distress. He represented that appropriate education intervention needs to be developed to improve the children and adolescents’ perception of their body weight status [12]. Zahedi represented that skipping breakfast was associated with depression, stress and psychological distress in all age groups [13]. Riahi...
reported that family dinner frequency has inverse relation with anxiety [14].

In the year 2011, Margreet et al., demonstrated that physical exercise in the leisure time is associated with a low prevalence of mental disorders, especially mood disorders and anxiety [15]. Rasouli et al., investigated Western, Mediterranean, and traditional food patterns using a questionnaire (FFQ), they also examined stress, depression, and anxiety using a questionnaire (DASS). The analysis of the ANOVA and ANCOVA tests indicates that adherence to the Mediterranean diet will reduce depression and stress, but the Western diet will increase them. However, no significant relationship was observed between dietary patterns and anxiety [16]. Ingram showed that changes in diet, sleep, and physical activity are associated with negative mood during covid-19 [17].

The process of increasing the volume of data requires data analysis and management. At a higher level, the discovery of knowledge represents the use of techniques such as data mining in the field of health more than before. As a result of using data mining on the students' mental health database, it is possible to discover the hidden relationships, trends, and patterns. Many studies have been conducted on the use of these techniques in mental health. In the year 2011, Biddle studied databases such as PubMed, Web of Science, and Medline. The search keywords are physical activities with at least one mental disorder such as depression, anxiety, and stress. These studies suggest that although a positive relationship between physical activity and mental health is evident in children and adolescents, most of the articles concerning this issue have poor design and limited effects. Moreover, a relationship has been found between screen viewing and poor mental health in children and adolescents [18]. Rahman demonstrated that data mining has numerous usages in mental health. He applied data mining techniques to the mental health dataset with 24 attributes and a binary class. The results indicated that family history also plays an important role in provoking mental health illness [19].

Mental health in this study is the total score that each person received from the health and illness part in the Caspian 5 study. This paper tries to identify nutrition, physical activity, and leisure time factors effects on mental health in 13 to 18 year-olds.

Materials and Methods

Romero [20] released the standard data mining process method. This study applied the standards for gathering and analyzing data.

2-1 Data Gathering

The dataset contains information of 14274 students aged 13 to 18 years registered in the Caspian V questionnaire available in the database of the Research Institute for Primary Prevention of Non-Communicable Diseases in Isfahan University of Medical Sciences. This study uses a purposeful sampling method, i.e, the following two criteria are met in the sample:

1. Only the information about students in the age group of 13 to 18 years is employed. Regarding the studies and the World Health Organization view, a majority of the symptoms of mental disorders begin at the age of 13, namely the beginning of adolescence. It is of note that the age range of 13 to 18 years is classified as adolescence.

2. The information about students living in the city is examined. Because during voice consultations with health professionals, diversity in urban lifestyle affects individuals' habits and mental health. Thus, the sample size is equal to 5283 students.

The data gathering method is extracting data of the Caspian V datasets which is available in the database of the Research Institute for Primary Prevention of Non-Communicable Diseases in Isfahan University of Medical Sciences. Based on the international model of student health care system, the Caspian study is a step-by-step approach of the World Health Organization and based on the GSHS questionnaire in five periods (2003-2005). These studies include two questionnaires of students and parents that differ only in some details.

The proposed method has four parts. The parts include data pre-processing, data cleaning, target class selection, and extracting rules.

Data Preprocessing

Data need to be preprocessed before applying classification techniques. In the data preprocessing step, the data need to be cleaned, the target class should be specified, and a subset of related attributes needs to be selected. Data related to demographic information, nutrition, physical activity, leisure activities, health, and illness, in the format of an SPSS file, are extracted from the database of the Research Institute for Primary Prevention of Non-Communicable Diseases in Isfahan University of Medical Sciences. Then, by summarizing the scores of data related to health and illness, mental health of each person is classified into one of the three groups including acceptable mental health, moderate mental health, and inappropriate mental health.

Data Cleaning

In this step, the missing values are detected. If the number of missing attributes of the record is high, the record is deleted [21]. In this data set, 1167 records are removed. In remaining records, the number of missing attributes is not high, so the most frequent values are used for missing attributes.

Target Class Selection

According to the coding of health and illness questions included feelings of worthlessness, irritability or bad temper, feelings of anxiety, insomnia, feelings of confusion and dizziness, feelings of sadness, feelings of
worry, health status from one’s perspective and life status from one’s perspective. According to the standard questionnaire, the sum of the scores of the questionnaire questions was calculated for each person and 33% quotas were calculated to classify people into three classes of inappropriate, moderate, and acceptable. Finally, mental health characteristics were determined as the target class.

Mental health has been selected as the goal for which three acceptable, moderate, and inappropriate are considered as a class. Therefore, classes in this study consist of acceptable mental health, moderate mental health, and inappropriate mental health.

Extracting Rules
A total of 23 attributes are considered as characteristics that can affect mental health. Data from 31 provinces were extracted from the main dataset. Then, for each province, a separate train data set, which includes 70% of that province's data set, was extracted and presented as input to the Rapid Miner software. For each province, rules were extracted separately by implementing three data mining methods. ID3 is a classifier based on decision-tree algorithms, where it selects the attribute with the highest information load [22]. Chaid [23] is the other method based on decision tree. Rule induction [24] tried to extract rules from the dataset. The rules whose confidence was less than 50 percent were removed. We extracted 5 rules for each province, thus, we had 155 rules. Among these rules, 98 rules had confidence of less than 50 percent. Among the 57 remaining rules, the top 10 rules were extracted.

Results
Three measures were used for evaluation. Accuracy was considered as the ratio of true values compared to all values. Sensitivity specified how many positive results were returned, and specificity indicated how many negative results were returned.

A total of 18 characteristics were investigated. These characteristics and their set of values are given in Table 1. Table 1 presents the properties by using a range of 4 to 5 for their set of values. Sometimes the set of values of these characteristics was the same, for example, the meals consisted of three items: breakfast, lunch, and dinner. Consumption of different foods such as dairy products, vegetables, sweets, soft drinks was in the range of daily, weekly, rarely, and never.

Table 1. Characteristics and values of the dataset

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of breakfast per week</td>
<td>Never, One day per week, Two days per week, Three days per week, Four days per week, Five days per week, Six days per week, Seven days per week</td>
</tr>
<tr>
<td>Number of lunch per week</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Number of dinner per week</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Cakes, cookies, pastries, biscuits, chocolate, packaged juices</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Puffs, chips, pickles</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Carbonated soda, diet soda, non-alcoholic beer</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Fresh and dried fruits and natural juices</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Fresh and cooked vegetables</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Dairy</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Sausages, pizzas and burgers</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Tea and coffee</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Sugar with Tea</td>
<td>Daily, Weekly, Rarely, Never</td>
</tr>
<tr>
<td>Exercise rate at school per week</td>
<td>daily, weekly, rarely, never</td>
</tr>
<tr>
<td>Means</td>
<td>School shuttle service</td>
</tr>
</tbody>
</table>
Characteristics | Values
--- | ---
used to commute to and from school | Bus, Family car, Walking, Bike
None, A little, Medium, Usually, Most of the time

Total physical activity during the week

The amount of TV watching per week

The amount of computer work per week

The amount of sleep per day

Table 2. Demographic information of respondents to the questionnaire

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender (%)</th>
<th>Grade(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>On Average</td>
<td>female</td>
<td>male</td>
</tr>
<tr>
<td>15</td>
<td>2206 (49%)</td>
<td>2210 (51%)</td>
</tr>
</tbody>
</table>

35 percent of the respondents approximately have low mood in mental health. 33% and 32% of the respondents have moderate and acceptable moods respectively.

Table 3. Accuracy of ID3, Chaid, and Rule induction

<table>
<thead>
<tr>
<th>Level of respondents' mental health</th>
<th>ID3</th>
<th>Chaid</th>
<th>Rule induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>55%</td>
<td>68%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Based on Table 3, CHAID method predicts more accurately than Rule induction and ID3. Based on Table 4, Chaid and Rule induction are more accurate in Moderate than other moods, but Id3 predicts more accurately in Inappropriate mood than other moods.

Table 4. Sensitivity and Specificity of ID3, Chaid, and Rule induction

<table>
<thead>
<tr>
<th></th>
<th>ID3</th>
<th>CHAID</th>
<th>Rule induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>Inappropriate</td>
<td>65%</td>
<td>63%</td>
<td>55%</td>
</tr>
<tr>
<td>Moderate</td>
<td>55%</td>
<td>66%</td>
<td>86%</td>
</tr>
</tbody>
</table>
Based on Table 5, if we drink carbonated soda, diet soda, and non-alcoholic beer rarely and have dinner seven days a week, the status of our mental health will be acceptable. Moreover, if we watch TV and play computer games 4 hours or more in a week the status of mental health will be inappropriate. If we have breakfast seven days a week and exercise at school 2 hours a week, the status of our mental health will be acceptable. If we drink tea weekly and have cakes, cookies, pastries, biscuits, chocolate, and packaged juices weekly, the status of our mental health will be inappropriate. Other rules are presented in Table 5.

### Table 5. The extracted rules

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Condition</th>
<th>Result</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule #1</td>
<td>Drinking Carbonated soda, diet soda, non-alcoholic beer rarely and having dinner seven days in a week</td>
<td>Acceptable Mental Health</td>
<td>0.82</td>
</tr>
<tr>
<td>Rule #2</td>
<td>Having breakfast seven days a week and exercising at school 2 hours a week</td>
<td>Acceptable Mental Health</td>
<td>0.77</td>
</tr>
<tr>
<td>Rule #3</td>
<td>Drinking tea weekly and having cakes, cookies, pastries, biscuits, chocolate, packaged juices weekly</td>
<td>Inappropriate Mental Health</td>
<td>0.74</td>
</tr>
<tr>
<td>Rule #4</td>
<td>Watching TV 4 hours or more in a week and playing computer games 4 hours or more in a week</td>
<td>Inappropriate Mental Health</td>
<td>0.73</td>
</tr>
<tr>
<td>Rule #5</td>
<td>Eating vegetables weekly and having fresh and dried fruits and natural juices weekly</td>
<td>Acceptable Mental Health</td>
<td>0.72</td>
</tr>
<tr>
<td>Rule #6</td>
<td>doing total physical activity during the week usually and watching TV 2 hours in a week</td>
<td>Acceptable Mental Health</td>
<td>0.69</td>
</tr>
<tr>
<td>Rule #7</td>
<td>having breakfast seven day a week and drinking tea and coffee weekly</td>
<td>Acceptable Mental Health</td>
<td>0.68</td>
</tr>
<tr>
<td>Rule #8</td>
<td>Drinking tea and coffee weekly and drinking tea with sugar weekly</td>
<td>Inappropriate Mental Health</td>
<td>0.68</td>
</tr>
<tr>
<td>Rule #9</td>
<td>Drinking Carbonated soda, diet soda, non-alcoholic beer weekly and having cakes, cookies, pastries, biscuits, chocolate, packaged juices weekly</td>
<td>Inappropriate Mental Health</td>
<td>0.65</td>
</tr>
<tr>
<td>Rule #10</td>
<td>Eating puffs, chips, pickles weekly and eating Sausages, pizzas and burgers weekly</td>
<td>Acceptable Mental Health</td>
<td>0.62</td>
</tr>
</tbody>
</table>

### Discussion

The best confidence rule indicates that if we drink carbonated soda, diet soda, and non-alcoholic beer rarely and have dinner seven days a week, the status of our mental health will be acceptable. Shi [25] examined the relationship between mental health and soda and found that soda would hurt mental health. The relationship between tea and mental health depends on the type of tea. Psychologists emphasize that having dinner with family members is acceptable for mental health.

The second best confidence rule suggests that if we have breakfast seven days a week and exercise at school 2 hours a week, the status of the mental health will be acceptable. The study [26] shows that there is a direct relationship between the breakfast meal and mental health. People who have breakfast will have a better mental health status than people who do not have breakfast. Elahi [27] examined the relationship between the physical fitness of the employees of a military university and mental health. The results show that physical activity has a significant impact on the mental health of employees.

The third best confidence rule demonstrates that if we drink tea and packaged juices weekly and eat cakes, cookies, pastries, biscuits, and chocolate weekly, the status of our mental health will be inappropriate. Shimbo [28] examined the relationship between green tea and mental health, where they found that green tea will have a positive effect on mental health.
The first three rules address the role of nutrition in mental health. Studies in Korea show that nutrition can affect mental health in such a way that some foods can have a devastating effect on mental health and even cause depression [28].

The fourth best confidence rule indicates that if we watch TV for 4 hours or more in a week and play computer games 4 hours or more in a week, our mental health status will be inappropriate. Mahmodi [29] reported that watching TV hurts mental health. Shobeiri [30] indicated that students who did not play computer games were in much better condition in terms of mental health in all dimensions (agression, depression, anxiety, interpersonal sensitivity, and physical symptoms) than students who played more or less.

The fifth best confidence rule shows that if we eat vegetables weekly and have fresh and dried fruits and natural juices weekly, the status of our mental health will be acceptable. Fallahzadeh demonstrated that those who consumed less fruit and vegetables had a higher degree of depression [31]. CHAID predicted more accurately than ID3 and rule induction. It can be concluded that CHAID has good performance on the nominal dataset.

**Conclusion**

An attempt for improving health in societies is one of the most important concerns of every country. The World Health Organization defines health as a set of complete physical, mental, and social well-being states. Dataset contains information of students aged 13 to 18 years registered in the Caspian V questionnaire. This questionnaire is available in the database of the Research Institute for Primary Prevention of Non-Communicable Diseases in Isfahan University of Medical Sciences. The dataset includes 14274 people. A total of 23 attributes are considered as characteristics that can affect mental health. Data from 31 provinces are extracted from the dataset. Some people have dinner with rarely carbonated soda, diet soda, and non-alcoholic beer. Rules express that if we do not have dinner with soda, the status of our mental health will be good. Having tea with cake is common worldwide, which is inappropriate for mental health. Moreover, having fruit and vegetables can promote mental health.

**Compliance with Ethical Standards**

**Funding**

This study was funded by Isfahan University of Medical Sciences Funding.

**Responsibilities of authors**

All authors accept responsibility for conducting the study and for the analysis and interpretation of the data. They helped write the manuscript and agree with the decisions concerning it. They meet the definition of an author as stated by the International Committee of Medical Journal Editors. They have seen and approved the final manuscript. Authors also certify that neither the article nor any essential part of it, including tables and figures, will be published or submitted elsewhere before appearing in the Journal.

**Ethical approval**

This article does not contain any studies with human participants or animals performed by any of the authors.

**Acknowledgments**

None.

**Conflict of Interest**

Authors declare no conflict of interest.

**References**


19. Rahman SS. An application of data mining of mental health data, Association for Information Systems AIS Electronic Library (AISeL), SAIS proceeding, 2019


How to Cite This Article:


Download citation:
BibTeX | RIS | EndNote | Medlars | ProCite | Reference Manager | RefWorks

Send citation to:
Mendeley | Zotero | RefWorks