An Application of Data Mining Algorithms for Analyzing Psychological Researches

Manar.A.Ali, Tareef K. Mustafa

Department of Computer Since, Collage of Since, University of Baghdad, Baghdad, Iraq

Abstract

Computer science has evolved to become the basis for evolution and entered into all areas of life where the use of computer has been developed in all scientific, military, commercial and health institutions. In addition, it has been applied in residential and industrial projects due to the high capacity and ability to achieve goals in a shorter time and less effort. In this research, the computer, its branches, and algorithms will be invested in the psychological field. In general, in psychological fields, a questionnaire model is created according to the requirements of the research topic. The model contains many questions that are answered by the individuals of the sample space chosen by the researcher. Often, these questions are long and tedious, and some questions are repeated and not useful. In this research, we will work on summarizing these forms without effecting their research results; eventually, these steps in the proposed approach will produce better forms. However, it leads to building a method using the Apriori algorithm and Association rule mining which, consequently, applies to each questionnaire from which unnecessary questions are to be removed.

Keywords: data mining, Apriori algorithm, association rule mining, psychological.

تطبيق خوارزميات التنقيب عن البيانات لتحليل البحوث النفسية

طريف كامل مصطفى، منار عباس علي

قسم علوم حاسبات، كلية العلوم، جامعه بغداد، بغداد، العراق

الخلاصة

الكمبيوتر هو العصب الرئيسي للتكنولوجيا في العصر الحالي لا يمكن الاستغناء عنه في جميع مجالات الحياة بأي شكل من الأشكال ، حيث أن الكمبيوتر موجود في جميع المؤسسات العلمية والصحية والعسكرية وجميع أنواع المؤسسات التجارية والمشاريع السكنية والصناعية ، بالاضافة لذلك استخدم الكمبيوتر لتسهيل هذه الإجراءات بتقليل الوقت والجهد المطلوب لتحقيقه .

في هذا البحث سيتم استثمار الكمبيوتر وفروعه وخوارزمياته في المجال النفسي ، بشكل عام في البحوث النفسية تصمم استبيانة البحث وفقا لمتطلبات موضوع البحث. تحتوي الاستبيانة على العديد من الأسئلة التي يجيب عليها أفراد عينة البحث التي اختارها الباحث. غالبًا ما تكون هذه الأسئلة طويلة ومملة ، وبعض الأسئلة تتكرر وليست مفيدة. هذا البحث ، يعمل على تلخيص هذه الاستبيانة دون التأثير على نتائج أبحاثهم ، وفي النهاية ، ستؤدي هذه الخطوات في النهج المقترح إلى نموذج استبانة أفضل. فإنه يؤدي إلى بناء طريقة

_____

*Email: manarameer22@gmail.com

باستخدام خوارزمية Apriori ووخوارزمية Association Rule Mining وبالتالي ، هذا المنهج المقترح يعمل
على إزالة الأسئلة الغير ضرورية منه.

1. Introduction

This research is an essential topic of study because researchers in the field of psychology usually use a questionnaire, which is often lengthy and has precise questions that require specialists to develop appropriate questions in accordance to the nature of the research to be finished. Thus, it requires extra time and effort. Psychologists are specialized in their field and, hence, they are unaware of the tools and algorithms of data mining that can help them in research [1].

Data mining is utilized to unravel the variations of cognitive issues, such as Trade Insights, Limitation Programming, and Data Recovery. The data mining procedure is utilized in data recovery point at finding information from a collection of records concurring to a users' inquiry. For instance, the classification, clustering, frequent item set/association-rule mining classifies new records, i.e. segments records, into comparative bunches and finds many terms from the collection of reports, separately [2].

In this method, an approach is created for consecutive steps, beginning with the utilization of the Apriori algorithm and then the Association rule mining, individually. The work of this approach aims to minimize the questionnaire and predict the associated and similar elements and delete them to make the questionnaire more excellent, flexible and straightforward as feasible. It also aids in avoiding the prolonged and repeated questions, along with the opportunity of making use of this approach to any kind of questionnaires in the future. Summarizing the research form using the proposed data mining approach will not affect the physiological results and findings.

Research specifications are appointed to describe the proposed strategy vicinity alongside with its obstacles, due to the fact that they are both phase and parcel from the main design. They are summarized as follows:

• An academic self-evaluation questionnaire was used by a researcher in the field of psychology at the University of Baghdad in 2017 in his research entitled "Conflict Types of Personality and Academic Self". One of the aims of this research was the academic self-knowledge of distinguished students through the application of the statistical tool (T-test for one sample) in the research sample. The results of the research showed that the students enjoyed a high academic level [3].
• WEKA 3.8.4 (2019) was used for its excellent performance in such data mining implementation analysis.
• The proposed approach uses 2 data mining algorithms which are Association rule mining and Apriori algorithm to offer a final improved decision support result.
• Visual Studio-2015 which contains C# was used for practical part applications.
• SPSS program was used to perform the T-test for one sample.

2. Literature review

In the literature, the topic has been extensively investigated in an attempt to take advantage of data mining algorithms in the psychological field.

Burman et al. [4] used the Association Rule Mining Strategy to evoke common sense data from the fundamental information collected. The work investigated the relationship between psychological factors and student's intellectual performance, with the utilization of the Apriori algorithm. It aimed at foreseeing and progressing the execution of instructional exercise of understudies through reshaping their psychological parameters.

Boyko et al. [5] suggested a method to inspect the behaviour of groups of people and show how to predict a person's place for the following month. The association rule mining

algorithms was used in this paper. The trouble of growing trade union's rules was additionally examined with the use of Apriori scalable algorithms to discover subtle rules. For analysis, they used a useful tool which is the standard mlxtend (machine learning extensions) library in python to collect statistics via block, user login, and time. In this article, they determined the Apriori and k-means algorithms for a sample person conduct analysis. In the process, they studied the problem of finding union rules and were capable of finding and describing patterns in massive information sets. The used Apriori scalable algorithms to find great rules.

Qin et al. [6] focused on the mental and psychological health of university and school students to strengthen high-quality education. For successful arrangement and decision-making, mental fitness instructions make a difference of ideas in deciding the right person to see in lifestyles and values, with remarkable importance paid to them. With the rapid improvement of science and innovation, mining has contributed to the development of students' mental health—the work focused on the application of mining information in mental health instructions for students. Through the intelligent software to extract data from the information, cutting-edge models of educating intellectual well-being for college learn about durations have been growing away from home that took excellent strides to grant a reference for teaching mental health in the future.

Morales et al. [7] aimed in their work at analyzing the relationship between students' study and their psychological state. This is made through using the association rules. These rules are introduced in an inter-active visible way, which allows the assessment to select those of interest. For this study, students of four careers in technologies from a university in southern Mexico had been considered. The aim of the study is to find a relation between the factor of student participation in the semester with other variables and their impact on students' decision to continue studying. For this purpose, Utretch Work Engagement Scale for Students instrument (UWES-S) was utilized. This instrument identifies the psychological association with the activities made by the students. After collecting the data using questionnaire which the students answered. The (Apriori) algorithm was applied on the data, which was collected from the students. The resulting association rules indicate that the dedication variable is the most prominent in this respect.

Kang et al. [8]. proposed the concept of massive information on security psychology (BDSP) and aimed at demonstrating the challenges of applying big data in security brain research. Their paper puts forward the concept of BDSP and analyzes the distinction between BDSP and general test information. The paper summarizes the classification standard and essential characteristic of BDSP, investigates the framework of BDSP, and builds a three-dimensional structure of BDSP. In other words, the paper deals with the challenges of utilizing BDSP. This consideration is of great help to security specialists to unravel mental issues within the security space. Such kind of studies focus on identifying and expecting patterns of human security behavior and the probability of committing crimes. This helps to support security specialists in their work and improves maintaining the security and safety in the countries.

Long et al. [9] state that the psychological troubles of college university students have aroused general concerns. All sorts of psychological health issues plague many college students. Psychological health problems caused many severe effects on students. Psychological appraisal measurements and the fundamental data assembled from 6500 amateurs are utilized to analyze alliance rules and characteristics of college students' mental variables. The side effect self-rating scale (SCL-90) was once compiled by utilizing L. R. Derogates in 1975, which contains 90 elements or coefficient.

The SCL-90 has been utilized in a wide assortment of psychiatric side effects. The SCL-90 incorporates ten variables, such as somatization, obsessive-compulsive side effects, interpersonal affectability, discouragement, uneasiness, antagonistic vibe, dread, suspicion,

psychosis, and other components. The PNARC model is brought in this paper to identify the factors effecting the creation of the student's psychology through data collected through SCL-90 in Chinese universities.

The back certainty gadget and the relationship check approach are procured to erase the negate affiliation policies and get terrific and frail association rules for successfully inspecting the beneficial relationship of SCL-90 variables. Association rule mining that built significant elements in forming psychological health for students was initially undiscovered and its impact is unclear.

Mustafa, T. K. [10] contends that stylometric creation attribution is one of the modern procedures within the textual substance mining and these days is the focus of numerous analysts and institutes because of its delicateness, especially for museums and authorized grand libraries which are very worried about the value of their books and literature assets educationally and financially. This worthiness, of course, depends on the certification and the authentication of these valuable books. This strategy is included in the analysis of writings, for example books and plays written by well-known creators by selecting some features that show creators in authorship, expecting that these authors have a certain method of authorship, that no other writer possesses; To achieve this, this paper examines some of the accounts that are used regularly to predict the identity of the author through his writings.

3. Method Plan and Implementation
3.1 Data Mining Algorithm

Aggarwal has presented association rule mining which is one of the preeminent basic information mining strategies. It points to find relative connections from the commercial exchange data set. In this way, association rule mining can offer assistance within the trade choice-making handle by letting the choice makers understand clients buying propensities. The rules are within the frame of $A \rightarrow B$, where A and B are thing sets and $A \cap B = \emptyset$. On the off chance that A happens, B is most conceivable to happen. The standard Association, rule mining calculations, return those rules that meet the edge values set for bolster and certainty measures which are the measures of interestingness. The least back and least certainty signify edge of back and certainty. These parameters are set by the researchers or those responsible for the study.

The back and certainty of an association run the show ($A \rightarrow B$) are characterized as follows:

$$Sup\ (A \rightarrow B) = \frac{\text{N of T contain } A\ \&\ B}{\text{N of T in the database}}$$

$$Conf\ (A \rightarrow B) = \frac{\text{N of T contain } A\ \&\ B}{\text{N of T contain A}}$$

The Apriori algorithm is considered one of the most famous mining algorithm, which was planned by Agrawal et al. in 1993. This algorithm can, by and large, be considered as a preparation having two steps. The primary step is to discover all frequent thing set. The second step is to produce robust association rules by the utilizing many things sets. The breaking point to inquire about is to deliver a transparent and reliable approach that can be utilized through mental analysts and assist them through clear steps to reach their points without losing time or endeavors [11]. The method steps suggested in this paper are as follows:

```
        ┌─────────────┐
        │    start    │
        └─────────────┘
               │
               ▼
        ┌─────────────────┐
        │  Coding dataset │
        └─────────────────┘
               │
               ▼
        ┌─────────────────┐
        │ Apriori algorithm│
        └─────────────────┘
               │
               ▼
        ┌───────────────────────┐
        │ Association rule mining│
        └───────────────────────┘
               │
               ▼
        ┌─────────────┐
        │    T-Test   │
        └─────────────┘
               │
               ▼
        ┌─────────────┐
        │     end     │
        └─────────────┘
```
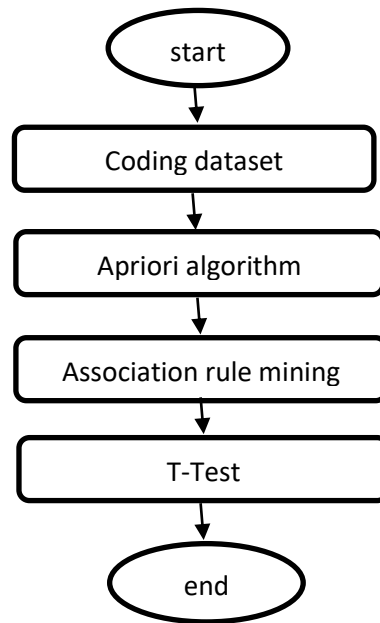
3.2   Research Tools:

3.2.1 Data set:

Using a self-academic scale, general psychological questionnaire form was designed and distributed on 350 students in distinguished high schools in Baghdad. The data has been gathered, and there were 270 correct forms entered into an excel sheet.

The academic self-measure was used by one of the researchers in the field of psychology at the University of Baghdad in 2017 to achieve the goal of academic self-knowledge of students and was applied by utilizing the statistical tool (T-test for one sample) to the research sample. A typical psychological questionnaire form was designed and dispensed on 270 students in 6 high schools in Baghdad to be filled with the aid of the students themselves, in addition to the assistance of the school management [4].

3.2.2 WEKA (The Waikato Environment for Knowledge Analysis):

Weka is an aggregation of machine learning algorithms for data mining operations. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The system is very reliable and robust with built-in features. WEKA facilitates the process of comparing various solution strategies depending on the same evaluation method and can determine the one that is most appropriate for the problem at hand. It diminishes the level of perplexity included in getting real-world information into an assortment of machine learning plans and assessing the yield of those plans. It has ensured flexibility for machine learning investigation in an instructive environment [12].

3.2.3   Language C# (C Sharp)

C # has been assembled to be suitable for programming purposes for both hosted and embedded systems, and from giant systems that use advanced frameworks to the simplest of small systems [13]. In this paper, used the WEKA library in C# language for the development of the program.

3.2.4. SPSS (Statistical Package for the Social Sciences).

SPSS was considered the most popular statistical program in the educational and commercial fields, which occupies the first place in terms of use. In addition, the SPSS package is a flexible package that allows for many types of statistical, analysis, mathematical

and economic operations [14]. In this paper, needed this software to behavior the T-test feature for a single sample to recognize whether or not splendid students have a high academic level or not.

3.3 Implementation Steps:
Step 1:
    Before applying a data mining algorithm, it is usually indispensable to perform some pre-processing tasks, which permit reworking the original data to a more appropriate form to be used by way of the specific algorithm.
1.   Coding the raw data collected from the school's questionnaire mentioned earlier in the dataset section [1], As shown in Table 1.
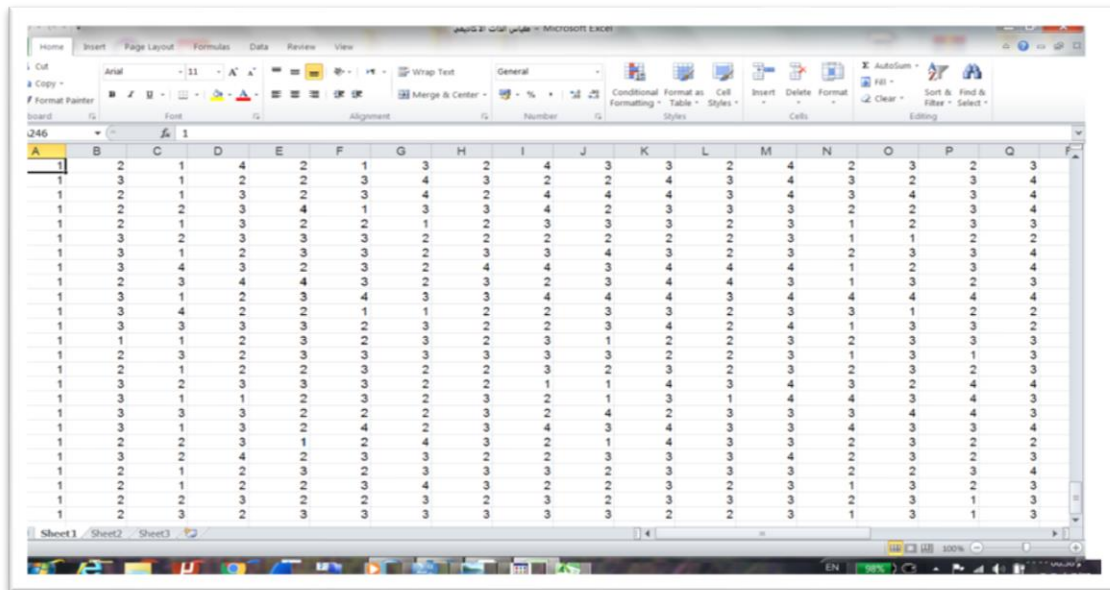
Figure 1-The dataset before the coding process.

Table -1-S.A = Strongly Agree, A= Agree, D= Disagree, S. D= Strongly Disagree

code	Attribute description	Instance description	symbol
A	Sex	Male	A1
		female	A2
B	friends are smarter than he	S. A	B1
		A	B2
		D	B3
		S. D	B4
C	have thoughts of leaving school	S. A	C1
		A	C2
		D	C3
		S. D	C4
D	suffer from distraction during lessons	S. A	D1
		A	D2
		D	D3
		S. D	D4
E	do homework faster than colleagues?	S. A	E1
		A	E2
		D	E3
		S. D	E4
F	can reach a high level by studying whatever the circumstances	S. A	F1
		A	F2
		D	F3
		S. D	F4
G	Read the article once and understand its meaning	S. A	G1

		A	G2
		D	G3
		S. D	G4
H	When encountering a study problem, find a solution quickly	S. A	H1
		A	H2
		D	H3
		S. D	H4
I	can create new ideas related to the subject of the study	S. A	I1
		A	I2
		D	I3
		S. D	I4
J	Spontaneously express unfamiliar thoughts	S. A	J1
		A	J2
		D	J3
		S. D	J4
K	Compare the different points of view and choose what matches the opinion	S. A	K1
		A	K2
		D	K3
		S. D	K4
L	am seeking scientific facts that no one previously found	S. A	L1
		A	L2
		D	L3
		S. D	L4
M	the feel could excel in all lessons	S. A	M1
		A	M2
		D	M3
		S. D	M4
N	like the extra duties that am assigned to	S. A	N1
		A	N2
		D	N3
		S. D	N4
O	find it easy to learn and understand what is new in the curriculum	S. A	O1
		A	O2
		D	O3
		S. D	O4
P	am ready for any sudden examination	S. A	P1
		A	P2
		D	P3
		S. D	P4
Q	Do love challenging items	S. A	Q1
		A	Q2
		D	Q3
		S. D	Q4
R	Do stand out from colleagues in the ability to relate ideas	S. A	R1
		A	R2
		D	R3
		S. D	R4
S	work to be highly distinguished from colleagues	S. A	S1
		A	S2
		D	S3
		S. D	S4
T	can find new alternatives to solve any problem	S. A	T1
		A	T2
		D	T3
		S. D	T4
U	have a problem understanding math-containing subjects	S. A	U1
		A	U2
		D	U3
		S. D	U4
V	teachers describe as a distinguished student	S. A	V1

|		A	V2
		D	V3
		S. D	V4
W	prepare scientifically to be innovative in the future	S. A	W1
		A	W2
		D	W3
		S. D	W4
X	can find new ways to accomplish routine duties	S. A	X1
		A	X2
		D	X3
		S. D	X4
Y	have reduced the use of vocabulary to express thoughts	S. A	Y1
		A	Y2
		D	Y3
		S. D	Y4
Z	Fellow students describe as a genius	S. A	Z1
		A	Z2
		D	Z3
		S. D	Z4
AA	feel that level of thinking is higher than the level of the subjects	S. A	AA1
		A	AA2
		D	AA3
		S. D	AA4
AB	feel that the curriculum restricts renewed ideas	S. A	AB1
		A	AB2
		D	AB3
		S. D	AB4
AC	love exams because it reveals the level of understanding of the subjects Feel joy when doing homework in class	S. A	AC1
		A	AC2
		D	AC3
		S. D	AC4
AD	try to help classmates with their scientific subjects	S. A	AD1
		A	AD2
		D	AD3
		S. D	AD4
AE	Feel the joy when doing homework in class	S. A	AE1
		A	AE2
		D	AE3
		S. D	AE4
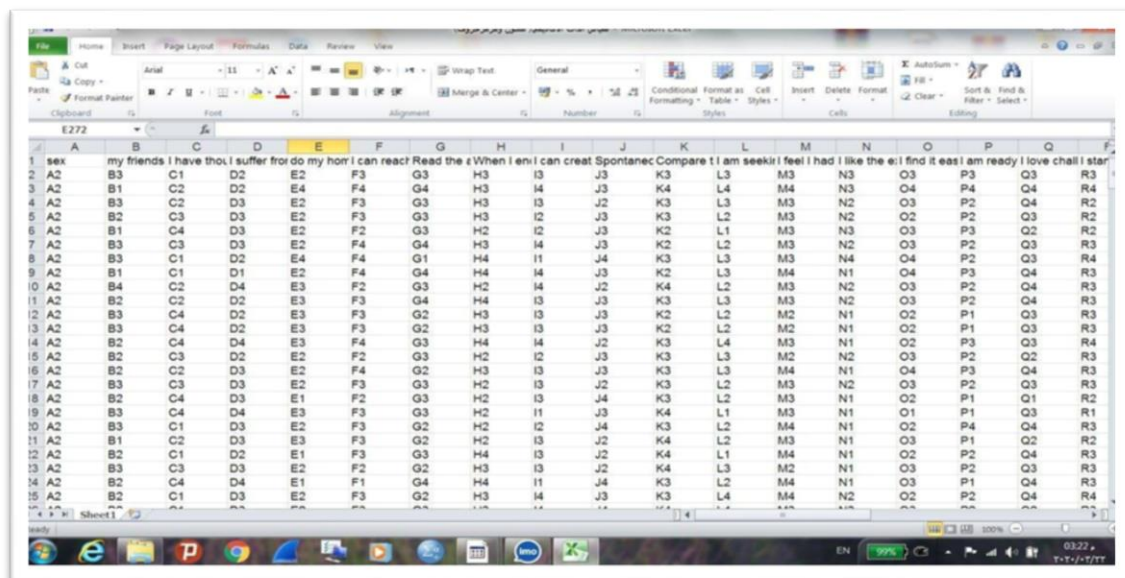
Figure 2-The result is post-coding

2. The dataset file format has been converted to a format acceptable to WEKA, which is the ARFF (attribute file format arrangement). The record is an ASCII content record that depicts a list of occurrences sharing a set of traits utilized with the Weka machine learning computer program. ARFF records were created by the Machine Learning Extend at the Department of Computer Science of the College of Waikato.

3. The Apriori function is executed on the data to extract the standard set and the superset.

4. Selecting the superset found on the last pruning level, which is the 4th.

5. Selecting the most frequent itemsets that override the estimated frequent   threshold equal and above 51.

Table 2- the result of the apriori algorithm

Apriori level	Frequent instances selected	frequency
Level3	E2 K3 O3	51
Level3	E2 K3 A1	52
Level3	E2 K3 R3	53
Level3	E2 K3 Z2	54
Level3	E2 K3 AD3	56
Level4	E2 K3 O3 Z2	34
Level4	E2 K3 Q3 S3	34
Level4	E2 K3 Q3 A1	34

Step 2:

They were using the previous output in step 1 (the Apriori result) as an input in this step to catch the relations and the associations that were found in the previous superset upon estimated threshold for Support >=25% and Confidence>=70% for each pair of attributes found in the Apriori result. In this paper, it was necessary to determine the support and confidence values. After many trials and probabilities were identified for the results, the support value 25 and confidence 70 were chosen. At these values, good results emerged.

Example: -

E2 $\longrightarrow$ ALL

E2	$\longrightarrow$	K3	Sup = 8/8 = 100 %	, Conf = 8/8 = 100 %
E2	$\longrightarrow$	O3	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
E2	$\longrightarrow$	A1	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
E2	$\longrightarrow$	R3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %
E2	$\longrightarrow$	Z2	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
E2	$\longrightarrow$	AD3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %
E2	$\longrightarrow$	Q3	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
E2	$\longrightarrow$	S3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %

K3 $\longrightarrow$ ALL

K3	$\longrightarrow$	E2	Sup = 8/8 = 100 %	, Conf = 8/8 = 100 %
K3	$\longrightarrow$	O3	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
K3	$\longrightarrow$	A1	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
K3	$\longrightarrow$	R3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %
K3	$\longrightarrow$	Z2	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
K3	$\longrightarrow$	AD3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %
K3	$\longrightarrow$	Q3	Sup = 2/8 = 25 %	, Conf = 2/8 = 25 %
K3	$\longrightarrow$	S3	Sup = 1/8 = 0.125 %	, Conf = 1/8 = 0.125 %

O3 $\longrightarrow$ ALL

O3	$\longrightarrow$	E2	Sup = 2/8 = 25 %	, Conf = 2/2 = 100 %
O3	$\longrightarrow$	K3	Sup = 2/8 = 25 %	, Conf = 2/2 = 100 %

O3 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
O3 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
O3 $\longrightarrow$ Z2   Sup = 1/8 = 0.125 %   , Conf = 1/2 = 0.50 %
O3 $\longrightarrow$ AD3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
O3 $\longrightarrow$ Q3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0 %
O3 $\longrightarrow$ S3   Sup = 0/8 = 0%   , Conf = 0/2 = 0 %
A1   ALL
A1 $\longrightarrow$ E2   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
A1 $\longrightarrow$ K3   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
A1 $\longrightarrow$ O3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
A1 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
A1 $\longrightarrow$ Z2   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
A1 $\longrightarrow$ AD3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
A1 $\longrightarrow$ Q3   Sup = 1/8 = 0.125 %   , Conf = 1/2 = 0.50 %
A1 $\longrightarrow$ S3   Sup = 0/8 = 0%   , Conf = 0/2 = 0 %
R3 $\longrightarrow$ ALL
R3 $\longrightarrow$ E2   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
R3 $\longrightarrow$ K3   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
R3 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
R3 $\longrightarrow$ O3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
R3 $\longrightarrow$ Z2   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
R3 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
R3 $\longrightarrow$ Q3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0 %
R3 $\longrightarrow$ S3   Sup = 0/8 = 0%   , Conf = 0/1 = 0 %
Z $\longrightarrow$ ALL
Z2 $\longrightarrow$ E2   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
Z2 $\longrightarrow$ K3   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
Z2 $\longrightarrow$ O3   Sup = 1/8 = 0.125 %   , Conf = 1/2 = 0.50%
Z2 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Z2 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Z2 $\longrightarrow$ AD3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Z2 $\longrightarrow$ Q3   Sup = 0/8 = 0%   , Conf = 0/2 = 0 %
Z2 $\longrightarrow$ S3   Sup = 0/8 = 0%   , Conf = 0/2 = 0 %
AD3 $\longrightarrow$ ALL
AD3 $\longrightarrow$ E2   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
AD3 $\longrightarrow$ K3   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
AD3 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
AD3 $\longrightarrow$ O3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
AD3 $\longrightarrow$ Z2   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
AD3 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
AD3 $\longrightarrow$ Q3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0 %
AD3 $\longrightarrow$ S3   Sup = 0/8 = 0%   , Conf = 0/1 = 0 %
Q3 $\longrightarrow$ ALL
Q3 $\longrightarrow$ E2   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
Q3 $\longrightarrow$ K3   Sup = 2/8 = 25 %   , Conf = 2/2 = 100 %
Q3 $\longrightarrow$ A1   Sup = 1/8 = 0.125 %   , Conf = 1/2 = 0.50 %
Q3 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Q3 $\longrightarrow$ Z2   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Q3 $\longrightarrow$ AD3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0%
Q3 $\longrightarrow$ O3   Sup = 0/8 = 0 %   , Conf = 0/2 = 0 %

Q3 $\longrightarrow$ S3   Sup = 1/8 = 0.125 %   , Conf = 1/2 = 0.50 %

S3 $\longrightarrow$ ALL

S3 $\longrightarrow$ E2   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
S3 $\longrightarrow$ K3   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
S3 $\longrightarrow$ A1   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
S3 $\longrightarrow$ R3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
S3 $\longrightarrow$ Z2   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
S3 $\longrightarrow$ AD3   Sup = 0/8 = 0 %   , Conf = 0/1 = 0%
S3 $\longrightarrow$ Q3   Sup = 1/8 = 0.125 %   , Conf = 1/1 = 100 %
S3 $\longrightarrow$ O3   Sup = 0/8 = 0%   , Conf = 0/2 = 0 %

    Relationships with Support were extracted above 25%, and Confidence is greater than 70% between these combinations for each pair of traits found after extracting the rules. Nine rules were chosen that met the threshold requirement estimated in this paper as follows:

Rule 1: E2 $\longrightarrow$ K3, Support = 100%, Confidence = 100%.
Rule 2: K3 $\longrightarrow$ E2, Support = 100%, Confidence = 100%.
Rule 3: O3 $\longrightarrow$ K3, Support = 25%, Confidence = 100%.
Rule 4: A1 $\longrightarrow$ E2, Support = 25%, Confidence = 100%.
Rule 5: A1 $\longrightarrow$ K3, Support = 25%, Confidence = 100%.
Rule 6: Z2 $\longrightarrow$ E2, Support = 25%, Confidence = 100%.
Rule 7: Z2 $\longrightarrow$ K3, Support = 25%, Confidence = 100%.
Rule 8: Q3 $\longrightarrow$ E2, Support = 25%, Confidence = 100%.
Rule 9: Q3 $\longrightarrow$ K3, Support = 25%, Confidence = 100%.

4. Results and Summary
4.1 The Results
Result 1:
    The results of the second step indicate that the first and second bases are the strongest and have the best Support. This medium is the opportunity to perform with all elements. Column (E) with the highest Support is omitted from the original file, and we apply a statistical process called t-test for one sample.

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation
30	270	80.4259	8.28265

One-Sample Test

	Test Value = 75		
	t	Df	Sig. (2-tailed)
30	10.764	269	.000

    The results showed that their average score on the scale reached (80.43) degrees and with a standard deviation of (8.28) degrees, and when balancing this average with the hypothetical average of the scale of (75) degrees, and using the T-test for one sample that shows the difference indicates that it is statistically in favour of the arithmetic mean, as the (calculated T value) was higher than the (tabular T value) of (1.96) with a degree of freedom (269) and the significance level (0.05).

Result 2:

Element columns that appeared with the letter E in step 2 were deleted and included (A, Q, O, and Z). Column k was left because it was parallel to column E. The T-test for one sample was applied again to the data after our previous deletion.

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation
26	270	68.7630	7.35525

One-Sample Test

	Test Value = 65		
	t	df	Sig. (2-tailed)
26	8.406	269	.000

The results showed that their average score on the scale reached (68.76) degrees and with a standard deviation of (7.355) degrees, and when balancing this average with the hypothetical average of the scale of (65) degrees, and using the T-test for one sample shows that the difference Indicates statistically and in favour of the arithmetic mean, as the (calculated T value) was higher than the (tabular T value) of (1.96) with a degree of freedom (269) and the significance level (0.05).

Result 3:

Column K is deleted because it equals column E by force. After one previous deletion, the T-test for one sample was reapplied to the data.

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation
25ﺍ	270	66.5111	7.02610

One-Sample Test

	Test Value = 62.5		
	t	df	Sig. (2-tailed)
25	9.381	269	.000

The results showed that their average score on the scale reached (66.51) degrees and with a standard deviation of (7.026) degrees, and when balancing this average with the hypothetical average of the scale of (62.5) degrees, and using the T-test for one sample shows that the difference Indicates statistically and in favour of the arithmetic mean, as the (calculated T value) was higher than the (tabular T value) of (1.96) with a degree of freedom (269) and the significance level (0.05).

4.2 Summary

Here are some experimental assessments of the new technology. Note the results of T-test for one sample, in the result 1. Column E was removed from the original file because it is the element with the highest level of Support and Confidence as the results of the second algorithm (Association rule mining) showed in the second step, and Column K was kept as an actor because it also has the highest Support and Confidence. T-test for one sample was

applied to the file, and the test results showed that students have a high academic. The deletion of column E did not affect the results, which means that there is no need for this question in the survey.

In result 2, all of the elements associated with E have been removed from the result of the Association rule mining, which has the highest Support (Sup) and Confidence (Conf) of the original file. Once again, the T-test was applied to one sample, and the result was that distinguished students have a higher academic level compared to other students, which means that there is no need for these questions in the questionnaire.

In result 3, column K was deleted and had the highest Support and Confidence from the excel file, and the application of the T-test was applied to one sample. It was found that distinguished students had a high academic level compared to other students. As a result, a useful approach has been taken for summarizing psychological questionnaires from questions that do not contribute to the usefulness of the research and do not change the outcome.

5. Conclusions

 In the psychological fields, the researchers deal with simple statistical tools such as the mean and the standard deviation, etc. These tools have specific limited functions. As for knowing the importance of each question or not, the results need psychological experts. The following represents the concluded results that were found in the proposed work:

• This approach designed an effective method using data mining techniques. In this research, a psychological questionnaire was summarized without the need for experts in this field and without losing efforts and time.

• The subsequent steps that were executed within the proposed work constitute a direct method for the analysts of the psychological field to be followed effectively and consistently.

• These steps in the proposed strategy will produce a better perception of the statistics and therefore, will lead to higher required and diverted effects that ought not to be extracted using the primitive and statistical conventional equipment.

• Summarizing the research form using the proposed data mining approach will not affect the physiological results and findings.

• Additional benefits arise upon using data mining analysis that will give a new understanding of the research subject and will lead to new original ideas.

References

[1] Mustafa, T. K. and Abd, M. S., "Proposed Approach for Analyzing General Hygiene Information Using Various Data Mining Algorithms". *Iraqi journal of science*, vol. 58, no. 1, pp. 337-344, 2017.

[2] Asma.B,Youcef. D, Jerry.C. L, Chongsheng. Z, Alberto. C., "Exploring Pattern Mining Algorithms for Hashtag Retrieval Problem". *IEEE.*, vol. 8, pp. 10569 – 10583, 2020.

[3] Hassan, S., The enneagram personality types and academic self among the distinguished students.Mc.s Thesis ,Department of Educational and Psychology Sciences, College of Education of Human Sciences,Univerity of Anbar, 2017.

[4] Burman, I. Som, S. Hossain, S. and Sharma, M., "Mining to Discover Association of Psychological Factors with Student Academic Performance". *International Conference on Information Systems and Computer Networks*, ISCON, pp. 432-435, 2019.

[5] Boyko, N. Komarnytska, H. Kryvenchuk, Y. and Malynovskyy, Y., "Clustering algorithms for economic and psychological analysis of human behavior". *CEUR Workshop Proceedings*, vol. 2588, pp. 1-13, 2019.

[6] Qin, H. Wu, Z. and Wang, M., "Application of data mining in cultivation of student mental health", *OP Conference Series: Materials Science and Engineering*, vol. 750, no. 1, pp. 1-9, 2020.

[7] Morales Mateos, E. López Garrido, M. Hernández Aguilar, J. Ochoa Ortiz, C. González González, O. and Corona Ferreira, A., "Visual Association Rules on the Psychological

Connection of University Students with their Studies". *Research in Computing Science*, vol. 148, no. 6, pp. 263-276, 2019.

[8] Kang, L.Wu, C. and Wang, B., " Principles, approaches, and challenges of applying big data in safety psychology research", *Frontiers in Psychology*, vol. 10, no. (JULY), pp. 1-12, 2019.

[9] Long, Z. Feng, H. Tiantian, X. and XiangJun, D., "Positive and Negative Association Rules Mining for Mental Health Analysis of College Students". *EURASIA Journal of Mathematics Science and Technology Education*, vol. 8223, no. 3501, pp. 5577-5587, 2017..

[10] Mustafa,T.K., "Textmining authorship detection methods development". *International Journal of Advances in Electronics and Computer Science*, no. December 2018, pp. 2394-2835, 2019.

[11] Neysiani, B. Soltani, N. Mofidi, R. Nadimi-Shahraki, M., "Improve Performance of Association Rule-Based Collaborative Filtering Recommendation Systems using Genetic Algorithm", *International Journal of Information Technology and Computer Science*, vol. 11, no. 2, pp. 48-55, 2019.

[12] Basma,S. Ahmed,S.Ali,A.and Lamees, S., "Analysis of Weka Data Mining Techniques for Heart Disease Prediction System". *International journal of medical reviews*, vol. 1, pp. 15-24, 2020.

[13] Ecma., {C#} *Language Specification ,5th Edition. Ecma International. Rue du Rhône 114. E-Book*, 2017.

[14] Arkkelin, D., "Using SPSS to Understand Research and Data Analysis". *Psychology Curricular Materials*. Valparaiso University ValpoScholar , vol. 2014, pp. 1- 194, 2014.