TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 19, No. 2, April 2021, pp. 444~453
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v19i2.16738

A hybrid analysis model supported by machine learning algorithm and multiple linear regression to find reasons for unemployment of programmers in Iraq

Mohamed A. Abdulhamed[1], Hadeel I. Mustafa[2], Zainab I. Othman[3]
[1,2,3]Computer Scince and Information System collage, University of Basra, Iraq
[1]Computer Science Department, University of Basrah, Iraq
[2,3]Computer Information System Department, University of Basrah, Iraq

Article Info

Article history:

Received May 19, 2020
Revised Aug 16, 2020
Accepted Oct 7, 2020

Keywords:

Apriori
Association rules
Data mining
Machine learning algorithm
Multiple linear regression
Weka

ABSTRACT

The problem of unemployment is one of the most important problems faced by most countries of the world, and it is one of the intractable problems in developing countries, and in Iraq unemployment occupies great importance due to its high rates. This problem in itself is a serious condition, because it results from mismanagement and the structure of the economy, and despite its great importance, it has not been carefully monitored. There are studies and strategies that deal with the analysis and study of those causes that lead to this problem, such as traditional statistical methods, various mathematical and statistical methods, in this research proposed a method uses machine learning methods to find the factors that affect the causes of this problem, as well as the multiple linear regression method.

Corresponding Author:

Zainab I. Othman
Computer Information System Department
Al-Ashar, Corniche St., Basrah, Iraq
Email: zai_nab_alfahad@yahoo.com

1. INTRODUCTION

Unemployment is currently one of the main problems facing most of the world, and unemployment is one of the intractable problems in developing countries, particularly the Arab countries, and the issue of unemployment in Iraq occupies a special importance [1]. given the high rates due to the restructuring of the economy and the many problems resulting from it, despite its importance However, it did not accurately monitor, and the evidence for this is the contradiction of official statistics among them, as well as its contradiction with what is published by Arab and international organizations. At a time when the World Bank statistics indicate that the unemployment rate in Iraq exceeds 50% [2]. The results of the survey conducted by the Ministry of Planning and Development Cooperation in cooperation with the Ministry of Labor and Social Affairs indicated that the unemployment rate in Iraq is 1.28%, and informal organizations have identified Unemployment rate in Iraq (40-60) % regardless of the conflict of numbers [3].

Economic theories indicate that 15% of the capable and searching workforce looks for a real crisis if the government, in cooperation with the private sector and international and civil organizations, does not take practical solutions to confront it [4]. Given the importance of the issue of unemployment and its accompanying repercussions, we chose a specific group of unemployed who are programmers, and we also adopted a mechanism for analyzing data to determine the most important causes and problems that lead to

unemployment. As the trends of modern science are in the possibility of using them to solve societal problems, the method used was to analyze data by linking it with artificial intelligence techniques and statistical processes in analyzing this data [5]. Data mining has attracted a lot of attention in the research community over the past decade, in an attempt to develop scalable algorithms and adapt to an increasing amount of data in the search for meaningful knowledge patterns [6, 7].

Packages of algorithms and software have grown significantly over the past decade. Data mining approaches can be divided into two basic types are [8, 9]: First: Descriptive exploration: which relies on reorganizing data to extract models in it and includes (associative rules, sequencing discovery, aggregation). Second: Predictive exploration: which tries to find the best predictions based on data and includes (classification, chronology, prediction) [10, 11]. This study included a description of demographic data for respondents represented by graduates of colleges of computer science in Basra Governorate, based on four Themes to study unemployment (the university education, the investment, the administrative factors and the graduate personality).

Then the stage of data analysis by relying on multiple linear regression analysis and using the spss program to determine the effect of demographic variables as independent factors on the dependent variable represented by the skills of programmers in the use of information technology and then rely on descriptive exploration to find the most important factors affecting the unemployment of programmers with the piriori algorithm which is one of the algorithms Distinguished in finding correlation relationships in data mining techniques to study the connections between study variables by relying on the weka program, then we discussed the most important conclusions that we obtained from this articale [12, 13].

Previous studies

Unemployment is very interested phenomena . there are many researches that related with. In [14] illustrated use panel data analysis methods that depend on cross-section dependency which give more reliable results. The results obtained as a conclusion that the impact of shocks on the unemployment rate are permanent. In [15] Apriori algorithm applied whith Visual Basic software as a tool for determin consumer purchase, as an output can be concluded, if the mini-support equal 15% while confidence equal 50% then 87 of rules will be generated as results. In [16] Weka software provided approximately all characteristics of data mining techniques. So, hat, the rule generated by Apriori provided market strategies for improve product quantities. In this article using Weka software to find list of the possible itemsets. In [17], machine learning algorithms are used to apply several scenarios to the problem of traffic congestion in Greece, in particular, a comparative test was conducted using four of the most common methods used in the field of machine learning (support vector regression (SVM), neural networks (NN), random forests (RF), and multiple linear regression (LR)), predict the traffic status. Where was obtained mean absolute error (MAE): 6.25, 6.57, 6.44 and 6.90 for SVM, NN, RF and LR methods respectively.

In [18] this study was conducted on individuals who previously had a job and then became unemployed, as a survey was conducted to obtain data in Turkey in order to predict the unemployed based on the use of machine learning algorithms and then compare with logistic regression analysis as econometric approach and shallow neural network, the results showed the superiority of machine learning algorithms over logistic regression and a shallow NN. So that an accuracy rate of 67% was obtained for the machine learning algorithm. In [19] phone records are relied upon to predict individual employment cases by using a survey on family records, as machine learning models were relied upon to predict and find those proportions for approximately 18 occupations in South Asia, where the result of the prediction accuracy was 70.4% Depending on the deep neural network models. In [20] proposed model to predict the value of the bid for bids and tenders for companies, where data of approximately 26 tenders has been entered to predict the price of the winning tender based on the linear regression model, where the proposed model showed distinct results to predict values with an error rate 3%, coefficient of determination R2=0.88167.

In [21], a model was made based on smart meters to collect data to monitor the state of electric power and the extent of its impact on some axes to design a smart model based on meter readings to predict the unemployed through the data collected by machine learning. The importance of research lies in:

- Decrease the important reasons that causes unemployment, by using data mining techniques, which benefits to findind the correlation between the and extract the knowledge patterns use to processing this problem.
- Decrease the governments efforts in restructure the ecomonic and supported the future plans building to counteract this phenomenon.
- Increase the works opportunities provident for graduates, such as depending relation and indicators, for defination the reasons practically, accurate and try to overcoming.

2. REGRESSION LINEAR ANALYSIS

Regression analysis is a statistical method used to analyze the relationship between one or more independent variables and a dependent variable. Regression analysis is mostly used for three purposes [12, 22]:

- Description: the linear regression model is used to describe the shape of the relationship between independent and dependent variables.
- Estimation and predication: the regression model is used linearity to predict the independent values of the dependent variable corresponding to the actual values of the independent variables. Estimation and forecasting are among the most important uses of regression analysis in the applied aspect.
- Control: means the interpretation of the change in the values of the dependent variable in terms of the change in the values of the independent variable on the basis of taking the independent as a controllable variable. Linear regression analysis is divided into two parts: simple linear regression and multiple linear regression. In these studies, we will rely on multiple regression analysis [23, 24].

Assuming that the variable y expresses the dependent variable and the variables (x_1, x_2, ..., x_k) express k from the independent variables and that the number of observations is n, then the dependent view i = 1,2, .. ., n, y_i can be expressed as a linear function in the Views group as follows:

$$Yi= \beta_{0+} \ \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i \quad (1)$$

where (β_0 , β_1 ,β_2, …, β_k) expresses the regression coefficients, e_i expresses the random error of viewing number i, i = 1,2,…, n, where n represents the number of observations, and equations can be formulated into matrices shown in (2).

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \ldots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{bmatrix} \quad (2)$$

Among the most important hypotheses of the regression model, there is independence between (x_1, x_2, ..., x_k) and random error e_i such that (3):

$$\text{Cov (X, E)} =0 \ , \ e_i \sim \text{N}(0 ,\sigma^2) \text{ , E} \sim \text{N}(0 ,\sum) \text{ and } \text{E}(e_i) =0 \quad (3)$$

where $\sum$ represents the variance matrix and is expressed by (4).

$$\sum = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} \quad (4)$$

By applying a method least squares (LOS) To find an estimate of the parameters of the regression model (2) that contains (K + 1) of the parameters is the vector $\beta = (\beta_0, \beta_1, \ldots, \beta_K)'$ where [24]:

$$\beta = (x'x) - 1 \ x'\hat{}Y \quad \ldots \quad (5)$$

3. ALGORITHM APRIORI

This algorithm is one of the distinctive algorithms in finding relationships for links in data mining operations, as this algorithm works to find interesting relationships between variables in large databases. Where the purpose is to find and define rules discovered in huge databases. This algorithm relies on two metrics to determine the associations strength they are confidence and support [16]. For mor explain See Algorithm 1 [25], that illustrate Apriori algorithm steps. At first Apriori algorithm pass simply counts item occurrences to determine the large 1-itemsets. Then a later pass, consists of two stages: Firstly, the large itemset L_{k-1} git it in the (k-1) the pass is used to generate the candidate itemset C_k, using the Apriori candidate generation function as show in the algorithm 1. finally, the database is discovered to be found support of candidates in C_k is counted [7, 11]. So that, the Apriori generation function takes L_{K-1} argument, the set of all large (k-1)-itemset extracted a superset [20, 21].

Algorithm 1. Pseudocode of Apriori algorithm

```
L1 = {large 1-itemsets};
For {K=2; Lk-1 ≠ 0; K++} do begin
       Ck = apriori-gen (Lk-1); // New candidates
```

```
  Forall transactions t ∈ D do begin
       Ct= subset (Ck , t); // Candidates contained in t
        Forall candidates c ∈ Ct do
               c.count ++;
  End
Lk = {c ∈ Ck | c.count ≥ minsup}
End
Answer = Uk Lk ;
```

In this algorithm, the LK represent Set of large k-itemset with minimum support, While CK represent Set of candidates k-itemset.

Algorithm 2. The Apriori-gen function

```
Insert into Ck
Select p.item1, p.item2, … , p.item k-1  , q.item k-2
From Lk-1  p, Lk-1 q
Where p.item1 = q.item1,…, q.itemk-2= q.itemk-2 ,
p.item k-1 < q.item k-1 ;
```

4. RESEARCH METHOD

In our proposed research, we used one of the machine learning algorithms to find the most influencing reasons for programmer unemployment in Iraq. We initially did a questionnaire for a group of the main axes that have a direct impact on unemployment in general. This survey included nearly 100 samples, where nine samples were excluded from the questionnaire in the primary treatment because the samples were excluded during the primary treatment process, the diagram in Figure 1 shown the steps were taken on the data to obtain the results.

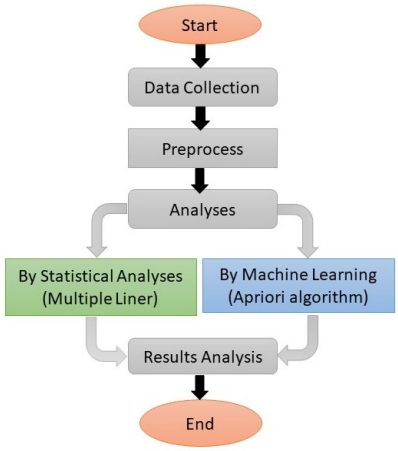

Figure 1. The proposed model flowchart for programmer's unemployment

4..1. Description of the data

That the method of collecting data in our research was the work of a questionnaire through Google Form, where the questionnaire was published throughout all of Iraq, and during a period of 7 days, the data represented by ninety responses were collected and the responses were different and varied, that the questionnaire is generally divided into two parts, the first section includes demographic information and some the most frequently asked questions. As for the second section, it included four axes, and each axis includes four paragraphs. We have adopted the stability and reliability test for the validity of the questionnaire is the calculation of the Cronbach's alpha which was 0.775 and reflects a good value for the approval of the questionnaire and the results of the research. The research included several variables related to graduates of colleges of computer science.

4.2. Preprocessing stage

Like any data analysis process, we must first Carrying out some primary treatments to improve the row data collected. This stage involved the completion of some basic tasks, which are summarized as follows:

- To facilitate the analysis process, the axes and their questions must be coded into short formulas and as in the Figure 2 that illustrates this process.
- Figure 3 describes the coding for the features presented by the questionnaire to find reasons for the unemployment of Iraqi programmers, so that attributes were divided into four important axes which are (University Education UE, INvestment IN, ADministrative factors AD and The CHaracter of the graduate CH) depending on the reality of life in Iraq.
- Then, we transformed the data format into a type called (ARFF), which represents the type of rules that can be dealt with it by using Weka Explorer.
- The third step of pre-treatment We have used filters that convert data in different ways. Since there are two types of filters: supervised and unsupervised filters, we chose one of the types of unsupervised filters that are called (Numeric To Nominal) filter so that use to converting numeric values to nominal values because of association rules in Weka software can only support nominal values.The Figure 3, illustrated the output of applied numeric to nominal filter for all attributes.

University Education (UE axes)	
Qquestions	**coding (Instances)**
The curricula are commensurate with the evolving requirements of the labor market	**UE1**
The use of laboratories, software, and advanced technology projects during the study phase supports the programmer's Scientific	**UE2**
The teaching personality affects the personality of the programmer and helps in understanding and creativity	**UE3**
Carrying out developmental courses for students and transferring experiences from outside the country to develop students 'abilities are good and improve their abilities	**UE4**
Investment (IN axes)	
Qquestions	**coding (Instances)**
The government's failure to invest in local factories helps greatly to exacerbate the problem of unemployment	**IN1**
Reliance on foreign workers and not giving the opportunity to graduates is an important reason for reducing job opportunities for graduates	**IN2**
Failure to provide assistance, support, and encouragement to graduates by the government, such as facilitating loans, increases unemploymen+D2:E2t	**IN3**
Opportunities for international companies to invest in Iraqi labor reduce unemployment	**IN4**
Administrative factors (AD axes)	
Qquestions	**coding (Instances)**
The increase in military spending reduces job opportunities for graduates	**AD1**
Putting people in high positions without competence contributes to not giving programmers their opportunities to work	**AD2**
Failure to develop a well thought out plan in terms of preparing students 'admissions in colleges of computer science with the needs of the labor market contributing to the graduation of large numbers of programmers	**AD3**
Misuse of some administrative positions and monopoly of positions on a certain group of people	**AD4**
The character of the graduate (CH axes)	
Questions	**coding (Instances)**
Your admission to the College of Computer Science was in your desired field	**CH1**
You consider yourself a good insider of the latest software and information systems	**CH2**
You have the ability to employ your specialty to serve other jobs	**CH3**
Accepting some job opportunities that may be offered to you as a programmer depends on your family's financial situation	**CH4**

Figure 2. The questionnaire coding process

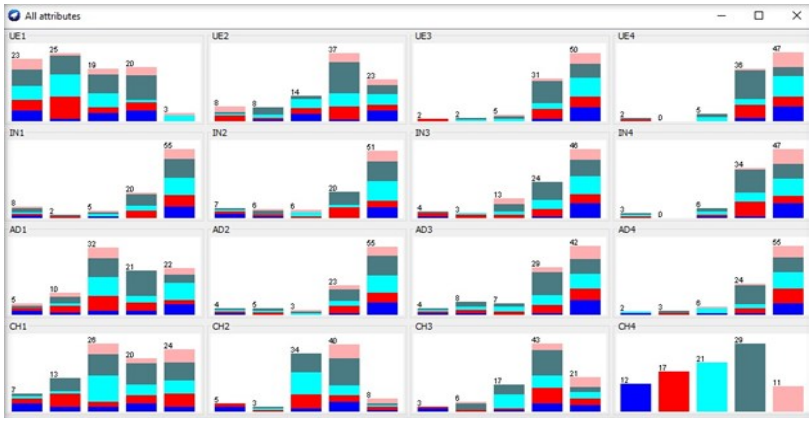

Figure 3. Some of result by applied numeric to nominal filter

4.3. First analysis based on multiple linear regression

Before using regression analysis, we will show demographic information for respondents: gender, age, graduation year, and type of work. The research includes studying the factors affecting unemployment of

programmers, regardless of the gender of the graduate, as the male and female questionnaire was addressed in different proportions, as in Table 1. The research dealt with different ages for males and females divided according to four categories as in Table 2 that the most responsive category was for ages between 23 and 27 years with a rate of 0.844 and that the lowest response is for the category between 38 and 42 and an average of 0.01.

In Iraq, there are three categories of youth classified as the category of workers in the government sector and the category of employees in the private sector and the last category is the unemployed. We will detail the categories as in Table 3 where the highest rate and by 57% is for the unemployed. The data was divided according to the graduation years into categories as shown in Table 4, the highest category by 54% is for graduates in the year 2018-2019, and the number of those who graduated before 2015-2016 increased by 22% of the total respondents.

Table 1. The performance of gender

Variable	Frequency	Percent	Mean
Male	34	37.8	0.378
Female	56	62.2	0.622
Total	90	100.0	

Table 2. The performance of age of respondents

Age	Frequency	Percent	Mean
23 – 27	76	84.4	0.844
28 - 32	8	8.9	0.089
33 - 37	5	5.6	0.056
38 - 42	1	1.1	0.011
Total	90	100.0	

Table 3. The performance of type the work

Type of the job	Frequency	Percent	Mean
state	27	30.0	0.3
private job	12	13.3	0.13
unemployed	51	56.7	0.57
Total	90	100.0	

Table 4. The performance of the graduation years

Graduation	Frequency	Percent	Mean
2018 – 2019	49	54.4	0.54
2017 – 2018	11	12.2	0.122
2016 – 2017	4	4.4	0.04
2015 – 2016	6	6.7	0.05
other	20	22.2	0.22
Total	90	100.0	

Is the best unbiased linear estimate (BLUE) by formulating study data as a multiple linear regression model as follows:

$$Y_i=\beta_{0+}\ \beta_1 x_{i1} + \beta_2 x_{i2}\ + \beta_3 x_{i3}\ + \beta_4 x_{i4} +\ e_i \qquad \textbf{w}\text{here, } i=1,2,\ldots,90 \tag{6}$$

Such that:
Y_i = The skills of programmers in preparing advanced software and it is the dependent variable.
X_{i1} = The gender of the programmer represents; it is an independent variable.
X_{i2} = The age of the programmer it is represented as an independent variable.
X_{i3} = The graduate year of the programmer represents an independent variable.
X_{i4} = The type of work that a programmer exercises after graduation and represents an independent variable.
The (6) showen the multiple linear regression model for the study is:

Using the statistical analysis program spss, where multivariate regression was analyzed, the results were as follows:

$$\hat{\beta} = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \beta_4\end{bmatrix} = \begin{bmatrix}1.767\\ 0.017\\ 0.126\\ -0.025\\ 0.93\end{bmatrix}$$

The (7) for our multiple linear regression is:

$$\hat{y_i} = 1.767 + 0.017x_{i1} + 0.126x_{i2} - 0.025x_{i3} + 0.93x_{i4} \tag{7}$$

Formulating the hypotheses of the multiple regression model for the study are as follows:

$$H0 = \beta_{0=}\ \beta_1 = \beta_2 = \beta_3 = \beta_4$$

4.4. Second analysis based on Apriori algorithm

In general, there are two steps for applied association rules. Firstly, finding all frequent item sets in a dataset based on minimum support, then find minimum confidence that use to construction of the best rules. The ARFF file that converted in preprocessing that include information regarding each programmer

graduate's, where we enter the programmes.arff file to the Weka explorer interface for applied Apriori algorithm by using configuration in Table 5.

So that, we want use the Apriori Algorithm to find the best association rules that have minimum support = 10% and minimum confidence= 90%. The result that obtained can be illustrated clearly by Figure 4. the parameters mentioned in Figure 4 can be explained as follows, (-N) is required num of rules extracted, while the min confidence for the rule is (-C), the (-D) delta at which the mini support is decreased at each cycle, (-U) Indicate upper bound for min support. Finally, the lower bound for the min support Indicate by (-M). the Figure 5 shown some generated sets of large item sets for associator model based on Apriori algorithm for full training set.

Table 5. The settings of Apriori algorithm by Weka

Properties	Setting value
Metric type	confidence
Number rules	100000
Minimum metric	0.9
Delta	0.05
Lower bound minimum support	0.1

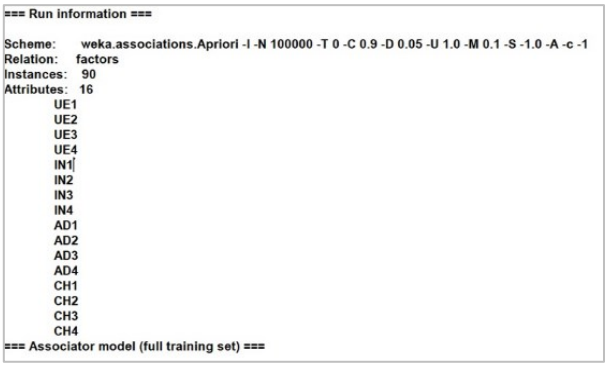

Figure 4. Aprioro information run by Weka

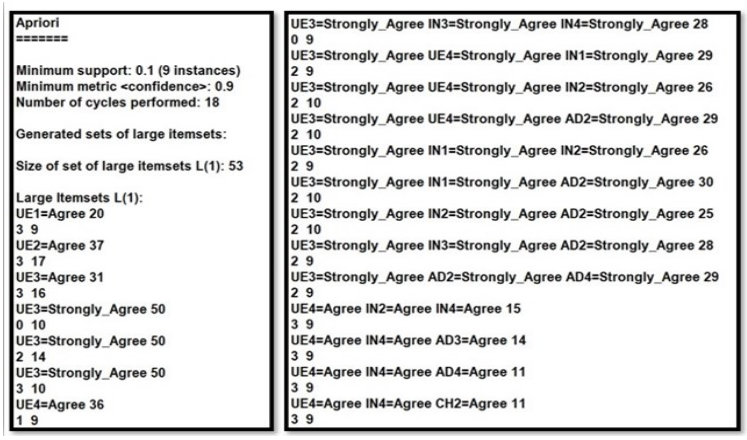

Figure 5. Apriori applied on full training set

For the analysis itemset that found, association rules Represented as X -> Y, so that the frequent itemset are generated based on Aproiri algorithm. The item sets (X) represent antecedent and Y are called consequent of the rule. Generally, Apriori algorithm controlled by the two metrics are support and confidence, to more clarify below some important criteria, so that P(X) and P(Y) represent the counting of total number of tuples at antecedent and consequent respectively. So, P(XY) = P(X□Y) = P (X U Y) = represent the total number of tuples that include both X and Y, the best rules obtained from applied Apriori algorithm in our dataset illustrated in Figure 6. In this programmer dataset, we can calculate the interest rules based on the Weka results for each generation association rules. In our analysis, As only one association rule

was obtained is (UE4=Agree AD3=Agree AD4=Agree 11) = P(X)= 11 itemset (CH4=Agree 10) = P(Y)= 10 itemset (UE4=Agree AD3=Agree AD4=Agree 11 ==> CH4=Agree 10) = P(XY)= P(X U Y)= 10.

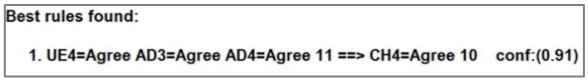

Figure 6. The best rule found in programmers' dataset

5. RESULTS AND DISCUSSION

This stage shows a summary of the results that were extracted depending on the application of the techniques used to analyze the data on a group of samples previously collected using a survey of a number of unemployment in Iraq after their graduation. The correlation rules were created using the Apriori algorithm and compared to Multiple linear regression. To clarify the results more, we will analyze the extracted results based on the methods previously suggested as follows:

5.1. Multiple linear regression

By using multiple linear regression analysis, the calculated value of F* is 0.721 at the confidence level 0.05 and the F-value at degrees of freedom = (k, n-k-1) is 2.49 indicating that the null hypothesis is accepted in the sense that independent variables such as gender, age, and year graduation in work in the field of computer science and psychology. And the year of graduation does not stand in the way of developing the skills of programmers.

5.2. Apriori Algorithm result

This part, showen the result from analysis of programmer dataset based on Apriori algorithm. The result that obtained can show in Table 6 that illustrate the final output for five important measures used by this tequniqe. To clarify more, the third column (equations) shows how to calculate the resulting value in the last column (output). For more clarification, the Figure 7 can be viewed plot matrix that calculated by Weka explorer software. That illustraet a 2-D plot of the current working relation.

From the above output, it can be concluded that the most influencing factors in the unemployment of programmers are represented in obtaining the best rule calculated by relying on the application of an Apriori algorithm. Where it appears that developing the capabilities of students during the university education stage and transferring experiences to them from outside Iraq has a clear impact on the axis of university education. On the other hand, the lack of a deliberate plan to accept students in the disciplines of computer and information technology, in addition to monopolizing positions for a specific category is the most influential aspect within the administrative axis. Finally, the graduate's personality axis also has an impact on unemployment, due to the computer's graduate's acceptance of any work granted to him depending on financial need and he may be away from his specialty.

In this artical, firstly, to use more properties, the Multiple Linear Regression was applied to the personal informatiom from our dataset. So that, Using the F test, the calculated value is 0.721 at the confidence level 0.05 and the F-value at degrees of freedom is 2.49 indicating that the null hypothesis is accepted. And the year of graduation does not stand in the way of developing the skills of programmers. Secondly, Association rules base mining is use to find hidden patterns, and the Apriori algorithm is used to find Associations rules between this attribute. By use 90 instances based on 16 attributes the minimum support and confidence calculated is 10% and 90% respectively.

Table 6. The final results for the best rule found in programmer dataset

Measures	Itemset	Equation	Output
Coverage	[UE4=Agree AD3=Agree AD4=Agree]	supp(X)= no. of transactions which contain the itemset X / total no. of transactions	11 / 90= **0.12**
Prevalence	[CH4=Agree]	supp(Y)= no. of transactions which contain the itemset Y / total no. of transactions	10/ 90= **0.11**
Confidence	[UE4=Agree AD3=Agree AD4=Agree ==> CH4=Agree]	Conf(X→Y) = Supp (X U Y)/Supp(X)	0.11/ 0.12= **0.90**
Lift	[UE4=Agree AD3=Agree AD4=Agree]	$Lift(X \rightarrow Y) = \frac{supp\ (X\ U\ Y)}{supp(Y)*sup(X)}$	0.11/ (0.12*0.11) = **8.3**
Leverage	[UE4=Agree AD3=Agree AD4=Agree ==> CH4=Agree]	lev (X→Y) = supp (X U Y) − sup(X). supp(Y)	0.11 - (0.12*0.11) = **0.09**

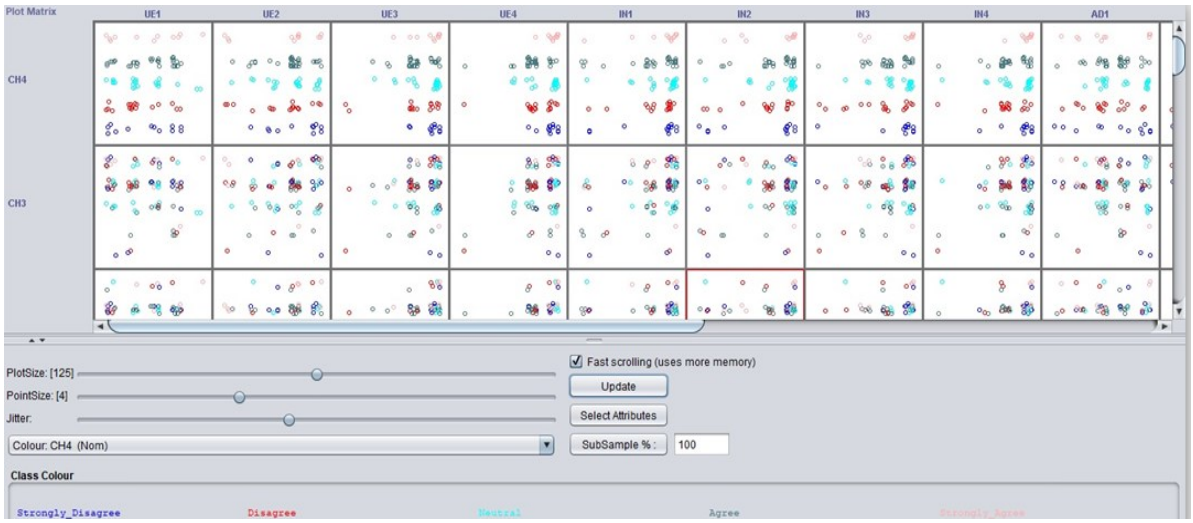

Figure 7. The Plot Matrix for the found association rules by Weka explorer

5 CONCLUSION

A hybrid method of statistical analysis by relying on one of the most important methods used to analyze data that is called multiple linear regression on the one hand, in contrast, using one of the most important methods of data mining in finding the factors with the highest impact based on machine learning, which determines patterns of reasons for not employing graduates of colleges of computer and information technology in Iraq, where patterns are being analyzed to explain how to find the causes for this issue. The mining rule was applied to a questionnaire that was published on a group of graduates of those colleges to analyze the collected results. In future work, we seek to have the questionnaire include more graduates on the one hand. And adding more factors on the other hand, also can be used several other types of algorithms that are used in the areas of finding rules of association such as Eclat or FP-growth algorithms

ACKNOWLEDGEMENTS

We must extend our thanks to everyone who helped us in carrying out this research, starting with every graduate from the Faculties of Computer Science and Information Technology who took us from their time to fill out the questionnaire. In addition to all those involved in providing information and support to us.

REFERENCES

[1] Lucy Anning, *et al.*, "Inflation, Unemployment and Economic Growth: Evidence from The Var Model Approach for The Economy of Iraq," *International Journal of Developing and Emerging Economies*, vol. 5, no. 1, pp. 26-39, 2017.
[2] World Bank Group, "Iraq Economic Monitor: from War to Reconstruction and Economic Recovery," *Spring* 2018.
[3] Al Jazeera, 2019. [Online]. Available at: https://www.aljazeera.net/
[4] Fatih Ayhan, "Youth unemployment as a growing global threat," *АКТУАЛЬНІ ПРОБЛЕМИ ЕКОНОМІКИ*, pp. 262-269, 2016.
[5] Kristina Lindemanna, *et al.*, "The intergenerational effects of unemployment: How parental unemployment affects educational transitions in Germany," *Research in Social Stratification and Mobility*, vol. 62, 2019.
[6] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," *Research Report RJ 9839*, IBM Almaden Research Center, San Jose, California, June 1994.
[7] Darshan M. Tank, "Improved Apriori Algorithm for Mining Association Rules," *Information Technology and Computer Science*, vol. 07, pp. 15-23, 2014.
[8] Adelaja Oluwaseun Adebayo, Mani Shanker Chaubey, "Data Mining Classification Techniques on The Analysis of Student's Performance," *Global Scientific Journals,* vol. 7, no. 4, pp. 79-95, April 2019.
[9] Diego Buenaño-Fernández, David Gil and Sergio Luján-Mora,"Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study," *Sustainability,* vol. 11, no. 10, pp. 1-18, 2019.
[10] Faisal Mohammed N. Ali, *et al.*, "Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents," *Journal of Information and Telecommunication*, vol, 2, no. 3, pp. 1-15, 2018.
[11] Ristianigrum, Sulastri, "Implementation of Data Mining Algorithm Using Apriori," *syntax Proceedings*, 2017.
[12] Petrovski Aleksandar, Petruseva Silvana, Zileska P. Valentina," Multiple Linear regression model for predicting bidding price," *Technics Technologies Education Management,* vol. 10, no. 3, pp. 386-393, 2015.
[13] G. Kaya Uyanik, Nese Guler, "A study on multiple linear regression analysis," *Procedia-Social and Behavioral Sciences,* vol. 106, pp. 234-240, 2013.

[14] Cem Doğan, Sinan Erdoğan, "An Empirical Analyses of Unemployment Hysteresis and Natural Rate of Unemployment Approaches for MENA Countries," *Optimum Journal of Economics and Management Sciences*, vol. 3, no. 2, pp. 41-50, 2016.
[15] Suprianto Panjaitan, *et al*., "Implementation of Apriori Algorithm for Analysis of Consumer Purchase Patterns," *The International Conference on Computer Science and Applied Mathematic,* 2019.
[16] Neeraj Bhargava and Nidhi Gupta, " Efficient Execution of Apriori Algorithm using WEKA," *International Journal of Advanced Research in Computer Science*, vol. 4, no. 7, 2013.
[17] Ch. Bratsas, K. Koupidis, Josep-M. Salanova,"A Comparison of Machine Learning Methods for the Prediction of Traffic Speed in Urban Places," *Sustainability*, vol. 12, no. 1, pp. 1-15, 2020.
[18] Yasin Kutuk and Bulent Guloglu, "Prediction of Transition Probabilities From Unemployment to Employment for Turkey Via Machine Learning and Econometrics: A Comparative Study," *Iktisat Araştırmaları Dergisi, Journal of Research in Economics*, pp. 58-75, 2019.
[19] Pål Sundsøy, *et al*., "Towards Real-Time Prediction of Unemployment and Profession," *International Conference on Social Informatics*, 2017.
[20] Mingjun Li, *et al.*, "An Empirical Comparison of Multiple Linear Regression and Artificial Neural Network for Concrete Dam Deformation Modelling," *Hindawi, Mathematical Problems in Engineering*, vol. 2019, pp. 1-13, 2019.
[21] Y. N. I. Sari, A. Triayudi and I. D. Sholihati, "Implementation of Data Mining to Predict Food Sales Rate Method using Apriori," *International Journal of Computer Applications*, vol. 178, no. 35, 2019.
[22] Casimiro A. Curbelo Montañez and William Hurst, "A Machine Learning Approach for Detecting Unemployment Using the Smart Metering Infrastructure," *IEEE Access*, vol. 8, pp. 22525-22536, 2020.
[23] Timothy Plotts, "A Multiple Regression Analysis of Factors Concerning Superintendent Longevity and Continuity Relative to Student Achievement," Seton Hall University Dissertations and Theses (ETDs), 2011.
[24] Gulden K. U and Nese G., "A Study On Multiple Linear Regression Analysis," *Elsevier, Procedia - Social and Behavioral Sciences*, vol. 106, pp. 234-240, 2013.
[25] Darshan M. Tank, "Improved Apriori Algorithm for Mining Association Rules," *I. J. Information Technology and Computer Science*, vol. 07, pp. 15-23, 2014.

BIOGRAPHIES OF AUTHORS

Mohamed Abdulrahman Abdulhamed, Assist. Lecturer. I have a master's degree in computer science since 2017 from Computer Scince department-college of science-university of Basrah. Currently I am working as a lecturer at Computer Science and Information technology collage, computer science department, university of Basrah, Iraq.

Hadeel Ismail Mustafa, Assist. Lecturer. I have a master's degree in mathmatics since 2014 from mathmatics department-college of Education for pure sciences-university of Basrah. Currently I am working as a lecturer at Computer Science and Information technology collage, computer information system department, university of Basrah, Iraq.

Prof. Dr. Zainab Ibrahim Othman, I have a master's degree in 1997 and I have a PHP's degree in computer science since 2007 from Computer Science Department–College of Science–university of Basra. Currently I am working as a Lecture at Computer Science and Information technology college, computer information system department, university of Basra, Iraq.