Contents lists available at Egyptian Knowledge Bank

Labyrinth: Fayoum Journal of Science and Interdisciplinary Studies

Journal homepage: https://lfjsis.journals.ekb.eg/

Data Mining Application to Detect Student's Exams Performance Factors

Hazem M. Hassan[a, *], Mohammed Khafaga [b], Mostafa Thabet [a]

[a] *Information system Department, Faculty of computers and Artificial Intelligence, Fayoum University, Fayoum 63514, Egypt*
b Computer Science Department, Faculty of Computers and Artificial Intelligence, Fayoum university, Fayoum 63514, Egypt

ARTICLE INFO

Keywords:
Students performance
Statistics
Standard deviation
Simple linear regression
Factors

ABSTRACT

The valuable procedures and evaluation couldn't reveal the valuable information concealed within the student's feedback on their achievements. Also, the rapid proliferation of academic failure has become the systemic challenge of the modern educational system. Therefore, identifying the variables that enhance student achievement and fulfilment was given top priority. Consequently, statistical analysis and data mining techniques were widely applied in a variety of disciplines, including education. Gender, father's occupation, student location, and school attendance may all have a role in this study's findings. This research set out to look into how these factors affected student performance. One hundred and ninety-nine Fayoum high school students were randomly selected for this descriptive correlational study. The information was gleaned via student surveys and the participants' own self-reported demographics. The information was gleaned via student surveys and the participants' own self-reported demographics. Descriptive statistics were calculated using mean and standard deviation; inferential statistics, such as Pearson correlation and linear regression analysis, were computed using SPSS v26. The outcomes indicated a statistically significant correlation (p 0.05) between a student's father's occupation and where he lived with the student's academic performance. A substantial (p 0.05) relationship was found between parent employment and student location, as well as academic performance. According to the results, there was a significant correlation between students' success and the occupation and location of their fathers.

1. Introduction

The level of education available in a society is directly correlated to its level of development, and the success or failure of an educational system may be evaluated by examining its components, its leadership, and how it handles the education of its pupils. Students' academic performance is influenced by a variety of social and cultural contexts, including their families' socioeconomic condition, the parents' occupations, their levels of education and the students' own backgrounds and experiences in school. Addiction among parents, parental absence, financial difficulties covering school costs and similar circumstances all contributed to teenagers' incapacity to succeed in school. Many studies investigated explored a variety of aspects affecting pupils' outstanding academic performance. Ding et al. [1] illusterated that poor academic performance was associated with inadequate physical and mental health.

Park et al. [2] and pong et al. [3] reported no correlation between authoritarian parenting and academic success on the part of their children and an inverse correlation between authoritarian parenting and the success of their children in school. Brian et al.[4] and Guay et al.[5] asserted that parents' attitudes towards education and their children's drive to succeed were major contributors to their children's academic. In recent years, researchers have looked for the factors that can aid in boosting students' performance in the classroom. Every person's learning was influenced by their own unique combination of mental and emotional processes, as is well documented. Not only did people differ in this regard due to their varying levels of intelligence, but also due to their unique sets of perspectives, assumptions, biases, emotions, worldviews and experiences. Research in several fields has focused for

* Corresponding author.
E-mail address: hm2079@fayoum.edu.eg (H. Hassan); Tel.: +201063780581

DOI: 10.21608/IFJSIS.2023.201596.1002
Received 21 March 2023; Received in revised form 12 June 2023; Accepted 3 August 2023
Available online 4 August 2023

a long time on trying to predict students' academic success [6]. In order to implement academic programmes or exams for students, it was necessary to have actual knowledge of the factors that influence the terms and conditions of their existence [7]. The educational performance level was used as a success indicator in scientific endeavours by all educational systems [8,9]. Researchers found that many other factors, such as students' lifestyles, cognitive ability, emotions, personal and familial attributes significantly influenced their academic outcomes.

Descriptive statistics is defined as the method of selecting, analysing, and interpreting numerical data using appropriate charts, graphs, and numerical indices [10]. Both data mining and statistics were related to learning from data. All of their efforts centred on finding patterns in data in an effort to transform raw information into useful knowledge. Despite the fact that both techniques served similar ends, statistics was restricted to data quantification, while data mining has traditionally relied on statistics to supply the essential tools. Data mining, on the other hand, used models to find connections and patterns in data [11].
In addition, SPSS (Statistical Package for the Social Sciences) was a popular tool for social science data analysis. Market researchers, health researchers, survey firms, government agencies, academics, businesses, and data miners all used it [12]. Some researchers considered it as "sociology's most influential tool" since it allowed even inexperienced researchers to perform statistical analysis on their own data. Data management and statistical analysis are both built into the core software. It included, in addition to statistical analysis, data administration (case selection, file reshaping, and constructing derived data) and data documentation [13]. In order to provide reports on the information data's predictor and criteria variables, these were typically arranged in a logical case figures, tables, and graphs. The subsequent representations of the following parameters were: minimum value, maximum value, sum of the values, scores, mean, mode, median, variance and standard deviation [14]. In recent years, educational data mining has received much interest. The main purpose of educational establishments was to offer students with a high-quality education in addition to improving their student's performance and to assure that the stdents generally who joined the system successfully completed their educational requirements [15, 16]. Data mining (DM) was the process of exploring and analysing large datasets for useful patterns and laws. It let users to examine data from many perspectives, classify it, and synthesise the results into useful knowledge. This data could be used to boost profits, reduce expenses, or do both. Data mining, from a statistical perspective, was the automated, computer-assisted exploration of big and complicated datasets. Data mining heavily utilised statistical ideas for analysing massive data sets and discovering patterns across many areas in huge relationship-based databases. Therefore, data mining was defined as the use of statistics in the form of exploratory data analysis and prediction models to ascertain trends and patterns in massive datasets [17]. Moreover, Knowledge Discovery in Databases (KDD) is a methodical approach to discovering actionable insights inside massive, heterogeneous datasets. The model was applied in information extraction, data analysis, and prediction.. A comprehensive evaluation of the clustering algorithm's applicability and effectiveness in the context of EDM were presented in on the basis of the assessed literature, future insights are detailed, and research goals are identified [18, 19] Fig.1.

Using data mining technologies, the performance of undergraduate students was analysed with a focus on two aspects of their performance. The initial stage was to forecast students' academic achievement after a four-year course [20]. The second step involves studying common progressions and merging them with prediction outcomes. There are two distinct groups of students: low scorers and hard workers. The findings suggested that low-achieving students can receive prompt warnings and assistance by concentrating on a small number of courses that are indicative of especially strong or poor performance, while high achievers may receive guidance and opportunities [21]. The purpose of this research was to investigate the association between certain variables (gender, father job, student place, student attendance) and educational achievement of students using Statistical Package for Social Sciences (SPSS) developed by IBM, is the most widely used software to analyze quantitative data.

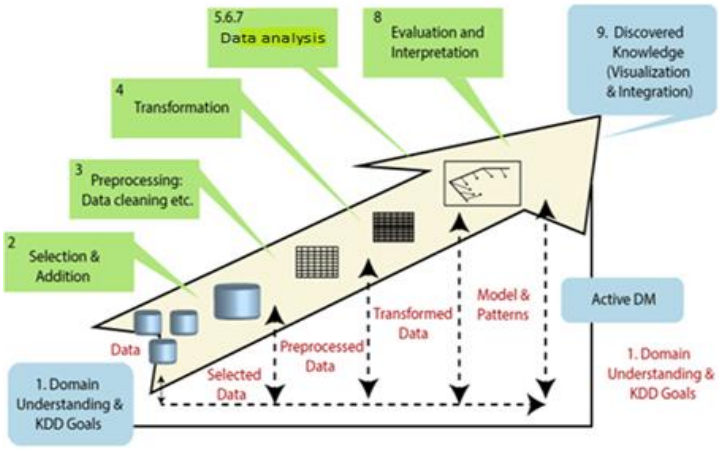

Fig1. The Knowledge Discovery in Databases Process opportunities [22].

2. Review of literature

A few representative studies were provided below, along with numerous reviews of the various individual, social, economic, psychological, and environmental variables influencing students' outcomes as well as the models that have been employed to estimate their performance. Ramaswami et al. [23] have employed a prediction model to analyse the interrelationship between variables utilised as a predictor for achievement in higher levels of secondary education. Variables like the medium of instruction, grades earned in secondary school, the location of the school, the living region, and the type of secondary education were the best indicators of a student's performance in higher secondary education.

Additionally, Studies have shown that students' abilities to describe and interpret a variable's distribution from a histogram, in the context of the data, is quite high even before taking a first course in statistics [24]. While qualitative explanations of distributions are useful for summarising a data set, students will eventually be expected to utilise statistics to mathematically define a distribution in terms of its centre, variability and shape. They quickly determined that central tendencies could be shown by means of the mean, median, and mode, and that variability could be described by means of the standard deviation and range.

Cen et al .[25] estimated that descriptive statistics is the study of, or the act of providing, a numerical description of the most significant charac-teristics of a set of data or a set of observations made by a group of people. A perfect 100 population of students was collected and just the grades of 20 of them, together with certain commonly-used measures to describe a data collection, were presented here. Numerous measures of average, standard deviation, and dispersion were taken. The mean, median, and mode were used to quantify central tendency, while the standard deviation (or variance), the minimum and maximum values of the variables, kurtosis, and skewness were used to measure variability.

3. Proposed Methodology

The objective of this study was to inquire about the social and personal factors influencing student performance.. Correspondingly, descriptive statistics were used to describe the distribution of variables and Pearson bivariate correlations and stepwise regression were utilized for data analysis. Mean, standard deviation and variance and demographic variables were used by Statistical Package for Social Sciences (SPSS) software.

3.1. Data set

This data set contains a total 99 evaluation scores. The social, demographic, student grades and school-related data characteristics present were obtained from school reports and surveys as described in [26]. The performance of different courses, English, Social, Science, Mathematics, Arabic language were represented in all datasets. The categorization and regression of 99 profiles and 4 attributes were explained in Table 1.

3.2 Data Preparation

Everything that was done to construct the final dataset (data that went into the design tool) was done during this stage. The dataset's variables were cleaned and organised in preparation for the following stage, whereby models would be created. The purpose of this analysis was to evaluate and contrast two groups of predictors for academic achievement. Therefore, the basic concept of data preparation was to produce two datasets, one containing only students' individual, social, and geographical details, and the other containing all of this information as well as academic variables including education courses, grades, and absences. Each dataset was assigned its own model, which provided for the evaluation and comparison of its achievements in order to specifically pre-dict if a student would succeed or fail at the end of the academic year. In other words, we may test the discriminative power of the associated dataset variables by comparing student performance to external benchmarks..

A questionnaire was used to collect primary data, which included questions about several personal, socioeconomic, psychological, and school-related characteristics that were hypothesised to affect student performance. Secondary data, such as grade details, were gathered from the schools. Every predictor and response variable extracted from the questionnaire are given in Table 1 for reference.

Table 1: Attributes of the dataset.

Attributes	Details
Gender	Male, Female
Father Job	Nominal: Agricultural engineer, Employee , Farmer, Teacher,Worker
Place	Binary: Abusir, Al-Masara, Aljaafra, Atsa, Dafno, Kafr Al-Zafarani, Sawfinh
Attendance	Absent, Attended

3.3 Model and Algorithm

To accomplish this, two methods are employed: predicting student outcomes and evaluating characteristics acquired at the start of the school year. Using this method to determine which students were most likely to fail at the end of the year. With this knowledge, education specialists could provide these pupils extra attention at the beginning of the school year. Subsequently, The results of this analysis were then applied to characterise the distribution of a single variable by computing its central tendency (mean, median, and mode) and dispersion (data-set, and measures of spread like the variance and standard deviation). Characteristics of a variable's distribution may also be depicted in graphical or tabular format, including histograms. Additionally, The bivariate Pearson correlation was measured to indicate a statistically significant linear relationship existed between two continuous variables. A positive r value expressed a positive relationship between the two variables while a negative r value indicated a negative relationship. A correlation coefficient of zero indicated no relationship between the variables at all. In the other hand, a regression model was estimated to model a linear representation of the relationship between a single independent and dependent variable. The dependent variable's mean often increases along with the value of the independent variable when the coefficient was positive. A negative coefficient indicates a downward trend in the dependent variable alongside an upward trend in the independent variable.

Then, Analysis of Variance (ANOVA) was performed to compare standard deviations across group means (or averages) and establish if there is a significant difference in group means. Under the assumption that there is no difference in group means, the P-value is the chance of receiving an F-ratio as great as or larger than the one observed. Significant P value of the ANOVA test indicates for at least one pair, between which the mean difference was statistically significant. Also, The Mann Whitney U test, was measured to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Finally, The When comparing continuous or ordinal dependent variables across many groups, the Kruskal-Wallis H test was shown to be the most appropriate statistical method for making these comparisons.

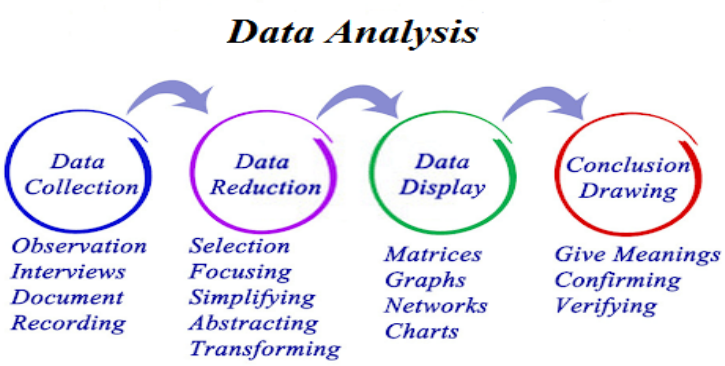

Fig 2. Data analysis steps[27].

4. Experiments and Results

Both descriptive and inferential statistics in this thesis were analysed with the use of the statistical package for the social sciences (SPSS V26) [28]. The method for creating dissimilar analytical models of student achievement is explained in this work. More especially, the steps and outcomes of the essential. Multiple linear regression were used for data modeling for the training dataset and predictions were created for it. The outcomes were analyzed and evaluated. (Table 2) showed the data describing the chosen variables, mainly, the minimum, maximum, mean (M), standard deviation (SD), and coefficient of variation (CV). The scores of English language were (20.51%), for the scores of social studies were (18.07%), for the scores of sciences were 24.08%), for the scores of math were (18.45%), for the scores of Arabic language were (15.06%).

Table 2: Descriptive statistics

Variables	Min	Max	Mean	SD	CV
English language	0	39.2	28.145	5.0848	20.51%
Social Studies	0	39.2	26.798	6.4517	18.07%
Sciences	0	40	30.024	5.5406	24.08%
Math	0	80	65.127	9.8085	18.45%
Arabic language	0	100	60.333	19.092	15.06%
Total score	0	298	210.419	43.1538	31.64%

Table 2: In particular, the min, max, mean (M), standard deviation (SD), and coefficient of variation (CV) were displayed for the specified variables. Furthermore, all correlations were statistically significant at 0.001, as can be shown. The observed coefficients fall within the range of 0.744-0.965. which indicate a significant strong multiple correlations between all variables in the study using Pearson bivariate correlations as illustrated in Table 3.

Table 3: Pearson bivariate correlations

Variable	Correlation	Degree	English language	Social Studies	Sciences	Math	Arabic language
English	Pearson Correlation	1	.939***	.940***	.965***	.890***	.956***
	Sig. (2-tailed)	-	0.000	0.000	0.000	0.000	0.000
Social studies	Pearson Correlation	-	1	.889***	.914***	.854***	.852***
	Sig. (2-tailed)	-	-	0.000	0.000	0.000	0.000
Science	Pearson Correlation	-	-	1	.920***	.793***	.877***
	Sig. (2-tailed)	-	-	-	0.000	0.000	0.000
Math	Pearson Correlation	-	-	-	1	.892***	.879***
	Sig. (2-tailed)	-	-	-	-	0.000	0.000
Arabic	Pearson Correlation	-	-	-	-	1	.744***
	Sig. (2-tailed)	-	-	-	-	-	0.000

Table 3: It showed a significant strong multiple correlations between variables using Pearson bivariate correlations

Table 4 displays the findings of the simple regression analysis. The alternative hypothesis, which states that all variables have a substantial positive effect on degree, is accepted, as the null hypothesis is rejected (P<0.001). The model's significance is confirmed by the F-test in the ANOVA table (F=546827.6,P0.001). Table 5.

Table 4: Regression results

Model	Unstandardized Coefficients		Standardized Coefficients	t	P-value
	B	Std. Error	Beta		
(Constant)	-.100	.197		-.509	.612
English language	1.012	.014	.119	71.578	.000***
Social Studies	.987	.012	.148	85.149	.000***
Sciences	1.019	.018	.131	56.576	.000***
Math	.995	.006	.226	157.666	.000***
Arabic language	.997	.003	.441	314.329	.000***

Table 4: All variables tested have a statistically significant (P<0.001) beneficial effect on graduation rates.

Table 5: Anova table.

Model	Sum of Squares	df	Mean Square	F	P-value
Regression	182494.646	5	36498.929		
Residual	6.207	93	.067	546827.614	.000***
Total	182500.854	98	-		

Table 5: The F-test in ANOVA table confirms the significance of the model since ($F = 546827.6, P < 0.001$).

Since gender was treated as a categorical variable with two independent categories, the Independent-Samples Mann-Whitney U Test revealed a statistically significant difference in students' grades based on their gender and attendance patterns. so the suitable test was nonparametric as showed in Table 6,7 and Fig 2 respectively.

Table 6: Descriptive statistics of degree regarding the gender.

Gender	N	Min	Max	Mean	Std. Dev
Male	92	.0	298.0	208.005	42.7725
Female	7	201.0	295.0	242.143	37.4986
Total	99	.0	298.0	210.419	43.1538

From this data, we may infer that females had a higher average degree (242.143) than males (208.005).

Fig 3. Bar chart for degrees of students regarding the gender

Fig 3: According to a bar chart comparing the average degrees of male and female students, the average degree of female students was higher than that of male students (242.143 vs. 208.005).

Table 7: Independent-samples mann-whitney u test summary.

Total N	99
Wilcoxon W	510.000
Mann-Whitney U	482.000
Test Statistic	482.000
Standardized Test Statistic	2.184
Standard Error	73.250
Asymptotic Sig.(2-sided test)	.029*

Table 7: The Mann-Whitney U test on independent samples found that there was a statistically significant difference in education levels between sexes (U=482,P0.05).

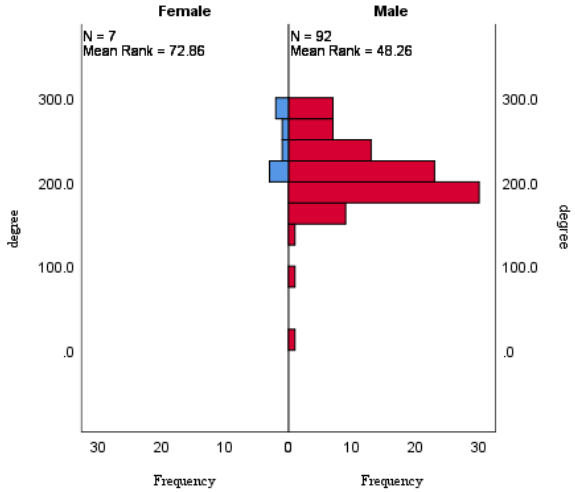

Fig 4. Independent-Samples Mann-Whitney U Test

Fig 4: The Mann-Whitney U test for independent samples indicates that there is a statistically significant difference in education levels between sexes (U=482,P0.05).

Attendance was treated as a categorical variable with two independent categories; thus to evaluate whether or not there was a significant difference in students' grades depending on gender and attendance, the Independent-Samples Mann-Whitney U Test was the suitable nonparametric test. as showed in (Table 8, 9) and Fig 3 respectively.

Table 8: Descriptive statistics of degree regarding the attendance of the students

Attendees	N	Minimum	Maximum	Mean	Std. Deviation
Absent	28	.0	201.0	166.696	38.6600
Attended	71	177.0	298.0	227.662	31.0553
Total	99	.0	298.0	210.419	43.1538

Table 8: The average number of participants (227.662) was higher than the number of people who were absent (166.696), as shown in the following table

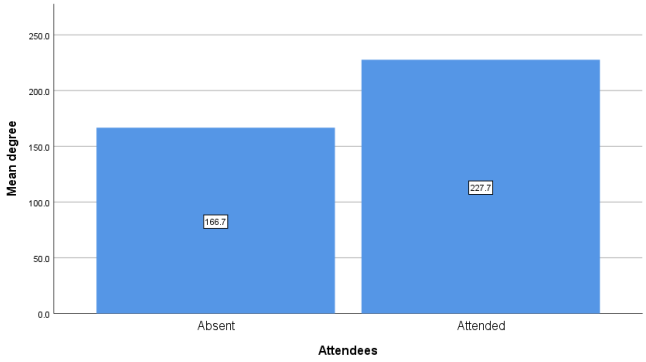

Fig 5. Bar chart for degrees of students regarding the attendance

Fig 5: The average percentage of present students (227.662) was higher than the percentage of absent students (166.696), as seen in the following bar chart.

Table 9: Explanation of the Mann-Whitney U Test for Independent Samples

Total N	99
Wilcoxon W	4504.000
Mann-Whitney U	1948.000
Test Statistic	1948.000
Standardized Test Statistic	7.413
Standard Error	128.698
Asymptotic Sig.(2-sided test)	.000***

Table 9: The Mann-Whitney U test for independent samples revealed a statistically significant difference in grades between present and absent students (U=1948, P0.001).

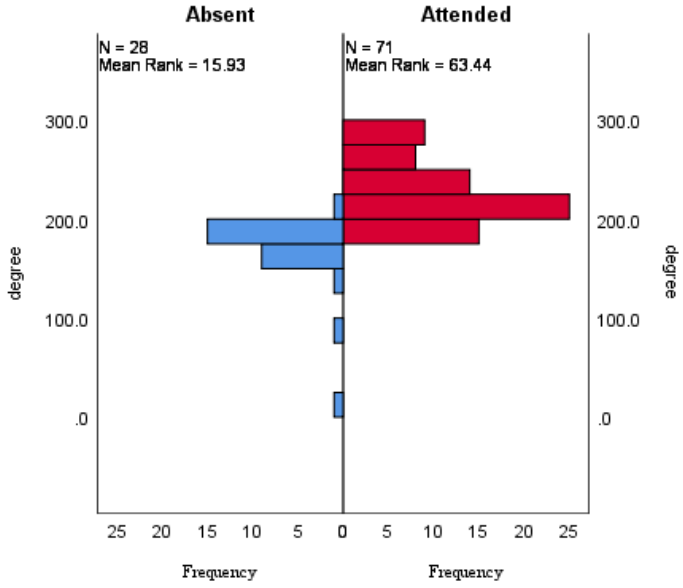

Fig 6. Independent-Samples Mann-Whitney U Test

Fig 6: The Mann-Whitney U test for independent samples revealed a statistically significant difference in grades between present and absent students (U=1948, P0.001).

Tables (10, 11) and Fig 4 showed that the student's father was evaluated using a Nonparametric Independent-Samples Kruskal-Wallis Test since it was treated as a categorical variable with more than two independent categories. Students' performance varied widely depending on their fathers' occupations. These results showed that these elements were significant predictors of students' academic outcomes and levels of achievement.

Table 10:Descriptive statistics of degree regarding the job of student's father

father job	N	Minimum	Maximum	Mean	Std. Deviation
Agricultural engineer	1	192.5	192.5	192.500	-
Employee	34	161.0	298.0	215.485	36.2026
Farmer	3	169.5	230.0	208.667	33.9644
Teacher	5	181.0	295.0	235.600	44.6744
Worker	56	.0	289.0	205.509	47.3356
Total	99	.0	298.0	210.419	43.1538

Table 10: Students whose fathers worked in education had the highest average degree (235.6), while agricultural engineers had the lowest (192.5%).

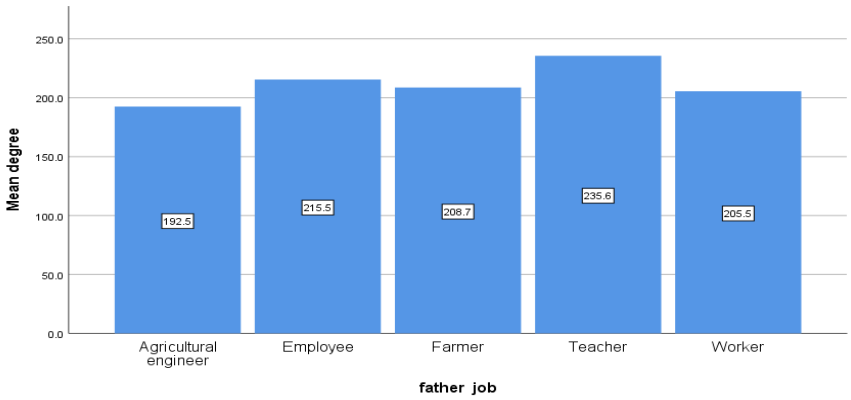

Fig 7. Bar chart for degrees of students regarding the father job

Fig 7: Bar chart for degrees of students regarding the father job that the average degree of students whose father was teacher (235.6) was the greatest while that of agricultural engineer (192.5) was the lowest.

Table 11: Independent-samples kruskal-wallis test summary

Total N	99
Test Statistic	2.342
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.023

Table 11: The results of independent-samples Kruskal-Wallis test showed that there was a significant difference in degrees of students regarding their father's job since P-value < 0.05.

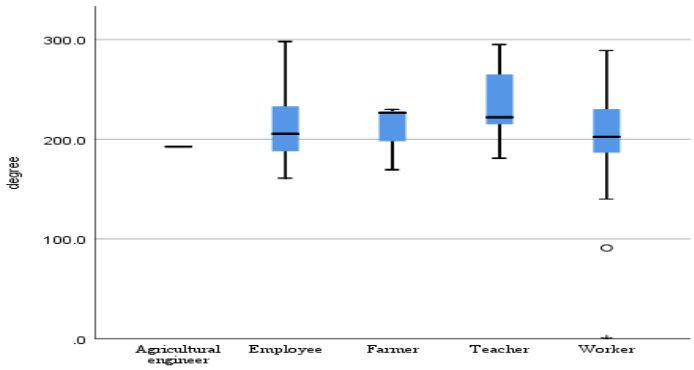

Fig 8. Independent-Samples Kruskal-Wallis Test

Fig 8: The Kruskal-Wallis test for independent samples indicates that there is no statistically significant difference in the students' degrees with respect to their father's occupation (KW=2.342,P>0.05). Furthermore, Table (12,13) and Fig 5 show that there is a significant difference in grades concerning the student's placement, so the Nonparametric Independent-Samples Kruskal-Wallis Test is used to evaluate the student's father, who is

seen as a categorical variable with more than two independent categories. These results showed that these elements are significant predictors of a student's academic performance and their level of success in the future.

Table 12: Descriptive statistics of degree regarding the place of the students

place	N	Minimum	Maximum	Mean	Std. Deviation
Abusir	1	222.0	222.0	222.000	-
Al-Masara	9	176.0	277.5	228.056	36.9750
Aljaafra	1	265.0	265.0	265.000	-
Atsa	60	.0	289.0	205.450	46.1652
Dafno	14	161.0	271.0	205.643	28.1065
Kafr Al-Zafarani	1	295.0	295.0	295.000	-
Sawfinh	13	170.0	298.0	214.692	41.2783
Total	99	.0	298.0	210.419	43.1538

Table 12: The table shows that the average student life expectancy is highest in Kafr Al-Zafarani (295), and lowest in Atsa (205.45).

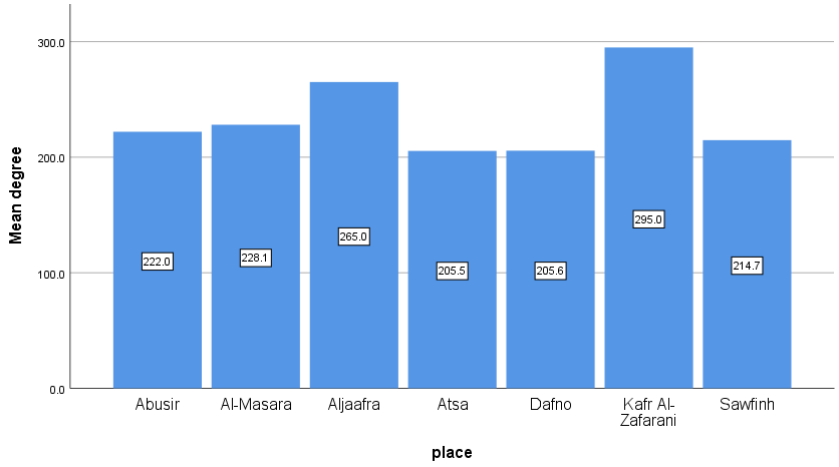

Fig 9. Bar chart for degrees of students regarding the place

Fig 9: The average degree of students in Kafr Al-Zafarani (295) was the highest, while that of Atsa (205.45) was the lowest in a bar chart for degrees of students.

Table 13: Independent-samples kruskal-wallis test summary

Total N	99
Test Statistic	6.999
Degree Of Freedom	6
Asymptotic Sig.(2-sided test)	.321 NS

Table 13: The independent-samples Kruskal-Wallis test findings demonstrate that there is no significant variation in students' degrees based on their placement because the P-value is greater than 0.05 (KW=6.999,P>0.05).

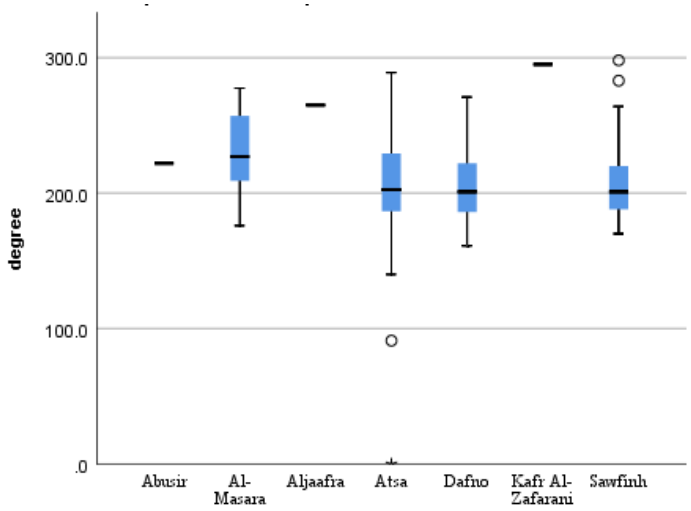

Fig 10. Independent-Samples Kruskal-Wallis Test

Fig 10: Kruskal-Wallis test findings for independent samples demonstrate that there is no statistically significant difference in students' degrees with respect to their placement (P-value > 0.05; KW=6.999, P>0.05).

Table 14: Hypothesis test summary

Null Hypothesis	Test	Sig.	Decision
The distribution of degree is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.029	Reject the null hypothesis.
The distribution of degree is the same across categories of father job.	Independent-Samples Kruskal-Wallis Test	.673	Retain the null hypothesis.
The distribution of degree is the same across categories of place.	Independent-Samples Kruskal-Wallis Test	.321	Retain the null hypothesis.
The distribution of degree is the same across categories of Attendees.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.

5. Conclusion

Every society's achievement and advancement in education is a direct result of the educational system's ability in identifying and prioritising the requirements of each student. Thus, when students' intellectual growth is at the highest level across all grade levels, an educational system is considered effective and successful. To have pupils with high academic success, related factors should be found. In order to do this, the present research was carried out. These data showed that the variables (gender, father employment, location, and student attendance) had an impact on a student's performance and level of success in the long run. Students that have a solid father work and place to live perform better in school and in their future careers. Additional factors that may affect student performance will be investigated in future research using well-known data mining techniques, including but not limited to: stress, anxiety, low entry grades, family support, accommodation, prior evaluation grade, student internal assessment grade, GPA, and students' e-learning activity.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank Fayoum University for supporting the publication of this work.

Author Contributions

All authors contributed to this work. H. Mohammed prepared the samples and completed the experimental measurements. Both M. Khafaga and M. Thabet shared writing and followed the performance of the experiments. M. Khafaga helped the first author complete the sample preparation. . H. Mohammed with M. Thabet completed the paper writing, analyzing the data, and validation. H. Mohammed followed the revision and submission of the manuscript for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] W. Ding, S. Lehrer, J. Rosenquist, J. Audrain. The impact of poor health on academic achievement: New evidence using genetic markers. J HealthEcon.,28(2009),578-97.
[2] YS. Park, BSK. Kim. Acculturation, enculturation, parental adherence to Asian-American college students. Asian Am J Psychol. 1 (2010) 67–79.
[3] S. Pong, J. Johnston, V. Chen. Authoritarian parenting and Asian adolescent school performance Insights from the US and Taiwan. Int J Behav Dev. 34 (2010) 62–72.
[4] M. Brian, J. Galla, Wood ET, Kim HA, Chiu DA. A longitudinal multilevel model analysis of the within-person and between-person effect of effortful engagement and academic self-efficacy on academic achievement. J School Psychol. 24 (2014) 123–30.
[5] F. Guay, CF. Ratelle, A. Roy, D. Litalien. Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. J Learn Individual Diff. 20 (2011) 644–53.
[6] Cohen, L., Manion, L., & Morrison, K. Research Methods in Education (6th ed.). London and New York, NY: Routledge Falmer (2007) sixth edition.
[7] R. Hacihasanoglu, A. Yıldırım, P. Karakur, R. Saglam. Healthy lifestyle behaviour in university students and influential factors in eastern Turkey. Int J Nurs Pract. 17 (2011) 43-51.
[8] M. Rispoli, M O'Reilly, R. Lang, W. Machalicek, T. Davis, G. Lancioni, et al. Effects of motivating operations on problem and academic behaviour in classrooms. J Appl Behav Anal.44 (2011) 187-92.
[9] J. Hauer, T .Quill. Educational needs assessment, development of learning objectives, and choosing a teaching approach. J Palliate Med. 14 (2011) 503–08.
[10] C. Romero and S. Ventura.: Educational data mining: a review of the state of the art, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40 (2010) 601-618.
[11] Alsharhan, Abdulla, Said A. Salloum, and Ahmad Aburayya. "Using e-learning factors to predict student performance in the practice of precision education." Journal of Legal, Ethical and Regulatory Issues 24 (2021) 1-14.
[12] Ong, Mohd Hanafi Azman, and Fadilah Puteh. "Quantitative data analysis: Choosing between SPSS, PLS, and AMOS in social science research." International Interdisciplinary Journal of Scientific Research 3 (2017) 14-25.
[13] Okagbue, Hilary I., et al. "Trends and usage pattern of SPSS and Minitab Software in Scientific research." Journal of Physics: Conference Series. 1734 (2021).
[14] Denis, Daniel J. SPSS data analysis for univariate, bivariate, and multivariate statistics. John Wiley & Sons, (2018) second edition.

[15] K. Bisharat, A. Christ and S. Kröner.: Detrimental effects of an economic crisis on student cognitive achievement–A natural experiment from Palestine, Intelligence, 79 (2020) .
[16] M. M. Arcinas , G. S. Sajja , S. Asif , S. Gour ,E. Okoronkwo, M. Naved.: Role Of Data Mining In Education For Improving Students Performance For Social Change, Turkish Journal of Physiotherapy and Rehabilitation, 32 (2021) 6519-6526
[17] S.N. Kargaonkar, A. Kadam, Review of Applications of Statistics in Data Mining, Conference: Emerging Trends in Computer Science & Application (2016)
[18] L.Z. Javier, L.T.J. Alfonso and R.M, Cristóbal, Early prediction of student learning performance through data mining: A systematic review, Psicothema, 33 (2021) 456-465
[19] S.O.Fonseca and A.A.Namen, Data mining on inep databases: An initial analysis aiming to improve brazilian educational system, Educação em Revista, 32 (2016 ١٣٣-١٥٧ (.
[20] A.Dutt, M.A.Ismail and T.Herawan, A systematic review on educational data mining, Ieee Access, 5(2017) 15991-16005 .
[21] R.Asif, A.Merceron, S.A.Ali and N.G.Haider, Analyzing undergraduate student's performance using educational data mining, Computers & Education, 113 (2017) 177-194
[22] E.Fernandes, R.Carvalho, M.Holanda, G.V.Erven. Educational data mining: Discovery standards of academic performance by students in public high schools in the federal district of brazil, World conference on information systems and technologies. Springer, Cham, 569 (2017) 287–296.
[23] M.Ramaswami, and R. Bhaskaran. "A CHAID based performance prediction model in educational data mining." arXiv preprint arXiv: (2010) 1002.1144 .
[24] Delmas, R., Garfield, J., Ooms, A., & Chance, B . Assessing students' conceptual understanding after a first course in statistics. Statistics education research journal, 6 (2007) 28-58.
[25] Cen, L., Ruta, D., Powell, L., Hirsch, B., & Ng, J. Quantitative approach to collaborative learning: Performance prediction, individual assessment, and group composition. International Journal of Computer-Supported Collaborative Learning, 11 (2016) 187-225
[26] Gbollie, Charles, and Harriett Pearl Keamu. Student academic performance: The role of motivation, strategies, and perceived factors hindering Liberian junior and senior high school students learning. Education Research International (2017)
[27] C.Jalota, R.Agrawal, Analysis of educational data mining using classification, presented at the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, (2019) 243-247
[28] V. Ramesh, P. Parkavi, K. Ramar, Predicting Student Performance: A Statistical and Data Mining Approach, International Journal of ComputerApplications,63(2013).