Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 12, Issue. 3, March 2023, pg.40 – 48

Determining the Relative Importance of Factors Affecting the Selection of High School Students for University Colleges Using Machine Learning Algorithms

Hassan M. Ibrahim[1*]; Ali N. Yousif[2]; Rajaa D. Resen[3]

[1, 2, 3]University of Information Technology and Communications (UOITC), Baghdad, Iraq
* hassan.m@uoitc.edu.iq; alinafea@uoitc.edu.iq; rajaa.alnidway@uoitc.edu.iq

DOI: https://doi.org/10.47760/ijcsmc.2023.v12i03.005

Abstract: Identifying the critical elements that impact students' choice of university majors is critical for leading the nation towards future success. This entails developing a cohort of talented science experts who can pioneer innovations in a variety of industries such as technology, industry, and transportation. To uncover these crucial aspects, this study uses artificial neural networks, a sophisticated analytical technique, to anticipate and assess their relative importance. The study used a descriptive analytical technique, administering a detailed questionnaire to students enrolled in Iraqi universities, resulting in a strong sample size of 238 people. The data was then meticulously analysed using artificial neural networks, a process well-known for its ability to find nuanced patterns and correlations within complicated datasets. The study's findings provide fascinating insights into the factors that influence students' university major selections. The quality of interactions with instructors and classmates during high school is among the most important of these characteristics, demonstrating the substantial influence of mentoring and social dynamics on academic trajectory. Following closely after are thoughts about the economic condition of students' parents, demonstrating the ubiquitous effect of socioeconomic issues on educational decision-making. In summary, this study provides useful information that may be used to guide strategic activities targeted at creating a favourable environment for the development of future leaders. By identifying the fundamental determinants underlying students' university major choices, policymakers and stakeholders may develop focused interventions to assure the nation's preparation for tomorrow's knowledge economy's problems and prospects.

Keywords: neural networks, decision tree, distance learning.

1. Introduction

The state has been interested in development in all areas of life, and among the interests of countries in the first place is education, developing the skills of individuals, and exploiting the entire human energies, as the effects of education are positively reflected on all economic, social and cultural fields, and among the interests of education is interested in higher education, especially in the university stage There are many specializations that people study, it is possible to join the study of arts, languages, science, engineering, nursing, medicine, and others [1]. Iraq is considered like other developed and developing countries alike, pays attention to education, and sets long and short-term plans with the aim of developing education and investing in the human element [2].

The student often faces difficulty in choosing the appropriate university specialization after completing the secondary stage, because the matter of choosing is a fateful thing that many things will depend on him, and the difficulty of choosing increases when there is a desire for the student to enter a specialization, another different desire for the parents, and another opinion about For friends and society, the individual's choice of specialization is considered one of the difficult readings in his life, as he must take into account his inclinations, capabilities, value and academic desires, and the work he wants to practice in the future. Hence this study came to learn about the relative importance of each of the factors affecting the choice of students for university specialization.

1.1 Study Problem

After higher education institutions in Iraq were subjected to several crises over the past few years, which prevented them from keeping up with changes in the demands and growth of society, higher education in Iraq is undergoing a deep shift [3]. One of the most dangerous of these crises is the adoption of a single stereotype by higher education institutions in Iraq through a single set of rules, regulations, procedures, and practices. This led to crowded classrooms and laboratories, which affected the quality of instruction and weakened the outputs, and the labor market lost faith in the majority of institutions' outputs, which led to unemployment; the Ministry of Higher Education had to adopt a new structure in terms of acceptance, absorption, quality, and opening new specialties in a way that suited the labor market to fulfill the objectives of higher education in Iraq [4].

The problem with the study is that the majority of students suffer a lot when choosing the type of academic specialization, as they have many fields of study available that qualify them for different future professions. The decision to choose an academic major is affected by several factors, including factors related to the student himself, factors related to the family, factors related to the teachers and their friends in high school, and factors related to the economic situation of the parents. Since the success of a student's choice in his choice is reflected in his educational career during the university stage, the research problem revolves around determining the relative importance of the factors influencing a student's choice of university major and the extent to which it corresponds to their satisfaction with the progress of their learning process in the majors they have chosen. Since machine learning algorithms are characterized by the ability to classify and determine the relative importance of variables, through the current study we will attempt to use machine learning algorithms to determine the relative importance of each of the factors influencing the selection process.

1.2 Importance of Studying

Making well-informed choices about their academic specialization influences a student's professional trajectory after graduation and influences their university experience. In addition

to improving academic achievement, a well-chosen specialization provides a strong basis for a prosperous career, allowing individuals to actively participate in the growth of society.

The advancement of educational initiatives designed to help students make wise academic decisions is greatly aided by this study. The research provides educational institutions and policymakers with significant information to plan targeted interventions by dissecting the fundamental elements impacting students' decisions regarding their university majors. With the help of these interventions, students should be able to successfully negotiate the challenging terrain of academic specialization.

In the end, this study's findings have the potential to develop a generation of students who are more capable of making thoughtful and calculated choices about their academic endeavors. Students who match their goals, interests, and abilities with appropriate university majors are better equipped to reach their full academic potential and make significant contributions to society's advancement. As a result, the study acts as a catalyst for improving educational results and equipping people to prosper in a world that is becoming more competitive and dynamic.

1.3 Objectives of the Study

The primary aim of this research is to ascertain the relative importance of several elements that influence students' choices of academic specializations in the context of Iraq. The research uses artificial neural networks—a cutting-edge type of machine learning algorithm—to accomplish this aim. These algorithms are effective analytical instruments that can translate complicated facts into meaningful patterns.

The study aims to provide light on the main influences influencing students' choices of academic majors by exploring the complex interactions between various elements. With a greater comprehension of the underlying dynamics driving students' educational decisions, this thorough investigation seeks to give priceless insights into the decision-making processes of Iraqi students.

Furthermore, a more accurate and focused examination is made possible by the use of machine learning algorithms, which allow for a detailed investigation of the relative relevance of different aspects. With this method, the study hopes to aid in the creation of customized interventions and tactics that assist students in choosing their academic paths in a way that is both beneficial and well-informed.

Ultimately, this research aims to provide educators, policymakers, and other stakeholders with the knowledge and resources needed to improve educational outcomes and foster academic success among students in Iraq by disentangling the complex web of factors influencing students' choice of academic majors.

1.4 The Study’s Limitations

- The human field: The study was applied to a sample of students at the University of Baghdad because it contains a large number of university specializations.
- Spatial area: University of Baghdad - Iraq.

2. Related Work

In this section, we will review prior research on students' choice of institutions once they graduate high school. In 2021, Lady D et al. used a machine learning system was to identify parameters influencing academic achievement in Peruvian high school pupils [5]. This study applies supervised machine learning algorithms to investigate factors negatively impacting academic performance among college students on probation at a major public university in

Oman. Using the Knowledge Discovery in Databases (KDD) methodology, data from 6514 students over 11 years (2009-2019) were analyzed. The Information Gain algorithm selected key features, and ensemble methods such as Logit Boost, Vote, and Bagging compared accuracy. Evaluated by metrics like accuracy, precision, recall, F-measure, and ROC curve, and validated via 10-fold cross-validation, results indicated that study duration at university and secondary school performance were top negative predictors. Gender, estimated graduation year, cohort, and academic specialization also significantly contributed. Theoretical and practical implications are discussed, with domain experts and students verifying some findings.

In 2022, Wenhan Ju. et al. study the factors influencing academic performance of university students [6]. This article examines the link between admission chances and GRE/TOEFL scores using the linear regression model and R-square, which are based on the work of Karl Pearson and Sir Francis Galton. The paper evaluates this association by fitting a simple linear regression model using Kaggle data on 386 student scores at an American institution. The examination of 375 graduate students' results shows a positive link between GRE/TOEFL scores and admission chances, as evidenced by R-square values. However, the model has certain disadvantages and limits. The paper explores how to increase model accuracy and reduce mistakes and omissions. Future research will aim on overcoming these restrictions.

In 2023, Masum N. et al. study the factors influencing academic performance of university students [7]. This study looks at the elements that influence university students' academic success, specifically the link between daily lifestyle choices and cumulative grade point average (CGPA). Data from a Google Form survey of a sample of students was analyzed using Python and statistical methods. Key findings show that students with better CGPAs study more hours per day, prefer morning or afternoon study sessions, participate in regular physical activity, and receive enough sleep. Conversely, regular video gaming and some leisure activities are associated with lower CGPA. Attendance, study habits, and academic help have a beneficial influence on academic achievement. Prioritizing excellent study habits and keeping a healthy lifestyle are among the recommendations.

In 2024, Jorge H. et al. suggest a study for predicting college students' academic performance using machine learning methods [8]. This project uses machine learning to predict academic achievement before the conclusion of the first year, with the goal of reinforcing collegiate self-efficacy and increasing student retention. The study used CRISP-DM methods to analyze a dataset of 6690 records that included 21 academic and socioeconomic factors. Preprocessing approaches and classification algorithms were tested, with XGBoost reaching an AUC of 87.75% and accurately categorizing 80% of instances, whilst a decision tree gave interpretability with 10 rules in 70% of cases. The findings highlight the necessity of timely interventions and support mechanisms for student retention. The primary restriction was evaluating algorithm fairness and discrimination. Future study will look at the implications of these results for university management.

3. Neural Network

Artificial Neural Networks (ANNs) are computing systems modelled after biological neural systems, including the human brain. ANNs use a distributed network of neurons to improve their output by learning from the input [9]. Neural networks approach problem-solving differently from traditional computers. Conventional computers tackle problems using an algorithmic technique, which involves following instructions. Computers cannot solve problems without specified procedures to follow. Conventional computers can only address issues that humans are familiar with [10].

Nowadays, ANN applications have gained popularity in a variety of human demands. Many organisations are investing in neural networks to tackle issues in many industries and the economy, which have historically been the job of operations research [11]. Because of its broad applicability, artificial intelligence is frequently recommended for data analysis by academics in social science and the arts, in addition to its utility in science and engineering [12]. Several methods are used in machine learning to address data-related issues. The optimum kind of algorithm to tackle a problem is not a one-size-fits-all solution, as data scientists like to point out. The type of method used relies on the type of issue you want to answer, how many variables there are, what kind of model works best for it, and other factors [13].The machine learning job of learning a function that translates an input to an output using exemplar input-output pairs is known as supervised learning. From labelled training data, which consists of a collection of training instances, it infers a function. The algorithms for supervised machine learning are those that require outside help. The train and test datasets are separated from the input dataset. An output variable from the train dataset has to be categorised or forecasted [14].In several domains, artificial neural networks (ANN) have shown to be an excellent tool for prediction, knowledge discovery, classification, time series analysis, modelling, and other purposes. Supervised learning, reinforcement learning, and unsupervised learning are the three categories under which ANN training falls. The use of supervised learning has some restrictions. The application of unsupervised learning techniques can help get over these restrictions [15].Unsupervised learning is a machine learning technique where computers only learn patterns from unlabeled data, as opposed to supervised learning. In this method, a machine learning model searches the data on its own for any patterns, structure, similarities, or contrasts. Human intervention is not required beforehand [16].

4. Methodology

Follow the descriptive analytical approach by describing everything related to the variables that have been studied and then analyzing the data obtained. The questionnaire was used as a tool to collect data from the studied community that represents university students in Iraq. The sample included 238 students from different university majors. artificial neural networks will be used in data analysis. We will use the program spss 22.

The amount of data is increasing exponentially, making it necessary to analyze such a massive amount of data to extract usable information from it. As a result, the field of data mining was created. Data mining is the process of extracting knowledge from a vast number of datasets [17]. Data mining tasks can be divided into two categories: Clustering and association rules are examples of descriptive tasks, whereas time series analysis, classification, and regression are examples of predictive jobs (Joshi et al., 2010). Data mining has numerous applications in a variety of industries, including banking, insurance, medical, finance, marketing, healthcare, and sales [18]. To determine an object's class, classification is a data mining technique. It effectively illustrates guided learning. The categorization (discrete, ordered) for the categorical label is predicted. The classification of data involves two processes. During the learning (or training) phase of the first phase, a classifier is created to characterize a preset set of data classes.

In the second stage, unclassified data is categorized using the first-stage model, and the accuracy of the classifier is estimated using test data. Among the many classification techniques available are decision trees, K nearest neighbors, naive Bayesian classifiers, and artificial neural networks [19]. To address categorization problems, machine learning techniques like artificial neural networks are used. It is a subset of artificial intelligence, or Artificial Neural Networks (ANNs), and is frequently referred to as the neural network model. Instead of setting an algorithm to carry out specific activities, the ANN learns the

system to carry out the mission to accomplish such synthetic jobs. The AI system is being developed as an operational model that may uncover hidden patterns in data that accurately and quickly replicate relevant knowledge. Given that AI systems should learn from data regularly in fields like medical diagnostics employing diverse data, artificial intelligence technologies are the best tools now accessible. Neural networks are one example of an AI scheme [20]. The various artificial neurons that make up an artificial neural network are connected by the network geometry specified during design. To transform inputs into significant outputs is the goal of a neural network. Unsupervised or supervised learning placements are both possible. When there is noise present, neural networks learn [21].

Artificial neural networks can be defined as Haykin defined them: A massively parallel distributed processor made up of simple processing units, known as a neural network, has a built-in predisposition to store and make use of experiential information. It is similar to the brain in two ways [22]:

- Through a learning process, the network picks up knowledge from its surroundings.
- Acquired knowledge is stored via synaptic weights, which are the strengths of interneuron connections.

5. Result

Proposed Artificial Neural Network Model: The questionnaire was divided into 4 factors affecting students' choice of scientific specialization (X1: factors related to the student, X2: factors related to teachers and high school friends, X3: factors related to the economic level of parents, X4: factors related to parents). We will predict the size of their impact or the percentage of their importance to the individual through artificial neural networks according to the following stages:

- **Description of the artificial neural network model and the division of the study sample:** The process of segmenting the study sample is considered one of the most important stages of building the model of artificial neural networks, as the strength of the model depends on the stability of the results between the test sample and the training sample. At this stage, the study sample was divided according to the table (1):

Table (1): Division of the Study Sample

		N	Percent
Sample	Training	177	74.4%
	Testing	61	25.6%
Valid		238	100.0%
Excluded		0	
Total		238	

From Table (1), we find that the training sample includes 177 viewers, or 74.4%, and the test sample includes 61 viewers or 25.6%.

- **Building the database:** The data has been converted to standard values instead of their natural values to avoid any distortion in the distribution of data that may lead to heterogeneity of the input values and thus a high average square error.
- **Determining the criteria and structure of the study model:** At this stage, the inputs and outputs of the model, the number of hidden layers, and the number of nodes were determined.

The layers were divided into: The first layer: the inputs, which are the variables of the study, as it included 4 nodes. The hidden layer contained one node per layer. The output layer included one node representing student satisfaction with the university major they are studying, which included 2 cases (Y: Satisfied, not satisfied). As shown in table (2).

Table (2): Network Structure

Input Layer	Covariates	1	x1
		2	x2
		3	x3
		4	x4
	Number of Units		4
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		1
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	y
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

- The network was trained several times to reach an appropriate model that provides accurate predictions. Table (3) shows the classification of students' satisfaction according to the proposed neural network.

Table (3): The Classification of Students' Satisfaction According to the Proposed Neural Network

		Predicted		
Sample	**Observed**	**Satisfied**	**Not Satisfied**	**Percent Correct**
Training	Satisfied	138	4	97.2%
	Not Satisfied	26	9	25.7%
	Overall Percent	92.7%	7.3%	83.1%
Testing	Satisfied	51	1	98.1%
	Not Satisfied	4	5	55.6%
	Overall Percent	90.2%	9.8%	91.8%

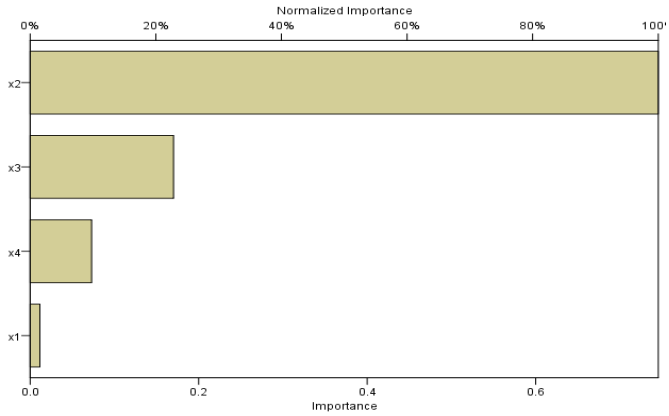
From the previous table, we find that the model:

- In the training phase, the model succeeded in classifying 138 of the first cases, while it failed to classify 4 cases, while it misclassified 26 viewers in the second case, and succeeded in classifying 9 cases.
- In the testing phase, the model succeeded in classifying 51 of the first cases, while it failed to classify 1 case, while it misclassified 4 viewers in the second case, and succeeded in classifying 5 cases.

- The relative importance of the study variables (factors affecting the choice of a university major): Table (4) and Figure (1) show the relative importance of the factors affecting the choice of a university major according to the proposed model.

Table (4): Relative Importance of the Study Variables

	Importance	Normalized Importance
x1	.011	1.5%
x2	.746	100.0%
x3	.170	22.8%
x4	.073	9.8%



From the table (4) we find that:

- The factor that gained the greatest importance, which was estimated at 100%, was the variable " factors related to teachers and high school friends ".
- The variable " factors related to the economic level of parents " came in second place, as it was of importance to the study sample by a percentage of 22.8%.
- In the last two ranks came " factors related to the parents " with a percentage of 9.8% and " factors related to the student " with a percentage of 1.5%.

To evaluate the accuracy of classification, we will use the press's Q test. To perform the test, the hypotheses are:

Primary hypothesis: The classification generated by the Artificial Neural Network has no statistical significance.

Alternative hypothesis: The classification generated by the Artificial Neural Network algorithm is statistically significant.

$$pressQ = \frac{[238-(203\times 2)]^2}{238(2-1)} = 118.89$$

By comparing the previous value with the tabular value of 4.51, we reject the null hypothesis and acknowledge that the classification resulting from the model is statistically significant.

Conclusion

The research dealt with determining the relative importance of the factors influencing student's choice of university majors in Iraq using artificial neural networks. Through the practical study, the research was able to determine the relative importance of the factors, as

the accuracy of prediction was determined by classifying the extent of students' satisfaction with the majors they are studying, and we obtained high prediction accuracy (greater than 83%) in both the testing phase and the training phase. The factor that gained the greatest importance, which was estimated at 100%, is the variable "factors related to teachers and secondary school friends." The variable "factors associated with the economic level of parents" came in second place, as its importance for the study sample was 22.8%. In the last two places were "family-related factors" with a rate of 9.8%, and "student-related factors" with a rate of 1.5%. It was found through testing the hypotheses that there is no statistically significant difference between classification according to artificial neural networks and students' satisfaction with the majors they are studying.

References

[1]. MILTON, Sansom. Higher education and post-conflict recovery. Springer, 2017.

[2]. BADRY, Fatima; WILLOUGHBY, John. Higher education revolutions in the Gulf: Globalization and institutional viability. Taylor & Francis, 2016.

[3]. AL-JANABI, Sufyan T. Faraj; ANDERSON, David. Reforming the Higher Education System in Iraq: Internationalization via Decentralization. Journal of Educational Technology, 2011, 8.3: 1-14.

[4]. BEDAN, Saddam. Higher Education in Iraq between the Past and the Present. Journal of Positive School Psychology, 2022, 4718-4731.

[5]. Lady Denisse Infante Acosta, Ing., Jonatán Edward Rojas Polo, Mg. Identification of factors that affect the academic performance of high school students in Peru through a machine learning algorithm. 19th LACCEI International Multi-Conference for Engineering, Education, and Technology, 2021.

[6]. JU, Wenhan; WANG, Heran; YE, Zi. Analysis of Relative Importance of Factors Affecting the Chance of Admission of Applying to Graduate Students. Journal of Education, Humanities and Social Sciences, 2022, 6: 99-109.

[7]. HASAN, Masum N.; AMEEN, Pranto F.; JOHN, Gomes A. Factors Affecting the Academic Performance of University Student, 2023.

[8]. GUANIN-FAJARDO, Jorge Humberto; GUAÑA-MOYA, Javier; CASILLAS, Jorge. Predicting Academic Success of College Students Using Machine Learning Techniques. Data, 2024, 9.4: 60.

[9]. O'SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458, 2015.

[10]. MAIND, Sonali B., et al. Research paper on basic of artificial neural network. International Journal on Recent and Innovation Trends in Computing and Communication, 2014, 2.1: 96-100.

[11]. ABIODUN, Oludare Isaac, et al. State-of-the-art in artificial neural network applications: A survey. Heliyon, 2018, 4.11.

[12]. BIHL, Trevor J.; YOUNG II, William A.; WECKMAN, Gary R. Artificial neural networks and their applications in business. In: Encyclopedia of Information Science and Technology, Fourth Edition. IGI Global, 2018. p. 6642-6657.

[13]. MAHESH, Batta. Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 2020, 9.1: 381-386.

[14]. JIANG, Tammy; GRADUS, Jaimie L.; ROSELLINI, Anthony J. Supervised machine learning: a brief primer. Behavior therapy, 2020, 51.5: 675-687.

[15]. DIKE, Happiness Ugochi, et al. Unsupervised learning based on artificial neural network: A review. In: 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS). IEEE, 2018. p. 322-327.

[16]. TYAGI, Kanishka, et al. Unsupervised learning. In: Artificial intelligence and machine learning for edge computing. Academic Press, 2022. p. 33-52.

[17]. KNERR, Stefan; PERSONNAZ, Léon; DREYFUS, Gérard. Handwritten digit recognition by neural networks with single-layer training. IEEE Transactions on neural networks, 1992, 3.6: 962-968.

[18]. AGATONOVIC-KUSTRIN, S.; BERESFORD, Rosemary. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 2000, 22.5: 717-727.

[19]. BOURLARD, Herve; MORGAN, Nelson. Continuous speech recognition by connectionist statistical methods. IEEE Transactions on Neural Networks, 1993, 4.6: 893-909.

[20]. ABD ALI, Doaa Mohsin; CHALOB, Donia Fadil; KHUDHAIR, Ameer Badr. Networks data transfer classification based on neural networks. Wasit J Comput Math Sci, 2022, 1.4: 207-225.

[21]. YOGITHA, R.; MATHIVANAN, G. Performance analysis of transfer functions in an artificial neural network. In: 2018 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2018. p. 0393-0397.

[22]. HAYKIN, Simon. Neural networks and learning machines, 3/E. Pearson Education India, 2009.

