

Article

Determination of the Factors Affecting King Abdul Aziz University Published Articles in ISI by Multilayer Perceptron Artificial Neural Network

Rashad A. R. Bantan ^{1,2}, Ramadan A. Zeineldin ^{2,3}, Farrukh Jamal ^{4,*}  and Christophe Chesneau ⁵

¹ Department of Marine Geology, Faculty of Marine Sciences, King Abdul Aziz University, Jeddah 21551, Saudi Arabia; rbantan@kau.edu.sa

² King Abdul Aziz University, Jeddah 21551, Saudi Arabia; rzainaldeen@kau.edu.sa

³ Faculty of Graduate Studies for Statistical Research, Cairo University, Giza 12613, Egypt

⁴ Department of Statistics, Government Postgraduate College Dera Nawab Sahib, Punjab 63351, Pakistan

⁵ Department of Mathematics, LMNO, Campus II, Science 3, Université de Caen, 14032 Caen, France; christophe.chesneau@unicaen.fr

* Correspondence: drfarrukh1982@gmail.com; Tel.: +92-3216802194

Received: 22 February 2020; Accepted: 6 May 2020; Published: 11 May 2020



Abstract: Deanship of scientific research established by the King Abdulaziz University provides some research programs for its staff and researchers and encourages them to submit proposals in this regard. Distinct research study (DRS) is one of these programs. It is available all the year and the King Abdulaziz University (KAU) staff can submit more than one proposal at the same time up to three proposals. The rules of the DSR program are simple and easy so it contributes in increasing the international rank of KAU. The authors are offered financial and moral reward after publishing articles from these proposals in Thomson-ISI journals. In this paper, multiplayer perceptron (MLP) artificial neural network (ANN) is employed to determine the factors that have more effect on the number of ISI published articles. The proposed study used real data of the finished projects from 2011 to April 2019.

Keywords: prediction; artificial neural network; research program; research productivity

1. Introduction

Deanship of scientific research (DSR) at King Abdulaziz University (KAU) supports staff members to develop high quality research in different scientific fields by providing some funded research programs. Among these research programs, there is the distinct research study (DRS). DRS program in KAU was initiated in 1432 h (2011) to support university staff and researchers to increase the productivity of high-quality research through publishing articles in Thomson-ISI Journals. Since the DSR program has been initiated, there are a lot of research articles published in Thomson-ISI Journals in different areas. Most of these articles are conducted by some academic departments in the KAU such as chemistry, mathematics, physics, statistics, etc., while other departments submit few projects. Publish articles in ISI journals has many advantages for the researchers; they can get a ward from their universities, financial incentives, and their work will be marketed in the international scientific community, and to be promoted to higher degree in their institutes or faculties. From the literature, we found that the research in this area is scares and the existing research comes from either questionnaire or from literature review. In this study, real data from the database of the DSR program at deanship of scientific research at KAU are analyzed. Due to their ability to solve complex numerical problem, we

focus our attention on artificial neural network methods, which are well-known to be suitable when other computational and statistical methods fall [1].

More precisely, in this research, distinct research study ISI published articles data is analyzed and multilayer perceptron (MLP) artificial neural network (ANN) is developed to determine, which variables have more effect on the number of ISI published articles and the rank of the journal. ANN has some advantages such as its ability to store information on the entire network, have a distributed memory and parallel processing capability. Additionally, ANN is suitable to handle our problem because it has two dependent variables; one of them is numerical and the other is string. The paper is structured as follows. Distinct research program is described in Section 2. An overview about ANN is presented in Section 3, Section 4 is assigned to the developed ANN, Section 5 is assigned to ANN results and conclusions and future research are discussed in Section 6.

2. Distinct Research Program (DRS)

Distinct research program is one of the research programs presented from DSR at KAU. This program is running over the year and all staff and researchers of KAU can submit on proposal or more. After publishing an article in ISI from each proposal, they got valuable financial incentives. All data of researchers, articles and journals are stored in deanship of scientific research database. So, data used in this research are got from the database, we focused on the finished projects that resulted in publishing articles in Thomson ISI and, at the same time, the articles information appearing on Web of Science from 2012 to April 2019.

Data include faculty name, department, authors (KAU staff members or external coauthors), journal category (means the area of research that the journal covers), project duration (the time between date of contract and data of publishing the article), budget of the project (it is divided into initial deposit, which is given when signing the contract and the final deposit, which is given after publishing the article and satisfying the rules of the distinct research study (DRS)), and contract category (which means the journal rank; Q1, Q2, Q3, and Q4, where the final deposit depends on this rank).

KAU has more than 70 faculties, centers, institutes and deanships; all these units have staff and researchers, and they can submit proposals for the DRS program.

Analyzing the data, it was found that the published articles were attended by the faculties of science, engineering, pharmaceuticals, computer engineering, and information technology, as well as the faculties of meteorology, environment, and agriculture with 77% of all the faculties, centers, and deanships. Additionally, it is found that departments of chemistry, mathematics, physics, statistics, pharmaceuticals, biochemistry, production engineering and mechanical system design, computer science, and electrical and computer engineering have participated in the published articles with 60% of all departments of the faculties, centers, and deanships. Regarding the rank of the journals, it was found that of published articles 7% of Q1, 14% of Q2, 18% of Q3, and 22% of Q4, and 40% of Q5.

Few studies in the literature tried to determine the factors affecting number of published research or research productivity. A brief state of the art is formulated below. The reference [2] developed a study to identify the factors affecting numbers of citations for papers in medicine based on the evidence-based approach. They found that journal rank subject, IF, h index of the authors, the notoriousness of the main author, SNIP (source normalized), and SJR (scientific journal ranks) had significantly positive correlations with paper citability. The reference [3] developed a study based on researcher-made questionnaire about study factors impacting on the research productivity of Iranian women in ISI. The reference [4] conducted a literature review and designed a questionnaire to determine the most influencing factors on the research performance of the university professor at different universities in Taiwan. They found that the most important factors are feeling in work, research funding, hardware climate, hardware and facility, human resources, and library resources. The reference [5] used questionnaires, interview and document analysis to analyze resource, institutional, and cultural factors that influence research productivity in Mwenge Catholic University in Tanzania. He found that lack of the staff, low salaries, teaching workload, not enough research budget, and the inadequate

computer application software for data analysis and plagiarism are the most factors affecting research productivity. The reference [6] conducted a study in 14 institutions under the control of university of agricultural sciences. They found that human factors (organizational, sociopsychological, and personal) and organizational related factors have an effect on research productivity. The reference [7] identified the opportunities and barriers that faculty members at Najran University can face when publishing in ISI journals. The reference [8] examined research productivity of faculty at two major Kenyan public universities. The reference [9] determined factors affecting the research productivity and the factors importance using statistical package LISREL (the linear structural relationship) and neural network analyses. The reference [10] found that financial incentives motivate the improving of the research productivity from the Scopus data base study of the assignment of Russian authors' affiliations to the branches of Russian science.

3. An Overview about ANN

Artificial neural networks (ANNs) are widely used in various fields and applications. The following overview provides some applications. The reference [11] developed a comprehensive study of ANNs, with discussions on the applications and contributions. ANN is used in various fields such as computer security, medicine, business, finance, banking, insurance, the stock market, power generation, management, the nuclear industry, mineral exploration, mining, and forecasting the quality of oil fractions. The reference [12] presented a study of neural networks of direct communication with random weights. The reference [13] reviewed ANNs uses in energy. The reference [14] presented a survey on randomized methods for training ANNs. The reference [15] investigated random single-hidden layer feedforward neural network based on metaheuristic and non-iterative learning approach. Application of recurrent ANNs in the field of statistical language modeling can be found in [16]. There are extensions and other developed types of ANN such as projection neural network. The reference [17] provided a comprehensive review of the projection neural networks for solving various constrained optimizations. Additionally, the reference [18] presented Zeroing neural networks to solve complex computation equations. The reference [19] applied ANN ensemble for the mapping of landslide susceptibility where this type has feature of higher accuracy.

4. Implementation of MLP ANN

IBM SPSS 25 was used to develop the proposed MLP neural network to determine the most important factors affects publishing in ISI journals. Table 1 shows the case processing summary; where the percentage of the included cases was 91.0% and the excluded cases was 9.0%. From the included cases 80.0% of the sample was employed for training and 20.0% was used for testing.

Table 1. Case processing summary.

		Percent	
Sample	Training	80.0%	
	Testing	20.0%	
	Valid	91.0%	100.0%
	Excluded	9.0%	
	Total	100.0%	

The proposed multilayer perceptron ANN structure had one input layer, one output layer, and two hidden layers. Table 2 shows MLP ANN information and it was discussed in the subsections. Two hidden layers were used because the data of the problem addressed was deterministic and two layers were enough for the NN to be learned and to find the most important factors affecting publication in ISI.

Table 2. Multilayer perceptron (MLP) artificial neural network (ANN) information.

Input Layer	Factors	1	Faculty
		2	Department
		3	JournalCateg
	Covariates	1	Budget
		2	InKAUStaff
		3	OutKAUStaff
		4	ProjDuration
Number of Units		363	
Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number		2
	Number of Units in 1 ^a		1
	Number of Units in 2 ^a		1
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	PublishYear
		2	ContractCateg
	Number of Units		6
	Rescaling Method for Scale Dependents		Adjusted normalized
	Activation Function		Hyperbolic tangent
Error Function		Sum of Squares	

^a The bias unit is excluded.

4.1. The Input Layer

The variables in the input layer were seven variables in two types; numeric variables (called covariates in SPSS) and nonnumeric variables (called factors in SPSS). The numeric variables were: budget of research, number of staff from KAU participated in these articles as principle authors or coauthors, number of staff out of KAU participated in these articles as principle authors or coauthors, and project duration (it is time between the contract date and the date of publishing the article on the web of science). The nonnumeric variables were the faculty of the author(s), the department at the faculty that the author(s) belongs to, and journal category that means the journal research area. The variables in the input layer were the independent variables. The output layer had two dependent variables; the first one was the contract category, which means the class of the journal where the article was published in A, B, C, D, or E, and the second variable was the number of published articles every year. The rescaling method for numerical variables was standardized. The ratio assigned to the three samples was 70% for training, 30% for testing, and 0% for holdout.

4.2. The Hidden Layers

We determined the number of hidden layers in layer 2, where the hidden layer contained network nodes (units). The first hidden unit is a function of the weighted sum of the input data, and the second hidden layer is a function of the weighted sum of units in the first hidden layer, and the activation function is the same for two hidden layers. The number of units in each hidden layer was automatically determined by SPSS.

The activation function used in this research was the hyperbolic tangent of the form:

$$f(u) = \frac{1 - e^{-\lambda u}}{1 + e^{-\lambda u}} \tag{1}$$

where λ is a scale parameter. For the output layer, which contains the dependent variables, the activation function is a hyperbolic tangent.

4.3. Types of Training

There were three types of training: batch, online, and mini-lot. The difference between the three types is the updating synaptic scales. In a batch training type, synaptic weights were updated after all training data records were completed. In an online workout, synaptic weights were updated after each workout data record. In the mini-lot type, the training data records were divided into groups of the same size and then update the synaptic weights after passing through the group. We tried the three types of training. For batch training type, we used the scaled conjugate gradient method to estimate the synaptic weights.

4.4. The Output Layer

The output layer contained two variables: the number of published articles and rank of the journals that the articles were published in. The activation function is defined by the hyperbolic tangent function. The error function is given by the sum of squares.

5. ANN Results

After determining the input and output variables and featuring the MLP ANN, the results are shown in Table 3.

Table 3. ANN model summary.

	Sum of Squares Error	185.885
	Average Overall Relative Error	0.766
Training	Percent Incorrect Predictions for Categorical Dependents	ContractCateg 43.4%
	Relative Error for Scale Dependents	PublishYear 0.999
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	0:00:00.22
Testing	Sum of Squares Error	42.322
	Average Overall Relative Error	0.748
	Percent Incorrect Predictions for Categorical Dependents	ContractCateg 44.4%
	Relative Error for Scale Dependents	PublishYear 0.978

^a The testing sample is used to determine the error computations.

Where the model summary shown in Table 3 displays a summary of the neural network results by partition and overall, including the error, the relative error or the percentage of incorrect predictions, the stopping rule used to stop training, and the training time. The error is the sum-of-squares error when the identity, sigmoid, or hyperbolic tangent activation function is applied to the output layer. It is the cross-entropy error when the softmax activation function is applied to the output layer.

Since one of the output variables (ContractCateg) is categorical, Table 4 shows the classification of this variable by sections and general parameters. This table shows the number of cases rightly or wrongly classified for a dependent variable of type ContractCateg. A percentage of the total number of cases that were classified correctly was also reported.

Table 4. Classification of ContractCateg variable.

Sample	Observed	Predicted					Percent Correct
		A	B	C	D	E	
Training	A	0	0	0	32	0	0.0%
	B	0	0	0	64	2	0.0%
	C	0	0	0	76	3	0.0%
	D	0	0	0	90	24	78.9%
	E	0	0	0	2	175	98.9%
	Overall Percent	0.0%	0.0%	0.0%	56.4%	43.6%	56.6%
Testing	A	0	0	0	8	0	0.0%
	B	0	0	0	18	0	0.0%
	C	0	0	0	19	4	0.0%
	D	0	0	0	16	3	84.2%
	E	0	0	0	0	49	100.0%
	Overall Percent	0.0%	0.0%	0.0%	52.1%	47.9%	55.6%

Dependent Variable: ContractCateg.

Figure 1 shows a graph predicted by the observed. It displays grouped rectangles with predicted pseudo-sequences for combined training and test samples for categorical dependent variables from ContractCateg. The x axis is for the observed response categories, and the legend is for the predicted categories.

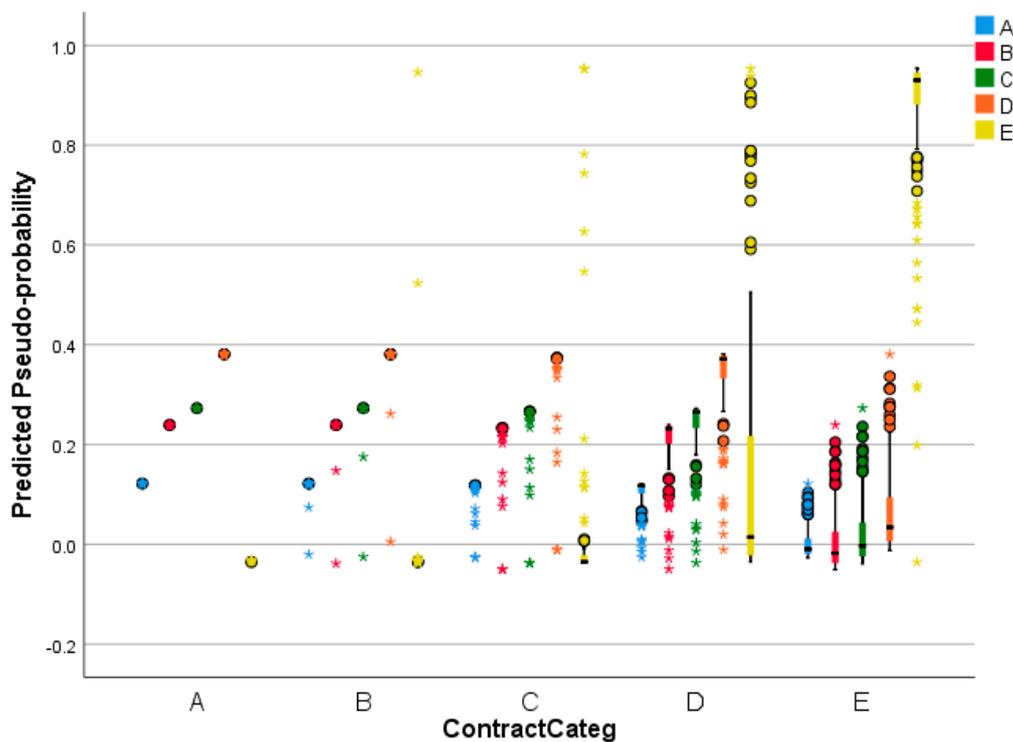


Figure 1. Predicted by the observed chart for ContractCateg.

Figure 2 shows a graph predicted by observables, in which this graph displays a graph of the variance of values predicted on the y axis over values observed on the x axis for combined training and testing samples for the dependent variable on the publish year scale.

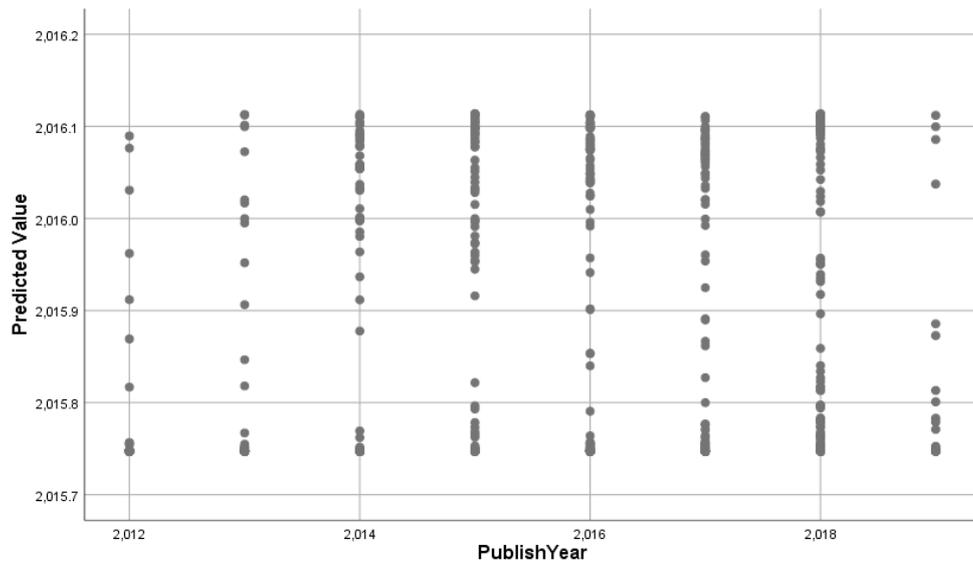


Figure 2. Predicted by the observed chart for the publish year variable.

Figure 3 shows the ROC curve (receiver performance) for the categorical dependent variable ContractCateg. It displays a curve for each category. Where the ROC is graph showing the performance of a classification model at all classification thresholds and serves to visualize and analyze the relationship between one or two performance measures and the threshold as well.

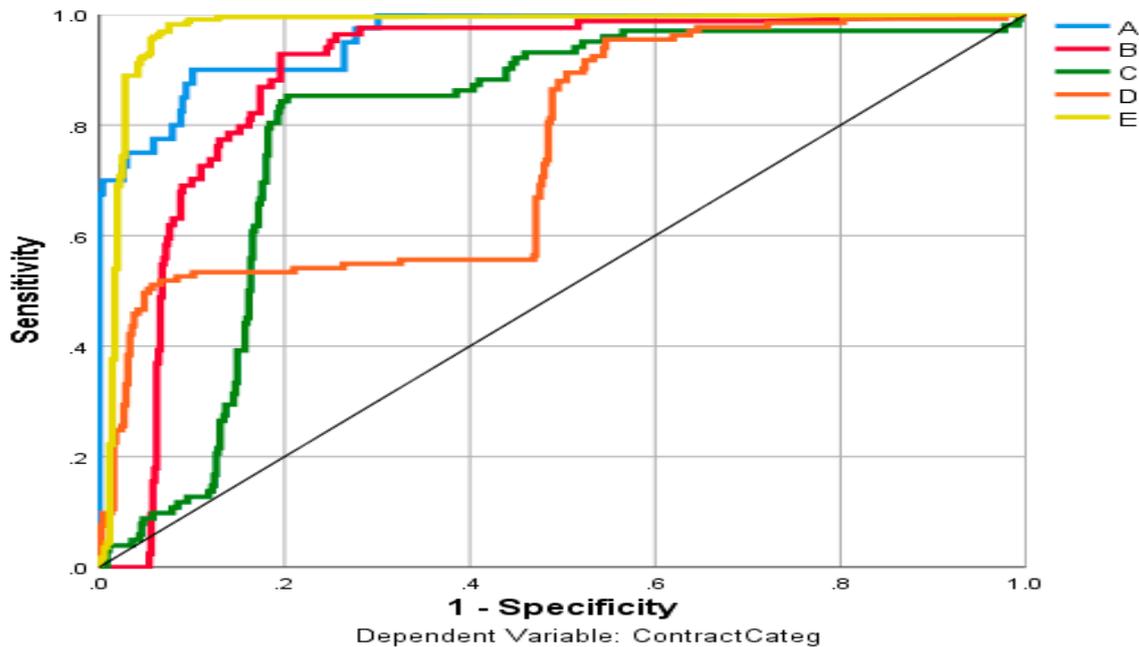


Figure 3. ROC curve for the ContractCateg variable.

Figure 4 displays a graph of total revenue for the dependent variable ContractCateg. It shows a percentage of the total number of cases in this “acquired” category, segmenting a percentage of the total number of cases.

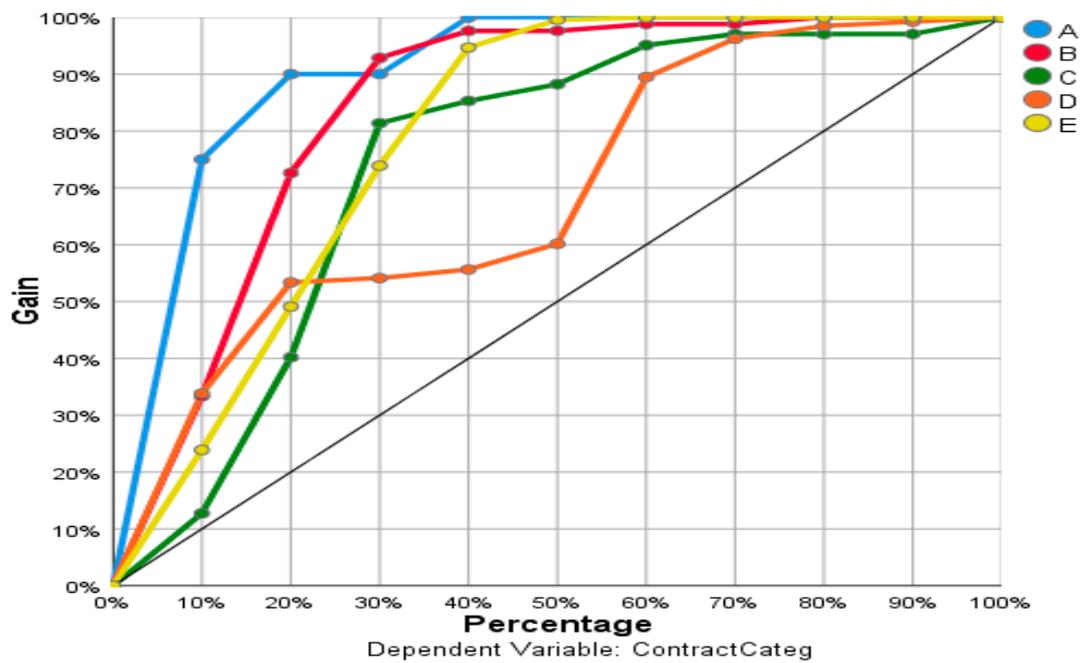


Figure 4. Cumulative gains chart for ContractCateg.

Figure 5 shows a height graph for the categorical dependent variable ContractCateg.

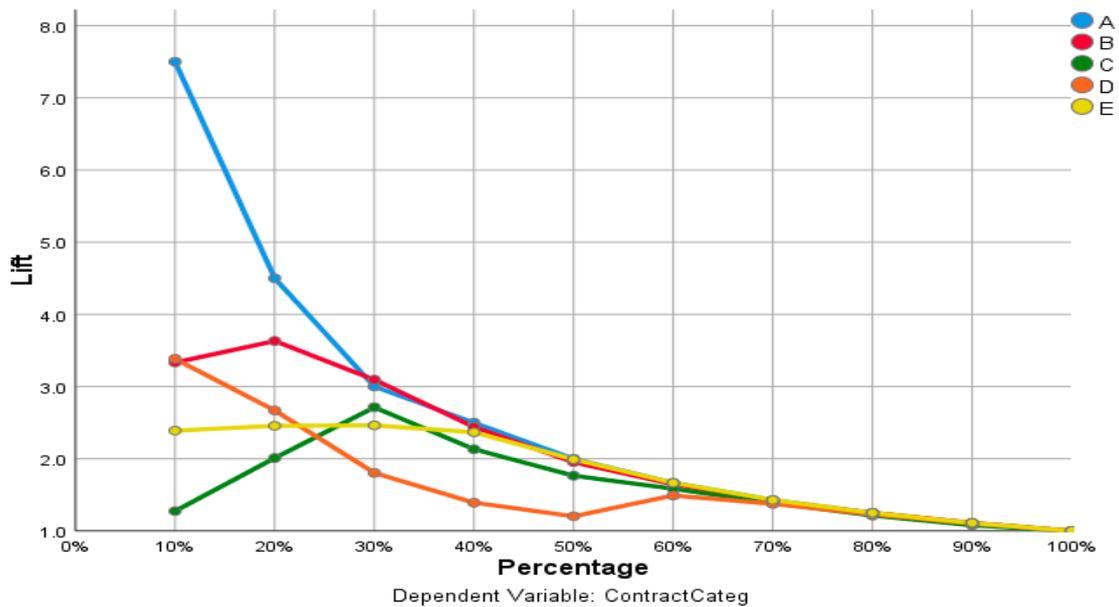


Figure 5. Lift chart for ContractCateg variable.

Table 5 indicates the area under each curve for the dependent variable ContractCateg.

Table 5. Area under the curve for ContractCateg.

ContractCateg	Area
A	0.958
B	0.889
C	0.795
D	0.752
E	0.975

Table 6 presents an independent variable importance analysis (the factors weights and their rank). Thus, a sensitivity analysis is performed, showing the significance of each predictor by the determination of the neural network.

Table 6. Factors weights and rank according to training type.

Factors	Batch		Online		Mini-Batch		Weights Average	Rank
	Weight	Rank	Weight	Rank	Weight	Rank		
Budget	0.520	1	0.543	1	0.461	1	0.508	1
ProjDuration	0.098	2	0.086	4	0.177	2	0.120	2
Department	0.090	3	0.097	3	0.100	4	0.096	3
OutKAUStaff	0.078	4	0.109	2	0.040	6	0.076	6
JournalCateg	0.077	5	0.063	6	0.104	3	0.081	4
InKAUStaff	0.069	6	0.039	7	0.018	7	0.042	7
Faculty	0.068	7	0.063	5	0.099	5	0.077	5

From Table 6, budget is the most important factor affecting the number of published articles and the class of the journal for all the training types.

When calculating the average of weights for the three types of training, the order of the ranks that affect publishing articles was as follows: budget, project duration, department, journal category, faculty, staff from outside the KAU, and finally staff from inside the KAU.

6. Conclusions and Future Research

In this a study, a multilayer perceptron ANN was developed to determine the factors affecting publishing articles of distinct research study program in KAU and their weights. The factors taken into consideration in this study such as project duration, number of inside, and number of outside staff were not considered in many similar studies.

The results show that the budget, as most of the similar research studies in this area, was the most important factor and the three different training types gave the same rank for the budget. The second important factor was the project duration, which is the available time to complete the project and publish the article in ISI journal. Two types of training gave this factor rank 2 while the third training approach ranked it 4. The third important factor is the department that the submitted project belongs to it. Two training methods ranked the department 3 while the third training method ranked it 4.

ANN was selected to handle this problem because it has the ability to store information on the entire network and parallel processing capability. Additionally, ANN is more suitable to handle our problem because we had two dependent variables; number of publications per year, which is numerical, and the contract category (A, B, C, D, or E), which is ordinal (string).

Determining the factors affecting publishing in ISI journals needs further large-scaled research that can handle all different research areas. This study can be extended to include the staff scientific rank and their number at each rank. It is recommended to use a neural network merged with other metaheuristics approaches such as genetic algorithm and particle swarm in to increase the performance of NN. Additionally, the problem can be dealt as a fuzzy problem for some variables.

Author Contributions: R.A.R.B., F.J., C.C., and R.A.Z. have contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Deanship of Scientific Research (DSR), King Abdul Aziz University, Jeddah, under grant no. (D-006-150-1439).

Acknowledgments: We warmly thank the three reviewers for their thorough and constructive comments. This work was funded by the Deanship of Scientific Research (DSR), King Abdul Aziz University, Jeddah, under grant no. (D-006-150-1439). The authors gratefully acknowledge the DSR technical and financial support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Khaze, S.R.; Masdari, M.; Hojjatkah, S. Application of artificial neural networks in estimating participation in elections. *Int. J. Inf. Technol. Model. Comput.* **2013**, *1*, 23–31.
2. Yaminfirooz, M.; Ardali, F.R. Identifying the Factors Affecting Papers' Citability in the Field of Medicine: An Evidence-based Approach Using 200 Highly and Lowly-cited Papers. *Acta Inf. Med.* **2018**, *26*, 10–14. [[CrossRef](#)] [[PubMed](#)]
3. Isfandyari-Moghaddam, A.; Hasanzadeh, M.; Ghayoori, Z. A study of factors affecting research productivity of Iranian women in ISI. *Scientometrics* **2012**, *91*, 159–172. [[CrossRef](#)]
4. Yang, J.C.-C. A Study of Factors Affecting University Professors' Research Output: Perspectives of Taiwanese Professors. *J. Coll. Teach. Learn.* **2017**, *14*, 11–20. [[CrossRef](#)]
5. Okendo, O.E. Constraints of research productivity in universities in Tanzania: A case of Mwenge catholic university, Tanzania. *Int. J. Educ. Res.* **2018**, *6*, 201–210.
6. Manjunath, L.; Shashidhara, K.K. Determinates of Scientific Productivity of Agricultural Scientists. *Indian Res. J. Ext. Educ.* **2011**, *11*, 7–12.
7. Omer, R.A. International Scientific Publication in ISI Journals: Chances and Obstacles. *World J. Educ.* **2015**, *5*, 81–90. [[CrossRef](#)]
8. Nafukho, F.M.; Wekullo, C.S.; Muyia, M.H. Examining research productivity of faculty in selected leading public universities in Kenya. *Int. J. Educ. Dev.* **2019**, *66*, 44–51. [[CrossRef](#)]
9. Wichian, S.N.; Wongwanich, S.; Bowarnkitiwong, S. Factors Affecting Research Productivity of Faculty Members in Government Universities: Lisrel and Neural Network Analyses. *Kasetsart J.-Soc. Sci.* **2009**, *30*, 67–78.
10. Kosyakov, D.; Guskov, A. Impact of national science policy on academic migration and research productivity in Russia. *Procedia Comput. Sci.* **2019**, *146*, 60–71. [[CrossRef](#)]
11. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2008**, *4*, e00938. [[CrossRef](#)] [[PubMed](#)]
12. Cao, W.; Wang, X.; Ming, Z.; Gao, J. A review on neural networks with random weights. *Neurocomputing* **2018**, *275*, 278–287. [[CrossRef](#)]
13. Mohandes, S.R.; Zhang, X.; Mahdiyar, A. A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing* **2019**, *340*, 55–75. [[CrossRef](#)]
14. Zhang, L.; Suganthan, P.N. A survey of randomized algorithms for training neural networks. *Inf. Sci.* **2016**, *364*, 146–155. [[CrossRef](#)]
15. Han, F.; Jiang, J.; Ling, Q.-H.; Su, B.-Y. A survey on metaheuristic optimization for random single-hidden layer feedforward neural network. *Neurocomputing* **2019**, *335*, 261–273. [[CrossRef](#)]
16. Mulder, W.D.; Bethard, S.; Moens, M.-F. A survey on the application of recurrent neural networks to statistical language modeling. *Comput. Speech Lang.* **2015**, *30*, 61–98. [[CrossRef](#)]
17. Jin, L.; Li, S.; Hu, B.; Liu, M. A survey on projection neural networks and their applications. *Appl. Soft Comput. J.* **2019**, *76*, 533–544. [[CrossRef](#)]
18. Jin, L.; Li, S.; Liao, B.; Zhang, Z. Zeroing neural networks: A survey. *Neurocomputing* **2017**, *267*, 597–604. [[CrossRef](#)]
19. Bragagnolo, L.; da Silva, R.V.; Grzybowski, J.M.V. Artificial neural network ensembles applied to the mapping of landslide Susceptibility. *Catena* **2000**, *184*, 104240. [[CrossRef](#)]

