Adaptive Network Based Fuzzy Inference System and the Future of Employability

Dr. Saada Abdel Gawad Khadragy*
The British University in Dubai

Prof. Sherief Abdallah
The British University in Dubai

Abstract
Educational data is considered by researchers and data scientists as an indicator for the future predictions. The current research study aims for classifying IT alumni students into employed and unemployed. The data collected from two universities in Jordan. 781 of IT alumni students in two universities in Jordan participate in the current study. Three classifiers are compared to determine the most suitable one for predicting the future of IT students' employability. The results show that Adaptive Network Based Fuzzy Inference System came as a suitable classifier for predicting IT students' employment in Jordan. As gender, programming skills, and communication skills came as the most effective factors affecting IT recruitment field, a set of recommendations is presented to the ministry of higher education based on the significant factors affecting IT graduates employment.

1. Introduction

the concept of employability is about being able to get and keep achieving particular work. In a more comprehensive way employability is the capacity to shift self-sufficiently into the work force market to recognize the ability through maintainable employment. In a personal view, employability is based on the knowledge, skills and viewpoints they have, the different way they apply those possessions and introduce them to employers, and the contexts within which they look for work, which is known as the personal status and the work force settings (Nauta et al., 2009). On the other hand, Employability is defined as having the ability to get a certain work as a type of employment, keep the same type of employment and attain new role if necessitated. Adding to the same point of view, employability for an individual focuses on how individuals can deploy their knowledge, skills, and attitudes in order to convince their employers (Vanhercke et al., 2014). Hence, the definition of employability is varied between individuals and institutional relationships with the labour market itself. As in a number of countries all over the world, the governmental sector concentrates more on the vocational skills than the soft skills (de Guzman and Choi, 2013). For all of the above, in this research study the concept of employability is highlighted for Computer Science, Computer Information System, and Computer Engineering graduates and the most effective factors affecting their abilities and skills to get a job. On the other hand, many tools and techniques are used to analyse the educational data in order to reach profitable decisions in the educational sector (Osofisan, Adeyemo and Oluwasusi, 2014). One of the major tasks of data mining in education is to analyse students' behaviours (Dawson and Dawson, 2019). Furthermore, recent years have witnessed a great rise in the application of electronic tools in the field of education. From nursery classes at the preschool stage to the postgraduate programs at the universities, electronic tools are being used extensively to support and enrich the quality of education. Although the implementation of computer networks is an integral characteristic of online learning, the face to face schools and universities are also using extensive network-connected electronic tools such as mobile phones, tablets, and computers, which directly and indirectly affect graduates' employability in all fields and majors. However, practitioners and academic administrators can gain from their colleagues in business and service industries a multifaceted system of methods and tools, usually indicating data mining which is being applied to analyse a huge dataset in decision-making. Scientists and researchers have begun paying attention to the use of data mining and data analytics to manage big data created by the educational sector. From the educational perspective, these tools are specifically defined as educational data mining (EDM) and learning analytics (LA). Generally, EDM seeks new patterns in big data and provides new algorithms and/or new models, while LA uses known predictive models in certain systems.

1.1 Research objectives:
Since Artificial Intelligence and data mining techniques are used in the field of graduates' employability with a limited amount in the Middle East region, this research study will cover this research gap which is considered as the major aim of the study. From all of the above, this research study aims at the following points:

I. To build an appropriate model using neuro-fuzzy, by sorting the CS and IT alumni graduates into three categories: employed, unemployed, and undetermined situation.

II. To classify the most effective factors influencing the current employment requirements. Those such as gender, CS modules grades, age, nationality, and self-employment are recognized as the major factors.

III. To predict an appropriate model for the newly registered students in both CS and IT in Jordanian universities. This model would help in classifying those students between will be employed, will not be employed, and undetermined situation.

IV. To recommend a required skills and knowledge basis for students and program designers in the higher education institutions. As explained in figure 1.2, the study will cover all of its questions after applying all procedures and techniques. 1.1.1 Heading 3

1.2 Research questions:

In order to fulfil the research objectives, a number of questions are addressed to highlight the major aims of the study as followed:

1. What is the most suitable classifier for predicting the future of graduates' employability?
2. Which factor most affect the CS, IT students' employability future?
3. Which factor most affect the CS and IT graduates' future employment rates?

The required data for this research study will be collected from different sources quantitatively and qualitatively. The sample population for the current study will be the alumni graduating students from different universities in Jordan. The data of the graduate students was collected from the graduates' data; demographics, soft skills, technical skills, and academic skills; which was built from a tracer study carried out by the Jordanian universities.

1.3 Problem statement:

Recently, educational data is generated with great amounts from all educational sectors. This data could be an indicator for the future predictions and decisions and guides to give a list of recommendations for future development as well. This research study aims to decide a suitable classifier to classify alumni graduates into employed and unemployed. Figure1 shows the educational process and the set of acquired skills from it. The acquired skills may help graduated students to find a job in the future. Those skills could be soft, academic, or communication skills. Also the number of research studies in the field of educational data mining and the future of employability is rare in the middle east which will help in giving reliability to the current research study.
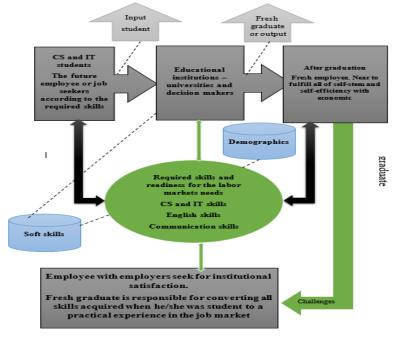

figure1. Problem statement of the study

Literature review:

Currently, employability has gained a lot of attention from higher education institutions. Employability data enables these institutions to better plan their educational strategies, enhance the curriculum, as well as improve

students' performance. Studying employability can be achieved by analysing the data extracted from educational resources. Nowadays, there are a lot of resources of educational data such as learning management systems (LMS), online learning system, admission systems, and social media, which provide a vast amount of data. Handling and analysing this huge amount of data cannot be carried out without sophisticated tools, and data mining is the most important technology for this purpose; the task of data mining predicts interesting and useful information that helps in making future strategy plans. Data mining has been used in many fields such as banking, stock exchange, marketing management, retail sales, health, and recently in education.

On the other hand, employers are often in search of skills that go beyond credentials, knowledge, and experience. An individual's education and experience may make him/her suitable to be hired and join a job, and to be effective in the majority of tasks, he/she will need certain skills which are likely to progress over time. Some skills will be precise to the job, but the greatest amount will be so-called 'soft skills' that can be needed in any job or employment fields. These soft skills are 'employability skills': they are what makes anyone employable (Lin, Korsakul and Korsakul, 2012). Furthermore, higher education has a basic role in the reinforcement of any people's economy as it is an industry branch in itself and it keeps the rest of the industry going by offering trained labourers (Laine, Leino and Pulkkinen, 2015). In the past, the basic issues for the educational institutions were the reduction in the students' success rates (Araque, Roldán and Salguero, 2009), reduction in students' retention (Pegrum, Bartle and Longnecker, 2015), growth in students shifting to other likely institutions and shortage of guiding students in subject choice. Since education has become more employment focused, graduating from any university has become an essential aspect in developing its reputation (Pool, Qualter and Sewell, 2014). Educational institutions produce and gather vast amount of data (Sedkaoui and Khelfaoui, 2019). This may be about students' academic progression, students' personal profile, their communication abilities, their web log interests and also graduates' profiles (Bharambe et al., 2017). The task of predicting students' employability may support identifying the students who are at a higher risk of unemployment and consequently management may make timely interference by introducing vital steps to train their students in order to develop their performance skills (Mishra, Kumar and Gupta, 2017). Computer scientists have tried to identify relations between academic achievement, socioeconomic surroundings, job skills of the graduates and employability, however, mostly by implementing statistical methods (Abas and Imam, 2016). Those statistical methods have a direct relation with the field of computer science and data analytics (Cao, 2017).

Educational Data Mining (EDM) is the process of applying Data Mining (DM) techniques to educational data (Wu et al., 2014). EDM has emerged as an important research area in recent years for researchers all over the world (García et al., 2011) . Due to the importance of data mining in the field of education, the International Educational Data Mining Society was founded in July 2011. This society provides much useful research related to EDM, as well as helpful datasets.

This chapter surveys the essential background to understand the DM and EDM. Firstly, here is discussed the data mining definitions as part of Knowledge Discovery in Databases (KDD), and also present data mining tasks and major applications. Furthermore, educational data mining (EDM) is illustrated as a main part of this research study. A lot of literature shown in this section demonstrates the most important research studies that have included the concept of EDM. Then, the most important techniques and applications of applying educational data mining in employability are investigated. Finally, a very important part of literature, which is neuro-fuzzy approach backgrounds and related works is illustrated.

- Data mining

This section deliberates the data mining definition as part of KDD, and also presents data mining tasks and its main applications.

- Definition

Data mining is the process of automatically discovering useful information in a huge dataset. Other definitions for DM are, "the non-trivial extraction of implicit, previously unknown and potentially useful information (such as rule, constraints and regularities) from data in a database" (Kaur and Madan, 2015), and "efficient discovery of previously unknown patterns in large datasets" (Anand, Bell and Hughes, 1996), and "process of employing one or more computer learning techniques to automatically analyse and extract knowledge from data contained within databases" (Pechenizkiy et al., 2011).

Not all tasks of information elicitation from databases are considered as data mining tasks. For instance, the retrieval of information about web pages using a search engine is considered an information retrieval task, not a data mining task.

- Data Mining Techniques Used for Employability:

Some researchers focus on data mining methods and algorithms to predict graduates' employability; others study the attributes that impact employability. The following literature studied attributes of employability.

Al-Janabi (2010) proposed an approach depending on features (knowledge areas) gained from the logged data of employment and university graduates. He presented a model for analysing data of the IT graduates according to the employability knowledge areas in order to predict feedback recommendations to enhance the IT programs'

teaching and learning resources and processes towards the improvement of the programs' learning outcomes.

Sriram, Srinivas and Thammi (2014) presented a model to predict the attributes which play the main role in the employability of students. They used Maximally Specific Hyporesearch study in order to reduce the representation of rules. These hypotheses can be used to identify the key attributes needed for employability among graduates.

Thakar and Mehta (2017) studied the role of secondary attributes to enhance the prediction accuracy of students' employability using data mining. They proved that prediction accuracy for students' employability can be enhanced with the applying of secondary attributes such as personal, social, psychological and other environmental variables in the dataset.

The following are related to work that focused on the data mining methods or algorithms to predict graduate employability.

Piad (2018) proposed a technique to predict the employability of IT graduates. His study defines the influential attributes for supervised learning using data mining methods. He conducted a comparison between several classification data mining algorithms. These algorithms are Naive Bayes, J48, Simple Cart, Logistic Regression and Chaid Algorithms. The author proved that the Logistic Regression achieved the highest accuracy, and he found that three possible predictors with a direct effect on IT employability are the IT_core Subjects, IT_professional subjects and gender (Piad, 2018).

Jantawan et al. (2013) used real data of graduate students of Maejo University in Thailand over three academic years. They conducted several experiments using algorithms of Bayesian Network and Decision Tree to predict whether a graduate has been employed, remains unemployed, or is in an undetermined situation after graduation (Jantawan and Tsai, 2013).

Sapaat et al. (2011) built the graduates employability model using a classification method in data mining. To perform the classification, they used extracted data from web-based survey system from the Ministry of Higher Education, Malaysia (MOHE) for the year 2009. Bayes algorithms were used to achieve classification. In addition, they compared the performance of Bayes algorithms against a number of tree-based algorithms. The comparison shows the superiority of the Decision Tree classification model over Bayes Network Classification Models (Sapaat et al., 2011).

Mishra et al. (2016) applied several classifiers to predict the employability of students and build an employability model based on proper classifiers. The authors used different classification methods of data mining such as Bayesian methods, Multilayer Perceptron's and Sequential Minimal Optimization (SMO), Ensemble Methods and Decision Trees. They conducted a comparison between the classifiers to find the best classifier. A comparative study shows that J48 (a pruned C4. 5 decision tree) is most suitable for employability (Mishra, Kumar and Gupta, 2016).

Rahman, Tan and Lim (2017) used supervised and unsupervised learning in data mining for employment prediction of fresh graduate students. These techniques were applied in features selection and determined the best model that can be used to predict the employment status of fresh graduates, either employed or unemployed. The algorithms in supervised and unsupervised learning, K-Nearest Neighbor, Naive Bayes, Decision Tree, Neural Network, Logistic Regression and Support Vector Machines, were compared to find which one achieved the best accuracy. They proved that K-Nearest Neighbor achieved the highest accuracy (Rahman, Tan and Lim, 2017).

Othman, Shan, Yusoff and Kee (2018) proposed a model that uses data mining techniques to discover the most important features that affect graduates' employability. They collected seven years of data (from 2011 to 2017) through Malaysia's Ministry of Education tracer study. The authors applied a set classification algorithms (three), Decision Tree, Support Vector Machines, and Artificial Neural Networks to develop the classification model, then compared them to reach the perfect performance. According to the authors, the decision tree J48 algorithm achieved higher accuracy compared to other algorithms, with a classification accuracy of 66.0651%, and it rose to 66.1824% after the process of parameter tuning. In their work, they discovered seven variables affecting graduate employability: age, faculty, the field of study, co-curriculum, marital status, industrial internship and English skill. In addition to these variables, attribute age, industrial internship and faculty hold the information that influences the employability status (Othman et al., 2018)

Methods

The main objective of the current research study is to categorize the graduate students according to their employability status. The data is taken from the graduates' dataset which is built from tracer studies carried out by various universities in Jordan. The dataset consists of 781 instances and 7 attributes that refer to graduates' profiles.

Studying the employability can be conducted by collecting graduates' data, then analyzing this data to extract useful information. Several methods are used to carry out analysis for different tasks. Data mining provides many techniques to build models that can extract important information from huge datasets (Lefebvre-

Ulrikson et al., 2016). The availability of a suitable data sample and diversity of data sources assists to perform data mining processes and apply different data mining techniques.

The methodology applied in this research study comprises the following steps:
1. Data collection
2. Dataset preprocessing
3. Implementation of Decision Tree, Bayesian methods, and ANFIS algorithms
4. Results evaluation and validation

Attribute	Description
Gender	Takes two linguistic values, male or female
University	University of graduate
Major	Program (computer science, computer information system and software engineering)
GPA	Final grade point average of the graduates
Age	The age divided into four intervals
Programming skills	The programming and technical skill divided into low, moderate or high
Educational degree	Level of certificate of graduate (bachelor, high diploma and master)

Table 3. study attributes

- *Confusion Matrix:*

Is a table that is used to evaluate the efficiency of a classification model on a group of test data for which the true values are defined. It visualizes the performance of an algorithm (Ting, 2017). It is a helpful tool to estimate the performance of classifiers; it gives us a better understanding not only of the mistakes being produced by a classifier but it gives us kinds of mistakes that are being produced. In this study we will use a confusion matrix to test our classifier and compare it with other classifiers.

Table 1. Confusion matrix

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

- True positives (TP): represents the true items that were rightly assigned by the classifier. Let TP be the number of true positives. For example, if we have two classes, the TP may be employed = yes where the negative items are employed = no.
- True negatives (TN): Represents the correct negative items that were assigned by the classifier.
- False positives (FP): Represents the incorrect negative items that were assigned as positive.
- False negatives (FN): estimates the positive items that were mislabeled as false.

The accuracy of the classifier can be calculated as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

The value of Recall provides us an idea about when it's really yes.:

Recall = TP / (TP + FN)

Precision value indicates when the prediction is yes.

Precision = TP / (TP + FP)

F-measure: We compute an F-measure which applies Harmonic Mean in place of Arithmetic Mean as it disciplines the extreme values more. The F-Measure will always be closer to the smaller value of Precision or Recall.

F-measure = (2 * Recall * Precision) / (Recall + Precision).

Table 2 shows how to utilize the confusion matrix by applying a real dataset on a classifier.

Table 2. Confusion matrix of applying real dataset on a classifier

Class	Employed = yes	Employed = no	Total
Employed = yes	169	67	236
Employed = no	54	310	364
Total	223	377	600

As shown in the above table, confusion matrix contains two classes; (Employed = yes) and (Employed = no) with a dataset of 600instances that represents graduates. The classifier classifies 169 graduate as (Employed = yes) out of 236 and classify 310 graduates as (Employed = no) out of 364.

There are two stages in the classification task, consisting of training and testing phases, of which the training phase was illustrated previously. The testing step is illustrated by determining the testing dataset for estimating the accuracy of the prediction. As was discussed previously, there are many popular testing strategies, but the most commonly used methods are the following four methods, that are used in WEKA and MATLAB:

Training set: this method is based on selecting a testing dataset from the training dataset randomly, or we can use whole training dataset as a testing dataset after creating the model. This method almost always causes overfitting problems and makes the accuracy result unreliable, so this method rarely is rarely used for testing purposes.

Supplied test set: this method is based on determining the training dataset and testing dataset separately. This method almost always gives reliable results about the accuracy of the prediction model because the testing data is not the same as the training data.

Cross-validation: in the cross-validation method, the dataset is divided into a number of portions that have equal size. One of the portions is used for testing the classifier, while the remaining portions are dedicated for training the classifier. The process is re-implemented k times (the folds); each time a different partition is used for testing (validation) and others for training. The results produced from the K testing process are combined to compute the final estimation. A 10-fold class validation is commonly used to get the perfect for measuring error. It has been widely applied on numerous datasets with different classification algorithms.

Percentage split: in this method the dataset is partitioned into two portions in which the first part is for training and the other for testing. Researchers adopt different percentages according the type of dataset and application; some of them use 70% for training, 30% for testing, and others use 50%, 50%.

Results

In this research study the 10-fold cross validation process is implemented to evaluate the accuracy of the model that was built using the neuro-fuzzy inference system (ANFIS). The ANFIS algorithm was applied 50 times for training the neural network architecture in order to reach the accurate model.

After building the above classifier, we needed to evaluate the performance of this model. We used the same data set to train several classifiers, such as decision tree, Naive Bayes, and ANFIS in order to compare them with our classifier. All classifiers were trained and tested using a 10-fold cross-validation method to avoid overfitting. Since we used 10-fold cross technique, all the observation of our data sample was used for training and testing. The accuracy of the classifier was specified by comparing the predicted class labels with the original consequence in the testing dataset. In this study we used several measures to evaluate our classification model. These measures include confusion matrix to compute the accuracy, Precision, Recall, False-Positive rate and recall. RMSE is another important measure used to test the efficiency of our model. Kappa measure is also used. Furthermore, as mentioned previously we adopted a technique based on attributes, an incremental approach to evaluate both the computational cost and efficiency of our classification model. We measured the computational cost by calculating the execution time for each classifier.

Table 4 Detailed accuracy by each class for three attributes classifiers.

Classifier	Class label	Recall (%)	False-Positive rate (%)	Prec. (%)	F-score (%)
ANFIS	Employed	73.4	3.4	73.8	74.6
	Not-employed	73.1	4.8	73.3	74.4
Decision tree	Employed	71.6	5.6	72.1	71.6
	Not-employed	71.3	6.5	70.5	71.4
Naïve Bayes	Employed	63.2	7.8	61.4	63.7
	Not-employed	63.3	6.7	60.3	62.5

Table 4 shows that the ANFIS classifier achieved the highest values for Recall, Precision, Recall, F-score; and lowest value for False-Positive rate. These values have scored in predicting both classes (Employed, Not-employed). On the other hand, lowest values for Recall, Precision, Recall, F-score; and highest value for False-Positive rate, which indicates again the superiority of ANFIS in predicting efficiency.

To complete the evaluation process of the above classifiers, the performance of classifier determined according two statistical measures that are the Root Mean Square Error and Kappa measures. In our experiment, we applied a well-known measure which is the Root Mean Square Error (RMSE) where the value of its must be in minimum level to say the classifier achieve good performance. Kappa statistic is used to compare between two raters that are in this thesis the classified data and the testing data

Table 5 RMSE and Kappa statistic values for each classifier applying three attributes

Classifier	RMSE	Kappa statistic
ANFIS	0.3489	0.7364
Decision tree	0.3912	0.7147
Naïve Bayes	0.6579	0.6218

As shown in Table 5 the ANFIS classifier achieve the lowest Root mean square error value of 0.3489, as well the Kappa measure value was 0.7364; followed by decision tree classifier comes out in second place with a root mean square error 0.3912 and Kappa measure value 0.7147; Naïve Bayes classifier achieved 0.6579 root mean square error and Kappa measure value was 0.6218, which is the worst algorithm in our experiments in this stage.

After comparing the three classifiers, ANFIS is found as the most suitable classifier for predicting the IT graduates' future employment in Jordan. The classifier shows that gender, programming skills, and GPA have the significant influence in the IT graduates' employment.

Conclusion and recommendations:

A set of recommendations is provided to the ministry of higher education in Jordan based on the research results as followed:

The admission process should follow a certain standard in accepting allocated percentage from both males and females in the majors of IT.

English language skills should be taken into considerations in accepting newly registered students in both of IT majors.

Decision makers and curriculum developers should enhance the courses given to IT students with more programming and soft skills.

References

Abas, M. C. and Imam, O. A. (2016) 'Graduates' Competence on Employability Skills and Job Performance', International Journal of Evaluation and Research in Education (IJERE). doi: 10.11591/ijere.v5i2.4530.

Anand, S. S., Bell, D. A. and Hughes, J. G. (1996) 'EDM: A general framework for Data Mining based on Evidence Theory', Data and Knowledge Engineering. doi: 10.1016/0169-023X(95)00038-T.

Araque, F., Roldán, C. and Salguero, A. (2009) 'Computers & Education Factors influencing university drop out rates', Computers & Education. doi: 10.1016/j.compedu.2009.03.013.

Bharambe, Y. et al. (2017) 'Assessing employability of students using data mining techniques', in 2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017. doi: 10.1109/ICACCI.2017.8126157.

Cao, L. (2017) 'Data science: A comprehensive overview', ACM Computing Surveys. doi: 10.1145/3076253.

Dawson, C. and Dawson, C. (2019) 'Educational data mining', in A–Z of Digital Research Methods. doi: 10.4324/9781351044677-18.

García, E. et al. (2011) 'A collaborative educational association rule mining tool', Internet and Higher Education. doi: 10.1016/j.iheduc.2010.07.006.

de Guzman, A. B. and Choi, K. O. (2013) 'The relations of employability skills to career adaptability among technical school students', Journal of Vocational Behavior. doi: 10.1016/j.jvb.2013.01.009.

Jantawan, B. and Tsai, C.-F. (2013) 'The Development of Educational Quality Administration: a Case of Technical College in Southern Thailand', International Journal of Computer Science and Information Security.

Kaur, J. and Madan, N. (2015) 'Association Rule Mining: A Survey', International Journal of Hybrid Information Technology. doi: 10.14257/ijhit.2015.8.7.22.

Laine, K., Leino, M. and Pulkkinen, P. (2015) 'Open Innovation Between Higher Education and Industry', Journal of the Knowledge Economy. doi: 10.1007/s13132-015-0259-2.

Lefebvre-Ulrikson, W. et al. (2016) 'Data Mining', in Atom Probe Tomography: Put Theory Into Practice. doi: 10.1016/B978-0-12-804647-0.00009-7.

Lin, R.-F., Korsakul, Nattawadee and Korsakul, Nattharuja (2012) 'Cultural Intelligence as Core Competency to Employability', The Business & Management Review.

Mishra, T., Kumar, D. and Gupta, S. (2016) 'Students' employability prediction model through data mining', International Journal of Applied Engineering Research.

Mishra, T., Kumar, D. and Gupta, S. (2017) 'Students' Performance and Employability Prediction through Data Mining: A Survey', Indian Journal of Science and Technology. doi: 10.17485/ijst/2017/v10i24/110791.

Nauta, A. et al. (2009) 'Understanding the factors that promote employability orientation: The impact of employability culture, career satisfaction, and role breadth self-efficacy', Journal of Occupational and Organizational Psychology. doi: 10.1348/096317908X320147.

Osofisan, A., Adeyemo, O. and Oluwasusi, S. (2014) 'Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database', African Journal of Computing & ICT ©.

Othman, Z. et al. (2018) 'Classification techniques for predicting graduate employability', International Journal on Advanced Science, Engineering and Information Technology. doi: 10.18517/ijaseit.8.4-2.6832.

Pechenizkiy, M. et al. (2011) Proceedings of the International Conference on Educational Data Mining (EDM) (4th, Eindhoven, the Netherlands, July 6-8, 2011), International Working Group on Educational Data Mining.

Pegrum, M., Bartle, E. and Longnecker, N. (2015) 'Can creative podcasting promote deep learning? The use of podcasting for learning content in an undergraduate science unit', British Journal of Educational Technology. doi: 10.1111/bjet.12133.

Piad, K. C. (2018) 'Determining the dominant attributes of information technology graduates employability

prediction using data mining classification techniques', Journal of Theoretical and Applied Information Technology.

Pool, L. D., Qualter, P. and Sewell, P. J. (2014) 'Exploring the factor structure of the CareerEDGE employability development profile', Education and Training. doi: 10.1108/ET-01-2013-0009.

Rahman, N. A. A., Tan, K. L. and Lim, C. K. (2017) 'Predictive analysis and data mining among the employment of fresh graduate students in HEI', in AIP Conference Proceedings. doi: 10.1063/1.5005340.

Sapaat, M. A. et al. (2011) 'A classification-based graduates employability model for tracer study by MOHE', in Communications in Computer and Information Science. doi: 10.1007/978-3-642-22389-1_25.

Sedkaoui, S. and Khelfaoui, M. (2019) 'Understand, develop and enhance the learning process with big data', Information Discovery and Delivery. doi: 10.1108/IDD-09-2018-0043.

Ting, K. M. (2017) 'Confusion Matrix', in Encyclopedia of Machine Learning and Data Mining. doi: 10.1007/978-1-4899-7687-1_50.

Vanhercke, D. et al. (2014) 'Defining perceived employability: A psychological approach', Personnel Review. doi: 10.1108/PR-07-2012-0110.

Wu, X. et al. (2014) 'Data mining with big data', IEEE Transactions on Knowledge and Data Engineering. doi: 10.1109/TKDE.2013.109.