Behavior Of Employability Indicators In University Graduates

Comportamiento de indicadores de empleabilidad en egresados universitarios

Comportamento dos indicadores de empregabilidade em graduados universitários

Nicole Valentina Chacon-Sánchez[1]
Daniel Esteban Casas-Mateus[2]
Luz Deicy Alvarado Nieto[3]

Artículo de investigación. https://doi.org/10.16925/2357-6014.2020.02.03

[1]   Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas.
       E-mail: nvchacons@correo.udistrital.edu.co
       ORCID: https://orcid.org/0000-0002-6295-3312

[2]   Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas.
       E-mail: decasasm@correo.udistrital.edu.co
       ORCID: https://orcid.org/0000-0002-1027-0413

[3]   Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas.
       E-mail: lalvarado@ udistrital.edu.co
       ORCID: https://orcid.org/0000-0002-1305-3123

Abstract

Introduction: This article is the result of research entitled the behavior of employability indicators in university graduates, developed at the Universidad Distrital Francisco José de Caldas in 2019.

Problem: The Emple-AP project promotes the creation of an observatory for labor insertion and the strengthening of employability in countries of the Pacific Alliance (PA), which particularly benefits Colombia, because one of its objectives with the PA is to overcome the socioeconomic inequality that exists among its inhabitants.

Objective: To identify the relationship between employability indicators through classification methods used in Artificial Intelligence.

Methodology: The indicators' behavior description involves data pre-processing, a formal global study in statistics and a specific formal study through comparison of classification methods.

Results: Descriptions of these employability indicators show characteristics of the situation in the studied population.

Conclusion: Given the analysis of the classification model, it is determined that the diversity and disparity of the dataset makes the RandomTree model the most accurate in this research, finding that the system has characteristic behaviors of an adaptative complex system.

Originality: Through this research, employability indicators were analyzed through data mining tools, additionally the analysis presented in this article could be replicated under particular conditions in other countries of the PA.

Limitations: The information comes from the Universidad Distrital Francisco José de Caldas graduate's office. A single source generates a limitation in the data and in the population studied.

Keywords: Behavior patterns, Comparative analysis, Complex system, Data mining, Employability indicators, University graduates.

Resumen

Introducción: El presente artículo es resultado del proyecto de investigación comportamiento de los indicadores de empleabilidad de egresados universitarios desarrollado en la Universidad Distrital Francisco José de Caldas en 2019

Problema: El proyecto Emple-AP promueve la creación de un observatorio para la inserción laboral y el fortalecimiento de la empleabilidad en países de la Alianza del Pacífico (AP), lo cual beneficia particularmente a Colombia, debido a que uno de sus objetivos con la AP es superar la desigualdad socioeconómica que existe entre sus habitantes.

Objetivo: Identificar la relación existente entre los indicadores de empleabilidad mediante métodos de clasificación empleados en Inteligencia Artificial.

Metodología: La descripción del comportamiento de los indicadores implica el pre- procesamiento de datos, el estudio formal global en la estadística y el estudio formal especifico a través de comparación de métodos de clasificación.

Resultados: Las descripciones de estos indicadores de empleabilidad evidencian características de la situación en la población estudiada.

Conclusión: Dado el análisis de modelos de clasificación se determina que la diversidad y disparidad del conjunto de datos hace que el modelo RandomTree sea el que tiene mayor precisión en esta investigación, encontrando que el sistema tiene comportamientos característicos de un sistema complejo adaptativo.

Originalidad: A través de esta investigación se analizó el comportamiento de la empleabilidad por medio de herramientas de minería de datos, adicionalmente el análisis presentado en este artículo podría ser replicado bajo condiciones particulares en los demás países de la AP.

Limitaciones: La información proviene de la oficina de egresados de la Universidad Distrital Francisco José de Caldas, siendo una sola fuente, lo que genera una limitante en los datos y en la población estudiada.

Palabras clave: Análisis comparativo, Egresados universitarios, Indicadores de empleabilidad, Minería de datos, Patrones de comportamiento, Sistema complejo.

Resumo

Introdução: Este artigo é o resultado do comportamento dos indicadores de empregabilidade do projeto de pesquisa de graduados universitários desenvolvido na Universidade Distrital Francisco José de Caldas em 2019.

Problema: O projeto Emple-AP promove a criação de um observatório para inserção trabalhista e o fortalecimento da empregabilidade nos países da Aliança do Pacífico (PA), o que beneficia particularmente a Colômbia, porque um de seus objetivos com a AP é superar a desigualdade socioeconômica que existe entre seus habitantes.

Objetivo: Identificar a relação entre indicadores de empregabilidade por meio de métodos de classificação usados em Inteligência Artificial.

Metodologia: A descrição do comportamento dos indicadores envolve pré-processamento de dados, estudo global formal em estatística e estudo formal específico através da comparação de métodos de classificação.

Resultados: As descrições desses indicadores de empregabilidade mostram características da situação na população estudada.

Conclusão: Dada a análise dos modelos de classificação, determina-se que a diversidade e disparidade do conjunto de dados tornam o modelo RandomTree o mais preciso nesta pesquisa, constatando que o sistema possui comportamentos característicos de um sistema complexo adaptativo.

Originalidade: por meio dessa pesquisa, o comportamento da empregabilidade foi analisado por meio de ferramentas de mineração de dados; além disso, a análise apresentada neste artigo pode ser replicada em condições particulares nos outros países da AP.

Limitações: As informações são do departamento de licenciados da Universidade Distrital Francisco José de Caldas, sendo que uma única fonte gera uma limitação nos dados e na população estudada.

Palavras-chave: padrões de comportamento, análise comparativa, sistema complexo, mineração de dados, indicadores de empregabilidade, graduados

1. INTRODUCTION

This article describes the behavior of employability indicators, analyzing it as a complex system. In the case of Colombia, several political and social actions have been created by the state, since they become a social theme which should be public and transparent. However, the information is distributed in different institutions of the Colombian government; to consider this topic as a system, the information must be

unified in order to analyze it thoroughly and frequently, so the results can be understood by any member of the community in general.

Through statistical and comparative analyses of data mining models [1], the behavior of employability indicators are described, to present them in a common language to the society and finally contrast them with employability [2] [3].

For this, information is collected from graduates of the Universidad Distrital Francisco José de Caldas (UDFJC) provided by the graduate's office of the same university, from the Observatorio Laboral para la Educación (OLE) [4]. Subsequently, a data preprocessing and preparation is performed to carry out analyses. The indicators provide data that reflect the system situation to be evaluated. This information is usually given quantitatively and it directly affects the system [5]. On the other hand, the importance of involving each of the employability actors, such as Higher Education Institutions (HEIs), companies, state, society, among others, is raised; presenting the information to contribute in the decision making of the actors involved.

In conclusion, information is obtained from government entities which represent the university's graduate population and through statistical analyses and the use of data mining [6] a description of the behavior of employability indicators is made [7], where these behaviors demonstrate that employability behaves as a complex system. Next, the theoretical framework and the proposed objectives are presented, followed by the design and implementation to finish with results and discussion of these.

1.1 Literature review

Globally, numerous studies related to employability and the analysis of various indicators for students and graduates have been carried out. Below, some of these studies are briefly presented.

The authors [8] establish, for 1400 students in the master's degree in computer applications of several universities in India, rules that indicate whether the student is likely to get a job or not. These rules are obtained through data collection and preprocessing in which missing, inconsistent and duplicate data are eliminated leaving just 1359 records. Subsequently, a comparative analysis of data mining classification methods is performed such as: J48 algorithm, Random Forest, Random Tree, Support Machine Optimization (SMO), Perceptron multilayer and Naive Bayes, where the method with the best response time is Naive Bayes. The previous process is carried out with the WEKA data mining tool which compares the area under the ROC (Receiver Operating Characteristic) curve information and finally a model is developed to deliver 11 behavior rules [9] with the J48 algorithm. This is selected because the Naive Bayes

accuracy is low with 62.86%. Additionally, SMO and multilayer perceptron have high compilation times whilst Random Forest does not allow for the visualization of parameters, which is the final objective of this study, and comparing them with Random Tree, J48 has a better accuracy percentage with 70.19%.

It is also observed that authors [10] predict the students' employability based on the admission test, where they give personal and academic information, which is analyzed through statistical techniques, clustering [11] and classification algorithms [12]. Among the algorithms used are X-Means which is a variation of K-Means, Naive Bayes and Support Vector Clustering (SVC). These algorithms were executed in Rapid Miner and it was obtained as a result that X-Means gives the best accuracy with 83 %, while SVC and Naive Bayes obtained 81% and 77% respectively.

On the other hand, [13] determined the work status of a graduate through the analysis of Tracer Studding which is a survey system of the Ministry of Higher Education in Malaysia (MOHE) which has 82290 data. To carry out this analysis, two groups of algorithms were used with the first based on Bayes and the second based on Trees.

Within the Bayes algorithms are WAODE, AODE, Naive Bayesian, Naive Bayesian Simple, BayesNet, AODEsr, Naive Bayes Updatable and HNB. Among those based on Trees are J48 Graft, J48, Simple Cart, Random Forest, LAD Tree, REPTree, Decision Stump, Random Tree and ID3. When each of the algorithms were executed in WEKA tool, it was evidenced that J48 Graft had an accuracy of 92.3%, being the best result for those based on Trees, while in Bayesian algorithms [14] the best was WAODE with a 91.3% accuracy. Therefore, for MOHE data, it is concluded that J48 Graft predicts the employability status of a graduate with greater precision, taking as a main factor the labor sector attribute contained in the profile of the graduate [15].

On the other hand, the studied systems characteristics are presented in a global and specific way [16] describing unexpected behaviors inside them, and due to the interconnections of their multiple parts, these are called complex systems. Likewise, different complex systems types are evidenced, such as: adaptive, connective, interdependent, emerging, among others.

In the same way, in 2011 a regional integration mechanism called Pacific Alliance (PA) was built, consisting of four Latin American countries: Colombia, Chile, Peru and Mexico. The initiative was born with the purpose of boosting the economy and development of the member countries, forming this mechanism into a new way of doing business, giving progress to the free movement of goods, services, capital and people [17].

For Colombia in particular, one of the objectives with this mechanism is to overcome the socioeconomic inequality that exists among its inhabitants together with the strengthening of the competences and capacities of the Colombian population, promoting access to high quality education. This objective fits the purpose of the Emple-ap project, in which several institutions belonging to the PA countries are currently participating and, among other things, tend to mobilize graduates from institutions belonging to countries that are members of the PA [18].

With the idea of supporting this proposal, an employability indicator's behavior description is made, which is information that shows the current situation or a particularity [19], providing parameters that affect employability, which is defined as Colombia's capacity, in this case, to increase the possibility of a graduate [20] finding or creating a job, keeping it and moving among roles through the articulation of the state, the Higher Education Institutions (HEIs), society and industry; materializing the contribution given to the employer and society [5].

In a complementary way, an analysis is carried out on the data of a graduates group from the Universidad Distrital Francisco José de Caldas (UDFJC) of Bogotá, where the labor insertion behavior is evidenced, which is an effective encounter between the opportunities that the labor market offers [21] and the graduate characteristics to take these opportunities [5]. Based on the fact that data mining is "the non-trivial process of identification that is valid, novel and potentially useful about understandable patterns that are hidden in the data" [22], which consists of data cleaning, noise removal, duplicate data removal, consistent data validation, and the data transformation [23], so they can be evaluated through descriptive methods in order to reveal relationships in this information, in order to find patterns of employability indicators behavior and present them within the Emple-ap project framework [24].

2. MATERIALS AND METHODS

In this section, the proposed research methodology design is detailed as follows:

2.1 General approach

Figure 1 shows the design and implementation process carried out, which begins with the data acquisition and collection. The preprocessing and preparation of this data is then carried out, to make the corresponding exploration and visualization at a statistical level [25] and evidence the classification results.
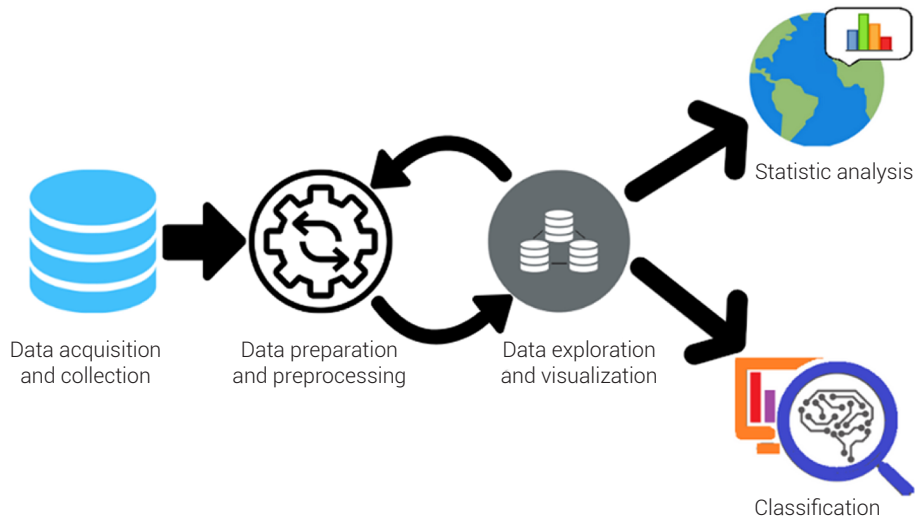
Figure 1. General Approach.
Source: own work

2.2 Data acquisition and collection

OLE collects data from Colombian university graduates. For the specific case from the UDFJC, these data are collected by the graduate's office, which requests the completion of the survey that is currently standardized for the periods between 2012 and 2018.

This data delivery entailes a privacy guarantee process, in addition to a feedback to the graduate office by the authors of this research. These surveys are digitized per year, so the first step was to consolidate data into a single dataset that initially had a total of 1641 records and 132 attributes (columns).

2.3 Data preprocessing

A data analysis was carried out, where duplicates were detected through two identifiers: the first is the national ID and the second a unique identifier generated by the system for each user.

The data set includes, in addition to the information required in this investigation, student graduate data from technological careers, therefore, a filter was made by graduates to obtain only those from professional careers, which constitute the target population of this study. On the other hand, it should be noted that in these last two steps the duplicity of university graduates with a previous technologist degree was

evidenced. It means that for people who have degrees as technologists and university professionals, only their professional degree is taken into account.

Additionally, a column with information of the survey completion percentage was found in the dataset, which means, since the survey was conducted virtually, graduates could leave it before completing it in its entirety. For this case, surveys that have a percentage equal or greater than 80% of completion were taken into account.

On the other hand, the columns with graduates' personal data were eliminated such as: name, address, telephone, among others; for data privacy and irrelevance in this research. Open response questions and opinion questions were also eliminated, which were not presented in a standardized way and therefore they did not contribute to the classification.

Then, through a correlation analysis, a close relationship of attributes was established such as: municipality and housing department, then the decision was made to leave only one of them in each case, leaving 1118 records at the end of the data preparation and 38 columns, where 30 of them were nominalized by ranges and 8 columns were described numerically for use in the different algorithms.

2.4 Data exploration and visualization

The data obtained are statistically analyzed [26] and with it, the classification is performed by data mining methods as follows:
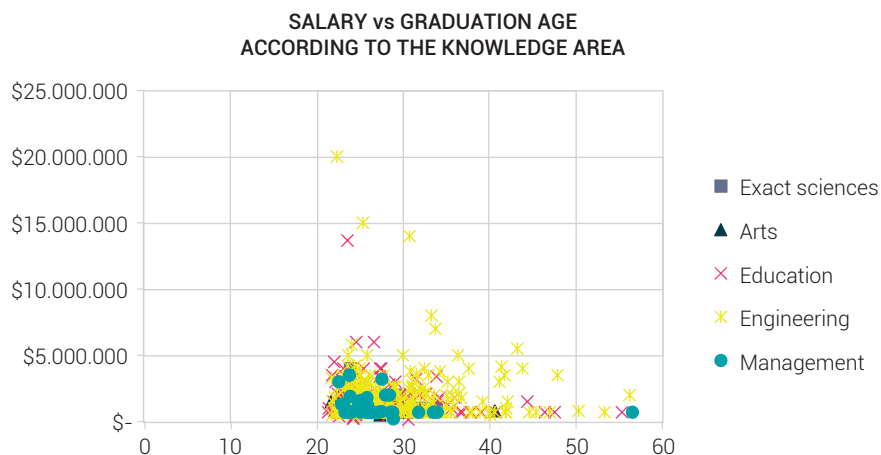
2.4.1 Statistical analysis

Figure 2. Wage vs. Graduation age.
Source: own work

Figure 2 shows a dispersion graph of wage vs. graduation age, where it is evident that the highest salaries correspond to those who obtained their degrees between 20 and 30 years. Additionally it is observed that the highest salaries are obtained by people in the engineering area.

SALARY vs SOCIOECONOMIC STRATA
ACCORDING TO THE KNOLEDGE AREA

Figure 3. Salary vs. Socioeconomic strata.
Source: own work

Figure 3 shows a graph comparing salaries, contemplated in a range of $ 700,000 COP to $ 2,050,000 COP, with respect to the graduates' socioeconomic strata separated by the area of knowledge. Here it is evident that the majority of graduates are in strata 0, 1, 2 and 3, fulfilling one of the objectives of the UDFJC: "to expand opportunities for access to higher education, for Colombians, particularly graduates of official schools of the District" [27].
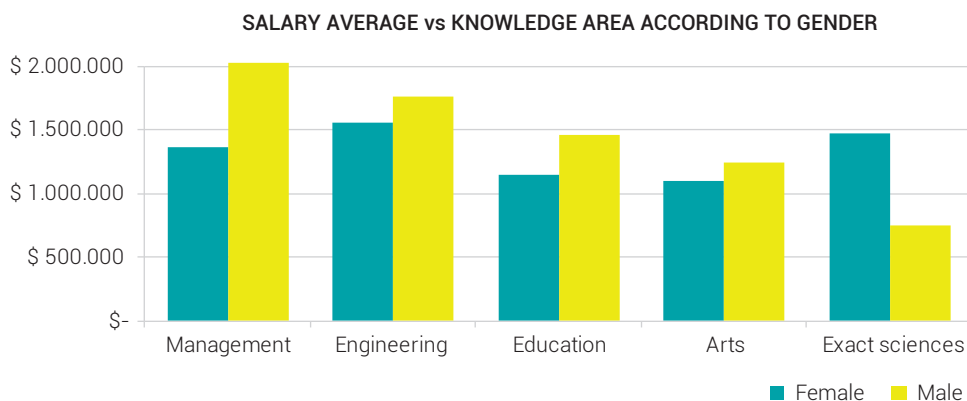
SALARY AVERAGE vs KNOWLEDGE AREA ACCORDING TO GENDER

Figure 4. Average salary vs. Knowledge area.
Source: own work

After the data consolidation, the salary averages are taken with respect to the knowledge areas, separating the population by gender. As Figure 4 shows, in four of the five areas, men have better salary averages. The Exact Sciences area is the only one where women have a better salary average. It is clear that Mathematics is the career with the highest number of graduates in the area of Exact Sciences.

2.4.2 Classification

For the classification, correlations were made between the variables with the help of the Associated tool provided by WEKA. In the end, the variable TRABAJO_RELACIONADO had a high correlation with 10 more variables, where four (4) of these were numerical and the other six (6) nominal. According to [28], for the results optimization in the classification, it is recommended that variables be numerical so the nominal ones are detected. The analysis was performed with a total of 5 variables as shown in Table 1.

Table 1. Variables and descriptions.

Variable	Description
TRABAJO_RELACIONADO	Relationship between work and profession
EDAD_ANOS	Graduation age in years
MESES_FIRST_JOB	Number of months from the graduation to obtain the first job
SALARIO	Accrued salary
HORAS_SEMANALES_FIRST_JOB	Number of hours worked in the first job

Source: own work

Where TRABAJO_RELACIONADO was the only nominal column to make the classification and took the values of YES or NO, EDAD_ANOS included a range between 21 and 57 years of age, MESES_FIRST_JOB from 0 to 24 months, salary from $ 150,000 Colombian Pesos (COP) up to $ 20,000,000 COP and HORAS_SEMANALES_FIRST_JOB from 20 to 130 hours.

3. RESULTS

The statistical graphs show the behaviors with respect to each of the knowledge areas, showing that it is in engineering where the best salaries are found. In 4 of the 5 areas men have a better salary than women and that UDFJC graduates belong mostly to socioeconomic strata 1, 2, and 3, thus fulfilling one of the fundamental University objectives.

To determine the classification method to be used, three of these Tree-based methods and one Bayesian method were evaluated, obtaining the accuracy and error results presented in Table 2.

Table 2. Classification methods comparison.

Classification method	Accuracy	Error
RandomTree	72.0036%	27.9964%
J48Graft	64.6691%	35.3309%
J48	62.7013%	37.2987%
NaiveBayes	61.3596%	38.6404%

Source: own work

When performing the classification with the characteristics mentioned above, the RandomTree confusion matrix is found in Table 3, where the true positive rate (TP) is 596 and the false positive rate (FP) is 209, giving an accuracy of 72.0036%, while the true negative rate (TN) is 199 and the false negative rate (FN) is 114, giving a 27.9964% error.

Table 3. RandomTree Confusion Matrix.

	YES	NO
YES	596	199
NO	114	209

Source: own work

As can be seen, the classification method with greater precision was RandomTree where its results show the data dependence obtained with respect to the first job and the graduation date to determine if this work was related to the profession. Additionally it was found that the Classification is volatile with respect to the input data, that means, when there are many entries, no associations are found between the variables and it is not possible to make predictions about the system [29].

From the classification model, system behavior rules are obtained, where it is found that the people for whom the first job is not related to their profession had salaries less than 325,000 COP and already worked before graduating. Additionally, the people who took more than 12 months to get their first job after graduation, started

earning less than 900,000 COP. On the other hand, if the work is related to the profession, there are three independent situations: a) earn more than 3,502,000 COP and are under 22.5 years, b) are under 22.5 years and work between 47, 5 and 77.5 hours per week, and c) are over 35.5 years working less than 33.5 hours per week.

From the descriptions presented in the statistical graphs of Figures 2, 3 and 4, the group behavior of university graduates can be evidenced as a formal global study of the system [16]. On the other hand, the RandomTree model describes the behavior and some interactions of a university graduate with the system, through collaborative analysis between mathematics, computing, data science and a good understanding of the problem which allows for a thorough formal study of it [16]. The process described above corresponds to the research carried out in this article.

4. CONCLUSIONS

Given the analysis of classification models, it is determined that the diversity and disparity of the dataset makes the RandomTree model the one that, for this case, is more accurate, although in the analysis performed by [8], the highest accuracy was obtained with J48 graft. Also, it is concluded that the attributes SALARIO and HORAS_SEMANALES_FIRST_JOB are determining factors in the classification to identify the relationship of the variable TRABAJO_RELACIONADO.

Finally, the parameters abstraction that involves the system as a whole must be taken into account, which must be provided by the articulated entities (HEI, state, company), since it is evident in the system:

1. It depends on the interaction of many parts
2. It is in constant transformation, which makes it non-linear and these transformations depend on the inputs such as the needs of companies, the professional personal skills of professionals and the involvement of HEIs, among others.
3. It grows exponentially as each of its parts involves more attributes becoming necessary within its subsystem, which implies an exponential connectivity.
4. It is in a constant process of evolution since the environment in which it develops, the labor market, is in permanent adaptation.

Finally, it is found that the characteristics evidenced above belong to a complex adaptive system more specifically a "living system" [30].

REFERENCES

[1]   Z. A. Bakar, R. Moheman, A. Ahmad y M. M. Deris, "A Comparative Study for Outlier Detection Techniques in Data Mining," *Proc. 2006 IEEE Conf. Cybernetics and Intelligent Systems,* pp. 1-6, 2006. [Online]. doi: http://dx.doi.org/10.1109/ICCIS.2006.252287

[2]   V. González Romá, J. P. Gamboa y J. Peiró, "University Graduates' Employability, Employment, Status, and Job Quality," *Journal of Career Development,* vol. 45, no. 2, pp. 132-149, 2018. [Online]. doi: http://dx.doi.org/10.1177/0894845316671607

[3]   M. R. Gabor, P. Blaga y C. Matis, "Supporting Employability by a Skills Assesment Innovative Tool - Sustainable Transnational Insights from Employers," vol. 11, no. 3360, p. 1, 2019. [Online]. doi: http://dx.doi.org/10.3390/su11123360

[4]   Ministerio de educación nacional - Colombia, "Observatorio Laboral para la Eduación," p. 1, 15 07 2019. [Online]. Available: https://www.mineducacion.gov.co/sistemasin-fo/Sistemas-de-Informacion-en-Ed-Superior-especializados/212301:Observator io-Laboral-para-la-Educacion.

[5]   L. Bohorquez y A. Sierra, "Aproximación al concepto de empleabilidad y sus indicadores caso proyecto EMPLEAP," p. 1. Work document, 2020.

[6]   M. Felt, "Social media and the social sciences: How researchers employ Big Data analytics," *Big Data & Society,* vol. 3, no. 1, p. 1, 2016. [Online]. doi: http://dx.doi.org/10.1177/2053951716645828

[7]   B. Oliver, L. Hunt, S. Jones, A. Pearce, S. Hammer, S. Jones y B. Whelan, "The graduate employability indicators: capturing broader stakeholder perspectives on the achievement and importance of employability attributes," *Proceedings of AuQF2010: Quality in Uncertain Times,* vol. 22, pp. 89-95, 2010. [Online]. Available: https://eprints.usq.edu.au/8273/

[8]   T. Mishra, D. Kumar y S. Gupta, "Mining Students' Data for Prediction Performance," p. 1, de *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 2014. [Online]. doi: http://dx.doi.org/10.1109/ACCT.2014.105

[9]   B. V. Balaji y V. V. Rao, "Improved Classification Based Association Rule Mining," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 2, no. 5, pp. 2211-2221, 2013. [Online]. Available: https://pdfs.semanticscholar.org/4d95/77616d427ed-36d8adf1ca0f1dee56f8c4a5b.pdf

[10] G. Sughanthi y M. Ashok, "Classification Techniques for Predicting Graduate Employability," *International Journal of Information Research and Review,* vol. 4, no. 2, pp. 3798-3801, 2017. [Online]. doi: http://dx.doi.org/10.18517/ijaseit.8.4-2.6832

[11] J. Hämäläinen, S. Jauhiainen y T. Kärkkäinen, "Comparison of Internal Clustering Validation Indices for Prototype-Based Clustering," *Algorithms,* vol. 10, no. 3, p. 105, 2017. [Online]. doi: http://dx.doi.org/10.3390/a10030105

[12] J. Du, Y. Liu, Y. Yu y W. Yan, "A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms," *Algorithms,* vol. 10, no. 2, p. 57, 2017. [Online]. doi: http://dx.doi.org/10.3390/a10020057

[13] M. Sapaat, A. Mustapha, J. Ahmad, K. Chamili y R. Muhamad, "A Classification-Based Graduates Employability Model for Tracer Study by MOHE," p. 1, *Communications in Computer and Information Science,* vol. 188, 2011. [Online]. doi: http://dx.doi.org/10.1007/978-3-642-22389-1_25

[14] T. Wang, C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl y P. MacNeille, "A Bayesian Framework for Learning Rule Sets for Interpretable classification," *Journal of Machine Learning Research,* vol. 18, no. 1, pp. 2357-2393, 2017. [Online]. Available: http://www.jmlr.org/papers/volume18/16-003/16-003.pdf

[15] A. A. Woya, "Employability among Statistics Graduates: Graduates' Attributes, Competence, and Quality of Education," *Education Research International,* p. 7, 2019. [Online]. doi: http://dx.doi.org/10.1155/2019/7285491

[16] S. d. Régules, *Caos y complejidad: la realidad como un caleidoscopio,* p. 1, Barcelona: Shackleton Books, 2019.

[17] Alianza del Pacífico, *¿Qué es la Alianza del Pacífico?,* p. 1, 05 08 2019. [Online]. Available: https://alianzapacifico.net/que-es-la-alianza/.

[18] Emple-AP, *Observatorio para la inserción laboral y fortalecimiento de la empleabilidad en países de la alianza del pacífico*, p. 1, 20 04 2019. [Online]. Available: http://www.emple-ap.com/es/emple-ap.

[19] J. C. Neffa, D. Panigo, P. Pérez y J. Persia, A*ctividad, empleo y desempleo: Conceptos y definiciones*, p. 1, Cuarta edición ed., Buenos Aires: Centro de Estudios e Investigaciones, 2014.

[20] E. Saito y T. Pham, "A comparative institutional analysis on strategies that graduates use to show they are 'employable': a critical discussion on the cases of Australia, Japan, and

Vietnam," *Higher Education Research & Development,* vol. 38, no. 2, pp. 369-382, 2019. [Online]. doi: http://dx.doi.org/10.1080/07294360.2018.1529024

[21]   L. E. Santana Vega, O. Gonzalez-Morales y L. F. García, "Percepción del empresariado de las competencias y características relevantes para el empleo," *Revista Española de Orientación y Psicopedagogía,* vol. 27, no. 1, pp. 29-46, 2016. [Online]. doi: http://dx.doi.org/10.5944/reop. vol.27.num.1.2016.17006

[22]   J. Riquelme, R. Ruiz y K. Gilbert, "Minería de Datos: Conceptos y Tendencias," *Revista Iberoamericana de Inteligencia Artificial,* vol. 10, no. 29, pp. 11-18, 2006. [Online]. Available: https://idus.us.es/xmlui/handle/11441/43290

[23]   L. Chen, "Statistical and Computational Methods for High-Throughput Sequencing Data Analysis of Alternative Splicing," *Statistics in Biosciences,* vol. 5, no. 1, pp. 138-155, 2013. [Online]. doi: http://dx.doi.org/10.1007/s12561-012-9064-7

[24]   Y. Cai, "Graduate employability: A conceptual framework for understanding employers' perceptions," *Higher Education,* pp. 457-469, 2013. [Online]. doi: http://dx.doi.org/10.1007/ s10734-012-9556-x

[25]   S. Jun, S.-J. Lee y J.-B. Ryu, "A Divided Regression Analysis for Big Data," *International Journal of Software Engineering and Its Applications,* vol. 9, no.° 5, pp. 21-32, 2015. [Online]. doi: http:// dx.doi.org/10.14257/ijseia.2015.9.5.03

[26]   P. Daas, M. Puts, B. Buelens y P. Van den Hurk, "Big Data as a Source for Official Statistics," *Journal of Official Statistics,* vol. 31, no. 2, pp. 249-262, 2015. [Online]. doi: http://dx.doi. org/10.1515/JOS-2015-0016

[27]   Universidad Distrital Francisco José de Caldas, *Objetivos - Direccionamiento Estratégico*, p. 1, 2017. [Online]. Available: https://www.udistrital.edu.co/objetivos.

[28]   J. Han, M. Kamber y J. Pei, *Data mining : concepts and techniques*. Michigan: Elsevier, p. 1, 2011.

[29]   K. P. Nuci y P. Kefalas, "A Recommender System Based on Hierarchical Clustering for Cloud e-Learning," *11th International Symposium on Intelligent Distributed Computing,* pp. 235-245, 2017. [Online]. doi: http://dx.doi.org/10.1007/978-3-319-66379-1_21

[30]   S.-a. Knight y G. Halkett, "Living Systems, Complexity & Information Systems Science," p. 1, de *21st Australasian Conference on Information Systems.*, 2010. [Online]. doi: http://dx.doi. org/20.500.11937/5433