Determining The Graduates Employability Using Data Analytics Model

Menentukan Kebolehkerjaan Graduan Menggunakan Model Analitik Data

Zenon A. Matos Jr.

J. H. Cerilles State College, Philippines

*Corresponding author: zenon.matos@jhcsc.edu.ph

ABSTRACT

The graduate's employability is one of the issues in the education sector. By molding its graduates with skills and knowledge in accordance to the industry, With the number of Bachelor of Science in Information Technology graduates increasing each year, we need to visualize and predict future employment. This will ultimately inform institutions and enable them to enhance their education delivery plans. These studies used a data analytics model to forecast the employability of graduates using the linear regression algorithm and the SMOreg algorithm, predict the number of graduates in the next 3 years in the BS Information Technology program, and determine the percentage of graduates who are employed and unemployed using the data sourced from the tracer study conducted. All data preparation and modeling are done using the Weka tool. The result shows that the majority of graduates are employed, whereas only 10% are unemployed. The majority of graduates employed are female, and 40% are male. Additionally, it is evident that there is an increasing pattern of the graduate's employability and that there will be approximately a 14% increase in the employment rate for the next 3 years.

Keywords: Crisp-DM, data analytics model, forecasting, machine learning, tracer study

ABSTRAK

Kebolehkerjaan graduan merupakan salah satu isu dalam sektor pendidikan. Dengan membentuk graduan dengan kemahiran dan pengetahuan yang selaras dengan industri, Dengan peningkatan jumlah graduan Ijazah Sarjana Muda Sains dalam Teknologi Maklumat setiap tahun, kita perlu membayangkan dan meramalkan pekerjaan di masa depan. Ini pada akhirnya akan memberi maklumat kepada institusi dan membolehkan mereka meningkatkan pelan penghantaran pendidikan mereka. Kajian ini menggunakan model analisis data untuk meramalkan kebolehkerjaan graduan menggunakan algoritma regresi linear dan algoritma SMOreg, meramalkan bilangan graduan dalam program Teknologi Maklumat BS dalam masa 3 tahun akan datang, dan menentukan peratusan graduan yang bekerja dan yang tidak bekerja menggunakan data yang diperoleh dari kajian penyurih yang dijalankan. Semua persediaan data dan pemodelan dilakukan menggunakan alat Weka. Hasilnya menunjukkan bahawa majoriti graduan bekerja, manakala hanya 10% yang tidak bekerja. Majoriti graduan yang

bekerja adalah perempuan, dan 40% adalah lelaki. Tambahan pula, jelas terlihat bahawa terdapat corak peningkatan dalam kebolehkerjaan graduan dan bahawa kadar pekerjaan akan meningkat kira-kira sebanyak 14% dalam masa 3 tahun akan datang.

Kata kunci: Crisp-DM, model analitik data, meramalkan, pembelajaran mesin, kajian penjejakan

INTRODUCTION

Academic institutions are one of the change agents that assist the people in the Philippines, (Guiamalon 2022) in advancing their social and economic standing. The state colleges and universities are trying to create an instant synergy between the needs and purposes of their education and training system to produce competent graduates that are ready for industry jobs. Employment matching is crucial for the industry, particularly as we transition to the Industry 4.0 framework. Industry 4.0 is the culmination of the digital transformation in how businesses produce, enhance, and distribute their goods by incorporating new technologies like the Internet of Things (IoT), cloud computing and analytics, as well as Artificial Intelligence (AI) and machine learning, into their manufacturing processes and across their entire organization. Both the national economy and the local and regional labor markets may be taken into account. This is largely due to a controversial international consensus that claims that in order to address the difficulties of a much more dynamic and rapidly changing world, individuals and organizations must regularly adopt new skills and knowledge.

Understandably, the greatest concern of the majority of our state college and university students is their ability to find work after graduation and in the long run (Ramirez, 2014). Thus, determining the graduate's employability and their future job employment is necessary for the institutions; employability is one way that higher education institutions can tell if their programs are relevant (Abarro 2021) to the industry's needs. Thus, the knowledge discovery and decision-making have greatly benefited from data mining and analytics in the process sector. Machine learning functions as a computational engine for data mining and analytics and provides fundamental tools for information extraction, data pattern detection, and forecasting (Ge et al 2017). The study aimed to conduct a tracer study on the graduate employability of J.H. Cerilles State College to
1. determine the percentage of graduates who are employed and unemployed;
2. predict the number of enrollees; and
3. forecast the future job employment of the BS Information Technology graduates
for the next 3 years using a data analytics model.

The J.H. Cerilles State College is the only state college in the province of Zamboanga Del Sur, with its mandate to provide quality, innovative, and culture-sensitive programs in the community. It has 3 organic campuses and 37 extension classes in the province of Zamboanga del Sur, offering 22 undergraduate degree programs and 4 graduate programs. The school is looking for some ways to further enhance their curriculum to equip their graduates with knowledge and skills that will match the needs of the industry. Considering the growing number of graduates every year in the Bachelor of Science in Information Technology program, the institution needs to visualize and predict future job employment, which will eventually give inputs to the institution for them to strengthen their plans in instruction delivery.

RELATED WORKS

The findings of a tracer study are critical for identifying and resolving curriculum problems, reasons why graduates are unemployed, graduates' employment status, monthly wage, and length of job search (Albina & Sumagaysay 2020), academic institutions are one of the change agents that assist the people in the Philippines in advancing their social and economic standing (Guiamalon 2022). Specifically, graduates of any school are, of course, expected to gain expertise in their chosen field and the confidence to pursue new opportunities and employment, particularly given the intensifying rivalry among coworkers (Mathed 2008). Siraye et al. (2018) identified and prioritized employability skills using the Borich training needs assessment model. Graduates and supervisors view employability skills differently. Supervisors valued problem-solving abilities, information technology skills, adaptability to change, and risk-taking abilities, whereas graduates valued the ability to identify problems and make decisions in a short period of time and the high need for curriculum enhancement. The curriculum needs to be evaluated to make sure that the ideas taught in the classroom are thoroughly improved, relevant, and likely to improve employability (Cervantes & Centeno 2019). In addition, in order to improve the staff and student knowledge, abilities, and attitudes, Balingbing (2014) recommended creating and implementing a program for faculty and student development for Camarines Sur Polytechnic College BS Information Technology graduates.

The CRISP-DM methodology is widely used in data mining projects. It is an organizational process model that is not limited to a single technology and in which multiple technologies can aid the process (Schröer et al. 2021). Wowczko (2015) found that data mining techniques proved the presence of a pattern with relation to work readiness among the jobless population registered with Irish state employment agencies. Due to its independence from specific industry tools, CRISP-DM has emerged as the top technique for data mining projects. Each CRISP-DM iteration has proven to be extremely valuable, as the obtained predictive performances have improved based on the study conducted by Laureno et al. (2011) on data mining for bank direct marketing using real-world and recent data from a Portuguese bank using CRISP-DM methodology to tune the DM model results.

The presentation of a machine learning-based time-series prediction model has been used to identify the disease curve and predict the likelihood of an Covid-19 epidemic (Balli 2020). Support vector machines (SVM), multi-layer perceptron, random forests, and linear regression were used as machine learning approaches. Time is a factor that is taken into consideration and used to assess other inputs in machine learning time series. The knowledge discovery and decision-making have greatly benefited from data mining and analytics in the process sector. Machine learning functions as a computational engine for data mining and analytics provides fundamental tools for information extraction, data pattern detection, and forecasting (Ge et al 2017). For theoretical development and empirical validation, Tsai et al. (2016) used three grey prediction models the GM (1,1) model, the NGBM (1,1) model, and the grey Verhulst model to represent the forecasted growth patterns in renewable energy using regression analysis to compare the prediction models' fitness and accuracy.

The WEKA tool is popular for data mining projects since it is free software under the GNU General Public License (Schrorer et al 2021). As a result, open-source technology in the field of data mining is capable of producing high-quality models for practical applications, which enables a decrease in the cost of data mining projects during the data preparation and modeling phases (Laureno et al. 2011). The tool is used to create machine learning approaches and use them to solve real-world data mining issues if you have the tools for data pretreatment, the

implementation of many learning algorithms, and visualization tools. Linear Regression and SMOreg machine learning algorithms are readily available in Weka.

MATERIALS AND METHODS

The data mining methodology was implemented using the information discovered at various stages of the Cross Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM can provide organized guidelines and clear execution of any project because of its industry and tool independence (Wowczko 2015) and has six phases: business understanding, data understanding, data preparation, modeling, evaluation, and conclusion. All the data analysis, implementation, and processing in this study were done using the WEKA tool. The Weka tool provides built-in machine learning algorithms that help in the application of different data analysis and forecasting techniques to obtain results in an easy and flexible process.

Business Understanding
The biggest challenge in evaluating job readiness proved to be the quality of the dataset and the process of graduate registration. The graduate will be evaluated based on the online forms, printed forms handed out, the interview, and the data provided during registration. Therefore, the judgments result in whether the graduates require further training or are ready to work. The CRISP-DM was used to achieve the following objectives: determined the graduate's employability as employed or unemployed; applied a data analytics model to predict the number of graduates in the BS Information Technology program and the graduate's employability in order to determine the underlying patterns for the next 3 years.

Data Preparation
The dataset used in this research was from the tracer study of BS Information Technology graduates of J.H. Cerilles State College. The gathering of data was done through online forms, handing out printed forms, and/or during the interview. The sample was derived from 5 academic years of graduates from 2016 through 2020. These include 9 graduate profiles such as the name of the respondent, gender, occupation classification, type of organization, year of graduation, name of organization, employment type, monthly salary, and employment status. Additionally, the organization is categorized into four categories: public, private, NGO, and non-profit organization. In the sample raw data displayed in Figure 1, 1380 instances related to graduate profiles are used.

FIGURE 1. Sample of raw data collected

Data Understanding

The data analysis was done using the WEKA tool since it is open-source technology in the field of data mining and capable of producing high-quality models for practical applications. The elements with a large number of missing values, suc
h as the monthly wage, organization address, and if this is your first employment after graduation, are deleted during the initial data assessment with the expert domain. The nature of the job and the number of years in the company are also discarded. The number of remaining attributes is only 7; these include:

1. Name of Respondent;
2. Year Graduated;
3. Name of Organization;
4. Occupation;
5. Employment Type;
6. Employment Status; and
7. Gender.

The remaining data is now converted into CSV file format and uploaded to the Weka tool for preprocessing. As indicated in Table 1, there are 1541 total graduates in the year 2016 up to 2020 in the BS Information Technology program, wherein 64.2% or 989 of graduate students are female while 552 or 35.8% are male.

TABLE 1. Profile of graduates in terms of gender

Year Graduated	Female		Male		Total
2016	172	60.8%	112	39.2%	286
2017	177	63.9%	100	36.1%	277
2018	255	65.9%	132	34.1%	387
2019	230	63%	135	37%	365
2020	153	67.7%	73	32.3%	226
Total	**989**	**64.2%**	**552**	**35.8%**	**1541**

TABLE 2. Profile of graduates employed

Year	# of respondents who responded		Total Respondents
2016	260	91%	286
2017	270	97%	277
2018	370	96%	387
2019	300	82%	365
2020	180	80%	226
Total	**1380**	**90%**	**1541**

The profile of graduates who responded to the tracer study is shown in Table 2. The table shows that 90% or 1380 out of 1541 respondents responded to the tracer study. Thus, there were only 10% of respondents who did not respond and were considered unemployed. Due to the COVID-19 pandemic, the number of employed 2019 and 2020 graduates is substantially declining; in 2019, only 82% of graduates were employed, and in 2020, only 80% of graduates were employed.

Modeling

The data analytics methodology being utilized to assess a graduate's employability was provided in this part. The obtained data is used as input to develop trends using time series analysis, and it is also used to generate forecasted graduate employability using Linear Regression and the SMOreg algorithm. The researcher chooses regression analysis because it allows you to understand the strengths of relationships between variables and can give a confidence interval for each regression coefficient that it estimates. SMOreg implements the support vector machine for regressions (Yang et al. 2007) used to forecast two typical time series models. The parameters can be learned using various algorithms by setting the regoptimizer. The time series analysis, linear regression, and SMOreg functions is executed through the Weka tool. The parameters included in the algorithms are

1. number of time units to forecast;
2. Time stamp;
3. Periodicity; and
4. Confidence intervals (95%).

The purpose of using machine learning in this study is not to compare the differences between linear regression and SMOReg methods, but rather to demonstrate the correctness and validity of the data being evaluated.

Figure 2 depicts the graduate employability plot, which shows that 90% of all graduates are in the workforce, with only 10% jobless. Thus, 548 of the graduates who are male are employed, compared to 832 graduates who are female, as shown in figure 3.
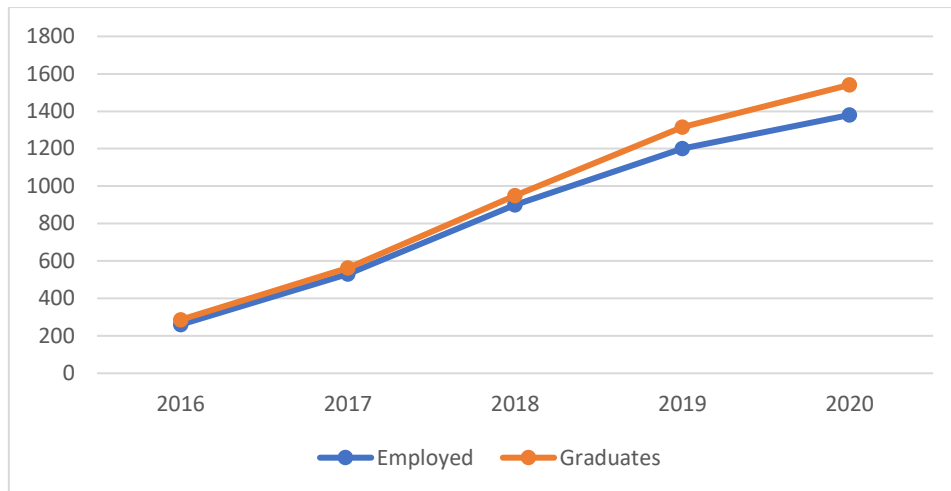
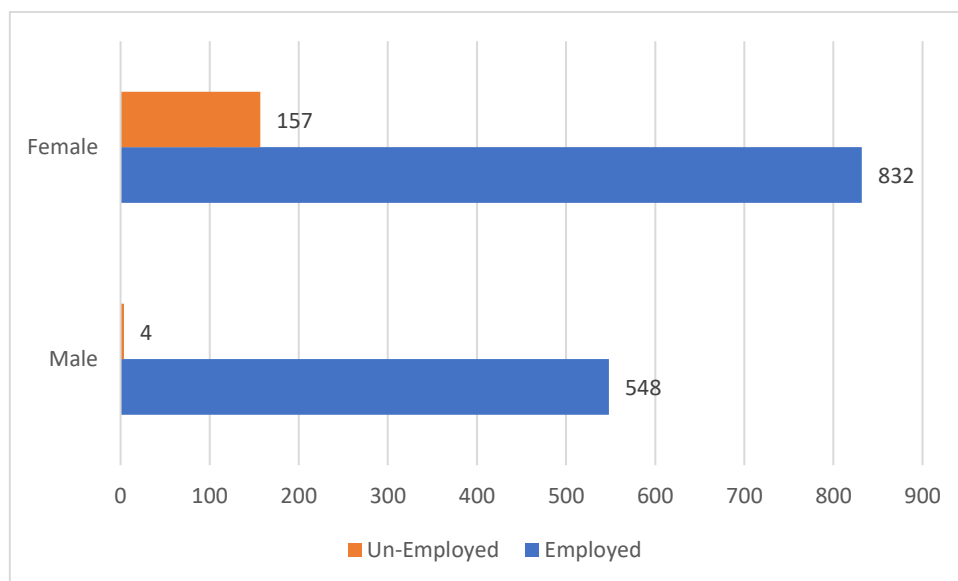FIGURE 2. Time series plot on graduate employability

FIGURE 3. Graduate's employment status by Gender

Predicting the employability of graduates at academic institutions has been shown to help assess their performance and determine what precautions to take. The time series plot presented below shows the trend in the graduate employment status of the BS Information Technology program at J.H. Cerilles State College from the year 2016 to 2020 using linear regression and the SMoreg algorithm, wherein the blue line plot indicates the number of graduates in the BS Information Technology program and the red line plot indicates the number of graduates employed. Additionally, the blue dotted line indicates the predicted number of graduates, and the red dotted line indicates the predicted number of graduates to be employed for the next 3 years in the BS Information Technology program of the J.H. Cerilles State College.
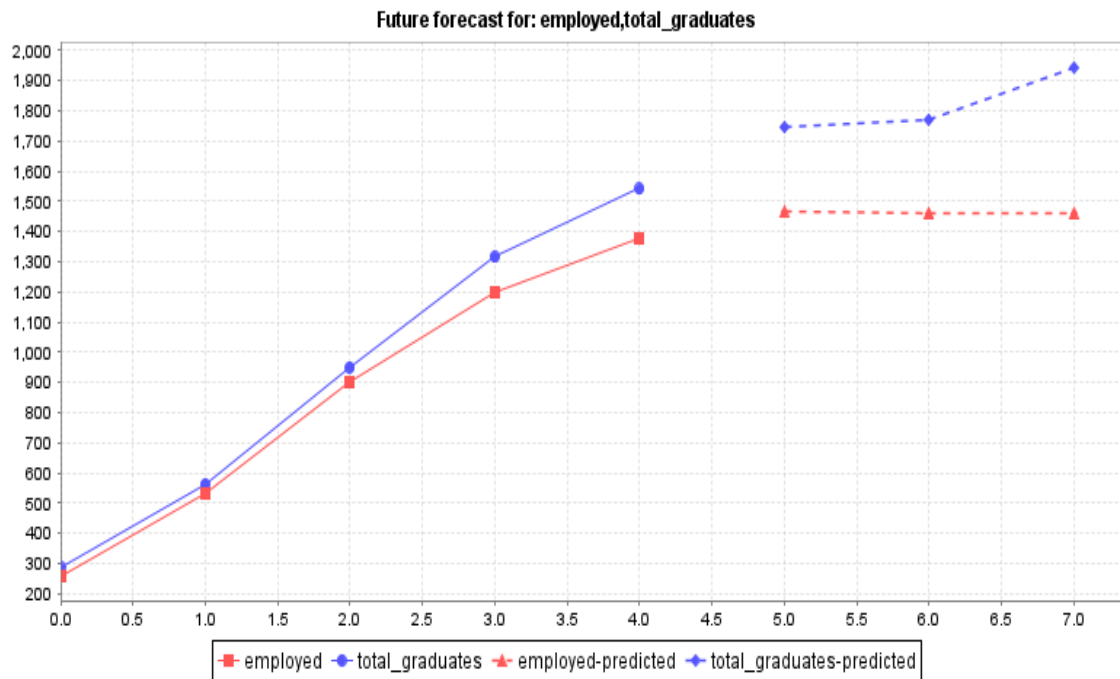
FIGURE 4. Graduate's employability using Linear Regression

The employability of BS Information Technology program graduates is projected in Figure 4 using a linear regression algorithm over the course of 3 years. The y-axis represents the number of graduates employed, and the x-axis represents the year gaps. The increase percentage is calculated based on the number of current employees divided by the number of previous employees, multiplied by 100 ((current_employed/prev_employed)*100). The graph illustrates the trend that there will be a 6% rise in 2021 and no increase in 2022 or 2023 in terms of the employability of graduates of the BS Information Technology program. Furthermore, there will be an approximately 10% increase in the number of graduates in the BS Information Technology program in the next 3 years.
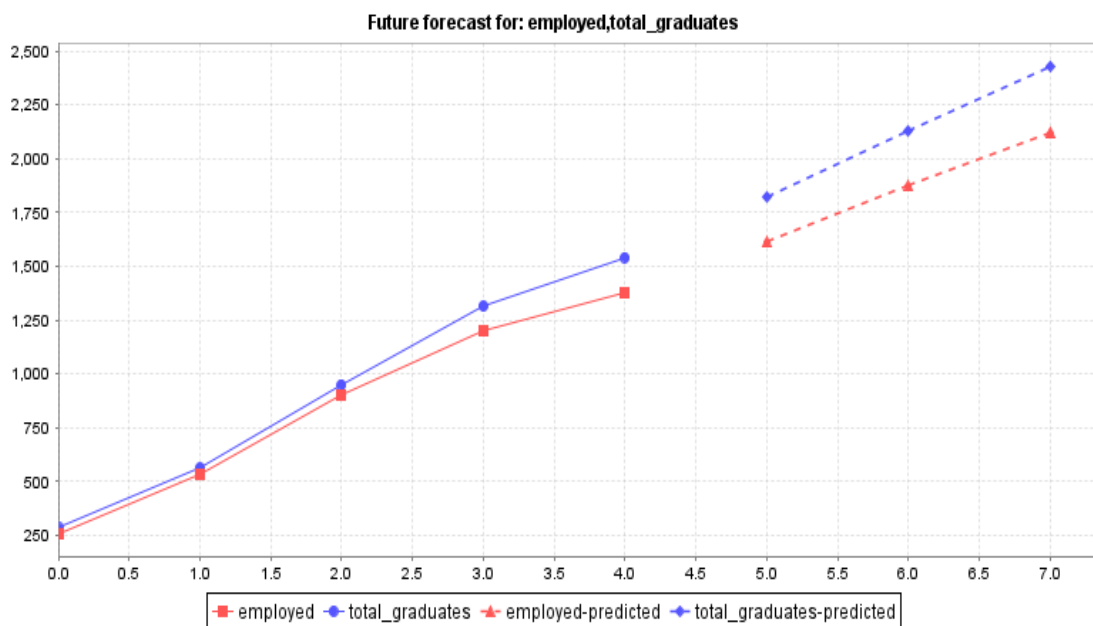

FIGURE 5. Graduate's employability using SMOreg

Figure 5 shows the future forecast of a graduate's employability for the BSIT program using the SMOREG algorithm for the next 3 years. The y-axis represents the number of graduates employed, and the x-axis represents the year gaps. The increase percentage is calculated based on the number of current employees divided by the number of previous employees, multiplied by 100 ((current_employed/prev_employed)*100). The graph displays the trends, showing that there will be an approximately 16% rise in the years 2021 and 2022 and a 13% increase in the year 2023 in terms of graduates' employability for BS Information Technology programs. Additionally, there is an approximately 16% increase in the number of graduates in the BS Information Technology program for the next 3 years.

EVALUATION AND CONCLUSION

Based on the evaluation of the research, the majority of the graduates from the year 2016 to 2020 in the BS Information Technology program are employed, only 10% of graduates are unemployed, and 90% of all graduates are in the workforce, wherein 548 of the graduates employed are male, compared to 832 graduates who are female. The employability of graduates and the anticipated enrollment of students in the three-year BS Information Technology program were predicted by the researcher using 2 data analytics models. The purpose of using machine learning in this study is not to compare the differences between 2 models, the linear regression and SMOReg methods, but rather to demonstrate the correctness and validity of the data being evaluated. The number of students enrolled in the BS Information Technology program is expected to increase by 13%. Consequently, it is obvious from the outcome of the study that there is a growing pattern of graduate's employability, with around 15% employability rate utilizing the SMOreg algorithm and only 5% rising when using the linear regression algorithm in the following 3 years.

It is recommended that the institution take into account asking the employer for input about the effectiveness and standing of their graduates after employment in order to overcome the competency gap. The job roles of graduates at each workplace may be taken into consideration in order to assess the relevance and alignment of their positions relative to the key job roles used as the basis for curriculum development. Furthermore, the percentage of the program's educational objectives attained or achieved by the graduates, the percentage of graduates who became successful entrepreneurs, and the number of graduates who are given awards and recognition for outstanding performance or contributions in the field of information technology may be determined.

ACKNOWLEDGEMENTS

REFERENCES

Abarro, Juan. 2017. Employability of bse and bee graduates of university of rizal system, antipolo city. *International Journal of Advanced Research.* 5. 1499-1504. 10.21474/IJAR01/6112.

Albina, A. & Sumagaysay L. (2020). Employability tracer study of Information Technology Education graduates from a state university in the Philippines. *Social Sciences & Humanities Open 2* (2020) 100055.

Balingbing, A. 2014. Tracer Study of BS in Information Technology (BSIT) Graduates of Camarines Sur Polytechnic Colleges, Nabua, Camarines Sur from 2004 to 2010. *Asia Pacific Journal of Multidisciplinary Research* Volume 2, No. 4, August 2014.

Ballı, S. 2021. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, Solitons & Fractals*, 142, 110512.

Ge, Z., Song, Z., Ding, S. X., & Huang, B. 2017. Data mining and analytics in the process industry: The role of machine learning. Ieee Access, 5, 20590-20616.

Guiamalon, Tarhata. 2022. *Social and economic development: state universities and colleges' (suc's) contribution creativity skills of the students in recycling*. 12. 104-110. 10.46360/globus.edu.220221015.

Laureno, R., et.al. 2010. Using Data Mining for Bank Direct Marketing: An Application of the Crisp-Dm Methodology.

Matthed, J. Professional graduates. www.yahoo.com.sg/status of graduates/2009/report /1.23.2008.

Schröer, C., Kruse, F., & Gómez, J. M. 2021. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.

Siraye, Zeleke & Abebe, Tiruneh & Melese, Mandefro & Wale, Teramaj & Asnakew, Zeleke. 2018. A tracer study on employability of business and economics graduates at Bahir Dar University. *International Journal of Higher Education and Sustainability.* 2. 45-63. 10.1504/IJHES.2018.10013675.

Tsai, Sang-Bing & Xue, Youzhi & Zhang, Jianyu & Chen, Quan & Liu, Yubin & Zhou, Jie & Dong, Weiwei. 2016. Models for forecasting growth trends in renewable energy. *Renewable and Sustainable Energy Reviews*. 77. 10.1016/j.rser.2016.06.001.

Wowczko, Izabela 2015. A Case Study of Evaluating Job Readiness with Data Mining Tools and CRISP DM Methodology. *International Journal for Infonomics* (IJI), Volume 8, Issue 3, September 2015.