Impact of Soft Skills Competencies to predict Graduates getting Jobs Using Random Forest Algorithm

by Adi Wibowo

Impact of Soft Skills Competencies to predict Graduates getting Jobs Using Random Forest Algorithm

1 Salamun Rohman Nudin
Doctoral Program of Information System
School of Postgraduate Studies
Diponegoro University
Semarang, Indonesia

1 *Department of Informatics Management*
Program of Vocational
State *University of Surabaya*
Surabaya, Indonesia

2 Budi Warsito
Department of Statistics
Faculty of Science and Mathematics
Diponegoro University
Semarang, Indonesia

2 *Doctoral Program of Information System*
School of Postgraduate Studies
Diponegoro University
Semarang, Indonesia

3 Adi Wibowo
Department of Informatics
Faculty of Science and Mathematics
Diponegoro University
Semarang, Indonesia

3 *Doctoral Program of Information System*
School of Postgraduate Studies
Diponegoro University
Semarang, Indonesia

Abstract— **Soft skills competencies is a special skill by a person in the field of self-management. This competency is very much needed by higher education in getting a job in the era of industry 4.0 and society 5.0. Analysis of the impact of soft skills competencies is currently not widely carried out, even though the tracer study instrument for higher education already exists. Prediction of graduate readiness to enter the work world is very important for higher education. It is necessary to carry out further analysis in this regard. The focus of this research is to analyze the relationship between soft skills competencies and the absorption of graduates in getting a job before six months using the random forest algorithm. The results of the testing using confusion matrix in the development of this prediction model resulted an accuracy of 64%.**

Keywords—competencies, soft skills, prediction, random forest algorithm, graduates

I. INTRODUCTION

In the era of industry 4.0 and society 5.0, higher education is not only required to create graduates who have expertise in accordance with the field of science or study program, namely hard-skill competence. However, higher education is also required to improve soft skills competencies for students [1].

Soft skills competencies are special skills by a person, including adapting to change, communicating with different groups of people, and working with people from different backgrounds [2]. In addition, communication, analytical, problem solving, critical thinking, visual, and oral skills are soft skills competencies [3]. In this research, there are 16 soft skills attributes that are used to predict the success of graduates in getting jobs which can be seen in table 1. The development of these soft skills attributes is based on the Ministry of Education, Culture, Research, and Technology tracer studies instruments in 2021.

Tracer Study (TS) is an effective media used to track the absorption of a higher education graduate in the work world. In addition, TS can be used to track the existence and condition of graduates after 1 and 2 years of graduation. TS also has an important role to collect various information as material for evaluation and development of higher education

institutions. TS results can be an illustration of the existence of a higher education [1]. Dataset of tracer study used for recommendations in building predictive models in this study is TS data for first and second cohort graduates at a higher education based on Lembaga Pendidikan Tenaga Kependidikan (LPTK).

In previous research, TS can be used to track graduate movement after graduation using the Mobility of the Highly Skilled method [4]. Movement of graduates can be three categorized: 1) cross-MSA (Metropolitan Statistical Area) moves, 2) cross-state moves, and 3) cross-regional moves with Geographic Mobility method [5]. So that higher education can find an graduate location and do geographic mapping.

TS Results show that in the era COVID-19 pandemic, the condition of nursing graduates just completed a professional program has worsened. There are three things that make concerned: 1) difficulty of getting a job, 2) delays in licensing, and 3) ambivalence of the profession in the public [6]. Results of interviews with 14 graduates from nursing students revealed that graduates from nursing students went through three specific stages during the program: 1) shocking reality, 2) accepting challenges, and 3) achieving desired results [7].

Currently, higher education has more difficult task than before the covid 19 pandemic. Higher education needs to analyze the readiness of graduates to enter the work world. Prediction model is a tool that can be used as a decision support system in the future. The Prediction Model can be used to provide recommendations for better learning outcomes in the education [8].

Prediction models in the education can be developed using technology and methods as Educational Data Mining [9] and Educational Big Data [10]. The technology used utilizes python programming [11]. The method used Random Forest Algorithm for the development of a predictive model for graduate absorption in the work world based on soft skills competencies upon graduation. The random forest algorithm is one of the solutions in developing models for prediction [12] and classification [11].

The results of this research are expected to provide recommendations to higher education in determining policies. So that it can improve the quality of graduates in the future.

II. METHODOLOGY

A. Data Analysis

Dataset in this research are the results of the 2019 and 2020 graduate tracer study surveys at higher education based on LPTK. The total data is 4549 records consisting of 16 attributes and 1 target. Attributes in this study are soft skills competencies by graduates when they graduate can be seen in table 1. The target is the waiting time for graduates to get a job.

TABLE I. ATRIBUT OF SOFT SKILLS COMPETENCIES

No	Atribut
1	Ethic
2	Communication
3	Cooperation
4	Time Management
5	Negotiation
6	Tolerance
7	Adaptation
8	Loyalty
9	Integrity
10	Leadership
11	Initiative
12	Responsibility
13	Analysis
14	Problem Solving
15	Critical Thinking
16	Self-Development

The questions used in this study consisted of two things: target and attribute. The target question is "Did you get a job <= 6 months or worked before graduating?" The answer to this question consists of two: (0) Yes and (1) No. The question related to soft skills competency in tracer study, "At the time of graduation, at what level did you master the competencies below?". The answers to these questions use a linkert scale consisting of five levels that can be seen in table 2.

TABLE II. SCALA LINKERT OF SOFT SKILLS COMPETENCIES

Scala	Description
1	Not Very Good
2	Not Good
3	Pretty Good
4	Good
5	Very Good

B. Random Forest

The ensemble method for classification and regression is called Random Forest. The ensemble method is an algorithm used to find the best prediction solution to significantly improve prediction performance. The Random Forest model is an extension of the Decision Tree model [12]. In general, the development of a Decision Tree Model consists of 3 steps:

1. Define attributes and targets.
2. Application of the split algorithm to attributes for separate the dataset into child nodes.
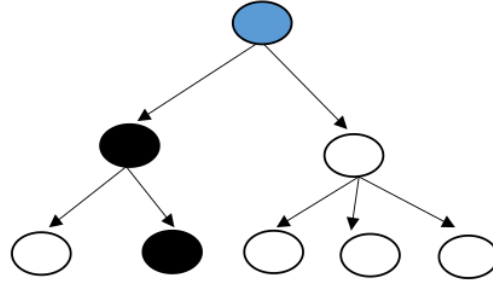3. Further separation, child nodes will be treated as parent nodes.

Fig. 1. Decision Tree Algorithm

C. Random Forest Model

There are two methods used in development the Random Forest model: bagging and random subspace. The process of taking bootstrap sampling and combining the learned model on each bootstrap sample is called bagging. Meanwhile, bootstrapping is a statistical random sampling method with substitutions to handle unbalanced data. The nodes in each decision tree in the Random Forest model are randomly assigned to select attributes in the random subspace. The application of bagging and random subspaces in the Random Forest model can handle overfitting better than a single decision tree [12]. In general, the development of an Random Forest Model consists of 4 steps:

1. Method of bootstrap as a random sample that has the same size as training dataset or full dataset.
2. Method of random subspace to chosen K attributes from a total of M attributes where $K << M$ (usually K is select equal to the square root from M)
3. Construct a decision tree with bootstrap sample and selecting attributes from steps 1 and 2.
4. Repeat steps 1 to 3 to develop the number of trees until the desired Random Forest is reached. The out-of-bag (OOB) error rate was used to determine the number of trees in the Random Forest.
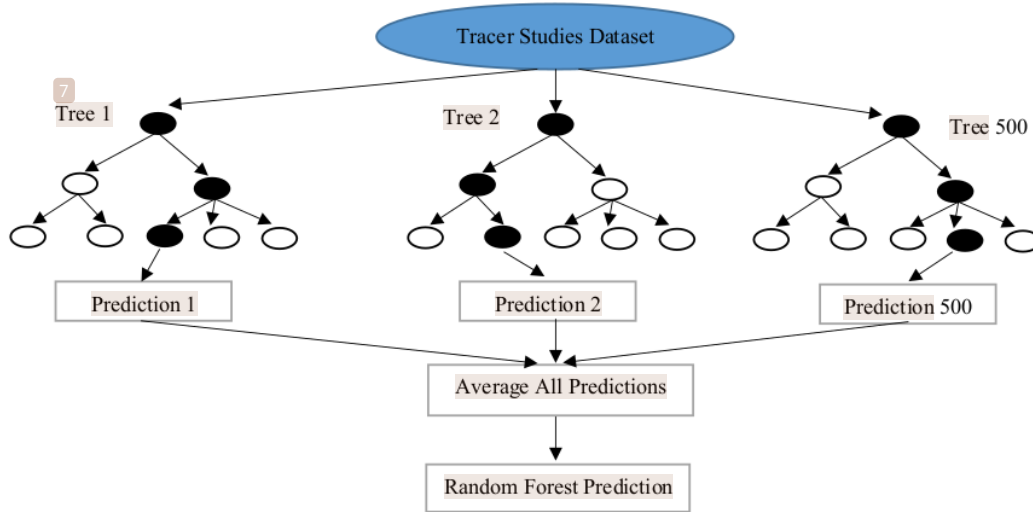
Fig. 2. Prediction of Random Forest Model

III. RESULT AND DISCUSSION

The results and discussion consist of three parts: implementation model, testing model, and analysis results.

A. Implementation Model

The model implementation in this research used python programming. Normalization of data types to integer is required before implementing the model so that the data can be processed optimally. The results of data type normalization can be seen in table 3.

Waiting time is a target that will be used as a goal in the prediction model. The targets are divided into two classifications: Yes or No. Yes, it means that graduates get a job before 6 months, while no means that graduates get a job after 6 months. In the research, the Yes is converted to 0 and not converted to 1.

B. Testing Model

Testing model to determine the reliability of the prediction model. In this research, were used 4549 records from tracer study survey that were divided into two parts: training data and testing data. Training data and testing data were separated randomly with a composition of 75% and 25%. The test results on the training data get an accuracy of 62%. The accuracy of the testing data is 64%.

C. Analysis Result

The results of the analysis consist of 2 parts: 1) target classification and 2) impact attributes on targets. The results of the classification of graduates getting a job before and after six months can be seen in Figure 3. The influence of each attribute of soft skills competence on waiting time can be seen in Figures 4 to 11. The results shown are eight of soft skills impact on the waiting time of graduates in got a job.

TABLE III.   INFORMATION OF DATASET COLUMNS

No	Column	Count	Data Type
0	Waiting Time	4549	int64
1	Ethic	4549	int64
2	Communication	4549	int64
3	Cooperation	4549	int64
4	Time Management	4549	int64
5	Negotiation	4549	int64
6	Tolerance	4549	int64
7	Adaptation	4549	int64
8	Loyalty	4549	int64
9	Integrity	4549	int64
10	Leadership	4549	int64
11	Initiative	4549	int64
12	Responsibility	4549	int64
13	Analysis	4549	int64
14	Problem Solving	4549	int64
15	Critical Thinking	4549	int64
16	Self-Development	4549	int64

Figure 3 shows the number of graduates getting a job before 6 months is 62.8% and the number of graduates getting a job after 6 months is 37.2%. In the future, it is expected that all graduates will get a job before 6 months. Improving the quality of the soft skills of graduate candidate with seminars and workshops is one way that can be used by higher education.
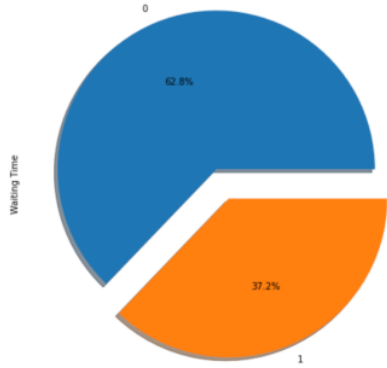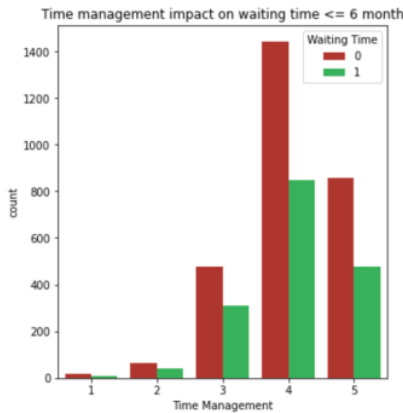
Fig. 3. Classification of Waiting Time

Fig. 4. Time Management Impact on Waiting Time

Time management is a soft skill competency that has the highest impact on graduates in getting a job. Figure 4 shows that good quality of time management the highest position and very good ranks second on the Linkert scale. Based on the figure, it can be concluded that the time management of graduates is needed in the industry.
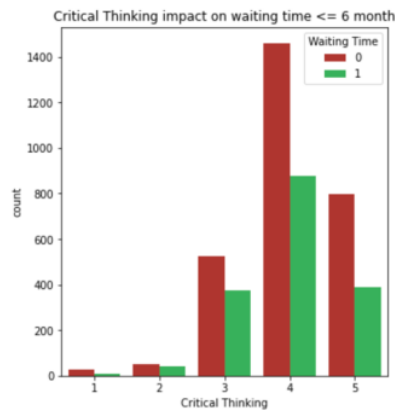

Fig. 5. Critical Thinking Impact on Waiting Time

Critical thinking is the second soft skill competency that has an impact on graduates in getting jobs. Figure 5 shows a unique phenomenon, that graduates who have good competence do not necessarily get a job on time. However, graduates have very good critical thinking soft skills are more likely to get a job before 6 months.
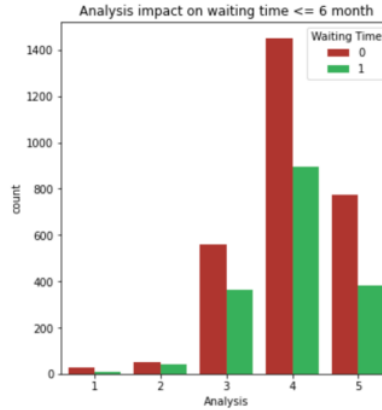

Fig. 6. Analysis Impact on Waiting Time

Analysis is the third soft skill competency that has an impact on graduates in getting a job. Figure 6 shows a unique phenomenon, that graduates who have good analytical competence do not necessarily get jobs on time. although graduates who have good analytical soft skills more chances to got a job before 6 months.
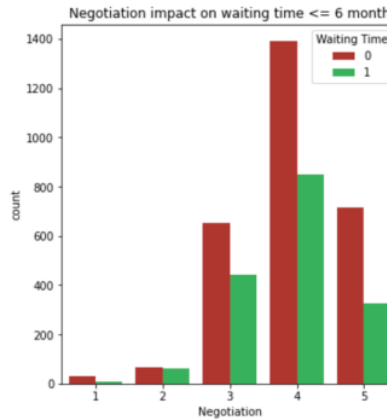

Fig. 7. Negotiation Impact on Waiting Time

Currently, graduates have very good negotiation soft skills are more likely to get a job before 6 months. Negotiation is the fourth soft skill competency that has an impact on graduates in getting a job. Figure 7 shows a unique phenomenon, that graduates who have good negotiation competence do not necessarily got a job on time. although graduates who have good negotiation soft skills more chances to got a job before 6 months.
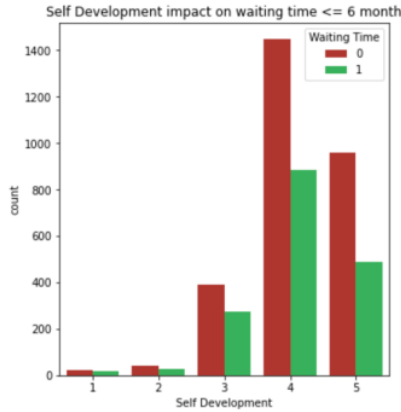
Fig. 8. Self Development Impact on Waiting Time

Fig. 10. Adaptation Impact on Waiting Time

Self-development is the fifth soft skill competency that has an impact on graduates in getting a job. Figure 8 shows that the quality of good self-development the highest position and very good ranks second on the Linkert scale. Based on the figure, it can be concluded that the self-development of graduates is needed in the industry. Currently, university graduates must be able to carry out self-development in order to complete work in accordance with industry expectations. Self-development improvement of graduates higher education can by participating in seminars and workshops on the topic of talent and personality.

Technological advances in the industrial era 4.0 in supporting work are very fast. So that the ability to adaptation in work is needed by graduates of higher education. Adaptation is the seventh soft skill competency that has an impact on graduates in getting a job. Figure 10 shows that the quality of good communication the highest position and very good ranks second on the Linkert scale. Based on the figure, it can be concluded that the adaptation of graduates is needed in the industry.

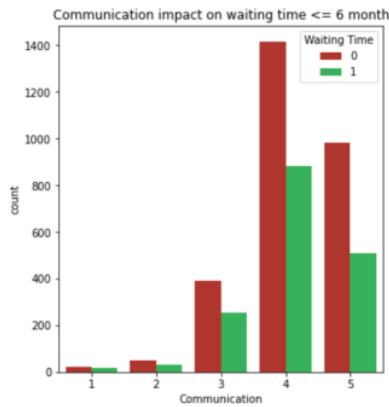Fig. 11. Leadership Impact on Waiting Time

Fig. 9. Communication Impact on Waiting Time

Today, communication is an important part of teamwork in formal work or informal work. Improving communication skills by graduates can be done by participating in seminars and workshops with the topic of public speaking. Communication is the sixth soft skill competency that has an impact on graduates in got a job. Figure 9 shows that good communication quality the highest position and very good ranks second on the Linkert scale. Based on figure, it can be concluded that graduate communication is very much needed in the work.
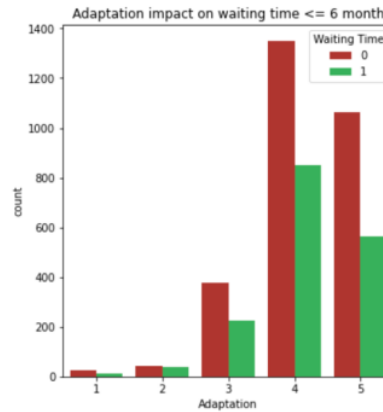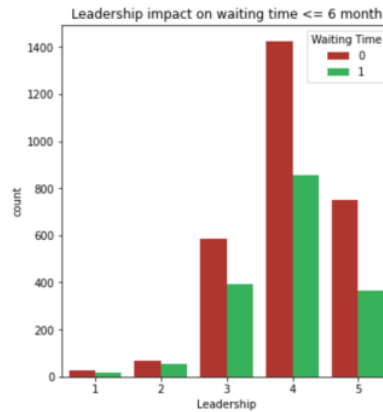
Leadership is the eighth soft skill competency that has an impact on graduates in getting a job. Figure 11 shows a unique phenomenon, that graduates who have good leadership competence do not necessarily got a job on time. However, graduates have very good leadership soft skills are more likely to get a job before 6 months.

Based on Figures 4 until 11, it can be concluded that soft skills competencies of graduates greatly impact the uptake of graduates in the industry. The better soft skills competence, the higher chance to got jobs on time before 6 months. Higher education institutions have an obligation to improve the soft skills competencies of graduates candidate in order to get a job on time.

TABLE IV. PRECISION AND RECALL

	Precision	Recall	F1-score	Support
0	0.64	1.00	0.78	733
1	0.83	0.01	0.02	403
Accuracy			0.65	1138
Macro avg	0.74	0.51	0.40	1138
Weighted avg	0.71	0.65	0.51	1138

Table 4 above is the result of testing the development of a prediction model using a precision and recall. Prediction of graduates got jobs before 6 months get 64% precision and graduates got jobs after 6 months get 83% precision.
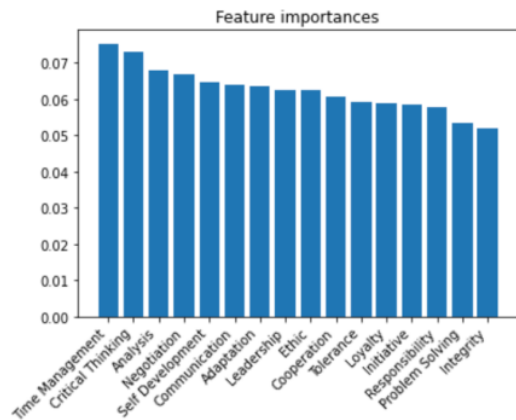

Fig. 12. Importance Soft Skills Competencies Impact on Waiting Time

Figure 12 above, it can be explained that the attribute that has a big impact is time management. The attribute that has little impact is integrity. This can be a recommendation for higher education leaders in decisions making to upgrade the quality of education and learning.

IV. CONCLUTIONS

The results of the development research of a prediction model using the random forest algorithm on 16 soft skills competencies attributes and 1 target present that the prediction accuracy 64%. The dataset used is TS for first and second cohort graduates at higher education based on LPTK. The accuracy results are still not optimal to be used as a predictive model. Future research is expected to improve prediction accuracy using other data mining algorithm or methods.

V. REFERENCES

[1] B. Heriyadi, N. Jalinus, N. Syah, U. Verawardina and T. E. Panggabean, "Tracer Study Analysis for the Reconstruction of the Mining Vocational Curriculum in the Era of Industrial Revolution 4.0," *Turkish Journal of Computer and Mathematics Education,* vol. 12, no. 3, pp. 3013-3019, 2021.

[2] S. Asonitou and T. Hassall, "Which skills and competences to develop in accountants in a country in crisis," *The International Journal of Management Education,* vol. 17, no. 3, pp. 1-19, 2019.

[3] N. Gunarathne, S. Senarathe and R. Herath, "Addressing the expectation–performance gap of soft skills in management education: An integrated skill-development approach for accounting students," *The International Journal of Management Education,* vol. 19, pp. 1-16, 2021.

[4] M. Kahn, T. Gamedze and J. Oghenetega, "Mobility of Sub Sahara Africa Doctoral Graduates from South Africa - A Tracer Study," *International Journal of Educational Development,* vol. 68, pp. 9-14, 2019.

[5] Z. He, N. Zhen and C. Wu, "Measuring and exploring the geographic mobility of American professors from graduating institutions: Differences across disciplines, academic ranks, and genders," *Journal of Informatics,* vol. 13, pp. 771-784, 2019.

[6] D. Crismon, K. J. Mansfield, S. O. Hiatt, S. S. Christensen and K. G. Cloyes, "COVID-19 pandemic impact on experiences and perceptions of nurse graduates," *Journal of Professional Nursing,* vol. 37, pp. 857-865, 2021.

[7] D. Alshael, M. Albloushi, A. Aldawsari, A. Alhamed, H. Al-Anazi, B. A. Alenazy and E. Alzaharni, "Saudi Nursing Students' Graduate Program Experiences from Reality Shock to Reaching Desired Outcomes: A Qualitative Study," *Journal of Professional Nursing,* vol. 37, pp. 1049-1056, 2021.

[8] A. A. Mubarak, H. Cao and I. M. Hezam, "Deep analytic model for student dropout prediction in massive open online courses," *Computers and Electrical Engineering,* vol. 93, pp. 1-14, 2021.

[9] Y. K. Salal, S. M. Abdullaev and M. Kumar, "Educational Data Mining: Student Performance Prediction in Academic," *International Journal of Engineering and Advanced Technology,* vol. 8, no. 4c, pp. 54-59, 2019.

[10] X. Bai, F. Zhang, J. Li, T. Guo, A. Aziz, A. Jin and F. Xia, "Educational Big Data: Predictions, Applications and Challenges," *Big Data Research,* vol. 26, pp. 1-12, 2021.

[11] D. Dessì, G. Fenu, M. Marras and D. R. Recupero, "Bridging Learning Analytics and Cognitive Computing for Big Data Classification in Micro-Learning Video Collections," *Computers in Human Behavior,* vol. 92, pp. 468-477, 2019.

[12] X. Zhou, P. Lu, Z. Zheng, D. Tolliver and A. Keramati, "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree," *Reliability Engineering and System Safety,* vol. 200, pp. 1-9, 2020.

Impact of Soft Skills Competencies to predict Graduates getting Jobs Using Random Forest Algorithm

development approach for accounting students", The International Journal of Management Education, 2021
Publication

7   Submitted to University of Stellenbosch, South Africa   1%
     Student Paper

8   pubmed.ncbi.nlm.nih.gov   <1%
     Internet Source

9   Prathapchandran Kannimuthu, Janani Thangamuthu. "A TRUST AWARE SECURITY   <1%
     MECHANISM TO DETECT SINKHOLE ATTACK IN RPL-BASED IOT ENVIRONMENT USING
     RANDOM FOREST – RFTRUST", Computer Networks, 2021
     Publication

10   ijessr.com   <1%
     Internet Source

11   ntnuopen.ntnu.no   <1%
     Internet Source

12   www.freepatentsonline.com   <1%
     Internet Source

13   Armando E. Marques, Pedro A. Prates, André F. G. Pereira, Marta C. Oliveira, José V.   <1%
     Fernandes, Bernardete M. Ribeiro. "Performance Comparison of Parametric and
     Non-Parametric Regression Models for

Uncertainty Analysis of Sheet Metal Forming Processes", Metals, 2020
Publication

14   Submitted to Middlesex University   <1%
      Student Paper

15   Zekai He, Ni Zhen, Chaojiang Wu. "Measuring   <1%
      and exploring the geographic mobility of
      American professors from graduating
      institutions: Differences across disciplines,
      academic ranks, and genders", Journal of
      Informetrics, 2019
      Publication

16   www.semanticscholar.org   <1%
      Internet Source

17   edisciplinas.usp.br   <1%
      Internet Source

18   slidetodoc.com   <1%
      Internet Source

19   Ali Al-Zawqari, Dries Peumans, Gerd   <1%
      Vandersteen. "A flexible feature selection
      approach for predicting students' academic
      performance in online courses", Computers
      and Education: Artificial Intelligence, 2022
      Publication

20   C. E. Goodwin. "Lamprey (Lampetra fluviatilis   <1%
      and Lampetra planeri) ammocoete habitat
      associations at regional, catchment and

microhabitat scales in Northern Ireland",
Ecology Of Freshwater Fish, 4/10/2008
Publication

21	dergipark.org.tr Internet Source	<1%
22	media.fupress.com Internet Source	<1%
23	www.asianinstituteofresearch.org Internet Source	<1%
24	www.mdpi.com Internet Source	<1%
25	Ruihua Xu, Fan Luo. "Risk prediction and early warning for air traffic controllers' unsafe acts using association rule mining and random forest", Safety Science, 2021 Publication	<1%

Impact of Soft Skills Competencies to predict Graduates getting Jobs Using Random Forest Algorithm

FINAL GRADE

/0

GENERAL COMMENTS

Instructor