Available at https://journalenrichment.com/index.php/jr/

Enrichment: Journal of Multidisciplinary Research and Development

APPLICATION OF MULTIPLE LINEAR REGRESSION (MLR) METHOD IN CERTIFICATION ACTIVITIES AT ITCC ITPLN

Hendra Jatnika*, Luqman Luqman, Meilia Nur Indah Susanti, Petra Andriyani Mulyo Wibisono, Mulya Jefri
Institut Teknologi PLN, Indonesia
Email: h.jatnika@itpln.ac.id*, Petra1831028@itpln.ac.id

ABSTRACT

Multiple Linear Regression (MLR) is one of the algorithms in Machine Learning. Machine Learning that estimates the linear coefficient equations involved in one or more independent variables that can predict the value of the variable of interest. Algorithms used to predict the value of a variable based on the value of other variables. Based on 2021 data at the Information Technology Certification Center (TCC), it can be seen that the quality and quantity of Microsoft International Certification graduates is decreasing. In the pre-pandemic MOS certification, the percentage of passes was seventy-two percent (72%), while in the MOS certification during the pandemic the percentage of passes dropped to fifty-two percent (52%). Based on the results of the MLR trial test on the dataset of MOS-Word and MCF-AI certification test participants, a calculation formula is obtained as a benchmark in assessing the MOS-Word and MCF AI certification scores. . The study provides specific recommendations to optimize certification training programs by tailoring materials to address critical competencies and participant needs. Additionally, a predictive formula developed in this research can serve as a self-assessment tool for participants to evaluate their readiness for certification tests. These findings underscore the potential of MLR as a robust analytical tool for improving certification processes, enhancing training effectiveness, and ensuring better outcomes for participants. This research contributes to advancements in machine learning applications within education and professional development contexts.

Keywords: *ITCC; multiple linear regression; MOS Word; MCF AI self-assessment*

INTRODUCTION

In every implementation of the Microsoft International Certification test at ITCC, training is held, which must be attended as a prerequisite for the certification test, aiming to provide knowledge and learning related to the material in order to prepare participants facing the certification test (Buenaño-Fernández et al., 2019; Y. Huang et al., 2021; Pressman & Maxim, 2015). Trainees are required to have previous certification test material experience as a prerequisite. The implementation of training that runs at ITCC lasts several days doing practice questions or self-study, and for one day, one must attend training guided by certified and experienced trainers, training which is divided into two sessions. Each session lasts 7 hours and consists of material exposure and practice (Agtriadi et al., 2021; Jatnika et al., 2022; Jatnika, Rifai, & Primadhani, 2023). During the pandemic period, training activities were carried out online using Microsoft Teams, previously carried out onsite at the ITPLN laboratory campus, followed by the implementation of onsite certification tests in a special certification laboratory at ITCC (Purnama & Sofana, 2021; Rifai et al., 2020; Wibawanto & Ds, 2017).

Based on data from 2021, it can be seen that the quality and quantity of ITPLN student certification graduates have decreased. In the pre-pandemic certification, the percentage of passes was seventy-two percent (72%). In the pandemic certification, the percentage of passes decreased even more to fifty-two percent (52%), it is estimated that in addition to the conditions caused by the pre-certification test training pattern (Jatnika et al., 2019; Jatnika, Rifai, & Napitupulu, 2023; Jurnal, 2018).

Certification activities at ITCC ITPLN, particularly Microsoft International Certification tests, are essential for equipping students with technical competencies that align with industry standards. These certification programs require rigorous training sessions as prerequisites, which include material exposure and practice guided by certified trainers. However, during the pandemic, the transition from onsite to online training using platforms like Microsoft Teams revealed significant challenges in maintaining the quality and effectiveness of these sessions. Data from 2021 indicates a sharp decline in

certification pass rates—from 72% pre-pandemic to 52% during the pandemic—highlighting inefficiencies in the pre-certification training patterns (Arnesia et al., 2022; Kang & Zhao, 2020; Kumar et al., 2021; Maulany et al., 2021).

Previous research has demonstrated the utility of machine learning algorithms, such as Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Decision Trees, in predicting academic performance and improving educational outcomes (Novaliendry & Azkia, 2021; Rajendran & Balasubramanie, 2011; Sheldon et al., 2002). For instance, MLR has been successfully applied to estimate graduation grade point averages (GPA) based on input variables like high school GPA and demographic factors Pressman & Maxim (2015). Similarly, ANN models have shown higher predictive accuracy than MLR in estimating student performance metrics Agtriadi et al. (2021). Despite these advancements, limited studies have explored the application of MLR, specifically within certification activities, to optimize training systems and predict graduation outcomes.

This study introduces a novel application of MLR in certification training at ITCC ITPLN. By integrating self-assessment exercises with predictive algorithms, this research aims to address declining pass rates and improve training strategies. The innovation lies in leveraging MLR to analyze independent variables such as assessment scores, training duration, and trainer expertise to accurately predict certification outcomes. This approach builds upon existing literature and provides practical solutions for enhancing certification processes in educational institutions.

This study aims to address this gap by leveraging MLR to predict certification outcomes and improve training strategies at ITCC ITPLN. By analyzing factors such as assessment scores, training duration, and trainer expertise, this research seeks to provide actionable insights for improving pass rates. Beyond ITCC ITPLN, the findings have broader implications for enhancing certification systems globally, particularly in adapting to remote learning environments.

METHOD

To answer the needs of the certification and training implementation training system, a proposal was developed in the form of additional features in the existing training by applying the Multiple Linear Regression (MLR) method to the self-assessment exercises system (S. Huang, 2022; Schmidt & Finan, 2018; Wang et al., 2024).

Multiple Linear Regression is one of the algorithms in Machine Learning, an algorithm used to predict a variable's value based on another variable's value. The process is done by applying the following steps:

A. Determining the Hypothesis

Score prediction is designed based on the needs of the score calculation system, and graduation prediction is based on the performance/self-assessment scores of training and certification participants. The prediction value of the resulting pass becomes the dependent variable, with the hypothesis that the prediction results are influenced by independent factors (assessment value, duration of training, trainer, etc.).

B. Data Collection

Data taken from the ITCC database is from a series of MOS and Microsoft Certified Fundamental (MCF) certification activities, from registration and training to certification testing. MOS certification participants were 88 people, and MCF AI Fundamental participants were as many as 110.

C. Data Exploration

It is a stage in which the independent variables are understood and narrowed down. This is done by knowing the character of each feature, which will be used to test the relationship/correlation and impact on the dependent variable.

D. Data Processing

The process of tidying up the data so that it can be used as a dataset for training data and test models that will be made, data taken with a presentation of more than 70%, this process includes feature reduction by :

1. Remove fields that have many empty values
2. Removing fields with few variants
3. Removing fields that have a low level of correlation with the dependent variable

Table 1. MOS-Word filter fields and MCF-AI certification results

Activity Name		Field Name
Certification MOS-Word	**Certification MCF-AI**	Study program
		City of Origin
		Province of Origin
		Training Group
		Exam Group
		Trainer during Training
		Training Duration
		Score Analysis
		Certification Test Score
		Questionnaire
		Certification test history

Table 2. Fields Score Analysis

Program	Score Analysis
Certification MOS - Word	Manage Documents (MD)
	Insert and Format Text, Paragraphs, and Sections (IFPS)
	Manage Tables and Lists (MTL)
	Create and Manage References (CMR)
	Insert and Format Graphic Elements (IFGE)
Certification MCF-AI	Describe Artificial Intelligence Workloads and Considerations (DAIWC)
	Describe Fundamental Principles of Machine Learning on Azure (DFPMLA)
	Describe Features of Computer Vision Workloads on Azure (DFCVWA)
	Describe Features of Natural Language Processing Workloads on Azure (DFNLPWA)

E. Feature Engineering

The stage of determining the criteria that affect the target or dependent value involves reviewing the fields contained in the Dataset to determine the independent variables. This is done using the Forward Stepwise Regression algorithm.

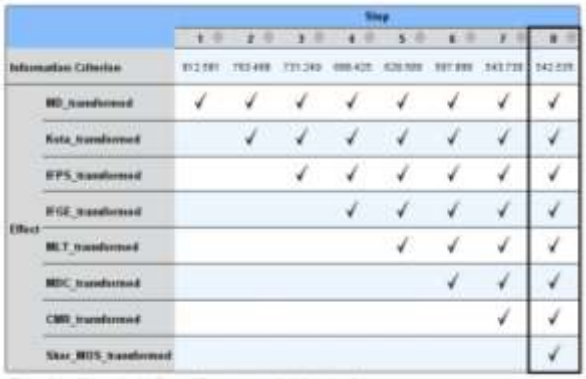

Figure 1. Model Summary of MOS-Word Certification

		Step				
		1	2	3	4	5
Information Criterion		981.623	876.977	802.159	770.333	765.044
Effect	DFPMLA_transformed	✓	✓	✓	✓	✓
	DFNLPWA_transformed		✓	✓	✓	✓
	DFCVWA_transformed			✓	✓	✓
	DAIWC_transformed				✓	✓
	Provinsi_transformed					✓

Figure 2. MCF-AI Summary Model

F. Training Model

At this stage, model training is carried out using the multiple linear regression method on the obtained variables. The transformed data will be divided into training data (70%) and test data (30%) of the total number of records; this process is carried out using IBM SPSS Statistics.

Table 3. Number of training and testing data

	Certification MOS -Word	Certification MCF-AI
Testing data (30%)	26	33
Training data (70%)	62	77
Total (100%)	88	110

Training results of MOS-Word certification data:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.990[a]	.980	.978	16.656

a. Predictors: (Constant), MDC, MLT, IFPS, CMR, IFGE, MD

Figure 3. MLR Summary Model of MOS-Word Certification

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	116.819	11.726		9.962	<,001
	MD	1.951	.139	.277	14.002	<,001
	IFPS	2.023	.120	.290	16.791	<,001
	MLT	1.954	.131	.257	14.918	<,001
	CMR	.719	.072	.174	10.046	<,001
	IFGE	1.736	.102	.315	16.995	<,001
	MDC	.563	.068	.137	8.266	<,001

a. Dependent Variable: Skor_BMPSI19

Figure 4. MLR result coefficient of MOS-Word certification

Training result of AI MCF data:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971[a]	.944	.942	29.495

a. Predictors: (Constant), DFNLPWA, DFPMLA, DFCVWA, DAIWC

Figure 5. MLR Summary Model of MCF-AI Certification

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-7.460	19.382		-.385	.701
	DAIWC	1.520	.248	.208	6.134	<,001
	DFPMLA	3.329	.228	.446	14.633	<,001
	DFCVWA	1.787	.199	.271	8.975	<,001
	DFNLPWA	3.361	.298	.300	11.273	<,001

a. Dependent Variable: Skor_MCF

Figure 6. MLR result coefficient of MCF-AI certification

Based on the Multiple Linear Regression formula:

$$Y_i = \beta_0 + \beta_1 x_1 + \varepsilon_1$$

$$Y_i = Varibel\ Dependen$$

$\beta_0 = Populasi\ pada\ axis\ y$
$\beta_1 = Populasi\ koefisien\ axis\ y\ variabel\ independe$
$x_1 = Variabel\ Independen$
$\varepsilon_1 = Komponen\ random\ error$

Then, the formula for calculating MOS-Word certification and MCF-AI certification is obtained as follows:

Table 4. MLR Formula of MOS-Word Certification Score and MCF-AI certification

MOS-Word Certification Score
$Y_{skor} = 116{,}619 + (1{,}521)MD + (2{,}023)IFPS + (1{,}954)MLT + (0.719)CMR + (1{,}736)IFGE + (0.563)MDC$
MCF AI Score
$Y_{skor=}(1.52)DAIWC + (3.329)DFPMLA + (1{,}787)DFCVWA + 3{,}361(DFNLPWA) - 7{,}460$

Model Evaluation

Model Evaluation is carried out using the Standard Error of the Regression (SER), also known as the Standard Error of the Estimate, which represents the average distance of data outside the regression line measuring the variability of the actual and estimated values of Y, where the correlation coefficient (R) value estimates the true linearity of the original data.

Table 5. MOS-Word certification table

N (Total population)	K (Number of *criterion*)	SSR	SER
62	6	20162,62	19,26

Based on the data above, it is concluded that the MLR model for MOS-Word certification assessment, with a population of 62 training data records with 6 dependent variables, has an accuracy percentage of 98% with a standard error based on the Certification value of 19.26.

Table 6. MCF-AI certification table

N (Total population)	K (Number of *criterion*)	SSR	SER
77	4	82374,72	33,59

Based on the data above, it is concluded that the MLR model for MCF-AI certification assessment, with a population of 77 training data records with 4 dependent variables, has an accuracy percentage of 94% with a standard error based on the SER value of 33.59.

Testing Model

The model testing stage is carried out by comparing the value of the prediction results on the testing dataset with the actual participant's score value. By entering the acquisition score analysis formula obtained, with the help of Microsoft Excel, rounding the results of two numbers behind the comma, resulting in the following data:

Table 7. MOS-Word certification model testing results table

No.	Participant Name	MOS Score 19	Prediction Score	Accuracy (%)
1.	Participant 1	792	782	99%
2.	Participant 2	825	831	99%
3.	Participant 3	907	908	100%
4.	Participant 4	800	807	99%
5.	Participant 5	925	935	99%
6.	Participant 6	861	857	100%
7.	Participant 7	509	525	97%
8.	Participant 8	953	958	99%
9.	Participant 9	950	960	99%
10.	10 participants	775	779	99%
11.	Participant 11	700	704	99%
12.	12 participants	884	880	100%
13.	Participant 13	861	858	100%
14.	14 participants	800	806	99%
15.	15 participants	723	705	98%
16.	Participant 16	792	802	99%
17.	17 participants	976	983	99%
18.	18 participants	792	777	98%
19.	Participant 19	792	782	99%
20.	20 participants	884	907	97%
21.	Participant 21	700	678	97%
22.	Participant 22	884	883	100%
23.	Participant 23	838	832	99%
24.	24 participants	907	911	100%
25.	25 participants	725	727	100%
26.	Participant 26	884	881	100%

Table 8. MCF-AI Certification model testing results table

No.	Participant Name	MCF Score	Prediction Score	Accuracy (%)
1	Participant 1	731	721	99%
2	Participant 2	684	684	100%
3	Participant 3	700	694	99%
4	Participant 4	905	905	100%
5	Participant 5	669	657	98%
6	Participant 6	810	797	98%
7	Participant 7	669	662	99%
8	Participant 8	639	630	99%
9	Participant 9	700	707	99%
10	10 participants	826	821	99%
11	Participant 11	936	928	99%
12	12 participants	778	777	100%
13	Participant 13	623	622	100%

14	14 participants	764	793	96%
15	15 participants	517	516	100%
16	Participant 16	639	631	99%
17	17 participants	715	718	100%
18	18 participants	747	730	98%
19	Participant 19	684	692	99%
20	20 participants	747	751	99%
21	Participant 21	608	594	98%
22	Participant 22	563	572	98%
23	Participant 23	842	872	96%
24	24 participants	826	825	100%
25	25 participants	857	850	99%
26	Participant 26	889	878	99%
27	Participant 27	654	564	86%
28	Participant 28	857	853	100%
29	29 participants	654	656	100%
30	30 participants	410	408	100%
31	31 participants	763	763	100%
32	32 participants	873	882	99%
33	Participant 33	608	609	100%

RESULTS AND DISCUSSION

In this research, information is obtained regarding the percentage of influence of each competency on the certification test score, factors that influence the certification test score, and the calculation formula for the certification test score. The influence of each factor is shown in the following graph.

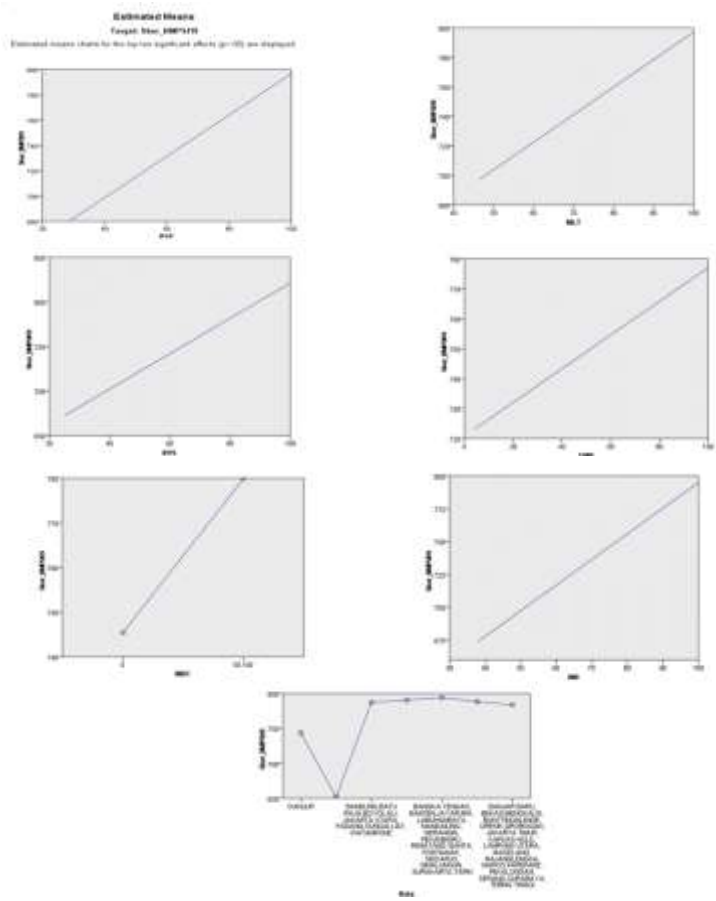

Figure 7. Forecasting results of independent variables (score analysis, city) and dependent variables (MOS -Word certification score)

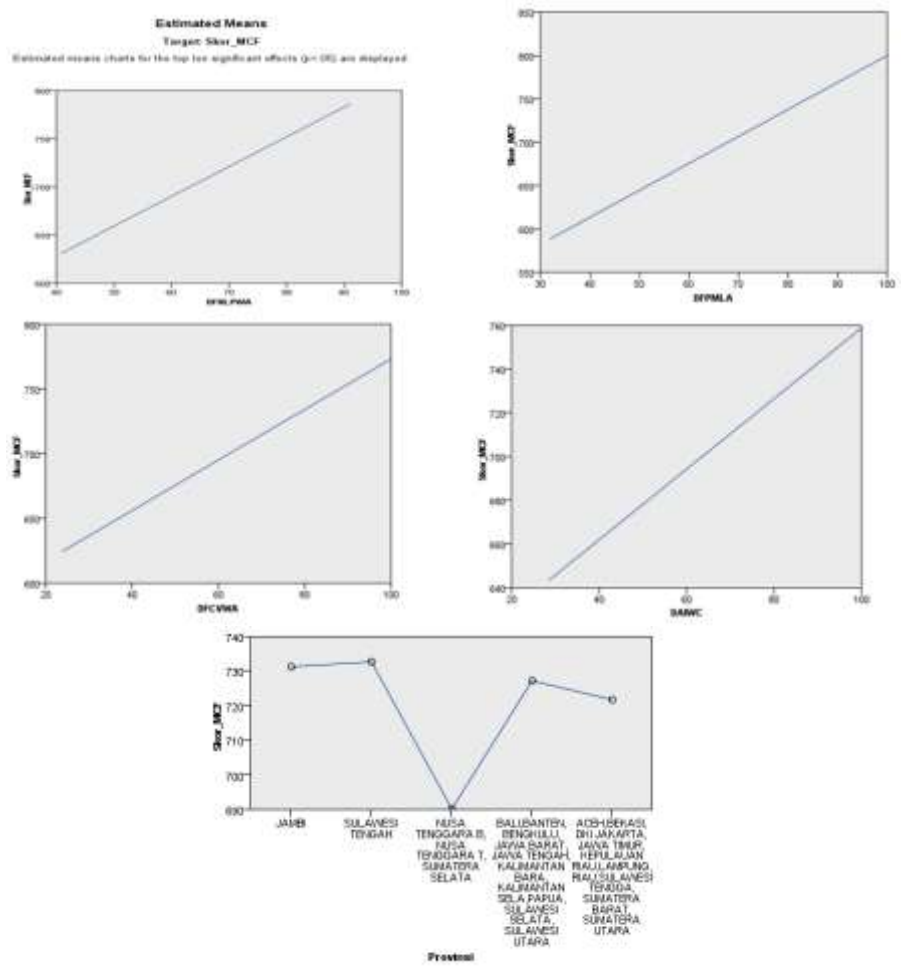

Figure 8. Forecasting results of independent variables (score analysis, province) and dependent variables (MCF-AI certification score).

The percentage of influence of each competency on the certification test score is also obtained, which can be used as a benchmark for preparing materials and self-assessment exercises, as follows:

Table 9: Percentage of MOS-Word certification competency

Criteria	Description	Percent
MDC	Manage Document Collaboration	6%
CMR	Create and Manage References	8%
IFGE	Insert and Format Graphic Elements	19%
MD	Manage Documents	22%
MLT	Manage Tables and Lists	22%
IFPS	Insert and Format Text, Paragraphs, and Sections	23%
	Total	**100%**

Table 10. Percentage of MCF-AI certification competencies

Criteria	Description	Percent
DAIWC	Describe Artificial Intelligence Workloads and Considerations	15%
DPCVWA	Describe Fundamental Principles of Machine Learning on Azure	18%
DFPMLA	Describe Features of Natural Language Processing Workloads on Azure	33%
DFNLPWA	Describe Fundamental Principles of Machine Learning on Azure	34%
	Total	**100%**

The calculation formula that can be used in calculating the self-assessment score is as follows:

Table 11. MOS-Word certification score and MCF-AI certification calculation formula

MOS Word 2019 Score
$Y_{skor} = 116{,}619 + (1{,}521)MD + (2{,}023)IFPS + (1{,}954)MLT + (0.719)CMR + (1{,}736)IFGE + (0.563)MDC$
MCF AI Score
$Y_{skor=}(1.52)DAIWC + (3.329)DFPMLA + (1{,}787)DFCVWA + 3{,}361(DFNLPWA) - 7{,}460$

CONCLUSION

Based on the results of research and discussion, several conclusions can be drawn, including a.) The application of the MLR method can provide the calculation of competency test score criteria, with accuracy above 98% and 94% for the MOS-Word certification program and MCF-AI certification, respectively. This method runs by mapping each independent variable to the dependent variable into a linear form so that it can be mapped and get a quadratic equation, which has a coefficient value that affects the results of the dependent variable being sought, namely the 2019 MOS-Word certification score and MCF-AI certification. b.) Several factors affect the results of certification test scores, namely score analysis, city and province of origin of participants. This score analysis is a benchmark for competencies tested at MOS-Word and MCF-AI certifications. c.) The calculation formula is also obtained by using score analysis as a score prediction calculation variable.

Furthermore, the study identified key factors affecting certification test scores. These include score analysis and geographic factors such as participants' city and province of origin. These findings highlight the importance of tailoring training programs to address specific participant needs and regional contexts. The score analysis, particularly, serves as a benchmark for evaluating competencies tested in MOS-Word and MCF-AI certifications.

Additionally, the research developed a predictive formula based on score analysis that can be used for self-assessment and performance prediction. This formula provides a practical tool for participants to gauge their readiness for certification tests and for trainers to refine their instructional strategies. ITCC ITPLN can enhance its certification processes and improve participant outcomes by integrating this predictive capability into the training system. In conclusion, this study illustrates the potential of Multiple Linear Regression (MLR) as a formidable analytical instrument for optimizing educational and certification systems. It enhances pass rates and offers actionable insights for designing more effective training programs specifically tailored to the needs of participants. The findings emphasize the importance of machine learning techniques in further developing education and professional certification practices.

REFERENCES

Agtriadi, H. B., Jatnika, H., Fitriani, Y., Sihabudin, Z. V., Khanib, M. N., Sembiring, D. G., & Situmorang, O. N. M. (2021). Implementation of WYSIWYG in the Development of ITCC ITPLN Letter Management Information System. *Jurnal E-Komtek (Elektro-Komputer-Teknik)*, *5*(2), 310–318.

Arnesia, P. D., Pratama, N. A., & Sjafrina, F. (2022). Aplikasi Artificial Intelligence Untuk Mendeteksi Objek Berbasis Web Menggunakan Library Tensorflow Js, React Js Dan Coco Dataset. *JSiI (Jurnal Sistem Informasi)*, *9*(1). https://doi.org/10.30656/jsii.v9i1.4243

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*, *11*(10), 2833. https://doi.org/10.3390/su11102833

Huang, S. (2022). Linear regression analysis. In *International Encyclopedia of Education: Fourth Edition*. https://doi.org/10.1016/B978-0-12-818630-5.10067-3

Huang, Y., Talwar, A., Chatterjee, S., & Aparasu, R. R. (2021). Application of machine learning in predicting hospital readmissions: a scoping review of the literature. *BMC Medical Research Methodology*, *21*(1). https://doi.org/10.1186/s12874-021-01284-z

Jatnika, H., Putra, R. I., Al-Khair, M. Z., Stendel, A. L. B., & Akbar, M. T. (2022). Application of Genetic Algorithm with Roulette Wheel Selection Method for International Certification Scheduling Optimization at ITCC ITPLN. *Jurnal E-Komtek (Elektro-Komputer-Teknik)*, *6*(2), 360–369.

Jatnika, H., Rifai, M. F., & Aritonang, I. Y. (2019). Rancang Bangun Aplikasi Sistem Registrasi Dan Penjadwalan Sertifikasi Berbasis Web Menggunakan Metode Algoritma Fisher Yates Shuffle (Studi Kasus: Laboratorium ITCC STT-PLN). *Petir*.

Jatnika, H., Rifai, M. F., & Napitupulu, L. T. E. (2023). Implementation of the Rational Unified Process (Rup) Method in Designing a Web-Based Certification Scheduling Application (Citation) on Itcc Itpln. *Syntax Idea*, *5*(4). https://doi.org/10.46799/syntax-idea.v5i4.2188

Jatnika, H., Rifai, M. F., & Primadhani, V. R. (2023). Application of iso/iec 9126 on quality measurement of web-based records management applications at ITCC ITPLN. *Jurnal Scientia*, *12*(03), 2665–2676.

Jurnal, R. T. (2018). Penerapan metode Value chain dan Four Stage Life Cycle untuk menentukan kandidat aplikasi pada Laboratorium ITCC STT-PLN. *PETIR*, *10*(2). https://doi.org/10.33322/petir.v10i2.25

Kang, H., & Zhao, H. (2020). Description and Application Research of Multiple Regression Model Optimization Algorithm Based on Data Set Denoising. *Journal of Physics: Conference Series*, *1631*(1). https://doi.org/10.1088/1742-6596/1631/1/012063

Kumar, S., Raja, R., Tiwari, S., & Rani, S. (2021). Cognitive behavior and human computer interaction based on machine learning algorithms. In *Cognitive Behavior and Human Computer Interaction Based on Machine Learning Algorithms*. https://doi.org/10.1002/9781119792109

Maulany, R., Hasan, B., Abdullah, A. G., & Rohendi, D. (2021). Design of learning applications using the Rapid Application Development method. *IOP Conference Series: Materials Science and Engineering*, *1098*(2). https://doi.org/10.1088/1757-899x/1098/2/022090

Novaliendry, D., & Azkia, A. (2021). Design Of Web-Based Infromation System For The Bung Hatta University Professional Certification Institute. *Jurnal Teknologi Informasi Dan Pendidikan*, *14*(1). https://doi.org/10.24036/tip.v14i1.446

Pressman, R. S., & Maxim, B. R. (2015). Software Engineering: a Practitioner'S Approach, Eighth Edition. In □□□□□□.

Purnama, B., & Sofana, I. (2021). Implementasi artificial intelligence dan machine learning. *Bandung: Informatika Bandung*.

Rajendran, T., & Balasubramanie, P. (2011). An efficient multi-agent-based architecture for Web Service registration and Discovery with QoS. *European Journal of Scientific Research*, *60*(3).

Rifai, M. F., Jatnika, H., Sudirman, M. Y. D., Octaviasari, A. N., Rusdi, R. P. A., Faros, F., & Soniansyah, M. R. (2020). Implementation of Spiral Modeling in the Design and Construction of the Paperless Office System for ITCC ITPLN Activity Documents. *Jurnal E-Komtek (Elektro-Komputer-Teknik)*, *4*(2), 239–248.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. In *Journal of Clinical Epidemiology* (Vol. 98). https://doi.org/10.1016/j.jclinepi.2017.12.006

Sheldon, F. T., Jerath, K., Kwon, Y. J., & Baik, Y. W. (2002). Case study: Implementing a web based auction system using UML and component-based programming. *Proceedings - IEEE Computer Society's International Computer Software and Applications Conference*. https://doi.org/10.1109/cmpsac.2002.1044554

Wang, Y. A., Huang, Q., Yao, Z., & Zhang, Y. (2024). On a class of linear regression methods. *Journal of Complexity*, *82*. https://doi.org/10.1016/j.jco.2024.101826

Wibawanto, W., & Ds, S. S. M. (2017). *Desain dan pemrograman multimedia pembelajaran interaktif*. Cerdas Ulet Kreatif Publisher.