

Comparing Machine Learning and Human Judge in SATU Indonesia Awarding Processes

Onno W. Purbo

Institut Teknologi Tangerang Selatan, Komplek Komersial BSD Jalan Raya Serpong, Tangerang Selatan, Indonesia

ARTICLE INFO

Article history:

Received November 09, 2021

Revised November 25, 2021

Accepted December 23, 2021

Keywords:

Machine Learning;
Random Forest;
Orange Data Mining

ABSTRACT

For more than ten years, SATU Indonesia Awards, with PT. Astra International Tbk's support is given to inspiring young Indonesians. Every year, more than 10,000 nominations must be short-listed to 90 nominations within one week with five (5) assessment parameters. The research contributions are (1) creating a machine learning mechanism for the awarding process from ten years of the SATU Indonesia Awards nomination archive, (2) creating two (2) models of training data for the five (5) assessed parameters, namely motivation, obstacle, outcome, outreach, and sustainability, and (3) compare machine learning prediction with 2021 judge's assessment. TEMPO Data and Analysis Center (PDAT) extracts the corpus training data from ten years' SATU Indonesia Awards data in six months. The corpus training data contains nomination texts with Judges' scores on motivation, obstacle, outcome, outreach, and sustainability. Two (2) corpus training data and two models were generated with, namely, (1) the average Judges' parameter value per instance and (2) the Judges' smallest value and stored in two (2) corpus of 1220 instances each. The classification model was generated by Random Forest, which has the slightest error among the classification algorithms tested. The first model aims to predict the nomination assessment parameters. The second model is to detect the outlier in the incoming nominees for extraordinary nominees. The machine learning predictions were compared and found to be similar to the 2021 judge's assessment in the awarding processes at SATU Indonesia Awards. The average Judges' pre-final 2021 nominees' scores are compared to the Random Forest's predictions and found to be reasonably similar, with a small RMSE error around 1.1 to 1.6 for all assessment parameters. The smallest RMSE was obtained in the Sustainability parameter. The Obstacle parameter was found to have the largest RMSE.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



Corresponding Author:

Onno W. Purbo, Institut Teknologi Tangerang Selatan, Tangerang Selatan, Indonesia

Email: onno@itts.ac.id, onno@indo.net.id

1 INTRODUCTION

PT. Astra International Tbk. has fully supported SATU Indonesia Awards for more than ten years to inspire young Indonesians [1][2]. The judges for the SATU Indonesia Awards are some of the finest in Indonesia, including former ministers, such as Prof. Emil Salim, Prof. Nila Moeloek, Prof. Fasli Jalal, Ms. Tri Mumpuni, Ms. Dian Sastrowardoyo, Mr. Riza Deliansyah, Mr. Boy Kelana Soebroto, Ms. Diah Suran Febrianti, Mr. Arief Zulkifli, Mr. Billy Boen, and the author. The total yearly nomination increases so that in 2021, received more than 10,000 nominations.

The awarding processes begin with the public submission of nominations in written form via the Internet. The judges assess the nomination write-up. The assessment is focused on five parameters, namely, motive, obstacle, outcome, outreach, and sustainability, with discrete scores of 1, 3, 7, and 9. The most significant problem in the SATU Indonesia Awards is to select 90 pre-finalist nominations out of more than

10,000 submitted nominations in less than one week. To get the pre-finalist nomination marking consistently in a short time requires the help of machine learning for the classification processes.

Thus, machine learning mechanisms need to be developed. Machine Learning in award prediction is rare and difficult to perform [3][4]. In this work, the machine learning predictions were compared and found to be similar to the 2021 judge's evaluation of the awarding processes at the SATU Indonesia Awards. The corpus training data is extracted from ten years of historical data, nominations write-up, and the corresponding Judge's score. TEMPO Data & Analysis Center (PDAT) for six months, specifically Ms. Ai Mulyani, Ms. Rum Hayati, Ms. Ratih Virgorini (decd), and Mr. Arief Priandono, extract more than six thousand instances out of tens of thousands of archive data. The corpus is reprocessed to obtain two (2) types of corpus training data: average score and minimum score from the Judges, each consisting of 1220 instances of corpus training data.

1.1 Text Marking and Text Rating

The effort to automatically evaluate writings is exciting research. Many machine-learning marking methods have been developed, especially for precision education [5]–[10]. Some researchers use deep learning to regress numerical ratings from review texts or recommendation systems in supervised learning [11]–[15]. For lengthy text, more rigorous machine learning for text classification was developed [13][15]–[18]. Some researchers have gone so far as to make it more automated by using English Natural Language Processing (NLP) to assess short answers [11][13][19]. The last technique may be challenging to do as the Indonesian NLP is not as mature as English [20][21].

The SATU Indonesia Awards machine learning process avoids NLP and utilizes regressions from thousands of previous historical texts. The machine learning at the SATU Indonesia Awards has many distinctions, among others are,

- Indonesian NLP is not as mature as English NLP. Consequently, The Indonesian corpus text regression is used.
- The corpus of SATU Indonesia Awards nominations is a collection of long texts of thousands of nominees.
- In this work, two types of assessment classification models were developed and used, namely, (1) continuous with an average Judges' assessment value per instance, and (2) discrete with a minimum Judges' assessment value with possible values of 1, 3, 7, 9 per instance.

2 METHOD

2.1 ORANGE to enable non-data scientist

The most uncomplicated, intuitive machine learning techniques with a graphical interface are necessary for implementing machine learning in the SATU Indonesia Awards. ORANGE data mining is easy, very intuitive, and free to download from the Internet. Furthermore, ORANGE is attractive because it has many plugins for various complex data analysis, such as text analysis and image analysis, freely available on the Internet [22]–[24]. Thus, the ORANGE data mining application with a graphical interface enabled SATU Indonesia Awards personnel to conduct the necessary analysis.

Machine Learning used for SATU Indonesia Awards was simply a regression of the history text data with corresponding Judge's assessment value of each parameter to obtain a model to predict the necessary assessment classification.

2.2 Corpus Training Data Preparation

Machine learning requires us to generate corpus training data. The corpus contains extracted data from the nomination archive in the last ten years. It includes the text of nominees' activities with corresponding Judge's assessment on motive, obstacle, outcome, outreach, and sustainability parameters of each nominee.

TEMPO PDAT had to work for more than six (6) months to extract tens of thousands of historical data for the SATU Indonesia Awards to produce more than 6 thousand instances of corpus data. The resulting corpus training data is an excel file tagged for text analysis in the ORANGE data mining application. The data must be sorted and processed to separate motive, obstacle, outcome, outreach, and sustainable parameter values as separate corpus training data.

For each instance, the nominee is usually judged by several Judges at the same time. Thus, there are several choices of Judge's scores per instance. Although, there are some possible modeling techniques. In this work, two (2) approaches were used, namely,

- Corpus training data uses the average Judge's assessment value to predict the nominee's value. In this approach, the average Judge's value is used as an instance in training data. This technique is used to

obtain a model to predict the precise nomination parameter's value. The predicted value is between 1 to 9. Hence, linear regression can be employed [25]–[28].

- Corpus training data use the Judge's minimum assessment value to predict the outliers. In this approach, the Judge's minimum assessment value of each instance is used as corpus training data. There are only four (4) possible classification values, namely, 1, 3, 7, and 9. Thus, it is reasonably similar to logistic regression. The prediction results will not precisely predict nominee assessment value, instead of detecting unique outliers [29]–[33] from the average nominees. These exceptional outliers are usually those who perform differently from the average.

These two (2) approaches represent two (2) corpus training data stored in two (2) different files, with each having 1220 instances. The corpus training data is an excel file tagged for ORANGE data mining.

2.3 Machine Learning to Evaluate SATU Indonesia Awards Nomination

There are several small ORANGE graphical programs developed for SATU Indonesia Awards purposes, namely,

- To test the best algorithm for model generation.
- To generate a model from corpus training data.
- To predict parameter value by knowing nominee activities in the text.
- To compare Random Forest predicted value vs. 2021 Judge's value.

It is crucial to test the best algorithm for our data prior to any machine learning processes [34][35]. The ORANGE program to test the best algorithm is shown in Fig. 1. The test and score widget compares several algorithms for the same test data to assess the best algorithm for model generation. Text from uploaded Corpus processed by the Bag of Words widget to obtain a corpus with word counts for each data instance before being analyzed in the test & score widget. In the above program, seven (7) algorithms are tested and compared. The two (2) corpus training data are used to find the best algorithm.

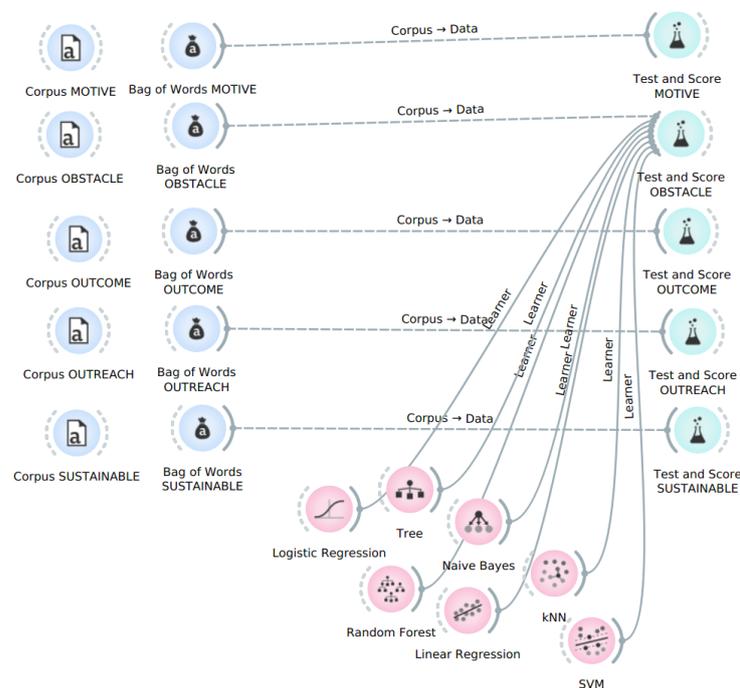


Fig. 1. Program to test the best algorithm for model generation.

ORANGE program to generate and save the resulting model is relatively simple, as shown in Fig. 2. The loaded corpus training data is processed through pre-processed text widget to remove stop words, then to the bag of words widget to obtain corpus with word counts for each data instance. Corpus with word counts process through Random Forest [36]–[39] to generate the required model and save for the next prediction or classification processes. Random Forest is used because it produces minor errors compared to other algorithms [37][38][40].

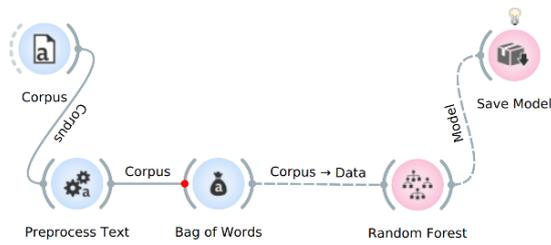


Fig. 2. Program to generate a model from corpus training data.

The incoming nominee corpus text data is loaded into the ORANGE program in Fig. 3 to predict the nominee's parameter value. The loaded Corpus is then processed through preprocess text and a bag of words widget, and the data is injected into the Prediction widget, which gets the model from the load model widget. The output of the scores can be viewed or saved in a spreadsheet file that can be combined with other new nomination data.

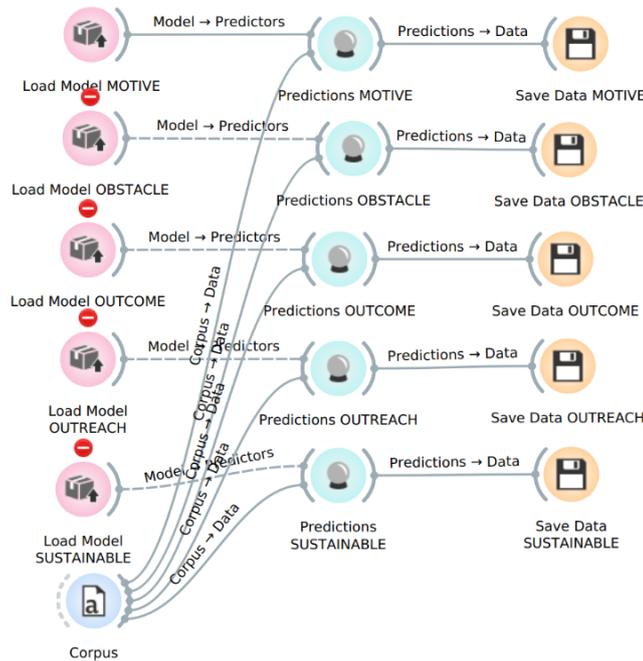


Fig. 3. Program to predict parameter value by knowing nominee activities in the text.

The next ORANGE program in Fig. 4 is specifically made to compare the Random Forest's predicted values with the 2021 average Judge's assessment value. In this particular example program, the evaluated parameter is the nominee's motive. The corpus training data on motive is processed into the pre-processed text and a bag of words widget before passing it into the test & scores widget as data used to generate a model and prediction using Random Forest. The 2021 Judge's assessment values are loaded and processed before being injected into the test & score widget as test data. The test and score widget will evaluate the predicted value from data and the test data and report the error or the deviation.

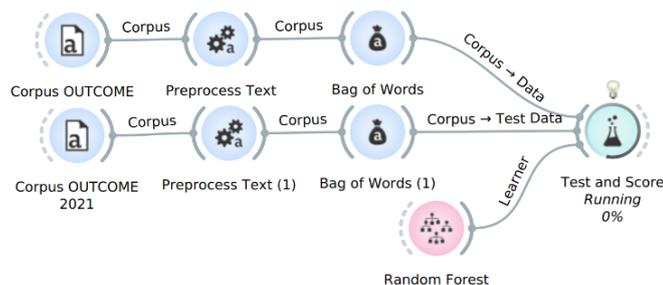


Fig. 4. Program to compare Random Forest predicted value vs. 2021 Judge's value.

3 RESULTS AND DISCUSSION

In this section, we will discuss the results of the implementation of machine learning at the SATU Indonesia Awards, especially the resulting accuracy.

3.1 Error measurement of the machine learning algorithm

The algorithm test on corpus training data with average Judge's assessment values is shown in [Table 1](#). It turns out that only three (3) algorithms can perform linear regression analysis, namely, kNN [36], Random Forest [36]–[38][40]–[42], and Linear Regression [25][28][43]. The results obtained are shown in the table below. Random Forest [36][38][41] provides the best MSE, RMSE, MAE, and R1 values compared to other algorithms.

Table 1. Model Error Estimate for corpus training data with average Judge's assessment value.

Algorithms	MSE	RMSE	MAE
kNN	2.583	1.607	1.245
Random Forest	2.262	1.504	1.151
Linear Regression	2.501	1.582	1.246

As shown in [Table 2](#), only six (6) algorithms can process the corpus training data with the minimal Judge's parameter value. The algorithms able to perform discrete value classification, namely, 1, 3, 7, and 9, are kNN [36], Tree [36][38], SVM [15][41], Random Forest [41][42], Naive Bayes [16][23][41], Logistic Regression [36][42]. The results obtained are shown in the table below. Random Forest provides the highest AUC, CA, F1, recall, and especially precision values compared to other algorithms. Therefore, Random Forest [36][38] will be used in both types of modeling. For different kinds of data, Random Forest may have disadvantages [40].

Table 2. Estimated Error of corpus training data with minimum Judges's parameter value.

Algorithm	AUC	CA	F1	Precision
kNN	0.535	0.424	0.332	0.349
Tree	0.553	0.402	0.402	0.403
SVM	0.555	0.337	0.326	0.367
Random Forest	0.568	0.450	0.419	0.436
Naive Bayes		0.191	0.155	0.140
Logistic Regression	0.573	0.451	0.280	0.203

3.2 Comparing machine learning vs. 2021 Judges'

The 2021 predicted parameter value based on corpus training data is compared to the 2021 Judge's assessment on 90 pre-finalist nominations to evaluate the model's accuracy. The program shown in [Fig. 4](#) is used. The model was generated using Random Forest [42] from the average Judge's assessment corpus training data. The 2021 Judge's assessment value and its predicted value are compared in the test & scores widget. The test & score widget will then report the error between the Random Forest predicted values and the Judges' values on the motive, obstacle, outcome, outreach, and sustainable parameters are shown in [Table 3](#).

Table 3. The error between Random Forest predicted values and 2021 Judge's values

Parameters	MSE	RMSE	MAE
Motive	2.200	1.483	1.296
Obstacle	2.534	1.592	1.402
Outcome	2.420	1.556	1.360
Outreach	2.363	1.537	1.205
Sustainability	1.300	1.140	0.950
Sustainability (Linear Regression)	4.435	2.106	1.689

It clearly shows that the 2021 Random Forest predicted values vs. the 2021 average Judge's assessment values are not much different, and the error is relatively small with RMSE around 1.1-1.6. The slightest error is obtained in the Sustainability assessment, while the highest error is found in the Obstacle assessment. Thus, we can use Random Forest [42] in machine learning to predict the SATU Indonesia Awards nomination value with a small RMSE error of around 1.1 to 1.6 for all assessment parameters.

We found Linear Regression algorithm produces a twice higher error as compared with Random Forest. Tests on the Sustainability value show RMSE Linear Regression 2,106, while RMSE Random Forest is much smaller at 1,140.

A slight difference between predicted and Judge's value in sustainability may be because, in reality, it is very challenging for most nominees to reach long-term sustainability in their activities. It is easily detected by all Judges, including in the corpus history data. Thus, the deviation of the Sustainability value by 2021 Judge's and machine learning predicted is slight.

On the other hand, the obstacles experienced by the nominees are more subjective. Thus, depending on the Judge's understanding of the obstacle faced by the nominee, it is more difficult to obtain an agreement from all Judges. It is shown by more volatile obstacle values between judges and a higher RMSE in machine learning prediction results.

4 CONCLUSION

Machine learning has been successfully predicting the parameter values of the SATU Indonesia Awards nominee with a small RMSE error of around 1.1 to 1.6 for all assessment parameters. Random Forest regression generates a model of the nominee's text and its corresponding Judges' parameter values. TEMPO Data and Analysis Center has to spend more than six long months extracting the corpus training data from the ten-year assessment archive. Two models generated from two types of training data were developed and used. The model based on the average Judge's assessment values accurately predicts the nominee's parameter values. At the same time, the model based on Judge's minimum assessment value is to classify the outliers in the nominees. It is found that the 2021 Random Forest predicted results vs. the 2021 average Judge's values are not much different, and the error is relatively small. The slightest error is obtained in the Sustainability assessment, while the highest error is found in the Obstacle assessment.

Acknowledgments

The author would like to thank PT. Astra International Tbk., TEMPO Data & Analysis Center, the Judges of the Satu Indonesia Awards, and the organizers for the effort and sharing the data of the SATU Indonesia Awards

REFERENCES

- [1] M. Fathony, A. Khaq, and E. Endri, "The effect of corporate social responsibility and financial performance on stock returns," *Int. J. Innov. Creat. Chang.*, vol. 13, no. 1, 2020. https://www.ijicc.net/images/vol_13/13120_Fathony_2020_E_R.pdf
- [2] N. D. Hidayati, "Pattern of corporate social responsibility programs: A case study," *Soc. Responsib. J.*, vol. 7, no. 1, 2011. <https://doi.org/10.1108/174711111111114576>
- [3] Y. Liu, R. Huang, and J. Yu, "Towards award prediction based on big data co-author network," 2019. <https://doi.org/10.1109/ICCCBDA.2019.8725612>
- [4] J. Wu *et al.*, "Product Design Award Prediction Modeling: Design Visual Aesthetic Quality Assessment via DCNNs," *IEEE Access*, vol. 8, 2020. <https://doi.org/10.1109/ACCESS.2020.3039715>
- [5] M. Kokoç, G. Akçapınar, and M. N. Hasnine, "Unfolding Students' Online Assignment Submission Behavioral Patterns Using Temporal Learning Analytics," *Educ. Technol. Soc.*, vol. 24, no. 1, 2021. <https://eric.ed.gov/?id=EJ1292999>
- [6] D. Tempelaar, B. Rienties, and Q. Nguyen, "The Contribution of Dispositional Learning Analytics to Precision Education," *Educ. Technol. Soc.*, vol. 24, no. 1, 2021. <http://oro.open.ac.uk/74065/8/74065VOR.pdf>
- [7] H. Luan and C. C. Tsai, "A Review of Using Machine Learning Approaches for Precision Education," *Educ. Technol. Soc.*, vol. 24, no. 1, pp. 250–266, 2021. <https://eric.ed.gov/?id=EJ1292868>
- [8] C. C. Y. Yang, I. Y. L. Chen, and H. Ogata, "Toward Precision Education: Educational Data Mining and Learning Analytics for Identifying Students' Learning Patterns with Ebook Systems," *Educ. Technol. Soc.*, vol. 24, no. 1, 2021. <https://eric.ed.gov/?id=EJ1292957>
- [9] X. Chen, H. Xie, D. Zou, and G.-J. Hwang, "Application and theory gaps during the rise of Artificial Intelligence in Education," *Comput. Educ. Artif. Intell.*, vol. 1, p. 100002, 2020. <https://doi.org/10.1016/j.caeai.2020.100002>
- [10] J. Y. Wu, C. C. Y. Yang, C. H. Liao, and M. W. Nian, "Analytics 2.0 for Precision Education: An Integrative Theoretical Framework of the Human and Machine Symbiotic Learning," *Educ. Technol. Soc.*, vol. 24, no. 1, 2021. <https://eric.ed.gov/?id=EJ1292867>

- [11] Y. Lan, Y. Hao, K. Xia, B. Qian, and C. Li, "Stacked Residual Recurrent Neural Networks with Cross-Layer Attention for Text Classification," *IEEE Access*, vol. 8, 2020. <https://doi.org/10.1109/ACCESS.2020.2987101>
- [12] J. Du, C. M. Vong, and C. L. Philip Chen, "Novel Efficient RNN and LSTM-Like Architectures: Recurrent and Gated Broad Learning Systems and Their Applications for Text Classification," *IEEE Trans. Cybern.*, vol. 51, no. 3, 2021. <https://doi.org/10.1109/TCYB.2020.2969705>
- [13] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," *IEEE Access*, vol. 8, 2020. <https://doi.org/10.1109/ACCESS.2020.2982538>
- [14] J. Zheng and L. Zheng, "A Hybrid Bidirectional Recurrent Convolutional Neural Network Attention-Based Model for Text Classification," *IEEE Access*, vol. 7, 2019. <https://doi.org/10.1109/ACCESS.2019.2932619>
- [15] Y. S. Mehanna and M. Bin Mahmuddin, "A Semantic Conceptualization Using Tagged Bag-of-Concepts for Sentiment Analysis," *IEEE Access*, vol. 9, 2021. <https://doi.org/10.1109/ACCESS.2021.3107237>
- [16] K. Liu and L. Chen, "Medical Social Media Text Classification Integrating Consumer Health Terminology," *IEEE Access*, vol. 7, 2019. <https://doi.org/10.1109/ACCESS.2019.2921938>
- [17] H. Tang, Y. Mi, F. Xue, and Y. Cao, "An Integration Model Based on Graph Convolutional Network for Text Classification," *IEEE Access*, vol. 8, 2020. <https://doi.org/10.1109/ACCESS.2020.3015770>
- [18] K. Fiok *et al.*, "Text Guide: Improving the Quality of Long Text Classification by a Text Selection Method Based on Feature Importance," *IEEE Access*, vol. 9, 2021. <https://doi.org/10.1109/ACCESS.2021.3099758>
- [19] C. N. Tulu, O. Ozkaya, and U. Orhan, "Automatic Short Answer Grading with SemSpace Sense Vectors and MaLSTM," *IEEE Access*, vol. 9, 2021. <https://doi.org/10.1109/ACCESS.2021.3054346>
- [20] O. J. Ying, M. M. A. Zabidi, N. Ramli, and U. U. Sheikh, "Sentiment analysis of informal malay tweets with deep learning," *IAES Int. J. Artif. Intell.*, vol. 9, no. 2, 2020. <https://doi.org/10.11591/ijai.v9.i2.pp212-220>
- [21] A. Amalia, O. S. Sitompul, E. B. Nababan, M. S. Lydia, and N. Rahmatunnisa, "Bahasa Indonesia text corpus generation using web corpora approaches," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 24, 2019. <http://www.jatit.org/volumes/Vol97No24/14Vol97No24.pdf>
- [22] B. B. Kadaru, M. Umamaheswararao, and C. Science, "An Overview of General Data Mining Tools," *Int. Res. J. Eng. Technol.*, vol. 4, no. 9, 2017. <https://www.irjet.net/archives/V4/i9/IRJET-V4I9165.pdf>
- [23] A. Naik and L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," in *Procedia Computer Science*, 2016, vol. 85. <https://doi.org/10.1016/j.procs.2016.05.251>
- [24] J. Demšar *et al.*, "Orange: Data mining toolbox in python," *J. Mach. Learn. Res.*, vol. 14, 2013. <https://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf>
- [25] D. N. Gujarati, *Linear Regression: A Mathematical Introduction*. 2020. <https://doi.org/10.4135/9781071802571>
- [26] A. Pant, "Introduction to Linear Regression and Polynomial Regression," *Towards Data Science*, 2019.
- [27] M. Pal and P. Bharati, "Introduction to Correlation and Linear Regression Analysis," in *Applications of Regression Techniques*, 2019. <https://doi.org/10.1007/978-981-13-9314-3>
- [28] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, Third edition, Sage publications, 2019.
- [29] M. Kwak and S. B. Kim, "Unsupervised Abnormal Sensor Signal Detection with Channelwise Reconstruction Errors," *IEEE Access*, vol. 9, 2021. <https://doi.org/10.1109/ACCESS.2021.3064563>
- [30] T. S. Buda, M. Khwaja, and A. Matic, "Outliers in Smartphone Sensor Data Reveal Outliers in Daily Happiness," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 1, 2021. <https://doi.org/10.1145/3448095>
- [31] H. O. Marques, R. J. G. B. Campello, J. Sander, and A. Zimek, "Internal Evaluation of Unsupervised Outlier Detection," *ACM Trans. Knowl. Discov. Data*, vol. 14, no. 4, 2020. <https://doi.org/10.1145/3394053>
- [32] N. Reunanen, T. Rätty, and T. Lintonen, "Automatic optimization of outlier detection ensembles using a limited number of outlier examples," *Int. J. Data Sci. Anal.*, vol. 10, no. 4, 2020. <https://doi.org/10.1007/s41060-020-00222-4>
- [33] H. Wang, M. J. Bah, and M. Hammad, "Progress in Outlier Detection Techniques: A Survey," *IEEE Access*, vol. 7, 2019. <https://doi.org/10.1109/ACCESS.2019.2932769>
- [34] M. Kang and E. Choi, *Machine Learning*. WORLD SCIENTIFIC, 2021. <https://doi.org/10.1142/12037>
- [35] M. Kubat, *An Introduction to Machine Learning*. Cham: Springer International Publishing, 2021. <https://doi.org/10.1007/978-3-030-81935-4>
- [36] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis," *IEEE Access*, vol. 8, 2020. <https://doi.org/10.1109/ACCESS.2020.3015966>
- [37] H. Zhang, Z. Fu, and K. I. Shu, "Recognizing Ping-Pong Motions Using Inertial Data Based on Machine Learning Classification Algorithms," *IEEE Access*, vol. 7, 2019. <https://doi.org/10.1109/ACCESS.2019.2953772>
- [38] Y. Nieto, V. Gacia-Diaz, C. Montenegro, C. C. Gonzalez, and R. Gonzalez Crespo, "Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions," *IEEE Access*, vol. 7, 2019. <https://doi.org/10.1109/ACCESS.2019.2919343>
- [39] I. Kaur and A. Kaur, "A Novel Four-Way Approach Designed with Ensemble Feature Selection for Code Smell Detection," *IEEE Access*, vol. 9, 2021. <https://doi.org/10.1109/ACCESS.2021.3049823>
- [40] C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," *IEEE Access*, vol. 5, 2017. <https://doi.org/10.1109/ACCESS.2017.2762418>
- [41] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *Int. J. Res. Mark.*, vol. 36, no. 1, 2019. <https://doi.org/10.1016/j.ijresmar.2018.09.009>

- [42] A. Mamgain, "Guidance to Data Mining in Python," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3, no. 6, 2018. <https://ijsrseit.com/CSEIT1836128>
- [43] D. Lilja, *Linear Regression Using R: An Introduction to Data Modeling*. 2016. <https://doi.org/10.24926/8668/1301>

BIOGRAPHY OF AUTHORS



Onno W. Purbo holds a Ph.D. in Electrical Engineering from the University of Waterloo, Canada. He is a copyleft, educator, and ICT evangelist. He has published 50+ books. He led the first Internet connection at the Institute of Technology in Bandung (ITB) in 1993-2000 and built the first Indonesian education network. He liberates WiFi frequencies and introduces neighborhood networks, wajanbolic (wokbolic) antenna, and OpenBTS cellular network. He led the first community telephony network over the Internet. He actively introduces e-learning, runs the most massive Indonesian free e-learning servers <https://lms.onnocenter.or.id/moodle/> with 54,000+ students. He is active on many social media. In 2021, he is active as a senior lecturer and vice-rector at Institut Teknologi Tangerang Selatan (ITTS). Internationally, he serves as a committee member at World Economic Forum on AI for Humanity. He received many awards, including the Johnathan Postel Service Award from the Internet Society. His activities were supported by many international organizations, such as IDRC Canada, Facebook, Google, InterLab AIT, ISIF Asia, ANU Australia, Timor Leste. Email: onno@indo.net.id