www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)
Budapest Institute

Talent Acquisition Implementation with People Analytic Approach

Atyoko Utomo[1], Dian Indiyati[2], Gadang Ramantoko[3]
[1,2,3]*Telkom University, Bandung, Indonesia*
atyoko.utomo@gmail.com

Abstract

The current human resource (HR) fulfillment conditions in this company are still quite low. This can be seen from the percentage of HR fulfillment of approximately 60% of the total HR needs. The strategy of fulfilling human resources through the recruitment and selection process must be done quickly and optimally. The problem that arises is related to the optimization of the talent acquisition process carried out, so that the results obtained are in accordance with the target and have quality that meets the required. In this study, data analysis was used using the random forest method. The method is used to develop a model that can predict the pass level of participants in recruitment and selection quickly and precisely in accordance with the profile of each participant, and can provide insight on the projected achievement of individual performance on each participant if passed at the company, to assist management in making decisions about the participants accepted in the recruitment and selection process. The data population used is data on recruitment and selection participants in 2018. To carry out the process of predicting the graduation rate of prospective employees, data for prospective employees who register for the recruitment and selection process will be used with a total of 17,294 people. The analytical tool in this study uses a people analytic approach. The conclusion of this study is that making people analytics on the process of talent acquisition can be done using the Random Forest Classification method. This method aims to determine the class of each predicted data. Modeling has been made to predict performance achievements, but the performance of the model is still not showing the level of significance in accordance with the standard level of confidence, which is still below 0.05.

Keywords
recruitment and selection; talent acquisition; people analytic; classification; random forest

I. Introduction

Based on research by Gavin and William (2018) regarding Talent Rising: People Analytics and Technology Driving Talent Acquisition Strategy, it is stated that many large or advanced organizations and companies have successfully used people analytics as a tool to deal with challenges related to HR problems, such as talent acquisition, talent pipeline planning, organizational development, engagement, and learning and talent development. However, the use of technology must be very careful, considering that human factors are the main factors that can guarantee long-term success. Based on Randhawa's research (2017), it is stated that talent acquisition is a strategic approach to identify, attract, and get the best talent to meet dynamic business needs effectively and efficiently. Described in Hasibuan (2012: 46), with the implementation of a good selection, employees who are accepted will be more qualified so that coaching, development, and employee management will be easier.

Business activities can grow and develop for a long period of time is the goal of each company. Competitiveness, innovation, creativity, and the quality of the products produced must be in accordance with the needs of consumers and can adapt to a dynamic environment (Rosmadi, 2018). Kuswati (2019) stated that in the world of work, employees are required to have high work effectiveness. Organizational effectiveness is usually interpreted as the success achieved by an organization in its efforts to achieve predetermined goals. According to Werdhiastutie et al (2020) the development of human resources should focus more on increasing productivity and efficiency. This can be realized because today's competition, especially among nations, is getting tougher and demands the quality of strong human resources as managers and implementers in an organization or institution.

The concept of the implementation of recruitment and selection or what is currently more commonly referred to as talent acquisition is not a new thing for the company. The process of implementing talent acquisition itself is carried out because there is an imbalance between supply and demand as well as the need for certain specifications from the Company.

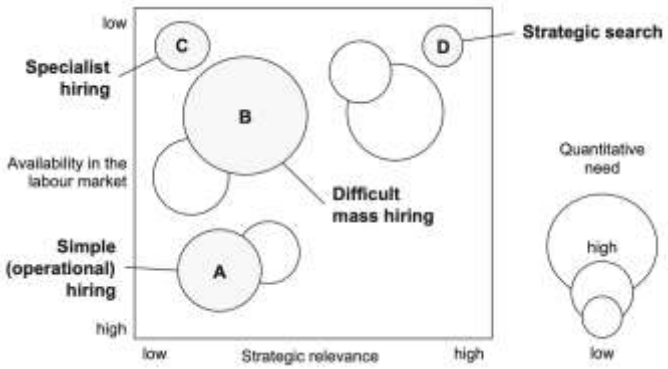
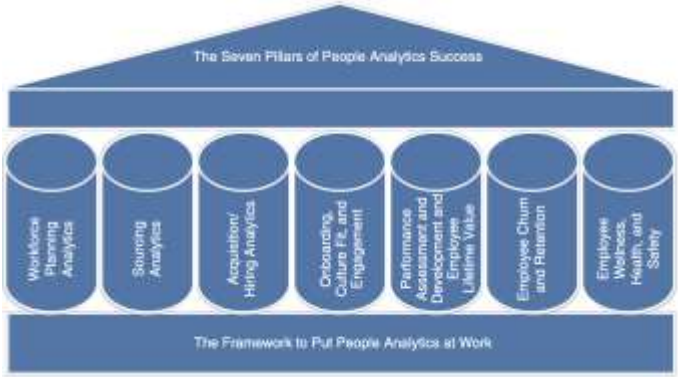


Figure 1. *Four Strategy of Talent Acquisition*
Source: Human Resources Strategies, Armin Trost, 2019

Based on Figure 1 above, it shows that the talent acquisition process is carried out differently depending on the quality and quantity expected to be received at the time of carrying out the process. In order to minimize bias at all stages in talent acquisition, the role of technology is needed so that the results obtained are faster and more accurate. With the rapid development of technology, the application of data analytics in the field of HR management or what is more commonly referred to as people analytics has been frequently carried out and applied both in the context of research and implementation in the real world.

Basically, the application of people analytics can be divided into 7 (seven) main pillars, as follows:



Figures 2. *The 7 Pillars of People Analytics Success*
Source: People Analytics in the Era of Big Data, Isson J. P., Harriott J. S., 2016

Based on Figure 2 above, it is stated that the talent acquisitions process is one of the pillars in implementing people analytics, so that the concept of implementing talent acquisitions using a data analytics approach can be carried out.

According to Isson and Harriott (2016: 177), talent acquisition is a practice to get new talents after the candidate search process is complete. Talent acquisition analytics is the practice of adding predictive analytics to the process of acquiring new talents. Analytics are used to determine which candidate fits the company's needs from all the candidates that are accommodated in the recruitment process.

With limited time, companies are required to find candidates who have the right abilities and at the right time. The challenge that arises is in selecting the method that has the best level of effectiveness and accuracy. The problem that arises is related to optimizing the implementation of the talent acquisition process so that the results obtained are in accordance with the target and have the quality according to the required specifications. Speed and accuracy are the most important indicators in the process of implementing talent acquisitions, so the application of data analytics through people analytics in the talent acquisition process is needed as a decision support system. Then the research questions are how is the best talent acquisition model to obtain a candidate employee profile according to the needs of the company and how is the best model that can predict performance achievements.

The problem statement of this research is how the talent acquisition model can predict the profile of prospective employees according to the company's needs and how the model can predict the performance achievements of the company. To answer the problem statement then the purpose of this study is to obtain a talent acquisition model that can obtain a candidate employee profile that suits Company needs by studying recruitment patterns using predictive techniques and obtaining models that can predict the performance achievements of prospective employees for the Company.

II. Research Methods

In this study, researchers used a data analytic approach in conducting research stages. The stages of the research consist of, among others, Data Collection, Data Cleansing, Data Processing, Data Modeling, and Output Recommendations. This stage can be explained in Figure 3.

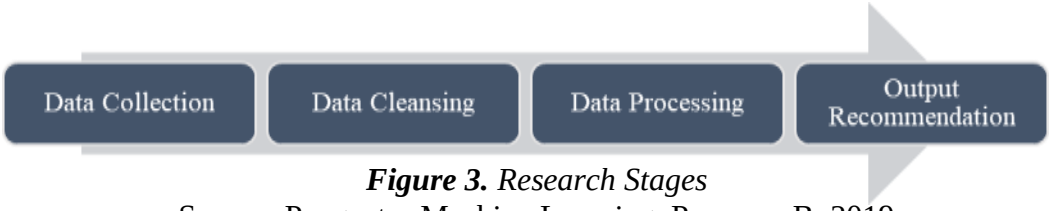


Figure 3. *Research Stages*
Source: Pengantar Machine Learning, Purnama B, 2019

1. Data Collection: at this stage, data collection was carried out on all required attributes, the data collection process was carried out through the recruitment and selection system. The data taken is data on the recruitment and selection process in 2018.
2. Data Cleansing: at this stage, a check is carried out on the data that will be used and there is no effort to ensure that there is no data that does not have one of the attributes in it. The data that will be used is data that has complete attributes as specified in the Operational Variable.
3. Data Processing: at this stage, data processing is carried out using the Random Forest method to obtain classification results in accordance with the variables and attributes described in the Variable Operations section.
4. Analysis of Recommendations: after obtaining the model and the results from data processing, the next step is to analyze the recommendations that arise from the results of the data processing. This is done to assist the decision-making process for management by using information obtained from the results of data processing.

This study used data analysis using the random forest method. Random Forest or random decision forest is a machine learning method introduced by Leo Breiman and Adele Cutler. According to Breiman (2001), "Random forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest". Random forest can be explained as a combined decision tree. The number of decision trees will affect the accuracy of the overall random forest. If the decision tree is used in large enough data sets, it tends to "remember" more than "learn". However, the decision tree is quite accurate if it is re-applied to the same or relatively the same data set.

The decision tree converts data into a tree (decision tree) and rules (decision rules). The decision tree learns through a set of if /then (if/else) or yes/no (yes/no) questions or other questions, which form a hierarchical tree. Every decision causes another decision or forms a prediction. The technique used by the decision tree is very similar to the way humans make decisions, so that the decision tree is more natural for humans than other models.

To measure the results of the decision tree, it can be done by measuring the impurity of the decision tree modeling results. To measure impurity in a decision tree, a metric called entropy and gini index can be used, as follows:

1. Entropy is the amount of information needed to describe a sample accurately. If sample is homogeneous (all elements are similar), then the entropy is 0. If all samples are not homogeneous and equally divided, then the entropy is 1. The maximum value of entropy is 1.
2. Gini Index is a measure of inequality (inequality) in the sample. The value is between 0 and 1. A Gini index with a value of 0 means that the sample is perfectly homogeneous or all elements are the same. Meanwhile, the Gini Index with a value of 1 means that all elements are not the same (inequality).

III. Results and Discussion

Based on table 1, it shows that the final results of the participants who were accepted did not match the predetermined target, where the targeted number of passes was 250 but after going through all the selection stages only 81 or 32% of the targets had been accepted. The number of participants who passed in each function is also still far from the target set in each function.

Table 1. Distribution of Recruitment and Selection Process Data

FUNCTION	TARGET	PARTICIPANTS	PASSED THE SELECTION					
			ADM.	POTENTION TEST (1)	POTENTION TEST (2)	USER INTERVIEW	MCU	FINAL INTERVIEW
ALO	30	750	127	35	6	5	4	2
CBD	10	1.699	79	22	8	6	6	2
ELC	45	1.089	227	85	29	27	25	13
ELT	50	1.630	260	89	37	31	29	16
FIN	10	2.479	57	18	11	8	8	3
HC	10	1.925	55	21	8	8	8	2
IFM	40	2.267	226	84	47	41	40	20
IT	25	2.195	138	54	19	16	15	11
MEC	15	1.530	82	34	17	16	16	10
STY	15	1.730	79	27	11	8	8	2
TOTAL	250	17.294	1.330	469	193	166	159	81

Source: Recruitment & Selection Process 2018

Model to predict the pass rate of participants was carried out using the Random Forest method, while model to predict performance was carried out using the Linear Regression method. The model process will be carried out separately based on objectives, phases, approaches and scenarios. The total model made in this study is according to table 2.

Table 2. List of Model Created

MODEL	PHASE	APPROACH 1	APPROACH 2
PASS SELECTION LEVEL PREDICTION MODEL	PHASE 1	Scenario 1	Scenario 1
		Scenario 2	Scenario 2
		Scenario 3	Scenario 3
	PHASE 2	Scenario 1	Scenario 1
		Scenario 2	Scenario 2
		Scenario 3	Scenario 3
	PHASE 3	Scenario 1	Scenario 1
		Scenario 2	Scenario 2
		Scenario 3	Scenario 3
PERFORMANCE PREDICTION	PHASE 1	Scenario 1	
	PHASE 2	Scenario 1	

Source: Data processed by researcher

Based on table 2, in this study, 18 (eighteen) types of models will be made for the prediction model and for the performance prediction model, 2 (two) types of models will be made. Each of these models will be compared the level of accuracy to get the best model to predict both graduation rates and performance.

There are a total of 18 (eighteen) types of models for prediction models of passing rates and 2 (two) types of models for prediction models of performance. Each model has its level of accuracy measured so that results can be compared to determine which model is the best for predicting graduation and predicting performance according to table 3.

Table 3. Recapitulation of the Accuracy Level of the Pass Selection Level Prediction Model

MODEL	PHASE 1		PHASE 2		PHASE 3	
	Approach 1	Approach 2	Approach 1	Approach 2	Approach 1	Approach 2
Scenario 1	0,9026	0,9285	0,9647	0,9639	0,9378	0,9793
Scenario 2	0,9269	0,946	0,9782	0,9797	0,9741	1
Scenario 3	0,9514	0,9631	0,9887	0,9925	1	1

Source: Calculated with Confusion Matrix on Python

Based on the results of measuring the importance level of features, level of accuracy, ROC Curve, and Area Under Curve (AUC), the best model in the short term can use a model made with approach 1 and scenario 3 is a model with the best level of accuracy and can provide much added value compared to other models. However, in the long term, if the data in each class (function) is sufficient, it is advisable to use a model with approach 2 and scenario 3. This is because the model with approach 2 is more in accordance with the selection process in the company seen from the test results of the level of importance features used in modeling.

The strengths and weaknesses of each model that have been made in this study are as follows:

1. Pass Selection Level Prediction Model with Approach 1
    a. Advantages :
        1) The model can be used in the long term and is not affected by the functions opened during the recruitment and selection process;
        2) The amount of data in each class is quite large, so that it can produce a more stable output;
        3) The computation process is faster because the class division is not too much.
    b. Weakness:
        1) The information generated from the model is limited to the pass rate only;
        2) The model does not pay attention to the specifications that have been determined because it is carried out in general.
2. Pass Selection Level Prediction Model with Approach 2
    a. Advantages :
        1) The model can provide recommendations up to the probability in each function;
        2) Information and insights that can be used by the management or recruitment and selection committee are numerous and varied;
        3) The model takes into account the specifications that have been set by the company, for example, is the participant's education style, because each function has different specifications.
    b. Weakness:
        1) The computation process takes longer than approach 1;
        2) The model will be disturbed if there are additional functions outside of the functions that have been arranged in the system;
        3) The amount of data in each class is small, so it is possible that the model will change if there is new data and is heterogeneous from the current data.

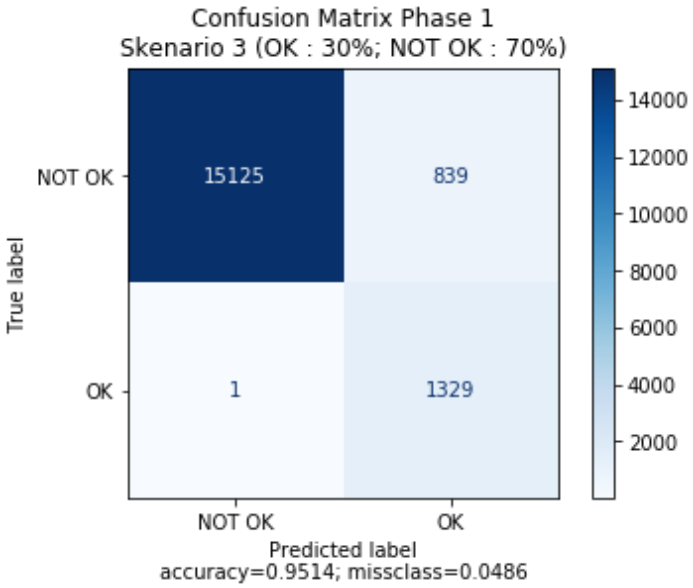


Figure 4. *Confusion Matrix Pass Selection Level Model Phase 1*
Source: Generated with Scikit-Learn on Python

Based on the calculation of confusion matrix for the model made in phase 1 or the administrative selection stage which can be seen in Figure 4. shows that in the model there are 1 data that fall into the False Positive category and 839 data that fall into the False Negative category. Data that falls into the False Negative category means that there are 839 candidates whose profiles match the profiles of candidates who passed the administrative selection stage. It can be interpreted that there is a potential loss of superior candidates who are not included in the next selection process. Based on the measurement of features importance, information is obtained that the most important feature at this stage is GPA.

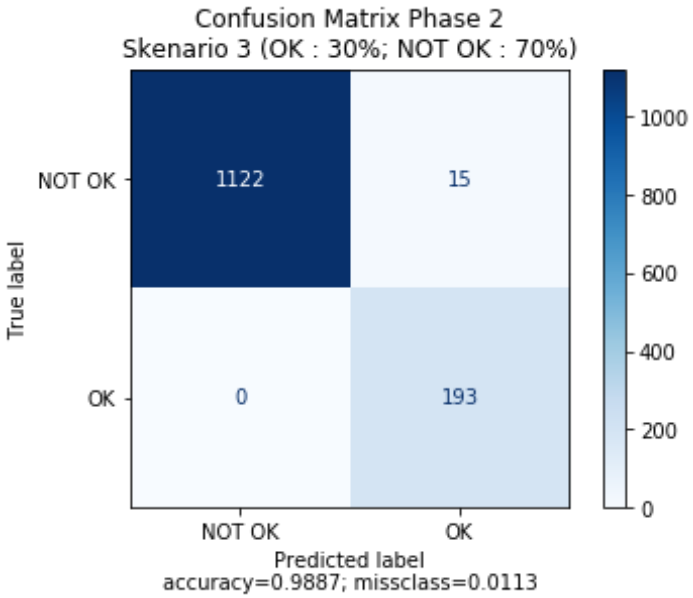


Figure 5. *Confusion Matrix Pass Selection Level Model Phase 2*
Source: Generated with Scikit-Learn on Python

Based on the calculation of confusion matrix for the model created in phase 2 or the selection stage of the potential test which can be seen in Figure 5., it shows that in the model there are 15 data that fall into the False Negative category, but none are included in the False Positive category means that the level of precision of this model is 1. Data that falls into the False Negative category means that there are 15 candidates whose profiles match the profiles of candidates who passed the administrative selection stage. It can be interpreted that there is a potential loss of superior candidates who are not included in the next selection process. Based on the measurement of features importance, information is obtained that the most important features at this stage are related to salaries and personalities of the participants.

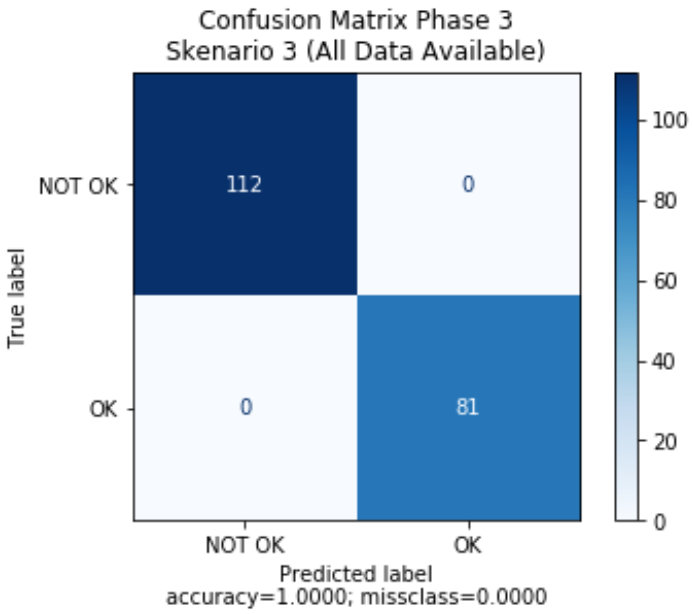


Figure 6. *Confusion Matrix Pass Selection Level Model Phase 3*
Source: Generated with Scikit-Learn on Python

Based on the calculation of confusion matrix for models made in phase 3 or the final selection stage which can be seen in Figure 3., it shows that in the model there is no data that falls into the False Negative or False Positive categories so that the level of precision and sensitivity of this model Very good. Based on the measurement of features importance, information is obtained that the most important features at this stage are related to the results of medical check-ups and the educational background of the participants.

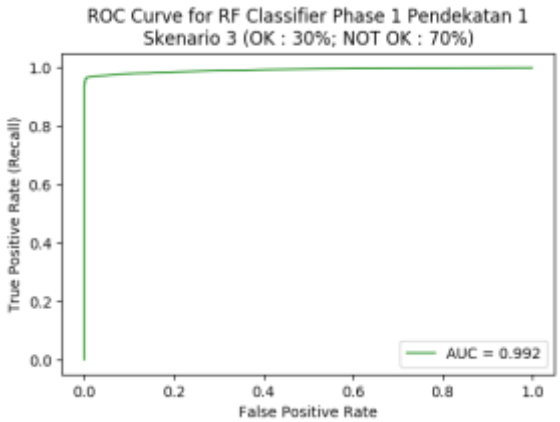


Figure 7. *ROC Curve Pass Selection Level Model Phase 1*
Source: Generated with Scikit-Learn on Python

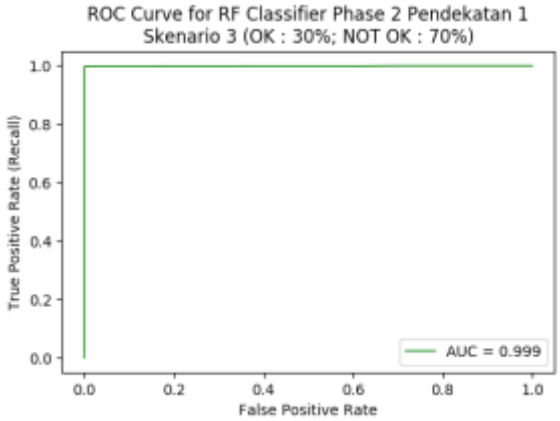


Figure 8. *ROC Curve Pass Selection Level Model Phase 2*
Source: Generated with Scikit-Learn on Python

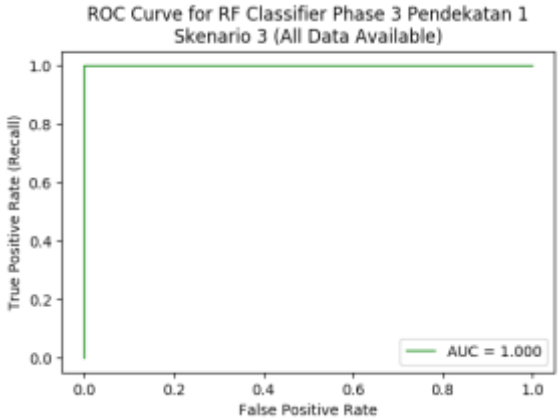


Figure 9. *ROC Curve Pass Selection Level Model Phase 3*
Source: Generated with Scikit-Learn on Python

A ROC (Receiver Operator Characteristic Curve) can help in deciding the best threshold value. It is generated by plotting the True Positive Rate (y-axis) against the False Positive Rate (x-axis) as you vary the threshold for assigning observations to a given class.ROC curve will always end at (1,1). The threshold at this point will be 0. This means

that we will always classify these observations falling into class 1(Specificity will be 0. False-positive rate is 1). One should select the best threshold for the trade-off you want to make. According to the criticality of the business, we need to compare the cost of failing to detect positives vs cost of raising false alarms.

Table 4. Recapitulation of the Accuracy Level of the Performance Prediction Model

METHOD	ACCURACY SCORE	
	PHASE 1	PHASE 2
Mean Absolute Error	5,47198	4,63553
Root Mean Squared Error	6,28185	5,52126

Source: Generated with Python

Based on Table 4 regarding the level of accuracy of each model made for performance prediction, the model that has the best level of accuracy is the model in phase 2 where the model is made using all available features from the beginning to the end of the recruitment and selection process. However, based on the results of the significance test, it is found that the independent variable that is owned still does not significantly affect the dependent variable, namely the Individual Achievement Value score.

To be able to answer the research question, what is the model that can predict performance achievements for the company, that the model that has been made in this study can already predict the performance achievement only with an error rate above the confidence level. This happens because the individual Achievement Value score data used is not normally distributed or tends to lean to the right. This condition occurs because the individual performance appraisal process in the company still has too many subjective factors, so that the final results of the assessment are difficult to justify. For this reason, it is necessary to improve the prediction model by multiplying the data used in the prediction model making process. In addition, because the training data used in the modeling process is still too little, so it cannot produce models and predictions with the expected level of confidence.

IV. Conclusion

From a series of data collection processes, data processing, to discussion of research results, the following conclusions can be drawn:

1. The selection process is a series of steps carried out to determine the suitability of the qualifications of the participants with the predetermined specifications. The success of the selection process will only be seen after the participants who pass the selection have served at the Company, so that the decision making in the selection process cannot be wrong;
2. In its development, the selection process can be assisted by the application of data analytics through people analytics in the talent acquisition process, which is indispensable as a decision support system. The use of data analytics is used to improve the speed and accuracy of decision making;
3. Making people analytics on the talent acquisition process can be done using the Random Forest Classification method. This method aims to determine the class of each predicted data. In this study, the use of Random Forest Classification was carried out to create a prediction model for passing rates. For the short term, the best passing rate prediction model based on the data available in this study is a model using approach 1 and scenario 3, where the accuracy level produced in Phase 1 is 0.9514, Phase 2 is

0.9887, and Phase 3 is 1. Whereas in the long run, the prediction model for passing rates using approach 2 and scenario 3 is very suitable for the selection process as seen from the results of the feature importance level test used in the model.

4. To determine the level of success of the selection process, it is necessary to look at the performance achievements of each participant resulting from the selection process. Therefore it is necessary to make a model that can predict performance achievements, so that the decision-making process of management or recruitment and selection committees is better. In this study, a model has been made to predict performance achievements, but the performance of the model has not shown a significance level in accordance with the standard level of confidence, which is still below 0.05. When viewed from the level of accuracy of the model that can predict performance achievement, the best model is the model in phase 2 with a mean squared error of 4.6355.

Suggestions

Based on the results of this study, there are several suggestions that can be given both for further research and for companies, as follows:

1. For further research, in the process of making a prediction model, it is hoped that data can be used over a period of more than 1 (one) period of the recruitment and selection process. This will be useful for comparing the accuracy of the model in each period of the recruitment and selection process, because basically the recruitment and selection process is an independent process from each implementation period. If the resulting model has a good level of accuracy even though it is used in different periods, it can be said that the model can be used in the long term and continuously.
2. For companies, historical data is one of the important factors in the modeling process so that data management is expected to be one of the factors that need to be of concern to the company. The process of individual performance appraisal needs to be re-validated using other factors, for example applying the 360 degree concept or sociometry so as to reduce the subjectivity factor.

References

Bogen M., Rieke A. (2018). An Examination of Hiring Algorithms, Equity, and Bias. Upturn.

Brownlee J., (2020). How to Calculate Feature Importance With Python. [online]. https://machinelearningmastery.com/calculate-feature-importance-with-python/ [27 Desember 2020]

Deters J., (2018). Global Leadership Talent Selection As An Integrated Acquisition Process. Advances in Global Leadership, 73-96.

Foley M., Williamson S. (2018). Does anonymising job applications reduce gender bias?: Understanding managers' perspectives. Gender in Management: An International Journal, 33 (8), 623-635.

Geron A., (2019). Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow. O'Reilly.

Goldstein H., Pulakos E., Passmore J., Semedo C. (2017). Handbook of the Psychology of Recruitment, Selection and Employee Retention. Wiley Blackwell.

Isson J. P., Harriott J. S. (2016). People Analytics in the Era of Big Data. Wiley.

Joy O. E., Ugochukwu U. I., Ndubuisi U. (2015). The Impact of Recruitment and Selection Criteria on Organizational Performance. Global Journal of Human Resource Management, 3 (2), 22-33.

Kuswati, Y. (2019). Motivation Role in Improving Work Effectiveness. Budapest International Research and Critics Institute-Journal (BIRCI-Journal). P. 281-288.

Lepisto L., Ihantola E. (2018). Understanding the recruitment and selection processes of management accountants: An explorative study. Qualitative Research in Accounting & Management, https://doi.org/10.1108/QRAM-11-2016-0080

Purnama B. (2019). Pengantar Machine Learning. Informatika.

Rosmadi, M.L.N. (2018). Factors in Developing Creative Industry. Budapest International Research and Critics Institute-Journal (BIRCI-Journal). P. 64-69.

Randhawa N. (2017). The Changing Dynamics of Talent Acquisition. Imperial Journal of Interdisciplinary Research, 3 (6), 1137-1142.

Sahay P. (2015). Lean Six Sigma tools in the hiring process. Strategic HR Review, 14, 22-29.

Sahay P. (2014). Design thinking in talent acquisition: a practitioner's perspective. Strategic HR Review, 14, 170-180.

Sequeira A., Attupuram P. J., Gopalakrishnan S. (2015). Talent Acquisition Process in a Multinational Company: A Case Study. DOI: 10.2139/ssrn.2708086

Trost A. (2019). Human Resoruces Strategies: Balancing Stability and Agility in Times of Digitization. Springer.

Walford-Wright G., & Scott-Jackson W. (2018). Talent Rising; people analytics and technology driving talent acquisition strategy. Strategic HR Review. https://doi.org/10.1108/SHR-08-2018-0071.

Werdhiastutie, A., et al. (2020). Achievement Motivation as Antecedents of Quality Improvement of Organizational Human Resources. International Research and Critics Institute-Journal (BIRCI-Journal). P. 747-752.

Yen C. (2017). The tradeoff between fit perceptions across recruitment stages by new job seekers. International Journal of Contemporary Hospitality Management, https://doi.org/10.1108/IJCHM-05-2016-0295