http://ejournal.seaninstitute.or.id/index.php/Ekonomi
Jurnal Ekonomi, Volume 12, No 02, 2023
ISSN: 2301-6280 (print) ISSN: 2721-9879 (online)

USING FEATURE ENGINEERING IN LOGISTIC REGRESSION AND RANDOM FOREST METHODS TO IMPROVE EMPLOYEE ATTRITION PREDICTION IN KIMIA FARMA

Lani Asep Sutisna
Faculty of Information Technology, Budi Luhur University, Jakarta

ARTICLEINFO

Keywords :
employee attrition, machine learning, logistic regression, random forest, feature engineering, selectKBest

E- mail :
Laniasep.sutisna@gmail.com

ABSTRACT

This study aims to analyze the effect of Feature Engineering on the Logistic Regression and Random Forest methods on the prediction of employee attrition at PT Kimia Farma Tbk. In addition to knowing the most effective method in increasing employee attrition prediction at PT Kimia Farma Tbk. The results of this study indicate that feature engineering significantly affects performance in predicting employee attrition at PT Kimia Farma Tbk. using Logistic Regression and Random Forest models. It can be seen that the application of feature engineering can affect the accuracy, precision, recall, and F-Score of the two methods. The Recursive Feature Elimination (RFE) method with the Logistic Regression model has an accuracy of 0.866, a precision of 0.5, a recall of 0.159, and an F-Score of 0.259. Meanwhile, the RFE with the Random Forest model has an accuracy of 0.886, a precision of 0.916, a recall of 0.25, and an F-Score of 0.392. The SelectKBest method with the Logistic Regression model has an accuracy of 0.88, a precision of 0.9, a recall of 0.204, and an F-Score of 0.333. Meanwhile, SelectKBest with the Random Forest model has an accuracy of 0.87, a precision of 0.818, a recall of 0.204, and an F-Score of 0.327. According to the results of the performance comparison, the RFE (Recursive Feature Elimination) method with the Random Forest model can be said to be the best method in terms of accuracy and precision. Although the recall of this method is slightly lower, the performance of this method still meets the criteria as a good method. Therefore, the Recursive Feature Elimination method with the Random Forest model was chosen as the best method for this case.

1. INTRODUCTION

Employee attrition is the loss of employees through natural processes, such as retirement, resignation, position deletion, personal health, and termination of employment (layoff). This is a challenge for management to be able to manage its employees in order to predict these problems. HR (Human Resources) plays an important role in all strategic decisions for every organisation. Therefore, it must be able to analyze why employees with outstanding performance resign prematurely. Attrition is an inevitable part of any business. There will come a time when an employee wants to leave the company either for personal or professional reasons.

Losing employees through attrition and staff turnover is costly. Between the costs associated with separation, loss of productivity, hiring, interviewing, training, and onboarding, the loss of one employee is estimated to cost the business from that individual's annual salary. Replacement costs can be lower for entry-level roles while replacement costs are significantly higher for professional, technical, and supervisory positions.

From the data of PT. Kimia Farma Tbk. in 2021 showed that around 245 people out of a total of 1500 employees experienced attriton as evidenced by the last activity in the employee management application. The employee has not logged in from 2020 to 2021. This means that there are around 16.2% of employees who experience attrition from all divisions at PT. Kimia Farma Tbk. This means that every year employees of PT. Kimia Farma Tbk. always experiences high attrition and there are several factors that affect the occurrence of attrition.

The main problems that occur in the organization of PT. Kimia Farma Tbk. is a company management that has not been able to predict the needs of employees in the organization due to employee attrition. So that the high level of entry and exit of employee which results in expensive recruitment and training costs. In addition, the lack of career opportunities causes employees to feel there are no long-term prospects in the company. Furthermore, the problem of lack of training and development of employees that causes employees to feel unprepared to handle their jobs. Therefore, there is a need for an analysis model that can provide predictions for employees who will resign or leave the company.

In a previous Indonesian research journal entitled "Classification of Factors Affecting Employee Reduction in "XYZ" Company" written by Ainun Umami from Institut Teknologi Sepuluh Nopember (ITS). The study raised the main problem: the factors causing termination of employment (PHK) in Indonesia. The study analyzed the classification of factors that cause employee attrition in company "XYZ" using the(Umami, 2018) Naïve Bayes, SVM, Logistic Regression, MLP, Gradient Boosting, KNN, Random Forest, Decision Tree methods. After research and analysis, the best methods for classification are Naïve Bayes, SVM, Logistic Regression, and MLP. The dataset used in this study is a public dataset.

In a previous study, an international journal entitled "Machine Learning Approach for Employee Attrition Analysis" was written by Dr. R. S. Kamath1, Dr. S . S. Jamsandekar, Dr. P. G. Naik in 2019. The journal explains how the comparison between several algorithms to predict employee attrition. The algorithms used in the study include Decision Tree, Random Forest, Support Vector Machine, Linear Regression. Of the four algorithms used in the study, Random Forest produced better accuracy than the others. The accuracy generated by Random Forest is 0.9773 higher than the other three algorithms.(Kamath et al., 2019) The dataset used in the journal uses a public dataset.

In this study, a model is proposed to predict employee attrition which can later be used by HR Staff (Human Resource) in analyzing trends regarding employee turnover in the company's organization. Currently, HR staff only analyzes based on the needs of employees who have resigned and when there is no replacement, there is no prediction when employees will resign. The use of this employee attrition prediction model uses machine learning techniques. Where the amount of existing data will be studied to find knowledge from the data.

Machine learning is carried out with a series of methods and algorithms that are staged, the process of data mining can also be assessed for accuracy. In this proposal, two algorithms are used to be compared to obtain high accuracy and combined with the use of feature engineering to be able to improve the accuracy of the recommendations provided by this system. The models used in this study are Logistic Regression Model and Random Forest both of these algorithms will be combined with Feature Engineering methods. The dataset used is sourced from PT. Kimia Farma Tbk Human Capital division as data training and data testing. This dataset is commonly used by researchers to conduct research related to employee attriton.

2. LITERATURE REVIEW

In a case study of *employee attrition* at PT. Kimia Farma, a predictive model generated from the data mining process must be evaluated to determine the level of accuracy. One method used in this evaluation is the *confusion matrix*, which shows the number of true and incorrect predictions of a model. In addition, other measures used in this evaluation are accuracy, sensitivity, specificity, and AUC *(Area Under the Curve).* By analyzing these measures, it can be determined the level of accuracy of the *employee attrition* prediction model that is built, so that improvements or changes can be made to the model if needed.

1. *Confusion Matrix*
   Confusion matrix can be used to determine the predicted probability of employees leaving the company or employees who will remain employed. This is done using *a cut value* of 0.5 or $\pi(Y) = 0.5$. That is, if the resulting probability value of $\pi(Y) < 0.5$ then the employee is categorized as not leaving the company and if the resulting probability value of $\pi(Y)$ is ≥ 0.5 then the employee is predicted to leave the company. *Confusion matrix* will produce accuracy, *precision* and *recall*. Employees who leave the company are referred to as "*attrition*" and employees who remain employed are referred to as "*non-attrition"*.
2. Accuracy
   Accuracy is an evaluation method used to measure how well a model classifies data. This method measures how much data is properly classified compared to the amount of data provided. In this study, *accuracy* was used to evaluate the model used in the study. It is usually used to evaluate a model on test data to find out how well the model classifies data that has not been seen before. However, *accuracy* is not always a good measure for evaluating a model, especially if the data has an unbalanced distribution between classes, or if the model is more important in detecting one class than another.
3. ROC & AUC

To evaluate the performance of *employee attrition* prediction models, one method that can be used is ROC (*Receiver Operating Characteristic)* and AUC *(Area Under the Curve).* ROC is used to evaluate model performance by comparing the False Positive Rate (FPR) with *the True Positive Rate* (TPR) or *Sensitivity*. AUC is the area under the ROC curve that measures the model's ability to distinguish between positive and negative classes. However, in the case of *employee attrition* at PT. Kimia Farma, AUC values can differ depending on the model used and the results of the tests performed. In the context of *employee attrition*, AUC is used to measure how well the model distinguishes between employees who will *resign* or not. A perfect AUC score is 1, which means the model can distinguish between employees who are about to *resign* and not perfectly. An AUC value lower than 0.5 indicates that the model is less effective in distinguishing between employees who will *resign* and who will not.

4. *Grid Search*
   Grid search is a method used to find the best parameters of a model by evaluating the model with various combinations of predefined parameters. In the process of *tuning* the *employee attrition* model, *grid search* can be used to find the best parameters of the model used, such as the number of trees in the *Random Forest* or regularization in *Logistic Regression* . This method will test each combination of predefined parameters, and determine the combination of parameters that gives the best results based on the specified metric such as accuracy, precision, recall or F1-score. This can help in improving the performance of the model and avoiding *overfitting***.**

3. METHOD

In this study, the methodology used is the CRISP-DM (Cross Industry Standard Process for Data Mining) method, which consists of four main stages: Business Understanding, Data Understanding, Data Preparation, and Modeling.

The data understanding that researchers currently have comes from 2021 with a total of 1500 employees. This data is collected from different sources such as employee management apps and employee attendance records. The process of extracting knowledge from this data was carried out to identify the level of attrition in 2021. In addition, below are some contexts of employee attrition research at PT. Kimia Farma to support the process of understanding data, as follows.

The sampling technique in this study uses Purposive Sampling which is one of the sampling techniques that is often used in research. The data used in this study was sourced from PT. PT. Kimia Farma Tbk. The file has csv extension format and contains text with various column attributes provided earlier. This file has 35 column attributes with 1500 rows of data that will later be used for this study.

Data collection is carried out to obtain samples of data that will be used for experiments in the testing process. The data collection method used is secondary data collection. Using surveys or questionnaires, interviews, observations and company documents.

4. RESULT AND DISCUSSION

Model Comparison & Selection

In the process of employee attrition research, a model is needed to predict the possibility of *employee attrition* well. One way to determine the best model is to compare the results of several models used.

Table 1 Model Comparison Results

Type	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	0.852	0.69	0.15	0.25
Random Forest	0.88	0.7	0.31	0.43

In table 1 above, this study uses two models: Random *Forest* and *Logistic Regression*. From the results obtained, it was obtained that the *Random Forest* model had an accuracy of 0.88, precision of 0.7, recall of 0.31, and F1-Score of 0.43. While the *Logistic Regression* model has an accuracy of 0.852, precision of 0.69, *recall* of 0.15 , and F1-Score of 0.25.

From these results, it can be seen that the *Random Forest* model is better than the *Logistic Regression* model. This can be seen from the higher accuracy, precision, *recall*, and F1-Score values in the *Random Forest* model. However, keep in mind that these results only apply to the data used in this study and do not guarantee the same results for other data.

Evaluation

The process of *testing (evaluation)* on *employee attrition* data is carried out to evaluate the performance of the model that has been created. In this process, the data used is test data that is not used in the model creation process. There are several methods that can be used in this testing process, including:

1. *Grid Search*: is a method used to find the best parameter value from the model created. *Grid Search* will experiment by combining multiple values from each parameter and evaluating the model's performance with each combination of parameters.
2. ROC AUC *Curve*: is a method used to evaluate the performance of binary classification models. ROC AUC shows how well the model can distinguish between two existing classes.
3. *Confusion Matrix*: is a method used to determine the level of accuracy of the model by calculating the number of correct and incorrect predictions from the model made.

All of the above methods are used to evaluate the performance of the model created and determine the *Random Forest* model to be used in the *employee attrition* research process. Because only selected models will be evaluated. The next process is to evaluate using some of the methods above. The following is a detailed presentation table related to the results of the evaluation of the model.

Evaluation of the Random Forest Model

The purpose of evaluating the *Random Forest* model is to determine the performance of the model in predicting new data. Model evaluation is carried out using several methods such as *grid search, accuracy*, ROC-AUC, and *confusion matrix*. *Grid search* is used to find the best parameters of the *Random Forest* model. Then, the results of the best parameters are used to predict the test data.

The results of these predictions will be compared with actual data to find out how well the *Random Forest* model predicts new data. By knowing the performance of the model, we can evaluate the model and make improvements if needed.

It can also be used to assess model quality and determine which model is best for use in a particular application. The following is a table of the evaluation results of the *Random Forest* model.

Table 2 *Random Forest Model* Evaluation Results

Evaluation Method	Result
Grid Search	*Best parameters found: {'max_depth': 20*, 'n_estimators': 200} *Best accuracy found*: 0.8574999999999999
Accuracy	0.883
ROC AUC	0.611
Confusion Matrix	[255 1] [34 10]

Table 2 above shows the results of the evaluation of the *random forest* model using the *Grid Search* method. In this method, the best parameter search for the *random forest* model is carried out by evaluating a combination of '*max_depth*' and '*n_estimators'* parameters. From the results of *Grid Search*, the best parameters are '*max_depth*' = 20 and '*n_estimators'* = 200.

Furthermore, the results of the model evaluation in Table 4.30 above are shown using the best parameters. The evaluation results show that the model has an accuracy of 0.883 and an ROC AUC of 0.611. ROC *AUC (Receiver Operating Characteristic - Area Under Curve)* is a method for evaluating binary classification models by measuring how well the model differentiates between two classes.

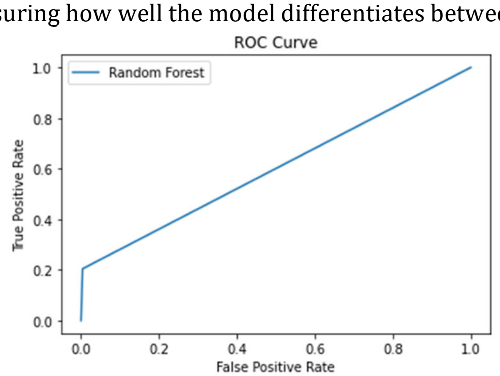

Figure 1 ROC *Curve Random Forest Results*

The ROC curve is a graph that shows the relationship between *the false* positive rate (FPR) and *true* *positive rate* (TPR) of a classification model. FPR is the sum of false positives divided by the total number of *negatives*, and TPR is the number of true positives divided by the total number of *positives*.

The results of Figure 4.2 ROC Curve above show show that the initial FPR is 0 and the initial TPR is also 0. Then, when the *threshold* is lowered to 1, FPR increases by 0.003906 and TPR increases by 0.204545. At *threshold* 0, FPR becomes 1 and TPR becomes 1. From these results, it can be interpreted that the classification model used in employee attrition performs fairly well in predicting the classification of *employees* who will experience *attrition*. However, please note that these results only apply to the test data used and do not necessarily apply to other data.

Finally, the results of the *Confusion Matrix* are shown in table 4.30 above, which is a matrix that shows the number of correct and incorrect predictions from the model. In table 4.20, the model successfully predicted 255 true cases and 1 false case from the class 'attrition = 0', as well as 10 cases true and 34 cases false from the class 'attrition = 1'.

From the results of the evaluation above, it can be concluded that the *random forest* model with the best parameters obtained from *grid search* is quite good at evaluating *employee attribute* data. But can still be improvements and further model development.

Deployment (Prototype Design)

Deployment or design of application prototypes is the final stage in this research process, where the model that has been developed and tested is implemented in the form of an application that is ready to be used by users. In the case of employee *attrition*, the application prototype is created using the *Random Forest* model and selected features such as "Age", "Daily Rate", "Distance From Home", "Job Role", "Monthly Income", "Monthly Rate", "Overtime", "Total Working Year" which will be used to predict the level of employee propensity to *resign*. In this process, the training and test data are split using Stratified Sampling so that the distribution of classes in the training and test data is the same.

Framework Flask

The Flask framework is one of the popular *frameworks* in web application development. In our research on employee attrition, we used the Flask framework to create an application that can predict the likelihood of employee attrition. *This framework* allows us to quickly implement our *Machine Learning* algorithms and present the results in the form of intuitive visualizations through the application dashboard. Flask also allows us to create *endpoints* that can receive input from users and output prediction results in *real-time*. The use of the Flask framework makes this application development process more efficient and easy to use by end users. Table 4.32 below shows the default structure of the Flask folder.

Table 3 Flask Folder Structure

Folder	Description
.app	Contains files related to application logic, such as *views*, *controllers*, and *models*
static	Contains static files such as CSS, JavaScript, and images
Templates	Contains HTML template files used by the application
tests	Contains test files to test the application
Venv	Contains virtual *environment* files used by the application
.env	Contains the *environment* configuration file used by the application
.flaskenv	Contains the *flask environment* configuration file used by the application
requirements.txt	Contains a list of *libraries* required by the application
app.py	Files used to run the application

Menu Dashboard

The following in figure 2 below shows a dashboard menu consisting of several analyzes related to the level of *employee attrition* or leaving the company. The analysis displayed in this dashboard includes employee age analysis, the relationship between business travel and employee attrition rate, the relationship between department and employee attrition level, the relationship between distance from home and employee attrition level, the relationship between education and employee attrition level, and the relationship between monthly income and *employee attrition* level. The data used in this analysis is expected to provide a clear and informative picture of the factors that affect *employee attrition* in the company.

http://ejournal.seaninstitute.or.id/index.php/Ekonomi
Jurnal Ekonomi, Volume 12, No 02, 2023
ISSN: 2301-6280 (print) ISSN: 2721-9879 (online)

Figure 2 Dashboard Page

Employee Attrition Prediction Menu

The employee attrition prediction menu is a feature in the application that is used to predict the chances of *exit (attrition)* from employees. This feature asks for input from 10 features that influence employees' decisions to leave the company. Such features are: "Age", "Daily Rate", "Distance From Home", "Hourly Rate", "Job Role", "Monthly Income", "Monthly Rate", "Percent Salary Increase", "Total Years Worked", Year in the Company".

After these features are inputted, the application will process the data and issue the results of predicting employee exit opportunities in the form of percentages. This feature can be used by company management to make decisions in reducing employee exit rates.

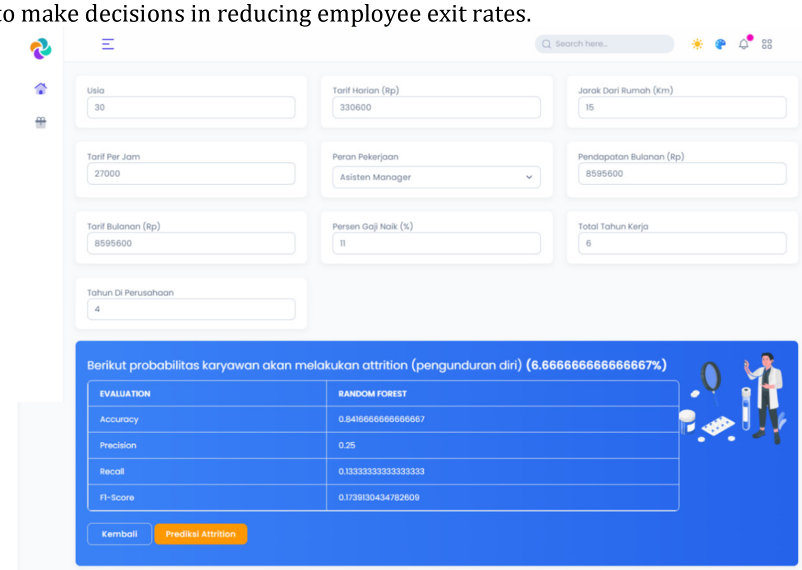

Figure 3 Employee Attrition Prediction Menu Page

From the results of figure 3 above the application analysis that has been done, it can be concluded that the features "Age", "Daily Rate", "Distance From Home", "Hourly Rate", "Job Role", "Monthly Income", "Monthly Rate", "Percent Salary Increase", "Total Years Worked", "Year In Company" have a significant influence on the prediction of *employee attrition*. In addition, through the prediction feature that the author has provided, it can be seen that *employee attrition probability* can be estimated with a fairly good level of accuracy. However, keep in mind that the results of this analysis only apply to the data used in this application and cannot be generalized to other data.

5. CONCLUSION

The results of this study show that feature engineering significantly influences performance to predict employee attrition at PT Kimia Farma Tbk. using Logistic Regression and Random Forest models. It can be seen that the application of feature engineering can affect the accuracy, precision, recall, and F-Score of both methods. The Recursive Feature Elimination (RFE) method with the Logistic Regression model has an accuracy of 0.866, precision of 0.5, recall of 0.159 , and F-Score of 0.259. Meanwhile, RFE with the Random Forest model has an accuracy of 0.886, precision of 0.916, recall of 0.25, and F-Score of 0.392. The SelectKBest method with the Logistic Regression model has an accuracy of 0.88, precision of 0.9, recall of 0.204 , and F-Score of 0.333. Meanwhile, SelectKBest with the Random Forest model has an accuracy of 0.87, precision of 0.818, recall of 0.204 , and F-Score of 0.327. According to the performance comparison results, the RFE (Recursive Feature Elimination) method with the Random Forest model can be said to be the best method in terms of accuracy and precision. Although the recall of this method is slightly lower, the performance of this method still meets its criteria as a good method. Therefore, the Recursive Feature Elimination method with the Random Forest model was chosen as the best method for this case.

REFERENCE

[1] A. K. Singh and G. K. Kaur. (2009), "An overview of feature selection techniques in bioinformatics", *Journal of Biomedical Informatics*, Vol. 23 No. 19, pp. 2507–2517.
[2] A.M. Yusoff. (2020), "The effect of work-life balance on employee turnover intention: The mediating role of job satisfaction", *International Journal of Human Resource Management*, Vol. 13 No. 34, pp. 23–25.
[3] Arifin, A.L. (2022), "Human Resource Planning and Selection Strategy at PT Kimia Farma, tbk", *Majalah Sains Bijak*, Vol. 19 No. 1, pp. 34–46.
[4] Breiman, L. (2001), "Random Forests", *Machine Learning*, Vol. 45, pp. 5–32.
[5] Choe, J.Y., Lee, C.U. and Kim, S.K. (2023), "Association between Novel Hematological Indices and Measures of Disease Activity in Patients with Rheumatoid Arthritis", *Medicina (Lithuania)*, MDPI, Vol. 59 No. 1, doi: 10.3390/medicina59010117.
[6] C.M. Caruana and R.Y. Niculescu-Mizil. (2015), "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable", *Communications of the ACM*, Vol. 3 No. 6, pp. 12–18.
[7] D. Domingos. (1998), "An Empirical Comparison of Supervised Learning Algorithms", *Proceedings of the Fifteenth International Conference on Machine Learning*, Vol. 5 No. 19, pp. 31–45.
[8] D. W. Hosmer and S. Lemeshow. (1989), "An Introduction to Logistic Regression", *Journal of the American Statistical Association*, Vol. 2 No. 14, pp. 24–36.
[9] Fallucchi, F., Coladangelo, M., Giuliano, R. and de Luca, E.W. (2020), "Predicting employee attrition using machine learning techniques", *Computers*, MDPI AG, Vol. 9 No. 4, pp. 1–17, doi: 10.3390/computers9040086.
[10] Gao, G., Wang, M., Huang, H. and Tang, W. (2021), "Agricultural Irrigation Area Prediction based on Improved Random Forest Model ", *Journal Research Square*, Vol. 34 No. 32, pp. 14–15.
[11] Gareth Dwyer. (2019), "A Beginner's Guide to Building Web Applications with Flask", *PyCon India*, Vol. 1 No. 21, pp. 43–45.
[12] H. Liu and H. Motoda. (2008), "A survey on feature selection methods", *International Journal of Knowledge Discovery and Data Mining*, Vol. 13 No. 18, pp. 23–25.
[13] H.-P. Kao and C.-J. Lin. (1998), "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3 No. 17, pp. 12–19.
[14] J. Bergstra and Y. Bengio. (2012), "A Practical Guide to Hyperparameter Tuning in Machine Learning", *Journal of Machine Learning Research*, Vol. 16 No. 25, pp. 67–70.
[15] Jain, P.K., Jain, M. and Pamula, R. (2020), "Explaining and predicting employees' attrition: a machine learning approach", *SN Applied Sciences*, Springer Nature, Vol. 2 No. 4, doi: 10.1007/s42452-020-2519-4.
[16] J.C. Scott and T.L. Cummings. (1997), "The impact of pay dissatisfaction on employee turnover", *Academy of Management Journal*, Vol. 8 No. 16, pp. 36–38.

[17] J.E. Dutton and S.B. Roberts. (1976), "The influence of career development opportunities on employee turnover", *Academy of Management Journal*, Vol. 54 No. 33, pp. 67–78.
[18] J.E. Dutton and S.R. Dukerich. (1991), "The effects of physical work environment on employee turnover", *Administrative Science Quarterly* , Vol. 16 No. 23, pp. 125–132.
[19] Juvitayapun, T. (2021), "Employee Turnover Prediction: The impact of employee event features on interpretable machine learning methods", KST 2021 - 2021 *13th International Conference Knowledge and Smart Technology*, Institute of Electrical and Electronics Engineers Inc., pp. 181–185, doi: 10.1109/KST51265.2021.9415794.
[20] K. K. Liu and H. Motoda. (2003), "A comparison of feature selection techniques for text classification", *International Journal of Intelligent Systems*, Vol. 15 No. 27, pp. 25–46.
[21] Kamath, R.S., Jamsandekar, S.S. and Naik, P.G. (2019), "Machine Learning Approach for Employee Attrition Analysis", *International Journal of Trend in Scientific Research and Development-IJTSRD*, Vol. 34 No. 17, pp. 62–67.
[22] L. Davis and M. Goadrich. (2006), "A systematic analysis of performance measures for classification tasks", *Journal of Machine Learning Research*, Vol. 3 No. 16, pp. 57–69.
[23] M. Ivancevich and Michael T. Matteson. (1980), "An examination of employee turnover", *Journal of Applied Psychology*, Vol. 45 No. 67, pp. 56–90.
[24] Mansor, N., Kebangsaan, U., Bangi, M. and Aliff, M.M. (2021), "Machine Learning for Predicting Employee Attrition", *IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 12 No. 11, pp. 435–445.
[25] Mhatre, A., Mahalingam, A., Narayanan, M., Nair, A. and Jaju, S. (2020), "Predicting Employee Attrition along with Identifying High Risk Employees using Big Data and Machine Learning", *Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN* 2020, Institute of Electrical and Electronics Engineers Inc., pp. 269–276, doi: 10.1109/ICACCCN51052.2020.9362933.
[26] Ott, R.L. and Longnecker, M.T. (2011), *An Introduction to Statistical Methods and Data Analysis*, 7th Edition., Duxbury Press.
[27] P.A.M. Hessels. (2021), "The impact of job insecurity on turnover intentions: The role of psychological contract fulfillment", *Journal of Occupational Health Psychology*, Vol. 34 No. 42, pp. 45–67.
[28] Ponnuru, S.R., Merugumala, G.K., Padigala, S., Vanga, R. and Kantapalli, B. (2020), "Employee Attrition Prediction using Logistic Regression", International Journal for Research in Applied Science and Engineering Technology, *International Journal for Research in Applied Science and Engineering* Technology (IJRASET), Vol. 8 No. 5, pp. 2871–2875, doi: 10.22214/ijraset.2020.5481.
[29] Pratt, M., Boudhane, M. and Cakula, S. (2021), "Employee Attrition Estimation Using Random Forest Algorithm", *Baltic Journal of Modern Computing*, University of Latvia, Vol. 9 No. 1, pp. 49–66, doi: 10.22364/BJMC.2021.9.1.04.
[30] Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R. and Alghamdi, H.S. (2021), "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods", *International Journal of Machine Learning and Computing*, Vol. 11 No. 2, pp. 110–114, doi: 10.18178/ijmlc.2021.11.2.1022.
[31] R.D. Griffeth, P.E.H. and M.S. Gaertner. (2000), "An examination of the relationships between job satisfaction and organizational commitment with turnover intentions and turnover: A meta-analysis", *Journal of Applied Psychology*, Vol. 5 No. 27, pp. 46–79.
[32] Sarah S. Alduayj and Kashif Rajpoot. (2018), "Predicting Employee Attrition using Machine Learning", *13th International Conference on Innovations in Information Technology (IIT)*, IEEE, pp. 93–98.
[33] T. Fawcett. (2005), "ROC Graphs: Notes and Practical Considerations for Researchers", *Journal Machine Learning*, Vol. 23 No. 56, pp. 34–46.
[34] Umami, A. (2018), "Classification of Factors Affecting Employee Reduction in 'XYZ' Company", *Academia Journal*, Surabaya, pp. 4–17.
[35] Yahia, N. ben, Hlel, J. and Colomo-Palacios, R. (2021), "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction", *IEEE Access*, Institute of Electrical and Electronics Engineers Inc., Vol. 9, pp. 60447–60458, doi: 10.1109/ACCESS.2021.3074559.