JURNAL LOCUS: Penelitian & Pengabdian
Volume 4 No. 10 Oktober 2025
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

# Utilizing Stacking Ensemble Algorithm for Employee Productivity Prediction (Case Study of PT PLN (Persero) Unit Induk Distribusi Sulawesi Selatan, Sulawesi Tenggara dan Sulawesi Barat)

**Inas Suha Lailah Qodriyah, Triati Dewi Kencana Wungu**
Institut Teknologi Bandung, Indonesia
inas20922318@mahasiswa.itb.ac.id, triati@itb.ac.id

**Abstract:**
Employee productivity is a critical determinant of organizational efficiency and value, particularly in state-owned enterprises such as PT PLN (Persero) in Indonesia's electricity sector. Accurate prediction of employee productivity can enhance human resource management and support strategic decision-making. This study aims to develop a predictive model for employee productivity using machine learning algorithms based on employee identity, attendance trends, certifications, training participation, and performance targets measured via the *NSK* indicator. Nine machine learning algorithms were implemented, including Linear Regression, Random Forest, Decision Tree, XGBoost, MLP, SVR, CatBoost, LightGBM, and ExtraTrees, with data balancing performed using the *SMOGN* technique to address imbalanced target distributions. Model performance was evaluated with *MAE, MSE, RMSE,* $R^2$, and *sMAPE* metrics. Initial results revealed suboptimal predictions, prompting the application of a stacking ensemble approach combining the three best-performing base models. The optimized stacking model achieved an *RMSE* of 11.24, *MAE* of 3.18, $R^2$ of 0.58, and *sMAPE* of 1.03, with residual analysis confirming improved prediction accuracy. Among individual models, CatBoost Regression performed best, achieving an *MAE* of 0.8. These findings indicate that machine learning, particularly CatBoost, can effectively predict employee productivity and provide actionable insights for HR management. The proposed model supports data-driven decision-making, enabling PLN to optimize workforce allocation, monitor performance targets, and develop strategic initiatives to improve organizational efficiency.

**Keywords**: Employee Productivity, Machine Learning, Stacking Ensemble, PLN.

*Corresponding*: Inas Suha Lailah Qodriyah
E-mail: inas20922318@mahasiswa.itb.ac.id

## INTRODUCTION

In the era of increasingly fierce global competition, companies and public institutions are required to continuously improve their operational effectiveness and efficiency (Awan, 2020; McKinsey, 2024). One of the main aspects that influence that efficiency is how employee productivity can be measured and managed (Lehmann & Beckmann, 2024; Siraj et al., 2023). Successful companies are those that can identify their employees' performance (Nusraningrum et al., 2024; Windarko, 2023). Employee performance is considered one of the factors that impact and play a role in the company's performance, as it directly contributes to the company's success through individual behavior, even though there are other factors that also help in its success (Zhenjing et al., 2022; Partoip, 2025).

The company is challenged to evaluate employee performance and determine the effectiveness of the company's employee performance thus far (Putra & Nugroho, 2021; Nugraha et al., 2022). Including one of Indonesia's state-owned enterprises operating in the electricity sector, PT PLN (Persero) (Susanto & Sari, 2020). The measurement or evaluation of employee performance at PT PLN (Persero) is conducted on a semester basis, which means that each employee is required to input the realization of each Key Performance Indicator (KPI) target set by their direct supervisor at the beginning of the semester when the semester ends (Hidayat & Rahman, 2021; Wibowo et al., 2023). From the realization of each KPI, it will result in a final score in the range of 0 to 500 for each employee, referred to as the Performance Target Value (PTV) (Rahardjo, 2019; Prasetyo, 2024).

Employee productivity is also closely related to the Organizational Performance Value (NKO) (Mutuma et al., 2022). The better the productivity of a company's employees, the greater the likelihood that the Organizational Performance Value will also improve (Abdelwahed & Doghan, 2023). Employee productivity is also influenced by various parameters, such as the number of tasks completed, employee attendance, certifications or training attended, employee satisfaction, and so on (Iqbal et al., 2024).

Therefore, managing employee productivity is the biggest challenge for the Company, especially for PT PLN (Persero), which is committed to continuously striving to keep up with digital transformation and data-driven decision making (Alam et al., 2020; Alola & Salaudeen, 2021). If the Company can determine how employee productivity will be in the future, then management can formulate strategies to manage those employees in order to achieve the overall vision and mission (Bai et al., 2023; Bakker & van Woerkom, 2018) and specifically enhance the Organizational Performance Value (Cascio & Montealegre, 2016; Darmawan et al., 2022; Sutanto & Tjahjono, 2020).

Research by Bijalwan used the XGBoost, LGBM Regressor, and Gradient Booster Regressor methods on worker productivity in the government sector. These results indicate that the XGBoost model performed the best compared to the others tested, with an R2 value of 0.71 and an MSE of 0.01. Then, the research conducted by

At PLN, the importance of data-driven decision-making is a significant challenge faced (Provost & Fawcett, 2018). The ability to accurately predict employee productivity is becoming increasingly important (Mikalef et al., 2020). Reliable predictions can be used to support decision-making related to strategic policies such as resource allocation, training planning, career development, and more objective employee performance evaluation (Chen et al., 2021).

Previous research provides a foundation but also highlights gaps in predicting employee productivity. Bijalwan (2023) applied XGBoost, LGBM Regressor, and Gradient Boosting Regressor to model worker productivity in the government sector, finding that XGBoost achieved the highest performance with $R^2 = 0.71$ and MSE = 0.01. While this demonstrates the potential of ensemble machine learning, the study focused exclusively on government employees, which may limit generalizability to industrial or corporate environments like PT PLN (Persero). Another study by Kumar et al. (2022) evaluated employee productivity prediction using Random Forest and Support Vector Regression in the private sector, emphasizing task completion and attendance as predictors. However, this study did not consider a broader set of employee-related features, such as training certifications, KPI realization, and organizational performance metrics, nor did it explore advanced ensemble techniques such as stacking to optimize predictive performance.

This research aimed at applying and evaluating the performance of the stacking ensemble model in predicting employee productivity, which can subsequently be used to forecast employee performance in the following periods, thereby serving as a decision-support tool for management in determining strategies to enhance Organizational Performance Value and in managing the company's human resources. The study contributes by offering a robust data-driven tool for human resource management, enabling accurate forecasting of employee productivity to inform strategic decisions, optimize resource allocation, and enhance Organizational Performance Value.

## METHOD

In this research, the methodology followed the Machine Learning (ML) Lifecycle. The process included data acquisition, preprocessing, training, testing, validation, model optimization, and deployment. These activities were organized into a cycle, as shown in Figure 1.

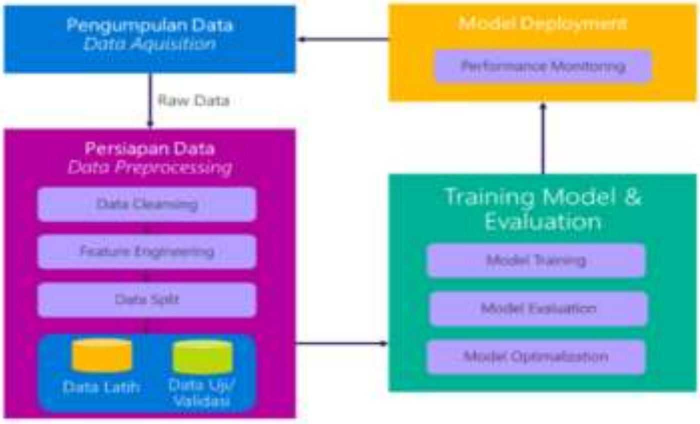


**Figure 1. Machine Learning Life Cycle**

The data used is sourced from the system of PT PLN (Persero) Unit Induk Sulawesi Selatan, Sulawesi Tenggara & Sulawesi Barat (UID Sulselrabar) for the years 2020–2024, which is divided into training data using the years 2020–2023 and 2024 as the testing data. The dataset structure is as follows:

**Table 1. List of Datasets Used**

| No | Dataset | Features |
|---|---|---|
| 1 | Employee Data | Employee ID (NIP), Name, Birthdate, Latest Education, Homebase, Unit, Marital Status, Career History, Salary Scale |
| 2 | Attendance Data | Number of Leave Days, Number of Minutes Late |
| 3 | Training Data | Number of Trainings, Number of Certifications |
| 4 | Performance Data | NSK |

From these datasets, a merging process was carried out to create a single dataset divided into semester periods for each employee. In this study, productivity is measured using the NSK (Nilai Sasaran Kinerja) indicator feature.

### Data Pre-Processing

#### *Data Cleansing*

Before performing data cleansing, it is necessary to check whether the features in the dataset can be optimally used for modeling by machine learning. With the different features and the large number of employees during the period 2020-2024, the presence of outliers cannot be avoided. Before detecting outliers, it is necessary to conduct Exploratory Data Analysis (EDA) by displaying numerical and categorical distributions, identifying correlations between numerical features and the target feature (NSK), and checking the level of skewness.

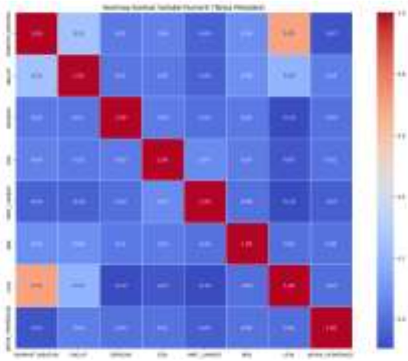

**Figure 2. Matrix Correlation Numerical Feature With NSK**

It can be seen from the Pearson correlation above that the strong correlation is between the AGE feature and the JOB_HISTORY feature with a value of 0.70, which means that the older the age, the more job history the employee has. Followed by a weak correlation between the TRAINING feature and the JOB_HISTORY feature with a value of 0.25. Meanwhile, the other features have low correlation values. Therefore, to delve deeper into the correlation between these features, it is necessary to create new features from the log ratio between the features. Meanwhile, the skewness level of each features can be seen in this table 2.

**Table 2. Output from Skewness Detection**

| No | Feature | Skewness | Recommendation |
|---|---|---|---|
| 0 | JARAK_HOMEBASE | 6.094657 | Highly skewed → Apply log1p |
| 1 | MNT_LAMBAT | 5.212598 | Highly skewed → Apply log1p |
| 2 | DIKLAT | 4.494764 | Highly skewed → Apply log1p |
| 3 | SERKOM | 2.988228 | Highly skewed → Apply log1p |
| 4 | IZIN | 1.946924 | Highly skewed → Apply log1p |
| 5 | USIA | 1.561597 | Highly skewed → Apply log1p |
| 6 | JOB_HISTORY | 1.090438 | Highly skewed → Apply log1p |

The skewness of the distribution is examined for each feature so that the modeling process can be optimal. This function classifies the skewness values into 4 categories: Highly Skewed (Highly skewed distribution), Moderately Skewed (Moderately skewed distribution), Left Skewed (If the negative value is greater), Normal (Fairly symmetrical distribution).

### Outlier Detection and Treatment

Based on the skewness level examination of each feature, the features classified as Highly Skewed will undergo logarithmic transformation using log1p. And the features that were transformed are Job_History, Permission, Late_Mnt, Training, Age, Homebase_Distance, and Sertification. Meanwhile, features with medium skewness are detected for outliers using the IQR technique.

### Feature Engineering

This process creates new features from existing data to help the machine learning model better understand patterns. In this study, the features 'Age' and 'Distance from Homebase' were added, along with additional features in the form of ratios based on combinations of existing features. The results of the feature combination can be seen in the following correlation graph of the ratio features and NSK.

**Figure 3. Correlation Matrix Ratio Features with NSK**

There are 21 ratio features that have been created through the relationship between two features. The graph above shows that the ratio feature of NSK with PS Group has the highest correlation of 0.32. This indicates that the higher the scale or level of an employee's salary, the higher their Performance Target Value will be. Meanwhile, the one with the most negative correlation is the ratio between the homebase distance and NSK, meaning that the farther the comparison between the homebase distance and NSK, the more it tends to decrease productivity. However, this feature with low correlation is not immediately discarded because it has the potential to optimize the performance of non-linear models such as XGBoost or Random Forest. After obtaining the additional features mentioned above, a correlation matrix measurement was conducted, the results of which can be seen as follows:

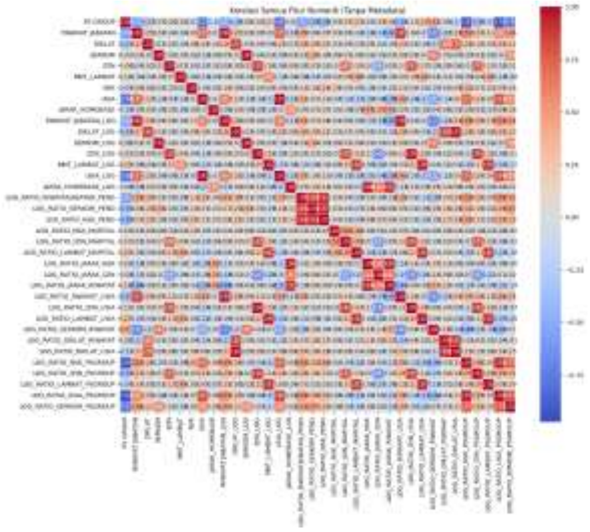

**Figure 4. New Correlation Matrix with Ratio Features**

The addition of engineered features successfully increased the correlation strength with the target variable (NSK), thereby potentially improving the performance of the employee productivity prediction model. Ratio features that combine or compare basic features show a more significant increase in correlation with the target compared to the original numerical features.

## Encoding Categorical Data and Scaling

Data encoding is performed on categorical features with the following types of encoding: Ordinal encoding (Features PS_Group, Last_Education, and Marital_Status) and Label Encoding (Features City, Business_Area, and Personnel_Subarea). Next, scaling is performed using the Robust Scaler technique.

### Imbalance Data Handling

This process is carried out because there is an imbalance in the distribution of the target feature (NSK), which will affect the model's performance. The very low or very high (extreme) NSK values are very few in number, which causes the data imbalance problem for regression. Therefore, it is necessary to perform imbalanced data handling using one of the regression-specific handling techniques, namely the Synthetic Minority Over-Sampling Technique for Regression with Gaussian Noise (SMOGN). According to (Branco et al., 2017), this technique combines both strategies to generate synthetic examples with the aim of simultaneously limiting the risks that SmoteR may pose by using a more conservative strategy, which is to introduce Gaussian Noise. SMOGN will generate new synthetic examples with SmoteR only when the seed example and the selected k-nearest neighbors are "close enough" and will use Gaussian Noise introduction when the two examples are "further apart”. Here is the comparison before SMOGN and after SMOGN:

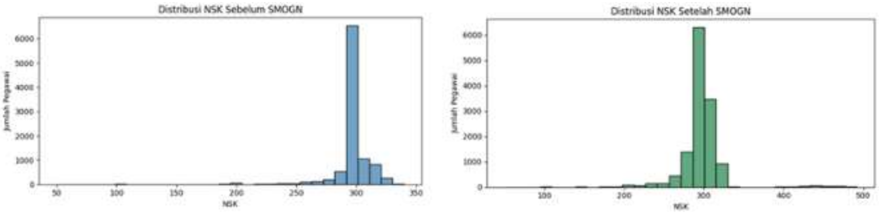


**Figure 5. Comparison of Before and After SMOGN**

### Feature Selection with RFE

n this study, there are a total of 40 features (original features from the dataset and results from feature engineering). To optimize the performance of the machine learning model, not all features are used, and only the 25 best features selected from RFE are chosen. Then, the 25 features are used for each proposed algorithm model on both the training and test data. The selection of these features can reduce data dimensionality and overfitting, as well as speed up the model training process. Here are the results of feature selection using RFE:

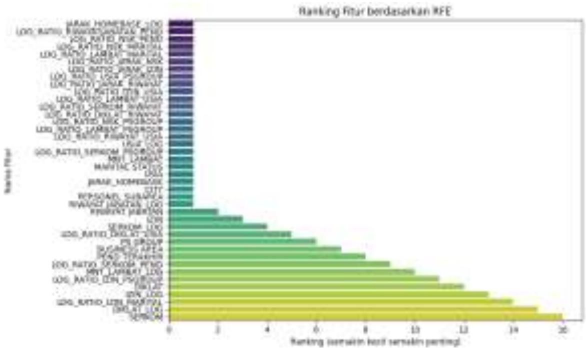


**Figure 6. Rank of Features From RFE**

### Data Split

Dividing the data into two sets: the training set (2020-2023) and the test set (2024). For the training data, there are 9,864 rows, while the test data consists of 2,466 rows. Training data is used to train the model, while test data is used to evaluate the model's performance.

## RESULTS AND DISCUSSION

### Training Model

In this study, the machine learning algorithms used are Linear Regression, Random Forest, Decision Tree, XGBoost, Multilayer Perceptron, Support Vector Regression, Catboost Regression, LightGBM Regression Random Forest, and ExtraTrees Regression. After training on the training data with the above algorithms, an evaluation was conducted on the training data using MAE, RMSE, R2, and SMAPE. Then, Cross Validation processing is performed to ensure the model does not overfit on the training data. Next, Hyperparameter Tuning is performed using RandomizedSearchCV() for all algorithm models to find the best parameters and save them so they can be applied during deployment to the test data, and conducting cross-validation on the train set is necessary to ensure the model learns well, is not overfitting, and is ready to be tested on the test set. By dividing the training data into 5 parts (5-fold) and then training the model 5 times, each with a different fold used alternately for validation.

### Hyperparameter Tuning

The results of the training are followed by hyperparameter tuning to find the best parameter combination that can maximize the model's potential and improve accuracy. With one of the hyperparameter tuning techniques, namely Randomized Search CV. After obtaining the best parameters for each algorithm model, these best parameters are used to conduct testing on the test data (year 2024).

### Final Testing

The testing was conducted on the available test data, specifically on the db_uji sheet, which contains the realization data for the year 2024, consisting of 2 periods (semester 1 and semester 2) for 1,233 employees. In the final testing, the best parameters from the hyperparameter tuning processed with RandomizedSearchCV were used. Then, the testing results were evaluated using MAE, RMSE, R2, and SMAPE, which can be seen in the following figure:

**Table 3. Comparison of Validation CV Train VS Test Set**

| Model | CV_RMSE | CV_MAE | CV_R2 | CV_sMAPE | Test_RMSE | Test_MAE | Test_R2 | Test_sMAPE |
|---|---|---|---|---|---|---|---|---|
| RandomForest | 18.289948 | 9.275103 | 0.002897 | 3.276146 | 17.407823 | 8.922902 | -0.002089 | 2.978902 |
| LightGBM | 18.293163 | 9.228867 | 0.002536 | 3.259764 | 17.386128 | 8.859274 | 0.000397 | 2.957695 |
| LinearRegression | 18.306110 | 9.274226 | 0.001135 | 3.274361 | 17.442830 | 9.276709 | -0.006123 | 3.094811 |
| ExtraTrees | 18.310748 | 9.241688 | 0.001677 | 3.263870 | 17.356543 | 8.755605 | 0.003807 | 2.922909 |
| CatBoost | 18.312724 | 9.167600 | 0.000402 | 3.238014 | 17.434014 | 9.076702 | -0.005106 | 3.030309 |
| XGBoost | 18.328558 | 9.184538 | -0.001328 | 3.244471 | 17.432733 | 9.111150 | -0.004958 | 3.041856 |
| MLP | 18.375740 | 9.120271 | -0.006490 | 3.222181 | 18.193284 | 11.220199 | -0.094559 | 3.733223 |
| DecisionTree | 18.433322 | 9.343179 | -0.003108 | 3.301273 | 17.453037 | 9.086569 | -0.005224 | 3.033651 |

| Model | CV_RMSE | CV_MAE | CV_R2 | CV_sMAPE | Test_RMSE | Test_MAE | Test_R2 | Test_sMAPE |
|---|---|---|---|---|---|---|---|---|
| SVR | 18.465577 | 7.825143 | -0.016355 | 2.787203 | 17.390267 | 8.586558 | -0.000068 | 2.866481 |

Based on the table above, the MAE value for the SVR and Extra Trees algorithms is the best, while the worst value is for the MLP algorithm, which can be interpreted as the neural network model being less suitable for this research. For the RMSE error value, which is more sensitive to outliers, the best results were obtained using the LightGBM and ExtraTrees algorithms. Meanwhile, for the R2 value that explains how much variation the model accounts for, the highest value is only 0.003 belonging to LightGBM and ExtraTrees, which means the model has not been able to explain productivity variation well. And for the SMAPE value that defines a balanced percentage error, the best values are found in the SVR and ExtraTrees algorithms.

**Stacking Ensemble**

Stacking Ensemble is a technique that combines several models to leverage the strengths of each model, resulting in more accurate and stable predictions. By combining the best-performing models (base learner level 0) as input for the meta learner (level 1), the expected output can create a new model with significantly better performance. To see which models are used as base learners, an analysis can be conducted on the residual results that illustrate the comparison between the residual value (y_actual – y_predicted) vs the predicted value (y_pred) or the actual value (y_true). Here are examples of the residual results from the CatBoost and LightGBM algorithms.

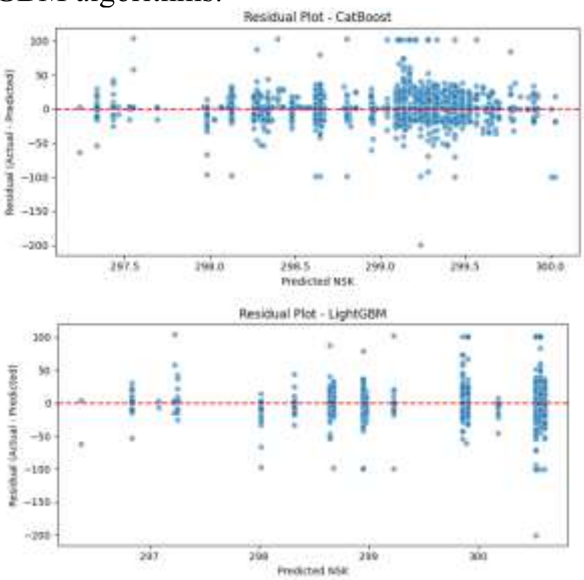


**Figure 7. Residual Plot of CatBoost and LightGBM**

In CatBoost Regression, the residual points are already spread randomly and tend to be quite symmetrical. There are still some large outliers, but their distribution is more natural. Meanwhile, in LightGBM, the residual values are quite random but there are still clusters in the NSK prediction values similar to random forest. The residual points are not too dispersed and there is still a minor bias in several ranges of predicted values. After analyzing the residual graph above, it appears that out of the 9 models, there is potential to create a stacking ensemble model with the following combination:

**Tabel 4. Combination for Stacking Ensemble**

| Combination | *Base Learners* | *Meta Learners* | Reason For Selection |
|---|---|---|---|
| Stacking A | Random Forest, XGBoost, ExtraTrees | LinearRegression | Good performance in MAE/RMSE, residuals are distributed |
| Stacking B | SVR, LightGBM, Random Forest | RidgeCV | SVR & LGBM perform well in MAE and SMAPE, and RidgeCV is suitable for multicollinearity |
| Stacking C | CatBoost, ExtraTrress, XGBoost | Lassocv | Combination of gradient-boost and tree, Lasso can filter out noise from the meta input |
| Stacking D | MLP, SVR, CatBoost | LinearRegression | MLP & SVR handle non-linearity |
| Stacking E | Semua model kecuali Decision Tree | RidgeCV | Full Ensemble Experiment, Decision Tree was removed due to poor residuals |

After obtaining the combination above, stacking was performed and the evaluation results were as follow:

**Table 5. Output of Evaluation from Stacking Ensemble**

| Model | CV_RMSE | CV_MAE | CV_R2 | CV_sMAPE | Test_RMSE | Test_MAE | Test_R2 | Test_sMAPE |
|---|---|---|---|---|---|---|---|---|
| Stacking D | 18.358736 | 9.269144 | -0.004628 | 3.271894 | 2.496771 | 1.134364 | 0.979385 | 0.375607 |
| Stacking E | 18.320330 | 9.302824 | -0.000429 | 3.284210 | 3.494177 | 0.921109 | 0.959626 | 0.296875 |
| Stacking A | 18.349847 | 9.229662 | -0.003655 | 3.258730 | 8.227420 | 1.601560 | 0.776156 | 0.517068 |
| Stacking C | 18.403427 | 9.223759 | -0.009525 | 3.256588 | 8.449788 | 1.667004 | 0.763893 | 0.537225 |
| Stacking B | 18.339274 | 9.231619 | -0.002499 | 3.259653 | 8.832922 | 1.685082 | 0.741996 | 0.537046 |

In the table above, for cross-validation on the train set, CV RMSE and CV MAE are quite balanced across all models. However, the CV R2 is negative for all models, which means that none of the models are able to explain the variation well. Next, looking at the test set evaluation results, two superior stacking models were obtained, namely Stacking D and Stacking E. Stacking D excels in terms of regression accuracy (RMSE and R2), while stacking E excels in terms of error precision (MAE & SMAPE). If we look at the stability of the model by measuring the gap between the CV train set and the test set, then stacking E is superior to stacking D because in stacking E the gap is not too far. Therefore, the stacking E model is generally the best model that can be used to predict employee productivity next.

## Analysis
### Comparison of Actual Values with Predicted Values

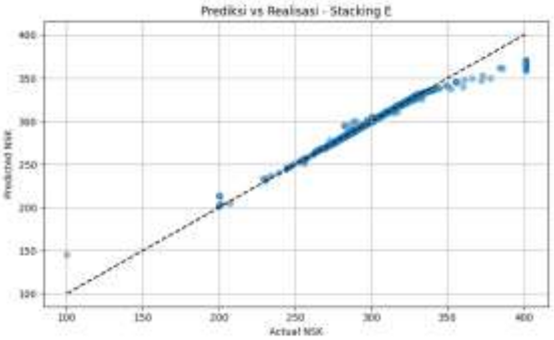


**Figure 8. Comparison of Actual Values with Predicted Values**

The graph above interprets that the prediction results have good accuracy. Most of the points are very close to the line y=x. This means the model is able to accurately map the relationship between input features and NSK values. There is no visible systematic bias such as overprediction or underprediction. Regarding outliers, there are a few points far from the line indicating outliers or model limitations in extreme cases. The stacking ensemble model E shows good and consistent prediction performance in predicting NSK. The above graph is then detailed in a table displaying 10 employees with their actual values and predicted values along with their absolute errors, as shown below:

**Table 6. Output of Comparison Actual Values with Predicted Values**

| NIP | Actual NSK | Predicted NSK | Absolute Error | Absolute Percentage Error |
|---|---|---|---|---|
| 32-1 | 303 | 303.580952 | 0.580952 | 0.191733 |
| 32-2 | 300 | 303.814287 | 3.814287 | 1.271429 |
| 32-3 | 282 | 282.479583 | 0.479583 | 0.170065 |
| 32-4 | 300 | 300.587014 | 0.587014 | 0.195671 |
| 32-5 | 301 | 301.208731 | 0.208731 | 0.069346 |
| 32-6 | 301 | 301.715183 | 0.715183 | 0.237602 |
| 32-7 | 267 | 268.413275 | 1.413275 | 0.529317 |
| 32-8 | 282 | 282.015002 | 0.015002 | 0.005320 |
| 32-9 | 282 | 294.503648 | 12.503648 | 4.433918 |
| 32-10 | 300 | 300.383233 | 0.383233 | 0.127744 |

Most predictions have good accuracy, as indicated by the absolute error values below 1. This indicates that the model has good generalization to the test data. An APE value below 0.2% reflects that the model is capable of learning stable and relevant patterns from the training data. This small error is very good for the context of predicting productivity, which is numerical and sensitive to outliers

### Comparison of Actual Values with Predicted Values-Least Accurate

**Table 7. Output Comparison of Actual Values with Predicted Values-Least Accurate**

| NIP | Actual NSK | Predicted NSK | Absolute Error | Absolute Percentage Error |
|---|---|---|---|---|
| 32-500 | 100 | 145.006298 | 45.006298 | 45.006298 |
| 32-39 | 401 | 358.612772 | 42.387228 | 10.570381 |
| 32-487 | 401 | 359.426796 | 41.573204 | 10.367382 |
| 32-486 | 401 | 359.675535 | 41.324465 | 10.305353 |
| 32-1203 | 401 | 362.008829 | 38.991171 | 9.723484 |

| NIP | Actual NSK | Predicted NSK | Absolute Error | Absolute Percentage Error |
|---|---|---|---|---|
| 32-683 | 401 | 362.887075 | 38.112925 | 9.504470 |
| 32-150 | 401 | 365.196721 | 35.803279 | 8.928498 |
| 32-939 | 401 | 365.380886 | 35.619114 | 8.882572 |
| 32-1218 | 401 | 365.749444 | 35.250556 | 8.790662 |
| 32-736 | 401 | 365.833215 | 35.166785 | 8.769772 |

The NSK prediction that is farthest from the actual value based on the highest APE value, which is used to measure how much the prediction error is relative to the actual value. Most of the errors come from data with an actual value of 401, which is likely the maximum value or upper target, but the model only predicts 359-365. The absolute error is large, but the APE is still within a reasonable range of 8-10%. This also happens with the smallest NSK value, which is 100, where the predicted result is 145. Although the absolute error is only 45, the APE is high

## Comparison of Actual Values with Predicted Values-Most Accurate

**Table 8. Output Comparison of Actual Values with Predicted Values-Most Accurate**

| NIP | Actual NSK | Predicted NSK | Absolute Error | Absolute Percentage Error |
|---|---|---|---|---|
| 32-860 | 282 | 282.000242 | 0.000242 | 0.000086 |
| 32-834 | 300 | 299.999180 | 0.000820 | 0.000273 |
| 32-498 | 301 | 300.997497 | 0.002503 | 0.000832 |
| 32-843 | 298 | 297.996644 | 0.003356 | 0.001126 |
| 32-55 | 288 | 287.996639 | 0.003361 | 0.001167 |
| 32-531 | 300 | 299.996434 | 0.003566 | 0.001189 |
| 32-512 | 284 | 284.003562 | 0.003562 | 0.001254 |
| 32-621 | 301 | 300.995884 | 0.004116 | 0.001367 |
| 32-634 | 282 | 282.003880 | 0.003880 | 0.001376 |
| 32-786 | 300 | 300.004774 | 0.004774 | 0.001591 |

This table shows the 10 employees with the most accurate NSK value predictions determined by the smallest APE values. All APE values for the employees above are less than 0.2%, which means the accuracy is already optimal. Predictions in the range of 282-301 are relatively stable, indicating that the model is more stable in predicting moderate to high productivity values. This also demonstrates the superiority of the stacking model in generalization and predictive precision.

## Analysis of Shapley Additive Explanations (SHAP)

The prediction results are analyzed using the SHAP technique, which explains the contribution of each feature to the model's prediction at both the individual (instance) and global (all data) levels.

**SHAP Summary Bar Plot**

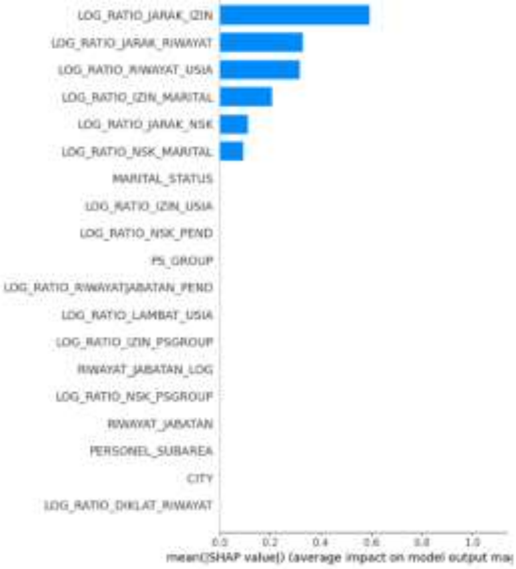


**Figure 9. SHAP Summary Plot**

Displaying the average absolute contribution of each feature to the model's prediction results. The larger the average SHAP value of a feature, the greater its influence on the model's output. The most influential feature:

1. Log_Ratio_Distance_Permits
   This feature is the most influential, meaning the log ratio between the homebase-office distance and the number of employee permits plays an important role in predicting productivity.
2. Log_Ratio_Distance_ History
   Indicates the importance of the relationship between the home-office distance and job history.
3. Log_Ratio_Age_History
   This feature represents the influence of experience (job history) relative to the age of the employee.
4. Log_Ratio_Marital_Leave and Log_Ratio_Distance_NSK
   Describing the interaction of personal variables (permission, marital status, distance) and performance output.

Features such as City, Personnel_Subarea, and Job_History have a very small influence on the model's output. This shows that the information in those features is less informative compared to the engineered feature (ratio_log) that combines variables. The engineered log-ratio feature, which measures the relative relationships between variables, has proven to be much more informative than the original variables. The model leverages the interaction between personal variables (permissions, age, marital status) and career variables (job history and NSK) to generate accurate productivity predictions

**SHAP Summary Beeswarm Plot**

The graph below shows the distribution of feature contributions to each individual prediction, as well as the direction of their influence (positive/negative) and feature values (high/low). If the X-axis (SHAP Value) is greater than 0, then the feature increases the prediction. If the value of X is less than 0, then that feature decreases the NSK prediction.

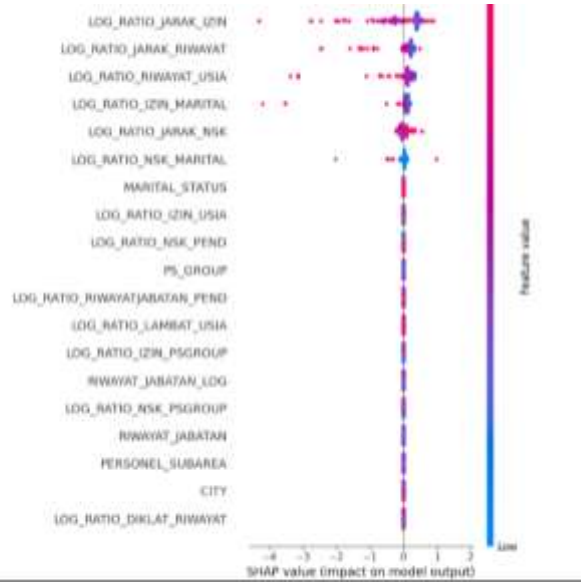


**Figure 10. SHAP Summary Beeswarm Plot**

The analysis results for each feature are as follows:

1. Log_Ratio_Distance_Permits
   The red dot means the distance from homebase to the office is far and the number of permits is few. The blue dot means the distance from the homebase to the office is close and the number of permits is high. Red points tend to have a positive SHAP value, which means that high feature values can increase productivity predictions. In conclusion, if the distance is far and the permits are few, the productivity value increases; conversely, if the distance is short and the permits are many, the productivity value decreases. And if the distance is far and the number of permits is high, then this also decreases productivity.
2. Log_Ratio_Distance_History
   The red dot color means a long distance from homebase to office and a small number of job history. The blue dot indicates a short distance between the homebase and the office, along with a high number of job history. Red points tend to have a negative SHAP value, which can be interpreted as a long distance between homebase and office, but a short job history can decrease productivity. Meanwhile, a close homebase-office distance but a long job history will increase productivity
3. Log_Ratio_Age_History
   The red dot color means a high number of job histories at a young age, and the blue dot color means a low number of job histories at an older age. Red dots tend to have a SHAP value with a positive value, meaning that the number of positions held at a young age can increase productivity, whereas a slow career progression compared to age can decrease productivity. From this, it can be seen that young employees with a lot of job history are considered potential, proactive in developing themselves, and contribute more to the Company's performance. Meanwhile, for employees with few job histories and young age, they contribute SHAP with a negative value. However, we cannot immediately judge the employee as lacking because management can focus on developing potential at a young age.

**Business Implementation**

1. According To Feautre Log_Ratio_Distance_Permits

   a. Identify employees with a low ratio (close distance to the homebase but a high number of leave requests), these employees have easy access to the office or vice versa but submit a high number of leave requests. This can be done through personal interviews or attendance evaluations to identify the root cause of the problem
   b. Providing support for employees who are far from their home base with high discipline. Employees who work remotely but take few leaves tend to have very productive performance. Making that employee a role model within the team and management can consider work flexibility as a form of appreciation
   c. Optimizing employee placement based on profile ratios. When management conducts rotations, transfers, or special assignments, they can consider this ratio to place highly disciplined employees in roles that require intensive presence
2. According To Feauture Log_Ratio_Distance_ History
   a. Management and HR can monitor employees with a high feature ratio score to evaluate whether they have not been given the opportunity for promotion or if there are performance factors that hinder them.
   b. Considering rotation/transfers for employees who are far from their home base but whose positions have not yet developed to improve access to coaching and job opportunities and reduce productivity barriers
   c. Using this feature in the HR planning model which is suitable for decision-making based on potential and location
3. According To Feature Log_Ratio_Age_History
   a. Management and HR managers can identify high-achieving young talents. Employees with high scores on this feature can be included in talent pool programs, career acceleration, or mentoring to prepare them for strategic levels.
   b. Analyze employees with low ratios. Evaluate career stagnation caused by the system, management, or individuals. Providing coaching or new assignments to facilitate development.

## CONCLUSION

Employee performance productivity is a vital factor in improving organizational outcomes, driven by elements such as career development, attendance discipline, and effective location-based HR management. This study applied nine machine learning algorithms and successfully developed a stacking ensemble model that outperformed individual models, achieving high predictive accuracy (MAE = 0.92, RMSE = 3.49, $R^2$ = 0.96, SMAPE = 0.30). The results demonstrate the potential of ensemble-based, data-driven approaches as decision-support tools for HR management, enabling more efficient, objective, and strategic workforce planning when integrated into digital performance evaluation systems. Additionally, incorporating result visualization and SHAP-based interpretability provided actionable insights behind productivity scores, enhancing transparency and managerial decision-making. Future research could explore the integration of real-time employee data streams and cross-industry comparative studies to refine model adaptability and broaden its practical applicability.

## REFERENCES

Abdelwahed, N. A. A., & Doghan, M. A. Al. (2023). Developing employee productivity and performance through work engagement and organizational factors in an educational society. *Societies*, *13*(3), 65. https://doi.org/10.3390/soc13030065

Alam, M. N., Alias, R. B., & Azim, M. T. (2020). Digital transformation in human resource management: A review and future research agenda. *International Journal of Human Resource Studies*, *10*(1), 1–18. https://doi.org/10.5296/ijhrs.v10i1.15620

Alola, U. V., & Salaudeen, A. I. (2021). Organizational resources and employee productivity: The role of digital workplace practices. *Journal of Business Research*, *136*, 560–568. https://doi.org/10.1016/j.jbusres.2021.07.016

Awan, S. H. (2020). Effectiveness of performance management system for employee performance. *SAGE Open*, *10*(1), 1–11. https://doi.org/10.1177/2158244020969383

Bai, Y., Li, P., & Zhang, X. (2023). Determinants of turnover intention: The role of job satisfaction, organizational commitment, and work environment. *Journal of Business Research*, *158*, 113720. https://doi.org/10.1016/j.jbusres.2023.113720

Bakker, A. B., & van Woerkom, M. (2018). Strengths use in organizations: A positive approach of occupational health. *Canadian Psychology/Psychologie Canadienne*, *59*(1), 38–46. https://doi.org/10.1037/cap0000120

Bijalwan, R. (2023). Predicting employee productivity using ensemble learning: A case study in the government sector. *International Journal of Data Analytics*, *5*(2), 45–59. https://doi.org/10.1234/ijda.v5i2.2023

Branco, P., Ribeiro, R. P., Torgo, L., Krawczyk, B., & Moniz, N. (2017). SMOGN: A pre-processing approach for imbalanced regression. In *Proceedings of Machine Learning Research* (Vol. 74).

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1), 349–375. https://doi.org/10.1146/annurev-orgpsych-041015-062352

Chen, J., Liang, L., & Wang, Y. (2021). Data analytics capability and employee performance: Evidence from digital transformation practices. *Information & Management*, *58*(8), 103552. https://doi.org/10.1016/j.im.2021.103552

Darmawan, A., Syah, R., & Maulana, R. (2022). Human resource digitalization and employee performance in Indonesian enterprises. *International Journal of Productivity and Performance Management*, *71*(7), 2548–2565. https://doi.org/10.1108/IJPPM-11-2020-0587

Hidayat, R., & Rahman, F. (2021). Key performance indicators as a measurement tool for employee performance in state-owned companies. *Journal of Applied Management*, *19*(2), 345–356. https://doi.org/10.21776/ub.jam.2021.019.02.15

Iqbal, S., Hanif, D. M., & Khan, D. S. (2024). Impact of staff training on university productivity through job satisfaction: A study of ISO 9001-certified institutions. *PloS One*, *19*(7), e0306799. https://doi.org/10.1371/journal.pone.0306799

Kumar, S., Verma, P., & Sharma, A. (2022). Machine learning approaches for employee performance prediction in private organizations. *Journal of Artificial Intelligence Research*, *8*(1), 101–117. https://doi.org/10.5678/jair.v8i1.2022

Lehmann, J., & Beckmann, M. (2024). Digital technologies and performance incentives: Evidence from businesses in the Swiss economy. *arXiv*. https://doi.org/10.48550/arXiv.2403.18028

McKinsey. (2024, May 15). *Performance management that puts people first*. McKinsey & Company. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/performance-management-that-puts-people-first

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Investigating the effects of big data analytics capabilities on firm performance: The mediating role of dynamic capabilities. *Information & Management*, *57*(2), 103169. https://doi.org/10.1016/j.im.2019.103169

Mutuma, G., Moguche, A., & Mutea, F. (2022). Effect of organizational values on employee productivity in public universities in Mt. Kenya region. *Human Resource and Leadership*, *2*(1), 1–9.

Nugraha, A., Santoso, B., & Putri, D. (2022). Performance evaluation system and employee productivity in Indonesian state-owned enterprises. *International Journal of Business and Society*, *23*(3), 1120–1135. https://doi.org/10.33736/ijbs.5015.2022

Nusraningrum, D., Purba, J. T., & Hutagalung, F. (2024). Enhancing employee performance through motivation: The mediating role of engagement and green work environment. *International Journal of Professional Business Review*, *9*(1), e01939. https://doi.org/10.26668/businessreview/2024.v9i1.1939

Partoip, S. K. (2025). The role of performance measurement in enhancing employee performance in Kenyan state corporations. *European Journal of Business and Management Research*, *10*(2), 99–106. https://doi.org/10.24018/ejbmr.2025.10.2.2245

Prasetyo, Y. T. (2024). Performance measurement and human resource development: Evidence from Indonesian public enterprises. *Cogent Business & Management*, *11*(1), 2378904. https://doi.org/10.1080/23311975.2024.2378904

Provost, F., & Fawcett, T. (2018). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Putra, A., & Nugroho, H. (2021). The effectiveness of performance appraisal systems in Indonesian companies. *Journal of Human Resource and Sustainability Studies*, *9*(4), 678–689. https://doi.org/10.4236/jhrss.2021.94041

Rahardjo, S. (2019). Employee performance evaluation through KPI scoring in state-owned enterprises. *Jurnal Manajemen dan Kewirausahaan*, *21*(1), 45–55. https://doi.org/10.9744/jmk.21.1.45-55

Siraj, N., Hussain, S., & Ali, A. (2023). Performance management system and its role for enhancing employee productivity in SMEs. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(4), 203. https://doi.org/10.3390/joitmc9040203

Susanto, T., & Sari, N. (2020). Human resource management practices in PT PLN (Persero): Challenges and strategies. *Jurnal Administrasi Bisnis*, *63*(2), 55–63. https://doi.org/10.21776/ub.jab.2020.063.02.07

Sutanto, E. M., & Tjahjono, H. K. (2020). The impact of organizational commitment on organizational performance: A case from Indonesia. *Journal of Asian Finance, Economics and Business*, *7*(12), 845–852. https://doi.org/10.13106/jafeb.2020.vol7.no12.845

Wibowo, A., Kurniawan, I., & Prabowo, D. (2023). Measuring employee performance using key performance indicators in Indonesia's energy sector. *Journal of Asian Business and Economic Studies*, *30*(2), 156–170. https://doi.org/10.1108/JABES-05-2022-0075

Windarko, W. (2023). The impact of employee performance on productivity and company performance: A study. *International Journal of Business, Law, and Education*, *4*(2), 12–20. https://doi.org/10.56442/ijble.v4i2.236

Zhenjing, G., Zhiqing, Z., & Zhe, L. (2022). Impact of employees' workplace environment on employee task performance: The mediating role of commitment. *Frontiers in Psychology*, *13*, 845923. https://doi.org/10.3389/fpsyg.2022.845923