Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

Cracking Overtime: Unleashing Machine Learning at PT XYZ with Linear Regression, Neural Networks, and Random Forests

Tifani Dewi Christia[1)]*, Dr. Imam Yuadi[2)]
[1,2)]*Human Resources Development, Faculty of Human Resources, Airlangga University, 60286, Surabaya, Indonesia*
[1)] tiffanytambunan@gmail.com, [2)] imam.yuadi@fisip.unair.ac.id

ABSTRACT
Excessive overtime at PT XYZ is a significant issue for the organization. Besides the significant financial repercussions, they may also affect employee health and productivity. This research aims to forecast future overtime hours, facilitating strategic planning, mitigating excessive overtime, and developing more effective overtime policies. This study employs an overtime realization dataset encompassing many characteristics that influence overtime determinations. The dataset is partitioned into training and testing data to serve as inputs for the three predictive algorithm models: linear regression, artificial neural network, and random forest. The random forest model demonstrates superior performance, evidenced by a mean squared error (MSE) of 158.78, which is proximate to the actual value. The root mean squared error (RMSE) of 12.601 is lower than that of the other two models, indicating a reduced average prediction error. The mean absolute error (MAE) of 8.931 reflects the average deviation from the actual value, while the mean absolute percentage error (MAPE) of 0.336 indicates a prediction error of 34%. Furthermore, the coefficient of determination (R^2) of 0.914 signifies that approximately 91.4% of the variation in overtime hours is accounted for, in contrast to the other models, which accounted for 78.8% and 79.6%, respectively. The results indicate that the random forest model demonstrates superior predictive accuracy compared to the other two algorithms, owing to its capacity to handle non-linear data and outliers. Consequently, the random forest model is advocated as the most efficacious method for forecasting the amount of supplementary working hours in the future.
Keywords: Linear Regression, Neural Networks, Overtime, Prediction, Random Forest

INTRODUCTION
For businesses, overtime is a difficult issue with regard to financial effects, production targets, workplace culture, labor policies, and other aspects. The level of effort required in a given period of time is largely influenced by strict operational criteria. Extra time is required here. PT XYZ's assessment of overtime use indicates that it is only permitted for specific job levels and work types, that is, for positions inside the central corporate process. Nonetheless, the evaluation and mapping of overtime realization depicted in the scatter plot graph revealed that the quantity of overtime performed across structural and functional positions was nearly equivalent, with a tendency for structural positions to accumulate more overtime hours than functional positions. The financial value allocated by the organization is influenced by the position or grade of an individual, with higher-ranking personnel receiving greater overtime compensation. This incident underscores the necessity for precise overtime forecasting to enhance labor planning and operational efficacy.

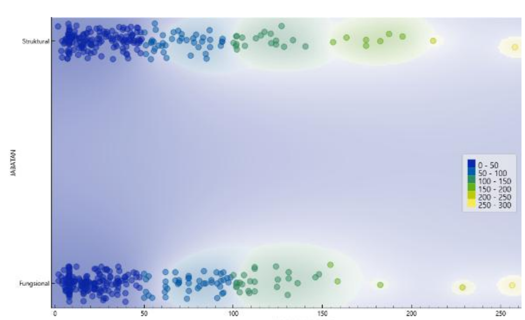

Fig. 1 Trends Overtime by Job Function (Structural & Fungsional)
(Source: Internal Data)

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

Much research has used linear regression, neural networks, and random forests among other predictive modeling methods to address these challenges. Linear regression is a direct and understandable method for separating the dependent variable (Y) from the independent variable (X). This method is appropriate for predictive modeling with structured or linear information; yet, it underfits for complex datasets since it fails to suitably represent relations in fundamental non-linear (Teng & Yuan, 2020). While too much training and insufficient data diversity may lead overfitting, which adversely affects performance when the model is evaluated with fresh data beyond the training set, neural networks are used to model complex and non-linear data that conventional linear models struggle to address. The random forest model is much valued for its great learning capability, effective handling of non-linear interactions, multiple complex variables, and outliers in predictive analysis (Jin et al., 2020). Exact calibration of these three models for comparison analysis helps to increase the accuracy of overtime estimations, thereby generating outcomes in line with corporate objectives (Zhang et al., 1998).

This study principally aims to identify whether machine learning model—linear regression, neural network, or random forest—attaches the best prediction accuracy and is both exact and dependable for forecasting overtime data at PT XYZ. Through better staff planning and future overtime laws development, this will enable management to create appropriate budgeting plans and take out strategic initiatives to lower too high overtime. First define the merits and shortcomings of every model concerning interpretability, scalability, and computational efficiency; second, evaluate which of the three models—linear regression, neural network, and random forest—yields the best predictive accuracy and performance in analyzing overtime data at PT XYZ; and third, provide management actionable insights and recommendations to choose the most effective model to improve workforce management and operational efficiency.

This work aims to evaluate random forest models, neural networks, and linear regression in overtime prediction thereby combining contemporary and traditional machine learning methods. The goal is to project overtime most effectively by means of the labor overtime data at PT XYZ. This work aims to overcome challenges with model interpretability and data complexity. This work aims to provide a foundation for upcoming research on machine learning linked to labor analysis and associated practical applications, thereby enhancing labor management techniques and lowering of overtime expenses and random forest models in predicting overtime. Forecasting overtime most successfully using the workforce overtime data at PT XYZ is the aim here. This work tries to solve problems including model interpretability and data complexity. This work seeks to give a basis for future studies on machine learning connected to labor analysis and associated applications, so improving labor management techniques and reducing of overtime costs.

LITERATURE REVIEW

In predictive analytics with data, linear regression, neural networks, and random forests are among the important machine learning methods applied.

Linear regression is a linear method for illustrating the relationship between independent and dependent variables by means of linear equations(Gede Surya Mahendra et al., 2023). Linear regression employs a linear predictor function to describe relationships; in this approach, the model parameters are unknown, necessitating estimates derived from the available data. The initial type of regression extensively examined was linear regression, which is commonly employed in practical applications (Crocker & Seber, 1980). Linear regression is a quite attractive model because of its fundamental presentation. Numerical ally expressed by a linear equation for a given set of input values (x), the expected output value for the corresponding input values (y) is found by the input values (x) (Gede Surya Mahendra et al., 2023). In previous studies using linear regression such as temperature forecasting, and soil moisture prediction although these models are straightforward, they often show lower prediction accuracy compared to more complex models such as random forests and neural networks. For example, in temperature prediction, linear regression was found to be less effective than random forests and neural networks (Manwal et al., 2024). Similarly, when predicting soil moisture, linear regression achieved 92% accuracy, slightly lower than the 94% accuracy of random forests (Tynchenko et al., 2024).

Neural network learning techniques are generally employed for discrete, real, or vector-based problems. The designed neural network can detect non-linear interactions via layers of interconnected nodes, mirroring the architecture of the human brain. Two illustrative uses of the approach are workforce behavior prediction and time series analysis. Both methods can record complex traits and exhibit remarkable adaptability (Kim, 2022; Zhang et al., 1998). Large processing capabilities and thorough hyperparameter tuning are prerequisites for neural networks to achieve optimal performance and prevent overfitting (Mathieu et al., 2016; Schmidhuber, 2015). Recent studies have employed neural networks for various predictive tasks, including temperature forecasting and the prediction of

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

superconductor critical temperatures. Neural networks frequently surpass linear regression owing to their capacity to represent intricate and non-linear interactions. Neural networks outperform linear regression in forecasting the critical temperature of superconductors (Nieto et al., 2024). Neural networks are susceptible to hyperparameter variations and therefore necessitate comprehensive tweaking to attain maximum performance (Nadh et al., 2023).

Random Forests is an ensemble learning method that integrates the outcomes of multiple decision trees to enhance predictive accuracy. This method is acknowledged for its ability to manage non-linear data, outliers, and missing values (Jin et al., 2020; Liaw & Wiener, 2002). This approach has been widely employed in workforce management, encompassing employee turnover modeling and overtime prediction. This strategy demonstrates more accuracy compared to simpler modelss (Biau & Scornet, 2016). In another investigation on predicting soil moisture, random forests demonstrated greater performance with an accuracy of 94%, in contrast to linear regression, which achieved only 92% accuracy (Tynchenko et al., 2024). Furthermore, random forest demonstrated superior accuracy and reduced susceptibility to overfitting compared to artificial neural networks in forecasting concrete strength (Rizvon & Jayakumar, 2022).

METHOD

This research employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (Chapman et al., n.d.). CRISP-DM is a methodical and systematic approach designed to mitigate the risk of failure in data mining operations or research endeavors.

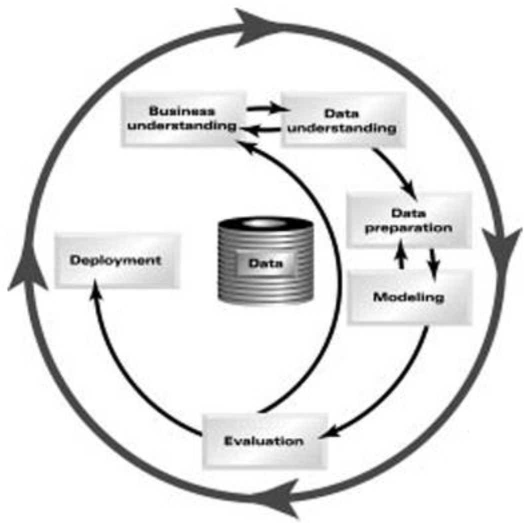

Fig. 2 CRISP-DM Framework
Source: (Ncr et al., n.d.)

This study employs the five stages of CRISP-DM, that is:

1. Business Understanding
   This phase underscores a thorough understanding of corporate operations, and the relevant requirements associated with this research.
2. Data Understanding
   At this stage, the process of identifying essential data, collecting data, assessing data quality, and understanding the structure and relationships within the data is undertaken to determine if the data meets the study needs.
3. Data Preparation
   This phase includes all actions required for the construction of the dataset employed in modeling. This includes data cleansing, feature selection, data transformation, and/or the creation of new features.
4. Modeling
   At this stage, various machine learning (ml) models are developed using the prepared data. The choice of a model depends on the research objectives, interpretive needs, and the characteristics of the issue at hand.
5. Evaluation

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

The final stage involves evaluating the model's performance to ensure it meets the defined research objectives and analytical criteria. Evaluation involves measuring the model's performance by testing it against the data with diverse scoring metrics.

RESULT

Business Understanding

The implementation of overtime in October 2024 created a significant financial burden on Company A. In just two months of post-implementation, overtime costs have accounted for 1.7 billion of the 3.7 billion annual budgets. This expenditure will also affect employee well-being, productivity, and overall corporate performance. Consequently, it is imperative to examine the current overtime data using a predictive model, enabling the organization to forecast overtime hours and formulate methods to alleviate excessive overtime.

Data Understanding

The data derives from the final overtime compensation assigned to employees, generated by the human capital service sector over the two-month period post-implementation, specifically in October and November 2024. The overtime dataset has 294 records for October and 181 entries for November.

Data Preparation

Data cleansing is conducted on the preliminary dataset, which comprises name, employee ID, position, position grade, overtime job name, overtime job type, overtime date, working day clock-in, working day clock-out, total working day overtime hours, and total working day clock-out. The data is streamlined as the foundational input. The data simplification procedure was notably time-consuming due to the inconsistency of the data, particularly with dates and times presented in non-uniform formats. The deletion of data was executed since it was unnecessary for the data processing, including position grade defined in the position data and the name of the overtime job specified in the type of overtime work. Furthermore, data was consolidated for total workday hours and total holiday hours, with dates transformed into a count of days, as the basic data requirement specified the number of overtime days and hours within a month.

Table 1 Dataset

No	number of overtime hours	NIP	Name	Jabatan	Type overtime work	Month **Work overtime**	Summary of Work overtime
1	8	9818038TAY	Iqbal Ramadhan	Fungsional	Preventive	November	1
2	8	9818042TAY	Aswadi Agus	Fungsional	Preventive	November	3
3	8	7292002P3B	Syarifudin D	Fungsional	Preventive	November	2
4	8	9219868ZY	Adi Susilo	Fungsional	Preventive	November	2
5	8	9618001TAY	E. Muisan Jabar	Fungsional	Preventive	November	2
6	8	7493315K3	Januardi Salam	Fungsional	Preventive	November	3
7	8	9319265ZY	Muhammad Alif Dwi	Fungsional	Preventive	November	4
8	8	0022390ZY	Sarin Nurdin	Fungsional	Preventive	November	1
9	8	002239ZY	Fajri Nurman	Fungsional	Preventive	November	1
10	8	7094164K3	Seza Hadi Susilo	Fungsional	Preventive	November	2
11	8	9516003BAY	Diwangga Afandi	Fungsional	Preventive	November	4
12	8	9616007TAY	Arbi Syirif	Fungsional	Preventive	November	2
13	8	9618003TAY	Kevin Dwi Ronaldo	Fungsional	Preventive	November	4
14	8	9920974ZY	Muhammad Majid Hidayatulah	Fungsional	Preventive	November	4
15	8	9820596ZY	Sigit Ismanda	Fungsional	Preventive	November	3
16	8	0022394ZY	Milad Djanuardi	Fungsional	Preventive	November	2
17	8	7393463K3	Muhammad Fikri	Fungsional	Preventive	November	2
18	8	7393786K3	Riki Subagja Danuar	Fungsional	Preventive	November	2
19	8	9818044TAY	Gatot Tri Rubian	Fungsional	Preventive	November	3
20	8	9619622ZY	Wiranto Hamad	Fungsional	Preventive	November	1

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

Modeling

This research employs orange data mining to forecast the number of overtime hours by evaluating three model types for optimal predictive value. The specified features and target data are as follows:

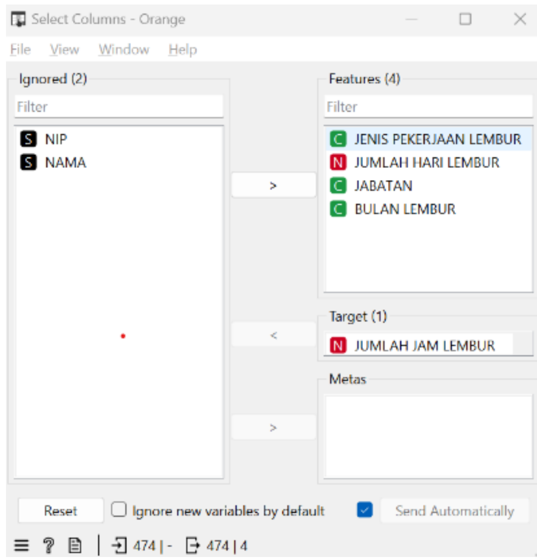

Fig. 3 Attribute Data

Feature data encompasses the kind of overtime work, the quantity of overtime days, the job title, and the month in which the overtime occurred. The objective data pertains to the quantity of overtime hours. Metadata comprises employee ID and name and is excluded from the data to be processed as it is irrelevant to the prediction procedure to be executed. This study requires only training data, as the objective is to identify the optimal predictive model among three options: linear regression, neural networks, and random forests.

Evaluation

Supervised learning methodologies, including linear regression, neural networks, and random forests, are thereafter linked to the predictions widget to assess the processing outcomes of each model in comparison to the original data. This data indicates which model closely approximates the original data. To examine the evaluation findings in greater depth, one can connect the predictions widget, which is linked to the trained model, to the data table.

Deployment

The model with the optimal value is chosen to forecast the number of overtime hours for the subsequent month. This model will furnish firm management with an overview to ascertain future overtime initiatives and procedures.

In this section, the researcher will explain the results of the research obtained. Researchers can also use images, tables, and curves to explain the results of the study. These results should present the raw data or the results after applying the techniques outlined in the methods section. The results are simply results; they do not conclude.

This is a model generated with the application of orange data mining techniques:

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

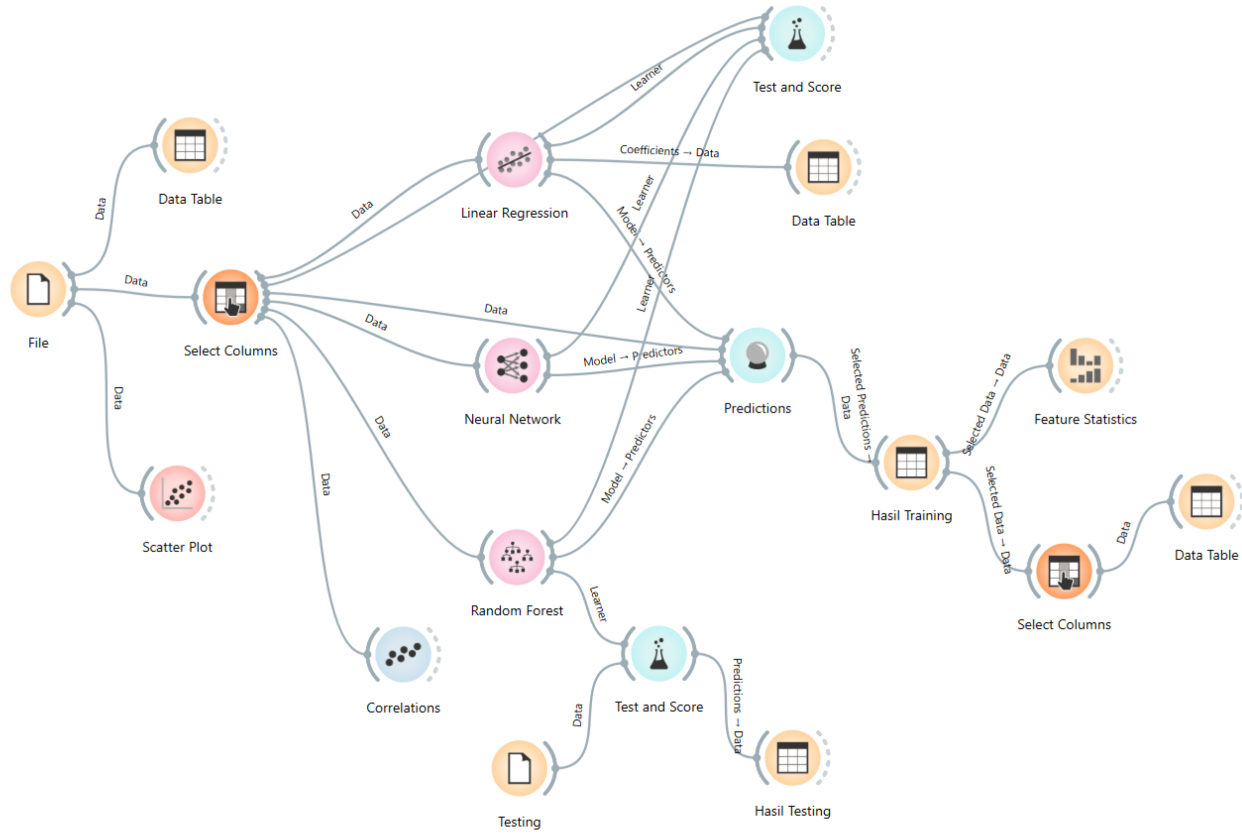

Fig. 4 Design Model

Data that has been preprocessed is subsequently evaluated to identify the optimal predictive model. The testing and evaluation of the orange application using a specific dataset yielded the prediction results of the three specified methods.

Table 2 Model Prediction Results

Model	MSE	RMSE	MAE	MAPE	R2
Neural Network	393.437	19.835	14.287	0.691	0.788
Linear Regression	377.843	19.438	14.283	0.622	0.796
Random Forest	158.78	12.601	8.916	0.336	0.914

The neural network model exhibits the poorest performance relative to the other models, evidenced by the greatest MSE (393.437) and the highest MAE (14.287). The R^2 value of 0.788 signifies that the model accounted for 78.8% of the data's variability, albeit less effectively than the other models. The linear regression model outperformed the neural network somewhat, with a lower MSE of 377,843 and a nearly equivalent MAE of 14,283. An R^2 value of 0.796 indicates that this model explains 79.6% of the variability in the data. The Random Forest model had exceptional performance, producing the lowest error values for all metrics: MSE (158.78), RMSE (12.601), MAE (8.916), and MAPE (0.336). The R^2 value of 0.914 signifies that this model explains 91.4% of the observed variability, demonstrating higher performance compared to alternative models.
Derived from the linear regression method, the comparison of actual and expected overtime hours indicates that the prediction outcomes are respectable, especially with general tendencies or linear patterns in the dataset. Nonetheless, a substantial prediction error exists in the high-variation segment of the data, since the model fails to grasp intricate relationships within the data.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

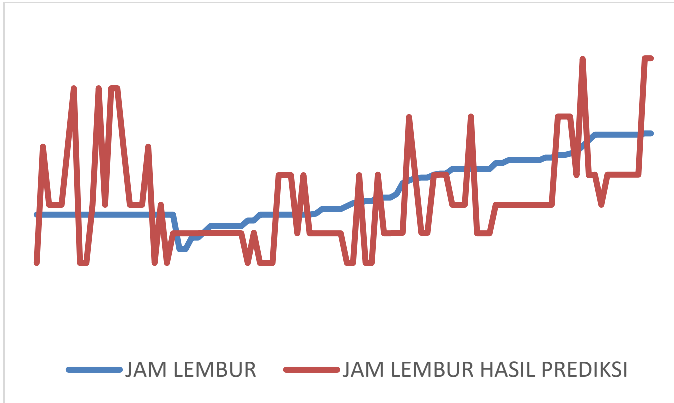

Fig. 5 Prediction Results with Linear Regression

The Neural Network model shows the main trend of the dataset well, especially in parts where the data has a consistent and straight-line pattern. In these areas, the model makes predictions that are close to the actual values, showing that the neural network can recognize and combine the main patterns from the input data. However, when looking at data segments that show a lot of ups and downs—like sudden spikes or quick drops—the model's ability to predict accurately goes down a lot. The difference in performance shows that the model doesn't really pick up on small changes or unusual patterns in the data. The expected values in some areas are different from the actual values, which causes a bigger prediction error.

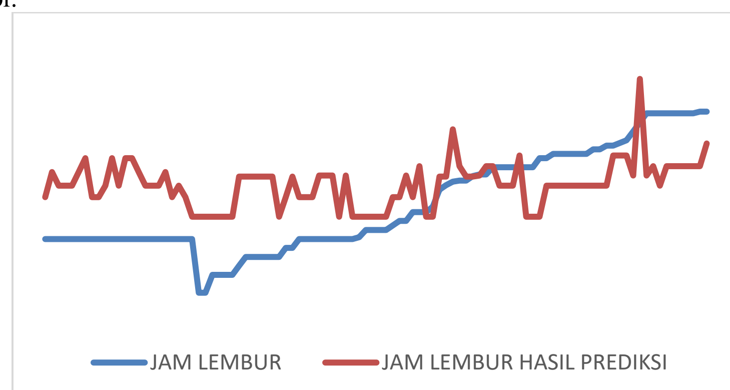

Fig.6 Prediction Results with Neural Network

Figure 7 depicts the Random Forest model's accurate representation of the complex interaction between input and target variables. In several segments of the graph, the projected values closely correspond with the actual values, indicating exceptional accuracy. In datasets characterized by substantial oscillations (notable peaks or troughs), Random Forest outperforms linear algorithms like Linear Regression in detecting oscillatory patterns. Nevertheless, there were occasions when the predictions were notably off, particularly during substantial peaks.

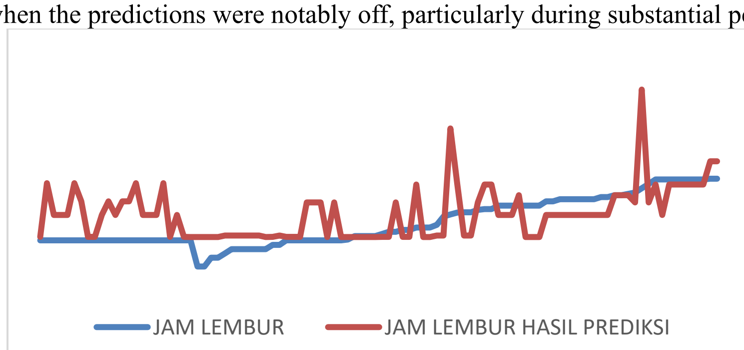

Fig.7 Prediction Results with Random Forest

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 2, April 2025
https://doi.org/10.47709/cnahpc.v7i2.5678

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

In comparisons among the three models, Random Forest regularly outperforms the others in both quantitative metrics and visual forecasts. The Random Forest model accurately reflects the actual trend of additional hours, particularly in regions characterized by significant variability. Conversely, Neural Network and Linear Regression models frequently underperform in variable data environments, struggling to identify distinct peaks or troughs. Their elevated error rates (MAE ~14.28) and low R^2 scores (0.788 and 0.796, respectively) demonstrate this. The Random Forest model exhibits the lowest error rates (MSE: 158.78, MAE: 8.916) and the greatest R^2 value of 0.914, indicating it is the most accurate and dependable model for predicting employee overtime compared to the alternatives.
.

DISCUSSIONS

This study utilized overtime prediction modeling through three distinct algorithms—Linear Regression, Neural Network, and Random Forest—to determine the most effective model for forecasting employee overtime based on historical data. The investigation's findings reveal that the Neural Network (78.8% accuracy) and Linear Regression (79.6% accuracy) models demonstrate subpar prediction performance for overtime relative to the Random Forest model (91.4% accuracy). This study's practical implications indicate that an effective predictive model can facilitate the early detection of potential workload surpluses for budgetary efficiency and evaluate job allocation to improve employee work-life balance.

This computational approach is thus recommended for next overtime hour estimates. Utilizing precise predictions from the Random Forest model, organizations may optimize human resource allocation, ensuring equitable work distribution and minimizing excessive overtime. This allows management to develop more precise budgeting tactics, minimizing unnecessary labor costs while maintaining productivity. Moreover, deliberate strategies may be utilized to reduce excessive reliance on overtime, such as adjusting shift schedules, recruiting additional staff as necessary, or improving workload forecasting tools. These strategies ultimately promote a more sustainable and employee-focused work environment while improving operational efficiency.

Despite the above-described research findings, the utilized dataset presents specific constraints. The dataset insufficiently includes all possible elements affecting overtime, such as demographics, performance, and others. This study is based on historical data from a particular firm, hence limiting the model's application across several industries. Future research could augment this work by creating a dataset that includes a wider array of factors affecting overtime and by extending the historical data, as well as utilizing different algorithms for more varied and adaptable forecasts.

CONCLUSION

In comparisons among the three models—Neural Network, Linear Regression, and Random Forest—Random Forest regularly outperforms the others in both quantitative metrics and visual forecasts. The Random Forest model accurately reflects the actual trend of additional hours, particularly in regions characterized by significant variability. This indicates its capability to identify non-linear patterns and manage intricate data dynamics. Conversely, Neural Network and Linear Regression models frequently underperform in variable data environments, struggling to identify distinct peaks or troughs. Their elevated error rates (MAE ~14.28) and low R^2 scores (0.788 and 0.796, respectively) demonstrate this. The Random Forest model exhibits the lowest error rates (MSE: 158.78, MAE: 8.916) and the greatest R^2 value of 0.914, indicating it is the most accurate and dependable model for predicting employee overtime compared to the alternatives.

REFERENCES

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Chapman, P., Kerber, R., Clinton, J., Khabaza, T., Reinartz, T., & Wirth, R. (n.d.). *The CRISP-DM Process Model*.

Crocker, D. C., & Seber, G. A. F. (1980). Linear Regression Analysis. In *Technometrics* (Vol. 22, Issue 1). https://doi.org/10.2307/1268395

Gede Surya Mahendra, Lely Priska D. Tampubolon, M. MSI Herlinah, M. S. S. A., & Lalu Puji Indra Kharisma, Mochzen Gito Resmi, M.Kom I Gede Iwan Sudipa, Khairunnisa, Anak Agung Gede Bagus Ariana, Syahriani Syam, E. (2023). Sistem Pendukung keputusan Teori dan Penerapannya dalam berbagai metode. In *Jurnal Ilmu Pendidikan* (Vol. 7, Issue 2).

Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes*

* Corresponding author

Submitted : Mar 12, 2025
Accepted : Mar 25, 2025
Published : Apr 04, 2025

in Artificial Intelligence and Lecture Notes in Bioinformatics), *12343 LNCS*, 503–515. https://doi.org/10.1007/978-3-030-62008-0_35

Kim, H. (2022). Deep Learning. *Artificial Intelligence for 6G*, *22*(4), 247–303. https://doi.org/10.1007/978-3-030-95041-5_6

Liaw, A., & Wiener, M. (2002). The R Journal: Classification and regression by randomForest. *R Journal*, *2*(3), 18–22. http://www.stat.berkeley.edu/

Manwal, S., Bhandari, A. S., Narang, H., Vats, S., Sharma, V., & Yadav, S. P. (2024). Temperature Prediction: A Comparative Comprehensive Study Between Machine Learning Algorithms. *1st International Conference on Electronics, Computing, Communication and Control Technology, ICECCC 2024*. https://doi.org/10.1109/ICECCC61767.2024.10593978

Mathieu, M., Couprie, C., & LeCun, Y. (2016). Deep multi-scale video prediction beyond mean square error. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, *2015*, 1–14.

Nadh, B. S. S., Sha, A., Addanki, V., Yerramreddy, D. R., & Nisha, K. S. (2023). Performance Analysis of Machine Learning-Based Scheduling Policies for Operating Systems. *7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings*, 771–776. https://doi.org/10.1109/ICECA58529.2023.10395782

Ncr, P. C., Spss, J. C., Ncr, R. K., Spss, T. K., Daimlerchrysler, T. R., Spss, C. S., & Daimlerchrysler, R. W. (n.d.). *Step-by-step data mining guide*.

Nieto, P. J. G., Gonzalo, E. G., García, L. A. M., Prado, L. Á.-D., & Sánchez, A. B. (2024). Predicting the critical superconducting temperature using the random forest, MLP neural network, M5 model tree and multivariate linear regression. *Alexandria Engineering Journal*, *86*, 144–156. https://doi.org/10.1016/j.aej.2023.11.034

Rizvon, S. S., & Jayakumar, K. (2022). Strength prediction models for recycled aggregate concrete using Random Forests, ANN and LASSO. *Journal of Building Pathology and Rehabilitation*, *7*(1). https://doi.org/10.1007/s41024-021-00145-y

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Teng, J., & Yuan, Y. (2020). *Inject Machine Learning into Significance Test for Misspecified Linear Models*. http://arxiv.org/abs/2006.03167

Tynchenko, V., Kukartseva, O., Degtyareva, K., Khrapunova, Y., & Anisimov, P. (2024). Application of machine learning methods to predict soil moisture based on meteorological and atmospheric data. *BIO Web of Conferences*, *130*. https://doi.org/10.1051/bioconf/202413002003

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

* Corresponding author