GARMENT EMPLOYEE PRODUCTIVITY PREDICTION USING RANDOM FOREST

Imanuel Balla [1]; Sri Rahayu [2*)] ; Jajang Jaya Purnama [3]

[1)]Ilmu Komputer; [2)]Informatika; [3)]Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
www.nusamandiri.ac.id
14002208@nusamandiri.ac.id [1]; srirahayu.rry@nusamandiri.ac.id [2*)]; jajang.jjp@nusamandiri.ac.id [3]
(*) Corresponding Author

Abstract— Clothing which also means clothing is needed by humans. Besides the need for clothing in terms of function, clothing sales or business is also very potent. About 75 million people worldwide are directly involved in textiles, clothing, and footwear. In this case, a common problem in this industry is that the actual productivity of apparel employees sometimes fails to reach the productivity targets set by the authorities to meet production targets on time, resulting in huge losses. Experiments were conducted using the random forest model, linear regression, and neural network by looking for the values of the correlation coefficient, MAE, and RMSE. This aims to predict the productivity of garment employees with data mining techniques that apply machine learning and look for the minimum MAE value. The results of testing the proposed algorithm on the garment worker productivity dataset obtained the smallest MAE, namely the random forest algorithm, namely 0.0787, linear regression 0.1081, and 0.1218 neural networks.

Keyword: Employee Productivity, Garment, Random Forest.

Abstrak— *Sandang juga berarti pakaian adalah apa yang dibutuhkan manusia. Disamping kebutuhan pakaian dari segi fungsinya, penjualan atau bisnis pakaian juga sangat berpotensi. Sekitar 75 juta orang terlibat langsung dalam tekstil, sektor pakaian, dan alas kaki di seluruh dunia. Masalah umum industri dalam hal ini adalah produktifitas aktual dari karyawan garmen terkadang tidak memenuhi produktifitas yang ditargetkan yang telah ditetapkan oleh otoritas untuk memenuhi tujuan produksi pada waktunya sehingga menimbulkan kerugian yang sangat besar. Eksperimen yang dilakukan menggunakan model random forest, regresi linier dan neural network dengan mencari nilai dari correlation coefficient, MAE, dan RMSE. Hal ini bertujuan untuk melakukan prediksi produktifitas karyawan garmen dengan teknik data mining yang menerapkan machine learning di dalamnya dan mencari nilai MAE yang paling minimal. Hasil dari pengujian algoritma yang diusulkan pada dataset garment worker productivity didapatkan MAE terkecil yaitu algoritma random forest yaitu 0.0787, regresi linier 0.1081 dan neural network 0.1218.*

Kata Kunci*: Garmen, Produktifitas Karyawan, Random Forest.*

INTRODUCTION

The words clothing, food, and shelter are basic human needs that are familiar to the ears. As we all know, the demand for clothing also means that clothing is what humans need (Mubarok, 2017). It is inconceivable if these needs were not met. Besides the need for clothing in terms of function, clothing sales or business is also very potent. Apart from an economic standpoint, the garment industry is one of the most labor-intensive industries in the world. About 75 million people are directly involved in the textile, clothing, and footwear sectors worldwide (United, 2021).

The ready-to-wear clothing industry is a major part of manufacturing production, employment, and trade in many developing countries, for example, Bangladesh, which is now the second-largest exporter of apparel in the world after China (Chaerani, 2018). According to the recently released Bureau of Export Promotion Data, Bangladesh's export revenue from the ready-to-wear sector is around $ 30.61 billion and ready-to-wear clothing holds nearly 14.07% of Bangladesh's GDP as well as 81% of total export revenue. (M Saiful Islam et al., 2019).

With the increasing demand for clothing needs around the world, the increase in the quality of production in the garment industry must be maintained and improved. One of the business performance measurement tools used is productivity, where the definition of productivity itself is a comparison between output and all sources used (input) (Sri & Margareta, 2020). Technology is a tool used to accelerate productivity (Afani, Utari Nur. dan Solovida, 2019). With better employee productivity with employee productivity

getting better, the quality and quantity of the product produced will not be doubted every time it is produced.

A common problem in the industry in this regard is that the actual productivity of garment employees sometimes does not meet the targeted productivity that has been set by the authorities to meet production objectives in time, resulting in enormous losses. Before increasing the productivity of employee performance, it is necessary to know in advance what factors affect and how to predict employee productivity, especially garment employees that are being discussed. Several previous studies that support the discussion of employee work productivity are outlined in table 1 below :

Table 1. *Research Literature*

Author	*Research Problem*	*Literatur Supports*
Abu Hamja, Malek Maalouf & Peter Hasle (Hamja et al., 2019)	Lean is increasingly being used by garment manufacturers to increase productivity to reduce costs and waiting times. However, it has not been documented in the review whether lean increases productivity, and it is an open question whether lean can increase productivity without compromising health and safety work.	The solution is to explore and collect a systematic review of the available review literature Research on Lean in Clothing and Its Associated Effects on Productivity and Safety.
Lila Ayu Ratna Winanda (Winanda, 2010)	The problems of construction workers cannot be separated from the resulting productivity. Therefore, this article discusses the methods of factors affecting the productivity of construction workers, so that in the end we can estimate the productivity of construction workers themselves.	The solution is to take a probabilistic approach to the neural networks used for mapping the productivity and productivity factors of the workers themselves.
Doni Efriza & Iswandi Idris (Doni Efriza, 2018)	Measurement of employee productivity is used as a management tool to analyze and encourage efficiency so that increased productivity will provide a greater ability for companies to increase employee wages, which in turn will stimulate employee morale and morale.	The solution is to distribute questionnaires to bank employees interviewed in Medan City to see the variables of motivation, knowledge, skills, and income levels of the factors that affect work efficiency.

Source : (Balla et al., 2021)

In addition to the cases described in the previous paper on solving productivity problems, a researcher from Indonesia, (Gunawan et al., 2010) has identified that increased managerial ability to monitor finances and evaluate activities in medium-scale Indonesian garment factories is essential to sustain the Indonesian economy.

However, it is not enough to increase managerial abilities, because of the limitations of each person in the managerial ranks itself. Need help with data mining processing or data mining that can make predictions in this case. As has been done by (Imran et al., 2019) in their research which aims to solve the productivity gap problem by predicting the actual productivity of employees who are currently appointed as the main reference paper, presented in table 2 below.

Table 2. Main Reference Paper

Title	Measurement results
Deep Neural Network Approach for Predicting the Productivity of Garment Employees (Imran et al., 2019)	MSE = 0.086
	MAE = 0.018
	MAPE = 15.932

Source : (Imran et al., 2019)

By using the Deep Neural Network (DNN) model, the experimental results of this study have shown that the proposed model produces promising predictive performance with a minimum Mean Absolute Error of 0.086 which is less than the basic performance error of 0.15. Such predictive performance can help producers to set accurate targets, minimize production losses and maximize

profits. By using the same dataset, in this study, we will discuss experiments using the random forest model, linear regression, and neural network by looking for the values of the correlation coefficient, MAE, and RMSE. It aims to predict the productivity of garment employees using data mining techniques that apply machine learning and look for the minimum MAE value.

MATERIALS AND METHODS

In a study, of course, the main material is a dataset to be used as machine learning material using algorithms. In this study, the dataset used is garments worker productivity, which is a public dataset because it is taken from the UCI repository website.

The dataset used in this study was published in 2020 with 15 attributes including date, day, quarter, department, team_no, no_of_workers, no_of_style_change, targeted_productivity, SMV, wip, over_time, incentive, idle_time, idle_men, actual_productivity with continuous actual_productivity classes. , has 1197 instances. In table 3, we can see from the specification of the garments worker productivity dataset, which has 15 attributes and 1197 data.

Table 3. Dataset Specifications

Dataset Name	Number of Attributes	Amount of data
garments worker productivity	15	1197

Source : (Imran et al., 2019)

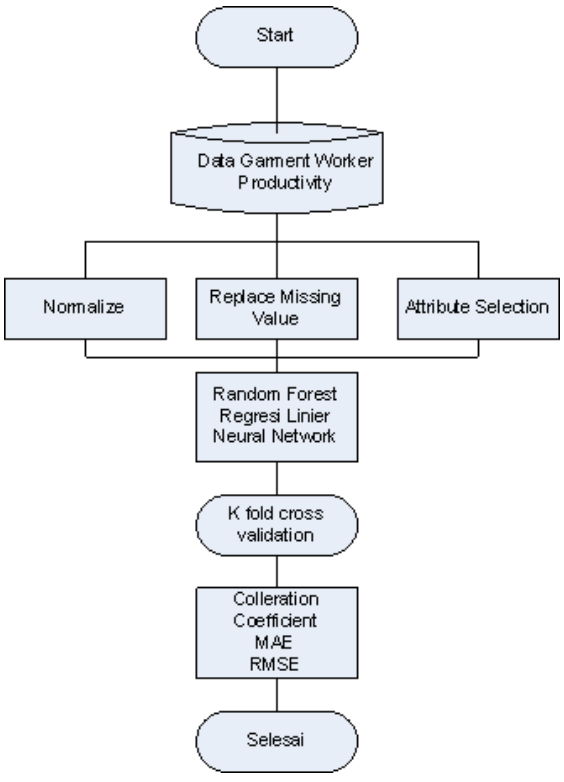
With a description of each attribute is described in table 4 as follows :

Table 4. Dataset Description

No	Attribute	Description
1	*Date*	The date is in MM-DD-YYYY format
2	*Day*	Days of the week
3	*Quarter*	Part of this month. One month is divided into four parts
4	*Department*	The department is associated with the instance
5	*team_no*	The team number associated with the instance
6	*no_of_workers*	The number of workers on each team
7	*no_of_style_change*	The number of changes to a specific product style
8	*targeted_productivity*	The targeted productivity is set by the Authority for each team for each day
9	*SMV*	Standard Minute Value, this is the time allocated for a task
10	*wip*	Work in Progress includes the number of unfinished items for the product
11	*over_time*	Represents the amount of overtime by each team in minutes
12	*incentive*	Represents the number of financial incentives (in the UDB) that enable or motivate certain actions
13	*idle_time*	The length of time the product has stalled for several reasons
14	*idle_men*	The number of unemployed workers due to production disruptions
15	*actual_productivity*	The actual percentage of productivity generated by workers. It ranges from 0-1

Source : (Imran, 2020)

From table 4, it can be seen each information on the attributes in the dataset, this dataset is then carried out by machine learning using the random forest algorithm for prediction. The stages of the research carried out are described in the following chart :



Source : (Balla et al., 2021)

Figure 1. Research Method

Figure 1, describes the research carried out on the garment worker productivity dataset, namely before the algorithm is applied, preprocessing is carried out first such as the normalize technique, the replace missing value technique and the attribute selection technique.

```
In [21]: peekup  =df.head(5)
         print(peekup)

         date   quarter  department       day  team  targeted_productivity  \
0  1/1/2015  Quarter1      sweing  Thursday     8                   0.80
1  1/1/2015  Quarter1  finishing   Thursday     1                   0.75
2  1/1/2015  Quarter1      sweing  Thursday    11                   0.80
3  1/1/2015  Quarter1      sweing  Thursday    12                   0.80
4  1/1/2015  Quarter1      sweing  Thursday     6                   0.80

      smv          wip  over_time  incentive  idle_time  idle_men  \
0  26.16  1108.000000       7080         98        0.0         0
1   3.94  1190.465991        960          0        0.0         0
2  11.41   968.000000       3660         50        0.0         0
3  11.41   968.000000       3660         50        0.0         0
4  25.90  1170.000000       1920         50        0.0         0

   no_of_style_change  no_of_workers  actual_productivity
0                   0           59.0             0.940725
1                   0            8.0             0.886500
2                   0           30.5             0.800570
3                   0           30.5             0.800570
4                   0           56.0             0.800382
```

Source : (Balla et al., 2021)

Figure 2. Sample Data

Figure 2.shows the initial data sample from the garments worker productivity dataset of five records from each of the attributes in the dataset using the Jupyter notebook tool (python 3.7.7) for normalization.

Normalization is carried out in the hope of equalizing the frequency value from 0 to 1, replacing missing values is also done to change the blank data to the average data of the attributes, and attribute selection is carried out for attribute dates because it is represented by attribute day.

After normalizing, replacing the missing value and attribute selection, the next step is machine learning using the random forest algorithm (Li et al., 2020), linear regression (Asohi & Andri, 2020), and neural network (Zhou et al., 2020). The three algorithms were chosen because they were seen from the characteristics of the dataset which were supervised learning with a continuous label. The results obtained a pattern from machine learning that predicts the productivity of garment employees. Furthermore, the deployment stage can be utilized by implementing the pattern into a machine learning application. In the following, the data is presented to check the blank data :

```
In [3]: import numpy as np
        import pandas as pd

In [4]: df = pd.read_csv("garments_worker_productivity.csv")
        # Check for Null Data
        df.isnull().sum()

Out[4]: date                       0
        quarter                    0
        department                 0
        day                        0
        team                       0
        targeted_productivity      0
        smv                        0
        wip                      506
        over_time                  0
        incentive                  0
        idle_time                  0
        idle_men                   0
        no_of_style_change         0
        no_of_workers              0
        actual_productivity        0
        dtype: int64
```

Source : (Balla et al., 2021)

Figure 3. *Check for Null Data*

Figure 3.explains checking all attributes for null or empty data using the Jupyter notebook tool (python 3.7.7), 506 empty data is found in the WIP attribute, as many as 506 null or blank data, after it is known that there is null data, do replace the missing values.

```
In [6]: #Replace missing using median
        mean=df['wip'].mean()
        df['wip'].fillna(mean, inplace=True)
```

Source : (Balla et al., 2021)

Figure 4. *Replace missing using median*

Figure 5 shows the results of replacing missing values using the median of the WIP attribute, and these results are automatically generated in the garments worker productivity dataset, there are no missing values anymore.

```
In [7]: df.isna().sum() # Sums all the missing values

Out[7]: date                       0
        quarter                    0
        department                 0
        day                        0
        team                       0
        targeted_productivity      0
        smv                        0
        wip                        0
        over_time                  0
        incentive                  0
        idle_time                  0
        idle_men                   0
        no_of_style_change         0
        no_of_workers              0
        actual_productivity        0
        dtype: int64
```

Source : (Balla et al., 2021)

Figure 5. The Result of *Replace Missing Values*

The platform used in this study uses the weka 3.8 application with the following personal computer device specifications :

Table 5. Device Specifications

Processor	intel intel® core™ i7-8565U
CPU	1.80 GHz 1.99 GHz
RAM	8.00 GB

Source : (Balla et al., 2021)

This machine learning uses the k fold cross-validation technique where 10 repetitions are carried out, namely 1 time testing 9 times training repeatedly 10 times alternating subsets, the results are seen from the correlation coefficient, MAE, RMSE, the more MAE is closer to 0, the less error level.

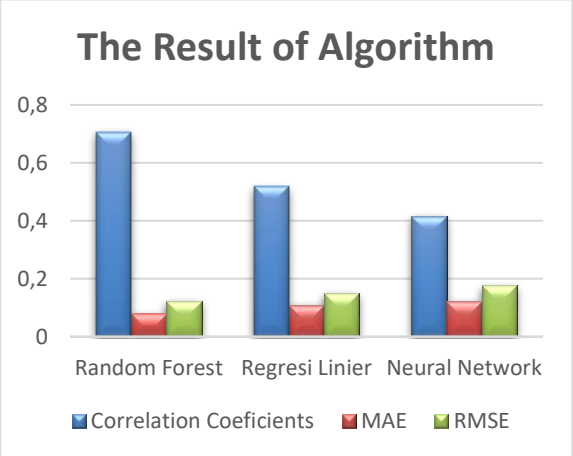
RESULTS AND DISCUSSION

The pattern that is formed results from the results of preprocessing, namely normalize, replace the missing value and attribute selection as well as algorithmic testing using random forest, linear regression, and neural network. The results of each algorithm test can be seen in Table 3 below.

Table 6. Algorithm Testing Results

Algorithm	*Correlation Coeficients*	MAE	RMSE
Random Forest	0.7071	0.0787	0.1236
Regresi Linier	0.5173	0.1081	0.1494
Neural Network	0.4169	0.1218	0.1763

Source : (Balla et al., 2021)



Source : (Balla et al., 2021)

Figure 6. Algorithm Results

From table 6, it can be seen that the results of testing the random forest algorithm, linear regression, and neural network on the garment worker productivity dataset obtained the smallest MAE, namely the random forest algorithm, namely 0.0787, linear regression 0.1081, and 0.1218 neural networks.

CONCLUSION

Fulfilling the need for clothing that is utilized based on its function also has economic potential in the industrial world. Quality garment products cannot be separated from the performance of employees who meet the targeted productivity set by the authorities to meet production goals on time. In predicting the productivity of garment employees, technology is needed. Currently, data mining with the application of machine learning is a solution. This research was conducted using random forest algorithm testing, linear regression, and neural network on the 2020 garments worker productivity dataset which consists of 15 attributes with 1 class, namely actual_productivity which is continuous. This study produces the correlation coefficient, MAE, and RMSE values for each of the models applied. Because the aim is to find the smallest MAE value, the random forest model, in this case, is most appropriate, each MAE value obtained is the random forest algorithm of 0.0787, linear regression of 0.1081, and the neural network of 0.1218. Further research is suggested to develop by making deployment applications to implement the results of the resulting pattern.

REFERENCE

Afani, Utari Nur. dan Solovida, G. T. (2019). LINGKUNGAN (Studi Kasus UMKM Rumah Pemotongan Hewan di Semarang). *Jurnal Sustainable Competitive Advantage*, *9*(51), 51–59.

Asohi, Y., & Andri, A. (2020). Impelementasi Algoritma Regresi Linier Berganda Untuk Prediksi Penjualan. *Jurnal Nasional Ilmu Komputer*, *1*(3), 149–158. https://doi.org/10.47747/jurnalnik.v1i3.161

Balla, I., Rahayu, S., & Purnama, J. J. (2021). Prediksi Produktivitas Karyawan Garmen Menggunakan Random Forest. *Jurnal TECHNO Nusa Mandiri*, *1*, 1–6.

Chaerani, N. (2018). Peran International Labour Organization Terhadap Peningkatan Lingkungan Kerja Di Sektor Industri Garmen Di Bangladesh. *Universitas Hasanuddin*, *151*(2), 10–17.

Doni Efriza, I. I. (2018). Produktivitas kerja karyawan perbankan di kota medan. *Jurnal BIS-A*, *05*(02), 49–53. http://ejurnal.plm.ac.id/index.php/BIS-A/article/view/164/145

Gunawan, A., Wahdan, M., & van den Herik, H. J. (2010). Increasing the managerial capabilities in Indonesian garment manufacturing. *International Journal of Economic Policy in Emerging Economies*, *3*(4), 346–367. https://doi.org/10.1504/IJEPEE.2010.037582

Hamja, A., Maalouf, M., & Hasle, P. (2019). The effect of lean on occupational health and safety and productivity in the garment industry–a literature review. *Production and Manufacturing Research*, *7*(1), 316–334. https://doi.org/10.1080/21693277.2019.1620652

Imran, A. Al. (2020). *Productivity Prediction of Garment Employees Data Set*. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees

Imran, A. Al, Amin, M. N., Islam Rifat, M. R., & Mehreen, S. (2019). Deep neural network approach for predicting the productivity of garment employees. *2019 6th International Conference on Control, Decision and Information Technologies, CoDIT 2019*, 1402–1407. https://doi.org/10.1109/CoDIT.2019.8820486

Li, J., Tian, Y., Zhu, Y., Zhou, T., Li, J., Ding, K., & Li, J. (2020). A multicenter random forest model for effective prognosis prediction in collaborative clinical research network. *Artificial Intelligence in Medicine*, *103*(September 2019), 101814. https://doi.org/10.1016/j.artmed.2020.101814

M Saiful Islam, Rakib, M. A., & Adnan, A. (2019). Ready-Made Garments Sector of Bangladesh: Its Growth, Contribution, and Challenges. *Economics World*, *7*(1). https://doi.org/10.17265/2328-7144/2019.01.004

Mubarok, N. (2017). Strategi Pemasaran Islami Dalam Meningkatkan Penjualan Pada Butik Calista. *I-Economics*, *3*(1), 73–92.

Sri, D., & Margareta, C. (2020). Pengaruh Pelatihan Kewirausahaan , Kemampuan Memanfaatkan Teknologi Dan Pendidikan Terhadap Produktifitas Wanita. *Economic and Education Journal*, *42*, 142–158.

United, F. (2021). *Global fashion industry statistics - International apparel*. Https://Fashionunited.Com/. https://fashionunited.com/global-fashion-industry-statistics/

Winanda, L. (2010). Estimasi Produktivitas Pekerja Konstruksi Dengan Probabilistic Neural Network. *Spectra*, *8*(15), 40–50.

Zhou, G., Moayedi, H., Bahiraei, M., & Lyu, Z. (2020). Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. *Journal of Cleaner Production*, *254*, 120082. https://doi.org/10.1016/j.jclepro.2020.120082