Syntax Literate: Jurnal Ilmiah Indonesia p–ISSN: 2541-0849
e-ISSN: 2548-1398
Vol. 9, No. 3, Maret 2024

PREDICTIVE MAINTENANCE USING LINEAR REGRESSION

Benedict Ariel Kurnianto, Rudy Hartono Prayogo*, Nidia Pialina Nababan, Suharjito
Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering, Bina Nusantara University, Indonesia
Email: rudy.prayogo@binus.ac.id*

Abstract

Problems regarding machine damage often occur in many industries, especially the manufacturing industry, which causes large losses for companies. This is of course influenced by various factors such as engine temperatures that are too high, engine rotation that is too fast, poor engine torque values, and so on. This research aims to provide predictive analysis results regarding engine conditions that have the potential to experience damage. To achieve this goal, this research will carry out predictive maintenance analysis using a linear regression analysis approach in which two linear regression models will be carried out where the first model involves PCA preprocessing and the second model is carried out without PCA. This research will use the predictive maintenance dataset from the conference (Matzka, 2020). It is known that the MSE, RMSE, MAE, and R2 values of the two methods have the same values, namely 0.909, 0.953, 0.806, and 0.772 respectively. Based on this research, it is concluded that whether PCA is performed or not, it does not significantly affect the results of the regression analysis. This outcome can be attributed to the artificial nature of the dataset, rendering it ideal. Moreover, the retained PCA value of 98% is close to the number of attributes in the original dataset.

Keywords: Data Science, Predictive Maintenance, Manufacturing Industry, Linear Regression, Principal Component Analysis, Orange Data Mining Software

Introduction

The advancement of technology allows us to easily acquire large volumes of data, often referred to as big data. This is undoubtedly positive news for companies, as they can leverage this data to enhance their performance. To harness the power of this data, a more in-depth analysis is necessary, and this is where data science comes into play. According to (Aalst, 2016), data science is an interdisciplinary field with the goal of turning data into tangible value. Data science simplifies the analysis of big data, ensuring accurate results in line with the acquired data (Aalst, 2016).

Data science analysis is commonly applied across various sectors to examine the data at hand. One crucial sector where data science plays a significant role is manufacturing, particularly in the context of predictive maintenance. The manufacturing industry often faces issues stemming from various factors. These issues are critical to address because equipment breakdowns can result in reduced production efficiency, ultimately leading to unfulfilled customer demands (Pranoto, Matondang, & Siregar, 2013). This is why predictive maintenance is of great importance, as it helps minimize equipment failures, thus preventing production stoppages (Pranoto, Matondang, & Siregar, 2013).

In response to these challenges, this research focuses on conducting predictive

How to cite:	Kurnianto, et al. (2024) Predictive Maintenance Using Linear Regression. *Syntax Literate. (9)3.* http://dx.doi.org/10.36418/syntax-literate.v9i3
E-ISSN:	2548-1398

maintenance analysis using regression analysis methods. Additionally, two approaches will be implemented: one involving Principal Component Analysis (PCA) as an initial step and the other without Principal Component Analysis (PCA). This research is significant, as it provides an opportunity for the academic community to contribute new knowledge across various disciplines. It serves as a valuable source of insight into the application of cutting- edge technologies like machine learning and data processing algorithms in data science. Furthermore, it offers the broader public a new technique for addressing their machine maintenance issues, thus reducing potential losses.

Predictive Maintenance

Many studies on maintenance-related issues have been conducted in recent decades (Selcuk, 2017). According to Heizer and Render, the definition of maintenance includes all activities involved in keeping the equipment system in line with the work order, with maintenance activities divided into two types: preventive maintenance and corrective maintenance. Corrective maintenance, as defined by Hendarsin, is maintenance performed only when a component/system fails. On the other hand, according to (Ebeling, 2010), preventive maintenance is scheduled maintenance activities, generally performed periodically, involving activities like inspection, repair, replacement, cleaning, lubrication, adjustment, and alignment.

In addition to these two types of maintenance, predictive maintenance, or prediction-based maintenance, is receiving increasing attention in research related to data acquisition, infrastructure, storage, distribution, security, and intelligence (Zonta, et al., 2020). Predictive maintenance, as known, is a maintenance strategy that leverages technology, data analysis, and an understanding of equipment performance to predict when equipment is likely to fail or experience issues. Its goal is to identify potential issues or failures before they occur so that corrective or maintenance actions can be taken promptly (Senanayaka, et al., 2022). Predictive maintenance is considered to be a more efficient strategy compared to other maintenance approaches like preventive or corrective maintenance because it is only performed when necessary. Over time, research in predictive maintenance has advanced due to its ability to generate predictions based on equipment performance or condition, which is crucial for the future of the industry (Wu, Jennings, Terpenny, & Kumara, 2016).

The key requirement for effective predictive maintenance is to have an adequate amount of data from every part of the manufacturing process. Properly conducted analysis can reduce maintenance costs, lead times, and improve productivity and quality (Kiangala & Wang, 2018).

Regression Analysis

Regression analysis and predictive maintenance are two complementary concepts in the world of industry and maintenance management. A statistical technique called regression analysis is used to determine how one or more independent variables, or predictors, relate to a dependent variable. In the context of predictive maintenance, regression analysis can assist in predicting the remaining lifespan of equipment based on certain factors (Montgomery, Peck, & Vining, 2021). On the other hand, predictive maintenance is an approach that enables organizations to predict when equipment is likely to fail, allowing maintenance to be performed before failure occurs (Mobley, 2002).

Regression analysis is a statistical technique used to understand the relationship between independent variables (predictors) and a dependent variable (response). It allows measuring to what extent changes in one variable will affect another variable.

Regression is a widely used tool in various disciplines, including social sciences, economics, natural sciences, and business. One crucial concept in regression analysis is the regression coefficient. This coefficient measures how much the independent variable influences the dependent variable. In a simple linear regression equation ($Y = a + bX + \varepsilon$), "b" is the regression coefficient. This coefficient describes how changes in the independent variable (X) will affect changes in the dependent variable (Y). But in order to apply regression analysis effectively, certain presumptions must be met: the independent and dependent variables must have a linear relationship; changes in the independent variable must affect the dependent variable proportionately; homoscedasticity—that is, the variance of the random errors must remain constant along the regression line—must be assumed; and, lastly, the residuals, or errors, are supposed to have a normal distribution.

In the context of predictive maintenance, regression analysis can play a vital role. Regression analysis can be used to model the relationship between monitoring variables (such as vibration or temperature) and equipment failure trends. By analyzing historical data, regression can be used to understand how monitoring variables correlate with equipment condition changes. For example, regression analysis can be used to identify the monitoring variables most strongly correlated with equipment failures. This can help build a stronger predictive model for forecasting the remaining lifespan of equipment. By knowing the relationship between monitoring variables and failures, early signs of damage can be identified, and preventive maintenance can be performed before more severe damage occurs. Moreover, regression analysis can also be used in evaluating the impact of maintenance that has been conducted. For instance, regression analysis can be used to understand whether the maintenance performed has resulted in a significant improvement in equipment performance. This can help assess the effectiveness of predictive maintenance programs.

Principal Component Analysis (PCA)

A statistical technique called principal component analysis (PCA) uses advanced mathematical concepts to break down a group of possibly correlated variables into a smaller set of variables called principal components. The goal of PCA is to handle high-dimensional data by reducing the complexity of the data (Richardson, 2009). There are several advantages to working with a smaller set of data rather than the original high-dimensional data. According to (Kherif & Latypova, 2020), these advantages include:

1. The ability to visualize data in 2D or 3D
2. Reduced storage space
3. Elimination of collinearity
4. Reduction of noise

New components are produced by PCA analysis. The projection of the original data onto the first main axis provides the first principal component, which captures the majority of the variation in the data. The majority of the variation in the data that the first principal component was unable to explain is explained by the second principal component, which is the projection of the original data onto the second principal axis. Until the whole data matrix is deconstructed, each succeeding principle component explains the majority of the variance under the restriction that it is orthogonal to the preceding principal components.

Creating disease models that satisfy these requirements can be greatly aided by PCA. Furthermore, PCA has the distinct benefit of offering a condensed representation of data that captures the essential elements of individual variations in the age of Big Data and growing interest in customized treatment.

Review of Previous Research

This research will also conduct a review of several previous studies where 8 studies were obtained which are shown in Table 1 along with the author, title, method and main results.

Table 1. Review of previous research

Author	Method	Results
(Chazhoor, et.al, 2020)	Using various classification models.	Logistic Regression is the suitable model for the dataset.
(Susto, et.al, 2015)	Employing multiple classifiers such as SVM and KNN for classification.	Multiple Classifier PdM- SVM exhibits better performance compared to other approaches.
(Bukhsh, et.al, 2019)	Using Decision Tree, Random Forest, and Gradient Boosted Tree.	GBT Model is the most optimal results with an accuracy of 86% for maintenance needs prediction.
(Abidi, Umer, et.al, 2020)	Applying a variety of machine learning models such as NN, SVM, KNN, and RNN.	Proposed Method SH-WOA is the best model.
(Hadi, Hady, et.al, 2023)	Using Random Forest, Gradient-Boosting Classifier, Extra Trees, Light Gradient-Boosting Machine, and Extreme Gradient Boosting.	Random Forest achieves high accuracy with 99.7% on the test set.
(Senanayak a, et al., 2022)	Using ANN, SVM, TLNN, and SiMuS-TL.	SiMuS-TL attains maximum classification accuracy above an unknown target, supporting the effectiveness of the proposed method.
(Selvaraj, et.al, 2022)	Using KNN, XGBoost, ITD-SVM, GAN-SAE, and a Proposed Model.	The proposed model can produce good results without the need for time-consuming feature engineering and complex signal-transmitting tasks.
(Kim, Park, & Jung, 2021)	Comparing SVM and Random Forest with CS_SVM and CS_RF.	One-Class Method proves more effective at detecting errors when faulty data is scarce compared to binary methods.

In previous research such as that conducted by (Chazhoor, Y, M, Sanjana, & R, 2020) it was discovered that the algorithms used to model classification in carrying out predictive maintenance were Random Forest, Logistic Regression, Decision Tree, and Multi-layered Perception. Apart from that, research conducted by (Susto, Schirru, Pampuri, McLoone, & Beghi, 2015) only used two algorithms to analyze classification, namely SVM (Support Vector Machine) and KNN (K-Nearest Neighbors). It is known that most studies directly carry out classification analysis by comparing several different algorithms.

From several studies, it is known that most of the research was carried out using classification methods. Apart from that, there are not many studies that carry out the PCA process first before making predictions. Based on this, two research gaps were formed that differentiate this research from previous research: (1) this research will carry out regression analysis which aims to increase maintenance efficiency, reduce unplanned machine downtime, and optimize the use of maintenance resources, and (2) this research will carry out a comparison where the methods being compared are using the PCA method first and then carrying out regression analysis and the second method is without using PCA and directly carrying out regression analysis. Moreover, this research aims to provide predictive analysis results regarding engine conditions that have the potential to experience damage.

Research Methods

Research Flow

Research will be carried out by formulating the problem and background regarding predictive maintenance. After that, we will determine the objectives of the research that will be carried out so that the research has the right objectives and answers the problems that are currently occurring. In order for research to be reliable, it is necessary to study previous research, after which gaps will be determined that can be used as a differentiator between this research and previous research. Data collection will be carried out using a dataset regarding predictive maintenance originating from the conference (Matzka, 2020). After that, the analysis will be carried out twice, where the first analysis is carrying out the PCA stage first and after that the linear regression analysis is carried out. Then the second analysis is without doing PCA, but directly doing linear regression analysis. The analysis will be carried out in the Orange application. After the analysis is carried out, conclusions can be drawn and determine the appropriate linear regression for the dataset that has been obtained. The research framework can be seen in Figure 1.

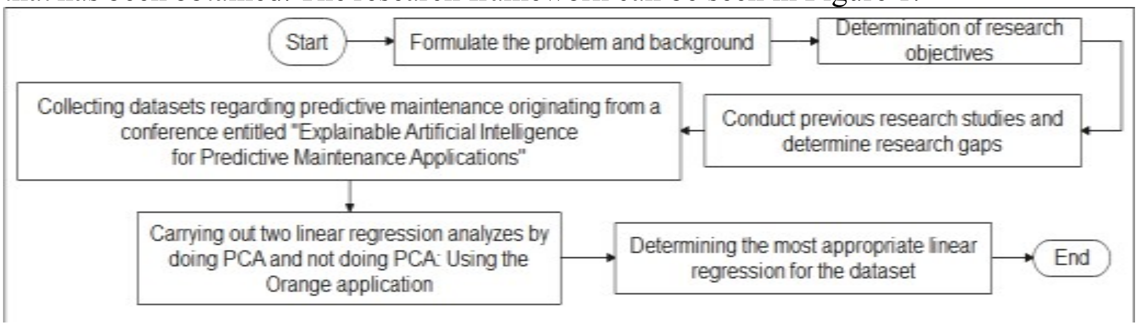


Figure 1. Research flow

Data Collection

In this research, a dataset regarding predictive maintenance taken from a conference (Matzka, 2020) will be used. From this dataset, it is known that there are 10,000 instances with a total of 14 attributes consisting of:

1. UID: unique identifier that ranges from 1 to 10000
2. Product ID: consists of the letters L, M, or H for low (50% of all products), medium (30%), and high (20%) product quality as a product quality variant and a specific serial number for that variant
3. Type: only product types L, M, or H from column 2
4. Air Temperature [K]: produced by a random walk procedure, standardized at around 300 K to a standard deviation of roughly 2 K.
5. Process Temperature [K]: produced by adding 10 K to the air temperature plus a random walk procedure standardized to a 1 K standard deviation.
6. Rotational Speed [rpm]: computed by adding normally distributed noise to 2860 W of electricity.
7. Torque [Nm]: There are no negative values and the torque value is generally

dispersed around 40 Nm with an SD of 10 Nm.

8. Tool Wear [min]: The process tools used in H/M/L quality versions have an additional 5/3/2 minutes of wear.
9. A value of 1 in the "Machine Failure" column denotes a machine failure, whereas a value of 0 denotes none at all.
10. Tool Wear Failure (TWF): 120 times in our data set, the tool will either fail or be replaced at a randomly chosen tool wear time of 200–240 minutes. The tool was changed 69 times during this period, and it failed 51 times (chosen at random).
11. Heat Dissipation Failure (HDF): If the tool rotation speed is less than 1380 rpm and the difference between the process and air temperatures is less than 8.6 K, heat dissipation will lead to process failure. This happens in 115 different data points.
12. Power Failure (PWF): the outcome of torque and rotational speed (measured in rad/s) that match the process's power requirements. The process will fail if this power is less than 3500 W or more than 9000 W, as observed in 95 instances in our data set.
13. Overstrain Failure (OSF): The process will fail as a result of excess stress if the product of torque and tool wear for product variant L (12,000 M, 13,000 H) exceeds 11,000 minNm. 98 data points are covered by this.
14. Random Failures (RNF): Regardless of the process parameters, there is a 0.1% chance that any process will fail. Less frequently than one might anticipate for the 10,000 data points in our data set, this only happens on 5 of them.

Descriptive Analysis

The descriptive results of the analysis of each attribute are given, where the results of the distribution, mean, mode, median, dispersion, minimum, maximum and missing values of each attribute are known. The results of the descriptive analysis can be seen in Figure 2.

	Name	tributi	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
N	Air temperature [K]		300.005	300.7	300.1	0.007	295.3	304.5	0 (0 %)
C	HDF			0		0.0628			0 (0 %)
C	Machine failure			0		0.148			0 (0 %)
C	OSF			0		0.0551			0 (0 %)
N	Process temperature [K]		310.006	310.6	310.1	0.005	305.7	313.8	0 (0 %)
C	PWF			0		0.0537			0 (0 %)
C	RNF			0		0.0138			0 (0 %)
N	Rotational speed [rpm]		1538.78	1452	1503	0.12	1168	2886	0 (0 %)
N	Tool wear [min]		107.95	0	108	0.59	0	253	0 (0 %)
N	Torque [Nm]		39.987	40.2	40.1	0.249	3.8	76.6	0 (0 %)
C	TWF			0		0.0293			0 (0 %)
C	Type			L		0.898			0 (0 %)
N	UDI		5000.50	1	5000.50	0.58	1	10000	0 (0 %)

Figure 2. Descriptive analysis of dataset

Data Analysis

The initial stage of data processing is using the correlations widget to determine the relationship between variables, after which imputation will be carried out on the dataset to determine whether there are missing values or not, which will be analyzed for the presence or absence of outliers so that they will be removed using the concatenate widget. After that, two analyzes will be carried out, where first the PCA process will be carried out, after which a linear regression analysis will be carried out. Then the second analysis does not go through the PCA process, but directly carries out linear regression analysis. The two analyzes will be compared to find out the best results. Orange software will be used to perform linear regression analysis. The working framework for Orange can be seen in Figure 3.

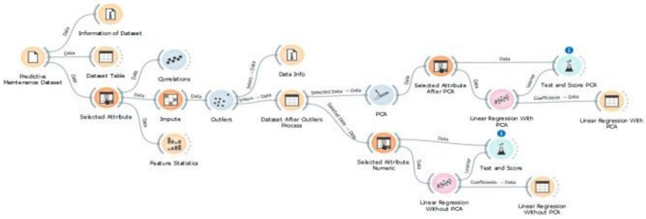
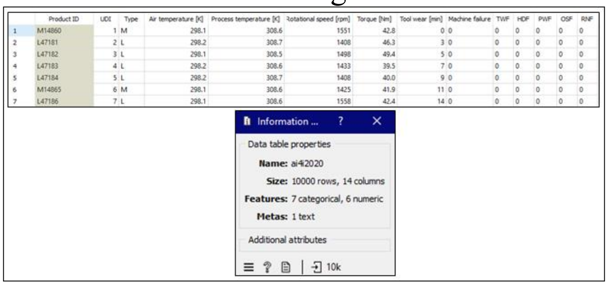


Figure 3. Linear regression used for predictive maintenance

Result and Discussion

Import csv File

The first step in carrying out the analysis is to enter the Predictive Maintenance dataset obtained from the Kaggle website from the conference (Matzka, 2020). In this dataset, it is known that there are 10,000 instances with a total of 14 attributes. Detailed information about the dataset can be seen in Figure 4.



	Product ID	UDI	Type	Air temperature [K]	Process temperature [K]	Rotational speed [rpm]	Torque [Nm]	Tool wear [min]	Machine failure	TWF	HDF	PWF	OSF	RNF
1	M14860	1	M	298.1	308.6	1551	42.8	0	0	0	0	0	0	0
2	L47181	2	L	298.2	308.7	1408	46.3	3	0	0	0	0	0	0
3	L47182	3	L	298.1	308.5	1498	49.4	5	0	0	0	0	0	0
4	L47183	4	L	298.2	308.6	1433	39.5	7	0	0	0	0	0	0
5	L47184	5	L	298.2	308.7	1408	40.0	9	0	0	0	0	0	0
6	M14865	6	M	298.1	308.6	1425	41.9	11	0	0	0	0	0	0
7	L47186	7	L	298.1	308.6	1558	42.4	14	0	0	0	0	0	0

Information ...
Data table properties
Name: ai4i2020
Size: 10000 rows, 14 columns
Features: 7 categorical, 6 numeric
Metas: 1 text
Additional attributes
10k

Figure 4. Dataset of predictive maintenance

Data Transform

Before conducting the analysis, data transformation was performed on the attributes Machine Failure which were originally categorical types, converted into numeric to enable a proper linear regression analysis. Figure 5 displays the data transformation's outcomes.

	Name	Type	Role	Values
8	**Machine failure**	N **numeric**	**feature**	**0, 1**

Figure 5. Transformation Data

Select Numeric Attribute

Next, start selecting the attributes that will be used. In this case, attributes that have a subset of numerical categories will be selected for the next analysis process. This process can be seen in Figure 6.

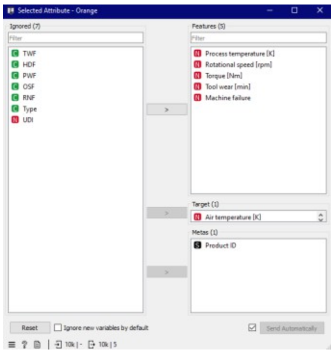


Figure 6. Selecting numeric attribute

Correlations Analysis

After selecting the attributes you want to use, it is important to first know the relationships that occur between variables to make the analysis process easier. The relationship formed can be seen in Figure 7.

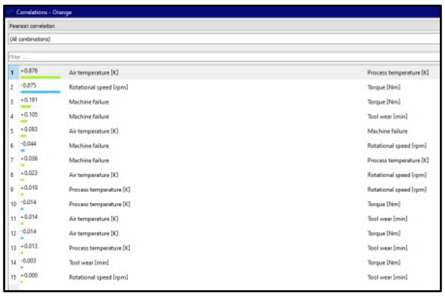


Figure 7. Correlations analysis

Impute

At this stage, an analysis of the missing values in the dataset will be carried out, where the missing values will be replaced with the average number of the column. The settings can be seen in Figure 8.

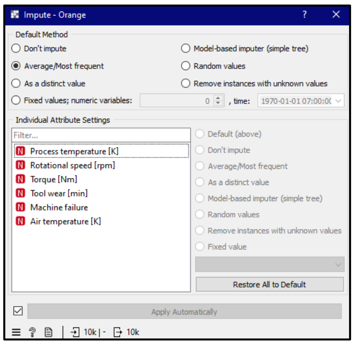


Figure 8. Impute Analysis

Descriptive Analysis of Numeric Attribute

In this case, an analysis will be carried out on each previously selected attribute to determine the distribution, mean, mode, median, dispersion, minimum, maximum and missing values of each attribute. These results can be seen in Figure 9.

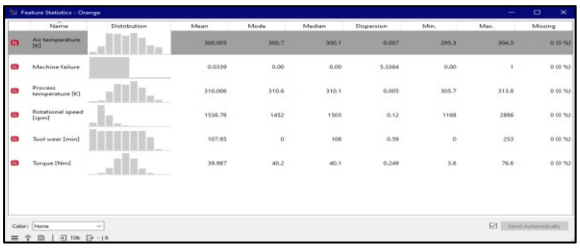


Figure 9. Descriptive analysis of numeric attribute

Outliers

The next step is to analyze outliers on the attributes selected in the previous step. At this stage, the method used is Local Outlier Factor with a contamination parameter

is

of 10% with neighbors 20 and Euclidean metrics. After processing the outliers, it was discovered that the number of instances was reduced to 9103 data. In Figure 10, the outlier process displayed.

	name	coef
1	intercept	300.003
2	PC1	0.0331478
3	PC2	0.873914
4	PC3	-1.43509
5	PC4	-0.48168

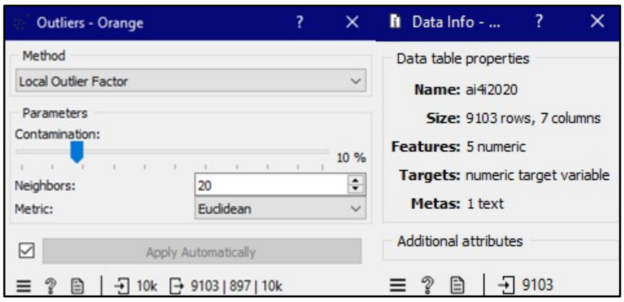


Figure 10. Outlier's analysis

Analysis Linear Regression

Next, the analysis will be carried out in two ways, namely through the PCA process and not through the PCA process, where the results will be compared to determine the best result.

1. *Linear Regression Analysis with PCA*

   At this stage the process will begin by carrying out a PCA analysis where the 5 attributes will be reduced to 4 attributes or principle components with the names PC1, PC2, PC3, and PC4 which maintain 98% of the data variation so that it can meet the requirements. Then the cumulative variance formed is 0.980 and the component variance is 0.184. The PCA process can be seen in Figure 11.

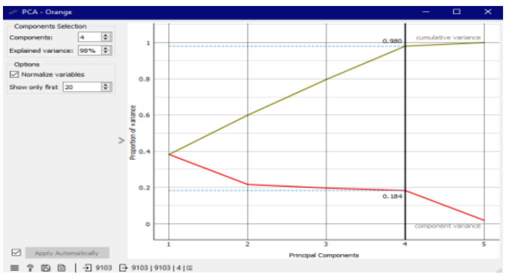


Figure 11. Principal component analysis

   After that, attribute selection is carried out using the select column widget where Air Temperature will be used as the target or dependent variable (Y), while PC1 (X1), PC2 (X2), PC3 (X3), and PC4 (X4) will be used as features or independent variables (X). Details of attribute selection can be seen in Figure 12.

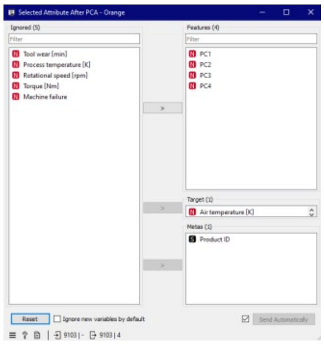


Figure 12. Selecting main components of PCA

After attribute selection is carried out, linear regression analysis can be carried out. The results of linear regression can be seen in Figure 13.

Model	MSE	RMSE	MAE	R2
Linear Regression With PCA	0.909	0.953	0.806	0.772

Figure. 13. Result of linear regression with PCA

From these results it is known that the linear regression equation formed is:
$Y = 300.003 + 0.033X1 + 0.874X2 - 1.435X3 - 0.482X4$

From this equation it is known that Air Temperature, which is the dependent variable, is influenced by 4 independent variables, namely PC1, PC2, PC3, and PC4. PC1 has a coefficient of 0.033. There is a positive correlation between PC1 and air temperature, as indicated by this positive coefficient. It can be concluded that if the PC1 value increases, the Air Temperature value will increase too. The coefficient for PC2 is 0.874. There is a positive correlation between PC2 and air temperature, as indicated by this positive coefficient.

It can be concluded that if the PC2 value increases, the Air Temperature value will increase too. PC3 has a coefficient of -1.435. This negative coefficient indicates that there is a negative relationship between PC3 and Air Temperature. It can be concluded that if the PC3 value increases, the Air Temperature value will tend to decrease. PC4 has a coefficient of – 0.482. This negative coefficient indicates that there is a negative relationship between PC4 and Air Temperature. It can be concluded that if the PC4 value increases, the Air Temperature value will tend to decrease. Apart from that, it is also known that the intercept value found when the PC1, PC2, PC3, and PC4 values are 0, Air Temperature is found to have a positive value of 300.003.

In this analysis, it is also known that the MSE (Mean Squared Error) value obtained is 0.909, then the RMSE (Root Mean Squared Error) value is 0.953, then the MAE (Mean Absolute Error) value is 0.806, and the R-squared (R^2) 0.722.

2. *Linear Regression Analysis without PCA*

At this stage the linear regression process will be carried out without carrying out PCA analysis, where the 5 attributes will be retained for linear regression analysis. In this stage, the Air Temperature (Y) attribute will be used as the target or dependent variable, while Rotational Speed (X1), Torque (X2), Tool Wear (X3), Process Temperature (X4), and Machine Failure (X5) will be used as features or independent variables. Details of attribute selection can be seen in Figure 14.

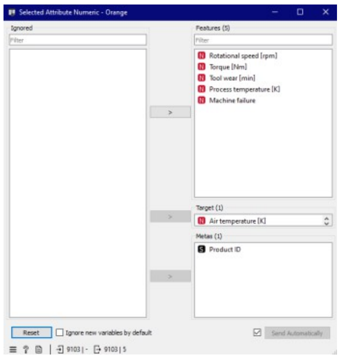


Figure 14. Selecting target

After attribute selection is carried out, linear regression analysis can be carried out. The results of linear regression can be seen in Figure 15.

	name	coef
1	intercept	-64.7421
2	Rotational spee...	-0.00015771
3	Torque [Nm]	-0.00518357
4	Tool wear [min]	-3.46494e-05
5	Process temper...	1.17797
6	Machine failure	0.828752

Figure. 15. Result of linear regression without PCA

From these results it is known that the linear regression equation formed is:

$$
\begin{aligned}
Y = -64.7421 - 0.00015771X_1 - 0.00518357X_2 \\
- 3.46494e05X_3 + 1.17797X_4 \\
+ 0.828752X_5
\end{aligned}
$$

From this equation it is known that Air Temperature, which is the dependent variable, is influenced by 5 independent variables, namely Rotational Speed, Torque, Tool Wear, Process Temperature, and Machine Failure. Rotational Speed has a coefficient of -0.00015771. This negative coefficient indicates that there is a negative relationship between Rotational Speed and Air Temperature. It can be concluded that if the Rotational Speed value increases, the Air Temperature value will tend to decrease. Torque has a coefficient of -0.00518357. This negative coefficient indicates that there is a negative relationship between Torque and Air Temperature. It can be concluded that if the Torque value increases, the Air Temperature value will tend to decrease. Tool Wear has a coefficient of -3.46494e05. This negative coefficient indicates that there is a negative relationship between Tool Wear and Air Temperature. It can be concluded that if the Tool Wear value increases, the Air Temperature value will tend to decrease. Process Temperature has a coefficient of 1.17797. The positive coefficient suggests a positive correlation between the air temperature and the process temperature. It can be concluded that if the Process Temperature value increases, the Air Temperature value will also increase. The coefficient for machine failure is 0.828752. The positive coefficient suggests a positive correlation between Air Temperature and Machine Failure. It can be concluded that if the Machine Failure value increases, the Air Temperature value will also increase. Apart from that, it is also known that the intercept value is found when the value of all independent variables is 0, so the Air Temperature is found to have a negative value of -64.7421.

In this analysis, it is also known that the MSE (Mean Squared Error) value obtained is 0.909, then the RMSE (Root Mean Squared Error) value is 0.953, then the MAE (Mean Absolute Error) value is 0.806, and the R-squared (R^2) value is 0.772. Information regarding these results can be seen in Figure 16.

Model	MSE	RMSE	MAE	R2
Linear Regression Without PCA	0.909	0.953	0.806	0.772

Figure. 16. Evaluation linear regression without PCA

Conclusion

This study utilizes linear regression analysis techniques to investigate predictive maintenance. Two models were employed: one involved conducting principal component analysis (PCA) before linear regression, while the other skipped PCA and used the data directly. Surprisingly, both models yielded identical values for key metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). This suggests that the inclusion or exclusion of the PCA process did not significantly impact the analysis, likely due to the characteristics of an artificial dataset created by the author, which could be considered ideal. Moreover, PCA revealed a high explained variance of 98%, retaining 4 out of 5 attributes that were reduced, thus explaining the similarity in evaluation results between the two models. Moving forward, future research should consider expanding the scope of data incorporation, including relevant attributes, and ensuring the utilization of up-to-date datasets. Exploring alternative analytical techniques such as time series analysis and clustering may also offer deeper insights. Embracing diverse analytical approaches is crucial to achieving a comprehensive understanding of the investigated phenomenon. The author expresses gratitude to the editor and reviewers for their invaluable feedback, which significantly improved the paper's quality.

BIBLIOGRAPHY

Aalst, W. V. (2016). Data Science in Action. Berlin: Springer. doi:https://doi.org/10.1007/978-3-662-49851-4_1.

Ebeling, C. (2010). An Introduction to Reliability and Maintability Engineering. New York: MC Graw Hill.

Kherif, F., & Latypova, A. (2020). Chapter 12 - Principal component analysis. Machine Learning Methods and Applications to Brain Disorders, 209–225. doi:10.1016/B978-0-12-815739-8.00012-2.

Kiangala, K. S., & Wang, Z. (2018). Initiating predictive maintenance for a conveyor motor in a bottling plant using industry 4.0 concepts. International Journal of Advanced Manufacturing Technology, 97(9- 12), 3251–3271.

Kim, S.-G., Park, D., & Jung, J.-Y. (2021). Evaluation of One-Class Classifiers for Fault Detection: Mahalanobis Classifiers and the Mahalanobis–Taguchi System. Processes, 9.

Matzka, S. (2020). Explainable Artificial Intelligence for Predictive Maintenance

Applications. 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), (pp. 69-74).
Mobley, R. K. (2002). An introduction to predictive maintenance (2 ed.). Amsterdam: Elsevier Science.
Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis (6 ed.). New York: John Wiley & Sons, Inc.
Pranoto, J., Matondang, N., & Siregar, I. (2013). Implementasi Studi Preventive Maintenance Fasilitas Produksi Dengan Metode Reliability Centered Maintenance Pada PT. XYZ. Jurnal Teknik Industri USU, 1(3), 18-24.
Richardson, M. (2009). Principal component analysis. 6(16), 4. Retrieved from http://people. maths. ox. ac. uk/richardsonm/SignalProcPCA.pdf
Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 231(9), 1670–1679. doi:https://doi.org/10.1177/0954405415601640
Senanayaka, A., Mamun, A. A., Bond, G., Tian, W., Wang, H., Fuller, S., Bian, L. (2022). Similarity-based Multi-source Transfer Learning Approach for Time Series Classification. International Journal of Prognostics and Health Management.
Wu, D., Jennings, C., Terpenny, J., & Kumara, S. (2016). Cloud-based machine learning for predictive analytics: Tool wear prediction in milling. In Proceedings - 2016 IEEE international conference on big data, (pp. 2062-2069).
Zonta, T., Costa, C. A., Righi, R. d., Lima, M. J., Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering.

Copyright holder:
Benedict Ariel Kurnianto, Rudy Hartono Prayogo, Nidia Pialina Nababan, Suharjito (2024)

First publication right:
Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under: