JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 4, October 2024, pp. 1026-1039

Selection Dominant Features Using Principal Component Analysis for Predictive Maintenance of Heave Engines

Suryasatriya Trihandaru[1], Hanna Arini Parhusip[2], Adrianus Herry Heriadi[3], Petrus Priyo Santosa[3], Yohanes Sardjono[4], Lea[1]
[1]Satya Wacana Christian University, Indonesia
[2]Departemen Matematika dan Sains Data, Indonesia
[3]PT APSI Jakarta, Indonesia
[4]BRIN Jogjakarta, Indonesia
suryasatriya@uksw.edu

Article History:
Received : 11-04-2024
Revised : 01-08-2024
Accepted : 05-08-2024
Online : 01-10-2024

Keywords:
PCA;
SOS;
CoreDataQ;
Predictive Maintenance.

ABSTRACT

This article aims to identify the dominant features that have a significant impact on the health of a heavy machine that relates to the digital infrastructure of a company. The importance of this research is that the authors define predictive maintenance based on Principal Component Analysis (PCA), which is the novelty of this article. The novel contribution of this research lies in the application of Principal Component Analysis (PCA) for predictive maintenance of heavy machinery, which has not been integrated into the Scheduled Oil Sampling (SOS) procedures. The recorded data are called Scheduled Oil Sampling (SOS) and historical data from an equipment called CoreDataQ, which works for recording many features from heavy machine activities. The data contain two sets data. The method is Principal Component Analysis (PCA). This method leads to obtain a maximum of 20 significant features on data based on SOS. The results have been confirmed and agreed upon by the manager who owned CoreDataQ to consider the selected dominant features for further related maintenance.

Crossref
https://doi.org/10.31764/jtam.v8i4.22854

This is an open access article under the CC–BY-SA license

—————————— ◆ ——————————

A. INTRODUCTION

Predictive maintenance of machines is becoming increasingly crucial with the advancement of digital infrastructure, particularly in the context of Industry 4.0 (Zonta et al., 2020). Digital infrastructure is one of the most pressing needs of this century due to the necessity for integrating acquired digital information. This integration is vital not only for the industry but also for a country's overall digital framework (Schade & Schuhmacher, 2022). Enhancing digital infrastructure supports the continuation of Industry 4.0 processes, aiming for processes and products that align with sustainable development goals (Systems, 2022). Consequently, digital infrastructure has become a significant requirement for various large industrial sectors (Dabića et al., 2023)(Qiu et al., 2023), including in Indonesia. PT Artha Puncak Semesta Indonesia (APSI), which operates in the mining industry and utilizes heavy machinery around the clock, is one such industry that greatly benefits from these advancements.

PT Artha Puncak Semesta Indonesia (APSI), headquartered in Jakarta and founded in 2007, specializes in producing CoreDataQ. This technology serves as a digital infrastructure for activities related to mining exploration, including planning, communication, and education. APSI provides the latest technologies such as machine health data analysis, multi-platform data aggregators, data download and storage on machine units, automation, integration, and interactive data visualization to support decision-making systems (Luo, 2019)(Dugger et al., 2022), Numerous academics have also developed and researched such technologies (Lee et al., 2022)(Zhang et al., 2022).

Predictive maintenance remains a significant challenge, particularly when dealing with heavy machinery in mining operations (Stodola & Stodola, 2020). These industries typically rely on Scheduled Oil Sampling (SOS) as the standard procedure for maintenance. However, this approach has limitations when it comes to predicting machinery failures and optimizing maintenance schedules. To address this issue, the article proposes enhancing the SOS procedure by implementing Principal Component Analysis (PCA). PCA is utilized to identify the dominant features that contribute the most to predictive maintenance, thereby improving the accuracy and efficiency of maintenance decisions.

The primary objective of this research is to renew the existing SOS procedure by integrating PCA to streamline the maintenance process for heavy machinery. The novelty of this approach lies in the fact that PCA has never been incorporated into the SOS framework before. By reducing the dimensionality of the features involved, the research aims to make the maintenance of heavy engines more efficient. Although each engine's analysis takes only a few minutes, the overall cost savings could be substantial, leading to more cost-effective and timely maintenance practices.

B. METHODS

1. Principal Component Analysis In Predictive Maintenance

There are several methods to renew the SOS procedure. For instance, the Recursive Feature Elimination (RFE) is one of the feature selection methods used to select important features that contribute the most to predicting target variables, which has been extended using random forest for classification (Bahl et al., 2019), or using Support Vector Machine(SVM) (Escanilla et al., 2018)(Alisneaky, 2019). This method works by iteratively eliminating features that have the lowest weight until the desired number of features is reached. RFE has been used for cervical cancer classification (Hamada et al., 2022). RFE is particularly useful in data analysis because it has many features, such as image data, genomic data (Peterson & Coleman, 2006), or stroke data (Chourib et al., 2022). RFE can be used with a variety of machine learning models, such as logistic regression, SVM, or other models. The other method is implementing the Principal Component Analysis. PCA is a powerful technique for reducing the dimensionality of high-dimensional data while preserving the most important information. It is widely used in fields such as image processing (Ng, 2017), genetics (de Góes Maciel et al., 2024), and finance (Beliavsky et al., 2023), among others. Previous research dealt with visualizing data with several features after the Principal Component Analysis was employed (Parhusip et al., 2022). The main idea behind PCA is to identify the principal components (linear combinations of the original features) that capture the most amount of variation in the data.

2. Data Sources

The study uses Principal Component Analysis (PCA) on two different datasets, referred to as Case 1 and Case 2, to identify significant features for predictive maintenance of heavy machinery. Here's how PCA was applied to each dataset and the differences in approach:
Case 1: February 2021 Dataset
Data Description: This dataset consists of 20,656 rows and 94 columns, recorded by CoreDataQ from five heavy vehicles.

The data for February 2021 consists of 20656 rows x 94 columns, which means there are 20656 rows with 94 features. Data was recorded on February 10, 2021, by the CoreDataQ infrastructure. After preprocessing the data, there are 49 features resulting from observations of five heavy vehicles in the mining area,i.e.:
'incline_x', 'incline_y', 'gps_status', 'ecm_status', 'imu_status', 'powersup_status', 'barometric_press', 'vehicle_speed', 'clutch_switch', 'cruise_cont_state', 'enable_switch', 'set_speed', 'pto_speed', 'fuel_level_1', 'act_engine', 'eng_speed', 'eng_torq_mode', 'eng_demand', 'eng_current_speed', 'nom_friction', 'eng_des_op_speed', 'eng_des_op_speed_ass', 'est_eng_par_loss', 'tot_eng_hours', 'revolutions', 'eng_cool_temp', 'eng_fuel_temp', 'eng_oil_temp', 'fuel_cons', 'fuel_rate', 'inst_fuel_ec', 'tot_dist', 'high_res', 'eng_press', 'eng_temp', 'tach_output_shaft_speed', 'tach_vehicle_speed', 'trans_sel_gear', 'trans_cur_gear', 'trans_actual', 'eng_oil_press', 'bat_vol', 'keyswitch_batpot', 'act_retarder', 'ret_torq_mode', 'act_max_avret', 'eng_moment' 'prog_shift', 'est_fan_speed'

Case 2. August 2019 to January 2022 Dataset
Data Description: This dataset includes data manually recorded from 43 heavy equipment units, with 52 features each, over the period from August 8, 2019, to January 22, 2022. The names of all the features in the data file read by Python, i.e.,:
Index(['unitno', 'model', 'maker', 'compartid', 'compart', 'oiltypeid', 'oilgradeid', 'coolantid', 'notesos', 'label', 'last_interpret_date', 'sampledate', 'processdate', 'oiladded', 'oilhours', 'actual_fluid_hours', 'meterread', 'oilchanged', 'filterchanged', 'evalcode', 'dateReported', 'Confirmed', 'Problem Solved', 'Action Taken', 'SMU', 'Si', 'Al', 'Cr', 'Fe', 'Pb', 'Cu', 'Sn', 'Ni', 'Na', 'K', 'Mo', 'Zn', 'Mg', 'Ca', 'P', 'B', 'ST', 'OXI', 'NIT', 'SUL', 'V100', 'TBN', 'W', 'FP', 'GF', 'PQ', 'Visual'], dtype='object'. We will find the dominating features to be considered for further analysis.

3. Data Preprocessing

Cleaning: Removed rows with missing or NaN values and columns with constant values or that served only as identifiers. For each, removed rows with missing values and constant-value columns.
Standardization: Standardized the data similarly to Case 1. Standardized the data to have a mean of zero and a standard deviation of one. PCA works best on data that has been standardized (i.e., mean-centered and scaled by the standard deviation). Mathematically, one writes as:

$$X_{i,new} = \frac{X_i - \mu_i}{\sigma_i}, \quad i\text{: } i\text{-th feature.}$$

Finally, we implement the PCA to the standardized data. PCA Application:

a. Covariance Matrix Calculation: Computed to understand feature relationships. The covariance matrix shows how the different features in the data are related to each other, which is shown here, i.e.

$$C = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_p) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & Cov(X_p, X_2) & \cdots & Var(X_p) \end{bmatrix}$$

b. Eigenvalues and Eigenvectors: Calculated to identify principal components. Compute the eigenvectors and corresponding eigenvalues of the covariance matrix. The eigenvectors represent the principal components, while the eigenvalues represent the amount of variance explained by each principal component (Jollife & Cadima, 2016). The eigenvalues and eigenvectors can be computed by the following formula, i.e., $CX_C = \lambda X_C$ where X_C is the eigenvector associated with the computed eigenvalues of covariance matrix C.
c. Principal Components Selection: Selected based on cumulative variance explained. Initially examined the top two components.
d. Feature Identification: Analyzed loadings (coefficients) to identify significant features, with the highest loadings being 'incline_y' and 'act_max_avret'. Loadings on principal components were used to identify significant features in both cases, but the specific features varied due to differences in the datasets. Case 1 highlighted features like 'incline_y' and 'act_max_avret', while Case 2 identified metal levels as significant.
e. Further Feature Selection: A cumulative variance plot indicated that 20-30 features contributed significantly, thus up to 30 significant features were identified.
f. Summary and Differences.

In essence, while the overall approach to PCA was similar for both datasets, the specific features identified as significantly differed, reflecting the unique characteristics of each data set.

C. RESULT AND DISCUSSION

1. Data Preprocessing Before Analysis

The data preprocessing process is essential to ensure that the dataset is clean, meaningful, and suitable for analysis. The steps taken before performing Principal Component Analysis (PCA) are as follows:

a. Initial Data Inspection and Cleaning:

Removing Unused Features: Features that contain rows with zero values, NaN values, constant values during the observation, or those representing status or identity are not used in the data processing. For example, features like 'id', 'timestamps', 'ip_address', 'unit_type', and 'veh_name' are discarded. Filtering Constant Columns: Features with constant values across all observations (both numeric and string) are excluded from the analysis. Examples include 'gps_status', 'ecm_status', 'imu_status', and 'powersup_status'.

b. Identifying and Handling Missing Values:
Missing Data Identification: The dataset is checked for missing or null values, and a summary is generated to identify features with missing data. For example, features like 'oil graded' and 'coolants' have missing values that are accounted for. Discarding Rows with Missing Values: Rows with missing values in critical features are discarded to ensure the quality and completeness of the data.

c. Feature Selection:
Feature Importance and Variability: Initial feature selection is based on the variability and importance of the features. Features that do not contribute significantly to the analysis are removed. Selection of Relevant Features: The relevant features retained for analysis include 'incline_x', 'incline_y', 'barometric_press', 'vehicle_speed', 'fuel_level_1', 'act_engine', 'tot_eng_hours', 'revolutions', 'eng_fuel_temp', 'fuel_cons', 'tot_dist', 'high_res', and others.

d. Data Standardization:
Standardizing the Data: The data is standardized to have a mean of zero and a standard deviation of one. This step ensures that all features contribute equally to the analysis and prevents features with larger scales from dominating the results.

e. Handling Redundant Features:
Removing Redundant Features: Features that are highly correlated or redundant are identified and removed to reduce multicollinearity and improve the efficiency of the PCA.

f. Transforming Data:
Feature Engineering: Additional transformations and feature engineering may be performed to enhance the dataset's quality. This can include creating new features based on existing ones or transforming features to better represent the underlying data.
By meticulously following these preprocessing steps, the data is cleaned, standardized, and prepared for PCA. This ensures that the analysis is performed on a high-quality dataset, leading to more accurate and reliable results. Finally, we observe that the used features are:
'incline_x', 'incline_y','gps_status', 'ecm_status', 'imu_status', 'powersup_status','barometric_press','vehicle_speed', 'clutch_switch','cruise_cont_state','enable_switch','set_speed', 'pto_speed', 'fuel_level_1', 'act_engine', 'eng_speed','eng_torq_mode', 'eng_demand', 'eng_current_speed','nom_friction', 'eng_des_op_speed', 'eng_des_op_speed_ass', 'est_eng_par_loss', 'tot_eng_hours', 'revolutions', 'eng_cool_temp', 'eng_fuel_temp', 'eng_oil_temp', 'fuel_cons', 'fuel_rate', 'inst_fuel_ec', 'tot_dist', 'high_res', 'eng_press', 'eng_temp', 'tach_output_shaft_speed', 'tach_vehicle_speed', 'trans_sel_gear', 'trans_cur_gear', 'trans_actual', 'eng_oil_press', 'bat_vol', 'keyswitch_batpot', 'act_retarder', 'ret_torq_mode', 'act_max_avret', 'eng_moment', 'prog_shift', 'est_fan_speed', 'Area'

Furthermore, we can do an observation process where the data finally has the size [20656 rows x 49 columns]. Of all the 49 features, we want to choose which feature dominates, which is done with PCA. We need to establish how many of the selected features above are considered to dominate. Firstly, we will use PCA as the standard method. Note that there are features that are not enumerative but do not give the necessary meaning. These need to be cleaned or removed, meaning unnecessary columns are not considered.

2. Result on Case 1 using Principal Component Analysis (PCA)

Implementing Principal Component Analysis (PCA) for case 1 is done by determining the first two dominating features to guarantee that the algorithm is working well. The data were initially standardized. We observe that the first two greatest eigenvalues are 0.35747157 and 0.1731355. Then, we use the relationship of this variance or eigenvalues with the covariance matrix and the values in the eigenvector to get the two feature names sought from the original feature name. From the computation, the two dominant features are incline_y and act_max_avret. The result has been confirmed by the company, which also agrees with the result. However, the company may require more features to be considered. The company needs the maximum number of features that can be kept as dominant features. Therefore, we propose cumulative variances to identify how many features might be considered. Figure 1 depicts the potential features that determine the number.

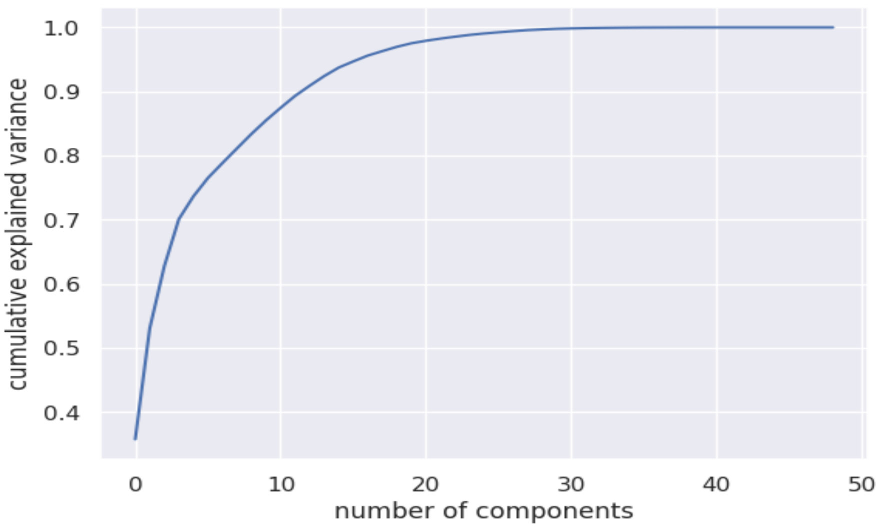

Figure 1. Illustration of cumulative explained variance based on the number of components (features)

Figure 1 suggests that 20-30 features contribute to variance because, after 30 components, (features) there is no more change in variance. Thus, analysis can be done to obtain 20-30 features. The process in PCA takes too long compared to the following methods, where we don't need to create principal components in other variables before getting the actual features. The following 20 features were selected and the importance of each feature as shown in Figure 2.

Figure 2. The 20 selected features and the importance of each feature are indicated by the height of the bar.

In fact, although 20 features may be significant in dominating the data, some of them do not contribute their importance in Figure 2. The height of the importance bar that looks not zero is about 11 only, or we round it to 10 even though Figure 1 shows that we can set it up to 20. By observing the result, we get 4 features that in the 10 are not listed in the 20 selected features, i.e. 'inst_fuel_ec', 'eng_press', 'trans_actual', and 'act_retarder'. The list can be the same if we use the same number of estimators in the random forest classifier. Therefore, we vary the number of estimators, i.e., 30, to get the 10 selected features, i.e. 'barometric_press', 'vehicle_speed', 'fuel_level_1', 'act_engine', 'tot_eng_hours', 'revolutions', 'eng_fuel_temp', 'fuel_cons', 'tot_dist', 'high_res'] The list has different features indicating that the number of estimators is significant to be constant. Furthermore, we depict the importance of the selected features in Figure 3.

Figure 3. Diagram of the importance of features dominating data.

If we set there are 10 features, it turns out that there are only about 7 features with a non-zero level of dominance as shown in Figure 3. One may conclude that the selected number of features is first given, and the algorithm determines the result.

3. Result on Case 2

As mentioned in the method section, the CoreDataQ digital infrastructure has been developed, and there needs to be data acquisition of the stored historical data results. This article only shows some of the obtained data, namely historical data from August 8, 2019, to January 22, 2022. Several simple studies were produced to be one of the decisions on predictive maintenance. By listing data based on oil type (written as the oiltypeid on the list above), we obtained: DEO 37, DEO CH4 5, UNKNOWN 1, indicating that DEO is the oil type used at most. Thus, there are two types of used oil and there is one that is not recognized. A visual distribution of one of the features can also be done, as shown in Figure 4.

<matplotlib.axes._subplots.AxesSubplot at 0x7f0acce45d90>

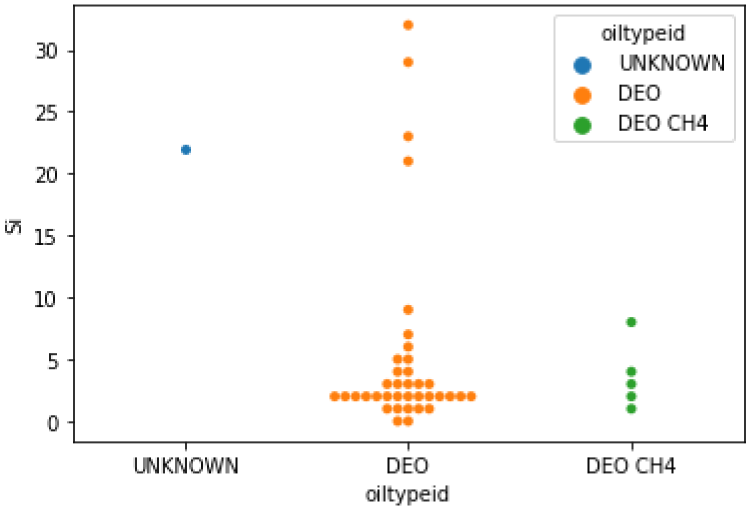

Figure 4. Visualization of the use of oil on one of the features (Si)

Furthermore, we observe whether there is missing data or null values. This step is necessary for the process of cleaning data. We obtained output results stating that some features are missing, shown as follows:
unitno 0; models 0; maker 0; compared 0; compart 0; oiltypeid 0; oil graded 39; coolants 43; notes 0; label 0; last_interpret_date 0; sample date 0; process date 0; oil added 1; oil hours 0; actual_fluid_hours 0; meterread 0; oilchanged 3; filterchanged 43; evalcode 0; dateReported 0; Confirmed 0; Problem Solved 0; Action Taken 0; SMU 0; Si 0; Al 0; Cr 0; Fe 0; Pb 0; Cu 0; Sn 0; Ni 0; Na 0; K 0; Mo 0; Zn 0; Mg 0; Ca 0; P 0; B 0; ST 0; OXI 0; NIT 0; SUL 0; V100 0; TBN 0; W 0; FP 0; GF 0; PQ 0; Visual 0; dtype: int64.

The output results are in 2 columns, where column 1 is the feature name, and column 2 is the amount of missing data. For example, an oil graded 39 means that there are 39 missing oil graded. After that, we discard all data rows that are missing and those that contain rows and features with NaN and NaT because they are not meaningful. This means that the number of observed units is reduced, and the number of meaningless features is discarded. We read the data based on certain features with histograms. For example, by paying attention to Fe to oilchanged, one yields the histogram result as depicted in Figure 5.

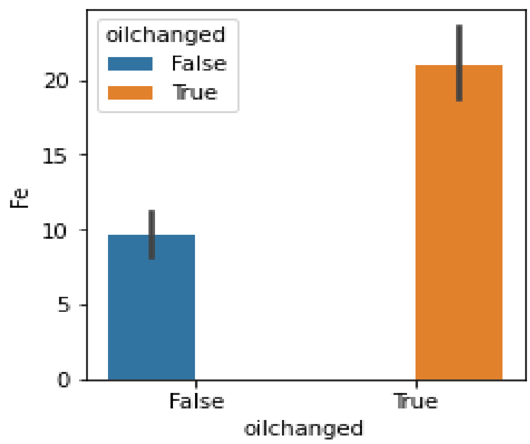

Figure 5. The relationship between the amount of Fe content and oil changed.

In Figure 5, the blue bars are the number of Fe levels around 10, so oil changed (denoted as oilchanged) which is marked false and those who change the oil (colored orange) based on Fe levels around 20, where the engine must change the oil. In fact, from the standard table, it is critical to change the oil if the Fe is > 200. Therefore, this means changing the oil due to another feature. According to PT APSI technicians, oil replacement is very dominant due to the influence of the large SMU. To explore further, several other features are used to get visualization, which is illustrated in Figure 6, where SMU is the dependent variable, as shown in Figure 6.

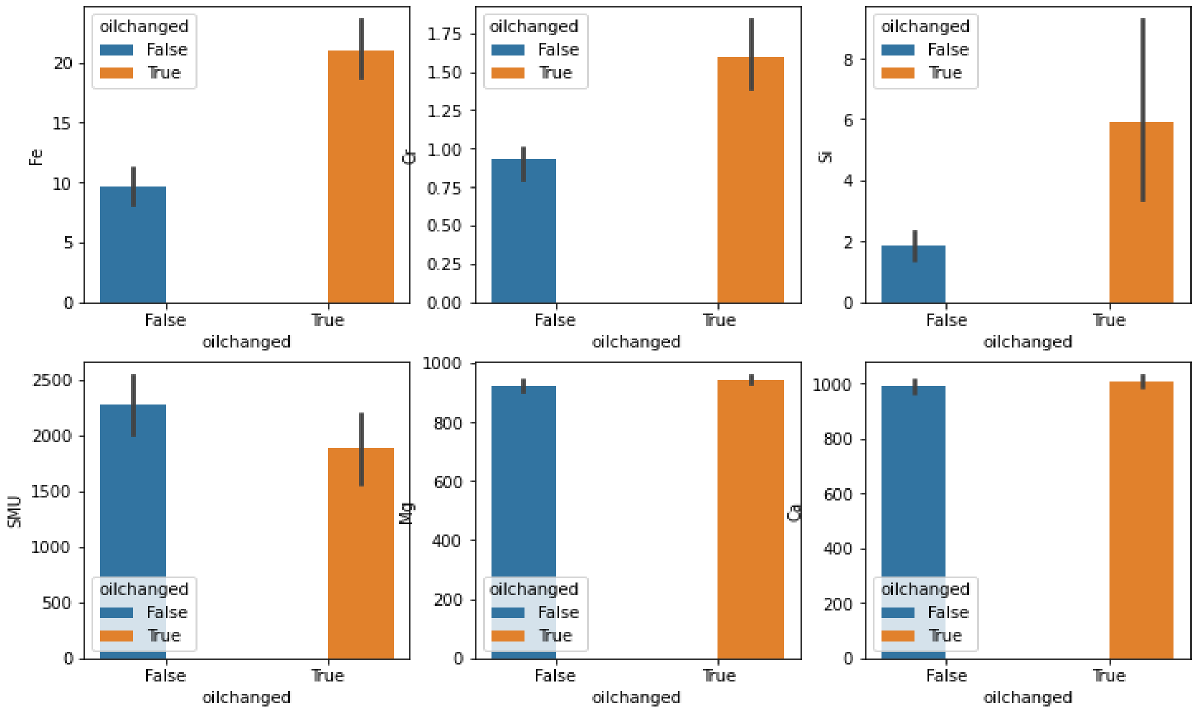

Figure 6. The relationship between the amount of metal content,i.e.. Fe., Cr, Su (row 1); SMU, Mg, and Ca (2nd row), for oil changes (typed as oilchanged)

In Figure 6, for all the depicted features between oilchanged and other metal features such as Fe, Cr, Su, SMU, Mg, and Ca, the orange bar is shorter when the selected vertical feature is SMU. This provides empirical evidence of the need to pay attention to SMUs in oilchanged based on data. Additionally, there are 2 distributions of data groups where around Fe is 7-13 and with data distribution on Fe > 13. Additional study is addressed, i.e., service meter unit (SMU) vs. oilchanged. Both can be used for comparison because SMU (Service Meter Unit) is a number that shows the number of working hours (operating hours) of heavy equipment from starting the engine on and off. This SMU can later be determined to perform periodic service (periodic maintenance) as well as the standard oil changed.

Further studies related to SOS have been carried out for several heavy equipment units where Tableau is used for visualization of various existing feature relationships grouped into 3 categories, namely wear metal, contaminants, and metal additives (Parhusip et al., 2022) to determine the dominant chemical elements that may improve the monitoring of the productivity and efficiency of heavy engines in 2015-2021 in the company. The result has shown that various types of analysis may be used to contribute to the equipment health analysis. From these features, we do not know the correlation between existing features. Because SMU is considered to affect the metal concentration that occurs, we pay attention to the correlation between features. We code by preferring the correlation of each, as shown in Figure 7. We choose features whose correlation is greater than 0.5 to SMU, which means that those features are factors that need attention. However, we want to select 10 features from all features with a correlation > 0.5, that are more influential than other features. For this reason, Principal Component Analysis (PCA) is used to obtain these ten features, namely: 'Pb', 'K', 'Cu', 'Fe', 'Si', 'Al', 'Mo', 'Ca', 'Sn', 'B'.

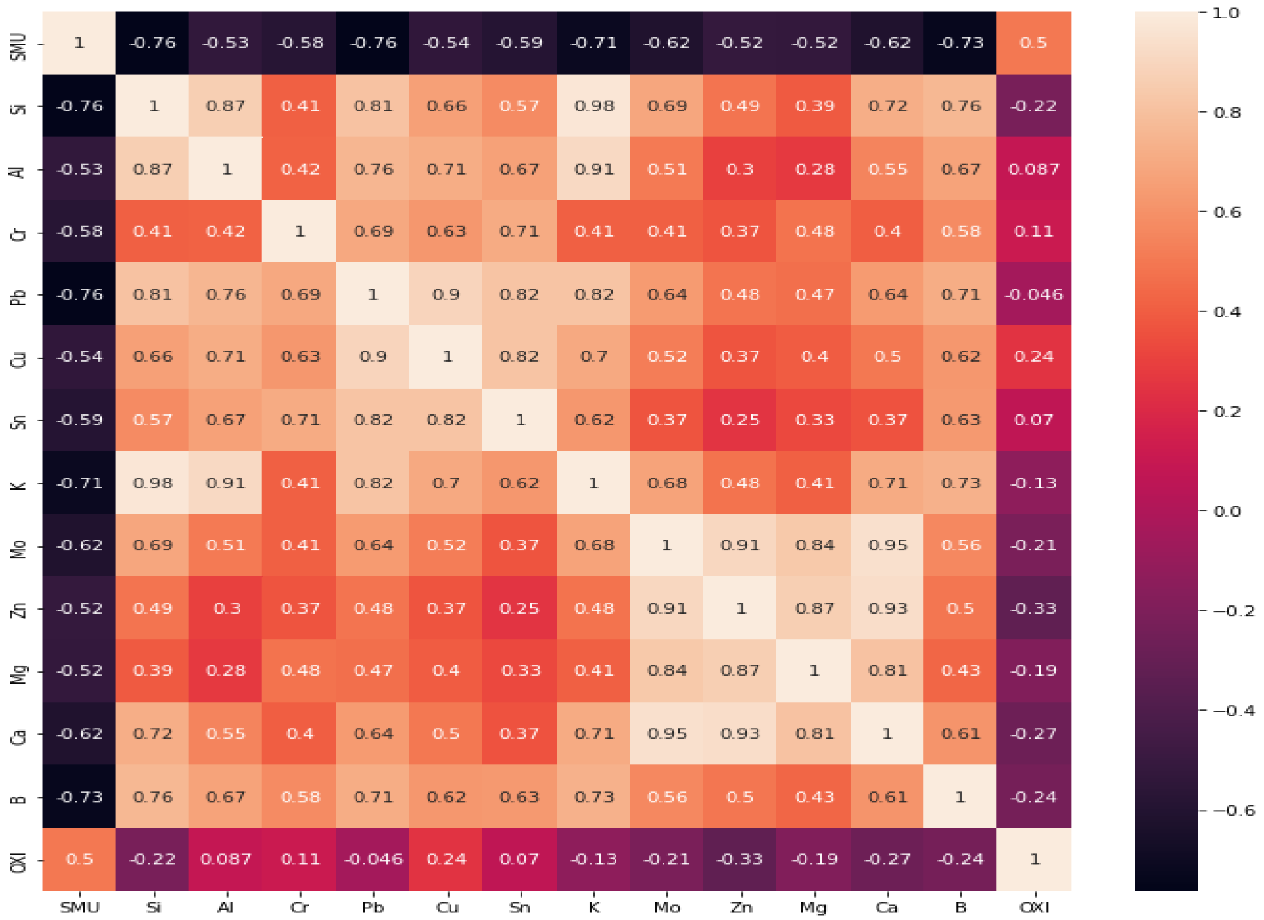

Figure 7. Correlation between 2 features
for all features.

Because SMU is considered very important in the process of changing the oil, we observe the fluctuations in the features above during the observation time, as shown in Figure 8. The visualization in Figure 8 indicates that the fluctuation of SMU is not significantly increasing for each feature. Therefore, the prediction that the 10 features correlating more than 0.5 are thought to cause fluctuations in the SMU value is not true. Finally, it is necessary to carry out further studies of the observed features.

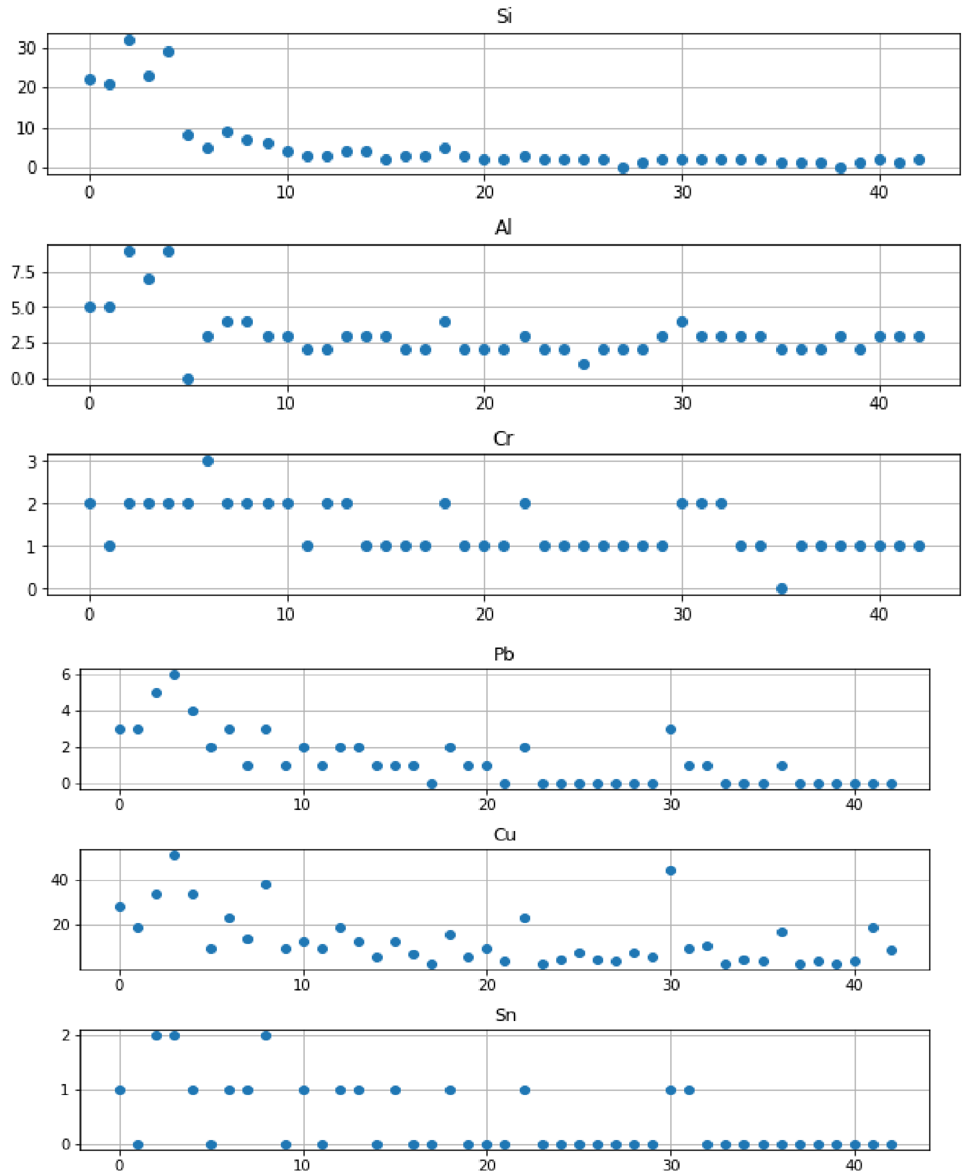

Figure 8. Examples of feature distribution (Si, Al.Cr.Pb, Cu, Sn) (horizontal on each image and versus SMU (vertical)

D. CONCLUSION AND SUGGESTIONS

This article explores predictive maintenance for heavy machinery using data from PT APSI's digital infrastructure, CoreDataQ. The analysis was conducted using Principal Component Analysis (PCA) and machine learning techniques. Initially, PCA identified two dominant features, incline_y and act_max_avret, but further analysis revealed that around 30 features could be significant. These features were compared against the Service Meter Unit (SMU), but no significant correlations were found. Despite this, CoreDataQ demonstrated robust performance, supported by historical oil sampling data, and provided valuable recommendations for decision-makers. Specific recommendations from the data analysis include focusing on the identified significant features for maintenance scheduling and refining the predictive models used in CoreDataQ. Implementing these recommendations can enhance the efficiency of predictive maintenance practices at PT APSI by targeting the most impactful features and optimizing maintenance schedules accordingly. Future research should focus on enhancing feature selection using methods like Recursive Feature Elimination (RFE) and comparing the effectiveness of various machine learning models for predictive maintenance. Additionally, integrating advanced feature selection with CoreDataQ's business processes, developing Big Data Analysis (BDA) and IoT methods, and examining network topology modeling will be crucial for improving system performance and connectivity.

ACKNOWLEDGEMENT
The article here is funded by the Ministry of Higher Education and Culture Research and Technology under the grant No. 001/LL6/PB/AL.04/2024, 048/SPK-PTM/RIK/6/2024,

REFERENCES

Alisneaky, C. B.-S. 4. . (2019). *Intuitively understanding SVM and SVR*. Medium. https://www.machinecurve.com/index.php/2019/09/20/intuitively-understanding-svm-and-svr/#how-does-support-vector-regression-work

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., & Haase, A. (2019). Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, *15*(March), 1–34. https://doi.org/10.1016/j.impact.2019.100179

Beliavsky, G., Danilova, N., & Yao, K. (2023). Principal component analysis and optimal protfolio. *Journal of Mathematical Sciences*, *38*(4), 368–377. https://doi.org/10.1093/llc/fqad054

Chourib, I., Guillard, G., Farah, I. R., & Solaiman, B. (2022). Stroke Treatment Prediction Using Features Selection Methods and Machine Learning Classifiers. *Iirbm*, *43*(6), 678–686. https://doi.org/10.1016/j.irbm.2022.02.002

Dabića, M., Jane Frances, Maleyd Jadranka, Š., & Počekg, J. (2023). Future of Digital Work: Challenges for Sustainable Human Resources Management. *Journal of Innovation & Knowledge*, *8*(2).

de Góes Maciel, F., O'Rourke, S., Jones, M., Hemstrom, W., Miller, M. R., Schmaedecke, G., Tambosi, L. R., Pires Baptista, M. S., Keuroghlian, A., Nava, A. F. D., Nardi, M. S., de Almeida Jácomo, A. T., Silveira, L., Furtado, M. M., Tôrres, N. M., & Biondo, C. (2024). Loss of genetic diversity and isolation by distance and by environment in populations of a keystone ungulate species. *Conservation Genetics*, 1–24. https://doi.org/10.1007/s10592-024-01614-w

Dugger, Z., Halverson, G., McCrory, B., & Claudio, D. (2022). Principal Component Analysis in MCDM: An exercise in pilot selection. *Expert Systems with Applications*, *188*(February). https://doi.org/10.1016/j.eswa.2021.115984

Escanilla, N. S., Hellerstein, L., Kleiman, R., Zhaobin, K., James, D., & Shull†, D. P. (2018). Recursive Feature Elimination by Sensitivity Testing. *Proc Int Conf Mach Learn Appl.*, 40–47. https://doi.org/10.1109/ICMLA.2018.00014.Recursive

Hamada, J. J. T. M., Hassan, M., Kakudi, H., & Abiodun, J. O. (2022). A Machine Learning Method for Classification of Cervical Cancer. *Electronics*, *11*(2), 463. https://doi.org/https://doi.org/10.3390/electronics11030463

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202

Lee, C. H., Trappey, A. J. C., Liu, C. L., Mo, J. P. T., & Desouza, K. C. (2022). Design and management of digital transformations for value creation. *Advanced Engineering Informatics*, *52*(April), 3–5. https://doi.org/10.1016/j.aei.2022.101547

Luo, W. (2019). User choice of interactive data visualization format: The effects of cognitive style and spatial ability. *Decision Support Systems*, *122*(July), 1–9. https://doi.org/10.1016/j.dss.2019.05.001

Ng, S. C. (2017). Principal component analysis to reduce dimension on digital image. *Procedia Computer Science*, *111*, 113–119. https://doi.org/10.1016/j.procs.2017.06.017

Parhusip, H. A., Trihandaru, S., Heriadi, A. H., Santosa, P. P., & Puspasari, M. D. (2022). Data Exploration Using Tableau and Principal Component Analysis. *International Journal on Informatics*

Visualization, *6*(4), 911–920. https://doi.org/10.30630/joiv.6.4.952

Peterson, L. E., & Coleman, M. A. (2006). Comparison of gene identification based on artificial neural network pre-processing with k-means cluster and principal component analysis. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3849 LNAI*(September), 267–276. https://doi.org/10.1007/11676935_33

Qiu, Z., Wang, S., Hou, Y., & Xu, S. (2023). What Drives Infrastructure Participants to Adopt Digital Technology: A Nexus of Internal and External Factors. *Sustainability*, *15*(23), 16229. https://doi.org/10.3390/su152316229

Schade, P., & Schuhmacher, M. C. (2022). Digital infrastructure and entrepreneurial action-formation: A multilevel study. *Journal of Business Venturing*, *37*(5), 1–6. https://doi.org/10.1016/j.jbusvent.2022.106232

Stodola, P., & Stodola, J. (2020). Model of predictive maintenance of machines and equipment. *Applied Sciences (Switzerland)*, *10*(1). https://doi.org/10.3390/app10010213

Systems, C. (2022). Internet of Things and Cyber-Physical Systems Sustainability 4 . 0 and its applications in the field of manufacturing. *Internet of Things and Cyber-Physical Systems*, *2*, 82–90. https://doi.org/https://doi.org/10.1016/j.iotcps.2022.06.001

Zhang, J., Cao, G., Peng, Q., Tan, R., Liu, W., & Zhang, H. (2022). A systematic knowledge-based method for design of transformable product. *Advanced Engineering Informatics*, *52*(April), 2022–2023. https://doi.org/10.1016/j.aei.2022.101638

Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers and Industrial Engineering*, *150*(December), 1–7. https://doi.org/10.1016/j.cie.2020.106889