JOIV : Int. J. Inform. Visualization, 6(4) - December 2022 911-920

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Data Exploration Using Tableau and Principal Component Analysis

Hanna Arini Parhusip [a,*], Suryasatriya Trihandaru [a], Adrianus Herry Heriadi [a,b], Petrus Priyo Santosa [a,b], Magdalena Dwi Puspasari [a]

[a] *Master of Data Science, Universitas Kristen Satya Wacana, Salatiga, Indonesia*
[b] *PT APSI Jakarta, Indonesia*

*Corresponding author: *hanna.parhusip@uksw.edu*

Abstract*—This study aims 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 method used is usually Scheduled Oil Sampling. This article proposes a new approach. The research problems are analyzing the recorded chemical elements that are produced by heavy engines and visualizing them through the Tableau program. The basic design of the study is learning the given data after visualization and using the Principal Component Analysis. This method is to obtain chemical elements that affect engine wear during each engine's use in the 2015-2021 period. Because there are three categories in each element in the oil sample, namely wear metals, contaminants, and oil additives, a technique is needed to obtain these elements using Principal Component Analysis. Therefore, Oil Sampling Analysis through data exploration using Tableau resulted in a new approach to data analysis of elements recorded by heavy vehicles. The main findings as a result of the analysis are given by the visualization of Tableau, in which there are five machines analyzed to obtain the main components that cause engine wear. From the visualization results, it is shown that there is one engine coded MSD 012 that experienced wear and tear in 2018 and 2019. This shows where two main components, Ca and Mg, dominate engine wear. These results have been confirmed with the related companies. The company then carried out further studies on the machine to get special treatment because of these results.

***Keywords*— Oil; heavy equipment; Tableau; principal component analysis.**

Manuscript received 7 Jun. 2022; revised 30 Oct. 2022; accepted 8 Nov. 2022. Date of publication 31 Dec. 2022.

I. INTRODUCTION

Oil is one of the fluids that are often found in engines and serves as a lubricant to protect engine components from friction and rust [1]. Oil is an important component of maintaining the engine's health, so choosing the type and brand of used oil for the engine must be done carefully and thoroughly. In addition to maintaining engine health, oil can also be used as an indicator to determine the health condition of an engine through analysis of the oil sample used by the engine, such as in a power plant, and planning and repairing equipment/facilities to acceptable standards. Using the oil sample, conclusions can be drawn about the health condition of the oil engine. Over time the use of oil, the viscosity of the oil will decrease (become more liquid). In this condition, the oil will be contaminated with elements that mix with the oil [2]. Therefore, there is a standard procedure called Scheduled Oil Sampling to overcome the problem above.

Scheduled Oil Sampling is an activity carried out to detect wear and tear or damage to a heavy equipment machine [3] which engineers already know in related fields. The process to analyze oil samples from heavy equipment machines is based on the date of oil sampling, hours of operation, type of used oil, quality of used oil, time of used oil, oil addition, oil change, and filter replacement. The waste oil should also be maintained properly due to its impact on the environment where serious efforts undertaken by researchers toward the management of oil wastes employ different chemical and physicochemical methods. Through oil sample analysis, it is known the elements contained in oil samples from heavy equipment engines. The elements contained in the oil sample are divided into 3 categories: Metals, Contaminants, and Additives. Based on these three categories, the elements contained in the oil sample are separated based on the nature and type of these elements which can be considered crude oils. To check the wear and possible damage that will occur is done by analyzing the value of the concentration of the elements in the oil sample for the life cycling of the machine which may arise during the combustion engine[1]. If there is an increase in the concentration value of the elements that exceed the normal limit that has been mutually agreed, then there will be wear and tear on certain parts of the heavy equipment engine during the oil use period.

In this study, the scheduled oil sampling data will be visualized using the Tableau program, which the company does not yet do. Additionally, the company wanted to know whether the researchers could give valuable information by visualization without understanding the technology that developed the data. The visualization aims to see whether an increased concentration value exceeds a certain limit. The elements in question are wear elements, namely Iron (Fe), Copper (Cu), Chromium (Cr), Aluminum (Al), Lead (Pb), Tin (Sn), and Nickel (Ni) which are considered pollutants. Wear and tear on certain parts of the engine can be an indicator to predict the damage that will occur to the engine, particularly in heavy machines.

This article proposes the Principal Component Analysis (PCA) method using the R program to find the correlation between the elements in the oil sample and oil changes performed on heavy equipment engines. Thus, the elements that affect the oil change of heavy equipment engines are known. The novelty of this research is indicated when data are explored using Tableau without knowing the clear background of the data to get insight more freely information governed by the data, and the Principal Component Analysis (PCA) is used to do further analysis. The result gives a new perspective to the study related to Scheduled Oil Sampling which has its procedure to do the analysis.

II. Material and Method

A. Data source

PT Artha Puncak Semesta Indonesia obtained the Oil Sampling data from 2015 to 2021 (PT APSI). The data are data from the analysis of the content in the oil sample from heavy equipment, i.e., elements in the oil sample, time to use oil, type of used oil, quality of used oil, hours of operation of heavy equipment, and unit number of heavy equipment. Scheduled oil sampling data is used to check engine wear and predict failure that will occur. Component wear can be determined by examining chemical elements and measuring the amount of wear on metals such as Iron (Fe), Copper (Cu), Chromium (Cr), Aluminum (Al), Lead (Pb), Tin (Sn), and Nickel (Ni) which usually are known wear and some of 12 elements [2],[3]. The increase in the concentration of wear elements becomes an indicator of wear in certain parts of the engine. Table 1 shows three categories of elements from Oil Data Sampling, namely Wear metals, Contaminants, Oil additives, and the location of each element in the engine. The complete description of these is shown in Table 1.

TABLE I
DATA 3 CATEGORY OIL SAMPLING (CONTRIBUTED FROM PT APSI JAKARTA)

Element	Symbol	Metals	Contaminant	Additive	Found in
Iron	Fe	X	X		Gears, roller bearings, cylinder/liners, shafts
Copper	Cu	X	X	X	Brass/bronze bushes, gears, thrust washers, oil cooler cores, internal coolant leaks
Chromium	Cr	X			Roller Bearings, piston rings
Tin	Sn	X			Bronze bushes, washers, and gears
Aluminum	Al	X	X		Piston, journal bearings, dirt
Lead	Pb	X			Journal bearings, grease, petrol contamination
Silicon	Si		X	X	Dirt, grease, additive
Sodium	Na		X	X	Internal coolant leaks, additives, seawater contamination
Boron	B		X	X	Additive, internal coolant leak, brake fluid contamination
Magnesium	Mg		X	X	Additive, seawater contamination
Zinc	Zn	X		X	Additive (anti-wear)
Phosphorus	P			X	Additive (anti-wear, extreme pressure)
Molybdenum	Mo			X	Piston rings, additive, solid additive (Mo-di)
Nickel	Ni	X			Roller Bearings, camshafts and followers, thrust washers, valve stems, valve guides
Silver	Ag	X			Silver solder, journal bearings (seldom)
Lithium	Li		X		Grease
Sulfur	S			X	Lubricant base stock, additive
Calcium	Ca			X	Detergent Dispersant Additive, Water Contaminant, Airborne Contamination
Potassium	K		X		Coolant Leak, Airborne Contaminant

B. Dataset Parameter (Attribute) Information

Table 2 indicates the names of the attributes in the data and an explanation of the names of these attributes.' The attribute information is listed and described based on the attribute's name, which is shown in Table 2 below.

TABLE II

THE ATTRIBUTES IN THE DATA AND THE DESCRIPTION FOR EACH ATTRIBUTE

Attribute Info	Description
unitno	Number of units of Heavy Equipment (HE)
Serial no	Serial number HE or backend number HE
cust_name	Customer name /owner HE
model	Type of vehicle (TRUCK2 = Dump Truck)
compart	Compartment/Part of vehicle HE ex. Engine
oiltypeid	Type oil (DEO,CAT DEO, CALTEX, PERTAMINA dll)
Oil graded	Grade oil Id (SAE 15W-40,20W,30W,40W, dll)
oilsampling_desc	Description sampling process
Sample date	Date, Month, and year took the sampling oil
Oil added	Oil which is added
oilhours_erp	Time on Oil ERP
actual_hours	Actual time oil
SeviceMeterUnit	SMU: operational hour HE recorded by Equipment Service Meter (hour)
Oil replacement	Oil is done to be replaced (FALSE or TRUE)
filter_replacement	The filter is changed (FALSE or TRUE)
Si	Silicon (found in Dirt, grease, and additive)
Al	Aluminum (found in Pistons, journal bearings, and dirt)
Cr	Chromium (found in Roller Bearings, and piston rings)
Fe	Iron (found in Gears, roller bearings, cylinder/liners, and shafts)
Pb	Lead (found in Journal bearings, grease, petrol contamination)
Cu	Copper (found in Brass/bronze bushes, gears, thrust washers, oil cooler cores, and internal coolant leaks)
Sn	Tin (found in Bronze bushes, washers, and gears)
Ni	Nickel(found in Roller Bearings, camshafts, and followers, thrust washers, valve stems, and valve guides)
Na	Sodium (found in Internal coolant leaks, additives, and seawater contamination)
K	Kalium (found in Coolant Leak, Airborne Contaminant)
Mo	Molybdenum(found in Piston rings, additive, solid additive (Mo-di))
Zn	Zinc (found in additive (anti-wear))
Mg	Magnesium (found in Additive, seawater contamination)
Ca	Calcium (found in Detergent Dispersant Additives, Water Contaminants, and Airborne Contamination)
P	Phosphorous (found in Additive (anti-wear, extreme pressure))
B	Boron (found in Additive, internal coolant leak, brake fluid contamination)
ST	Soot
OXI	Oxidation
NIT	Nitration
SUL	SUL=Sulfation
V100	V100=Viscosity 100°C (WIN.MP.05)
TBN	TBN=Total Base Number - ASTM D 4739 (mg KOH /g)

C. Data visualization steps using Tableau

In performing data visualization using the Tableau program, the following steps are carried out:

1) Import the oil sampling data in excel format into the Tableau program.
2) Entering table data into the section provided by the Tableau program.
3) Make sheets for the three categories: Metal, Contaminant, and Additives.
4) Visualize the metal category firstly by entering Fe, Cr, Pb, Cu, Sn, Al, and Ni data into the column section, and Sample date, Oil typeid, and oil graded data into the row section.
5) Perform the same steps as step 4 for the visualization of the Contaminant and Additives categories.
6) The value for numeric data in the column section is the maximum value of the data.
7) The following filters are used for all visualizations, namely
   a) Sample date uses year data with the years used only from 2016 to 2020.
   b) The used units are MSD012, MSD013, MSD014, MSD016, and MSD017. Where a visualization of 3 categories of elements is made for each unit so that the image can be seen clearly.
   c) The used Oil type is DEO.

d) The used Oilgraded is SAE 15W-40

8) Next, the sheets of the three categories of elements based on the used unit are compiled into one dashboard and named according to the unitno.

D. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a technique commonly used to reduce data due to the number of variables being considerably many. PCA is used to select the dominant variable so that many variables can be reduced. Another method, such as Gradient Boosting Decision Tree Algorithm, may also be combined to improve PCA for predicting the main controlling factors of cycle oil production [4]. Additionally, PCA can also be used for discovering anomaly detection. Heavy metal contents and some agrochemical parameters which characterize the soil samples from the contaminated and uncontaminated area were observed by PCA as an example of this case[5]. PCA has been implemented to predict the price of crude oil after dimensional reduction[6]. One observed that the sample size could impact the Eigen structure leading to the resulting reduction [7]. The PCA method simplifies complex data by identifying a small number of principal components (PCs) that have the maximum variance and can represent the entire data set. This simplification process is carried out by making sample projections in two-dimensional or three-dimensional axes (principal components) as a new system of variables [8]. To assess oil and gas production, PCA is one of the useful methods that has already been implemented[9], including the scaled principal component analysis (s-PCA), to improve the oil price predictability with technical indicators [10].

The principal component (PC) is a linear combination of the original variables, and each PC is orthogonal to one other. PC 1 explains more data variation than PC 2, PC 2 explains more data variation than PC 3, and so on. Using several PCs can make it easier for us to interpret the data and find out the distribution of data into several clusters of data that are correlated with each other. Hence, the final result of PCA can be seen in what attributes have the most role in explaining the phenomena that exist in the dataset while maintaining the characteristics of the data (maximum variance). Because PCA is part of unsupervised learning, PCA does not require dependent variables. The first step to performing PCA is to express the data in a dimensionless form by dividing the data by the maximum value of each data row and converting the data into a matrix form. Scaling variables leading to dimensionless form in PCA can be done for oil price predictability with technical indicators. Then we look for the matrix using the average value of the data matrix rows and the deviation matrix. After obtaining the covariance matrix from the data, the eigenvalues and eigenvectors can be determined. The eigenvectors are orthogonal to each other and have a value of 1, so they can be used as a basis for Y_i (principal component) [11]. Finally, the correlation value between Y_i and $X_{k,i}$ is calculated (data variable) to determine the attributes that affect the phenomenon being studied.

III. Results and Discussion

A. Visualization Restrictions Using Tableau

Maintaining heavy equipment needs accurate information about the use and wear of the machines, and therefore dynamic data acquisition techniques are required to assess suitability to achieve both diagnostic and prognostic requirements of condition-based maintenance. It concludes that hydraulic oil contamination analysis, namely the detection of metallic particulates, offers a reliable way to measure the real-time wear of hydraulic components[12]. A more rigorous study on oil scheduling for planning engine management is set up into a nonlinear optimization problem [13]. However, the article here focuses on the Oil Sampling analysis through data exploration using Tableau, leading to a new approach to data analysis of the elements recorded by heavy vehicles. Each element in the oil sample is categorized into three categories: wear metals, contaminants, and oil additives.

The study of critical wear metals has been investigated [14]. In this article, we have limited some components. The following is a list of elements from the three categories:

1) Wear Metal: Fe, Sn, Al, Cu, Pb, and Ni. The last three elements are known as heavy metals [15]. Heavy metal accumulation in high amounts within the body can have severe health implications.
2) Contaminants: Si, Na, and K.
   These contaminant elements must be considered in the Scheduled Oil Sampling due to excessive wear on the components of the vehicle leading to repair with a high cost and unexpected downtime.
3) Metal Additives: Mg, Zn, P, Zn, and Ca.
   The metal additives usually utilized in engine oils act as an anti-wear or friction reduction media [16] where each element has its heat capacity. Therefore, these metals are also required to consider since the presence of these elements may cause an unexpected effect on the engine oil's heat capacity and viscosity.

These categories will give three statuses of the heavy engine, i.e., normal, abnormal, and critical. The visualization carried out on Tableau is made based on the three categories of elements contained in the heavy equipment oil sample. In the visualization, five units of heavy equipment are used as representatives of all data. In addition to the three categories of elements, visualization is also carried out regarding the operating hours of the heavy equipment, the time of using oil, and the addition of oil. The data used are in the year 2016 to 2020 because the data therein is complete from January to December for each year.

B. Normal Value for Concentration of Elements

Wear elements such as Iron (Fe), Copper(Cu), Chromium(Cr), Aluminum(Al), lead(Pb), Tin(Sn), and Nickel(Ni) have different normal limit values. Larger element values are categorized as abnormal and critical values. If there is an increase in abnormal and critical values, one concludes that there is wear on certain parts of the heavy equipment machine. Table 1 shows all wear elements' normal, abnormal, and critical values.

TABLE III
TABLE OF LIMIT VALUES FOR EACH WEAR ELEMENT (CONTRIBUTED FROM PT APSI JAKARTA)

		Normal	Abnormal	Critical
Fe	Iron	<100	100 – 200	>200
Pb	Lead	<30	30 – 75	>75
Cu	Copper	<30	30 – 75	>75
Cr	Chromium	<10	10 – 25	>25
Al	Aluminum	<20	20 – 30	>30
Ni	Nickel	<10	10 – 20	>20
Ag	Silver	<3	3 – 15	>15
Sn	Tin	<20	20 – 30	>30
Na	Sodium	<50	50 – 200	>200
Si	Silicon	<20	20 – 50	>50
Fuel Dilution%		<2	2 – 6	>6
Soot%		<2	2 – 6	>6

Note: Boron, Potassium, Phosphorus, Zinc, Calcium, Barium, Magnesium, Titanium, Molybdenum, and Cadmium have no established limits. Compare with new oil reference sample and plot trends.

Note: This table should be used as a *general guide only*. it is subject to continual change, and *should not be considered current official OEM limits, guidelines, or recommendations of any kind*, as it is derived from many different sources, OEMs do not always make their most current, official limits readily available to commercial oil analysis labs. Effective date 2/15/01.

C. Analysis of Visualization Results Using Tableau

Using the Tableau program, the following graphic images are generated. The pictures below show a dashboard image in the Tableau program for each heavy equipment unit number: MSD 012, MSD 013, MSD 014, MSD 016, and MSD 017. The resulting pictures consider the normal concentration values for elements that are included in the wear element. Note that the increase in the concentration of elements that have a value greater than normally indicates wear and tear in certain parts of the heavy equipment machine. The increase in value is the increase in the value of the concentration of elements that are included in the critical condition. Meanwhile, for the value of element concentration, which is in an abnormal condition, it is still acceptable and concluded that there had not been any wear and tear on certain parts of the heavy equipment machine.

The first analysis was carried out on the results of visualizing the value of wear elements contained in oil samples from heavy equipment engines with unit number MSD 012 shown in Figure 1. The concentration values of Fe, Pb, Cr, Al, Ni, Sn, Na, and Si elements had normal values (below the values for the normal limit). However, the Cu element has a Critical element concentration value, which has increased from 2018 to 2019. It can be concluded that there is wear on certain parts of the heavy equipment machine with the unit number MSD 012. Note that Copper (Cu), as an essential metal, plays a crucial role in biochemical reactions and physiological regulations [17].

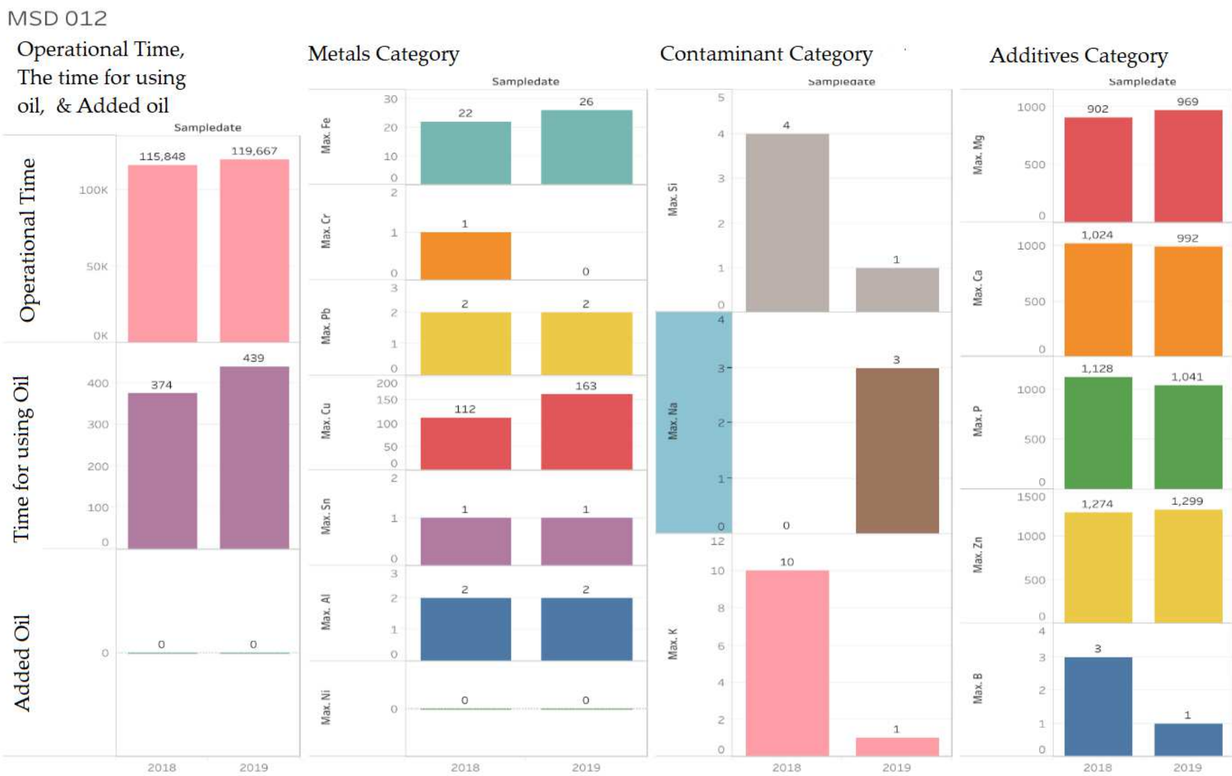

Fig. 1 Tableau visualization for equipment MSD 012 for attributes metal category, contaminant category, additives category

The second analysis was carried out on the results of visualizing the value of the wear element contained in the oil sample from the heavy equipment engine with the unit number MSD 013. Na and Si are in the normal value category. Meanwhile, the value of the Cu element has a value greater than the normal limit value and is included in the category of abnormal value. However, because there is no increase in value, the condition is still accepted and based on these conditions, it is concluded that there has not been any wear and tear on the machine with unit number MSD 013 as depicted in Figure 2.

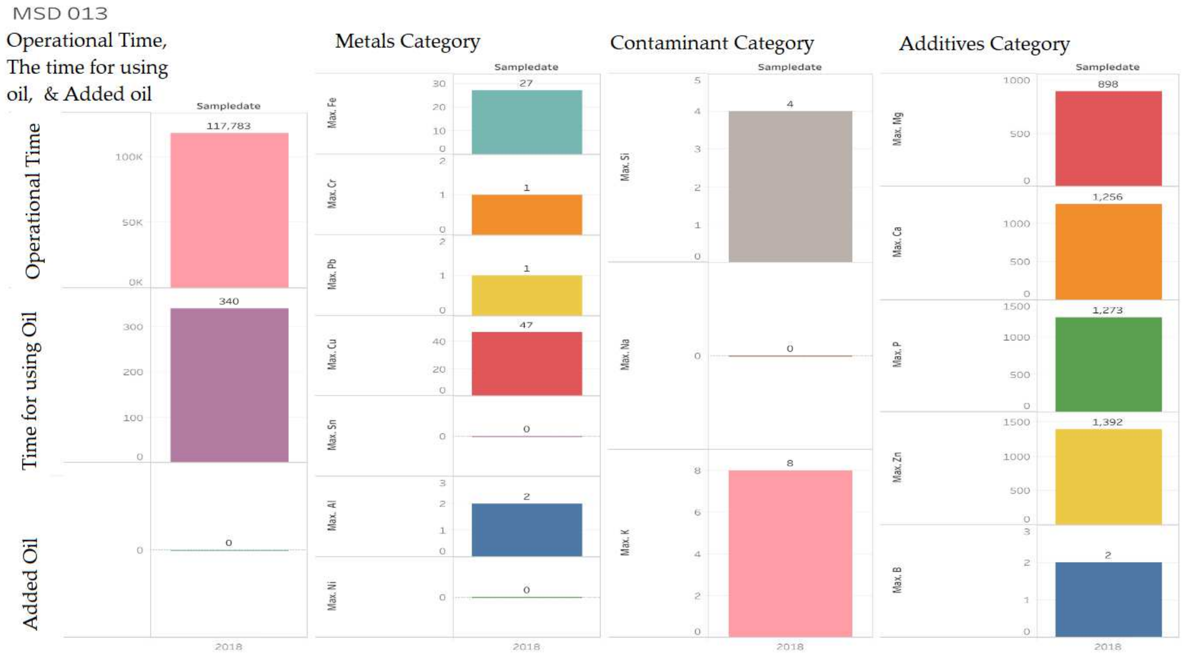

Fig. 2 Tableau visualization for equipment MSD 013 for attributes metal category, contaminant category, additives category

Further analysis was carried out on the results of visualizing the value of the wear elements contained in the oil sample from the heavy equipment engine with the unit number MSD 014. By paying attention to the graph of the data visualization results using the Tableau program, as shown in Figure 4, it was found that all wear elements of the heavy equipment engine oil sample MSD 014 were of normal value. There is no wear and tear on heavy equipment machines with unit number MSD 014.

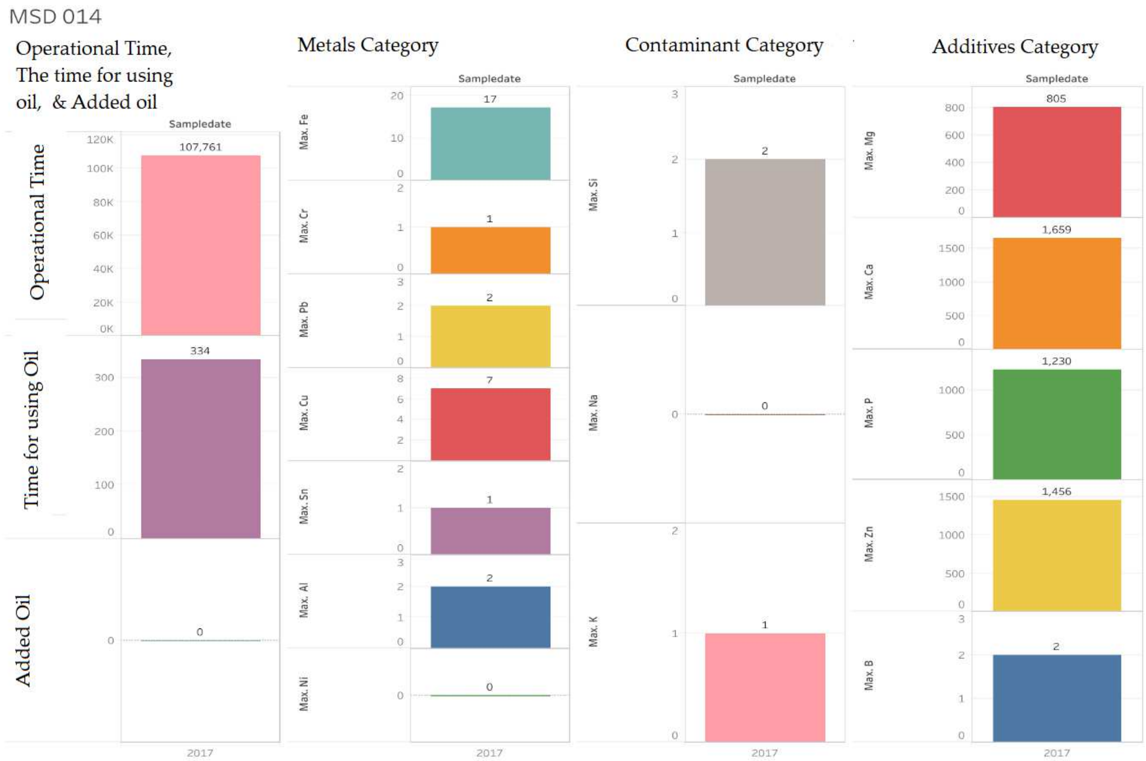

Fig. 3 Tableau visualization for equipment ID MSD 014 for attributes metal category, contaminant category, additives category

The next analysis was carried out on the results of visualizing the value of the wear element contained in the oil sample on the heavy equipment engine with the unit number MSD 016. In the MSD 016 engine oil sample, it was also found that the values of all wear elements were within normal values. We conclude that there is no wear and tear on the MSD 016 heavy equipment machine, and one may observe this in Figure 4 below.

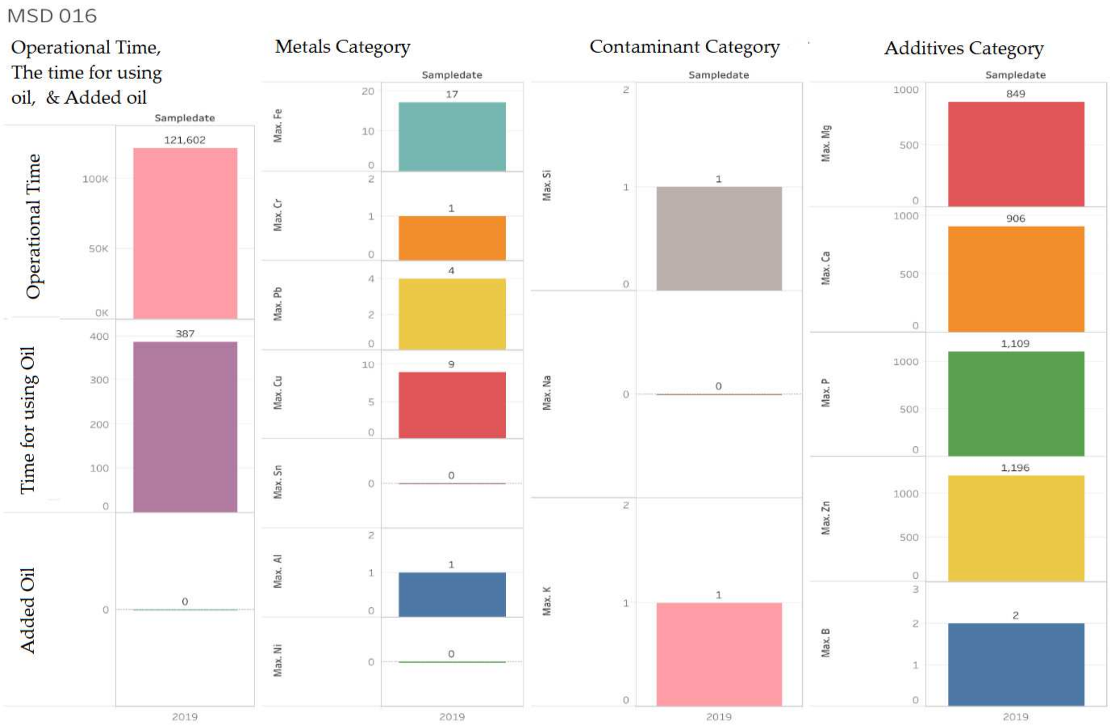

Fig. 4 Tableau visualization for equipment MSD 016 for attributes metal category, contaminant category, additives category

The last analysis was carried out on the results of visualizing the value of the wear element of the oil sample from the heavy equipment engine with the unit number MSD 017 as shown in Figure 5. Based on the visualization results using the Tableau program, it was found that all the wear element values in 2017 were within normal limits, while for 2018 all the elements were within normal limits, except for the Cu element. There was an increase in the Cu concentration value from 2017 to 2018, wherein in 2018, the Cu element value was above the normal (abnormal) limit. However, this condition is still acceptable because it is not classified as a critical value. Hence, there has not been any wear and tear on certain parts of the heavy equipment machine with unit number MSD 017.

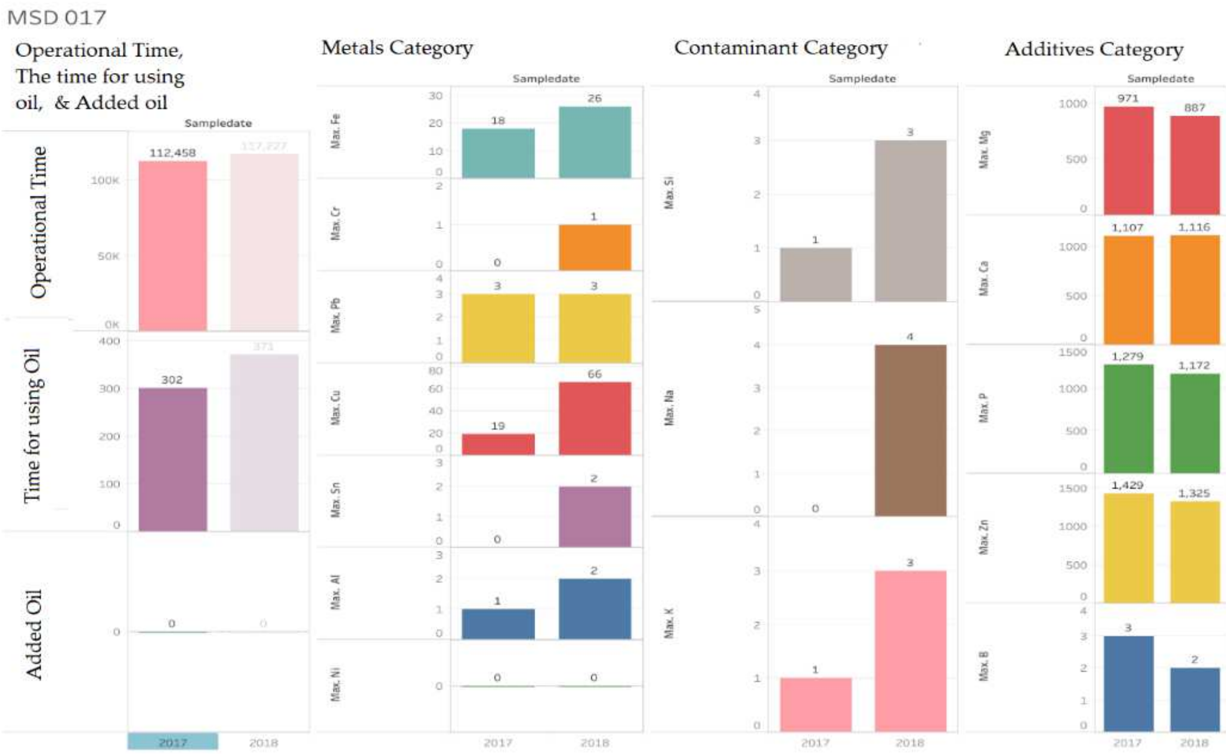

Fig. 5 Tableau visualization for equipment MSD 017 for attributes metal category, contaminant category, and additives category.

Thus, all tested units in the tested period have no significant wear and tear elements. We will investigate the analysis based on PCA.

D. Discussion

In the first exploration data, we observe behavior data with the help of Tableau. The PCA method is applied using the R program and produces a scree plot which is shown in Figure 6. The scree plot in Figure 6 is a plot of the eigenvalues of the main components to be extracted. The plot shows the relationship between the eigenvalues and the main component to be extracted with the eigenvalues as the *y*-axis (vertical direction) and the main component as the *x*-axis (horizontal direction). Many of the principal components are determined by the slope of the eigenvalue plot. When the scree plot becomes horizontal and the eigenvalues are less than one, the principal component extraction process stops. The percentage of PC1 is greater than PC2 because PC 1 explains more data variation than PC2. PC 2 has a larger percentage than PC3 because PC 2 explains more data variations than PC 3. Similarly, the same goes for the next PC, namely PC3 with PC4, PC4 with PC5, and so on.

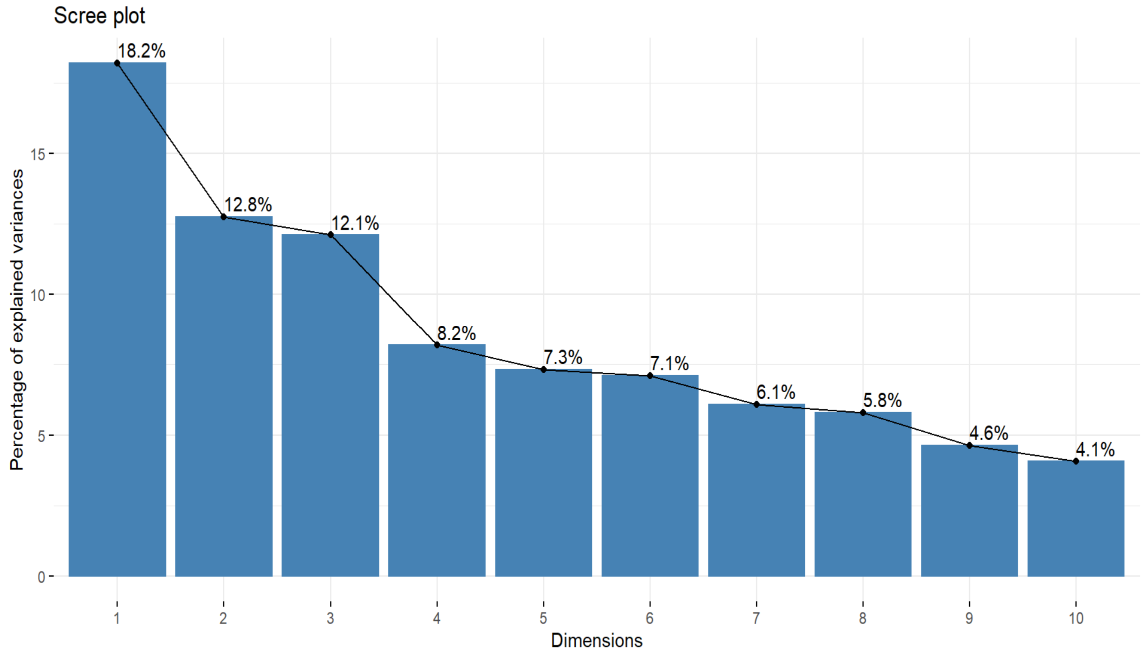

Fig. 6 The final result of the PCA

E. Principal Component Analysis (PCA) Final Results

With the value of the components obtained previously, the correlation between the main component and the variables from the original data can be determined. A correlation between 2 attributes is strong if the coefficient of correlation is close to 1, indicating positively correlated and close to -1 indicating negatively correlated.

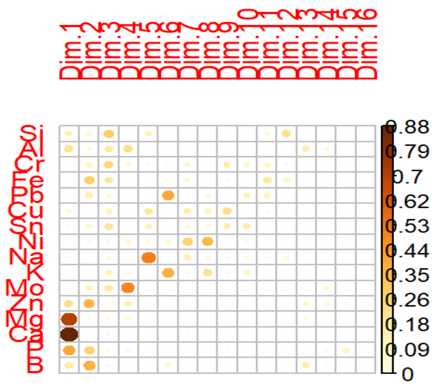

Fig. 7 Correlation plot among features

The final PCA result is obtained by knowing the variables that affect oil replacement. The element of Ca has the largest correlation value among other variables with PC 1, variable B has the largest correlation value among other variables with PC 2, and the correlation value between the main components and other original data variables is shown in the correlation plot image Figure 7.

Because the data has 16 dimensions, it is necessary to use a PCA biplot to see the observation points and the distribution pattern of the real data. The observation points that are close to each other are observation points with similar characteristics. Similarly, the observation points far apart from the collection of observations (clusters) show that the observation points have different characteristics from the data set. Identifying the correlation between the variable and the principal component is done by looking at the length of the vector (arrow) concerning the axis of the biplot (principal component). The vector length shows the variable's correlation value with the main component. The correlation value is greater when the vector of the variable is getting longer. In addition to the correlation between variables and the main component, the direction of the vector shows the correlation between data variables. If the two variables have nearly the same vector direction, then the two variables are positively correlated. Meanwhile, the two variables are negatively correlated if the two variables' vector directions are opposite. The smaller angle between the two vectors shows that the two variables have a strong and positive correlation.

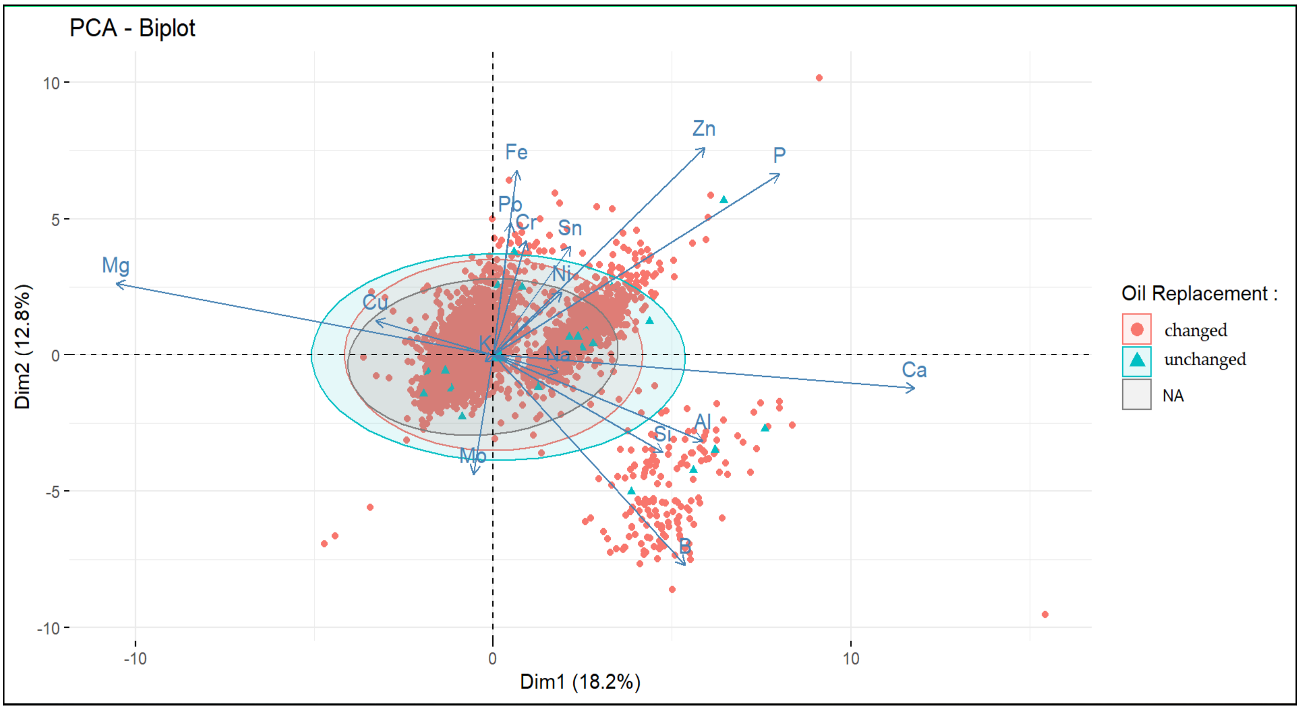

Fig. 8 PCA Biplot

Based on the biplot Figure 8, there are 3 data clusters based on oil changed, unchanged, and NA (meaning unidentified). Oil changes are influenced by the variables Ca and Mg, which have the highest correlation value with PC1. Of the 3 existing data clusters, the cluster was not converted into a cluster with the most observation points. Four groups of variables are positively correlated, which are indicated with the same direction of the vectors, i.e., with the first group being Fe, Pb, Cr, Sn, Ni, and Zn, the second group being Ca, Al, B, and Na, the third group being K and Mo, and the last group being Mg and Cu. Where the first group is negatively correlated with the third group and the second group is negatively correlated with the fourth group. Variable pairs Fe with Pb, Ni with Zn, Cu with Mg, and Na with Al are two variables that are strongly and positively correlated by looking at the angles of the two vector variables where the angle between every two elements is exceedingly small. Shortly, we will relate those results with the viscosity of oil as has been done by other literature.

There are much more data that are available provided by the company for doing the analysis. The analysis here has not considered the different periods for the use of the heavy engine. Other authors observed that the heavy metals would increase related to the time for operating the heavy auto mechanic machines [18]. Unlike the data used here, the engines were operated in the same period. The given data are considerably big, so we will apply machine learning and artificial intelligence. This idea refers to some authors who have implemented these in tribology strongly related to lubricants on heavy equipment with friction [19]. The process shown in this article is one of the entire processes to be termed condition-based maintenance (CBM)[20]. Further research will be done on other factors in the maintenance of heavy engines. One idea refers to Engine Health Management (EHM) for aircraft as one heavy engine where oil Monitoring is also stated as an important factor. Sensor Technology, Data Management Technology, Algorithms, Fault Diagnosis, and Prediction Technology are several important factors in EHM [21] of the monitored heavy engines that are not yet studied in this research.

IV. Conclusion

This article aims to show the analysis of data obtained from a company in the form of recording various chemical components that arise from the activities of 5 machines operating at the company's heavy equipment in the 2015-2021 period. In general, a standard procedure called Scheduled Oil Sampling is used. However, this article demonstrates data visualization using Tableau to recognize data patterns. In addition, the dominant element is sought using Principal Component Analysis because this dominant element affects engine performance. This provides a new step in the process commonly used to evaluate operating machines' wear. Data exploration with this visualization provides convenience in the implementation of engine performance evaluation. As an extension of the results of this research, the company can use this process to carry out further assessments on other heavy machinery in the company and at different periods regularly to be able to take action on heavy equipment machines that are experiencing wear and tear. The wear machine is indicated by the presence of Ca and Mg components that dominate the number of chemical components recorded.

Acknowledgment

We thank PT APSI Jakarta for permits to use their data for this article. The article is also supported by the Ministry of Research and Higher Education for the grant PPS PTM year 2022 No. 0267/E5/AK.04/2022 and the grant of Matching Fund 2022 entitled CoE-AIOT Mining Beta (shorten). This article is the collaborative research between PT APSI Jakarta and the Master of Data Science under the AIOT Laboratory in UKSW.

REFERENCES

[1] V. Kumar, S. K. Sinha, and A. K. Agarwal, "Tribological Studies of an Internal Combustion Engine," *J. Tribol.*, vol. 141, no. 3, pp. 3–13, 2019, doi: 10.1115/1.4041762.

[2] N. Y., M. A.B, B. B.U, A. M, S. M., and S. Y.M, "Investigation of the Extent of Wear Metals in Five Different Lubricating Oils Before and After Exposure to Engine Stress," *IOSR J. Appl. Chem.*, vol. 9, no. 8, pp. 75–78, 2016, doi: 10.9790/5736-0908017578.

[3] S. A. Latip, S. Kasolang, S. K. Alias, S. Yunus, A. H. Abdullah, and N. Jenal, "Wear elemental spectrometric quantitative analysis of used perodua automatic transmission fluid-3 series (ATF-3)," in *Procedia Engineering*, 2013, vol. 68, no. August 2015, pp. 193–198, doi: 10.1016/j.proeng.2013.12.167.

[4] H. Liu, J. Gu, Y. Wang, and Z. Wei, "Prediction method of heavy oil horizontal well cycle oil production based on PCA and gradient boosting decision tree," in *2021 3rd International Conference on Intelligent Control, Measurement and Signal Processing and Intelligent Oil Field, ICMSP 2021*, 2021, pp. 276–280, doi: 10.1109/ICMSP53480.2021.9513392.

[5] M. Noura, "Interoperability in Internet of Things : Taxonomies and Open Challenges," pp. 796–809, 2019.

[6] M. He, Y. Zhang, D. Wen, and Y. Wang, "Forecasting crude oil prices: A scaled PCA approach," *Energy Econ.*, vol. 97, no. May, pp. 4–7, 2021, doi: 10.1016/j.eneco.2021.105189.

[7] S. S. Shaukat, T. A. Rao, and M. A. Khan, "Impact of sample size on principal component analysis ordination of an environmental data set: Effects on eigenstructure," *Ekol. Bratislava*, vol. 35, no. 2, pp. 173–190, 2016, doi: 10.1515/eko-2016-0014.

[8] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, doi: 10.1098/rsta.2015.0202.

[9] G. R. Igtisamova, N. N. Soloviev, F. A. Ikhsanova, D. S. Nosirov, and A. A. Abdulmanov, "Principal component analysis for assessing oil and gas production (the case of the Kogalym field)," in *IOP Conference Series: Earth and Environmental Science*, 2019, vol. 378, no. 1, doi: 10.1088/1755-1315/378/1/012113.

[10] M. Rhanoui, S. Yousfi, M. Mikram, and H. Merizak, "Forecasting financial budget time series: Arima random walk vs lstm neural network," *IAES Int. J. Artif. Intell.*, vol. 8, no. 4, pp. 317–327, 2019, doi: 10.11591/ijai.v8.i4.pp317-327.

[11] S. P. Mishra, S. Uttam, T. S. D. Subhash, R. Saikhom, Devi Prasanna Swain, S. Panda, and M. Laishram, "Principal components analysis," *Int. J. Livest. Res.*, vol. 12, no. 6, pp. 333–338, 2017, doi: 10.5455/ijlr.20170415115235.

[12] F. Ng, J. A. Harding, and J. Glass, "Improving hydraulic excavator performance through in line hydraulic oil contamination monitoring," *Mech. Syst. Signal Process.*, vol. 83, pp. 176–193, 2017, doi: 10.1016/j.ymssp.2016.06.006.

[13] H. Yang *et al.*, "Integration of crude-oil scheduling and refinery planning by Lagrangean Decomposition," *Comput. Chem. Eng.*, vol. 138, p. 106812, 2020, doi: 10.1016/j.compchemeng.2020.106812.

[14] A. Devaraju, "A critical review on different types of accident morbidity studies," *Courrier*, vol. 6, no. 11, pp. 77–83, 2015.

[15] A. M. Freije, "Heavy metal, trace element and petroleum hydrocarbon pollution in the Arabian Gulf: Review," *J. Assoc. Arab Univ. Basic Appl. Sci.*, vol. 17, pp. 90–100, 2015, doi: 10.1016/j.jaubas.2014.02.001.

[16] A. Roslan, A. S. Ibrahem, and A. Hadi, "Metal additives composition and its effect on lubricant characteristic," in *AIP Conference Proceedings*, 2016, vol. 1774, no. October 2016, doi: 10.1063/1.4965083.

[17] S. Catalani *et al.*, "Free copper in serum: An analytical challenge and its possible applications," *J. Trace Elem. Med. Biol.*, vol. 45, no. November 2017, pp. 176–180, 2018, doi: 10.1016/j.jtemb.2017.11.006.

[18] C. O. Ikese, P. A. Adie, C. Adah, R. Amokaha, G. Abu, and T. Yager, "Heavy metal levels in spent engine oils and fingernails of auto-mechanics," *Ovidius Univ. Ann. Chem.*, vol. 32, no. 1, pp. 28–32, 2021, doi: 10.2478/auoc-2021-0004.

[19] M. Marian and S. Tremmel, "Current trends and applications of machine learning in tribology—a review," *Lubricants*, vol. 9, no. 9, 2021, doi: 10.3390/LUBRICANTS9090086.

[20] A. Ali and A. Abdelhadi, "Condition-Based Monitoring and Maintenance: State of the Art Review," *Appl. Sci.*, vol. 12, no. 2, 2022, doi: 10.3390/app12020688.

[21] W. Abousada, "Design Technology Research of Aircraft Engine Health Management (EHM) Technologies," *Adv. Aerosp. Sci. Technol.*, vol. 06, no. 01, pp. 9–23, 2021, doi: 10.4236/aast.2021.61002.