International Journal of
Innovation in Mechanical Engineering & Advanced Materials (IJIMEAM)
https://publikasi.mercubuana.ac.id/index.php/ijimeam

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

THE IMPLEMENTATION OF PANDAS PROFILING AS A TOOL FOR ANALYZING MECHANICAL PROPERTIES DATA OF NICKEL-BASED SUPERALLOYS BASED ON ALLOY CHEMICAL COMPOSITION

D. Leni[1], Y. P. Kusuma[2], Muchlisinalahuddin[1], R. Sumiati[3], and H. C. Mayana[3]*

[1] Department of Mechanical Engineering, Faculty of Engineering, Universitas Muhammadiyah Sumatera Barat, Padang 25586, INDONESIA
[2] Department of Mechanical Engineering and Biosystem, Faculty of Agricultural Technology, Universitas Andalas, Padang 25175, INDONESIA
[3] Department of Mechanical Engineering, Politeknik Negeri Padang, Padang 25164, INDONESIA

Abstract

The purpose of this study is to evaluate the mechanical properties of nickel-based superalloys with variations in alloy chemical compositions using the Exploratory Data Analysis (EDA) method with the assistance of the pandas profiling library on Google Colab. In this study, data from 312 tensile tests of nickel-based superalloys were used as research samples, with alloy chemical compositions including carbon (C), manganese (Mn), silicon (Si), chromium (Cr), nickel (Ni), molybdenum (Mo), vanadium (V), nitrogen (N), niobium (Nb), cobalt (Co), tungsten (W), aluminum (Al), and titanium (Ti), as well as mechanical properties such as yield strength (YS), tensile strength (TS), and elongation (EL). The methodology used in this study was the EDA method with the assistance of the pandas profiling library on Google Colab, which enables the automatic creation of a dataset report, presenting information on various aspects such as data structure, descriptive statistics, correlation, distribution, and missing values. The results show that yield strength has a fairly high correlation with titanium (0.51), medium correlations with nickel (0.25), vanadium (0.2), and cobalt (0.2). Tensile strength in nickel-based superalloys has a fairly high correlation with yield strength (0.88), carbon (0.49), and cobalt (0.55), and medium correlations with titanium (0.25) and vanadium (0.25). Elongation in nickel-based superalloys has a negative and fairly high correlation with tensile strength (-0.62) and yield strength (-0.58). Some warnings for missing data and zero values in some variables were identified. These results indicate that the pandas profiling library can be used as a tool to analyze the data of mechanical properties of nickel-based superalloys quickly and easily, and provide clear information on data patterns, data structure, and correlation among data.

Keywords: Mechanical Properties, Nickel-Based Superalloys, Exploratory Data Analysis, Pandas Profiling

*Corresponding author: Tel. +62 751 72772
E-mail: yudaperdana.kusuma@gmail.com

DOI: 10.22441/ijimeam.v4i3.19439

1. Introduction

Nickel-based superalloys are alloy metals that consist of 10 or more chemical elements, with nickel being the dominant element among others[1]. This type of metal is commonly used in structural components in the aerospace industry that operate under high-temperature conditions, such as aero-engine blades, turbine disks, and combustion chambers[2]. The application of nickel-based superalloy materials in various industries has prompted researchers in the field of materials to develop their mechanical properties to meet the needs of modern industries[3]. Currently, various variations and types of chemical composition of nickel-based superalloy alloys have been developed to enhance the mechanical properties of materials according to specific requirements, such as the addition of Cr and Al elements for material stability at high temperatures[4]. The abundance of research results on variations in composition and percentage of chemical elements of nickel-based superalloy alloys stored in material databases is expected to facilitate researchers in the future in developing nickel-based superalloy materials according to standards and requirements. Material databases have been developed extensively, such as Open Quantum Material Database, Material Project, Computational Materials Repository, Harvard Clean Energy Project, Anorganic Crystal Structure Database, and Aflowlib, which have been used for computational materials science[5]. However, material databases usually provide only general information about the material, and in other

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

cases, only data is available without clear information. Therefore, to obtain clear information from data, data analysis is necessary. However, performing complex data analysis requires sufficient expertise in data analysis and long processing times. Hence, a fast and easy computational tool that provides clear information on data patterns, data structure, and data correlation is required.

Exploratory data analysis (EDA) is a method used to find patterns, structures and correlations between data plotted in graphical form to make it easier for researchers to understand data. EDA is a field of statistics and data analysis that explores data using descriptive statistics and visualization techniques without making assumptions[6]. It aims to uncover patterns and analyze and model data[7]. EDA is applied in various disciplines for data analysis. Kurzl has applied EDA to analyze geochemical data obtained from Austria's regional river sediment survey[8]. Jansen and Kelkar have used EDA to analyze production data to determine the relationship between wells[9]. Vieira et al. have used EDA to determine the origin of crude oil in the Espírito Santo sedimentary basin located on the southeastern coast of Brazil. Flumignan et al. have shown the application of EDA to determine whether the quality of gasoline from cars in Brazil meets the specifications set by the government[10]. Kumar et al. have extensively analyzed the data collected by their soil sensors via EDA to develop an expert system to predict various fungal diseases in plants[11]. Ogunsina et al. have implemented EDA to analyze historical flight scheduling and operating data to determine the causes of flight schedule disruptions[12]. The rapid use of EDA in various fields of science is inseparable from the era of digitalization, which allows for the digital storage of experimental test results. EDA is mostly done using the Python programming language.

The pandas profiling library is a python module that makes it possible to perform EDA practically and interactively in visualizing data. This study conducted an EDA experiment by utilizing the material database as a dataset to analyze the mechanical properties of nickel-based superalloys based on the chemical composition of the alloy using the pandas profiling library run on Google Colab. Using pandas profiling has several advantages, such as being accessible online using Google Colab, speeding up the analysis process, low cost, and user-friendliness.

2. Experimental and Procedures

Exploratory Data Analysis (EDA) on tensile test results for high alloy steel based on the percentage of alloying chemical elements using the pandas profiling library run with Google Colab. The results of the panda's profiling analysis report from Google Colab were reprocessed to obtain clearer and easier-to-understand information regarding the mechanical properties of nickel-based superalloys. Before evaluating the mechanical properties of nickel-based superalloys, data processing was first performed using the pandas profiling library on Google Colab. The data that has been input is then analyzed to find out data patterns, data structures, and correlations between data using the describe, correlations, and alerts features Fig 1.

Fig 1. Research scheme

2.1 Material Data Preparation

The data used in this study are data on the mechanical properties of nickel-based superalloys, which consist of the percentage of the chemical composition of the alloy consisting of the chemical elements carbon (C), Manganese (Mn), Silicon (Si), Chromium (Cr), Nickel (Ni), Molybdenum (Mo), Vanadium (V), Nitrogen (N), Niobium (Nb), Cobalt (Co), Tungsten (W), Aluminum (Al), and Titanium (Ti) and mechanical properties such as Yield Strength

Table. 1. Dataset statistics

Material (%)	Data Type	Min	Max	Mean
C (%)	Input	0	0.43	0.096
Mn (%)	Input	0.01	3	0.146
Si (%)	Input	0.01	4.75	0.221
Cr (%)	Input	0.01	17.5	8.044
Ni (%)	Input	0.01	21	8.184
Mo (%)	Input	0.02	9.67	2.766
V (%)	Input	0	4.32	0.184
N (%)	Input	0	0.15	0.006
Nb (°C)	Input	0	2.5	0.035
Co	Input	0.01	20.1	7.009
W	Input	0	9.18	0.161
Al	Input	0	1.8	0.239
Ti	Input	0	2.5	0.311
YS (Mpa)	Output	1.005.900	2.510.300	1.420.998
TS (Mpa)	Output	1.019.000	2.570.000	1.641.653

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

(YS), Tensile Strength (TS), and Elongation (EL), these data were obtained[13]. Gareth Conduit created the dataset from the University of Cambridge and Intelligence with a total of 312 data for data which can be seen in Table 1.

2.2 Eda Pandas Profiling

The dataset of nickel-based superalloy tensile test results was analyzed using the pandas profiling library, which runs on Google Colab. The pandas profiling library allows researchers to carry out exploratory data and visualize data easily and practically without having to understand the python programming language; pandas profiling has four main features consisting of[14]:

a. Overview of data structures

This feature can display statistical data and data structure of high alloy steel tensile test results, such as the number of variables, number of data lines, missing values, data size, correlation between variables, and data processing time.

b. Variables

The variable feature provides detailed information about each variable or data column in the dataset, such as the percentage of missing data, the ratio of different values, frequency, mean, and negative data, as well as table information on Quantile statistics, Descriptive statistics, and histogram visualizations.

c. Interaction

This feature serves to visualize the relationship between one variable and another using scatter plots.

d. Correlation

Correlation in pandas profiling is visualized in the form of a Correlation Heatmap, which is a visualization of the strength of the relationship between numerical variables. Correlation plots are used to understand which variables are related to each other.

2.3 Enriched Data and Visualization

The dataset of nickel-based superalloy tensile test results that have been inputted into the pandas profiling library is then analyzed; if there is a lot of missing data, then the Enriched data process is carried out to facilitate the data visualization process. Enriched data or data enrichment enhances existing information and completes missing or incomplete data with relevant context obtained from additional sources[15]. Data visualization is a form of graphic presentation that provides information in data, usually displayed in visual elements such as charts and graphs so that researchers can easily see patterns and trends in the data[16].

3. Results and Discussion

3.1 Result of Eda Pandas Profiling

This research aims to evaluate the mechanical properties of nickel-based superalloys with variations in chemical composition using the Exploratory Data Analysis (EDA) method with the help of the pandas profiling library on Google Colab. In this study, data from 312 nickel-based superalloy tensile tests were used as samples, with alloying chemical elements such as carbon (C), manganese (Mn), silicon (Si), chromium (Cr), nickel (Ni), molybdenum (Mo), vanadium (V), nitrogen (N), niobium (Nb), cobalt (Co), tungsten (W), aluminum (Al), and titanium (Ti), as well as mechanical properties such as yield strength (YS), tensile strength (TS), and elongation (EL). The use of pandas profiling for analyzing the chemical composition data of nickel-based superalloys has several advantages, such as analyzing data quickly, generating reports, and providing information on data distribution, variable correlation, and outliers.

The results of the Data Structure Review show that the Dataset that is uploaded to the pandas profiling library will automatically display the general data set structure, such as the number of variables in the Dataset which totals 17 variables, the number of data rows is 312, there are no duplicate data, the data size capacity is 41.6 KB, and has a data type of numbers, as shown in Fig. 2.

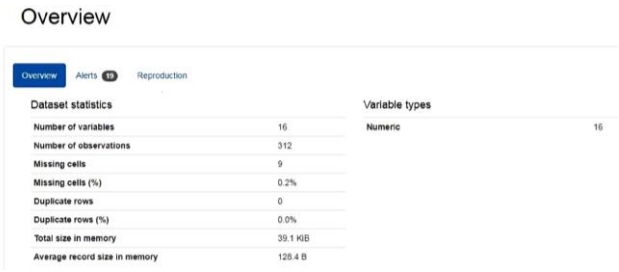

Fig. 2. General view of the dataset

The alerts feature in pandas profiling displays warning messages that appear in reports on data analysis results, such as high correlations between variables, most values containing 0 values, and missing values. The high correlation between variables in the mechanical properties dataset of this nickel-based superalloy includes yield strength (YS), which strongly correlates with the other six variables consisting of C, Ni, V, Co, Ti and TS, then Tensile Strength (TS) has a high correlation. Strong with seven other variables such as C, Ni, Mo, Co, Al, Ti, YS, and Elongation, then elongation has a strong correlation with five other variables such as C, Ni, Co, Ti, and TS. Missing data warnings are also seen in the

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

elongation variable of 9 data, the total value of 0 in variable C is 26 data, and variable N is 269 data, as shown in Fig. 3.

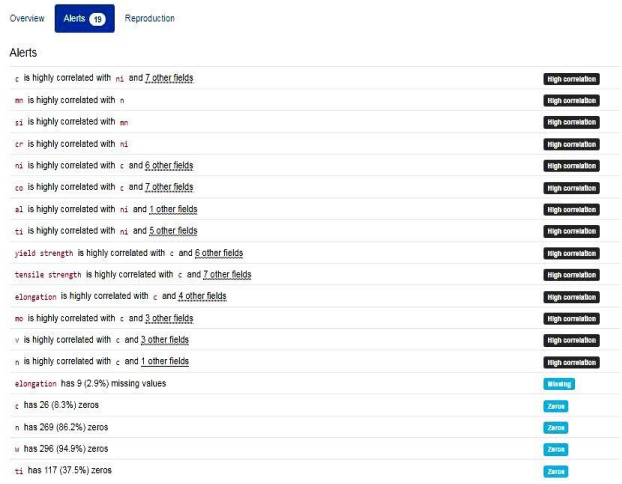

Fig. 3. Display alerts feature

The reproduction feature in pandas profiling displays information on data processing time, data processing time, software version information and configuration downloads, as shown in Fig. 4.

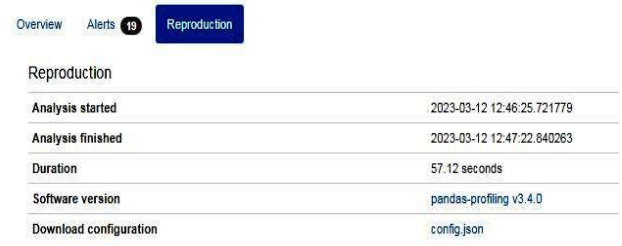

Fig. 4. Reproduction view

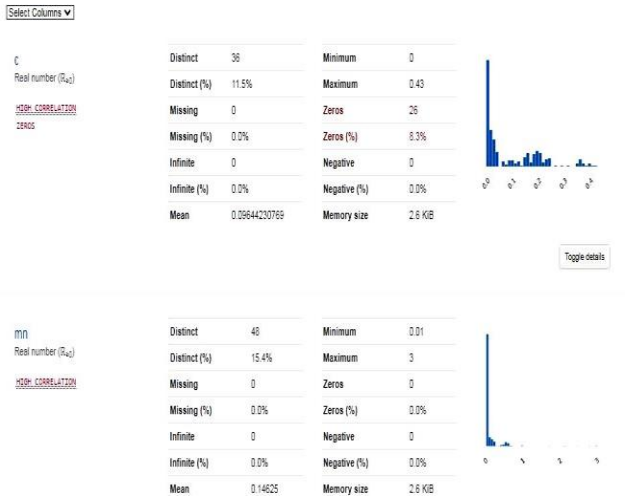

Fig.5. Display of variable features in pandas profiling

The variable feature provides detailed information about each variable or column of data in the dataset, both chemical element data and mechanical property data, information displayed such as percentage of missing data, percentage of different values, frequency, mean, negative data, and Quantile statistics table information, Descriptive statistics, and histogram visualization as shown in Fig. 5 which displays detailed information on the element Carbon (C).

3.2 Result of Enriched Data and Visualization

Interaction features are useful for identifying more complex relationships between variables in the data, such as determining whether there are significant interactions between the selected variables and how these interactions affect the results of the analysis; in pandas profiling, data interactions are visualized using a scatterplot so that it allows researchers to see the interactions of each variable as display in Fig. 6. This feature also allows it to display alloy chemistry that strongly correlates with nickel-based superalloys such as C, Ni, V, Mo, Co, Ti, Al and mechanical properties such as YS, TS and EL.

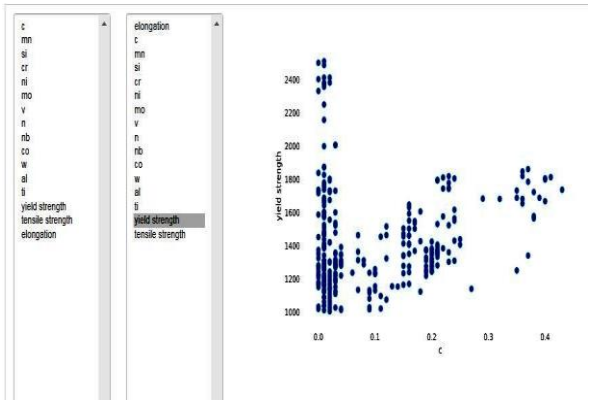

Fig 6. Visualization of interactions between variables

The correlation coefficients in pandas profiling are Pearson's correlation coefficient and Spearman's correlation coefficient. Spearman's correlation coefficient measures the monotonic relationship between two variables. Spearman's correlation coefficient values also range from -1 to 1, with a value of -1 indicating a strong negative relationship, 0 indicating no connection, and 1 indicating a strong positive relationship. The form of the correlation table can be seen in Fig. 7 and Fig. 8.

The result of the visualization of the correlation coefficient table using pandas profiling shows that the correlation table can provide clear visual information on the relationship between data through colour and the value of the correlation coefficient. This information is similar to the information displayed in the alerts feature but can provide a more detailed

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

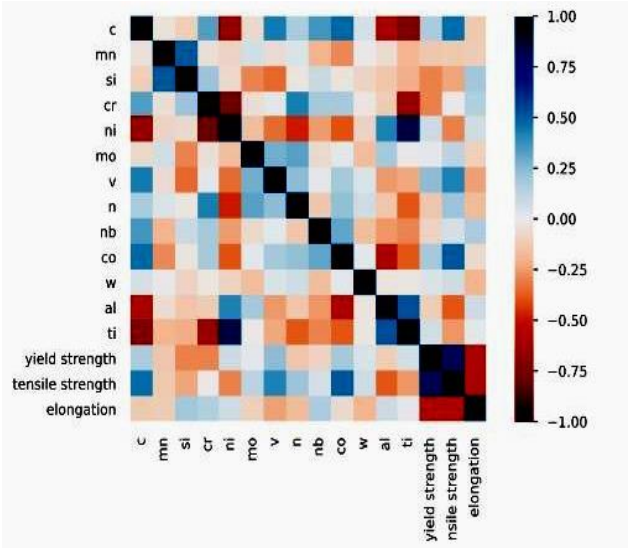

Fig 7. Pandas profiling correlation table display

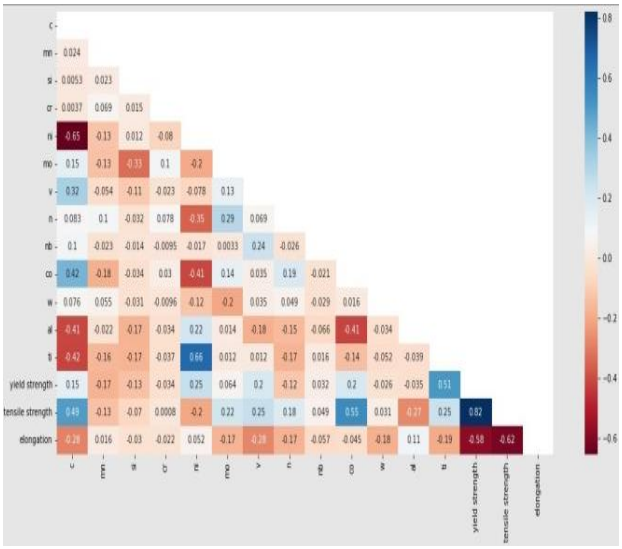

Fig 8. Display of the correlation coefficient table

picture of the relationship between data. The results showed that the yield strength of nickel-based superalloys had a fairly high correlation with titanium (0.51) and a moderate correlation with nickel (0.25), vanadium (0.2), and cobalt (0.2). The tensile strength of the nickel-based superalloys has a relatively high correlation with the yield strength (0.88) and a moderate correlation with the elements carbon (0.49), cobalt (0.55), titanium (0.25), and vanadium (0.25). Elongation of nickel-based superalloys has a relatively high negative correlation with tensile strength (-0.62) and yield strength (-0.58). This shows that the higher the tensile strength and yield strength value, the lower the elongation value. Conversely, the lower the tensile strength and yield strength value, the higher the elongation value. Thus, these elements significantly influence the mechanical properties of nickel-based superalloys.

Conclusions

Based on the analysis of the mechanical properties of nickel-based superalloys according to their chemical composition using the pandas profiling library, the following conclusions were obtained. The use of Exploratory Data Analysis (EDA) method with the pandas profiling library on Google Colab can assist in analyzing the mechanical properties of nickel-based superalloys quickly and easily, as well as providing clear information about data patterns, data structure, and data correlations. The analysis results show that the mechanical properties of nickel-based superalloys are influenced by the chemical composition of the alloy. Yield strength has a significant correlation with titanium (0.51), moderate correlation with nickel (0.25), vanadium (0.2), and cobalt (0.2). Tensile strength has a moderate correlation with carbon (0.49) and cobalt (0.55), and weak correlation with titanium (0.25) and vanadium (0.25). On the other hand, elongation has a significant negative correlation with tensile strength (-0.62) and yield strength (-0.58). These results also demonstrate the usefulness of the pandas profiling library in analyzing the mechanical properties of nickel-based superalloys quickly and easily, providing clear information about data patterns, structure, and correlations.

References

[1] Pollock, T. M., & Tin, S. (2006). Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *Journal of propulsion and power*, *22*(2), 361-374.

[2] Yeh, A. C., Sato, A., Kobayashi, T., & Harada, H. (2008). On the creep and phase stability of advanced Ni-base single crystal superalloys. *Materials Science and Engineering: A*, *490*(1-2), 445-451.

[3] Mabruri, E. (2015). Peranan unsur refraktori didalam nickel-based superalloys: suatu review. *Metalurgi*, *26*(2), 67-78.

[4] Caron, P., & Khan, T. (1999). Evolution of Ni-based superalloys for single crystal gas turbine blade applications. *Aerospace Science and Technology*, *3*(8), 513-523.

[5] Wei, J., Chu, X., Sun, X. Y., Xu, K., Deng, H. X., Chen, J., ... & Lei, M. (2019). Machine learning in materials science. *InfoMat*, *1*(3), 338-358.

Vol.4 (No.3). 2022. pp. 86-91
Published online: December 1, 2022
ISSN: 2477-541X

[6] Martinez, W. L., Martinez, A. R., & Solka, J. (2017). *Exploratory data analysis with MATLAB*. CRC Press.

[7] Martinez, W. L., & Martinez, A. R. (2015). *Computational statistics handbook with MATLAB* (Vol. 22). CRC press.

[8] Kürzl, H. (1988). Exploratory data analysis: recent advances for the interpretation of geochemical data. *Journal of Geochemical Exploration*, *30*(1-3), 309-322.

[9] Jansen, F. E., & Kelkar, M. G. (1996, March). Exploratory data analysis of production data. In *Permian Basin Oil and Gas Recovery Conference*. OnePetro.

[10] Flumignan, D. L., Anaia, G. C., de O. Ferreira, F., Tininis, A. G., & de Oliveira, J. E. (2007). Screening Brazilian automotive gasoline quality through quantification of saturated hydrocarbons and anhydrous ethanol by gas chromatography and exploratory data analysis. *Chromatographia*, *65*, 617-623.

[11] Kumar, M., Kumar, A., & Palaparthy, V. S. (2020). Soil sensors-based prediction system for plant diseases using exploratory data analysis and machine learning. *IEEE Sensors Journal*, *21*(16), 17455-17468.

[12] Ogunsina, K., Bilionis, I., & DeLaurentis, D. (2021). Exploratory data analysis for airline disruption management. *Machine Learning with Applications*, *6*, 100102.

[13] Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N. E., Bajaj, S., Wang, Q., ... & Jain, A. (2018). Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, *152*, 60-69.

[14] Brugman, S. (2019). *Pandas-profiling*. https://github.com/pandas-profiling/pandas-profiling-data

[15] Nguyen, Q. V., Simoff, S., Qian, Y., & Huang, M. L. (2016, September). Deep exploration of multidimensional data with linkable scatterplots. In *Proceedings of the 9th International Symposium on Visual Information Communication and Interaction* (pp. 43-50).

[16] Allen, M., & Cervo, D. (2015). *Multi-domain master data management: Advanced MDM and data governance in practice*. Morgan Kaufmann.