Original Research Article: full paper

(2024), «EUREKA: Physics and Engineering»
Number 5

OPTIMIZING STAINLESS STEEL TENSILE STRENGTH ANALYSIS: THROUGH DATA EXPLORATION AND MACHINE LEARNING DESIGN WITH STREAMLIT

Desmarita Leni
Department of Mechanical Engineeringi
Universitas Muhammadiyah Sumatera Barat
4 Pasir Jambak str., Pasie Nan Tigo, Kec. Koto Tangah, Kota Padang,
Sumatera Barat, Indonesia, 25586

Arwizet Karudin ✉
Department of Mechanical Engineering
Universitas Negeri Padang
Prof. Dr. Hamka str., Air Tawar Padang, Sumatera Barat, Indonesia, 25132
arwizet@ft.unp.ac.id

Muhammad Rabiu Abbas
Department of Mechanical Engineering
Federal University of Transportation – Daura
28JF+J29, Daura, Katsina, Nigeria, 824101

Jai Kumar Sharma
Department of Mechanical Engineering
ITM University
NH-44, Bypass Burari, Jhansi Road Gwalior (M.P.), India, 475001

Adriansyah Adriansyah
Department of Mechanical Engineering
Politeknik Negeri Padang
Kampus str., Limau Manis, Kec. Pauh, Kota Padang, Sumatera Barat, Indonesia, 25164

✉**Corresponding author**

Abstract

The use of Exploratory Data Analysis (EDA) and machine learning in material science has rapidly advanced in recent years. EDA enables researchers to thoroughly explore and analyze material datasets, while machine learning allows for the development of predictive models capable of understanding complex patterns within the data. This study aims to develop an optimization tool to enhance the analysis of tensile strength in stainless steel by leveraging integrated data exploration and machine learning approaches within the Streamlit framework. The developed tool consists of four main features: data visualization, correlation analysis, 3D visualization, and machine learning. The developed machine learning model has 14 input variables, including chemical elements and heat treatment temperatures. In this research, the machine learning features comprise three models: Decision Tree, Random Forest, and Artificial Neural Network. The research findings indicate that the optimization tool can automatically display stainless steel tensile strength data using available pandas profiling in the visualization feature. The correlation feature can illustrate the relationship between chemical elements and heat treatment temperatures concerning stainless steel tensile strength. The 3D visualization feature can be utilized to identify optimal values of chemical elements and heat treatment temperatures according to desired tensile strength. Meanwhile, the machine learning feature can accurately predict stainless steel tensile strength based on chemical composition and heat treatment temperatures. This is evident from the performance evaluation metrics of the Random Forest model, which achieved MAE of 10.36, RMSE of 14.44, and R-squared of 0.97.

Keywords: stainless steel, tensile strength, exploratory data analysis, machine learning, Streamlit.

DOI: 10.21303/2461-4262.2024.003296

1. Introduction

The conventional analysis of the mechanical properties of materials, especially alloy steels, often involves a series of experimental physical tests and laboratory testing [1]. This process is time-consuming and requires significant costs. The initial stages involve the preparation of varied

samples, requiring a considerable amount. Subsequently, testing is conducted using specialized equipment, such as tensile testing machines and hardness testing machines. The data generated from these tests then needs to be manually analyzed to extract information about the desired mechanical properties of the material, such as strength, toughness, and hardness [2, 3]. This conventional approach is often prone to human errors due to the complexity of the data required to obtain accurate analysis results.

Research indicates [4, 5], that when designing the mechanical properties of alloy steels, variables such as chemical composition percentages, heat treatment processes, and microstructure need to be considered. Generally, alloy steels consist of a mixture of specific metal elements in appropriate proportions. Changes in the percentage of chemical elements can affect the strength, toughness, and hardness of alloy steels; for example, adding carbon can increase the hardness of alloy steels, while adding chromium can enhance corrosion resistance [6]. Heat treatment is a critical step in shaping the mechanical properties of alloy steels. This process involves heating and cooling alloy steels at specific temperatures to alter their microstructure and mechanical properties. Heat treatment can influence the level of hardness, strength, and toughness of alloy steels. Research suggests [7, 8] that rapid heat treatment can increase the hardness of alloy steels, while slow heat treatment with gradual cooling can enhance strength and toughness. The microstructure of alloy steels refers to the arrangement of crystal grains and phases formed during the formation and heat treatment processes. This microstructure significantly affects the mechanical properties of materials by determining stress distribution and dislocation movement within the material. A homogeneous microstructure with fine crystal grains tends to provide higher strength, whereas a heterogeneous microstructure or the presence of defects in crystal grains can reduce strength and toughness [9]. These three variables are interrelated and have a highly complex relationship when influencing the mechanical properties of alloy steels. Therefore, a profound understanding of the relationship between these variables is crucial in designing alloy steels with the desired mechanical properties.

The field of materials science and engineering has undergone three stages of development over the years, starting from the experimental stage, moving to the theoretical stage, and progressing to the computational simulation stage [10, 11]. In recent years, the data generated from experiments and simulations has rapidly evolved, as evidenced by the proliferation of material databases such as the Open Quantum Material Database, Material Project, Computational Materials Repository, Harvard Clean Energy Project, Inorganic Crystal Structure Database, Matmach, MatWeb, and Aflowlib [12, 13]. This phenomenon enables the active utilization of material data for designing, simulating, and predicting the mechanical properties of materials based on data using an approach called exploratory data analysis (EDA) integrated with the Streamlit framework. EDA is an initial data analysis process aimed at understanding the characteristics, structure, and essential components of data before further analysis or predictive modeling [14, 15]. This analysis involves descriptive statistical approaches and various 2D and 3D visualization techniques such as histograms, scatter plots, box plots, bar plots, line plots, and other visualization methods [16]. On the other hand, Streamlit is an open-source framework that allows developers to use the Python programming language to create interactive web applications easily and quickly. By integrating EDA into Streamlit, researchers can create interactive user interfaces (UI) for direct data analysis within a web browser. Various EDA techniques like data visualization, descriptive statistics, and exploration of patterns in data can be implemented into web-based Streamlit applications [17].

The integration of EDA with Streamlit has proven to be one of the tools capable of optimizing data-driven analysis, as demonstrated in the study [18]. They developed an interactive tool named QuickViz, a web application leveraging the Streamlit framework. QuickViz is designed to simplify the complex data analysis process and make it more accessible to various stakeholders, including users with no programming background. The use of this tool allows researchers and data analysts to upload their datasets into the web application without having to write complex code or additional commands. The research findings indicate that using QuickViz can save time in data analysis, as the analysis results are automatically presented in an interactive web dashboard. This facilitates stakeholders in understanding of the findings of the data analysis and aids in making better decisions. In another study conducted [19], they integrated machine learning models with the Streamlit framework as a platform to build interactive web applications that enable users to easily access developed

prediction models. The research results show that machine learning models integrated with Streamlit can be accessed in real-time by users, allowing them to take necessary preventive actions quickly [20] combined EDA and machine learning methods integrated with Streamlit to create a predictive model for consumer loan approval. Using EDA, loan approval data was prepared and evaluated to identify patterns and relationships among variables. Steps like feature engineering and feature selection were performed to prepare the data before training the model. ML modeling was conducted using various algorithms, including Logistic Regression (LR), Random Forest (RF), and XGBoost (XGB). The research results indicate that this combination allows users to input loan application details and obtain prediction results directly. Based on the findings of previous research, it can be seen that EDA integrated with Streamlit can be a solution to optimize data-driven analysis.

With the rapid growth of material data in the past and the development of open-source frameworks that can be integrated with various methods such as data analysis techniques, visualization, machine learning models, and web technologies, this research aims to develop a supporting tool to optimize the design of stainless steel tensile strength based on a web platform using the Streamlit framework. In this study, EDA and machine learning methods are integrated with the Streamlit framework to analyze the influence of chemical elements, heat treatment methods, cooling methods, and heat treatment temperatures on the tensile strength of stainless steel. Additionally, machine learning methods can also be used to predict and simulate how the variables of chemical elements and heat treatment can affect the tensile strength of stainless steel. The use of this framework allows researchers and analysts to conduct stainless steel tensile strength analysis more efficiently, as it can be accessed in real-time through a web browser.

2. Materials and methods

In this study, the methods of exploratory data analysis (EDA) and machine learning are combined using the Streamlit framework. This combination produces a web-based optimization tool that can be used to enhance the tensile strength of stainless steel. This tool enables interactive and efficient analysis of data and prediction of stainless steel tensile strength. In general, the design of this stainless steel tensile strength optimization tool is divided into two stages:

1) the design of the machine learning model;

2) the integration of all features into the Streamlit framework, as shown in **Fig. 1.**

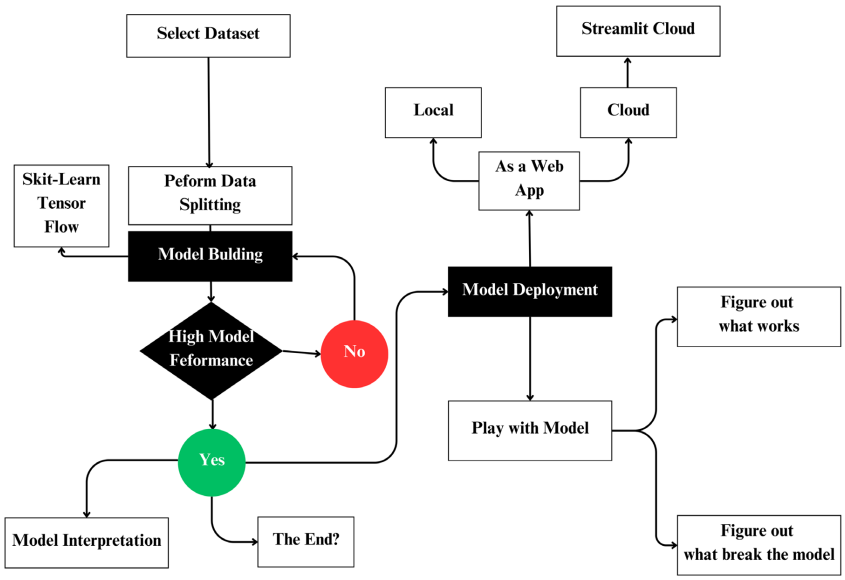

Fig. 1. Research scheme

2. 1. Dataset

The dataset used in this study refers to tensile test results from various types of austenitic stainless steels (ASS), such as SUS 304, SUS 316, SUS 321, SUS 347, and NCF 800 H. The entire dataset consists of 2180 samples, including information on mechanical properties, alloy chemical

elements, heat treatment temperatures, and cooling methods. The data source is from the Creep Data Sheet of Steel (No. 4B, 5B, 6B, 14B, 15B, 26B, 27B, 28B, 32A, 42, and 45) published by NIMS MatNavi and BSCC High Temperature Data from the British Steelmakers Creep Committee [21]. This data has been collected by the Material Algorithm Project (MAP) [22], a research project conducted by the University of Cambridge, and is available for research and educational purposes.

2. 2. Design of machine learning models

The machine learning method used in this research consists of three models, namely decision trees, random forests, and artificial neural networks. The three algorithms were set using the default parameters provided by the scikit-learn library version 1.2.2 [23] for *DT* and *RF* and Keras version 2.12.0 [24] for ANN. By varying these three models, users can optimize the prediction of stainless steel tensile strength by leveraging the strengths and advantages of each model. This provides flexibility and higher prediction accuracy, allowing users to select the model that best fits their needs and the characteristics of the data they possess. Each model is evaluated using several evaluation metrics, such as mean absolute error (MAE), root mean square error (RMSE), and R-squared. These evaluation metrics are commonly used in the context of predicting the mechanical properties of alloy steels [25–27].

MAE measures the average of the absolute differences between the predicted values and the actual values, providing an insight into how accurate the model's predictions are in capturing the data variation. *MAE* can be calculated using the following equation:

$$MAE = \frac{i}{N}\Sigma\left|y_i - z_i\right|, \tag{1}$$

where *i* represents the index of the data sample, *N* signifies the total number of samples, *y_i* denotes the actual value of the *i*-th data point, and *z_i* indicates the predicted value by the model for the *i*-th data point.

RMSE measures the root of the mean square of the differences between the predicted values and the actual values, providing an insight into how well the model estimates the true values of the data. *RMSE* can be calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f\left(X_i\right) - Y_i\right)^2}, \tag{2}$$

where *n* is the number of data points used to test the model, *f*(*X_i*) is the value predicted by the model for the *i*-th data point, and *Y_i* is the actual value for the *i*-th data point.

R-squared measures how well the model fits the data by comparing the variation explained by the model to the total variation in the data, thus providing information about how well the model explains the data variability. *R*-squared can be calculated using the following equation:

$$R = \frac{\sum_i^n = 1\left(f\left(X_i\right) - f\left(\bar{X}\right)\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(f\left(X_i\right) - f\left(\bar{X}\right)\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}, \tag{3}$$

where $f(X_i)$ denotes the predicted value of the dependent variable (Y) based on the independent variable (X) for the i-th observation, $f\left(\bar{X}\right)$ represents the average of all predicted values $f(X_i)$ across all observations, Y_i indicates the actual observed value of the dependent variable for the i-th observation, $\bar{Y}$ signifies the average of all observed values Y_i across all observations, and n represents the total number of observations.

2. 3. Streamlit integration

The integration process involves developing an interactive web application using Streamlit, which will combine the *EDA* techniques and pre-prepared machine learning models. The *EDA* methods implemented in this optimization tool consist of Pandas profiling, data visualization techniques, and correlation heatmaps. The use of pandas profiling enables the tool to automatically present statistical summaries of the data, including distributions, the presence of missing values, and correlations between variables [28]. Meanwhile, visualization techniques will assist users in

understanding hidden patterns in the data more easily through various types of interactive graphs and plots [29]. Additionally, the correlation heatmap will provide a clear visual insight into the relationships between variables in the form of a color matrix [30]. Thus, users can quickly identify significant relationships between variables and make informed decisions based on the available information. This will be key in optimizing the strength of stainless steel's tensile pull, as it enables users to take appropriate actions based on a comprehensive understanding of the data. Correlation heatmap can be calculated using the Pearson correlation equation as can be seen in the following equation:

$$r_{xy} = \frac{\sum xy}{(n-1)S_x\, S_y}, \tag{4}$$

where *r_xy* represents the Pearson correlation coefficient, Σ*xy* denotes the sum of the products of *x* and *y*, *n* indicates the sample size, *x* stands for the independent variable, y represents the dependent variable, and *S* signifies the standard deviation [13]. The value of the correlation coefficient ranges from –1 to 1. A value of –1 indicates a strong negative correlation between the two variables, a value of 0 indicates no correlation, and a value of 1 indicates a strong positive correlation [30, 31].

In this way, users can directly access the optimization tool through a web browser. This integration enables users to easily utilize EDA and machine learning technologies in analyzing data and making real-time predictions, thus supporting faster and more accurate decision-making. With a user-friendly interface, integration with the Streamlit framework ensures that the results of this research can be easily and effectively utilized in practical situations.

3. Results and Discussion

Essentially, the tensile strength of stainless steel is influenced by several key factors, including the chemical elements contained in the material, the heat treatment processes applied, and the microstructure of the material itself. These three factors play a significant role in determining the mechanical performance of stainless steel. The chemical composition of stainless steel can affect its mechanical properties, such as strength, ductility, and corrosion resistance. On the other hand, heat treatment processes such as heating, cooling, and recrystallization can alter the microstructure of the material, which in turn affects its mechanical properties [32, 33]. Therefore, the optimization tool designed aims to improve the analysis process of stainless steel tensile strength by considering these factors holistically. This tool is designed to enable users to optimize the analysis process by considering the composition of chemical elements and the heat treatment applied to the material. Thus, users can identify the optimal combination of chemical elements and heat treatment that can produce the desired tensile strength for specific applications. With this approach, the optimization tool aims to provide a more efficient and precise solution in the development and processing of stainless steel, thereby enhancing product quality and overall manufacturing process efficiency.

3. 1. Dataset

The dataset initially consisted of 2180 samples, but after cleaning, only 986 samples with complete information remained. The original dataset contained various other information, such as grain size, melting type, and bar size. However, this research only utilized 14 input variables and one output variable, as shown in **Table 1**.

Next, the stainless steel dataset is divided into two parts: for training and external testing. The total number of cleaned samples reached 986, then was reduced by 50 samples for external testing purposes. Thus, the number of training datasets used is 936. This training data is further divided into two parts, with 80 % for the training process and 20 % for validation purposes. To obtain maximum results, the external testing data must represent the distribution of the training data, so a Kolmogorov-Smirnov (KS) test is needed. The Kolmogorov-Smirnov (KS) test results indicate how similar the distribution of variables is between the training dataset (data used to train machine learning models) and the external testing dataset (data used to test model performance). The smaller the KS statistic value, the more similar the distribution between the two datasets. Conversely, the larger the p-value, the less significant the difference between the distributions of the two datasets [34]. The KS test values obtained range from 0.060 to 0.132, indicating relatively small differences between the

distributions. Furthermore, the p-values are calculated for each test, ranging from 0.20 to 0.98. These results indicate that the low KS statistic values and high p-values are indicative of the similarity in variable distribution between the training dataset and the external testing dataset. The similarity in distribution between the training and external testing datasets helps ensure that machine learning models trained using the training dataset can provide consistent and accurate predictions when applied to previously unseen data, such as external testing data. This greatly assists in objectively measuring model performance and ensuring the reliability of model predictions in real-world situations. **Fig. 2** illustrates the comparison of the distribution of training data and external testing data.

Table 1

Variables of the austenitic stainless steel (ASS) dataset

No.	Variables	No.	Variables
1	Chromium (Cr, wt %)	8	Carbon (C, wt %)
2	Nickel (Ni, wt %)	9	Boron (B, wt %)
3	Molybdenum (Mo, wt %)	10	Phosphorus (P, wt %)
4	Manganese (Mn, wt %)	11	Sulfur (S, wt %)
5	Silicon (Si, wt %)	12	Solution treatment temperature (*Ts*, K)
6	Niobium (Nb, wt %)	13	Solution treatment time (*ts*, s)
7	Titanium (Ti, wt %)	14	Temperature (K)

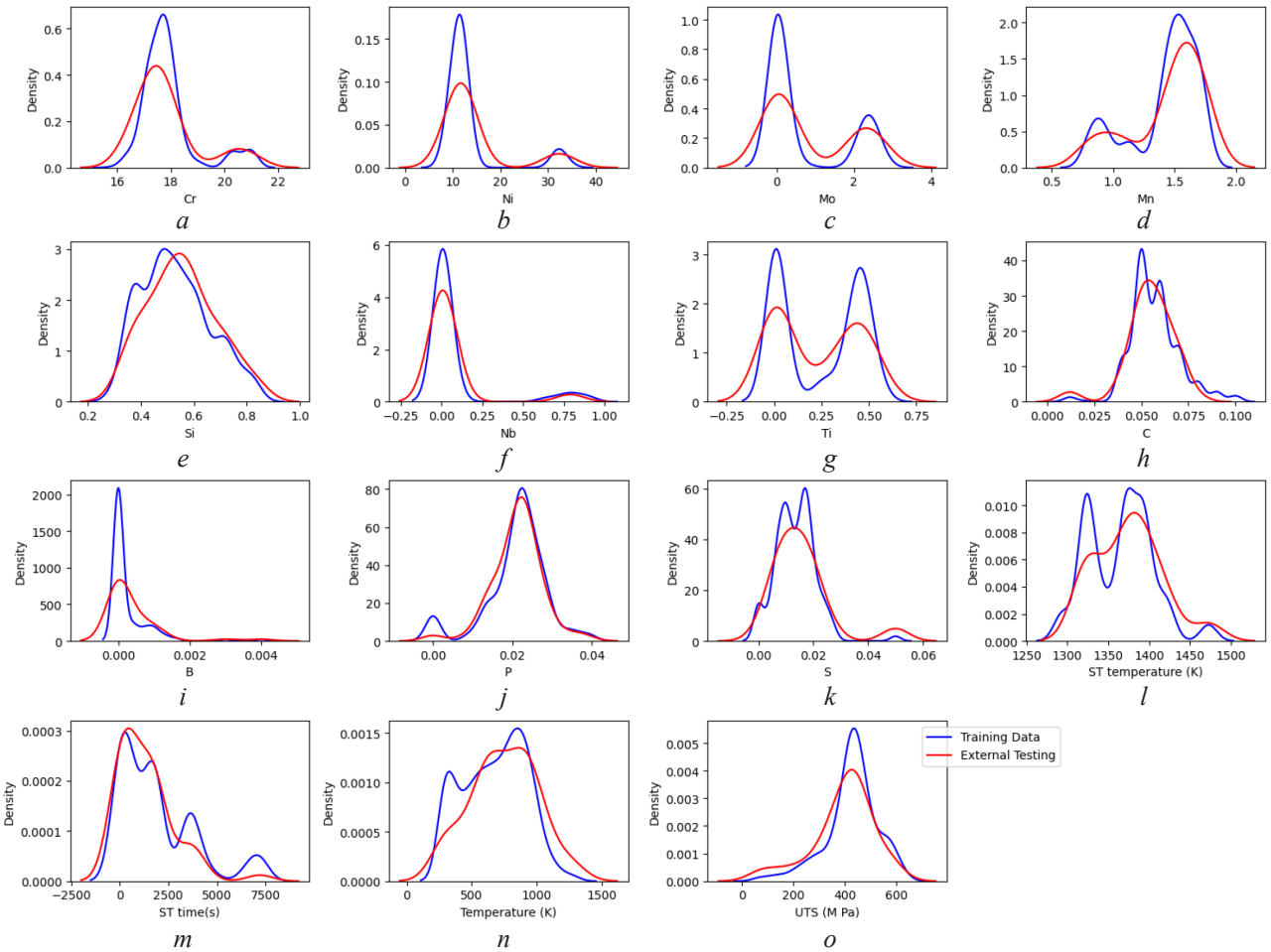

Fig. 2. Comparison of training and external testing data distributions where:
a – Chromium; *b* – Nickel; *c* – Molybdenum; *d* – Manganese; *e* – Silicon; *f* – Niobium; *g* – Titanium; *h* – Carbon; *i* – Boron; *j* – Phosphorus; *k* – Sulphur; *l* – solution treatment temperature; *m* – solution treatment time; *n* – temperature; *o* – ultimate tensile strength

3. 2. Design of machine learning models

In this study, three predictive machine learning algorithms are employed to predict the tensile strength of stainless steel, namely Decision Tree (DT), Random Forest (RF), and Artificial Neural Networks (ANN). The use of these three algorithms aims to provide variation and flexibility in modeling, as well as enable users to leverage the strengths and advantages of each model. Decision Tree (DT) is chosen for its intuitive and easy-to-interpret nature. DT works by dividing the dataset into smaller subsets based on simple decision rules [35]. This allows researchers to understand the relationships between variables and directly identify factors that affect the tensile strength of stainless steel. Random Forest (RF) is selected for its ability to address overfitting issues and improve prediction accuracy. RF works by combining predictions from multiple different decision trees, resulting in more stable and reliable predictions [36]. The use of RF can help improve the accuracy and reliability of models for predicting the tensile strength of stainless steel. Meanwhile, artificial neural networks (ANN) are chosen for their ability to handle complex and non-linear patterns in data. ANN operates with a network structure inspired by the human brain, allowing it to identify complex relationships between input and output variables [37]. The use of ANN can help reveal patterns that are difficult for humans to understand or explain. In this research, the three algorithms were set using the default parameters provided by the scikit-learn library version 1.2.2 for *DT* and *RF* and Keras version 2.12.0 for ANN. These default values and parameters can be seen in **Table 2**:

Table 2

Parameters and values of machine learning models for each algorithm

DT		RF		ANN	
Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Criterion	Mse	n_estimators	100	Epochs	100
Splitter	best	criterion	Mse	Batch	32
Min Samples Split	2	max_depth	None	Model	Sequential
Min Samples Leaf	1	min_samples_split min_samples_leaf max_features boot-strap	2 1 Auto True	Dense (layer 1) Dense (layer 2) Dense (output layer)	Units: 32 Activation: relu Units: 16 Activation: relu Units: 1 Activation: linear

The designed machine learning model consists of 14 input variables and 1 output variable, as seen in **Table 1**. Additionally, to address the issue of overfitting, the method of cross-validation is employed with a fold count of 10. This approach helps improve the model's generalization and avoids the tendency for the model to memorize the training data excessively, resulting in more consistent and reliable predictions for new unseen data [38]. Furthermore, each model is evaluated using three evaluation metrics: mean absolute error (MAE), root mean square error (RMSE), and *R*-squared. The evaluation results of training the three models can be observed in **Fig. 3**.

Based on the training model evaluation results, RF demonstrates the best performance with MAE of 9.42, RMSE of 11.68, and an *R*-squared value of 0.98. This is followed by DT with MAE of 15.19, RMSE of 18.75, and R-squared of 0.95. Meanwhile, ANN shows lower performance with MAE of 40.80, an RMSE of 50.19, and R-squared of 0.70. From these results, it can be concluded that RF has more accurate predictive capabilities compared to DT and ANN in predicting the tensile strength of stainless steel based on chemical composition and heat treatment.

Next, model testing is conducted using external data. Model testing with external data refers to the process of evaluating the performance of a machine learning model using a dataset not used in the training phase. This external dataset is specifically separated from the dataset used to train the model and is used to test how well the model can generalize patterns in new data that it has not seen before [39]. In other words, testing the model with external data helps ensure that the model

has good predictive capabilities for the mechanical properties of stainless steel using new data it has never encountered before. The practice of testing models using external data has become a common standard in evaluating the performance of machine learning models, as outlined in studies [25–27]. This approach aims to validate the results of the proposed model, ensuring that the model is reliable and has strong generalization capabilities in various real-world situations. Thus, the use of external data in model testing becomes a critical step in ensuring the reliability and superiority of the proposed machine learning model.

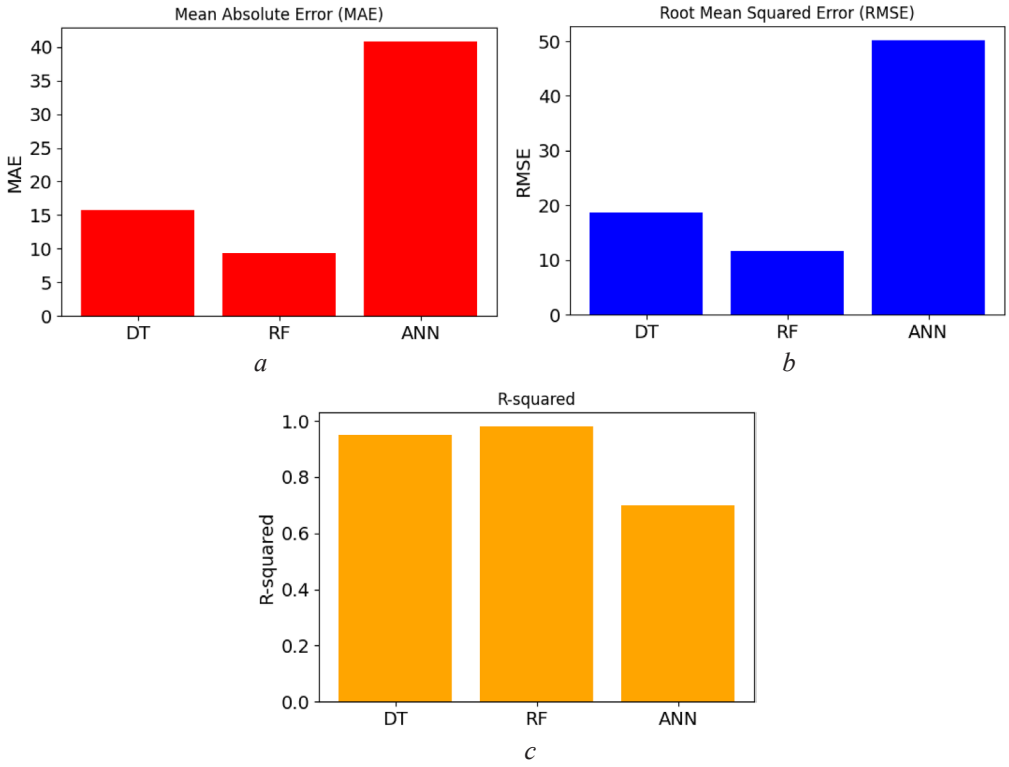

Fig. 3. Comparison of model evaluation results:
a – mean absolute error; *b* – root mean squared error; *c* – R-squared

In this stage, each model is tested using a new dataset consisting of 50 data samples that were not used in the model training process. The performance evaluation of the models using this new dataset can be found in **Table 3**, while the prediction results of each model can be seen in **Fig. 4.**

Table 3
Comparison of training and testing evaluation metrics using new data

Model	Evaluation metrics	Training	New data
DT	MAE	15.77	16.12
	RMSE	18.75	19.09
	R-squared	0.95	0.94
RF	MAE	9.42	10.36
	RMSE	11.6	14.44
	R-squared	0.98	0.97
ANN	MAE	40.8	42.26
	RMSE	50.19	58.97
	R-squared	0.7	0.69

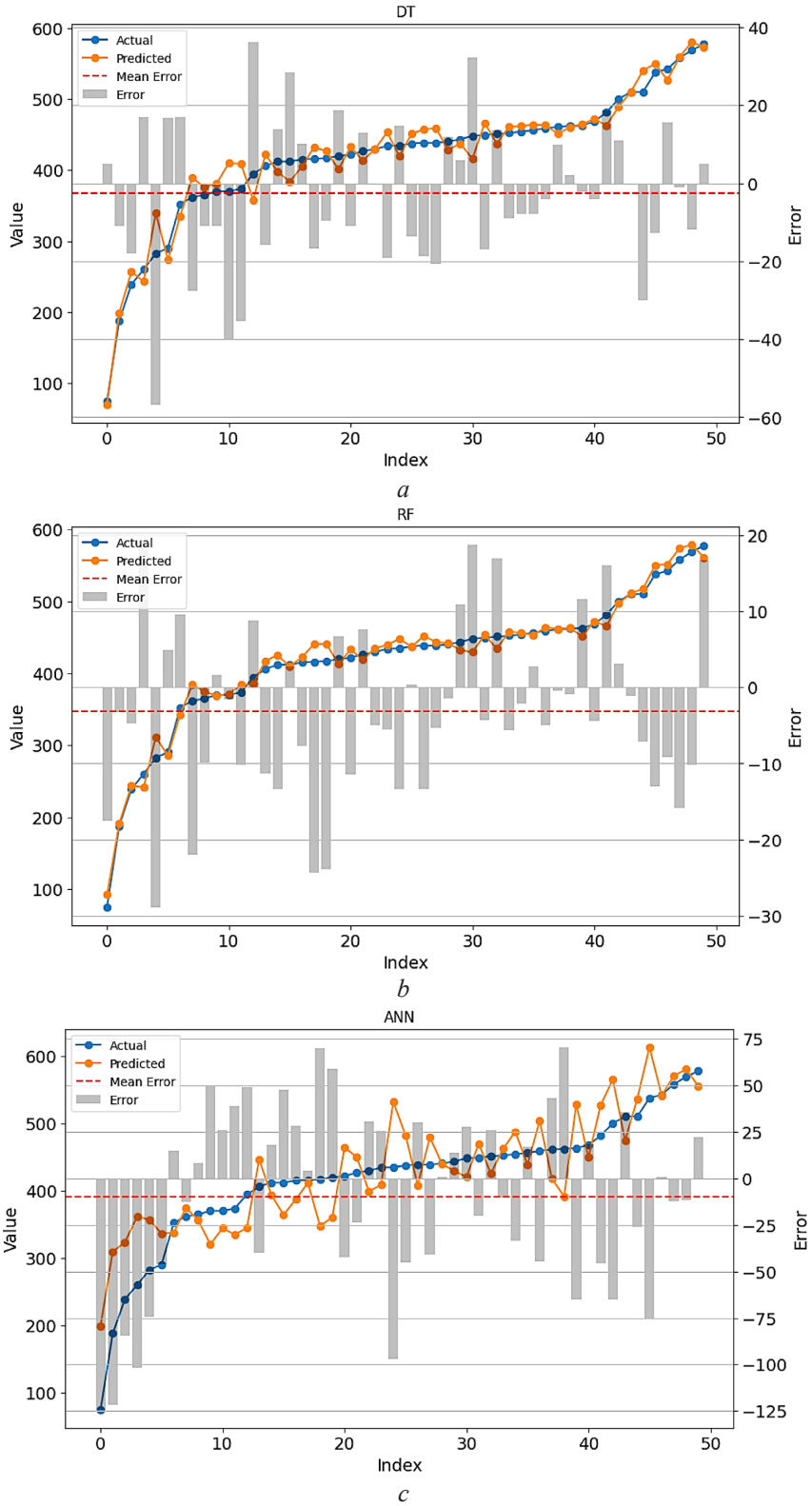

Fig. 4. Comparison of predicted and actual results:
a – decision tree; *b* – random forest; *c* – artificial neural network

The evaluation results indicate that, overall, there is no significant difference between the evaluation metrics in training and testing for each model. This indicates that the models have good capabilities for generalizing patterns from new data that were not seen in the training dataset.

The testing results with the new data also show that RF still outperforms DT and ANN. These findings are consistent with previous research [40], which shows that Random Forest (RF) tends to exhibit better performance. One reason is the application of the ensemble learning concept, where the combination of a large number of decision trees in an ensemble can reduce variance and significantly improve prediction accuracy. Additionally, RF also applies random feature selection when building each decision tree, allowing for more effective handling of datasets with a large number of features or redundancy. Another advantage is RF's ability to handle imbalanced datasets without relying on feature normalization or standardization because it does not require measuring the distance between features [41]. Overall, these metric factors give RF an advantage in various situations, especially in complex datasets with many features such as chemical composition and heat treatment.

Based on **Fig. 4**, it can be seen that random forest (RF) obtains the smallest error values when predicting the tensile strength of stainless steel, ranging from –30 MPa to 20 MPa. This indicates that RF tends to produce predictions that are close to the actual values of tensile strength. However, although RF has the smallest error values, there is still variability in its predictions, as reflected by the range of error values. Furthermore, Decision Tree (DT) also shows relatively good performance, with an error range of –60 MPa to 40 MPa. Although this range is slightly wider than RF, DT is still capable of producing predictions that are relatively close to the actual values. Meanwhile, artificial neural networks (ANN) experience higher error levels, with a range of –125 MPa to 75 MPa. This indicates that ANN tends to have difficulty predicting the tensile strength of stainless steel with high accuracy. In the context of predicting the tensile strength of stainless steel, positive error indicates that the model's prediction is lower than the actual value, while negative error indicates that the model's prediction is higher than the actual value [26]. Therefore, the different ranges of errors for each model provide an indication of how well the model can predict tensile strength, as well as whether the model tends to underestimate or overestimate.

3. 3. Streamlit integration

At this stage, the exploratory data analysis (EDA) method and machine learning are combined using the Streamlit framework. This combination results in a web-based optimization tool that can be used to enhance the tensile strength of stainless steel. This tool enables interactive and efficient data analysis and prediction of the tensile strength of stainless steel. The stainless steel tensile strength optimization tool is designed with four main features:

1) data visualization;
2) correlation;
3) 3D visualization;
4) machine learning.

The EDA method implemented in this optimization tool consists of three features: data visualization, correlation feature, and 3D visualization feature. These features are created using various Python libraries, such as Pandas profiling, Matplotlib, and Seaborn, which are integrated into the Streamlit framework for web-based interface access. **Fig. 5** illustrates the interface view of the visualization features of the stainless steel tensile strength optimization tool.

Based on **Fig. 5**, it can be seen that the visualization feature consists of various visualization techniques, ranging from automatic ones like Pandas profiling to manual ones like boxplot, lineplot, and scatterplot. Pandas Profiling is a Python library used for comprehensive and automated data analysis of datasets using Pandas DataFrame. Its function is to provide a general overview of the dataset's structure and characteristics in depth without requiring much code or manual steps. By using Pandas Profiling, users can quickly gain insights into various aspects of the dataset, including descriptive statistics, variable distributions, and variable correlations, as well as detecting missing or duplicate values. This enables users to save time and effort in the data exploration process, especially when working with large or complex datasets. Additionally, the reports generated by Pandas Profiling are presented in an easily understandable format and can be visualized well, making it easier for users to understand the dataset's characteristics comprehensively. Thus, Pandas profiling becomes a highly useful tool in the initial stage of data analysis to identify patterns, anomalies, and trends that can aid in better decision-making.

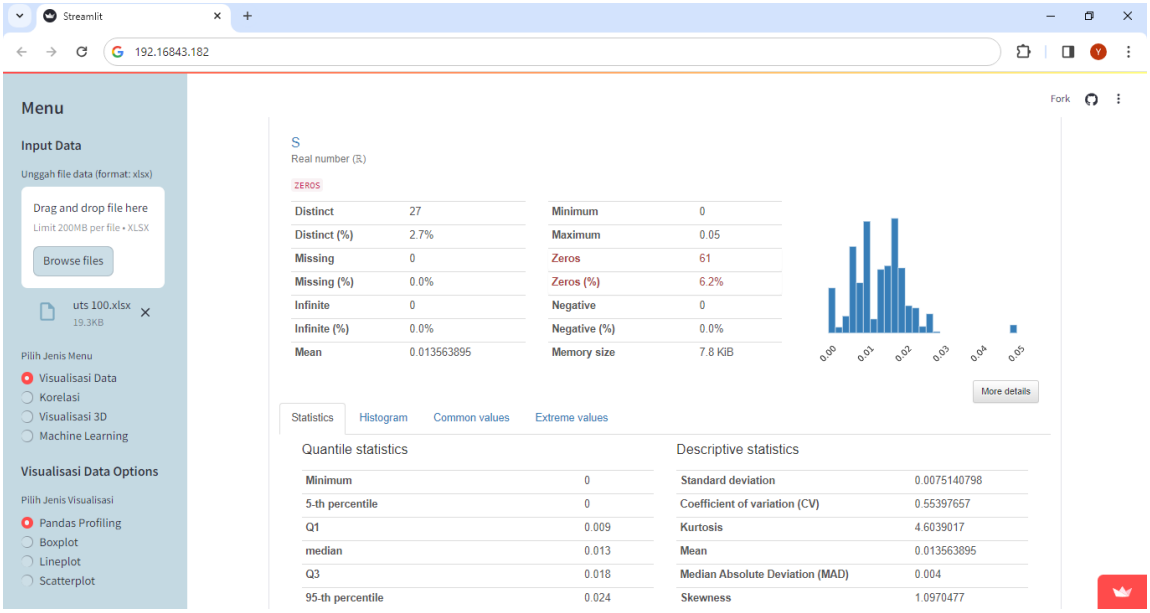

Fig. 5. Interface display of data visualization features using Pandas profiling

In the analysis of alloy steel mechanical properties based on data, it often takes a considerable amount of time to inspect the data structure in detail. This process involves examining the value distribution, identifying outliers, analyzing variable correlations, and so on. All of these steps are important for understanding the dataset well before proceeding to further analysis. Therefore, Pandas profiling becomes one of the most effective solutions to address these challenges. By using Pandas Profiling, users can easily view the value distribution of each variable, including the chemical composition and heat treatment temperature optimal for obtaining the desired tensile strength in stainless steel.

The second feature of this stainless steel tensile strength optimization tool is the correlation feature. This feature plays a significant role in providing a deep insight into the relationship between chemical composition and heat treatment temperature and the tensile strength of stainless steel. This feature not only displays the statistical correlation between variables but also allows users to explore the complex dynamics behind the interaction of various factors. With this capability, users can quickly see how variations in chemical composition or changes in heat treatment temperature affect the tensile strength of steel. With a deeper understanding of this correlation, users can identify patterns or trends that may be hidden in the data, enabling them to make better decisions in the analysis of stainless steel tensile strength. Thus, the correlation feature not only provides additional insights but also enriches users' understanding of the factors contributing to the tensile strength of stainless steel, thereby improving the quality and accuracy of their overall analysis. The display of this feature can be seen in **Fig. 6**.

The third feature of this stainless steel tensile strength optimization tool is the 3D visualization feature. By leveraging the 3D visualization feature, users can explore data more deeply and intuitively, thus presenting complex information in an easily understandable manner. This feature consists of two models, namely the surface area and the contour diagram. The 3D visualization feature not only allows users to explore data more deeply and intuitively but also provides the capability to perform optimization analysis of the percentage values of chemical elements and heat treatment temperature according to the desired tensile strength. By using the surface area and contour diagram models, users can visually identify the optimal combination of chemical elements and heat treatment temperature that will result in the desired tensile strength of stainless steel according to the established standards or goals. Thus, the 3D visualization feature not only facilitates a better understanding of the data but also enables users to make more precise decisions in optimizing the production process and producing products with the desired quality.

Various previous studies have highlighted the important role of 3D visualization in supporting, facilitating, and even enhancing the results of steel mechanical property analysis. With its ability to present data more intuitively and deeply, 3D visualization has proven to be a highly useful tool

for understanding material characteristics. One study conducted [42], showed that 3D visualization can be used as a medium to enhance a better understanding of the complex relationship between the chemical composition and mechanical properties of steel. By using 3D visualization models, researchers can easily explore variations in crystal structure and material phase distribution, which in turn help identify factors influencing mechanical properties. In another study conducted [43], it was highlighted that 3D visualization plays a key role in presenting a more in-depth analysis of the effects of heat treatment on the microstructure of steel. By presenting the steel microstructure in a realistic three-dimensional representation, researchers can effectively observe structural changes that occur during the heat treatment process. These findings support the idea that 3D visualization has a positive impact on deepening understanding of steel mechanical properties, which in turn can aid in designing stronger, more durable, and more cost-efficient materials. Therefore, the integration of the 3D visualization feature into the stainless steel tensile strength optimization tool is a suitable step. The display of the 3D visualization feature can be seen in **Fig. 7**.

Fig. 6. Interface display of correlation analysis using heatmap

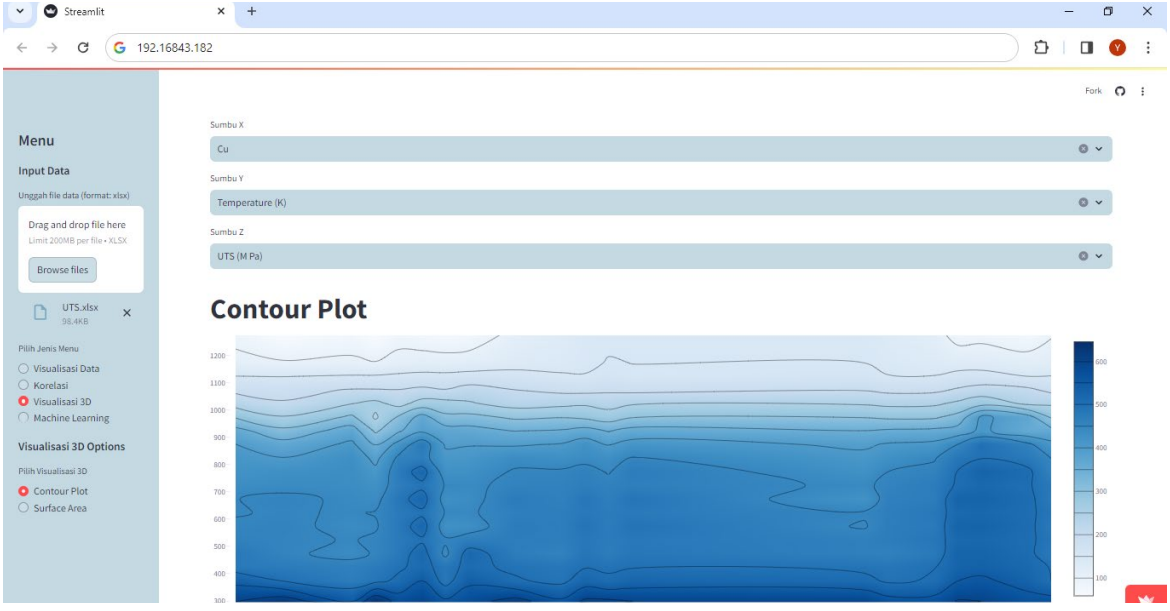

Fig. 7. Interface display of the 3D visualization feature

The last feature of this optimization tool is the prediction model using machine learning. This feature is designed to predict the tensile strength of stainless steel using 14 input variables that include chemical elements such as chromium (Cr), nickel (Ni), molybdenum (Mo), manganese (Mn), silicon (Si), niobium (Nb), titanium (Ti), carbon (C), boron (B), phosphorus (P), and sulfur (S), as well as heat treatment factors such as initial temperature (ST temperature), heat treatment time (ST time), and tempering temperature. Users can utilize this feature by inputting chemical element and heat treatment temperature data through the sidebar, as seen in **Fig. 8**. By leveraging machine learning technology, the process of predicting the tensile strength of stainless steel can be done faster, more cost-effectively, and more environmentally friendly. The machine learning models implemented in this optimization tool consist of three machine learning models: DT, RF, and ANN. This is intended to allow users to choose the model that best suits their needs and data characteristics.

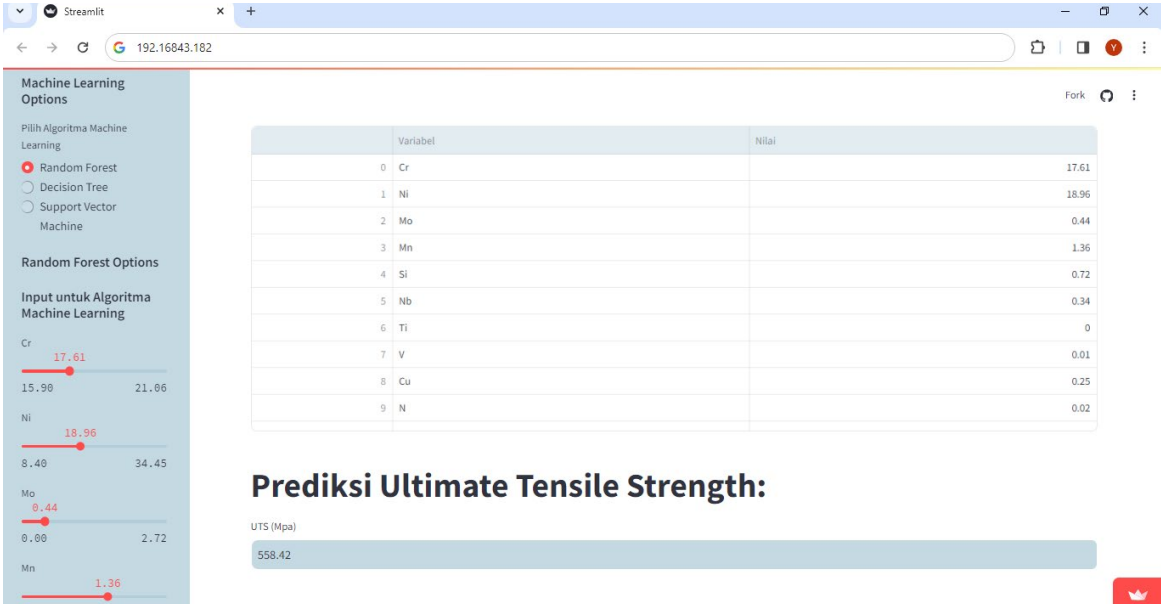

Fig. 8. Interface display of the machine learning feature

Based on several previous studies, machine learning methods have proven to be highly effective in designing new materials, predicting the mechanical properties of steel, and optimizing steel production processes. Research conducted [44], highlights that the use of machine learning techniques in material design allows researchers to explore a wider design space, identify complex patterns in material structures, and predict material mechanical properties. Another study [45], shows that machine learning methods have succeeded in predicting the mechanical properties of steel with high accuracy. By leveraging data related to chemical composition, heat treatment, and mechanical characteristics, machine learning models can produce predictions that closely approximate the actual values of the tensile strength, hardness, and toughness of steel. Studies conducted [27], emphasize that machine learning can be used to optimize steel production processes. By analyzing data from various production parameters such as temperature, time, and material composition, machine learning models can provide better recommendations to improve the efficiency, quality, and reliability of steel production processes. From these studies, it can be concluded that machine learning methods have great potential for supporting material research and development as well as improving operational efficiency in the steel industry. Integrating this technology into the optimization tool for the tensile strength of stainless steel is a fitting step in supporting practitioners to achieve better, faster, and more efficient material analysis and design.

This optimization tool plays a very important role in the overall analysis of the tensile strength of stainless steel. However, it should be acknowledged that this tool has limitations in predicting the tensile strength of stainless steel for more complex or extensive materials. When used to predict the properties of more complex materials, the results obtained may not be as optimal as expected.

Limitations of the data used in the study can also affect the accuracy and precision of the model. Additionally, the machine learning models used only utilize default parameters, so there is potential to improve model performance by adjusting parameters more specifically. Therefore, for future research, it is recommended to conduct more thorough parameter tuning for each model to enhance the quality of prediction results.

4. Conclusions

Based on the results of designing the optimization tool to improve the analysis of the tensile strength of stainless steel using data exploration and machine learning methods using Streamlit, it can be concluded that this tool can be used to improve efficiency in the analysis of the tensile strength of stainless steel. The optimization tool for the tensile strength of stainless steel is designed with four main features:

1) data visualization;
2) correlation;
3) 3D visualization;
4) machine learning.

These features can help users identify complex patterns between chemical elements and heat treatments with the tensile strength of stainless steel. Machine learning models are designed with 14 input variables and one output variable. Input variables consist of chemical elements such as chromium (Cr), nickel (Ni), molybdenum (Mo), manganese (Mn), silicon (Si), niobium (Nb), titanium (Ti), carbon (C), boron (B), phosphorus (P), and sulfur (S), as well as heat treatments such as initial temperature (ST temperature), heat treatment time (ST time), and tempering temperature (Temperature). The machine learning algorithms used include decision Tree (DT), random Forest (RF), and artificial neural networks (ANN). Each model is evaluated using three evaluation metrics, such as MAE, RMSE, and R-squared. Based on the training model evaluation results, RF demonstrates the best performance with MAE of 9.42, an RMSE of 11.68, and an R-squared value of 0.98. This is followed by DT with MAE of 15.19, RMSE of 18.75, and R-squared of 0.95. Meanwhile, ANN shows lower performance with MAE of 40.80, an RMSE of 50.19, and an R-squared of 0.70. From these results, it can be concluded that RF has more accurate predictive capabilities compared to DT and ANN in predicting the tensile strength of stainless steel based on chemical composition and heat treatment. Thus, this tool can be an effective solution for designing new materials, understanding sensitivity to certain factors, and improving the overall quality of steel production processes.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

[1] Kumar, W., Sharma, U. K., Shome, M. (2021). Mechanical properties of conventional structural steel and fire-resistant steel at elevated temperatures. Journal of Constructional Steel Research, 181, 106615. https://doi.org/10.1016/j.jcsr.2021.106615

[2] Yildiz, A. S., Davut, K., Koc, B., Yilmaz, O. (2020). Wire arc additive manufacturing of high-strength low alloy steels: study of process parameters and their influence on the bead geometry and mechanical characteristics. The International Journal of Advanced Manufacturing Technology, 108 (11-12), 3391–3404. https://doi.org/10.1007/s00170-020-05482-9

[3] Jafarian, H. R., Sabzi, M., Mousavi Anijdan, S. H., Eivani, A. R., Park, N. (2021). The influence of austenitization temperature on microstructural developments, mechanical properties, fracture mode and wear mechanism of Hadfield high manganese steel. Journal of Materials Research and Technology, 10, 819–831. https://doi.org/10.1016/j.jmrt.2020.12.003

[4] Zavdoveev, A., Poznyakov, V., Baudin, T., Rogante, M., Kim, H. S., Heaton, M. et al. (2021). Effect of heat treatment on the mechanical properties and microstructure of HSLA steels processed by various technologies. Materials Today Communications, 28, 102598. https://doi.org/10.1016/j.mtcomm.2021.102598

[5] Baharuddin, M. Y., Salleh, S., Suhasril, A. A., Zulkifly, A. H., Lee, M. H., Omar, M. A. et al. (2014). Fabrication of Low-Cost, Cementless Femoral Stem 316L Stainless Steel Using Investment Casting Technique. Artificial Organs, 38 (7), 603–608. https://doi.org/10.1111/aor.12222

[6] Sun, J., Tang, H., Wang, C., Han, Z., Li, S. (2021). Effects of Alloying Elements and Microstructure on Stainless Steel Corrosion: A Review. Steel Research International, 93 (5). https://doi.org/10.1002/srin.202100450

[7] Yadav, P., Saxena, K. K. (2020). Effect of heat-treatment on microstructure and mechanical properties of Ti alloys: An overview. Materials Today: Proceedings, 26, 2546–2557. https://doi.org/10.1016/j.matpr.2020.02.541

[8] Hafeez, M. A., Inam, A., Farooq, A. (2020). Mechanical and corrosion properties of medium carbon low alloy steel after cyclic quenching and tempering heat-treatments. Materials Research Express, 7 (1), 016553. https://doi.org/10.1088/2053-1591/ab6581

[9] Nuswantoro, N. F., Gunawarman, G., Saputra, M. R., Nanda, I. P., Idris, M. H., Arafat, A. (2018). Microstructure Analysis of Hydroxyapatite Coating on Stainless Steel 316L Using Investment Casting Technique for Implant Application. International Journal on Advanced Science, Engineering and Information Technology, 8 (5), 2168–2174. https://doi.org/10.18517/ijaseit.8.5.5808

[10] Agrawal, A., Choudhary, A. (2016). Perspective: Materials informatics and big data: Realization of the «fourth paradigm» of science in materials science. APL Materials, 4 (5). https://doi.org/10.1063/1.4946894

[11] Rajan, K. (2005). Materials informatics. Materials Today, 8 (10), 38–45. https://doi.org/10.1016/s1369-7021(05)71123-8

[12] Leni, D., Earnestly, F., Sumiati, R., Adriansyah, A., Kusuma, Y. P. (2023). Evaluasi sifat mekanik baja paduan rendah berdasarkan komposisi kimia dan suhu perlakuan panas menggunakan teknik exploratory data analysis (EDA). Dinamika Teknik Mesin, 13 (1), 74. https://doi.org/10.29303/dtm.v13i1.624

[13] Leni, D. (2023). The Influence of Heatmap Correlation-based Feature Selection on Predictive Modeling of Low Alloy Steel Mechanical Properties Using Artificial Neural Network (ANN) Algorithm. Journal of Energy, Material, and Instrumentation Technology, 4 (4), 152–162. https://doi.org/10.23960/jemit.v4i4.203

[14] Elman, C., Gerring, J., Mahoney, J. (Eds.) (2020). The Production of Knowledge. Cambridge University Press. https://doi.org/10.1017/9781108762519

[15] Milo, T., Somech, A. (2020). Automating Exploratory Data Analysis via Machine Learning: An Overview. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. https://doi.org/10.1145/3318464.3383126

[16] Scheiner, S. M., Gurevitch, J. (Eds.) (2001). Design and Analysis of Ecological Experiments. Oxford Academic. https://doi.org/10.1093/oso/9780195131871.001.0001

[17] Leni, D., Kesuma, D. S., Maimuzar, Haris, Afriyani, S. (2024). Prediction of Mechanical Properties of Austenitic Stainless Steels with the Use of Synthetic Data via Generative Adversarial Networks. The 7th Mechanical Engineering, Science and Technology International Conference. https://doi.org/10.3390/engproc2024063004

[18] Pitroda, H. (2022). A Proposal of an Interactive Web Application Tool QuickViz: To Automate Exploratory Data Analysis. 2022 IEEE 7th International Conference for Convergence in Technology (I2CT). https://doi.org/10.1109/i2ct54291.2022.9824068

[19] More, M., Panda, R., Bandgar, B., More, M. (2023). Bankruptcy Prediction Using Machine Learning: A New Technological Approach to Prevent Corporate Bankruptcy Through Well Deployed Streamlit Based Application. 2023 International Conference for Advancement in Technology (ICONAT). https://doi.org/10.1109/iconat57137.2023.10080089

[20] Shukla, S., Maheshwari, A., Johri, P. (2021). Comparative Analysis of Ml Algorithms & Stream Lit Web Application. 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). https://doi.org/10.1109/icac3n53548.2021.9725496

[21] Kodur, V., Dwaikat, M., Fike, R. (2010). High-Temperature Properties of Steel for Fire Resistance Modeling of Structures. Journal of Materials in Civil Engineering, 22 (5), 423–434. https://doi.org/10.1061/(asce)mt.1943-5533.0000041

[22] Materials Algorithms Project. Available at: https://www.phase-trans.msm.cam.ac.uk/map/

[23] S. learn. Scikit-learn (BSD License).

[24] Keras. K. Keras, 'TensorFlow Core', TensorFlow 2.12.

[25] Leni, D. (2023). Prediction Modeling of Low Alloy Steel Based on Chemical Composition and Heat Treatment Using Artificial Neural Network. Jurnal Polimesin, 21 (5), 54–61. https://doi.org/10.30811/jpl.v21i5.3896

[26] Wang, Y., Wu, X., Li, X., Xie, Z., Liu, R., Liu, W. et al. (2020). Prediction and Analysis of Tensile Properties of Austenitic Stainless Steel Using Artificial Neural Network. Metals, 10 (2), 234. https://doi.org/10.3390/met10020234

[27] Xie, Q., Suvarna, M., Li, J., Zhu, X., Cai, J., Wang, X. (2021). Online prediction of mechanical properties of hot rolled steel plate using machine learning. Materials & Design, 197, 109201. https://doi.org/10.1016/j.matdes.2020.109201

[28] Clemente, F., Ribeiro, G. M., Quemy, A., Santos, M. S., Pereira, R. C., Barros, A. (2023). ydata-profiling: Accelerating data-centric AI with high-quality data. Neurocomputing, 554, 126585. https://doi.org/10.1016/j.neucom.2023.126585

[29] Suyatna, A., Anggraini, D., Agustina, D., Widyastuti, D. (2017). The role of visual representation in physics learning: dynamic versus static visualization. Journal of Physics: Conference Series, 909, 012048. https://doi.org/10.1088/1742-6596/909/1/012048

[30] Djajadi, A., Jaya, P. U., Muhida, R. (2017). Human Behavior Tracking with Vision Heatmap Toward Smart Building System. JTEC, 9 (2-8), 159–163. Available at: https://jtec.utem.edu.my/jtec/article/view/2648

[31] Adler, J., Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A, 77A (8), 733–742. https://doi.org/10.1002/cyto.a.20896

[32] Jia, S., Tan, Q., Ye, J., Zhu, Z., Jiang, Z. (2021). Experiments on dynamic mechanical properties of austenitic stainless steel S30408 and S31608. Journal of Constructional Steel Research, 179, 106556. https://doi.org/10.1016/j.jcsr.2021.106556

[33] Zieliński, A., Golański, G., Sroka, M. (2020). Evolution of the microstructure and mechanical properties of HR3C austenitic stainless steel after ageing for up to 30,000 h at 650–750 °C. Materials Science and Engineering: A, 796, 139944. https://doi.org/10.1016/j.msea.2020.139944

[34] Gemechu, W. F., Sitek, W., Batalha, G. F. (2024). Improving Hardenability Modeling: A Bayesian Optimization Approach to Tuning Hyperparameters for Neural Networks Regression. https://doi.org/10.20944/preprints202402.0312.v1

[35] Abellán, J., López, G., de Oña, J. (2013). Analysis of traffic accident severity using Decision Rules via Decision Trees. Expert Systems with Applications, 40 (15), 6047–6054. https://doi.org/10.1016/j.eswa.2013.05.027

[36] Sadorsky, P. (2021). A Random Forests Approach to Predicting Clean Energy Stock Prices. Journal of Risk and Financial Management, 14 (2), 48. https://doi.org/10.3390/jrfm14020048

[37] Munahar, S., Setiyo, M., Brieghtera, R. A., Saudi, M. M., Ahmad, A., Yuvenda, D. (2023). Fuel Control System on CNG Fueled Vehicles using Machine Learning: A Case Study on the Downhill. Automotive Experiences, 6 (1), 173–187. https://doi.org/10.31603/ae.8107

[38] Xiong, Z., Cui, Y., Liu, Z., Zhao, Y., Hu, M., Hu, J. (2020). Evaluating explorative prediction power of machine learning algorithms for materials discovery usingk-fold forward cross-validation. Computational Materials Science, 171, 109203. https://doi.org/10.1016/j.commatsci.2019.109203

[39] Brajer, N., Cozzi, B., Gao, M., Nichols, M., Revoir, M., Balu, S. et al. (2020). Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. JAMA Network Open, 3 (2), e1920733. https://doi.org/10.1001/jamanetworkopen.2019.20733

[40] Zayed, R. A., Ibrahim, L. F., Hefny, H. A., Salman, H. A., AlMohimeed, A. (2023). Using Ensemble Method to Detect Attacks in the Recommender System. IEEE Access, 11, 111315–111323. https://doi.org/10.1109/access.2023.3321322

[41] Wang, Z., Cao, C., Zhu, Y. (2020). Entropy and Confidence-Based Undersampling Boosting Random Forests for Imbalanced Problems. IEEE Transactions on Neural Networks and Learning Systems, 31 (12), 5178–5191. https://doi.org/10.1109/tnnls.2020.2964585

[42] de Souza, R. P., Sierra-Franco, C. A., Santos, P. I. N., Polonia Rios, M., de Mattos Nascimento, D. L., Barbosa Raposo, A. (2020). Automatic Deformation Detection and Analysis Visualization of 3D Steel Structures in As-Built Point Clouds. Lecture Notes in Computer Science, 635–654. https://doi.org/10.1007/978-3-030-49059-1_47

[43] Rodgers, T. M., Lim, H., Brown, J. A. (2019). Three-Dimensional Additively Manufactured Microstructures and Their Mechanical Properties. JOM, 72 (1), 75–82. https://doi.org/10.1007/s11837-019-03808-x

[44] Guo, K., Yang, Z., Yu, C.-H., Buehler, M. J. (2021). Artificial intelligence and machine learning in design of mechanical materials. Materials Horizons, 8 (4), 1153–1172. https://doi.org/10.1039/d0mh01451f

[45] Pollice, R., dos Passos Gomes, G., Aldeghi, M., Hickman, R. J., Krenn, M., Lavigne, C. et al. (2021). Data-Driven Strategies for Accelerated Materials Design. Accounts of Chemical Research, 54 (4), 849–860. https://doi.org/10.1021/acs.accounts.0c00785

Received date 11.03.2024
Accepted date 03.07.2024
Published date 10.09.2024

How to cite: *Leni, D., Karudin, A., Abbas, M. R., Sharma, J. K., Adriansyah, A. (2024). Optimizing stainless steel tensile strength analysis: through data exploration and machine learning design with Streamlit. EUREKA: Physics and Engineering, 5, 73–88. https://doi.org/10.21303/2461-4262.2024.003296*