JOIV : Int. J. Inform. Visualization, 8(3) - September 2024 1134-1145

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

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

Design of Tools for Visualizing Thermodynamic Concepts in Steam Power Plant Trainer Processes with Web-Based Exploratory Data Analysis (EDA)

Arwizet Karudin [a,*], Desmarita Leni [b], Remon Lapisa [a], Yuda Perdana Kusuma [c], Muhammad Rabiu Abbas [d]

[a] *Department of Mechanical Engineering, Engineering Faculty, Universitas Negeri Padang, Padang, Indonesia*
[b] *Department of Mechanical Engineering, Engineering Faculty, Universitas Muhammadiyah Sumatera Barat, Padang, Indonesia*
[c] *Department of Agricultural Engineering and Biosystems, Faculty of Agricultural Technology, Andalas University, Padang, Indonesia*
[d] *Department of Mechanical Engineering, Federal University of Transportation, Daura, Katsina, Nigeria*

*Corresponding author: *arwizet@ft.unp.ac.id*

Abstract*— Thermodynamics is considered one of the most complex and challenging subjects for many students. This is primarily due to comprehending abstract concepts such as entropy, enthalpy, and energy flow, which involve complex mathematical equations and are rarely accompanied by tangible visualizations. This research aims to design, develop, and test a data-based visualization tool for thermodynamics testing results. This study collected and processed data from thermodynamics testing and simulations, such as the mini-steam power plant trainer used as a teaching aid in thermodynamics education, as the foundation for designing a data-based visualization tool for thermodynamics concepts. The visualization tool was created using the Python programming language integrated with the web-based Streamlit framework. The designed visualization tool encompasses various features, including automated data reporting, visualization of variable correlations using correlation heatmaps, Sankey diagrams for visualizing energy flow, and the capability to predict electrical output using machine learning integrated with three different machine learning algorithms. The visualization tool was evaluated by thermodynamics experts using a Likert scale. Based on the results obtained, the experts gave an average score of 4 in the information accuracy aspect in the good category. This shows that the information displayed in this visualization tool is by thermodynamics learning at Padang State University. In the visualization aspect, experts gave an average score of 4.25, which is in the Good and Very Good range. In alignment with the education aspect, experts gave an average score of 3.75, which is close to the good category. This shows that this aspect is considered suitable for studying thermodynamics, although shortcomings still need to be corrected. Experts gave a relatively high assessment of the Ease-of-Use aspect, with an average score of 4.5, with a range of Good and Very Good. This enables students to better understand complex patterns, cause-and-effect relationships, and parameter changes within thermodynamics concepts.

***Keywords*— Design; visualization; thermodynamics; exploratory data analysis.**

Manuscript received 22 Sep. 2023; revised 19 Jan. 2024; accepted 18 Feb. 2024. Date of publication 30 Sep. 2024.

I. INTRODUCTION

Thermodynamics is a branch of physics that deals with systems in equilibrium, which can be used to determine the amount of energy required to change a system from one state of equilibrium to another. However, thermodynamics cannot be used to determine the rate of change that occurs during processes when the system is not in equilibrium. Such a system can change because of the surrounding environmental conditions. The material system in thermodynamics can receive heat energy or energy in various forms. A good understanding of thermodynamics allows students to optimize processes such as steam power plants, engine cooling, and the design of renewable energy systems [1], [2]. Field observations at the State University of Padang found that in the thermodynamics course, there is teaching material that examines the working process of steam turbines. However, previous research has found that thermodynamics is often considered one of the more complex and challenging themes for many students to understand [3], [4], [5]. This is partly due to understanding abstract concepts such as entropy and enthalpy, complex mathematical equations, and the lack of concrete visualizations. As a result, students must rely on mathematical representations that are sometimes difficult to grasp. Therefore, more interactive learning methods and effective teaching media are needed to visualize thermodynamic concepts.

The rapid and dynamic development of computer technology and data science is greatly facilitated by widespread internet usage and web-based technologies within the academic community. This phenomenon enables the integration of thermodynamics concepts with data in a web-based format, providing students with authentic experiences of how changes in one parameter can impact others. According to research findings [6], there has been a shift in learning styles among students due to extensive exposure to computers and the Internet. As a result, students tend to prefer learning with the aid of visualizations [7]. This has prompted educators to adapt teaching methods to align with visual learning methods familiar to technology-savvy students. To enhance thermodynamics education in an era where students are accustomed to technology-driven learning styles, there is a need for a tool that visualizes the data from thermodynamics testing and simulations. The primary goal is to improve students' understanding of thermodynamics concepts, particularly in the context of mini-steam power generation. In the process of steam power plant trainer apparatus, which serves as an educational tool in thermodynamics, monitoring parameters such as temperature, pressure, volume, energy, and other variables is often required.

Exploratory Data Analysis (EDA) in statistics and data science is a visual and descriptive data-exploration process aimed at drawing conclusions and raising questions about a given dataset before performing analysis [8],[9],[10]. It aims to uncover patterns, relationships, anomalies, and insights present in the data that might have been difficult to spot before the analysis occurred [11]. In web-based EDA, system performance can be visualized in real-time or as a series of intervals, allowing students to observe the change in parameters over time [12]. In EDA, it's also possible to build an interactive simulation through which a user can directly manipulate key system parameters. For example, in EDA for a thermodynamic system simulation, the user can set the values of temperature and pressure and observe how system performance changes due to these changes. This capability can deepen students' understanding of the concepts associated with the thermodynamic system, such as the role of temperature, pressure, and energy changes.

Taking that into account, understandably, we run into issues with the concepts of temperature and pressure, as explained earlier. All in all, the main objective of this research is to create a tool for visualizing thermodynamic concepts in the form of a web-based interface based on the knowledge from thermodynamic testing. This way, students can grasp more challenging concepts through a visual representation of data by getting the chance to observe and recognize relations of thermodynamic variables, find correlations between variables, and run simulations based on machine learning methods to predict the electric energy output of a steam turbine. Being able to play with data in this manner will help students explore the relationships between thermodynamic variables, understand more complex concepts, and apply that knowledge to run simulations for predicting the outcomes of steam turbines. Consequently, this research will be approached using a visualization tool for the most effective and supportive form of learning when it comes to thermodynamic concepts, thus giving students an exciting experience to comprehend thermodynamic phenomena and their application to the process of generating steam power.

II. Materials and Method

This research is research and development that aims to design, develop, and test thermodynamic concept visualization tools. This research focuses on creating and evaluating visualization tools integrated with web-based EDA to visualize thermodynamic concepts. This visualization tool was created using the Python programming language, while the web interface was designed using the Streamlit framework [13], [14]. The data used in this research are the steam turbine trainer testing results by mechanical engineering students at Padang State University. This visualization tool has three main features, namely: (1) automatic checking using Pandas profiling [15], so that users can quickly identify potential errors or anomalies in the dataset; (2) visualization with various diagrams, including Sankey diagrams to visualize energy flows and correlation analysis between variables using correlation heatmaps; (3) the electricity output prediction feature using machine learning methods; this allows students to test various scenarios and see how changes in input variables can affect the resulting electricity output. The design of this visualization tool consists of several stages, as can be seen in Figure 1.

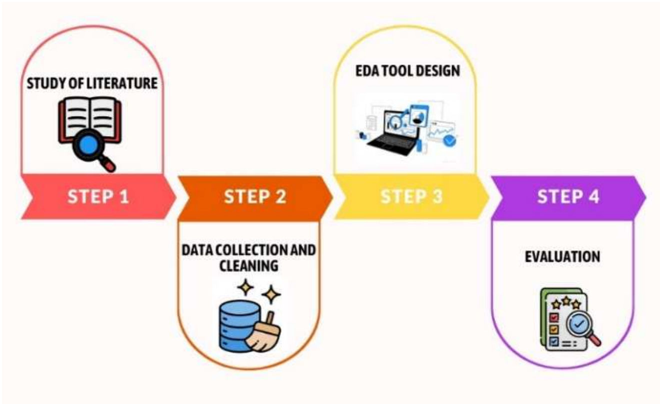

Fig.1 Research Method

A. *Literature Review*

Literature study is a series of activities related to methods of collecting library data, reading and taking notes, and managing research materials, such as looking for various reference sources related to thermodynamic concepts in steam power plants, EDA, and understanding how to apply the Streamlit framework with the Python library. This stage aims to understand better the theoretical and practical basics related to the visualization tools that will be designed.

B. *Data Collection and Data Cleaning*

Data collection and cleaning are the initial stages of research, where good data collection can ensure adequate data quality for analysis. Data cleaning helps ensure that the data is free from errors or anomalies that could affect the analysis results [16], [17]. This research creates a breakthrough by utilizing data from students' independent practice testing as a means of interactive learning. So far, data from testing and experiments in the steam turbine process has only been used as basic knowledge without being collected or used actively to improve students' understanding of thermodynamic concepts. The data used in this research comes from the steam turbine trainer testing results carried out by mechanical engineering students at Padang State University over the last three years, from 2020 to 2023. This data is a routine assignment for students taking thermodynamics courses at Padang State University. Students use a steam turbine as a medium to simulate thermodynamic phenomena such as energy changes, entropy, and the Rankine cycle as in the process of steam power generation. In the learning process, students are generally divided into groups of 5 students, where each student works together to test and collect data on various components of steam power plants such as boilers, fuel, turbines, and condensers for further analysis. In this research, this data was used as a basis for creating a thermodynamic visualization tool for steam power plants. However, before further analysis is carried out, the data is cleaned first, such as removing invalid values.

C. *Design of Visualization Tools*

The subsequent phase involves the design and integration of visualization features into a web framework that is based on Streamlit, following the acquisition of the data and its subsequent cleansing. The visualization features include machine learning models, correlation analysis with correlation heatmaps, energy flow visualization with Sankey diagrams, and automatic data examination with the Pandas profiling and Sweetviz modules. Two stages comprise the design of this visualization tool, as illustrated in Figure 2.

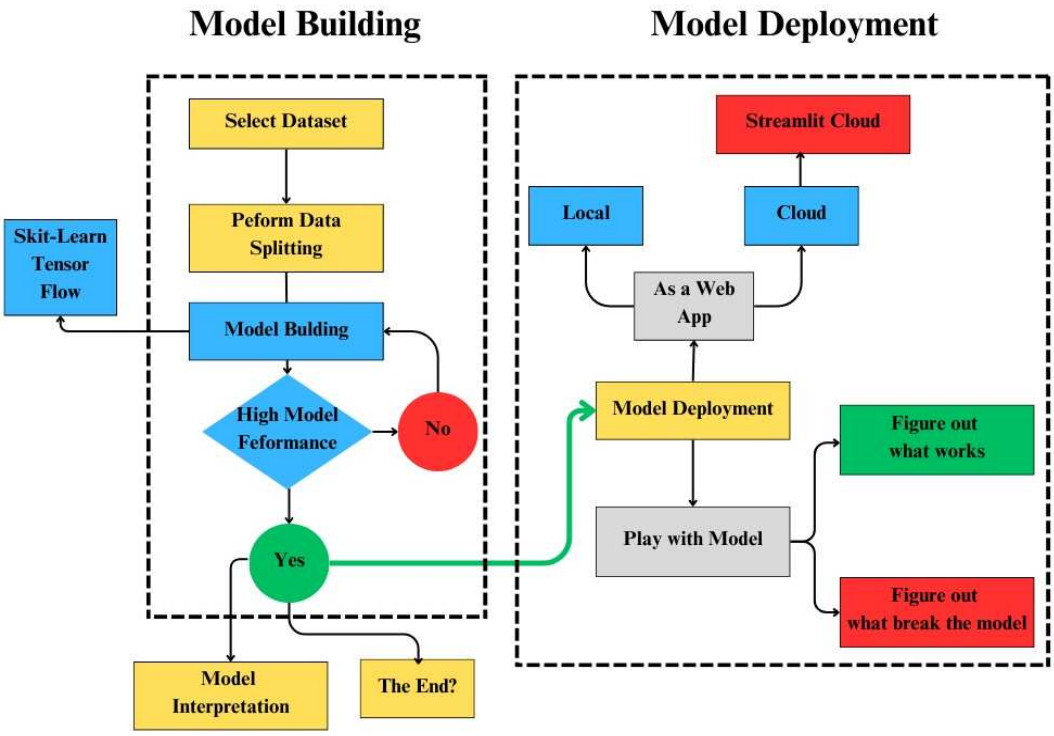

Fig. 2 Schematic of the visualization tool design

In the first stage, a machine learning model was designed using 3 algorithms commonly used for prediction, namely Decision tree (DT), Random Forest (RF), and Artificial neural network (ANN). This algorithm is implemented using the default parameters provided by the scikit-learn library version 1.2.2 [18] and Keras version 2.12.0 [19] for ANN, as can be seen in Table 1. The data used for training and testing the model is divided into two parts, namely 80% training data and 20% testing data. Each machine-learning model is validated using cross-validation. This technique allows the training data to be divided into several subsets or folds, and iteration is carried out on each subset to be used as test data. In contrast, the other subset is used as training data. The selection of default parameters in the design stage of this machine learning model is based on considerations aimed at providing ease of use and a good understanding for students in utilizing machine learning features. This is intended to enable students to comprehend the fundamental concepts of machine learning without delving too deeply into complex parameter tuning. With these default parameters, it is expected that students can focus their attention on the learning process of machine learning concepts and the interpretation of model prediction results.

TABLE I
MODEL PARAMETERS

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	Optimizer	Adam
Min Samples Leaf	1	min_samples_split	2	Learning rate	0.001
		min_samples_leaf	1	Model	Sequential
		max_features	Auto	Dense (layer 1)	Units: 32 Activation: relu
		bootstrap	True		
				Dense (layer 2)	Units: 16 Activation: relu
				Dense (output layer)	Units: 1 Activation: linear

The machine learning model is evaluated using evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2). The selection of these three metrics is designed to provide holistic and comprehensive information about the performance of the electricity power prediction model. These evaluation metrics have also been employed in previous research on predicting electricity output in steam power plants [20], [21], [22]. MAE provides an understanding of overall prediction accuracy, RMSE emphasizes the handling of large errors. At the same time, R-squared gives an insight into how well the model can explain variations in power data. By utilizing these three metrics, the model evaluation can be conducted more effectively from various perspectives relevant to the research objectives and practical applications in the field. Model evaluation metrics can be calculated using the following equations [23], [24]:

1) Mean Absolute Error (MAE): MAE measures the average of the absolute differences between the model's predictions and the target values. The lower the MAE value, the better the model makes predictions. MAE can be calculated using Equation 1.

$$MAE = \frac{i}{N}\sum|y_i - z_i| \quad (1)$$

Where '*i*' represents the index of data samples, 'N' stands for the total number of samples, '*y_i*' corresponds to the actual data values for the i-th sample, and '*z_i*' represents the predicted values generated by the model for the *i*-th sample.

2) Root Mean Square Error (RMSE): RMSE is the root of the mean squared difference between the model's predictions and the target values. RMSE measures the same units as the target variable and is generally more sensitive to significant differences. Lower RMSE values indicate better predictive performance by the model. RMSE can be calculated using equation 2.

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

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

3) R-squared (R^2): R-squared (R^2) is the coefficient of determination that provides information about how well the model fits the data. R^2 is the ratio of the total variation the model explains to the total variation present in the data. The R^2 value ranges from 0 to 1, with higher R^2 values indicating a better model's ability to explain the variation in the data. R-squared can be calculated using equation 3.

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

where f (Xi) represents the predicted value of the dependent variable (Y) based on the independent variable (X) for the i-th observation, f ($\bar{X}$) is the mean of all predicted values f (Xi) across all observations, Yi is the actual observed value of the dependent variable for the i-th observation, $\bar{Y}$ is the mean of all observed values Yi across all observations, and n is the total number of observations.

In the second stage, all visualization and machine learning features are integrated into the web-based Streamlit framework. The visualization tool's interface is user-friendly, with a simple layout, making it accessible to users who may not be familiar with data exploration and machine learning.

D. Evaluation

The thermodynamics visualization tool that has been designed will be evaluated through a series of tests involving two main aspects: (1) internal testing to ensure that all functions and components work as expected and (2) assessment to be submitted to experts in the field of thermodynamics to evaluate the accuracy of information, visualization quality, alignment with educational objectives, and ease of use using a Likert scale.

III. RESULTS AND DISCUSSION

Based on field observations at the State University of Padang in the mechanical engineering education program, the thermodynamics course includes learning materials related to the operation of steam turbines. Typically, thermodynamics learning is conducted through in-class sessions with theoretical content delivery and a miniature power plant trainer, such as the one depicted in Figure 3.

However, up to this point, the data from testing and experiments on the mini power plant trainer has been used primarily for bare knowledge and has not been actively collected or utilized to enhance students' understanding of thermodynamics concepts. Data can actively serve as an interactive learning tool [25].

Fig. 3 Steam Power Plant Trainer

With increasingly advanced information technology and visualization aids, students can easily access and visualize the data obtained from mini-power plant experiments through graphs, diagrams, and tables. This enables students to more profoundly observe and comprehend complex patterns, cause-and-effect relationships, and parameter variations within thermodynamic concepts. The active use of data in visualization also empowers students to engage in self-directed exploration [26]. Furthermore, this approach aligns with the contemporary trend where data is actively employed for decision-making, analysis, and innovation across various industries and academic disciplines [27].

A. Data Collection and Data Cleaning

Data was collected from the results of testing and previous student simulation assignments regarding steam turbine trainers. This data amounted to 312 samples with 9 variables consisting of 8 input variables and 1 output variable, as can be seen in Table 2.

TABLE II
STATISTICS OF STEAM POWER PLANT TRAINER TEST DATASET

Parameter	Variable	Min	Max	Mean
Steam pressure inside the boiler (bar)	Input	3.44	4.36	3.9
Boiler steam temperature (°C)	Input	141.37	151.31	147.62
Fuel consumption (L/h)	Input	50	51	50
Turbine RPM	Input	1247.77	1285.164	1260.44
Inlet turbine temperature (°C)	Input	109	123	113
Outlet turbine temperature (°C)	Input	96	106	100
Inlet turbine pressure (bar)	Input	2.79	3.58	2.72
Outlet turbine pressure (bar)	Input	0.05	1	0.4
Generator Output (Watt)	Output	2.22	3.41	2.73

However, after cleaning the data, the numbers were reduced to 300 samples. This was caused by missing data, duplicates, and a scale much different from other data, so it was removed from the dataset. Apart from the steam turbine trainer test data, energy flow data was also obtained for each PLTU component, such as the boiler, turbine, condenser, and generator. Removing data from a dataset is an essential stage in data processing, known as data cleaning. This process aims to ensure that the data used in analysis or modeling is of good quality and reliable [16], [17].

B. Design of Visualization Tools

Exploratory Data Analysis (EDA) is an initial investigation process to identify patterns, discover anomalies, test hypotheses, and examine assumptions. Through EDA, users can detect errors early on, identify outliers, understand relationships between data, and explore essential factors within the data. The EDA process includes calculating various basic statistical values, visualization, hypothesis formulation, assumption checking, and storytelling and reporting. Additionally, EDA involves handling missing values, outliers, dimensionality reduction, clustering, transformation, and data distribution.

After obtaining the data, the next step is to perform data analysis before creating the machine learning model. The test data from the steam power plant trainer is analyzed using a correlation heatmap to visualize the relationships between the input variables and the output variable. A correlation heatmap is a visual tool used to display the correlation between multiple variables by using different colors to indicate the strength and direction of the correlation between two variables in the data [28], [29]. The correlation heatmap uses the Pearson correlation coefficient as its correlation measure. The Pearson correlation coefficient can range from -1 to 1, where -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship [30]. The correlation results for the steam power plant trainer test data can be seen in Figure 4.

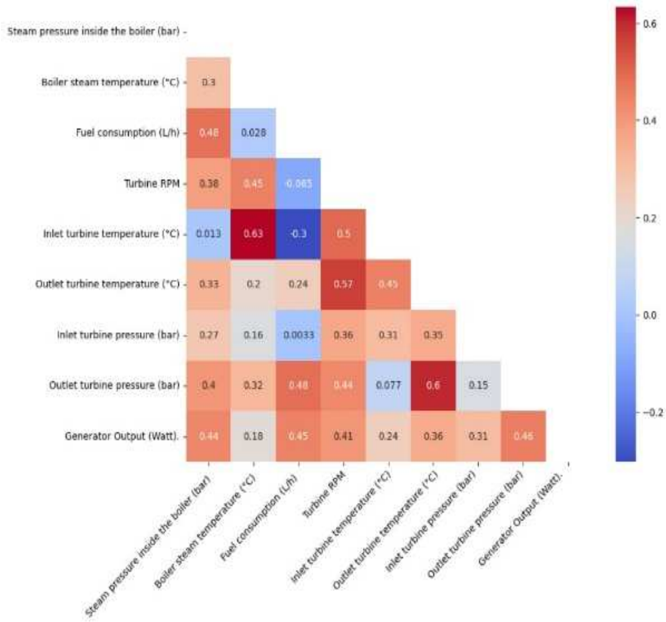

Fig. 4 Heatmap of correlation for steam power plant trainer test data

Based on Figure 3, it can be seen that almost every input variable has a positive correlation with the output variable (Generator Output). The outlet turbine pressure and turbine RPM parameters have strong positive correlations with the electrical power generated by the generator, with values of 0.46 and 0.41, respectively. Meanwhile, the parameter with the smallest correlation to the input variables is the steam temperature inside the boiler, with a correlation of 0.18. This is consistent with the findings of a previous study [31], [32], [33] which indicated that the outlet turbine pressure can influence the turbine's rotational speed. Higher turbine rotational speeds result in greater electrical output from the generator.

The next step involves machine learning modeling to predict the electrical power output generated by the steam power plant trainer using three algorithms: DT, RF, and ANN. Default parameters provided by scikit-learn version 1.2.2 [18] and Keras version 2.12.0 for ANN [19] are used for modeling. Machine learning modeling is designed based on the previously obtained data, which consists of 8 input variables and 1 output variable. The data is divided into two parts for training and testing: 240 and 60 samples for training. Cross-validation with K=10 is used for evaluation, and three evaluation metrics are employed: MAE, RMSE, and R-squared. The prediction results for the three machine learning models can be seen in Figure 5, and the cross-validation results for the three machine learning models are shown in Figure 6.

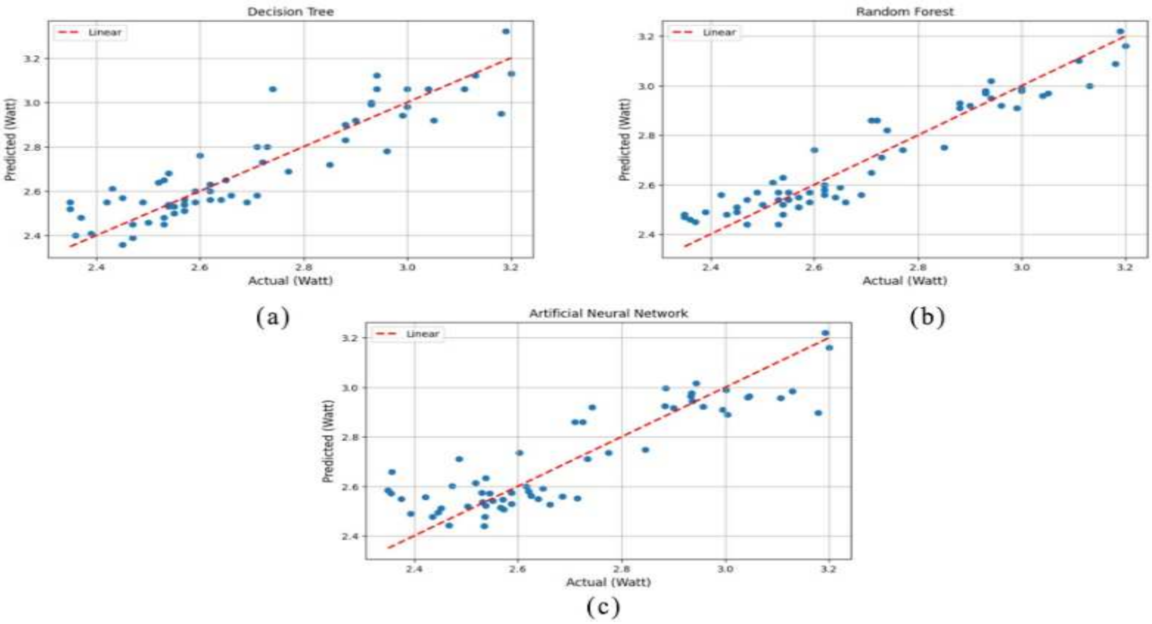

Fig. 5 Prediction results of (a) DT, (RF), and (c) ANN models

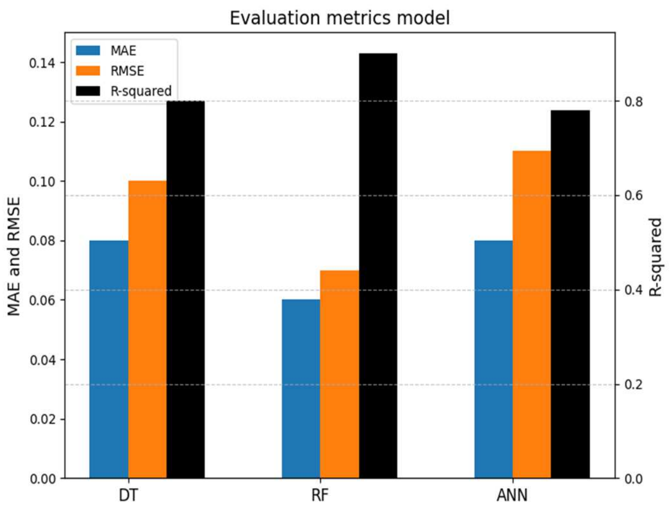

Fig. 6 Model Evaluation Results Using Cross-Validation K=10

Based on the validation results using the cross-validation method with K=10, this research indicates that out of Random Forest (RF), Decision Tree (DT), and Artificial Neural Network (ANN), RF performs the best and has the most efficient results. We can visualize this by looking at the chart below: RF scores higher than both DT and ANN in the R-squared value, reaching 0.90, while DT and ANN record R-squared values of 0.80 and 0.78, respectively. If we transfer this data into percentages, we will see that RF predicts generator output with a higher accuracy when compared to the other two models, by an R-squared difference of 12% compared to DT and 12% compared to ANN. In addition, Random Forest scores the lowest in MAE and RMSE values, with each point being 0.06 and 0.07, respectively. DT comes next to it, with an MAE of 0.08 and RMSE of 0.10, and ANN has the highest MAE and RMSE values – 0.08 and 0.11, respectively. This proves that RF not only predicts generator output with a higher degree of accuracy but also carries a lower error rate when compared to both DT and ANN.

It is obvious from the chart that RF is capable of high generalization, which we can also notice from the R-squared value of 0.90, indicating that RF can explain up to 90% of the variation in present data. This explains an ability to learn underlying patterns in the data. RF is far superior to DT and ANN. It mainly shines in ensemble, model stability, and scalability. The ensemble is one of the core features of RF, where we already know DT has a considerable downside to the effect that the decision tree can easily be influenced by noise or arbitrary statistical/ accidental relationships in the training data and pick up those patterns as well. This causes the model to overfit. Luckily, RF introduces us to an ensemble of DTs trained separately, and a voting or averaging procedure is done for the final result, reducing overfitting and improving model performance. DT overfits because it relies on a single DT, unlike RF, which relies on an ensemble of DT [34]. In addition to this benefit, RF is also known to be more model stable than most decision tree models. Model stability is even more important than determining if the model is good. Sometimes, small changes in input at random (usually by adding or removing a few observations), even when these samples were not included before, can cause a dramatic change in a model based on a single decision tree [34] DT changes as we change the training data slightly, but RF sustains the results. This means that RF is less likely to be a trick of the current data at hand since it provides strong results even with minor changes in data. Finally, RF is more accessible to implement by providing good results even with small data. RF is also scalable and uses many DTs. On the opposing side, ANN requires a large amount of data for successful training, and since it has deeper architectures, more computational resources are needed, too.

These results align, moreover, with the exploration of solar power plant prediction by random forests [35] That indicates that RF uses ensemble learning concepts since the model is based on many decision trees combined randomly. This further increases the robustness of the model to overfitting. Furthermore, in the case of making predictions about hourly solar power generation, the weather input in solar power plant prediction contributes to uncertainty and errors. Therefore, RF proves robust in such cases as it is relatively stable to errors in weather input data. Random Forest quantity response was stable on diverse weather data. The reliability of the RF approach is further confirmed by the algorithm's robustness to variations and diversity in weather data. Since the algorithm has random subsets of features and training data (data samples), using RF can reduce the model's variance and consequently help improve performance and provide a better-fitting model on unseen data.

Finally, another study [36] the RF approach can discover the most significant and explanatory features required for an hour-ahead power price prediction in the electricity market. This study automatically quantifies and evaluates the relative importance of various weather features to electricity prices through the RF algorithm. By doing so, the study identifies the most relevant variables in determining electricity prices. According to the results, the RF algorithm can model the power consumption data and demonstrate explanatory variables with primacy in determining electricity prices. The analyzed variables include load, hydro, thermal production, and wind energy production. The research confirms these results [37] that explores the performance of RF as an option for predicting country-level biomass energy indicators. The study points out that RF proves effective in predicting several human activities, including crop production or malaria diseases, that cause data to be varied in types, including categorical and numerical. All these findings are also consistent with previous studies [35], [38] that report the robustness of the Random Forest approach in achieving superb performance in electricity generation prediction. These earlier studies reported that RF's success in handling complexity and non-linearity in data content can influence the goodness of the predictions made with this algorithm.

After completing the modeling process using three machine learning algorithms, the next step involves integrating all features of this data-driven thermodynamics concept visualization tool into the Streamlit framework. This is in line with a study [39] that indicates using simulations can enhance the quality of learning but requires modifications to the simulation design and lesson sequence to maximize its impact. The study discusses the utilization of computational simulations in online learning of advanced placement physics at the high school level. The focus is on how students use simulations to generate quantitative data and then analyze, interpret, and model thermodynamic concepts.

The research results show that some students experience an improvement in conceptual understanding of the properties of matter particles and diffusion. Students find simulations helpful in explaining complex concepts, and the data generated by simulations facilitates understanding. In another study [40], there's mention of a new e-learning-based educational package called TermolabUA, consisting of three programs: Volcontrol for steady-state flow device analysis, CarnotCycle for analyzing reversible and irreversible processes, and CombustionUA for studying combustion processes. This educational package is designed to help undergraduate students achieve cognitive competence in understanding various thermodynamics topics. The t-student test results indicate that the average scores obtained by students using the software are higher than the average scores without using the software.

Therefore, this web-based visualization tool is designed with an interface that includes a sidebar and content page. The sidebar navigates, enabling users to easily access various features and pages, such as the Exploratory Data Analysis (EDA) page. The EDA page comprises visualization features designed to provide a deep dataset understanding. These features involve automated data inspection using pandas profiling modules, allowing users to identify potential errors or anomalies in the dataset quickly. According to a study [41] Using Google Colab notebooks to teach thermodynamics, students are guided to code to facilitate learning, such as creating simulations and visualizing problems. The research findings indicate that Google Colab notebooks can be an effective tool for enhancing thermodynamics learning.

Additionally, there are correlation heatmap features that visualize relationships between variables, pair plots to explore data distributions and manual graphs like scatterplots that offer flexibility in data analysis. Users can freely access these features by using the sidebar as a navigation guide to gain better insights into thermodynamics concepts. The EDA page features can be seen in Figure 7, which can provide a comprehensive learning experience for students in understanding thermodynamics concepts through data analysis and informative visualization.

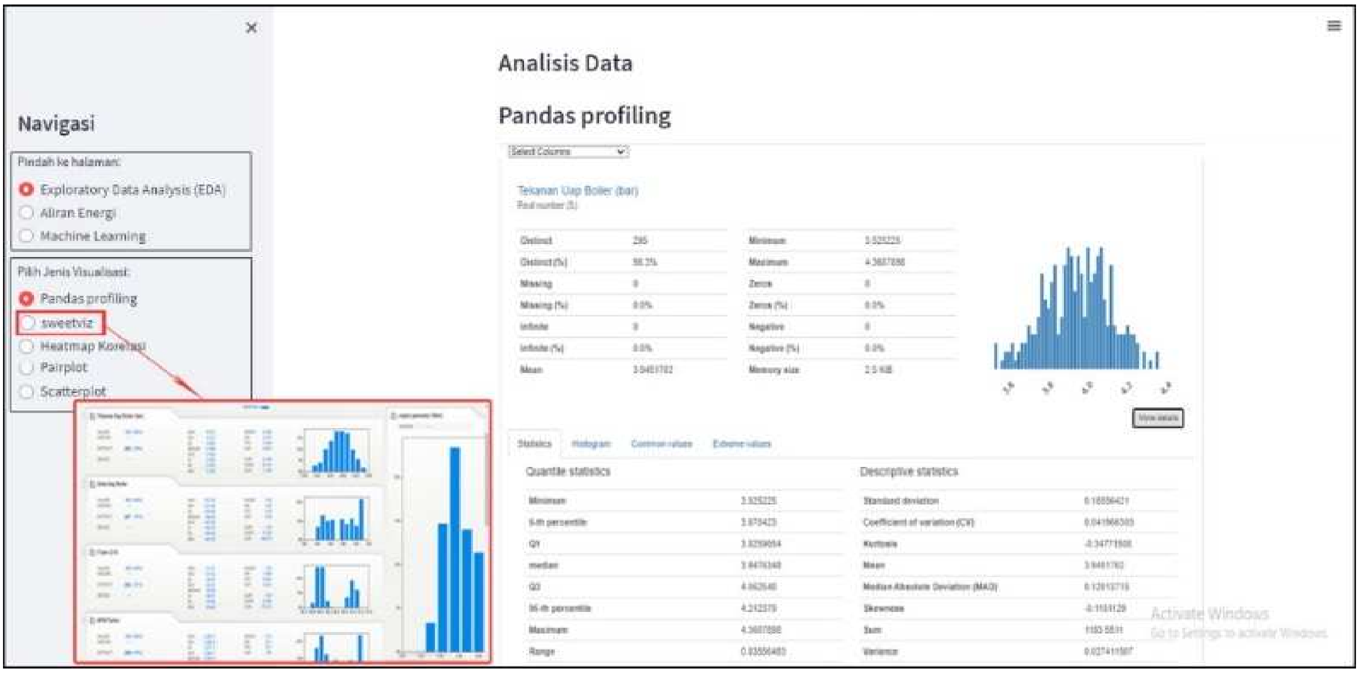

Fig. 7 Exploratory Data Analysis (EDA) Page

On the second page or in other features of this visualization tool, there is a Sankey diagram graph feature, a visual element that plays a very significant role in understanding thermodynamics concepts, especially in the energy flow processes within a system. The Sankey diagram plays a crucial role in providing a visual representation of how energy enters and flows through a thermodynamic system [42], [43]. In this representation, the thickness of the arrows on the Sankey diagram reflects the amount of energy flow, allowing students to visualize how energy moves through various system components. The Sankey diagram graph in this visualization tool is designed to be an interactive learning tool, enabling students to understand the extent to which energy is used or lost in each component of the thermodynamic system. By examining the different widths of lines on the Sankey diagram, students can gain profound insights into energy distribution and how it transfers between system components. The view of the Sankey diagram feature can be seen in Figure 8, serving as one of the visual elements that support the learning of thermodynamics concepts in a clear and informative manner.

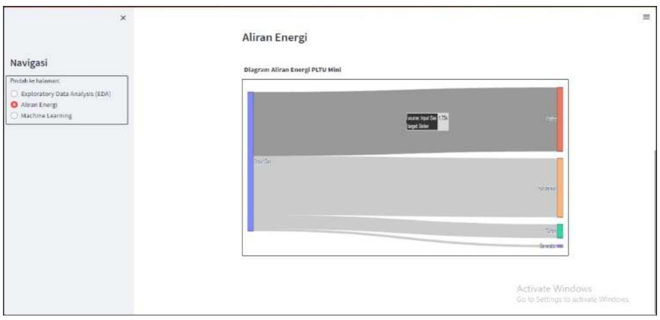

Fig. 8 Sankey Diagram Feature Display

On the third page of this visualization tool, there is a machine-learning modeling feature specifically designed to predict the electrical power output generated by the generator. Users can choose the desired algorithm in this feature, such as Decision Tree, Random Forest, and Artificial Neural Network. The page design is deliberately simple, focusing on user-friendliness, especially for those unfamiliar with machine learning methods. The sidebar on this page displays input variables that users can adjust, such as steam pressure in the boiler, boiler steam temperature, fuel consumption, turbine RPM, turbine inlet temperature, turbine outlet temperature, turbine inlet pressure, and turbine outlet pressure. Users can select the algorithm that suits their needs on the right side of the page. The results of the generator's electrical power prediction will appear in the content section of the page, accompanied by units of measurement in watts. Furthermore, this page also provides information about model evaluation metrics, such as MAE, RMSE, and R-squared. Visual representations of these evaluation metric values can be found in Figure 9, offering users direct insights into how well the machine learning model can predict the generator's electrical power.

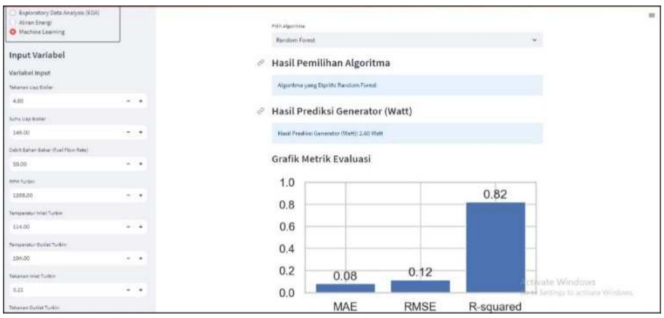

Fig. 9 Machine Learning Model Page Display

With the combination of these features, this page aims to provide an informative and in-depth learning experience for students in understanding thermodynamics concepts through the practical application of machine learning modeling. This visualization tool embodies the concept that data originating from practical work, testing, and scientific simulations, as found in thermodynamics concepts, can play a critical role in interactive learning. The interactive learning approach is an educational method in which students are not merely passive recipients of knowledge but actively participate in the entire learning process [44]. This visualization tool is designed to provide students with insights showing that data is not merely the result of experiments but also an active learning resource. Through this visualization tool, students are expected to comprehend that data plays a role beyond being the output of an experiment. Data also serves as the foundation students use to gather information, clean and analyze data, and make decisions. Therefore, this visualization tool encourages students to take an active role in their learning process, build analytical skills, and understand the significance of data as a dynamic and interactive learning tool. Specifically, integrating data from experiments and simulations of thermodynamics learning devices, such as steam turbine trainers, with a web-based thermodynamics concept visualization tool establishes a robust connection with previous research. This method enables students to comprehend concepts theoretically and experience their practical application through data visualization, acting as a bridge between theory and practical experience.

The outcome was that machine learning works to analyze problems in thermodynamics, especially in modeling nonlinear relations and multi-dimensional data, out of the reach of traditional mathematical functions. The outcome of this research was similar to Ding [45] who has used machine learning methods to predict the thermodynamic properties of several compounds from available data? The results show that machine learning can be applied in molecular thermodynamics to predict and investigate properties and behaviors commonly design issues in chemical engineering real-life applications. Such machine learning methods can better understand the role of ions and molecules in complex systems. Nonetheless, implementing machine learning in chemical engineering is still at the beginning due to the costs of acquiring datasets as one of the primary application areas in this research field. In another study, the title of the paper was based on extraction research, Funai [46] about how the Restricted Boltzmann Machine (RBM) can model phases of matter in thermodynamics. RBM was trained by data samples obtained from spin configurations taken by repeated sampling from the Ising Hamiltonian at various temperatures in the presence of an external magnetic field by using Monte Carlo methods. The results allow an understanding of how physical phase transitions can be identified through the machine learning method and then by identifying the distinctive properties of configurations, such as maximization of specific heat or correlation length. The outcome of this manuscript makes it clear that the RBM does not directly relate to the renormalization group (RG) flow and its fixed points. Jirasek [47] discussed using machine learning methods in thermodynamics to predict activity coefficients of binary liquid mixtures. Activity coefficients are calculated from Gibbs-Duhem relations, resulting in Henry's law constant prediction, which is a nonlinearity issue in systems with excess entropy. Using a probabilistic matrix factorization modeling, the outcome of this study shows an advantage without physical descriptors. This predictive method for activity coefficients in many binary liquid mixtures, including solutions of hydrofluorocarbons that are widely used as refrigerants, is far better than current methods whose methodology has been developed during the previous decades, without using physical descriptors for its modeling components.

Using machine learning without physical descriptors provides a new perspective on the possibility of predicting thermodynamic properties without relying on traditional models that may require complex physical descriptors. Based on the findings of this research and previous studies, it can be observed that machine learning methods have a significantly positive impact on modeling, visualizing, and explaining thermodynamic issues very effectively. Integrating machine learning with Streamlit has opened new opportunities for data-driven learning models in thermodynamics. Previously, data from experiments, testing, and simulations were only used for validation, but now, this data can be utilized for interactive learning about steam power plants in thermodynamics courses.

This visualization tool not only holds significant potential in the context of formal education but can also make a substantial contribution beyond academic environments, particularly in industry, research, and development. The visualization tool can be utilized to train the workforce in specific industries, helping them grasp thermodynamics concepts and processes relevant to their work [48]. It serves to understand and analyze data in industrial research. For instance, in industries involving thermal processes, such as steam power plants, this tool can assist researchers in comprehending changes in energy, entropy, and other thermodynamic factors. The use of machine learning algorithms to predict electrical power output can also be applied in industrial settings, aiding companies in planning and optimizing power plant operations, as shown in [20], [21], [22]. By combining this visualization tool's strengths, providing clear visual insights and robust data analysis capabilities, it is expected to have a positive impact in various contexts beyond formal education.

The visualization tool integrated with the Streamlit framework significantly impacts both industry and research, particularly in the context of steam power plants. Essentially, the demand in the steam power plant industry is to achieve maximum electrical power production even under various environmental conditions. Variability in conditions such as temperature fluctuations, air pressure, humidity, and other factors significantly affects the performance of critical components in steam power plant systems [49],[50]. High air temperatures, for instance, can result in increased temperatures in various key components such as turbines and condensers. This temperature rise can detrimentally impact heat transfer efficiency and overall system performance, potentially leading to a decrease in power production. Conversely, low-temperature conditions can cause significant fuel wastage, especially when heating water in boilers, which may increase operational costs.

C. Evaluation of Visualization Tool

The evaluation of these tests was categorized into two aspects, namely internal testing and assessment conducted by experts, who are the structural relationships that mainly learn the subject material of thermodynamics, which is by correlation of the structural relationships that mitigate this tool for the students and anyone who desires to understand thermodynamics. Internal testing refers to the evaluation addressing whether the stimuli visual in this project is in line with the target for this visual, which is being able to present a range of data associated with thermodynamics from a clear, easy, and concise viewpoint so that the students can easily apply the pieces of information related to thermodynamics to understand the subject of thermodynamics [51], [52]. External testing refers to the evaluation and assessment conducted by experts of strong relationships who learn the subject material of thermodynamics to see whether the students and anyone who utilizes the visual can carry out the theoretical calculations related to thermodynamics based on the data provided by this tool. The calculation is performed following the calculations formula of thermodynamics that has been fixed since the early twentieth century, which is used for a wide range of materials by various experts related to the chemistry and physics fields. This type of visualization tool was assessed and tested by the two subjects above, including the tutors, to see whether the visual stimuli of this project match the purpose of the core impression and whether the visual can visualize stimulation of relations of data of thermodynamics, which are derived from two sources, which are data simulation and data experimentation, for the students. We tested the tool of the visual and received good results. It can be compiled quickly and efficiently. The time it takes to load and transition between features is not long and is happening smoothly without any crashes, moving back to places where they came from, or repeating two particular errors in the tool. In more detail, the tool shows stability clearly in morphology. Data compilations of stimuli experiments, whether from the simulation or experiments that the students conducted, were carried out successfully with no problem.

The second phase evaluates the visualization tool using a Likert scale by experts to assess the correctness of the information being presented, the effectiveness of visualizations, the integration of the educational aspect, and the ease of use of the visualization tool. Four lecturers of thermodynamics courses at the University of Padang conducted the evaluation process. Figure 10 shows experts' evaluation of the visualization tool.

Based on the results obtained, it can be seen that the experts gave an average score of 4 in the Information Accuracy aspect in the good category. This shows that the information displayed in this visualization tool is by thermodynamics learning at Padang State University. In the Visualization aspect, experts gave an average score of 4.25, which is in the Good and Very Good range. In alignment with the education aspect, experts gave an average score of 3.75, which is close to the good category. This shows that this aspect is considered appropriate for thermodynamics learning, but shortcomings still need improvement. Experts give a relatively high score to the Ease-of-Use aspect, with an average score of 4.5, in the Good and Very Good range. Overall, the experts provided very positive assessments, indicating that the visualization tool has good quality in conveying thermodynamics concepts, especially regarding information accuracy and ease of use. Although alignment with learning approaches is a good category, there is still room for improvement in enhancing relevance to the learning context.

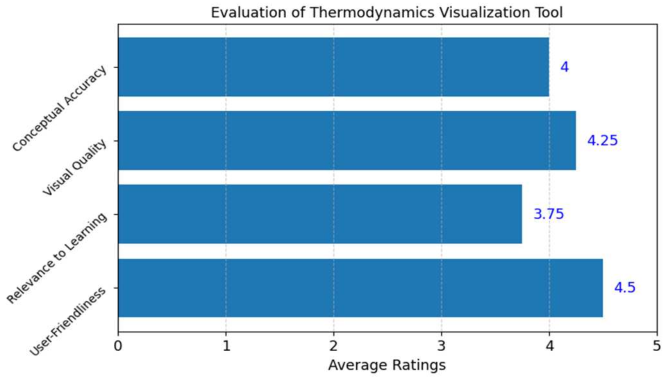

Fig. 10 Expert Evaluation Results of the Visualization Tool

This visualization tool is specially designed for learning in thermodynamics courses. It assists students in illustrating how thermodynamic phenomena occur in steam power plant systems. The data used for machine learning modeling is derived from testing and simulation in small-scale steam power plant systems. Therefore, to implement machine learning on an industrial scale, adjustments to the dataset are needed to meet specific industrial requirements. It is essential to emphasize that the data used to train the model is still very limited. This is mainly due to the previous paradigm that viewed data merely as a tool to validate research results [53]. In future research, it is expected that campus laboratories, particularly those involved in thermodynamics testing, will consider data an active learning source, especially in interactive learning using this visualization tool. A new understanding of the role of data can open opportunities to expand and enrich the tool's usage in various learning scenarios [54].

Based on previous research, it was found that a common challenge in implementing machine learning methods is the limitation of data. Privacy issues cause this [55], high costs, and limited access to related databases [56]. Nevertheless, in some cases, it has been discovered that this issue can be addressed by using synthetic data [57],[58]. Synthetic data is intentionally created or generated to increase the quantity or variation of available data. This data does not come from observations or real-world data collection but is created using various techniques and methods, such as generative algorithms or data manipulation. The primary goal of using synthetic data is to enhance the number of samples available for model training, especially in the context of machine learning [59], [60]. Research on synthetic chemistry [12] identified a constraint: one of the main challenges in applying machine learning is the limitation of available data.

IV. CONCLUSION

Based on the results of designing a web-based exploratory data analysis (EDA) tool to visualize thermodynamic concepts in the steam power plant trainer process, it can be concluded that the experts gave an average score of 4 in the information accuracy aspect in the good category. This shows that the information displayed in this visualization tool is by thermodynamics learning at Padang State University. In the visualization aspect, experts gave an average score of 4.25, which is in the good and very good range. In alignment with the education aspect, experts gave an average score of 3.75, close to the good category. This shows that this aspect is considered suitable for studying thermodynamics, although shortcomings still need to be corrected. Experts gave a relatively high assessment of the Ease-of-Use aspect, with an average score of 4.5, with a range of Good and Very Good. With this visualization tool, students can easily access and visualize data from steam power plant trainer experiments in the form of graphs, diagrams, and predictive modeling using machine learning methods. This allows students to observe and understand complex patterns, cause-effect relationships, and parameter changes in thermodynamic concepts more deeply. Active use of data in visualization also allows students to explore independently. Additionally, this approach aligns with current trends where data is actively used for decision-making, analysis, and innovation across industries and academic disciplines.

REFERENCES

[1] E. Cueto and F. Chinesta, "Thermodynamics of Learning Physical Phenomena," *Arch. Comput. Methods Eng.*, vol. 30, no. 8, pp. 4653–4666, Nov. 2023, doi: 10.1007/s11831-023-09954-5.

[2] J. Zhang, H. Cho, and P. J. Mago, "Improving Student Learning of Energy Systems through Computational Tool Development Process in Engineering Courses," *Sustainability*, vol. 13, no. 2, p. 884, Jan. 2021, doi: 10.3390/su13020884.

[3] S. A. Finkenstaedt-Quinn *et al.*, "Capturing student conceptions of thermodynamics and kinetics using writing," *Chem. Educ. Res. Pract.*, vol. 21, no. 3, pp. 922–939, 2020, doi: 10.1039/C9RP00292H.

[4] J.-C. De Hemptinne *et al.*, "A View on the Future of Applied Thermodynamics," *Ind. Eng. Chem. Res.*, vol. 61, no. 39, pp. 14664–14680, Oct. 2022, doi: 10.1021/acs.iecr.2c01906.

[5] M. J. Brundage and C. Singh, "Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 2, p. 020112, Aug. 2023, doi:10.1103/PhysRevPhysEducRes.19.020112.

[6] S. A. Chaturvedi, T. Abdel-Salam, and R. Arora, "A web-based simulation and visualization module for an undergraduate thermodynamics course," *Comput. Educ. J.*, vol. 5, no. 2, pp. 58–67, 2014.

[7] M. R. D. Biasi, G. E. Valencia, and L. G. Obregon, "A New Educational Thermodynamic Software to Promote Critical Thinking in Youth Engineering Students," *Sustainability*, vol. 12, no. 1, p. 110, Dec. 2019, doi: 10.3390/su12010110.

[8] T. Milo and A. Somech, "Automating Exploratory Data Analysis via Machine Learning: An Overview," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, Portland OR USA: ACM, Jun. 2020, pp. 2617–2622. doi:10.1145/3318464.3383126.

[9] Dr. T. H. S. Tariq and P. S. Aithal, "Visualization and Explorative Data Analysis," *SSRN Electron. J.*, 2023, doi: 10.2139/ssrn.4400256.

[10] P. Ma, R. Ding, S. Han, and D. Zhang, "MetaInsight: Automatic Discovery of Structured Knowledge for Exploratory Data Analysis," in *Proceedings of the 2021 International Conference on Management of Data*, Virtual Event China: ACM, Jun. 2021, pp. 1262–1274. doi:10.1145/3448016.3457267.

[11] J. Hullman and A. Gelman, "Designing for Interactive Exploratory Data Analysis Requires Theories of Graphical Inference," *Harv. Data Sci. Rev.*, Jul. 2021, doi: 10.1162/99608f92.3ab8a587.

[12] M. Ma, J. Yang, P. Wang, W. Liu, and J. Zhang, "Light-Weight and Scalable Hierarchical-MVC Architecture for Cloud Web Applications," in *2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/ 2019 5th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, Paris, France: IEEE, Jun. 2019, pp. 40–45. doi:10.1109/CSCloud/EdgeCom.2019.00017.

[13] D. Leni, F. Earnestly, R. Sumiati, A. Adriansyah, and Y. P. Kusuma, "Evaluasi sifat mekanik baja paduan rendah bedasarkan komposisi kimia dan suhu perlakuan panas menggunakan teknik exploratory data analysis (EDA)," *Din. Tek. Mesin*, vol. 13, no. 1, p. 74, Apr. 2023, doi:10.29303/dtm.v13i1.624.

[14] J. M. Nápoles-Duarte, A. Biswas, M. I. Parker, J. P. Palomares-Baez, M. A. Chávez-Rojo, and L. M. Rodríguez-Valdez, "Stmol: A component for building interactive molecular visualizations within streamlit web-applications," *Front. Mol. Biosci.*, vol. 9, p. 990846, Sep. 2022, doi: 10.3389/fmolb.2022.990846.

[15] J. Peng *et al.*, "DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python," in *Proceedings of the 2021 International Conference on Management of Data*, Virtual Event China: ACM, Jun. 2021, pp. 2271–2280. doi:10.1145/3448016.3457330.

[16] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Front. Energy Res.*, vol. 9, p. 652801, Mar. 2021, doi:10.3389/fenrg.2021.652801.

[17] M. S. Smitha Rao, M. Pallavi, and N. Geetha, "Conceptual Machine Learning Framework for Initial Data Analysis," in *Computing and Network Sustainability*, vol. 75, S.-L. Peng, N. Dey, and M. Bundele, Eds., in Lecture Notes in Networks and Systems, vol. 75. , Singapore: Springer Singapore, 2019, pp. 51–59. doi: 10.1007/978-981-13-7150-9_6.

[18] S. learn, "Scikit-learn (BSD License)." [Online]. Available: https://scikit-learn.org/stable/auto_examples/release_highlights/plot_release_highlights_1_3_0.html

[19] K. Keras, "TensorFlow Core," TensorFlow 2.12. [Online]. Available: https://blog.tensorflow.org/2023/03/whats-new-in-tensorflow-212.html

[20] R. Siddiqui *et al.*, "Power Prediction of Combined Cycle Power Plant (CCPP) Using Machine Learning Algorithm-Based Paradigm," *Wirel. Commun. Mob. Comput.*, vol. 2021, pp. 1–13, Dec. 2021, doi:10.1155/2021/9966395.

[21] D. Chary, "Prediction of Full Load Electrical Power Output of a Base Load Operated Combined Cycle Power Plant Using Machine Learning Methods," *SSRN Electron. J.*, 2021, doi: 10.2139/ssrn.3945086.

[22] N. S. Santarisi and S. S. Faouri, "Prediction of combined cycle power plant electrical output power using machine learning regression algorithms," *East.-Eur. J. Enterp. Technol.*, vol. 6, no. 8 (114), pp. 16–26, Dec. 2021, doi: 10.15587/1729-4061.2021.245663.

[23] D. Leni, Y. P. Kusuma, M. Muchlisinalahuddin, F. Earnerstly, R. Muharni, and R. Sumiati, "Perancangan Metode Machine Learning Berbasis Web Untuk Prediksi Sifat Mekanik Aluminium," *J. Rekayasa Mesin*, vol. 14, no. 2, pp. 611–626, Aug. 2023, doi:10.21776/jrm.v14i2.1370.

[24] D. Leni, A. Karudin, M. R. Abbas, J. K. Sharma, and A. Adriansyah, "Optimizing stainless steel tensile strength analysis: through data exploration and machine learning design with Streamlit," *EUREKA Phys. Eng.*, no. 5, pp. 73–88, Sep. 2024, doi: 10.21303/2461-4262.2024.003296.

[25] M. D. B. Castro and G. M. Tumibay, "A literature review: efficacy of online learning courses for higher education institution using meta-analysis," *Educ. Inf. Technol.*, vol. 26, no. 2, pp. 1367–1385, Mar. 2021, doi: 10.1007/s10639-019-10027-z.

[26] A. Stanciulescu, F. Castronovo, and J. Oliver, "Assessing the impact of visualization media on engagement in an active learning environment," *Int. J. Math. Educ. Sci. Technol.*, pp. 1–21, Mar. 2022, doi: 10.1080/0020739X.2022.2044530.

[27] B. Aquilani, M. Piccarozzi, T. Abbate, and A. Codini, "The Role of Open Innovation and Value Co-creation in the Challenging Transition from Industry 4.0 to Society 5.0: Toward a Theoretical Framework," *Sustainability*, vol. 12, no. 21, p. 8943, Oct. 2020, doi:10.3390/su12218943.

[28] S. Babicki *et al.*, "Heatmapper: web-enabled heat mapping for all," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W147–W153, Jul. 2016, doi:10.1093/nar/gkw419.

[29] Z. Gu, "Complex heatmap visualization," *iMeta*, vol. 1, no. 3, p. e43, Sep. 2022, doi: 10.1002/imt2.43.

[30] A. J. Bishara and J. B. Hittner, "Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches.," *Psychol. Methods*, vol. 17, no. 3, pp. 399–417, Sep. 2012, doi: 10.1037/a0028087.

[31] A. Mambro, F. Congiu, and E. Galloni, "Influence of stage design parameters on ventilation power produced by steam turbine last stage blades during low load operation," *Therm. Sci. Eng. Prog.*, vol. 28, p. 101054, Feb. 2022, doi: 10.1016/j.tsep.2021.101054.

[32] A. Patil *et al.*, "Two-phase operation of a Terry steam turbine using air and water mixtures as working fluids," *Appl. Therm. Eng.*, vol. 165, p. 114567, Jan. 2020, doi: 10.1016/j.applthermaleng.2019.114567.

[33] A. Joseph Omosanya, E. Titilayo Akinlabi, and J. Olusegun Okeniyi, "Overview for Improving Steam Turbine Power Generation Efficiency," *J. Phys. Conf. Ser.*, vol. 1378, no. 3, p. 032040, Dec. 2019, doi: 10.1088/1742-6596/1378/3/032040.

[34] S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., *Encyclopedia of bioinformatics and computational biology*. Amsterdam Boston Heidelberg London New York Oxford Paris San Diego San Francisco Singapore Sydney Tokyo: Elsevier, 2019.

[35] A. Khalyasmaa *et al.*, "Prediction of Solar Power Generation Based on Random Forest Regressor Model," in *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, Novosibirsk, Russia: IEEE, Oct. 2019, pp. 0780–0785. doi: 10.1109/SIBIRCON48586.2019.8958063.

[36] C. González, J. Mira-McWilliams, and I. Juárez, "Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests," *IET Gener. Transm. Distrib.*, vol. 9, no. 11, pp. 1120–1128, Aug. 2015, doi: 10.1049/iet-gtd.2014.0655.

[37] M. W. Ahmad, J. Reynolds, and Y. Rezgui, "Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees," *J. Clean. Prod.*, vol. 203, pp. 810–821, Dec. 2018, doi:10.1016/j.jclepro.2018.08.207.

[38] V. Veeramsetty, K. R. Reddy, M. Santhosh, A. Mohnot, and G. Singal, "Short-term electric power load forecasting using random forest and gated recurrent unit," *Electr. Eng.*, vol. 104, no. 1, pp. 307–329, Feb. 2022, doi: 10.1007/s00202-021-01376-5.

[39] J. Rosenberg and M. Lawson, "An Investigation of Students' Use of a Computational Science Simulation in an Online High School Physics Class," *Educ. Sci.*, vol. 9, no. 1, p. 49, Mar. 2019, doi:10.3390/educsci9010049.

[40] J. G. Acevedo, G. Valencia Ochoa, and L. G. Obregon, "Development of a new educational package based on e-learning to study engineering thermodynamics process: combustion, energy and entropy analysis," *Heliyon*, vol. 6, no. 6, p. e04269, Jun. 2020, doi:10.1016/j.heliyon.2020.e04269.

[41] W. Vallejo, C. Díaz-Uribe, and C. Fajardo, "Google Colab and Virtual Simulations: Practical e-Learning Tools to Support the Teaching of Thermodynamics and to Introduce Coding to Students," *ACS Omega*, vol. 7, no. 8, pp. 7421–7429, Mar. 2022, doi:10.1021/acsomega.2c00362.

[42] V. Kumar, B. Pandya, and V. Matawala, "Thermodynamic studies and parametric effects on exergetic performance of a steam power plant," *Int. J. Ambient Energy*, vol. 40, no. 1, pp. 1–11, Jan. 2019, doi:10.1080/01430750.2017.1354326.

[43] B. Rudiyanto *et al.*, "Energy and Exergy Analysis of Steam Power Plant in Paiton, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 268, no. 1, p. 012091, Jun. 2019, doi: 10.1088/1755-1315/268/1/012091.

[44] K. Munn, S. Goh, M. Basson, and D. Thorpe, "Asset management competency requirements in Australian local government: a systematic literature review," *Australas. J. Eng. Educ.*, vol. 26, no. 2, pp. 167–200, Jul. 2021, doi: 10.1080/22054952.2021.1934262.

[45] J. Ding *et al.*, "Machine learning for molecular thermodynamics," *Chin. J. Chem. Eng.*, vol. 31, pp. 227–239, Mar. 2021, doi:10.1016/j.cjche.2020.10.044.

[46] S. S. Funai and D. Giataganas, "Thermodynamics and feature extraction by machine learning," *Phys. Rev. Res.*, vol. 2, no. 3, p. 033415, Sep. 2020, doi: 10.1103/PhysRevResearch.2.033415.

[47] F. Jirasek *et al.*, "Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion," *J. Phys. Chem. Lett.*, vol. 11, no. 3, pp. 981–985, Feb. 2020, doi:10.1021/acs.jpclett.9b03657.

[48] S. Jafarinejad, L. E. Beckingham, M. Kathe, and K. Henderson, "The Renewable Energy (RE) Industry Workforce Needs: RE Simulation and Analysis Tools Teaching as an Effective Way to Enhance Undergraduate Engineering Students' Learning," *Sustainability*, vol. 13, no. 21, p. 11727, Oct. 2021, doi: 10.3390/su132111727.

[49] S. Mahmoudi Herris *et al.*, "Identification and Assessment of the Effective Factors on the Occurrence of the Environmental Events Caused by the Construction and Operation of Gas and Steam Power Plants (Case Study: MAPNA Group(," *J. Occup. Hyg. Eng.*, vol. 6, no. 4, pp. 10–17, Feb. 2020, doi: 10.52547/johe.6.4.10.

[50] R. Alizadeh Kheneslu, A. Jahangiri, and M. Ameri, "Interaction effects of natural draft dry cooling tower (NDDCT) performance and 4E (energy, exergy, economic and environmental) analysis of steam power plant under different climatic conditions," *Sustain. Energy Technol. Assess.*, vol. 37, p. 100599, Feb. 2020, doi:10.1016/j.seta.2019.100599.

[51] A. Yousif, N. Drou, J. Rowe, M. Khalfan, and K. C. Gunsalus, "NASQAR: a web-based platform for high-throughput sequencing data analysis and visualization," *BMC Bioinformatics*, vol. 21, no. 1, p. 267, Dec. 2020, doi: 10.1186/s12859-020-03577-4.

[52] S. H. Chae, Y. Kim, K.-S. Lee, and H.-S. Park, "Development and Clinical Evaluation of a Web-Based Upper Limb Home Rehabilitation System Using a Smartwatch and Machine Learning Model for Chronic Stroke Survivors: Prospective Comparative Study," *JMIR MHealth UHealth*, vol. 8, no. 7, p. e17216, Jul. 2020, doi: 10.2196/17216.

[53] A. Weidlich, G. Gust, and M. Schäfer, "Conference on Energy Informatics:," *Energy Inform.*, vol. 4, no. S4, pp. 26, s42162-021-00184–2, Sep. 2021, doi: 10.1186/s42162-021-00184-2.

[54] C. Bordin, S. Mishra, A. Safari, and F. Eliassen, "Educating the energy informatics specialist: opportunities and challenges in light of research and industrial trends," *SN Appl. Sci.*, vol. 3, no. 6, p. 674, Jun. 2021, doi: 10.1007/s42452-021-04610-8.

[55] P. Himthani, G. P. Dubey, B. M. Sharma, and A. Taneja, "Big Data Privacy and Challenges for Machine Learning," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India: IEEE, Oct. 2020, pp. 707–713. doi: 10.1109/I-SMAC49090.2020.9243527.

[56] S. Koyejo and Neural Information Processing Systems Foundation, Eds., *36th Conference on Neural Information Processing Systems (NeurIPS 2022): New Orleans, Louisiana, USA, 28 November-9 December 2022*. in Advances in neural information processing systems, no. 35. Red Hook, NY: Curran Associates, Inc, 2023.

[57] F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, and F. Glorius, "Machine learning the ropes: principles, applications and directions in synthetic chemistry," *Chem. Soc. Rev.*, vol. 49, no. 17, pp. 6154–6168, 2020, doi: 10.1039/C9CS00786E.

[58] T. Xu, G. Coco, and M. Neale, "A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning," *Water Res.*, vol. 177, p. 115788, Jun. 2020, doi: 10.1016/j.watres.2020.115788.

[59] S. I. Nikolenko, *Synthetic data for deep learning*. in Springer optimization and its applications, no. volume 174. Cham: Springer, 2021. doi: 10.1007/978-3-030-75178-4.

[60] D. Leni, D. S. Kesuma, Maimuzar, Haris, and S. Afriyani, "Prediction of Mechanical Properties of Austenitic Stainless Steels with the Use of Synthetic Data via Generative Adversarial Networks," in *The 7th Mechanical Engineering, Science and Technology International Conference*, MDPI, Feb. 2024, p. 4. doi:10.3390/engproc2024063004.