 p-ISSN 2827-8585
e-ISSN 2827-8062

International Journal of Multidisciplinary Research and Literature
IJOMRAL
Vol. 5, No. 3, May 2026 pp. 432-444
Journal Page is available at http://ijomral.esc-id.org/index.php/home

PREDICTION OF MECHANICAL PROPERTIES OF CAST TITANIUM BASED ON THE CHEMICAL ELEMENTS OF THE ALLOY USING A MACHINE LEARNING MODEL

Desvita Irwan[1], Jana Hafiza[2], Desmarita Leni[3*], Yassirli Amri[4], Ade Usra Berli[5]
[1,2,3,4] Department of Mechanical Engineering, Universitas Muhammadiyah Sumatera Barat, Indonesia
Corresponding Email: *desmaritaleni@gmail.com*[3]

Abstrack

Cast titanium is widely used in advanced engineering applications due to its high strength-to-weight ratio and good corrosion resistance. However, its mechanical properties are highly sensitive to variations in the alloy's chemical composition, making conventional testing less efficient in terms of time and cost. This study aims to predict the mechanical properties of cast titanium, namely yield strength and tensile strength, based on the alloy's chemical composition using a machine learning approach. The dataset was obtained from the Materials Algorithms Project (MAP), which includes 100 cast titanium specimens developed through small, controlled variations in chemical composition to represent realistic manufacturing conditions. Modelling was performed using Decision Tree and Random Forest algorithms, with data splits of 60:40, 70:30, and 80:20, and k-fold cross-validation. Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The results of the Pearson correlation analysis showed that Al, V, and Fe-X elements exhibit strong positive correlations with mechanical properties. In contrast, Ti and O show a significant negative correlation, consistent with the theory of titanium alloy metallurgy. The modelling results show that the Random Forest algorithm achieves better performance, with lower prediction errors and greater stability, than the Decision Tree. This study proves that the machine learning approach, especially the Random Forest algorithm, is effective in predicting the mechanical properties of cast titanium based on chemical composition, with the best performance shown by the RMSE value of 70.95 and MAE of 47.95, thus potentially supporting the design and optimisation of cast titanium alloys based on data.

Keywords: Cast titanium, chemical composition, machine learning, Decision Tree, Random Forest

INTRODUCTION

Titanium and its alloys have been widely used in advanced engineering applications, particularly in the aerospace, medical, and energy industries, due to their high strength-to-weight ratio, excellent corrosion resistance, and good biological compatibility (Pushp dkk, 2022). In its application, the casting process is an efficient manufacturing method for producing titanium components with complex geometries and lower production costs than forging or advanced machining processes (Anwar dkk, 2021). However, the mechanical properties of cast titanium are very sensitive to variations in the alloy's chemical composition, especially substitutional alloying elements (such as Al and V) and interstitial elements (such as O, C, and N) (Babaremu dkk, 2022). Small changes in the content of these elements can cause significant variations in yield and tensile strengths, which directly impact component performance and reliability (Al-hajiri dkk, 2024).

Conventional determination of material mechanical properties is generally performed through experimental laboratory testing, such as tensile and hardness tests. These methods are time-consuming,

expensive, and require specialised facilities and specimens, making them less efficient when the number of material composition variations increases significantly. Therefore, an alternative approach is needed to predict material mechanical properties quickly and accurately without relying entirely on physical testing (Z. Fu dkk, 2022).

The development of machine learning (ML) in recent years has opened new opportunities in materials engineering, particularly for modelling the nonlinear relationship between chemical composition and the mechanical properties of materials (Guo dkk, 2021). Various studies have shown that machine learning algorithms can accurately predict mechanical properties from material composition and process data (Leni, 2022). This data-driven approach enables more efficient exploration of the alloy design space than conventional trial-and-error methods. However, studies on predicting the mechanical properties of cast titanium based on chemical composition are still relatively limited. Most previous studies have focused on wrought or forged titanium, used limited datasets, and have not specifically compared the performance of Decision Tree and Random Forest algorithms across different data-sharing scenarios. Furthermore, discussions on the influence of dominant elements such as Al, V, Fe-X, O, and Ti on mechanical properties are often not explicitly linked to their metallurgical interpretation (Shugurov, 2021).

Given these conditions, this study makes several main contributions. First, this study develops a cast titanium dataset based on the average chemical composition, which is then expanded to small, controlled variations to represent realistic manufacturing conditions. Second, the performance of the Decision Tree and Random Forest algorithms in predicting yield and tensile strength is compared across several data-sharing scenarios. Third, the influence of dominant chemical elements on mechanical properties is analysed and interpreted by linking statistical correlation and machine learning modelling results with the principles of titanium alloy metallurgy. With this approach, this study is expected to contribute to the development of accurate and efficient prediction models for the mechanical properties of cast titanium. It can support data-driven decision-making in the design and optimisation of cast titanium alloys.

METHOD

The research data were obtained from the Materials Algorithms Project (MAP) available at makeitfrom.com, which provides selected material data covering the chemical composition and mechanical properties of various titanium alloys. The research methodology is structured in stages, from a literature review to the analysis and discussion of the results, as shown in the flowchart below **(Figure 1).**

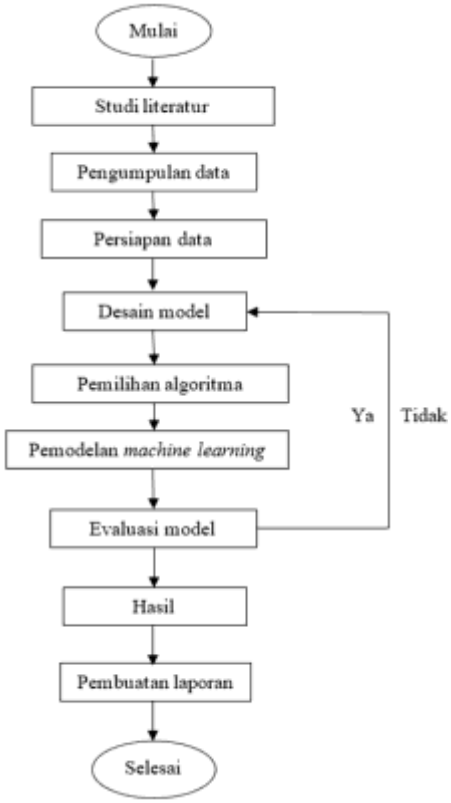

Figure 1. *Flow Chart*

The initial dataset consists of 100 cast titanium specimens with chemical composition values expressed as mean values for each alloy element (Table 1). Small, controlled variations were then applied to each specimen to represent realistic manufacturing conditions, without increasing the dataset size, so that the total dataset remained 100 samples.

Table 1. Titanium dataset statistics

Keterangan	Tipe Data	Min	Max	Mean
Ti (%)	Input	70	99.60	94.15
O (%)	Input	0.070	0.200	0.123
Fe (%)	Input	0	8	0.275
C (%)	Input	0.050	0.050	0.050
Ni (%)	Input	0.015	0.250	0.046
H (%)	Input	0.007	0.007	0.007
Al (%)	Input	0	8	2.040
V (%)	Input	0	15	1.305
Fe-X (%)	Input	0	0.250	0.129
Sn (%)	Input	0	10	0.340
Pd (%)	Input	0	5	0.105
YS (Mpa)	Output	190	1250	593.50
TS (Mpa)	Output	270	1350	680

This dataset was developed through small, controlled variations in chemical composition to represent fluctuations common in the casting process, without deviating from the valid range of cast

titanium grades. The major and major alloying elements (Ti, Al, V, Fe, and Fe-X) were varied by ±2–3% of the mean value, while the interstitial elements (O, C, N, and H) were varied by ±5% or kept constant if their values were very small. Where α is a variation factor adjusted for the type of element (Leni, 2023). To maintain the validity of the composition, each specimen resulting from the variation is then normalised so that the total elemental fraction remains satisfactory.

The distribution used is uniform, so each value within the range has an equal chance of being selected. After the variation process, composition normalisation is performed to ensure that the total fraction remains 100% (Pramana dkk, 2024). Additionally, each variation specimen is validated to remain within a realistic cast titanium composition range and within engineering material tolerances, so that the resulting data set does not deviate from actual manufacturing conditions (X. Fu dkk, 2025). This procedure ensures that the resulting dataset remains metallurgically consistent and suitable for use as model training data.

The input variables in this study are the chemical composition of cast titanium, including the main elements and alloying elements, namely Ti, Al, V, Fe, O, C, N, and other relevant elements, as specified in the dataset. The predicted output variables are the main mechanical properties, namely yield strength (YS) and tensile strength (TS), both expressed in MPa.

The data preprocessing stage included checking data completeness, standardising numerical formats, and performing descriptive statistical analysis to identify the minimum, maximum, and average values for each variable. Pearson correlation analysis was performed to evaluate the initial relationship between chemical composition and mechanical properties before machine learning modelling (Xu dkk, 2020)

The dataset used for model training and evaluation is split into two parts: training and testing. The testing data uses a training-test split, with three split ratios: 60:40, 70:30, and 80:20. The testing data is used only for final model evaluation and is not involved in the training or parameter tuning process. On the training data, k-fold cross-validation (k = 5) is performed, which is used exclusively for parameter tuning (hyperparameter tuning) and selecting the best model (Wijiyanto dkk, 2024). With this scheme, the model evaluation process becomes more stable and avoids bias introduced by single data splitting.

Modelling was carried out using the Decision Tree and Random Forest algorithms due to their ability to model nonlinear relationships between chemical composition and the mechanical properties of materials (Leni, 2023). In the Decision Tree algorithm, the main parameter varied is the maximum depth. In the Random Forest algorithm, the parameters that vary include the number of trees and the maximum depth. The selection of the best parameters is based on cross-validation results on the training data.

Model performance was evaluated using three main metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). (Kuswanto & Hakim, 2025). Can be expressed with the equation 1,2, dan 3 (Leni, 2023) :

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{1\text{-}}\ y_1)^2} \quad (1)$$

$$MAE = \frac{1}{n}\ \sum_{i=1}^{n}|y_{i\text{-}}\ y_i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\ y_i\text{-}y_i\ )^2}{\sum_{i=1}^{n}(\ y_i\ .\ \ y)^2} \quad (3)$$

RMSE is used to measure the overall magnitude of the prediction error by giving a larger penalty to high errors. In contrast, MAE measures the average absolute difference between the predicted and actual values. The coefficient of determination (R^2) indicates the level of agreement between predicted and actual values. These three-evaluation metrics are applied. consistently throughout all stages of model evaluation, both during the train–test split process and during the final model evaluation (Leni dkk, 2023).

RESEARCH RESULTS AND DISCUSSION

The Pearson correlation heatmap in Figure 3 shows that Al, V, and Fe-X elements exhibit strong positive correlations with the yield strength (YS) of cast titanium, with coefficients of 0.721, 0.618, and 0.723, respectively, indicating their role in increasing the material's resistance to initial plastic deformation. In contrast, the Ti element shows a very strong negative correlation with YS (−0.826), indicating that the dominance of pure Ti tends to reduce the yield strength of the material. The O element also has a fairly strong negative correlation (−0.603), indicating that increasing oxygen content can increase brittleness, thereby reducing the stability of initial plastic deformation.

Figure 2. *Heatmap Correlation Yield Strength*

Figure 3. *Heatmap Correlation Tensile Strength*

The Pearson correlation heatmap in Figure 4 shows that Al, V, and Fe-X elements exhibit strong positive correlations with the tensile strength (TS) of cast titanium, with coefficients of 0.725, 0.630, and 0.729, respectively, indicating their contribution to increasing the material's maximum tensile strength. In contrast, the Ti element shows a very strong negative correlation with TS (−0.834), while the O element also has a significant negative correlation (−0.613), indicating that the dominance of pure Ti and increasing oxygen content tend to decrease the maximum tensile performance of cast titanium.

The positive correlation between Al and V elements in the yield and tensile strengths of cast titanium can be explained by their roles in phase stabilisation. Aluminium functions as an α-stabiliser, increasing strength through solid-solution strengthening and stabilising the α phase, thereby inhibiting dislocation movement. Vanadium acts as a β-stabilizer that promotes the formation of a two-phase microstructure (α+β), thereby increasing the material's tensile and yield strengths.

Fe-X elements also show a positive correlation because they act as minor β-stabilisers, increasing strength through solid-solution strengthening without significantly reducing ductility at low levels. In contrast, oxygen (O), as an interstitial element, shows a negative correlation: increasing its content tends to increase brittleness (embrittlement) and reduce the stability of plastic deformation. The negative correlation of Ti indicates that the dominance of pure Ti without alloying element reinforcement results in lower mechanical strength than that of titanium alloys.

Although the correlation patterns for tensile strength and yield strength are generally similar, the differences in correlation strength indicate that the chemical element contributions to both mechanical properties are not completely identical, so the analysis was conducted separately for TS and YS. The results of this correlation heatmap provide an important basis for machine learning modelling because they show a significant, nonlinear relationship between chemical composition and mechanical properties and align with the Random Forest results, which identified Al, V, Fe-X, and O as dominant features in the prediction.

To obtain optimal predictions of tensile and yield strengths in cast titanium, the Decision Tree and Random Forest algorithms were used with varying parameter settings. The Decision Tree and Random Forest algorithms were chosen because they can model nonlinear relationships in small-to-medium-sized datasets and are highly interpretable. It allows analysis of the contribution of alloying elements to the mechanical properties of cast titanium, thus focusing not only on prediction accuracy but also on understanding its metallurgy. In the Decision Tree algorithm, the parameter varied was the maximum depth, whereas in the Random Forest algorithm, both the maximum depth and the number of trees were varied. Details of the parameter variations used in each algorithm are presented in **Table 2**, which includes the algorithm type, the parameter set, and the parameter values used in the modelling process.

Table 2. Pengaturan parameter untuk mencari algoritma terbaik

No	Algorithm	Parameter Name	Parameter Settings
1.	*Decision Tree*	*Criterion*	*Least square*
		Maximal depth	1,2,3,4,5,6,7,8,9,10,11,12
		Ratio	60:40, 70:30, 80:20
2.	*Random Forest*	*Number of trees*	20,40,60,80,100,120,140,160.180,200
		Criterion	*Least square*
		Maximal depth	1,2,3,4,5,6,7,8,9,10,11,12
		Ratio	60:40, 70:30, 80:20

The best parameter search results are determined based on the smallest RMSE and MAE values and the highest coefficient of determination (R^2). The results are as follows:

1. Decision Tree

Figure 5 shows the effect of varying the maximum depth on the performance of the Decision Tree algorithm, as measured by RMSE, MAE, and RRSE. At maximum depth = 1, the prediction error value is very high, indicating underfitting due to low model complexity. As the tree depth As the number increases, the RMSE and MAE decrease significantly and reach their best performance at maximum depth = 4, with an RMSE of 75.63, an MAE of 49.07, and an RRSE of 0.259.

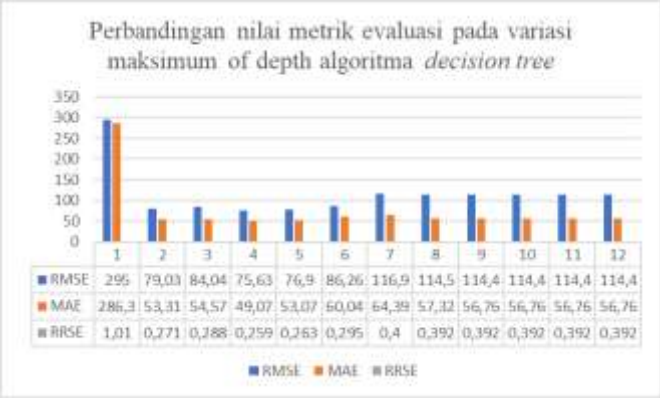

Figure 4. Comparison chart of the results of the variation test maximum depth value of the Decision Tree

After exceeding this depth, increasing the maximum depth no longer improves model performance; instead, it tends to plateau or increase, indicating potential overfitting. Therefore, a maximum depth of 4 was determined to be the optimal configuration for the Decision Tree algorithm.

2. *Random Forest*

Figure 6 shows that increasing the maximum depth in the Random Forest algorithm consistently reduces the RMSE and MAE values until it reaches optimal performance at a maximum depth of 10 with an RMSE of 70.95, an MAE of 47.95, and an RRSE of 0.243. After this depth, increasing the maximum depth does not provide significant performance improvements, indicating that the model has reached an optimal level of complexity. These results indicate that Random Forest has better stability and generalisation capabilities than Decision Tree.

Feature importance analysis in the Random Forest algorithm indicates that Al, V, Fe-X, and O are the most dominant features for predicting the yield and tensile strengths of cast titanium. The dominance of these features is consistent with the results of the Pearson correlation analysis and the metallurgical interpretation, which indicate that α- and β-stabiliser elements, as well as interstitial elements, significantly influence mechanical strength. The consistency between feature importance and material theory strengthens the validity of the developed prediction model.

Figure 5. Comparison chart of the test results of the maximum variation of the depth of the Random Forest

Figures 7 and 8 compare the performance of Random Forest and Decision Tree in predicting tensile strength (TS) and yield strength (YS) across the 60:40, 70:30, and 80:20 data splitting schemes using cross-validation. Random Forest consistently produces lower RMSE, MAE, and R2 values than Decision Tree across all scenarios, indicating better generalisation. The best performance of Random Forest for TS prediction is obtained at a ratio of 70:30. In contrast, for YS prediction, it is obtained at a ratio of 80:20. These results confirm the superiority of Random Forest, which is supported by an ensemble learning mechanism that can reduce variance and increase prediction stability.

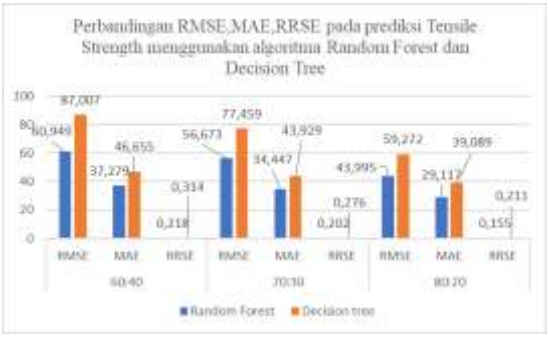

Figure 6. Comparison of the performance of the Decision Tree and Random Forest algorithms on TS prediction

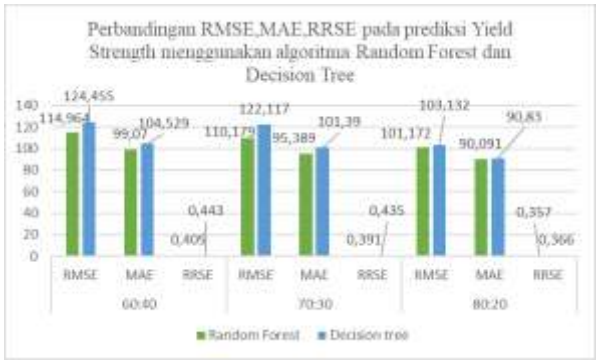

Figure 7. Comparison of the performance of the Decision Tree and Random Forest algorithms on YS prediction

The graphical results show that Random Forest produces lower RMSE, MAE, and RRSE values than Decision Tree across all data ratios, indicating lower prediction error and more stable model performance. For tensile strength prediction, Random Forest's best performance was obtained at a 70:30 ratio, while for yield strength, at an 80:20 ratio. Conversely, the Decision Tree shows higher error values and greater performance variation across data ratios. It indicates that the Decision Tree is more sensitive to changes in training data. At the same time, the Random Forest has better generalisation capabilities because it combines multiple decision trees during prediction.

Model evaluation was conducted to measure the model's reliability in producing accurate predictions. Based on modelling results comparing two algorithms, it was found that the Random Forest algorithm with a 70:30 data split provided superior performance in predicting yield and tensile strength values compared to the Decision Tree algorithm.

Figure 9 and Figure 10 show a comparison of the actual and predicted values of yield strength and tensile strength using the Random Forest algorithm with a data ratio of 70:30. The graph results show that the predicted values are close to the actual values with a relatively small error distribution, indicating good model accuracy and stability.

Figure 8. Comparison chart of predicted Yield Strength data with actual data using the 70:30 Random Forest algorithm

Figure 9. Comparison chart of predicted tensile strength data with actual data using the 70:30 Random Forest algorithm

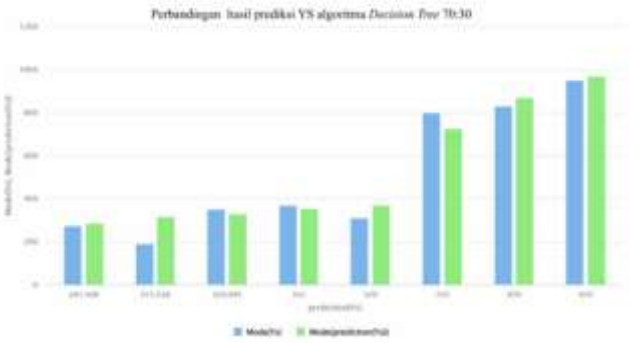

Figure 10. Comparison chart of predicted Yield Strength data with actual data using the 70:30 Decision Tree algorithm

Figure 11. Comparison chart of predicted tensile strength data with actual data using the 70:30 Decision Tree algorithm

In Figures 11 and 12, using the Decision Tree algorithm, a larger deviation between the predicted and actual values was observed compared to Random Forest. Quantitatively, the Decision Tree produced higher errors, with an RMSE of 75.63, an MAE of 49.07, and an RRSE of 0.259 in its best configuration. Conversely, the Random Forest algorithm performed better, with an RMSE of 70.95, an MAE of 47.95, and an RRSE of 0.243, indicating lower prediction error and better model stability.

The performance difference between Decision Tree and Random Forest can also be explained by the bias–variance trade-off (Talekar, 2020). Decision Trees with low depth tend to experience underfitting (high bias), while too much depth has the potential to cause overfitting (high variance) (Firmansyach dkk, 2023). In contrast, Random Forest reduces variance through ensemble learning by combining many decision trees, producing more stable and accurate predictions (Indahyanti dkk, 2022). It explains why Random Forest consistently performs better across all data sharing scenarios.

CONCLUSION

The results of the Pearson correlation analysis show that Al, V, and Fe-X elements exhibit strong positive correlations with the yield and tensile strengths of cast titanium. In contrast, Ti and O elements show significant negative correlations. This finding is consistent with the theory of titanium alloy metallurgy, where Al acts as an α-stabiliser, V and Fe-X as β-stabilisers that increase mechanical strength, and O, as an interstitial element, tends to increase brittleness and decrease plastic deformation stability.

Machine learning modelling shows that the Random Forest algorithm performs better and more consistently than the Decision Tree across all data-sharing scenarios. Random Forest produces lower prediction errors and better generalisation, with optimal performance at data ratios of 70:30 for tensile strength and 80:20 for yield strength. These results indicate that Random Forest is an effective predictive

approach for the mechanical properties of cast titanium alloys based on chemical composition and has the potential to support data-driven design and optimisation of cast titanium alloys.

REFERENCES

Al-hajiri, M., Alshammari, Y., Yang, F., & Bolzoni, L. (2024). Gaining understanding of the effect of alloy composition on the properties of Ti–Cu–Mn–Al alloys. *Materials Science and Engineering: A*, *908*, 146726. https://doi.org/10.1016/j.msea.2024.146726

Anwar, F. M., Hidayat, M. I. P., & Ramadhani, M. (2021). Simulasi Aliran pada Proses Pengecoran Connecting Rod Berbahan Aluminium 7075 untuk Mesin Motor 150 cc dengan Variasi Tinggi Sprue Cetakan dan Temperatur Penuangan Menggunakan Metode Elemen Hingga. *Jurnal Teknik ITS*, *9*(2), F171–F176. https://doi.org/10.12962/j23373539.v9i2.54627

Babaremu, K. O., Jen, T.-C., Oladijo, P. O., & Akinlabi, E. T. (2022). Mechanical, Corrosion Resistance Properties and Various Applications of Titanium and Its Alloys: A Review. *Revue des composites et des matériaux avancés*, *32*(1), 11–16. https://doi.org/10.18280/rcma.320102

Firmansyach, W. A., Hayati, U., & Arie Wijaya, Y. (2023). Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 262–269. https://doi.org/10.36040/jati.v7i1.6329

Fu, X., Li, B., Fu, B., Dong, T., & Li, J. (2025). Composition Design of a Novel High-Temperature Titanium Alloy Based on Data Augmentation Machine Learning. *Materials*, *18*(13), 3099. https://doi.org/10.3390/ma18133099

Fu, Z., Liu, W., Huang, C., & Mei, T. (2022). A Review of Performance Prediction Based on Machine Learning in Materials Science. *Nanomaterials*, *12*(17), 2957. https://doi.org/10.3390/nano12172957

Guo, K., Yang, Z., Yu, C.-H., & Buehler, M. J. (2021). Artificial intelligence and machine learning in the design of mechanical materials. *Materials Horizons*, *8*(4), 1153–1172. https://doi.org/10.1039/D0MH01451F

Indahyanti, U., Azizah, N. L., & Setiawan, H. (2022). Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa. *Jurnal Sains dan Informatika*, *8*(2). https://doi.org/10.34128/jsi.v8i2.459

Kuswanto, J., & Hakim, L. (2025). Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa. *Decode: Jurnal Pendidikan Teknologi Informasi*, *5*(1), 262–270. https://doi.org/10.51454/decode.v5i1.1103

Leni, D. (2022). *Perbandingan Algoritma Machine Learning Untuk Prediksi Sifat Mekanik Pada Baja Paduan Rendah: Vol.5 No.2* (hlm. 167–174). Jurnal Rekayasa Material, Manufaktur dan Energi. DOI: https://doi.org/10.30596/rmme.v5i2.11407

Leni, D. (2023). Pemilihan Algoritma Machine Learning Yang Optimal Untuk Prediksi Sifat Mekanik Aluminium. *Jurnal Engine: Energi, Manufaktur, dan Material*, *7*(1), 35. https://doi.org/10.30588/jeemm.v7i1.1490

Leni, D., Dwiharzandis, A., Sumiati, R., Haris, H., & Afriyani, S. (2023). Seleksi Fitur Berdasarkan Korelasi Pearson dalam Pemodelan Efisiensi Energi Bangunan. *Teknika Sains: Jurnal Ilmu Teknik*, *8*(2), 103–115. https://doi.org/10.24967/teksis.v8i2.2525

Pramana, I. W. S., Kotama, I. N. D., & Adnyana, A. A. G. O. K. (2024). Analisis Komprehensif Normalisasi dalam Weighted Product Model: Implikasinya terhadap Pengambilan Keputusan. *Jurnal Krisnadana*, *3*(2), 112–119. https://doi.org/10.58982/krisnadana.v3i2.551

Pushp, P., Dasharath, S. M., & Arati, C. (2022). Classification and applications of titanium and its alloys. *Materials Today: Proceedings*, *54*, 537–542. https://doi.org/10.1016/j.matpr.2022.01.008

Shugurov, A. (2021). Microstructure and Mechanical Properties of Titanium Alloys. *Metals*, *11*(10), 1617. https://doi.org/10.3390/met11101617

Talekar, B. (2020). A Detailed Review of Decision Tree and Random Forest. *Bioscience Biotechnology Research Communications*, *13*(14), 245–248. https://doi.org/10.21786/bbrc/13.14/57

Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa. *Jurnal Algoritma*, *21*(1). https://doi.org/10.33364/algoritma/v.21-1.1618

Xu, G., Cheng, H., Zhang, K., Liang, B., Cheng, Y., Hu, J., & Liu, C. (2020). Modelling of the damage behaviour of carbon fibre-reinforced plastic composites and interference bolting with a sleeve. *Materials & Design*, *194*, 108904. https://doi.org/10.1016/j.matdes.2020.108904