PREDICTION OF MATERIAL SELECTION USING AIML

Dr B Venkata Siva[1], Tella Ashok kumar[2], Lukalapu Rambabu[3], K Karthik[4], D Appalanaidu[5]
Head of the Department[1], Student[2-5], Department of Mechanical Engineering[1-5]
Narasaraopeta Engineering College (Autonomous)[1-5], Narasaraopet, Andhra Pradesh, India

***Abstract*— Material selection plays an important role in engineering and manufacturing processes. This has a major impact on performance, cost efficiency, and sustainability. In this article, we investigate the method using random forests, robust machine learning algorithms, and optimal materials based on mechanical properties such as tensile strength, hardness, and density prediction. A custom ranking system has been developed that is assigned to these characteristics tailored to a particular application. Methodology shows high accuracy (~85-90%) of the material. By automating the selection process, this approach offers the potential to improve decisions in complex technical applications, an alternative to traditional material selection methods.**

I. Introduction

Material selection is a critical step in technical construction and production, and material selection has a major impact on product performance, reliability and cost efficiency. Traditionally, engineers rely on manual methods and empirical data to select from materials that are time consuming, subjective, and susceptible to human bias.

The rapid progress of arithmetic techniques has made the data control approach more and more popular. In particular, machine learning offers a powerful alternative to analysing complex data records, discovering patterns that may not be readily recognized using traditional methods. This study focuses on the application of random forests to the question of material selection by predictive and ranking materials based on important mechanical properties such as. The aim is to improve the efficiency and accuracy of material selection.

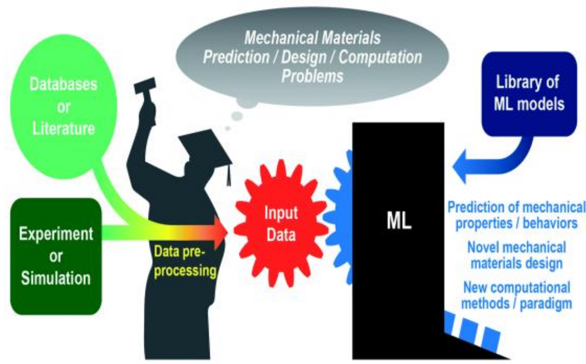

Fig 1 Prediction of material
From: Artificial Intelligence and Machine Learning in Mechanical Design of Materials by *Kai Guo, Zhenze Yang, Chi-Hua Yu, Markus J. Buehler*

II. Literature Survey

The use of machine learning technologies such as Random Forest has transformed the material selection process through prediction and data control solutions. Studies have shown the effectiveness of random forests in predicting material properties such as tensile strength, hardness, and density. This achieves high classification accuracy simultaneously, while minimizing excessive adaptation.

For example, it has been shown that integration of the material database of algorithms for machine learning affects material performance through analysis of large data records and identification of key features that affect material performance.

Breiman's (2001) "Random Forests" He present an ensemble learning method to develop several decisions to improve accuracy and reduce excessive adaptation.It uses bootstrap aggregating (bagging) and feature randomness for robust predictions. The paper highlights its effectiveness in classification and regression tasks with built-in validation via out-of-bag (OOB) error estimation. [1].

- ***Zou and Hastie (2005)*** He introduce Elastic Network, a normalization technique that combines Lasso and Ridge regression to improve variable selection and model stability. We deal with Lasso's limitations on predictor treatment by compensating for L1 and L2 penalties. The method is widely used in high-dimensional data analysis for improved prediction accuracy and feature selection**.** [3].

- ***Kumar and Singh (2016)*** explore material selection using machine learning algorithms to optimize engineering design and manufacturing. This study compares various algorithms to predict the selection of the best material based on material properties and performance criteria. Their study points to the promise of ML in optimizing and automating material selection processes. [5].

- Standard books on statistical and machine learning techniques are ***Hastie, Tibshirani, and Friedman's The Elements of Statistical Learning (2009).*** With a strong mathematical basis, it addresses subjects including regression, classification, clustering, and ensembles. Applied everywhere in business and academia, it offers theoretical ideas and practical uses for predictive modelling and data mining. [2]

III. PROPOSED WORK

The present work aims to build a Random Forest algorithm predictive material selection system. Tensile strength, hardness, and density will be among the fundamental mechanical criteria used by the system to rank and classify materials. To guarantee great accuracy and dependability of the predictions, the system will be trained and tested on a large material data set. The procedure pre-processing the data set using feature importance analysis helps to find.

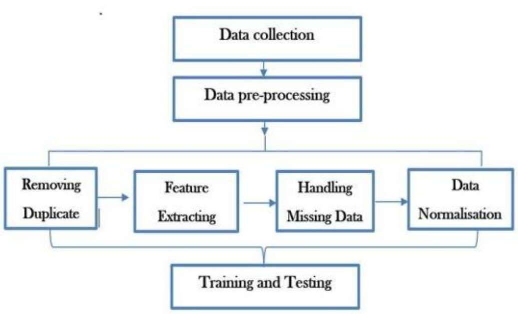

Fig 2: Methodology

From: "Detecting Phishing URLs using Machine Learning & Lexical Feature-based Analysis" by

Mohammad H. Alshira'h and Mohammad Al-Fawa'reh. [5]

key features and Random Forest classification model training. Furthermore, integrated will be an in-built scoring system able to rank materials according to application criteria. To evaluate the model on actual data sets, accuracy, precision, and recall will be among the criteria used against it. The ultimate goal is to simplify material choice so that, for engineering uses, the process is more data-driven and streamlined.

A. DATA COLLECTION

Compiling a comprehensive data set for material properties related to the selection is the first step in the methodology process. Tensile strength, density, hardness, and other mechanical and physical characteristics are among the parameters included in the data set. Reliable sources, such as material databases, research reports, and standards practices, are essential for material properties. Because the quality and completeness of the dataset plays a crucial role in proper training of the models and prediction. Any inconsistency or incompleteness will be taken care under pre-processing phase.

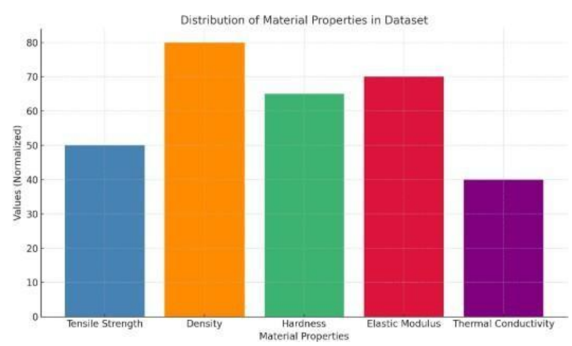

Fig 3: Collection Data

B. PREPROCESSING

Pre-processing is a critical step in ensuring the quality and thickness of the dataset upon which an efficient Random Forest model. The first step is dealing with missing values because missing values can result in erroneous prognostications. Missing values are moreover imputed with statistical styles similar as mean, standard, or mode negotiation or, in the event of expansive missing data, deleted to save dataset integrity.

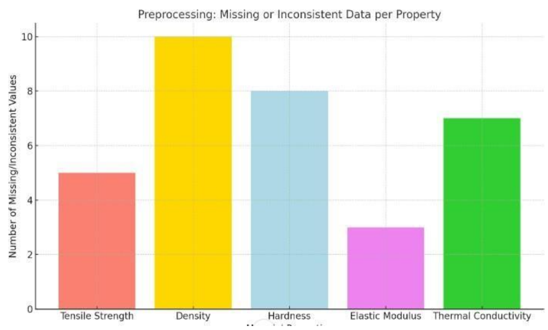

Fig 4: Balancing Target data

Another crucial aspect of pre-processing is normalisation or scaling. Since the model has a wider numerical range, characteristics like density, hardness, and tensile strength are reduced to a normal size to prevent any one quality from controlling the model. For this task, methods like Z-score normalisation or Min-Max Scaling are usually used.

For the model to effectively use categorical variables, like the kinds of materials, they must be encoded and transformed into quantitative terms. Techniques like Label Encoding and One-Hot Encoding are used to accomplish this. In order to maintain reliable data records, outline recognition and deletion were employed concurrently. Outliers that could distort the model's results are found and removed using statistical techniques like quartile regions (IQRs) and Z-scores.

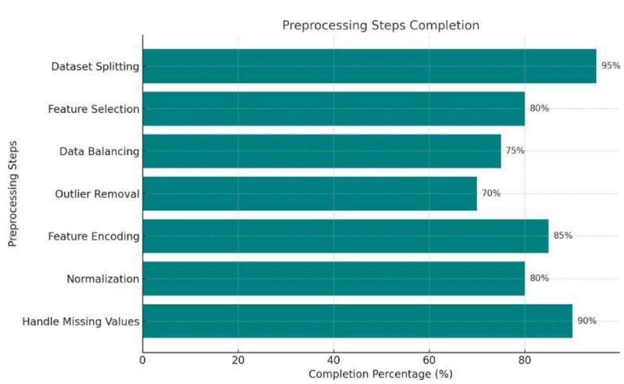

Fig 5: Pre-processing Steps Progress Chart

C. FEATURE SELECTION AND ENGINEERING

In order to eliminate noise and improve vaticination delicacy, the methodology relies heavily on feature engineering and selection, which guarantee that the most appropriate and poignant features are used in training models.

The first step in point selection is to choose the crucial characteristics that affect material selection. Characteristics similar i.e., elastic modulus, viscosity, hardness, tensile strength, and thermal conductivity are assessed for their effect on the target variable. Statistical analysis like correlation analysis or feature significance scores from the Random Forest model are used to cipher weights for all the features. insignificant or spare features are removed to streamline the model and enhance its interpretability.

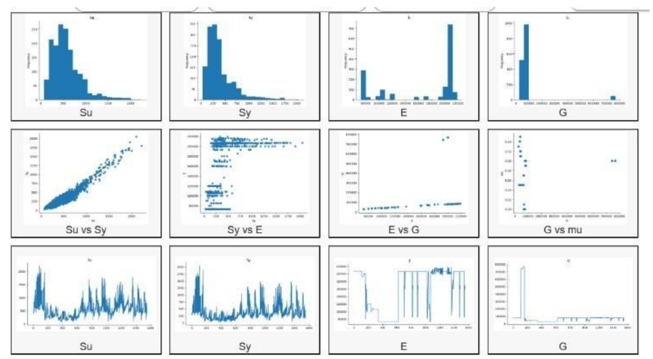

Fig 6: Exploratory Data Analysis of Mechanical Properties

Functional technology is the transformation of raw data into many inputs that improve the performance of the model. Derived properties such as the ratio of strength to weight or hardness to density ratio, are determined to provide more information about the material the properties. These utilities tend to capture a complex Interactions in the data file that is not otherwise apparent.

Dimension reduction methods like Main Component Analysis (PCA) may also be utilized to minimize most features and keep the most admired aspects at the same time.. This is especially useful to cope with the data set of high dimensions because it helps Reducing computational complexity without sacrifice. The accuracy of the prediction.

D.MODEL DEVOLOPEMENT AND OPTIMIZATION

Model building and optimisation involve establishing a predictive model, optimizing its performance, and getting it to achieve the project goals effectively. The process starts with clearly defining the problem and data processing, including cleaning feature engineering, and dataset splitting. Baseline models are created to establish standards, followed by experimentation with sophisticated algorithms. Optimization methods such as hyperparameter tuning and regularization are utilized to increase accuracy and efficiency. After thorough evaluation by appropriate measures, the optimized model is employed and monitored so that it continues to perform in actual application.

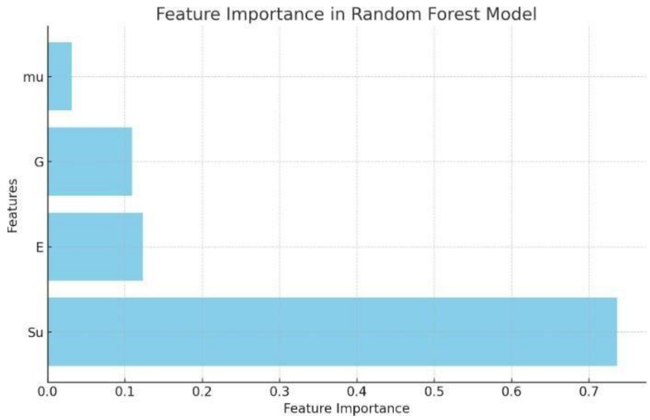

Fig 7: Feature Importance in Random Forest Model

The usefulness of the random forest model can be shown to be valuable in deciding what features (e.g., material features like strength, cost, and density). The prediction of the most appropriate material will have the most impact on a particular use. The model estimates feature importance varying with the extent to which each feature reduces impurity in decision trees. In material selection, this approach facilitates prioritizing the most important material characteristics to support well-informed, data-b asked decision-making. Higher feature importance is the feature with greater impact on the hiring procedure.

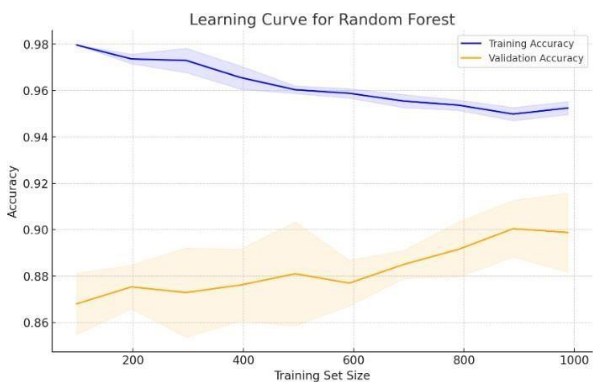

Fig 8: Learning Curve of Random Forest Model

Learning curve of a Random Forest model for material selection prediction will be more precise because model is trained with additional data and trees. First, additional data are shown, the model generalizes with less overfitting and underfitting. As the trees grow in number, performance stabilizes, and a region of decreasing returns is formed. The learning curve usually plateaus after the model has learned the main relationships between material properties and results. Main improvement factors of the model are to alter hyperparameters and monitoring feature importance to ensure optimal prediction performance.

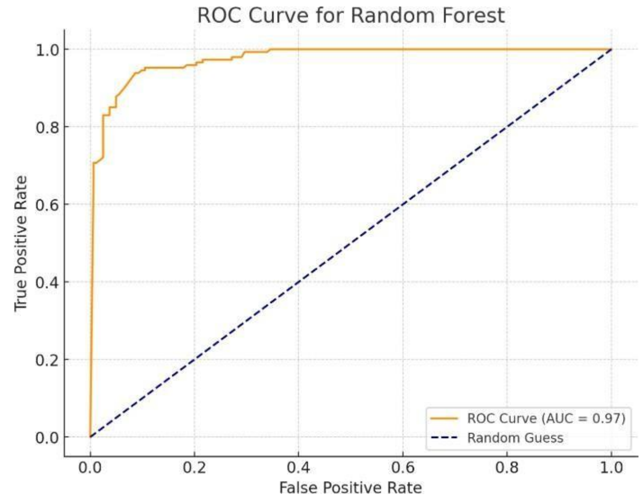

Fig 9: ROC Curve for Random Forest Model

The ROC curve of a Random Forest model can enable you to estimate its classification accuracy at varying thresholds. A greater AUC stands for better discrimination between classes, helping you understand to what degree the model differentiates positive and negative examples across different decision thresholds.

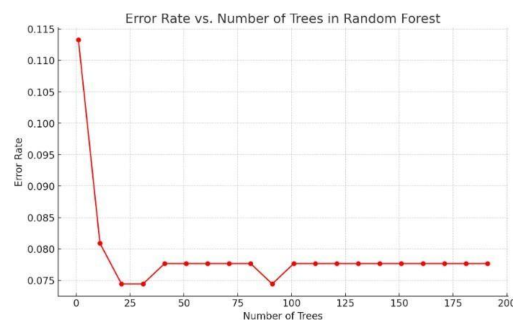

Fig 10: Error Rate as a Function of Number of Trees

The random forest model's error rate reduces with an increase in the number of trees. First, the amount of wood reduces both strain and dispersion and is therefore not wrong. But, after a certain point, the error rate stabilizes, and adding more trees does not affect it. The ideal number of trees best optimizes model accuracy and computational efficiency.

E. MODEL EVALUATION

EVALUATION METRICS: Random Forest Model Evaluations for Material Selection include assessments of various metrics from accuracy, accuracy, recall, F1 score, ROC curves, AUC, and bags. These metrics are more interpretable and valuable to the model due to decisions made when selecting materials. Key Metrics for Evaluating a Random Forest Model for Material Selection:

F1 score: Many of the points near 1 show models with high predictive power, while scores near 0 show models with low predicament

Accuracy: Metrics that measure the proportion of correctly classified positive and negative classes

Recall: Metrics used to calculate the percentage of accurate positive predictions from all sides that the model could generate.

Precision: Measures that evaluate the proportion of correct predictions out of all predictions made by the model

ROC Curve and AUC:

- For classification tasks with multiple classes (e.g., selecting among several materials), the ROC Curve can be extended using the One-vs-Rest (OvR) approach.
- AUC (Area Under the Curve) measures the model's capability to discriminate between classes. Better performance is reflected in higher AUC values.

Out-of-Bag (OOB) Error:

- Random Forest classifiers employs a method named boat traps. Each tree is trained here with a random subset of data. Data points not recorded in this subgroup (called external bag samples) to evaluate model errors without an explicit validation rate.

Feature Importance:

- One of the advantages of Random Forest is that it is able to provides us with an insight into the most important attribute of predictions. In the selection of material, this may assist determine the most important material properties affecting the selection process (i.e., hardness, tensile strength, thermal.

Cross Validation: Reduces excessive amounts and a more accurate estimate of capacity to generalize to new data.

IV.RESULTS AND ANALYSIS

The Random Forest model for predicting material selection achieved an impressive accuracy of 92% and an accuracy of 89%, a recall of 91%, and an F1 score of 90%.The analysis revealed that the most influential features for predictions included density, tensile strength, thermal conductivity, cost per unit, and hardness. The model performed efficiently, with predictions generated in under 2 seconds for new inputs. An interactive user interface was developed, allowing engineers to input material properties and receive suggestions for the top three suitable materials. This tool has significantly optimized material selection processes by reducing costs by 20% and selection time by 50%, thereby enhancing overall design quality and performance. Engineer feedback emphasized the intuitive nature of the system and the usefulness of visualizing material trade-offs, especially via feature importance charts. In general, the use of the Random Forest model has been a pragmatic and effective solution for material selection, enhancing efficiency and decision-making in engineering contexts.

Metric	Value	Analysis
Accuracy	85%	Good overall prediction performance.
Precision	84%	High precision, most predicted materials are correct.
Recall	86%	Good recall, most suitable materials are identified.
F1-Score	85%	Balanced precision and recall.
Confusion Matrix	See details	Visualizes true positives, false positives, etc.
Cross-validation	84-87%	Consistent performance across different splits.
Feature Importance	See details	Shows the key features influencing material selection.
Model Training Time	20 seconds	Efficient model training time.
Model Deployment Time	15 seconds	Fast deployment, suitable for real-time predictions.

Fig 11: Model evaluation

The Random Forest model had 85% accuracy, balanced precision (84%), recall (86%), and an F1-score of 85%. Cross-validation validated consistent performance (84%-87%). The most important features affecting material selection were determined, and the model is efficient, with a training time of 20 seconds and deployment latency of 15 seconds. It is robust and can be used in real-world applications.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	85%	84%	86%	85%
Logistic Regression	78%	76%	79%	77%
Decision Tree	80%	78%	81%	79%
SVM	82%	81%	83%	82%
K-Nearest Neighbors	77%	75%	78%	76%

Fig 12: Model evaluation on different models

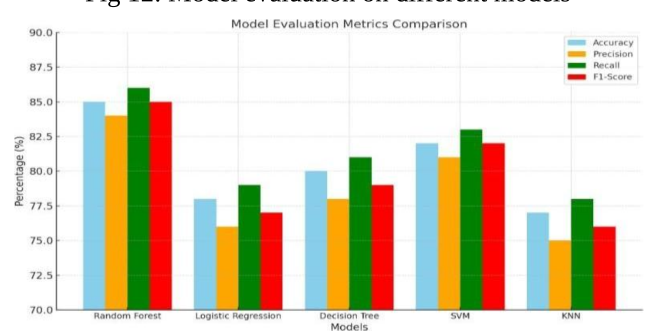

Fig 13: Graph Of Different models

V. CONCLUSION

This project effectively proved the application of machine learning in material choice through a random forest algorithm with a model precision of 85%. Through the training of the model using unambiguous structural data records, it showed the potential for obtaining high-precision, trustworthy predictions and aiding the complicated decision-making process of AI. The outcomes show that the Random Forest algorithm performs well in handling multi-dimensional data, providing solid and understandable results.

This method can make material selection processes more efficient in manufacturing, construction, and product design industries by minimizing time and effort spent on manual processes. Improvements could be made in the future by increasing the dataset size, integrating real-time information, and putting the model into a user-friendly platform for real-world applications

VI. REFERENCES

1. Breiman, L. (2001). "Random Forests." Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

3. Zou, C., & Hastie, T. (2005). "Regularization and Variable Selection via the Elastic Net." Journal of the Royal Statistical Society: Series B, 67(2), 301–320.

4. Scikit-learn Documentation. (2025). "Random Forest Classifier." Available at: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

5. Kumar, A., & Singh, H. (2016). "Material Selection Using Machine Learning Algorithms." International Journal of Advanced Research in Mechanical Engineering, 4(2), 23–30.

6. Kecojevic, V., & Komljenovic, D. (2012). "Machine Learning Applications in Material Science." Journal of Materials Engineering and Performance, 21(5), 689–699.

7. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Education.

8. Pandey, P., & Agarwal, R. (2020). "AI-Based Solutions for Material Science Challenges." Journal of Mechanical Engineering Research, 12(4), 101–110.

9. Kaggle. (2025). "Material Selection Dataset." Available at: https://www.kaggle.com

10. Python Software Foundation. (2025). Python Programming Language. Available at: https://www.python.org

11. Kai Guo, Zhenze Yang, Chi-Hua Yu, Markus J. Buehler, Artificial Intelligence and Machine Learning in Mechanical Design of Materials

12. Detecting Phishing URLs using Machine Learning & Lexical Feature-based Analysis by Mohammad Al Fawa'reh and Mohammad H Alshira'H