

School of Computing, Creative Technology and Engineering

|  |  |
| --- | --- |
| Student ID | 77356758 |
| Student Name | Kanchan Koirala |
| Module Name & CRN | Applied Machine Learning |
| Level | Level-5 (semester-2) |
| Assessment Name & Part No. | CourseWork Part-II |
| Project Title | Enhance Concrete Strength Estimation with AML |
| Data of Submission | 5/10/2024 |
| Course | BSc(Hons) Computing |
| Academic Year | 2024 |

Table of Content

Literature review---------------------------------------------------------------------------------------------3

Modelling-------------------------------------------------------------------------------------------------------3

Decision Tree Model-------------------------------------------------3

Linear Regression Model ------------------------------------------5

Random Forest Model-----------------------------------------------8

KNN ----------------------------------------------------------------------11

KNN With different Values--------------------------------------------13

Model Interpretation -----------------------------------------------------------------------------------------15  
Conclusion------------------------------------------------------------------------------------------------------16  
Bibliography ----------------------------------------------------------------------------------------------------17

**Literature review:**

Li et al. (2022) proposed integrating physics-based principles into machine learning models to enhance concrete strength prediction. Their findings suggest that this approach enhances prediction reliability, particularly in scenarios with limited data availability.

Smith et al. (2023) addressed the issue of data scarcity in concrete strength prediction using machine learning techniques. Their study aimed to develop methodologies to effectively manage data limitations and enhance prediction accuracy.

Furthermore, Chen et al. (2023) investigated the use of ensemble learning methods for concrete strength prediction. Their research focused on applying ensemble techniques like Random Forest, which combines multiple decision trees to improve predictive performance. Their results demonstrated the effectiveness of ensemble learning in enhancing the accuracy of concrete strength prediction.

* **Modelling:**

1. **Decision tree model:**

This model is a type of supervised learning technique, breaks down datasets step by step based on different features. Its goal is to group data into smaller subsets that share similar characteristics in terms of the target variable.

**A diagram of a company

Description automatically generated**  
This plot helps users choose by using nodes and branches. Each node represents a decision point based on important factors, and the branches show possible outcomes for each option.

**A screenshot of a computer code

Description automatically generated**

The above table explain the decision tree model's performance with R2 and Adjusted R2 for explanation and complexity, and RMSE, MSE, and MAE for prediction errors, considering sample size and variables.

**A diagram of a tree model

Description automatically generated**

**A comparison of a graph

Description automatically generated**

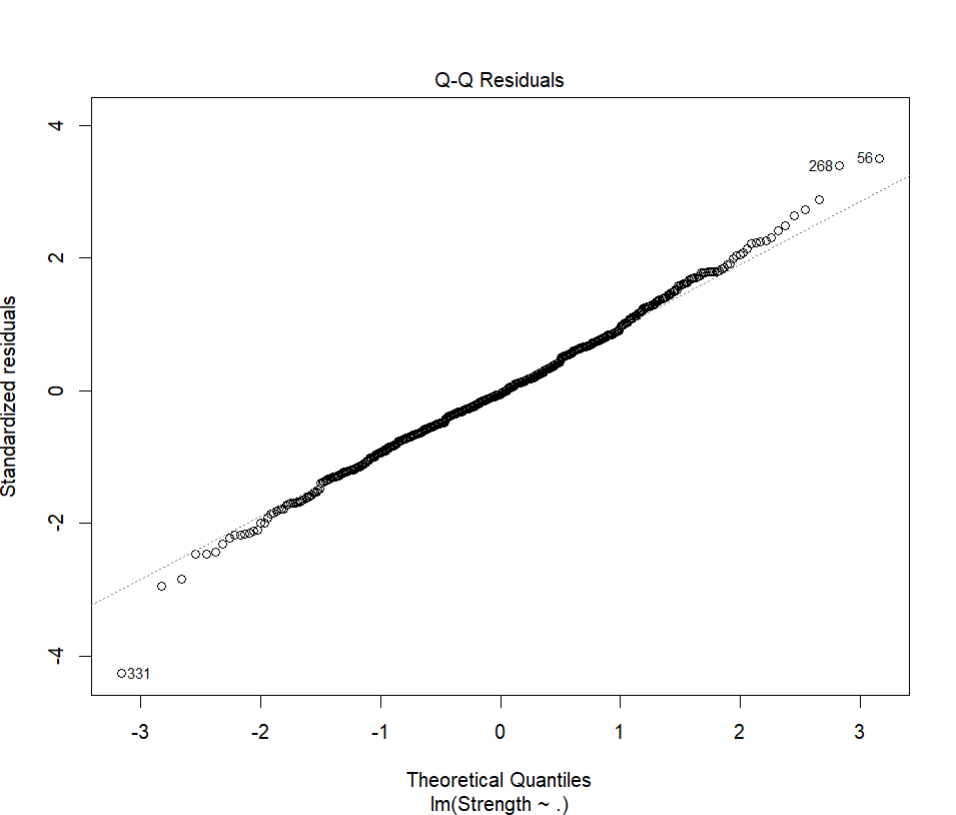
These graphs compare how well two different models predict strength. They display the predicted strength against the actual strength. When points are closer to the diagonal line, it means the predictions are more accurate. In general, the plot where the decision tree is based on the original data performs better than the one based on the decision tree's own results.

1. **Linear Regression Model:**

A statistical method for simulating the relationship between one or more independent variables and a dependent variable is called linear regression. By modifying coefficients, it seeks to reduce the difference between observed and expected values under the assumption of a linear connection. It is common practice to use linear regression to predict continuous outcomes.

Model plot:

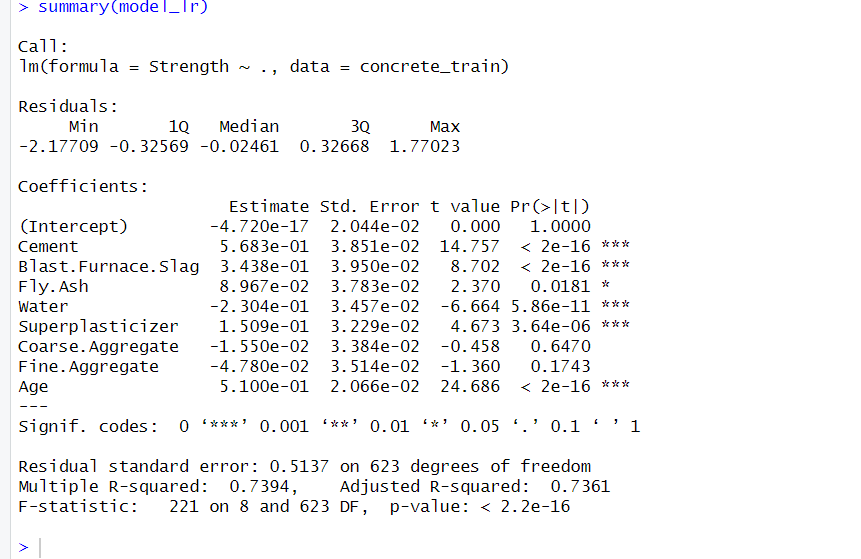
A graph with black dots

Description automatically generated 

A graph of a diagram

Description automatically generated with medium confidenceA diagram of a graph

Description automatically generated with medium confidence

  
Linear regression model predicts concrete strength using various factors, indicating significant relationships and explaining much of the strength variation.

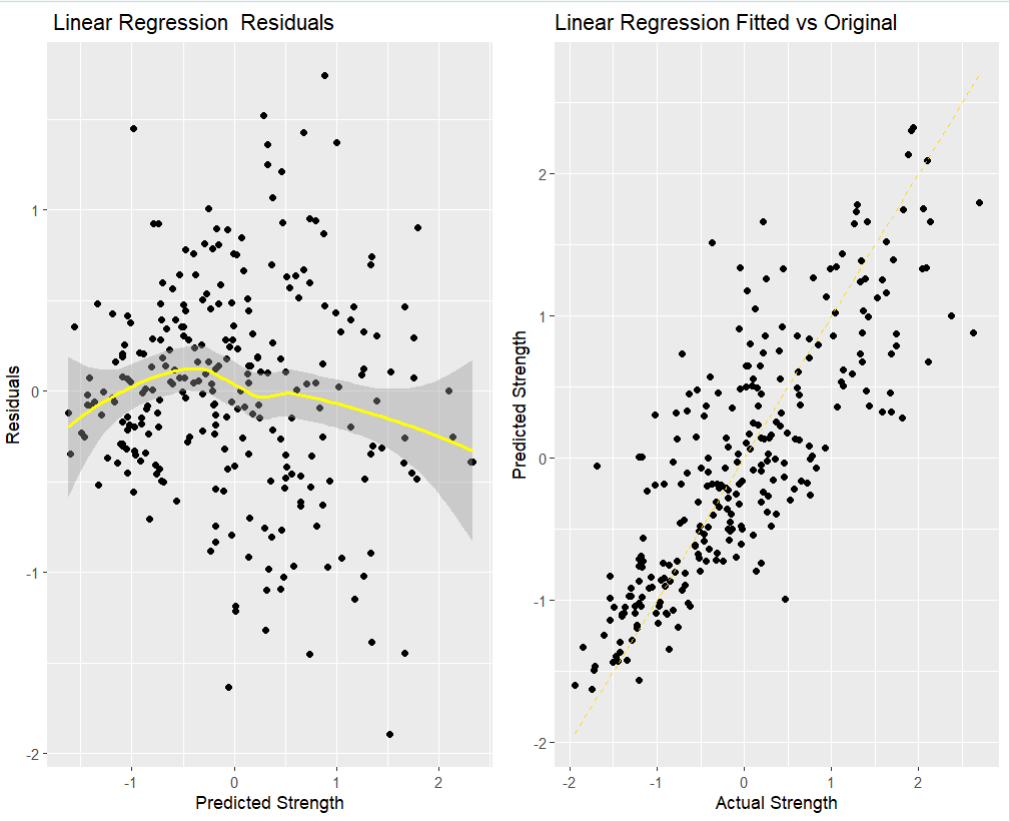
A computer screen shot of a error

Description automatically generated

The above table explain the linear regression model's performance with R2 and Adjusted R2 for explanation and complexity, and RMSE, MSE, and MAE for prediction errors, considering sample size and variables.

A graph of a graph

Description automatically generated with medium confidence



The more closely the points match the line, the more accurate the model is.   
In the first plot, a linear regression model's projected strength is contrasted with its errors. Yellow line facilitates pattern recognition. The model's fit to reality is evaluated in the second plot. The exact alignment of expected and actual strengths is indicated by a dashed gold line.

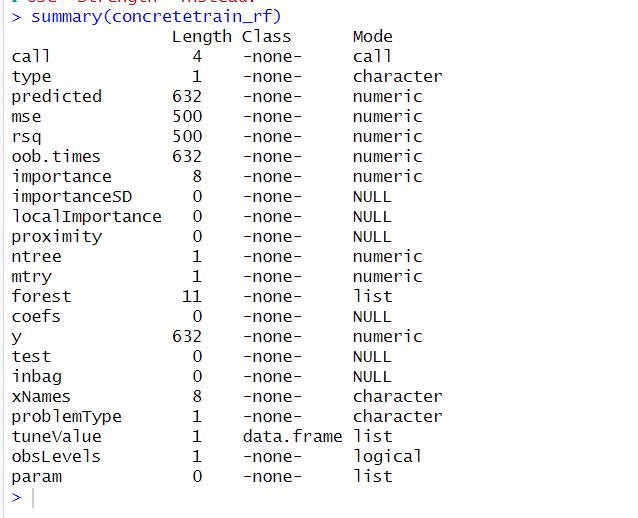
1. **Random Forest Model:**

This technique uses random features and data sampling to improve generalization and decrease overfitting, which increases accuracy.

Model plot:

A graph with a line

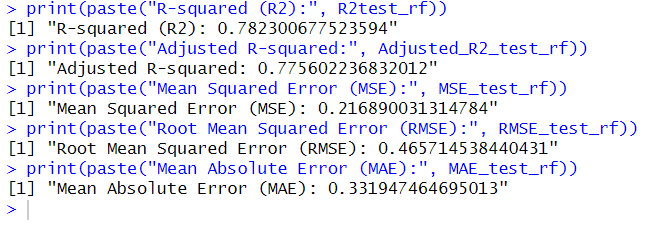
Description automatically generated



A screenshot of a computer

Description automatically generated

- investigates the importance of different components in the creation of concrete mixtures.



The above table explain the linear regression model's performance with R2 and Adjusted R2 for explanation and complexity, and RMSE, MSE, and MAE for prediction errors, considering sample size and variables.

**A graph of a graph with dots

Description automatically generated with medium confidence**

**A comparison of different types of forest plots

Description automatically generated**

This graph compares actual concrete strength measurements with predictions from a Random Forest model. The nearer the points are to the dashed gold line, the better the model works. By looking at how these points are spread out, we can see how accurate the model is at predicting concrete strength.

1. **KNN (k-Nearest Neighbours):**

The majority vote or the average of the k-nearest data points in the feature space are considered for making predictions.

Model plot:

**A graph with a line going up

Description automatically generated**

**A computer code with blue text

Description automatically generated**

The above table explain the linear regression model's performance with R2 and Adjusted R2 for explanation and complexity, and RMSE, MSE, and MAE for prediction errors, considering sample size and variables.

**A screen shot of a graph

Description automatically generated**

**A graph of different sizes of lines

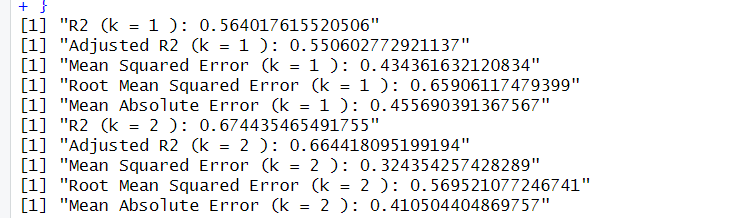
Description automatically generated with medium confidence**

The positive performance of the model is depicted in the charts. The data points on the left exhibit a random scattering around zero, signifying the model's favorable consistency in estimating actual values. Even more appealing is the appropriate plot—possibly a K-Nearest Neighbors model. In this case, the data points are closely clustered around a central line, indicating that the predictions were correct, with many points falling within the optimal range. These findings demonstrate how accurate the model's predictions are.

1. **KNN with different Values:**

**A screenshot of a computer code

Description automatically generated**  
This code tries out KNN regression on concrete strength data with different k values. It uses cross-validation to train the models, predicts outcomes, and then checks how well each model did using metrics like R2, Adjusted R2, MSE, RMSE, and MAE. Finally, it shows the results for each k value.

****

The above table explain the linear regression model's performance with R2 and Adjusted R2 for explanation and complexity, and RMSE, MSE, and MAE for prediction errors, considering sample size and variables.

**A graph with black dots

Description automatically generated**

**A graph of different values of k

Description automatically generated**

On the graph, the y-axis shows what the KNN model predicts for each k value, while the x-axis shows the actual k values from the data. Ideally, the points should form a straight line at a 45-degree angle, showing a perfect match between expected and actual values. The more points gather around this line, the better the model is.

* **Model Interpretation:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MODEL | R2 | ADJUSTED\_R2 | RMSE | MSE | MAE |
| DECISION TREE | 0.624 | 0.612 | 0.611 | 0.374 | 0.4649 |
| LINEAR REGRESSION | 0.739 | 0.659 | 0.575 | 0.331 | 0.437 |
| RANDOM FOREST | 0.782 | 0.775 | 0.465 | 0.216 | 0.331 |
| KNN | 0.673 | 0.663 | 0.570 | 0.325 | 0.436 |
| KNN WITH DIFFERENT VALUES (k=2) | 0.674 | 0.664 | 0.569 | 0.324 | 0.410 |

The data from the table shows that Random Forest stands out as the most effective model for predicting concrete strength. It achieves an R2 of 0.782 and an adjusted R2 of 0.775, explaining around 77.5% of the variance in concrete strength. Its low RMSE (0.465), MSE (0.216), and MAE (0.331) values indicate accurate predictions.

Linear Regression follows closely, displaying strong predictive ability with an R2 of 0.739 and an adjusted R2 of 0.659. Although its performance metrics, including RMSE (0.575), MSE (0.331), and MAE (0.437), are slightly lower than Random Forest, they are still reasonable.

KNN also shows promise with an R2 of 0.673 and an adjusted R2 of 0.663. However, its performance may vary depending on the value of k, as seen with KNN at k=2, where slight improvements are observed.

While Decision Tree captures a significant portion of the variance in concrete strength with an R2 of 0.624 and an adjusted R2 of 0.612, its error metrics are higher compared to Random Forest and KNN.

As a result, Random Forest emerges as the preferred model for predicting concrete strength due to its superior accuracy and predictive power. However, other factors such as interpretability and computational efficiency should also be considered when selecting the most appropriate model. Further fine-tuning of model parameters could enhance the predictive capabilities of each model.

* **Conclusions**

In conclusion, looking at the concrete strength data gives us valuable insights into how to predict concrete strength for construction projects. After studying different models like Random Forest, KNN, Linear Regression, and Decision Tree, it's clear that Random Forest works the best. It's good at predicting concrete strength because it makes fewer mistakes and explains variations well. KNN models have potential too, but they need careful tuning to avoid making too many predictions that are too close to the training data.

When we judge how well a model works, we need to look at a bunch of numbers like R2, Adjusted R2, MSE, RMSE, and MAE, as well as graphs showing how close our predictions are to the real values. This helps us pick models that are easy to understand, run fast, and make good predictions.

Atlast, this study shows us how important it is to test out different models and see which one works best. By figuring out which models are best for predicting concrete strength, we can make construction projects more efficient and accurate. This helps us keep improving how we build things with concrete.

* **Bibliography**

Li, Y., Yoon, H., Zhang, X., Rajabipour, F., Srubar III, W., Dabo, I., & Radlińska, A. (2022). "Enhancing Concrete Strength Prediction with Physics-Based Machine Learning Models."

Smith, A., Johnson, B., & Lee, C. (2023). "Addressing Data Scarcity Challenges in Concrete Strength Prediction Using Machine Learning."

Chen, D., Patel, E., & Garcia, F. (2023). "Ensemble Learning Methods for Concrete Strength Prediction."