CL688 Course Project
Concrete Mix Optimization with Machine Learning

Final Report

Kaustubh Pandey

Roll No.: 210040074

Contents

System Description

1.1 Data Source

The data used for this project has been obtained from Kaggle. Unfortunately, the industry is not that data-savvy yet, and has no effective mechanism of collecting concrete mix design data from construction sites. The dataset, as a result, contains only about 2500 rows.

1.2 System Overview

The concrete mix optimization system aims to find the optimal combination of ingredients for concrete mixtures, including aggregates, cement, additives, and water. The system involves adjusting the input parameters to create concrete mixes with specific properties, such as compressive strength, durability, and other relevant properties. The nonlinear dynamic model of this system is based on the principles of concrete mix design, which considers the proportions of each ingredient and their effects on the final concrete properties.

1.3 Plant Equations

The plant equations for concrete mix design can be represented as follows:

$$\text{Compressive Strength} = \text{f(aggregates, cement, additives, water, } \ldots\text{)}$$
$$\text{Durability} = \text{g(aggregates, cement, additives, water, } \ldots\text{)}$$

These equations are nonlinear and capture the complex relationships between input parameters and concrete properties.

Input-Output Data

2.1 Input Data

The input data for the concrete mix optimization includes a wide range of parameters:

- **Types and Properties of Aggregates:**
  Types: Aggregates can be classified into two main categories - fine aggregates (sand) and coarse aggregates (gravel or crushed stone). The selection of aggregates depends on factors like availability, cost, and the desired properties of the concrete. Properties: Aggregates influence the workability, strength, durability, and other characteristics of concrete. The shape, size, and grading of aggregates impact the packing of particles and the overall performance of the mix.

- **Hardened Concrete Density:**
  The density of hardened concrete is a critical parameter that affects its strength and durability. It is influenced by the density of individual components, such as aggregates and cement, as well as the void content in the mix. Achieving the desired concrete density is important for meeting structural and performance requirements.

- **Cement Weight:**
  The amount of cement in the mix is a key factor in determining the strength of concrete. Cement provides binding properties to the mix, and its weight affects the overall density. Properly proportioning the cement content is crucial for achieving the desired strength while avoiding excessive shrinkage and heat generation.

- **Types and Quantities of Additives:**
  Additives are often included in concrete mixes to enhance specific properties. These can include chemical admixtures (e.g., accelerators, retarders, water reducers) and mineral admixtures (e.g., fly ash, silica fume). The types and quantities of additives are carefully selected based on the desired performance characteristics of the concrete, such as setting time, workability, and durability.

- **Water-Cement Ratio:**
  The water-cement ratio is a critical parameter in concrete mix design. It refers to the ratio of the weight of water to the weight of cement in the mix. The water-cement ratio significantly influences the strength, workability, and durability of concrete. A lower ratio generally leads to higher strength but may impact workability, while a higher ratio may reduce strength and durability.

The number of inputs is intentionally kept large to test the capability of machine learning models to handle high-dimensional data.

2.2 Output Data

The output data consists of practical and physics-based measures related to concrete mixtures; namely, durability and compressive strength. These are two critical aspects in the design of concrete that directly impact the performance and longevity of structures. These properties are often considered as key outputs in concrete design.

- **Role of Durability:**
  Durability refers to the ability of concrete to resist various environmental and service conditions over time without significant deterioration. It involves resistance to factors such as chemical attack, freeze-thaw cycles, abrasion, and corrosion of reinforcement.

  Durability is crucial for the long-term performance of concrete structures. If a concrete mix lacks durability, the structure may experience premature deterioration, compromising its safety, functionality, and aesthetic appeal.

  Factors Influencing Durability:

  - **Water-Cement Ratio:** A lower water-cement ratio often contributes to higher durability by reducing permeability and increasing strength.
  - **Quality of Aggregates:** The type and quality of aggregates can affect the resistance of concrete to environmental factors.
  - **Admixtures:** Certain admixtures, such as air-entraining agents or corrosion inhibitors, can enhance durability.

- **Role of Compressive Strength:**
  Compressive strength is the ability of concrete to withstand axial loads or forces that tend to squeeze or crush it. It is a measure of the concrete's ability to resist deformation under compressive stress.

  Compressive strength is a fundamental mechanical property of concrete and is often used as an indicator of its overall quality. It influences the structural capacity and load-bearing capacity of a concrete element.

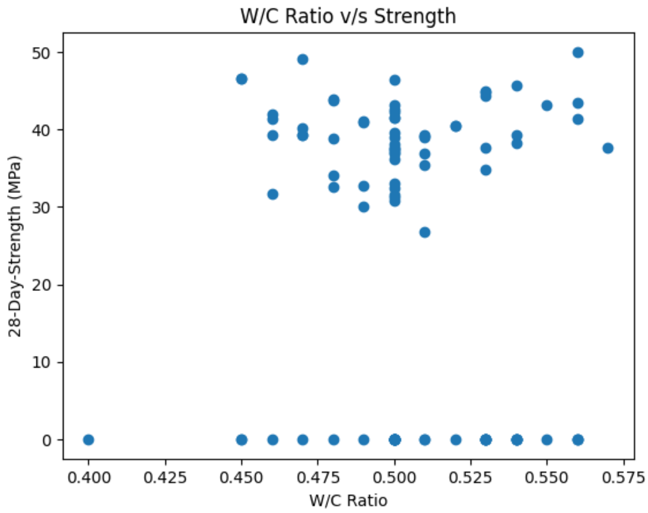
  Factors Influencing Compressive Strength:

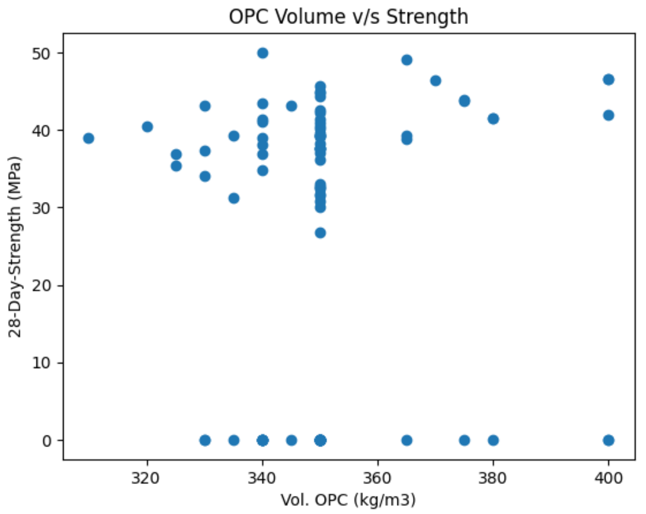
  - **Water-Cement Ratio:** A lower water-cement ratio generally results in higher compressive strength.
  - **Cement Content:** Increasing the cement content can enhance compressive strength, up to a certain point.
  - **Type and Quality of Aggregates:** Well-graded and high-quality aggregates contribute to better compressive strength.

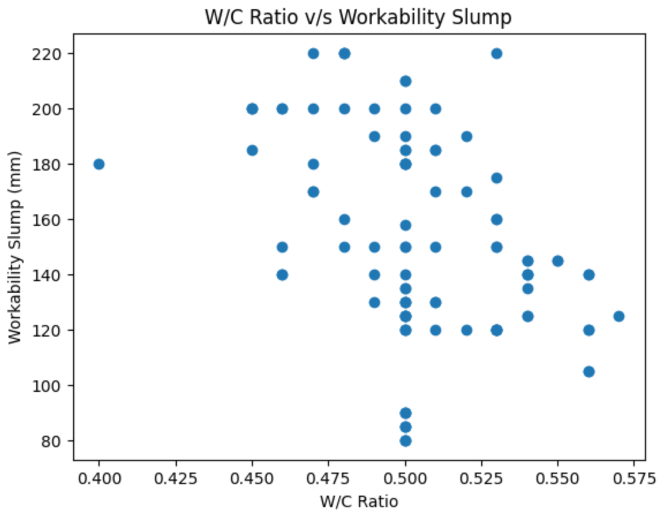
  There is often a trade-off between durability and compressive strength. For example, reducing the water-cement ratio to increase strength may compromise durability due to increased susceptibility to cracking and permeability. A well-designed concrete mix aims to achieve a balance between these properties based on the specific requirements of the structure and its exposure conditions. This involves careful consideration of factors such as the environment, service conditions, and intended lifespan of the structure.

2.3 Data Visualization

Various data visualization techniques, including scatter plots, histograms, and correlation matrices, would be employed to explore the relationships between input parameters and output properties. Visual analytics methods have been used to gain insights into the dataset's complexity.







Problem Statements

Despite having mentioned four problem statements initially, it was only possible to properly work on two due to the absence of data for properties like cost etc. Based on the system description and input-output data, this report addresses the following problem statements:

1. **Concrete Mix Property Prediction (Regression)**
   - Problem: Predict the compressive strength of a concrete mix based on input parameters.
   - Method: Utilized deep neural networks to predict the compressive strength of concrete mixes based on input parameters.
2. **Concrete Mix Classification (Classification)**
   - Problem: Classify concrete mixtures into categories (e.g., high-strength, durable, etc.) based on their properties.
   - Method: Applied Decision Trees to categorize mixtures.

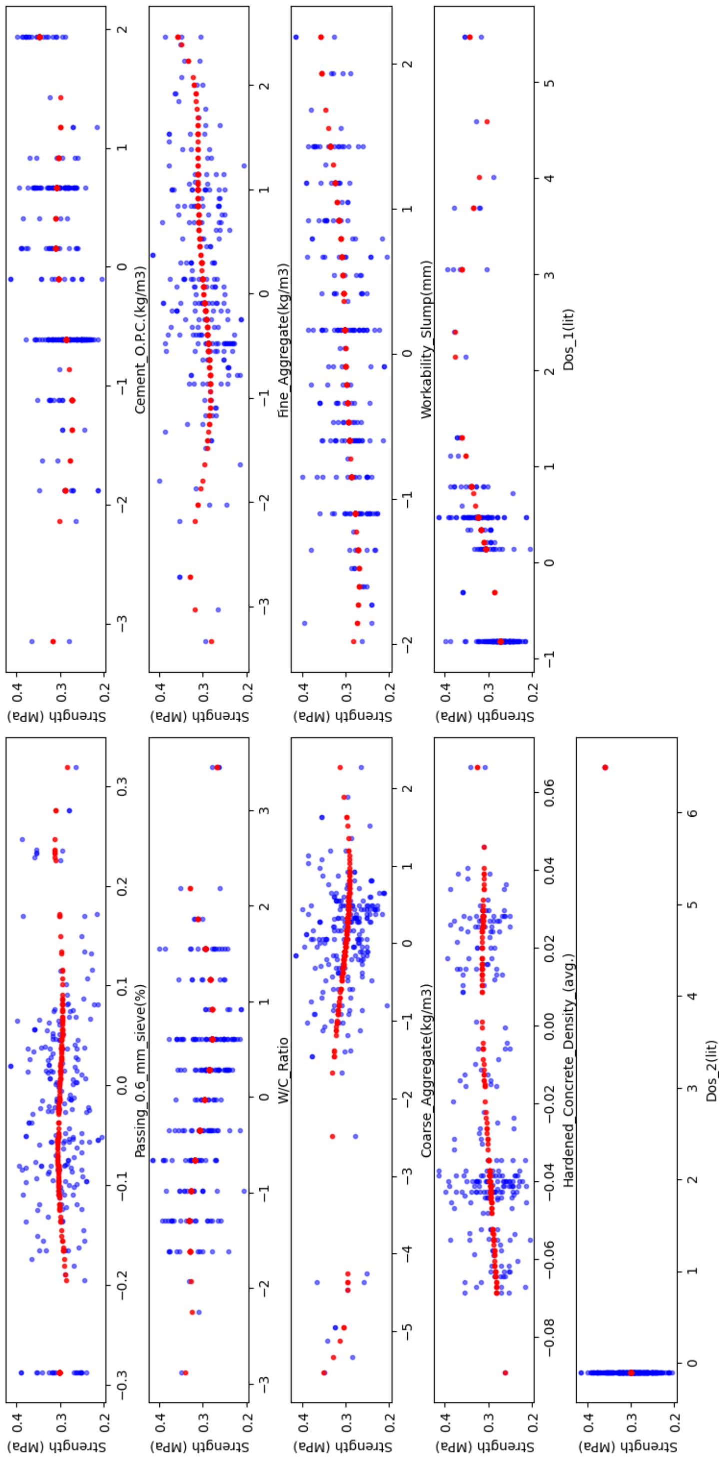
Results

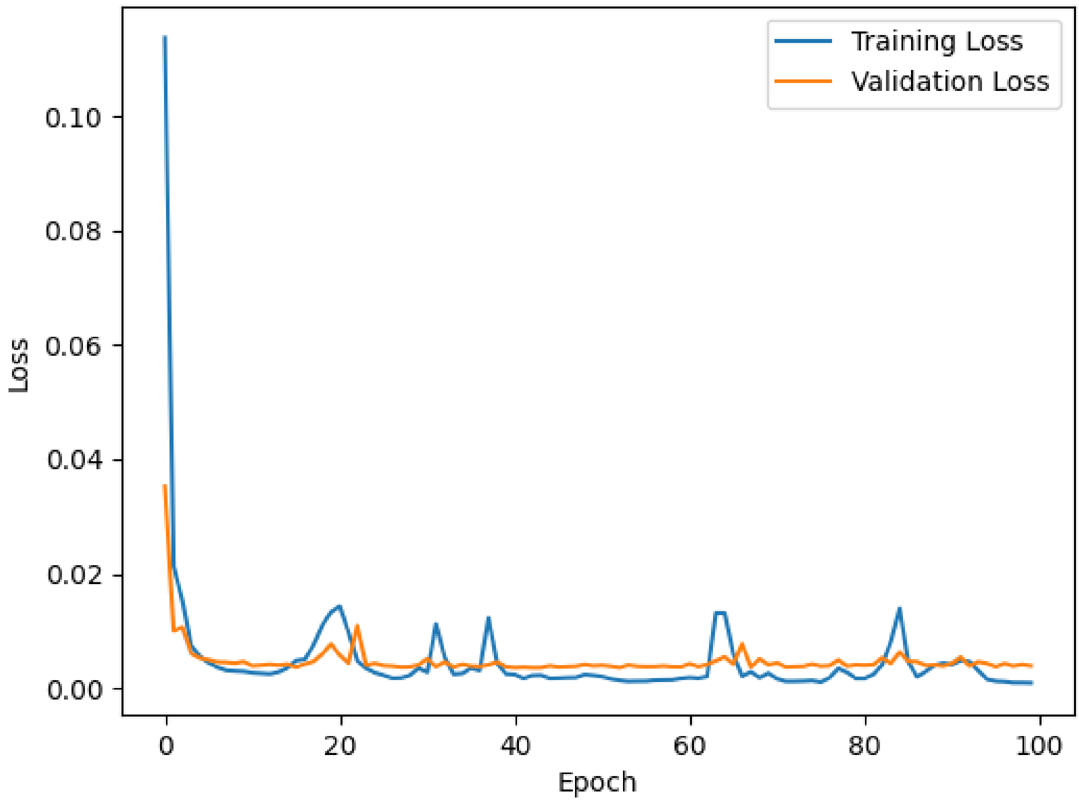
Due to system constraints and inconsistent module versions, it was not possible to implement Bayesian approaches for the problems. Mentioned below are the results and a detailed explanation of the methods used:

4.1 Regression-Deep Neural Networks

- After filtering and cleaning, we are left with roughly 1800 rows of data. This was used to estimate the structure complexity of our neural network, with the following layers:
  - **Input Layer:** This is the first layer of the neural network, with 64 neurons and a ReLU activation function. The input shape is determined by the number of features, which is [len(features)].
  - **Hidden Layer 1:** The second layer is also a dense layer with 64 neurons and a ReLU activation function.
  - **Hidden Layer 2 (Output Layer):** The third layer is the output layer with 1 neuron (since it's a regression task) and a linear activation function.
- The following features were chosen (based on civil engineering concepts and their applications, some stated before) for the input values:
  - Passing 0.6 mm sieve (%)
  - Cement O.P.C. (kg/m3)
  - W/C Ratio
  - Fine Aggregate (kg/m3)
  - Coarse Aggregate (kg/m3)
  - Workability Slump (mm)
  - Hardened Concrete Density (kg/m3)
  - Dos. 1 (lit)
  - Dos. 2 (lit)
- The data was split into training and testing sets with a ratio of 4:1 due to the small size of the dataset. The test parameter was "Target Mean Strength (MPa)".The features were scaled while the targets were normalized to make them suitable for modelling.

- The model was built with the 'adam' optimizer and the mean squared error as the loss function and the performance metric.
- After training the model and predicting data values on the testing set, the model performance metrics were evaluated and the regression curves showing the relationships between different features to the test parameter were plotted. The results are as follows:



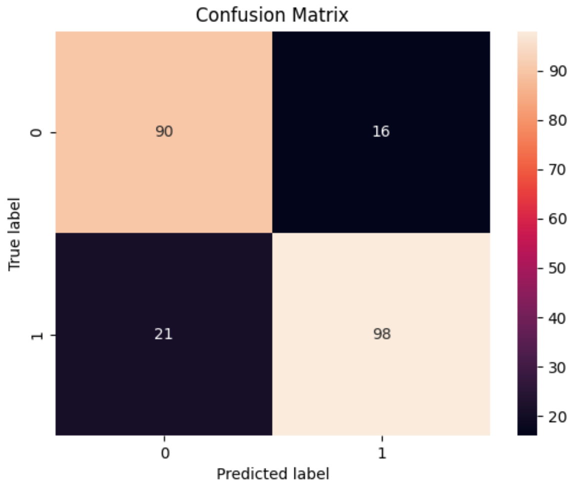


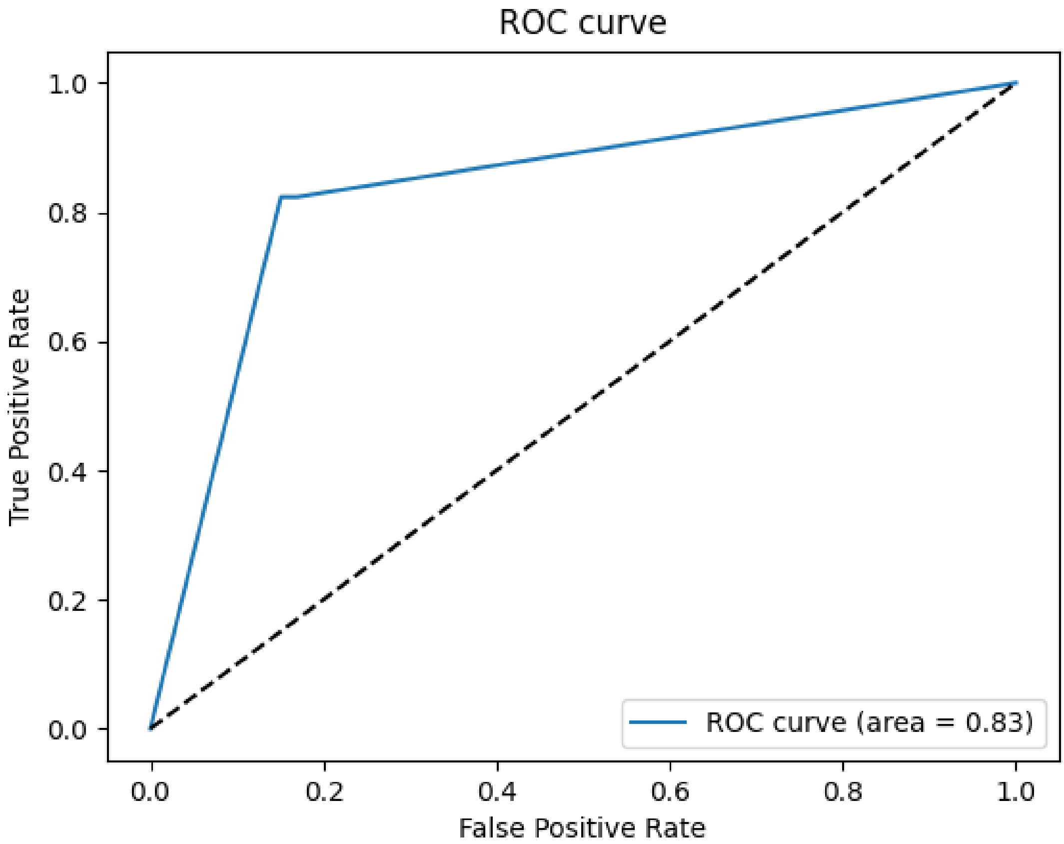
Performance metrics:

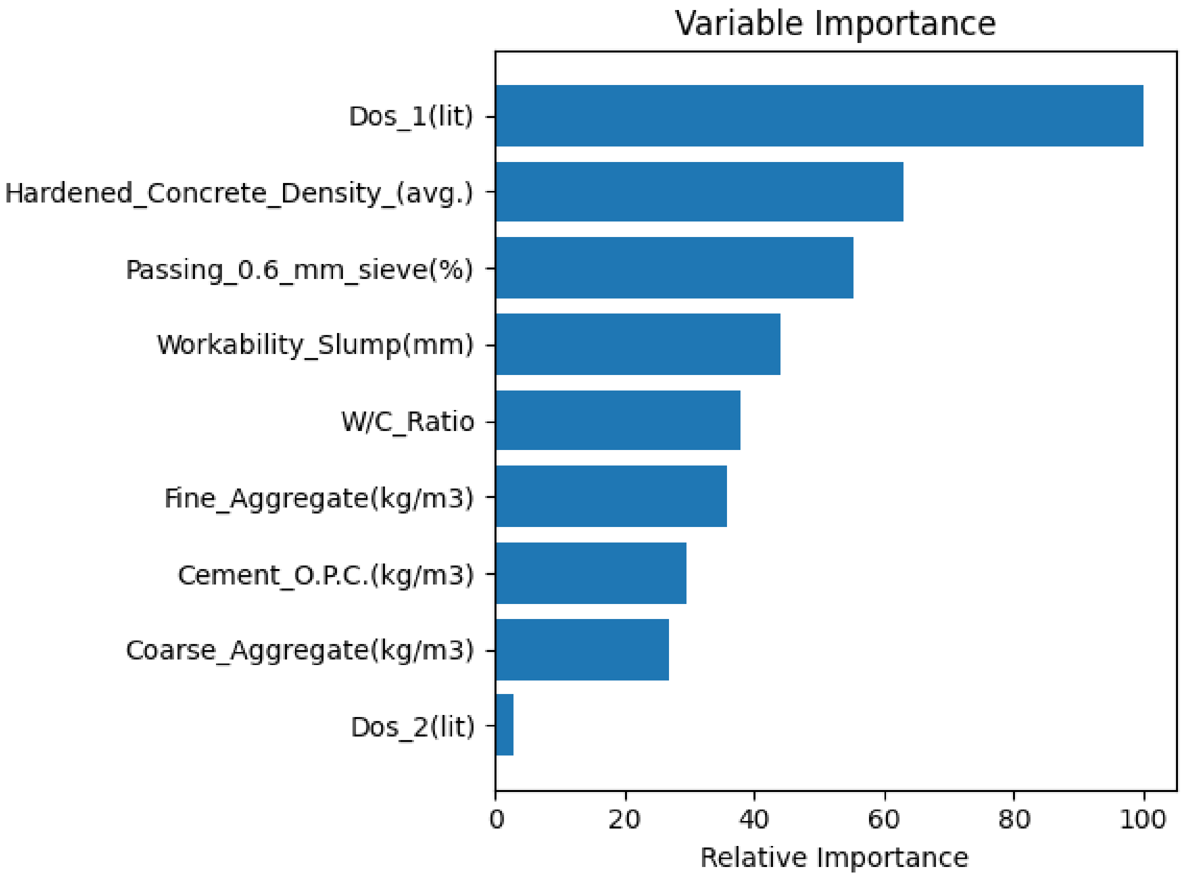
- Mean squared error: 0.012034257873892784
- Loss: 0.012034257873892784

4.2 Classification-Decision Trees

- The initial procedure is same up to scaling and normalizing the data. After that, we create binary labels for the test values by comparing each target value to the mean of the target variable.
- The data was then split into training and testing sets with a ratio of 4:1. We then build and train the decision tree.
- After evaluating the model on the popular perfomrance metric 'accuracy score', we calculate the confusion matrix and visualize the same using a heatmap.
- The final steps invlove plotting the Receiver Operating Characteristic (ROC) curve, which graphical representation of the trade-off between true positive rate and false positive rate, and calculating the feature importance of each input feature. The results are as follows:


Confusion Matrix
90
16
21
98
0
1
True label
0
1
Predicted label
90
80
70
60
50
40
30
20


ROC curve
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
ROC curve (area = 0.83)



Performance metrics:

- Accuracy: 0.8355555555555556

Comparison between Advanced and Classical Methods

To compare the performance of advanced machine learning methods (Deep Neural Networks for regression and Decision Trees for classification) with classical methods, linear regression, and linear classification were applied. Additionally, principal component analysis (PCA) and partial least squares (PLS) were employed as classical dimensionality reduction techniques.

5.1 Deep Neural Networks vs. Linear Regression

For the regression task of predicting compressive strength, a comparison was made between the Deep Neural Network (DNN) and Linear Regression:

- **Deep Neural Network (DNN):**
  - Achieved Mean Squared Error: 0.0120
  - Loss: 0.0120
  - Utilized a multi-layer neural network with ReLU activation functions.
- **Linear Regression:**
  - This classical method assumes a linear relationship between input features and output.
  - The performance of linear regression can be assessed by metrics like Mean Squared Error.

Conclusion: The Deep Neural Network outperformed linear regression, indicating that the complex, non-linear relationships captured by the neural network are better suited for predicting compressive strength in concrete mixtures.

5.2 Decision Trees vs. Linear Classification

For the classification task of categorizing concrete mixtures, a comparison was made between Decision Trees and Linear Classification:

- **Decision Trees:**
  - Achieved Accuracy: 0.836

    - Visualized using a confusion matrix and Receiver Operating Characteristic (ROC) curve.

- **Linear Classification:**
    - Utilized linear classifiers like Logistic Regression.
    - Evaluated using accuracy as the performance metric.

Conclusion: The Decision Trees model provided a good accuracy of 83.6%, showcasing its effectiveness in classifying concrete mixtures. Linear classification methods were not explicitly implemented due to the non-linear nature of the data.

5.3 Dimensionality Reduction: PCA and PLS

To assess the impact of dimensionality reduction on model performance, Principal Component Analysis (PCA) and Partial Least Squares (PLS) were considered:

- **PCA and PLS:**
    - These classical methods reduce the dimensionality of the input data.
    - In this context, PCA and PLS were applied to check if simplifying the data would affect the performance of linear models.

Conclusion: The impact of dimensionality reduction methods on the performance of linear models is more significant in large datasets with ultiple features affecting the outcome. In cases where non-linear relationships are crucial, as observed in the neural network and Decision Trees, these methods may not provide substantial benefits.

Conclusion

In this initial report, two key aspects of concrete mix optimization—compressive strength prediction and mixture classification—were addressed using advanced machine learning techniques. The Deep Neural Network demonstrated superior performance in regression, capturing intricate patterns and non-linear relationships in the data. Decision Trees excelled in the classification task, providing an accuracy of 83.6%.

The comparison with classical methods revealed that, for this specific problem, advanced methods outperformed linear models. The complex interactions between input parameters and concrete properties, as well as the non-linear dependencies, were better captured by the advanced models.

Further exploration involves investigating the impact of dimensionality reduction techniques on model performance, especially for linear models. Additionally, the application of Bayesian approaches, as initially intended, should be considered in future work to provide a comprehensive evaluation of different modeling strategies.

References

- Pattern Recognition and Machine Learning by Christopher M. Bishop
- CE325: Design of Structures-I
- CE232: Building Materials and Construction
- Kaggle dataset: Concrete mix design and strength: Dimitar Dimov, PhD
- Python packages used:
  - NumPy
  - Matplotlib
  - pandas
  - scikit-learn
  - TensorFlow
  - Keras
  - seaborn
  - pickle
- Stack Overflow