



**Module Code & Module Title**

**Level 6 – Artificial Intelligence**

**Assessment Type**

**Semester**

**2024/25 Autumn**

**Student Name: Jay Binayak Baral**

**London Met ID: 22067892**

**College ID: np01cp4a220271@islingtoncollege.edu.np**

**Assignment Due Date: Wednesday, December 25, 2024**

**Assignment Submission Date: Wednesday, December 25, 2024**

**Submitted To: Dipeshor Silwal**

**Word Count (Where Required):3500**

*I confirm that I understand my coursework needs to be submitted online via MST Classroom under the relevant module page before the deadline in order for my assignment to be accepted and marked. I am fully aware that late submissions will be treated as non-submission and a mark of zero will be awarded.*

*A screenshot of a computer

Description automatically generated*

*A screenshot of a web page

Description automatically generated*

Table of Contents

[1.Introduction: 5](#_Toc187373555)

[1.1 Introduction to AI concepts: 6](#_Toc187373556)

[1.2 Introduction to Prediction system: 7](#_Toc187373557)

[1.2.1 Motivation: 9](#_Toc187373558)

[1.2.2 Aims : 10](#_Toc187373559)

[1.2.3 Objective of the system: 10](#_Toc187373560)

[2.Background: 11](#_Toc187373561)

[2.1 Research on Problem scenario: 12](#_Toc187373562)

[2.2 Description of selected Domain: 14](#_Toc187373563)

[2.4 System Architecture: 16](#_Toc187373564)

[2.5 Advantages of working around the problem domain/ algorithm implementations: 17](#_Toc187373565)

[2.6 Considered drawbacks revolving the problem domain/ algorithm implementations: 18](#_Toc187373566)

[2.7 Dataset: 19](#_Toc187373567)

[2.7.1 Dataset design: 19](#_Toc187373568)

[2.8 Review and analysis of similar system: 34](#_Toc187373569)

[3.Solution: 38](#_Toc187373570)

[3.1 Approach to solving problems: 38](#_Toc187373571)

[3.2 Algorithm and Functions: 40](#_Toc187373572)

[3.3 Pseudocode: 47](#_Toc187373573)

[3.4 Algorithm and flowchart: 48](#_Toc187373574)

[3.5 Flowchart: 49](#_Toc187373575)

[3.6 Development Procedure: 50](#_Toc187373576)

[3.7 Development platform : 51](#_Toc187373577)

[3.8 Language used: 52](#_Toc187373578)

[3.9 Module/ library used: 53](#_Toc187373579)

[3.10 Data Visualisation: 54](#_Toc187373580)

[4.Conclusion: 55](#_Toc187373581)

[4.1 Analysis of the work done: 55](#_Toc187373582)

[4.2 How the solution addresses real world problems: 56](#_Toc187373583)

[4.3 Further Work: 56](#_Toc187373584)

[References 57](#_Toc187373585)

**Table of figures:**

[Figure 1: lasso regression 15](#_Toc186017140)

[Figure 2: ridge regression 16](#_Toc186017141)

[Figure 3:support vector regression 17](#_Toc186017142)

[Figure 4: Decision tress regression 18](#_Toc186017143)

[Figure 5: formula of mean absolute error 19](#_Toc186017144)

# 1.Introduction:

The purpose of this coursework was to conduct research on AI-related topics and provide a conceptual solution for the selected issue utilizing the relevant diagrams and pseudocode. According to need in the current situation, Prediction System was selected among other topics for this coursework.

This report presents an analysis of the Concrete Strength Prediction Model and its dataset, which contains information about various components of concrete and their compressive strength. The goal of this analysis is to build a model that can predict the compressive strength of concrete based on its components.

Based on a number of input variables, including mix proportions, material characteristics, and ambient circumstances, the Concrete Strength Prediction System forecasts the compressive strength of concrete. As a basic building material, concrete is essential to guaranteeing the longevity, safety, and affordability of constructions. Predicting its strength accurately is crucial for maximizing material use, guaranteeing safety regulations, and cutting project expenses.

The link between the input factors (such as the cement content, water-to-cement ratio, aggregate type, curing time, etc.) and the concrete's compressive strength is usually modeled in this system using machine learning methods. The main objective is to create a model that can forecast concrete strength based on these characteristics, providing a more effective and dependable substitute for time-consuming and resource-intensive old experimental methods.

Large datasets of past concrete mix designs and their accompanying strengths may be analyzed by these systems thanks to developments in artificial intelligence, especially in machine learning techniques like neural networks, decision trees, and regression models. This makes it possible to make accurate projections, which enhances construction industry decision-making and guarantees the best possible material performance in a range of circumstances.

## Introduction to AI concepts:

The ability of technology, particularly computer systems, to mimic human intelligence processes is known as artificial intelligence. Natural language processing, speech recognition, machine vision, and expert systems are a few examples of AI applications. Large volumes of labeled training data are usually processed by AI systems, which then look for patterns and connections to forecast future events. An image recognition algorithm may learn to recognize and describe items in photographs by examining millions of instances, much how a chatbot trained on text chats can have realistic discussions with users. (Crabtree, 2024)

The creation of models and algorithms that let computers learn from data and make judgments or predictions without explicit programming is the main goal of machine learning. The following are important aspects of machine learning:

* Feature Engineering: To help the algorithm provide precise predictions, machine learning specialists manually engineer or choose pertinent features from the input data.
* Unsupervised and Supervised Learning: Machine learning algorithms fall into two categories:

unsupervised learning, in which algorithms find patterns and structures in unlabeled data, and supervised learning, in which models learn from labeled data with known outcomes.

* Wide Range of Use: Machine learning methods are used in a number of fields, such as recommendation systems, natural language processing, and picture and audio recognition. (simplilearn, 2025)

Using data mining techniques, predictive modeling is produced, which aids in forecasting future trends and creates a prediction system. This can occur in a variety of fields, including corporate intelligence, stock markets, weather, and health effects. By calculating the likelihood of an event occurring or not, forecasting in this manner aids in decision-making and lowers risk.

In order to gain deeper insights, predictive modeling can produce intricate simulations that go beyond basic statistics like standard deviation. Prediction systems depend on gathering problem-related data, so it's critical that the data be cleansed and prepared before being combed for precise predictions (keele, 2025)

Artificial Intelligence (AI) is a key component of prediction systems because it uses machine learning algorithms to evaluate past data, spot trends, and generate precise forecasts. In order to forecast results (like concrete strength), artificial intelligence (AI) models like regression, decision trees, and neural networks learn from input features (such concrete mix proportions and material attributes). As more data is added to these algorithms over time, their predictions can get better and better. In the end, artificial intelligence (AI) improves productivity, accuracy, and decision-making across a range of areas by automating complicated decision-making processes and offering quick, dependable, and scalable solutions that would be challenging or time-consuming for humans to accomplish manually.

## 1.2 Introduction to Prediction system:

Artificial intelligence refers to the ability of technology, especially computer systems, to replicate human intelligence processes. Examples of AI applications include machine vision, speech recognition, natural language processing, and expert systems. AI systems typically process vast amounts of labeled training data, identifying patterns and relationships to predict future outcomes. Just as a chatbot trained on text conversations can engage in realistic dialogues with users, an image recognition algorithm can learn to identify and describe objects in images by analyzing millions of examples. (teams, 2024)

The Concrete Strength Prediction System forecasts the compressive strength of concrete based on a number of variables, including mix proportions, material qualities, ambient conditions, and curing methodologies, using sophisticated machine learning algorithms. The strength of concrete, a basic building material, is essential to guaranteeing the security, longevity, and economy of structures. Optimizing material utilization, adhering to safety regulations, and cutting project expenses all depend on accurate concrete strength prediction.

Concrete strength is traditionally determined by time-consuming, expensive, and labor-intensive experimental testing. By automating the forecast process and enhancing decision-making in building projects, machine learning algorithms provide a more dependable and effective substitute.

Large volumes of historical data from concrete mix designs and their related compressive strengths can be processed by machine learning algorithms, which can then identify complex patterns and correlations between input factors and the concrete's final strength. The system can produce extremely accurate predictions based on new input parameters by training these models on sizable datasets, offering important insights for quality control and concrete mix design. Additionally, as more data is gathered, machine learning models can adjust and get better over time, continuously improving their predictions and helping to create more effective concrete design. (nature, 2025)

The construction sector can profit from quicker decision-making, less material waste, and optimal material costs by putting our Concrete Strength Prediction System into practice. Early in the design process, engineers and builders can obtain accurate strength estimates, which enables them to modify concrete mixes to satisfy particular performance requirements, resulting in safer and more affordable constructions.Recent studies have demonstrated the effectiveness of machine learning in predicting concrete strength. For instance, a study evaluated the performance of various machine learning models, including regression methods such as Linear, Ridge, and LASSO, for concrete strength prediction. (nature, 2025)

This system models the relationship between input factors and concrete's compressive strength using five machine learning algorithms:

* A straightforward and understandable model known as linear regression shows a linear relationship between the goal variable (concrete strength) and input variables (such as cement content and water-to-cement ratio). It acts as a reference model for assessing how well more intricate algorithms function.
* By reducing part of the coefficients to zero, Lasso Regression, a linear regression extension that uses L1 regularization, aids in feature selection. Preventing overfitting and managing high-dimensional data are two benefits of this method.
* Ridge Regression: Like Lasso, Ridge regression penalizes large coefficients using L2 regularization, which helps stabilize the model, especially when features are multicollinear.
* A non-linear model that divides the data into subsets according to feature values is called a decision tree regression. It is more adaptable than linear models because it can capture intricate correlations between characteristics and concrete strength.
* A potent ensemble model that combines the advantages of random forests and XGBoost is called XGBRFRegressor (XGBoost Random Forest Regressor). It handles intricate non-linear correlations in the data by constructing numerous decision trees and using their outputs to provide predictions that are more reliable and accurate. (nature, 2025)

### 1.2.1 Motivation:

The shortcomings of traditional testing techniques are the driving force behind the creation of a Concrete Strength Prediction System. In order to conduct experimental testing, samples must be made, cured under carefully monitored circumstances, and then tested for compressive strength—a procedure that may take up to 28 days. Due to the requirement for specialized tools and supplies, this delay not only lengthens project deadlines but also raises expenses. Furthermore, the performance of concrete under various circumstances or in large-scale applications may not always be predicted by these conventional techniques. By automating the prediction process, machine learning algorithms present a viable solution that is quicker, more dependable, and less expensive than traditional testing.

### 1.2.2 Aims :

To develop a precise, effective, and trustworthy model that can forecast the compressive strength of concrete based on a variety of input parameters, including mix proportions, material characteristics, environmental factors, and curing techniques, is the main aim of the Concrete Strength Prediction System. The system aims to improve decision-making in quality control, construction planning, and concrete mix design by utilizing machine learning approaches, which will ultimately result in safer and more economical constructions.

### 1.2.3 Objective of the system:

The objective of creating a concrete strength prediction system is to accurately estimate the compressive strength of concrete based on its mix design and environmental conditions.

This serves several key purposes:

* Optimization of Material Usage:

Predicting concrete strength helps in determining the optimal proportions of cement, water, aggregates, and additives to achieve the desired strength while minimizing material costs and waste.

* Quality Control

Ensures that the concrete produced meets the required specifications for safety and durability, reducing the risk of structural failure.

* Improved Design Efficiency

Helps engineers design structures with precise load-bearing capacities by providing accurate predictions of the concrete's performance.

* Time Savings

Eliminates the need for extensive trial-and-error testing by using data-driven methods to predict outcomes, which accelerates project timelines.

* Sustainability

Facilitates the use of alternative materials (e.g., fly ash, slag, recycled aggregates) by predicting how they affect strength, encouraging eco-friendly practices in construction.

* Risk Mitigation

Reduces the likelihood of overdesign or underdesign, ensuring that the structure is neither excessively costly nor prone to failure.

* Automation and Decision Support

Integrates predictive analytics into construction workflows, enabling better decision-making through tools like machine learning models and AI systems.

* Economic Benefits

Optimizing material usage and reducing testing costs ultimately lead to significant cost savings in construction projects.

# 2.Background:

Following the study and critique of previous projects, numerous research tasks on the selected topic were assigned for this course. The topic of Concrete strength prediction systems was chosen for this coursework to address issues A new buyer or worker may have about the quality of concrete . Maintaining and recommending a system that manages the excess of information by selecting the most important data based on the data supplied by a user and other factors that take the user's preferences and interests into account. It determines whether a user and an object are compatible, and then it makes suggestions based on the assumption that they are. Machine learning has a subclass known as prediction engines that often rank or rate people or products to predict wanted outcome. A prediction system, broadly defined, is a system that anticipates the ratings a user would give to a certain item. The user will subsequently be given a ranking of these forecasts.

Numerous well-known businesses, including Google, Instagram, Spotify, Amazon,

Reddit, Netflix, etc. frequently employ them to boost user and platform engagement. For instance, Spotify will suggest tracks that are like those you've loved or listened to a lot so that you can keep utilizing their platform to listen to music. Amazon uses recommendations to make product suggestions to different consumers based on the user data they have gathered. (Ghazanfar, 2024)

Prediction systems use statistical models and historical data to forecast or estimate future trends or events. Predictions by themselves, though, might not be enough to help you make wise choices. For example, it might not be sufficient to only purchase a stock based on a financial analyst's estimate that its value will increase. To make a wise choice, further considerations must be made, including market trends, economic indicators, and the company's financial standing. Decision assistance systems are useful in this situation. In order to give users a more complete picture of the issue and help them make wise decisions, these systems combine predictive models with extra data and analytical tools. These computer programs or algorithms, which are commonly employed in industries including banking, healthcare, sports, and weather forecasting, rely on past data and statistical models to estimate future events. (Srivastava, 2024)

## 2.1 Research on Problem scenario:

**Concrete strength prediction system addresses following problems:**

* Testing Takes Time: Conventional techniques for assessing the strength of concrete entail making samples, allowing them to cure under carefully monitored circumstances, and then testing their compressive strength—a procedure that can take up to 28 days. Due to the requirement for specialized tools and supplies, this delay not only lengthens project deadlines but also raises expenses.
* Resource Intensiveness: Materials, personnel, and equipment are only a few of the resources needed for experimental testing, which raises project costs and may cause delays.
* Limited Predictive Capability: Conventional techniques would not always be able to forecast how concrete will behave in various scenarios or in large-scale applications, which could jeopardize structural integrity.
* Data Complexity: A variety of factors, including mix proportions, material qualities, curing techniques, and environmental conditions, affect the strength of concrete. The intricate relationships between these variables may be difficult for traditional approaches to account for. (Hearns, 2025)

The Concrete Strength Prediction System seeks to automate the prediction process by utilizing machine learning algorithms, offering quicker, more dependable, and more economical substitutes for conventional testing techniques. In the end, this method can lead to safer and more economical structures by improving decision-making in quality control, construction planning, and concrete mix design.

Recent research has shown how well machine learning predicts the strength of concrete. For example, a study assessed how well several machine learning models performed in predicting concrete strength, including regression techniques like Linear, Ridge, and LASSO. This suggests that engineers might optimize concrete mix designs and construction methods by using machine learning algorithms to produce precise forecasts.

The construction sector can produce safer and more resilient structures by incorporating machine learning into the concrete strength prediction process, which will also result in more economical and efficient outputs. (nature, 2025) (Hearns, 2025)

## 2.2 Description of selected Domain:

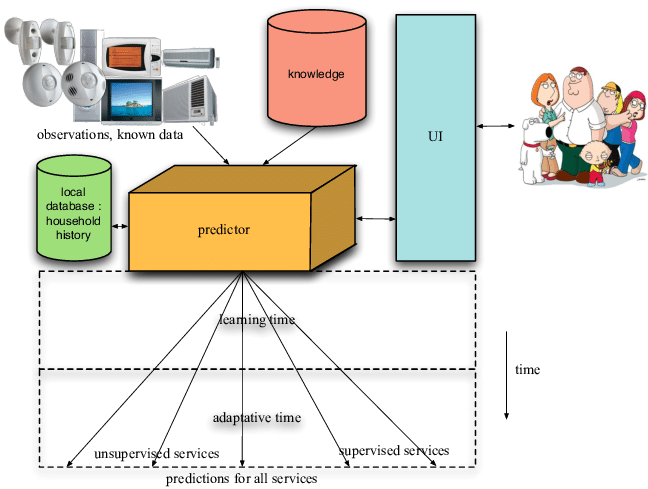


Figure : prediction system

By using sophisticated machine learning algorithms, the Concrete Strength Prediction System predicts the compressive strength of concrete based on a number of input parameters, including mix proportions, material characteristics, environmental factors, and curing techniques. Making precise predictions about the strength of concrete is crucial for maximizing material use, guaranteeing safety regulations, and cutting project expenses.

Important Elements of the Domain:

* Designing a concrete mix entails choosing and adjusting components such as cement, water, aggregates, and admixtures to provide the required results.
* Concrete strength is influenced by material properties, which include attributes like cement type, aggregate size, and additive content.
* Environmental Conditions: Elements that impact the hydration process and, in turn, the development of concrete's strength, such as temperature, humidity, and curing techniques.
* Machine Learning Algorithms: To model intricate correlations between input factors and concrete strength, methods like neural networks, decision trees, and linear regression are used.

1.linear Regression:

An method known as linear regression predicts the future course of events by establishing a linear relationship between an independent variable and a dependent variable. It is a statistical technique for predictive analysis in data science and machine learning.The predictor or explanatory variable that stays constant as a result of changes in other variables is also known as the independent variable. However, when the independent variable fluctuates, so does the dependent variable. The response or result variable under analysis or study is the dependent variable, and the regression model forecasts its value.

Thus, a supervised learning approach called linear regression models a mathematical relationship between variables and generates predictions for continuous or numerical variables like product price, age, sales, and wage. (Kanade, 2025)

2.Lasso Regression:

One of the most crucial regression analysis techniques for variable selection and regularization is the Lasso Regression, which is based on the Least Absolute Shrinkage and Selection Operator. In order to avoid overfitting, it eliminates features from the data that aren't significant, and by reducing the coefficients toward zero, it makes features with little influence easier to identify. (geeksforgeeks, 2025)

3. Ridge Regression:

Ridge regression, sometimes referred to as Tikhonov regularization, is a linear regression technique that tackles the issue of multicollinearity amongst predictor variables. High correlation between independent variables in a regression model is known as multicollinearity, and it can result in unstable and inaccurate regression coefficient estimations. (geeksforgeeks, 2025)

By penalizing big coefficients and hence reducing their variation, ridge regression addresses this problem by incorporating a regularization component into the ordinary least squares (OLS) objective function. (geeksforgeeks, 2025)

4. XGBRFRegressor:

One effective method for creating supervised regression models is XGBoost. By understanding its (XGBoost) objective function and base learners, it is possible to deduce the truth of this claim. A regularization term and a loss function are included in the objective function. It describes the discrepancy between actual and anticipated values, or how far the model's output deviates from reality. The most popular loss functions in XGBoost are reg:linear for regression tasks and reg:logistics for binary classification. XGBoost is one of the ensemble learning techniques that entails training and merging separate models, also referred to as base learners, to provide a single prediction. (geeksforgeeks, 2025)

5.DecisionTreeRegressor:

Decision Tree is a method for making decisions that employs a tree structure analogous to a flowchart or is a model of decisions and all of their potential outputs, including utility, input costs, and outcomes.

The decision-tree algorithm belongs to the class of algorithms used in supervised learning. It is applicable to both continuous and categorical output variables. (geeksforgeeks, 2025)

## 2.4 System Architecture:

## 2.5 Advantages of working around the problem domain/ algorithm implementations:

The Concrete Strength Prediction System's performance and dependability are improved by the following benefits of integrating different machine learning algorithms:

* Increased Prediction Accuracy: By employing a variety of algorithms, the system is able to identify different patterns and connections in the data, which results in predictions that are more accurate. In prediction tasks, for example, it has been demonstrated that ensemble learning approaches—which integrate several models—perform better than individual models. (medium, 2025)
* Robustness and Reliability: The advantages and disadvantages of various methods differ. The system's overall robustness can be improved by implementing numerous algorithms, which reduces the danger of depending solely on one model that might not function effectively in specific situations.
* Detailed Insights: Every algorithm may draw attention to a particular feature of the data, offering a more thorough comprehension of the variables affecting concrete strength. Better mix design and quality control decisions can be made with the help of this comprehensive analysis.
* Flexibility and Adaptability: By using a variety of methods, the system can adjust to diverse data distributions and levels of complexity, guaranteeing that it is applicable to a wide range of scenarios and datasets.
* Improved Model Validation: By comparing the results of several algorithms, it is possible to select the best model for a given prediction task and validate it more thoroughly, which results in better model selection procedures.

## 2.6 Considered drawbacks revolving the problem domain/ algorithm implementations:

There may be several difficulties when integrating various machine learning algorithms into the Concrete Strength Prediction System.

* Enhanced Complexity: Organizing and combining several algorithms can make the architecture of the system more complex, which makes it more challenging to create, maintain, and debug. (Akshaj, 2025)
* Greater Computational Resources: Using ensemble methods or many algorithms at once requires more memory and computing capacity, which could result in longer processing times and higher operating expenses.
* Data Requirements: For certain machine learning models to function well, sizable and superior datasets are necessary. Regardless of the number of algorithms utilized, erroneous predictions might result from inadequate or low-quality data. (Devshatwar, 2025)
* Risk of Overfitting: When several models are combined without adequate validation, the system may perform well on training data but badly on unknown data, which limits its generalizability. (riskaware, 2025)
* Implementation Difficulties: Setting up a system with several algorithms could take more effort and specialized knowledge, possibly delaying projects.

## 2.7 Dataset:

### 2.7.1 Dataset design:

The dataset that is used in development of this project is commonly known as Concrete Compressive Strength Dataset that is widely known resource in civil engineering and machine learning. It is frequently used dataset for predictive modes for determining strength of concrete body on the basis of its ingredients and its age.

This dataset consists of 1030 instances that each represent unique concrete sample. Each instance includes eight input variables corresponding to the quantities of concrete components and the age of the sample, and one output variable indicating the concrete's compressive strength. these variables are:

* Cement: it represents the amount of cement used
* Blast Furnace Slag: Quantity of blast furnace slag
* FlyAsh: amount of by product of coal combustion
* Water: Quantity of water
* Super plasticizer: amount of additive that enhances the workability of concrete
* Coarse Aggregate: quantity of coarse like gravel
* Fine Aggregate: quantity of fine aggregate like sand
* Age: age of concrete sample
* Concrete Compressive strength: target variable representing strength of the concrete sample

A screenshot of a computer

Description automatically generated

Concrete's compressive strength is a critical parameter in construction, influencing the durability and safety of structures. Accurately predicting this strength based on the mix proportions and age can lead to more efficient and effective construction practices. The dataset was donated to the UCI Machine Learning Repository by I-Cheng Yeh in 2007 and has since been widely used for modelling and analysis purposes. (Yeh, 2007)

**2.7.2 EDA:**

**.head()**

A screenshot of a computer

Description automatically generated

**.shape()**

A white rectangular box with black text

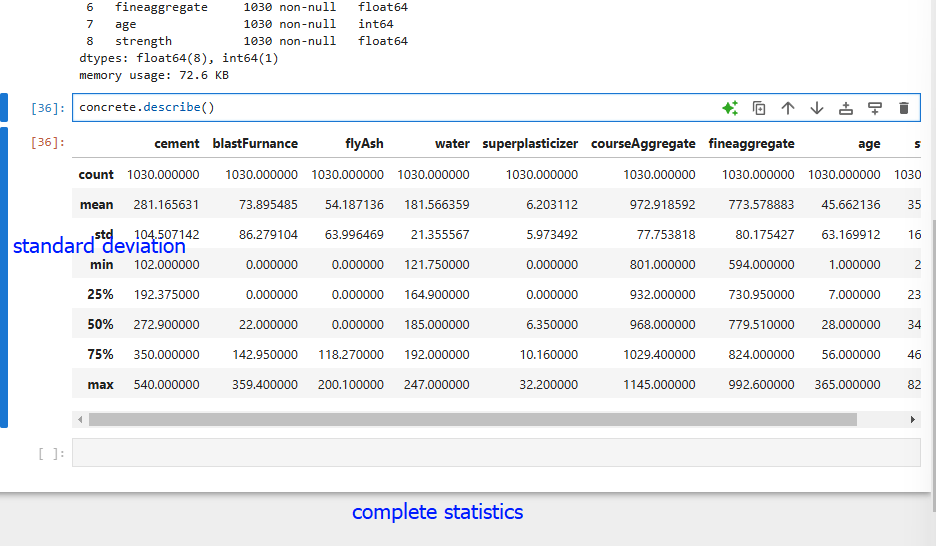
Description automatically generated

**Duplicated(), isnull()**

A screenshot of a computer

Description automatically generated

**.describe()**

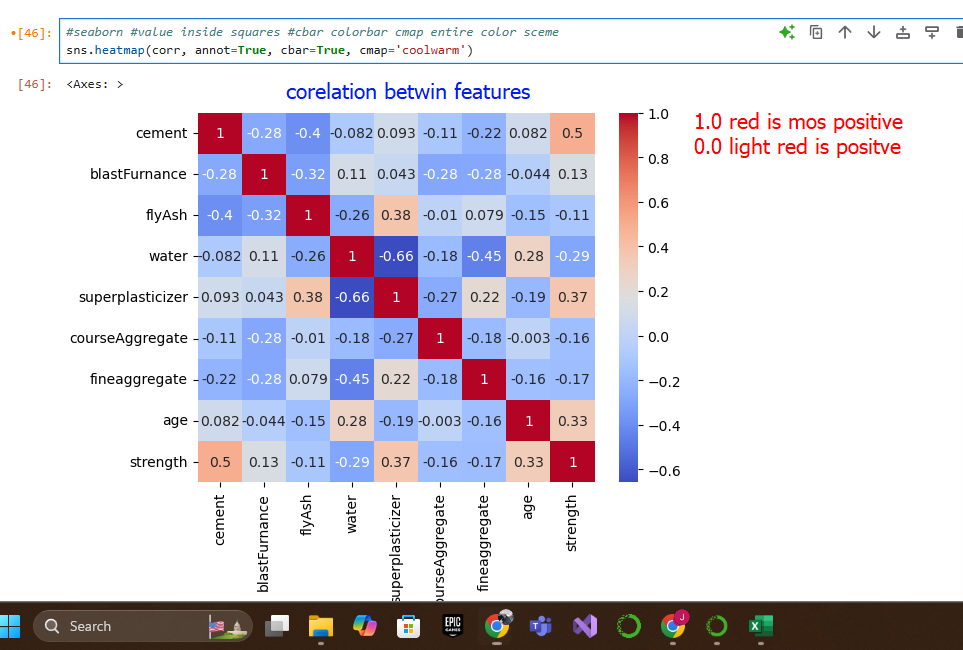


**Corelation between the features**

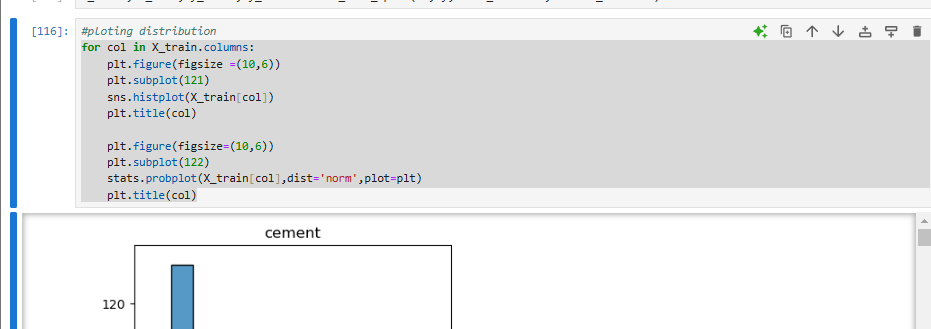
**A screenshot of a computer

Description automatically generated**

**Corelation using colormap from seaborn**

****

**Plotting distributions with histograms**

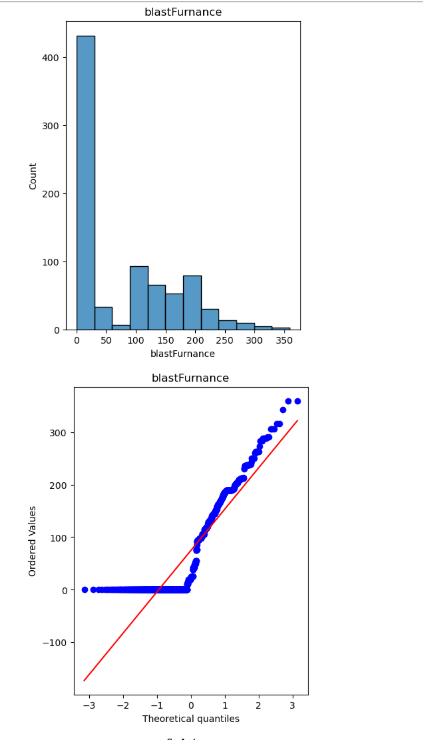


**Cement histogram**

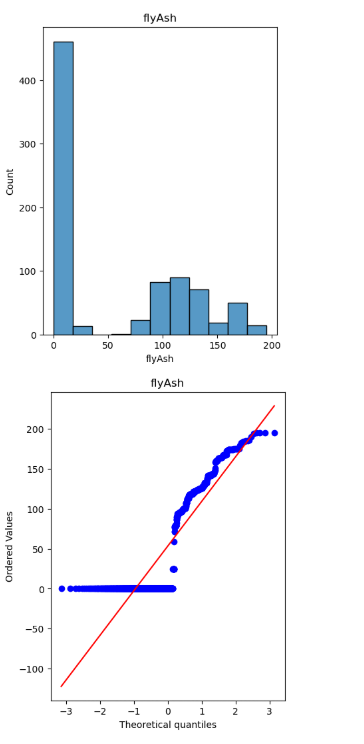
A graph and a graph

Description automatically generated

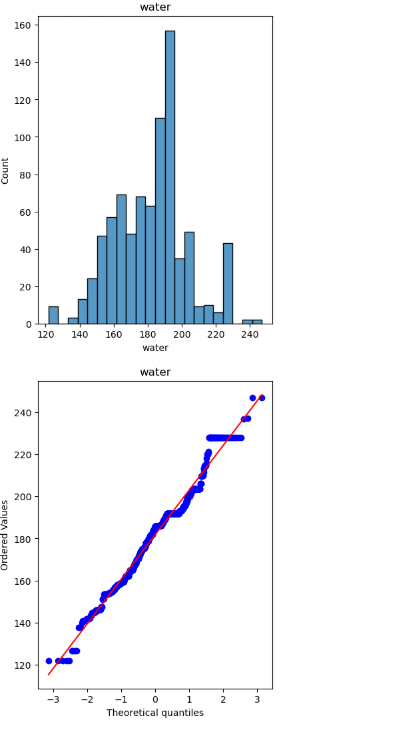
**Blast furnace histogram**

****

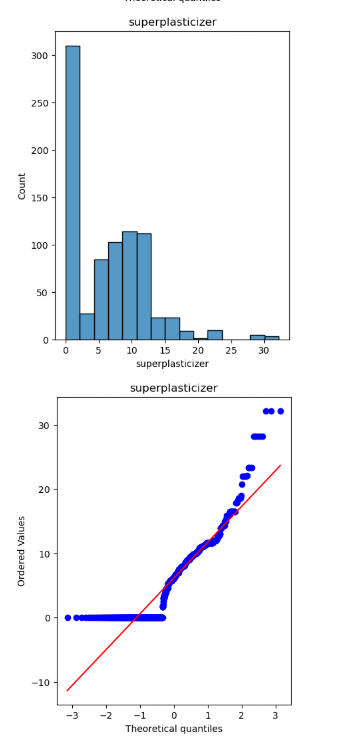
**Fly Ash Histogram**

****

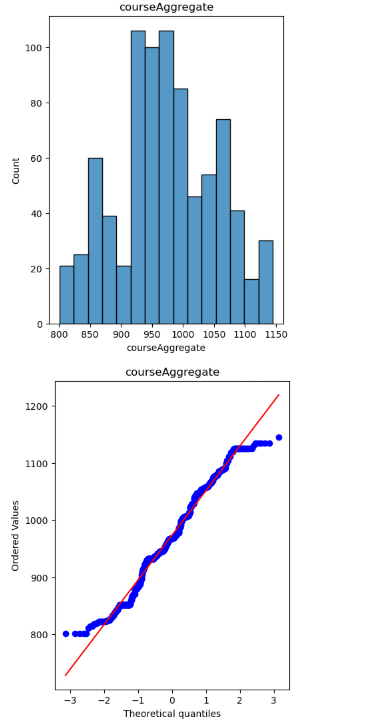
**Water**

****

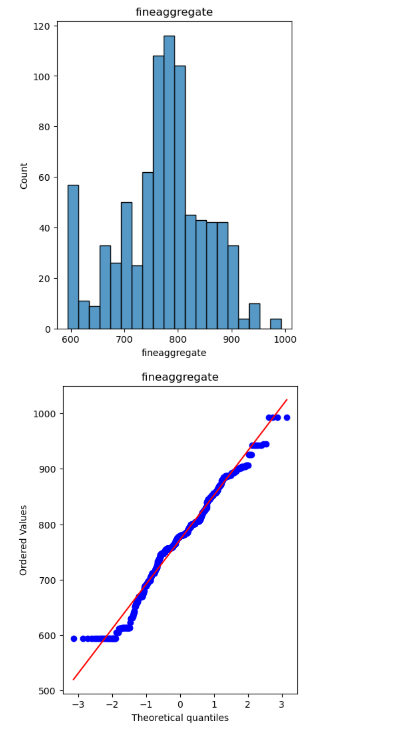
**Super Plasticizer Histogram**

****

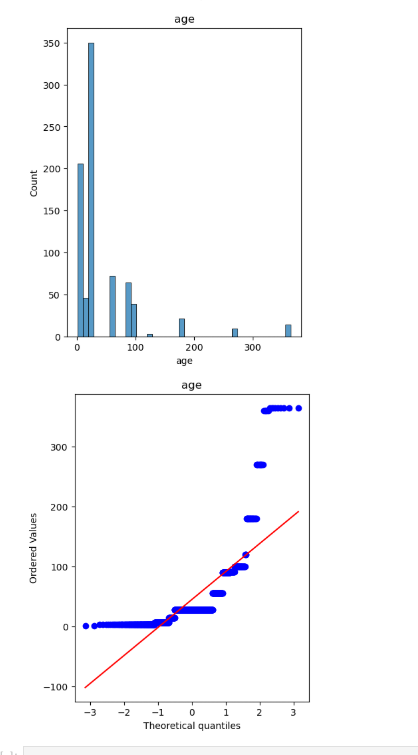
**Coarse Aggregate**

****

**FineAggregate histogram**

****

**Age histogram**

****

## 2.8 Review and analysis of similar system:

**Some of the concrete strength prediction system are explained below:**

**Article 1:**

* Concrete's compressive strength (CS) has a major impact on its structural integrity, thus it is essential to estimate CS precisely when building. However, the erratic mechanical and physical characteristics of concrete and its constituent parts make this work difficult. This study presents an ideal artificial neural network (ANN) model for CS forecasting in order to get over the drawbacks of conventional laboratory testing techniques. For this goal, 776 datasets from earlier research were gathered. Training and testing sets were randomly selected from the preprocessed dataset. By adjusting the right hyperparameters, the best ANN model was created, and the Adaptive Optimization Algorithms (Adam) optimizer was used to reduce overfitting and validation loss. With an R-squared value of 0.87 and errors in MAE, MSE, and RMSE of 3.419 MPa, 21.909 MPa, and 4.68 MPa, respectively, the ANN model's predictions demonstrated strong performance. Furthermore, the SHAP values showed that the volume of fly ash and water had the most significant negative effect on CS, whereas the volume of cement and water had the largest positive impact. The ideal ANN model is very practical in the design and construction of concrete-based infrastructure since this work demonstrates the potential of machine learning approaches for the prompt and effective prediction of concrete compressive strength**.** (Thapa, 2024)

**Article 2:**

* Because of its advantageous engineering qualities, concrete is frequently employed in architecture and construction. It has a high compressive strength, is reasonably priced, and is constructed from a variety of raw materials. Coarse aggregate, fine aggregate, cement, and water are the four primary ingredients of concrete. It is a common choice in building due to its economic benefits and ease of availability in local markets. Concrete can be made with little effort, which is another important advantage over other materials like steel. To improve its mechanical qualities, additional elements including fly ash (PFA), blast furnace slag (GGBS), silica fume, and other industrial byproducts are occasionally added. In addition to increasing the strength of concrete, adding these waste elements has positive environmental effects and increases the durability and longevity of concrete constructions. Compressive strength is the most important of the concrete's qualities since it directly impacts structural safety and is used to evaluate a structure's performance from design to evaluation. Due to cost and local material availability, choosing the appropriate components and precisely forecasting concrete's mechanical properties—particularly its compressive strength—can be difficult. The danger of non-compliant concrete during construction can be decreased and experiment costs can be decreased by early development of trustworthy predictive models using available input-output data. Time and money can be saved by identifying the best material combinations with the use of appropriate models. Despite their widespread use, empirical and statistical models such as linear and nonlinear regression can necessitate substantial experimental labor.(Vimal Rathakrishnan, 2024)

**Article 3:**

* Because of its high-strength, stable properties, concrete has emerged as the most popular building material in recent years. Nowadays, a variety of other cementitious materials, including fly ash, blast furnace slag, and chemical additives like superplasticizers, are frequently utilized in addition to the four fundamental components of cement, coarse aggregate, fine aggregate, and water. Since Portland cement, the main ingredient in concrete, is the most costly, these other ingredients not only enhance the concrete's qualities but also have financial advantages. Concrete's workability, strength, and durability are all improved by additives, but predicting the compressive strength becomes more difficult as a result. Because of this increasing complexity, traditional modeling tools frequently fail to generate reliable forecasts.

Although strength tests are usually carried out 28 days following the casting of concrete, this waiting period may cause the next stages of construction to be delayed. However, skipping testing would restrict the ability to maintain quality control in intricate, large-scale construction projects. Therefore, even in the early stages of design, it is essential to anticipate concrete's compressive strength promptly and accurately in order to maintain quality control. When the concrete strength does not match standards, it can save a lot of time and money if the mixture proportions can be changed quickly. Determining the best time to remove formwork, planning projects, and maintaining quality control all depend on early concrete strength prediction.

Mathematical modeling becomes complicated due to the nonlinear interaction between concrete components and attributes. The empirical equations used in the current standard codes for determining compressive strength were created for concrete without additional cementitious ingredients. Optimizing the concrete mix requires an understanding of the connection between the mix and strength. Even though a lot of research has been done via experimental tests, these approaches are expensive and time-consuming. As a result, a novel modeling technique that can accurately forecast concrete's compressive strength without the need for experimental is required.

Since they are more accurate than conventional approaches, artificial intelligence (AI) techniques are being used more and more to solve classification and regression problems. Based on a number of factors, this study investigates the application of AI algorithms to precisely forecast the compressive strength of concrete. Yeh gathered experimental data from the University of California, Irvine's machine learning library and utilized it to forecast the compressive strength of HPC. Three prediction methods—artificial neural networks (ANN), support vector machines (SVM), and linear regression (LR)—were used in Clementine to create the AI models. Both single and ensemble model constructs are used in the study. By utilizing the advantages of each classifier, integrating several classifiers in ensemble models helps to offset individual errors and enhance overall performance. (af, 2020)

# 3.Solution:

## 3.1 Approach to solving problems:

The solution for the concrete strength prediction system is to write the code for a program having different python library like NumPy and packages like pandas, Scikit-learn etc. along with the flowchart and pseudocode of the program. To write the program and the code, python programming language will be used.

**Data cleaning:**

The data will be cleaned by renaming some of the feature columns to make them more readable, and by checking for and dropping any duplicate rows.

**Data preprocessing:**

The data set will be preprocessed by performing several steps, such as handling missing values, scaling the data, and splitting the data into training and testing sets.

**Model selection:**

Several models were trained and tested on the data set, including Linear Regression, Lasso Regression, Ridge Regression, Support Vector Regression, Decision Tree Regression, Random Forest Regression, and XGBoost Regression. The model with the best performance was selected based on various evaluation metrics, such as mean squared error (MSE), mean absolute error (MAE), and R-squared score.

**Model evaluation:**

The selected model was evaluated using the test data set, and its performance was analyzed using various evaluation metrics. The results were visualized using various plots, such as scatter plots and residual plots.

**Python:**

Python is a powerful and versatile programming language that plays a critical role in a wide variety of technological solutions. From web applications, search engines, and games to animation software and even other programming languages, Python is at the heart of innovation. In recent years, Python has seen a surge in popularity, becoming one of the most widely used programming languages across the globe. Its applications are expanding into new and exciting areas, such as artificial intelligence, machine learning, and data science. (Worsley, 2024)

Python does not require compilation into machine code prior to running because it is an interpreted language. Alternatively, the programmer can utilize it directly to execute the program. Because of this characteristic, Python may be effectively used with an emulator or virtual machine that uses native code that the hardware can understand. Python is frequently utilized in intricate situations and is regarded as a high-level programming language. The capacity to handle arrays, variables, objects, complicated arithmetic, Boolean statements, and other abstract notions in computer science is what makes high-level languages like Python more adaptable and useful. (javatpoint, 2024)

## 3.2 Algorithm and Functions:

* **Linear regression**
* By fitting a linear equation to the observed data, linear regression seeks to demonstrate the link between two variables. The two variables are regarded as independent and dependent, respectively. For example, a person's weight and height can be linearly correlated, indicating that weight tends to rise in tandem with height.

A linear regression line's equation is expressed as follows: Y = a + bX

In this formula:

The independent variable, represented by the x-axis, is X.

Plotted on the y-axis, Y is the dependent variable, and b is the line's slope, which shows how quickly Y varies as X changes.

When X equals 0, the value of Y is represented by the intercept, a.

(Botolan campus, 2024)

* **Lasso Regression:**

Lasso (Least Absolute Shrinkage and Selection Operator) is a popular machine learning tool for automatically choosing pertinent characteristics from high-dimensional data. This is accomplished by adding a penalty term and scaling the residual sum of squares (RSS) with a regularization value (lambda or λ). The amount of regularization that is used is controlled by this option. Automatic feature selection results from higher lambda values since they increase the penalty and cause more coefficients to shrink towards zero, which lessens the significance of some features or removes them from the model. Lower lambda values, on the other hand, lessen the cost and preserve more features in the model. (ibm, 2024)

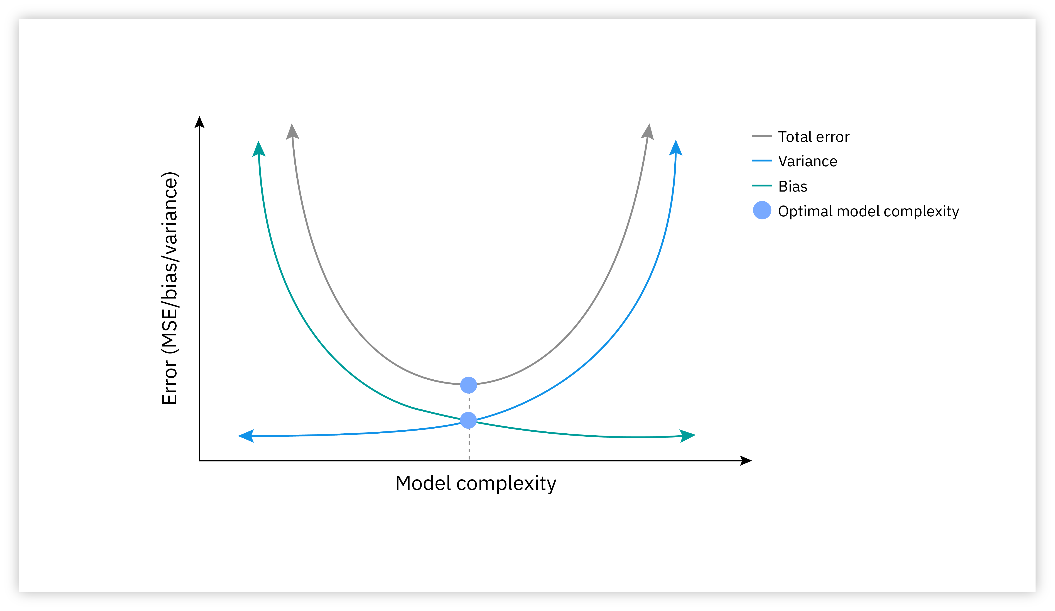


Figure : lasso regression

* **Ridge Regression**

The feature weights in ridge regression never go to zero, even if they could get extremely small. Multicollinearity can be addressed by feature selection, which is the act of reducing a coefficient to zero, which effectively eliminates the associated predictor from the model. Ridge regression does not carry out feature selection, which is frequently regarded as a drawback of this technique, because it does not lower coefficients to zero. Furthermore, when faced with extreme multicollinearity, ridge regression finds it difficult to distinguish between the impacts of predictors. In comparison, another type of regularization in linear regression is called lasso regression, or L1 regularization. By reducing some of the coefficients to zero, L1 regularization eliminates those independent variables from the model. Ridge regression lessens the impact of each independent variable, whereas lasso regression completely removes some of them. Both techniques contribute to the reduction of model complexity. (ibm, 2024)

**A graph showing the amount of blue dots

Description automatically generated**

Figure : ridge regression

* **Support Vector Regression**

One machine learning technique for regression applications is Support Vector Regression (SVR). It is a Support Vector Machine (SVM) variation that is especially made to predict continuous numerical values, which makes it perfect for jobs like stock price prediction and time series forecasting. SVR seeks to balance data fitting with overfitting prevention by developing a regression model that keeps a buffer around the projected values. By choosing the right kernel functions, this method may be tailored for various problem domains and is particularly useful for managing non-linear interactions. (Verma, 2024)

A graph of support vector regression

Description automatically generated

Figure :support vector regression

* **Decision Tress Regression**

Decision Tree Regressors are a powerful, interpretable, and non-linear method used widely for regression tasks in machine learning. Unlike linear regression, decision trees partition the feature space in a hierarchical, rule-based way that enables them to capture complex, non-linear relationships. In this post, we’ll delve deeply into the intuition, structure, and mechanics of decision trees for regression, exploring how they work, why they’re useful, and what makes them a versatile choice for various prediction tasks.

(K, 2024)

A diagram of a decision tree

Description automatically generated

Figure : Decision tress regression

* **Mean Squared Error**

Mean Squared Error (MSE) is a loss function that is used in regression tasks. These are tasks where an example can be predicted as a continuous value, and the model must determine this value. Mean Squared Error is widely used in many applications of machine learning, including:

1.Regression tasks, where a continuous value needs to be predicted based on input data.

2.Time series forecasting tasks, where future values need to be predicted based on past and present data.

3.Any other type of machine learning task that involves predicting a continuous value. (Sorokin, 2024)

* **Mean absolute Error**

A simple yet useful indicator for evaluating the precision of regression models is Mean Absolute Error (MAE). The average absolute difference between the goal values and the forecasted values is computed. In contrast to other metrics, MAE treats all errors equally, whether they are positive or negative, because it does not square the errors. Because of this feature, MAE is particularly useful for determining the magnitude of mistakes without concentrating on whether they are overestimations or underestimations. (Ahmed, 2024)

A black background with white text

Description automatically generated

Figure : formula of mean absolute error

* **R-squared Error**

The coefficient of determination, or R-squared, shows what proportion of the variance in the dependent variable can be accounted for by the independent variables. Stated differently, it demonstrates how well the regression model fits the data.(medium, 2024)



Mean Squared Error (MSE) calculates the average of the squared differences between the actual and predicted values. It provides an indication of the size of the errors in the model's predictions. (medium, 2024)

A black background with a black square

Description automatically generated with medium confidence

MSE shows the magnitude of the prediction errors, whereas R-squared shows how well the model takes into account the variability of the target variable. When combined, these measures provide a comprehensive evaluation of the model's effectiveness. (T, 2024)

## 3.3 Pseudocode:

The pseudocode for the proposed solution is as follows:

START

IMPORT libraries

IMPORT dataset

READ dataset

IF repeated names

RENAME repeated column names

ELSE

END IF

PROCESS dataset

SPLIT the dataset

PLOT the distribution without transformation

POWERTRANSFORMER to transform data in symmetric

PLOT the distribution with transformation

SCALE the data

TRAIN the model

CALCULATE prediction from the data

END

## 3.4 Algorithm and flowchart:

Algorithm:

Step 1: START

Step 2: IMPORT libraries

Step 3: IMPORT dataset

Step 4: READ dataset

Step 6: IF repeated names

Step 7: IF True - RENAME repeated column names

ELSE - Continue to the next step 8

Step 8: PROCESS dataset

Step 9: SPLIT the dataset

Step 10: PLOT the distribution without transformation

Step 11: POWERTRANSFORMER to transform data in symmetric

Step 12: PLOT the distribution

Step 13: SCALE the data

Step 14: TRAIN the model

Step 15: CALCULATE prediction

Step 16: END

## 3.5 Flowchart:

**A diagram of a process flow

Description automatically generated**

Figure : flowchart diagram

## 3.6 Development Procedure:

* Importing

A screenshot of a computer

Description automatically generated

* Renaming features



* EDA

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* Train test split

A screenshot of a computer

Description automatically generated

* PowerTransformer

A white rectangular object with text

Description automatically generated

* EDA after power transformer
  + cement

A comparison of a graph

Description automatically generated

* + blast furnace

A graph of a function

Description automatically generated with medium confidence

* + fly ash

A graph of a function

Description automatically generated with medium confidence

* + water

A comparison of a graph

Description automatically generated

* + super plasticizer

A graph of a function

Description automatically generated with medium confidence

* + course aggregate

A graph of a function

Description automatically generated with medium confidence

* + fine aggregate

A graph of a number of data

Description automatically generated with medium confidence

* + Age

A screenshot of a graph

Description automatically generated

* Scaling data to prioritize equally

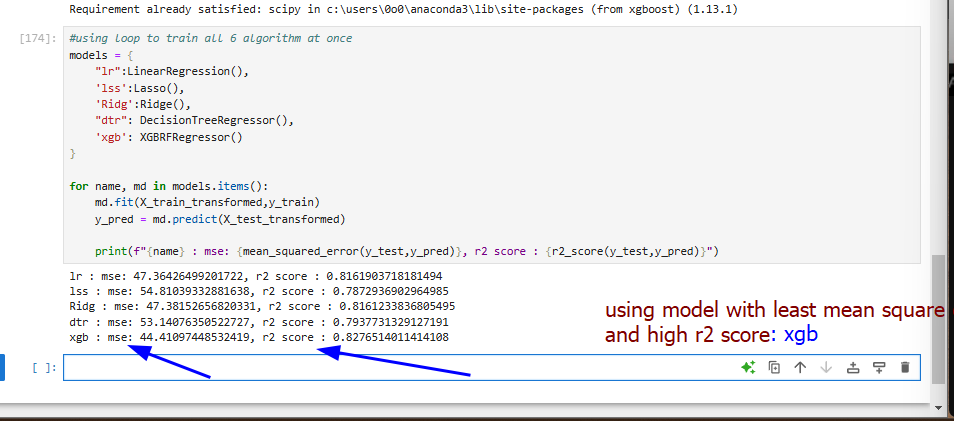
A screenshot of a chat

Description automatically generated

* Training all algorithms mean squared error

A screenshot of a computer program

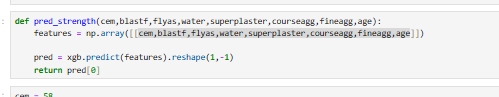
Description automatically generated



* We take algorithm with least mean squared error and most R2 score i.e. xgboost



* Defining a function to predict on the basis of input features



* Inputing the sample inputs to test to workflow

A screenshot of a computer

Description automatically generated

* Saving our trained module using pickle package for application backend

A white rectangular object with red and yellow text

Description automatically generated

* Coding front end for application

A screen shot of a computer

Description automatically generated

A computer screen shot of a person's face

Description automatically generated

A screen shot of a computer program

Description automatically generated

A computer screen shot of a program code

Description automatically generated

* Loading module and implementing backend

A computer screen with white and green text

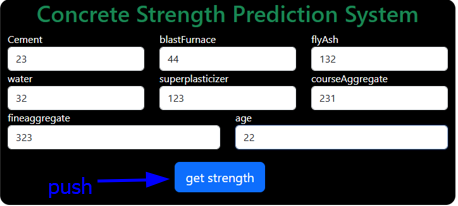
Description automatically generated

* Creating prediction function for UI

A computer screen shot of a program code

Description automatically generated

* Running css
  + InputA screenshot of a computer

    Description automatically generated
  + Push button

A screenshot of a computer

Description automatically generated

## 3.8 Language used:

Python

Css

# 4.Conclusion:

The proposed system for concrete strength prediction utilizes advanced data preprocessing and machine learning techniques to provide accurate and reliable predictions. By implementing a systematic approach that includes data cleaning, transformation, scaling, and predictive modelling, the system ensures that the input data is well-prepared for robust analysis.

## 4.1 Analysis of the work done:

This coursework is mostly related with the research work that carried out on the AI topic. In depth research about the chosen topic were necessary according to the criteria of the assignments and those criteria were fulfilled. Among research domain concrete strength prediction system was selected.

One of the many things this project will be finished on I a system that can predict whether the given concrete is reliable or not i.e strength prediction system. This system aids those people with little experience and knowledge of concretes and also those who are learning about the topic, this project will also be helpful for people who are uncertain about their house or any other structure with concrete’s maintenance

Numerous challenges and misunderstandings arose while completing this assignment, particularly when studying the AI topic and conducting the review and analysis sections. Countless difficulties were encountered while doing the research. Numerous studies on Python programming have been conducted to solve all the problems, clear up the confusion, and finish all the jobs. The slides were also quite helpful. It was essential to contact with the teachers frequently.

## 4.2 How the solution addresses real world problems:

* Efficient Handling of Raw Data: The system is designed to handle raw datasets by identifying and resolving inconsistencies, such as repeated column names, ensuring data integrity.
* Enhanced Data Representation: Transformation techniques like the PowerTransformer improve the symmetry of data distribution, reducing skewness and ensuring better model performance.
* Optimized Model Performance: Through proper scaling and preprocessing, the predictive model can achieve higher accuracy and generalizability, making it suitable for diverse concrete compositions.
* User-Friendly Insights: Visual plots of data distribution before and after transformation help users better understand the impact of preprocessing.
* Practical Applicability: The system is designed to assist engineers and researchers in predicting concrete strength quickly and efficiently, reducing the need for extensive laboratory testing.

## 4.3 Further Work:

The development of a working prediction system that can foretell events based on previous data is another goal of the project's research and suggested solution. The system may provide users with an intuitive grasp of the forecasted data by efficiently visualizing the expected outcomes through understandable and instructive plots.

# References

af, H. M. e., 2020. *Purpose LED Publishing.* [Online]   
Available at: https://iopscience.iop.org/article/10.1088/1742-6596/1625/1/012018/pdf  
[Accessed 2024].

Ahmed, M. W., 2024. *medium.* [Online]   
Available at: https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df  
[Accessed 25 december 2024].

Botolan campus, 2024. *coursehero.* [Online]   
Available at: https://www.coursehero.com/file/159948189/Linear-Regressionpptx/  
[Accessed 25 decmber 2024].

Crabtree, M., 2024. *datacamp.* [Online]   
Available at: https://www.datacamp.com/blog/what-is-ai-quick-start-guide-for-beginners?utm\_source=google&utm\_medium=paid\_search&utm\_campaignid=19589720824&utm\_adgroupid=152984013054&utm\_device=c&utm\_keyword=&utm\_matchtype=&utm\_network=g&utm\_adpostion=&utm\_creative=7248

Ghazanfar, Z., 2024. *Linkedin.* [Online]   
Available at: https://www.linkedin.com/pulse/what-prediction-systems-how-enhance-them-make-better-zoya-ghazanfar

ibm, 2024. *ibm.* [Online]   
Available at: https://www.ibm.com/think/topics/lasso-regression?utm\_id=  
[Accessed 25 december 2024].

ibm, 2024. *ibm.* [Online]   
Available at: https://www.ibm.com/think/topics/ridge-regression?fbclid=IwZXh0bgNhZW0CMTAAAR26Vst8pNTbyerR0IrsjD9bbGG4\_w3ylcu  
[Accessed 25 december 2024].

javatpoint, 2024. *javatpoint.* [Online]   
Available at: https://www.javatpoint.com/role-of-python-in-artificial-intelligence

K, F., 2024. *Medium.* [Online]   
Available at: https://farshadabdulazeez.medium.com/understanding-decision-tree-regressor-an-in-depth-intuition-a1d3af182efd#:~:text=Introduction,complex%2C%20non%2Dlinear%20relationships.

maths, b., 2024. *byjus.* [Online]   
Available at: https://byjus.com/maths/linear-regression/

medium, 2024. *medium.* [Online]   
Available at: https://medium.com/@priyanka.talla09/understanding-r-squared-and-mean-squared-error-in-regression-tree-model-performance-494ffbf5b5ac  
[Accessed 25 december 2024].

Sorokin, M., 2024. *Medium.* [Online]   
Available at: https://medium.com/@vergotten/mean-squared-error-an-in-depth-analysis-and-application-in-regression-tasks-237735c52245

Srivastava, N., 2024. *E2ecloud.* [Online]   
Available at: https://www.e2enetworks.com/blog/from-prediction-to-decision-support-systems  
[Accessed 25 december 2024].

teams, I., 2024. *IBM.* [Online]   
Available at: https://www.ibm.com/think/topics/lasso-regression#:~:text=Lasso%20regression%E2%80%94also%20known%20as,W%20)%20%2B%20%7C%7Cw%7C%7C1

teams, N., 2024. *NVIDIA.* [Online]   
Available at: https://www.nvidia.com/en-us/glossary/recommendation-system/

Thapa, J., 2024. *nep jol.* [Online]   
Available at: https://www.nepjol.info/index.php/joeis/article/view/65288

Thapa, J., 2024. *nepjol info.* [Online]   
Available at: https://www.nepjol.info/index.php/joeis/article/view/65288

T, P., 2024. *Medium.* [Online]   
Available at: https://medium.com/@priyanka.talla09/understanding-r-squared-and-mean-squared-error-in-regression-tree-model-performance-494ffbf5b5ac

Verma, N., 2024. *Medium.* [Online]   
Available at: https://medium.com/@nandiniverma78988/an-introduction-to-support-vector-regression-svr-in-machine-learning-681d541a829a

Vimal Rathakrishnan, s. b. a. n. A., 2024. *Nature.* [Online]   
Available at: https://www.nature.com/articles/s41598-022-12890-2

Worsley, S., 2024. *datacamp.* [Online]   
Available at: https://www.datacamp.com/blog/all-about-python-the-most-versatile-programming-language?utm\_source=google&utm\_medium=paid\_search&utm\_campaignid=19589720824&utm\_adgroupid=152984014214&utm\_device=c&utm\_keyword=&utm\_matchtype=&utm\_network=g&utm\_adpostion=&utm