DATA CURATION SYSTEM AND PREDICTIVE MODEL FOR UNCONFINED CONCRETE STRENGTH

Prateek Kakkar, Ramgopal Reddy Putta, Xuanzhi Li, Mentored by Dr. Ariana Mendible
Team MSDS 23.3, MSDS, Seattle University

DESIMONE

SPONSOR

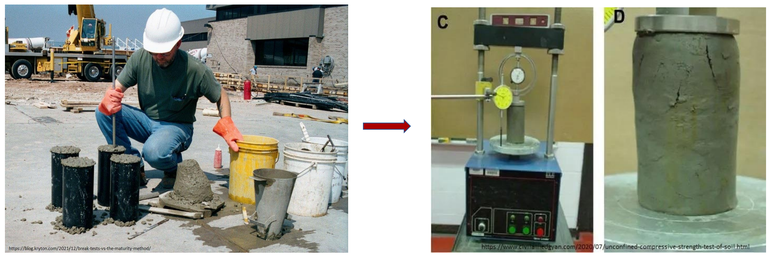
DeSimone Consulting Engineering
Liaisons: Jeff Dragovich, PHD, PE, SE, F.ACI
Matthew Cummins, P.E., S.E.

ABSTRACT

This project aims to enhance data management in the construction industry through a comprehensive solution that generates clean data and a machine learning model to predict the concrete strength and identify key factors affecting it. The solution extracts data from PDF files, transforms it into a structured format, and develops a machine learning model to predict concrete strength. Data visualization is also used to gain insights. This project enables DeSimone Consulting Engineering (DCE) to make better business decisions and improve overall efficiency by gaining better insights into concrete strength and its influential variables.

INTRODUCTION

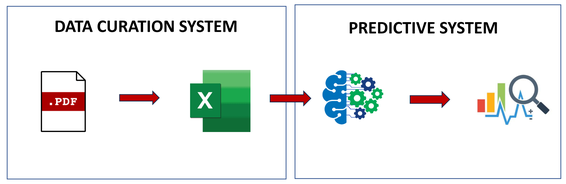
Concrete is a vital material used in many structures. To guarantee the safety and longevity of these structures, it is important to measure the strength of the concrete being used. *Unconfined Concrete Strength* (UCS) is the highest compressive stress obtained by a standard test specimen. To test this, wet concrete samples are collected from construction sites, molded into a cylinders, and subjected to strength testing. UCS is critical for engineers to design safe structures and assess concrete quality and strength over time.



DCE is a multinational engineering firm that receives UCS test reports from independent testing agencies. In addition to UCS, the test reports include essential data points that impact it, such as Required strength, Site and Concrete Temperature, Specimen Age during testing, Concrete Mix Design etc. These factors, along with site conditions, and curing process, significantly affect the concrete's strength, durability, and overall performance. The reports are presented in various PDF formats and stored on the DCE network. To streamline the data extraction process and gain valuable insights into concrete strength, DCE tasked us with building a data curation system and developing a predictive model for UCS.

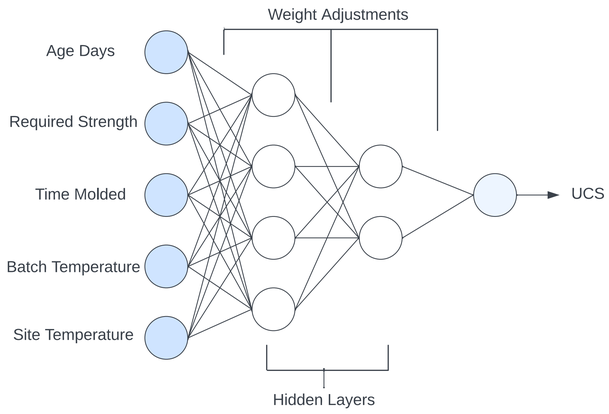
METHODOLOGY

We developed a data curation system in Python to extract and analyze unconfined concrete strength test data stored in PDF files. By utilizing text scraping techniques and Python libraries, our system offers a streamlined solution for extracting and structuring data from concrete compression test lab reports. This results in the creation of clean and structured Excel files, saving time and resources. The tool enables analysts and engineers to efficiently process large amounts of data and focus on generating insights rather than data management.



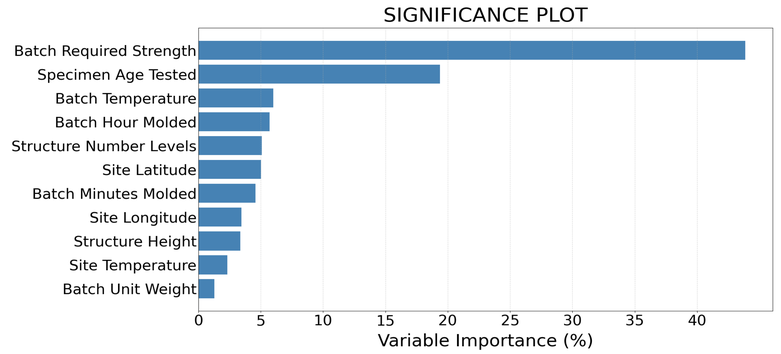
We developed a system using a Neural Network to predict the unconfined concrete strength over various time periods (Specimen Curing Days) and determine the significant variables that impact the strength. To gain valuable insights from the model and data, we also utilized Python-based data visualization techniques.

ARTIFICAL NEURAL NETWORK

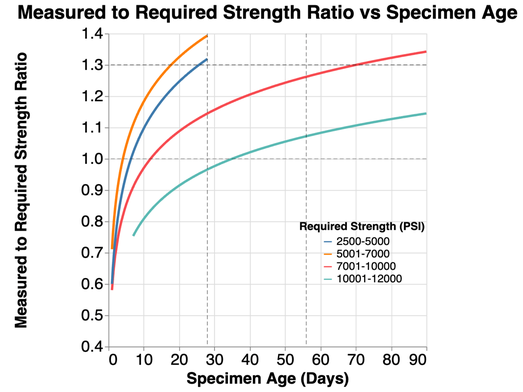


Artificial Neural Network (ANN) is a machine learning model inspired by the human brain. It is made up of layers of interconnected neurons. The first layer takes in the input data, and the last layer produces the output. The intermediate layers, known as hidden layers, perform computations on the data as it passes through the network. Each neuron has a weight that decides how important it is. When we train the ANN, it adjusts the weights to make better predictions. Backpropagation is the process of calculating how much each weight contributed to the difference between the network's output and the expected output, and then adjusting them accordingly for improvement.

RESULTS



Our machine learning model found that the Required Strength and Age at Testing are the most important factors in predicting the UCS. This led us to investigate the relationships between these variables. The following plot shows one of the relationships.



The vertical lines mark the specimen age, typically on the 28th and 56th day. The horizontal lines indicate the required strength (y=1) and the industry's 1.3 times requirement (y=1.3). All specimens surpass the required strength within 56 days. Specimens with lower required strength meet the industry requirement within 28 days, while specimens with higher required strength take longer.

CONCLUSION

We conducted a project to study the strength of concrete. We established a structured way to manage data and extracted important information from PDF files. We used a neural network model and data visualization techniques to identify the factors that affect concrete strength. Our findings can help improve the composition of concrete for different uses.

ACKNOWLEDGEMENT

We thank the Seattle University Project Centre, Jeff Dragovich, Matthew Cummins from DCE, and our faculty advisor, Dr. Ariana Mendible, for their invaluable support throughout our project.