2021

High Level Design (HLD)

CEMENT CONCRETE COMPRESSIVE STRENGTH PREDICTOR

VISHVESH SHIROLKAR

Contents

Abstract

Now-a-days we see that infrastructure is growing at the maximum rate. Development in recent building trends is happening rapidly. This needs a firm and long lasting concrete material. Cement concrete is at the core of any construction. It's compressive strength is a big concern. If we can create a program and predict the compressive strength of a concrete mixture before using it to actual construction site, it will be a big achievement towards building high strength and massive infrastructure with minimal chances of failure. This work discusses the solution to how to predict a concrete mixture's strength from its components' proportions.

1 Introduction

1.1 Why High-Level Design document?

The purpose of the HLD is to help learners/readers understand the project with detailed knowledge. Also, it is useful to understand the background and outline of a project. HLD is intended to describe following points:

- General description of project
- Design details
- KPI's
- Conclusion

1.2 Scope

HLD document shows structure of a project like process flow, architecture, technology implemented etc in simple terms which are learner understandable.

1.3 Definitions

TERM	DESCRIPTION
FEATURES	ATTRIBUTES OF CEMENT MIXTURE
NON-NULL VALUES	WHOLE NUMBERS/FLOAT VALUES
MODEL	A MACHINE LEARNING ALGORITHM
R2 SCORE	R SQUARED VALUE

2 Description

2.1 Product perspective

The Cement Concrete compressive strength predictor is a machine-learning based compressive strength prediction model which helps to predict strength of the cement mixture from its components' proportions.

2.2 Problem statement

The quality of concrete is determined by its compressive strength, which is measured using a conventional crushing test on a concrete cylinder. The strength of a concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength which is a long period. So, to save a lot of time, we can use data science and machine learning tools to predict the strength well ahead of time.

2.3 Proposed solution

The solution proposed here, the project, is we can use a machine learning algorithm to train the dataset on various models and from the multiple models, we can select the best fit or high scoring model to predict the strength.

2.4 Further improvements

The models that are used in this project are within the scope of machine learning and not beyond that. So, algorithms like deep learning are out of scope. If one uses them, the model will improve automatically based on certain conditions.

2.5 Technical requirements

This topic discusses various requirements that a machine learning project should have technically.

- The project should discuss multiple ML models or algorithms
- Features of datasets should be studied thoroughly for importance.
- It should take into account the accuracy and goodness of fit for each model
- Training and testing/validation datasets should be separately used.
- Model optimization techniques should be discussed.

2.6 Data requirements

Data requirements of the project are as follows:

- At least 1000 data records are required to train-test a model effectively.
- Non-null values in dataset are required
- A dataset of various features is required to train a model
- Data should be well organised within the columns and rows

2.7 Tools used

Python programming language and various data science & machine learning libraries such as NumPy, Pandas, Scikit-learn are used. Kaggle online notebook, Git-hub.

- Kaggle notebook is used for writing python code
- Pandas is used for data extracting and loading
- NumPy is used for data transformation
- Scikit-learn is used for machine learning models, evaluation metrics, pre-processing of extracted data

A virtual machine environment or physical computer desktop having at least 8Gb of RAM, 64 bit is used. This is necessary to swiftly execute the model.

2.8 Constraints

The program should give direct prediction of a test case by only inputting components' amount in units. It should be user friendly and not complex.

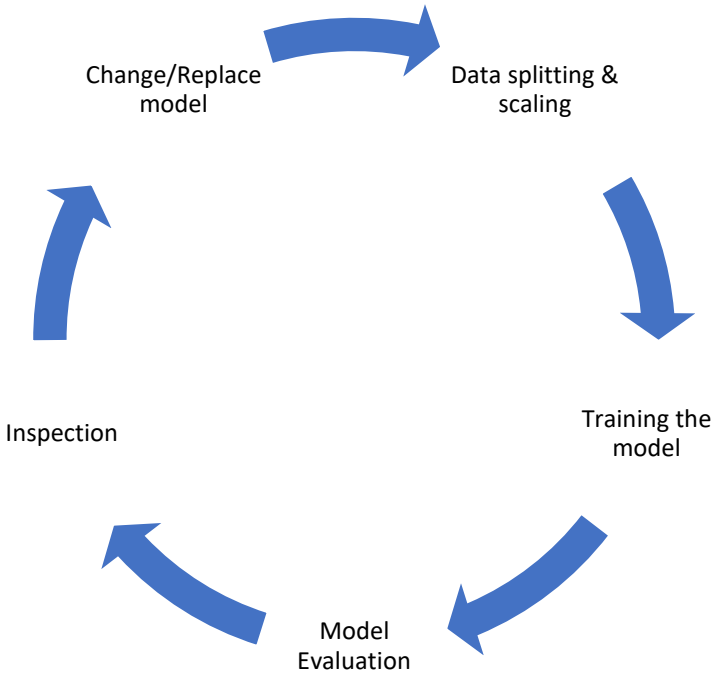
2.9 Assumptions

The project assumes that all data entries used to train the model are flawless and need not require inspection before the use. All the data entries used in the project are real-time observations and not just theoretical records. The project only gives predictions based on quantity of each component used in the mixture and not based on the quality of component used; e.g. the quality of cement used, company name of the cement etc.

3 Design details

3.1 Process Flow

3.1.1 Model training & evaluation



3.1.2 Deployment process

3.2 Performance

The project is predictor of compressive strength of concrete material. The performance of the model can be checked with the scikit-learn metrices like r2 score, MAE, root mean square value, etc.

3.2.1 Code Restructure

The project code can be modified based on the user requirements to get desired outcome.

4 KPI's (Key Performance Indicators)

KPI's include,

- R2 score
- MAE (Mean Absolute Error)
- Root mean square
- Max error
- Mean squared error
- Mean squared log error

5 Conclusion

This project intends to predict the compressive strength of cement concrete with the help of features or components involved in the concrete mixture. So, one can modify the amounts of components in the mixture to obtain desired compressive strength. Hence, it also reduces the time required to measure the strength.

6 References

- http://scikit-learn.org/stable/modules/model_evaluation.html Model evaluation metrices
- http://scikit-learn.org/stable/supervised_learning.html Machine learning models