ML-project-internship

Predict Compresive strength of Concrete

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 the concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength, which is a long period.

I solved this problem using Data science and Machine learning technology, developed a web application which predicts the "Concrete compressive strength" based on the quantities of raw material, given as an input. Sounds like this saves a lot of time and effort right !

Data source: https://www.kaggle.com/datasets/elikplim/concrete-compressive-strength-data-set

Approach:

1. Loading dataset using Pandas and performed basic checks like the data type of each column and having missing values.

2. Performed Exploratory data analysis:

    - First viewed distribution of target of the feature, "concrete compressive strength", which was in Normal distribution with very little right skiewness.
    - Visualized each predictor or independent feature with the target feature and found that there's a direct proportionality between cement and the target while there's an inverse proportionality between water and result as above.
    - Checked for the presence of outliers in all the columns and found that the columns 'age' is having more no. of outliers. Removed outliers using IQR technique, in which I considered both including and excluding the lower and upper limits into two separate dataframes and merged both into single dataframe. This has increased the data size so that a Machine learning model can be trained efficiently.

3. Experimenting with ML alogrithms:

- First , tried with Linear regression model and features selections.
- The models were like Lasso Regression and Ridge Regression.
- The spliting the data into the models and then measuring the performence with using methods like RMSE, MSE, MAE and R^2.
- Next Algorithms tried is DecisionTree Regressor and RandomForest Regressor
- Again spliting data in each models and measuring the performance.
- After that It's model evaluation in which we combining all the model's variables into the list and checking it out to see which model is giving better accuracy.
- To which RandomForest is performing better and the rest.
- So save the model into the pickle file

4. Deployment: Deployed the Randomforest Regressor model using Flask, which works in backend part while for frontend UI web page, used HTML5

At each step in both development and development parts, logging operation is performed which are stored in the development_logs.log and deployment_logs.log files respectively.

So, now we can find the Concrete compressive strength quickly by just passing the quanitites of the raw materials as an input into the web application

Deployment:

The project is deployed on local machine 127.0.0.1:5000/predict
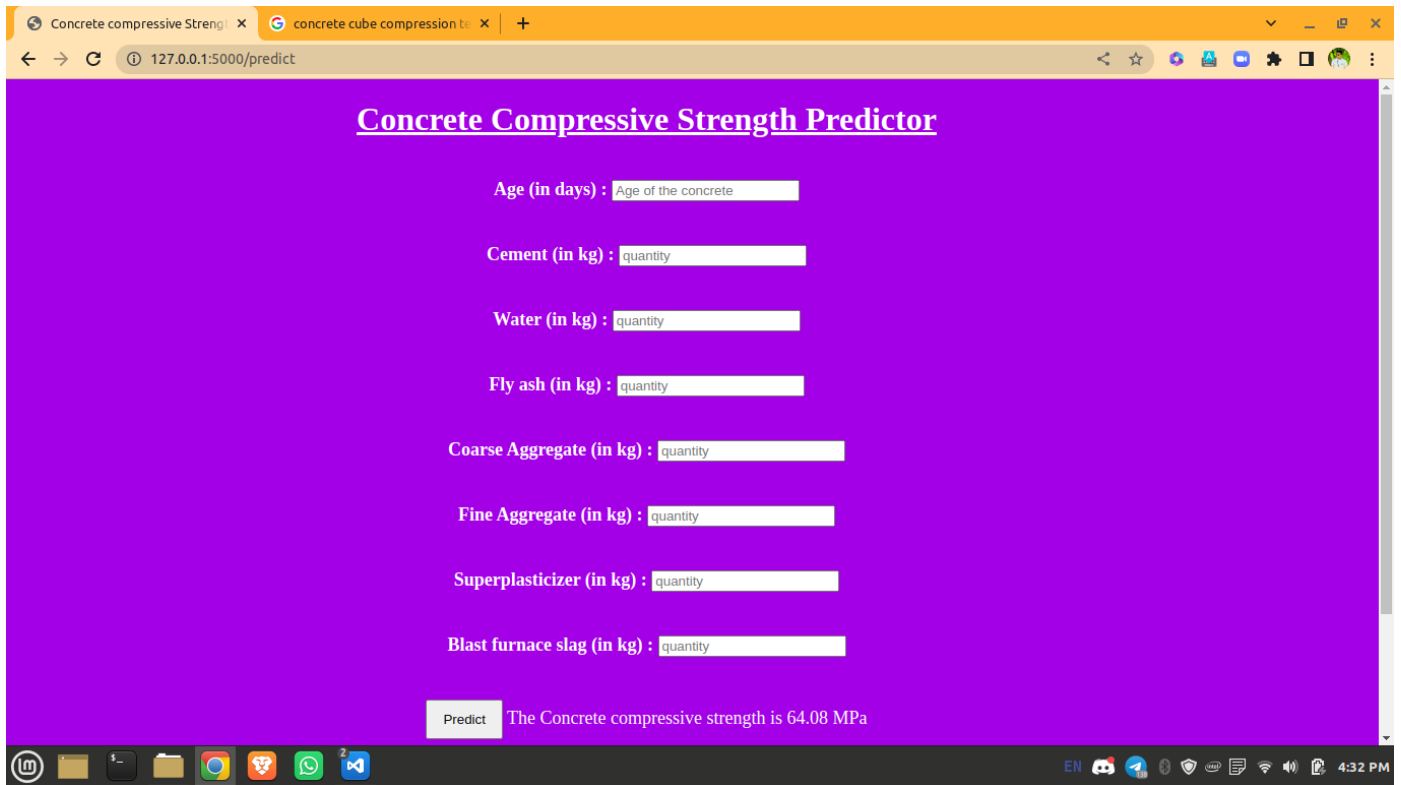
Stack used:

- Jupyter-notebook, python, vscode, html, css, matplolib, seaborn, numpy, pandas, scikit learn, statsmodels, git, gunicorn, joblib, flask, linux.

High Level Design:

Low Level Design:

Architecture:

Detail Project Report:

Wireframe Document:

Demo video:

project demo video link

Refrences:

- Concrete Basics: Essential Ingredients For A Concrete Mixture
- Applications of Fly ash
- Blast furnace slag cement
- Applications of Superplasitcizer in concrete making
- Factors that affect strength of concrete
- Feature selection with sklearn and pandas

- [HTML, CSS tutorials](#)

Author:

- Yash Dawande [LinkedIn](#)