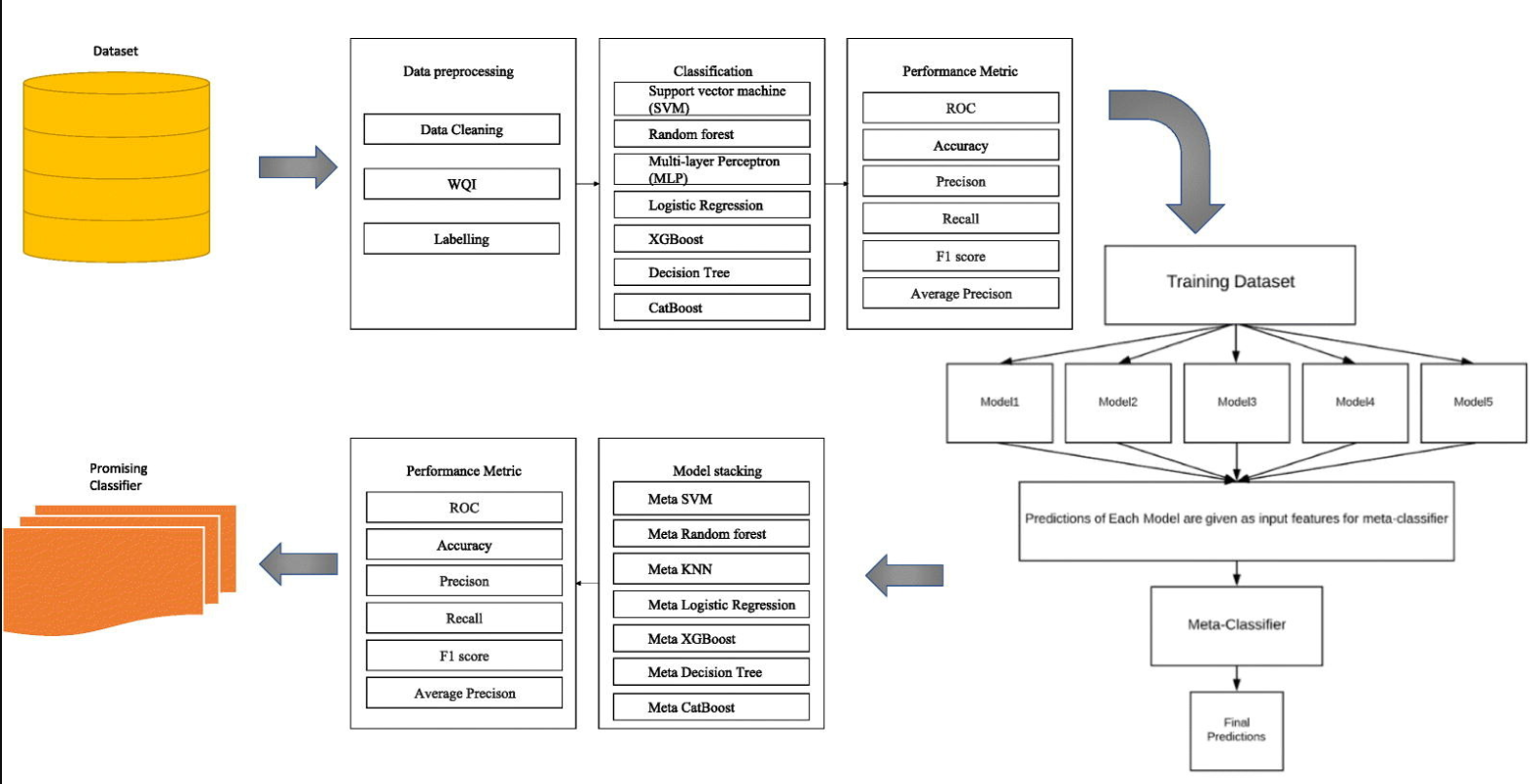
River Water Quality Forecasting using Time Series(Forecasting)

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions.

**Technical Architecture:**

****

**Project Flow:**

● User interacts with the UI to enter the input.

● Entered input is analysed by the model which is integrated.

● Once model analyses the input the prediction is showcased on the UITo accomplish this, we have to complete all the activities listed below,

● Define Problem / Problem Understanding

The problem is to develop an effective and reliable system for river water quality forecasting. This involves predicting the water quality of a river at a specific location and time based on various environmental factors and parameters. The goal is to provide accurate and timely information to stakeholders, environmental agencies, and the public to facilitate better decision-making regarding water usage, environmental conservation, and public health.

● Data Collection & Preparation

○ Collect the data-set

○ Data Preparation

● Exploratory Data Analysis

○ Visual Analysis

● Model Building

○ train and test split

○ Apply the Forecasting algorithms

● Performance Testing & Hyperparameter Tuning

○ Testing model with multiple evaluation metrics

○ Comparing model accuracy

● Model Deployment

○ Save the best model

○ Integrate with Web Framework

● Project Demonstration & Documentation

○ Project Documentation-Step by step project development procedure

**Prior Knowledge:**

You must have prior knowledge of following topics to complete this project.

ML Concepts

o Supervised learning: https://www.javatpoint.com/supervised-machine-learning o Unsupervised learning: https://www.javatpoint.com/unsupervised-machine-learning

Evaluation metrics: https://www.analyticsvidhya.com/blog/2019/08/11-important-model evaluation-error-metrics/

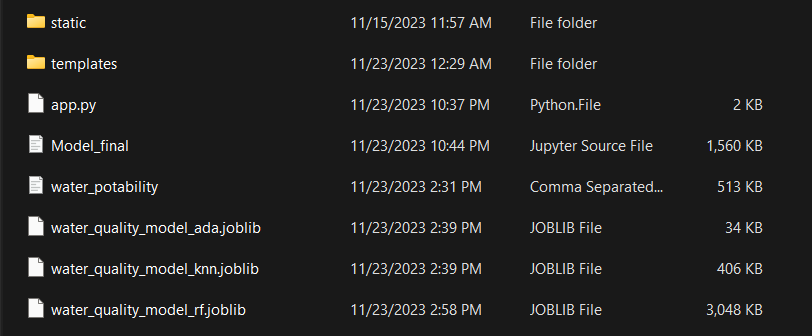
Flask Basics : https://www.youtube.com/watch?v=lj4I\_CvBnt0

**Project Structure:**



**Create the Project folder which contains files as shown below:**



****

We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.

● water\_quality\_model\_rf is our saved model. Further we will use this model for flaskintegration.

● Data Folder contains the Data-set used

● The Notebook file contains procedure for building the model.

**● Milestone 1: Define Problem / Problem Understanding: ● Activity 1: Specify the Genetic Classification problem :**

Access to safe drinking-water is essential to health, a basic human right and a component of effective policy for health protection. This is important as a health and development issue at a national, regional and local level. In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions.

Analysis of data dependencies, including EDA.

1. Prediction of the data in the certain station by data from upstream stations with the highest accuracy.

**Activity 2: Literature Survey :**

**The data-set can be utilized for training and evaluating for forecasting models and machine learning algorithms. Researchers and data scientists can leverage this data-set to develop automated systems for water quality forecasting, and early detection of water quality**.

**Milestone 2: Data Collection & Preparation:**

**Activity 1: Collect the dataset**

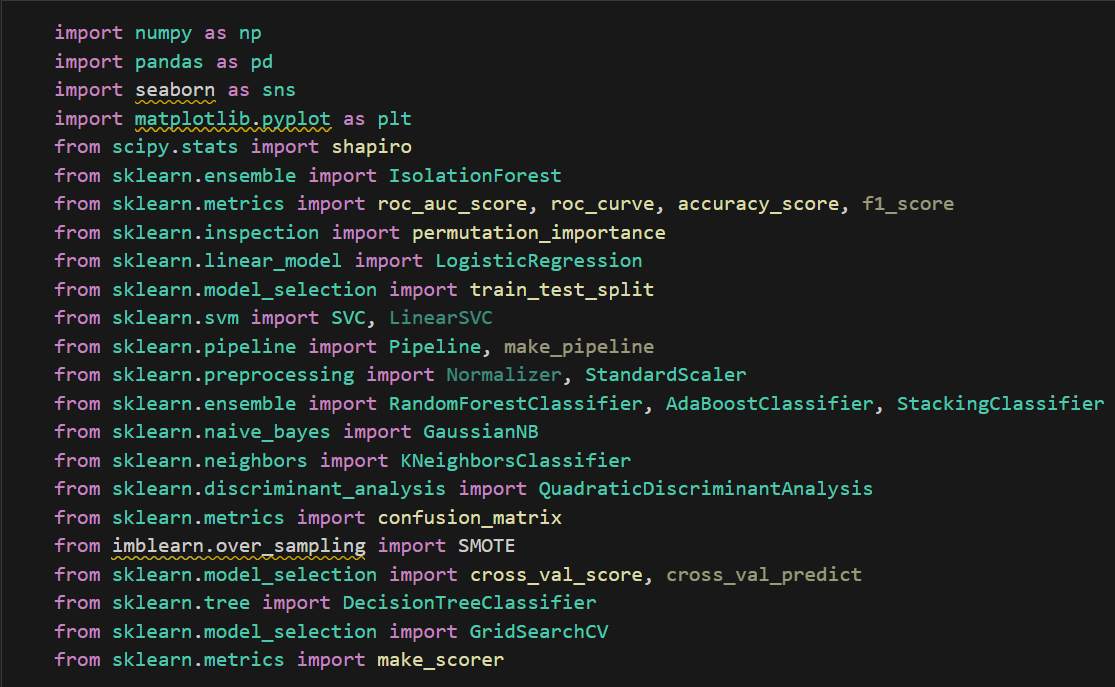
There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.

In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.

Link: <https://www.kaggle.com/datasets/adityakadiwal/water-potability>

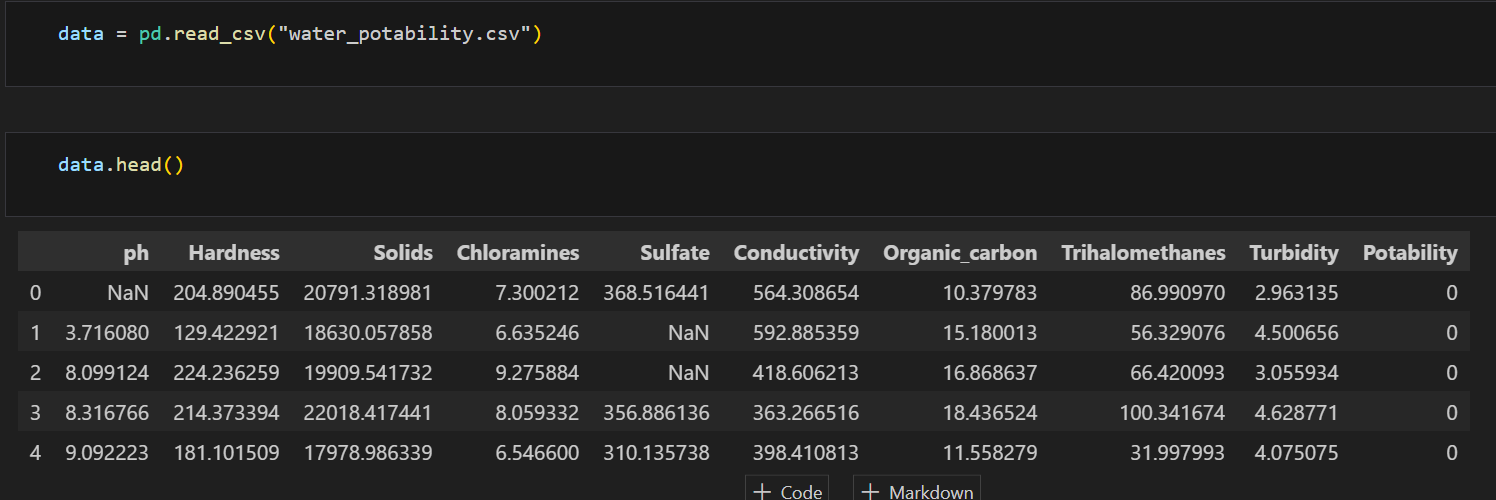
As the dataset is downloaded. Let us read and understand the data properly with the help of some visualisation techniques and some analysing techniques.

**Activity 1.1: Importing the libraries**: Import the necessary libraries as shown in the image.



**Activity 1.2: Read the Dataset :**

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.In pandas we have a function called read\_csv() to read the dataset. As a parameter we have to give the directory of the csv file.



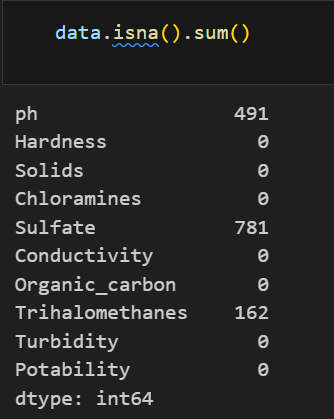
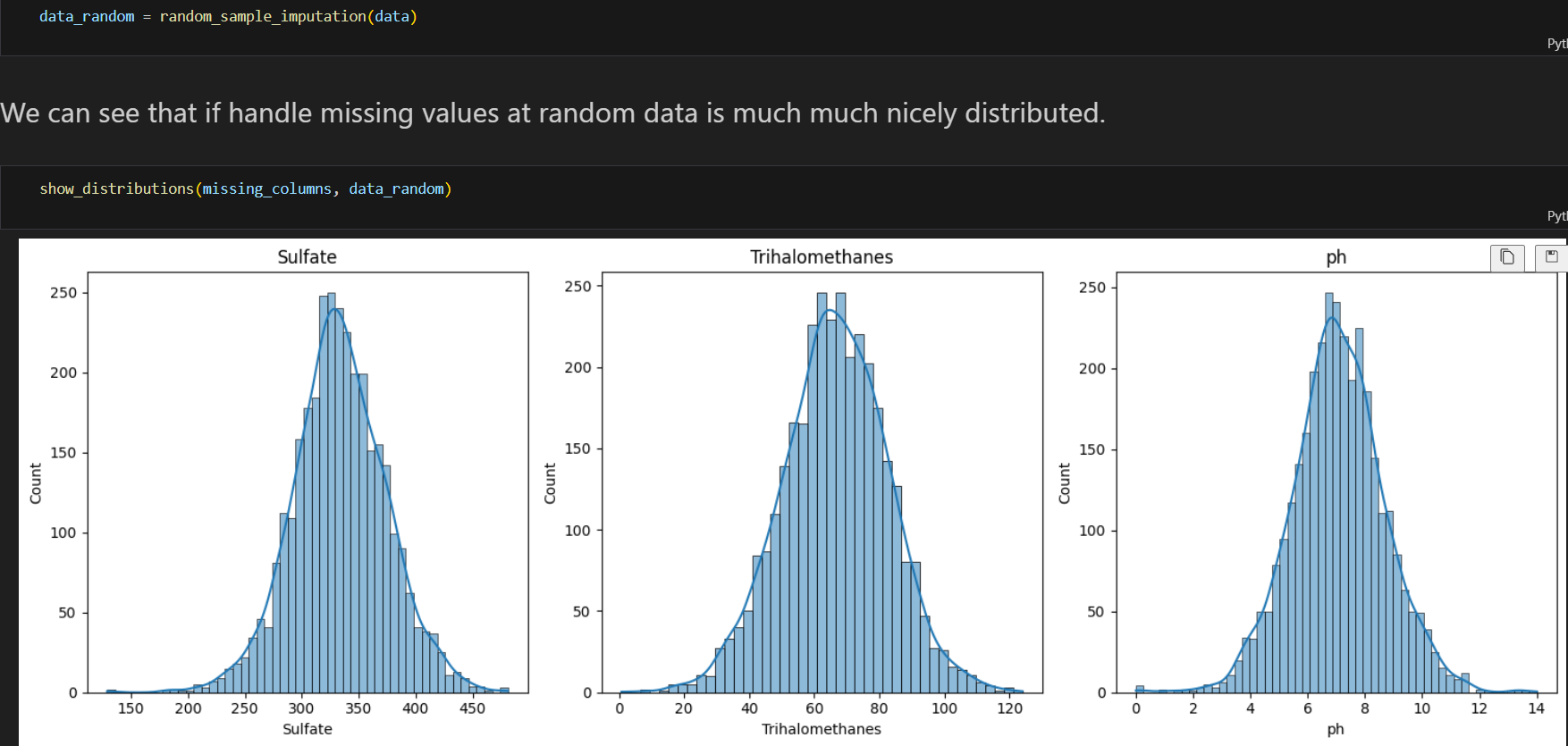
**Activity 2: Data Preparation:**As we have understood how the data is, let's pre-process the collected data.The download data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

● Handling missing values

● Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

**Activity 2.1: Handling missing values:**

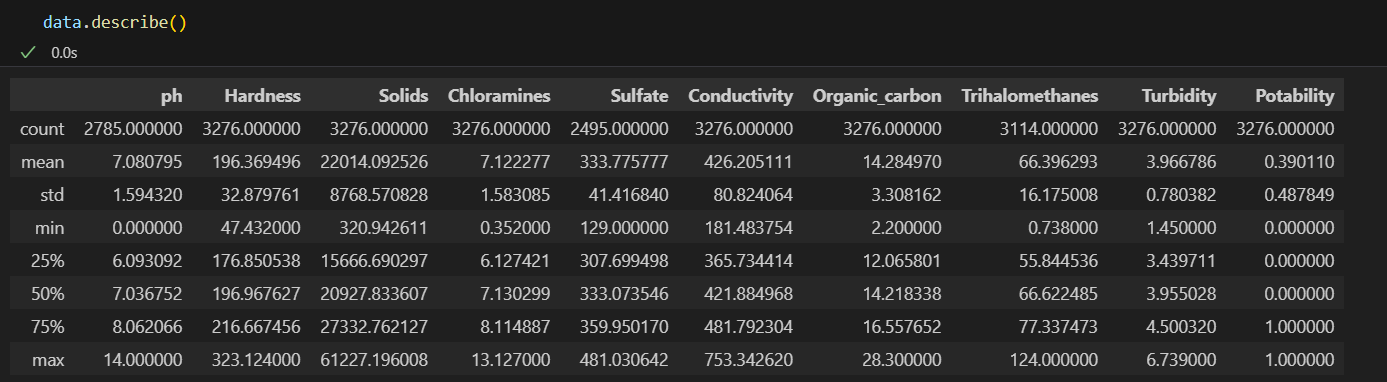
● For checking the null values, df.isna().any( ) function is used. To sum those null values we use .sum() function. From the below image we found that there are null values present in our dataset.

**** ****

**Milestone 3: Exploratory Data Analysis Activity 1:**

**Descriptive statistical**

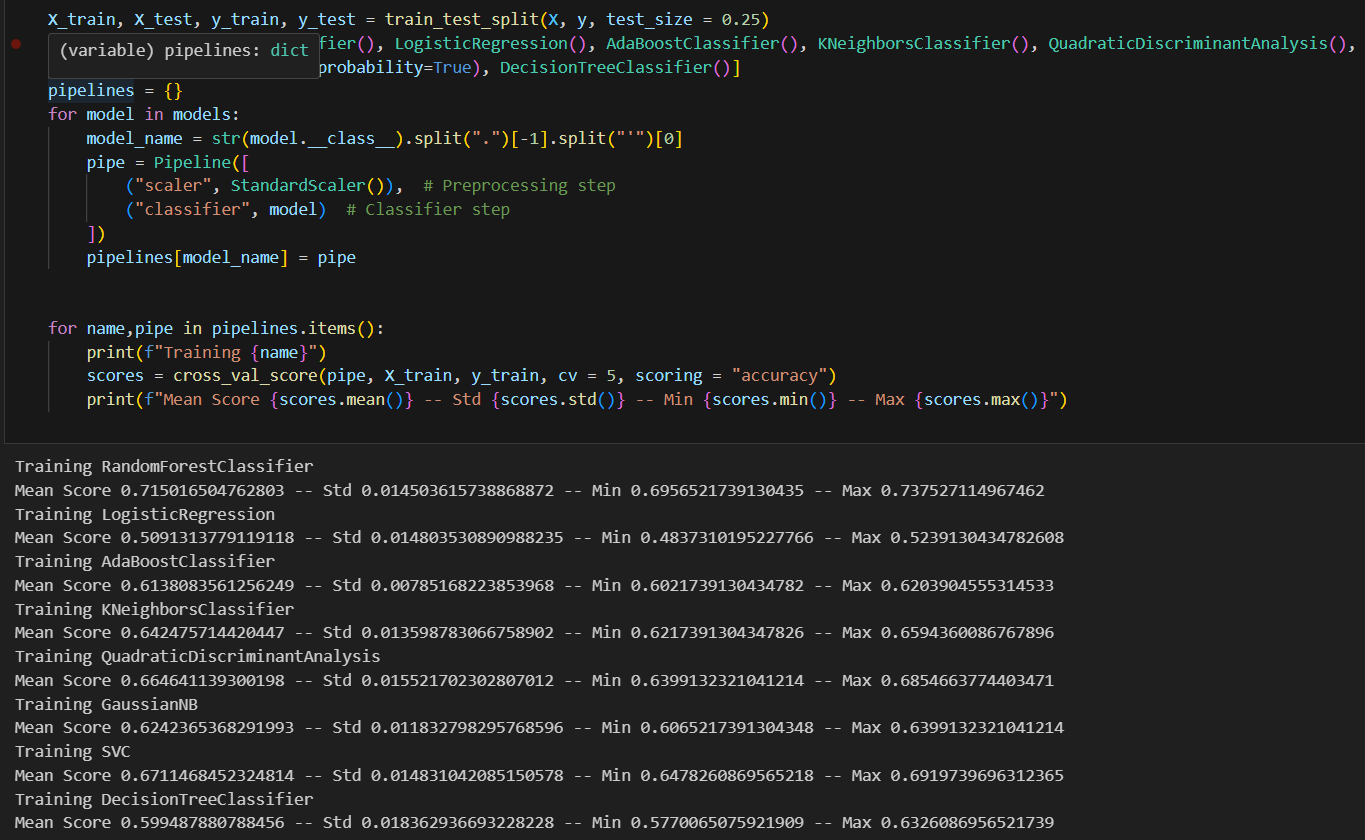
Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.



**Activity 2: Visual analysis**

Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data. It is a way to quickly identify patterns, trends, and outliers in the data, which can help to gain insights and make informed decisions

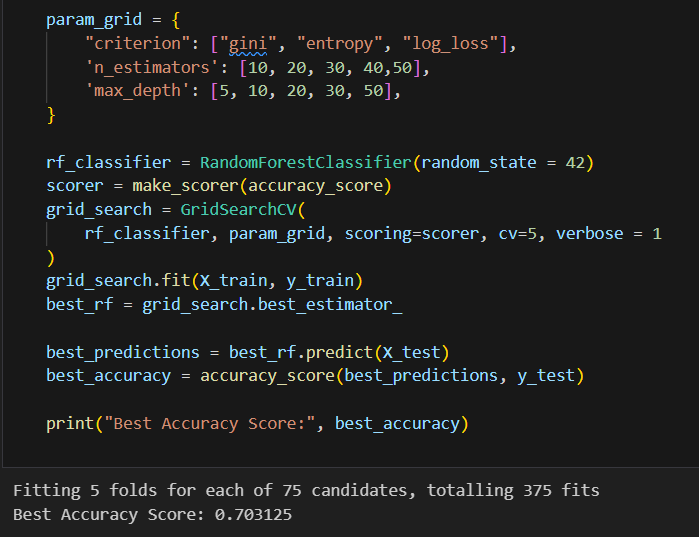
Spliting of the data:



**Milestone 4: Model Building**

**Activity 1: Training the model in multiple algorithms**

Now our data is cleaned and it’s time to build the model. We can train our data on different algorithms. For this project we are applying three classification algorithms. The best model is saved based on its performance.

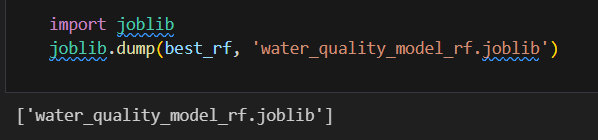


**Milestone 6: Model Deployment Activity**

**1: Save the best model**

Saving the best model after comparing its performance using different evaluation metrics means

selecting the model with the highest performance.This can be useful in avoiding the need to retrain the model every time it is needed and also to be able to use it in the future.



**Activity 2: Integrate with Web Framework**:In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks Building

HTML Pages

● Building server-side script

● Run the web application

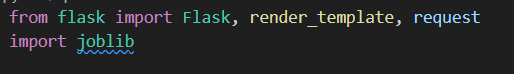
**Activity 2.1: Building Html Page:**

For this project create HTML file namely

* index.html

and save them in the templates folder.

**Activity 2.2: Build Python code:** Import the libraries



Load the saved model. Importing the flask module in the project is mandatory. An object of Flask class is our WSGI application. Flask constructor takes the name of the current module ( name ) as argument

Here we are routing our app to predict() function. This function retrieves all the values from the HTML page using Post request. That is stored in an array. This array is passed to the model.predict() function. This function returns the prediction. And this prediction value will be rendered to the text that we have mentioned in the submit.html page earlier.

Main Function:

#run the app

if name\_\_ =='\_\_main\_\_':

app.run()

**Activity 2.3: Run the web application**

● Open anaconda prompt from the start menu

● Navigate to the folder where your python script is.

● Now type “python app.py” command

● Navigate to the localhost where you can view your web page.

● Click on the predict button from the top left corner, enter the inputs, click on the submit button, and see the result/prediction on the web.

Now,Go the web browser and write the localhost url (http://127.0.0.1:5000) to get the index.html

