

**W25-PAK-INP-AI-06**

**Submitted by**

*Nakhsheeb Zahoor*

**Supervised by**

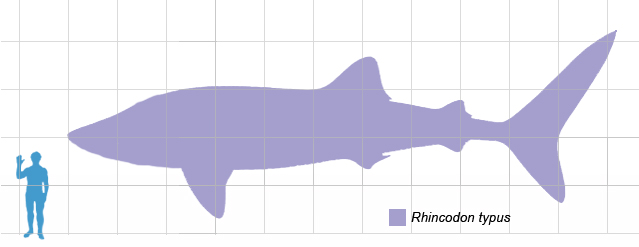
*Zreen Fatima*

**Introduction**

The idea is to create a machine learning model from scratch which means collecting data set. Choosing targeted model based on problem statement and then creating a model of our own. The aim is to predict the weight of certain species of fish. With that being said one should know what fish itself is

A **fish**  is an [aquatic](https://en.wikipedia.org/wiki/Aquatic_animal), [an amniotic](https://en.wikipedia.org/wiki/Anamniotes), [gill](https://en.wikipedia.org/wiki/Gill)-bearing [vertebrate](https://en.wikipedia.org/wiki/Vertebrate) [animal](https://en.wikipedia.org/wiki/Animal) with swimming [fins](https://en.wikipedia.org/wiki/Fish_fin) and [hard skull](https://en.wikipedia.org/wiki/Craniate), but lacking [limbs](https://en.wikipedia.org/wiki/Limb_(anatomy)) with [digits](https://en.wikipedia.org/wiki/Digit_(anatomy))

Fish range in size from the huge 16-metre (52 ft) [whale shark](https://en.wikipedia.org/wiki/Whale_shark) to some tiny teleosts only 8-millimetre (0.3 in) long, such as the cyprinid [*Paedocypris progenetica*](https://en.wikipedia.org/wiki/Paedocypris_progenetica) and the [stout infantfish](https://en.wikipedia.org/wiki/Stout_infantfish).



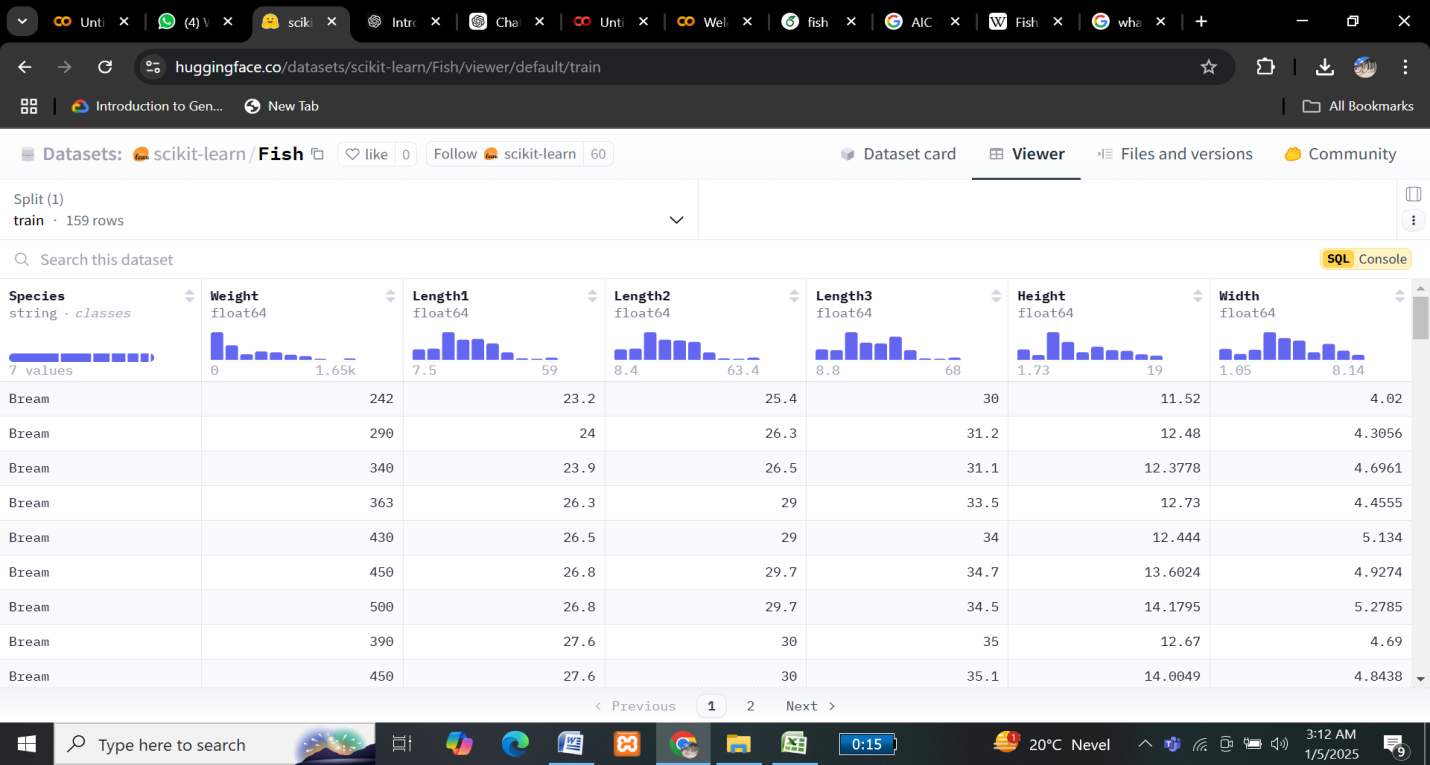
Whale shark in comparison to human

 Smallest fish on human finger 

**Gathering Data set**

The basics of good machine learning model is good data

Data set was taken from Hugging face



Reference

<https://huggingface.co/datasets/scikit-learn/Fish?sql=--+The+SQL+console+is+powered+by+DuckDB+WASM+and+runs+entirely+in+the+browser>

Firstly necessary libraries were installed

Then data set was imported

After that actual values and predicted values were assigned to x and y

**Model**

We took inspiration from random forest regressor and created our own model

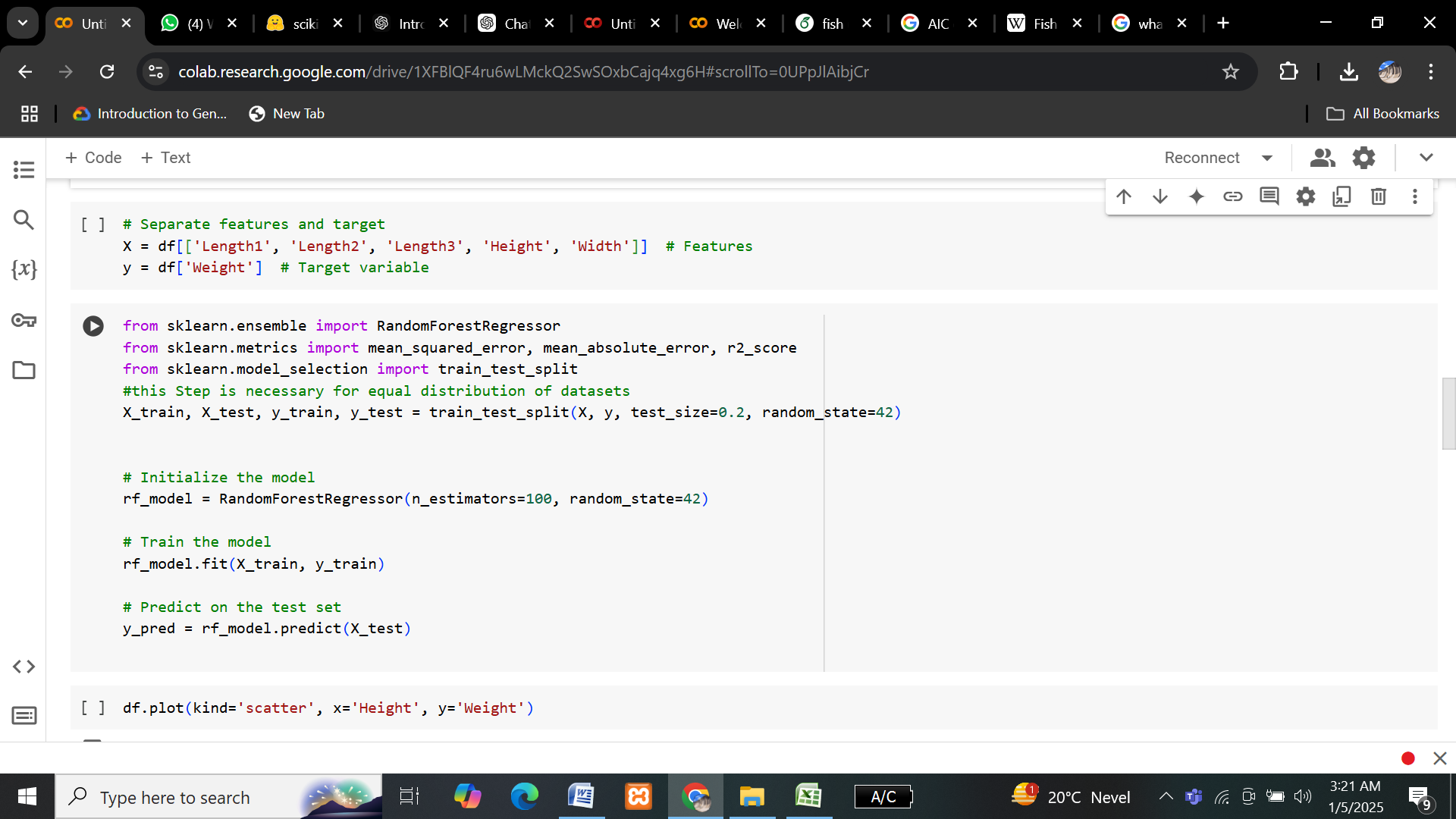
Random forest

A random forest is an ensemble learning method that combines the predictions from multiple decision trees to produce a more accurate and stable prediction. It is a type of supervised learning algorithm that can be used for both classification and regression tasks

It could be a regreesor or classifier

We used regressor

[Regression](https://www.geeksforgeeks.org/types-of-regression-techniques/) is a statistical approach used to analyze the relationship between a dependent variable (target variable) and one or more independent variables (predictor variables). The objective is to determine the most suitable function that characterizes the connection between these variables



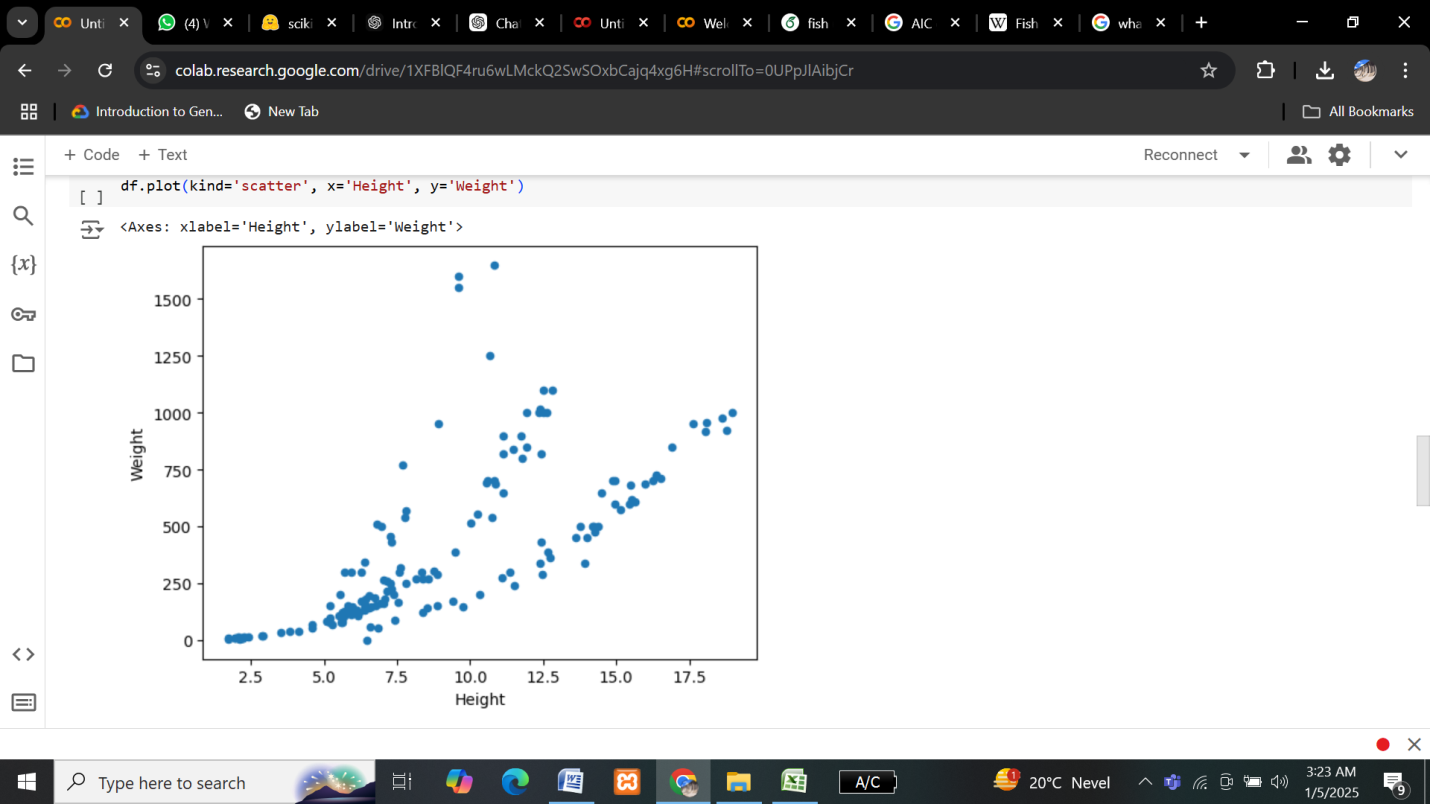
The steps involved were training and testing data i.e splitting it

Then initializing the model

Training model (putting our values)

Then lastly

Predicting



X and y taken as height and weight

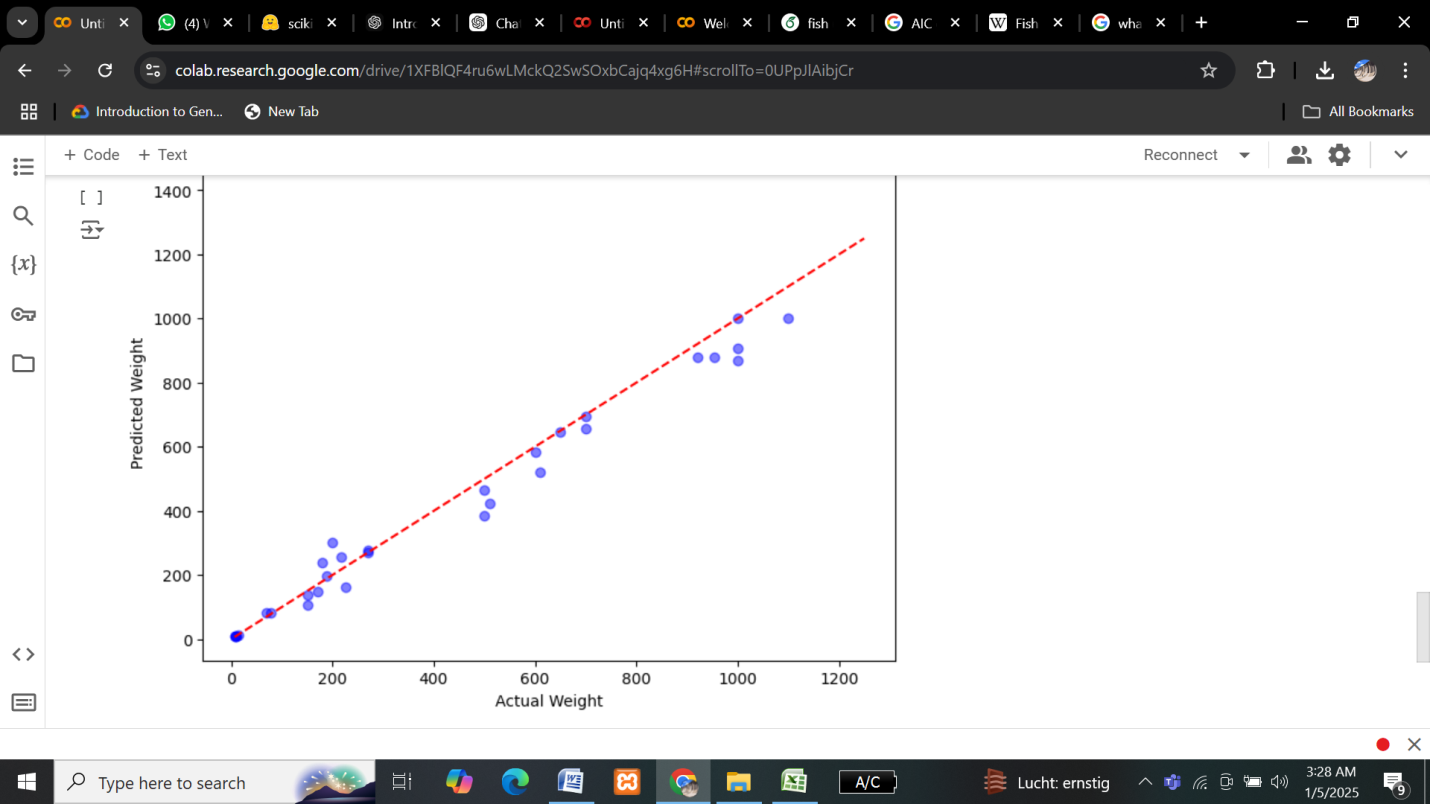
And them plotting our points on Cartesian plan (showed by blue dots)

Note: showing no signs of under fitting or over fitting

**Evaluation**

Then model was evaluated to check credibility of results

Which was up to 96%



In the end best fit line was created

**Best fit line**

A straight line that depicts the trend of the given scattered data plots on a graph.

**Conclusion**

The purpose of the model is to predict weight of species of fish. We can change parameters to our desired targeted specie and can find out its weight.