Documentation

I-Introduction:

Estimating sediment discharge using machine learning models, including PyTorch-based models.
The goal is to predict the total sediment load based on various input features such as channel width, flow velocity, and slope.

II-Data Loading and Preprocessing:

EDA:

- Although the data is nominal but it was saved as objects .
- The Data has some Duplicates so I dropped it .
- Data Describe for statical info .
- Removed the outliers .
- the data was not normaly distibuted and had high skewness.
  So , we did some transformation such as (logs , boxcox , sqrt).

III -Modeling :

- Ensamble learning :
  It was a right choise as we use multiple models and choose by majority voting the best model .

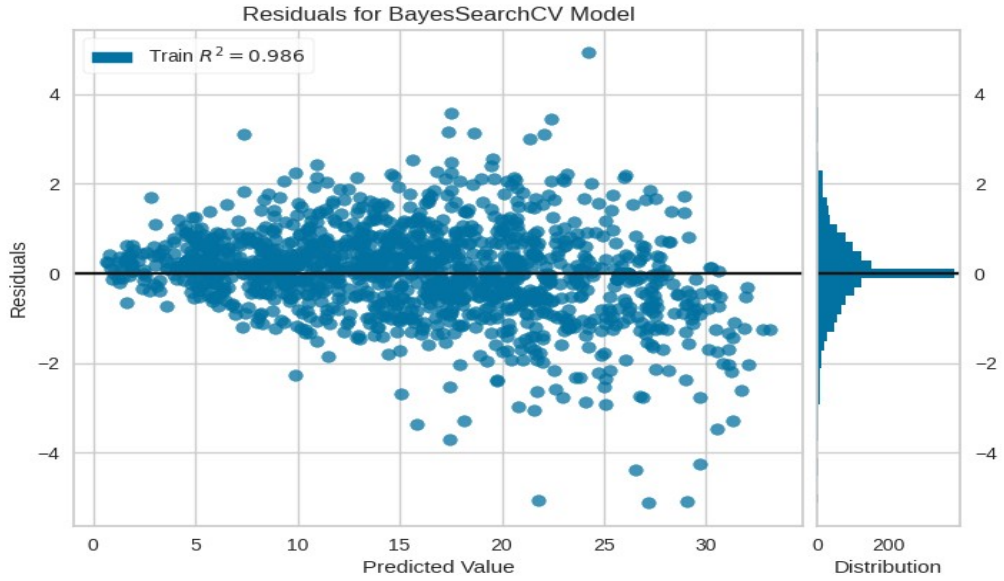
-First , the Random Forest also used Hyperparameter tunning "bayes search" to get the best parameters that will give us the best fit for our model

Parameters like :

```
params = {
    'n_estimators': Integer(10, 500),
    'max_depth': Integer(1, 50),
    'min_samples_split': Integer(2, 20),
    'min_samples_leaf': Integer(1, 20)
}
```

And the result as follows :

The evaluation metrics	R^2 score	MSE	RMSE
Result	0.9121435983524147	6.059913613326676	2.461689178862083



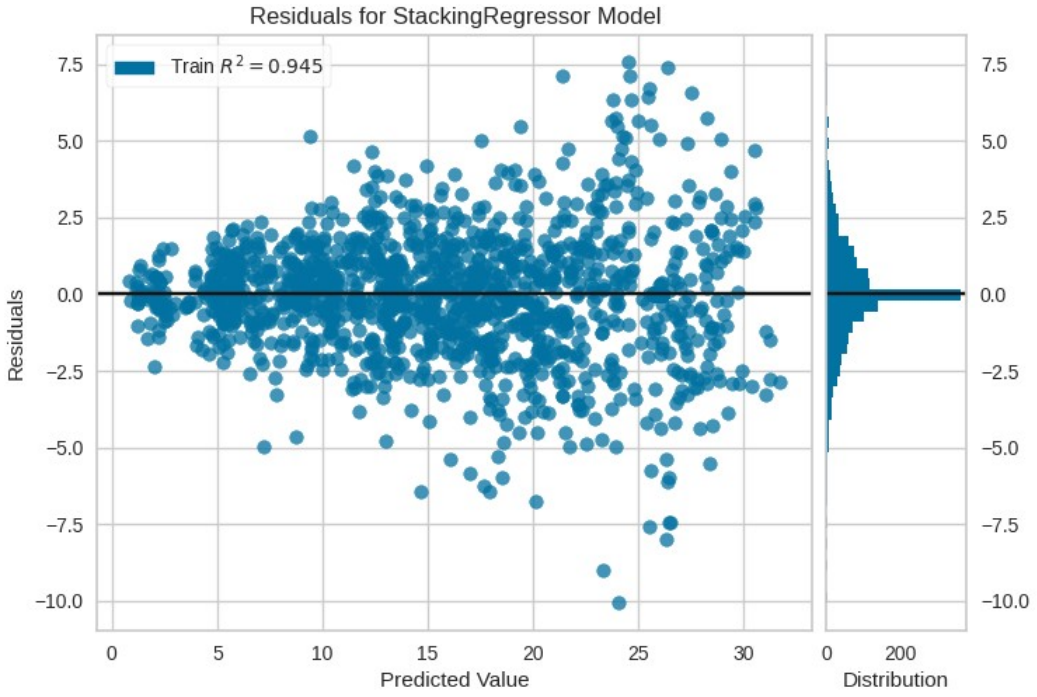
- the second model is Stacking as it works sequentially and train final model based on the best model

And the chosen models was :
1-Adaboost "as it is also ensamble learning model".
2-Gradientboost "this one also as previous one "
3-RandomFores .
4-And the final model is linear Regression .

Also used Hyperparameter Tuning "Bayes search".

And the result as follows :

The evaluation metrics	R^2 score	MSE	RMSE
Result	0.9121435983524147	6.969239698761439	2.6399317602471166



- Deeplearning models

First , TensorFlow
Used two Dense hidden layers , adam optimizer , 250 epochs .

Evaluation metrics	Test loss	R^2 score
Result	9.260442733764648	0.8657424483055332

Secondly , pytorch :
I made an initial model and than made another enhanced model .

That model with more hidden layers , activation functions , adam optimizer and used 1000 epochs.

Evaluation metrics	Test loss	R^2 score	Mean Squared Error
Result	9.0105	0.8694	9.0105