

Spectrum Club,
College of Engineering and Technology,
Bhubaneswar
www.spectrumcet.com

Dear Intern,

Welcome to **Data Science and Machine Learning Task 2 part 2**. We are finally ready to start predicting the final grade of a student..

Technology Stack to be used:

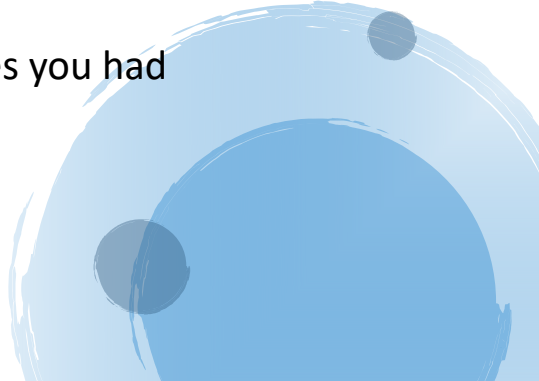
In this final task, we will mainly be working on. The libraries we will mainly be using will be :-


- sklearn
- numpy
- Pandas
- statsmodel.api

Final stage:

Tasks:

Billy has his data refined. It's ready now to predict his marks based on his attributes.

1. Create a linear regression model by using `sklearn.linear_model` library.
 2. Fit the model with the `x_train` and `x_test` values you had previously created.
- 

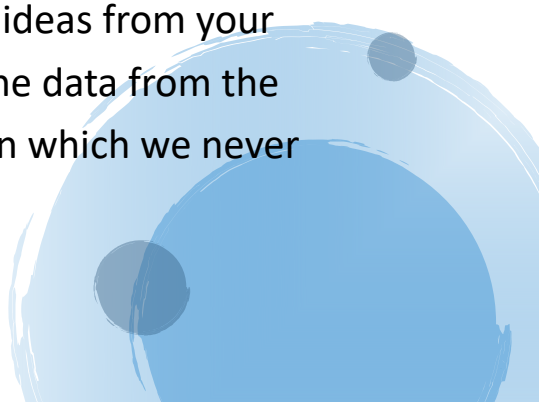
- 
3. Calculate the accuracy by using the score attribute of the linear regression model on the `x_test` and `y_test` values.
 4. Using its `predict` attribute predict the values of all `x_test`, and then plot a scatter plot between the true and predicted values of `x_test`.


We don't need all the features of the input, some features affect the accuracy a lot while some don't affect it at all. Now we will be using backward elimination method to find the most important attributes which affect our prediction from that dataset.

5. Using `statsmodel.api`'s OLS(Ordinary Least Squares) find how each feature affects the prediction and the model. Then based on that eliminate the features and keep on doing so till you achieve highest accuracy with suitable features are required.

All interns are required to make a short project report, stating in general about the task and the project which you worked on in the spectrum internship drive. **This report shall be written in your `readme.md` file in your github repo.**

You can try to implement any interesting ideas from your side post project completion, maybe analyze the data from the dataset and find an interesting angle or relation which we never





thought about. Extracting interesting and useful information from the dataset and visualizing them may well lead you to find some quite interesting facts from the data, and who knows, maybe your prediction and inference can land you a spot on the spectrum internship drive page as well as a certificate of merit.

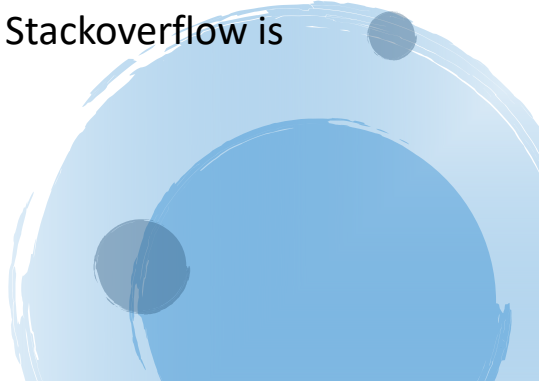
Extra tasks/future possibilities:

You can use various other algorithms like Decision Tree/ Random Forest/ SVM, etc. Each will have their own different type of tuning and feature selection. Try to get even higher accuracy with the help of these.

Based on your final feature selection list you can build an interface(like a chatbot or website or app) which takes the desired input from the user and predicts the marks he/she will get.

RESOURCES: -

The resources provided here include both documentation and Youtube tutorials for the above tasks. If any case of any confusion, do Google once and search around or ask in the forum, but do understand the concepts behind the functions properly. Remember, Stackoverflow is your best friend for programming.



- 
- **Backward Elimination –**
(Example)

<https://www.javatpoint.com/backward-elimination-in-machine-learning>

(Video Tutorial):

<https://www.youtube.com/watch?v=Nv6e9S8fwzA>

The dataset has been provided with this zip file, and the attributes, and the type of data they contain(numeric, non-numeric/binary, nominal), etc. have been provided too.

Good luck!

LAST DATE OF SUBMISSION: 31th May-2020



WARM REGARDS,
SPECTRUM, CET-B

