

MACHINE LEARNING USING PYTHON TO PREDICT THE FISH WEIGHT

We, Chetana Patel, Nikita V Ambaldhage, Rashmi R Pai in the course of the Internship has been provided training all the concepts of Technologies involved in Industry 4.0. During the Internship period, we have been assigned for the project on Machine Learning using python to predict the fish weight. We have successfully completed the project under the mentorship of Mr. Avinash from Gustovalley Technovations.

This report is prepared as requirement for the internship of BE program of Computer Science and Engineering Department, KLE Institute of Technology, Hubballi.

INDUSTRY 4.0

We're in the midst of a significant transformation regarding the way we produce products thanks to the digitization of manufacturing. This transition is so compelling that it is being called Industry 4.0 to represent the fourth revolution that has occurred in manufacturing. From the first industrial revolution (mechanization through water and steam power) to the mass production and assembly lines using electricity in the second, the fourth industrial revolution will take what was started in the third with the adoption of computers and automation and enhance it with smart and autonomous systems fuelled by data and machine learning.

This manufacturing revolution will increase productivity, shift economics, foster industrial growth, and modify the profile of the workforce—ultimately changing the competitiveness of companies and regions. As the adoption of Industry 4.0 is increasing day by day across various industries, the scope for the skill sets will be enormous for the driving technologies of Industry 4.0. The technologies of Industry 4.0 include Integrated System, IOT, Data Management, Cloud Computing, Cyber security, Artificial Intelligence, Additive Manufacturing, Augmented Reality.

About the Company

Gustovalley Technovations is an automation-based company working towards the goal of Digital transformation. Gusto Learn is the initiative of Gustovalley technovations which focuses on offering Industry training to the students.

The Students will be good exposed to the latest technologies and will have the opportunity to work on solutions for the Realtime problems.

Our course is designed in such a way that each student will get complete hands on experience and will have deeper understanding of concepts by working on projects and assignments to solve real-time problems.

Problem Statement

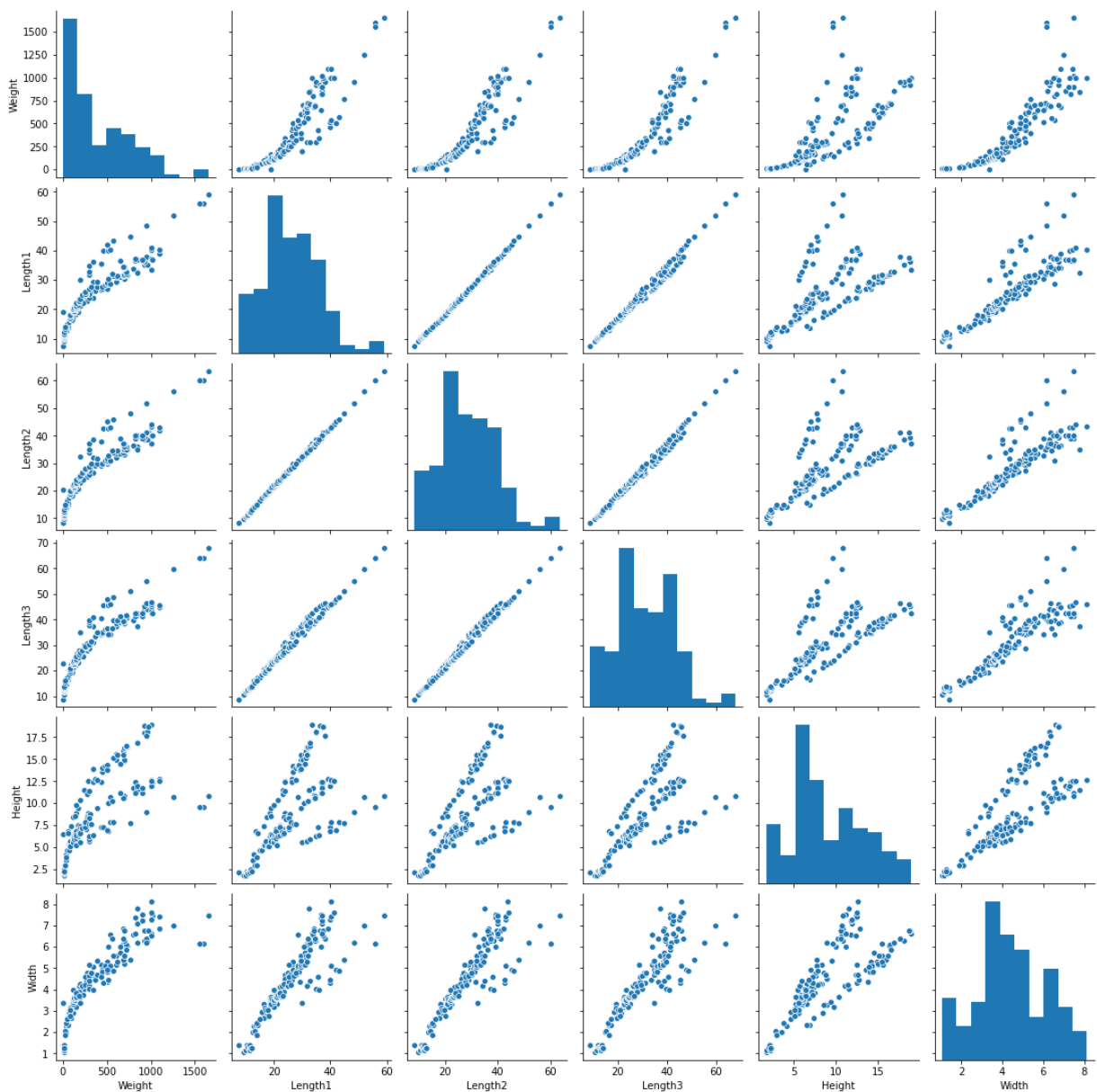
Analyse the data set provided, do Exploratory Data Analysis, Visualize the data and find all principal components from dataset. From the pre-processed dataset create a ML model (Linear Regression) to predict the output and calculate the error associated with the model created. (Predict the fish weight)

Methodology

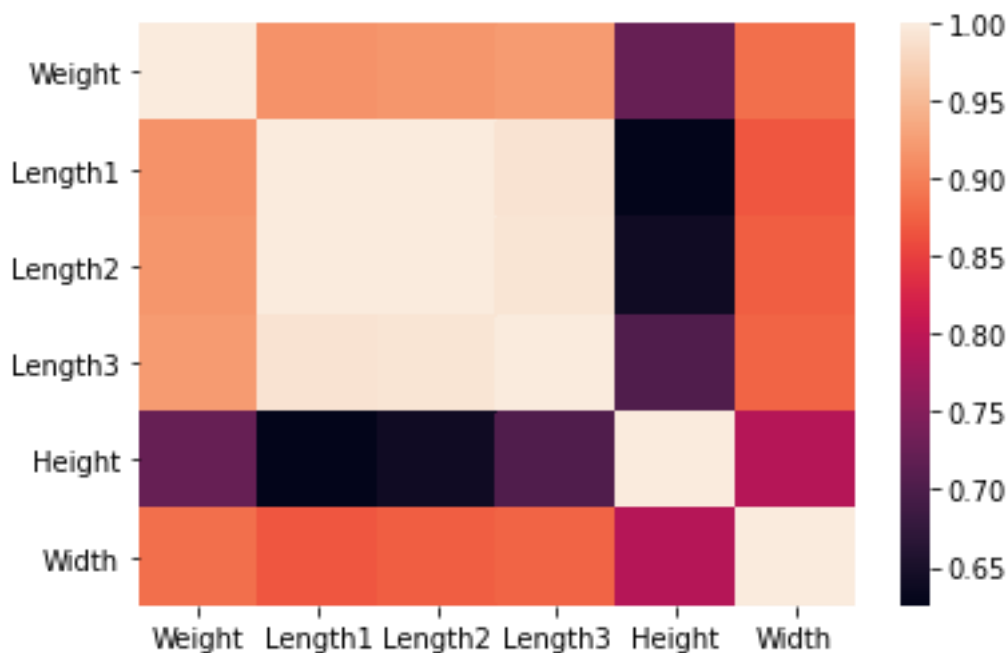
We used Data processing libraries Numpy and Pandas for numerical array operations. We understood the Data processing techniques by working on the datasets.

We then learned the concepts of Exploratory Data Analysis (EDA) and applied the learnings of data pre-processing from Numpy and Pandas library to our fish weight data set and visualized our dataset using Matplotlib and Seaborn library. From the visualization of our dataset we have found several insights and identified the principal components for our machine learning process.

Visualization of the fish weight data set using pair plot :



Identifying the principle component:



Machine learning techniques:

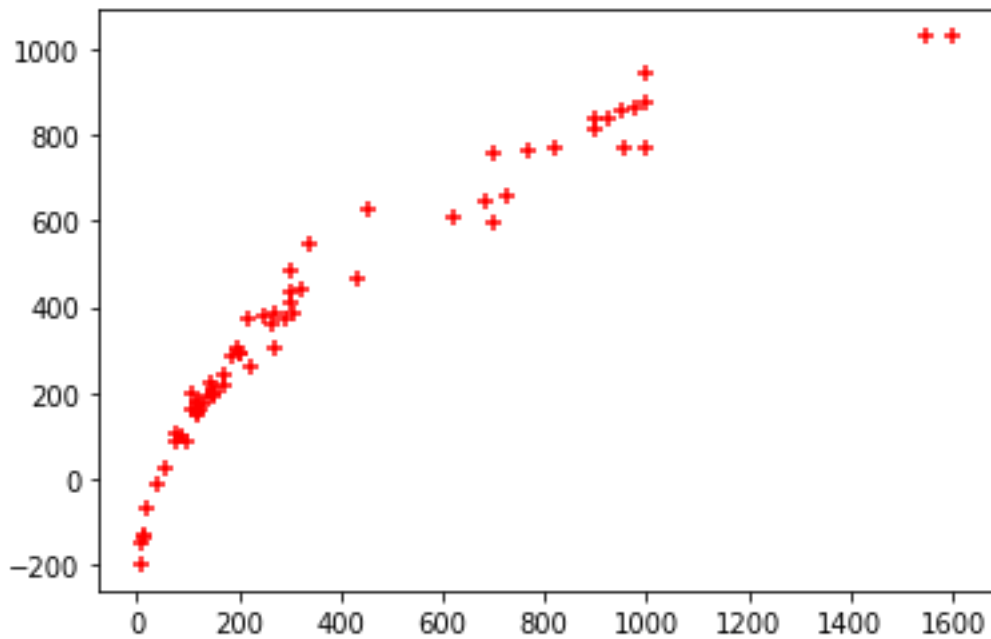
We learned about the types of machine learning techniques. There we went deeper in to Linear Regression techniques starting from the principle, working and algorithm.

We then worked on Scikitlearn Library for developing a prediction model to predict the fish weight based on the length, height and width.

After pre-processing and visualization of our data. We divided our data set in to training data (70%) and testing data (30%) and using Scikitlearn Library we developed a Machine learning model using the training data and calculated the coefficient and the intercept.

After development of the model we then compared the result from the Training model and actual data by feeding the test data set in to the model.

We found that our model is 97% accurate by considering the Root Mean Squared Error.



Conclusion

a) DATA VISUALIZATION:

The given dataset is visualized(fish.csv) using python and python libraries (numpy, pandas, matplotlib) as a tool to find all the principal components from the dataset

b) MACHINE LEARNING MODEL:

Linear Regression Model is used to predict the weight of the fish as the regression line makes the estimation of a dependent variable more accurate.