

Name: Sharon Sam S

Company: Learn and Build

Duration: 11th May, 2022 – 10th July, 2022

This summer, I have been lucky to work as an intern here at LEARN AND BUILD. Through this journey, I have learned more than I could possibly layout in this blog.

This blog will serve as an overview of my internship experience and the things I learned along the way at Learn and Build.

I have registered my application for an opportunity to do an internship in the domain of the emerging and trending field of Data Science. Gratefully Learn and Build considered my application and given an opportunity to carry out my learning in this field.

Before tasking with ample of projects, I was given a training videos for a period of month. I gained knowledge from those videos which where explained in a way one could easily understand and grasp the concepts quickly. From those training session I could able to gain the knowledge of many concepts including Python programming language and Machine learning algorithms.

After the completion of the training session, I was tasked with filling out a list of goals for the summer consists of three projects. These goals guided my experiences and provided me a direction for learning and aids in gaining deeper insights about machine learning algorithms.

The first task of the internship is to build a Phone Directory using a data structure in python known as dictionary. I used python dictionary to create a phone directory to add, update, delete, select one or all contacts. Initially I have created a menu to carry out all the necessary operations and then I have declared each operation as a function in python to define the set of rules for each operation. Finally tested it with some inputs and the output is given.

The second task is all about the understanding the concepts of statistical models preferred as Machine Learning algorithms and develop a model for the given problem statement. The problem statement is given that to Train a machine learning model so that it's capable of predicting the profit of a company after their

expense. Below is the agenda that I have performed to develop a model. I have initially Imported the necessary libraries and dataset required to develop a model. Then I have explored the data by printing the properties of the dataset given. Extraction of input and output features from dataset is done. Some preprocessing steps like converting categorical feature to numeric, split the dataset with 20% as test data, normalizations using Standard Scaler and MinMax Scaler have been done. Finally created a Model using Linear, Ridge, Lasso Regressor models and the models are evaluated using r2 score and root mean squared error. Model Evaluation is also done without Categorical features. At last the best Model which yields an accuracy of 94 % is selected and visualization of Regression Model is drawn. Thus the predicted values are outputted.

The third task is similar to the second task except two models were created for the desired dataset to work upon of one which is regression problem and another is classification problem. Below is the agenda that I have performed to develop a model.

For regression task, I have used the Chennai House Price prediction dataset as the price money is continuous in nature. I have initially Imported the necessary libraries and dataset required to develop a model. The dataset is downloaded from Kaggle. Then I have explored the data by printing the properties of the dataset given. Some preprocessing steps like handling missing values, handling categorical features, extraction of input and output features from dataset, splitting the dataset into train and test data with 20% as test data and with some normalizations used such as Standard Scaler and MinMax Scaler have been done. Finally created a Model using Linear, Ridge, Lasso and KNeighbors Regressor models and the models are evaluated using r2 score and root mean squared error. At last the best model is selected and visualization of Regression Model is drawn. Thus the predicted values are outputted.

For Classification task, I have used the Animal Class Type prediction dataset which is a multiclass problem. I have initially Imported the necessary libraries and dataset required to develop a model. The dataset is downloaded from Kaggle. Then I have explored the data by printing the properties of the dataset given. Some preprocessing steps like extraction of input and output features from dataset, splitting the dataset into train and test data with 33% as test data and with some normalizations used such as Standard Scaler and MinMax Scaler have been done. Finally created some Classifiers using Logistic Regression Classifier, Decision

Tree Classifier and Random Forest Classifier models and the models are evaluated using accuracy metrics. At last the best model is selected and visualization of Classifier Model is drawn.

In Regression Task for prediction of Chennai House Price, Lasso Regression with MinMaxScaler with highest r^2 score = 0.673687 and with lowest RMSE = 56.381385 is the best model. In Classification Task for classifying the Animals based upon its type, Decision Tree Classifier with highest Accuracy of 97% is the best model.

Thus I have successfully completed all of the three tasks given within the duration of 2 months. Thanks to Learn and Build for providing me with this opportunity.