**DengAI : Disease Spreading Prediction**

|  |  |
| --- | --- |
| Name | Y.Adarsha |
| Project Title | DengAI:Disease Spreading Prediction |
| Problem Statement | <https://www.kaggle.com/qianyigang129/dengai-dataset> |
| Data set used and its link | Three datasets are used  1)train 2)test 3)train labels 4)submission format  <https://www.kaggle.com/qianyigang129/dengai-dataset?select=DengAI_Predicting_Disease_Spread_-_Training_Data_Features.csv> |
| How many rows and columns in the data set(s) and what are its data types | 1456 rows and 24 columns in the train dataset ,416 rows and 24 columns in the test dataset. These both have20 columns have numerical data and 4 columns are factors. The train label dataset has 2 numerical and 2 factors. |
| Goal of the Project | Predicting a column called ‘total\_cases’, it represents the number of cases of dengue for each city on a weekly basis. |
| Challenges Identified in your project | Most of the challenging part was the train dataset with 548 missing values and the columns having high value numeric data,which was difficult when we were reducing the skew. |
| Pre-processing techniques used | 1)Inter Quartile for treating outliers  2)skewness verification  3)Mice package. |
| What Machine Learning Algorithms Implemented and their accuracies | Before Pre-processing, ML Accuracies:  r-squared value(linear model) :0.24  mean absolute error(svm model) : 13.8  After Pre-processing, ML Accuracies:  r-squared value : 0.2  mean absolute error(for svm model) : 7.8 |
| Use of Machine Learning Algorithms | I have implemented two ML algorithms linear model and support vector regression model(svm), the svm model gave higher accuracies because for SVM has supported for reducing cross-validation errors and provided many methods known as kernels useful for controlling mean absolute errors apart from linear model there are many functions useful for predictions. |
| Final Conclusion | Finally I predicted the total\_cases column for the test dataset but during prediction we encountered some negative values for the column generated due to our pre-processing errors and these values are converted to positive using abs() function and the values are submitted through ‘submission.csv’ file name. |