Dengue Outbreak Visualization

Divyang Soni (divyang.soni@sjsu.edu)
Varun Upadhyay (varun.upadhyay@sjsu.edu)
Vinod Katta (vinod.katta@sjsu.edu)

Charles Davidson College of Engineering @San Jose State University
One Washington Square, San Jose, California -95192-0080, USA

Abstract

Dengue is a deadly disease that affects 33% of the world population. A lot of lives are lost due to dengue outbreaks in various parts of the world. A nearly prediction of such outbreak could save many lives. Along with a prediction, a proper representation and visualization of these cases in a affected region would be very helpful for the rescue/medical teams to determine their action plan to help affected people.

KeyWords: Dengue outbreak, Prediction analysis, Visualization

1. Introduction

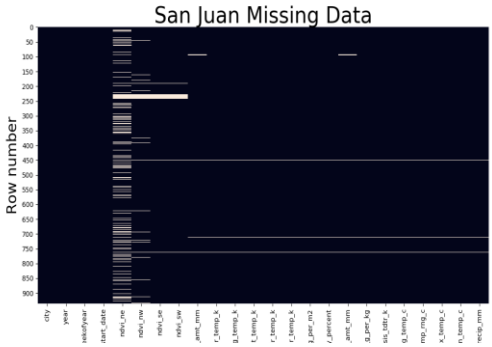
Dengue Outbreak Visualization is containing 2 separate modules named Dengue Prediction and Dengue Outbreak Visualizer. Dengue Prediction module aims to predict future dengue outbreaks per year and weeks while Dengue Outbreak Visualizer is a web application with aim to give a geographical visualization of dengue outbreaks using previous predictions.

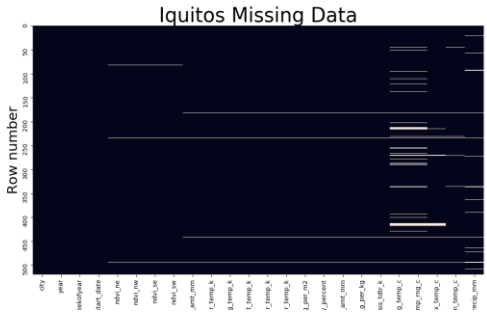
We implemented these features using available weather and dengue cases data of two cities and represented geographically to provide proper visualization of outbreak. This paper will give you detailed description about the functionalities that we have implemented in this application.

2. Analysis

By referring various websites, we found that there are many websites which have past dengue data available and some other websites are having weather related data for those years. Also, some websites contain both kind of data sets, but we cannot find proper relationships shown between these two different datasets. The next analysis has result that there is not proper visualization available for dengue cases which can be helpful for the respective services to use them. So, we decided to use these two different datasets and give a future prediction about the dengue in different regions and show the predictive results into geographical distributed format for proper visualization.

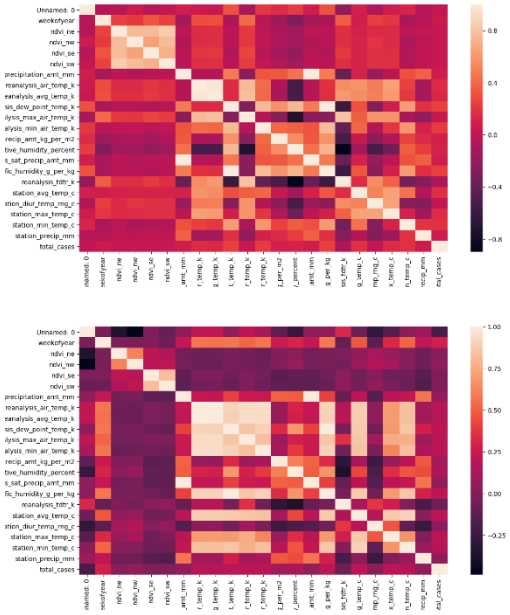
We gathered data from various government and other websites which provide datasets of dengue outbreaks and their trend according to the weather data in those areas.



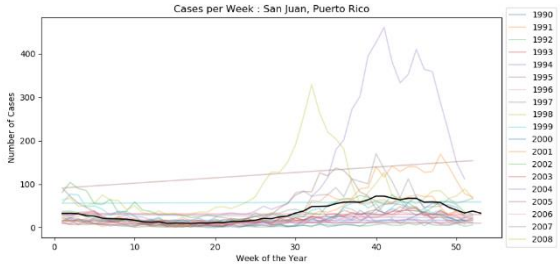


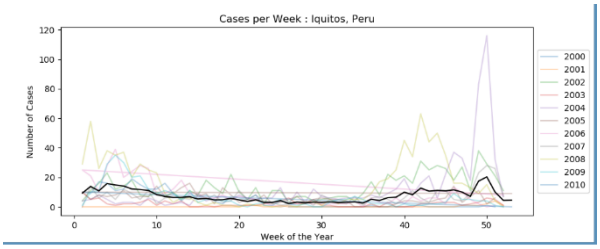
We found most of the satellite data(natural vegetation) is missing in the San Juan dengue data. And, in Iquitos temperature data is missing.

We cleaned the data using various python libraries viz. pandas, numpy. We filled missing values in those datasets using with the mean values. These filled datasets are used for dengue prediction. We used linear regression algorithms to create various prediction models. To visualize our prediction models and verify our results we used matplotlib and seaborn charts. Our primary goal is to correlate the dengue outbreak with the weather data we have.

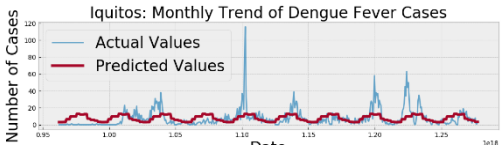


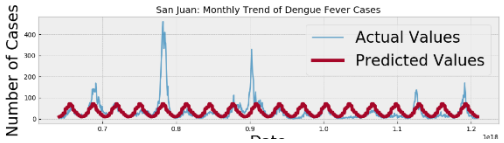
The weather factors that are in correlation with the dengue cases are considered for the weekly, monthly prediction of the data.





We used the 70% of the data to train the prediction model and used rest 30% data for testing the prediction model. We got the results as follows:

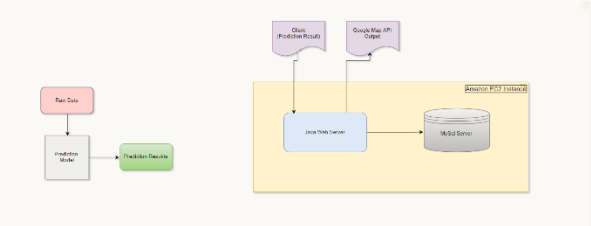




Once we are done with the prediction, our next step is to visualize these results. We have designed a Java web application and hosted the Tomcat server on an Amazon EC2 instance.

The client would input the prediction results in our web application in CSV format. The database is then read by the application and stored it in a MySQL database which is then read by our Java application which then displays the results using the Google Maps API. The results display the year-wise and week-wise visualization of number of cases over the two cities.

3. ARCHITECTURE



A. Datasets and Database:

We got datasets for weather and dengue cases from a Hackathon contest of dengue prediction. These datasets were in unstructured format and had a lot null values. By using pandas min fill technique, we have cleaned the datasets to use for prediction. After successful prediction, we are generating a CSV file containing prediction related data for year and weeks. We have used java programs to transform and load the prediction into RDBMS (**MySQL**) for further processing of data.

B. Prediction Algorithm

We have used python to cleaning data and analyze it. We have separate py files which we need to run to get the prediction records from the given raw data. Each and every steps used is creating several analysis and prediction related file and in last we are getting the CSV file as our final output.

C. Server implementation

Server is providing a web services to consume prediction data and put that in proper format in database. Server is implemented in **Java (JSP/Servlet)** to get optimal performance. Server uses JDBC connector module to communicate with database for exchanging data. Clients applications directly communicates with java server by invoking required web services or by calling appropriate JSP file.

D. Application Deployment

Complete application is configured and deployed on **Amazon EC2 linux instance**. Amazon EC2 is configured with elastic IP, which can be used from anywhere to connect to the application. Application can be scaled up to handle client's requests as per requirement.

4. Future Enhancement

After developing a more advanced prediction model, website can be integrated with the prediction model so that the users of website can directly input the data and get the prediction results.

More visualization graphs can be developed to show the dengue prediction results to the users for better analyzing the dengue cases. Users can perform different analysis through the prediction results in the website.

5. Conclusion

We analyzed the previous years data for dengue cases and related it to the weather data. We created a model of relating dengue cases to the weather conditions and did our predictions based on this model. We further created a visualization tool using Google Maps API where we fed our prediction results and then visualized our results for various regions affected by dengue. We have also hosted the application on Amazon web services.

6. Contributions

Divyang Soni : Student in MS Software Engineering at SJSU. Developed google maps interface to visualize the dengue cases. Performed initial analysis on the datasets and plotted the results.

Varun Upadhyay：Student in MS Software Engineering at SJSU. Developed linear regression model to predict the dengue. Deployed the project on AWS.

Vinod Katta：Student in MS Software Engineering at SJSU. Gathered dengue and weather data for the project and cleaned the datasets. Tested the prediction model.

7. References

-Wikipedia:
www.wikipedia.com

-DrivenData:
https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/page/82/

-Ebola Outbreak Study:
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152438