An Industry Oriented Mini Project Report on

DENGUE PREDICTION THROUGH CLIMATIC DATA

Submitted to the Department of Computer Science & Engineering, GNITS in the partial fulfillment of the academic requirement for the award of B.Tech (CSE) under JNTUH

By

M. Akhila	**17251A0580**
B. Ruchira	**17251A0584**
R. Sri Laxmi Ragini	**17251A05B3**

under the guidance of

Mrs. K Sneha Reddy

Assistant Professor

Department of Computer Science & Engineering

G. Narayanamma Institute of Technology & Science (for Women)

Accredited by NBA & NAAC &Affiliated to JNTUH

Hyderabad-500 104, T.S, INDIA

February, 2021

Department of Computer Science & Engineering
G. Narayanamma Institute of Technology & Science (for Women)
Accredited by NBA & NAAC &Affiliated to JNTUH
Hyderabad-500 104, T.S, INDIA

Certificate

This is to certify that the Mini Project report on "**Dengue Prediction Through Climatic data**" is a bonafide work carried out by M.Akhila(17251A0580), B.Ruchira(17251A0584),R. Sri Laxmi Ragini(17251A05B3) in the partial fulfillment for the award of B.Tech degree in Computer Science & Engineering, G. Narayanamma Institute of Technology & Science (For Women), Shaikpet, Hyderabad, affiliated to Jawaharlal Nehru Technological University, Hyderabad under our guidance and supervision.

The results embodied in the Industry Oriented Mini Project have not been submitted to any other University or Institute for the award of any degree or diploma.

Internal Guide

Mrs. K Sneha Reddy
Assistant Professor
Department of CSE

Head of the Department

Dr. M. Seetha
Professor and Head
Department of CSE

External Examiner

Acknowledgements

We would like to express our sincere thanks to **Dr K. Ramesh Reddy, Principal** , GNITS, for providing the working facilities in the college.

Our sincere thanks and gratitude to **Dr. M.Seetha, Professor and HOD**, Dept. of CSE, GNITS for all the timely support and valuable suggestions during the period of our mini project.

We are extremely thankful to **Mrs. R. Pallavi Reddy, Asst. Prof, Mrs. S. Sandhya, Asst. Prof, Mr. T. Anil, Asst. Prof**, Dept. of CSE, GNITS for their encouragement and support throughout the project.

We are extremely thankful and indebted to our internal guide, **Mrs. K Sneha Reddy, Assistant professor,** Dept. of CSE, GNITS for her constant guidance, continuous advice, encouragement and moral support throughout the miniproject.

Finally, we would also like to thank all the faculty and staff of CSE Department who helped us directly or indirectly, parents and friends for their cooperation in completing the miniproject work.

M. Akhila (17251A0580)

B. Ruchira (17251A0584)

R. Sri Laxmi Ragini (17251A05B3)

ABSTRACT

Dengue fever (DF) is of huge public health problem that causes mortality and morbidity worldwide. Currently, dengue fever is the most crucial infectious disease. DF is a disease caused by a family of viruses, known as Flaviviridae. It is a mosquito-borne viral infection that is being rapidly transmitted by the Aedes mosquitoes. The World Health Organization indicated that early detection of DF and appropriate treatment will decrease fatality rates from more than 20% to less than 1%.

This project is aimed to develop a model that could predict dengue outbreak using climate data and detect early signs and symptoms of dengue fever and also serve as a practical system for notification of the disease.

Dengue depends mainly on climatic factors and that data can be used to predict its possibility. The data gathered is trained using Machine Learning algorithms. Based on the prediction of algorithm, user will be asked for further details as a series of questions. Based on severity appropriate suggestions will be given. The questions asked are relevant to symptoms of dengue and these questions are gathered from real doctors and symptoms generally observed in patients

SYSTEM REQUIREMENTS:

Software Requirements:

Platform	-Windows 10
Front End	-HTML, Bulma CSS
Development Tool	-Jupyter Notebook
Back End	-Flask, SQLite3

Hardware Requirements:

Name of the Processor	–Intel core i5
Hard Disk Capacity	-1 TB
RAM Capacity	-8.00GB

Contents

Abstract……………………………………………………………………………...**iv**

1. INTRODUCTION..**1**

1.1. OBJECTIVES 1

1.2. METHODOLOGY 2

1.3. ORGANIZATION OF THE PROJECT... 3

2. LITERATURE SURVEY **4**

2.1. EXISTING SYSTEMS 4

2.2. PROPOSED SYSTEM 5

2.3. ADVANTAGES 6

3. SYSTEM DESIGN..**7**

3.1. ARCHITECTURE.. 7

3.2. DATA PREPARATION 8

3.3. DENGUE FORECAST... 9

3.4. DENGUE SYMPTOMS IDENTIFICATION 13

3.5. STATISTICS OF DENGUE AFFECTED PEOPLE 15

4. IMPLEMENTATION..**17**

4.1. TECHNOLOGY USED.. 17

4.1.1. Python.. 17

4.1.2. Python Packages 17

4.1.3. Flask 18

4.1.4. Bulma CSS 19

4.1.5. HTML... 19

4.1.6. Anaconda.. 19

4.1.7. Openweathermap Onecall API 19

4.2. UML DIAGRAMS.. 20

4.2.1. Use Case Diagram:.. 20

4.2.2. Sequence Diagram: 22

4.2.3. Activity Diagram: 23

5. RESULTS..**26**

5.1. DISCUSSION ON RESULTS 26
5.2. SCREENSHOTS 29
6. CONCLUSION AND FUTURE WORK **34**
6.1. CONCLUSION 34
6.2. FUTURE WORK 34
6.3. REFERENCES 34

1. INTRODUCTION

Dengue is a most infectious disease in the world, there are about 50-100 million infections per years and 0.7 million deaths per years from around the world Dengue fever, which is transmitted to humans by Aedes mosquito, is one of the most serious infectious disease.

There are 4 serotypes of the virus that causes dengue. These are known as DEN-1, DEN-2, DEN-3, DEN-4. Severe dengue is a potentially lethal complication which can develop from dengue infections. Symptoms of dengue fever are high fever, headache, muscle and joint pain, vomiting, and skin rash about2-7 days. These symptoms are expressed in 3-4 days after infection. In the worst case, patients have blood bleeding, low levels of blood platelets, plasma and blood pressure, shock, organ failure, and finally dead. These worst symptoms have period about 24-48 hours.

Dengue is intensively influenced by the climate conditions, and usually circulates near tropical or subtropical regions. As dengue fever has strong transmissibility and difficult to cure, it is able to cause serious morbidity and mortality, greatly threatening people’s healthy life as well as placing heavy burdens on health care systems. Thus, in addition to the basic precaution steps such as awareness campaigns, education to the community, more efforts especially in dengue fever forecast are necessary and urgent.

1.1. Objectives

The main objectives of the proposed Model are

- To develop a model for early dengue forecast
- To analyze symptoms of user to check for dengue.
- To provide statistics to the user about dengue affected people in their region
- To consider various climate factors and predict outbreak in the region

1.2. Methodology

The project aims to develop a model for dengue forecast, aids user in identifying their symptoms early on to avoid late diagnosis. In addition, the user can know about the statistics of people affected in their location.

Dengue fever is geographically distributed in tropical and subtropical regions. Climate change is likely to expand the geographical distribution of several vector-borne human infectious diseases. The risk of dengue transmission is increased by warming climates, as the growth and development of mosquitoes are significantly influenced by temperature and humidity.

Ambient temperature and rainfall (Humidity, dewpoint) are important factors that directly affect the development of dengue virus in major mosquito vectors Ae. aegypti and Ae. albopictus. Climate change will not only affect the rate of mosquito development, but also the virus incubation time. In other words, climatic factors influence dengue ecology both directly and indirectly by affecting mosquito growth dynamics, virus replication, and mosquito–human interactions.

For Dengue Forecast temperature is significant because it is associated with dengue fever incidence. Temperature is an important climatic factor affecting biological processes of mosquitoes, including their interactions with viruses. Temperature is also positively associated with pre-adult mosquito maturation, oviposition rate, and virus incubation rate in mosquitoes. Using Pearson’s correlation and feature selection algorithms attributes of weather data can be ranked from most to least prominent.

Furthermore, Location (latitude and longitude) changes the climate of the region. Based on that number of cases change per region. It will be helpful to user to know about statistics of dengue affected people based on their current location.

Lastly to identify symptoms of dengue fever, certain combinations of symptoms are defined so that one can say about the type of dengue particularly. If symptoms are not assessed properly dengue identification will be difficult.

1.3. Organization of the project

Dengue predictions are made by using different models and considering various types of factors. Many experiments are being conducted to make the predictions more reliable as to take measures accordingly.

The accuracy of the system depends on the algorithm used for prediction. The project flow is organized as follows:

- The climatic data is collected from the user.
- The data is sent to model for prediction.
- Prediction of number of cases is made according to the given data
- Statistics of dengue affected people are displayed on screen of user according to their location
- Then, Symptoms are collected from the user.
- Suggestions are given according to the given symptoms.

The rest of the document deals with the previous work study on dengue predictions in Chapter 2, the architecture and design models of the proposed system in Chapter 3, the utilizations of different modules in Chapter 4, discussion on results in Chapter 5 and Future Enhancements of the project in the last Chapter of the document.

2. LITERATURE SURVEY

2.1. Existing Systems

Geographical statistics:

This study presents spatial analysis of Dengue Fever (DF) outbreak using Geographic Information System (GIS) in the state of Selangor, Malaysia. DF is an Aedes mosquito-borne disease. The aim of the study is to map the spread of DF outbreak in Selangor by producing a risk map while the objective is to identify high risk areas of DF by producing a risk map using GIS tools.

Using rainfall data and dengue cases:

In Malaysia a good dengue surveillance system was designed using a Neural Network Model (NNM) and Nonlinear Regression Model (NLRM) using different architectures and parameters incorporating time series, location and rainfall data to define the best architecture for early prediction of dengue outbreak.

Using number of cases:

In this model, the dengue outbreak was predicted by knowing the predicted number of cases that occur. The prediction was made using the Seasonal Autoregressive Integrated Moving Average, (SARIMA or Seasonal ARIMA), model resulting in high accuracy prediction. However, this study involved only one factor, which is the number of cases. It causes other affecting factor conditions not to be found. Temperature, rainfall, and sunlight play an essential role in the cycle of dengue transmission

Using number of cases with other factors:

Prediction using the number of DF cases approach was also carried out in a model using a linear regression method. The results show that temperature, precipitation, and sunlight correlate significantly in several regions. However, the correlation only sees the effect of factors partially.

Predicting the outbreaks using Fuzzy combined with Logistic Regression have an average accuracy around 79.93%. But the predictions developed using this algorithm is bound to compromise with imbalanced data.

Guo constructed a dengue search index and combined it with the weekly dengue cases and climate factors to use several machine learning algorithms as candidate models to predict dengue incidence approaches predict the dengue outbreak only based on one or two climate factors

Using meteorology data:

In Thailand, meteorology data is also combined with the patient's clinical data to predict outbreaks. The model used is the Generalized Additive Model. The results show that the model resulting in weak predictions for other data.

2.2. Proposed System

Previous work in dengue forecast has lacked in considering various climatic factors. This system covers most of the drawbacks of the existing system and it helps the people to predict number of cases, provide regional statistics and provides an interface to check symptoms.

The advancement in the field of Machine learning has reduced the complexity of this task. Machine Learning technologies, have been rapidly developed in recent years, especially the feature selection algorithms like Recursive Feature elimination which select a subset of relevant features for further use in model construction and decision tree algorithm to train the model. It is promising to use the state-of-art Machine learning techniques to help people in the ease of dengue forecast. This system uses the DengAI dataset to uncover relations between various climatic factors and their dependence on number of cases of dengue.

Further, the user can check for symptoms for dengue and statistics of dengue effected people in their region. Statistics are provided with the information of user’s latitude and longitude. Furthermore, the user can check their past symptoms experienced on the designed system.

2.3. Advantages

- In this model, the use of recursive feature elimination eliminates the unnecessary features beforehand.
- Regional statistics are provided with user specific location.
- In terms of performance decision tree out performs other machine learning algorithms.
- The model provides an interface to check for symptoms for dengue.
- System has a user-friendly interface.

3. SYSTEM DESIGN

3.1. Architecture

Dengue prediction through climatic data is mainly used to predict the number of cases possible in a region according to the climatic data of the region. In this model the process takes input as climatic data of 7 parameters manually or though the user location. In addition to predicting the number of cases it also gives statistics of affected people around the location of user and also says about the possibility of dengue fever by considering the symptoms of the client.

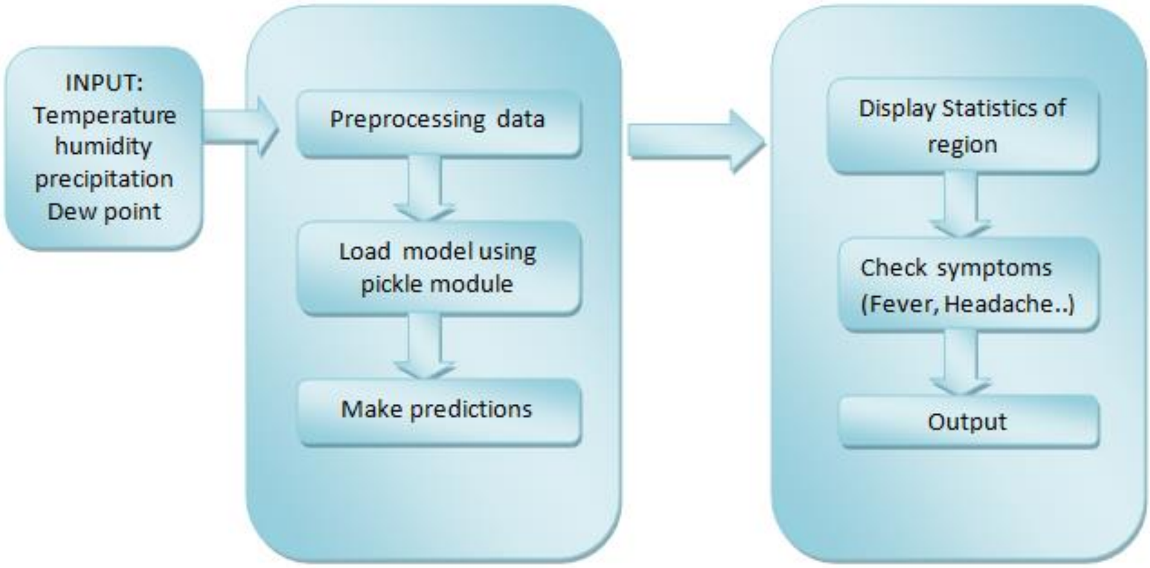

Fig 3.1.1 Architecture of the model

Training Data:

For dengue forecast model is trained using of DengAI: Predicting Disease Spread dataset.

Input Data:

For predicting number of cases, the input is taken from the user manually or through calculating their current location using latitude, longitude. This location data is used to collect weather data automatically by using Openweathermap One call API. The input to predict about fever is done by checking the symptoms through a check list

Output:

The climatic data taken as the input is given to a trained machine learning model to predict the number of cases. To say about the wellbeing status of the client the symptoms experienced by the client are processed and compared with standard side effects of dengue fever and dengue tall fever.

3.2. Data Preparation

Dengue forecast model is trained using of DengAI: Predicting Disease Spread dataset. The dataset has total of 21 independent variables and one dependent variable which is the output of number of cases.

```
week_start_date
ndvi_ne
ndvi_nw
ndvi_se
ndvi_sw
precipitation_amt_mm
reanalysis_air_temp_k
reanalysis_avg_temp_k
reanalysis_dew_point_temp_k
reanalysis_max_air_temp_k
reanalysis_min_air_temp_k
reanalysis_precip_amt_kg_per_m2
reanalysis_relative_humidity_percent
reanalysis_sat_precip_amt_mm
reanalysis_specific_humidity_g_per_kg
reanalysis_tdtr_k
station_avg_temp_c
station_diur_temp_rng_c
station_max_temp_c
station_min_temp_c
station_precip_mm
```

Figure 3.2.1 Independent variables in the data

All these independent variables contribute to rise or dip in cases of dengue. The outline of parameters in the dataset are listed below.

Temperature is a physical quantity that expresses hot and cold. It is the manifestation of thermal energy, present in all matter, which is the source of the occurrence of heat, a flow of energy, when a body is in contact with another that is colder or hotter. On a particular day Maximum, Minimum, Average, Diurnal temperature can be reported. Diurnal temperature is the difference between maximum and minimum temperature.

Humidity is the concentration of water vapor present in the air. Humidity depends on the temperature and pressure. **Relative humidity**, expressed as a percentage, indicates a present state of absolute humidity relative to a maximum humidity given the same temperature. **Specific humidity** is the ratio of water vapor mass to total moist air parcel mass.

Precipitation is any product of the condensation of atmospheric water vapor that falls under gravity from clouds the main forms of precipitation include drizzle, rain, sleet, snow, ice pellets, graupel and hail. Precipitation occurs when a portion of the atmosphere becomes saturated with water vapor (reaching 100% relative humidity), so that the water condenses and "precipitates" or falls.

The **dew point** is the temperature to which air must be cooled to become saturated with water vapor. When cooled further, the airborne water vapor will condense to form liquid water (dew).

3.3. Dengue Forecast

The machine learning algorithms used for estimating number of dengue cases that are included in sklearn are given as follows.

Feature selection algorithms:

In machine learning and statistics, feature selection is the process of selecting a subset of relevant features (variables, predictors) for use in model construction.

Feature selection techniques are used for several reasons like simplification of models to make them easier to interpret by researchers/users, shorter training times, to avoid the curse of dimensionality, reduce overfitting.

The following algorithms are used for feature selection:

Recursive feature elimination:

This is a wrapper-based method. The goal of recursive feature elimination (RFE) is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained either through a coef_ attribute or through a feature_importances_ attribute. The estimator used here is Multiple Linear Regression Then, the least important features are pruned from current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to select is eventually reached.

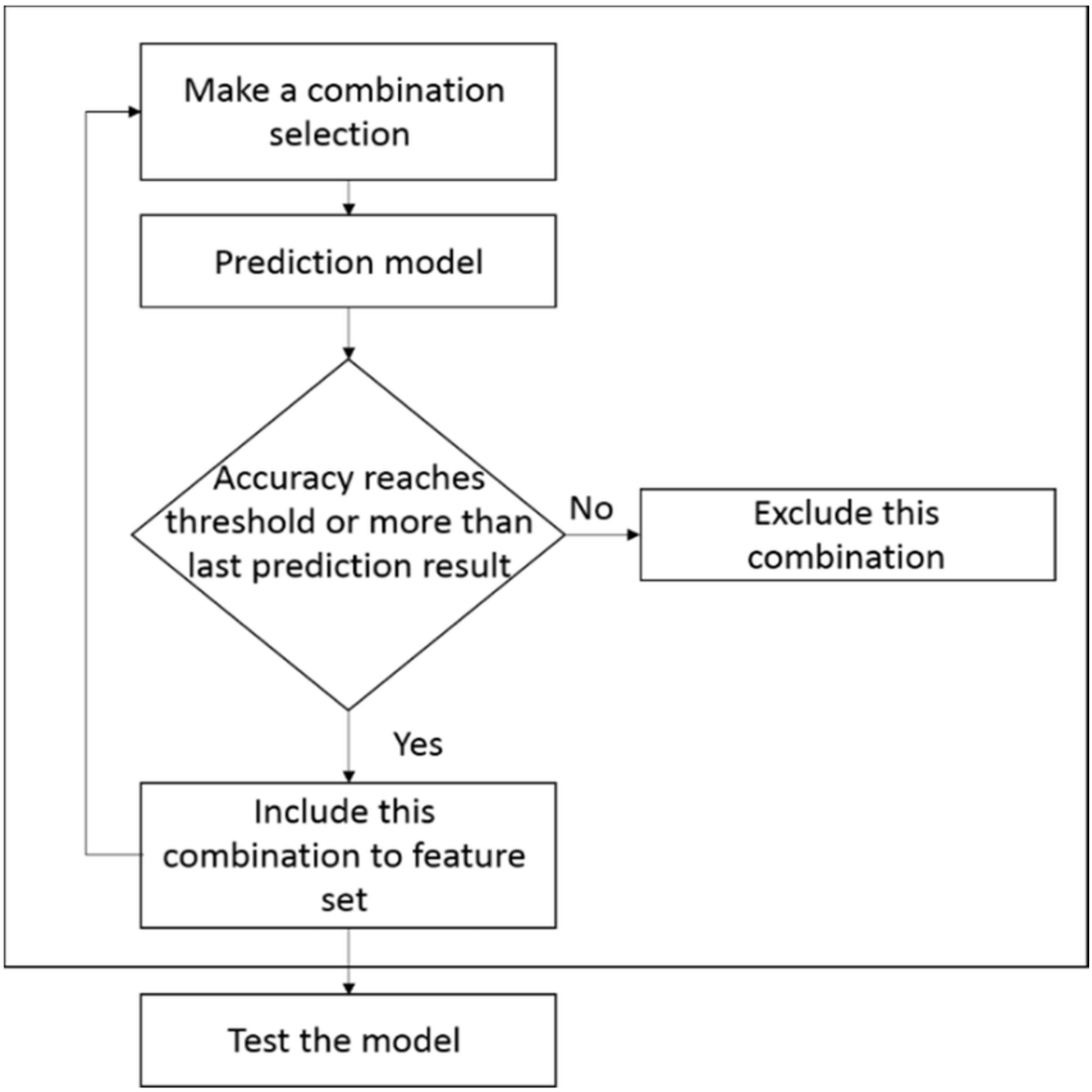

Figure 3.3.1 Recursive feature elimination procedure

Pearson Correlation: This is a filter-based method where the absolute value of the Pearson's correlation between the target and numerical features in our dataset. We keep the top n features based on this criterion.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Figure 3.3.2 Pearson's corelation formula

Common features from both the feature selection algorithms are selected

Features selected from Pearson's Correlation and Recursive feature elimination
reanalysis_specific_humidity_g_per_kg
reanalysis_relative_humidity_percent
reanalysis_dew_point_temp_c
reanalysis_avg_temp_c
reanalysis_max_air_temp_c
reanalysis_air_temp_c
reanalysis_tdtr_k

Algorithms for Model training

1) **Linear Regression:**

   Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and has a constant slope. It's used to predict values within a continuous range, (e.g., sales, price) rather than trying to classify them into categories (e.g., cat, dog).

In this model multiple linear regression (MLR) is used to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. This algorithm is used in feature selection for ranking top features in the dataset.

2) **K nearest neighbors:**

K nearest neighbors is a simple algorithm that stores all available cases and predict the numerical target based on a similarity measure (e.g., distance functions). A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbors.

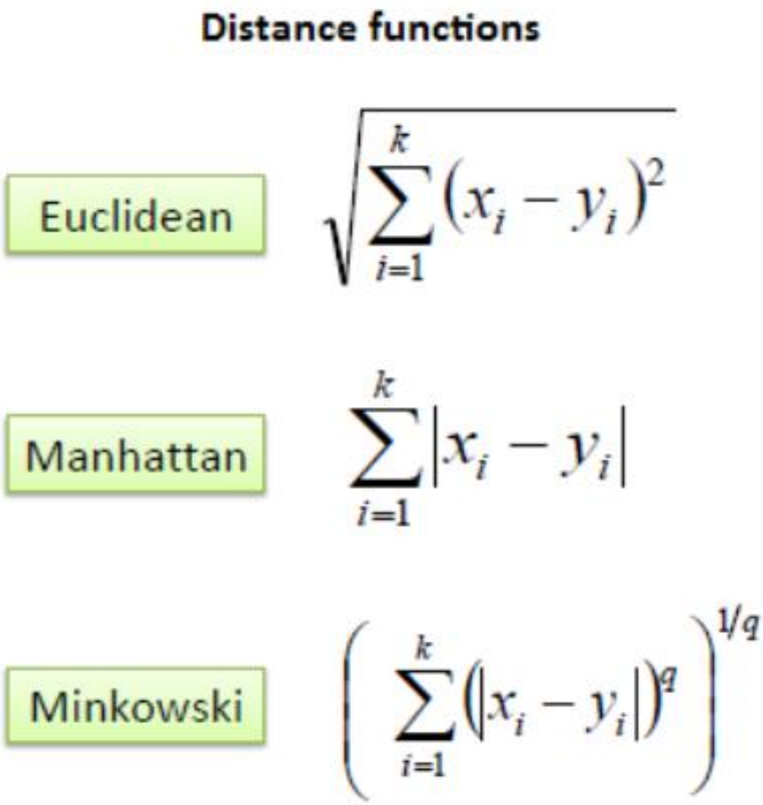

Figure 3.3.3 Distance functions in K-nearest neighbours

3) **Support vector machines:**

Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification, with only a few minor differences.

First of all, because output is a real number it becomes very difficult to predict the information at hand, which has infinite possibilities. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem.

4) **Decision Tree:**

Decision tree builds regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches each representing values for the attribute tested. Leaf node represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Among all the algorithms Decision Tree outperforms the others.

The model is converted to pickle file for the ease of use by the client or the user. Pickle operation is used to serialize your machine learning algorithms and save the serialized format to a file. This converted file is stored in some directory. Later the model is loaded using which the predictions are to be made on the input/data.

3.4. Dengue symptoms identification

Life threatening complication can occur rapidly in delay of proper medical care. In a study it is revealed that delay in medical treatment is found to be significantly associated with complications leading to severe dengue. World Health Organization indicated that early detection of DF and appropriate treatment will decrease fatality rates from more than 20% to less than 1%.

There are many differences between Dengue fever and severe dengue. The main difference is symptoms of the both the fevers are different. The type of symptom the user is experiencing and the number of symptoms user is feeling are used to know the type of dengue fever user may have chance of getting affected. The symptoms taken from the user and the duration of symptoms are used to determine the result.

Dengue Fever	Dengue High Fever
High fever	Swollen glands
Headaches	Severe abdominal pain
Pain behind the eyes	Bleeding gums
Severe joint and muscle pain	Vomiting blood
Nausea	Rapid breathing
Vomiting	Fatigue
Skin rash	
No appetite	

Figure 3.4.1 Symptoms for respective type of dengue

The system flow chart represents the general processes of this system. This flow chart roles as a guideline for the dengue patients' diagnostic system and notify whether they need to seek for medical attention.

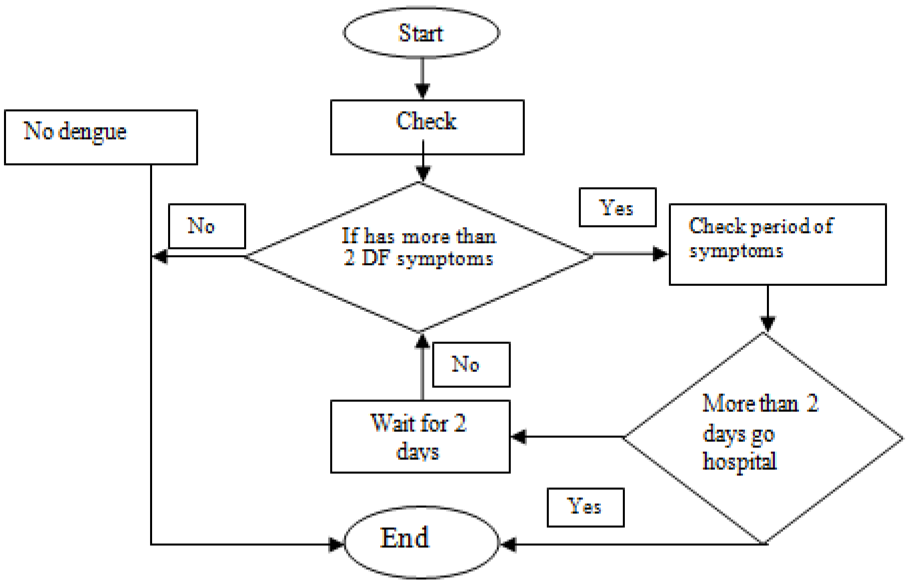

Figure 3.4.2 Flowchart for type of dengue identification

General rule:

High fever symptom + more than 1 other symptoms + Period of infection more 2 days = Positive

Else = Negative

Figure 3.4.3 General rule of dengue

The general rule seen above is used to calculate the symptoms. Two experienced doctors have been interviewed for getting the correct knowledge about dengue. The time symptoms are seen is included as the chances of getting affected increases by 10% if the symptoms are seen from more than 2 days.

3.5. Statistics of Dengue affected people

One of factor that greatly influences dengue is location. There are many locations where disease is common in many popular tourist destinations in the Caribbean (including Puerto Rico), Central and South America, Southeast Asia, and the Pacific Islands. In the United States, local cases and limited spread of dengue does occur periodically in some states with hot, humid climates and Aedes mosquitoes.

In geography, latitude is a geographic coordinate that specifies the north–south position of a point on the Earth's surface. Latitude is an angle (defined below) which ranges from 0° at the Equator to 90° (North or South) at the poles and is a geographic coordinate that specifies the east–west position of a point on the Earth's surface, or the surface of a celestial body. Lines of constant latitude, or parallels, run east–west as circles parallel to the equator. Latitude is used together with longitude to specify the precise location of features on the surface of the Earth

In this model, the statistics of people who are affected by dengue are provided to the user based on their location. Latitude and longitude of the user is used to calculate distances from other users using the model. The **haversine formula** determines the great-circle distance between two points on a sphere given their longitudes and latitudes. Important in navigation, it is a special case of a more general formula in spherical trigonometry, the law of haversines, that relates the sides and angles of spherical triangles.

The haversine formula allows the haversine of Θ (that is, hav(Θ)) to be computed directly from the latitude and longitude of the two points: To solve for the distance d, apply the arc haversine (inverse haversine) to h = hav(Θ) or use the arcsine (inverse sine) function:

$$\begin{aligned} d &= 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\operatorname{hav}(\lambda_2 - \lambda_1)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \end{aligned}$$

Figure 3.5.1 Haversine Formula for distance calculation

Here r=6371 for calculating distance in kilometre.

4. IMPLEMENTATION

4.1. Technology used

4.1.1. Python

Python features a dynamic type system and automatic memory management. It supports multiple Programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library. Python interpreters are available for many operating systems. Python, the reference implementation of Python, is open-source software and has a community-based development model, as do nearly all of its variant implementations.

4.1.2. Python Packages

Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Num array into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatterplots, etc., with just a few lines of code.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

Sklearn

Scikit-learn (formerly scikit-learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Sqlite3

SQLite is a C library that provides a lightweight disk-based database that doesn’t require a separate server process and allows accessing the database using a nonstandard variant of the SQL query language. Some applications can use SQLite for internal data storage. It’s also possible to prototype an application using SQLite and then port the code to a larger database such as PostgreSQL or Oracle.

Geocoder

Geocoding is the computational process of transforming a physical address description to a location on the Earth’s surface (spatial representation in numerical coordinates). Geocoder is a simple and consistent geocoding library written in Python.

4.1.3. Flask

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. The microframework Flask is based on the Pocoo projects, Werkzeug and Jinja2.

Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

4.1.4. Bulma CSS

Bulma CSS is a free, open-source framework that provides ready-to-use frontend components that you can easily combine to build responsive web interfaces.

4.1.5. HTML

Hypertext Mark-up Language (HTML) is the standard mark-up language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets (CSS) and scripting languages such as JavaScript.HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets.

4.1.6. Anaconda

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 13 million users and includes more than 1400 popular data-science packages suitable for Windows, Linux, and MacOS.

- **Jupyter Notebook**

Jupyter Notebook is an open-source cross-platform IDE for scientific programming in the Python language. It comes installed with anaconda. If not, install it using anaconda navigator.

4.1.7. Openweathermap Onecall API

An application programming interface (API) is a computing interface that defines interactions between multiple software intermediaries. It defines the kinds of calls or requests that can be made, how to make them, the data formats that should be used, the conventions to follow, etc. One API call can be used get current, forecast and historical weather data. Weather data can be Minute forecast for 1-hour, Hourly forecast for 48 hours and Daily forecast for 7 days

4.2. UML Diagrams

4.2.1. Use Case Diagram:

Use case diagram are used for modelling the behaviour of a system, subsystem, or a class. They depict the dynamic aspects of a system. They are important for visualizing, specifying, and documenting the behaviour of the system. The Use case diagram for prediction of number of cases is given below.

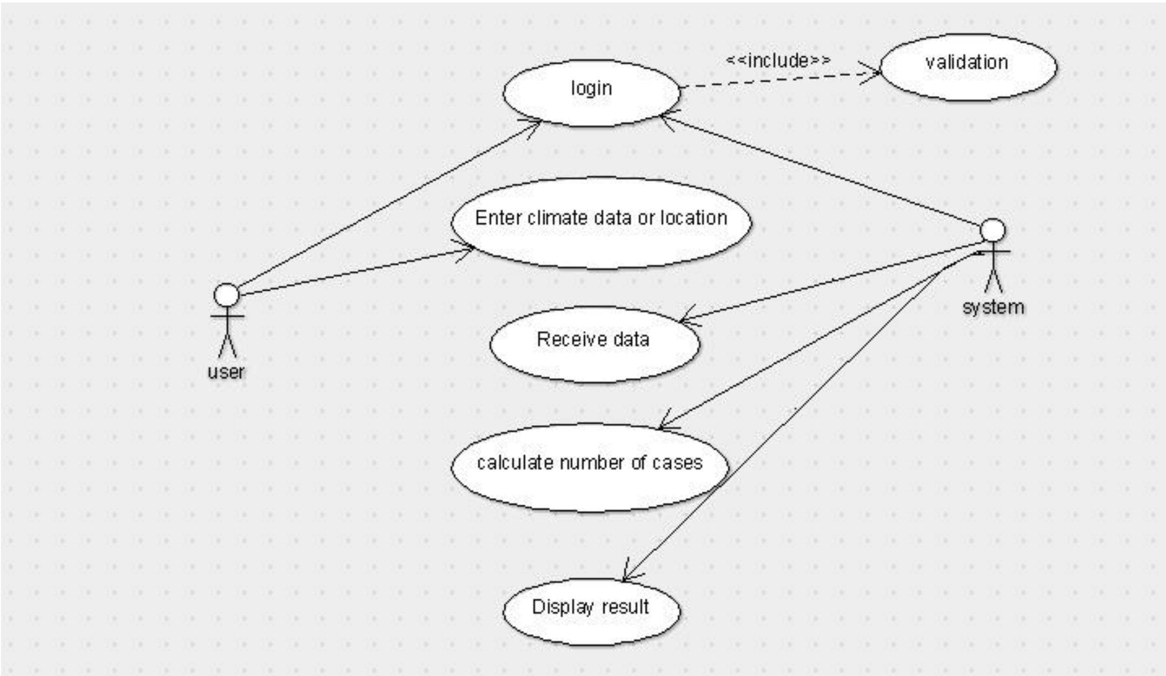

Figure 4.2.1 Use case for dengue forecast

Actors:

User: A user may be person requesting services from the system.

System: The system is a combination of hardware and software components, which has the capability to predict the dengue cases from climate data.

Use cases:

Login use case: In Login use case user, gives login details to start the session.

Validation use case: In Validation use case user, credentials are validated by checking data in database.

Enter climate data or location use case: In Enter climate data or location use case, users can enter the climatic data of location manually or can enter latitude and longitude of the location.

Receive data use case: In Receive data use case the data given by the user is sent to system.

Calculate number of cases use case: In Calculate number of cases use case from the received data system calculates the number of cases using machine learning model.

Display result use case: After calculating the number of cases, the system displays it to the user.

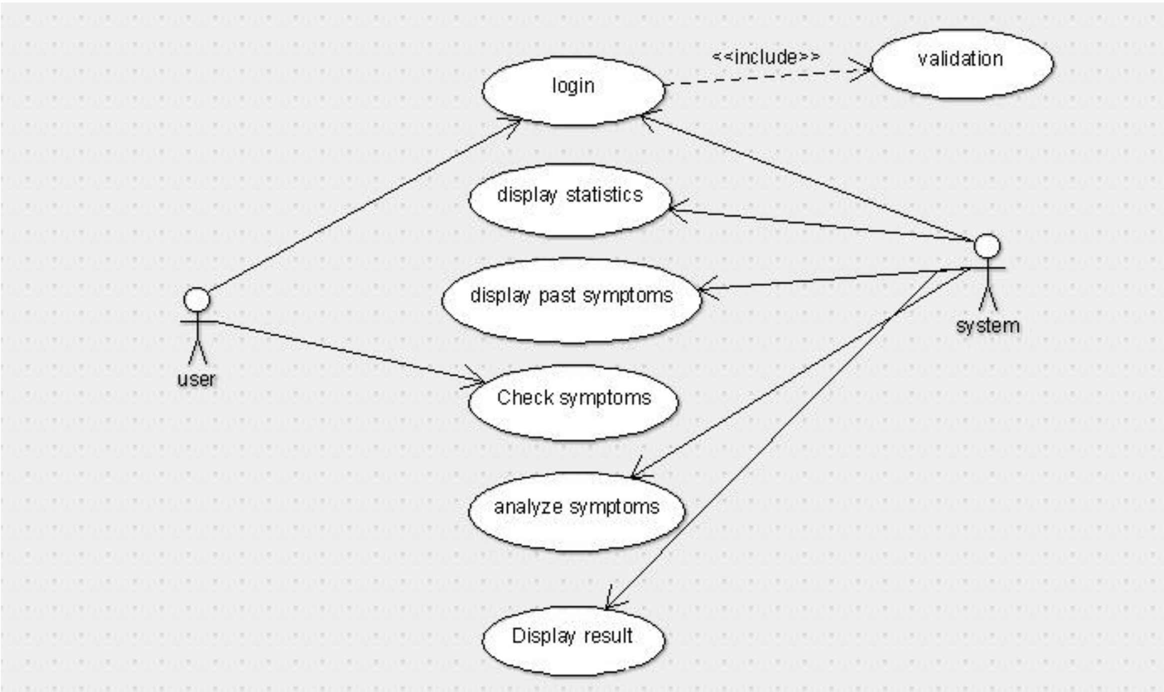

Figure 4.2.2 Use case for dengue symptom identification

Actors:

User: A user may be person requesting services from the system.

System: The system is a combination of hardware and software components, which has the capability to predict the dengue from symptoms data given by user.

Use cases:

Login use case: In Login use case user, gives login details to start the session.

Validation use case: In Validation use case user, credentials are validated by checking data in database.

Display statistics use case: It shows the number of persons affected in the location range of user.

Display past symptoms use case: In Display past symptoms use case it shows the user their previous history of symptoms.

Check symptoms use case: In Check symptoms use case the check list with different symptoms is checked by the user.

Analyse symptoms use case: In Analyse symptoms use case the symptoms checked by the user are analysed according to model.

Display result use case: In Display result use case, the result after analysing the symptoms is displayed to the user.

4.2.2. Sequence Diagram:

Sequence diagram emphasis on time ordering of messages is shown in the figure 4.2.3 and figure 4.2.4 for the model.

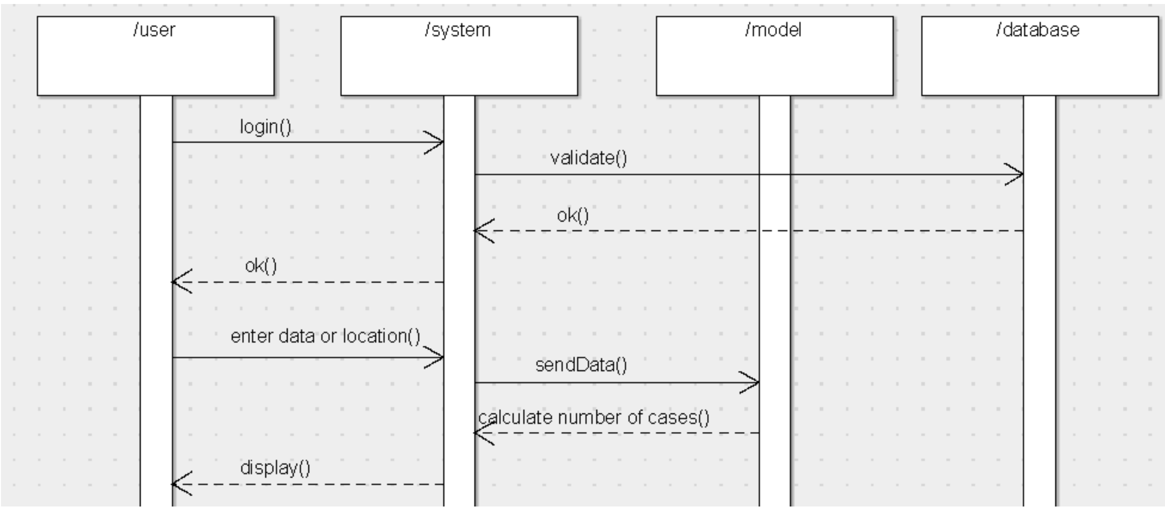

Fig 4.2.3 Sequence diagram for dengue forecast

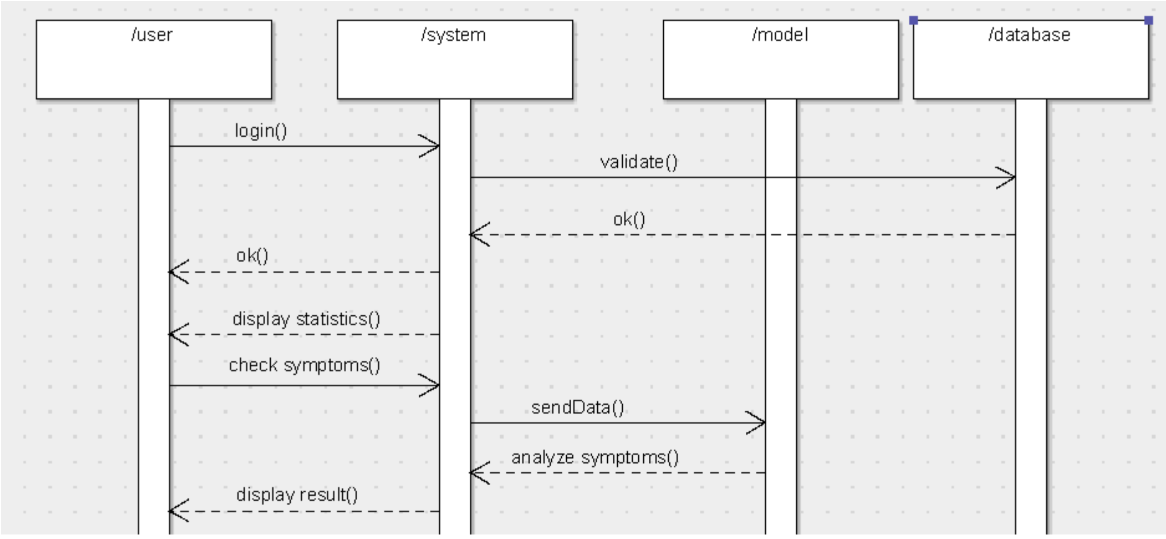

Fig 4.2.4 Sequence diagram for dengue symptom identification

4.2.3. Activity Diagram:

Activity diagram is essentially a flow chart, showing flow of control from activity to activity. Figure 4.2.5 and 4.2.6 shows the activity diagram for the model.

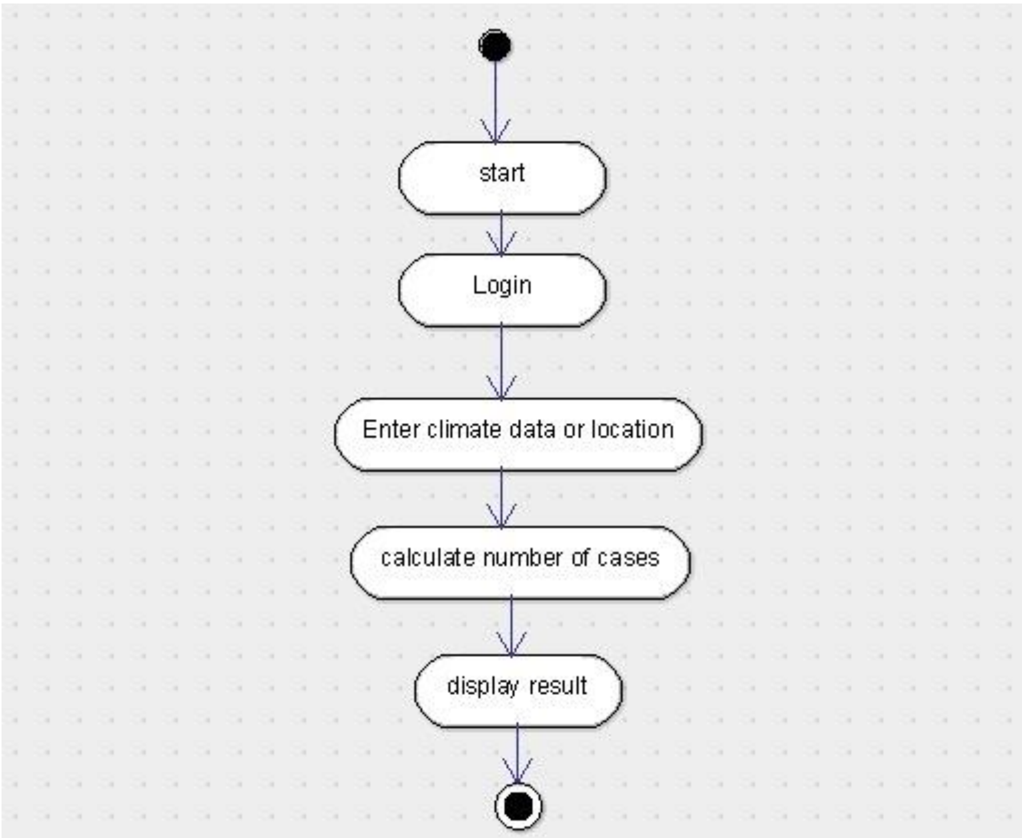

Fig 4.2.5 Activity diagram for dengue forecast

Action States:

Start: The system loads the webpage as it gets started.

Login: The user gives login credentials which are verified against database.

Enter climate data or location: Enter the climatic data manually or through latitude and longitude.

Calculate number of cases: According to the climatic data given number of cases are calculated.

Display result: After calculating the number of cases using the model the result is displayed.

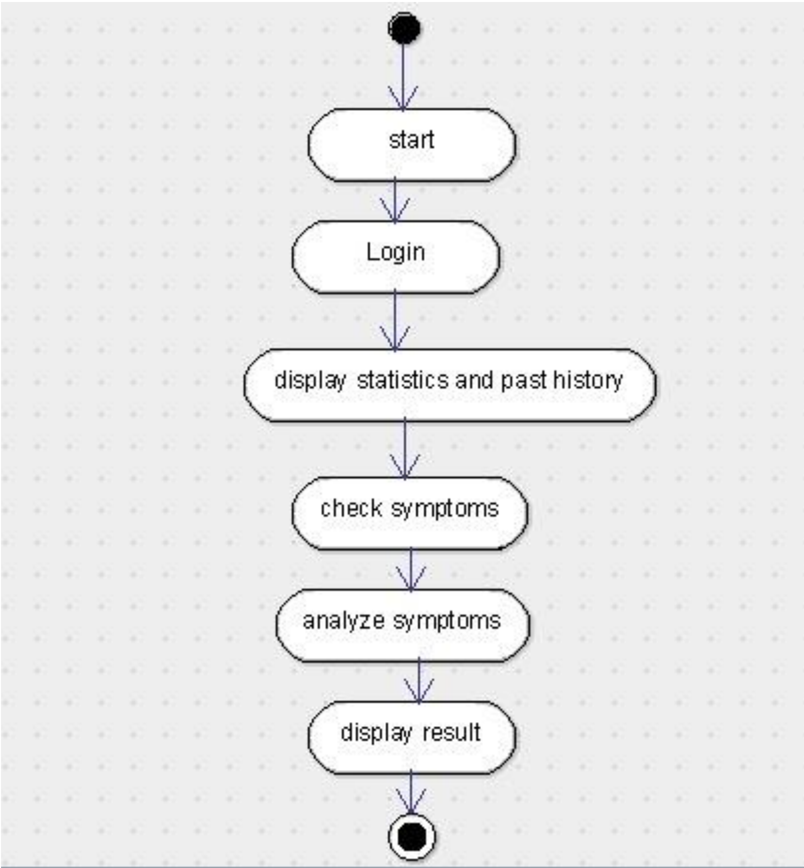

Fig 4.2.6 Activity diagram for dengue symptom identification

Action states for figure 4.2.6:

Start: The system loads the page as it gets started.

Login: The user gives login credentials which are verified against database.

Display statistics and past history: The number of people affected at nearby location of user is displayed and the past symptoms history of the user is displayed in a table.

Check Symptoms: The user should check the check list of different symptoms listed.

Analyse Symptoms: The given symptoms are analysed according to the model.

Display result: After analysing the symptoms the result is displayed to the user.

5. RESULTS

5.1. Discussion on results

In the Implementation the dengue forecast from climate data the dataset of DengAI has been used. Different models were used for training and testing the data.

Model Evaluation was performed by calculating mean absolute error (MAE), Root mean squared error (RMSE), r2 score. Further, the predicted and test values are plotted to validate the model's performance for optimization.

Algorithm name	Mean absolute error	Root mean squared error	R2_Score
Linear Regression	22.46	51.86	0.1
KNeighbours	7.26	11.68	-0.25
Support vector regression	6.32	10.98	-0.1
Decision Tree	0.03	0.34	0.99

Figure 5.1.1 Metrics used and their results on given algorithms

Mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. The Root mean squared error (RMSE) represents the square root of the square of the differences between predicted values and observed values.

R2 score (coefficient of determination) is regression score function. Best possible score is 1.0. It is the amount of the variation in the output dependent attribute which is predictable from the input independent variable(s).

From the above figure, it is clear thar decision tree is the top performer with greater than zero r2 score and less error is observed in predicted and test values compared to that of other models. Decision is slightly better due to its r2 score being close to 1.

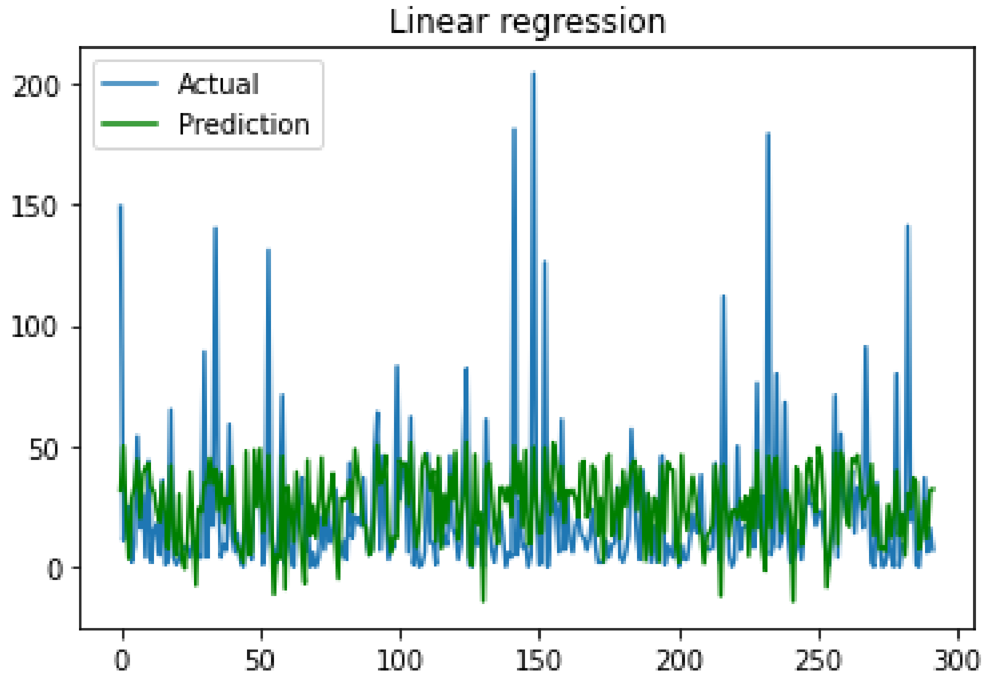

Figure 5.1.2 Performance of Linear Regression Actual vs Prediction

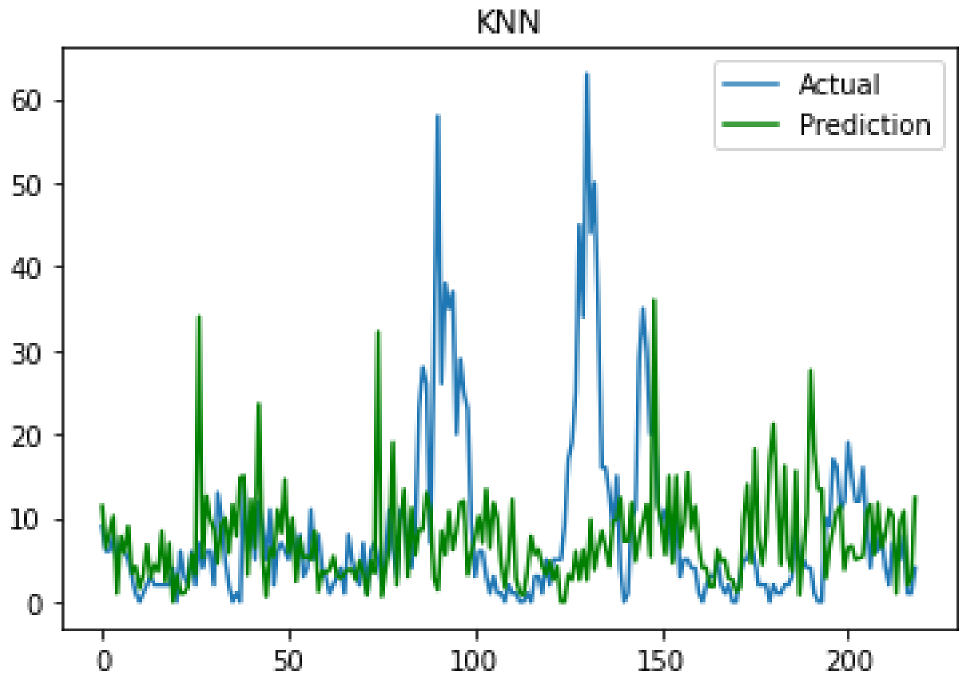

Figure 5.1.3 Performance of KNN Actual vs Prediction.

The above graphs are for Actual vs Predicted number of cases for linear regression and KNN.

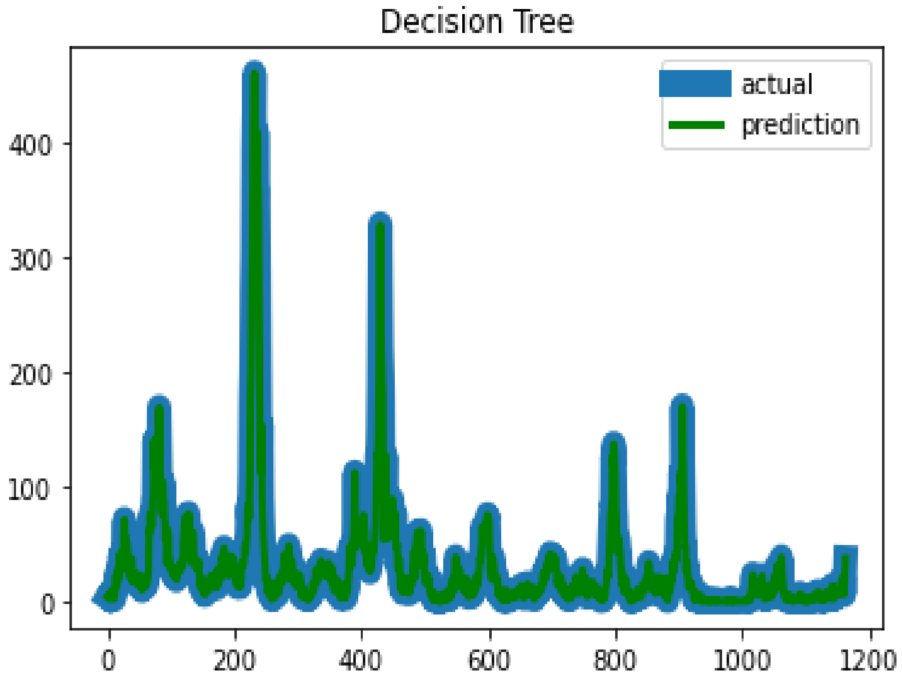

Figure 5.1.4 Performance of Decision tree Actual vs Prediction

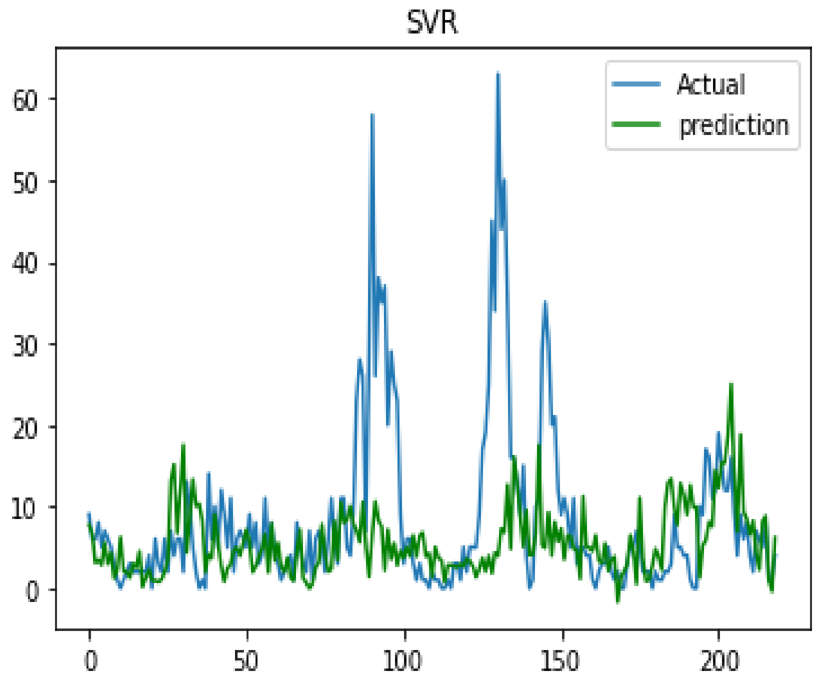

Figure 5.1.5 Performance of SVR Actual vs Prediction.

The above graphs are for Actual vs Predicted number of cases for Decision Tree and SVR. Decision Tree is almost predicting the cases correctly whereas the next best model SVR is able to predict low spikes in number of cases.

The above all figures are plotted using matplotlib where green colour depicts predicted data and blue colour is test data. Fig 5.1.4 is of decision tree where test and predicted values almost overlap each other which means that less error is occurred.

For displaying statistics of affected people in a region latitude and longitude are collected using geocoder package of python and displayed accordingly. In analysing symptoms of user, the model clearly distinguishes the input received and acts accordingly.

5.2. Screenshots

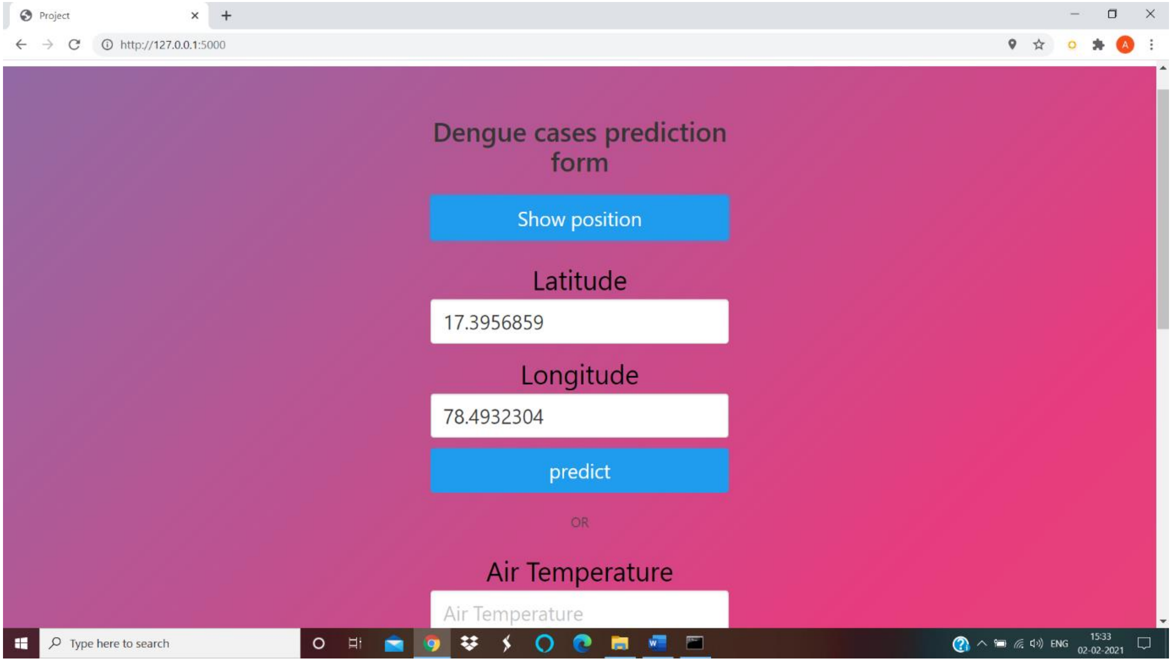

Figure 5.2.1 Input of latitude and longitude

In Figure 5.2.1 user is giving their latitude and longitude so system can automatically collect weather data for the given location.

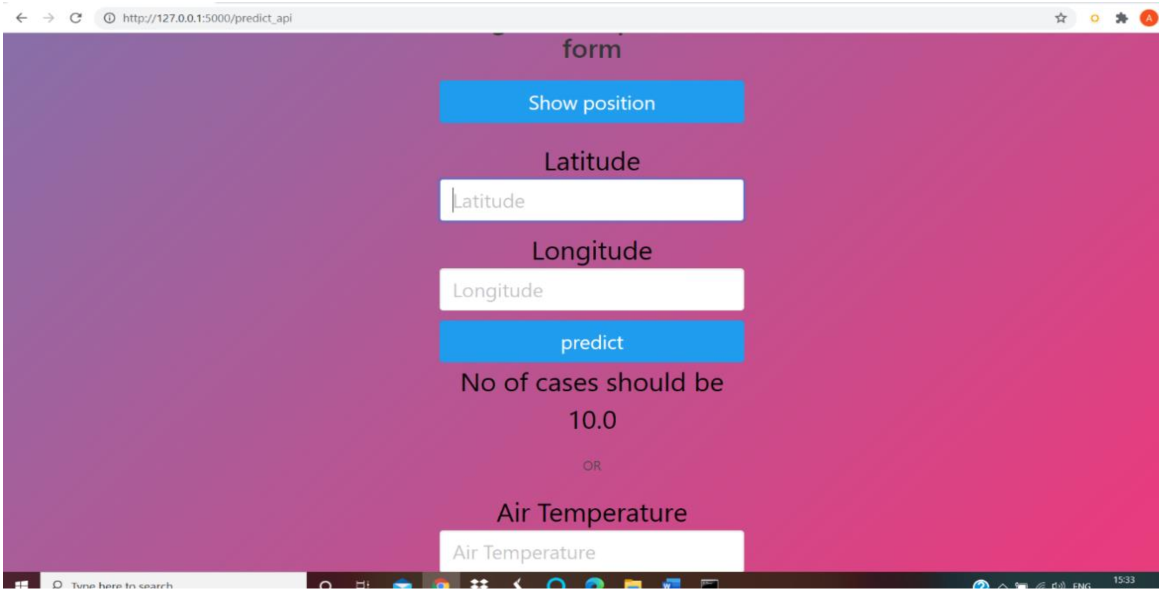

Figure 5.2.2 Output of automatic input of weather data

Figure 5.2.2 shows output of estimated number of cases for automatic weather data based on location.

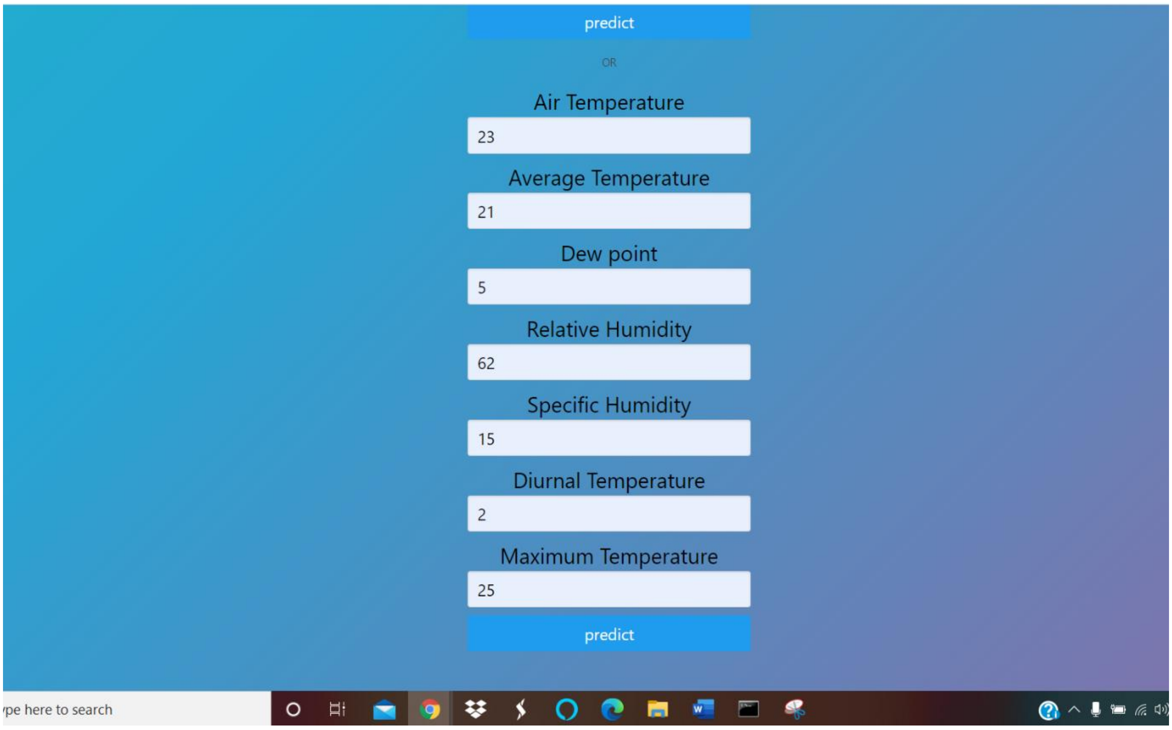

Figure 5.2.3 Manual Entering of climate data

In Figure 5.2.3 user enters climate data manually

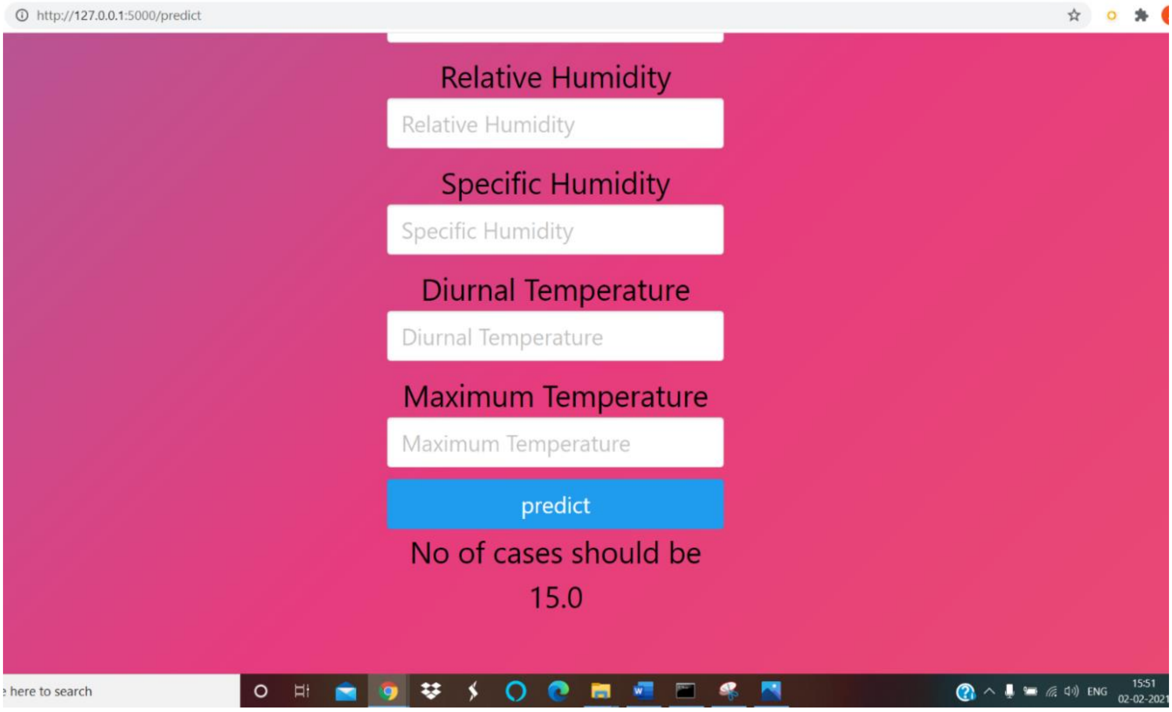

Figure 5.2.4 Output of manual climate data

Figure 5.2.4 shows output of estimated number of cases for manual weather data.

Figure 5.2.5 Statistics of affected people around a region

Figure 5.2.5 shows the statistics of affected people around the region of user using haversine formula

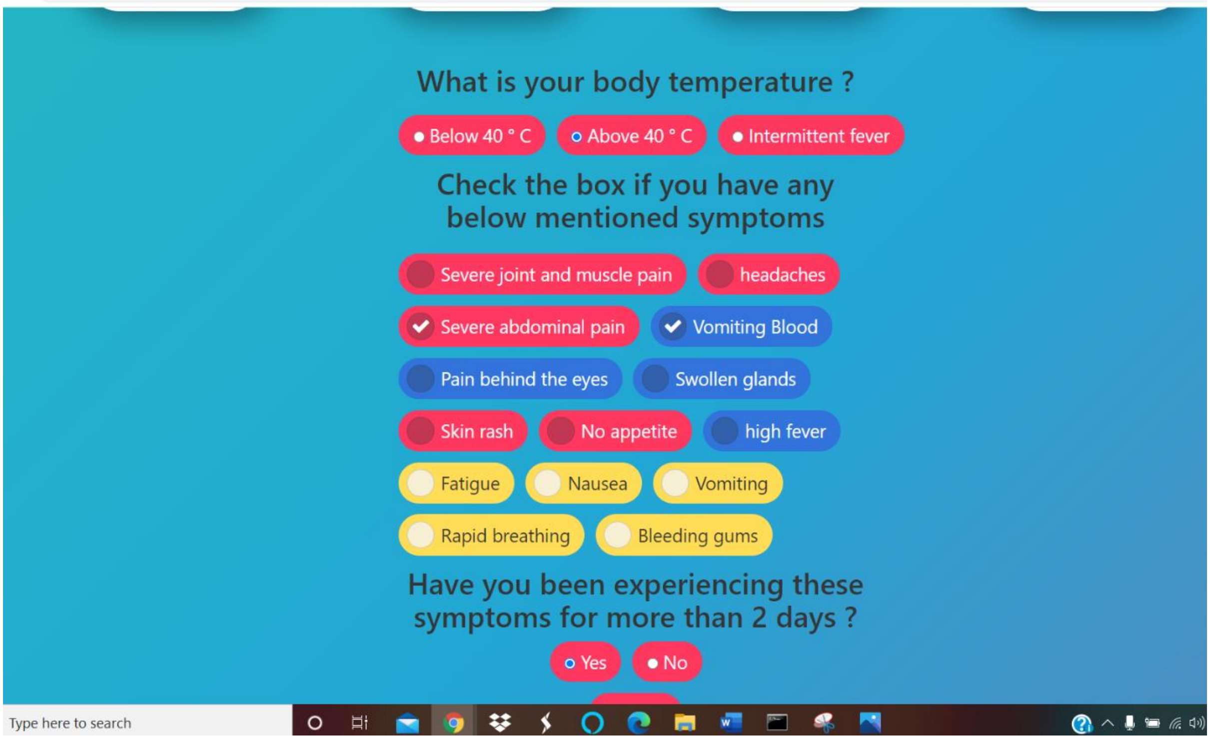

Figure 5.2.6 Checklist given to user for symptoms analysis

In Figure 5.2.6 user checks off the symptoms to check for dengue and severe dengue.

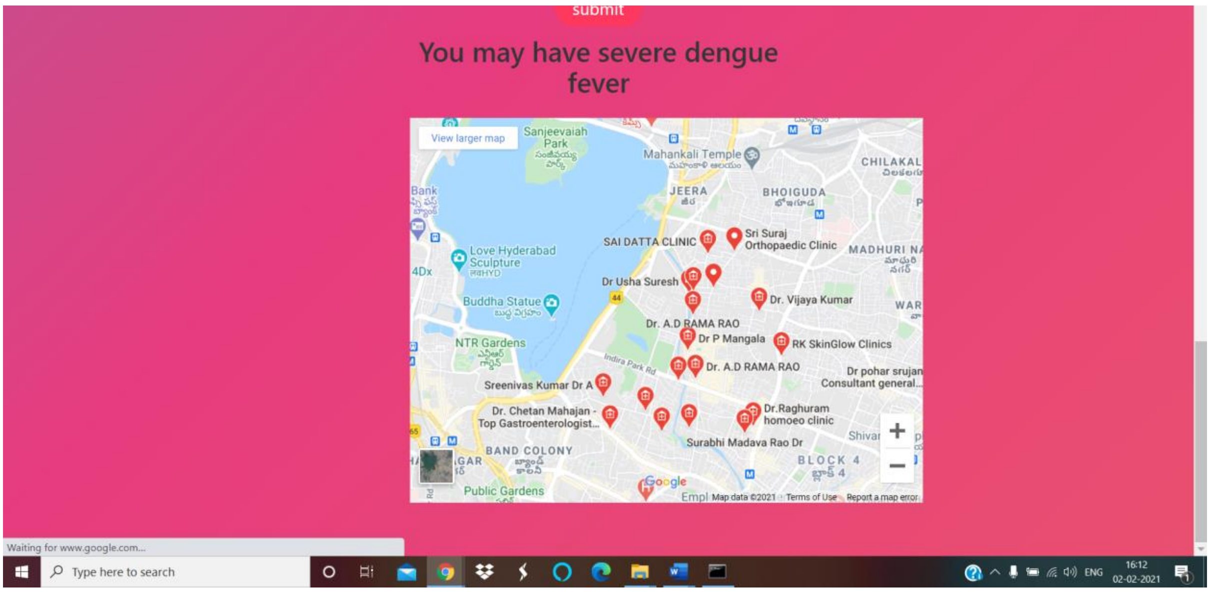

Figure 5.2.7 Output after symptom analysis

In Figure 5.2.7 output is displayed accordingly based on the given user symptoms. Here user is diagnosed with severe dengue and map for nearby doctors for further check-up is provided

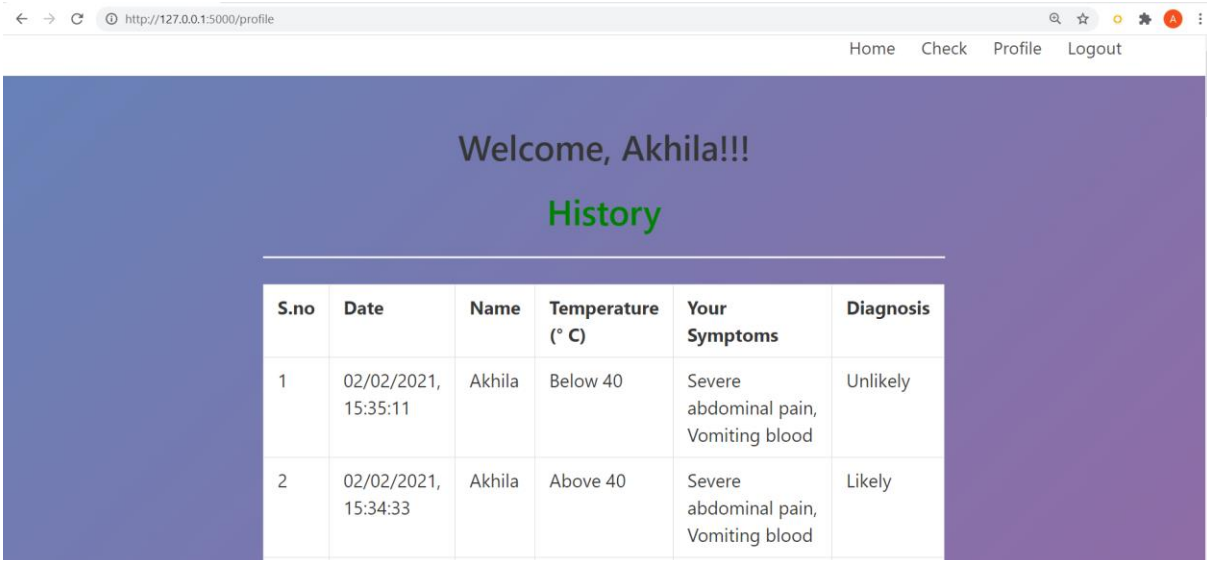

Figure 5.2.8 User symptom history

Figure 5.2.8 shows the user can check all their past experienced symptoms

6. CONCLUSION AND FUTURE WORK

6.1. Conclusion

Accurate dengue predictions would help public health workers and people around the world take steps to reduce the impact of these epidemics. But predicting dengue is a hefty task that calls for the consolidation of different data sets on disease incidence, weather, and the environment. As humans have littered the environment with plastic containers, which provide an ideal breeding ground for mosquitoes the chances for dengue cases increased. The prediction of number of cases to be cautious and checking the symptoms to prevent the risks because of late analysis are important. This system predicts number of cases according to the weather data like temperature, humidity, precipitation using a machine learning algorithm. It also provides an interface to check the health status along with the statistics of affected people in the neighbourhood.

6.2. Future Work

The high awareness of society to maintain appropriate sanitation will make the chances of mosquitoes breed into small. It will certainly reduce the mosquito population. Reduced mosquito populations may decrease the probability of an outbreak. So, Hygiene factors can be considered in future work

6.3. References

[1] A. F. B. Gabriel, A. P. Alencar, and S. G. E. K. Miraglia, “Dengue outbreaks: unpredictable incidence time series,” Epidemiol. Infect., vol. 147, p. e116, 2019, doi: 10.1017/S0950268819000311.Felestin Yavari Nejad, Kasturi Dewi Varathan

[2] Felestin Yavari Nejad, Kasturi Dewi Varathan, “Identification of Significant Climatic Risk Factors and Machine Learning Models in Dengue Outbreak Prediction,” In Review, preprint, Oct. 2019. doi: 10.21203/rs.2.15755/v1.

[3] Li C, Wang X, Wu X, et al. “Modeling and projection of dengue fever cases in Guangzhou based on variation of weather factors”[J]. Science of the Total Environment, 2017, 605-606:867- 873.

[4] Mustaffa Z, Yusof Y. “A Comparison of Normalization Techiques in Predicting Dengue Outbreak”[J]. International Proceedings of Economics Development & Research, 2011.

[5] Pham D N, Aziz T, Kohan A, et al. “An Efficient Method To Predict Dengue Outbreaks In Kuala Lumpur”[c], International Conference on Artificial Intelligence and Computer Science, 2015.

[6] P. Siriyasatien, S. Chadsuthi, K. Jampachaisri, and K. Kesorn, “Dengue Epidemics Prediction: A Survey of the State-of-the-Art Based on Data Science Processes,” IEEE Access, vol. 6, pp. 53757– 53795, 2018, doi: 10.1109/ACCESS.2018.2871241.

[7] T.-W. Chuang, L. F. Chaves, and P.-J. Chen, “Effects of local and regional climatic fluctuations on dengue outbreaks in southern Taiwan,” PLOS ONE, vol. 12, no. 6, p. e0178698, Jun. 2017, doi: 10.1371/journal.pone.0178698.

[8] W. Farooqi, and S. Ali, “A Critical Study of Selected Classification Algorithms for DF and DHF”, Proceedings of the 2013 11th International Conference on Frontiers of Information Technology, (IEEE, 2013), p. 140-145.

[9] Y.-H. Lai, “The climatic factors affecting dengue fever outbreaks in southern Taiwan: an application symbolic data analysis,” Nov 2018.

[10] Youngjo C, Tang C S, Lachlan M I, et al. “Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia”:[J]. Bmc Public Health 2016, 16(1):241.

[11] Jia Nan, Xianyi Liao, Jie Chen, Xiangping Chen, Jing Chen, Guanghui Dong, Kangkang Liu, Gang Hu, “Using Climate Factors to Predict the Outbreak of Dengue Fever”,2018 7th International Conference on Digital Home (ICDH)

[12] Wiwik Anggraeni, Surya Sumpeno, Eko Mulyanto Yuniarno, Reza Fuad, Rachmadi, Agustinus Bimo Gumelar, Mauridhi H Purnomo, “Prediction of Dengue Fever Outbreak Based on Climate Factors Using Fuzzy-Logistic Regression”, 2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)