DATA ANALYTICS MEASURE FOR REPORTING DISEASE OUTBREAK IN A REGION

[1]Dr.T.Subha, [2]Sindu R, [3]Mahalakshmi R, [4]Nivedaa A
[1]Asso.Prof, Sri Sai Ram Engineering College,
[2,3,4] B.Tech Information Technology, Sri Sai Ram Engineering College
[5]Software Engineer, Costar Group, Washington DC

ABSTRACT:

Big data analytics is the process of evaluating large and varied data sets to uncover information that includes hidden patterns, unknown correlations, market trends and preferences of the customers which could be helpful for the organizations to make informed business decisions. With the use of Analytic system and software, the Big data analytics offers various business benefits which includes new revenue opportunities, more effective marketing ,best customer services, better operational efficiency and competitive advantages. It is a new era of modern technology. This project deals with mosquito borne diseases that rises seasonally and regionally. Nearly 700 million people get affected due to mosquito born illness each year resulting in over one million deaths. Numerous patients lost their lives because of unhygienic conditions in their surroundings that lead to mosquito-borne and diseases that are caused due to unhygienic conditions. To overcome this problem, the data of patients from numerous hospitals from various localities is been collected and then the following data is been processed on the data analytics platform and resultant information is used to identify the disease outbreak in a particular region. The NGO and the higher officials of that particular region uses the output to take necessary steps to prevent the spread of diseases.

Keywords: Accuracy, Classification, Dataset, Machine learning, Python, Prediction

INTRODUCTION:

The project is based upon the big data analytics where the details of patients are collected from several hospitals across various locations. The processing of the data is done with the help of machine learning algorithm which is a platform independent one. The ultimate aim of the project is to collect patient details and to categorize their details based upon the locations. The location helps to easily identify the areas which are getting affected the most. The commonly affected areas are due to improper hygienic facilities and hence they should be concentrated to clean more in order to avoid certain diseases which are caused due to improper hygiene. The reports that are gathered are given to several NGO's and other organisations who are responsible for cleaning that area. The python language provides methods foranalytics and the details are analysed and reports have been generated. A webpage is been created at the front end where the login and disease details are provided and they are correspondinglyshown to the respected authorities.

RELATED WORKS:

The existing system makes use of MATLAB which is used for analytical process. The MATLAB is integrated with thingspeak which provides the analytical data in form of graphs. The major drawback of MATLAB is it does not process unstructured data. The data that is fed inside the MATLAB must be structured and static where its well organized.Another drawback is it has a lack of specificity where it is a primary limitation of this system. It provides limited evidence of the ability of the systems to detect emerging threats before signals from more traditional system. Data sources don't provide in-depth evaluation, especially with respect to false

positives and gaps in coverage. The major disadvantage is MATLAB software is dependent on a particular platform.

PROPOSED SYSTEM

The various hospitals from different location would be collected in the back end. This would involve collecting each and every person information who were been admitted in the hospital for viral and mosquito borne diseases would be tracked with the help of location of the hospital in which they are admitted. The data entry of each and every person who was admitted in the hospital is must for this project . These information updating is done with the help of excel sheet. And various excel sheet is imported to machine learning to analyze the data efficiently. This would analyze the various details and would give a graphical representation of the mosquito affected area and the number of people who have been affected for the same infection . These details would be fed regularly using the web application portal so that the NGO who are responsible for that locality would do the needful measure such as cleaning the area , stagnant water cleanup in that locality etc.... And this website is also open for the regular public, so that if any volunteers were ready for the cleanup activity, it could also be helpful in serving and making the environment clean

ARCHITECTURE DIAGRAM:
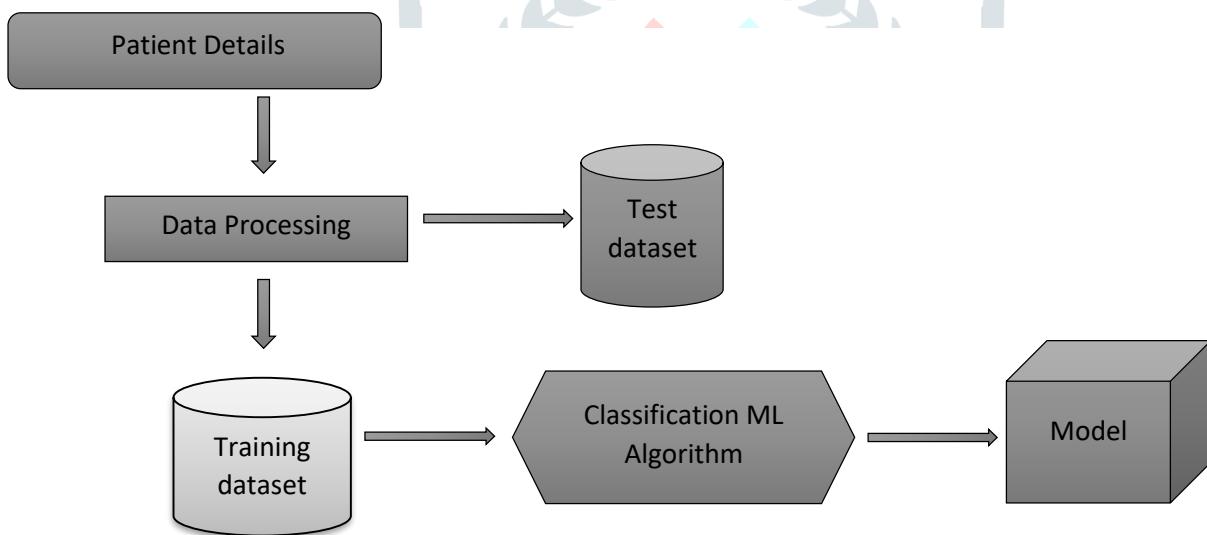
Fig.1. Architecture of Proposed System

SYSTEM ARCHITECTURE:
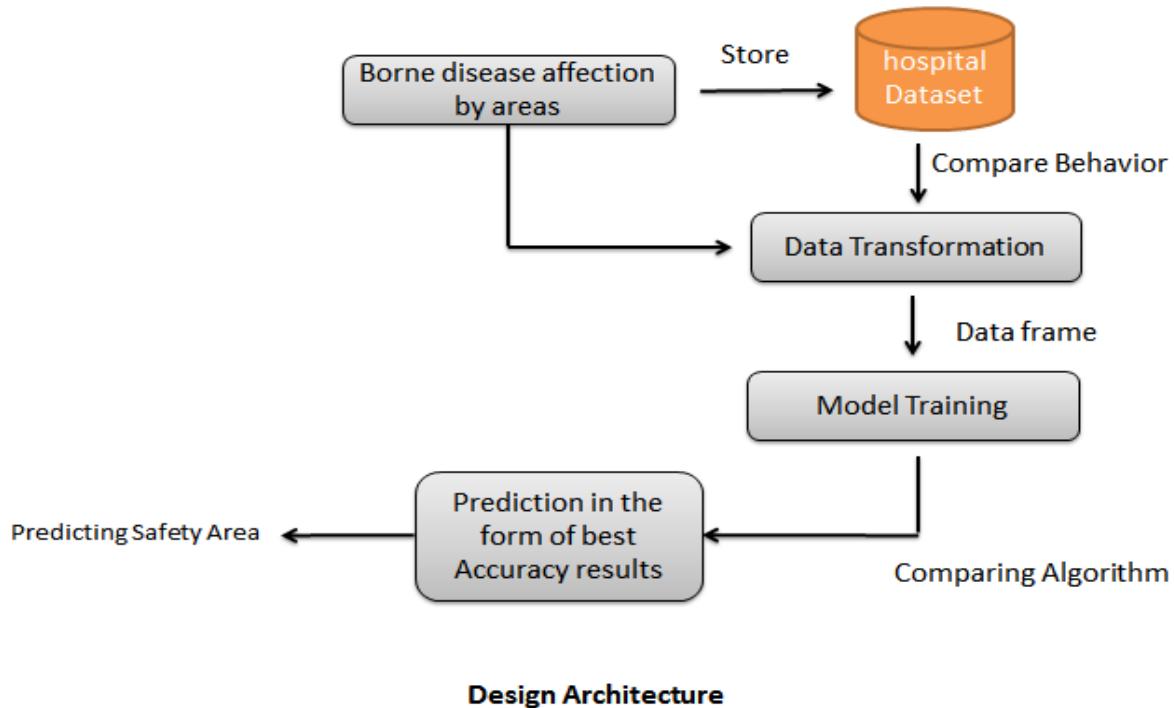
Design Architecture

Fig.2.Design Architecture

The early warning functions of surveillance are fundamental for national, regional and global health security. Recent outbreaks such as the severe acute respiratory syndrome (SARS) and avian influenza, and potential threats from biological and chemical agents, demonstrate the importance of effective national surveillance and response systems. The International Health Regulations (IHR) underscore the commitment to the goal of global security and request all Member States to establish and implement effective surveillance and response systems to detect and contain public health threats of national and international importance. It helps all others department to carried out other formalities. It have to find Accuracy of the training dataset, Accuracy of the testing dataset, Specification, False Positive rate, precision and recall by comparing algorithm using python code. The following Involvement steps are,

➢ Define a problem
➢ Preparing data
➢ Evaluating algorithms
➢ Improving results
➢ Predicting result

DATA COLLECTION:

The data set collected for predicting loan customers is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest , logistic , Decision tree algorithms etc. are applied on the Training set and based on the test result accuracy, Test set prediction is done.

DATA PREPROCESSING:

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done. Based on the correlation among attributes it was observed that attributes that are significant individual.

BUILDING THE CLASSIFICATION MODEL:

It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables. It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.
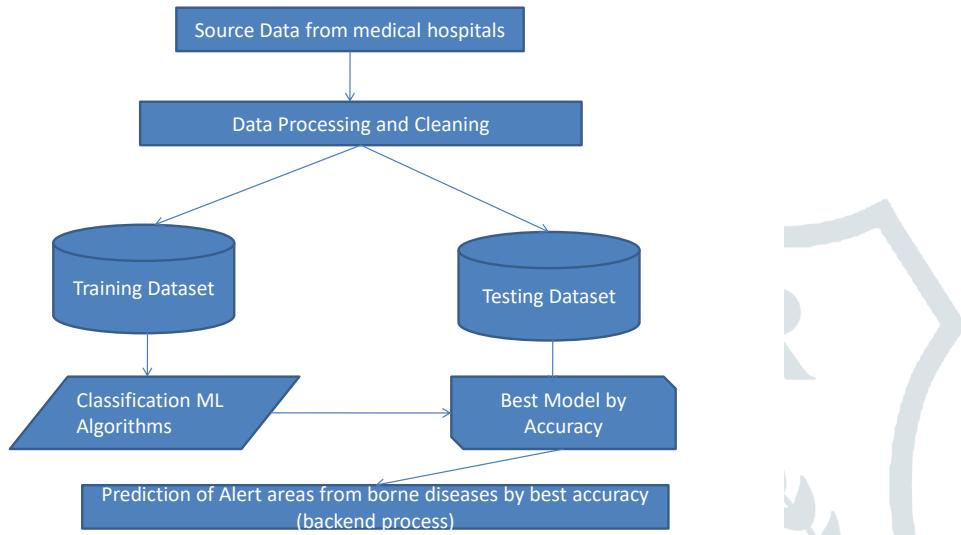

Fig.3. Classification Model

RESULTS

Comparison result for best Accuracy

Algorithm	Precision	Recall	F1-Score	Accuracy (100%)
LR	0.85	0.80	0.79	80
DT	1	1	1	100
SVC	0.28	0.53	0.37	53.33
RF	1	1	1	100
KNN	0.67	0.67	0.66	66.66
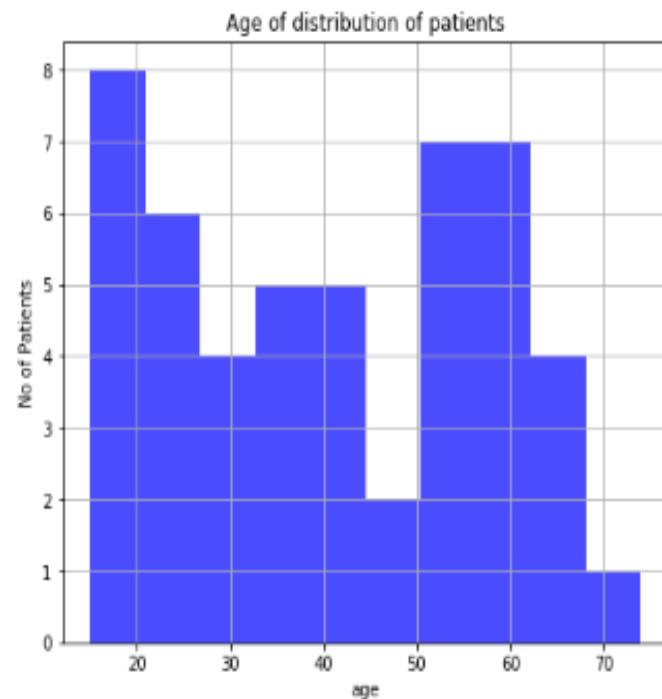
Fig.4. Age distribution of each patient
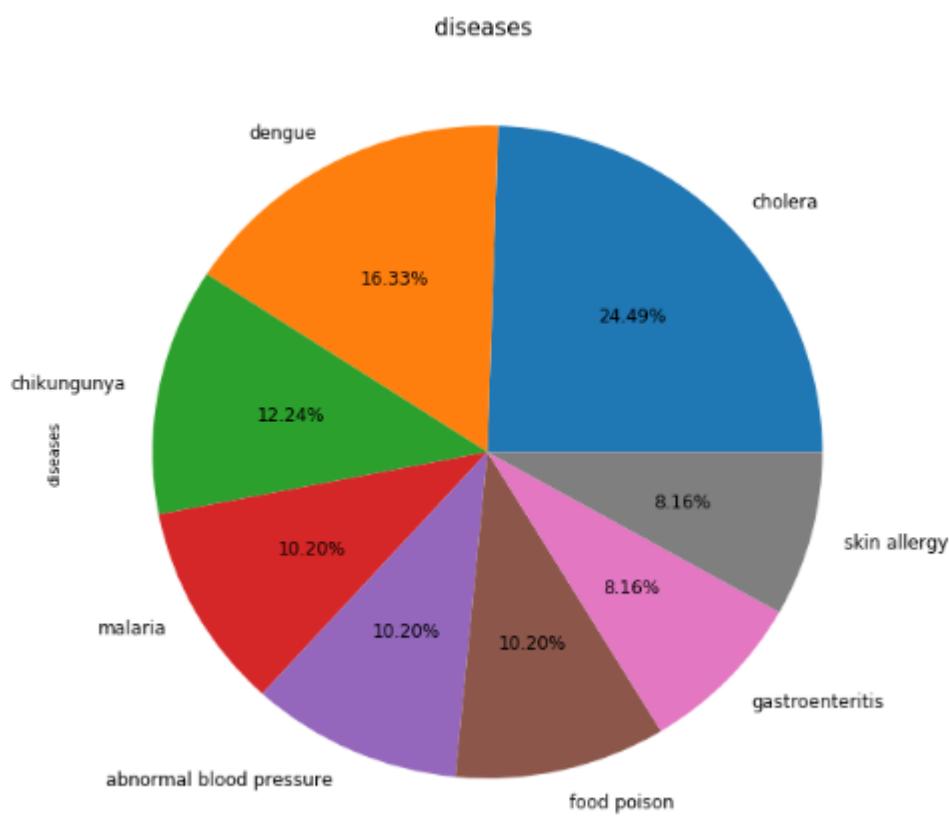
Fig.5. Most area affected disease
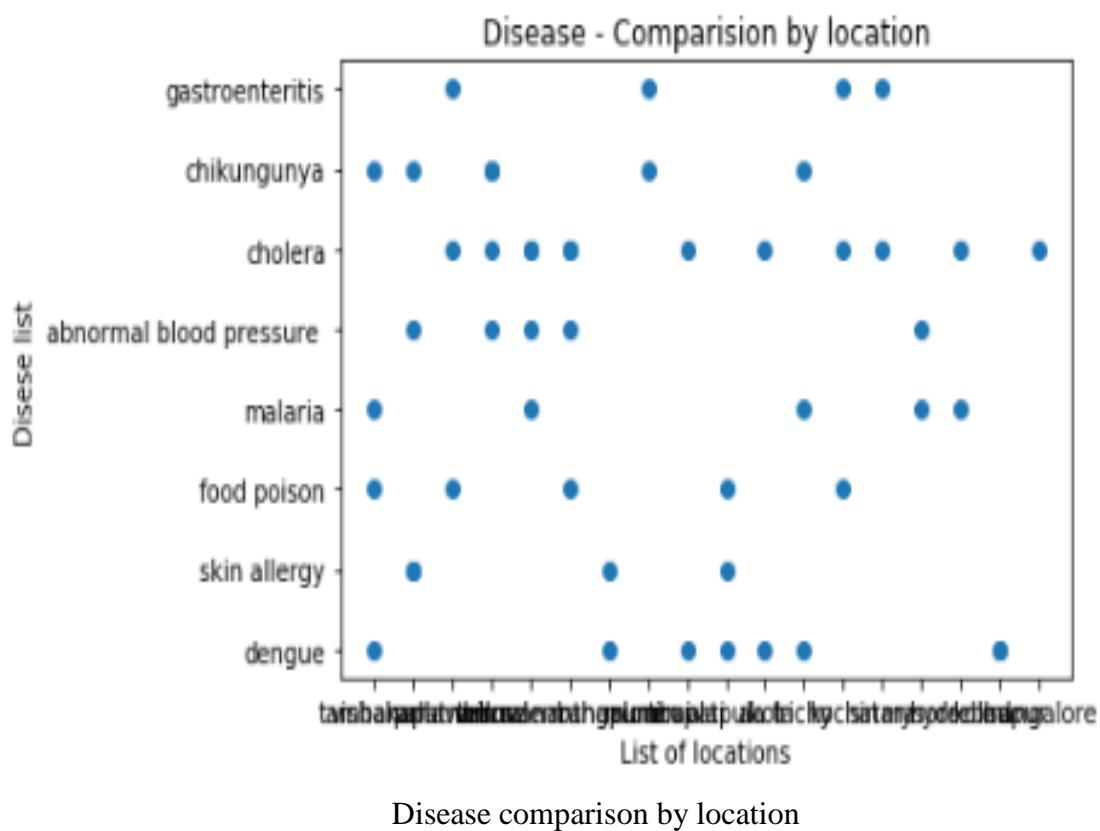
Fig.6.

Disease comparison by location
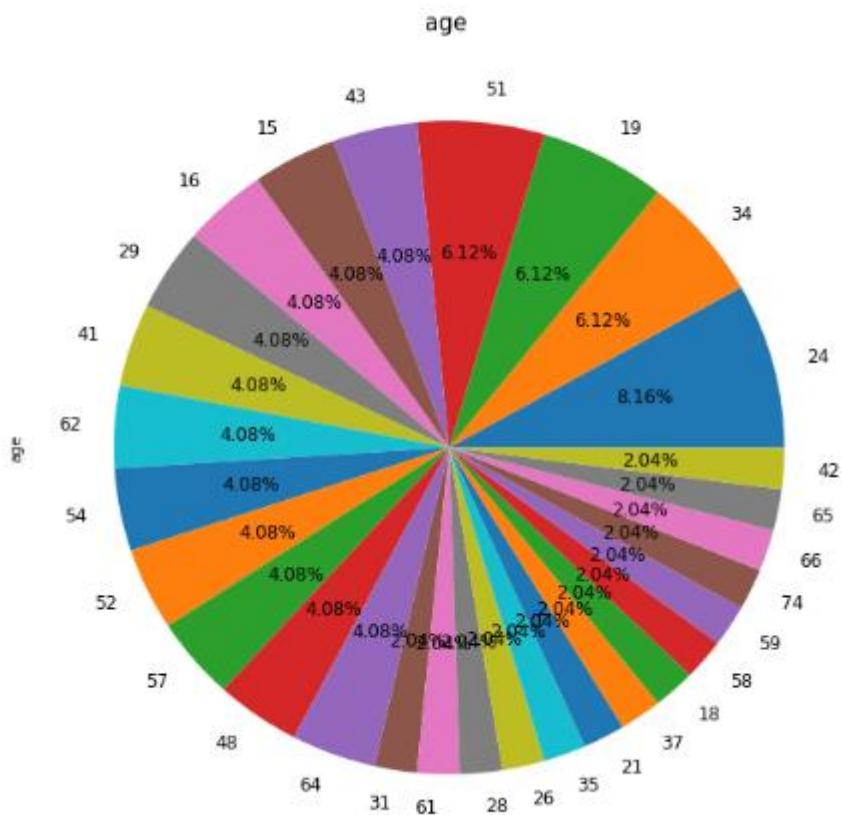
Fig.7. Most area affected patient age

CONCLUSION

In this system we have concentrated to avoid the spread of more hygiene related diseases which provide a clear picture of denoting the areas which is been infected more. The NGO's and other healthcare parties are viable to take necessary steps to clean the particular areas. Patient, individuals, medical clinics and several other medical authorities are maintained in a particular region by using this system. The system uses a user friendly screen andan easier way of analysis.

REFERENCES:

[1] Govindakarnavar, Arunkumar Radhakrishnan, Chandni Devendra et al. , "The Journal of InfectiousDiseases" Outbreak Investigation of Nipah Virus Disease in Kerala, India, 2018

[2] https://www.afro.who.int/health-topics/disease-o[1] Govindakarnavar, Arunkumar Radhakrishnan, Chandni Devendra et al. , "The Journal of InfectiousDiseases" Outbreak Investigation of Nipah Virus Disease in Kerala, India, 2018

[2] https://www.afro.who.int/health-topics/disease-outbreaks

[3] https://ieeexplore.ieee.org/document/7912315

[4] https://www.slideshare.net/aranli/safety-in-hospitals

[5] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, Lin Wang, "Disease Prediction by Machine Learning Over Bigdata from Healthcare Communities", IEEE Access on Healthcare and Bigdata, 2017, Vol.5.

[6] P.Siriyasatien, S.Chadsuthi, K.Jampachaisri, K.Kesorn, "Dengue Epidemics Prediction: A survey of the state- of–the-art based on data science processes, IEEE Access, August 2018.