

Dengue Fever Prediction using Datamining Classification Technique



R.Anusha

Abstract: Dengue is a life threatening disease in all the developed countries like India. This is a virus borne disease caused by breeding of Aedes mosquito. Dengue is caused by female mosquitoes. A predictive system which can identify and minimize the loss due to this problem can be constructed. Datasets used is here the body temperature, vomiting, metallic taste, joint pain etc.. the main objective of this paper is to classify data and to identify the maximum accuracy to predict the dengue fever using description like yes/no. So the classification techniques used here is Bayes classification, nearest neighbor (knn), naïve bayes, rule bayes, id3, and decision tree. From the classified algorithms Naïve bayes had occurred maximum accuracy of 72%. Rapid miner is the data mining tool used to classify the data mining techniques.

Keywords: Decision tree, Naïve Bayes, Rapidminer.

I. INTRODUCTION

Dengue fever is a mosquito-borne disease caused by the dengue virus. The symptoms generally start to work after three to fourteen days of infection. The typical symptoms of dengue are high fever, headache, vomiting, muscle and joint pains, and a characteristic skin rash. Recovery generally takes two to seven days. In the worst case, the disease develops into severe dengue, also known as dengue hemorrhagic fever, resulting in bleeding, low levels of blood platelets and blood plasma leakage, or into dengue shock syndrome, where dangerously low blood pressure occurs. There is no specific treatment or antibiotic for dengue as such. It usually tends to attack children for the age group below 10 years. Dengue is caused by several types of female mosquitoes of the Aedes type. We can prevent dengue fever by eradicating the mosquitoes that carry viruses that cause dengue. When an Aedes mosquito bites a healthy person which carries viruses in its saliva enters the body which then gets mixed with the white blood cells. Followed by the dengue virus reproducing inside

the white blood cells and thus initiating the dengue virus cycle. According to the data from National Vector Borne Disease Control Programme (NVBDCP) and National Health Profile 2018. From less than 60,000 cases in 2009, cases increased to 188,401 in 2017—more than a 300 per cent spike. When compared to 75,808 cases in 2013, it is more than a 250 percent spike.

1.1 CHARACTERISTICS OF DATA MINING

Data Mining explores voluminous data to identify patterns and relationships between variables. and then use the inferences to detect patterns. predictive data mining can be applied in almost all business applications.

1.2 RESEARCH OBJECTIVE

- To predict and identify dengue disease infected person by using the training dataset which consists of the factors causing dengue fever as attributes and applying the pattern to the test data set.
- Applying classification techniques to identify dengue affected person by using key attributes
- Validating the results by measuring the accuracy of the classification techniques

II. REVIEW OF LITERATURE

kashisharashakil, shadmaan and mansafalam^[1], have classified the dengue data set and then compared the different data mining techniques in weka used 18 attributes and compared the accuracy of different classification algorithms. PrapatSuriyaphol, NuanwanSoonthornphisaj^[2] used two classification techniques such SVM and Decision tree for predicting dengue fever. R. Sanjudevi, D. Savitha^[3] predicted dengue fever using attributes like temperature, humidity, pressure and evaluated the performances of classification algorithms like random forest and support vector machine. Vanny Narita^[4] analyze the performance of dengue early warning system [DEWS] by testing its accuracy of predictions, using data of environmental factors, climate and surveillance in District Cempaka Putih. Naïve Bayes was chosen as Dengue outbreak predictor. BuchadeOmka^[5] has used the dataset collected from various hospitals and to achieve the maximum accuracy, they used the algorithm like spider monkey optimization using WEKA tool where they achieved a maximum accuracy of 90.9%. Bhavani, M., et al^[6] emphasised on attributes like fever, bleeding, no of RBC cells.

Manuscript published on November 30, 2019.

* Correspondence Author

Dr.R.Anusha*, Asst.Professor, Department of Computer Science, M.O.P.Vaishnav College For Women (Autonomous), Chennai

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Dengue Fever Prediction using Datamining Classification Technique

Since the main objective of this study is to calculate the accuracy through various classification Techniques and compare their performance. The study used classification techniques such as REP Tree, J48, SMO, ZeroR and Random Tree and the performance was compared by plotting graphs using WEKA tool.

III. CAUSES OF DENGUE FEVER

Dengue fever is caused by any one of four types of dengue viruses spread by mosquitoes that survive in and near human lodgings. A person bit by mosquito gets infected with dengue virus as it enters the bloodstream. After recovery the person may be immune to the infected virus but still prone to infection by other viruses. The prevention of dengue fever requires control or eradication of the mosquitoes carrying the virus

IV. DENGUE SYMPTOMS AND SIGNS

Primary symptoms of dengue appear three to 15 days after the mosquito bite and include the following:

- high fever and severe headache
- with severe pain behind the eyes that is apparent when trying to move the eyes.

Other associated symptoms are:

- joint pain
- muscle and bone pain
- rash
- mild bleeding
- low back pain
- reduce of WBC
- reduce of platelets

V. METHODOLOGY

For prediction of dengue disease dengue data set was used. It consists of 11 attributes and 60 records for dengue disease classification and prediction which has been collected through survey as questionnaire format, circulated between various age groups. Dataset was given as input to Classification and prediction algorithms and their accuracy was compared using rapidminer tool.

	A	B	C	D	E	F	G	H	I	J	K
1	name	age	will platelets reduce	headache	joint pain	metallic taste	abdominal pain	vomiting	bleeding	major symptom	dengue
2	madhu	19	yes	no	yes	no	yes	no	no	reduction of wt	yes
3	ram	19	no	no	no	yes	no	no	no	bleeding	no
4	keerthana	20	yes	yes	yes	yes	no	yes	yes	joint pain	yes
5	shalini	24	no	no	no	yes	yes	no	no	body temperati	no
6	yedhudev	17	yes	yes	no	no	yes	yes	yes	reduction of wt	yes
7	diwarakesi	31	yes	yes	yes	no	no	no	no	bleeding	no
8	krishtna	20	yes	no	yes	no	no	yes	yes	joint pain	yes
9	rohit	24	yes	yes	no	no	no	no	no	reduction of wt	no
10	venu	21	yes	no	yes	no	no	yes	yes	body temperati	yes
11	lavanya	21	yes	no	no	yes	no	no	no	reduction of wt	no
12	lasiya	22	no	no	no	yes	no	yes	no	body temperati	yes
13	nandini	23	yes	yes	no	no	no	no	no	body temperati	no
14	moleesh	24	no	no	yes	yes	no	yes	no	reduction of wt	yes
15	syed	25	yes	yes	no	no	yes	no	yes	reduction of wt	no
16	vignesh	26	no	no	yes	yes	yes	yes	no	reduction of wt	yes
17	yuvraj	27	yes	no	no	no	yes	no	yes	reduction of wt	no
18	kesav	27	yes	yes	yes	yes	no	yes	no	reduction of wt	yes
19	pooja	19	yes	no	no	yes	no	no	yes	bleeding	no
20	sree	18	yes	no	yes	no	no	no	yes	bleeding	yes
21	jaythnsa	29	yes	yes	no	no	yes	no	no	bleeding	no
22	pavithra	40	yes	yes	no	no	no	no	yes	bleeding	yes
23	geetha	46	yes	no	no	no	yes	yes	no	body temperati	no
24	mevin	39	yes	yes	yes	yes	no	yes	yes	reduction of wt	yes
25	antony	55	yes	no	yes	yes	yes	yes	no	bleeding	yes

The attributes used here is body temperature, reduction of platelets count and various symptoms like vomiting, bleeding, joints pain based on which the classification algorithms are applied. For this dataset, different types of classification algorithms like KNN, naïve bayes, rule-based and ID3 are applied. Then these algorithms are implemented using rapid miner tool to analyse algorithm accuracy.

5.1 DATA MINING TECHNIQUES:

The data mining technique have been used to predict dengue disease. Predictions have been done using RAPID MINER , a data mining tool for classification and accuracy by applying different classification algorithms approaches.

DECISION TREE ALGORITHM:

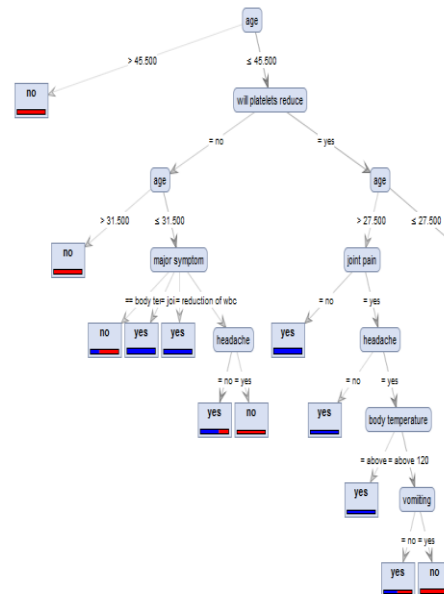


Table View Plot View

accuracy: 55.56%

	true yes	true no	class precision
pred. yes	10	8	55.56%
pred. no	0	0	0.00%
class recall	100.00%	0.00%	

NAÏVE BAYES ALGORITHM:

Naïve Bayes is one of the algorithms that works as a probabilistic classifier of all attributes contained in data sample individually and then classifies data problems. Running the algorithms using Naïve Bayes we analyse the classifier output to make a prediction of each instance of the dataset.

TABLE 1 :PREDICTION USING NAÏVE BAYES ALGORITHM

Table View Plot View

accuracy: 72.22%

	true yes	true no	class precision
pred. yes	8	3	72.73%
pred. no	2	5	71.43%
class recall	80.00%	62.50%	

After running this algorithm, it has been observed that classification accuracy of 72.2% is achieved.

KNN ALGORITHM:

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. KNN has been used in statistical estimation and pattern recognition. For this dengue dataset, knn algorithm has been applied to analyse the accuracy.

TABLE 2 :PREDICTION USING KNN ALGORITHM

Table View Plot View

accuracy: 50.00%

	true yes	true no	class precision
pred. yes	6	5	54.55%
pred. no	4	3	42.86%
class recall	60.00%	37.50%	

The accuracy for KNN algorithm is found to be 50% which is much lower when compared to other classification algorithms.

RULE-BASED CLASSIFICATION:

Rule-based classifier makes use of a set of if-then rules for classification. For this dengue dataset, rule-based algorithm has been applied to compare the accuracy.

TABLE 3 :PREDICTION USING RULE BASED ALGORITHM

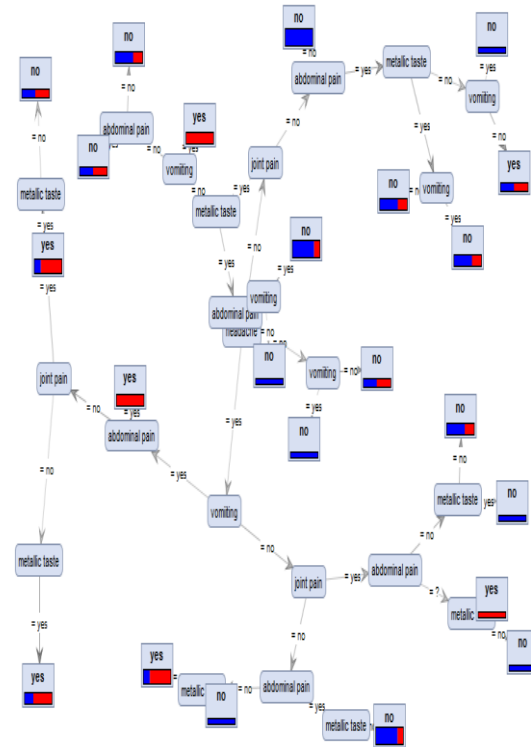


Table view Plot view

accuracy: 64.71%

	true yes	true no	class precision
pred. yes	6	1	85.71%
pred. no	5	5	50.00%
class recall	54.55%	83.33%	

The accuracy generated by rule-based classification algorithm is 64.71%. It is evident that the accuracy of rule-based algorithm is much higher than the other algorithms except naïve bayes.

ID3 ALGORITHM

ID3 - Iterative Dichotomiser 3 is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset.

TABLE 4 :PREDICTION USING ID3 ALGORITHM

accuracy: 66.67%

	true no	true yes	class precision
pred. no	9	4	69.23%
pred. yes	2	3	60.00%
class recall	81.82%	42.86%	

The accuracy generated by ID3 algorithm is 66.67%.

DECISION TREE GENERATED BY ID3 ALGORITHM:

than other classification algorithms in predicting the factors of dengue fever. The results of comparison are tabulated below:

TABLE 5 :Comparison of accuracies of different classification algorithms

Name of the algorithm	Accuracy generated
Rule based classification	64.71%
K nearest neighbor	50%
ID3 algorithm	66.67%
Naïve bayes	72.2%
Decision tree	55.56%

VI. RESULTS:

Five data mining classification algorithms were used to classify the dengue dataset. After the analysis of each classification techniques, comparison has been made between these five algorithms based on their accuracy rate. Their performance was analysed, and from that analysis it is been observed that naïve bayes algorithm performs better

VII. CONCLUSION:

The main aim of this paper is to predict dengue disease using rapid miner tool by taking into various factors as attributes. It is also observed that of all the symptoms dengue fever mostly varies on high fever temperature followed by joint pain and reduction of platelets count. The main objective is toward prediction of dengue infection using classification techniques. Of all classification techniques like rule based, KNN, naïve bayes and Bayesian classification it is been concluded that the accuracy of naïve bayes algorithm is more than that of all other classification techniques.

REFERENCES:

1. kashisharashakil, shadmaan and mansafalam, "Dengue disease prediction using weka data mining tool" Department of Computer Science, JamiaMilliaIslamia ,New Delhi, India, Published in ArXiv 2015
2. PrapatSuriyaphol, NuanwanSoonthornphisaj, "Data Mining of Dengue Infection Using Decision Tree", Latest Advances in Information Science and Applications, ISBN: 978-1-61804-092-3
3. R. Sanjudevi, D. Savitha, dengue fever prediction using classification techniques, International Research Journal of Engineering and Technology ,ISSN:2395-0056
4. Vanny Narita, " Dengue Outbreak Prediction for GIS based Early Warning System", 2015 International Conference on Science in Information Technology .
5. Buchade Omkar, Dalsania Preet, Deshpande Swarada, Doddamani Poonam, "Dengue Fever classification using SMO Optimization Algorithm", International Research Journal of Engineering and Technology ,ISSN:2395-0056
6. Bhavani.M,"A Datmining approach for precise diagnosis of dengue fever",International journal of latest trends in Engineering and Technology",vol7,issue 4,e-ISSN:2278-62IX