International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Hybrid Classification Method for Dengue Prediction

Prashansa Taneja, Nisha Gautam

Abstract*: *Data mining is defined as the process in which useful information is extracted from the raw data. In order to acquire essential knowledge it is essential to extract large amount of data.. In this existing work, the technique of SVM is applied for the prediction of dengue. The SVM classifier has less accuracy and high execution time for the prediction. To improve the accuracy of prediction the voting based classification approach will be applied for the dengue prediction. The proposed method will be implemented in python and results will be analyzed in terms of accuracy, precision, recall and execution time.

Index Terms*: *Dengue Prediction, Hybrid Classifier, SVM

I. INTRODUCTION

Data mining is defined as the process in which useful information is extracted from the raw data. In order to acquire essential knowledge it is essential to extract large amount of data. This process of extraction is also known as misnomer. Currently in every field, there is large amount of data is present and analyzing whole data is very difficult as well as it consumes a lot of time. This present data is in raw form that is of no use hence a proper data mining process is necessary to extract knowledge [1]. The process of extracting raw material is characterized as mining. This is a world where having a lot of information leads to power and success and this is possible only because of sophisticated technologies such as satellites, computers. With the advent in the technology in the mass digital storage and computers it becomes easy to handle large amount of information by which different types of data is stored. In the cluster analysis, image processing, market research, data analysis and pattern recognition are some major application of this technique. In the clustering technique, customer categorized group and purchasing patterns has been done in order to discover their customer's interest by the marketers. It is also utilized in biology as it derives the plant and animal taxonomies and also categorizes genes with similar functionality. In geology this technique is used to identify the similar houses and lands areas. Information clustering can be used to discover new theories that classify all documents available on Web.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Prashansa Taneja*, (Research Scholar, CSE, AP Goyal Shimla University, Shimla, India)
Nisha Gautam (HOD – CSE, AP Goyal Shimla University, Shimla, India)

Clusters, group of objects are created by the unsupervised data clustering classification as in different clusters objects are not identical to each other and objects that are in same clusters are identical to each other. Cluster analysis, in the data mining is considered as the essential method such as knowledge of discovery (KDD) method. Clusters are defined as the set of disjoint classes which are grouped into data objects. There is resemblance between the objects within a class to each other and in sometimes they are dissimilar to each other [5]. When the clustering is not dependent on the pre-defined classes and trainings is called as unsupervised clustering. In the area of statistics pattern recognition and unsupervised learning is different from each other, which is known as decision analysis. It is also called as discriminate analysis in which objects are classified from given set of object.

A. Dengue Prediction

The prediction analysis is most useful type of data which is performed today. To perform the prediction analysis the patterns needs to generate from the dataset with the machine learning. The prediction analysis can be done by gathering historical information to generate future trends. So, the knowledge of what has happened previously is used to provide the best valuation of what will happen in future with predictive analysis. The predication analysis models are designed according to application type. The model is trained using the sample data that includes known attributes. The new data can be analyzed and its behavior can be determined using this trained model. A dengue virus is spread by the Aedes mosquito to cause a disease called dengue. Mainly, the tropical regions which have the highest mosquitoes-sustaining environment are more prone to this disease. The female mosquitoes are only responsible for spreading dengue. When a patient suffers from dengue, the pain caused is very similar to that of breaking a bone due to which it is also known as breakbone fever. The best precaution one can take to save themselves from dengue is to protect themselves from the mosquito bites [7]. After 3 to 15 days of mosquito bite the primary symptoms of dengue start appearing. Following are some of the commonly found symptoms of this disease:

- The diseased person suffers from severe headache and high fever.
- When moving the eyes, a severe pain is suffered behind the eyes by the patient.

- Pain in joints is another common symptom.
- The bone and muscle pains are also complained by the diseased.
- The diseased body might have rash.
- Sometimes mild bleeding is also complained by the patient.

Based on the severity of symptoms, the dengue fever is categorized. The dengue fever and dengue hemorrhagic fever are difficult to be differentiated at the initial stages. To predict the dengue fever, various data mining techniques have been proposed by researchers.

II. LITERATURE SURVEY

P. Manivannan, et.al (2017) proposed a novel method of household clustering for the prediction of dengue outbreak [8]. In the proposed approach, dengue patterns were utilized on the basis of age cluster with the help of k-means clustering approach. The effectiveness of output was improved with the help of k means clustering approach. This approach was considered to be the most effective approach for the prediction of dengue patients with several patterns. The data sample implemented in the projected approach was completely clustered. The researcher was given attention to mainly four stages. These stages were preprocessing, characteristic election, clustering and forecasting of dengue disease. For the preprocessing of dengue household data sample, a tool named R3.3.2 was utilized. A new technique named D win was implemented for the generation of overflowed data samples through alternating every misplaced assessment for ostensible and arithmetic features in association with mode and mean assessment. For the forecasting of dengue germ distinct data mining approaches could be utilized. The major objective of this study was the prediction of dengue affected patients on the basis of age groups with the employment of k means clustering approach.

ChienHsing Wu, et.al (2018) highlighted the utilization of data mining approaches for dengue prediction [9]. The proposed research served mainly three objectives. The first aim of this study was the use of data mining approaches for the recognition of dengue disease. The second aim of this study was the incorporation of various open information origins. Sourcing of realistic approaches from experimental qualitative study of realm experts and professionals was the third aim of this study. Various investigative results declared that the categorization based data miming approaches could be significantly used for Taiwan's dengue information samples. The recognized capability of hereditary variables with the help of perceptible investigation was decrypted. Utilization of exposed acquaintance through experimental study proved beneficial for the generation of dengue disease manage plans and tactics. For this, a depth consideration of dengue disease was required. It was also identified that the change in environmental conditions could be the reason of this disease. In future, more researches will be carried out for observing the effects of other parameters in case of dengue disease. Olav Titus Muurlink, et.al (2018) presented a review of dengue epidermis transmitted through mosquitoes [10]. The research work was carried out in Bangladesh. This disease was the main threat to human being's life in hot and cold areas globally. This investigation incorporated novel amalgam factors like irregular proceedings, operating averages; cyclic variables relied on conventional Bengal six period yearly almanacs and insulated time span of one year in forecasting continuation or the scale of every dengue plague. For showcasing the prediction of distinct variables, a comparatively new and inclusive data mining technique was used in this research work. It was also predicted that the rainy reason of some months and the regular least temperature were the main reason of dengue outbreak. The scale of associations among dampness of six, seven and eight months before the eruption suggested that the association was not merely because of the repetitive character of existing topological surroundings. Vandana Rajput, et.al (2017) suggested the use of a new methodology for the prediction of dengue epidermis [11]. In the presented study, the utilization of genetic algorithm was performed. This approach was implemented for calculating the real time heaviness of characteristics and after this FP-growth through real heaviness was implemented. Various surveys and experimental outcomes depicted that the customized methodology was capable for the recognition of practical connotation of characteristics in prerequisites of their heaviness. The presented approach was premeditated. For the attainment of most possible prediction reports, certain constraints were utilized. The outcomes of various experiments indicated that the proposed approach produced enhanced forecast. The major aim of this study was the generation of a significant plan and the delimitation of previous outcomes. In future, actual information utilizing distinct features in association with other advanced methodologies on certain other constraints will be used for defining the suitable architectural design for timely forecasting. Dini Rahmawati, et.al (2016) presented a novel approach of linear expansion and RBF Kernel operation of C-Support Vector categorization for the prediction of arithmetical dengue disease eruption [12]. This dengue fever outbreak was associated with region, environmental conditions and regular rainfall. These characteristics could be the reason of dengue outbreak which included health related impendent. For the resolution of hyper parameter election issues in erudition methodologies, grid discovering system was implemented. The simulation outcomes verified that the proposed approach after constraint escalation attained superior forecasting precision on dengue disease epidemic in comparison with some conventional used approaches. In the presented approach most superior Kernel model was utilized for the classification of mutual environmental information. Although the better outcomes were achieved with the help of proposed approach, but there was still the need of more researches in this area. Many appealing problems will also be investigated in the near future related to dengue fever outbreak. In future, implementation strategy of spatio-temporal systems will also be carried out as this is necessary for making an association in the apparition scheme. This study will prove beneficial in future in categorization of spatio-temporal information for the enhancement of anticipation task on certain ailments.

Shermon S. Mathulamuthu, et.al (2016) presented a dengue occurrence forecasting methodology. The main purpose of this research was the avoidance of dengue outbreak with the help of topological systems in actual scenario [13]. For the attainment of highly suitable deterioration curvature, some data mining approaches like numerous regressions and clustering techniques were implemented. Actual time flexible calculative software was matured for the prediction of dengue occurrence instantly. Utilization of R in association with information base execution was fundamental because it had the capability of working with huge amount of information and extremely organized information and at the same time was very effective also. The procedure of gathering information from distinct huge climate data samples was implemented via withdrawal of weather mock-up outlines. Data mining approaches in association with machine learning algorithms helped in the attainment of dengue forecaster. For the achievement of good investigative outcomes, data was clustered into tiny clusters and the development of regression level was carried out near cluster lines. It was also suggested that with the help of online techniques more precise outcomes could be achieved.

III. RESEARCH METHODOLOGY

The prediction analysis is the technique which can predict the future possibilities from the existing data. The prediction analysis techniques are based on the clustering and classification. The machine learning algorithms are the most popular algorithms which are applied for the dengue prediction. The SVM is the most common and widely used classification algorithm for the dengue prediction. The dengue disease dataset are very complex in nature means it has number of attributes due to which SVM classifier give less accuracy.

Following are the various phases of research process:-

1. **Pre-Processing**:- The first phase of the research process is the pre-processing in which dataset is loaded which is collected from the UCI repository. The input data is cleaned in this phase means missing values are removed from the dataset

2. **Apply Mutiple Classifiers:-** In this phase, the multiple classifiers are applied for the predication analysis. The multiple classifiers are SVM, KNN, and decision tree, naïve bayes

3. **Voting Method: -** In the last phase, the voting method is applied for the generation of final predicted results. The voting method is applied which can select best performing classifier from the multiple classifier and generate predicted result

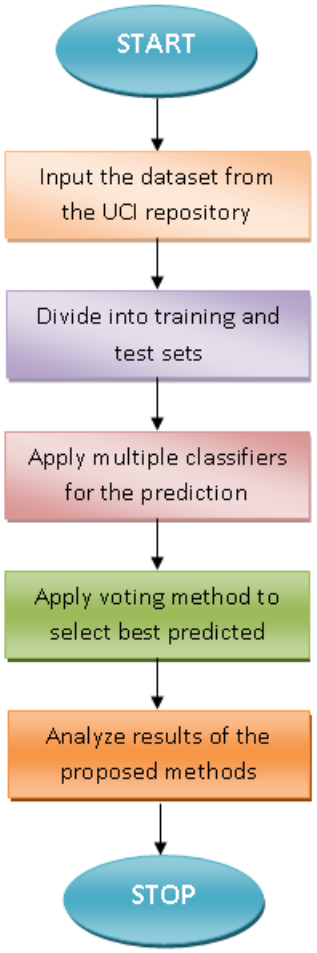

Fig 1: Proposed Flowchart

IV. RESULTS AND DISCUSSION

The dataset of the fake news detection is collected from the kaggle. The dataset does not have any missing or redundant or missing values. The performance of the proposed model is tested in terms of certain parameters which are precision, recall, and accuracy

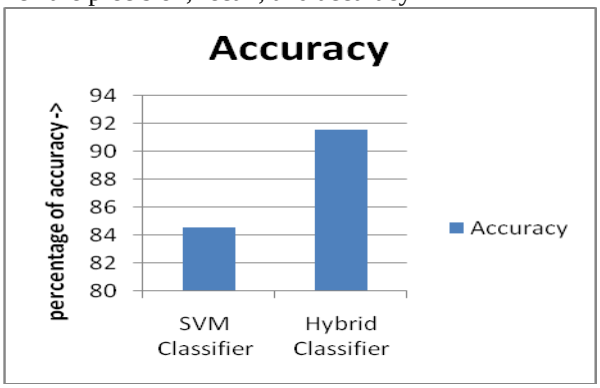

Fig 2: Accuracy Analysis

As shown in figure 2, the accuracy of the SVM classifier is compared with the hybrid classification. The hybrid classification model is the combination of naïve bayes and decision tree. It is analyzed that when the hybrid classification model is used accuracy is increased up to 8 percent

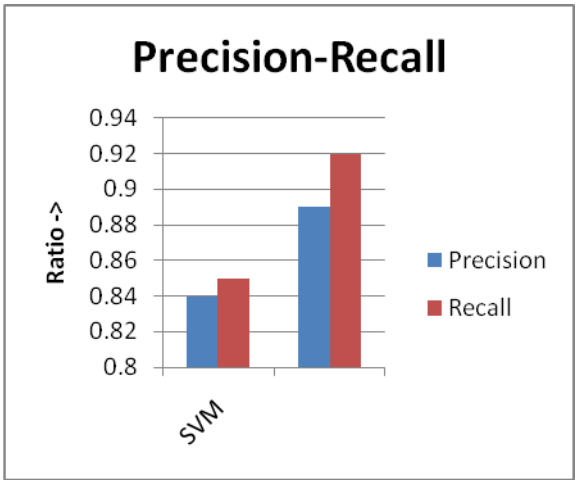

Fig 3: Precision-Recall Analysis

As shown in figure 3, the precision-recall value of the SVM classifier is compared with the hybrid classification The precision-recall value of the hybrid classifier is high as compared to SVM classifier for the fake news detection.

V. CONCLUSION

The dengue prediction is the problem of prediction analysis. The dengue prediction analysis can be done in two steps which are features extraction and classification. In this work, it is concluded that in the existing work, the SVM classification approach is applied for the prediction analysis. The SVM classifier has less accuracy, as it is not able to drive relationship between the attributes and target set. To increase accuracy of dengue prediction the technique of SVM can be replaced with the voting based classifier. The proposed method will be implemented in python and results will be analyzed in terms of accuracy, precision, recall

REFERENCES

1. Abdelghani Bellaachia and Erhan Guven (2010), "Predicting Breast Cancer Survivability Using Data Mining Techniques", Washington DC 20052, vol. 6, 2010, pp. 234-239.
2. Oyelade, O. J, Oladipupo, O. O and Obagbuwa, I. C (2010), "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance", International Journal of Computer Science and Information Security, vol. 7, 2010, pp. 123-128.
3. Azhar Rauf, Mahfooz, Shah Khusro and Huma Javed (2012), "Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity", Middle-East Journal of Scientific Research, vol. 12, 2012, pp. 959-963.
4. Osamor VC, Adebiyi EF, Oyelade JO and Doumbia S (2012), "Reducing the Time Requirement of K-Means Algorithm" PLoS ONE, vol. 7, 2012, pp-56-62.
5. Azhar Rauf, Sheeba, Saeed Mahfooz, Shah Khusro and Huma Javed (2012), "Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity," Middle-East Journal of Scientific Research, vol. 5, 2012, pp. 959-963
6. Kajal C. Agrawal and Meghana Nagori (2013), "Clusters of Ayurvedic Medicines Using Improved K-means Algorithm", International Conf. on Advances in Computer Science and Electronics Engineering, vol. 23, 2013, pp. 546-552.
7. P. Sathya, Mrs. A.Sumathi, "Predicting Dengue Fever Using Data Mining Techniques", International Journal of Computer Science Trends and Technology (IJCST) – Volume 6 Issue 2, 2018
8. P. Manivannan, Dr. P. Isakki @ Devi, "Dengue Fever Prediction Using K-Means Clustering Algorithm" ,2017, IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNIQUES IN CONTROL, OPTIMIZATION AND SIGNAL PROCESSING
9. ChienHsing Wu, Shu-Chen Kao, Chia-Hung Shih, Meng-Hsuan Kan, "Open Data Mining for Taiwan's Dengue Epidemic",2018, ACTROP 4618
10. Olav Titus Muurlink, Peter Stephenson, Zahirul Islam, Andrew Taylor-Robinson, "Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach", 2018, IDM 70
11. Vandana Rajput, Asst. Prof. Amit Manjhvar, "A Model for Forecasting Dengue Disease Using Genetic based Weighted FP-Growth",2017, International Conference on Trends in Electronics and Informatics, ICEI
12. Dini Rahmawati and Yo-Ping Huang, "Using C-support Vector Classification to Forecast Dengue Fever Epidemics in Taiwan", 2016, International Conference on System Science and Engineering (ICSSE)
13. Shermon S. Mathulamuthu, Vijanth S. Asirvadaml, Sarat C.Dass , Balvinder S. Gile, Loshini T "Predicting Dengue Incidences Using Cluster Based Regression on Climate Data",2016, IEEE International Conference on Control System, Computing and Engineering,

AUTHORS PROFILE

Prashansa Taneja is a post graduate student (M.Tech) in Computer Science and Engineering at AP Goyal Shimla University, where her area of Research is Data Mining. She is Graduated from AP Goyal Shimla University, with her bachelor degree in computer science in 2017.

Ms. Nisha Gautam is currently serving as Head of the department in Computer Science & Engineering Department. She is also serving as Assistant Professor in the Department of Computer Science & Engineering. She is having total of 6.8 Years of academic experience in the field of Computer Science & Engineering. She has guided various students in their M. Tech dissertations. She is an active participant of various Seminars, Induction and faculty development programs.