DESIGNING DISEASE PREDICTION MODEL USING MACHINE LEARNING APPROACH

Prof. Sarika Aundhakar, Lokesh Girase, Rishikesh Khakal, Vaishnavi Wattamwar

Department of Computer engineering, Smt. Kashibai Navale College of Engineering. Maharashtra, Pune 412307

ABSTRACT:

Today, human beings face diverse sicknesses because of the environmental situation and their residing conduct. So the prediction of sickness at an in advanced degree will became a vital task. But the correct prediction on the idea of signs will become too hard for doctors. The accurate prediction of sickness is the maximum hard task. To triumph over this trouble, information mining performs a vital function to are expecting the sickness. Medical technological know-how has a massive quantity of information increase according to year. Due to the multiplied quantity of information increase with inside the clinical and healthcare area the correct evaluation of clinical information has been cashing in on early affected person care. With the assistance of sickness information, information mining reveals hidden sample facts in a large quantity of clinical information. We proposed fashionable sickness prediction primarily based totally at the signs of the affected person. For the sickness prediction, we use Naive bayes and Random Forest gadget studying set of rules for the correct prediction of sickness. For sickness prediction required sickness signs dataset. In this fashionable sickness prediction, the residing conduct of someone and checkup facts don't forget for the correct prediction. The accuracy of fashionable sickness prediction via way of means of the use of RANDOM FOREST is 84.5% that is extra than the NAIVE BAYES set of rules. And the time and the reminiscence requirement also are extra in NAIVE BAYES than RANDOM FOREST, DECISION TREE. After fashionable sickness prediction, this device is ready t deliver the danger related to a fashionable sickness that is a decrease danger of fashionable sickness or higher.

***Key Words*: RANDOM FOREST, NAIVE BAYES, DECISION TREE, Machine learning, Disease Prediction.**

1. INTRODUCTION

Artificial Intelligence made laptops extra wise and can permit the laptop to think. AI examine don't forget system getting to know as a subfield in severe studies paintings. Different analysts' sense that without getting to know, perception cannot be created. There are severe types of Machine Learning Techniques like Unsupervised, Semi-Supervised, Supervised, Reinforcement, Evolutionary Learning,

and Deep Learning. These learnings are used to categorize massive statistics very fastly. So, we use Naive bayes and Random Forest system getting to know set of rules for immediate category of large statistics and correct prediction of sickness. Because clinical statistics is growing each day so utilization of that for predicting accurate sickness is an essential venture however processing large statistics may be very essential in popular so statistics mining performs very critical position and category of big dataset the use of system getting to know turns into so easy. It is crucial to realize the correct analysis of patients with the aid of using medical exam and evaluation. For compelling dedication choice guide structures that rely on laptop can also additionally count on a vital job. The health care field creates sizeable statistics approximately medical evaluation, report with regard to patient, cure, next meet-ups, remedy and so forth. It is tricky to orchestrate appropriately. The quality of the statistics affiliation has been stimulated because of improper control of the statistics. Upgrade within side the degree of statistics wishes a few valid manners to pay attention and process statistics viably and efficiently. One of the numerous machine learning programs is applied to assemble such a classifier that can separate the statistics primarily based totally on their characteristics. The data set is partitioned into or extra than classes. Such classifiers are applied for clinical statistics research and sickness prediction. Today system gets to know a gift anywhere so that without understanding it, you can actually in all likelihood use it normally a day. RANDOM FOREST makes use of each of the based and unstructured statistics of a health center to do category. While different systems getting to know algorithms handiest paintings on based statistics and time required for computation is excessive additionally they're lazy due to the fact they store whole statistics as a schooling dataset and makes use of the complicated approach for calculation. Segment I explain the Introduction of popular sickness prediction the use of category approach which includes NAIVE BAYES and RANDOM FOREST. Section II gives the literature overview of present structures and Section III gift proposed device implementation info Section IV gives experimental analysis, outcomes, and dialogue of proposed device. Section V concludes our proposed device. While on the give-up a listing of references paper is presented.

2. LITERATURE SURVEY

M. Chen Proposed [1] a brand new multimodal disease hazard prediction set of rules primarily based totally on Random Forest with the aid of using the use of prepared and unorganized records of hospital. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang Discovered disorder prediction device for diverse regions. They achieved disorder prediction on 3 extraordinary illnesses inclusive of diabetics, cerebral infraction and coronary heart disorder. The disorder prediction is achieved on prepared records. Prediction of coronary heart disorder, diabetes and highbrow infraction is achieved with the aid of using the use of diverse system mastering set of rules like naïve bayes, Decision tree and NAIVE BAYES set of rules. The final results of Decision tree set of rules plays higher than NAIVE BAYES set of rules and Naïve bayes. Also, they predict that both a affected person enjoy from the excessive hazard of cerebral infarction or minimal hazard of cerebral infarction. They used RANDOM FOREST primarily based totally multimodal disorder hazard prediction on textual content records, for the hazard prediction of cerebral infraction. The accuracy contrast takes region among RANDOM FOREST primarily based totally unimodal disorder hazard predictions against RANDOM FOREST primarily based totally multimodal disorder hazard prediction set of rules. The accuracy of disorder prediction resulted as much as the 94.8% with greater speedy velocity than RANDOM FOREST primarily based totally unimodal disorder hazard prediction set of rules. Step of comparable as that of the RANDOM FOREST-UDRP set of rules the RANDOM FOREST primarily based totally multimodal disorder hazard prediction set of rules step simplest the testing steps carries of extra steps. Given paper paintings on each the form of dataset like prepared and unorganized records. Author labored on unorganized records. While preceding

paintings simplest primarily based totally on prepared records, none of the writer labored on unorganized and semi- prepared records. But this device proposed paintings is relying on prepared in addition to unorganized records. B. Qian, X. Wang, N. Cao, H. Li, and Y.- G. Jiang [2] deliberate the Alzheimer disorder hazard prediction device with the help of EHR data of the affected person. Here they used energetic mastering context to address a authentic problem persisted with the aid of using the affected person. In this the hazard version changed into construct. For that energetic hazard prediction set of rules is used the hazard of Alzheimer disorder. IM. Chen, Y. Mama, Y. Li, D. Wu, Y. Zhang, and C. Youn [3] proposed wearable 2.zero device wherein configuration eager cleanable material that improves the QoE and QoS of the next-technology healthcare device. Chen dependent new IoT primarily based totally records series device. In that new sensor primarily based totally clever cleanable garments created. By the applied of this garments, professional stuck the affected person physiological condition. What's greater, with the help of the physiological records evaluation occur. In this reversal of cleanable clever material consisting of a couple of sensor, wires and cathode with the help of this component factor person can geared up to collect the physiological nation of affected person in addition to emotional fitness repute data used of cloud primarily based totally device. With the help of this material, it stuck the physiological nation of the affected person. Also, for the exam reason, this data is applied. Examined the troubles which might be confronting even as designing wearable 2.zero architecture. The troubles in present device encompass physiological records amassing, poor intellectual impacts, ant wireless for frame quarter networking and Sustainable massive physiological records accumulation and so on. The severe sports achieved on statistics like exam on records, tracking and prediction. Again author classify the useful additives of the clever apparel representing Wearable 2.zero into sensors Integration, electrical-cable-primarily based totally networking, virtual modules. In this, there are various packages pointed out like continual disorder tracking, aged humans care, emotion care etc. Y. Zhang, M. Qiu, C.- W. Tsai, M. M. Hassan, and A. Alamri [4] designed cloud-primarily based totally fitness –Cps device wherein offers with the massive degree of biomedical records. Y. Zhang tested big degree of data improvement with inside the medicinal field. The data is made in the much less degree of time and the ordinary for data is positioned away in diverse configuration so that is the issue that the problem recognized with the massive records. In this designed the Health-Cps device in that improvements lean one is cloud and 2d one is massive records technology. Cloud-like records evaluation, tracking and prediction of records. With the help of this device, an character receives greater records approximately a way to cope with and cope with the great degree of biomedical data within side the cloud. The 3 layers remember records series layer, records control layer and data oriented layer. The records amassing layer positioned away allotted garage and parallel computing. The records control layer used for allotted garage and parallel computing. By this framework diverse obligations are completed with the help of Health-cps gadget, the numerous Health-cps systems. Related to healthcare know through this gadget. L. Qiu, K. Gai, and M. Qiu in [5] proposed telehealth gadget and tested a way to address plenty of health facility information within side the cloud. This paper creator proposed strengthen within side the telehealth gadget that is for the most element depending on the sharing information amongst all of the telehealth offerings over the cloud. Yet, the data sharing at the cloud confronting hundreds of problems like community capability and digital system switches. In this proposed the information sharing on cloud technique for the higher sharing of data via the information sharing ideas. Here deliberate the appropriate method for telehealth sharing model. this model, creator consciousness on transmission probability, community abilities and timing constraints. For this writer concocted new huge information sharing set of rules. By this calculation, customers get the appropriate association of coping with biomedical information. Ajinkya Kunjir, Harshal Sawant, Nuzhat F.Shaikh [6] proposed a great scientific selection-making gadget which predicts the disorder primarily based totally on historic information of patients. In this anticipated numerous sicknesses and inconspicuous instance of affected person condition. Designed a great scientific selection- making gadget applied for

the genuine disorder prediction at the historic information. In that moreover determined numerous sicknesses idea and hid instance. For the notion purpose on this used 2D/three-D graph and pie Charts.And 2D/three-D graph and pie charts illustration purpose. S.Leoni Sharmila, C.Dharuman and P.Venkatesan [13] offers a comparable research of numerous system getting to know method such Fuzzy logic, Fuzzy Neural Network and selection tree. They remember information set to categorise and do examine approximately nearly. As indicated through examine Fuzzy Neutral Network offers 91 %Acuracy for category in liver disorder dataset than different system getting to know set of rules. Author applied Simplified Fuzzy ARTMAP in modified nature of utility domain names and is succesful to carry out category all round productively and giving enormously advanced performances. Author have reasoned that system getting to know algorithms for instance, Naive Bayes and Apriori [14] are very precious for disorder analysis at the given information set. Here little extent information applied for prediction like signs and symptoms or beyond getting to know were given from the bodily analysis. Confinement of this paper they could not remember big dataset, currently a day's medicinal information is growing so wishes to categorise that and category of that data is challenging. Shraddha Subhash Shirsath [15] proposed a RANDOM FOREST-MDRP set of rules for a disorder prediction from an big extent of health facility's prepared and unstructured data. Utilizing a system getting to know set of rules (Neavi- Bayes) Existing set of rules RANDOM FOREST-UDRP simply makes use of an prepared data but in RANDOM FOREST-MDRP middle round each prepared and unstructured data the accuracy of disorder prediction is greater and quick while contrasted with the RANDOM FOREST-UDRP. Here they consider remember huge information.

3. SYSTEM ARCHITECTURE

3.1 Block diagram

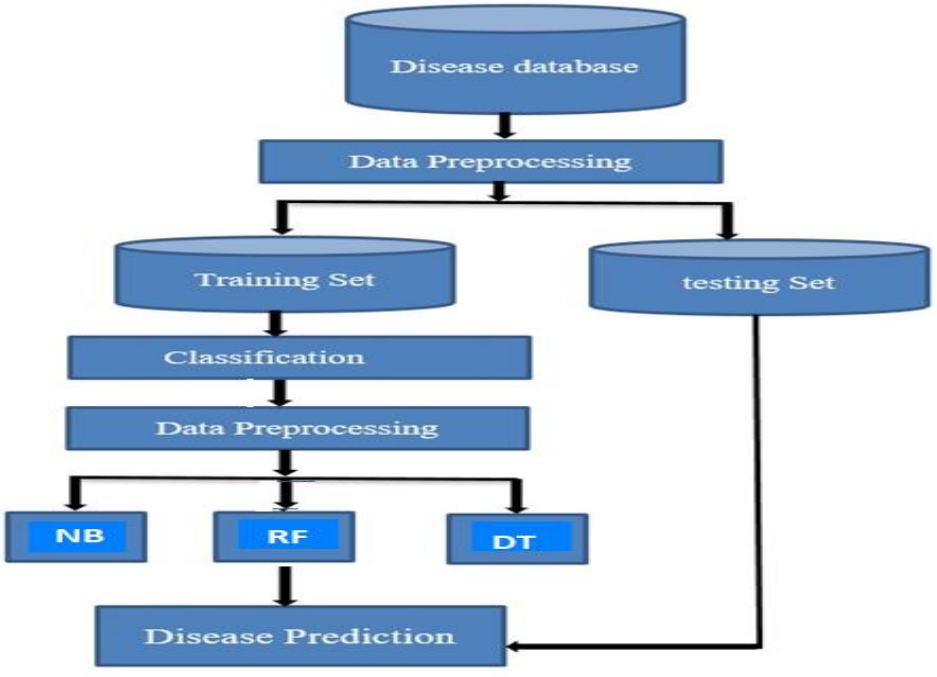

Fig. 1. Block diagram

Initially, we take ailment dataset from UCI gadget studying internet site and this is with inside the shape

of ailment listing with its symptoms. After that preprocessing is achieved on that dataset for cleansing which is putting off commas, punctuations, and white places. And this is used as an education dataset. After that characteristics are extracted and selected. Then we classify that data the usage of type strategies together with DECISION TREE, NAIVE BAYES and RANDOM FOREST. Based on gadget studying we will expect the correct ailment.

3.2 Algorithms and methods

1) Naive Bayes

2) Random Forest

3) Decision Tree

4. RESULT AND DISCUSSIONS

A. Experimental Setup:

All the experimental instances are carried out in Python in congestion with Flask equipment and python interpreter as backend, algorithms and strategies, and the competing type technique at the side of diverse feature extraction technique, and run in surroundings with System having the configuration of Intel Core i5-6200U, 2.30 GHz Windows 10 (sixty-four bit) system with 8GB of RAM

B. Dataset:

Patient disease dataset downloaded from UCI or Kaggle machine learning website.

C. Results:

This segment provides the overall performance of the NAIVE BAYES, RANDOM FOREST, DECISOIN TREE algorithms in phrases of time required and reminiscence and different overall performance measures which include FP measure, precision, recall, and accuracy.

1. SOFTWARE

Python IDE: Pycharm / Jupyter notebook
Python Interpreter/compiler: python 3.8/3.9

- Arduino IDE
- Embedded c

2. ADVANTAGES

- Health status detection becomes easy & accurate
- This system is economical and compact.
- False health diagnosis can be reduced
- Common or repeated health problems can be tracked and recorded

3. DISADVANTAGES

- Continues update in data & algorithms libraries is required
- It required high maintenance.
- It is expensive.
- Need to handle by experienced people in the ML domain

4. APPLICATIONS

- Hospitals to diagnosis before any major treatment
- Small clinics to operate new patient.

5. CONCLUSION

We proposed a disease prediction machine primarily based totally on symptoms. For disease prediction primarily based totally on symptoms, we used a gadget studying set of rules this is DECISION TREE, NAIVE BAYES and RANDOM FOREST. We carried out disease prediction through the use of the NAIVE BAYES set of rules and RANDOM FOREST set of rules.

We evaluate the outcomes among the DECISION TREE, NAIVE BAYES and RANDOM FOREST set of rules and the accuracy of the RANDOM FOREST set of rules is 94% which is extra than the NAIVE BAYES set of rules. We were given accurate disease prediction as output, through giving the enter as sufferers file which assist us to apprehend the degree of prediction. This machine may also lead to low time intake and minimal fee feasible for disease prediction. In the future, we can upload extra sicknesses and are expecting the chance which affected person suffers from the precise disorder.

REFERENCES

[1]. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang,"Disease prediction by machine learning over big data from healthcare communities", ," IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2]. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Springer Data Mining Knowl. Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[3]. IM. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," IEEE Common., vol. 55, no. 1, pp. 54–61, Jan. 2017.

[4]. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "HealthCPS: Healthcare cyberphysical system assisted by cloud and big data," IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

[5]. L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for telehealth in cloud computing," in Proc. IEEE Int. Conf. Smart Cloud (Smart Cloud), Nov. 2016, pp. 184–189.

[6]. Disease and symptoms Dataset –www.github.com.

[7]. Heart disease Dataset-WWW.UCI Repository.com

[8]. Ajinkya Kunjir, Harshal Sawant, Nuzhat F.Shaikh, "Data Mining and Visualization for prediction of Multiple Diseases in Healthcare," in IEEE big data analytics and computational intelligence, Oct 2017 pp.2325.

[9]. Shanthi Mendis, Pekka Puska, Bo Norrving, World Health Organization (2011), Global Atlas on Cardiovascular Disease Prevention and Control, PP. 3– 18. ISBN 978-92-4-156437-3.

[10]. Amin, S.U.; Agarwal, K.; Beg, R., "Genetic neural network based data mining in prediction of heart disease using risk factors", IEEE Conference on Information & Communication Technologies (ICT), vol., no., pp.1227-31,11-12 April 2013.

[13]. Alex M. Goh and Xiaoyu L. Yann, (2021), "A Novel Sentiments Analysis Model Using Perceptron Classifier" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 4, pp. 01-10, DOI 10.30696/IJEEA.IX.IV.2021.01-10

[14]. Dolly Daga, Haribrat Saikia, Sandipan Bhattacharjee and Bhaskar Saha, (2021), "A Conceptual DesignApproach For Women Safety Through Better Communication Design" Int. J. of Electronics Engineering andApplications, Vol. 9, No. 3, pp. 01-11, DOI 10.30696/IJEEA.IX.III.2021.01-11

[15]. Alex M. Goh and Xiaoyu L. Yann, (2021), "Food-image Classification Using Neural Network Model" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 3, pp. 12-22, DOI 10.30696/IJEEA.IX.III.2021.12-22

[16]. Jeevan Kumar, Rajesh Kumar Tiwari and Vijay Pandey, (2021), "Blood Sugar Detection Using Different Machine Learning Techniques" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 3, pp. 23-33, DOI 10.30696/IJEEA.IX.III.2021.23-33

[17]. Nisarg Gupta, Prachi Deshpande, Jefferson Diaz, Siddharth Jangam, and Archana Shirke, (2021), " F- alert: Early Fire Detection Using Machine Learning Techniques" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 3, pp. 34-43, DOI 10.30696/IJEEA.IX.III.2021.34-43

[18]. Reeta Kumari, Dr. Ashish Kumar Sinha and Dr. Mahua Banerjee, (2021), "A Comparative Study Of Software Product Lines And Dynamic Software Product Lines" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 01-10, DOI 10.30696/IJEEA.IX.I.2021.01-10

[19]. MING AI and HAIQING LIU, (2021), "Privacy-preserving Of Electricity Data Based On Group Signature And Homomorphic Encryption" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 11-20, DOI 10.30696/IJEEA.IX.I.2021.11-20

[20]. Osman Goni, (2021), "Implementation of Local Area Network (lan) And Build A Secure Lan System For Atomic Energy Research Establishment (AERE)" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 21-33, DOI 10.30696/IJEEA.IX.I.2021.21-33.

[20]. XIAOYU YANG, (2021), "Power Grid Fault Prediction Method Based On Feature Selection And Classification Algorithm" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 34-44, DOI 10.30696/IJEEA.IX.I.2021.34-44.

[21]. Xiong LIU and Haiqing LIU, (2021), "Data Publication Based On Differential Privacy In V2G Network"Int. J. of Electronics Engineering and Applications, Vol. 9, No. 2, pp. 34-44, DOI 10.30696/IJEEA.IX.I.2021.45-53.