International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2283105

Disease Prediction Based on Symptoms by Using Decision Tree And Random Forest In Machine Learning

Ch. Esther*[1], S. Nayana Sai[2], S. Sushma[3], B. V. R. Gupta[4], Mr. G. Srinivasa Rao[5]
[*1,2,3,4] Department of Computer Science & Engineering, NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh, India
[*5]Assistant Professor Department of Computer Science & Engineering, NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh, India

Article Info
Volume 8, Issue 3
Page Number : 419-427

Publication Issue :
May-June-2022

Article History
Accepted: 10 June 2022
Published: 25 June 2022

ABSTRACT

The medical care space is one of the unmistakable examination fields in the ongoing situation with the fast improvement of innovation and information. Dealing with the colossal measure of information of the patients is troublesome. Taking care of this information through Big Data Analytics is simpler. There are a ton of methodology for the treatment of different infections across the world. AI is an arising approach that aides in expectation, determination of an illness. This venture portrays the expectation of illness in light of side effects utilizing AI. AI calculations, for example, Support Vector Machine, Decision Tree and Random Forest are utilized on the gave dataset and anticipate the sickness. Its execution is finished through the python programming language. The task exhibits the best calculation in light of their exactness. The exactness of a not entirely set in stone by the presentation on the given dataset.

Keywords : CNN, KNN, Machine learning, Disease Prediction.

I. INTRODUCTION

Medication and medical care are probably the most essential pieces of the economy and human existence. There is a colossal measure of progress on the planet we are living in now and the world that existed half a month back. Everything has turned frightful and unique. In this present circumstance, where everything has turned virtual, the specialists and medical attendants are investing most extreme amounts of energy to save individuals' lives regardless of whether they need to peril their own. There are likewise a few distant towns which need clinical offices. Virtual specialists are board-ensured specialists who decide to rehearse online by means of video and telephone arrangements, as opposed to face to face arrangements yet this is beyond the realm of possibilities on account of crisis. Machines are constantly viewed as better compared to people as, with no human mistake, they can perform undertakings all the more productively and with a reliable degree of precision.

Sickness indicator can be known as a virtual specialist, which can foresee the infection of any persistent with no human blunder. Additionally, in conditions like

COVID-19 and EBOLA, a sickness indicator can be a gift as it can distinguish a human's illness with no actual contact. A few models of virtual specialists do exist, yet they don't include the expected degree of precision as every one of the boundaries required are not being thought of. The essential objective was to foster various models to characterize which one of them gives the most reliable expectations. While ML projects differ in scale and intricacy, their general construction is something similar. A few rule-based procedures were attracted from AI to review the turn of events and sending of the prescient model. A few models were started by utilizing different AI (ML) calculations that gathered crude information and afterward bifurcated it as indicated by orientation, age gathering, and side effects. The informational collection was then handled in a few ML models like Decision trees, Logistic relapse, support vector machines, Random Forest.

As indicated by ML models, the exactness differed. While handling the information, the info boundaries informational collection was provided to each model, and the infection was gotten as a result with divergent exactness level. The model with the most noteworthy precision has been chosen.

Information mining is cycle of examining mass measure of information to consequently find the intriguing consistencies or affiliations which with regards to go lead to worked on comprehension of the first cycles [1]. There are two classifications of information mining: 1. Information mining in Descriptive, 2. Information mining in Predictive. Clear information mining sums up or sums up the overall properties of the information in the data set. Prescient information mining look through the induction on the current information to make expectations [2]. Information mining has a few undertakings, for example, affiliation rules, grouping, expectations and bunching and so on. Arrangement are administered learning strategies which groups information into predefined class name. It is quite possibly of the most valuable method in information mining; this strategy is generally used to construct models that anticipate future information patterns. The fundamental point of the characterization methods is to investigate the info information and to anticipate the exactness for the future work. In clinical field information mining assumes an essential part to find the connection between persistent information and clinical informational index from the huge data set.

II. LOGISTIC REGRESSION MODELING ON CROP DATA

Relapse examination is a strategy for researching utilitarian connections among factors. The relationship is communicated as a situation or a model interfacing the reaction or ward variable and at least one illustrative or indicator factors. At the point when the reaction variable is quantitative, the standard hypothesis of various straight relapse (MLR) investigation holds great. A large portion of the factors in this model are quantitative in nature. Assessment of boundaries in this relapse model depends on four essential presumptions. In the first place, reaction or ward variable is directly related with illustrative factors. Second, model mistakes are freely and indistinguishably conveyed as typical variable with mean zero and normal fluctuation. Third, autonomous or illustrative factors are estimated without blunders. The last supposition that is about equivalent unwavering quality of perceptions. Be that as it may, circumstances where the reaction variable is subjective are likewise very normal and happen widely in measurable applications. For instance to decide the gamble factors for malignant growth in people, information could be gathered on a few factors, for example, age, sex, smoking, diet and so on. The reaction variable here, is dichotomous that either the individual has malignant growth (Y=1), or didn't have disease (Y=0).In such cases, the standard

MLR hypothesis isn't fitting. Rather, the measurable model liked for the examination of such twofold (dichotomous) reactions is the double calculated relapse model, grew principally by Cox (1958) and Walker and Duncan (1967).Thus Logistic Regression is a numerical demonstrating approach that can be utilized to portray the relationship of a few free factors to (say) a paired (dichotomous) subordinate variable. Later on, the models to manage polytomous (multinomial) reactions developed. In the event that, our reaction variable in model is subjective in nature, then probabilities of falling this reaction variable in different classifications can be demonstrated instead of reaction variable itself, utilizing same model yet there are number of requirements as far as suppositions of numerous relapse model. In the first place, since the scope of likelihood is somewhere in the range of 0 and 1, while, right hand side capability in the event of various relapse models is unbounded. Second, mistake term of the model can take just restricted values and blunder fluctuation are not constants but rather relies upon likelihood of falling reaction variable in a specific classification. For the most part, ordinary hypothesis of different straight relapse (MLR) examination has been applied for a quantitative reaction variable, while for the subjective reaction variable or all the more explicitly for twofold reaction variable considering elective models is better. With respect to model, taking into account following situations:

- A pathologist might be intrigued whether the likelihood of a specific illness can be anticipated utilizing culturing practice, soil surface, date of planting, climate factors and so forth as indicator or free factors.
- A business analyst might be keen on deciding the likelihood that an agro-based industry will bomb given various monetary proportions and the size of the firm (for example enormous or little).

Typically discriminant examination could be utilized for tending to every one of the above issues. Be that as it may, in light of the fact that the free factors are combination of downright and ceaseless factors, the multivariate ordinariness suspicion may not hold. Underlying relationship among different subjective factors in the populace can be measured utilizing number of elective procedures. In these strategies, essential interest lies on subordinate element which is reliant upon other free factors. In these cases the most ideal method is either probit or calculated relapse examination as it makes no suspicions about the appropriation of the autonomous factors. The reliant element is known as reaction factor. In this model structure process, different log chances connected with reaction factors are demonstrated. As an exceptional case, on the off chance that reaction factor has just two classes with probabilities p1 and p2 individually then the chances of getting classification one is (p1/p2). In the event that log (p1/p2) is displayed utilizing Analysis of Variance (ANOVA) sort of model, it is called logit model. Once more, on the off chance that a similar model is being treated as relapse type model, it is called calculated relapse model. From a genuine perspective, logit and calculated are names of changes. In the event of logit change, a number p between values 0 and 1 is changed with log {p/(1-p)}, though if there should arise an occurrence of calculated change a number x between - ∞ to + ∞ is changed with {ex/(1 + ex)} capability. It very well may be seen that these two change are opposite of one another for example if logit change is applied on calculated change capability, it offers some benefit x and likewise, assuming strategic change is applied to logit change capability it offers some incentive p. A decent record of writing on strategic relapse are accessible, to refer to a couple, Fox(1984), Klienbaum (1994) and so forth.

Direct relapse is viewed as to make sense of the requirements in utilizing such model when the reaction variable is subjective. Consider the accompanying basic direct relapse model with single indicator variable and a parallel reaction variable:

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, i= 1, 2, … , n

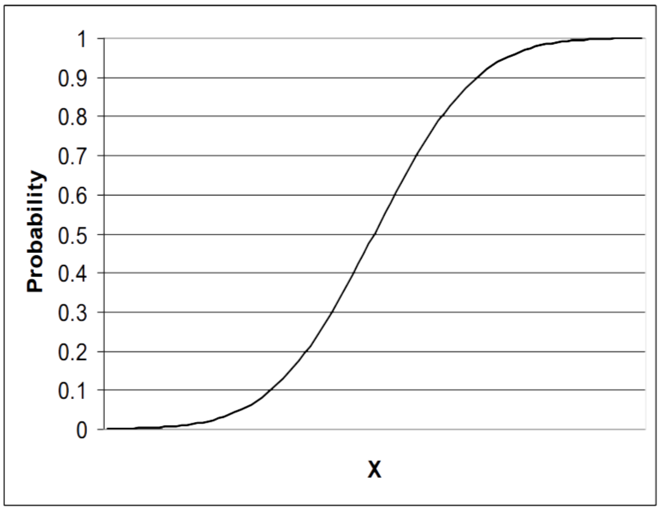
where the result Yi is double (taking qualities 0,1), εi ~ N(0,σ2ε) , and are free and n is the quantity of perceptions.
Let πi indicate the likelihood that Yi =1 when Xi = x, for example
πI =P(Yi =1|Xi = x) =P(Yi =1)
thus P(Yi =0)=1-πi
Under the presumption E(εi) 0 , the normal worth of the reaction variable is
E(Yi)=1.(πi)+0.(1-πi)= πi
In the event that the reaction is parallel, the mistake terms can take on two qualities, specifically,
εI 1πi when Yi =1 εi 1πi when Yi =0
Since the blunder is dichotomous (discrete), ordinariness supposition that is disregarded. Also, the mistake change is given by:
V(εi) = πi (1-π)2 + (1-π)(- π)2
= πI (1-π)
It very well may be seen that fluctuation is a component of I π 's and it isn't consistent. Subsequently the presumption of homoscadasticity (equivalent fluctuation) doesn't hold.

Strategic relapse is ordinarily suggested when the autonomous factors don't fulfill the multivariate ordinariness presumption and simultaneously the reaction variable is subjective. Circumstances where the reaction variable is subjective and free factors are combination of unmitigated and consistent factors, are very normal and happen broadly in factual applications in farming, clinical science and so on. The factual model liked for the examination of such twofold (dichotomous) reactions is the paired calculated relapse model, grew basically by a scientist named Cox during the last part of the 1950s. Processes creating sigmoidal or extended S-formed bends are very normal in horticultural information. Calculated relapse models are more proper when reaction variable is subjective and a non-straight relationship can be laid out between the reaction variable and the subjective and quantitative elements influencing it. It resolves the very questions that discriminant capability investigation and different relapse do yet with no distributional suppositions on the indicators. In strategic relapse model, the indicators need not need to be regularly disseminated, the connection among reaction and indicators need not be direct or the perceptions need not have equivalent change in each gathering and so on The issue of non-ordinariness and heteroscadasticity prompts the non appropriateness of least square assessment for the straight likelihood model. Weighted least square assessment, when utilized as another option, can cause the fitted qualities not obliged to the span (0, 1) and hence can't be deciphered as probabilities. Also, a portion of the blunder change might emerge to be negative. One answer for this issue is just to compel π to the unit span while holding the straight connection among π and regressor X inside the stretch. Along these lines

Nonetheless, this obliged direct likelihood model has specific ugly highlights, for example, unexpected changes in slant at the limits 0 and 1 making it hard for fitting similar on information. A smoother connection among π and X is by and large more reasonable. To address this issue, a positive droning (for example non-diminishing) capability is expected to change (β0 + β1xi) to unit stretch. Any combined likelihood dispersion capability (CDF) P, meets this necessity. That is, respecify the model as πi = P (β0 + β1xi). In addition, it is favorable on the off chance that P is completely expanding, for, the change is coordinated, so that model can be modified as P-1 (πi) = (β0 + β1xi), where P - 1 is the backwards of the CDF P. In this way the non-direct model for itself will become both smooth and symmetric, drawing nearer π = 0 and π = 1 as asymptotes. From that point most extreme probability technique for assessment can be utilized for model fitting.

III. PROPERTIES OF LOGISTIC REGRESSION MODEL

The Logistic reaction capability looks like a S-shape bend, a sketch of which is given in the accompanying figure. Here the likelihood π at first increments gradually with expansion in X, and afterward the increment speeds up, at long last balances out, yet doesn't increment past 1.



The state of the S-bend can be repeated on the off chance that the probabilities can be displayed with just a single indicator variable as follows:

π= P(Y=1|X= x)= 1/(1+e-z)

where z = β0 + β1x, and e is the foundation of the regular logarithm. Subsequently for mutiple (say r) logical factors, the likelihood π is displayed as π = P(Y=1|X1 = x1 ...Xr = xr) =1/(1+e-z)

where z= β0 +β1 x1 +...+βr xr .

This condition is known as the strategic relapse condition. It is nonlinear in the boundaries β0, β1… βr . Displaying the reaction probabilities by the strategic dissemination and assessing the boundaries of the model comprises fitting a calculated relapse. The strategy for assessment by and large utilized is the greatest probability assessment technique. To make sense of the fame of strategic relapse, let us consider the numerical structure on which the calculated model is based. This capability, called f (z), is given by

f (z) = 1/(1+e-z) , - ∞ < z < ∞

Presently when z = - ∞, f (z) =0 and when z = ∞, f (z) =1. In this manner the scope of f (z) is 0 to1. So the calculated model is famous in light of the fact that the strategic capability, on which the model is based, gives

- Gauges that lie in the reach somewhere in the range of nothing and one.
- An engaging S-formed depiction of the joined impact of a few illustrative factors on the likelihood of an occasion.

IV. DECISION TREE

Information Mining is an advancement that through examining and dealing with huge informational indexes or enormous, deficient and boisterous crude information from the data set, individuals can draw possibly valuable information, data, models and patterns which are obscure ,to have a profound comprehension of information and more successful utilization of data[1],[2]. Simultaneously, the course of information mining can be seen as a course of drawing models from the information, which can be delegated affiliation model, order model, relapse model, bunching model, exception investigation model and time series models as indicated by the models' functional applications. Among the models, the characterization model is primarily used to investigate the directed information, which can sum up a model that can recognize the information object ID by examining the preparation informational indexes. In the order model, the most well known strategy is the choice tree technique, choice tree for characterization and like a stream graph. A root hub of the choice tree frequently contains the most properties, and its inward hubs address a trial of each and every characteristic. The parts of the choice tree address test results and the tree leaf is in many cases the agent of the class. The most significant level of the tree is the root hub, which is the start of the choice tree. In addition, the quantity of youngster hubs of each and every hub has something to do with the calculation utilized by the choice tree. For instance, each choice tree hub has two branches in the event that CART calculation is utilized with the choice tree called twofold tree, and the choice tree is named multi-tree when its hub contains multiple youngster hubs or branches. Each part of the tree is either another choice hub, or the finish of the tree, which is known as the leaves. The most common way of utilizing the choice tree order the records in the training is really a crossing cycle from the top as far as possible along the choice tree. Various responses to each address experiencing on each hub will prompt various branches, till the leaf hub. This can be known as the choice tree's course of investigation. Utilizing the choice tree strategy for information order, it ordinarily makes two strides. Initial, an underlying choice tree ought to be produced from the preparation sets. Furthermore, the above choice tree will be changed and changed, which is against a case that a few parts of the underlying choice tree are built by the strange information of preparing test sets. Generally the pruning strategy is to utilize the measurable techniques to eliminate the most temperamental branches or youngster trees, so that to work on the speed of estimate and grouped recognizable proof and the capacity of accurately arrange new information.

ID3 calculation is the most impact calculation in Decision-Tree calculation, it previously proposed by Quinlan, ID3 is created from CLS calculation, in which calculation, property picked through the data gain esteem. The size of the obtained data gain is greater, the less the vulnerability is. With the goal that grouping effectiveness and quality are enormously raised, and extremely broad in actuality .It is by a wide margin the most famous calculation in choice tree region. ID3 calculation is a voracious calculation. It utilizes a hierarchical, gap and rule technique, through ceaseless cycle handling, bit by bit refinement, view as a generally exact until the choice tree. It's structural choice tree is the hierarchical like IF - Then manages tree. Utilizing this strategy the constitutive can be easier, structural tree structure in structural cycle estimation is lesser, and particularly reasonable for enormous scope informational collection's utilization to tackle issues. The fundamental thought of the ID3 calculations is as per the following:

(1) Select the entire preparation models' size of coordinating PN irregular subset of PN as of now.
(2) Based on the data entropy drop speed for the norm, the determination of each and every test ascribes, shaping the ongoing subsets of choice tree.
(3) Order to examine all preparing models; figure out the ongoing choice tree exemption, in the event that no special cases, preparing finished.
(4) Combining the ongoing subsets of some preparation for certain models in (3) the special case tracked down in shaping new subsets, turn (2).

The learning methodologies of ID3 calculation could be portrayed as the accompanying a few perspectives:

(1) When started to lay out dynamic tree, root hub contain all the preparation tests.
(2) If a hub tests are important for a similar class, the hub becomes leaf hubs, and labels for this classification; any other way, will the data entropy as motivation information to pick fitting to the branch trait, isolated into a few little subset tests, this property will turn into the relating hub testing credits. In the ID3 calculation, all the property estimations are discrete qualities, so if existing in the first information for esteem, it requirements to its separation.
(3) A test characteristic qualities are every one of the relating a will make branch, likewise compares to an ordered subset.
(4) Recursively apply the over (1) - (3) handling of information handling. So if a property in a hub to show up, so it will not show up on the hub tree

delivered child after the hub. ID3 calculation produces not contained the choice tree has rehashed choice child tree.

(5) stop condition is a hub of the examples of all have a place with a similar class; Or is without characteristics can apply to parcel the ongoing examples, in the event that seem this sort of situation, as per the rule of the minority is subordinate to the larger part will be mandatory for the ongoing hub leaf hubs. In the choice tree calculation, the construction of the utilization of the data gain regularly strategies to assist with figuring out which created when every hub is the right properties ought to pick, so they can choose the most elevated data gain, to be specific entropy decrease degree greatest characteristics as test credits of the ongoing hub, to make its acquired after the division of the preparation tests required least measure of data subset. That is, utilizing the characteristics of the ongoing hub tests contained partitions, will make all examples delivered the various classifications of blended subset to limit the data hypothesis, so utilizing this choice tree is developed to protest can be really diminished the quantity of characterization, subsequently guaranteeing a choice tree created reasonable and basic. ID3 calculation created by choice tree of the particular strides as follows: we accepted known PN, then, at that point, for preparing subset. In the event that every one of the models PN preparing subset is a model, are undeniably created a Yes hub and end; If prepared in all cases are PN child for counterexample, then produce a No hub and end; Otherwise the technique decision as per the calculation one quality set A worth kind A, for A1, A2,... An, and produce new hub. Will the case preparing subset as indicated by its property PN A worth division, age n A subset PN1, separately, for PN2 recall... PNn. Will this calculation in every subset recursively on application. Since in the ID3 calculation, the data entropy drops speed is one of the keys to choose the test credits of standard (the data entropy is the decay of assurance down). Data in light of entropy of quality the determination cycle is as per the following: Assume a PN contains P preparing subset of models and N a positive illustration of a negative, then we put PZ and PF is preparing subset of 2 preparation PN PZ called the subsets, which is the positive model sets, and is known as the PF counterexample set of preparing subset. Is an illustration of positive model sets having a place with likelihood for PZ p/(p + n), have a place with the model set PF likelihood for a n/(p + n), then the data entropy can be communicated as:

(1) We pick a property An as choice tree roots test credits with m an alternate, A discrete qualities, A1, A2... Am, they will prepare PN into m a subset PN2 subsets, PN1... PNi PNm, expecting that have A positive models and PI ni A negative model, PNi data entropy for subset I (PI, ni), basis with property A test credits assumptions for the data entropy for.

(2) Thus, the data gain is From condition (3), which should be visible when the condition (2) the worth of (A) the hours, the data gain more prominent increase (A), for example quality A characterized data to give, the more noteworthy the choice An after which the more modest the vulnerability. In the ID3 calculation, we will choose data gain greatest increase (A) the greatest properties as the root, and afterward A choice tree to new division recursively subset of the activity of comparative, can create need choice tree.

Many elements influence agrarian result, the light from the actual information we can't clear what variables are the fundamental variables, and what it means for the farming result esteem. In this way, we are worried about the improvement of rural creation worth might think the unified design factors are dissected, and afterward a choice tree to cut by choice tree will influence the moderately little out factors, the fundamental variables recognized, producing rules. In this venture, we picked the Duration of the harvest, Average Temperature, pH Values, Average Rainfall, Nitrogen(N), Phosphorus(P), Potassium(K), like choice precipitation in agrarian result worth of information characteristic, on the grounds that horticultural result esteem is as per the entire province rural result worth of to figure, it isn't vital area factor, the evacuation of the first information [8] got in the above table 4.1. Then, at that point, information structure is summed up handling, the low degrees of unique information into an elevated degree of idea to deal with information mining.

V. SUPPORT VECTOR MACHINES

SVM is a one of the well known AI calculations for relapse, characterization. A managed learning calculation investigations information utilized for order and relapse. SVM displaying includes two stages, right off the bat to prepare an informational collection and to get a model and then, to utilize this model to foresee data of a testing informational collection. A Support Vector Machine (SVM) is a discriminative classifier officially characterized by an isolating hyper-plane where SVM model addresses the preparation data of interest as focuses in space and afterward planning is done so the focuses which are of various classes are separated by a hole that is essentially as wide as could be expected. Planning is finished in to similar space for new data of interest and afterward anticipated on which side of the hole they fall.

In SVM calculation, plotting is finished as every information thing is taken as a point in n layered space where n is number of elements, with the worth of each component being the worth of a specific direction. Then, at that point, order is performed by finding the hyper-plane that isolates the two classes quite well.

VI. CONCLUSION

Anticipating the illness, the in beginning phase in light of side effects are essential for the treatment. We have gathered a dataset from the Kaggle and we have prepared the 4 different AI model Logistic relapse, Decision tree, Random Forest and backing vector machines. Among every one of them irregular backwoods gets the precision of the 93%. Practically all the ML models gave great exactness values. As certain models were reliant upon the boundaries, they couldn't anticipate the illness and the precision rate was very low. When the illness is anticipated, we could without much of a stretch man-age the medication assets expected for the treatment. This model would assist in bringing down the expense expected in managing the sickness and would likewise further develop the recuperation cycle.

VII. REFERENCES

[1]. A. Mir, S.N. Dhage, in 2018 Fourth International conference on Computing Communication Control and Automation (ICCUBEA) (IEEE, 2018), pp. 1-6

[2]. Y. Khourdifi, M. Bahaj, Heart disease prediction and classification using Machine Learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst. 12(1), 242 (2019)

[3]. S. Vijayarani, S. Dhayanand, Liver disease prediction using svm and naïve bayes algorithms, International Journal of Science, Engineering and Research (IJSETR) 4(4), 816(2015)

[4]. S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7,81542(2015)

[5]. T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent Parkinson disease prediction using machine learning

algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568(2013)

[6]. A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, J. Schweinle, Clinical signs and symptoms predicting influenza infection, Archives of internal medicine 160(21), 3243 (2000)

[7]. R.D.H.D.P. Sreevalli, K.P.M. Asia, Prediction of diseases using random forest classification algorithm

[8]. D.R. Langbehn, R.R. Brinkman, D. Falush, J.S. Paulsen, M. Hayden, an International Huntington's Disease Collaborative Group, A new model for prediction of the age of onset and penetrance for huntington's disease based cag length, Clinical genetics 65(4), 267(2004)

[9]. K. Kourou, T.P. Exarchos, K.P. Exaechos, M.V. Karamouzis, D.I. Fotiadis, Machine Learning Applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13, 8 (2015)

[10]. T. Karayilan, O. Kili c, in 2017 International Conference on Computer Science and Engineering (UBMK) (IEEE, 2017), pp. 719-723

[11]. M. Chen, Y. Hao, K. Hwang, L.Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, IEEE Access 5, 8869(2017)

[12]. S. Chae, S. Kwon, D. Lee, predicting infectious disease using deep learning and big data, International Journal of environmental research and public health 15(8), 1596 (2018)

[13]. A.U. Haq, J.P. Li, M.H. Memon, S. Nazir, R. Sun, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mobile Information Systems 2018(2018)

[14]. Amalgam Approach to Computer Vision – Deep Learning and IoT Approaches for Aerial View Cultivating Monitoring – IEEE- COMSOC-MMTC Communications- Frontiers

Cite this article as :

Ch. Esther, S. Nayana Sai, S. Sushma, B. V. R. Gupta, Mr. G. Srinivasa Rao, "Disease Prediction Based on Symptoms By Using Decision Tree And Random Forest In Machine Learning ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 3, pp. 419-427, May-June 2022. Available at doi : https://doi.org/10.32628/CSEIT2283105
Journal URL : https://ijsrcseit.com/CSEIT2283105