ENGINEERING IN ADVANCED RESEARCH SCIENCE AND TECHNOLOGY

ISSN 2040-7467
Vol.04, Issue.02
December-2024
Pages: 574-584

Syncronova Health Intelligence: Exploring New Frontiers in Intelligent Health Predictions Using Machine Learning with Feature Analysis

1. Krishna yedukondalu, 2 Torlapati Venkateswara Rao

[1] *M. Tech., Dept of CSE, Sree vahini institute of science and technology Tiruvuru, JNTUK India,*
[2,] *M. Tech., Dept of, CSE, Sree vahini institute of science and technology Tiruvuru, JNTUK India,*
krishna.Y1086@gmail.com[1] , vinnuyadav907@gmail.com[2]

Abstract: -

The Project implements 3 linear models and one deep learning model: Naïve Bays. K-Nearest Neighbours network and Support Vector Machine to determine the efficiency of the two for diagnosing diabetes as well as heart diseases from datasets available from the UCI data repository. Thirdly on the performance of the algorithms, each algorithm has been incorporated into a prediction engine and the engine is served over an 'API'.to investigate their performance on diabetes and heart disease datasets obtained from the UCI data repository. In addition, on the comparison of the algorithms, each algorithm has been integrated into a prediction engine and exposed over an API. The project also incorporates a web-page working platform for tight cooperation of researches and the doctors. Based on the results indicating that our prediction engine can identify the existence of the disease and provide accurate prediction, it is also possible to enhance performance via complex deep learning methodologies.

Keywords: -Convolutional neural network, UCI data repository, machine learning, deep learning

Introduction: -

Liver is large and fleshy organ located in right side of the abdomen at the pit of the stomach. As big as a fist, heart is approximately 3 pounds in weight , it is reddish brown in colour and has feeling of rubber. The heart is divided into two large areas of the right lobe and the left lobe. The gallbladder is located beneath the heart, and parts of the pancreas or intestines. The heart and these organs collectively function for digestion, assimilation and metabolism of food. Generally, the heart is required to filter the blood that arrives from the digesting tract to be distributed to the rest of the body. The heart equally purifies chemicals and catabolises drugs. While it does so, the heart develops bile that turns back into the intestines. The heart has its own enzymes and protein synthesis factories forBlood clotting proteins or other proteins.

Heart disease is any disturbance of heart function that results in sickness of the organ. The heart performs many dangerous operations in the body and if the heart should become diseased or damaged, the loss of these functions can lead to serious harm to the body. There are noted that heart disease is also known as hepatic disease. Heart diseases can be defined as a broad term that would include all the possible complications that lead to heart dysfunction. Normally, heart condition worsens when more than 75% of the heart mass or three quarters of the cardiac tissue is damaged.

The lungs is a sort of muscular organ to pump blood in body part and the cardiovascular part of the body which is associated with the heart. Cardiovascular system also consist

of blood vessels like veins, arteries and capillaries. These blood vessels beheart blood in all parts of the body. Irritant or irregular circulation of blood within and from the heart leads to several categories of diseases, which are popularly called cardiovascular diseases (CVD). Cardiac illnesses are global leading causes of death. As confirmed by the recent WHO global survey, about 17,5 million combined end due to heart attacks and strokes. Majorities of cardiovascular diseases deaths, which is greater than seventy five percent happen in mostly middle-income and low-income countries. Further, 80% of CVD deaths are caused by stroke and coronary attack [1] . Thus, identification of cardiac abnormalities at the early stage as well as/relation of tools used to forecast heart diseases could significantly reduce the number of deaths and assist the physicians in the development of precise treatment strategies, which in its turn lowers the indicators of mortality from cardiovascular diseases. This is because most patient data is today available (Big Data in Electronic Health Record System) courtesy of the improvement of the advance healthcare system across the world and should be utilized when it comes to developing the predictive models of cardiovascular diseases. Machine learning or data mining is a discovery technique for understanding large data set from a diverse angle and packaging it into useful data. Data Mining is defined as: “the non-trivial process of identifying previously unknown, hidden patterns of data ” [3] . At present, a vast volume of information relining to ailments identification, people and so on are created by healthcare sectors. Data mining offers several methods which extract concealed information or relationships from data. Therefore in this paper a machine learning algorithm is put forward for the development of a heart disease prediction system that were tested on two open access heart disease prediction datasets.

Data mining is the discovery of useful information from large databases through the application of computer programs. Knowledge discovery is most beneficial in an explorative kind of analysis due to assertion of nontrivial from numerous bodies of evidence. Thus, there is a high possibility that using the concepts of medical data mining can clearly and efficiently unveil the hidden factors that exist in the datasets of the clinical scope. All these patterns can be used to diagnose health issues. However, the raw medical data available with most organizations are spread, large in number and heterogeneous in nature These data have to be assimilated systematically. This collected data can be then integrated to form an medical information system. Business questions and techniques for predicting the various diseases in the healthcare reconcile with data mining where Data mining is a user-oriented approach to novel and hidden patterns in the Data. Disease prediction is one of the areas in which data mining is very crucial. The results regarding the predictions of heart disease are discussed under the classification algorithms of this paper. It is established that these latent patterns could be employed in diagnosing health in health care data. The datum is a very effective way of getting comprehensive level of the latest and infinite patterns or relations in the datum. The above information which is identified can help the healthcare administrators to receive better services. This was seen most prominently in these areas as the leading killer the heart disease was prominent in point form among the victims the countries like India United States. Explorating the different kinds of heart – based problems, Data mining techniques include clustering, Association Rule Mining, classification algorithms include Decision tree [2], other techniques which are C4.5 algorithm, Naive Bayes [4]. These algorithms can be used to improve the storing of data for functional and legal requirements.

Literature Survey: -

Heart disease Patients have been rising overtime due to reckless drinking of alcohol, breathing of toxic gases, consumption of contaminated foods, pickles and drugs. Use of such aids for classification may also decrease the load on doctors. In this paper, the performance of the selected classification algorithms for the classification of some heart patient datasets is compared and assessed. The classification algorithms discussed in the present work are Naïve Bayes

classifier, C4.5, Back propagation Neural Network algorithm, and SVMs. These algorithms are evaluated based on four criteria: Positivity in the orchi's test has significant values like Accuracy, Precision, Sensitivity and Specificity.

Author: -Bendi Venkata Ramana

A variety of studies have been carried out to diagnose heart diseases using different data mining techniques, with each study achieving varying degrees of success. An example is the work of Sellappan Palaniappan et al., who developed an Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques such as Naive Bayes, Neural Networks, and Decision Trees. Each technique offers unique strengths in predicting heart disease. The IHDPS utilizes hidden patterns and relationships between data attributes to make predictions. This system is web-based, user-friendly, and easily expandable for future use, offering a valuable tool for heart disease prediction.

Similarly, Heon Gyu Lee et al. proposed a novel technique for developing multi-parametric features related to Heart Rate Variability (HRV), which can be both linear and nonlinear in nature. To achieve this, they employed various classifiers, including Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree), and SVM (Support Vector Machine). Their approach enhances the accuracy of heart disease prediction by analyzing a wide range of features and relationships within the dataset, enabling more precise classification of heart disease risk.

Niti Guru et al. focused on predicting heart disease, blood pressure, and sugar levels using neural networks. Their dataset contained records with 13 attributes, and they used a supervised neural network with a backpropagation algorithm for both training and testing the data. This work demonstrates the potential of neural networks in handling complex healthcare data and improving the prediction of various health conditions, including heart disease.

The work of Carlos Ordonez explored the identification of constrained association rules for heart disease prediction. In his approach, the dataset consisted of records from patients with heart disease, and three constraints were introduced to reduce the number of patterns in the data. These constraints helped refine the prediction model by focusing on the most relevant attributes. Additionally, Franck Le Duff et al. created a decision tree model based on a patient database for diagnosing medical conditions, including heart disease. Latha Parthiban et al. proposed a Coactive Neuro-Fuzzy Inference System (CANFIS) approach for heart disease prediction, combining neural network capabilities with fuzzy logic and genetic algorithms, offering a more adaptable and efficient method for prediction.

In another study, Kiyong Noh et al. developed a classification method based on multi-parametric features extracted from ECG-based HRV data, enhancing the accuracy of heart disease pattern classification. Furthermore, V. Manikantan and S. Latha applied medicinal data mining methods to analyze heart disease diagnosis. By employing techniques such as association rule mining and clustering algorithms like K-means, they enhanced the heart disease prediction process. These approaches also helped refine the data, improving the accuracy of predictions for heart disease and heart attacks.

These research efforts illustrate the growing significance of data mining and machine learning techniques in diagnosing heart disease. The various algorithms and models developed by these researchers contribute significantly to improving diagnostic accuracy and enhancing predictive capabilities in the medical field. By identifying meaningful patterns in healthcare data, these approaches have the potential to revolutionize the early detection and treatment of heart disease, offering valuable tools for medical professionals in managing patient health.

Authors: Sellappan Palaniappan and Rafiah Awang

SURVEY ON PREDICTION OF HEART MORBIDITY USING DATA MINING TECHNIQUES

Data Mining can be defined as the process of identification and discovery of covert, hitherto unknown relationships between data points and this process is not trivial in any sense of the term. Data mining technology offers more of a user perceptive of new patterns unthought-of and unable to be seen by humans. This paper gives information about different existent techniques, problems and challenges occurring in connection with them. The discovered knowledge can be used to enhance the quality of the service by the heath care administrators and also be utilized by the medical practitioners to reduce number of adverse drug effects, to provide least expensive therapeutically equivalent recommendations. In this paper we talk about the common Data mining methods that are Decision Tree algorithm, Naîve Bayes and Neural Network which are used for the prediction of disease.

Related Work: -

Algorithms and Implementations: -

SUPPORT VECTOR MACHINES

Support Vector Machines (SVM) is a type of supervised learning that is used for classification and as well as a regression technique. Yet it is mostly used in classification problems. In this algorithm, we bring each data item and make an n-dimensional point out of it with the value of each feature being the value of a certain coordinate. Then there is classification in which we find the hyper-plane that clearly separates the two classes – see the snapshot below. In the application of the SVM algorithm, a kernel is used in practice. The optimization of the hyperplane in linear SVM is conducted by changing the problem into some form of linear algebra that is beyond the scope of the discussion of this introduction to SVM. A very useful observation is that linear SVM can be reexpressed as using the inner product of any two given instances rather than the instances themselves. The inner product between two vectors by sum of the multiplication of each pair of the inputs. For instance, the dot-product of the vectors [2, 3] and [5, 6] is obtained as 2*5 + 3*6 which equals 28. The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

f(x) = B0 + sum(ai * (x,xi))

Naïve Bayes

Naive Bayes is a simple yet effective machine learning algorithm commonly used for classification tasks. It operates as a probabilistic model, leveraging Bayes' theorem to make predictions. This classifier is particularly popular in text classification tasks, such as spam detection, due to its simplicity, efficiency, and ability to handle large datasets with ease.

In probabilistic classification, the goal is to determine the likelihood that a set of features belongs to a particular class, and then assign the class with the highest probability. To calculate this, Bayes' theorem is employed, which updates the probability of a class based on new evidence (features). However, directly calculating the normalization factor (the total probability of the features) is often impractical, so this term is usually ignored. Instead, the focus is placed on calculating the likelihood of the features given the class and the prior probability of each class.

Naive Bayes simplifies this calculation by assuming that the features are conditionally independent given the class. This assumption allows the likelihood of the features to be expressed as the product of the individual probabilities of each feature given the class. While this assumption is rarely true in real-world data, the classifier tends to perform well despite it. This simplification makes Naive Bayes an efficient and scalable model, particularly when the relationship between features is not highly complex.

Although the assumption of feature independence is "naive," the Naive Bayes classifier is often surprisingly accurate, especially in cases where features are weakly correlated. It

provides a fast and effective way to make predictions, particularly in domains such as text classification, where the independence assumption is less likely to hinder performance**:**

$$P(c_i|x_0,\ldots,x_n) \propto P(x_0,\ldots,x_n|c_i)P(c_i)$$
$$\propto P(c_i)\prod_{j=1}^{n} P(x_j|c_i)$$

Calculating the individual P(x_j | c_i) terms will depend on what distribution your features follow. In the context of text classification, where the features are typically word counts or frequencies, the multinomial Naive Bayes model is often used. It assumes that each word in a document contributes independently to the document's classification, making it a suitable model for document categorization tasks such as spam detection or sentiment analysis. In such cases, the algorithm learns the probability of each word occurring in documents of each class, and during classification, it computes the likelihood of a document belonging to each class based on the words it contains.

The performance of Naive Bayes also depends on how well the features represent the underlying structure of the data. For instance, in text classification, the presence or absence of certain words may be highly indicative of the document's class, but if features are poorly chosen or irrelevant to the classification task, the model's performance may degrade. Nonetheless, Naive Bayes typically requires less feature engineering compared to more complex models like support vector machines (SVMs) or decision trees.

Despite its advantages, Naive Bayes does have some limitations. Its primary drawback is the assumption of feature independence, which can sometimes lead to suboptimal performance when the features are highly correlated. Additionally, the algorithm tends to perform poorly if the features are not adequately represented by the assumed distributions (e.g., assuming a Gaussian distribution for features that are not normally distributed). Moreover, Naive Bayes can struggle with rare or unseen features, as it may assign them a zero probability, leading to a failure in prediction. To mitigate this issue, smoothing techniques like Laplace smoothing are often used to handle the case of zero probability.

In conclusion, Naive Bayes is a simple, efficient, and highly effective classification algorithm. It excels in scenarios where the assumption of conditional independence holds or when the data is high-dimensional. Its ability to handle large datasets, combined with its minimal training requirements and robustness to noise, makes it a valuable tool in machine learning. While it is not always the most accurate model in every case, its speed, scalability, and ability to handle complex, high-dimensional data with ease ensure that it remains a staple in many machine learning pipelines. Whether for text classification, spam detection, or medical diagnosis, Naive Bayes continues to be a go-to algorithm for a wide range of classification tasks.

KNN Algorithm:-

The K-Nearest Neighbors (KNN) algorithm is a fundamental and widely-used supervised machine learning technique employed in both classification and regression tasks. Originally introduced by Evelyn Fix and Joseph Hodges in 1951, and later expanded by Thomas Cover, the algorithm has become a cornerstone of pattern recognition and classification problems due to its simplicity and effectiveness. In essence, the KNN algorithm classifies new data points based on the majority class of their nearest neighbors in a feature space, making it an intuitive approach for many machine learning applications.

KNN is part of the non-parametric class of algorithms, which means it does not assume any specific form or distribution for the underlying data, unlike other models such as Gaussian Mixture Models (GMM) that assume a Gaussian distribution of the input data. This non-parametric property is particularly advantageous in real-world scenarios where the data is complex and does not adhere to any known

distribution. Unlike parametric models, KNN does not require prior knowledge about the data's characteristics, making it highly flexible and applicable to a wide range of tasks.

The working principle of KNN is straightforward but powerful. Given a set of labeled training data, the algorithm predicts the class of an unseen test data point based on its proximity to the labeled data points in the training set. The proximity is typically measured using distance metrics like Euclidean distance, although other metrics such as Manhattan distance, Minkowski distance, or cosine similarity can also be employed depending on the problem. KNN relies on a crucial hyperparameter, K, which defines the number of nearest neighbors to consider when making a classification or regression decision.

In a classification task, the algorithm identifies the K closest training samples to the test point, and the test point is assigned the class label that is most frequent among these neighbors. In the case of regression, instead of a class label, KNN predicts the target value by averaging the values of the nearest neighbors. This simple approach works remarkably well in many applications, especially when the boundaries between different classes are not well-defined and are based on proximity rather than explicit decision rules.

One of the major advantages of the KNN algorithm is its simplicity and ease of use. It is considered a "lazy learner" because it does not require any explicit training phase. Instead of learning a model during training, KNN stores all the training data and defers computations until it is presented with a new data point for classification or prediction. This characteristic makes KNN highly adaptable for problems where the data is constantly changing or where real-time predictions are needed without the overhead of model training.

In the context of text classification, KNN can also be highly effective. Here, the features may be represented by word counts or other text-based features, and the task is to classify documents based on their content. In such cases, the features often follow a multinomial distribution, which is appropriate for representing the frequency of different words across documents. On the other hand, for problems where the features are continuous, such as in medical diagnosis or sensor data analysis, the features may follow a Gaussian distribution. The flexibility of KNN allows it to handle both discrete and continuous feature spaces effectively.

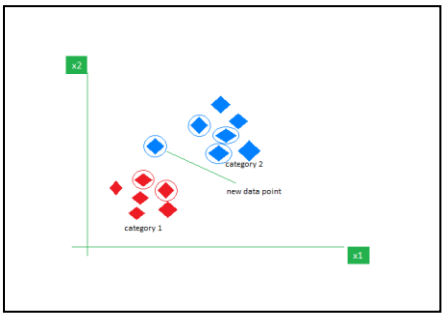
Another significant benefit of KNN is its performance with high-dimensional or large datasets. Unlike many other machine learning algorithms that require complex model training and parameter tuning, KNN's simplicity means that it can be applied quickly to high-dimensional datasets. The algorithm works by computing distances between data points, so the computational complexity is determined by the number of data points and features. In many cases, even with large datasets, KNN can yield reliable results. This is particularly valuable in domains like data mining, intrusion detection, and bioinformatics, where high-dimensional data is common.

However, KNN does come with a few challenges. One of the most well-known issues is the "curse of dimensionality," where the algorithm's performance degrades as the number of features increases. In high-dimensional spaces, data points become increasingly sparse, and the notion of proximity or distance becomes less meaningful. This can lead to poor performance unless dimensionality reduction techniques such as Principal Component Analysis (PCA) or feature selection are applied before using KNN. Another issue is the computational cost of finding the nearest neighbors, particularly with very large datasets. The brute-force method of calculating distances between every pair of points can be slow, but more efficient data structures like KD-trees or Ball-trees can help speed up the search process.

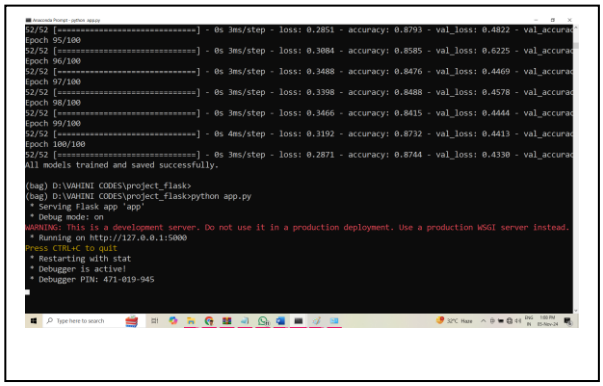
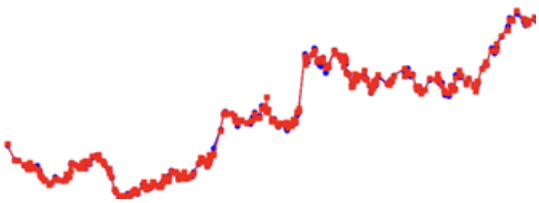
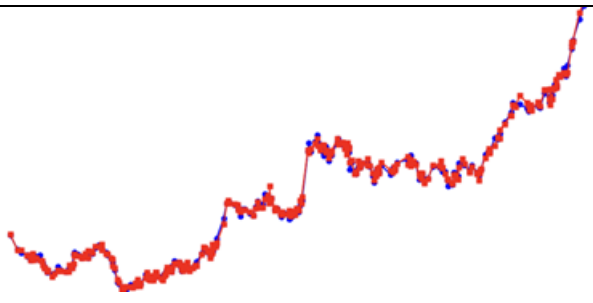
Despite these limitations, KNN remains one of the most popular and robust algorithms in machine learning due to its

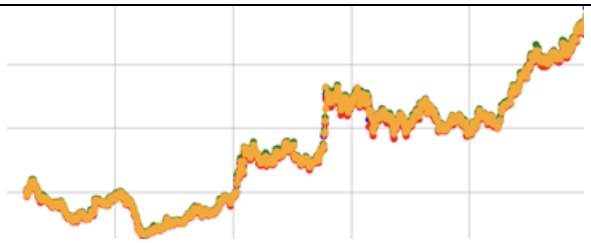
intuitive nature, ease of implementation, and ability to handle both classification and regression tasks with minimal assumptions about the data. Its use cases range from pattern recognition and speech recognition to recommendation systems and anomaly detection. The simplicity of KNN also means that it can serve as a strong baseline for more complex algorithms, helping to establish performance benchmarks for other models.

In summary, the KNN algorithm is a versatile, non-parametric method for classification and regression tasks, suitable for a wide variety of applications. While it can handle both high-dimensional and large datasets effectively, it does have challenges that require careful consideration, such as the curse of dimensionality and computational cost. Nonetheless, KNN remains a vital tool in machine learning, especially in cases where simplicity, interpretability, and flexibility are key.



Result and Methodology: -

CNN generates better results for diabetic and heart disease datasets	Accuracy results for the selected dataset is above 98.5%	
KNN algorithm generates accuracy prediction along with precision and recall values	Accuracy values variant between 89% and 90%. Finding nearest neighbour is very easy	
Support vector Machine is very easy Machine learning algorithm and generates better results	Accuracy values variant between 89% and 90%. Finding Its easy to calculate estimators	

Naive Bayes classifiers have been especially popular for text classification, and are a traditional solution for problems such as spam detection.	Accuracy values variant between 89% and 90%. Finding Its easy for text classification data.	

Deep Learning CNN: -

Step 1: Convolution Operation

Our targeted attack strategy starts with convolution operation as the first building block. In the course of this step, there is the question of the feature detectors which may be described as the neural network filters in a way, or more, precisely, the primal layers of the neural network used in corresponding classifications. We will also look at feature maps, learning of such maps, how patterns are detected, the layers of detection, and mapping the results.

Step 1(b): ReLU Layer

The second subspace of this step will be the Rectified Linear Unit or ReLU. I will discuss ReLU layers and also discuss how linearity works in Convolution Neural Networks.

Not really required for learning the basic of CNNs but a brief session won't hurt and you get better at it.

Step 2: Pooling

In this part, we are also going to be discussing pooling and by the time we are done we'll have got a fair understanding of how it usually operates. However, from this perspective, our focus will be a particular category of the pooling; max pooling. There are others though, including mean (or sum) pooling which need to be discussed and that is what we'll be doing in the next sections. This part will end with a demonstration made using a visual interactive tool that will definitely sort the whole concept out for you.

Step 3: Flattening This will be a quick guide of what flattening entails and how we transition from a pooled layer to a flattened layer in the implemention of Convolution Neural Networks

Step 4: Full Connection

In the end of this part, we will combine all that have been discussed throughout the section. By learning this, you will get to see a panoramic view of how exactly Convolution Neural Networks work and how the "neurons" that are finally produced come to learn the classification of images.

Existing System and Disadvantages: In the existing system of prediction of diseases, only one algorithm was in use because of the complexities. One technique does not offer improved results and quality for every disease type, while one classifier (Naïve Bayes) offers the highest performance on a given dataset, a different method or approach excels all others in other diseases. In the previous system also data mining techniques were not used which could enhance the percentage of the prediction of the diseases.

DISADVANTAGES

1)The first disadvantage is that the Naive Bayes classifier makes a very strong assumption on the shape of your data distribution, i.e. any two features are independent given the output class. Due to this, the result can be potentially very bad.

2)Another problem happens due to data scarcity. For any possible value of a feature, you need to estimate a likelihood value by a frequent approach. This can result in probabilities going towards 0 or 1, which in turn leads to numerical instabilities and worse results.

3)A third problem arises for continuous features due to which a lot of information can be lost.

PROPOSED SYSTEM

This part provides the imperatives of the proposed systems and shows all the elements, approaches and tools which are used in the development of the whole system. For creating the intelligent and user friendly diabetic and heart disease prediction system, an efficient software tool is required to training large number of data sets and to compare different algorithms of ML. After selecting the powerful algorithm with accurate and efficient performance indexes, the Smartphone-based application would be developed for the detection and early predicting of the levels of heart disease along with diabetic controlling. Other hardware components required for developing continuous patient monitoring system includes Arduino/Raspberry Pi, different biomedical sensors, display monitor, buzzer etc.

ADVANTAGES OF PROPOSED SYSTEM: -

1)This kind of advantage makes semantic analysis a very favorable technique since this method is able to look for relationships among words. Hence, adding semantic analysis helps fix one.

The proposed thus identify the relationship between the input data and predicts accuracy more than the existing system is; count victimizer and term frequency-inverse document frequency.

3) CNN is a more powerful Machine Learning algorithm that others.

Conclusion:

The major type of Machine Learning is classification which is widely used in health care systems for diagnosis of diseases and prediction of diseases. In this research work, two classification algorithms, namely, Naïve bayes and Support Vector Machine (SVM) has been applied for predicting heart diseases. Comparisons of these algorithms are made at and this is done based on the performance factors of classification accuracy and time. According to the experimental outcomes of this work, the chosen SVM classifier is regarded as the best algorithm due to its maximum classification rate. On the other hand, while comparing the execution time, Naïve Bayes classifier requires minimum time for execution. In this paper we paid about the heart disease prediction system with different classifier techniques to predict heart disease. The algorithms used are Naïve Bayes classifier and decision tree classifier we have seen that the accuracy of the decision tree classifier is better than naïve Bayes classifier. For higher performance of the classifier in future, Our plan is to work on Selective naïve Bayes classifier; It is a well-known fact that Naïve Bayesian classifier (NB) is highly efficient on some domains and wholly inefficient on other domains. Thus, the aim of this paper is to enhance the Naïve Bayesian classifier, to decrease its complexity, and choose only the significant for the classification attributes in the given data set. To accomplish this end-result, we employ trees built up by C4.5 and this strategy is referred to as the Selective Naïve Bayes classifier.

Future Work: -

In the project we suggested to use data collected from different labels and two datasets one is heart and diabetic collected data from UCI data repository has been used to predict two diseases heart and diabetic using machine learning algorithms. In order to determine accuracy we employed the deep learning algorithm. The convolutional neural network algorithm deals with large dataset value, and the values are collected from different UCI data repository. Moreover, there is no any mentioned of accuracy generation problem in CNN. Different machine learning based models are presented for each of the diseases in the context of this paper. Calculations point at the high efficiency of the system Commercial and industrial applications 4. The new designed trending models based on our well trained predictor for inferring the trend of a model on the basis of its established success ratio.

References

[1] P .K. Anooj, ―Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules‖; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40. Computer Science & Information Technology (CS & IT) 59

[2] Nidhi Bhatla, Kiran Jyoti"An Analysis of Heart Disease Prediction using Different Data Mining Techniques".International Journal of Engineering Research & Technology

[3] Jyoti Soni Ujma Ansari Dipesh Sharma, Sunita Soni. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction".

[4] Chaitrali S. Dangare Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888)

[5] Dane Bertram, Amy Voida, Saul Greenberg, Robert Walker, "Communication, Collaboration, and Bugs: The Social Nature of Issue Tracking in Small, Collocated Teams".

[6] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, ―Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm‖; International Journal of Engineering Science and Technology, Vol. 2(10), 2010.

[7] Erickson, T., Smith, D., Kellogg, W., Laff, M., Richards, J and Bradner, E. Socially translucent conversations: Social proxies, persistent conversation, and the design of "Babble."Proc. ACM CHI (1999), 72–79.

[8] Hollan, J., Hutchins, E. and Kirsh, D. Distributed cognition: Toward a new foundation for human computer interaction research. ACM TOCHI, 7(2),(2000), 174–

[9] Shantakumar B.Patil, Y.S.Kumaraswamy, ―Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network‖; European Journal of Scientific Research, ISSN 1450-216X Vol.31 No.4, 2009.

[10] Statlog database: http://archive.ics.uci.edu/ml/machinelearning-databases/statlog/heart

[11] Shantakumar B.Patil,Dr.Y.S.Kumaraswamy "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction" IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.2, February 2009

[12] ChandrakarOmprakash,Dr.kumarJatinder, Saini R.," Development of Indian Weighted Diabetic Risk Score (IWDRS) using Machine Learning Techniques for Type-2 Diabetes",ACM COMPUTE '16, October 21-23, 2016.

[13] Mohammed Imran, Alhanouf M. Al-Abdullatif, Bushra S. Al-Awwad, Mzoon M. Alwalmani, Sarah A. Al-Suhaibani, and Shahad A. Al-Sayah,"Towards Early Detection of Diabetic Retinopathy Using Extended Fuzzy Logic",International Journal of Pharma Medicine and Biological Sciences Vol. 5, No. 2, April 2016.

[14] Mohammed Imran, Alhanouf M. Al-Abdullatif, Bushra S. Al-Awwad, Mzoon M. Alwalmani, Sarah A. Al-Suhaibani, and Shahad A. Al-Sayah,"Towards Early Detection of Diabetic Retinopathy Using Extended Fuzzy Logic",International Journal of Pharma Medicine and Biological Sciences Vol. 5, No. 2, April 2016.

[15] R. Meza-Palacios et al., Development of a fuzzy expert system for the nephropathy control assessment in patients with type 2 diabetes mellitus, Expert Systems With Applications (2016),

[16] Polatkemal, SalihGüne,"An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease", Digital Signal Processing 17 (2007) 702–710.

[17] GowdaKaregowda Ashs1 , Manjunath A.S , Jayaram M.S.," APPLICATION OF GENETIC ALGORITHM OPTIMIZED NEURAL NETWORK CONNECTION WEIGHTS FOR MEDICAL DIAGNOSIS OF PIMA INDIANS DIABETES", International Journal on Soft Computing (IJSC), Vol.2, No.2, May 2011.

[18] ButwallMani,kumarShradha," A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier",International Journal of

Computer Applications (0975 – 8887) Volume 120 – No.8, June 2015.

[19] VeenaVijayan V.,C.Anjali," Prediction and Diagnosis of Diabetes Mellitus -AMachine Learning Approach", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015 | Trivandrum.

[20] E.P.Ephzibah," COST EFFECTIVE APPROACH ON FEATURE SELECTION USING GENETIC ALGORITHMS AND FUZZY LOGIC FOR DIABETES DIAGNOSIS", International Journal on Soft Computing (IJSC), Vol.2, No.1, February 2011.

[21] G. Thippa Reddy1, a , NeeluKhare," FFBAT-Optimized Rule Based Fuzzy Logic Classifier for Diabetes", International Journal of Engineering Rese

[22] VarmaKamadiV.S.R.P,RaoAllamAppab, ThummalaSitaMahalakshmia," A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (PCA) and modified fuzzy SLIQ decision tree approach", Applied Soft Computing.

[23] YoichiHayashi n, ShonosukeYukita," Rule extraction using Recursive-Rule extraction algorithm with J48graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the Pima Indian dataset", Informatics in Medicine Unlocked.

Krishna Yedukondalu is a student of Sree Vahini Institute of Science and Technology, Tiruvuru Present He is pursuing M. tech (CSE) from this college and received MCA Master Degree from JNTUH.

Torlapati Venkateswara Rao completed B.Tech from The Vazir Sultan College of Engineering, Khammam and M.Tech from Nova's Institute of Technology, Eluru. He has 08 Years of Teaching Experience. At Present he works as Assistant Professor at Sree Vahini Institute of Science and Technology, Tiruvuru, A.P. His areas of Interest Include Deep Learning, AI and Machine Learning, Programming in C, Java & Data Structures. He has Published 10 Papers in Various International Journals and Conferences.