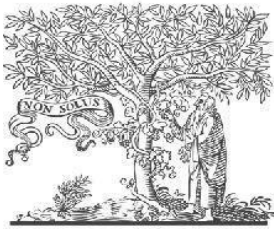


COPY RIGHT



ELSEVIER

SSRN

2024 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper; all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 8th Aug 2024. Link

<https://ijiemr.org/downloads.php?vol=Volume-13& issue=issue08>

DOI: 10.48047/IJIEMR/V13/ISSUE 08/19

Title Application of Predictive Analytics in Healthcare using Machine Learning Algorithms-
An Experimental Approach

Volume 13, ISSUE 08, Pages: 135 - 140

Paper Authors

Dr. G.V. Ramesh Babu , Turubati Vinay Kumar



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per **UGC Guidelines** We Are Providing A Electronic Bar code

Application of Predictive Analytics in Healthcare using Machine Learning Algorithms- An Experimental Approach

Dr. G.V. Ramesh Babu

Associate Professor, Department of Computer Science, Sri Venkateswara University, Tirupati
gvrameshbabu74@gmail.com

Turubati Vinay Kumar

Master of Computer Applications,
Sri Venkateswara University, Tirupati
vinayturubati1617@gmail.com

Abstract

If we have a large amount of data and any form of analysis has to be done the choice to do that is to go with machine learning algorithms. When digitization concepts go everywhere it happens to be in healthcare where a lot of health care practitioners of healthcare will be benefitted out of it. The application of machine learning algorithms helps the healthcare professionals to meet the growing needs and also can treat the patients effectively. They can also standardize their operations with lower costs. The predictions can be made accurate if there is application such algorithms which can save the lives of many patients. The purpose of this work is to reveal how machine learning algorithms are useful in healthcare sectors for making predictions and other operations.

An Introduction to Machine Learning

Machine Learning is a sub field of Artificial Intelligence which makes the software applications more accurate to predict the outcomes without making any extra programs to do such task. The historical data is used as an input to predict a particular task and come out with new values. There are a number of fields in which machine learning is applied and some of them are making recommendations for people in the form of machine learning, to be able to detect frauds happening in bank transactions, to filter the spam, to detect the malware in an application, to automate the business processes and finally to make predictions in health care based on past records of the patients. Machine learning becomes so important because of the following reasons. It can able to accelerate the business, can make predictions with the past historical data, able to make

recommendations in buying a new product and many more.

Types of Machine Learning

There are different forms of machine learning.

- Supervised Learning
- Unsupervised Learning
- Semi Supervised Learning
- Reinforcement Learning

The above categories are made based on how the algorithms are making more accurate predictions.

Supervised Learning

In this form, the algorithms are supplied with training data which are labelled as well as variables that has to be used in defining the correlations. The input and output of the algorithm are mentioned. Supervised learning is good doing the following.

- Classification of Binary target classes
- Can able to classify more than two classes of target variables i.e., target class.
- Predicting Continuous set of values in the form of regression modelling
- Can able to combine predictions of two or more algorithms in the form of ensembles which gives more accurate prediction.

Unsupervised Learning

In this form the algorithms are supplied with training data which is not labelled and the algorithm will be scanning for any meaningful connection in the datasets. The data that has been trained, the predictions, recommendations and the output they produce are predetermined. It has been working well for the following tasks.

- Can be to split the dataset upon similarity occurred between the data into different groups It is termed as clustering.
- Can able to detect abnormal data values in the data set. It is known as Anomaly Detection.
- Can able to detect the set of items which occurs frequently in a dataset. It is termed as association rule mining.
- Helps in reducing the dimensions of the dataset.

Semi Supervised Learning

It is a combination of the above two types. The data that has been supplied may be a labelled one but the output produced by the algorithm will be of its own. The data can also be read on its own. It works on the principle that it has been given a small amount of labelled data and with the help of that it can able to detect the data dimensions and all these can be applied to the newer unlabelled dataset. The algorithm performance can be improved when they are trained with the labelled data. The semi supervised learning is good at performing the following tasks.

- It can able to train the algorithms which can able to translate the language based on the dictionary. It is termed as machine translation.
- Can able to detect the frauds happening in the fraud transactions.
- The algorithms are trained with smaller datasets and they are obviously been applied

to larger datasets. It is termed as labelling of data.

Reinforcement Learning

It is used to solve the multiple step involved in the process for which the rules are defined properly. The algorithms are programmed in such a way that they have been supplied with all positive and negative aspects and the algorithm works on its own in most of the cases. Reinforcement learning can be applied in the fields of

- Robotic world which can able to perform the real world tasks on its own.
 - Teaching bots to play video games.
 - Helps enterprises in allocating resources.
- Thus all the above forms can able to applied according to the data and the problem that has been selected.

Machine Learning Applications in Healthcare

Machine learning applications can able to improve the accuracy of the patient's predictions and can able to do with the right choice of the algorithm. A Complex application of machine learning which can able to reflect the functions of a human brain and also been used widely in medical images and in the field of radiology. The application of neural networks will helps us to learn data with no supervision. The applications of deep learning can able to detect and analyse the cancerous parts from the predefined images. The speed in processing the application and intervention of cloud applications will help the machine learning applications to detect the anomalies in images where it can be detected using human eyes. The following are the three areas in which we can able to apply machine learning in health care.

Maintaining patient records

The application of machine learning in healthcare helps in maintaining the records of the patients in the form of electronic records. The application of Artificial intelligence to improve electronic Health Record Management can able to minimize the administrative costs, reduces the healthcare operations and helps in maximizing the patient care. An example of the above form will be inclusion of natural

language processing which can avoid manual process of taking notes and the notes can be captured automatically.

Integration of data

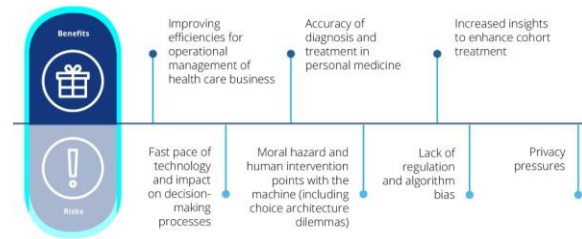
The knowledge gaps that occur in information provided by the health care professionals will lead to false predictions and also it has a negative impact on decision making process. Since the predictions made by machine learning algorithms are purely dependant on data they has to be integrated properly to have better decisions. Health care data providers are responsible for the integrity of data. The various activities performed by the health care data providers are data collection, data analysis, classification of data and data cleaning. All the above activities will be helpful in maintaining data integrity.

Predictive Analytics

Predictive analytics is the most important factor where patient historical data are analysed and decisions are made accordingly. It helps in predicting the disease at an early stage thus by saving the lives of the patients. It can able to alert the patient at an early stage and proper treatment can be provided. It can also able to cover the knowledge gap that occurs between the various data available for making prediction.

Predictive Analytics in Healthcare

Due to the digital era and everywhere the data is rolling up and down, health care sector also has improved a lot by providing quality treatment to the patients. Technology improvement has played a major role in health care sector thus by providing accurate treatment to the patients. Predictive analytics can be used as an advanced analytics where it has been used to provide predictions based on the data that is available. It will be helpful in making better decisions. It is a field where it applies a number of techniques such as data modelling, data mining, and artificial intelligence and will analyse the past and real time data. These data are useful in making predictions for the future.



Source: Deloitte analysis.

Figure 1. Benefits and risks of predictive analytics in healthcare

Predictive analytics can be termed as a logic which is derived from a set of theories which was derived by humans and it is made to fit into a hypothesis. From the hypothesis a set of rules and definitions are framed which is further developed into a formula and have calculations on it. Predictive analytics can also be a form of an unsupervised algorithm which does not have proper set of rules and generally depends on a set of algorithms to discover patterns from the dataset. After then it will cluster them into groups. The unsupervised learning generally it does not depends on any predefined structure simply goes through the dataset and identifies the complex process out of it. This set will be new to the humans and they cannot able to predict on their own. Both supervised and unsupervised are good forms of analytical tools to be used along with predictive analytics to gain newer insights. The various benefits of choosing predictive analytics is

- To improve the efficiency of the operational management in healthcare operations of business.
- To improve accuracy in diagnosis and providing better treatment to patients.
- To provide improved insights in providing treatment which are cohort.

Predictive analytics helps us to improve the operational efficiency with the integration of big data and predictive analytics. They both plays a major role in improving the efficiency of health care sectors. Reports are generated in real time as well as they provide newer insights of data and will be able to adjust dynamically the predefined predictive algorithms according to the current needs. The real time data helps us to determine the capabilities and will evaluate the performance of staff in real time. To provide

with an example in operational management, a useful insight that a predictive analytics can provide is to minimize the staff and determine how many staff are needed in providing better quality healthcare to the patients. This can be possible by making use of past data, data which are overflow from the facilities available nearby, data available as population, demographic, predictable disease and the sickness involved due to seasonal changes. All these will help us to analyse the past data and come out with better decisions in operational level.

The operational management can also provide us with the analysing the visiting time of the patients, technology to record the temperature readings, also reading the calendar and how far the patients are visiting on public holidays or on other days so that proper steps can be taken in mobilising the things at operational level. Predictive analytics also helps us in analysing the performance of candidate performance in interview with some of the historical data and can able to predict whether a candidate will be performing better or not.

As far as personal medicine is concerned, predictive analytics plays an important role in the individual level as well as in making diagnosis of a newer diseases and providing the treatments accordingly. This will help us to provide with a more accurate analysis and thus have will increase the prediction rate of the disease. This happens mainly because of the people who were complex and applications of certain investigations for certain people. Also the predictive analytics should work when there is an increased rate of diagnosis and all these differs from patient to patient. All these activities are not possible for a health care practitioner do to it manually. The application of predictive analytics helps us in discovering the unknown correlations, hidden information by proper examination of larger datasets and making predictions out of it. If all these are applied at an individual level the practitioner can able to provide with a specific treatment with specific set of people.

Application of predictive analytics and machine learning in diagnosing a certain disorder.

Predictive analytics is a field in which the future can be predicted based on the past history. Heart disease is a major issue in most of the humans and if it is predicted in advance it can save the life of many humans. If it is predicted in advance meaningful insights can be drawn and it will be useful for the doctors to diagnose in advance the existence of disease and treatment can be provided based on that.

The heart disease dataset has been taken for prediction and machine learning algorithms such as Support Vector machines, Decision Tree Classifier, K Nearest Neighbour and Random Forest has been applied on the dataset and the results are compared. The dataset has been downloaded from kaggle and loaded into Google Colab an interactive Laboratory provided by Google to run python codes. The dataset consists of 301 observations of 14 variables which is of mixed in types. Out of 14 variables one is a target variable which shows the existence of heart disease or not. The value of target class 1 indicates existence of disease and value of 0 indicates non-existence of disease.

The correlation between the different variables can be known with the help of correlation function and it can be visualised as follows

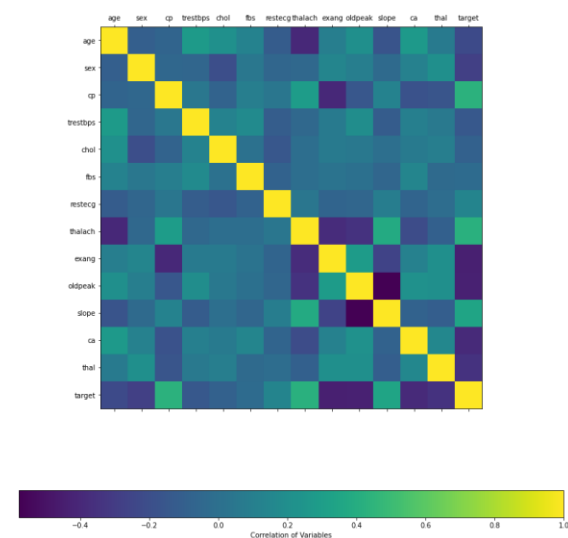


Figure 2 Correlation Matrix

The correlation matrix takes the value range between +1 and -1. A positive correlation denotes the variables are highly correlated and a negative value shows that the variables are not highly correlated. It has been observed from the above figure that no variable is highly correlated with the target value. There are also positive and negative correlation occurs between target and other variables. The target can also be viewed in the same way which gives the proportion of values in each of the two classes that is observations with heart disease and observations with no heart disease.

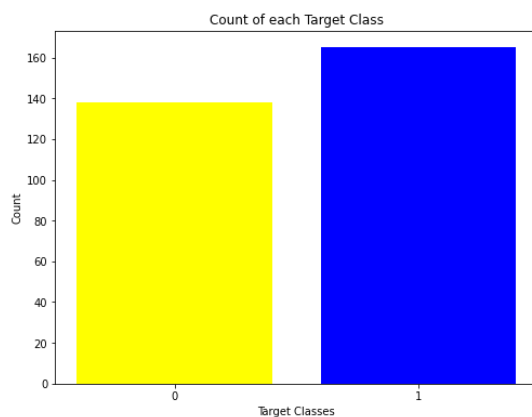


Figure 3: Count of Target Class

The above figure shows the count of target class with which we can observe that the data is highly balanced which means there data has been distributed in both the classes in equal number. Now the dataset can be split into train and test set to build the models and perform the evaluation. The following algorithms are analysed.

- Support Vector machines,
- Decision Tree Classifier,
- K Nearest Neighbour
- Random Forest

Considering the full dataset 80% has been taken as train dataset and 20% has been taken as test dataset. The above said models are built and the results are plotted as given below.

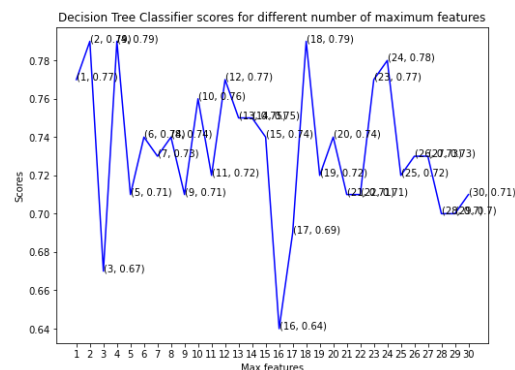
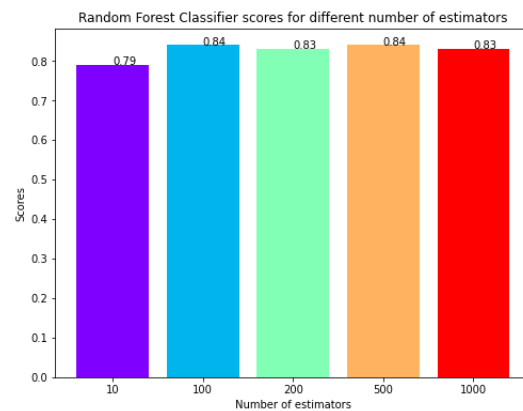
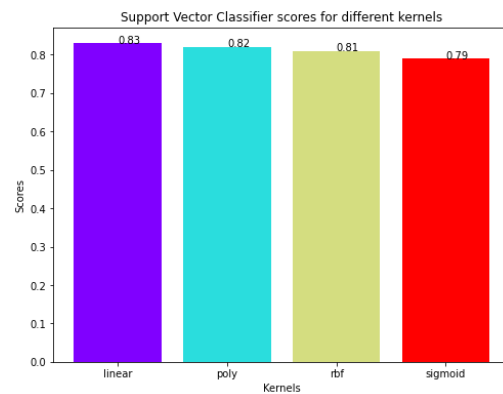
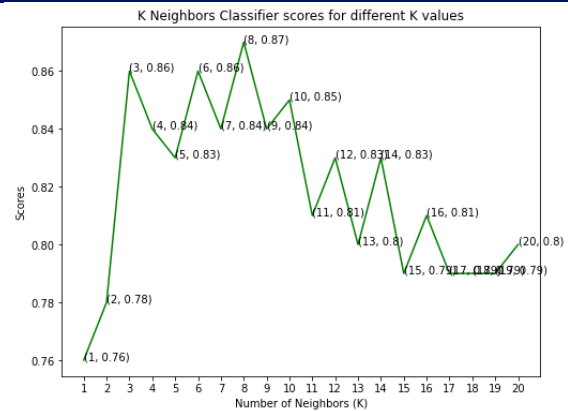


Figure 4 Results of the four Models

The above visual shows the results of four models. The KNN has been trained for different k values and produces an average accuracy of 87.9% with 8 neighbours. The Support Vector machine has been trained with four different kernels and out of which linear kernel performs the best. The accuracy of the model is 83%. The random forest model has been trained with 5 different estimators and the maximum value is achieved for estimators 500 and 100. The accuracy with these estimators will be 84%. The decision tree classifier has been trained with maximum of 30 features ranging from 1 to 30. The classifier reaches its best at feature value 2, 4 and 18. The accuracy of the classifier is 79%. Thus out of four models tested and trained kNN serves its best for the problem taken. The above models can be applied to different datasets and the results can be compared to choose the best one.

Conclusion

Thus predictive analytics plays an important role in healthcare by predicting the future and saving the life of the patients. The above work reveals the fact that the predictive analytics can be applied in different areas of the health care sector and can improve the sector with technology advancements. The different machine learning algorithms can be applied to various health care datasets and results are compared to get the best model out of it. Thus the improvements in technology and other sector helps in improvising the different areas of healthcare and can also helpful in predicting the future.

References

- [1] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.
- [2] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *RadioGraphics*, vol. 37, no. 2, pp. 505–515, 2017.
- [3] P. Radhika, R. Nair, and G. Veena, "A comparative study of lung cancer detection using machine learning algorithms," in *Proceedings of the IEEE International Conference on Electrical, Computer and Communication Technologies*, Coimbatore, India, November 2019.
- [4] A. Al-Zebari and A. Sengur, "Performance comparison of machine learning techniques on diabetes disease detection," in *Proceedings of the 1st International Informatics and Software Engineering Conference*, Ankara, Turkey, November 2019.
- [5] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using K-nearest neighbor machine learning algorithm," in *Proceedings of the 9th International Conference on Developments in eSystems Engineering*, Liverpool, UK, September 2016.
- [6] M. S. Yarabarla, L. K. Ravi, and A. Sivasangari, "Breast cancer prediction via machine learning," in *Proceedings of the 3rd International Conference on Trends in Electronics and Informatics*, Tirunelveli, India, April 2019.