ISBN 978-93-91286-40-8

KRISTU JYOTI COLLEGE OF MANAGEMENT & TECHNOLOGY
IQAC | Department of Computer Applications
RESEARCH HUB

CERTIFICATE
OF PRESENTATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Harithamol K H

OF SAINTGITS COLLEGE OF APPLIED SCIENCES, PATHAMUTTOM
FOR SUCCESSFULLY PRESENTING A PAPER AT THE **FIRST INTERNATIONAL CONFERENCE ON ADVANCE MODERN COMPUTING TRENDS AND TECHNOLOGY (ICAMCTT 2021)** ON 30TH & 31ST OF JULY 2021

Paper Title : Analysing the Impact of Online Classes on Student's Health using Machine Learning Techniques

REV. FR. JOSHY CHEERAMKUZHY CMI
Principal

SUSHEEL GEORGE JOSEPH
Conference Secretary

ROJI THOMAS
Conference Director

BINNY S
Conference Convenor

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

Analysing the Impact of Online Classes on Student's Health using Machine Learning Techniques

Alen James
Department of Computer Applications
Mahatma Gandhi University, Kottayam
alenj.bca1922@saintgits.org

Harithamol K. H.
Department of Computer Applications
Mahatma Gandhi University, Kottayam
harithamol.bca1922@saintgits.org

Vasudevan Nampoothiri
Department of Computer Applications
Mahatma Gandhi University, Kottayam
vasudevan.bca1922@saintgits.org

Asst.Prof. Johnsymol Joy
Department of Computer Applications
Mahatma Gandhi University, Kottayam
johnsymol.joy@saintgits.org

Abstract*—The onset of Covid 19 had an adverse effect in all areas of one's life and thereby made us adapt to new and unfamiliar ways to carry on with our daily life. Utilization of online mode of interaction became a necessity for most of us. One such shift was the transition from regular class to online class. The educational sector was one of the quickest to adapt to online mode of operation, using apps like google meet, zoom, microsoft teams,etc to conduct online interactive classes. In this paper we try to analyse the health concerning problems associated with this new mode of education. For this purpose we use machine learning to categorize the data collected and use it to analyse the student's health and check whether their health is more negatively affected on online classes when compared to offline classes. Machine learning classification techniques such as Naive Bayes, J48, Random Forest, Decision Table etc to conduct the analysis.

***Keywords*—COVID-19, Google Meet, Zoom, Microsoft Teams, Machine Learning, Naive Bayes, J48, Random Forest, Decision Table**

I. INTRODUCTION

Education is one of the most prominent as well as an important sector of a country. It impacts the future of the youth as well as the future of the whole country itself. In 2020 marked the arrival of a world-wide pandemic due to a deadly virus known as COVID-19 and most of the organizations and institutions had to be shut to suppress the spread of the virus. This pandemic has disrupted the traditional way of learning and studying and forced the education sector to adopt a new means of teaching so that the students can stay home and study [6]. This method of Online Learning was a remarkable measure to continue the education of the students, and the whole

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

sector did not come to a halt due to the pandemic.

The pandemic forced a sudden shift to online mode of teaching and learning which was new to the majority of students and teachers alike and had a great deal of starting trouble. But among all this the health related side effects of this was shortsighted and often ignored by many due to many reasons. Using devices like computers and smartphones for attending classes can cause strain to students who attend classes for a long period of time. Temporary as it may be, the harmful after effects can be carried on by the students for the rest of their life and can cause severe problems in the future. Therefore through this paper we try to identify these problems which students face in comparison to the offline class.

A student's academic performance could be affected by differences between economic, racial, availability of resources, stress and also anxiety caused due to the pandemic situation. These can all affect negatively and can impair the learner's ability to concentrate and study through online classes. Apart from that, not all teachers are used to teaching through these online services, rather be ready to do it swiftly[5]. These acts as physical barriers to efficient learning through online services

Online classes are attended through tools commonly called Learning Management Systems. Some of them are Next, Canvas, Moodle etc. YouTube acts as an efficient platform for tutorials and guides. Real-time online lectures are held by using the previously mentioned softwares like Microsoft Teams, Google Meet, Zoom, Cisco Webex etc[4]. Online assignments are required to be submitted by students, these are done using pdfs, created by PDF Scanners like Adobe Scan, CamScanner, Tiny Scanner, etc. The government in some states like Tamil Nadu have provided students with laptops and internet for free to promote and support the online learning system.

Therefore we use data mining to collect data and categorize data and use a variety of machine learning algorithms to find one with highest precision to drive an outcome . Machine learning is used in such cases because it becomes easier to process large data collection by teaching the computer and automating the task [8]. This is achieved by feeding the machine learning algorithm tokenized data during training sessions . The algorithm uses pattern recognition to learn and apply complex mathematical calculations to learn [7] . After this the algorithm can be used to operate on the actual data to predict any outcomes. Through machine learning we can predict data using models,patterns and algorithms and therefore has before an important feature in the field of data analytics [7].

We use the weka tool for testing different machine learning algorithms. Weka

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

is a software used for implementing data mining using Java. It has a different collection of tools and has a roster of different algorithms used to analyse data, model prediction, etc. It has a GUI(Graphical User Interface) which helps in easier navigation and usability of the software. It has a workbench that includes different methods to solve Data Mining problems such as clustering, classification, regression, etc. A single relational table is used as the input for all the algorithms and methods. This single relational table is created using the file or generated by a database query[1].

II. LITERATURE REVIEW

Machine learning algorithms use little to no human interventions to analyse the data and build a new data model. It can find complex patterns and irregularities and most of all learn from the data. Therefore we use different machine learning algorithms to analyse our real time data.

A. Random Forest

Random forest is a machine learning algorithm which is used to achieve accurate results most of the time. The random forest algorithm is a kind of supervised learning algorithm. The algorithm creates a bunch of decision trees which are trained with bagging method. In other words, the Random Forest algorithm builds two or more decision trees from the data and groups or merges them together to get an accurate prediction.

Leo Breiman proposed this scheme in the 2000's to build a predictor that operates on a set of decision trees that grow randomly in some selected subspaces of data. Using Breiman's approach, trees are formed in the collection by selecting a random node, and a small group of input coordinates to split, then calculating the best split based on these features(input coordinates) in the training set. This newly created tree is grown using CART methodology as maximum as possible without pruning. Bagging is blended across this subspace to resample and replace the tree with training data sets that grow a new individual tree[2].

For classification and regression problems, the random forest approach is commonly used. This algorithm is used in banking sectors to determine if a customer is eligible for a loan based on his account balance, payment status, etc. It is also used in healthcare sectors to identify the diseases suffered by a patient, or if they are at risk of cancer or other diseases, etc.

B. J48

J48 comes as an extension of the ID3 algorithm which was developed by the WEKA project team. Since It uses a

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

predictive machine learning model it used to to predict the target variable of a new dataset record. The J48 algorithm creates a decision tree based on attributes data set.The nodes on the decision tree are used to represent different instances of classification and the branches are used to show the results, and finally the leafs show the class label

C. Naive Bayes

Naive Bayes algorithm is an algorithm that uses Bayes' Theorem, which is a theorem that is used for calculating a probability by counting frequency and combinations of values in a data set that is previously taken. Naive Bayes algorithm uses conditional probabilities of each attribute of each class and uses this data for prediction from the data set[9]. Real world situations are more often predicted by using Naive Bayes because it performs much better in those conditions. One of the parameter estimations of Naive Bayes model is Maximum Likelihood and it is used very often. Also this algorithm only requires a small training data and then it can start estimating the parameters.

D. Decision table

Decision tables are one of the most simple machine learning algorithms which is also used for prediction based on data where decision is taken on each instance of data depending on a set of variables that are provided [10]. A decision table is divided into two fractions ,the first part contains all the attributes and their values that are possible and the second part of the decision table includes all the conclusions [11]. After that the attributes are cross checked according to the conclusions drawn from the decision table. But decision tables are rarely used because they cannot be used if the attributes have more outcomes and it also cannot be used for complex data.

III. METHODOLOGY USED

A. Data Collection

The primary means of data collection was done through google form which included a questionnaire which contained questions about health problems associated with online classes and whether the participants are facing more problems when compared to offline classes. The data collected were real time data and were collected from indian states of Kerala and Madhya Pradesh and a total of 207 responses were received.

The following attributes were included in the form for data collection

Refer [Table 1]

From Table 1, the 13th entry is used as the class label response for prediction.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

B. Results and Discussion

This survey and classification is done by using Weka software tool which is open source machine language. It is a graphical user interface based tool . It includes tools such as preprocessing, clustering, regression etc. From the collected dataset 120 samples were taken through Google forms and stored in MS Excel and converted as .arff files.

Weka software tool was used in classification of the data set from Google Forms. 207 responses were taken into account and were processed using this tool. The result is shown in Table 2.

Refer [Table 2]

Chart 1 shows some of the measures of the four algorithms

Refer [Chart 1]

Chart 2 shows the comparison of accuracy between the four algorithms

Refer [Chart 2]

Accuracy is used to machine learning algorithms to find which algorithm is best suited for the purpose depending on the relationships and patterns it can identify from the data. Accuracy is not one of the most appropriate performance metric in some unusual situations like in cases where target variable classes in the data set from or input are unbalanced[3].

$$\square\square\square\square\square\square = \frac{\square\square}{(\square\square + \square\square)}$$

Recall shows the amount of positive predictions made by the algorithm from all the possible predictions on the data. In other words, it specifies what number of relevant data items are selected[3].It's calculated by dividing the total number of true positives by the total number of true positives and false negatives in the data set.

$$\square\square\square\square\square\square\square\square\square\square = \frac{\square\square}{(\square\square + \square\square)}$$

Precision is the ratio between the positively predicted data and all the possible positive predictions. It means the number of selected data items from the data set that are relevant. Or it means the number of observations that the algorithm predicted to be positive, is actually positive in the data set[3]. The precision of an algorithm is therefore the no. of true positives divided by the sum of the true positives and false positives in the dataset founded by the algorithm.

$$\square - \square\square\square\square\square\square\square = \frac{\square\square}{\square\square + \square.\square(\square\square + \square\square)}$$

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

F-measure is used to test the accuracy of the algorithm, it combines precision and recall into a single entity. It is also known as F1-Score or F-Score. Its formula is the harmonic mean of precision and recall.

IV. CONCLUSION

Through this paper we try to analyse the effects of online classes on student's health with the help of real-time data-set. Online classes place a huge amount of strain and stress on students, and long and constant viewing of screens of devices leads to headaches. Using Weka tool and different machine learning algorithms were used to process the data collected from the google forms. Through testing of the data, the random forest was found to be the most accurate algorithm when compared to other algorithms and had an accuracy rate over 96 percent. The recall precision and the f-measure of random forest and all the other algorithms are shown in the graph. The data used was real time data and a total of 207 responses were recorded, this data had a small sample size and hence could not represent the entire student body. From the paper we can see that random forest is the best technique which can be used to analyse the students overall health during online classes.

References

[1] Alam, F. and Pachauri, S., (2017). "Comparative Study of J48, Naive Bayes and One-R Classification Technique for Credit Card Fraud Detection using WEKA" Advances in Computational Science and Technology, ISSN 0973-6107 Vol 10, Number 6 pp. 1731-1743

[2] Biau, G., (2012). "Analysis of a Random Forests Model" Journal of Machine Learning Research Vol 13 pp. 1063-1095

[3] Vakili, M., Ghamsari, M. and Rezaei, M., "Performance Analysis and Comparison of Machine and Deep Learning Algorithms or IoT Data Classification"

[4] Gaudiot, J.-L., and Kasahara, I. H., (2020). "Computer Education in the Age of COVID-19" Computer Volume: 53, Issue: 10, Oct. 2020

[5] Arora, M., Goyal, L. M., Chintalapudi, N., and Mittal, M., (2020) "Factors affecting digital education during COVID-19: A statistical modeling approach" 2020 5th International Conference on Computing, Communication and Security (ICCCS)

[6] Kexin, L., Yi, Q., Xiaoou, S. and Yan, L., (2020) "Future Education Trend Learned From the Covid-19 Pandemic: Take<< Artificial Intelligence>>Online Course As an Example" 2020 International Conference

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

on Artificial Intelligence and Education (ICAIE)

[7] Angra, S., and Ahuja, S., (2017) "Machine learning and its applications: A review." 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

[8] Halde, R. R. (2016). "Application of Machine Learning algorithms for betterment in education system." International Conference on Automatic Control and Dynamic Optimization Techniques

[9] Ray, S., (2019) "A Quick Review of Machine Learning Algorithms." 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)

[10] Qun-Feng Zhang, Xi-Zhao Wang, and Jing-Hong Wang. (n.d.)." A further study on simplification of decision tables". Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693).

[11] https://www.sciencedirect.com/topics/computer-science/decision-table

Tables and Charts

Attributes	**Possible Values**
Regular attendee	{Yes, No}
Feel better attending online class	{Yes, No}
Headaches after online class	{Yes, No}
Uses headset	{Yes, No}
Fatigue after online class	{Yes, No}
Nourishment during online class	{Yes, No}
Body pain due to sitting postures	{Yes, No}
Regular Exercises	{Yes, No}
Attends in well air-circulating room	{Yes, No}

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

Uses laptop/desktop for class	{Yes, No}
Uses mobile for class	{Yes, No}
Network issues	{Yes, No}
Online class has affected negatively	{Yes, No}

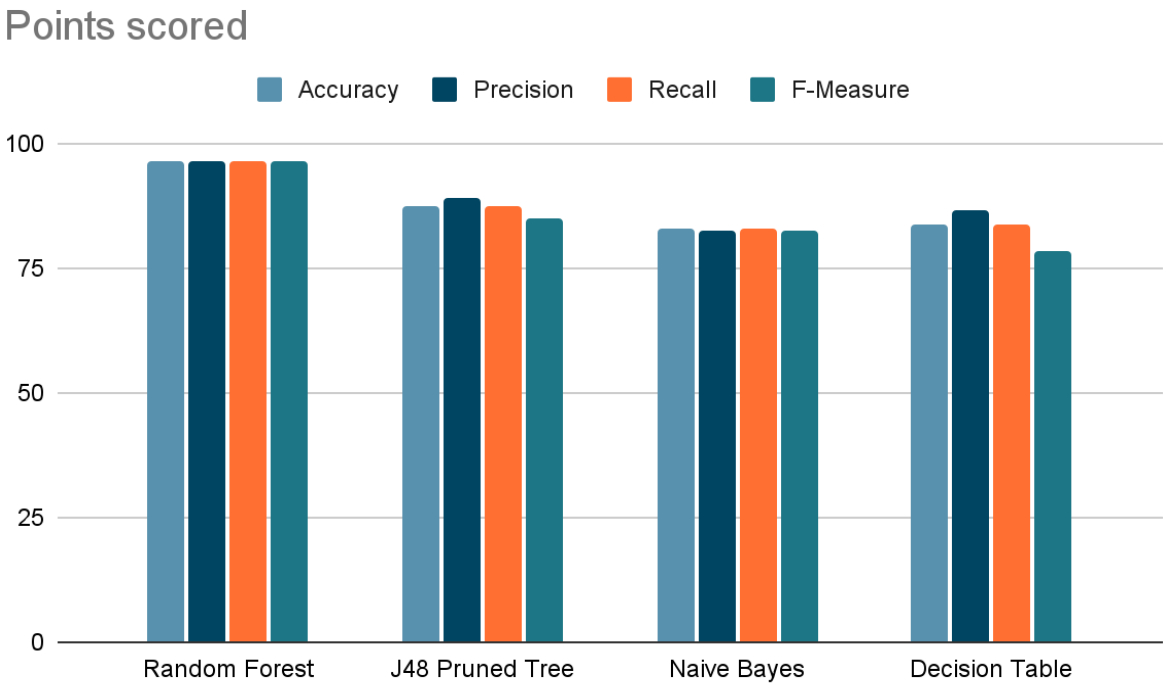
[Table 1] - Questionnaire used for data collection

Algorithms	**Accuracy**	**Precision**	**Recall**	**F-Measure**
Random Forest	96.618%	96.6%	96.6%	96.6%
J48 Pruned Tree	87.439%	89.1%	87.4%	85.0%
Naive Bayes	83.091%	82.4%	83.1%	82.7%
Decision Table	83.574%	86.4%	83.6%	78.5%

[Table 2] - Measures of each algorithm

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

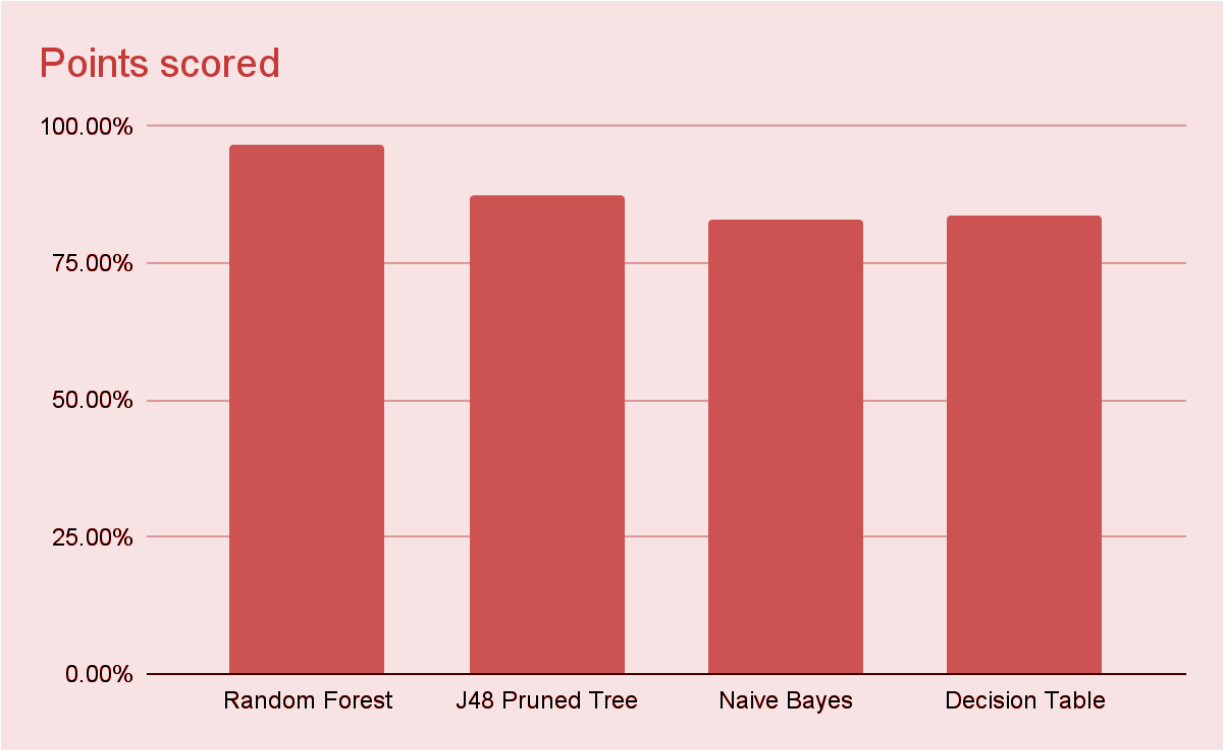
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021



[Chart 1] - Comparison of measures of four algorithms

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021



[Chart 2] - Comparison of accuracy between four algorithms