Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 3, 10297-10306

Impacts Of Covid-19 On Education Using Data Mining Techniques

Tara Yousif Mawlood[1] , Galawizh Muhammad Najeeb[2]

[1]*Sulaimani Polytechnic University MSc. Degree in Computer science.*
[2]*Sulaimani Polytechnic University MSc. Degree in Computer science.*

Abstract:

Machine learning systems that aggregate IoT apps and smart devices have enhanced e-learning systems by providing remote monitoring and screening of instruction and individual educational scores. This paper presents an ANN a deep learning model to detect technical aspects of teaching and e-learning in virtual education systems using data mining and propose a model to estimate execution of ANN which depend on COVID-19 effects. Association rules of data mining and supervised techniques are applied to detect performance of student before Covid-19 and after and Co-relation, Heat map is also added in the research.

Keywords: Internet of Things (IoT), ANN, Supervised techniques, data mining.

1. Introduction:

The technology in today's world is improving at a rapid speed in every area like IoT (Internet of things) IoT is technology which help us to connect the smart devices with each other through internet AI (Artificial Intelligence), ML (Machine learning) and many more. In e-learning most organizations are using applications like zoom, google meetings to teach the student their course work. The coronavirus pandemic circumstances have caused a significant difficulties and specialized issues to colleges, foundations and exploration focuses across all nations. The lack of access to information technology at educational systems has caused students and educators to deal with numerous issues in how to instruct and figure out the course ideas with regards to virtual schooling. IoT-based assisted learning frameworks has helped students study remotely using smart devices, RFID, actuators, and sensors (Alghamdi, 2021). An unexpected change school system has caused understudies and educators to deal with numerous issues in how to instruct and see course ideas with regards to virtual training. The data mining process involves the sorting of big data and analyzing the relationship and pattern in data to solve the problems which can be business problems or normal analysis problems. A technique that is used in data mining is the following. (1) Association (2) Data cleaning (3) Classification (4) Clustering (5) Data visualization. It depends on the problem and which technique you are using to solve your problem (Tan and Lin. 2021).

1.1. Aims and objectives:

The objectives and aims of the company in the project are to create a system that will help the user to fully understand the steps taken by the company to make their products reachable to everyone.

- To propose a model for prediction to estimate execution of VES (Virtual education system) which depend on COVID-19 effects.

- To perceive the behavioral characteristics of learning and teaching depend on Artificial Neural Network algorithm in virtual education systems.
- To estimate the behavioral factors based totally on ANN algorithm the usage of affiliation guidelines mining and supervised strategies to expect the satisfactoriness in virtual education system.

1.2. Research question:

Research question is a major component of our project. Deciding on a research question is a critical element of each quantitative and qualitative research. Investigation will require records collection and evaluation, and the method.

- How do we implement the prediction model to estimate execution of virtual education system which based on COVID-19 side effects?
- How do we understand the behavioral characteristics of learning and teaching depend on ANN algorithm in virtual education systems?
- How can we implement and estimate the behavioral factors based totally on an Artificial Neural Network algorithm.

1.3. Problem statement:

The Covid-19 pandemic has affected the world of education for several reasons. Students and staff are unsure if they can travel to school and the schools where they are located may be closed off to prevent the spread of the virus. Some people have chosen or had no choice but to stay home or relocate because they don't want their family members being infected with COVID-19. By this study answer is given to the various questions such as can we use non-traditional indicators that replicate the modern situation as a way to predict the performance of the scholar? And How has the pandemic and the following lockdown affected the academic lives of college students?

1.4. Significance:

It has been studied that the neural networks which are also called as artificial neural networks are considered to be an adaptive system which is learned using many of the associated nodes or neurons in a relatively layered structure. This has a resemblance with the human brain. It can also help to learn from the data which can be trained for the purpose of recognizing the patterns, classification of data and then predicting the future events. There are multiple applications which show the utilization of neural networks in machine learning such as the segmentation of images and videos in a semantic manner, detection of objects in images, detection of cancer and many more (Abiodun et al. 2019). Considering the applications of the neural networks, it is significant to apply the neural networks to study the impact of COVID-19 on the education systems. Further the study given by (Rodriguez et al. 2021), shows that artificial intelligence has contributed and thus builds the predictive models of academic performance of the students using the artificial neural networks. It provides the possibility of utilizing the association which is present between the variables to achieve the goal of estimating the result and thus process the capability of the model to obtain the prediction.

2. Literature Review:

Li and Jiang (2021) has discussed the impacts of COVID-19 that effected the tool-dependent academic research areas and mainly Educational Big Data (EBD) analysis for re-examine Internet Plus education system. Though, teachers learn and study the hindrances and thick and thins of online teaching as they taught face to face that has impacted their professional and personal dealings. They presented their work to bring in

account the most reliable research parameter for EBD in a more precise pattern and also discussed the +ve psychological paths variables for the stress handled by less controlled teachers. Their work included research correlation based analysis through CiteSpace 5.7 and VOSviewer that helps to extract the explicit ways and information design in information maps of scientific studies by Web of Science Core Collection. More than 1781 (Thousand seven hundred and eight) articles were analyzed and contains the data of education system patterns.

Research crossing 15 years was led to uncover that the information base has gathered emphatically after many states' drives starting around 2012 with a speeding up yearly development and diminishing geographic lopsidedness. The audit additionally distinguished a few powerful creators and diaries whose impacts will keep on having future ramifications. The creators recognized a few effective foci, for example, Data mining, understudy execution, learning climate and brain science, learning examination, and application. All the more explicitly, the creators distinguished the logical shift from information mining application to information protection and instructive brain research, from general sweep to explicit examination. Among the ends, the outcomes featured the significant combination of instructive brain science and innovation during basic times of instructive turn of events.

Safdari et al. (2021) reviewed the literature and published research paper to recognize the importance of successful n popular data mining techniques and to cover the space in information. As the pandemic threatened the world the concerns were identified in public health sector, to find hidden information researchers make use of data mining techniques. Systematic searches were carried out through Web of Science, Scopus, and PubMed databases. The retrieved search papers were analyzed in the steps in favor of Systematic Reviews reporting style and checklist was done for Meta-Analysis for particular research paper selection. Classification was done on obtained results which were inferred. Scoping review showed that 335 citations out of which only fifty research papers were recognized as eligible. The outcomes that were reviewed demonstrated the highest popular DM connected to Natural language processing twenty two percent (22%) and the most frequently introduced method was the reveling disease characteristics twenty two percent (22%). The COVID-19 was the most talked disease in favor of diseases. Supervised Learning Techniques were applied in nighty percent (90%) cases showing predominance in literature. In terms of Healthcare sector the concerns showed that infectious disease spread was determined as the most common hardly followed by epidemiology field. SPSS and R software's were mostly utilized by the researchers in research papers like twenty to percent (22%) and twenty percent (20%) respectively. The unrecognized paths of disease in pandemics were determined through some robust research as showed by outcomes of some popular research papers. But there will be more requirements of growth in studying the treatment and control the disease spread.

Parthiban et al. (2021) had also discussed the propagation of COVID-19 in days of March 2020 that was a highly considered issue for public health sector as nation was suffering badly. People's lives were highly effected and disturbed through the spread of Corona virus. Education systems suffers from the threatening impact all through all areas is to be sure a different extension influence. A total closure continues to add new issues for understudies to learn and furthermore for teachers to really deal with the class, prone to bring about the change of such a disconnected education sector into an online class. Their worked explored and demonstrated number of online learning stages and structures, also

teaching styles and resources distribution tool and modern tools were utilized to make sure that the learners can learn and study in the best form. Finally, online examinations were also considered and a suitable intimate environment were created. There were many hindrances and problems that were faced by teachers during online teaching methods, for example many learners and students concept of online learning and studying was so threatening and had a –ve approach towards their personality and socializing in society. Therefore, a method was introduced in favor of online teaching methods to give students the best online classroom learning experience, giving online teaching classrooms to be best as, if not good as, a one online teaching classroom. This research work arrows on day to day teaching techniques that focuses online teaching methods sustained by a machine grounded tool to give one individual with a most particular stress-free result.

Similarly Ahouz and Golabpour (2021) explained that the largest occurrence of COVID-19 that was found it as a new pandemic. Forecasting and predicting both for its occurrence and happens all over the world is very important and essential in terms to aid the public health sectors and professionals in making best and important decisions. They worked to forecast the prevalence of COVID 19 around a 2 week period (14 days) time to control the disease spread.

Awadh et al. (2021) studied that directly following the episode of the new COVID-19, the nations on the planet have battled to battle for controlling the disease spread and forced preventive measures to urge the populace to social removing, which prompted a worldwide emergency. COVID-19 spread control of 2019 were studied through different strategies for identification of factors. In their research work, the impacts of controlling plans on COVID-19 spread were discussed, also supervised machine learning tools were examined that was model based and recommended in favor of disease control in terms of accuracy and efficiency. In their study, three popular classifiers Naive Bayes, Multilayer Perceptron and J48 were examined on COVID-19 disease spread data that was a primary data collected through a questionnaire filled out by Basra City residents. That primary data collection through questionnaire contains almost 50 queries that linked to and had a major impact on the disease control of COVID-19, covers health care management, precautions, demographic, psychological and cognitive variables. The size of data set contains one thousand and seventeen (1017) entries of residents. Weka 3.8 tool was utilized for building a model. Outcomes clearly arrows that quarantine was a major factor in controlling the disease spread. J48 were declared the best choice among all three algorithms in respect of accuracy and efficiency.

3. Methodology:

The aim of this study is to evaluate the impact of COVID-19 on education worldwide. During the research, predictive analysis has been carried out based on a survey dataset which can be developed into an efficient forecasting system that will provide the upcoming pictures of school closures. These stages include processes such as initiation, pre-service, and post-service professionalization. Besides, there are many factors involved in the production of teachers including training and qualification, the availability of materials, curriculum formation, and teacher development. In this project, Machine learning and Ann both are used to implement the prediction problems. For this project ANN is used. Its flexibility and adaptability as well as scalability capacity made it appropriate for examination system evaluation (Faisal et al. 2021). The ANN is used for the implementation part you can see in the result and analysis chapter.

3.1. Data mining:

Data mining is an important technique when any organization or company is talking about big data. The data mining era can look for doubtlessly treasured information from a major measure of records, explicitly separated into insights practice and insights mining, and articulation and assessment of impacts. The records in the information base are handled and dissected by way of analyzing the strategies of the structure including storage, layout, control, and application of the database (Dogan and Birant. 2021).

3.2. Dataset:

One component that becomes clear is that we want to recall factors that mirror the mastering surroundings of the student in those times more as it should be. For this, we used the dataset of Jordanian university students as our number one dataset, which recorded responses from college students in a survey questionnaire form. This dataset is very useful as it now not best elicits responses from students on their use of virtual gear for studying however also takes into account the psychological effect as a result of their excessive use, which in flip will become an important component in a student's instructional performance.

First of all, we have to import all the necessary libraries like seaborn, pandas, matplotib, and TensorFlow. The dataset is of student surveys just checking the performance before and after covid in their academic records. The dataset is of CSV file before data processing you have to read the data and put it into data frames. The Category encode is mostly used for algorithm to understand the data or modify data that can help the algorithm to categories the complicated data into simple format. Down I the example this is used. The category encoders needs to be installed before making the data categories. After categorizing the data the dataset values will be changed to label data. The label data will be converted into a number from 1-5 each number representing the label data 'Strongly Disagree: 0, 'Disagree: 1, Uncertain: 2, Agree: 3, strongly Agree: 4, Agree: 3, Uncertain: 2, strongly agree: 4, strongly disagree: 0. After this sets the data is completely changed according to our need now some columns need to be removed like "Your cumulative average (GPA)". Smote function is used to fit x, and y values the basic function of smote is SMOTE synthesizes new minority times among present minority instances. It generates the digital schooling statistics with the aid of straight addition for the minority class. These fake preparation realities are created by arbitrarily picking one or more prominent of the OK closest companions for each model in the minority magnificence. After the oversampling process, the information is reconstructed and numerous types of models may be carried out for the processed information. After the sampling x and y, the splitting of test cases will be performed the data is split into training and testing sets after that import classification report, confusion matrix, and accuracy_score. The dataset and test and train set are ready for algorithm now we will apply the ANN classifier to the data. First initializing the model Ann. Then you have to add the input layer and the first hidden layer. Then add the second hidden layer and output layer and compile the model.

4. Results and analysis:

The analysis is performed on the dataset to show what plots, count and accuracy with respect to model and dataset. First a count is put with the help of count function the count function works on numeric data but before the data is label you cannot identify the each CGPA count so the count function is performed on label data of 1-4 CGPA. CGPA count means we have implement count function to the category data or simple GPA. See the methodology above after the confirmation of the dataset the count function will performed and it show the count per CGPA please see the figure 1 below.

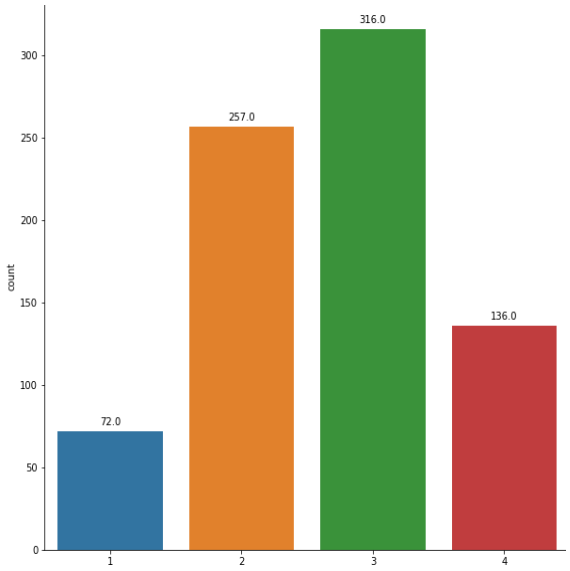

Figure 1 (Count of Different CGPA Categories)

This the graphical representation of the CGPA count. It simply explaining how KDE plots are generate. KDE plots displaying the distribution of 500 random capabilities at specific tiers of analysis. Every row represents a particular degree of evaluation (county, person, and message) and each column represents a specific sort of feature. The bar at the left of every plot represents the percentage of observations that are 0 for every feature wherein the shading represents the percent of functions accomplishing the given threshold. As the bar gets darker it approach more functions out of 500 are 0 in that percentage of people. The right portion of every plot is based totally on standardized relative frequencies of the variables.

The interpretation is basically measuring the probability density or probability of an event that can be accorded due to that particular value. It can be seen that students had a better risk of having a better CGPA if they spent more than 1–three hours on digital equipment for studying after the pandemic. However, in both curves, excessive use of online gaining knowledge of equipment leads to a consistent decline in educational performance. For this reason, intuitively it can be seen that excessive use of virtual equipment may additionally damage the student's academic overall performance. More on this later. See the figure 2 below.

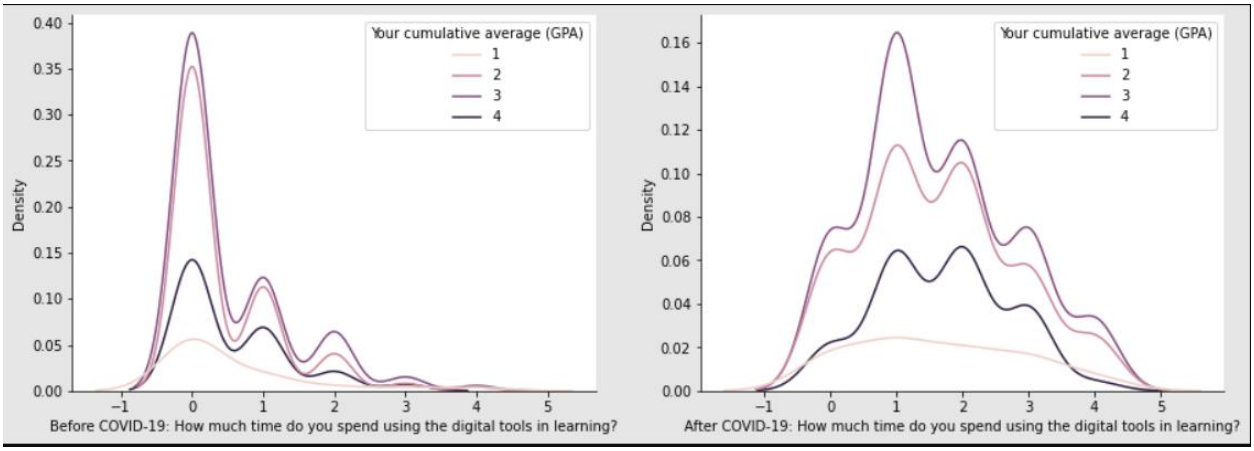

Figure 2 (Cumulative GPA Before and After Covid-19)

After this step feature selection is performed to prevent the model from over fitting. It will reduce on the bases of these 2 rules. First is the selected feature itself the correlation with the target variable and the other one is when the features are highly co-related with each other the rule is to keep only single feature to minimizing the risk for variance. The same KDE plotting was done on another factor for after_laptop via density. Please see the figure 3 below.

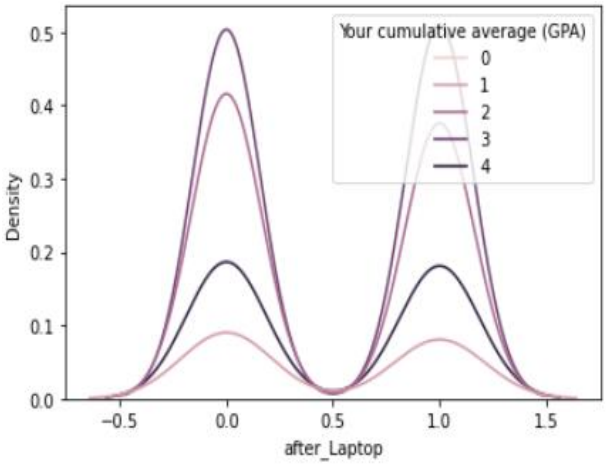

Figure 3 (Selected feature Plot)

Now to perform correlation on the original data before labeling and before changing it to another data frame data. Now what is this correlation is the correlation is the technique that appraises the alliance in-between variables or the dataset feature. All of these statistics are of high significance in technology and science and python is an important tool that helps us to calculate them. Python has different methods like Pandas correlation, NumPy, and SciPy to make the calculation faster, well documented, and complete. Corr() is how the function used in python code this function is used to ignore all the non-numeric variables and a minimum number of inspections is required per pair of columns to give the well-ground result (Sirunyan et al. 2018). Please see the figure 4 below the correlation of the complete dataset.

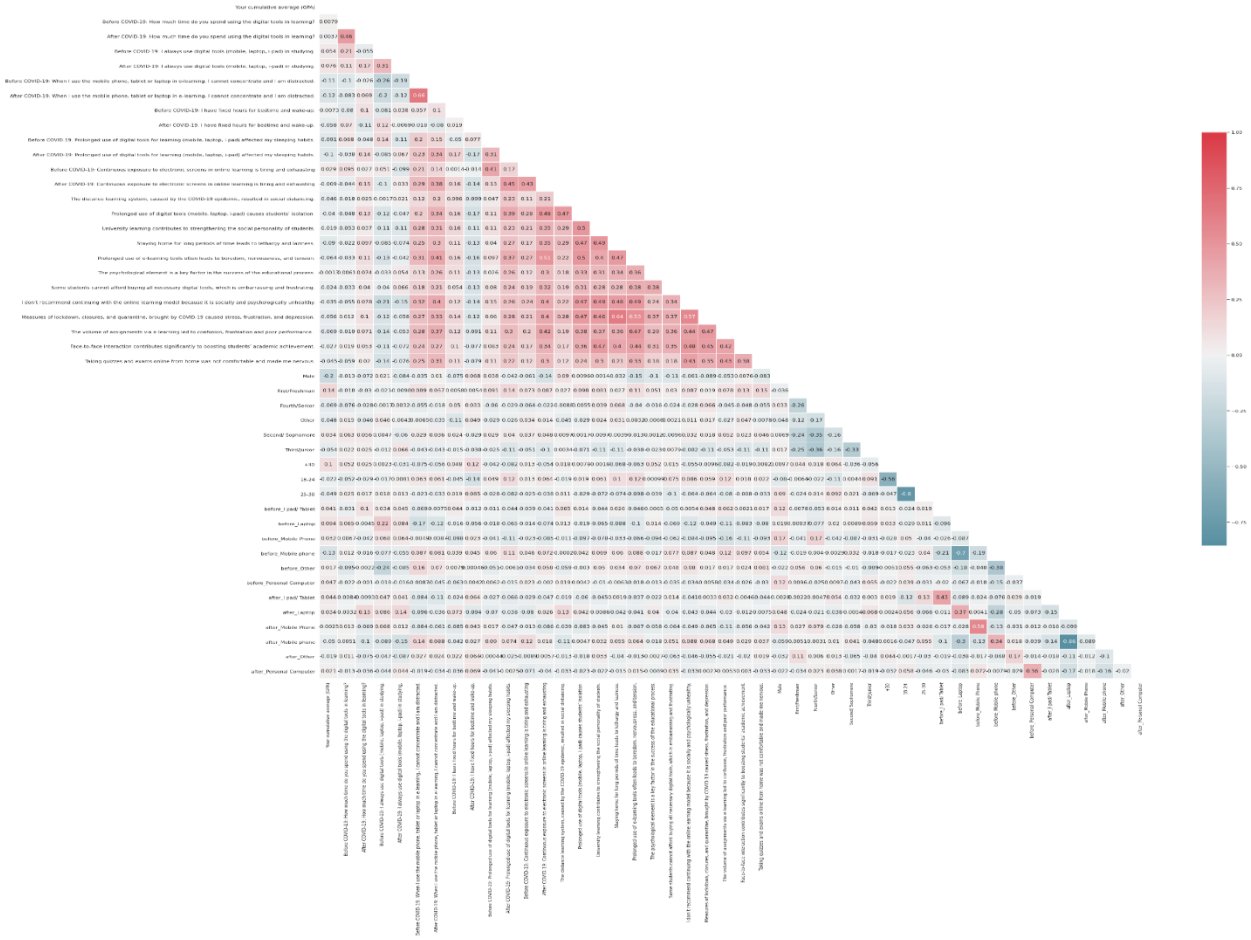

Figure 4 (Correlation of Dataset)

After the correlation of complete dataset now it time make a relation of the targeted variable that are used in this study. The relation with targeted variable can be shown using heat map. Heat map in python for each value to be plotted, a heatmap has values showing a few shades of a similar variety. The hazier tints of the outline as a rule compare to higher qualities than the lighter shades. Something else entirely can in like manner be used for an essentially unique items. Heatmaps use color changes such as hue, saturation, and brightness to depict data as 2-D colored maps. Heatmaps use colors instead of numbers to depict relationships between variables. See the figure 5 below that is show the heatmap of targeted variables.

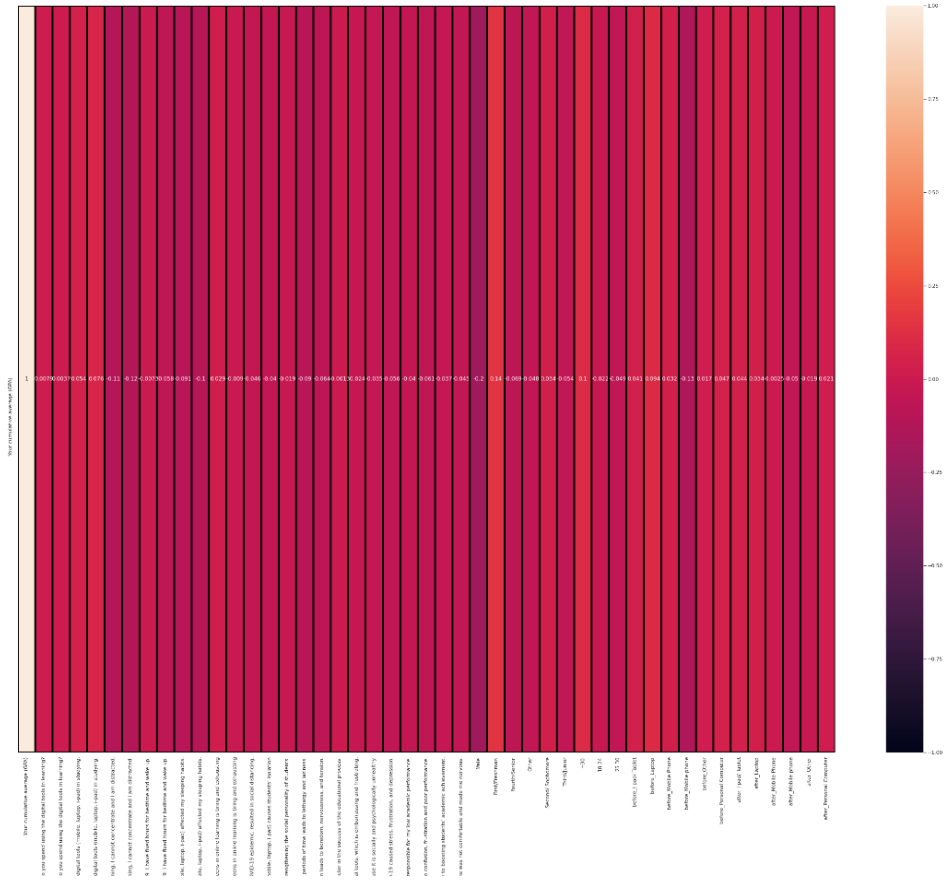

Figure 5 (Heatmap)

5. Conclusion:

The data mining era can look for doubtlessly treasured information from a big amount of records, specifically divided into statistics practice and statistics mining, and expression and evaluation of effects. For this, the dataset is used of Jordanian university students as our number one dataset, which recorded responses from college students in a survey questionnaire form. This dataset is very useful as it now not best elicits responses from students on their use of virtual gear for studying however also takes into account the psychological effect as a result of their excessive use, which in flip will become an important component in a student's instructional

performance. First of all, all the necessary is imported libraries like seaborn, pandas, matplotib, and TensorFlow. First is the selected feature itself the correlation with the target variable and the other one is when the features are highly co-related with each other the rule is to keep only single feature to minimizing the risk for variance. Please see the figure 3. After the correlation of complete dataset now it time make a relation of the targeted variable that are used in this study. The darker hues of the chart usually correspond to higher values than the lighter shades.

5.1 Future implications and recommendations

We have to moderate techniques and tools for improving our online education structures and system so that one can study in a stress-free environment with the robust teaching and learning techniques. Also, the moderate algorithms in terms of efficiency and accuracy of their implantation that can aid the systems for decision making and policy making in relevance of the vital pandemics in the world. To wrap things up, notwithstanding various limits of the dataset, absence of information about such an obscure sickness and changes in infectious prevention approaches in various nations during the period under a magnifying glass, the proposed model demonstrated viable in foreseeing the worldwide rate of COVID-19.

In this day and age, understudies are excessively upheld for their schooling by tutors and speakers. The examples, yet additionally consciousness of COVID-19 and lockdown measures. The essential obligations of a coach and a teacher are to help understudies in easing pressure and giving twofold kinship to their close to home sentiments. Because of COVID-19, the vulnerabilities connected with their assessments and development way via entry level positions, occupations, and so on, are a significant reason for mental pressure for the understudies. The outcomes show that the understudy accepts that a web-based class or virtual classes can be utilized to enhance information yet can't supplant face to face learning in a classroom because of one-on-one connection.

The main and significant obstruction is the shortage of unique dataset thus there are less experts that order the information intended for such a safe kind of human contamination. Misleadingly wise advances could be handily constructed then streamlined for integrating new AI models and utilizing the possibility to change and interface them with clinical preliminaries information to determine the arising COVID-19 local area and furthermore the inventive troubles related with that too.

The studies in this domain could assist researchers and scientists with coming too distributed explores in regards to DM methods and furious pandemics simpler. In this study, the large portion of these strategies have been created in the ongoing setting to forestall and foresee the COVID-19 plague.

6. References:

1. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Umar, A.M., Linus, O.U., Arshad, H., Kazaure, A.A., Gana, U. and Kiru, M.U., 2019. Comprehensive review of artificial neural network applications to pattern recognition. IEEE Access, 7, pp.158820-158846.
2. Ahouz, F., Golabpour, A. Predicting the incidence of COVID-19 using data mining. BMC Public Health **21,** 1087 (2021). https://doi.org/10.1186/s12889-021-11058-3
3. Alghamdi AA (2021) Impact of the COVID-19 pandemic on the social and educational aspects of Saudi university students' lives. PLoS ONE 16(4): e0250026.

https://doi.org/10.1371/journal.pone.0250026

4. Almodaresi, F., Ungar, L., Kulkarni, V., Zakeri, M., Giorgi, S. and Schwartz, H.A., 2017, July. On the distribution of lexical features at multiple levels of analysis. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 79-84).
5. Dogan, A. and Birant, D., 2021. Machine learning and data mining in manufacturing. Expert Systems with Applications, 166, p.114060.
6. Faisal, F., Nishat, M.M., Mahbub, M.A., Shawon, M.M.I. and Alvi, M.M.U.H., 2021, August. Covid-19 and its impact on school closures: a predictive analysis using machine learning algorithms. In 2021 International Conference on Science & Contemporary Technologies (ICSCT) (pp. 1-6). IEEE.
7. Li Jia and Jiang Yu hong, "The Research Trend of Big Data in Education and the Impact of Teacher Psychology on Educational Development During COVID-19: A Systematic Review and Future Perspective" , Frontiers in Psychology, vol. 12, 2021 https://www.frontiersin.org/article/10.3389/fpsyg.2021.753388, DOI=10.3389/fpsyg.2021.753388
8. Parthiban K, Pandey D, Pandey BK. Impact of SARS-CoV-2 in Online Education, Predicting and Contrasting Mental Stress of Young Students: A Machine Learning Approach. Augmented Human Research. 2021;6(1):10. doi:10.1007/s41133-021-00048-0
9. Rodríguez-Hernández, C.F., Musso, M., Kyndt, E. and Cascallar, E., 2021. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. Computers and Education: Artificial Intelligence, 2, p.100018.
10. Safdari, R., Rezayi, S., Saeedi, S. et al. Using data mining techniques to fight and control epidemics: A scoping review. Health Technol. **11,** 759–771 (2021). https://doi.org/10.1007/s12553-021-00553-7
11. Sirunyan, A.M., Tumasyan, A., Adam, W., Ambrogi, F., Asilar, E., Bergauer, T., Brandstetter, J., Brondolin, E., Dragicevic, M., Erö, J. and Flechl, M., 2018. Search for supersymmetry in events with one lepton and multiple jets exploiting the angular correlation between the lepton and the missing transverse momentum in proton–proton collisions at s= 13TeV. Physics Letters B, 780, pp.384-409.
12. Tan, C. and Lin, J., 2021. A new QoE-based prediction model for evaluating virtual education systems with COVID-19 side effects using data mining. Soft Computing, pp.1-15.
13. Wid Akeel Awadh et al 2021 J. Phys.: Conf. Ser. **1879** 022081