Depression and Anxiety Analysis and Prediction using Decision Tree and Logistic Regression

Mersiha Ćeranić*, Samed Jukić*
*International Burch University
mersiha.ceranic@stu.ibu.edu.ba
samed.jukic@ibu.edu.ba

Original research

Abstract: *COVID-19 pandemic brought many changes in people's lifestyles. Some of those changes hurt people's mental health in different age groups. This research is done to investigate which factors contributed most to the occurrence of depressive and anxiety symptoms during COVID-19 lockdown, and what type of people in terms of age, sex, level of education, place of living, was the most exposed to the appearance of mental health disorders. 1115 people (18-85 years old) from Poland joined the research process. They fulfilled online questionnaires which were used as a basis for further research of lockdown impact on mental health. Responses are evaluated by using ML tools predicting the group of participants with signs of depression and anxiety, based on their answers to the questionnaires, and the attributes of the participants. Based on the results given by the studies, the youngest population (age 18-29), which participated in the surveys, experienced more intense depression and anxiety symptoms than participants from other age groups.*

Keywords: Anxiety, covid-19, decision tree, depression, logistic regression, tableau, Weka.

1. Introduction

During the COVID-19 lockdown, many people were facing mental health issues, where the global pandemic, caused by the spread of the COVID-19 virus, has changed statistics for mental illnesses, depression, and anxiety. The percentage of the population who suffer from mental illnesses increased significantly during the COVID-19. There were different factors, such as difficulties in family relationships, lack of live communication within the community, insecurity as the virus spread, etc., causing people to feel depressive and anxiety symptoms. Various studies have shown that people at a younger age were more affected by having depression and anxiety symptoms than the older ones throughout the lockdown. Relationship difficulties and bad communication at home are determined to be the most common factors of having anxiety and depression symptoms among all participants throughout the pandemic, regardless of age. Besides recently mentioned, frequent factors of mental health issues between the youngest throughout the pandemic were impediments related to constraints outside [1].

This research is conducted to determine which factors have the greatest impact on the occurrence of depression and anxiety symptoms among the population, considering their age, gender, financial situation, etc. Besides that, we wanted to investigate the effects of the COVID-19 pandemic on the mental health of people aged 18-85, and the possible long-term consequences of the lockdown on it. This will help prevent the occurrence of further bad outcomes for people's mental health.

The result of this research process will be the implementation of Machine Learning models, which will classify the people who participated in the research process into two classes: those who have symptoms of depression and anxiety, and those who have not. Classification is based on the attributes/characteristics and the answers provided by the people who participated in these studies.

The following chapter, after “Introduction”, is “Literature review”, where we will make an overview of similar studies. Chapter “Methods” will represent the process of chain decisions and research that will shape the final solution. Surveys that are used for gathering participants’ characteristics, will be described here. Within this chapter, we will also give a brief overview of data used in this work, but also the steps that data needs to go through before creating a Machine Learning model. This chapter will give a brief overview of the Machine Learning classification algorithms that were used for the implementation of Machine Learning models. Besides that, this chapter is providing information about the tools used during the work. Chapter “Results” will present how accurate are Decision Tree and Logistic Regression ML models are in detecting which records, based on their features and answers to the online questionnaires, might have symptoms of depression and anxiety, and which records might not have. Besides that, this chapter also explains in detail the performance evaluation of implemented Machine Learning models, by using accuracy parameters. The chapter “Discussion” will be presented a comparison between our studies and similar ones. The “Conclusion” chapter will conclude our studies.

2. Literature Review

Mental wellbeing during coronavirus disease (COVID-19) has negatively impacted many people around the globe. Reports suggest that almost four times as many people stated clinically important signs of depression or anxiety during January 2021 than in January through June 2019 [2]. Prevention and control measures such as social distancing, closing down businesses and distance learning created negative conditions which lead to negative consequences for persons' mental health. The COVID-19 depression rate is different among different groups of people. For example, youths will more often notify depression or anxiety signals than older age groups, and people with lower income are more likely to declare that stress related to the COVID-19 period had a major negative influence on their mental wellbeing.

Alarming levels of unfolding and severity lead the World Health Organization (WHO) to declare the coronavirus (COVID-19) natural event a worldwide crisis on March 11, 2020. This caused several countries around the world to start out implementing countermeasures within the style of limiting people's movement and interactions like the closure of schools, kindergartens, and nurseries. Closing down schools and businesses, as well as isolation from friends and relatives, brought many concerns about the mental state of younger ones. Constraints caused by a pandemic impacted the economy negatively, which brought many difficulties in maintaining mental balance. Parental job loss and financial insecurity also contributed to psychological distress becoming a rising trend among youth.

The overall aim of these studies is to identify and explore the symptoms of anxiety and depression before and during the COVID-19 period. In this respect, the research objectives that arise from the review of the literature is to do more research on which factors contributed most to the occurrence of depressive and anxiety symptoms during COVID-19 lockdown, and what type of people in terms of age, sex, level of education, place of living, was the most exposed to the appearance of mental health disorders.

Women, especially younger ones are more exposed to perceiving anxiety and depression during elevated stress periods [3]. Job loss, financial insecurity, fear of contracting the virus, and consequences of lockdown restrictions were high contributors to the poor mental state between men and women, regardless of age [4]. Many people across the world are affected by mental issues. The types of mental illness are placed in the world for a long time. Depression and anxiety are the most common types of mental disorders [5].

Anxiety can include concerns about problems such as money, health, and/or family problems. People who suffer from anxiety are extremely worried about the mentioned or some other things, even when there is no need for them. They are very worried about how they will get through the day, think negatively, underestimate themselves, and think that all things will go wrong. Studies gave the results that in 2017, American patients, mostly affected by anxiety, were aged 30 – 44. The following group was 18 – 29, which presents 22.3%, and group 45 – 59 presents 20.6%. People aged 60+ were the least affected in 2017.

Illnesses like depression are feelings that cause an enduring feeling of unhappiness and/or loss of interest. Depression is additionally referred to as affective disorder or emotional. It greatly affects people's behavior, thinking, and feeling. Those that suffer from depression might have issues with daily activities, and additionally to the flow, they often feel that living isn't worthy. Quite 264 million people worldwide suffer from depression. Someone stricken by depression will suffer severely and perform poorly at work, college and family. Worst of all, a mental state like depression can result in suicide. About 800,000 people, typically aged 15 -19, die thanks to suicide yearly. Between 76% and 85% of individuals in poor income, counties have no treatment for their disorder [6].

Statistics for sickness, depression, and anxiety are being modified thanks to the COVID-19, wherever for the studies were participated 398,771 patients. The share of the population who suffer from mental illnesses was increased during the COVID-19, where the report for the sickness gave the results: 28% for depression, ~27% for anxiety, ~24% for post-traumatic stress symptoms, ~37% for stress, 50% for psychological stress and ~28% for insomnia issues.

A. Related work

In an article by Marco Delmastro & Giorgia Zamariola, we can find most likely the first study of the impact of COVID-19 on mental well-being on a random and representative sample of the population of Italy. As we all know COVID-19 had a tremendous influence on social, personal and many other aspects of our lives and Italy was one of the first countries in Europe that experienced large-scale issues in the health sector due to COVID. This research included more than five thousand individuals who shared their gender, age, location, living situation, and socioeconomic status were also considered. Italy was the first country in Europe that reported the first death related to COVID-19 which took place on February 21st, 2020. Following this Italy was also the first country in Europe which initiated full-state lockdown measures due to the rapid spreading of the infection and the number of deaths. The lockdown period included travel restrictions, closure of schools, and nonessential industries. Requests to stay at home and avoid live meetings with people to inhibit the spreading had the potential to impact people's mental state in an extremely negative way, fear of getting infected, worry about death, and anxiety about uncertainties all contributed to worsening of mental state. People's exposure to constant COVID-related news also produced negative effects since much of that information was misleading or harmful. This data was collected using online questionnaires and social media. The study focused primarily on adults and young adults who completed self-administered questionnaires. The data collection was conducted using a mix of CATI and CAWI methods to limit the risks of self-selection and sample distortion. All the research happened based on the Declaration of Helsinki, it was also approved by the ethics commission. Data that was used for research was provided by the national department of civil protection. COVID-19 crisis was leading to the sharp rise of the number of people with poor economic status. The study showed that mental health is just as important as the physical when it comes to one's well-being [7].

Article by Ayesha Kamran Ulhaq, Amira Khattak, Noreen Jamil et al. (2020), describes in which way Data Analytics and its features can contribute to handling mental disorders in the right way. Worldwide, many people are suffering from mental issues daily. Adequate therapy and prevention would make their lives much easier. This is where Big Data comes into place. Researchers are working hard to find appropriate solutions that will predict future data, so the outcome would serve its purpose – prevent the increase of mental health issues among the population and predict mental illness. Through this article, it is displayed how Data Science is helping in predicting mental illness by utilizing Artificial Intelligence and Big Data, and also how mental disorder occurrence can be predicted by using Smart Devices. By this research, it is presented how social media is playing a very important role in predicting mental disorders. They were using Twitter data to realize these studies. Besides, the authors also have shown what challenges we can face when working with Big Data. By using Machine Learning algorithms for predicting future states/data to know how mental states can be improved or worsened, many lives can be saved in the future. Properly dealing with mental health issues will reduce the death rate and save many people from harming themselves, or at worst committing suicide [8].

In the article by Robert Stewart and Katrina Davis (2016), great interest in researching 'Big Data' resources are found in healthcare. However, to date, the applications of 'Big Data in the mental states have been left significantly restricted. When it comes to Big Data, the big challenge comes with the size of data sets, the speed of data collection, and the diversity of data. Numerous examples can be used for health studies, including the ones that are taken from huge collections of biological samples, research with a lot of complexity, social media, and so on. Clinical notes are also potential resources for Big Data when transferring data from paper to electronic format. Over the years, the "routine" has become the use of mental health data, where one can speak of asylum records in the mid-late twentieth century. However, with digitization and modernization, most data today is accumulated in electronic format. A large range of Bid resources then emerges as "mental health research platforms". Inevitably, the features of the resources will raise questions about the availability of such data. Some databases ret full hospital information from the electronic health record from the hospital. Some of the are populated from particular info which is used from the healthy grieve service department for the research procedure. Also, some data vas unmodified administrative data, and some databases rely on patient statements and reports [9].

3. Methods

A. Questionnaires

Depression and anxiety symptoms existence was assessed by the subsequent questionnaires, respectively: Patient Health Questionnaire-9, Generalized Anxiety Disorder-7, Scale of Perceived Health and Life Risk of COVID-19, Social Support Scale, Scale of Pandemic-Related Difficulties. The following part represents the psychology analysis of questionnaires.

- The Patient Health Questionnaire-9

Consisting of 9 fundamental things and 1 extra unit, it is used as an assessment instrument of chance for depression occurrence. Participants in the study gave answers from 0 - not at all to 3 - nearly every day.

- The Generalized Anxiety Disorder-7

Consisting of 7 things, it is used as an assessment instrument of chance for any disturbance occurrence, with the focus on concerning the frequency of signals throughout the last 14 days. People who participated in the studies answered from 0 - no to 3 - almost every day.

- Scale of Perceived Health and Life Risk of COVID-19

This survey was based on the research of the following possible threats: 1) COVID-19 contagion; 2) health problems that can be caused by a virus contagion; 3) life hazard as the outcome of the contagion. Every research field mentioned above had focus on the following things: the first one relates to the participants personally and the sec one to their loved ones. They answered from 1 - very low to 5 - very high.

- Scale of Epidemic-Related Difficulties

In this survey, people who joined the questionnaires were focused to give the answers to which pandemic-related difficulty they are dealing with at the moment. They answered from 1 - not an issue for me to 4 - it is an issue for me at all.

- Social Support Scale

In this survey, people who participated in the studies gave the answers if they are getting any of the following three kinds of social support: mental, physical, or communication and gentle/kind support. They gave answers from 1 - not getting any kind of support to 5 - yes.

B. Participants' Features Overview

Depression and anxiety analysis was done among 1115 participants, aged 18-85. Studies enclosed 563 females (50.5%) and 552 males (49.5%).

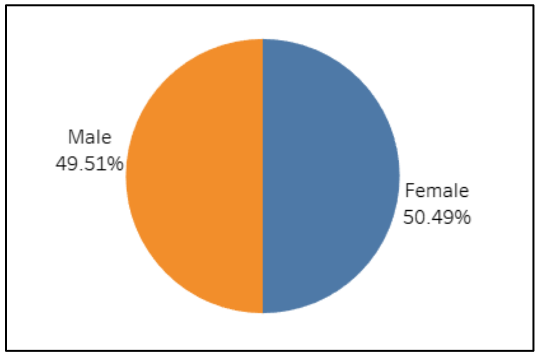

FIGURE 1. Percentage of participants by sex.

Studies brought together participants from four age groups: emerging adulthood (18-29 y.o.), established adulthood (30-44 y.o.), middle adulthood (45-59 y.o.), late adulthood (60-85 y.o.).

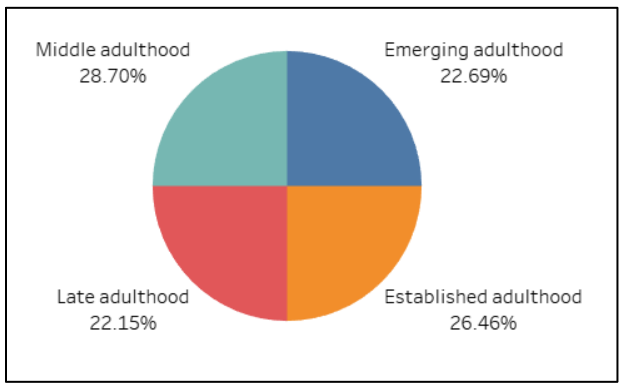

FIGURE 2. Percentage of participants by age group.

The education level of the people who participated in the studies is presented through the following levels: primary education, vocational training, secondary education, postsecondary education, university degree, no education.

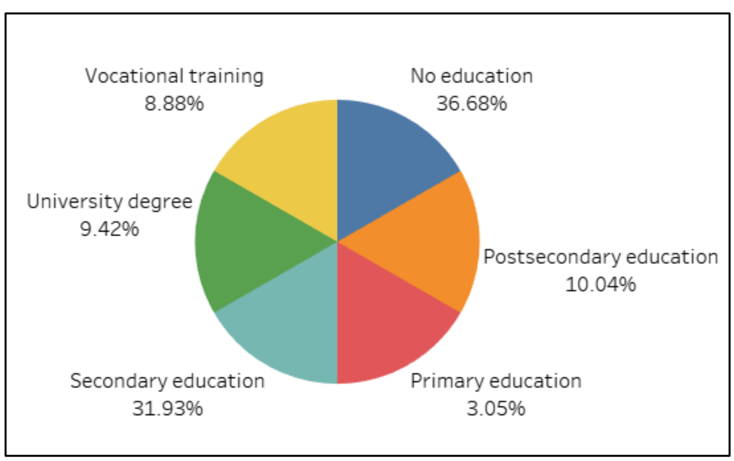

FIGURE 3. Percentage of participants by education level.

Participants were also asked whether, or not, there were any changes regarding the finances and income during the pandemic period. They gave the answers: 1 - has worsened to 7 - has improved. People who participated in the studies had also questions regarding income. They answered as no, yes, or didn't answer at all. Besides financial situation and income continuity, participants were asked whether they suffered from any pre-existing medical conditions that could cause or accelerate COVID-19 infection. They gave answers as no, yes, or don't know. As a part of questionnaires, participants were also asked about their employment status, or whether they are students or not. By checking the Class column, labeled by yes, or no, depending on whether the participant has signs of depression and anxiety, or not, we can see that 50.85% of participants has signs of depression and anxiety, while 49.15% of participants doesn't have any signs of depression and anxiety.

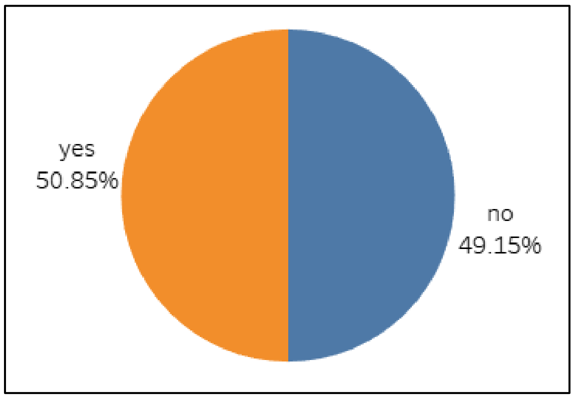

FIGURE 4. Percentage of participants depending on whether a participant has (yes), or has not showed (no) symptoms of depression and anxiety.

C. Preprocessing

During the work on these studies, after aggregation of the data, the next step was to convert them into an acceptable format for any work. Data preprocessing is the method of transforming raw data into a graspable format. Collected data is usually incomplete, noisy, inconsistent, and redundant. Preprocessing data is an important step to reinforce data effectiveness. This method involves numerous steps that facilitate converting raw data into a processed and wise format that is prepared for further work.

The first part of data preprocessing for this study is done in the Jupyter notebook. Data is cleaned by deleting unnecessary columns, and columns with empty fields are filled with a. The second part of the data preprocessing is done in the Excel tool. We were using Excel formulas to calculate values in the target column, which shows whether the particular record has/has not any signs of depression and anxiety, based on the answers to questionnaires. After data is preprocessed, it is prepared for any further training and testing model phases.

D. Machine Learning Methods

After data cleaning, pre-processing, and wrangling, the next step to be done is to feed it to the model and obtain output in probabilities. Since this is a classification problem, classification algorithms are used to predict if participants, aged 18-85, have symptoms of depression and anxiety, or not. Classification rules are determined within the training phase, and also the same doesn't seem to be modified within the validation/testing stage. Classification algorithms used to predict whether or not the participant has symptoms of depression and anxiety are:

- *Decision tree (J48)* is among the most popular decision tree algorithms in machine learning written works. If the decision tree is simply too populated a tree pruning methodology will be accustomed to remove inessential attributes and restart classification operation once more.
- *Logistic Regression* may be a linear algorithm that is straightforward and fast, however can be very effective on some kinds of problems. It works by predicting class probabilities instead of actual classes and uses logic transform to predict probabilities directly. The J48 algorithm for instance uses probabilities internally to assist with pruning. The logistic regression solely supports binary classification issues, although the Weka implementation has been changed to support one-vs-all classification issues.

While working on this research and master thesis, we were using two tools: *Weka*, to test and train the ML models, and *Tableau*, to visualize the data.

4. Results

A. Performance Evaluation

To measure accuracy for machine learning classification, we tended to use a *confusion matrix*. Within the confusion matrix, every row in a table points out the samples of predicted values whereas every column represents the samples of the actual values (or vice versa). Members of the dataset that are classified properly are set on the matrix's diagonal. The accuracy of the algorithms is calculated by the division of the total sum of all elements by a trace of the matrix (sum of the diagonal elements). If we have a case of binary classifier (which is in our case), then we will have only two labels: "Normal" and "Abnormal".

Detailed accuracy by class for all four samples (True Positive, True Negative, False Positive, False Negative) is presented through the following parameters: TPR/Recall, FPR, Precision, F-Score, MCC, ROC Area, PRC Area. The meaning of parameters is presented through the following bullet points.

TABLE 1. Confusion matrix illustrated with two-class system.

	Predicted: Abnormal	**Predicted:** Normal
Actual: Abnormal	True Positive	False Negative
Actual: Normal	False Positive	True Negative

- **TPR/Recall**

True Positive Rate (TPR) is used to measure the percentage of actual positives that are accurately determined as positive ones. TPR is calculated as a division of the total count of accurately determining positive samples (TP) and total count of positive samples (TP + FN):

$$TPR = \frac{TP}{TP + FN}$$

On the other side, *True Negative Rate (TNR)* is an outcome where the model is correctly predicting the negative class (actual negatives which are correctly identified). TNR is calculated as a division of the total count of accurately determining negative samples (TN) and total count of negative samples (TN + FP):

$$TNR = \frac{TN}{TN + FP}$$

- **FPRconstant value**

False Positive Rate (FPR) is used to measure the percentage of actual negatives which are inaccurately determined as positive samples. FPR is calculated as a division of the total count of negative samples inaccurately determined as positive samples (FP) and total count of negative samples (FP + TN):

$$FPR = \frac{FP}{FP + TN}$$

On the other side, *FNR (False Negative Rate)* is used to measure the percentage of actual positives that are inaccurately determined as negative samples. FNR is calculated as a division of the total count of positive samples inaccurately determined as negative samples (FN) and total count of positive samples (FN + TP):

$$FNR = \frac{FN}{FN + TP}$$

When examining model accuracy, typically two main measures that are considered are TPR and FPR.

- **Precision**

Precision is used to measure how precise, or accurate our model is - out of those predicted positive, how many of them are positive. Precision is calculated as a division of the total count of positive samples accurately determined as positive samples (TP) and total count of predicted positive samples (TP + FP):

$$Precision = \frac{TP}{TP + FP}$$

- **F-Score**

F-Score/ F1-Score is a criterion of a model's accuracy on a dataset. It is used when evaluating binary classification systems, where examples are classified into 'positive', or 'negative'. F-Score is the harmonic mean of the precision and recall:

$$F1\ Score = \frac{Precision * Recall}{Precision + Recall} * 2$$

- **MCC**

Matthews correlation coefficient (MCC) is a coefficient of correlation of the observed about predicted binary classifications. Outcome values are in a range from −1 to +1.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

From this formula for calculating MCC, we can conclude following:
- when FP = FN = 0 (perfect classifier), MCC = 1, indicating perfect positive correlation
- when TP = TN = 0 (classifier is misclassifying), MCC = -1, indicating perfect negative correlation

- **ROC Area**

ROC (Receiver Operator Characteristic) Area, or Area under the ROC (AUC-ROC) serves for measurement of the 2D area under the whole ROC in a range from (0,0) to (1,1). ROC displays the performance of the model at all classification inceptions.

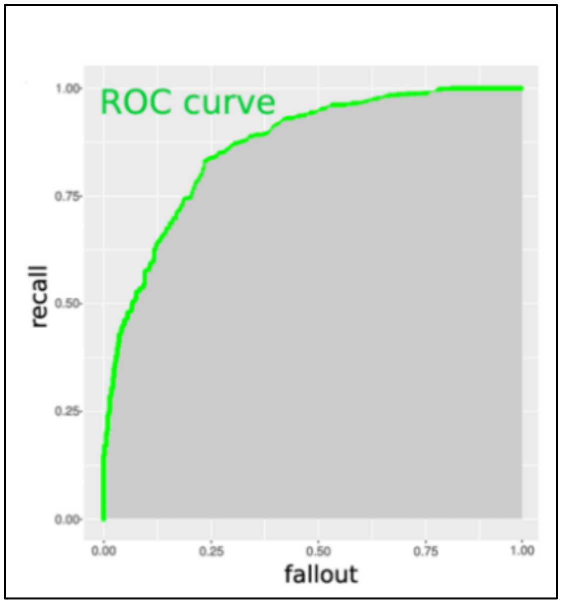

FIGURE 5. The grey area is AUC-ROC (Area under the ROC curve), where the recall (True Positive Rate) score is positioned on the y-axis and the fallout (False Positive Rate) score on the x-axis.

- **PRC Area**

PRC (Precision-Recall Curve) Area, or Area under the PRC (AUC-PRC) is used to measure 2D area under the whole PRC curve in a range from (0,0) to (1,1). The precision-recall curve (PRC) shows precision values for corresponding recall values.

FIGURE 6. The grey area is AUC-PRC (Area under the PRC curve), where the precision score is positioned on the y-axis and the recall score on the x-axis.

As a part of this research, results are tested to determine which method has more acceptable performance and if there is any difference in misclassification rates that is significant for statistics. Algorithms are trained and tested on the common dataset, on the true labels. The method that has better performance is shown as the more acceptable one.

A. Results by Classification algorithms

To obtain classification models results, we tended to use Decision Tree (J48) and Logistic Regression machine learning algorithms, where results were obtained using the Weka tool. Classification models are tested using percentage split choice, where the dataset was split in the ratio of 80:20. *That is 80% of data goes to the training set, 20% to the test set.* It is vital that the data that is being used to train and which is being used to test the model respect as similar statistical distribution as possible, w*hereas calculating performance measures,* precision, recall, and accuracy should be high as possible.
For the decision tree (J48) classification model, the accuracy score results were 86.0987%. Our Decision Tree (J48) model has a Precision of 0.861 - in other words, when our ML model predicts a participant has symptoms of depression and anxiety, it is correct 86.1% of the time. Our Decision Tree (J48) ML model has a Recall of 0.861 - in other words, it correctly identifies 86.1% of all symptoms of depression and anxiety. Additionally, if F1-Score has good value, that would be a sign of a fine Precision and a fine Recall value, too. In the case of our Decision Tree (J48) ML model, F1-Score = 0.861 indicates that both, Precision and Recall have good values. We have the value of 0.824 as the ROC Area (AUC-ROC), which is a good score. In simplest terms, this means that the model will be able to distinguish the participants with the signs of depression and anxiety and those with no symptoms 82% of the time. Just as for AUC-ROC, we got a good PRC Area (AUC-PRC) of around 78%. In Table 3. are interpreted confusion matrix values for the Decision Tree (J48) model, where a=no, b=yes.

TABLE 2. Detailed accuracy by class for Decision Tree (J48).

	TPR	FPR	Precision	Recall	F-Score	MCC	ROC Area	PRC Area	Class
	0.858	0.137	0.850	0.858	0.854	0.721	0.824	0.762	no
	0.863	0.142	0.871	0.863	0.867	0.721	0.824	0.803	yes
Weighted Avg	0.861	0.139	0.861	0.861	0.861	0.721	0.824	0.783	

TABLE 3. Confusion matrix for Decision Tree (J48) model.

	Predicted (a):	Predicted (b):
Actual (a):	91	15
Actual (b):	16	101

For the Logistic Regression classification model, the accuracy score results in 83.4081%, and in Table 4. The accuracy by class for the Logistic Regression classification model is elaborated in detail.

TABLE 4. Detailed accuracy by class for Logistic Regression.

	TPR	FPR	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.877	0.205	0.795	0.877	0.834	0.672	0.903	0.889	no
	0.795	0.123	0.877	0.795	0.834	0.672	0.903	0.924	yes
Weighted Avg	0.834	0.162	0.838	0.834	0.834	0.672	0.903	0.907	

Our Logistic Regression ML model has a Precision of 0.838 - in other words, when our ML model predicts a participant has symptoms of depression and anxiety, it is correct 83.8% of the time. Our Logistic Regression ML model has a Recall of 0.834 - in other words, it correctly identifies 83.4% of all signs of depression and anxiety. In the case of our Logistic Regression ML model, F1-Score = 0.834 indicates that both, Precision and Recall have good values. We get a value of 0.903 as ROC Area (AUC-ROC), which is a fine score. In simplest terms, this means that the model will be able to distinguish the participants with the symptoms of depression and anxiety and those with no symptoms 90% of the time. Just as for AUC-ROC, we got a good PRC Area (AUC-PRC) of around 90%. Table 5. shows confusion matrix values for the Logistic Regression classification model, where a = no, b = yes.

TABLE 5. Confusion matrix for Logistic Regression model.

	Predicted (a):	Predicted (b):
Actual (a):	93	13
Actual (b):	24	93

5. Discussion

This paper presents research based on changes in mental health caused by the irruption of the COVID-19 pandemic. By this, we can investigate more about the changes in a psychological state that people are coming through and find out in which way we can help people with mental disorders. The number of people who suffer from depression and anxiety increased significantly during the pandemic. Analysis of data is done by using ML tools predicting the group of participants with signs of depression and anxiety, based on his/her answers to the questionnaires, and the attributes of the participants. For the measurement of the performance of classification models we were using confusion matrices, and also did the comparison of the accuracy between decision tree and logistic regression model. The analysis reached 86.1 % and 83.4% classification accuracy for Decision Tree and Logistic Regression classification models, respectively.

Based on the results given by the studies, the youngest population (age 18-29), participated in the surveys, experienced more intense depression and anxiety symptoms than participants from other age groups. Results from the other studies showed that people from the oldest age groups more often experience depression and anxiety symptoms, while our studies provided conclusions that the oldest age groups were experiencing the lowest intensity of depression and anxiety signals throughout the COVID-19 period.

Both, our studies and related studies, which were based on the research of the impact of COVID-19 on mental health, found that females are more affected by depression, anxiety, and distress.

Our results show that the population with low income and financial instability were experiencing more severe symptoms of depression and anxiety, as found in previous research for population who experienced job loss and financial struggles. The pandemic period brought people online. This way of communication was more restrictive to the older generations since most of them are not pretty familiar with the technology. By this, the oldest age group felt abandonment and experienced a higher level of difficulties, related to lack of communication. In the end, lack of live meetings, building relationships, and a sense of abandonment was among the factors for depression and anxiety symptoms appearance between the younger ones (age 18-29).

6. Conclusion

Anxiety and depression symptoms were increasing significantly throughout the lockdown caused by the COVID-19 pandemic. The isolation and insecurity caused by the pandemic have made mental health even more impaired. Various factors that are predicting depression and anxiety symptoms within completely different age teams recommend that support should be distributed based on age since the performance of most tasks that were done daily has changed a lot throughout the COVID-19 period. This means that its performance needs to be adjusted, taking into account people's characteristics. This research process is used to indicate possible measurements, therapies, and solutions that will help those that were exposed to the risk the most. By being vigilant and with the possibility to identify potential threats, we will be prepared to come up with good solutions, to help in avoiding unhealthy outcomes.

7. References

[1] Gambin, M., Sekowski, M., Woźniak-Prus, M., Wnuk, A., Oleksy, T., Cudo, A., ... Maison, D. (2020, June 29). Generalized anxiety and depressive symptoms in various age groups during the COVID-19 lockdown. Specific predictors and differences in symptoms severity. https://doi.org/10.31234/osf.io/42m87

[2] Hafstad, G. S., Sætren, S. S., Wentzel-Larsen, T., & Augusti, E.-M. (2021). Adolescents' symptoms of anxiety and depression before and during the Covid-19 outbreak – A prospective population-based study of teenagers in Norway. The Lancet Regional Health - Europe, 5, 100093. https://doi.org/10.1016/j.lanepe.2021.100093

[3] Bjelland, I., Krokstad, S., Mykletun, A., Dahl, A. A., Tell, G. S., & Tambs, K. (2008). Does a higher educational level protect against anxiety and depression? The HUNT study. *Social science & medicine (1982)*, *66*(6), 1334–1345.

[4] Hammarberg, K., Tran, T., Kirkman, M., & Fisher, J. (2020). Sex and age differences in clinically significant symptoms of depression and anxiety among people in Australia in the first month of COVID-19 restrictions: A national survey. BMJ Open, 10(11), e042696. https://doi.org/10.1136/bmjopen-2020-042696

[5] Single Care Team, Anxiety statistics, https://www.singlecare.com/blog/news/anxiety-statistics/, 2021.

[6] Nochaiwong, S., Ruengorn, C., Thavorn, K., Hutton, B., Awiphan, R., Phosuya, C., Ruanta, Y., Wongpakaran, N., & Wongpakaran, T. (2021). Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: A systematic review and meta-analysis. Scientific Reports, 11(1), 10173. https://doi.org/10.1038/s41598-021-89700-8

[7] Delmastro, M., & Zamariola, G. (2020). Depressive symptoms in response to COVID-19 and lockdown: A cross-sectional study on the Italian population. Scientific Reports, 10(1), 22457. https://doi.org/10.1038/s41598-020-79850-6

[8] Kamran Ul haq, A., Khattak, A., Jamil, N., Naeem, M. A., & Mirza, F. (2020). Data Analytics in Mental Healthcare. Scientific Programming, 2020, 1–9. https://doi.org/10.1155/2020/2024160
[9] Stewart, R., & Davis, K. (2016). 'Big data' in mental health research: Current status and emerging possibilities. Social Psychiatry and Psychiatric Epidemiology, 51(8), 1055–1072. https://doi.org/10.1007/s00127-016-1266-8

Data Sharing
The dataset is administered by the *University of Warsaw, Social Sciences*. Dataset has been downloaded from the *Harvard Dataverse* site (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/0NP102).