Depression and Anxiety Diagnosis Using Unsupervised Learning Approach

Martín Di Felice, Ariel Deroche, Ilan Trupkin, Parag Chatterjee and María F. Pollo-Cattaneo

Universidad Tecnológica Nacional, Buenos Aires, Argentina

Abstract

With the increasing impact of artificial intelligence in the domain of healthcare, unsupervised learning techniques have been in discussion for the detection of hidden patterns or relationships between different data features. The domain of mental health is no exception. This paper presents an unsupervised machine learning approach to analyze the presence of depression and anxiety. It presents unsupervised models to analyze depression based on depression-related data of a student cohort containing demographic and academic data along with depression information collected through the Beck Depression Inventory questionnaire, in addition to scores such as the PHQ (Patient Health Questionnaire) score, GAD (Generalized Anxiety Disorder) score, and Epworth score, which provide insights into the severity and impact of depressive symptoms, anxiety symptoms, and daytime sleepiness, respectively. The methodology involves data collection and preparation, feature selection, modeling, anomaly detection, and analysis of the relationships between different features.

Keywords

Artificial Intelligence, Depression, Mental Health, Unsupervised Learning

1. Introduction

Depression, or depressive disorder, is a common mental health disorder that affects around 4% of the world population[1]. It can affect people of any age and its main symptoms are mood changes, loss of interest in daily activities, and sleep and appetite alterations among others[2]. It was studied that mood disorders, including different types of depression, are one of the most common comorbidities found among people who commit suicide [1]. Anxiety is another mental health disorder characterized by a significant disturbance in an individual's cognition, emotional regulation, or behavior, which generally causes excessive fear and worry[3]. Along

[1]Depressive disorder. https://www.who.int/news-room/fact-sheets/detail/depression

[2]Depression. https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007

[3]Generalized Anxiety Disorder | NEJM. https://www.nejm.org/doi/full/10.1056/nejmcp1502514

with depression, it is one of the most prevalent mental disorders[4]. On the other hand, Artificial Intelligence [2] (AI) is a branch of computer science that has become more relevant over the recent years in the domain of healthcare because of its ability to deal with problems that usually were left on human hands. There are plenty of different techniques and algorithms that can be considered part of the AI universe, but this study focuses on a particular set: unsupervised learning methods. Unsupervised learning [3] uses algorithms to analyze and segment unlabeled data sets into clusters. These algorithms discover hidden data clusters or patterns without human intervention.

Unsupervised learning makes sense when only the input data is known, and no output data corresponds to a given input. That is, only the structure of the data can be described, to try to find some pattern that simplifies the analysis. Clustering looks for groupings based on similarities, but nothing guarantees that these have any meaning. When discussing clustering techniques, one of the most popular techniques is the K-Means [4] algorithm, which is a clustering method that aims at partitioning a set of n observations into k groups in which each observation belongs to the group whose mean value is closest. According to those mentioned above, one of the main activities to be mentioned within the framework of artificial intelligence and unsupervised algorithms is anomaly detection, a process by which the characteristics or elements out of the common denominator are identified. In this way, it is possible to recognize new patterns with such out-of-the-ordinary behavior [5, 6]. Although machine learning methods are increasing their participation in mental health-related problems, the majority of these approaches use supervised learning techniques [7], so, even there exist studies that use unsupervised learning techniques to approach these kinds of problems [8, 9, 10, 11, 12], it is worth it to include this kind of algorithms which can help to face the problem from a different angle.

In this context, the objective of this paper is to use unsupervised modeling to analyze if the cohorts could be separated into different groups with distinguishing risk factors of the depression and anxiety dataset named "Depression and anxiety data" built by Shahzad [5] which focuses on college students. Although a dataset presents labeled data, the study does not use it because it aims to find hidden relationships among their values to gain better insights about depression and/or anxiety diagnoses. Besides this primary goal, a secondary objective is to analyze the relationships between the features and their impact on the development of distinguishing risk profiles in the cohorts.

The final goal of such research is to generate a model that could be included as a part of a decision support system capable of diagnosing mental health diseases such as depression or anxiety. A Clinical Decision Support System (CDSS) is a software tool that is designed to help healthcare professionals make decisions by giving personalized information about their patients[13]. The insights generated by this research are meant to be used as the basis for generating the models that given input in the form of patient data similar to the presented in the studied dataset can output probabilities about the patient belonging or not to a risk group. This paper introduces the definition of the methodology used and the results obtained, together with a comparative summary of the methods used, to finally present the conclusions and future

---

[4]Diagnostic and Statistical Manual of Mental Disorders. https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425787

[5]Depression and anxiety data. https://www.kaggle.com/datasets/shahzadahmad0402/depression-and-anxiety-data

lines of work.

2. Methodology

To achieve the proposed objective the programming language Python and its set of machine learning libraries are used, applying unsupervised learning modules of clustering and anomaly detection over the mentioned dataset. The mentioned dataset contains 783 records. Their characteristics are described below in Table 1.

To eliminate bias, and find the relationship between the different characteristics, the columns "Depression diagnosis" and "Anxiety Diagnosis" were eliminated from the dataset for both kinds of analyses. The data gathered here include age and body mass index information along with results from several psychological questionnaires: the Beck Depression Inventory[6] (a standardized method used to determine depression levels), the Beck Anxiety Inventory[7] (an analogous technique used to detect anxiety), the Patient Health Questionnaire (PHQ) [14] (another depression diagnosis test which can help to determine depression levels over the last two weeks), the Generalized Anxiety Disorder Assessment (GAD)[15] (technique used to determine anxiety levels) and the Epworth Questionnaire [16] (another test which can help determine sleepiness during daytime).

The process continues with the data preprocessing, which includes the elimination of null and duplicate values, and finally, with the data normalization using the Standard Scaling technique. Then it continues with the clustering process described in the following section, and, finally, the introduction of anomaly detection methods to see if it improves clustering results.

2.1. Clustering

Clustering is one of the most important areas within the unsupervised machine learning domain. The purpose of this process is to achieve the grouping of unlabeled data sets and to build subsets of data known as clusters. This way, each cluster within a network is formed by a collection of data similar to each other but having differential elements concerning others. This type of process is applied to unsupervised models, allowing data to be segmented into groups of similar dimensions. After obtaining the pre-processed dataset, a technique is used to determine the optimal number of clusters that will eventually be assigned using the K-Means algorithm. The method used to determine the optimal amount of clusters is the Silhouette, which measures the quality of clustering, using the separation distance between them. It tells us how close each point in a cluster is to points in neighboring clusters. This distance measure is in the range [-1, 1]. A high value indicates good clustering since it is far away from neighboring clusters. A negative value indicates that the sample might be assigned to the wrong cluster[17, 18]. For each one of the dataset versions, the Silhouette average value is calculated for a range of 2 to 10 clusters to see which performs better. Then the different clusters are analyzed to see how their values are close to the Silhouette average value.

---

[6]Beck Depression Inventory (BDI). https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/beck-depression

[7]Beck Anxiety Inventory. https://www.zotero.org/google-docs/?yIRpC0

Table 1
Characteristics of the dataset

Row	Characteristics
School year	First year: 35% Second year: 24% Third year: 23% Fourth year: 18%
Age (years)	20.26 ± 1.77
Sex (Female/Male)	Female: 52%, Male: 48%
BMI	23.41 ± 4.6
WHO BMI	Normal: 65% Overweight: 25% Underweight: 5% Class I Obesity: 4% Other: 1%
PHQ score	7.17 ± 4.42
Depression Severity	Mild: 44% None-minimal: 29% Moderate: 18% Moderately severe: 5% None: 2% Severe: 2%
Depressiveness	27% (true)
Suicidal	8% (true)
Depression diagnosis	8% (true)
Depression treatment	7% (true)
GAD score	6.88 ± 4.73
Anxiety severity	Mild: 38% None-minimal: 37% Moderate: 16% Severe: 9%
Anxiousness	25% (true)
Anxiety diagnosis	8% (true)
Anxiety treatment	8% (true)
Epworth score	6.4 ± 4
Sleepiness	18% (true)

After getting the optimal number of clustering and dividing the data into different clusters. As mentioned above, the K-Means algorithm is used to achieve such a goal. Being one of the
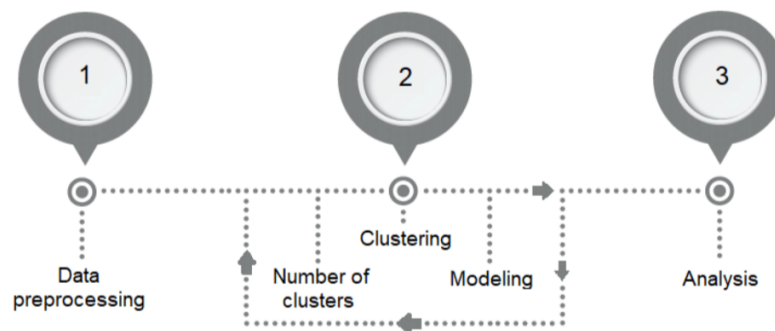
Figure 1: Clustering methodology pipeline

most powerful and popular methods for applying and understanding the grouping process [19], the K-Means algorithm works by calculating K centroids (that number, K, is the number of clusters) and for each one of the records, it calculates the Euclidean distance between them and those centroids, grouping the data in each nearest centroid [20].

After the data is grouped in different clusters, the target variables are introduced again. For each one of the clusters generated from both methods, the T-Test[8] is performed to determine if the data present significant differences among the different groups.

The clustering process is performed again, iteratively. The goal is to get the majority of depression and anxiety-diagnosed patients inside a single cluster. However, the process can finish sooner if some other conditions are met: if the unhealthy patients start to be separated among different clusters, or if the metrics for the new split do not show an improvement. This process, in case it is needed, is referred to as sub-clustering. Figure 1 summarizes the Methodology pipeline. After one cluster satisfies the conditions mentioned above, the last step is to analyze the results obtained by these processes.

2.2. Anomaly Detection

As explained above, Anomaly Detection is a technique for identifying events that are statistically different from the rest. Such anomalous behavior is usually translated as a problem. Anomaly Detection has a different standpoint, based on models and data predictions (see Figure 2).

Anomaly Detection is used for the detection of patients whose features do not match with the majority of the population, allowing analyzing them differently. This analysis can lead to seeking better metrics for the predictor models. Two groups are considered: the normal data (those values that are in match with the majority) and outliers (the anomalous values) and they are determined for each feature from the dataset, this way, it can be obtained the conclusion on the behavior of each feature. For each numeric or categorical feature, it is determined the range of normal values and its average. Also, there are identified the upper and bottom anomalous values closest to the normal ones. Besides, to improve the analysis, the average values for superior and inferior anomalous values are calculated. Taking into account the concept of

---

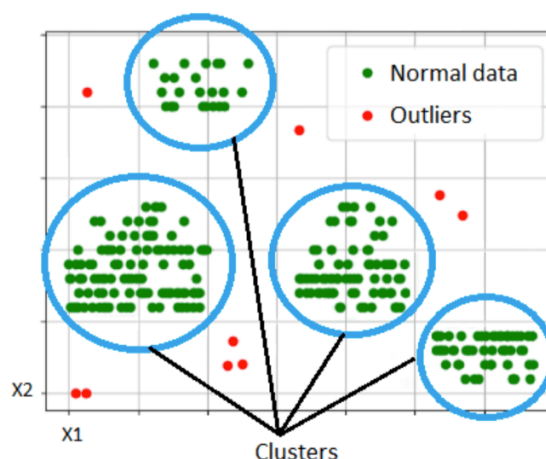[8]KoreaMed Synapse. https://synapse.koreamed.org/articles/1156170

Figure 2: Anomaly detection methodology

clustering, it can be represented with the following graph the situation in which certain data are grouped in several clusters, leaving out the so-called outliers.

Besides the detection of anomalous values, the process also includes analyzing the relationship between anomalous values from different features. Seeing whether anomalous values for a patient in a specific feature are also reflected in another different feature may help determine how those different features behave and find relationships between them. Discovering these relationships makes it possible to further justify decision-making on how to act with that patient. These are decisions that will later affect the tuning of the predictive model.

3. Results

Following a similar presentation to the previous section, the results for both approaches, Clustering and Anomaly Detection are presented.

3.1. Clustering

Before determining the optimal number of clusters, the dataset is normalized using the Standard Scaling method. Then, as mentioned in the Methodology section, the process continues by trying to obtain the optimal number of clusters. For each one of the cluster combinations, it is observed the differences between the cluster values and the Silhouette average value for that amount of clusters. For all the combinations, except the 9-cluster one, all the values are above the Silhouette average value, meaning that they are good candidates for the division.

Taking into account the Average Silhouette Width (ASW), which indicates the degree of similarity of an item in a cluster compared to others, the value ranges from -1 (the sample is not very similar to its neighbors and little different from those of other clusters) to 1 (the sample is very similar to its neighbors and very little similar to those of other clusters). According to the Silhouette analysis, the convenient number of clusters is selected in 2 to group the data, this

being the one with the highest value, considering a good cluster configuration. However, the 2-clusters division is decided because its average value is bigger.

Then, the T-Test method is used to check whether the difference between the clusters is significant enough. For the 2-cluster split, 78.95% of the P-Values for all the features are less than the selected threshold value (0.05), including 100% of the target variables, showing an important difference among the clusters, especially in the target variables which are the ones this procedure are trying to separate. Using that information, the 2-cluster division is concluded to separate the data effectively.

The data obtained for the different clusters are described following in Table 2. Table 3 shows how the target variables, once introduced again into each cluster, behave differently. Both features, in this case, fulfill the conditions analyzed by the T-Test.

It is observed that the first cluster (Cluster 0) groups all the patients with depression and anxiety levels above normal, while the second cluster (Cluster 1) tends to contain healthy patients. However, the first cluster also includes many healthy patients: it doubles the size of the second cluster (531 vs 242) while the original analysis visible in Table 1 presents only 10% of the cohort as diagnosed with one of the diseases. So, to differentiate unhealthy patients the same methodology (scaling, number of cluster determination, T-Test, clustering, and analysis) is applied again but only over this first cluster.

The Silhouette analysis indicates a value of 4 according to the obtained values for each possible number of clusters in conjunction with the chart analysis (value of 0.184).

The T-Test analysis is performed later which gives better significant differences between the features of the different sub-clusters in the 4-cluster combination (63.16%), but on the target variables, the test is passed in only 16.67% of the clusters. However, Table 4 shows how the sub-clusters are left having in Sub-cluster 1 a slightly larger proportion of unhealthy patients compared with the other groups. The T-Test analysis shows that Sub-cluster 1 has a 33.33% significant difference among the target variables with the other sub-clusters, rising to 66.67% in the case of depression diagnosis.

Table 5 shows how the target variables behave for each one of the sub-clusters once re-introduced. Again, Sub-cluster 1, groups the majority of unhealthy patients, being more significant in the depression diagnosis.

According to this grouping, and taking rows assigned to the Sub-cluster 0 as those identified as positive, a confusion matrix can be defined and, from there, determine the most popular metrics as it was a classification problem. Table 6 shows this.

Although the first attempt to group properly unhealthy patients, it creates a big set of healthy and unhealthy patients, leaving a small group of healthy patients in the smaller set, while the proportion in the dataset is exactly the opposite. When attempting to divide that second and larger cluster to identify better unhealthy patients, a minor sub-cluster is obtained which tends to group them, but the difference with the other sub-clusters is not so significant. The iterative process is stopped at this moment because the next iteration started to show that the unhealthy patients were not grouped into a single cluster anymore.

Table 2
Characteristics of the clusters

Row	Cluster 0 (n=531)	Cluster 0 (n=242)
School year*	First year: 36% Second year: 28% Third year: 15% Fourth year: 21%	First year: 33% Second year: 16% Third year: 25% Fourth year: 26%
Age (years)	20.19 ± 1.84	20.37 ± 1.59
Sex (Female/Male)*	Female: 58%, Male: 42%	Female: 38%, Male: 62%
BMI	23.19 ± 5.32	22.7 ± 4.95
WHO BMI	Normal: 63% Overweight: 26% Underweight: 5% Class I Obesity: 4% Other: 2%	Normal: 69% Overweight: 22% Underweight: 4% Class I Obesity: 3% Other: 2%
PHQ score*	9.12 ± 3.85	2.8 ± 1.7
Depression Severity*	Mild: 64% Moderate: 26% Moderately severe: 8% Severe: 2%	None-minimal: 92% None: 6% N/A: 2%
Depressiveness*	39% (true)	2% (true)
Suicidal*	12% (true)	1% (true)
Depression treatment*	9% (true)	5% (true)
GAD score*	8.03 ± 4.47	3.5 ± 2.8
Anxiety severity*	Mild: 44% Moderate: 22% None-minimal: 21% Severe: 12% N/A: 1%	None-minimal: 70% Mild: 26% Moderate: 2% Other: 2%
Anxiousness*	34% (true)	3% (true)
Anxiety treatment	9% (true)	5% (true)
Epworth score*	6.77 ± 4.26	4.69 ± 2.74
Sleepiness*	24% (true)	6% (true)

Features considered as significant according to the T-Test analysis.

3.2. Anomaly Detection

The Anomaly Detection analysis gives the results presented on the charts enumerated in Figure 3. School year, given it is a very balanced feature, does not show anomalous values, while the rest of the features do. Figures 3(a) through 3(d) represent features related to the patient body or

Table 3
Target variables

Row	Cluster 0 (n=531)	Cluster 1 (n=242)
Depression diagnosis	10% (true)	5% (true)
Anxiety diagnosis	9% (true)	5% (true)

Table 4
Characteristics of the sub-clusters

Row	Sub-cluster 0 (n=93)	Sub-cluster 1 (n=26)	Sub-cluster 2 (n=241)	Sub-cluster 3 (n=171)
School year	First year: 43% Second year: 33% Third year: 12% Fourth year: 12%	First year: 52% Second year: 31% Fourth year: 10% Third year: 7%	First year: 56% Second year: 43% Third year: 1%	Fourth year: 57% Third year: 38% Second year: 4% First year: 1%
Age (years)	19.7 ± 1.22	20.34 ± 2.02	19.19 ± 1.08	21.84 ± 1.79
Sex (Female/Male)	Female: 69%, Male: 31%	Female: 55% Male: 45%	Female: 50% Male: 50%	Female: 62%, Male: 38%
BMI	22.83 ± 3.98	31.59 ± 11.58	22.55 ± 4.04	22.86 ± 4.71
WHO BMI	Normal: 61% Overweight: 32% Underweight: 7%	N/A: 72% Class II Obesity: 17% Class III Obesity: 11%	Normal: 71% Overweight: 25% Underweight: 3% Class I Obesity: 1%	Normal: 65% Overweight: 28% Underweight: 6% Class I Obesity: 1%
PHQ score	13.96 ± 4.28	11.31 ± 4.89	7.73 ± 2.36	7.99 ± 2.57
Depression Severity	Moderate: 45% Moderately Severe: 35% Mild: 11% Severe: 9%	Mild: 48% Moderate: 28% Moderately Severe: 14% Severe: 10%	Mild: 79% Moderate: 20% Moderately Severe: 2%	Mild: 75% Moderate: 24% Moderately Severe: 1%
Depressiveness	91% (true)	58% (true)	24% (true)	28% (true)
Suicidal	32% (true)	34% (true)	7% (true)	5% (true)
Depression treatment	17% (true)	0% (true)	8% (true)	6% (true)
GAD score	13.54 ± 4.87	9.01 ± 4.04	6.58 ± 3.35	6.83 ± 3.1
Anxiety severity	Severe: 61% Moderate: 35% Mild: 3% None/Minimal: 1%	Mild: 41% Moderate: 35% Severe: 17% None/Minimal: 7%	Mild: 49% None/Minimal: 30% Moderate: 20% Other: 1%	Mild: 61% None/Minimal: 22% Moderate: 16% Other: 1%
Anxiousness	96% (true)	52% (true)	20% (true)	6% (true)
Anxiety treatment	15% (true)	3% (true)	9% (true)	6% (true)
Epworth score	9.64 ± 5	8.34 ± 5.83	5.91 ± 3.63	6.1 ± 3.53
Sleepiness	45% (true)	21% (true)	18% (true)	21% (true)

Table 5
Target variables in sub-clusters

Row	Sub-cluster 0 (n=93)	Sub-cluster 1 (n=26)	Sub-cluster 2 (n=241)	Sub-cluster 3 (n=171)
Depression diagnosis	18% (true)	4% (true)	7% (true)	10% (true)
Anxiety diagnosis	14% (true)	8% (true)	7% (true)	9% (true)

Table 6
Classification metrics

Approach	Accuracy	Recall	Precision	Specificity
epression diagnosis	0.81	0.06	0.04	0.88
Anxiety diagnosis	0.82	0.07	0.04	0.88

age, while Figures 3(e) through 3(i), which are shown following, indicate mental health-related features.

Most of the features present anomalous values above and below the set of normal values,
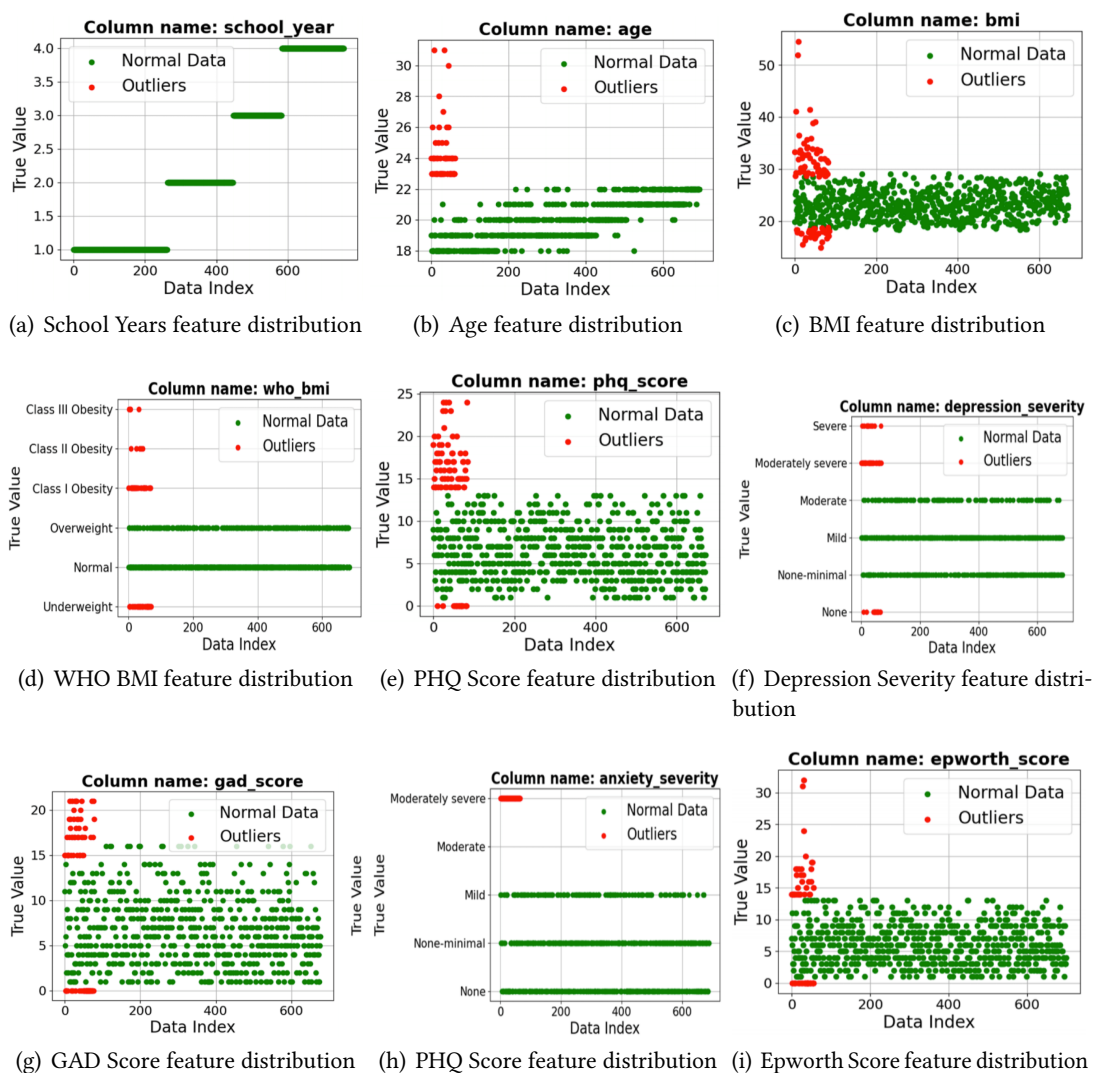

(a) School Years feature distribution   (b) Age feature distribution   (c) BMI feature distribution

(d) WHO BMI feature distribution   (e) PHQ Score feature distribution   (f) Depression Severity feature distribution

(g) GAD Score feature distribution   (h) PHQ Score feature distribution   (i) Epworth Score feature distribution

Figure 3: Anomaly Detection Analysis.

being the exceptions the features Age and Anxiety Severity, whose anomalous values are only above the normal data. Table 7 summarizes the number of normal data (normal values) and outliers (anomalous values) for all these features.

Table 7
Summary of normal data and outliers for categorical and numerical features

Feature	Normal data	Outliers
School Year	757	0
Age	695	62
BMI	674	83
WHO BMI	686	71
PHQ Score	688	69
Depression Severity	688	69
GAD Score	678	79
Anxiety Severity	691	66
Epworth Score	681	76

Also, binary columns with imbalanced data, such as Suicidal, Depression Treatment, and Anxiety Treatment are considered anomalous when the value is positive.

On the other hand, it was analyzed how different anomalous values relate among the different features. Table 8 shows these behaviors.

Table 8
Relationship between anomalous values

Feature	Feature with most mutual anomalous values	Anomalous values over total values
Age	BMI	0.18
BMI	WHO BMI	0.76
WHO BMI	BMI	0.88
PHQ Score	Depression Severity	1
Depression Severity	PHQ Score	1
GAD Score	Anxiety Severity	0.6
Anxiety Severity	GAD Score	0.71
Epworth Score	PHQ Score	0.24

It can be observed that in each case that the patient is detected as anomalous for the PHQ Score feature, it is also detected as anomalous for the feature Depression Severity and vice-versa. Something similar happens between GAD Score and Anxiety.

4. Conclusions

The clustering analysis returns good accuracy and specificity values in contrast with poor recall and precision values. This suggests that this kind of classification would be very helpful excluding healthy patients but still won't be able to precisely detect unhealthy patients. It is also observed many similarities between the depressive and anxious patient cohorts, giving the hint that both disorders share similar aspects. Anomaly Detection analysis informs which

patients may be altering the dataset so it can be part of future works on the usage of this kind of algorithm to decide what can be done with such patients for the classifier model to improve its metrics: they can be removed or analyzed separately. However, that analysis must be done on a feature level because some anomalous values could be very important to be analyzed as part of the dataset because they may indicate the presence of the disease, especially in such imbalanced data.

Another benefit that can be derived from the anomaly detection process is to find which patterns the anomalous records present to use them to classify the dataset. In this article, Anomaly Detection is used to analyze categorical and numerical features and their relationships, but in the future lines, the same analysis can be performed at a record level.

In conclusion, future open lines of work are mentioned, which include opening the same analysis to other datasets, and, after enough validation, the inclusion of these techniques in decision support systems that help mental health professionals to diagnose better and faster this kind of diseases. Also, the iterating process mentioned in the clustering process can be extended by making more indulgent the stop criteria and checking if that helps the detection of unhealthy patients.

As mentioned above, all this analysis, both the clustering and the anomaly Detection, are meant to be part of a Clinical Decision Support System pipeline, where both parts can help identify patients with diseases or at least anomalous, that may require a deeper inspection.

5. Acknowledgments

References

[1] Y. Conwell, P. R. Duberstein, C. Cox, J. H. Herrmann, N. T. Forbes, E. D. Caine, Relationships of age and axis i diagnoses in victims of completed suicide: a psychological autopsy study., The American journal of psychiatry 153 (1996) 1001–1008.
[2] M. A. Boden, Artificial intelligence, Elsevier, 1996.
[3] Z. Ghahramani, Unsupervised learning, in: Summer school on machine learning, Springer, 2003, pp. 72–112.
[4] J. Han, J. Pei, H. Tong, Data mining: concepts and techniques, Morgan kaufmann, 2022.
[5] K. G. Mehrotra, C. K. Mohan, H. Huang, K. G. Mehrotra, C. K. Mohan, H. Huang, Anomaly detection, Springer, 2017.
[6] E. O. Ogunseye, C. A. Adenusi, A. C. Nwanakwaugwu, S. A. Ajagbe, S. O. Akinola, Predictive analysis of mental health conditions using adaboost algorithm, ParadigmPlus 3 (2022) 11–26.
[7] M. Di Felice, P. Chatterjee, M. F. Pollo-Cattaneo, Depression diagnosis using text-based ai methods–a systematic review, 2022.
[8] Y. Chen, J. W. Stewart, J. Ge, B. Cheng, A. Chekroud, D. J. Hellerstein, Personalized

symptom clusters that predict depression treatment outcomes: A replication of machine learning methods, Journal of Affective Disorders Reports 11 (2023) 100470.

[9] Z. Yang, C. Chen, H. Li, L. Yao, X. Zhao, Unsupervised classifications of depression levels based on machine learning algorithms perform well as compared to traditional norm-based classifications, Frontiers in Psychiatry 11 (2020) 45.

[10] M. R. Morales, R. Levitan, Mitigating confounding factors in depression detection using an unsupervised clustering approach, in: Computing and Mental Health Workshop (CHI), San Jose, CA, USA, 2016, pp. 1–4.

[11] C. R. van Genugten, J. Schuurmans, A. W. Hoogendoorn, R. Araya, G. Andersson, R. M. Baños, T. Berger, C. Botella, A. Cerga Pashoja, R. Cieslak, et al., A data-driven clustering method for discovering profiles in the dynamics of major depressive disorder using a smartphone-based ecological momentary assessment of mood, Frontiers in psychiatry 13 (2022) 755809.

[12] J. Hernandez, K. Daza, H. Florez, Spiking neural network approach based on caenorhabditis elegans worm for classification., IAENG International Journal of Computer Science 49 (2022).

[13] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, K. I. Kroeker, An overview of clinical decision support systems: benefits, risks, and strategies for success, NPJ digital medicine 3 (2020) 17.

[14] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry, A. H. Mokdad, The phq-8 as a measure of current depression in the general population, Journal of affective disorders 114 (2009) 163–173.

[15] R. L. Spitzer, K. Kroenke, J. B. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: the gad-7, Archives of internal medicine 166 (2006) 1092–1097.

[16] M. W. Johns, A new method for measuring daytime sleepiness: the epworth sleepiness scale, sleep 14 (1991) 540–545.

[17] L. Nitya Sai, M. Sai Shreya, A. Anjan Subudhi, B. Jaya Lakshmi, K. Madhuri, Optimal k-means clustering method using silhouette coefficient, 2017.

[18] J. Hernandez, H. Florez, An experimental comparison of algorithms for nodes clustering in a neural network of caenorhabditis elegans, in: Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part IX 21, Springer, 2021, pp. 327–339.

[19] M. Ahmed, R. Seraj, S. M. S. Islam, The k-means algorithm: A comprehensive survey and performance evaluation, Electronics 9 (2020) 1295.

[20] J. A. Hartigan, Clustering algorithms john wiley & sons, Inc., New York, NY (1975).