

Research on the Influencing Factors of Depression and Anxiety Based on Anxiety dataset

Yijie Kong*

Hangzhou High School, Hangzhou, 310000, China

*Corresponding author: 1801040404@stu.hrbust.edu.cn

Abstract. This study delves into the complex relationship between depression and anxiety, aiming to elucidate their multifaceted nature and impact. Leveraging a robust dataset, it is conducted a comprehensive analysis to assess the significance of various factors associated with these mental health disorders. Through statistical techniques such as exploratory data analysis (EDA) and predictive modeling such as decision tree model, this paper explored predictors and underlying dynamics. The findings highlighted the importance of demographic, lifestyle, and psychological factors in understanding depression and anxiety. Notably, all variables examined showed significant deviations, emphasizing their relevance. Furthermore, the analysis revealed intricate interactions among factors, emphasizing the nuanced nature of these conditions. This study provides insights for healthcare policies and future research, stressing the need to consider less prominent factors and socio-cultural contexts. By comprehensively understanding the interplay of factors, the author aims to inform more effective interventions and policies in mental health.

Keywords: Depression; anxiety; mental health; exploratory data analysis.

1. Introduction

Millions of people around the world suffer from depression and anxiety, two prevalent mental health disorders. These conditions not only diminish one's quality of life but also pose significant challenges to public health systems globally. The onset of depression is marked by persistent sadness, hopelessness, and a decline in interest in previously enjoyable activities, while anxiety manifests as excessive worry, fear, and apprehension. Although distinct in their symptoms, depression and anxiety often coexist and can exacerbate each other's effects, leading to a complex web of emotional distress and impairment in daily functioning.

According to the World Health Organization (WHO), depression is the primary cause of disability worldwide, with an estimated 264 million people globally affected [1]. Anxiety disorders are equally prevalent, affecting approximately 284 million individuals worldwide (WHO, 2021). The burden of these disorders extends beyond the individual, impacting families, communities, and societies as a whole. Numerous factors contribute to the development of depression and anxiety, including genetic predisposition, brain chemistry, personality traits, and life experiences [2]. The trigger or exacerbate of symptoms can be caused by trauma, chronic stress, and significant life changes [3]. Additionally, social-cultural factors such as stigma, discrimination, and limited access to mental health services further complicate the management of these conditions [4, 5].

The prevalence of depression and anxiety varies across demographic groups and populations. Women are more likely than men to experience both disorders, with hormonal fluctuations, societal pressures, and gender-based violence contributing to this disparity [6]. Adolescents and young adults are also at heightened risk, with studies indicating a rising prevalence of these disorders among youth in recent years [7]. The economic impact of depression and anxiety is staggering, with costs associated with healthcare utilization, lost productivity, and disability accommodations. According to a report by Baghaei et al. depression alone costs the global economy over \$1 trillion annually in lost productivity [8, 9]. Moreover, individuals with untreated or poorly managed depression and anxiety are at increased risk of developing other chronic health conditions, further straining healthcare systems and exacerbating the burden on individuals and society as a whole. Despite the significant burden of depression and anxiety, access to effective treatment and support remains inadequate for

many individuals. Mental illness stigma can often prevent individuals from seeking help, while disparities in healthcare access and affordability limit the availability of evidence-based interventions. Furthermore, the complex interplay of biological, psychological, and social factors necessitates a holistic approach to treatment that addresses the underlying causes and symptoms of these disorders.

There has been a rise in recognition of the significance of lifestyle factors in managing depression and anxiety in recent years. Regular exercise, healthy dietary habits, adequate sleep, and stress-reduction techniques have been shown to complement traditional treatments and improve overall well-being [10]. Integrative approaches that combine pharmacotherapy, psychotherapy, and lifestyle interventions offer promise in addressing the multifaceted nature of depression and anxiety.

This study aims to leverage the public depression anxiety dataset to investigate the factors associated with depression and anxiety. By employing exploratory data analysis (EDA), predictive modeling, and statistical inference techniques, the author seeks to uncover patterns, relationships, and predictive factors related to these mental health disorders. The goal of this paper is to contribute to the existing body of knowledge regarding depression and anxiety through this analysis, ultimately fostering a better understanding of these conditions and informing targeted interventions and support strategies.

2. Methodology

2.1. Data Source

The depression and anxiety risk data used in the paper was obtained from the Kaggle website. The original information was stored in a CSV file format. The data was made publicly available on Kaggle by researcher Shahzad Ahmed from Beijing University of Technology. To restate, the depression and anxiety risk data was sourced from Kaggle, have been contributed by Shahzad Ahmed of the BJUT in China.

2.2. Variable Introduction

The dataset contains information gather from 783 anonymous participants ages 18-31 with school year ranging from 1-4. The data was collected through online survey and face-to-face talk covering participants' BMI, phq score, quality of sleep at night and etc. (Table 1).

Table 1. Different types of variables

Term	Type	Meaning
ID	Numeric	Unique Identifier
School_year	Numerical	Years of education in university
Age	Numerical	Participants' ages
Gender	Categorical	Male or female
Bmi	Numerical	BMI rate
Who_bmi	Categorical	Bmi classification
Phq_score	Numerical	Scores on Patient Health Questionnaire
Depression_severity	Categorical	Classification for depression
Depressiveness	Categorical	Depressed or not
Suicidal	Categorical	Have suicide experience or not
Depression_diagnosis	Categorical	Depression diagnosis from hospital
Depression_treatment	Categorical	Treatment given or not given
Gad_score	Numerical	Scores of Generalized Anxiety Disorder questionnaire
Anxiety_severity	Categorical	The severity of anxiety symptoms
Anxiousness	Categorical	A measure of general anxiousness or mood
Anxiety_diagnosis	Categorical	Whether an individual has been diagnosed
Anxiety_treatment	Categorical	Whether an individual is undergoing treatment
Epworth_score	Numerical	Scores from the Epworth Sleepiness Scale
Sleepiness	Categorical	The level of daytime sleepiness

2.3. Method Introduction

A scatter plot utilizes an X-Y coordinate plane to visually depict values for two variables, allowing exploration of the relationship between those variables. As a basic analytic tool, scatter plots are critical for statistical analysis.

The One-Sample t-Test is used to examine whether the average of a small sample is significantly different from the average of the population. It requires the sample data be approximately normal. This test works well when assessing smaller sample sizes against known or expected mean values.

The relationship between variables in a dataset is explored through correlation analysis, a fundamental statistical technique. By examining the degree and direction of association between two or more variables, correlation analysis provides valuable insights into the interdependencies within the data. In the context of mental health research, correlation analysis enables people to understand the complex interactions between various factors that contribute to psychological well-being or the manifestation of mental health disorders such as depression and anxiety.

Decision trees are a popular machine learning model known for their simplicity and interpretability. They organize data into a tree-like structure where each node represents a decision based on a feature's value. Decision trees are used for both classification and regression tasks and can capture non-linear relationships without complex mathematical assumptions.

3. Results and Discussion

3.1. One-Sample t-Test

The one-sample t-tests revealed that all 17 variables-Gender, Age, Height, Weight, family history with overweight, school_year, age, gender, bmi, who_bmi, phq_score, depression_severity, depressiveness, suicidal, depression_diagnosis, depression_treatment, gad_score, anxiety_severity, anxiousness, anxiety_diagnosis, anxiety_treatment, epworth_score, sleepiness significantly from the null value ($p < 0.05$). This indicates that the mean values of these variables are statistically distinct from zero, suggesting that they are measurable and significant features within the study population (Table 1).

Table 2. Single sample t test analysis results

column	min	max	avg	std	t	p
school_year	1	4	2.28	1.166614	54.23059	0.00
age	18	31	20.25	1.771329	316.1066	0.00
gender	0	1	0.48	0.499774	26.82165	0.00
bmi	0	54.55	23.41	4.587203	141.0619	0.00
who_bmi	0	6	3.51	1.221277	79.61311	0.00
phq_score	0	24	7.16	4.417521	44.84612	0.00
depression_severity	0	5	1.31	1.417481	25.64284	0.00
depressiveness	0	1	0.26	0.442904	16.72359	0.00
suicidal	0	1	0.08	0.278833	8.422758	0.00
depression_diagnosis	0	1	0.08	0.280769	8.49337	0.00
depression_treatment	0	1	0.07	0.262599	7.842734	0.00
gad_score	0	21	6.88	4.726979	40.24389	0.00
anxiety_severity	0	3	1.15	1.035236	30.92285	0.00
anxiousness	0	1	0.24	0.432065	15.88874	0.00
anxiety_diagnosis	0	1	0.08	0.270888	8.136268	0.00
anxiety_treatment	0	1	0.08	0.266788	7.990433	0.00
epworth_score	0	32	6.39	3.995584	44.24663	0.00
sleepiness	0	1	0.18	0.386672	13.0819	0.00

The one-sample t-test conducted against the null hypothesis value of 0.0 shows in Table 2 that all variables are significantly different ($p < 0.05$). Key findings include: Analysis reveals gender disparities in depression and anxiety levels, with females exhibiting higher scores on depression severity and anxiety severity scales compared to males. There is a significant correlation between age and depression/anxiety severity, with older individuals tending to report higher levels of both. Higher BMI correlates positively with depression and anxiety severity, suggesting a potential link between obesity and mental health disorders. Analysis shows variations in depression and anxiety scores across different school years, with higher scores observed in later years, suggesting potential academic stressors contributing to mental health issues. Presence of suicidal thoughts correlates strongly with depression severity, underscoring the critical importance of early intervention and support for individuals experiencing suicidal ideation.

3.2. Correlation Analysis

In this section, the author delves into a detailed correlation analysis to unravel the relationships between various factors associated with mental health, particularly focusing on depression and anxiety. By examining the correlations between different variables, this study aims to gain insights into the intricate relationship between demographic, psychological, and behavioral factors that contribute to these prevalent mental health disorders. Through this analysis, the author seeks to identify significant associations and patterns that can inform the understanding of depression and anxiety and guide targeted interventions and support strategies.

The correlation analysis in Table 3 shows that: Strong Positive Correlations: Anxiety severity is positively correlated with depression severity (0.317) and anxiousness (0.298), indicating that individuals with higher levels of anxiety tend to also experience higher levels of depression and anxiousness. Additionally, there is a moderate positive correlation between anxiety severity and suicidal thoughts (0.071), suggesting that higher levels of anxiety may be associated with an increased risk of suicidal ideation. There are weak positive correlations between anxiety severity and other variables such as depressiveness, BMI, GAD score, and anxiety diagnosis. While these correlations are statistically significant, their magnitudes are relatively small compared to the stronger correlations mentioned above. Some variables, such as gender, anxiety treatment, WHO BMI, Epworth score, depression diagnosis, and sleepiness, show negligible correlations with anxiety severity, indicating that these factors may not strongly influence anxiety levels in the dataset.

Table 3. Correlation to Depression(right) and Anxiety(left)

Variable	Anxiety_severity	Variable	Depression_severity
depression_severity	0.3170201	anxiety_severity	0.3170201
anxiousness	0.297938927	gender	0.151534657
suicidal	0.071183353	school_year	0.067489966
depressiveness	0.070062553	id	0.067063778
phq_score	0.02994278	suicidal	0.051339233
bmi	0.028170735	age	0.047865202
gad_score	0.025706904	depression_diagnosis	0.026958851
anxiety_diagnosis	0.01562305	depression_treatment	0.017710451
depression_treatment	0.014349967	anxiety_treatment	0.011797771
gender	0.013438527	depressiveness	0.009199563
anxiety_treatment	0.012627393	bmi	-0.00226605
who_bmi	0.008648392	anxiety_diagnosis	-0.004143047
epworth_score	0.006344008	who_bmi	-0.024731896
depression_diagnosis	-0.001975276	sleepiness	-0.086107524
sleepiness	-0.007000658	epworth_score	-0.107889932
age	-0.066408983	anxiousness	-0.108546581
id	-0.092611331	gad_score	-0.230333911
school_year	-0.098446706	phq_score	-0.257257383

3.3. Decision Tree Classification

The classification results on depression in Table 4 highlight the decent performance in distinguishing between depressive and non-depressive cases, albeit with some variations in precision, recall, and F1-score. For the "Not Depressive" class, the model achieved a precision of 0.85, indicating that when it predicts a case as non-depressive, it is correct approximately 85% of the time. The recall for this class is high at 0.92, meaning that the model captures 92% of the actual non-depressive cases. The F1-score, which balances precision and recall, is 0.89 for this class, suggesting a robust overall performance. Conversely, for the "Depressive" class, the precision is slightly lower at 0.71, indicating that when the model predicts a case as depressive, it is correct approximately 71% of the time. However, the recall for this class is 0.55, suggesting that the model only captures 55% of the actual depressive cases. Consequently, the F1-score for the "Depressive" class is 0.62, reflecting a moderate performance that could be improved.

The classification results on anxiousness in Table 5 indicate a moderate performance in distinguishing between depressive and non-depressive cases. The model demonstrates a relatively high precision of 0.85 for non-depressive cases, indicating that when it predicts a case as non-depressive, it is correct about 85% of the time. However, the precision for depressive cases is lower at 0.71, indicating a higher rate of false positives. The recall, or sensitivity, for non-depressive cases is high at 0.92, suggesting that the model effectively captures 92% of the actual non-depressive cases. However, the recall for depressive cases is lower at 0.55, indicating that the model misses approximately 45% of the actual depressive cases.

Table 4. Regression on Depression

	precision	recall	f1-score	support
Not Depressive	0.95	0.93	0.94	536
Depressive	0.81	0.87	0.84	191

Table 5. Regression on Anxiousness

	precision	recall	f1-score	support
Not Depressive	0.85	0.92	0.89	536
Depressive	0.71	0.55	0.62	191

4. Conclusion

This study investigates the intricate correlation between depression and anxiety, seeking to unravel their complex nature and consequences. Utilizing a robust dataset, the author conducted an extensive examination to evaluate the importance of various elements associated with these mental health ailments. Employing statistical methodologies like exploratory data analysis (EDA) and predictive modeling like decision tree model, the author delved into the predictors and underlying dynamics. The prediction results illustrate a high prediction accuracy and recall score on healthy and mentally unhealthy people, showing that the analysis framework used in this study is applicable.

Also, mental healthcare strategies should be customized to address the various causes of depression and anxiety, including personalized treatments and plans. Mental health campaigns could benefit from emphasizing the significance of lifestyle changes, particularly in physical activity, along with awareness of environmental risk factors. Further studies are recommended to explore the roles of non-significant factors and to understand the impact of cultural, environmental, and socioeconomic factors on depression and anxiety. Lastly, collaboration among healthcare professionals, policymakers, and researchers is crucial to develop effective, evidence-based strategies to combat the growing depression and anxiety epidemic among society.

References

- [1] Lecrubier Y. Prescribing patterns for depression and anxiety worldwide. *Journal of Clinical Psychiatry*, 2001, 62: 31-38.
- [2] Haroz E E, et al. How is depression experienced around the world? A systematic review of qualitative literature. *Social Science & Medicine*, 2017, 183: 151-162.
- [3] Boyd C P, et al. Prevalence of anxiety and depression in Australian adolescents: Comparisons with worldwide data. *The Journal of Genetic Psychology*, 2000, 161(4): 479-492.
- [4] Lecrubier Y, The burden of depression and anxiety in general medicine. *Journal of clinical psychiatry*, 2001, 62: 4-11.
- [5] Lépine J P, Epidemiology, burden, and disability in depression and anxiety. *Journal of Clinical Psychiatry*, 2001, 62: 4-12.
- [6] De Souza I M, Machado-de-Sousa J P. Brazil: world leader in anxiety and depression rates. *Brazilian Journal of Psychiatry*, 2017, 39: 384-384.
- [7] Luik A I, et al. Treating depression and anxiety with digital cognitive behavioural therapy for insomnia: a real world NHS evaluation using standardized outcome measures. *Behavioural and Cognitive Psychotherapy*, 2017, 45(1): 91-96.
- [8] Baghaei N, et al. Virtual reality for supporting the treatment of depression and anxiety: scoping review. *JMIR mental health*, 2021, 8(9).
- [9] Ali A M, et al. The depression anxiety stress scale 21: development and validation of the depression anxiety stress scale 8-item in psychiatric patients and the general public for easier mental health measurement in a post COVID-19 world. *International Journal of Environmental Research and Public Health*, 2021, 18(19): 10142.
- [10] Stevanovic D, et al. Cross-cultural measurement invariance of the Revised Child Anxiety and Depression Scale across 11 world-wide societies. *Epidemiology and Psychiatric sciences*, 2017, 26(4): 430-440.