



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

ESCUELA DE POSTGRADO

PROGRAMA DE MAESTRÍA EN DATA SCIENCE

***Proyecto:** Modelo Gráfico Probabilístico de Mental Health*

TRABAJO PARCIAL

***Curso:** Probabilistic Graphical Models*

DOCENTE

Patricia Reyes Silva

AUTOR(ES)

Neisser Alonzo Ale Ale

Nicolle Solange Pereira Bautista

Gabriela Edith Urquiza Ayala

Lima, junio del 2024

1. Descripción del Problema.....	3
2. Descripción del Conjunto de Datos.....	4
2.1. Origen de los Datos.....	4
2.2. Estructura de los Datos.....	4
2.3. Visualización de los Datos.....	5
2.3.1. Análisis Exploratorio de los Datos (EDA).....	5
3. Propuesta del Modelo.....	8
3.1. Objetivo del Modelo.....	8
3.2. Metodología del Modelo.....	8
3.2.1. Selección de variables para el Modelo.....	8
3.2.2. Representación del Modelo (DAG).....	9
3.2.3. Distribución de Probabilidad Condicional (CPD).....	10
3.2.4. Representación del Modelo Bayesiano.....	11
3.2.5. Comparación del Modelo Bayesiano con el Modelo de Aprendizaje Estructurado.....	11
3.3. Inferencias del Modelo Bayesiano.....	12
4. Validación de Resultados.....	13
5. Conclusiones.....	13
6. Referencias Bibliográficas.....	14
7. Anexos.....	15
7.1. Link del Repositorio en GitHub.....	15
7.2. Cuestionario.....	15

1. Descripción del Problema

Según la Organización Mundial de la Salud, la pandemia por el COVID-19 ha provocado un aumento del 25% en temas de ansiedad y depresión en todo el mundo. Este hecho ha llevado a los gobiernos a incluir la atención a la salud mental y dirigir planes de apoyo psicosocial en respuesta a la COVID-19.

Una de las posibles causas del aumento del estrés es por el aislamiento social debido a la pandemia. Adicionalmente, entre los factores de estrés causantes de ansiedad y depresión también se han mencionado la soledad, el miedo a contagiarse, sufrir y morir, o a que los seres queridos corran esa suerte, el dolor provocado por la pérdida de seres queridos y las preocupaciones económicas. Por otro lado, según Taylor (2020) también hay un miedo ante la peligrosidad de la COVID-19 la cual abarca preocupaciones como el miedo a entrar en contacto con objetos contaminados, preocupación por los costos socioeconómicos, temores xenófobos, síntomas de estrés traumático y conductas compulsivas.

Según el artículo publicado por Lancet (2021), señala que los jóvenes y las mujeres son los más afectados debido a la COVID-19, ya que, se subraya que hay mayor riesgo de que comportamientos autodestructivos afecten de manera desproporcionada a este grupo demográfico. También se señala que las mujeres se han visto mucho más afectadas que los hombres, ya que, las mujeres tienen más probabilidades de enfrentarse a desventajas financieras durante la pandemia, como salarios más bajos, menos ahorros y un empleo menos seguro en comparación con los hombres, lo que las hace más vulnerables a los impactos económicos de la COVID-19.

Por ello, en el dataset RHMCD-20 (2024) se centran en analizar los datos sobre la salud mental durante la pandemia de la COVID-19 en la población de Bangladesh. Este artículo contribuye proporcionando información valiosa sobre el análisis de datos sobre salud mental durante la pandemia de la COVID-19, y ofrece una comprensión exhaustiva de los factores que influyen en el bienestar.

La selección de este tema responde a la necesidad de entender los efectos que tuvo la pandemia en la salud mental. Creemos que analizar estos efectos puede ayudar a entender y por ende reducir los efectos que permanecen hoy.

2. Descripción del Conjunto de Datos

2.1. Origen de los Datos

Con 824 observaciones y 13 variables, el conjunto de datos RHMCD-20 (<https://data.mendeley.com/datasets/pxjmjyfdh2/2>) representa una valiosa herramienta para comprender los impactos psicológicos de la cuarentena por COVID-19 en Bangladesh.

Fue recopilado usando un enfoque de encuesta estructurada (Anexo 5.1) que se diseñó con preguntas sobre la salud mental.

Su objetivo principal fue medir las dificultades de salud mental que experimentaron las personas cuando estuvieron aisladas durante el periodo de cuarentena por COVID-19.

Esta encuesta abarcó una amplia gama de grupos de edad y antecedentes profesionales, se incluyeron adolescentes, estudiantes universitarios, amas de casa, profesionales corporativos, entre otros. Y las preguntas abordaban temas vinculados a salud mental, como cambios de comportamiento y cambio físicos, frustración e interacciones sociales.

2.2. Estructura de los Datos

Los datos se componen de la siguiente manera:

Nº	Variable	Tipo	Descripción
1	Age	STR	Representa la edad de los participantes (agrupado).
2	Gender	STR	Indica el género de los participantes.
3	Occupation	STR	Representa las ocupaciones de los participantes.
4	Days_Indoors	STR	Indica el número de días que el participante ha pasado sin salir de casa (agrupado).
5	Quarantine_Frustrations	STR	Frustraciones en las dos primeras semanas de cuarentena (Sí/Tal vez/No).
6	Changes_Habits	STR	Representa cambios importantes en los hábitos alimenticios y de sueño (Sí/Tal vez/No).
7	Mental_Health_History	STR	Precedente de trastornos mentales en la generación anterior (Sí/No).
8	Weight_Change	STR	Destaca los cambios en el peso corporal durante la cuarentena (Sí/Tal vez/No).
9	Mood_Swings	STR	Representa cambios extremos en el estado de ánimo (Bajo/Medio/Alto).

10	Coping_Struggles	STR	Incapacidad para lidiar con problemas diarios o estrés (Sí/Tal vez/No).
11	Work_Interest	STR	Representa si el participante está perdiendo interés en trabajar (Sí/No).
12	Social_Weakness	STR	Indica si el participante se siente mentalmente débil al interactuar con otros (Sí/No).
13	Growing_Stress	STR	Indica si el estrés del participante está aumentando día a día (Sí/No).

Tabla. 1 Diccionario de Datos del Dataset RHMCD-20

2.3. Visualización de los Datos

2.3.1. Análisis Exploratorio de los Datos (EDA)

Un vistazo a la composición del dataset se muestra a continuación:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 824 entries, 0 to 823
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   824 non-null    object
1   Gender                               824 non-null    object
2   Occupation                           824 non-null    object
3   Days_Indoors                         824 non-null    object
4   Growing_Stress                     824 non-null    object
5   Quarantine_Frustrations              824 non-null    object
6   Changes_Habits                       824 non-null    object
7   Mental_Health_History                824 non-null    object
8   Weight_Change                       824 non-null    object
9   Mood_Swings                         824 non-null    object
10  Coping_Struggles                     824 non-null    object
11  Work_Interest                        824 non-null    object
12  Social_Weakness                      824 non-null    object
dtypes: object(13)
memory usage: 83.8+ KB
```

Fig. 1 Estructura del conjunto de datos

Durante esta fase se llevó a cabo una revisión de las variables del conjunto de datos. Logramos observar que sus distribuciones se encontraban notablemente uniformes, lo que limitó la capacidad para discernir patrones significativos o identificar factores influyentes. (Anexo 5.2)

Asimismo, se realizó una matriz de correlación de variables y no se encontró una alta correlación entre la variable de estudio “Quarantine_Frustrations” con las demás.

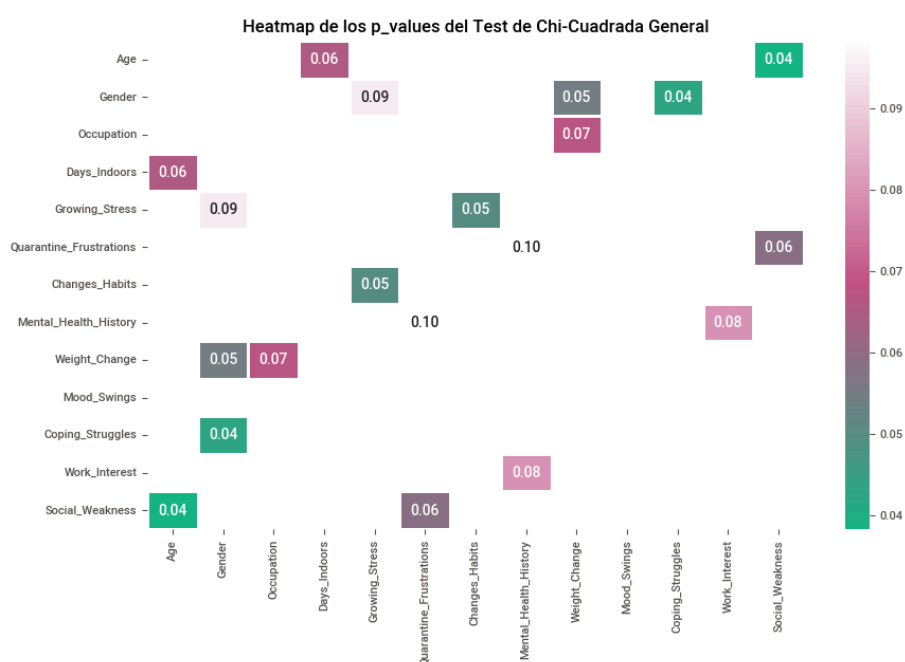


Fig. 2 Heatmap de los p_valores de la prueba de Chi Cuadrada

Esta uniformidad en las variables nos llevó a reconsiderar nuestro enfoque para realizar un análisis más efectivo, por lo que optamos por aplicar filtros que nos permitan explorar subconjuntos específicos de la población.

Por ese motivo hemos decidido filtrar el conjunto de datos para analizar únicamente información relacionada a mujeres.

Es importante reconocer que los factores socioculturales, económicos y de género pueden influir significativamente en la salud mental de las personas. Este enfoque nos permitirá profundizar en las experiencias y desafíos específicos que las mujeres pueden enfrentar en el contexto de la pandemia y su impacto en la salud mental.

Tras realizar este filtro, se realizó una nueva matriz de correlación donde a diferencia de la anterior, se ve una mayor correlación entre variables, lo que nos permitirá identificar relaciones más claras y significativas.

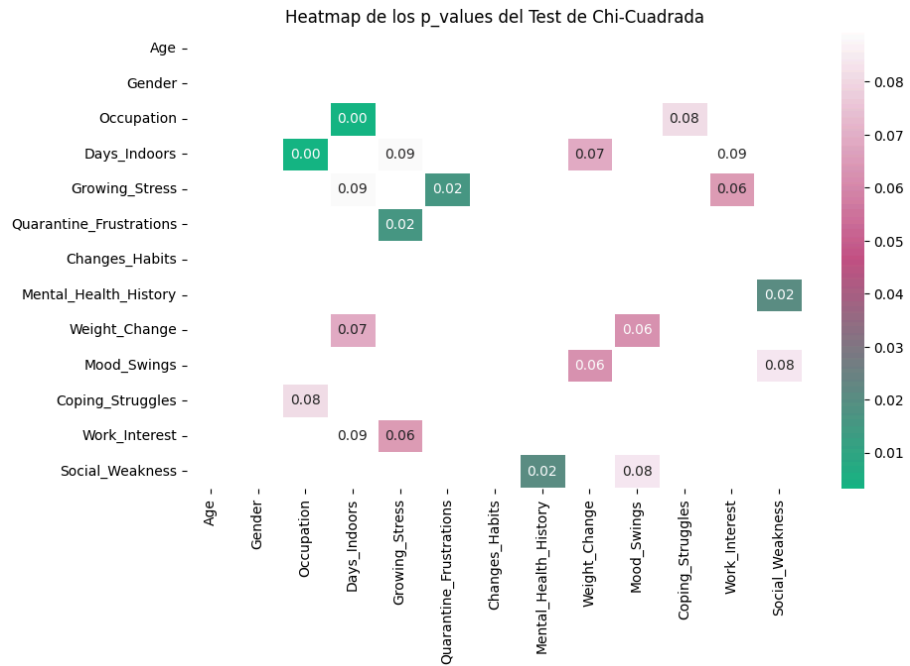


Fig. 3 Heatmap de los p_values de la prueba de Chi Cuadrada para la muestra de mujeres

También se realizó un análisis de las relaciones entre diversas variables y la variable "Quarantine Frustrations". Se han calculado y representado los porcentajes relativos en relación con la presencia o ausencia de frustración, lo que nos permite visualizar cómo varían estas distribuciones.

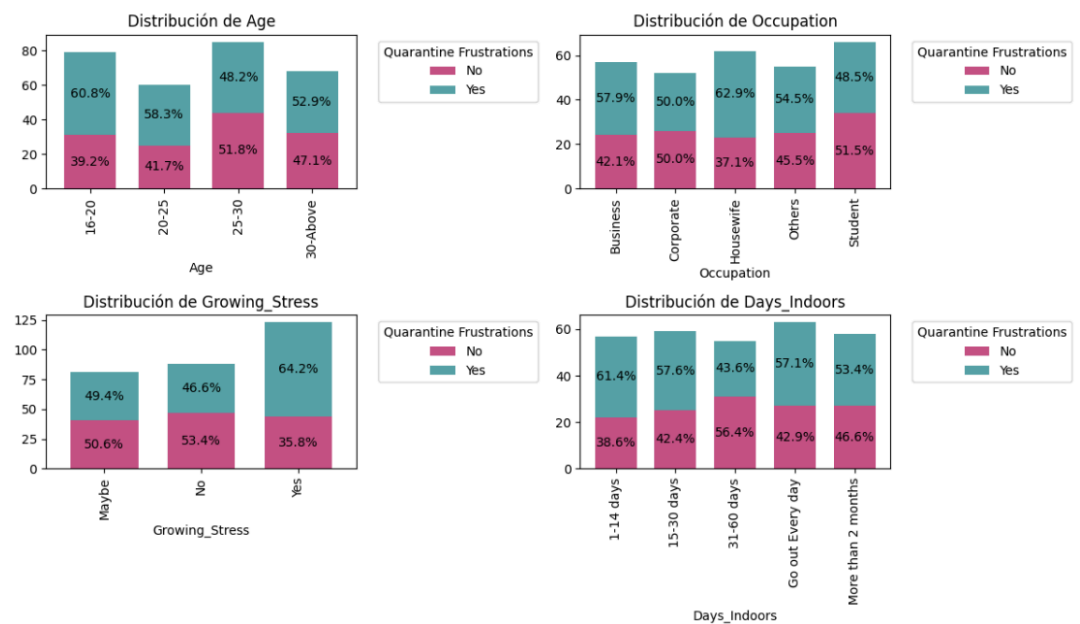


Fig. 4 Distribución de los datos según el "Quarantine Frustrations"

3. Propuesta del Modelo

3.1. Objetivo del Modelo

Desarrollar un modelo que permita predecir la probabilidad de si una persona es propensa o no es propensa al estrés ante una pandemia dado un conjunto de características.

- Analizar la distribución de los datos según el estrés
- Encontrar relaciones entre las variables del conjunto de datos
- Medir el rendimiento del modelo

3.2. Metodología del Modelo

En el presente trabajo, se usará Probabilistic Graphical Models (PGM) para representar distribuciones de probabilidad del conjunto de datos RHMCD-20 de acuerdo a las relaciones de las variables.

Con esta representación de PGM se obtendrá la probabilidad de si una nueva persona es propensa o no es propensa a un problema de estrés ante una pandemia.

3.2.1. Selección de variables para el Modelo

De acuerdo al análisis EDA y a la prueba de Chi-Cuadrada para hallar relaciones entre las variables se determinó usar las siguientes variables que guardan relación entre sí.

	count	unique	top	freq
Age	211	4	25-30	63
Gender	211	1	Female	211
Occupation	211	5	Student	49
Days_Indoors	211	5	1-14 days	48
Growing_Stress	211	2	Yes	123
Quarantine_Frustrations	211	2	Yes	120
Changes_Habits	211	3	Maybe	78
Mental_Health_History	211	3	Maybe	83
Weight_Change	211	3	No	77
Mood_Swings	211	3	High	79
Social_Weakness	211	3	Maybe	73

Fig. 5 Selección de variables finales para el análisis

3.2.2. Representación del Modelo (DAG)

A continuación, se muestra Grafo Acíclico Dirigido (DAG) del modelo de Red Bayesiana.

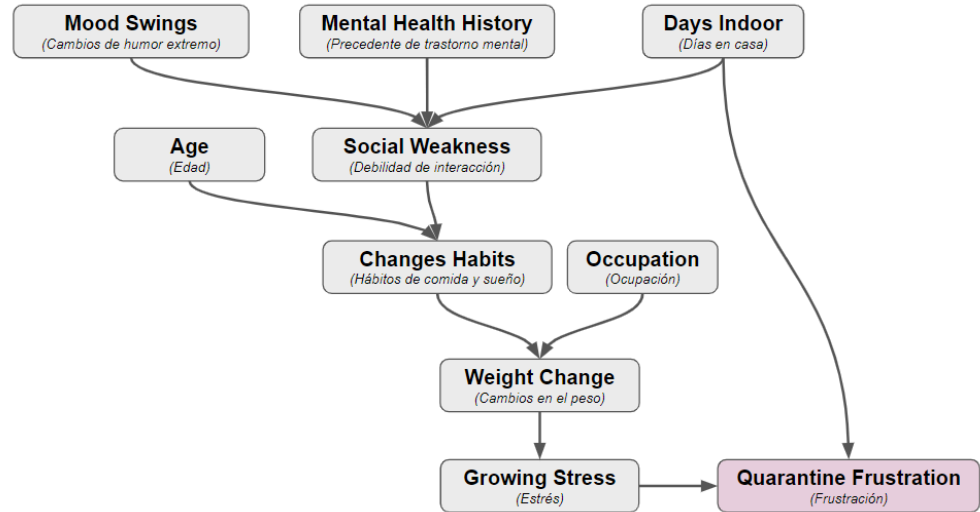


Fig. 6 DAG del Modelo PGM

3.2.3. Distribución de Probabilidad Condicional (CPD)

Para desarrollar el modelo PGM es necesario construir las Tablas de Distribución de Probabilidad Condicional (CPD) a partir del conjunto de datos.

A continuación, se detallan los CPDs para algunas variables del Modelo Bayesiano.

Weight_Change	Weight_Change(0)	Weight_Change(1)	Weight_Change(2)
Growing_Stress(0)	0.4411764705882353	0.45454545454545453	0.3484848484848485
Growing_Stress(1)	0.5588235294117647	0.5454545454545454	0.6515151515151515

Days_Indoors	...	Days_Indoors(4)	Days_Indoors(4)
Growing_Stress	...	Growing_Stress(0)	Growing_Stress(1)
Quarantine_Frustrations(0)	...	0.7857142857142857	0.2916666666666667
Quarantine_Frustrations(1)	...	0.21428571428571427	0.7083333333333334

Fig. 7 CPDs de las variables del Modelo Bayesiano

3.2.4. Representación del Modelo Bayesiano

A continuación se detalla muestra el modelo Bayesiano final:

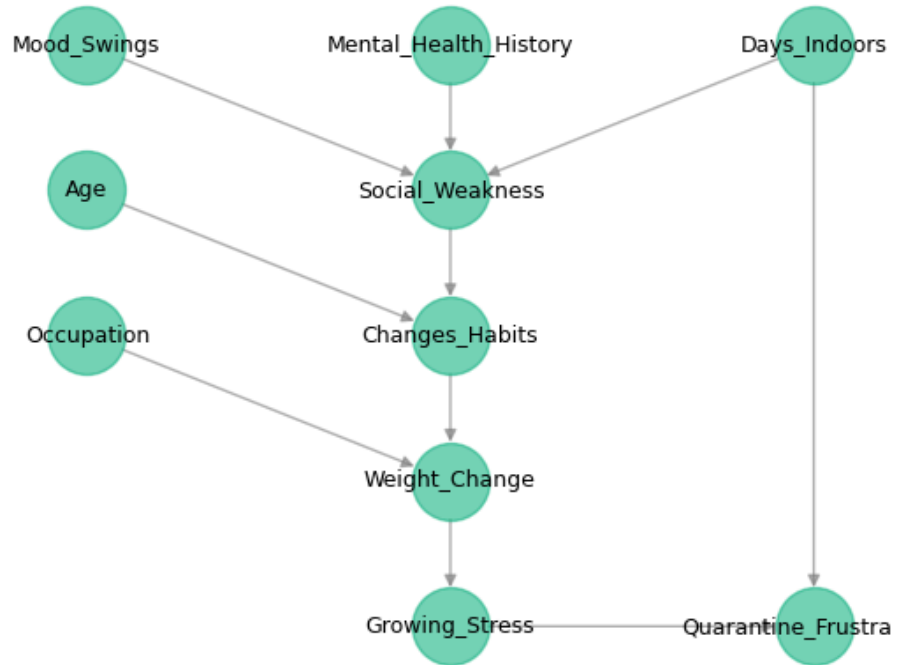


Fig. 8 Modelo Bayesiano para el Mental Health a partir del DAG propuesto

3.2.5. Comparación del Modelo Bayesiano con el Modelo de Aprendizaje Estructurado

Se construyó un modelo TreeSearch para que aprenda de la estructura de la data y encuentre el DAG a partir de los datos. Seguidamente, se calcularon los CPDs mediante un estimador de MaximumLikelihood. A continuación, se muestra el Modelo de Red Bayesiana a partir del DAG aprendido por el modelo.

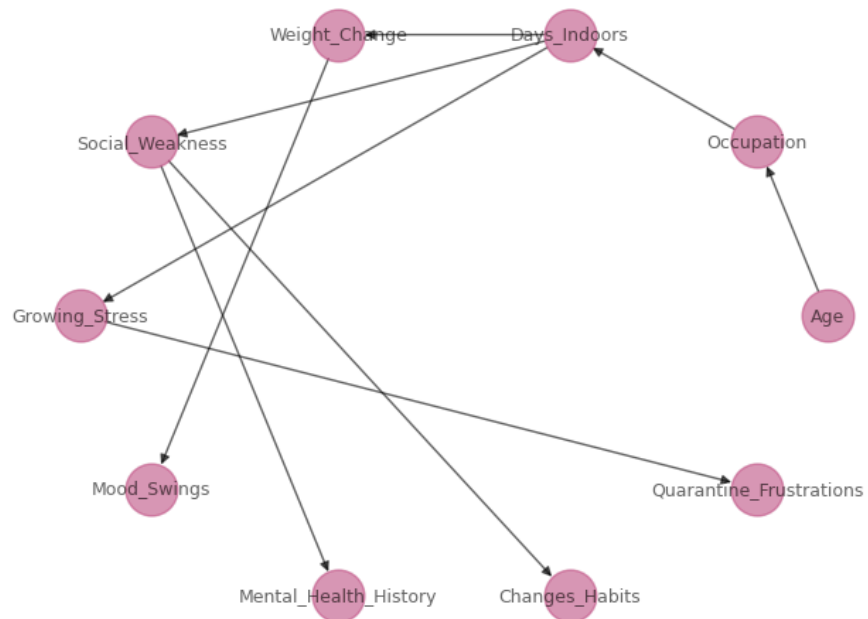


Fig. 9 Modelo Bayesiano para el Mental Health a partir del DAG del Aprendizaje Estructurado

Finalmente, se obtuvieron los accuracies para los dos modelos los cuales se muestran en la siguiente tabla.

Modelo	Accuracy
Modelo Bayesiano para el Mental Health a partir del DAG propuesto	63.51 %
Modelo Bayesiano para el Mental Health a partir del DAG del Aprendizaje Estructurado	59.72 %

Tabla. 2 Comparación del Accuracy del Modelo Bayesiano Propuesto y del Modelo Bayesiano por Aprendizaje Estructurado

3.3. Inferencias del Modelo Bayesiano

A continuación, se detalla una tabla con las potenciales preguntas que el modelo PGM podría responder:

Nº	Pregunta
1.	¿Cuál es la probabilidad de SÍ experimentar cambios en el peso corporal dado que una persona SI ha reportado frustración durante la cuarentena?

2.	¿Cuál es la probabilidad de SÍ tener frustración dado que la persona ha estado MÁS DE 2 MESES en casa, SI ha experimentado cambios de peso y SI tiene estrés.
3.	¿Cuál es la probabilidad de que una ama de casa experimente estrés creciente si ha reportado cambio de hábitos alimenticios y de sueño?

Tabla. 3 Listado de preguntas potenciales para ser respondidas por el modelo PGM

4. Validación de Resultados

Pregunta 1: ¿Cuál es la probabilidad de experimentar cambios en el peso corporal dado que una persona ha reportado frustración durante la cuarentena?

Entradas:

Quarantine_Frustrations = Sí

Salidas:

Modelo Propuesto:

Weight_Change(0): 32.44%

Weight_Change(1): 37.16%

Weight_Change(2): 30.40%

Modelo de Aprendizaje Estructurado:

Weight_Change(0): 32.20%

Weight_Change(1): 36.55%

Weight_Change(2): 31.26%

Interpretación:

Modelo Propuesto: Según el modelo propuesto, la probabilidad de que una persona experimente cambios en el peso corporal dado que ha reportado frustración durante la cuarentena es de 32.44% para no tener cambios, 37.16% para tener un cambio leve, y 30.40% para tener un cambio significativo.

Modelo de Aprendizaje Estructurado: Según el modelo de aprendizaje estructurado, la probabilidad de que una persona experimente cambios en el peso corporal dado que ha reportado frustración durante la cuarentena es de 32.20% para no tener cambios, 36.55% para tener un cambio leve, y 31.26% para tener un cambio significativo.

Pregunta 2: ¿Cuál es la probabilidad de SÍ tener frustración dado que la persona ha estado MÁS DE 2 MESES en casa, SI ha experimentado cambios de peso y SI tiene estrés?

Entradas:

Days_Indoors = Más de 2 meses

Weight_Change = Sí

Growing_Stress = Sí

Salidas:

Modelo Propuesto:

Quarantine_Frustrations(0): 78.57%

Quarantine_Frustrations(1): 21.43%

Modelo de Aprendizaje Estructurado:

Quarantine_Frustrations(0): 53.41%

Quarantine_Frustrations(1): 46.59%

Interpretación:

Modelo Propuesto: Según el modelo propuesto, la probabilidad de que una persona experimente frustración durante la cuarentena, dado que ha estado más de dos meses en casa, ha experimentado cambios de peso y tiene estrés, es de 78.57%.

Modelo de Aprendizaje Estructurado: Según el modelo de aprendizaje estructurado, la probabilidad de que una persona experimente frustración durante la cuarentena, dado que ha estado más de dos meses en casa, ha experimentado cambios de peso y tiene estrés, es de 53.41%.

Pregunta 3: ¿Cuál es la probabilidad de que una ama de casa experimente estrés creciente si ha reportado cambio de hábitos alimenticios y de sueño?

Entradas:

Changes_Habits = Sí

Salidas:

Modelo Propuesto:

Occupation(0): 19.91%

Occupation(1): 17.06%

Occupation(2): 18.96%

Occupation(3): 20.85%

Occupation(4): 23.22%

Modelo de Aprendizaje Estructurado:

Occupation(0): 19.87%

Occupation(1): 17.07%

Occupation(2): 18.98%

Occupation(3): 20.89%

Occupation(4): 23.19%

Interpretación:

Modelo Propuesto: Según el modelo propuesto, la probabilidad de que una ama de casa experimente estrés creciente si ha reportado cambio de hábitos alimenticios y de sueño es de 18.96%.

Modelo de Aprendizaje Estructurado: Según el modelo de aprendizaje estructurado, la probabilidad de que una ama de casa experimente estrés creciente si ha reportado cambio de hábitos alimenticios y de sueño es de 18.98%.

5. Conclusiones

El estudio confirmó la efectividad de los Modelos Gráficos Probabilísticos (PGM) para analizar la salud mental durante la pandemia de COVID-19 utilizando el dataset RHMCD-20. Se implementaron técnicas de construcción de Redes Bayesianas e inferencia probabilística para modelar las relaciones entre variables clave. El modelo Bayesiano propuesto, basado en un Grafo Acíclico Dirigido (DAG), alcanzó una precisión de 63.51 %, superando al modelo de aprendizaje estructurado que tuvo una precisión de 59.72 %. La metodología empleada resalta la importancia de una correcta selección de variables y sus interrelaciones.

La validación de resultados mostró que las variables `Quarantine_Frustrations`, `Days_Indoors` y `Changes_Habits` tienen una influencia significativa en la salud mental de los individuos durante la cuarentena. Por ejemplo, la probabilidad de experimentar frustración durante la cuarentena es significativamente mayor en el modelo propuesto (78.57%) en comparación con el modelo de aprendizaje estructurado (53.41%). Además, ambos modelos mostraron consistencia en la predicción de ciertos eventos como cambios en el peso corporal y estrés creciente en amas de casa, lo que refuerza la robustez de las inferencias realizadas.

Para futuras investigaciones, se recomienda integrar más datos y aplicar técnicas avanzadas de aprendizaje automático para mejorar la precisión y la generalización del modelo. Es esencial explorar diferentes estructuras de grafos y métodos de validación cruzada para garantizar la solidez de los resultados. Los hallazgos obtenidos pueden guiar el diseño de intervenciones específicas y políticas de salud pública para mitigar los efectos negativos de la cuarentena en la salud mental, proporcionando una base sólida para desarrollar estrategias de apoyo más efectivas.

6. Referencias Bibliográficas

- Amin N, Salehin I, Baten MA, Noman RA. RHMCD-20 dataset: Identify rapid human mental health depression during quarantine life using machine learning. *Data Brief*. 2024 Apr 5;54:110376. doi: 10.1016/j.dib.2024.110376. PMID: 38623549; PMCID: PMC11016953.
- Salehin, Imrus; Amin, Nazrul (2024), "The RHMCD-20 datasets for Depression and Mental Health Data Analysis with Machine Learning ", Mendeley Data, V2, doi: 10.17632/pxjmjyfdh2.2
- World Health Organization. (2022). *Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief*, 2 March 2022. World Health Organization. <https://iris.who.int/handle/10665/352189>. Licencia: CC BY-NC-SA 3.0 IGO
- Steven, Taylor., Caeleigh, A., Landry., Michelle, M., Paluszek., Thomas, A., Fergus., Dean, McKay., Gordon, J.G., Asmundson. (2020). COVID stress syndrome: Concept, structure, and correlates.. *Depression and Anxiety*, 37(8):706-714. doi: 10.1002/DA.23071
- COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. 2021 Nov 6;398(10312):1700-1712. doi: 10.1016/S0140-6736(21)02143-7. Epub 2021 Oct 8. PMID: 34634250; PMCID: PMC8500697.
- Wenham, C., Smith, J., Davies, S. E., Feng, H., Grépin, K. A., Harman, S., ... & Morgan, R. (2020). Women are most affected by pandemics—lessons from past outbreaks. *Nature*, 583(7815), 194-198.

7. Anexos

7.1. Link del Repositorio en GitHub

A continuación, se anexa el enlace del repositorio GitHub:

[neisserale/PGM-Mental-Health \(github.com\)](https://github.com/neisserale/PGM-Mental-Health)

7.2. Cuestionario

Introduction:

1. Welcome to the Mental Health During Quarantine Survey.
2. Your participation is voluntary, and your responses will be kept confidential.
3. Please provide the following demographic information:
 - a. Age: _____
 - b. Gender: Male / Female / Other
 - c. Occupation: _____

Quarantine Experience:

4. How many days have you been indoors during quarantine? _____
5. Have you experienced major changes in your eating habits and sleeping patterns?
 - a. No
 - b. Maybe
 - c. Yes
6. Did you face frustrations during the first two weeks of quarantine?
 - a. No
 - b. Maybe
 - c. Yes

Mental Health

7. How have you been coping with the challenges of quarantine? _____
8. Is your stress increasing day by day?
 - a. No
 - b. Maybe
 - c. Yes
9. Do you have a history of mental health issues?
 - a. No
 - b. Yes
10. Have you experienced extreme mood swings during quarantine?
 - a. Low
 - b. Medium
 - c. High
11. Are you struggling to cope with daily problems or stress?
 - a. No
 - b. Maybe
 - c. Yes
12. Are you losing interest in working?
 - a. No
 - b. Yes
13. Do you feel mentally weak when interacting with others?
 - a. No
 - b. Yes