



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Sistema de pronóstico usando machine learning para detección de la  
depresión en personas adultas del Condominio Ciudad Verde

**TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:**

Ingeniero de Sistemas

**AUTORES:**

Barriga Moreno, Elisbeth (orcid.org/0009-0002-0937-1732)

Muñoz Avalos, Flavio Stephano (orcid.org/0000-0001-5867-4858)

**ASESOR:**

Mgr. Quiñones Nieto, Yamil Alexander (orcid.org/0000-0003-4474-0556)

**LÍNEA DE INVESTIGACIÓN:**

Sistemas de Información y Comunicaciones

**LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:**

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2024



**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Declaratoria de Autenticidad del Asesor**

Yo, QUIÑONES NIETO YAMIL ALEXANDER, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: Sistema de Pronóstico usando Machine Learning para detección de la depresión en personas adultas del Condominio Ciudad Verde, cuyos autores son BARRIGA MORENO ELISBETH, MUÑOZ AVALOS FLAVIO STEPHANO, constato que la investigación tiene un índice de similitud de 19.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 18 de Julio del 2024

Apellidos y Nombres del Asesor:	Firma
QUIÑONES NIETO YAMIL ALEXANDER DNI: 42863390 ORCID: 0000-0003-4474-0556	Firmado electrónicamente por: YQUINONES el 18- 07-2024 17:50:12

Código documento Trilce: TRI - 0821232





**UNIVERSIDAD CÉSAR VALLEJO**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Declaratoria de Originalidad de los Autores**

Nosotros, BARRIGA MORENO ELISBETH, MUÑOZ AVALOS FLAVIO STEPHANO estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: Sistema de Pronóstico usando Machine Learning para detección de la depresión en personas adultas del Condominio Ciudad Verde, es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

<b>Nombres y Apellidos</b>	<b>Firma</b>
BARRIGA MORENO ELISBETH DNI: 46168342 ORCID: 0009-0002-0937-1732	Firmado electrónicamente por: EBARRIGA el 10-08-2024 19:51:28
MUÑOZ AVALOS FLAVIO STEPHANO DNI: 75162098 ORCID: 0000-0001-5867-4858	Firmado electrónicamente por: FMUNOZAV el 09-08-2024 19:11:32

Código documento Trilce: INV - 1691483



## Dedicatoria

De: Barriga Moreno Elisbeth

A nuestros padres. Este logro es un testimonio de su inmenso amor y dedicación. Valoro mucho las lecciones de vida que me han impartido y por el cariño que siempre me han brindado. Mi gratitud hacia ustedes es imposible de expresar completamente. Esta tesis es un tributo a su legado y a la eterna admiración que sentimos por ustedes. Gracias por ser los mejores padres del mundo.

De: Muñoz Avalos Flavio

A mi madre por ser el pilar que forjo mi futuro y su apoyo, ser la fuente de todos mis logros. Tu dedicación y sacrificio fue la inspiración y motivación en mi vida. A mi padre por su paciencia y apoyo incondicional. A mi abuelo Alberto, por su cariño y apoyo incondicional. A mis amigos siempre apoyándome en los momentos donde tuve dificultades y logré sobrellevarlos.

## Agradecimiento

A nuestro tutor, el Mgtr. Yamil A. Quiñones Nieto, extendemos nuestro sincero agradecimiento por su inestimable orientación y apoyo constante mientras fue nuestro tutor de tesis. Sus perspicaces comentarios y su incansable búsqueda de la excelencia han sido un faro de luz en el proceso de investigación. Ha sido un honor y un privilegio aprender bajo su tutela.

## Índice de contenidos

CARÁTULA	i
DECLARATORIA DE AUTENTICIDAD DEL ASESOR .....	II
DECLARATORIA DE ORIGINALIDAD DEL AUTOR(ES) .....	III
DEDICATORIA.....	IV
AGRADECIMIENTO .....	V
ÍNDICE DE CONTENIDOS .....	VI
ÍNDICE DE TABLAS.....	VII
ÍNDICE DE FIGURAS .....	VIII
RESUMEN.....	X
ABSTRACT .....	XI
I. INTRODUCCIÓN .....	1
II. METODOLOGÍA.....	12
III. RESULTADOS.....	24
IV. DISCUSIÓN .....	47
V. CONCLUSIONES .....	52
VI. RECOMENDACIONES .....	54
REFERENCIAS.....	57
ANEXOS .....	62

## Índice de Tablas

TABLA N°1: MÉTODOS MUESTRALES EN FUNCIÓN A TIPO DE MUESTREO .....	19
TABLA N°2: VALORES DE LAS CATEGORÍAS DE EDADES .....	25
TABLA N°3: VALORES DE CATEGORÍAS DE GÉNERO .....	26
TABLA N°4: AGITACIÓN .....	31
TABLA N°5: PENSAMIENTO O DESEOS SUICIDAS.....	31
TABLA N°6: AUTOCRITICA.....	32
TABLA N°7: DIFICULTAD DE CONCENTRACIÓN.....	33
TABLA N°8: SE MUESTRA ALGUNOS DATOS DE ENTRENAMIENTO (TRAIN) PARA LAS EJES "X" Y "Y" .....	34
TABLA N°9: SE MUESTRA ALGUNOS DATOS DE TEST (TEST) PARA LOS EJES "X" Y "Y" .....	35
TABLA N°10: MÉTRICAS DE ERRORES DE LAS (Y_TEST, PREDICCIONES).....	36
TABLA N°11: SE MUESTRA ALGUNOS DATOS DE ENTRENAMIENTO (TRAIN) PARA LOS EJES "X" Y "Y" .....	37
TABLA N°12: SE MUESTRA ALGUNOS DATOS DE TEST (TEST) PARA LOS EJES "X" Y "Y" .....	37
TABLA N°13: MÉTRICAS DE ERRORES DE LAS (Y_TEST, PREDICCIONES).....	38
TABLA N°14: SE MUESTRA ALGUNOS DATOS DE ENTRENAMIENTO (TRAIN) PARA LOS EJES "X" Y "Y" .....	39
TABLA N°15: SE MUESTRA ALGUNOS DATOS DE TEST (TEST) PARA LOS EJES "X" Y "Y" .....	39
TABLA N°16: MÉTRICAS DE ERRORES DE LAS (Y_TEST, PREDICCIONES).....	40
TABLA N°17: SE MUESTRA ALGUNOS DATOS DE ENTRENAMIENTO (TRAIN) PARA LOS EJES "X" Y "Y" .....	41
TABLA N°18: SE MUESTRA ALGUNOS DATOS DE TEST (TEST) PARA LOS EJES "X" Y "Y" .....	41
TABLA N°19: MÉTRICAS DE ERRORES DE LAS (Y_TEST, PREDICCIONES).....	42
TABLA N°20: CANTIDAD DE PERSONAS QUE NO TIENEN DEPRESIÓN Y QUE PRESENTAN SÍNTOMAS DE DEPRESIÓN .....	43

## Índice de figuras

FIGURA N°1: DISEÑO DE ESTUDIO DE CASO CON UNA SOLA MEDICIÓN .....	13
FIGURA N°2: FÓRMULA DE PRECISIÓN.....	14
FIGURA N°3: FÓRMULA DE TASA DE ERROR .....	15
FIGURA N°4: FÓRMULA DEL ÍNDICE DE GINI .....	15
FIGURA N°5: FÓRMULA RMSE .....	15
FIGURA N°6: FÓRMULA MAE.....	15
FIGURA N°7: MAPA DE PROCEDIMIENTO DEL PROYECTO DE INVESTIGACIÓN.....	21
FIGURA N°8: GRÁFICO DE LOS NÚMEROS DE REGISTROS (EJE “Y”) Y LAS COLUMNAS (EJE “X”) .....	25
FIGURA N°9: GRÁFICO DE LAS CATEGORÍAS DE EDADES (EJE “X”).....	26
FIGURA N°10: GRÁFICO DE CATEGORÍA DE GÉNERO (EJE “X”) .....	26
FIGURA N°11: GRÁFICO DE CATEGORÍA DE EDAD (EJE “X”) CON RESPECTO AL GÉNERO .....	27
FIGURA N°12: GRÁFICO DE CATEGORÍA DE GÉNERO (EJE “X”) CON RESPECTO AL NIVEL DE EDUCACIÓN .....	27
FIGURA N°13: GRÁFICO DE CATEGORÍA DE GÉNERO (EJE “X”) CON RESPECTO AL ESTADO CIVIL .....	28
FIGURA N°14: GRÁFICO DE INDICADOR “AGITACIÓN” (EJE “X”) CON RESPECTO AL GÉNERO .....	29
FIGURA N°15: GRÁFICO DE INDICADOR “PENSAMIENTO O DESEO SUICIDA” (EJE “X”) RESPECTO AL GÉNERO ....	29
FIGURA N°16: GRÁFICO DE INDICADOR “AUTOCRITICA” (EJE “X”) CON RESPECTO AL GÉNERO .....	30
FIGURA N°17: GRÁFICO DE INDICADOR “DIFICULTAD DE CONCENTRACIÓN” (EJE “X”) RESPECTO AL GÉNERO ...	30
FIGURA N°18: GRÁFICO DE INDICADOR “AGITACIÓN” (EJE “X”) .....	31
FIGURA N°19: GRÁFICO DE INDICADOR “PENSAMIENTOS O DESEOS SUICIDAS” (EJE “X”) .....	32
FIGURA N°20: GRÁFICO DE INDICADOR “AUTOCRITICA” (EJE “X”) .....	33
FIGURA N°21: GRÁFICO DE INDICADOR “DIFICULTAD DE CONCENTRACIÓN” (EJE “X”) .....	34
FIGURA N°22: ARRAY DE LAS PREDICCIONES QUE SE OBTUVIERON DE LAS X_TEST .....	35
FIGURA N°23: GRÁFICO DE REGRESIÓN LINEAL CON LA (Y_TEST, PREDICCIONES).....	36
FIGURA N°24: ARRAY DE LAS PREDICCIONES QUE SE OBTUVIERON DE LAS X_TEST .....	37
FIGURA N°25: GRÁFICO DE REGRESIÓN LINEAL CON LA (Y_TEST, PREDICCIONES).....	38
FIGURA N°26: ARRAY DE LAS PREDICCIONES QUE SE OBTUVIERON DE LAS X_TEST .....	39
FIGURA N°27: GRÁFICO DE REGRESIÓN LINEAL CON LA (Y_TEST, PREDICCIONES).....	40
FIGURA N°28: ARRAY DE LAS PREDICCIONES QUE SE OBTUVIERON DE LAS X_TEST .....	41
FIGURA N°29: GRÁFICO DE REGRESIÓN LINEAL CON LA (Y_TEST, PREDICCIONES).....	42
FIGURA N°30: PROPORCIÓN DE LA DETECCIÓN DE DEPRESIÓN EN MUJERES Y HOMBRES.....	43
FIGURA N°31: ÁRBOL DE DECISIÓN PARA LA DETECCIÓN DE LA DEPRESIÓN EN MUJERES Y HOMBRES.....	43
FIGURA N°32: DISTRIBUCIÓN DE CONJUNTO DE DATOS (ENTRENAMIENTO - VALIDACIÓN - PRUEBA) .....	44



FIGURA N°33: CORRELACIONES ENTRE CARACTERÍSTICAS SELECCIONADAS Y DEPRESIÓN.....	45
FIGURA N°34: CURVA ROC – TASA DE FALSOS POSITIVOS .....	45
FIGURA N°35: MATRIZ DE CONFUSIÓN DE VALORES DE PREDICCIÓN .....	46

## Resumen

El presente trabajo de investigación tiene como propósito el desarrollo de un sistema de pronóstico utilizando técnicas de Machine Learning para la detección de la depresión en personas adultas del Condominio Ciudad Verde. La problemática planteada surge de la baja atención a problemas de salud mental y los incidentes relacionados en la zona. Por ello, se propone la creación de un sistema de pronóstico basado en Machine Learning, que puede ser de gran ayuda para abordar estos problemas proporcionando diagnósticos no clínicos, pero altamente relevantes para el seguimiento oportuno de las personas involucradas.

Utilizando la técnica de árboles de decisión de Machine Learning sobre otras técnicas, se obtuvo una precisión del 97%. Esto permitió determinar que la implementación del sistema de pronóstico para la detección de la depresión ha proporcionado predicciones con una alta tasa de precisión, demostrando ser una herramienta eficaz y confiable.

**Palabras clave:** Sistema de pronóstico, machine learning, técnicas, detección de la depresión, regresión lineal, árbol de decisión.

## **Abstract**

The purpose of this research work is to develop a prognostic system using Machine Learning techniques for the detection of depression in adults from the Ciudad Verde Residential Condominium. The problem raised arises from the low attention to mental health problems and related incidents in the area. Therefore, the creation of a prognostic system based on Machine Learning is proposed, which can be of great help in addressing these problems by providing non-clinical diagnoses, but highly relevant for the timely follow-up of the people involved.

Using the Machine Learning decision tree technique over other techniques, an accuracy of 97% was obtained. This allowed us to determine that the implementation of the prognostic system for the detection of depression has provided predictions with a high rate of accuracy, proving to be an effective and reliable tool.

**Keywords:** Forecasting system, machine learning, techniques, depression detection, linear regression, decision tree.

## **I. INTRODUCCIÓN**

Nuestra investigación se centra en la depresión en personas adultas. A nivel internacional en el último antecedente reportado por la Organización Mundial de la Salud (2023), investigan que la depresión o también lo llaman trastorno depresivo, se caracteriza por ser un trastorno mental que se asocia al estado de ánimo, y esto conlleva a la devastación de placer o interés de actividades durante largo periodos pueden ser meses o años; también afirma en su investigación que, entre las mujeres y hombres, las mujeres son más vulnerables a tener depresión. La OMS según sus estudios realizados informa que la estimación del 3,8% de población en general experimenta depresión eso es aproximadamente 280 millones de personas, incluye el 5% adultos los cuales el porcentaje estimado es del 4% hombres y 6% mujeres; también el 5,7% de personas adultas mayores de 60 años sufren de depresión. Anualmente se suicidan más de 700 000 personas.

En el Perú la información reportada por el MINSA que para el 2022 tuvieron reportados 272,432 casos de depresión; la información que se brinda a la población es precaria y limitante, existen personas que no cuentan con la atención médica adecuada y no reciben el tratamiento para el cuadro de depresión que padecen; estos trastornos mentales en los últimos años se han encontrado en aumento por los acontecimientos en el cual se vive en el territorio patrio desde lo social hasta lo económico.

Nuestra investigación se enfocó en los acontecimientos que se han ido suscitando en el Condominio Ciudad Verde, donde las personas que habitan por diversos factores en su vida privada han llegado a tomar la decisión de aventarse al vacío y quererse suicidar, lo cual el trastorno mental está asociado a factores de la vida cotidiana con eventos de estrés. Es por eso que nuestra investigación se interesa en la elaboración de un Sistema de pronóstico usando técnica de Machine Learning para la detección de la depresión en personas adultas del Condominio Ciudad Verde.

Debido a la problemática en cuestión que estamos presentando, tenemos el alcance siguiente:

Problema general:

¿De qué manera influye un Sistema de pronóstico usando técnica de Machine Learning en personas adultas para la detección de la depresión?

Problema específico:

1. ¿De qué manera influye un Sistema de pronóstico usando técnica de machine learning para la detección de la depresión en el área afectiva de las personas adultas del Condominio Ciudad Verde?
2. ¿De qué manera influye un Sistema de pronóstico usando técnica de machine learning para la detección de la depresión en el área conductual de las personas adultas del Condominio Ciudad Verde?

Para Justificar nuestro proyecto de investigación evaluamos los siguientes criterios:

Nuestra investigación resulta ser conveniente, porque en la actualidad tanto internacional y nacional, tienen el deber de respetar la integridad y los lineamientos de la responsabilidad social. Por lo tanto, este desarrollo de investigación a nivel nacional aportará para detectar el trastorno de depresión, y logren tomar conciencia sobre la importancia de respetar a las personas, trabajadores y a la sociedad en unidad.

Para la relevancia social, cabe mencionar que este desarrollo de investigación tiene como principales beneficiarios a las personas adultas, trabajadores y a la sociedad del Condominio Ciudad Verde. Como efecto se obtendrá datos para la detección de la depresión, para que finalmente se establezca la calidad de vida de los residentes del Condominio Ciudad Verde e incluso se puede llegar a convertir en estímulo para que asistan a un centro médico.

Para la implicancia práctica, tenemos que con el desarrollo de esta investigación detectaremos la depresión en personas adultas del Condominio Ciudad Verde, nuestro criterio es saber la convivencia estable de las personas que residen dentro del condominio, teniendo en cuenta que en ocasiones han querido lograr suicidarse.

Para el valor teórico, esta investigación aporta a desarrollar un sistema de pronóstico usando técnica de machine learning para la detección de la depresión, esto conlleva el diagnóstico de la depresión en personas adultas y también al estudio de la situación real de la vivienda, para que luego prospere la calidad de vida de las personas y sociedad.

De manera tecnológico, nuestra investigación nos permite desarrollar un sistema de pronóstico usando técnica de machine learning, el cual nos servirá de aporte y apoyo

para detectar la depresión en personas adultas, beneficiando a la sociedad.

Por lo expuesto, tenemos lo siguiente:

Objetivo general:

De qué manera influye un Sistema de pronóstico usando técnica de Machine learning para la detección de la depresión en personas adultas del Condominio Ciudad Verde.

Objetivos específicos:

1. Determinar de qué manera influye un sistema de pronóstico usando técnica de machine learning para la detección de la depresión en el área afectiva de las personas adultas del Condominio Ciudad Verde.
2. Determinar de qué manera influye un sistema de pronóstico usando técnica de machine learning para la detección de la depresión en el área conductual de las personas adultas del Condominio Ciudad Verde.

Nuestros objetivos nos permiten plasmar las siguientes hipótesis:

Hipótesis general:

El Sistema de pronóstico usando técnica de Machine Learning permite detectar la depresión en personas adultas del Condominio Ciudad Verde.

Hipótesis específicas:

1. Un sistema de pronóstico usando técnica de machine learning permite detectar la depresión en el área afectiva de las personas adultas del Condominio Ciudad Verde.
2. Un sistema de pronóstico usando técnica de machine learning permite detectar la depresión en el área conductual de las personas adultas del Condominio Ciudad Verde.

En este proyecto de investigación realizamos las búsquedas de antecedentes de artículos internacionales y nacionales, también investigamos teorías, enfoques conceptuales que emplearemos para el desarrollo de la investigación, a continuación, detallaremos:

## **ANTECEDENTES**

Qingwen Jia, Yanhan Duan, Rui Gong, Meijun Jiang, Dianping You, Yi Qu (2023), en la investigación que realizaron tuvieron como variable dependiente la depresión de los adultos mayores y como variable independiente la disposición de vida, la cual su objetivo era explorar la relación entre la condición de vida y la depresión. Los datos que usaron fueron de la Encuesta Longitudinal China sobre Longevidad Saludable (CLHLS, por sus siglas en inglés) que fue realizado en el Centro de Estudios de Desarrollo y Envejecimiento Saludable; la cual el cuestionario hace referencia a información básica, estado de salud, condiciones de vivienda y otros contenidos. Este artículo explora las condiciones de vida y la depresión de los adultos mayores, la cual construyeron modelo de regresión logística binaria para el análisis donde exploraron la relación potencial entre las características sociodemográficas, las condiciones de vida reales y la depresión en el modelo 1. Para formar el modelo 2 y 3 agregaron valores del modelo 1 incluyeron las preferencias de condiciones de vida entre preferencias de arreglo de vida y realidad. Aplicaron el odds ratio (OR) y el intervalo de confianza del 95% (IC 95%). Los indicadores fueron analizados mediante SPSS 22.0, con una significancia de  $p < 0,05$  bilateral.

Concluyendo que los adultos mayores que realmente viven solos o solo con su conyugue tienen un mayor riesgo de depresión (los valores de OR en los modelos 1, 2 y 3 son 1,261, 1,838 y 1,732 respectivamente).

Los adultos mayores que prefieren vivir solos o solo con su conyugue tienen un riesgo relativamente bajo de depresión (los valores de OR en los modelos 2 y 3 son 0,553 y 0,611 respectivamente). Y los adultos mayores que la preferencia no coincide con la realidad tienen un mayor riesgo de depresión (OR = 1,320). Se destaca en esta investigación la relación de condición de vida y los problemas reales en la vida cotidiana que afronta las personas adultas mayores, donde la salud mental es referente e importante porque con el tiempo el problema puede llegar hacer complejo y agravante para las personas en general.



MALHOTRA y JINDAL(2022) en su artículo de investigación escribe sobre el suicidio y la detección de depresión en las redes sociales, la cual hace referencia que el arraigo del problema de salud mental es la depresión siendo así la más común y la principal causa del suicidio y conducta autolesiva. También informa que las redes sociales es un factor atractivo y medio de comunicación preferido de las personas en la cual se sienten apoyados para transmitir sus pensamientos, sentimientos y emociones. Su objetivo de esta investigación es estudiar las técnicas de aprendizaje profundo para la detección temprana y el diagnóstico predictivo no clínico de depresión, autolesión e ideación suicida a partir únicamente del contenido de redes sociales en línea. Realizaron una revisión sistemática de la literatura (SLR) de 96 estudios de investigación que han aplicado técnicas de aprendizaje profundo para detectar la depresión y el suicidio o el comportamiento de autolesión a partir del contenido de las redes sociales. Además, generan la confianza de los profesionales de la salud en las predicciones del modelo y aportan a comprender mejor los síntomas de la depresión para un diagnóstico y tratamiento integral, los modelos de redes neuronales profundas también deberían ayudar a extraer relaciones causales-efecto, motivos de depresión o factores estresantes/causas que conducen al suicidio o incidentes de autolesión en los usuarios.

TIANT, Gang, et al., (2022) en el artículo indica su objetivo que el estudio es explorar la asociación entre discapacidad, apoyo social y síntomas depresivos en los ancianos chinos. El estudio se realizó con una base de datos de 9,231 pacientes, mayores de 65 años. Siendo su objetivo explorar la asociación entre la discapacidad, el apoyo social y los síntomas de depresión, hallando el tema del apoyo social afecta los síntomas de la depresión. Pudiendo apoyar en nuestra investigación, en demostrar sobre la dimensión de área afectiva está involucrada en gran medida sobre como esta influye en los síntomas de la depresión y ser usado para poder detectar la depresión.

Chiyoung, Heewon, Gennady (2022), investigaron para poder predecir la depresión en personas adultas con hipertensión a través del uso de ML, en Estados Unidos. La

población se analizó a través de un dataset con datos de 8628 adultos mayores de 40 años. Los instrumentos que se usaron en la investigación fueron revisión de documentos. Los principales resultados obtenidos fueron no ponderados, pero demostraron la importancia del uso de machine learning y su uso como herramienta de predicción en el sector de la salud, demostrado usando diferentes modelos de machine learning como Random Forest, SVM, Adaboost y otros más. Se resalta la relevancia que tiene los diferentes modelos de Machine Learning, que se usan con el fin de ayudar a obtener resultados para detectar y predecir personas con depresión, demostrando la utilidad del aprendizaje automatizado para el sector salud.

Richter, Fishbain, Fruchter, Richter-Levin, Okon-Singer (2021), investigaron la creación de ML diseñado para diferenciar entre trastornos de ansiedad y depresión. La población fue de 86 pacientes psiquiátricos con uno o ambos trastornos mentales. Los principales resultados obtenidos fueron positivos al demostrar su utilidad como sistema de apoyo para el diagnóstico psiquiátrico. Usando el algoritmo ML Random Forest, se muestra una tasa de éxito de detección del 80,50% para la clasificación de la ansiedad y 66,46% de éxito para la clasificación de la depresión. Se demuestra la precisión con la que opera el algoritmo de Machine Learning para detectar y diferenciar los trastornos de ansiedad y depresión o en algunos casos se encuentran mentalmente estables; mostrando y enseñando la eficacia y la confiabilidad de la ML.

Wardenaar, Riese, Giltay, Eikelenboom, Van Hemert, Beekman, Schoevers (2021) investigaron determinar si existen factores comunes que afecten tanto a la depresión como a la ansiedad, así como factores específicos que influyan en cada uno de estos trastornos de manera particular. La población fue de 2981 personas entre los 18 a 65 años. Los instrumentos usados fueron análisis de documentación, entrevistas y herramientas de análisis. Principalmente los resultados obtenidos se encontró correlaciones de predicción bajas a altas ( $\rho_{pred}$ : 0,41-0,91, mediana = 0,73), puesto que en este estudio se encontró que la mayoría de la gravedad tanto de la ansiedad como la depresión a lo largo del tiempo son compartidos, pero también hay factores

específicos para cada trastorno.

National Institute of Mental Health (2021), detalla que el trastorno de la depresión es conocido como “depresión clínica” o “trastorno depresivo mayor” sus síntomas son graves en los individuos que alteran en la coordinación de sus actividades diarias. Mencionan 2 tipos de depresión frecuentes y son: Depresión mayor y trastorno depresivo persistente, pero también existen otras formas de depresión. Los síntomas asociados a la depresión mencionaremos algunos: sentimientos de tristeza, ansiedad o vacío, sentimientos de culpable, pérdida de interés en las actividades y pasatiempos, suicidio, entre otros. Para diagnosticar la depresión se debe tener 5 síntomas persistente durante dos semanas, uno de los síntomas debe ser un estado de ánimo depresivo o la pérdida de interés de las actividades que realizan. Los adultos de mediana edad son los que tienen más episodios de depresión y los que más sufren porque su libido reduce, sufren de desvelo a mitad de la noche o se despiertan en automático durante la madrugada, también se le complica con síntomas gastrointestinales como diarrea o estreñimiento.

## **CONCEPTOS:**

### **Depresión:**

Gil (2021), define que la depresión puede padecer cualquier persona en cualquier lugar del mundo, aunque se tiene sectores de la población que son de mayor riesgo, donde la depresión es grave y es asociado al suicidio. En estos sectores padecen las personas mayores y los adolescentes, y con frecuencia en mujeres que en hombres. La depresión está ligada con el envejecimiento del individuo, la vida en soledad y con patologías crónicas.

### **Pronostico:**

Hyndman y Athanasopoulos (2021), en su libro “Forecasting: Principles and Practice” indican que los pronósticos son la manera de obtener datos del futuro estudiando los

datos del pasado y también los métodos confiables que existen en estos tiempos para producir pronósticos fiables el cual se pueden aplicar en la mayoría de los ámbitos.

### **El inventario de Depresión de Beck-II (BDI-II):**

Maldonado, Castro y Cardona (2021), en su artículo de investigación informa que el instrumento "El inventario de Depresión de Beck-II (BDI-II)" se utiliza en la práctica clínica y de investigación, haciendo énfasis que realiza la medición de los síntomas depresivos. Usaron el instrumento para la evaluación de la confiabilidad y tener como evidencia una muestra de 409 estudiantes universitarios colombianos, En los resultados encontraron que en la fiabilidad obtuvieron un alfa de Cronbach = 0,91 y como prueba de validez encontraron correlaciones de Item-test que oscilan entre 0,31 y 0,67. Este artículo evidencia las puntuaciones obtenidas en el BDI-II que permiten inferir los síntomas depresivos de la población y también recomienda su uso clínico e investigador.

### **Machine Learning:**

Nwanganga, Chapple (2020), en su libro "Practical Machine Learning in R", el enfoque que dan a los computadores para poder aprender y analizar los datos que se le dan, poder identificar tendencias, predecir un resultado futuro con una precisión elevada. Hecho muy útil para las organizaciones y empresas que la información y tratamiento de datos ayudaran a una mejora toma de decisiones. Existen varios modelos y métodos de machine learning según el fin para el que lo estás usando, que van desde la clasificación de datos hasta la predicción de resultados. La que hay entre las diferentes técnicas y métodos es que la base de sus resultados es a través de fórmulas matemáticas, que difieren según el método.

Según Díaz-Martínez, Ahumada-Cervantes y Melo-Morín (2021), el árbol de decisión se define como "una prueba estadística de predicción cuya función objetivo es la de interpretar resultados a partir de observaciones y construcciones lógicas".

De manera teórica se describe cómo funciona el árbol de decisión, que es necesario

para poder hacer uso de ello y que resultados se podrán obtener del mismo.

### **Sistema de pronóstico:**

Makridakis, Wheelwright & Hyndman (1998), Un sistema de pronóstico busca predecir según los factores del entorno presente, y busca establecer logros de los objetivos y metas establecidos según el fin para el que se diseñó, buscando disminuir el azar y siendo más científico. Según la necesidad del usuario u empresa, se brinda el enfoque al sistema para que pueda predecir eventos inciertos y realice pronósticos, este sistema de pronóstico puede aplicarse en diferentes campos.

### **Indicadores de la depresión:**

Belloch & Rojas (1995, 1981), precisan que los síntomas de la depresión se manifiestan en el estado del ánimo, área cognitiva, conductas y relaciones sociales de los individuos.

### **Síntomas anímicos:**

Uno de los síntomas del ánimo para Rojas (1981), escribió que la tristeza es el síntoma nuclear más resaltante dentro del cuadro depresivo de las personas. Indicando también los síntomas como irritabilidad, sensación de vacío, nerviosismo, infelicidad, desesperanza y abatimiento.

### **Síntomas motivacionales y conductuales:**

Los síntomas como la apatía e indiferencia que son la capacidad de la disminución de disfrute de la vida cotidiana como asearse, levantarse, alimentarse, caminar o actividades que se realizan básicamente. Otro síntoma es la anhedonia que incapacita a experimentar placer. Todo esto concluye a que afecte un retardo psicomotor que significa baja motivación o predisposición a realizar actividades básicas.

### **Crisp-DM:**

Según Ayele (2020), CRISP-DM es un modelo de proceso de minería de datos

genérico que proporciona una descripción general de los ciclos de vida de los proyectos de minería de datos. Se ha vuelto popular tanto en la industria como en el mundo académico y se considera el estándar de facto para proyectos de descubrimiento de conocimientos y minería de datos. En el artículo se muestra relevancia al uso de esta metodología es el bajo costo y su facilidad de aplicación en proyectos innovadores.

## **II. METODOLOGÍA**

## 2.1 Tipo y diseño de investigación

### 2.1.1. Tipo de investigación

Según Ñaupas (2018, p. 108) escribió sobre la investigación aplicada que son aquellas técnicas u ciencias aplicadas a los problemas. Para realizar mejoras en diferentes áreas como procesos de información, vivienda y otras más.

En nuestra investigación fue de tipo aplicada, porque mejoramos procesos para el sector de vivienda, en la cual apoyamos con nuestra investigación básica al conocimiento tecnológico, por lo cual buscábamos integrar conocimientos teóricos y brindar solución al problema.

### 2.1.2. Diseño de investigación

Hernandez Siampieri (2014, p. 128) el término diseño hace referencia al planeamiento para la obtención y recopilación de la información por el cual responderá al planteamiento del problema.

Ñaupas (2018, p. 360) el diseño preexperimental no junta los suficientes requisitos para ser un experimento puro, sin embargo, el diseño cumple con mostrar un control mínimo.

Hernández Sampieri (2014, p. 141), preexperimentos; en el estudio de caso con una sola medición, atribuye un tratamiento a un grupo para después aplicar una medición a una o más variables. Su diseño de diagrama es el siguiente:



Fuente: Hernandez Siampieri (2014, p.141)

Figura N°1: Diseño de estudio de caso con una sola medición  
Nuestra investigación fue de diseño experimental, preexperimental.



Tomamos las variables siguientes:

G: Grupo experimental: Es el grupo de personas adultas (muestra) al cual se tendrá la dimensión del área afectiva y área conductual para la detección de la depresión.

X: Experimento (Machine Learning): Se implementará un sistema de pronóstico usando técnica de Machine Learning para detectar la depresión en personas adultas del Condominio Ciudad Verde.

O: Se mide a una porción experimental usando la técnica de Machine Learning para la detección de la depresión, para obtener los datos que están divididos en entrenamiento con un porcentaje de 70% y de test un 30%.

Realizaremos las comparaciones de las variables que usaremos para la detección de la depresión la cual la variable dependiente es medida como instrumento en la implementación con técnica de Machine Learning (ML).

Se considera la siguiente información para nuestro proyecto de investigación:

### ***Técnicas de Machine Learning:***

Regresión Lineal: Es un algoritmo de aprendizaje supervisado y se utiliza para el análisis de datos en estadísticas y machine learning. Modela la relación entre una variable escalar dependiente “Y” y una o más variables explicativas “X”.

Precisión: Para saber la probabilidad de acierto en la predicción.

$$\text{Precisión} = \frac{VP + VN}{VP + VN + FP + FN}$$

Figura N°2: Fórmula de precisión

Tasa de error: Para saber la probabilidad de error en la predicción.

$$\text{Error rate} = \frac{FP + FN}{Total}$$

Figura N°3: Fórmula de Tasa de error

Medida Gini: Formula usada por árbol de decisión para la elección y división de nodos.

$$Gini(D) = 1 - \sum_{i=1}^m P_i^2$$

Figura N°4: Fórmula del índice de Gini

MAE media del valor absoluto de los errores y RMSE raíz cuadrada de la media de los errores al cuadrado (RMSE): Diferencia de los errores entre los valores de predicción y valores reales de una variable.

$$RMSE = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}$$

Figura N°5: Fórmula RMSE

$$MAE = \frac{|(Y_i - Y_p)|}{n}$$

Figura N°6: Fórmula MAE

Árbol de Decisiones: Modelo de predicción que es utilizado en diversos ámbitos desde la inteligencia artificial hasta la economía. También es un método de clasificación de machine learning; que son diagramas de construcciones lógicas que sirven para categorizar una serie de condiciones que ocurren de forma sucesiva.

## **Herramientas para el desarrollo del Sistema de Pronostico**

### Configuración del entorno:

- Anaconda: Es una distribución popular de Data Science y Machine Learning, donde utilizaremos su herramienta Jupyter la cual es un entorno de desarrollo para Python, el cual es de código abierto que abarca una galería de aplicaciones, librerías, donde se muestra imágenes, escribir notas o comentarios, entre otros.

### Lenguaje de Programación

- Python: Es un lenguaje de programación de alto nivel de código abierto, orientado a objetos. Es utilizado en desarrollo de software, aplicaciones web, la ciencia de datos y machine learning. Solo se nombrará algunas librerías de Python:

Modulo NumPy: Extensión de Python para trabajar con vectores y matrices.

Modulo Pandas: Es una librería de código abierto escrita como extensión de NumPy para la manipulación y análisis de datos en Python.

Modulo Matplotlib: Es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación de Python.

Modulo Seaborn: Es una librería de visualización de datos en Python basada en matplotlib. Se puede realizar gráficos estadísticos de forma muy vistosa.

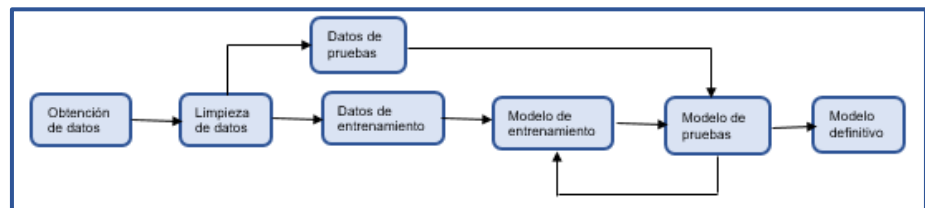


Figura N°4: Procesos con las técnicas de Machine Learning:

Fuente: Elaboración propia

## 2.2 Variables y operacionalización

Según Estrada (2023), redacta que la operacionalización de variables es útil para alinear objetivos de investigación con el diseño de instrumentos de investigación. Nuestro proyecto de investigación la variable es de tipo cuantitativo, se categoriza en variable independiente y dependiente.

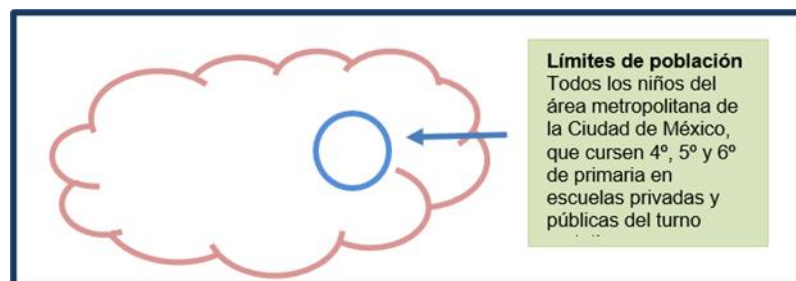
- Variable independiente: Sistema de pronóstico
- Variable dependiente: detección de la Depresión

Se detallará en la tabla de operacionalización de variables, se encontrará en (Anexo 2).

## 2.3 Población, muestra y muestreo

### 2.3.1. Población

Según Ñaupas, Valdivia, Palacios, Romero (2018, p. 334), se refieren a la población como Unidades de estudio, que se diferencian por características que son requeridas para la investigación. Estas características son: fenómenos, personas, objetos, hechos, etc.



Fuente: Hernandez, Fernandez, Baptista (2014, p.174)

Figura N°5 Ejemplo de delimitación de la muestra

Para Hernandez, Fernandez y Baptista (2014), informan que para evitar errores debemos de tener una adecuada delimitación del universo o población, todo va a depender de los objetivos de estudio que el investigador determine

de manera específica.

Nuestra investigación que realizamos, como población tuvimos a los residentes del Condominio Ciudad Verde, de lo cual, para detectar la depresión en el área afectiva y área conductual, como objeto de estudio tomamos como población a 1360 personas aproximadamente entre el rango de 20 a 60 años.

### 2.3.2. Muestra

Para Ñaupas, Valdivia, Palacios, Romero (2018, p 334, 335), define a la muestra como una fracción del poblamiento en la cual tienen características que los une para la investigación, permitiendo generalizar los resultados. La muestra controla el tamaño y la selección.

Cálculo para obtener la muestra:

$$n = \frac{N * Z^2 * p * q}{e^2 * (N - 1) + Z^2 * p * q}$$

Figura N°6 Fórmula de muestra

N: Población

Z: Nivel de Confianza al 95% (1.96)

P: Proporción Esperada (5%)

e: Precisión o el error (5%)

q: 1 – p (0.95)

$$n = \frac{(800) * (1.96)^2 * (0.5) * (1 - 0.5)}{(0.05)^2 * (800 - 1) + (1.96)^2 * (0.5) * (1 - 0.5)}$$

$$n = 301$$

### 2.3.3. Muestreo

Para Ñaupas, Valdivia, Palacios, Romero (2018, p.336), definen al muestreo como el procedimiento que selecciona las unidades de estudio para luego conformar la muestra. Comprende etapas que respetan y garantizan la confianza, para no sesgar la muestra. También existen tipos de muestreo que utilizan en función de la investigación.

También Ñaupas, Valdivia, Palacios, Romero (2018, p.337), escriben en su libro que los diseños explicativos según el criterio de la investigación necesitan una definición del método de muestreo, como también los diseños preexperimentales, experimentales y cuasi experimentales. También presentan un cuadro de los métodos muestrales en función al tipo de muestreo, como se observa en la siguiente figura:

Muestreos Probabilísticos	Muestreos No Probabilísticos
Aleatorio Simple	Por conveniencia
Al azar sistemático	Opinático
Estratificado	Por cuotas
Por conglomerados	Elección razonada

Fuente: Ñaupas, Valdivia, Palacios, Romero (2018, p.335)  
Tabla N°1: Métodos muestrales en función a tipo de muestreo

Existen 2 tipos de muestreo los cuales son: probabilístico y no probabilístico.

- **Probabilístico:** Es el tipo de muestreo que interviene el azar, donde todos los que componen una población tienen la posibilidad de ser elegido.
- **No probabilístico:** El investigador tiene el criterio para la selección de las unidades muestrales, dependiendo las características que requiera la investigación a desarrollar.

Para nuestra investigación emplearemos el muestreo No Probabilístico por conveniencia.

#### **2.3.4 Unidad de análisis**

La población que tenemos como objeto de estudio en la cual seleccionaremos son las *personas adultas* para que conforme la muestra.

### **2.4 Técnicas e instrumentos de recolección de datos**

Hernández, Duana (2020), Las técnicas de recolección de datos son métodos usados por el investigador para obtener datos e información necesaria para encontrar la respuesta adecuada para su investigación.

Para Hernandez Sampieri (2014 p.217), el cuestionario es un instrumento muy utilizado para recolectar datos que se basa en preguntas abiertas o cerradas que se utilizan en las encuestas para medir una o más variables.

En nuestro proyecto de investigación emplearemos el cuestionario para evaluar el área afectiva y área conductual de la depresión de los residentes del Condominio Ciudad Verde.

### **2.5 Procedimientos**

Para nuestra investigación el primer paso es establecer el título, luego establecer los puntos necesarios como la recolección de antecedentes y buscar en las bibliografías. Luego procederemos con la recolección de información que se llevará a cabo mediante el cuestionario de Inventario de Depresión de Beck-II (BDI-II) que es validado y utilizado previamente en estudios clínicos de investigación para detectar la depresión.

Para obtener la aceptación de la población del Condominio Ciudad Verde, se

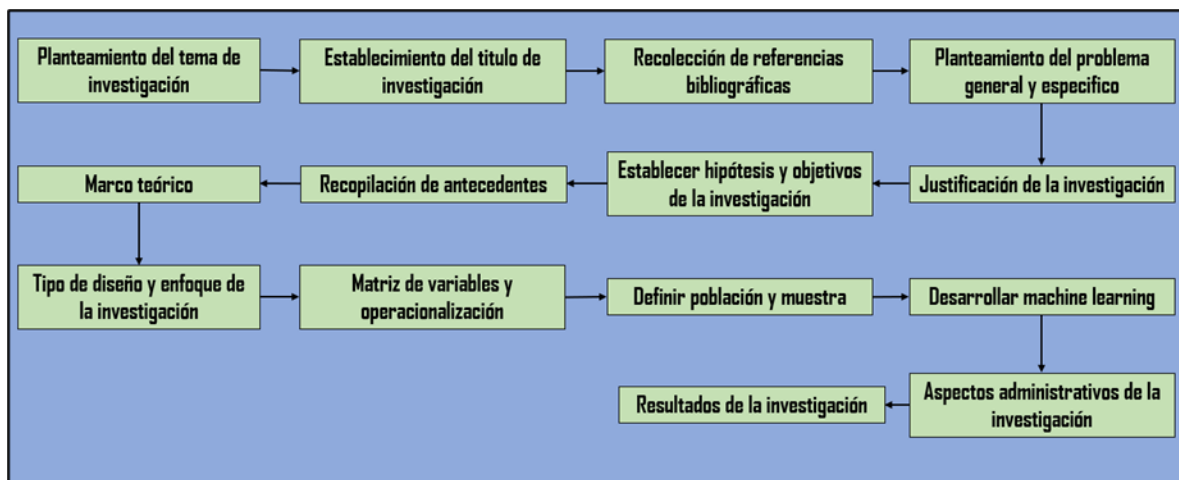
enviará una carta de autorización, solicitando el consentimiento para realizar nuestro desarrollo de investigación, cumpliendo todos los aspectos éticos y normas establecidos.

El proceso de recolección de datos realizaremos un Post - test.

Se permitirá el mapeo de las variables asociados a la depresión para luego evaluar los factores resultantes.

Asimismo, se implementará un sistema de pronóstico para la detección de la depresión, en la cual analizaremos los resultados obtenidos.

Mediante la figura siguiente detallamos el procedimiento que se está realizando en la investigación:



Fuente: Elaboración Propia

Figura N°7: Mapa de procedimiento del proyecto de Investigación

## 2.6. Método de análisis de datos

Hernandez Siampieri (2014 p.273,274), redacta que el análisis de datos se genera en una matriz de datos que en la actualidad existen varios programas computacionales.

Para nuestra investigación final usaremos las técnicas de Machine Learning



para el análisis de datos, la cual se obtendrá la recolección de datos usando como instrumento el cuestionario.

La data se exportará a la herramienta de Microsoft Excel por el cual el proceso de los datos se ordena y se selecciona según la conveniencia, para luego ser importado en Python para trabajar con la data en el entrenamiento y el test.

## **2.7. Aspectos éticos**

Para nuestra investigación se respetó los Códigos de éticas de la Universidad Cesar Vallejo. Se considera los capítulos y artículos más resaltantes los cuales son los siguientes:

En Capitulo III: Normas y éticas para el desarrollo de la investigación emplearemos los siguientes artículos:

El artículo 4° la investigación con seres humanos, donde salvaguardamos la integridad de la persona y la confidencialidad de sus datos personales, el cual será el anonimato para evitar la identificación de las personas.

El artículo 7° del consentimiento y asentamiento informado, donde incluiremos a las personas mayores de edad, brindaremos la información adecuada y comprensible de la investigación. Se adjunta la carta de aceptación del Condominio Ciudad Verde (ANEXO 4)

El artículo 10° de la originalidad de la investigación, la Universidad Cesar Vallejo nos promueve y exige que nuestro producto de investigación respete los derechos de autor.

El artículo 11 de los derechos de autor, la Universidad Cesar Vallejo tiene la ley

de Derecho de Autor y Reglamento de Propiedad Intelectual, por el cual consideramos este artículo si cometemos plagio o atentamos contra la ética de la investigación.

Por lo expuesto nosotros nos comprometemos a brindar información confiable, sin alterar nuestra investigación; y para investigaciones futuras se pueden apoyar y tomar como referencia la investigación.

### **III. RESULTADOS**

## Registros de la detección de depresión de personas adultas del Condominio Ciudad Verde

- Se verifica que la data importada en Python no tenga información con valores nulos o vacíos.

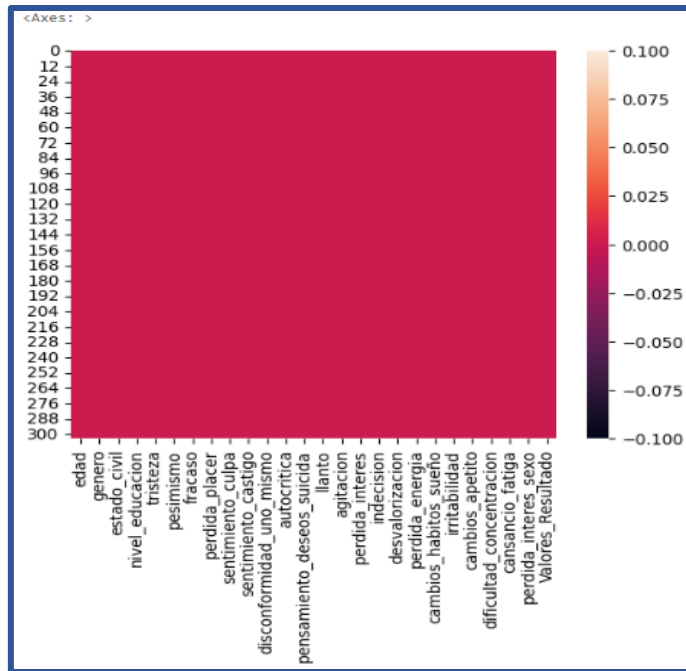


Figura N°8: Gráfico de los números de registros (eje “Y”) y las columnas (eje “X”) Fuente: Elaboración Propia

### Información demográfica:

Edad	Valores
Entre 20 y 29	1
Entre 30 y 39	2
Entre 40 y 49	3
Entre 50 y 59	4

Tabla N°2: Valores de las categorías de edades Fuente: Elaboración Propia

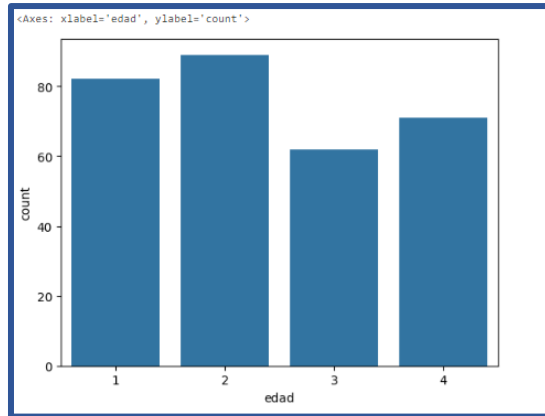


Figura N°9: Gráfico de las categorías de edades (eje “X”)  
Fuente: Elaboración Propia

Género	Valores
Mujer	1
Hombre	2

Tabla N°3: Valores de categorías de género  
Fuente: Elaboración Propia

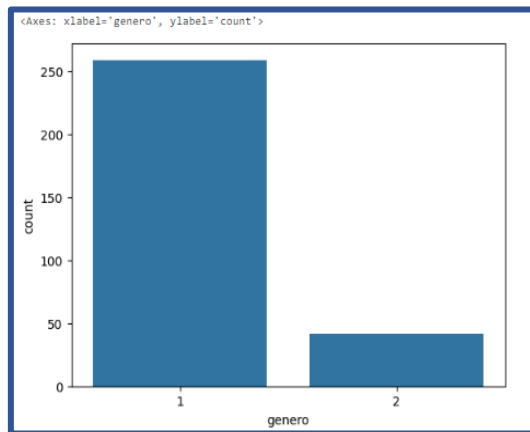


Figura N°10: Gráfico de categoría de género (eje “X”)  
Fuente: Elaboración Propia

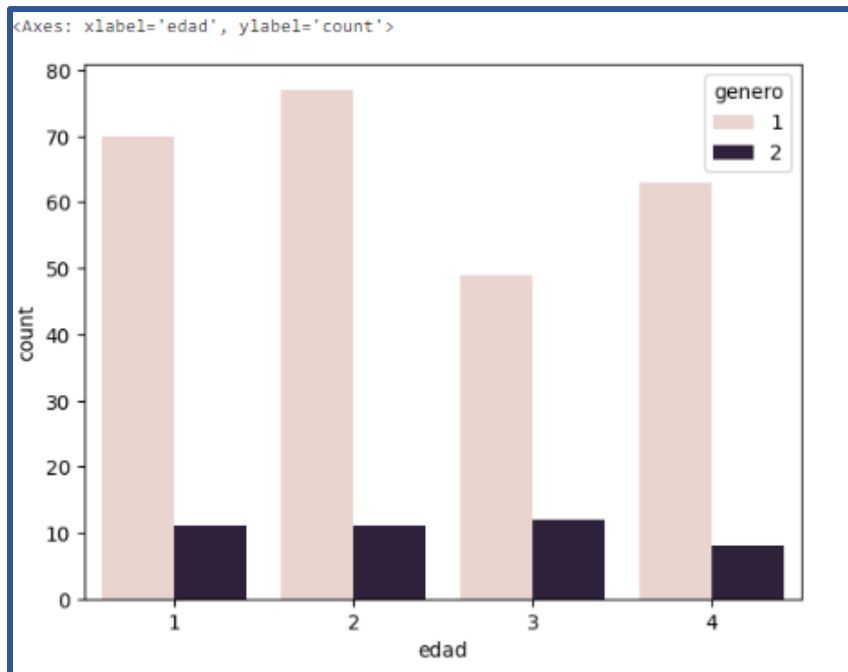


Figura N°11: Gráfico de categoría de edad (eje "X") con respecto al género  
 Fuente: Elaboración Propia

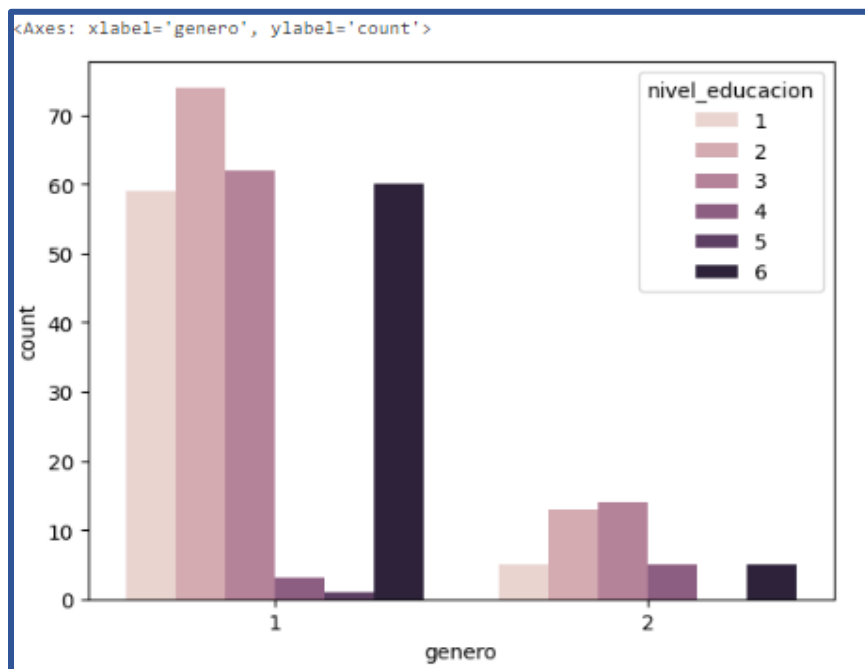


Figura N°12: Gráfico de Categoría de género (eje "X") con respecto al nivel de educación  
 Fuente: Elaboración Propia

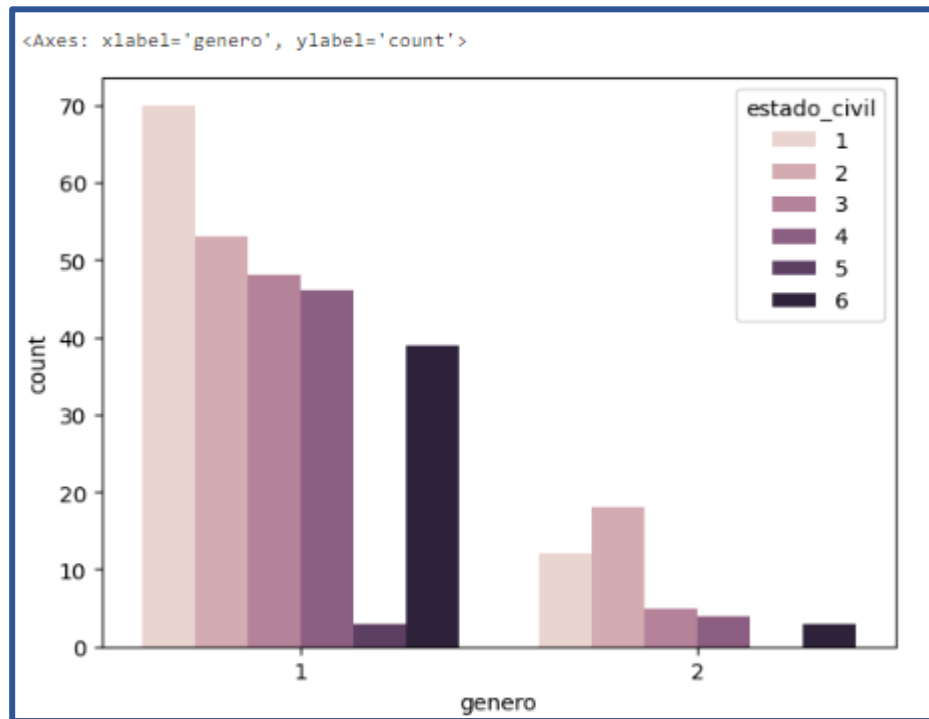


Figura N°13: Gráfico de Categoría de género (eje “X”) con respecto al estado civil  
Fuente: Elaboración Propia

### Información del Inventario de Depresión de Beck-II (BDI-II):

Se considera dos dimensiones: área afectiva y área conductual; cada área tomaremos dos indicadores.

- Área afectiva: Agitación, pensamientos o deseos suicidas.
- Área conductual: Autocritica, Dificultad de concentración.

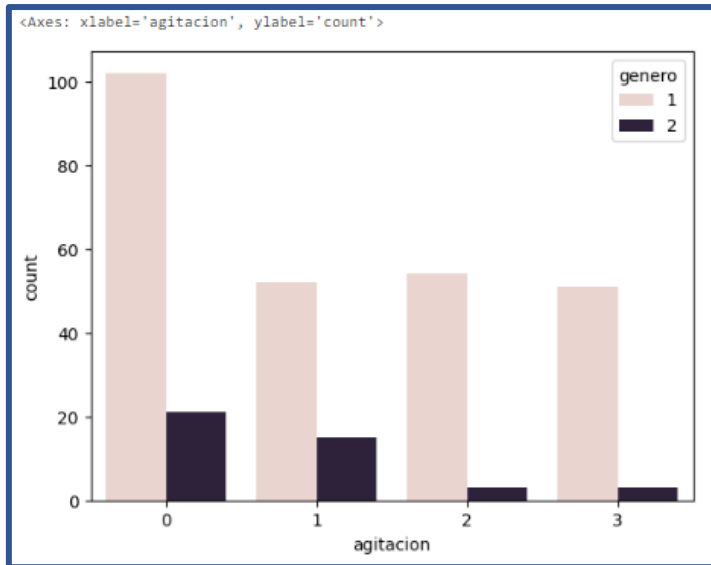


Figura N°14: Gráfico de indicador “Agitación” (eje “X”) con respecto al género  
Fuente: Elaboración Propia

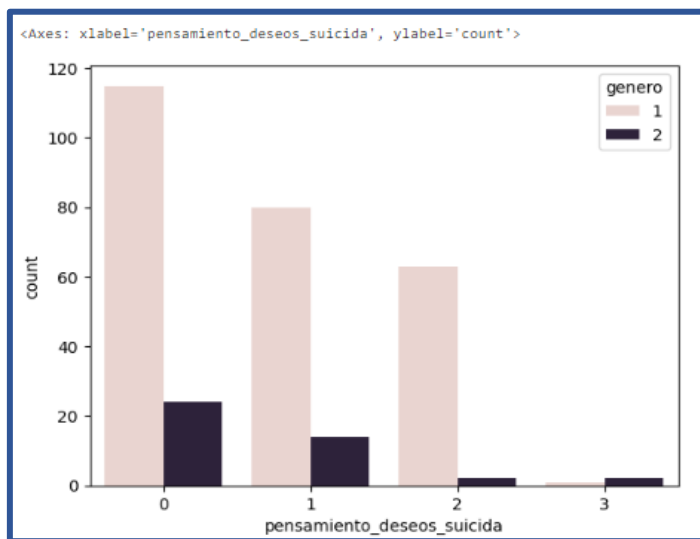


Figura N°15: Gráfico de indicador “Pensamiento o deseo suicida” (eje “X”) respecto al género  
Fuente: Elaboración Propia



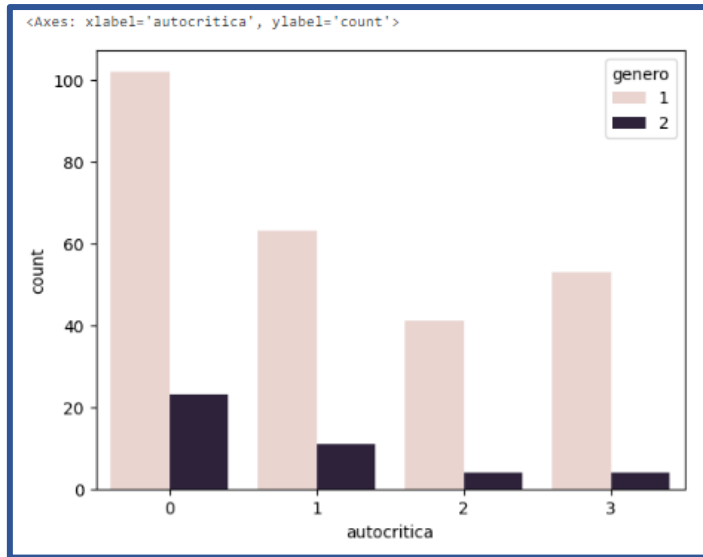


Figura N°16: Gráfico de indicador “Autocritica” (eje “X”) con respecto al género  
Fuente: Elaboración Propia

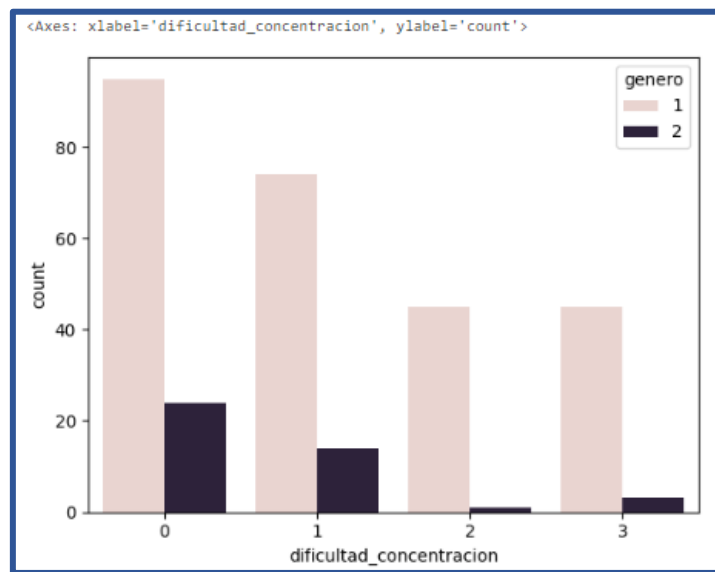


Figura N°17: Gráfico de indicador “Dificultad de concentración” (eje “X”) respecto al género  
Fuente: Elaboración Propia

## Resultados de Predicciones para la Depresión en personas adultas

### Inventario de Depresión de Beck – II (BDI - II)

**Agitación:** Tiene 4 opciones, se visualiza en el gráfico de “Agitación” que la primera opción tiene más proporción que las otras tres opciones, teniendo las tres opciones proporciones casi iguales.

<b>Agitación</b>	<b>Valores</b>
<i>No estoy más inquieto o tenso que lo habitual</i>	0
<i>Me siento más inquieto o tenso que lo habitual</i>	1
<i>Estoy tan inquieto o agitado que me es difícil quedarme quieto</i>	2
<i>Estoy tan inquieto o agitado que tengo que estar siempre en movimiento o haciendo algo</i>	3

Tabla N°4: Agitación  
Fuente: Elaboración Propia

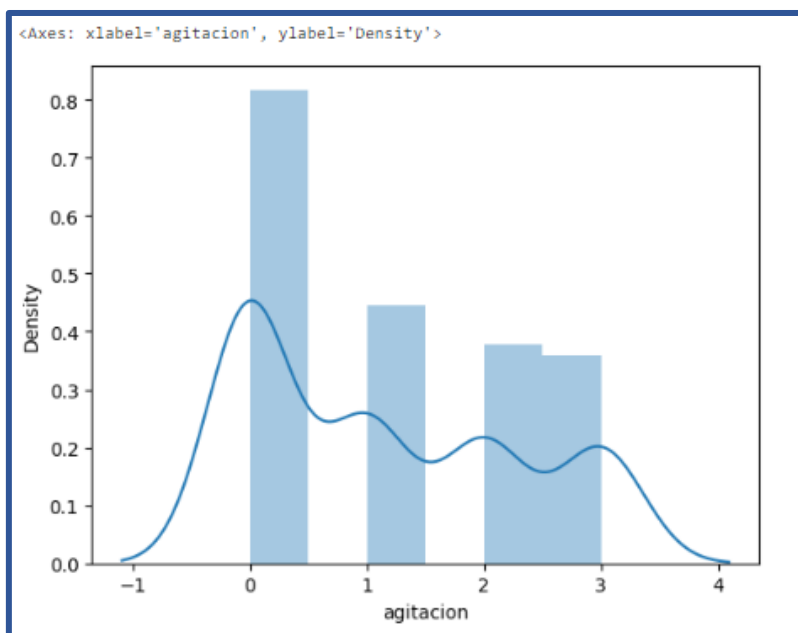


Figura N°18: Gráfico de indicador “Agitación” (eje “X”)  
Fuente: Elaboración Propia

**Pensamiento o deseos suicidas:** Tiene 4 opciones, se visualiza en el gráfico que la última opción “*Me mataría si tuviera la oportunidad de hacerlo*” tiene una proporción insignificativa para las otras opciones.

<b>Pensamientos o Deseos Suicidas</b>	<b>Valores</b>
<i>No tengo ningún pensamiento de matarme</i>	0
<i>He tenido pensamientos de matarme, pero no lo haría</i>	1
<i>Querría matarme</i>	2
<i>Me mataría si tuviera la oportunidad de hacerlo</i>	3

Tabla N°5: Pensamiento o deseos suicidas  
Fuente: Elaboración Propia

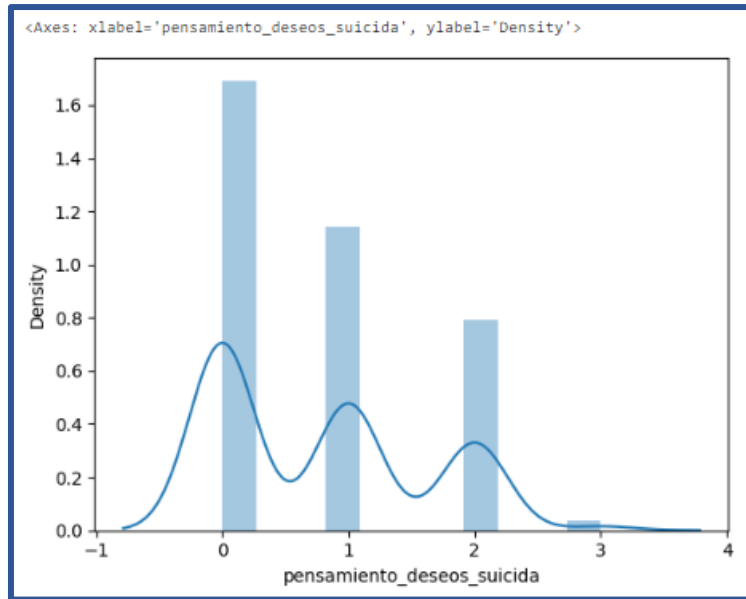


Figura N°19: Gráfico de indicador “Pensamientos o Deseos Suicidas” (eje “X”)  
Fuente: Elaboración Propia

**Autocrítica:** Tiene 4 opciones, se visualiza en el grafico que la tercera opción tiene una proporción más baja que las otras opciones.

<b>Autocrítica</b>	<b>Valores</b>
<i>No me critico ni me culpo más de lo habitual</i>	0
<i>Estoy más crítico conmigo mismo de lo que solía estarlo</i>	1
<i>Me critico a mí mismo por todos mis errores</i>	2
<i>Me culpo a mí mismo por todo lo malo que sucede</i>	3

Tabla N°6: Autocrítica  
Fuente: Elaboración Propia

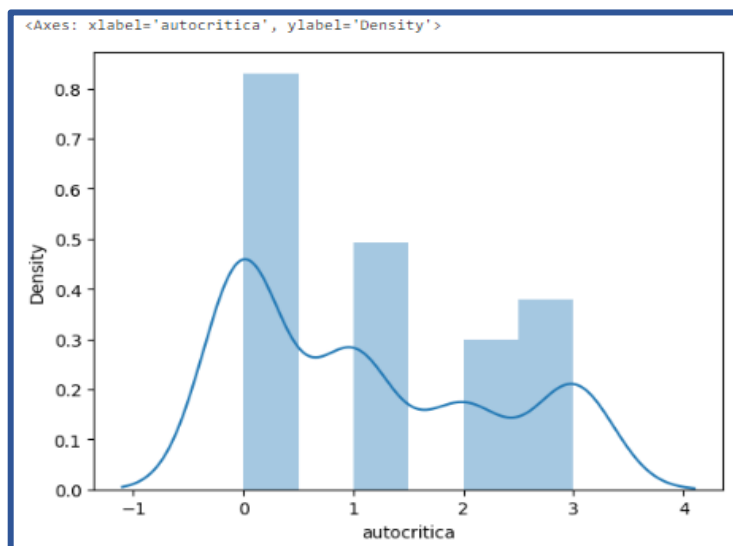


Figura N°20: Gráfico de indicador “Autocritica” (eje “X”)  
Fuente: Elaboración Propia

**Dificultad de concentración:** Tiene 4 opciones, se visualiza en el gráfico que la segunda opción tiene una proporción intermedia entre la primera opción y la tercera y cuarta opción.

<b>Dificultad de Concentración</b>	<b>Valores</b>
<i>Puedo concentrarme tan bien como siempre</i>	0
<i>No puedo concentrarme tan bien como habitualmente</i>	1
<i>Me es difícil mantener la mente en algo por mucho tiempo</i>	2
<i>Encuentro que no puedo concentrarme en nada</i>	3

Tabla N°7: Dificultad de concentración  
Fuente: Elaboración Propia

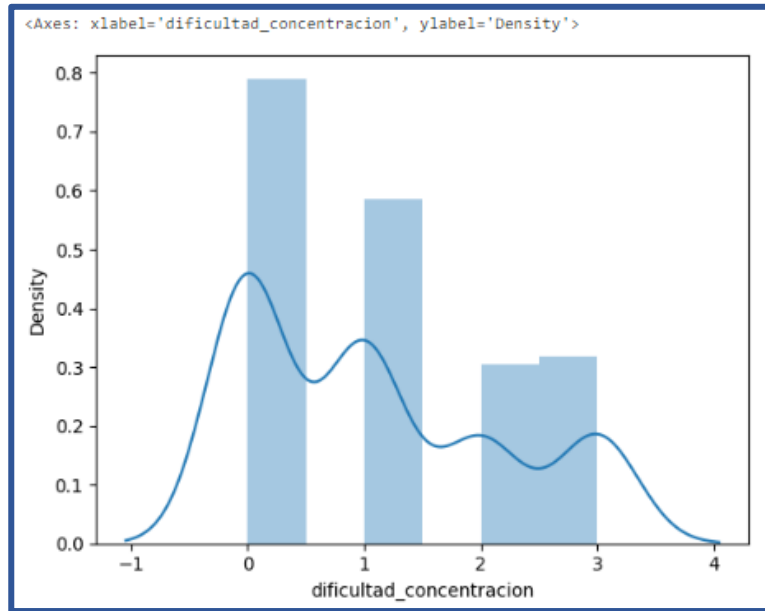


Figura N°21: Gráfico de indicador “Dificultad de concentración” (eje “X”) Fuente: Elaboración Propia

Por consiguiente, se presenta los resultados del Inventario de Depresión de Beck – II (BDI – II) obtenidos en la fase de entrenamiento y test para el análisis predictivo con Machine Learning.

**Regresión Lineal:**

**Agitación:**

X = edad, género  
Y = agitación

Se divide los datos para el train 70% y el test 30%. Obteniendo los datos aleatoriamente, 210 para el train y 91 para el test.

X_TRAIN			Y_TRAIN	
#	Género	Edad	#	Agitación
185	1	3	185	2
72	2	2	72	1
15	1	3	15	2
10	2	4	10	2
196	1	4	196	2

Tabla N°8: Se muestra algunos datos de entrenamiento (TRAIN) para las ejes “X” y “Y”

Fuente: Elaboración Propia

X_TEST			Y_TEST	
#	Género	Edad	#	Agitación
177	1	4	177	3
289	1	1	289	2
228	1	1	228	3
198	1	4	198	0
60	1	2	60	2

Tabla N°9: Se muestra algunos datos de test (TEST) para los ejes "X" y "Y"  
Fuente: Elaboración Propia

### Evaluación del modelo:

Predicción: Se tomo la X\_test para la predicción.

```
predicciones  
array([[1.44882979, 0.97483131, 0.97483131, 1.44882979, 1.13283081,  
1.44882979, 1.44882979, 0.97483131, 0.97483131, 1.13283081,  
1.00407323, 1.13283081, 0.84607374, 1.13283081, 0.97483131,  
0.97483131, 0.97483131, 0.84607374, 1.44882979, 1.44882979,  
1.2908303, 1.44882979, 0.97483131, 1.44882979, 1.13283081,  
1.13283081, 1.2908303, 1.13283081, 1.44882979, 1.44882979,  
1.13283081, 1.13283081, 1.44882979, 0.68807424, 1.44882979,  
1.44882979, 1.13283081, 0.97483131, 0.97483131, 0.68807424,  
0.84607374, 0.97483131, 0.97483131, 1.13283081, 0.97483131,  
1.2908303, 1.13283081, 1.13283081, 1.2908303, 0.97483131,  
1.44882979, 1.2908303, 1.13283081, 0.84607374, 0.97483131,  
1.13283081, 1.13283081, 1.44882979, 1.13283081, 0.97483131,  
0.97483131, 1.2908303, 0.97483131, 1.13283081, 1.00407323,  
1.44882979, 0.68807424, 1.2908303, 1.2908303, 0.97483131,  
1.00407323, 0.53007475, 1.13283081, 0.97483131, 1.2908303,  
1.2908303, 1.2908303, 1.13283081, 1.44882979, 1.2908303,  
1.44882979, 1.44882979, 0.68807424, 0.53007475, 0.97483131,  
0.97483131, 1.13283081, 0.53007475, 1.2908303, 0.84607374,  
1.44882979])
```

Figura N°22: Array de las predicciones que se obtuvieron de las X\_test  
Fuente: Elaboración Propia

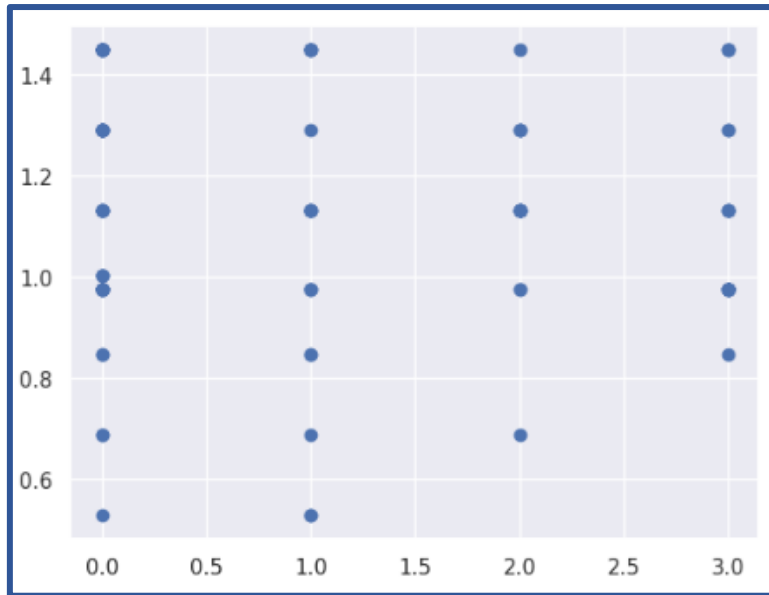


Figura N°23: Gráfico de Regresión Lineal con la (Y\_test, predicciones)  
Fuente: Elaboración Propia

Métricas	Valores
Media del valor absoluto de los errores (MAE)	1.02
Media de los errores al cuadrado (MSE)	1.35
Raíz cuadrada de la media de los errores al cuadrado (RMSE)	1.16

Tabla N°10: Métricas de errores de las (Y\_test, predicciones)  
Fuente: Elaboración Propia

El gráfico de regresión lineal en la dimensión del área afectiva siendo el indicador “agitación”, se obtuvieron el puntaje lineal de la MAE el 1.02, MSE 1.35, RMSE 1.16; siendo que la media del valor absoluto de los errores (MAE) tiene mayor capacidad de pronóstico, eficacia y un buen modelo para la predicción que las métricas de MSE y RMSE.

La media de los errores al cuadrado (MSE) para este modelo de pronóstico no es eficaz para la predicción de los síntomas de la depresión.

### **Pensamiento o Deseos Suicidas:**

X = edad, género

Y= pensamiento\_deseos\_suicida

Se divide los datos para el train 70% y el test 30%. Obteniendo los datos aleatoriamente, 210 para el train y 91 para el test.

X_TRAIN			Y_TRAIN	
#	Género	Edad	#	Pensamiento o Deseos Suicidas
185	1	3	185	0
72	2	2	72	1
15	1	3	15	1
10	2	4	10	0
196	1	4	196	1

Tabla N°11: Se muestra algunos datos de entrenamiento (TRAIN) para los ejes "X" y "Y"  
Fuente: Elaboración Propia

X_TEST			Y_TEST	
#	Género	Edad	#	Pensamiento o Deseos Suicidas
177	1	4	177	1
289	1	1	289	1
228	1	1	228	0
198	1	4	198	1
60	1	2	60	2

Tabla N°12: Se muestra algunos datos de test (TEST) para los ejes "X" y "Y"  
Fuente: Elaboración Propia

## Evaluación del modelo

Predicción: Se tomo la X\_test para la predicción.

```

predicciones
array([[0.8430592 , 0.82991022, 0.82991022, 0.8430592 , 0.83429322,
        0.8430592 , 0.8430592 , 0.82991022, 0.82991022, 0.83429322,
        0.56286054, 0.83429322, 0.55847755, 0.83429322, 0.82991022,
        0.82991022, 0.82991022, 0.55847755, 0.8430592 , 0.8430592 ,
        0.83867621, 0.8430592 , 0.82991022, 0.8430592 , 0.83429322,
        0.83429322, 0.83867621, 0.83429322, 0.8430592 , 0.8430592 ,
        0.83429322, 0.83429322, 0.8430592 , 0.55409456, 0.8430592 ,
        0.8430592 , 0.83429322, 0.82991022, 0.82991022, 0.55409456,
        0.55847755, 0.82991022, 0.82991022, 0.83429322, 0.82991022,
        0.83867621, 0.83429322, 0.83429322, 0.83867621, 0.82991022,
        0.8430592 , 0.83867621, 0.83429322, 0.55847755, 0.82991022,
        0.83429322, 0.83429322, 0.8430592 , 0.83429322, 0.82991022,
        0.82991022, 0.83867621, 0.82991022, 0.83429322, 0.56286054,
        0.8430592 , 0.55409456, 0.83867621, 0.83867621, 0.82991022,
        0.56286054, 0.54971156, 0.83429322, 0.82991022, 0.83867621,
        0.83867621, 0.83867621, 0.83429322, 0.8430592 , 0.83867621,
        0.8430592 , 0.8430592 , 0.55409456, 0.54971156, 0.82991022,
        0.82991022, 0.83429322, 0.54971156, 0.83867621, 0.55847755,
        0.8430592 ]])

```

Figura N°24: Array de las predicciones que se obtuvieron de las X\_test  
Fuente: Elaboración Propia



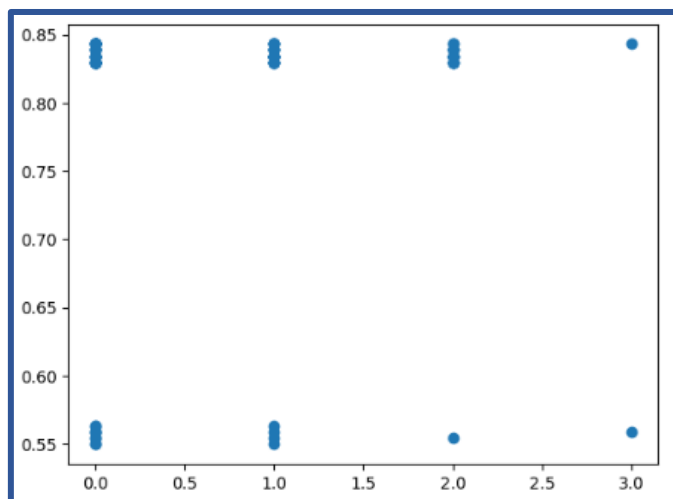


Figura N°25: Gráfico de Regresión Lineal con la (Y\_test, predicciones)  
Fuente: Elaboración Propia

Métricas	Valores
Media del valor absoluto de los errores (MAE)	0.73
Media de los errores al cuadrado (MSE)	0.71
Raíz cuadrada de la media de los errores al cuadrado (RMSE)	0.84

Tabla N°13: Métricas de errores de las (Y\_test, predicciones)  
Fuente: Elaboración Propia

El gráfico de regresión lineal en la dimensión del área afectiva siendo el indicador “Pensamiento o Deseo suicida”, se obtuvieron el puntaje lineal de la MAE el 0.73, MSE 0.71, RMSE 0.84; siendo que la media de los errores al cuadrado (MSE) tiene mayor capacidad de pronóstico, eficacia y un buen modelo para la predicción que las métricas de MAE y RMSE.

La raíz cuadrada de la media de los errores al cuadrado (RMSE) para este modelo de pronóstico no es eficaz para predecir los síntomas de la depresión.

### **Autocritica:**

X = edad, género  
Y= autocritica

Se divide los datos para el train 70% y el test 30%. Obteniendo los datos aleatoriamente, 210 para el train y 91 para el test.

X_TRAIN			Y_TRAIN	
#	Género	Edad	#	Autocritica
185	1	3	185	3
72	2	2	72	0
15	1	3	15	2
10	2	4	10	2
196	1	4	196	2

Tabla N°14: Se muestra algunos datos de entrenamiento (TRAIN) para los ejes "X" y "Y"  
Fuente: Elaboración Propia

X_TEST			Y_TEST	
#	Género	Edad	#	Autocritica
177	1	4	177	1
289	1	1	289	1
228	1	1	228	3
198	1	4	198	1
60	1	2	60	2

Tabla N°15: Se muestra algunos datos de test (TEST) para los ejes "X" y "Y"  
Fuente: Elaboración Propia

## Evaluación del modelo

Predicción: Se tomo la X\_test para la predicción.

```

predicciones
array([[1.2986364 , 1.08647413, 1.08647413, 1.2986364 , 1.15719488,
        1.2986364 , 1.2986364 , 1.08647413, 1.08647413, 1.15719488,
        0.89564571, 1.15719488, 0.82492495, 1.15719488, 1.08647413,
        1.08647413, 1.08647413, 0.82492495, 1.2986364 , 1.2986364 ,
        1.22791564, 1.2986364 , 1.08647413, 1.2986364 , 1.15719488,
        1.15719488, 1.22791564, 1.15719488, 1.2986364 , 1.2986364 ,
        1.15719488, 1.15719488, 1.2986364 , 0.75420419, 1.2986364 ,
        1.2986364 , 1.15719488, 1.08647413, 1.08647413, 0.75420419,
        0.82492495, 1.08647413, 1.08647413, 1.15719488, 1.08647413,
        1.22791564, 1.15719488, 1.15719488, 1.22791564, 1.08647413,
        1.2986364 , 1.22791564, 1.15719488, 0.82492495, 1.08647413,
        1.15719488, 1.15719488, 1.2986364 , 1.15719488, 1.08647413,
        1.08647413, 1.22791564, 1.08647413, 1.15719488, 0.89564571,
        1.2986364 , 0.75420419, 1.22791564, 1.22791564, 1.08647413,
        0.89564571, 0.68348343, 1.15719488, 1.08647413, 1.22791564,
        1.22791564, 1.22791564, 1.15719488, 1.2986364 , 1.22791564,
        1.2986364 , 1.2986364 , 0.75420419, 0.68348343, 1.08647413,
        1.08647413, 1.15719488, 0.68348343, 1.22791564, 0.82492495,
        1.2986364 ]])

```

Figura N°26: Array de las predicciones que se obtuvieron de las X\_test  
Fuente: Elaboración Propia

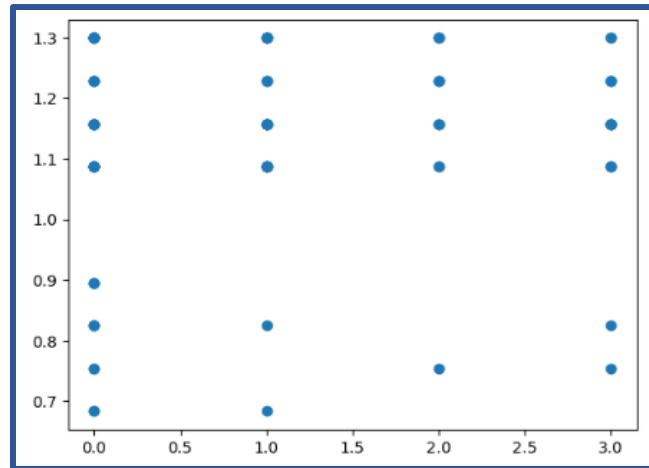


Figura N°27: Gráfico de Regresión Lineal con la (Y\_test, predicciones)  
Fuente: Elaboración Propia

Métricas	Valores
Media del valor absoluto de los errores (MAE)	0.90
Media de los errores al cuadrado (MSE)	1.17
Raíz cuadrada de la media de los errores al cuadrado (RMSE)	1.08

Tabla N°16: Métricas de errores de las (Y\_test, predicciones)  
Fuente: Elaboración Propia

El gráfico de regresión lineal en la dimensión del área conductual siendo el indicador “autocrítica”, se obtuvieron el puntaje lineal para la MAE el 0.90, MSE 1.17, RMSE 1.08; siendo que la media del valor absoluto de los errores (MAE) tiene mayor capacidad de pronóstico, eficacia y un buen modelo para la predicción que las métricas de MSE y RMSE.

La media de los errores al cuadrado (MSE) para este modelo de pronóstico no es eficaz para predecir los síntomas de la depresión.

### **Dificultad de concentración**

X = edad, género

Y= dificultad\_concentracion

Se divide los datos para el train 70% y el test 30%. Obteniendo los datos aleatoriamente, 210 para el train y 91 para el test.

X_TRAIN			Y_TRAIN	
#	Género	Edad	#	Dificultad de concentracion
185	1	3	185	1
72	2	2	72	1
15	1	3	15	2
10	2	4	10	1
196	1	4	196	0

Tabla N°17: Se muestra algunos datos de entrenamiento (TRAIN) para los ejes "X" y "Y"

Fuente: Elaboración Propia

X_TEST			Y_TEST	
#	Género	Edad	#	Dificultad de concentracion
177	1	4	177	0
289	1	1	289	3
228	1	1	228	1
198	1	4	198	3
60	1	2	60	1

Tabla N°18: Se muestra algunos datos de test (TEST) para los ejes "X" y "Y"

Fuente: Elaboración Propia

## Evaluación del modelo

Predicción: Se tomo la X\_test para la predicción.

```

predicciones
array([[1.34307332, 1.03708958, 1.03708958, 1.34307332, 1.13908416,
1.34307332, 1.34307332, 1.03708958, 1.03708958, 1.13908416,
0.65147245, 1.13908416, 0.54947787, 1.13908416, 1.03708958,
1.03708958, 1.03708958, 0.54947787, 1.34307332, 1.34307332,
1.24107874, 1.34307332, 1.03708958, 1.34307332, 1.13908416,
1.13908416, 1.24107874, 1.13908416, 1.34307332, 1.34307332,
1.13908416, 1.13908416, 1.34307332, 0.44748329, 1.34307332,
1.34307332, 1.13908416, 1.03708958, 1.03708958, 0.44748329,
0.54947787, 1.03708958, 1.03708958, 1.13908416, 1.03708958,
1.24107874, 1.13908416, 1.13908416, 1.24107874, 1.03708958,
1.34307332, 0.44748329, 1.24107874, 1.24107874, 1.03708958,
0.65147245, 0.34548871, 1.13908416, 1.03708958, 1.24107874,
1.24107874, 1.24107874, 1.13908416, 1.34307332, 1.24107874,
1.34307332, 1.34307332, 0.44748329, 0.34548871, 1.03708958,
1.03708958, 1.13908416, 0.34548871, 1.24107874, 0.54947787,
1.34307332])

```

Figura N°28: Array de las predicciones que se obtuvieron de las X\_test

Fuente: Elaboración Propia

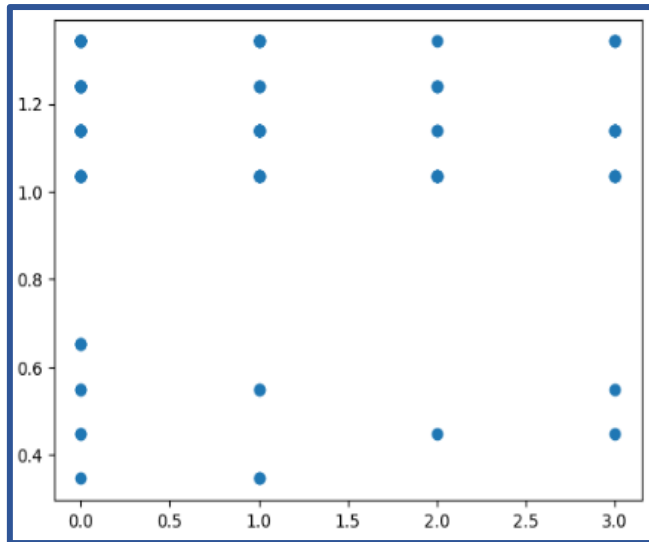


Figura N°29: Gráfico de Regresión Lineal con la (Y\_test, predicciones)  
Fuente: Elaboración Propia

Métricas	Valores
Media del valor absoluto de los errores (MAE)	0.93
Media de los errores al cuadrado (MSE)	1.23
Raíz cuadrada de la media de los errores al cuadrado (RMSE)	1.11

Tabla N°19: Métricas de errores de las (Y\_test, predicciones)  
Fuente: Elaboración Propia

El gráfico de regresión lineal en la dimensión del área conductual siendo el indicador “dificultad de concentración”, se obtuvieron el puntaje lineal para la MAE el 0.93, MSE 1.23, RMSE 1.11; siendo que la media del valor absoluto de los errores (MAE) tiene mayor capacidad de pronóstico, eficacia y un buen modelo para la predicción que las métricas MSE y RMSE.

La media de los errores al cuadrado (MSE) para este modelo de pronóstico no es eficaz para la predicción de los síntomas de la depresión.

Por último, los resultados en general de las métricas de errores para la MAE, MSE, RMSE son de 0 al infinito, para nuestra investigación de la predicción en la depresión usando el algoritmo de regresión lineal, muestra que es un modelo eficaz ya que los valores están cerca de 0 y el error es casi mínimo.

**Árbol de decisión:** Para el análisis tomaremos los siguientes datos:

Género	Sin depresión	Posible depresión
Hombres	21	21
Mujeres	63	196

Tabla N°20: Cantidad de personas que no tienen depresión y que presentan síntomas de depresión  
Fuente: Elaboración Propia

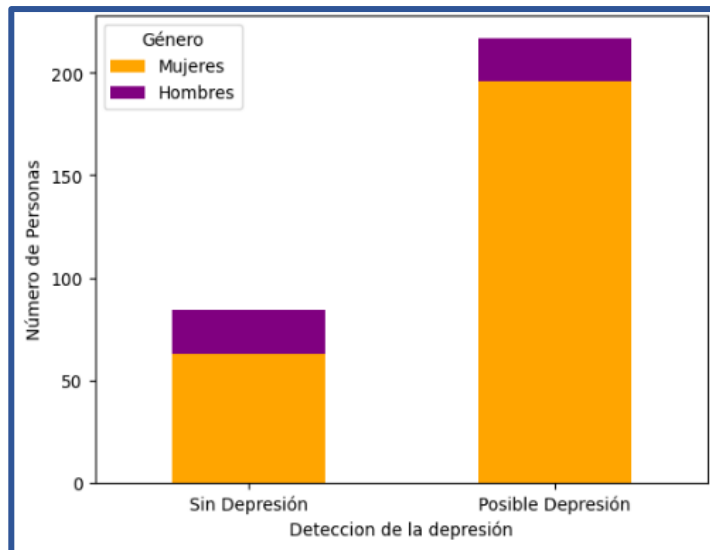


Figura N°30: Proporción de la detección de depresión en mujeres y hombres  
Fuente: Elaboración Propia

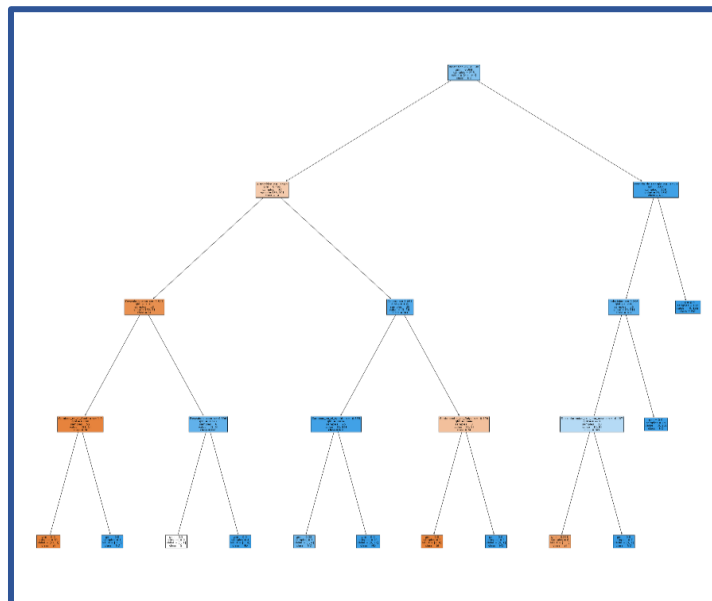


Figura N°31: Árbol de decisión para la detección de la depresión en mujeres y hombres  
Fuente: Elaboración Propia

## Segmentación de datos

En el proceso de entrenamiento de un modelo de aprendizaje automático, es importante dividir nuestros datos en conjuntos separados para diferentes propósitos. Primero, dividimos nuestros datos en tres conjuntos principales:

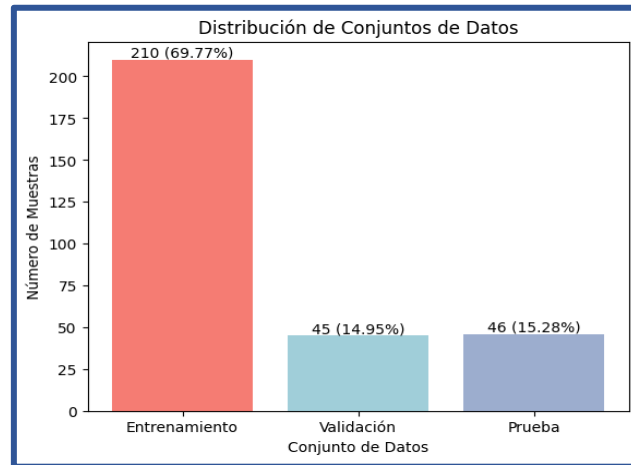


Figura N°32: Distribución de conjunto de datos  
(Entrenamiento - Validación - Prueba)  
Fuente: Elaboración Propia

## Correlación

La correlación entre características y la variable de depresión es una medida estadística que indica la fuerza y la dirección de la relación lineal entre cada característica (o variable independiente) y la variable de depresión (o variable dependiente). Observando la Grafica X, se puede percibir correlación positiva entre las diferentes características.

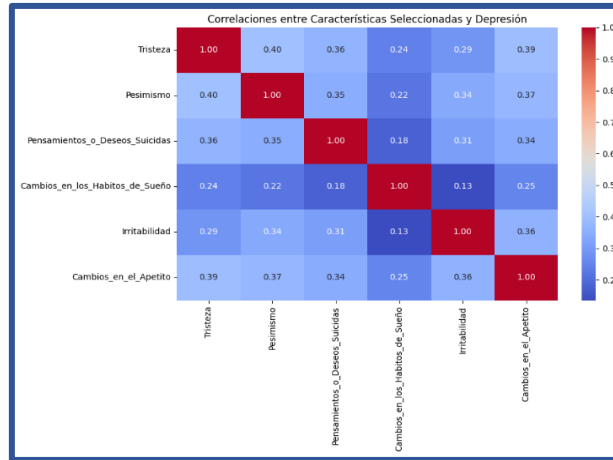


Figura N°33: Correlaciones entre características seleccionadas y depresión  
Fuente: Elaboración Propia

### Curva ROC

El resultado de Curva ROC, según la gráfica "X" es de 0.97, teniendo un valor cercano al 1, indicando un excelente poder discriminativo, necesario para predecir de manera correcta las clases positivas y negativas.

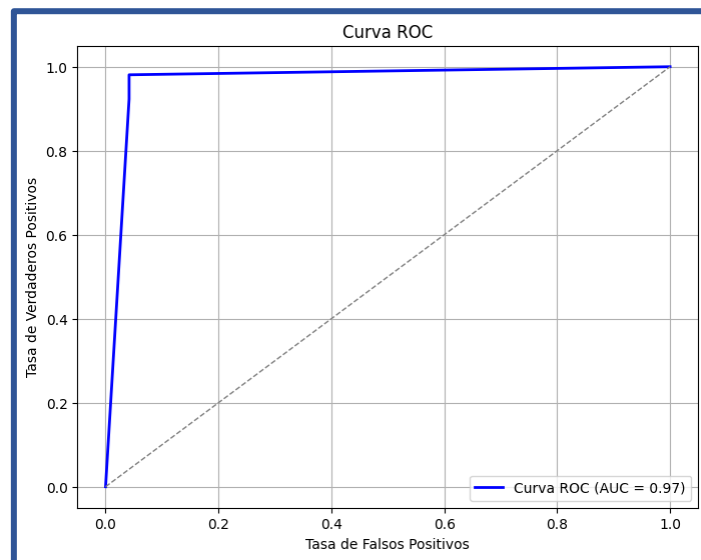


Figura N°34: Curva ROC – Tasa de Falsos Positivos  
Fuente: Elaboración Propia



## Matriz de confusión

**Verdaderos Positivos (TP):** Estos son los casos en los que el modelo predijo correctamente que una persona tiene depresión y en realidad tiene depresión. En este caso, hay 23 casos donde el modelo clasificó correctamente a personas con depresión como positivos.

**Falsos Positivos (FP):** Estos son los casos en los que el modelo predijo incorrectamente que una persona tiene depresión cuando en realidad no la tiene. En este caso, hay 1 caso donde el modelo clasificó erróneamente a una persona sin depresión como positiva.

**Falsos Negativos (FN):** Estos son los casos en los que el modelo predijo incorrectamente que una persona no tiene depresión cuando en realidad la tiene. En este caso, hay 1 caso donde el modelo clasificó erróneamente a una persona con depresión como negativa.

**Verdaderos Negativos (TN):** Estos son los casos en los que el modelo predijo correctamente que una persona no tiene depresión y en realidad no la tiene. En este caso, hay 51 casos donde el modelo clasificó correctamente a personas sin depresión como negativas.

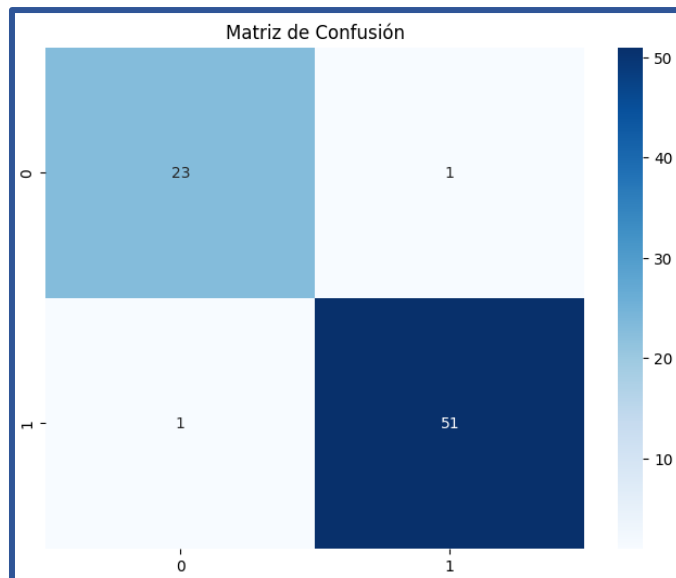


Figura N°35: Matriz de Confusión de valores de predicción  
Fuente: Elaboración Propia

## **IV. DISCUSIÓN**

En este capítulo se realiza el consolidado de los resultados en base a los antecedentes, teorías relacionadas y cuestionarios realizados determinando las diferencias para la investigación.

En primer lugar, el artículo “Machine learning-based predictive modeling of depression in hypertensive populations” de Chiyong L. et al (2022), su objetivo de estudio es el desarrollo de modelos predictivos basados en ML para la depresión en personas con hipertensión, usando seis diferentes algoritmos de ML, siendo ANN, random forest, AdaBoost, stochastic gradient boosting, XGBoost y SVM. En comparación con el estudio que estamos realizando el cual tiene el mismo fin de detectar la depresión usando modelos de ML, pero reduciendo en medida los algoritmos usados, los cuales se usaron 2 algoritmos, Regresión Lineal y Árbol de decisión.

La población que usaron fue de 8,628 adultos mayores de 40 años, una población más amplia comparada a la población usada en nuestra investigación de solo 1360 con una muestra de 301 personas en el rango de 20 a 60 años. Y la herramienta usada para la recolección de datos ha sido el uso de cuestionarios de detección de depresión, específicamente el Patient Health Questionnaire-9 (PHQ-9), siendo usado la misma herramienta para la recolección de datos, pero haciendo uso del cuestionario de Inventario de Depresión de Beck-II (BDI-II).

Para la curva AUC obtenido en el artículo de los modelos entrenados fue de ANN = 0.81, random forest = 0.77, AdaBoost = 0.76, stochastic gradient boosting = 0.803, XGBoost = 0.808 y SVM = 0.760, mientras en la realización del Árbol de decisión se obtuvo el 0.97, siendo mayor a los modelos anteriormente dicho.

Se demostró en el estudio del artículo que el uso de los modelos Machine Learning desarrollados pueden ayudar a la detección de la depresión. Siendo el modelo SVM la herramienta con la precisión más alta, con un 97% de precisión. Apoyando a nuestra investigación en el uso de modelos de aprendizaje automatizados son confiables en la

detección de la depresión, demostrando una precisión del 97%.

En segundo lugar, el artículo “Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders” realizado por Thalia Ritcher et al. (2021), cuenta con el objetivo de desarrollar un sistema de diagnóstico usando machine learning para diferenciar y detectar a las personas con ansiedad y depresión. Teniendo una similitud al objetivo de nuestra investigación, ambos estudios tienen como meta mejorar la identificación y atención de los trastornos mentales mediante el uso de herramientas tecnológicas avanzadas, diferenciando en el enfoque dado.

La población fue 111 participantes con diagnóstico de depresión y/o ansiedad los cuales se tomó una muestra inicial de 101 pacientes y recibieron una compensación monetaria, mientras en nuestra investigación tuvo una población de 1360 con una muestra de 301 pobladores y el tiempo de seguimiento fue de menos de un año y se hizo un muestreo no probabilístico por conveniencia.

La herramienta que se usó para recolectar datos en la investigación fue el inventario de Depresión de Beck-II (BDI-II) para luego apreciar la gravedad de la depresión. Mientras que en la investigación propia se hizo el uso del mismo cuestionario. Ambos cuestionarios han sido validados para permitir detectar y medir los síntomas depresivos.

En conclusión, la precisión de detección en el grupo de depresión fue del 66.46% según el algoritmo de ML que se utilizó en el estudio usando Random Forest, mientras en nuestro estudio se obtuvo un 97% de precisión con Árbol de decisión, pero en comparación con la técnica de Machine Learning del primer estudio es el Árbol de decisión es más sencillo y fácil de usar, mientras que el Random Forest ofrece un modelo más robusto, pero más complicado de interpretar.

En tercer lugar, en el artículo “Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA).” Realizado por Klaas J Wardenaar et al.

(2021) tiene como objetivo identificar las principales determinantes entre la depresión y ansiedad tanto puros como combinados, en comparación con el estudio realizado con el fin de identificar y usar las determinantes para detectar la depresión. La población del estudio fue de 2981 pacientes con diagnóstico de depresión y/o ansiedad, con una edad entre los 18 y 65 años, obteniendo una muestra de 1701, realizando un seguimiento de 2, 4, 6 y 9 años. Mientras que en nuestro estudio la población fue de 1360 con una muestra de 301 pobladores de unos 20 a 60 años, realizando un seguimiento menor a la de un año. La herramienta de recolección de datos fue de entrevista del tipo estructurada, mientras que en la propia se hizo el uso de cuestionarios.

Se usó el enfoque Superlearner, el cual se basa en el uso de varios Machine Learner, usando 6 tipos de ML, los cuales demostró su utilidad para predecir pacientes con depresión y ansiedad, logrando correlaciones significativas entre 0.41 y 0.91. En cambio, en la investigación realizada se hizo el uso de técnicas de Machine Learning como Regresión Lineal y Árbol de decisión donde se obtuvo correlaciones significativas entre las variables predictoras y la predicción obtuvo de 0.29 y 0.40.

Por otro parte, en el estudio “Living arrangements and depression of the older adults—evidence from the Chinese longitudinal healthy longevity survey”. Desarrollado por Qingwen Jia et al. (2023) estudio la relación de la depresión de los adultos mayores y su condición de vida. Teniendo la única similitud en el tema de depresión y la manera de detección de esta.

La importancia principal de la diferencia empieza en la dimensión de la muestra y los habitantes estudiados. Mientras que el estudio de Qingwen Jia et al. incluyó a 15,874 adultos mayores chinos, nuestro estudio se centró en una muestra más pequeña de 301 pobladores de 20 a 60 años. La dimensión de la muestra y sus edades de los participantes puede influir en la generalización de resultados a diferentes sectores de población. En ambos casos la herramienta de recolección de datos fueron el uso de

cuestionarios.

Otra diferencia importante se encuentra en las técnicas de análisis utilizadas. Mientras que el estudio de Qingwen Jia et al. empleó regresión logística binaria y regresión lineal múltiple, nuestro estudio se enfocó en la regresión lineal bajo el modelo de Machine Learning. Ambos planteamientos consideran sus ventajas y desventajas, y la selección de la técnica de análisis puede depender de los objetivos específicos del aprendizaje y los tipos de datos.

En discusión, al comparar los resultados y conclusiones de ambos estudios, es crucial tener en cuenta estas diferencias en el diseño del estudio, la población estudiada y las técnicas de análisis utilizadas. Esto nos ayuda a contextualizar e integrar las afinidades y disimilitud en los hallazgos, proporcionando una perspectiva más completa sobre el tema de estudio.

## **V. CONCLUSIONES**

Concluimos que la confiabilidad de las encuestas mediante la recopilación de artículos e investigaciones, que indican un alfa de Cronbach = 0,91 en la población usada en la misma. Otorgando la confiabilidad para su uso en la recopilación de datos necesario para el entrenamiento de Machine Learning usado.

Concluimos que la precisión del sistema de pronóstico usando Machine Learning para la detección de la depresión en personas adultas en el condominio Ciudad Verde de manera pre-experimental, obteniendo una precisión del 97% de confiabilidad dentro de los algoritmos usados como regresión lineal y árbol de decisiones.

El sistema de pronóstico demostró ser eficaz en la detección de la presencia de síntomas afectivos de la depresión en personas adultas. La precisión del 97% sugiere que los síntomas afectivos pueden ser identificados de manera confiable mediante el uso de árboles de decisión. De manera similar, la eficacia del sistema en la detección de la presencia de síntomas conductuales de la depresión también fue alta. La comparación entre la precisión de los árboles de decisión y la regresión lineal destaca la superioridad del primero para este propósito.

Concluimos que la eficacia del sistema de pronóstico para la detención de la depresión que realizamos es preexperimental; obteniendo los resultados de los errores de los modelos con alta confiabilidad siendo así menor a 2, considerando que para los modelos de métrica de errores tienen como parámetro de 0 al infinito.



## **VI. RECOMENDACIONES**

Al término de nuestra investigación nosotros recomendamos el uso del sistema de pronóstico usando técnicas de machine learning para la detección de la depresión en personas adultas, por la cual nuestra referencia es muy buena y nos permite a través de ella y sus algoritmos realizar toma de decisiones, análisis de estudio y llegar a brindar la solución a la problemática; los algoritmos y técnicas que empleamos son regresión lineal y árbol de decisiones. Esta última técnica nos favorece en los resultados, obteniendo una mejor predicción usando la matriz de confusión obteniendo como objeto de estudio la población.

También consideramos y proponemos la continuidad del tema de investigación a futuro. Nuestras recomendaciones son las siguientes:

Como primer requerimiento es utilizar mayores conjuntos de datos y mayores rangos de fecha para la mejora en el entrenamiento; con el fin de mejorar los resultados de la evaluación de pronósticos.

Recomendamos aplicar en estudios del área de salud, para validar la efectividad de las técnicas de machine learning para la detección de la depresión usando diferentes cuestionarios, la cual nosotros aplicamos el Inventario de Depresión de Beck-II (BDI-II) para nuestro sistema de pronóstico no clínico. También las personas que se dedican o quieran dedicarse a la investigación pueden aplicar las técnicas de machine learning investigando a detalle la enfermedad de la depresión, aportando para el beneficio de la población.

Con mira a proponer la continuidad del tema de investigación realizado para estudios e investigaciones futuras se detalla las siguientes recomendaciones:

Se recomienda el uso de más técnicas de machine learning, para mejorar la precisión y confiabilidad de los resultados obtenidas, entre los modelos recomendados, se indica el uso de Random Forest, ofreciendo un buen desempeño y estabilidad. Otro modelo recomendado sería el de redes

neuronales por su escalabilidad y flexibilidad.

Se recomienda la recopilación de una mayor cantidad de datos, para mejorar el entrenamiento y precisión del pronóstico.

Mayor tiempo de seguimiento a los participantes de la investigación, para poder realizar una mejor validación de los resultados, considerando también estudios longitudinales para observar los cambios y tendencias en los datos.

Además del uso de encuestas, como la usadas para esta investigación, se recomienda el uso de otras herramientas como entrevistas y recopilación de datos clínicos de los participantes

En nuestra investigación, por último, se pueden aplicar a futuro otras técnicas de machine learning, ya que en estos tiempos nos encontramos en constante cambios y avance en la tecnología, donde se busca optimizar recursos y brindar utilidad.

## REFERENCIAS

Alocución de apertura del Director General de la OMS en el panel sobre salud mental en el trabajo celebrado en el marco del Foro Económico Mundial - 18 de enero de 2023. World Health Organization (WHO) [en línea]. 18 de enero de 2023 [consultado el 21 de noviembre de 2023]. Disponible en: <https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-mental-health-at-work-panel--world-economic-forum---18-january-2023>

ACHARYA, Shwetha. What are RMSE and MAE? Towards Data Science [en línea]. 13 de mayo de 2021 [consultado el 8 de octubre de 2023]. Disponible en: <https://towardsdatascience.com/what-are-rmse-and-mae-e405ce230383>

ARANDA B., Carolina y Manuel PANDO M. Conceptualización del apoyo social y las redes de apoyo social. Revista de Investigación en Psicología [en línea]. 2014, 16(1), 233 [consultado el 21 de noviembre de 2023]. ISSN 1609-7475. Disponible en: [doi:10.15381/rinvp.v16i1.3929](https://doi.org/10.15381/rinvp.v16i1.3929)

AYELE, Workneh Y. Adapting CRISP-DM for Idea Mining: A Data Mining Process for Generating Ideas Using a Textual Dataset. International journal of advanced computer science & applications [online]. 2020, 11(6), 20-. ISSN 2158-107X. Dostupné z: [doi:10.14569/IJACSA.2020.0110603](https://doi.org/10.14569/IJACSA.2020.0110603)

BELLOCH, Sandin Ramos. Manual de psicopatología - volumen 1. McGraw-Hill Interamericana, 1995. ISBN 9788448118341.

BELLOCH, Sandin Ramos. Manual de psicopatología - volumen 2. McGraw-Hill Interamericana, 1995. ISBN 9788448117764.

CHAPPLE, Mike y Fred NWANGANGA. Practical machine learning in R. Wiley & Sons, Limited, John, 2020. ISBN 9781119591542.

Depresión. National Institute of Mental Health (NIMH) [en línea]. 30 de junio de 2021 [consultado el 21 de noviembre de 2023]. Disponible en: <https://www.nimh.nih.gov/health/publications/espanol/depression-sp>

Depresión. World Health Organization (WHO) [en línea]. [sin fecha] [consultado el 19 de noviembre de 2023]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/depression>

DÍAZ-MARTÍNEZ, Marco Antonio, María De los Ángeles AHUMADA-CERVANTES a Julia Patricia MELO-MORÍN. Árboles de Decisión como Metodología para Determinar el Rendimiento Académico en Educación Superior. Revista lasallista de investigación [online]. 2021, 18(2), 94–104. ISSN 1794-4449. Dostupné z: doi:10.22507/rli.v18n2a8

DURAN, Farfán Pablo Albert. Estrés y depresión en estudiantes de medicina de una universidad pública de Lima Metropolitana. En: Cybertesis UNMSM [base de datos en línea]. Bachelor's thesis, Universidad Nacional Mayor de San Marcos, 2021 [consultado el 3 de diciembre de 2023]. Disponible en: <https://hdl.handle.net/20.500.12672/16964>

HERNÁNDEZ SAMPIERI, Roberto, Carlos FERNÁNDEZ COLLADO y Pilar BAPTISTA LUCIO. Metodología de la investigación. 3a ed. México: McGraw Hill, 2004. ISBN 9789701036327.

HERNÁNDEZ SAMPIERI, Roberto, Carlos FERNÁNDEZ COLLADO y Pilar BAPTISTA LUCIO. Metodología de la investigación. 6a ed. México: McGraw Hill, 2004. ISBN 9781456223960.

HU, Rongyao et al. Robust SVM with adaptive graph learning. World Wide Web [en línea]. 2019, 23(3), 1945–1968 [consultado el 21 de noviembre de 2023]. ISSN 1573-1413. Disponible en: doi:10.1007/s11280-019-00766-x

HYNDMAN, Rob J. y George ATHANASOPOULOS. Forecasting: principles and practice. OTexts, 2018. ISBN 9780987507112.

HYNDMAN, Rob J. y George ATHANASOPOULOS. Forecasting: principles and practice. OTexts, 2021. ISBN 9780987507136.

HYNDMAN, Rob y George ATHANASOPOULOS. Forecasting: Principles and Practice [en línea]. 3a ed. Australia: OTexts, 2021 [consultado el 19 de noviembre de 2023]. ISBN 0987507133. Disponible en: <https://otexts.com/fpp3/index.html>

IBM. Decision trees [en línea]. [s.f.]. [Consultado el 29 de mayo de 2024]. Disponible en: <https://www.ibm.com/es-es/topics/decision-trees>

JARA, César. Redes sociales de apoyo social y salud mental. Cuadernos Médico Sociales [en línea]. 2021, 61(1), 49–53 [consultado el 21 de noviembre de 2023]. ISSN 2735-7759. Disponible en: [doi:10.56116/cms.v61.n1.2021.24](https://doi.org/10.56116/cms.v61.n1.2021.24)

JIA, Qingwen, Yanhan DUAN, Rui GONG, Meijun JIANG, Dianping YOU et al. Living arrangements and depression of the older adults— evidence from the Chinese longitudinal healthy longevity survey. En línea. BMC Public Health, vol. 23 (septiembre de 2023), n.º 1. ISSN 1471-2458. Disponible en: <https://doi.org/10.1186/s12889-023-16730-4>. [consultado el 10/11/2023].

La depresión: un trastorno de salud mental que también afecta a niñas, niños y adolescentes, y requiere la participación de la familia. Plataforma digital única del Estado Peruano [en línea]. 19 de enero de 2023 [consultado el 19 de noviembre de 2023]. Disponible en: <https://www.gob.pe/institucion/minsa/noticias/690010-la-depresion-un-trastorno-de-salud-mental-que-tambien-afecta-a-ninas-ninos-y-adolescentes-y-requiere-la-participacion-de-la-familia>

La evaluación basada en el Factor de Impacto de la revista pone en jaque los objetivos

de la investigación. <https://www.redalyc.org/articulo.oa?id=10458194011> [en línea]. 2 de julio de 2019 [consultado el 21 de noviembre de 2023]. Disponible en: <https://www.redalyc.org/articulo.oa?id=10458194011>

LEE, Chiyong y Heewon KIM. Machine learning-based predictive modeling of depression in hypertensive populations. PLOS ONE [en línea]. 2022, 17(7), e0272330 [consultado el 21 de noviembre de 2023]. ISSN 1932-6203. Disponible en: doi:10.1371/journal.pone.0272330

LEE, Ji-Yoon, Doyeon WON y Kiheon LEE. Machine learning-based identification and related features of depression in patients with diabetes mellitus based on the Korea National Health and Nutrition Examination Survey: a cross-sectional study. PLOS ONE [en línea]. 2023, 18(7), e0288648 [consultado el 21 de noviembre de 2023]. ISSN 1932-6203. Disponible en: doi:10.1371/journal.pone.0288648

MAKRIDAKIS, Spyros G. Forecasting: methods and applications. 3a ed. New York: John Wiley & Sons, 1998. ISBN 0471532339.

MALDONADO-AVENDAÑO, Natalia, Rubby CASTRO-OSORIO y Pilar CARDONA-GÓMEZ. Propiedades psicométricas del Inventario de Depresión de Beck-II (BDI-II) en población universitaria colombiana. En línea. Revista Colombiana de Psiquiatría, noviembre de 2021. ISSN 0034-7450. Disponible en: <https://doi.org/10.1016/j.rcp.2021.08.007>. [consultado el 28/11/2023].

MALHOTRA, Anshu y Rajni JINDAL. Deep learning techniques for suicide and depression detection from online social media: a scoping review. En línea. Applied Soft Computing, octubre de 2022, p. 109713. ISSN 1568-4946. Disponible en: <https://doi.org/10.1016/j.asoc.2022.109713>. [consultado el 28/11/2023].

Metodología de la investigación: cuantitativa, cualitativa y redacción de la tesis - 5.

edicion. Ediciones de la U, 2018. ISBN 9789587628760.

RICHTER, Thalia et al. Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research* [en línea]. 2021, 141, 199–205 [consultado el 21 de noviembre de 2023]. ISSN 0022-3956. Disponible en: doi:10.1016/j.jpsychires.2021.06.044

ROJAS MONTES, Enrique. *Psicopatología de la depresión*. Barcelona: Salvat, 1981. ISBN 84-345-2093-1.

SUN, Chenjian et al. A novel study for depression detecting using audio signals based on graph neural network. *Biomedical Signal Processing and Control* [en línea]. 2024, 88, 105675 [consultado el 21 de noviembre de 2023]. ISSN 1746-8094. Disponible en: doi:10.1016/j.bspc.2023.105675

TIAN, Gang et al. Association between disability, social support and depressive symptoms in Chinese older adults: a national study. *Frontiers in Public Health* [en línea]. 2022, 10 [consultado el 21 de noviembre de 2023]. ISSN 2296-2565. Disponible en: doi:10.3389/fpubh.2022.980465

WANG, Cancan et al. A degradation condition assessment technique based on online\_isse degradation feature and logistic regression model. *The International Journal of Acoustics and Vibration* [en línea]. 2021, 26(3), 202–211 [consultado el 21 de noviembre de 2023]. Disponible en: doi:10.20855/ijav.2021.26.31738

WARDENAAR, Klaas J. et al. Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the netherlands study of depression and anxiety (NESDA). *Journal of Affective Disorders* [en línea]. 2021, 293, 295–304 [consultado el 21 de noviembre de 2023]. ISSN 0165-0327. Disponible en: doi:10.1016/j.jad.2021.06.029



## ANEXOS

### Anexo 1. Tabla de operacionalización de variables

VARIABLES DE ESTUDIO	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIÓN	INDICADORES	ESCALA DE MEDICIÓN
<b>Independiente:</b> Sistema de pronóstico	Maya (2022) destaca que los métodos de pronóstico deben usarse para objetivos, mientras que los métodos de predicción son adecuados para subjetivos. Enfatiza que los investigadores actuales se esfuerzan por lograr los mejores resultados en problemas de pronóstico, y señala que para pronósticos a corto o largo plazo se debe entrenar un modelo neuronal de Machine Learning (ML).	Evaluación mediante modelos de Machine Learning entrenados con datos obtenidos mediante cuestionarios.			
<b>Dependiente:</b> Detección de la depresión	Instituto Nacional de la Salud Mental (2021), informa que, para diagnosticar la depresión, la persona debe tener 5 síntomas de depresión todos los días, durante 2 semanas.	Evaluación utilizando un sistema de Machine Learning entrenado con datos de cuestionarios validados (como el Inventario de Depresión de Beck - BDI) para detectar la depresión.	Área afectiva  Área conductual	<ul style="list-style-type: none"> <li>• Agitación</li> <li>• Pensamientos o deseos suicidas</li> <li>• Autocrítica</li> <li>• Dificultad de concentración</li> </ul>	De razón

## Anexo 2. Instrumentos de recolección de datos

La presente encuesta es para recolectar información de los usuarios del Condominio Ciudad Verde, la cual está dividida en dos secciones.

1. Información Demográfica
2. Inventario de Depresión de Beck-II (BDI-II), consta de 21 preguntas.

### URL:

[https://docs.google.com/forms/d/e/1FAIpQLSc5LZrZqxBz7zTyvO1ZH6lvj3pkMq8owsS2t-0UXURCBn-lpg/viewform?usp=sf\\_link](https://docs.google.com/forms/d/e/1FAIpQLSc5LZrZqxBz7zTyvO1ZH6lvj3pkMq8owsS2t-0UXURCBn-lpg/viewform?usp=sf_link)



**TEST**


Estimados participantes del Condominio Residencial Ciudad Verde, tenemos el agrado de dirigimos a ustedes por este medio, somos estudiantes de la carrera profesional de Ingeniería de Sistemas de la Universidad César Vallejo, nuestros nombres son: Barriga Moreno Elisabeth y Muñoz Avalos Flavio Stephano. Actualmente, estamos realizando un proyecto de investigación en la cual presentamos un pre-test para medir los niveles de Depresión.

En caso de no comprender algún enunciado o pregunta en este formulario, por favor comunicarse inmediatamente con los encargados.

Por confiabilidad y estándares de las Normas ISOS la información es reservada, solo para fines de la investigación respectiva.

Finalmente, se agradece su participación e interés.

### Anexo 3. Reporte de similitud en software Turnitin



## Recibo digital

Este recibo confirma que su trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.

La primera página de tus entregas se muestra abajo.

Autor de la entrega:	ELISBETH BARRIGA MORENO
Título del ejercicio:	Turnitin 2
Título de la entrega:	INTRODUCCIÓN_turnitin 01 de junio.docx
Nombre del archivo:	INTRODUCCIÓN_turnitin_01_de_junio.docx
Tamaño del archivo:	881.15K
Total páginas:	43
Total de palabras:	8,392
Total de caracteres:	44,602
Fecha de entrega:	01-jun.-2024 08:02p. m. (UTC-0500)
Identificador de la entrega:	2393324935

**INTRODUCCIÓN**

Este es un documento de introducción al curso de la asignatura de estadística. El texto describe los objetivos del curso, el contenido de los capítulos y el formato de las entregas. También se mencionan los recursos de apoyo y el contacto de los docentes.

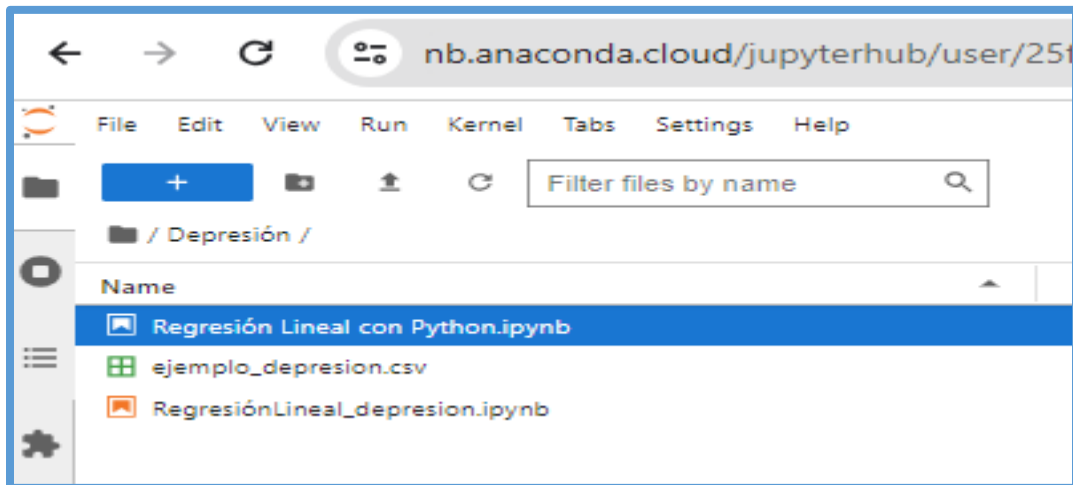
El curso está diseñado para proporcionar a los estudiantes una comprensión sólida de los conceptos estadísticos y su aplicación en el mundo real. El curso se divide en tres secciones principales: estadística descriptiva, inferencia estadística y estadística aplicada. Cada sección incluye teoría, ejemplos y ejercicios prácticos.

Los estudiantes deben completar las entregas de manera puntual y seguir las instrucciones de formato. El profesor se reserva el derecho de modificar el contenido del curso sin previo aviso.

Derechos de autor 2024 Turnitin. Todos los derechos reservados.

## Anexo 4. Análisis complementario

### Back-end: Regresión Lineal



```
Machine Learning

[38]: from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LinearRegression

[40]: depre.columns

[40]: Index(['edad', 'genero', 'estado_civil', 'nivel_educacion', 'tristeza',
          'pesimismo', 'fracaso', 'perdida_placer', 'sentimiento_culpa',
          'sentimiento_castigo', 'disconformidad_uno_mismo', 'autocritica',
          'pensamiento_deseos_suicida', 'llanto', 'agitacion', 'perdida_interes',
          'indecision', 'desvalorizacion', 'perdida_energia',
          'cambios_habitos_sueño', 'irritabilidad', 'cambios_apetito',
          'dificultad_concentracion', 'cansancio_fatiga', 'perdida_interessexo',
          'Resultado', 'Valores_Resultado'],
          dtype='object')

[42]: X = depre[['genero', 'edad']]

[44]: Y = depre['agitacion']

[46]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3, random_state=42)

[48]: lrm = LinearRegression()
      lrm.fit(X_train, Y_train)

[48]: LinearRegression
      LinearRegression()
```

## ▼ Evaluamos el modelo

```
[264]: from sklearn import metrics
```

```
[266]: predicciones = lrm.predict(X_test)
```

```
[268]: predicciones
```

```
[268]: array([[1.34307332, 1.03708958, 1.03708958, 1.34307332, 1.13908416,
          1.34307332, 1.34307332, 1.03708958, 1.03708958, 1.13908416,
          0.65147245, 1.13908416, 0.54947787, 1.13908416, 1.03708958,
          1.03708958, 1.03708958, 0.54947787, 1.34307332, 1.34307332,
          1.24107874, 1.34307332, 1.03708958, 1.34307332, 1.13908416,
          1.13908416, 1.24107874, 1.13908416, 1.34307332, 1.34307332,
          1.13908416, 1.13908416, 1.34307332, 0.44748329, 1.34307332,
          1.34307332, 1.13908416, 1.03708958, 1.03708958, 0.44748329,
          0.54947787, 1.03708958, 1.03708958, 1.13908416, 1.03708958,
          1.24107874, 1.13908416, 1.13908416, 1.24107874, 1.03708958,
          1.34307332, 1.24107874, 1.13908416, 0.54947787, 1.03708958,
          1.13908416, 1.13908416, 1.34307332, 1.13908416, 1.03708958,
          1.03708958, 1.24107874, 1.03708958, 1.13908416, 0.65147245,
          1.34307332, 0.44748329, 1.24107874, 1.24107874, 1.03708958,
          0.65147245, 0.34548871, 1.13908416, 1.03708958, 1.24107874,
          1.24107874, 1.24107874, 1.13908416, 1.34307332, 1.24107874,
          1.34307332, 1.34307332, 0.44748329, 0.34548871, 1.03708958,
          1.03708958, 1.13908416, 0.34548871, 1.24107874, 0.54947787,
          1.34307332])
```

## Media del ERROR

```
[88]: metrics.mean_absolute_error(Y_test, predicciones)
```

```
[88]: 1.0759637459053057
```

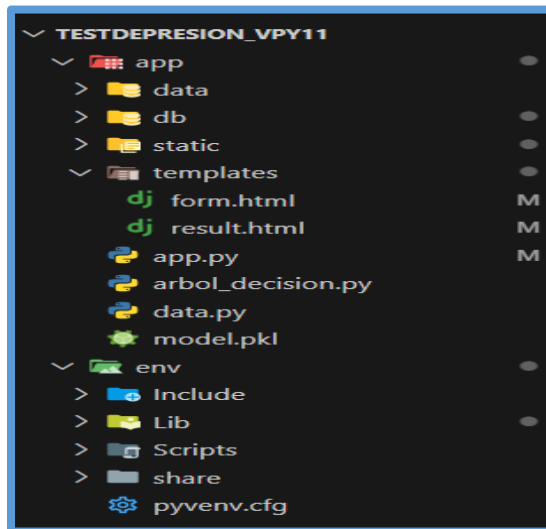
```
[90]: metrics.mean_squared_error(Y_test, predicciones)
```

```
[90]: 1.4910109654462498
```

```
[92]: np.sqrt(metrics.mean_squared_error(Y_test, predicciones))
```

```
[92]: 1.2210695989362153
```

## Back-end: Árbol de decisiones



```
EXPLORER
...
arbol_decision.py M X
app > arbol_decision.py > ...
1 import pandas as pd
2 import numpy as np
3 from sklearn.tree import DecisionTreeClassifier
4 from sklearn.model_selection import train_test_split
5 from matplotlib import pyplot as plt
6 from sklearn import tree
7 from sklearn.preprocessing import StandardScaler
8 from sklearn.metrics import confusion_matrix
9 #Obtener datos del dataset
10 pacientes = pd.read_csv("app\data\Dataset-Depresion.csv")
11 #Variables predictoras
12 X = pacientes.iloc[:,1:24]
13 #Variable a predecir
14 Y = pacientes.iloc[:,25]
15 #X_train y Y_train para entrenamiento
16 #Y_test y Y_test para prueba
17 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, train_size=0.75, random_state=42)
18 #llamamos al constructor del arbol de decision
19 arbol = DecisionTreeClassifier(max_depth=4)#max_depth es la maxima cantidad de ramificaciones que se quieren ver al generar
20 #Entrenamos el modelo
21 arbol_enfermedad = arbol.fit(X_train, Y_train)
22 tree.plot_tree(arbol_enfermedad, feature_names=list(X.columns.values),
23               class_names=list(Y.values), filled=True)
24 #Predict the response for test dataset
25 Y_pred = arbol_enfermedad.predict(X_test)
26 #Creación de la matriz de confusión
27 Matriz_de_confusion = confusion_matrix(Y_test, Y_pred)
28 #Estandarizar los datos
29 #calculamos la precisión
30 sc = StandardScaler()
31 sc.fit(X_train)
32 X_train_std = sc.transform(X_train)
33 X_test_std = sc.transform(X_test)
34 # Ajuste del modelo
35 arbol.fit(X_train_std, Y_train)
```

```

EXPLORER
TESTDEPRESSION_VPY11
  app
  data
  db
  static
  templates
    form.html
    result.html
  app.py
  arbol_decision.py
  data.py
  model.pkl
env
  Include
  Lib
  Scripts
  share
  pyenv.cfg
arbol_decision.py M
data.py M
app > data.py > ...
1 import pandas as pd
2 import numpy as np
3 from sklearn.tree import DecisionTreeClassifier
4 from sklearn.model_selection import train_test_split
5 from matplotlib import pyplot as plt
6 from sklearn import tree
7 from sklearn.preprocessing import StandardScaler
8 import warnings
9 import pickle
10 warnings.filterwarnings("ignore")
11 #Obtener datos del dataset
12 pacientes = pd.read_csv('app\data\Dataset-Depression.csv')
13 #Analisis de predictores
14 X = pacientes.iloc[:,0:22]
15 #Variable a predecir
16 Y = pacientes.iloc[:,24]
17 #X_train y Y_train para entrenamiento
18 #Y_test y Y_test para prueba
19 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, train_size=0.75, random_state=42)
20 arbol = DecisionTreeClassifier(max_depth=4) #max_depth es la maxima cantidad de ramificaciones que se quieren ver al generar la grafi
21 arbol_enfermedad = arbol.fit(X_train, Y_train)
22 #log_reg.fit(X_train, y_train)
23 input=[int(x) for x in "1 1 2 1 1 1 1 0 3 1 2 2 3 3 2 1 1 2 1 1".split(' ')]
24 final=np.array(input)
25 b = arbol.predict_proba(final)
26 pickle.dump(arbol,open('model.pkl','wb'))
27 model=pickle.load(open('model.pkl','rb'))
28 print(b)

```

```

EXPLORER
app.py M X
TESTDEPRESSION_VPY11
  app
  data
  db
  static
  templates
    form.html
    result.html
  app.py
  arbol_decision.py
  data.py
  model.pkl
env
  Include
  Lib
  Scripts
  share
  pyenv.cfg
app.py M X
1 from flask import Flask, render_template, request
2 import pickle
3 import numpy as np
4 import os
5 app = Flask(__name__, static_folder='static')
6 # Obtener la ruta del directorio actual del script
7 dir_path = os.path.dirname(os.path.abspath(__file__))
8 modelo_path = os.path.join(dir_path, 'model.pkl')
9 # Cargar el modelo
10 with open(modelo_path, 'rb') as file:
11     model = pickle.load(file)
12 @app.route('/')
13 def home():
14     return render_template('form.html')
15 @app.route('/predict', methods=['POST'])
16 def predict():
17     # Obtener los datos del formulario
18     data = request.form.to_dict()
19     # Lista de campos que deben ser convertidos a enteros
20     fields_to_convert = [
21         'Genero', 'Tristeza', 'Pesimismo', 'Fracaso', 'Perdida de Placer', 'Sentimientos de culpa',
22         'Sentimientos de Castigo', 'Disconformidad con uno mismo', 'Autocritica',
23         'Pensamientos o Deseos Suicidas', 'Llanto', 'Agitacion', 'Perdida de Interes',
24         'Indecision', 'Desvalorizacion', 'Perdida de Energia', 'Cambios en los Habitos de Sueño',
25         'Irritabilidad', 'Cambios en el Apetito', 'Dificultad de concentracion', 'Cansancio o fatiga',
26         'Perdida de Interes en el Sexo'
27     ]
28     input_data = [int(data[field]) for field in fields_to_convert]
29     final = np.array(input_data)

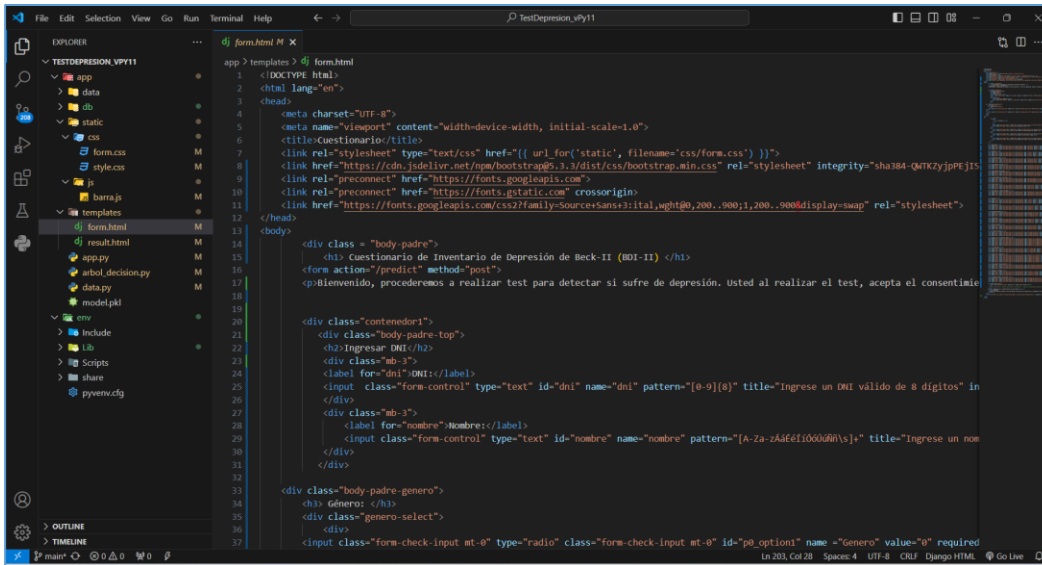
```

```

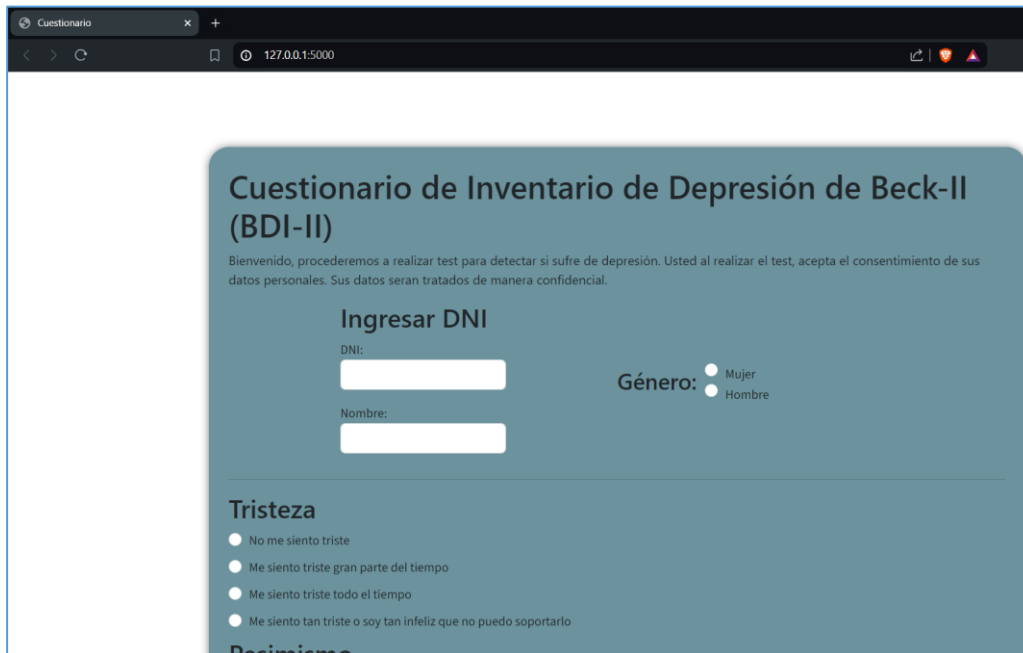
TESTDEPRESSION_VPY11
  app
  data
  db
  static
  templates
    form.html
    result.html
  app.py
  arbol_decision.py
  data.py
  model.pkl
env
  Include
  Lib
  Scripts
  share
  pyenv.cfg
app.py M X
16 def predict():
17     input_data = [int(data[field]) for field in fields_to_convert]
18     final = np.array(input_data)
19     # Hacer la prediccion
20     prediction = model.predict(final)
21     probability = model.predict_proba(final)[1, 1] # Probabilidad de que sea positivo
22     probability = np.array(probability)
23     # Determinar el valor de Depresion
24     # Determinar el valor de depresion valor según la prediccion
25     # Depresion valor = "SI" if prediction != "Depresion minima" else "no"
26     afectivo = ['Tristeza', 'Pesimismo', 'Fracaso', 'Perdida de Placer', 'Sentimientos de culpa', 'Sentimientos de Castigo', 'Disconfor
27     # Inicializar la suma
28     suma_total = 0
29     valor_base = 63
30     # Iterar sobre los campos y sumas sus valores
31     for campo0 in afectivo:
32         suma_total += int(request.form[campo0])
33     porcentaje = (suma_total / valor_base) * 100
34     porcentaje_afectivo = "{:.2f}".format(porcentaje)
35     conductual = ['Cambios en los Habitos de Sueño', 'Cambios en el Apetito', 'Dificultad de concentracion', 'Cansancio o fatiga', 'Per
36     for campo2 in conductual:
37         suma_total += int(request.form[campo2])
38     porcentaje = (suma_total / valor_base) * 100
39     porcentaje_conductual = "{:.2f}".format(porcentaje)
40     # Condicion para determinar si hay depresion minima
41     if prediction == "Depresion minima":
42         result = "Este resultado no es Clinico, sin embargo no presenta sintomas de depresion."
43     elif prediction == "Depresion leve":
44         result = "La probabilidad de tener DEPRESION es Minima, se recomienda acceder a un especialista."
45     elif prediction == "Depresion moderada":
46         result = "La probabilidad de tener DEPRESION es Moderada, se recomienda acceder a un especialista."
47     else:
48         result = "La probabilidad de tener DEPRESION es Alta, este resultado no es Clinico pero se recomienda acceder con prioridad alta
49     probability_porcentaje = probability * 100
50     return render_template('result.html', prediction=result, probability=probability, probability_porcentaje=probability_porcentaje,porce
51     if __name__ == '__main__':
52         app.run(debug=True)

```

## Front-end:



```
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4 <meta charset="UTF-8">
5 <meta name="viewport" content="width=device-width, initial-scale=1.0">
6 <title>Cuestionario</title>
7 <link rel="stylesheet" type="text/css" href="{% url 'static', filename='css/form.css' %}">
8 <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.3/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-QMKZyypPp15"
9 <link rel="preconnect" href="https://fonts.googleapis.com">
10 <link rel="preconnect" href="https://fonts.gstatic.com" crossorigin>
11 <link href="https://fonts.googleapis.com/css2?family=Source+Sans+3:ital,wght@0,200..900;1,200..900&display=swap" rel="stylesheet">
12 </head>
13 <body>
14
15 <div class="body-padre">
16 <h1>Cuestionario de Inventario de Depresión de Beck-II (BDI-II) </h1>
17 <form action="/predict" method="post">
18 <p>Bienvenido, procederemos a realizar test para detectar si sufre de depresión. Usted al realizar el test, acepta el consentimiento
19
20 <div class="contenedor1">
21 <div class="body-padre-top">
22 <div class="mb-3">
23 <input class="form-control" type="text" id="dni" name="dni" pattern="[0-9]{8}" title="Ingrese un DNI válido de 8 dígitos" in
24 </div>
25 <div class="mb-3">
26 <label for="nombre">Nombre:</label>
27 <input class="form-control" type="text" id="nombre" name="nombre" pattern="[A-Za-zÀáÊëÏîÏóÛüñ(s)+]" title="Ingrese un nom
28 </div>
29 <div class="mb-3">
30 <label for="genero">Género:</label>
31 <div class="genero-select">
32 <input class="form-check-input mt-0" type="radio" class="form-check-input mt-0" id="p0_option1" name="Genero" value="0" required
33 </div>
34 </div>
35 </div>
36 </div>
37 </div>
```





## **Metodología CRISP-DM**

### 1. Comprensión del negocio:

En ese sentido, en la fase inicial del proyecto, primero comprendimos la situación que sucede en el Condominio Cabo Verde, para definir las metas y los objetivos con las necesidades del condominio.

El Condominio Ciudad Verde se encuentra afectada por los casos de intentos de suicidio dentro del mismo, informado por los mismos residentes del lugar. Los cuales se les toma poca o nula importancia, por la falta de apoyo a las personas afectadas o programas que ayuden a este problema.

En conocimiento de los hechos, los alumnos de la Universidad Cesar Vallejo, se propone a implementar un proyecto de Machine Learning con el objetivo de detectar la depresión en los residentes. Con el deseo de ayudar a tomar más conciencia del mismo y poder dar el primer paso a que los residentes tomen importancia del tema, pudiendo buscar ayuda por iniciativa propia. Cumpliendo nosotros con el tema de confidencialidad en todo momento, por tratarse de un tema delicado.

### 2. Comprensión de los datos:

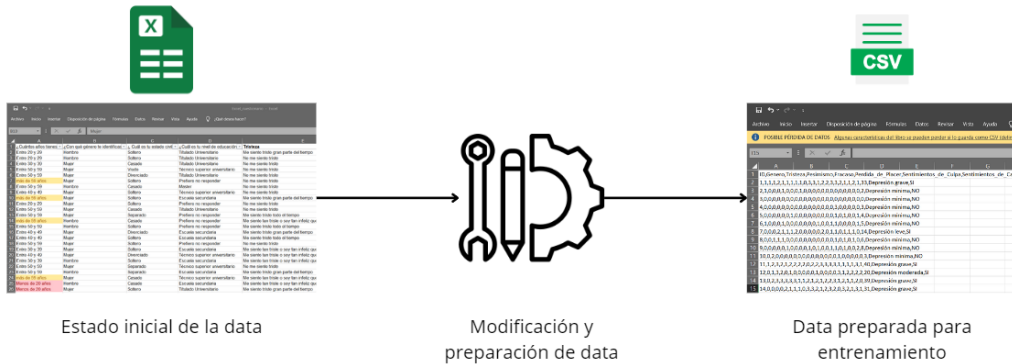
Se llevo a cabo la recolección de datos, primero de investigaciones que abordan del tema sobre la depresión, comprender su validez para poder medir y detectar la misma, teniendo como primera opción el uso de cuestionarios validados por otras investigaciones.

Se recopilaron los datos mediante cuestionarios del Inventario de Depresión de Beck-II realizadas a los residentes del condominio para entender mejor los aspectos afectivos y conductuales de la depresión.



### 3. Preparación de los datos:

Los datos obtenidos, se realiza una verificación para no introducir al entrenamiento del ML valores vacíos o nulos. Se hizo la codificación de variables, es decir, la conversión a valores numéricos para su medición y la normalización de datos para mejorar el rendimiento y precisión de los modelos de ML.



### 4. Modelado:

Se hizo una revisión y elección de los modelos para hacer la correcta elección para desarrollar un Sistema de Pronóstico utilizando técnicas de Machine Learning. Se entrenó el modelo con los datos recopilados y se evaluó su rendimiento. Las técnicas seleccionadas para este propósito son:

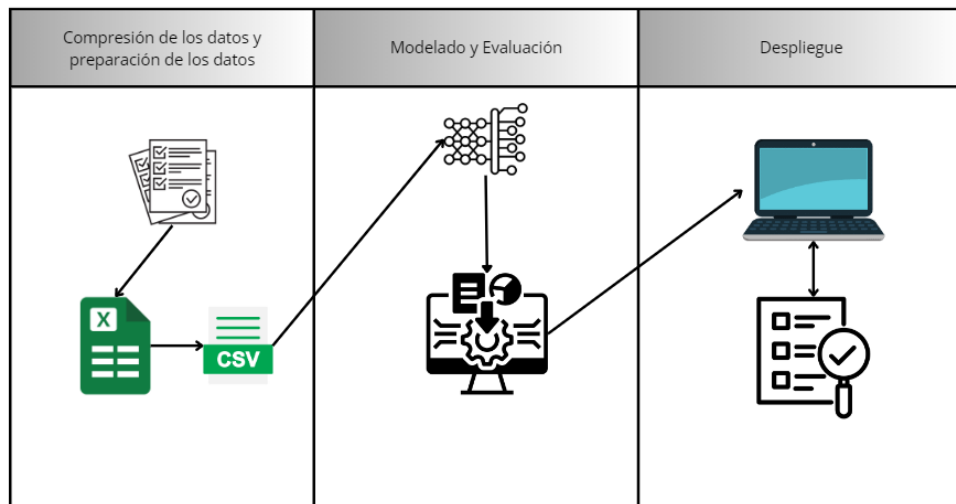
- Árboles de decisión
- Regresión Lineal

5. Evaluación:

Se evaluó el rendimiento del modelo utilizando métricas como la precisión, sensibilidad y especificidad para la elección de cuál de los modelos utilizados era el mejor para nuestro objetivo. En este caso, se encontró que el modelo tenía una precisión del 97% en la detección de la depresión en el área afectiva y conductual de las personas adultas del condominio.

6. Despliegue:

Se implementó el modelo desarrollado como parte de un sistema de pronóstico en el Condominio Ciudad Verde para ayudar en la detección temprana de la depresión en los residentes. Se usó el lenguaje de programación de Python junto al framework Flask para el desarrollo de una página web para poder interactuar con el cuestionario y poder dar resultados a las personas que la usen.



## Anexo 5. Autorizaciones para el desarrollo del proyecto de investigación

**ANEXO 4: Modelo del consentimiento o asentimiento informado UCV**

### **CARTA DE AUTORIZACIÓN DE INVESTIGACIÓN**

Lima, 20 de noviembre del 2023

Señores

**COMITÉ DE PROGRAMA DE INVESTIGACIÓN  
Universidad Cesar Vallejo**

Presente


De mi mayor consideración:

Tengo el agrado de dirigirme a Ud., saludarlo y a la vez hacer de su conocimiento que los alumnos Elisbeth Barriga Moreno con Código de estudiante 6000008826, DNI: 46168342 y Flavio Sthepano Muñoz Avalos con Código de estudiante 7002272363, DNI: 75162098.

En este sentido, autorizamos a los alumnos para que realicen el desarrollo de la investigación cuyo título es "*Sistema Inteligente con Machine Learning para la predicción de la depresión en personas adultas del Condominio Residencial Ciudad Verde*", permitiendo colaborar en este proceso ofreciendo información necesaria para el desarrollo.

Conociendo y aceptando el reglamento y disposición sobre la realización de grado de la Universidad Cesar Vallejo.

Cordialmente,

JUNTA DE PROPIETARIOS DEL CONJUNTO  
RESIDENCIAL CIUDAD VERDE  
  
GINA MONTENEGRO GUADALUPE  
PRESIDENTA

---

Presidenta

Condominio Residencial Ciudad Verde

**Anexo 6. Otras evidencias**

