UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

TESIS

Para Optar el Título Profesional de Licenciado en Estadística

TITULO

"REGRESION LOGÍSTICA DICOTÓMICA VERSUS RED NEURONAL PARA PREDECIR DEPRESIÓN EN ADULTOS MAYORES. HOSPITAL PROVINCIAL DOCENTE "BELÉN" Y CIAM LAMBAYEQUE, 2019."

INVESTIGADORES:

Casas Leguía Ninive Joel.

Guzmán Reyes Rowell Enrique.

ASESOR:

Msc. Antón Pérez Juan Manuel.

Lambayeque, 2020

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

"REGRESION LOGÍSTICA DICOTÓMICA VERSUS RED NEURONAL PARA PREDECIR DEPRESIÓN EN ADULTOS MAYORES. HOSPITAL PROVINCIAL DOCENTE "BELÉN" Y CIAM LAMBAYEQUE, 2019."

TESIS

PARA OPTAR POR EL TITULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

PRESENTADO POR:

BACH. CASAS LEGUÍA NINIVE JOEL

AUTOR

BACH. GUZMÁN REYES ROWELL ENRIQUE

AUTOR

MS.C. ANTÓN PÉREZ JUAN MANUEL

ASESOR

LAMBAYEQUE – PERU

2020

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

"REGRESION LOGÍSTICA DICOTÓMICA VERSUS RED NEURONAL PARA PREDECIR DEPRESIÓN EN ADULTOS MAYORES. HOSPITAL PROVINCIAL DOCENTE "BELÉN" Y CIAM LAMBAYEQUE, 2019."

TESIS

PARA OPTAR POR EL TITULO PROFESIONAL DE:

LICENCIADO EN ESTADÍSTICA

APROBADO POR:

Dra. Paredes López Lilian Roxana.
Presidenta

Dr. Acosta Piscoya Jorge Antonio.
Secretario

Lic. Est. Luis Enrique Tuñoque Gutiérrez.
Vocal

LAMBAYEQUE – PERU

2020

UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS
DECANATO
Ciudad Universitaria - Lambayeque

ACTA DE SUSTENTACIÓN Nº 001-2020-D/FACFyM
(Sustentación Autorizada por Resolución N° 1696-2019-D/FACFyM)

En la ciudad de Lambayeque, siendo las 11 a.m del día 07 de Enero del 2020 se reunieron en la Sala de Sustentación Videoteca - FISICA los miembros del Jurado designados mediante Resolución Nº 595-2019-D/FACFyM, los docentes:

Dra. Lilian Roxana Paredes López Presidente
Dr. Jorge Antonio Acosta Piscoya Secretario
Lic. Estad. Luis Enrique Tuñoque Gutiérrez Vocal

Para recibir la tesis titulada:
"Regresión Logística Dicotómica Versus Red Neuronal para predecir Depresión en Adultos Mayores. Hospital Provincial Docente "BELEN" y "CIAM" Lambayeque 2019"

desarrollada por los Bachilleres en Estadística, **Casas Leguia Ninive Joel y Guzmán Reyes Rowell Enrique.**

Después de escuchar la exposición y las respuestas a las preguntas formuladas por los miembros del Jurado, se acordó APROBAR el trabajo por UNANIMIDAD con el calificativo de MUY BUENO

En consecuencia, los Bachilleres en referencia quedan aptos para recibir el Título Profesional de **Licenciado en Estadística,** de acuerdo a la Ley Universitaria, el Estatuto y Reglamento de la Universidad Nacional Pedro Ruiz Gallo de Lambayeque.

Observaciones:

Para constancia del hecho firman.

Dra. Lilian Roxana Paredes López
Presidente

Dr. Jorge Antonio Acosta Piscoya
Secretario

Lic. Estad. Luis Enrique Tuñoque Gutiérrez
Vocal

DECLARACIÓN JURADA DE ORIGINALIDAD

Nosotros, Ninive Joel Casas Leguia y Rowell Enrique Guzmán Reyes investigadores principales y Juan Manuel Antón Pérez asesor del trabajo de investigación "REGRESION LOGÍSTICA DICOTÓMICA VERSUS RED NEURONAL PARA PREDECIR DEPRESIÓN EN ADULTOS MAYORES. HOSPITAL PROVINCIAL DOCENTE "BELÉN" Y CIAM LAMBAYEQUE, 2019". Declaramos bajo juramento que este trabajo no ha sido plagiado, ni contiene datos falsos. En caso se demostrara lo contrario, asumimos responsablemente la anulación de este informe y por ende el proceso administrativo que hubiera lugar. Que puede conducir a la anulación del título o grado emitido como consecuencia de este informe.

Lambayeque, 2020

Investigadores:

BACH. CASAS LEGUÍA NINIVE JOEL

AUTOR

BACH. GUZMÁN REYES ROWELL ENRIQUE

AUTOR

MS.C. ANTÓN PÉREZ JUAN MANUEL

ASESOR

DEDICATORIA

A Dios,

Mi creador, Todopoderoso que me brindo sabiduría y que gracias a su voluntad hoy cumplo uno de sus planes que tiene destinado para mí.

A mis padres y hermanos por ser la razón de mi vida y el apoyo incondicional ya que gracias a eso logro esta nueva meta. A mis hermanos por ser el ejemplo a seguir por sus buenas actitudes, que hoy hacen de mí una persona y un profesional de bien con una gran satisfacción darles infinitas gracias.

A Enrique

Por ser mi compañero y fiel amigo en esta aventura llena de retos y que, hoy cumplimos de manera satisfactoria.

Ninive Joel Casas Leguía

A Dios,

Mi creador, por la vida, salud y todas las maravillas que nos brinda que me llena de satisfacción cumplir su voluntad.

A mi familia, a mi madre, por su amor incondicional desde mi concepción, a mi padre, por su afecto, cariño y ejemplo.

A mi hermana, por ser mi mayor motivación en conjunto agradecerles por todas las cosas que han hecho de mí una persona de valores y un profesional de buena ética los amo de manera que las palabras describen poco su gran valor.

A Joel

Gracias, por compartir altos y bajos en este reto que con esfuerzo hoy nos brinda satisfacción.

Rowell Enrique Guzmán Reyes

AGRADECIMIENTO

A Dios por darnos la salud, fuerza e inteligencia para poder realizar este trabajo; para poder sobresalir frente a las adversidades presentadas tras los días de este camino y poder salir victoriosos.

A nuestro idóneo Ms.C. Juan Manuel Antón Pérez por ser imagen, mentor y y guía en cada paso de este proyecto, por su vocación, dedicación, comprensión y paciencia además de su entera disposición y desinterés para apoyarnos en cualquier interrogante que teníamos.

A los diferentes maestros de la Facultad de Ciencias Físicas y Matemáticas, los cuales formaron parte de nuestra etapa universitaria que con sus enseñanzas nos guiaron en el camino para lograr ser unos buenos profesionales de la carrera de Estadística

Los autores

INDICE GENERAL

INDICE DE TABLAS

INDICE DE FIGURAS

RESUMEN

La presente investigación tuvo como objetivo general determinar si una red neuronal con función logística proporciona un modelo de pronóstico de depresión en adultos mayores con menor error que la Regresión Logística Dicotómica, de acuerdo a las variables predictoras consideradas, "Hospital Provincial Docente Belén y CIAM" de Lambayeque 2019. Utilizando como método de estudio de nivel observacional-analítico y transversal, con 128 adultos mayores que acudieron al "Hospital Provincial Docente Belén y CIAM" de Lambayeque, los cuales se clasificaron con depresión: Leve y Moderada-Severa. Para definir los factores asociados a la depresión se incluyeron variables independientes (edad, género, estado civil, ingreso económico, actividad física, actividad laboral, actividad sexual, atención familiar, autoridad familiar y recreación), se utilizaron los modelos de regresión logística y red neuronal. En los resultados se demostró la prevalencia de depresión fue de 34,4% moderada-severa y 65,6% leve. Se encontraron asociaciones significativas respecto a la depresión: género, actividad física, actividad sexual, atención familiar, pérdida de autoridad y pasatiempo con $p<0,05$. En el modelo de regresión logística mostró que no tener atención familiar (OR=26,55, IC95% 8,41-98,81) y no tener pasatiempo (OR=6,59,IC95% 2,01–24,21) fueron considerados como factores de riesgo, asimismo en la red neuronal se encontraron los mismos factores de riesgo con un penalizador decay = 5e-2 y rang = 0,7. Se concluyó que la técnica multivariante de regresión logística dicotómica y la red neuronal proporcionan características similares de pronóstico: alta precisión (91,67%), f1-score (89,53%), alta sensibilidad (87,50%), especificidad (93,75%) y bajo error (8,33%).

Palabras clave: Depresión, red neuronal, regresión logística.

ABSTRACT

The present investigation had as a general objective to determine if a neural network with logistic function provides a model of depression prognosis in older adults with less error than the Dichotomous Logistic Regression, according to the predictive variables considered, "Belén Provincial Teaching Hospital and CIAM" of Lambayeque 2019. Using as an observational-analytical and cross-sectional study method, with 128 older adults who went to the "Provincial Teaching Hospital Belén and CIAM" of Lambayeque, which were classified as depression: Mild and Moderate-Severe. To define the factors associated with depression, independent variables were included (age, gender, marital status, economic income, physical activity, work activity, sexual activity, family care, family authority and recreation), the logistic regression and network models were used. neuronal The results showed that the prevalence of depression was 34.4% moderate-severe and 65.6% mild. Significant associations were found regarding depression: gender, physical activity, sexual activity, family care, loss of authority and hobby with $p < 0.05$. In the logistic regression model, he showed that he had no family care (OR = 26.55, 95% CI 8.41-98.81) and no hobby (OR=6.59, 95% CI 2.01-24.21) were considered as risk factors, also in the neural network the same risk factors were found with a penalty decay = 5e-2 and rang = 0.7. It was concluded that the multivariate dichotomous logistic regression technique and the neural network provide similar prognostic characteristics: high accuracy (91.67%), f1-score (89.53%), high sensitivity (87.50%), specificity (93.75%) and low error (8.33%).

Keywords: Depression, neural network, logistic regression.

INTRODUCCIÓN

El presente trabajo de investigación los autores realizan la comparación de dos modelos, uno de ellos estadístico (regresión logística) y el otro matemático computacional (red neuronal) para definir cuál de estos genera un mejor pronóstico de depresión en adultos mayores que acuden al Hospital Provincial Docente "Belén" y "C.I.A.M" Lambayeque en el año 2019.

Mayorga, en el 2013 expresa, que el modelo de regresión logística es una de las herramientas estadísticas con mejor capacidad para poder analizar datos en diferentes áreas, usualmente en investigaciones clínicas, teniendo como objetivo principal crear un modelo de pronóstico y el cómo influye en la aparición de un suceso para identificar factores de riesgo en diversos tipos de enfermedades.

Morales, en el 2010 expresa: Las redes neuronales son sistemas de procesamiento que fueron inspiradas en las propias neuronas de manera biológica, estas consisten en elementos simples de procesamiento conectados entre sí (pesos), su objetivo final de esto es reproducir la habilidad cognoscitiva del ser humano; este modelo es utilizado para resolver problemas de clasificación, estimación funcional y optimización en general.

La técnica de redes neuronales se sustenta en el perceptrón multicapa y en el algoritmo de retropropagación. Se explica el uso de la técnica mostrando sus diferentes pasos: establecimiento de la estructura, la función de activación sigmoidea, el paradigma de aprendizaje, el factor de aprendizaje, la regla de aprendizaje, el algoritmo de aprendizaje retropropagación, el entrenamiento y la evaluación de la red neuronal. (Palmer & Montaño, 2002)

"En ocasiones se utilizan diversas técnicas estadísticas para un mismo propósito, cada modelo hallado tendrá sus propias características y errores de precisión, que ponen en disyuntiva al tomador de decisiones dado que deberá elegir un modelo para su labor profesional" (Barón & Téllez, 2004).

Para identificar factores de riesgo de alguna patología, está probado una técnica estadística que es la de Regresión logística binomial o multinomial; sin embargo, surgen otras alternativas de orden matemático computacional que ofrecen reportes que anuncian mejores estimadores, como la teoría de la red neuronal. (Barón & Téllez, 2004)

Los autores de la presente investigación no han encontrado antecedentes que comparen estas 2 técnicas para identificar factores de riesgo o estimar probabilidades de contraer la depresión.

Se plantea la presente investigación teniendo en cuenta la polémica que genera este tipo de investigaciones y el poco uso de estos modelos en paridad, regresión logística y red neuronal, que permiten dar a conocer la significancia de los factores de riesgo identificados y que se asocian al pronóstico de depresión en adultos mayores.

Ante esta situación, es que en la presente investigación se plantea el siguiente **problema**: ¿Cuál de los modelos de Regresión Logística dicotómica o de red neuronal con función logística, reporta pronósticos de depresión en adultos mayores con menor error, de acuerdo a las variables predictoras consideradas, Hospital Provincial Docente Belén y CIAM de Lambayeque 2019?

Teniendo en cuenta la problemática se genera la siguiente **hipótesis**: El modelo de red neuronal con función logística proporciona un modelo de pronóstico de depresión del adulto mayor con menor error que el modelo de regresión logística dicotómico, de acuerdo a las variables predictoras consideradas.

Por tanto, en esta investigación se formula el siguiente **objetivo general:** Determinar si una red neuronal con función logística proporciona un modelo de pronóstico de depresión en adultos mayores con menor error que la Regresión Logística Dicotómica, de acuerdo a las variables predictoras consideradas, Hospital Provincial Docente Belén y CIAM de Lambayeque 2019. **Objetivos específicos: 1.** Construir un modelo de Regresión Logística Dicotómico para el pronóstico de depresión en adultos mayores de acuerdo a las variables predictoras edad, género, estado civil, ingreso económico, actividad física, actividad laboral, actividad sexual, atención

familiar, autoridad familiar y recreación. **2.** Construir un modelo de red neuronal para el pronóstico de depresión en adultos mayores de acuerdo con las variables predictoras, edad, género, estado civil, ingreso económico, actividad física, actividad laboral, actividad sexual, atención familiar, autoridad familiar y recreación. **3.** Calcular y comparar el error de pronóstico de depresión en adultos mayores con ambos modelos. **4.** Identificar los factores de riesgo y probabilidades de depresión con el modelo de mejor ajuste.

Después de lo antes mencionado para lograr el desarrollo de esta investigación, lo planteamos por capítulos que a continuación detallamos:

CAPITULO I.- En este capítulo se desarrolló el diseño teórico, en la cual se habla sobre el concepto de la variable a ser estudiada.

CAPITULO II.- En este capítulo se realizó los conceptos en relación a los modelos a comparar en esta investigación y se hará mención a los materiales que se requieren.

CAPITULO III.- En este capítulo se realizó las metodologías que se emplean en la ejecución de esta investigación.

CAPITULO IV.- En este capítulo se desarrolló el análisis e interpretación de los resultados obtenidos y observados en el transcurso del uso de los modelos empleados en la investigación.

CAPITULO V.- En este capítulo proponemos la solución al problema brindando recomendaciones para las futuras investigaciones.

ANTECEDENTES DE ESTUDIO

Ávila (2015), en su investigación *Determinantes sociales relacionados a la depresión del Adulto mayor en el centro de salud de la Parroquia San Juan Cantón Gualaceo Provincia del Azuay 2015*, reporta que:

> La prevalencia de depresión fue de 53,6%; de esto, 44% fue moderada y 9,6% fue severa, pero se encontraron diferencias entre sexos, pero se halló una tendencia progresiva ascendente con respecto a la edad, en cuanto, la disfuncionalidad familiar fue un factor fuertemente asociado a la prevalencia de depresión en este estudio (x^2 =18,171; p<0,001). Esta incrementó progresivamente a mayor disfuncionalidad familiar, mientras que tener confianza sobre sí mismo y disfrutar las actividades diarias se comportaron como factores protectores significativos para la depresión. El factor de riesgo para el modelo de regresión logística fue: disfunción familiar (OR=2,76; IC95%: 1,02-7,45) con p=0,04, la investigación fue trabajada con una muestra de 125 adultos mayores que acudieron al centro de salud de San Juan del Cantón Gualaceo, asimismo se utilizó un modelo de regresión logística múltiple para la determinación de los factores de riesgo.

Ramírez, Bedoya, Correa, y Villada (2015), en su investigación *Riesgo de Depresión y Factores Asociados en Adultos Mayores Institucionalizados en la red de asistencia social al Adulto Mayor. Medellín. 2015*, reportaron que:

> De los 644 adultos mayores que participaron del presente estudio a quienes se les había aplicado la escala de valoración Yesavage, el 65% no presentaron depresión, mientras que el 35% se encontraron con depresión leve (29%) y depresión grave (6%), respecto al sexo se encontró que el 21% de hombres tienes el riesgo de tener depresión y el 14% en mujeres; El modelo de regresión logística los factores de riesgo que se encontraron fueron: enfermedades neurológicas (OR=1,94, IC95% 1,14 – 3,31), a funcionalidad dependiente (OR=1,97, IC95% 1,30 - 2,98), la discapacidad auditiva (OR=2,21, IC95% 1,34 - 3,63) y el deterioro cognitivo (OR=1,54, IC95% 1,01 -2,35) con p<0,5, se utilizó toda la población de estudio ,pertenecientes a las 15 instituciones registradas en la Red de asistencia social al adulto mayor de Medellín y corregimientos aledaños. Ésta investigación realizó un modelo

de regresión logística binomial donde se comparó el riesgo de depresión con las variables independientes que en el análisis bivariado presentaron una p ≤0,25 según criterio de Hosmer-Lemeshow.

Francia (2010), en su investigación *Factores biopsicosociales que influyen en los niveles de depresión de los adultos mayores del C.S. Materno Infantil Tablada de Lurín, 2010* reportó que:

> Se evidenció que la mayoría presentó nivel depresión leve (59,1%), seguidos por los adultos mayores con nivel normal de depresión (31,8%), y finalmente, la minoría presentó nivel de depresión severa (9,1%), Los factores de riesgo finales para el modelo de regresión logística fueron: discriminación (OR=13, IC95% 1,1 -151,9) y sentimiento de soledad (OR=112,7, IC95% 7,5 1701,8) con p<0.5; Se trabajó con toda la población de estudio que fueron 47 adultos mayores, 14 varones y 33 mujeres que pertenecen a El Club de Adultos Mayores "Edad de Oro", asimismo para la investigación se utilizó un modelo de regresión logística.

Aldana y Pedraza (2012), en su investigación *Análisis de la depresión en el adulto mayor en la encuesta nacional de demografía y salud 2010, Colombia* reportaron que:

> La prevalencia de global de depresión en el grupo analizado fue de 9,5%, con menor prevalencia en hombres 6,9% con relación a las mujeres 11,5%. Se encontró un (OR= 1,74, IC 95% 1,56-1,94) para el sexo femenino además que se observó que había una disminución de la prevalencia de depresión relacionada con el aumento del nivel educativo, presentando un menor OR en la educación superior comparado con los que no tenían educación 5,6% (OR= 0.14, IC 95% 0.090.24). Inclusive al ajustarlo por las otras variables demográficas (OR ajustado 0.18, IC 95% 0.13-0.30), asimismo el ser viudo/a 11,5% con OR=1,36, IC95% 1,18-1,52 por el contrario el ser soltero/a 7,9% con OR=0,88, Ic95% 0,72—1,08 siendo este considerado como factor de confusión. La muestra de estudio fue de 17.574 adultos mayores encuestados pertenecientes a 6 regiones, 16 subregiones y 32 departamentos de Colombia. Se realizó una regresión logística binaria y múltiple para establecer las variables asociadas a la depresión.

Carmona y De los Santos (2012), en su investigación *Prevalencia de depresión en hombres y mujeres mayores en México y factores de riesgo* reportaron que:

> Las 3920 personas entrevistadas presentaban depresión; es decir, el 74.3% de la población consultada se ha sentido deprimida, infeliz, sola, cansada, sin energía, triste, agobiada, que no disfrutaba de la vida, con sueño intranquilo, síntomas presentes durante dos o más semanas seguidas (1734 hombres y 2180 mujeres). Estas cifras sobrepasan las estimaciones propuestas por la ENSANUT 2012 y por la OMS, 2011. Es relevante advertir que el 55.8% de las mujeres y el 44.2% de los hombres reveló estos síntomas, por lo que la situación concuerda con la evidencia empírica que señala mayor prevalencia de depresión en mujeres mayores; Los factores de riesgo del modelo de regresión logística binomial fueron: edad en sus diferentes categorías, escolaridad excepto la categoría secundaria con $p>0,5$ estado civil excepto las categorías soltero y casado o unido ($p>0,5$), contar con servicio médico considerado como un factor protector con $p>0.5$, tener capacidad funcional excepto la categoría sin dificultades ($p>0,5$), ocupación, contar con ayuda económica y no económica, actividad física como factor protector ($p<0,5$).

Fumero y Navarrete (2014), en su estudio *Personalidad y Malestar Psicológico: Aplicación de un Modelo de Redes Neuronales en América Latina, el Caribe, España y Portugal* reportó los siguientes resultados:

> La capacidad predictiva para clasificar correctamente a las personas que presentarán malestar en ese sentido fue del 86.50% con un error de 0.51, utilizando siete variables por este orden: edad, neuroticismo, activación antes situaciones de estrés, trastorno esquizoide de la personalidad, trastorno límite de la personalidad, control externo y afecto negativo. En otras palabras, deberíamos utilizar información que no señalara tanto la edad como la inestabilidad emocional, la vulnerabilidad a los trastornos de la personalidad que suponen la incapacidad para disfrutar con cosas agradables y la desregulación emocional, la tendencia a percibir falta de control sobre los acontecimientos y la afectividad negativa. En el caso de la disfunción social los resultados muestran que locus de control externo con depresión,

pesimismo y trastorno esquizoide de la personalidad permiten predecir correctamente al 75% con un error 0.77, de las personas que perciben disfunción social en sus actividades diarias. La predicción de esta disfunción debe considerar elementos críticos, la ausencia de sensación de control sobre los acontecimientos, la expectativa que lleva a esperar lo negativo y la tendencia a evitar vincularse y compartir intimidad con otros.

Bustos, Fernández, y Astudillo (2017), en su investigación *Autopercepción de la salud, presencia de comorbilidades y depresión en adultos mayores mexicanos: propuesta y validación de un marco conceptual simple logística* reportó los siguientes resultados:

Una asociación directa entre la autopercepción positiva de la salud y la presencia de comorbilidades (OR=0,48; IC95% 0,42-0,55), la discapacidad (OR=0,35; IC95% 0,30-0,40) y la depresión (OR=0,38; IC95% 0,34-0,43), así como con la variable exógena de pobreza (OR=0,87; IC95% 0,76-0,98); Respecto a la muestra de estudio que fueron 8.874 adultos mayores que tenían 60 o más años, y se ajustó preliminarmente un modelo de regresión logística convencional.

Sendra, Asensio y Vargas (2017), en su investigación *Características y factores asociados a la depresión en el anciano en España desde una perspectiva de género* reportó los siguientes resultados:

Del análisis multivariante señalan la presencia de cuatro variables comunes para ambos sexos asociadas de forma positiva a la depresión, cuyas categorías son: el estado de salud percibido regular y malo/muy malo (ORh =6,7; ORm=3,8), la permanencia en cama las últimas 2 semanas, el no poder caminar o hacerlo con mucha dificultad sin ayuda, y los grados severos (ORh =3,5; ORm=2) y extremo (ORh =5; ORm=3,9) de dolor en las últimas 4 semanas. De forma diferencial, se relacionan positivamente con los trastornos depresivos en la mujer el no saber leer ni escribir (ORm=2,5), la presencia de una enfermedad crónica (ORm=5,7), caminar con alguna dificultad (ORm=1,6) y la ausencia o poco interés por parte de otras personas (ORm=4), y en el hombre el grado moderado de dolor (ORm=1,9). Por el contrario, el incremento de la edad se asocia negativamente a la presencia de depresión únicamente en la

mujer (ORm=0,98). Los factores de riesgo para el modelo de regresión logística fueron: el no saber leer ni escribir (ORaj=2,48, IC95% 1,57-4,82), el sentirse regular (ORaj=2,28, IC95% 1,56- 3,32), malo/muy malo (ORaj=3,84, IC95% 2,55-2,80), enfermedad o problema crónico (ORaj=5,66, IC95% 1,77-18,16), permanencia en cama últimas 2 semanas (ORaj=1,68, IC95% 1,22-2,30), tener dificultad para caminar 500 metros sin ayuda con IC95% >1 y el tener poco o nada interés por parte de otras personas (ORaj=3,99, IC95% 2,11-7,53).

Fernández, Marrero, Mesa, Santiesteban, y Rojas, (2011), en su investigación *Depresión post–ictus: frecuencia y factores determinantes* se obtuvo los siguientes resultados:

Se observa que ninguno de la tres: la edad del paciente, los años de estudio y el tiempo de evolución de la enfermedad, mostró diferencias estadísticamente significativas; el cambio, la puntuación del MMSE fue menor y la puntuación de la NIHSS fue mayor en los pacientes con DPI. El modelo de regresión logística presentó como factor de riesgo al deterioro cognitivo (OR=2,62, IC95% 1,02—6,71), mientras que la afectación neurológica grave vs moderada, afectación neurológica grave vs leve, afectación frontal y dependencia para las AVD el modelo los considera como riesgo o protección; La población de estudio fue de 120 adultos mayores quienes se hospitalizaron consecutivamente para rehabilitación en el Hospital Julio Díaz (HJD) de La Habana y se desarrolló un modelo de regresión logística.

CAPITULO I: DISEÑO TEÓRICO

1. MODELO DE REGRESIÓN LOGÍSTICA

Reseña Histórica

En 1960 surge la regresión como alternativa al procedimiento de estimación de los mínimos cuadrados ordinarios que usualmente es usado en regresión lineal. El origen de la regresión logística se debió a las limitaciones de predicciones simples que presentaban las herramientas o las técnicas tales como Redes neuronales artificiales y los perceptrones. (López y García, 2011, p.14)

En 1984 Lilienfeld y Pyne, declararon que el tamaño de la muestra influía en la fiabilidad de las estimaciones, en la exactitud de la estimación de β y en la exactitud del contraste de hipótesis H0: $\beta = 0$. Es decir, encontraron que la desviación típica para las estimaciones de β en muestras simuladas se reducía a medida que aumentaba el tamaño de la muestra. (López y García, 2011, p.14)

Whittermore (1981), con la finalidad de minimizar el efecto de la fiabilidad respecto al tamaño de muestra, desarrollo un procedimiento para calcular tamaños óptimos para el análisis de regresión logística usando el método de máximo verosímil asintótica a partir de la matriz de Hess. (López y García, 2011, p.15)

1.1. MODELO DE REGRESIÓN LOGÍSTICA BINOMIAL

Los modelos de regresión logística son útiles para los casos donde la investigación requiera saber la presencia o ausencia de una característica o resultado según o en relación a las variables predictoras. Los coeficientes obtenidos se pueden utilizar para estimar la razón de las ventajas (Odds Ratio) de las variables independientes (Rodas, 2009). La variable respuesta es discreta (generalmente toma valores 1,0) mientras que las variables explicativas pueden ser cuantitativas o cualitativas por ende el modelo no es lineal sino exponencial sin embargo puede ser una regresión lineal al aplicar una transformación logarítmica. (Salcedo, 2002)

La regresión logística es una técnica estadística multivariante que permite estimar una variable dependiente usualmente dicotómica no métrica en relación con conjunto de variables independientes métricas o no métricas. (Barón & Téllez, 2004)

La ventaja de usar el modelo regresión logística es el uso conjunto de variables cuantitativas y cualitativas en una única ecuación. (Salcedo, 2002)

El objetivo primordial del modelo de regresión logística es el modelar cómo influyen las variables regresoras en la probabilidad de ocurrencia de un suceso, así mismo, como objetivo general es pronosticar o predecir la probabilidad de que ocurra un fenómeno o evento de interés en una investigación e identificar las variables predictoras que serían útiles para la investigación. (Salcedo, 2002)

A) Codificación de variables

- **Variable dependiente**, codificado como 1 si ocurre el evento de interés y 0 como la ausencia.
- **Variables independientes**:
  - ✓ **Caso dicotómico,** codificado como 1 si ocurre el evento de interés y 0 como la ausencia.
  - ✓ **Caso categórico,** codificado con variables indicadoras (dummy) dado que la variable independiente toma más de dos valores.

B) Formulación del modelo

B.1. Modelo Logit (Binario)

Usado cuando se quiere modelar una variable dependiente de tipo cualitativa.

Definición:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k \quad [1]$$

Donde, p es la probabilidad de que ocurra el evento de interés, xi son las variables independientes y β son los coeficientes asociados con cada variable independiente. (Rojo, 2007)

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i \, x_i^n \quad [2]$$

Aplicando Antilogaritmo

$$\text{OR} = \frac{p}{(1-p)} = e^{\beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta k Xk} \quad [3]$$

Despejamos p

$$p = (1- p)\, e^{\beta 0 + \beta 1 X1 + \beta 2 X2 + \ldots + \beta k Xk} \quad [4]$$

Cambio de variable

$$p = (1- p)e^{z}$$

$$p + pe^{z} = e^{z}$$

$$p(1+e^{z}) = e^{z}$$

$$p = \frac{e^{z}}{1+e^{z}} \qquad [5]$$

Dividimos entre e^{z}de lo cual se deduce

$$p = \frac{1}{1 + e^{-z}} \qquad [6]$$

B.2. Riesgo o Ventaja relativa (Odds Ratio)

Es una forma de expresar la probabilidad, denominada también como razón de oportunidades, razón de posibilidades o razón de productos cruzados. (Rojo, 2007)

Definición:

$$\text{Odds} = \frac{p}{(1-p)} \qquad \begin{array}{l}\text{Si Odd} > 1 \quad , \quad \text{es factor de riesgo} \\ \text{Si } 0 < \text{Odd} < 1, \quad \text{es factor protector}\end{array} \qquad [7]$$

Con $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k$

Por tanto, el odds ratio es el número de veces que es más probable que ocurra el fenómeno o suceso frente a que no ocurra. (Rojo, 2007)

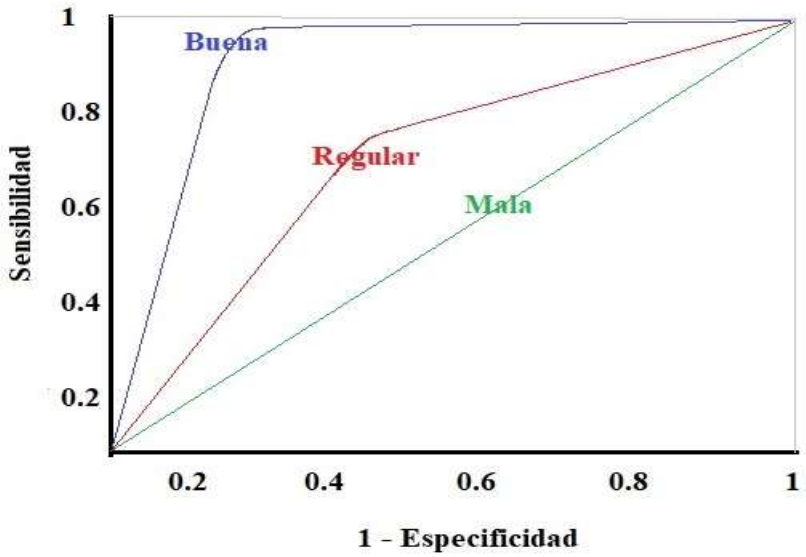

Figura 1. Tipo de curva ROC. Elaboración propia

Estimación de parámetros

Sea una muestra de n elementos, donde se ha observado la variable respuesta Y (toma dos valores: cero y uno) y la variable X. (Rojo, 2007)

La función de probabilidad de una observación cualquiera es:

$$P(Y = 1/ x) = p \quad [8]$$

$$P(Y = 0 / x) = 1- p \quad [9]$$

Donde:

$$P(Y / x) = p^{y} (1 – p)^{1-y} \quad [10]$$

Función de probabilidad:

$$P(y_1,y_2,….,y_n) = \prod_i p^{yi}(1-p)^{1-yi} \quad [11]$$

Aplicando logaritmo:

$$\ln P(Y) = \sum_i^n y_i * \ln(\frac{p}{1-p}) + \sum_i^n \ln(1-p) \quad [12]$$

Expresando en función a los parámetros:

$$L(\beta) = \sum_i^n Y_i * X_i * \beta - \sum_i^n \ln(1 + e^{Xi*\beta}) \quad [13]$$

Derivando se obtendrá:

$$\beta_a = \beta_o + (\frac{\partial^2 L(\beta)}{\partial\beta * \partial\beta r})^{-1} * \frac{\partial L(\beta)}{\partial\beta} = 0 \quad [14]$$

B.3. Medidas de Bondad de Ajuste

Pseudo R^2: Cox y Snell, R^2 de Nagelkerke

Al analizar los datos con un modelo de regresión logística, no existe una estadística que se asemeje al coeficiente de determinación (R^2). Sin embargo, para evaluar la bondad de ajuste de modelos de regresión logística se desarrollaron Pseudo R-Cuadrados. "Pseudos" R-cuadrados son denominados de esta manera dado que las estimaciones del modelo a partir de una regresión logística son estimaciones de máxima verosimilitud obtenidas a través de un proceso iterativo. No están calculados para minimizar la varianza pero si están en una escala similar al R-cuadrado convencional entre 0 a 1, aunque algunos Pseudo R-cuadrado nunca alcanzan 0 o 1. (Long, 1997)

Cox y Snell

$$R^2 = 1 - \frac{L(b_0)}{L(b_0,b_1,\ldots,b_k)} \qquad [15]$$

$$R^2 = 1 - \left(\frac{L(b_0)}{L(b_0,b_1,\ldots,b_k)}\right)^{2/N} = 1 - e^{\left(\frac{L(b_0,b_1,\ldots,b_k)-L(b_0)}{N}\right)} \qquad [16]$$

Este coeficiente está acotado entre:

$0 \leq R^2 < 1$

Es decir, no alcanza el valor 1.

R^2 De Nagelkerke

$$R^2 = \frac{R^2}{R^2_{máx}} = \frac{1-e^{\left(\frac{L(b_0,b_1,\ldots,b_k)-L(b_0)}{N}\right)}}{1-e^{\frac{-L(b_0)}{N}}} \qquad [17]$$

Donde $R^2_{máx} = 1 - (L(b_0))^{2/N}$

Para así poder alcanzar el valor 1.

Estos coeficientes tratan de medir la variabilidad explicada, sin embargo, van a ser mucho más bajos que en regresión lineal. (Long, 1997)

1.2. MODELO DE REGRESIÓN LOGÍSTICA MULTINOMIAL

Los modelos de regresión logística dicotómico o binario pueden ser generalizados para el caso de una investigación que desea obtener más de dos opciones, dando origen a los Modelos de Regresión Logística Múltiple. El modelo estadístico de regresión logística lo que desea explicar es la relación que existe entre la variable respuesta en función a las variables explicativas o independientes y estas pueden ser cualitativas o cuantitativas(Gómez y Palacios, 2013).

2. REDES NEURONALES ARTIFICIALES

Fundamentos Biológicos

En el Siglo XX (1888) el científico Santiago Ramón y Cajal desarrolló la idea de las neuronas como el componente más pequeño en la estructura del cerebro. (Serrano, Soria y Martín, 2010)

Ramón y Cajal (1888), demostró que "el sistema nervioso está compuesto por una red de células individuales, las neuronas, interconectadas entre sí, de la cual el proceso de información fluye atravesando el soma desde las dendritas hacia el axón. (Serrano, Soria y Martín, 2010)

Estructura de la neurona biológica

A partir del estudio y conocimiento del sistema nervioso, los investigadores trataron de encontrar respuestas a los estímulos profundizando acerca de los tejidos y órganos que lo conforman, en especial el cerebro. (Lara, s.f)

La neurona, consta como toda célula de una membrana exterior, que sirve y limita como órgano de intercambio, así mismo, conformada de un citoplasma, siendo el cuerpo principal de la célula donde radican sus funciones, y el núcleo siendo quien es el que contiene la mayor parte del genético celular. (Lara, s.f)

El citoplasma presenta dendritas que son órganos receptores y donde termina el gran número de fibras que son los conductores encargados de llevar la información o impulso nervioso hacia la neurona. Las fibras transmiten señales de una neurona a otra a través de la sinapsis que es un corpúsculo en donde ésta termina. (Lara, s.f)

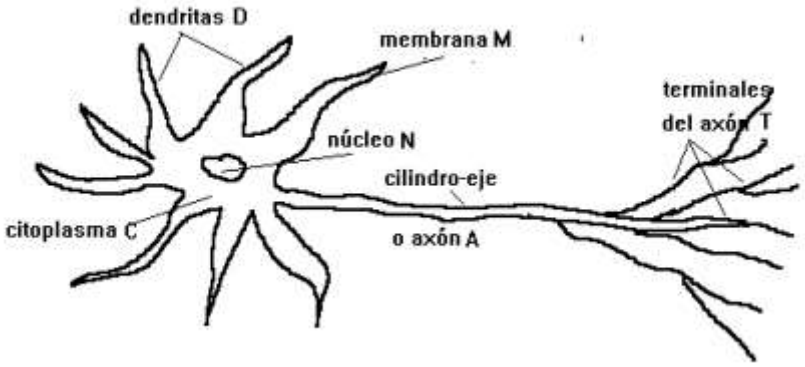

Figura 2. Esquema de una neurona. (Lara, s.f). Fundamentos de redes neuronales artificiales

Información biológica:

- ✓ Neurona: $100 \cdot 10^6$ neuronas en el cerebro
  - Cuerpo celular o soma (de 10 a 80 micras de longitud)
  - Del soma surge un denso árbol de ramificaciones (árbol dendrítico) formado por las dendritas.
  - Del soma parte una fibra tubular denominada axón (longitud de 100 micras hasta un metro).
  - El axón se ramifica en su extremo final para conectar con otras neuronas.
- ✓ Neuronas como procesadores de información sencillos. De manera simplista:
  - Las dendritas constituyen el canal de entrada de la información.
  - El soma es el órgano de cómputo.
  - El axón corresponde al canal de salida, y a la vez envía información a otras neuronas. Cada neurona recibe información de aproximadamente 10,000 neuronas y envía impulsos a cientos de ellas.
  - Algunas neuronas reciben la información directamente del exterior.
- ✓ El cerebro se modela durante el desarrollo de un ser vivo. Sus cualidades no son innatas, sino que se adquieren por la influencia de la información que del medio externo se proporciona a sus sensores.

  Diferentes maneras de modelar el sistema nervioso:
  - Establecimiento de nuevas conexiones.
  - Ruptura de conexiones.
  - Modelado de las intensidades sinápticas (uniones entre neuronas).
  - Muerte o reproducción neuronal.

Evolución Histórica

En 1943, McCulloch y Walter Pitts lanzaron una teoría acerca de cómo trabajan las neuronas, modelaron una red neuronal simple mediante circuitos eléctricos. (Flórez y Fernández, 2008)

En 1958, Frank Rosenblatt comenzó el desarrollo del perceptrón, siendo la red neuronal más antigua que se utiliza como una aplicación para reconocer patrones, así mismo este modelo después de un proceso llamado aprendizaje puede reconocer patrones similares. Posteriormente, Widrow y Hoff (1960), desarrollan variaciones en el algoritmo de aprendizaje llamada, "Ley de Widrow-Hoff", que dio nombre a ADALINE (Adaptative Linear Elements), que fue la primera en construir una RNA aplicada a un problema real. (Flórez y Fernández, 2008)

Características de las redes neuronales

Topología

Al hacer una clasificación topológica de las RNAs se suelen distinguir:

1. Según estructura de capas
    1.1. Redes monocapa: Establece conexiones compuestas por una única capa. Las redes monocapa se utilizan especialmente para tareas relacionadas conocidas como auto asociación.
    1.2. Redes multicapa: Disponen de neuronas agrupadas en varios niveles que están organizadas en varias capas. Cada capa de una RNA se distingue por está conformada por el conjunto de neuronas que comparten el origen de las señales de entrada y señales de salida. (Flórez y Fernández, 2008)
2. Según el flujo de datos en la red
    2.1. Redes unidireccionales o de propagación hacia adelante (feedforward), en las que ningún resultado de salida neuronal es la entrada de la misma capa o de las capas anteriores, donde la información circula en un solo sentido, desde las neuronas de entrada hacia las de salida.
    2.2. Redes de propagación hacia atrás (Backpropagation), en la que los resultados de las salidas pueden servir como entradas del mismo nivel (conexiones laterales). (Flórez y Fernández, 2008)

A) Elementos y características de una red neuronal artificial

A.1. La neurona artificial

La neurona artificial es una unidad procesadora con cuatro elementos funcionales:

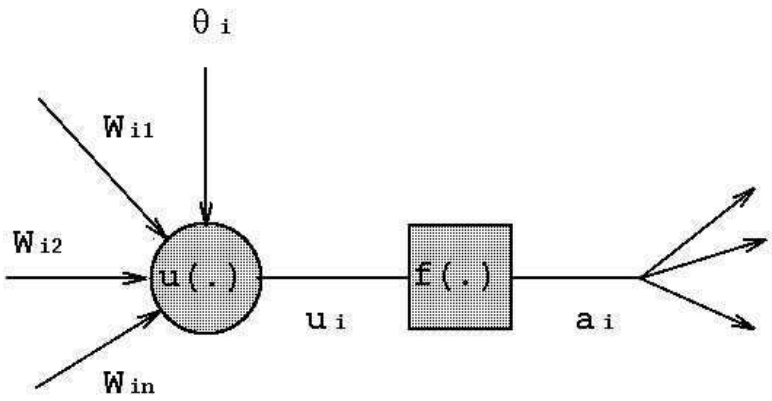

Figura 3. Esquema de una neurona artificial. (Marín, 2012). Introducción a las Redes Neuronales Aplicadas.

1. El elemento receptor o conexiones, a donde llegan una o varias señales de entrada xi, que generalmente provienen de otras neuronas y que son atenuadas o amplificadas cada una de ellas con arreglo a un factor de peso ***wi*** que constituye la conectividad entre la neurona fuente de donde provienen y la neurona de destino en cuestión. (Flórez y Fernández, 2008)

2. La función de red o elemento sumador efectúa la suma algebraica ponderada de las señales de entrada, ponderándolas de acuerdo con su peso, aplicando la siguiente expresión:

$$u_i(w, x) = \sum_{j=1}^{n} w_{ij} x_j \qquad [18]$$

3. El elemento función de activación, aplica una función no lineal de umbral (que frecuentemente es una función escalón o una curva logística) a la salida del sumador para decidir si la neurona se activa, disparando una salida o no. (Flórez y Fernández, 2008)

4. El elemento de salida que es el que produce la señal, de acuerdo con el elemento anterior, que constituye la salida de la neurona. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona siguiente. Normalmente, no cualquier valor es permitido como una entrada para una neurona, por lo tanto, los valores de salida están comprendidos en el rango

[0, 1] o [-1, 1]. También pueden ser binarios {0, 1} o {-1, 1}. (Flórez y Fernández, 2008)

Este modelo neuronal es el utilizado en casi todas las Redes Neuronales artificiales variando únicamente el tipo de función activadora. (Flórez y Fernández, 2008)

B) Redes neuronales supervisadas y no supervisadas

B.1. Reglas de entrenamiento Supervisado

Las redes neuronales de entrenamiento supervisado son las más populares. El conjunto de datos de entrenamiento está conformado por varios pares de patrones de entrenamiento de entrada y de salida. El hecho de conocer el valor de salida implica que el entrenamiento es beneficiado con la supervisión de un maestro. Dado un nuevo patrón de entrenamiento, en la etapa (m+ 1)-ésima, los pesos se adaptan de la siguiente forma:

$$w_{ij}^{m+1} = w_{ij}^{m} + \Delta w_{ij}^{m} \qquad [19]$$

Diagrama de un sistema supervisado:

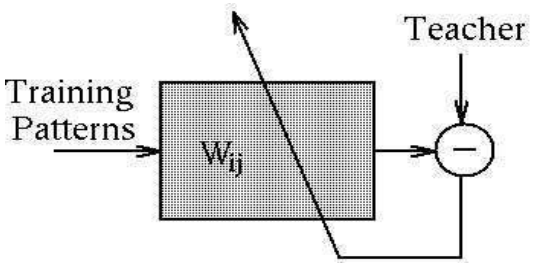

Figura 4. Sistema supervisado.(Marín, 2012). Introducción a las Redes Neuronales Aplicadas.

B.2. Reglas de Entrenamiento No Supervisado

Para este modelo no existe una supervisión de un maestro por lo que la red aprende a adaptarse sola a través de experiencia que recoge de los patrones de entrenamiento anteriores, asimismo el conjunto de datos de entrenamiento solo tiene patrones de entrada. (Marín, 2012)

Diagrama de un sistema no supervisado:

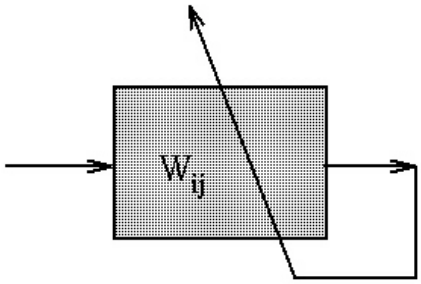

Figura 5. Sistema no supervisado. (Marín, 2012). Introducción a las Redes Neuronales Aplicadas.

B.3. Función de Activación

Cada nodo o neurona de la red proporciona un valor $\mathbf{y_j}$ a su salida, este valor se dispersa a través de conexiones hacia los otros nodos de la red, estas conexiones son unidireccionales y están asociadas a pesos sinápticos $\{\mathbf{w_{ij}}\}$ que es el que determina el efecto del nodo j-ésimo sobre el nodo i-ésimo. (Marín, 2012)

Las entradas al nodo i-ésimo que provienen de los otros nodos se acumulan junto con el valor umbral $\boldsymbol{\theta_i}$, posteriormente aplicándose a la función base **f**, obteniendo $\mathbf{u_i}$. La salida final $\mathbf{y_i}$ se obtiene aplicando la función de activación sobre $\mathbf{u_i}$. (Marín, 2012)

B.4. Función de Activación (Función de neurona)

El valor de red, expresado por la función de base, u(w, x), se transforma mediante una función de activación no lineal. La función de activación más común es la función sigmoidal, ésta función con forma de S, es acotada y no decreciente, la cual provee una respuesta no lineal:

- Función sigmoidal $f(u_i) = \frac{1}{1+e^{-u}}$ [20]

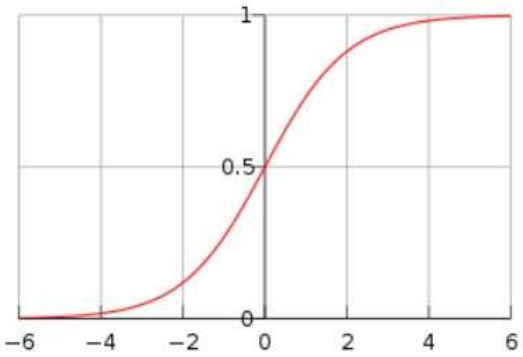

Figura 6. Función sigmoidal. (Academic, 2010)

- Función tangente hiperbólica $f(u_i) = \frac{e^{u} - e^{-u}}{e^{u} + e^{-u}}$ [21]

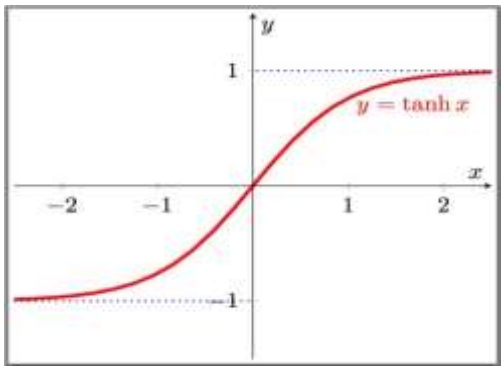

Figura 7. Función tangente hiperbólica.(Grill, 2014). Plotting the graph of hyperbolic tangent

2.1. PERCEPTRÓN MULTICAPA

El perceptrón multicapa actúa como como un aproximador haciendo el uso de backpropagation propuesta por Rumelhart (1986) que al menos conteniendo una capa oculta aproxima a cualquier tipo de función o relación entre las variables de entrada y salida. Se denomina regla de delta generalizada. (Marín, 2012)

Arquitectura

La arquitectura de perceptrón multicapa se caracteriza por tener en cada nivel capas de neuronas agrupadas, se distinguen tres tipos de capas: capa de entrada, capas ocultas y capa de salida. (Marín, 2012)

Las neuronas que están en la capa de entrada solo reciben datos, información o señales que provienen directamente del exterior para así propagar dichas señales al conjunto de neuronas que está en la siguiente capa. La última capa devuelve la respuesta de la red para cada uno de las señales de la capa de entrada y las neuronas de las capas ocultas realizan procesos no lineales de los patrones recepcionados. (Marín, 2012)

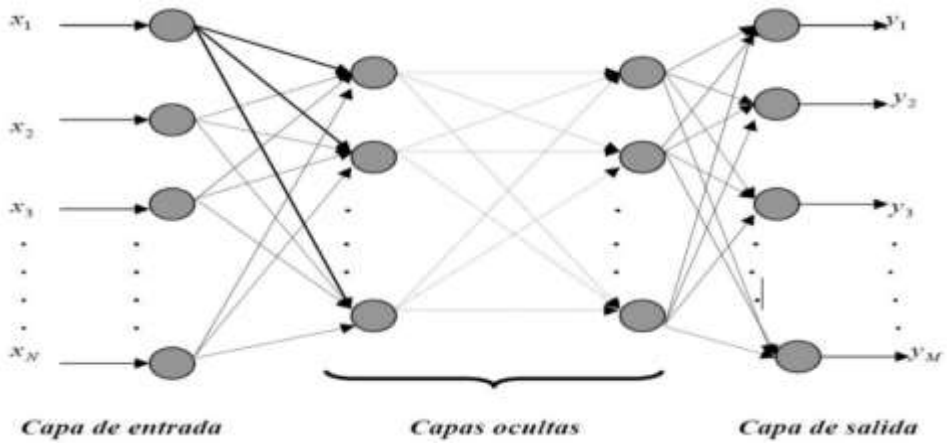

Figura 8. Esquema del perceptrón multicapa

Fuente: (Marín, 2012). Introducción a las Redes Neuronales Aplicadas.

Se puede observar que las conexiones de las capas son hacia adelante, desde una neurona de la capa anterior hacia la neurona de la siguiente capa, no existen conexiones laterales o hacia atrás por ende la información siempre será trasmitida desde la capa de entrada hacia la capa de salida. (Marín, 2012)

A) Algoritmo de backpropagation

Es la etapa en la que se modifican los pesos a través de un aprendizaje o entrenamiento que recibe la red en la capa de entrada para luego procesar la información hacia las nuevas capas hasta llegar a obtener una salida, estos pesos se modificarán hasta obtener el resultado que coincidan con el deseado por el investigador. (Marín, 2012)

B) Etapa de funcionamiento

El patrón de entrada p está expresado como un vector pX: $x_{p1,}$…., x_{pi} ,…, $x_{pN,}$ se trasmite a través de los pesos $\boldsymbol{w_{ji}}$ de la capa de entrada hacia la capa oculta, la entrada neta que recibe una neurona oculta **j, $\mathbf{netp_j}$** , es:

$$net_{pj} = \sum_{i=j}^{N} w_{ij} x_{pi} + \theta_j \qquad [22]$$

Donde $\boldsymbol{\theta_j}$ es el umbral de la neurona que se supone como un peso asociado a una neurona ficticia con valor de salida igual a 1. (Marín, 2012)

Las neuronas de esta capa intermedia transforman las señales recibidas mediante la aplicación de una función de activación (sigmoidea generalmente) y se obtiene así un valor de salida:

$$y_{pj} = f(net_{pj}) \qquad [23]$$

Donde $\mathbf{y_{pj}}$ es el valor de salida de la neurona j. Este valor se transfiere a través de los pesos $\mathbf{v_{kj}}$ hacia la capa de salida:

$$net_{pk} = \sum_{i=j}^{H} v_{kj} y_{pj} + \theta_j \qquad [24]$$

Por último, en la capa de salida se aplica la misma operación que en la capa anterior, las neuronas de esta última capa proporcionan la salida, y_{pk}, de la red:

$$y_{pk} = f(net_{pk}) \qquad [25]$$

C) Etapa de aprendizaje

En esta etapa el objetivo es minimizar el error entre la salida por la red y la salida deseada por el investigador ante el conjunto de patrones llamado grupo de entrenamiento. De esta manera se convierte en aprendizaje supervisado. (Marín, 2012)

La función para minimizar el error por cada patrón está dada por:

$$E^p = \frac{1}{2}\sum_{k=1}^{M}(d_k^p - y_k^p)^2 \quad [26]$$

Donde d_{pk} es la salida deseada para la neurona de salida k ante la presentación del patrón p. (Marín, 2012)

A partir de la expresión se puede obtener una medida general de error mediante:

$$E = \sum_{p=1}^{P} E^p \quad [27]$$

Tomando en cuenta como base el algoritmo de backpropagation que modifica los pesos, este método o técnica es conocía como descenso de gradiente. (Marín, 2012)

Como es función de todos los pesos de la red, el gradiente es un vector igual a la derivada parcial de en respecto a cada uno de los pesos, asimismo el gradiente toma la dirección que determina el incremento más rápido en el error, mientras que la dirección opuesta, es decir, la dirección negativa, determina el decremento más rápido en el error. Por tanto, el error puede reducirse ajustando cada peso en la dirección:

$$-\sum_{p=1}^{P} \frac{\partial E^p}{\partial w_{ji}} \quad [28]$$

Al realizar este método surge un inconveniente es que el aprendizaje coincida con un mínimo local, sin embargo, es poco visto al usar datos en la realidad. (Marín, 2012)

D) Algoritmo de pesos de conexión (CW)

En 2002, Olden y Jackson crearon el algoritmo de conexión de pesos que visualiza la importancia relativa de las variables predictoras en contribución a la variable dependiente de una red neuronal. (Ibrahim, 2013)

El algoritmo de ponderaciones de conexión como la suma de productos de los pesos finales de las neuronas de la capa de entrada a las neuronas de la capa oculta. Ibrahim (2013) la importancia relativa de una variable de entrada está definida como:

$$RI_x = \sum_{y=1}^{m} w_{xy} w_{yz} \quad [29]$$

Dónde RI_x es la importancia relativa de la neurona de entrada x, $\sum_{y=1}^{m} w_{xy} w_{yz}$ es la suma del producto de los pesos finales de la conexión de la neurona de entrada a las

neuronas ocultas con la conexión de las neuronas ocultas a la neurona de salida, y es el número total de neuronas ocultas, y z es neuronas de salida.

CAPITULO II: MÉTODOS Y MATERIALES

A) Definición y operacionalización de variables

VARIABLE	DIMENSION	INDICADOR	ESCALA	INSTRUMENTO DE MEDICION
VARIABLE DEPENDIENTE				
DEPRESIÓN EN ADULTOS MAYORES	Sicológica	• 1 = Depresión moderada - severa • 0 = Depresión leve	Nominal	TEST DE YESAVAGE
VARIABLE INDEPENDIENTE				
EDAD	Cronológica	• Años cumplidos al momento de la encuesta	Razón	Encuesta
GÉNERO	Fisiológica	• Masculino • Femenino	Nominal	
INGRESO FAMILIAR	Económica	•No tiene ingresos •De 100 a 400 soles •400 a 800 soles •800 a 1200 soles •Más de 1200 soles	Nominal	
ACTIVIDAD LABORAL		•Si trabaja •No trabaja	Nominal	
ACTIVIDAD FISICA	Motriz	•Si realiza actividad •No realiza actividad	Nominal	
FRECUENCIA DE VIDA COITAL	Tiempo	•No •Algunas veces	Nominal	
ATENCIÓN FAMILIAR	Protección	•Si •No	Nominal	
PÉRDIDA DE AUTORIDAD FAMILIAR	Psicológica	•Si •No	Nominal	
PASATIEMPO	Recreación	•Si •No	Nominal	

B) Diseño de contrastación de hipótesis

La presente investigación es de nivel Observacional - Analítico de tratamiento multivariante. Además, es transversal, ya que se recolectará la información en un solo momento del tiempo único de datos multivariantes. Prospectivo.

C) Población y muestra

C.1. Población

Conjunto de adultos mayores que acuden al Servicio de Geriatría del Hospital Provincial Docente Belén y CIAM de Lambayeque durante el periodo de Marzo - Abril del 2019, y que se estima sean 128.

Criterio de inclusión

Ambos géneros

Pacientes del área de geriatría del hospital "Belén"

Desde 60 años a 90 años

Criterios de exclusión

Adultos mayores con problemas para responder el test y encuesta.

Adultos mayores que se niegan a responder los instrumentos de recolección de datos.

C.2. Muestra

La muestra de estudio no probabilística estará constituida por toda la población de estudio con sus criterios de inclusión y de exclusión, y que se estima sean 128.

D) Técnicas e instrumentos de recolección de datos

D.1.Técnica

Para la variable en estudio se utilizará la técnica de observación estructurada de evaluación psicométrica.

D.2. Instrumento

Se utilizará el Test de Yesavage que considera tres dimensiones: Estado de ánimo depresivo, la anergia o vaciamiento de impulsos y la discomunicación, cada ítem se valora como 0 o 1, según corresponda.

El test corto de Yesavage

- Si marcan con NO las preguntas 1, 5, 7, 11 y 13 es la respuesta correcta.
- De la misma forma si marcan con SI las preguntas 2, 3, 4, 6, 8, 9, 10, 12, 14, 15 es la respuesta correcta.
- Dando un valor de 1 a la respuesta correcta y de 0 a la respuesta incorrecta.

El puntaje total corresponde a la suma de los ítems, con un rango de 0 -15. Para ésta versión de 15 ítems los puntos de corte propuestos la escala es la siguiente:

- Depresión leve 0-5 puntos
- Depresión moderada 6-9 puntos
- Depresión severa 10-15 puntos

Este test fue validado por Bacca, Gonzáles y Uribe (2004). Para la estandarización se utilizaron las medidas de tendencia central y los coeficientes de correlación punto biserial para cada ítem. El coeficiente de confiabilidad de la escala es del 0,7268, indicando que la escala GDS-15 es altamente confiable; 14 de tos 15 ítems estadísticamente predicen moderadamente el constructo de depresión.

D.3. Confiabilidad

Confiabilidad

La confiabilidad significa precisión, consistencia, estabilidad en repeticiones. Una definición conceptual bastante ilustrativa indica que un instrumento es confiable si aplicado en las mismas condiciones a los mismos sujetos produce los mismos resultados. (Prieto y Delgado, 2010)

En el 2006 para Aribay, la confiabilidad está vinculada con los errores de medición, sin embargo si se maximiza el valor verdadero el instrumento será más confiable.

D.4. Análisis estadístico

Para determinar el pronóstico de depresión se utilizará el modelo estadístico de regresión logística dicotómica y la técnica matemática computacional de redes neuronales.

Para establecer la confiabilidad de consistencia interna del Test de Yesavage se utilizará el coeficiente de Kuder Richardson dado que dicho test en escala dicotómica.

Respecto a la construcción del Modelo de Regresión Logística, éste será dicotómico, considerando la depresión leve como la categoría de no daño y a las categorías moderada-severa como la categoría de daño (con depresión perjudicial para la salud) y se considerarán variables predictoras las indicadas en los objetivos específicos.

Respecto al Modelo de Red Neuronal se utilizará en su arquitectura una capa oculta, la función sigmoidal y como inicio se sembrará una semilla.

CAPITULO III: RESULTADOS Y DISCUSIÓN

RESULTADOS

Confiabilidad del test de Yesavage (Kuder Richardson 20)

Tabla 1
Confiabilidad del test

Alpha de Cronbach	0.70057
Número de ítems	15
Número de examinaciones	128

Fuente: Elaboración propia

Del resultado anterior aplicado al test de Yesavage se obtuvo un coeficiente de consistencia interna de los ítems igual a 0.70057, que indica una alta confiabilidad.

1. Lectura de datos en R

Se obtuvo el siguiente resultado:

Tabla 2
Variables de estudio en escalas

'data. frame': 128 obs. of 11 variables:				
Genero	Factor	w/2 levels	“Femenino”, “Masculino”	2 2 2 2 2 2…
Edad	Num		70 79 67 64 65 …	
Estado_civ	Factor	w/2 levels	“Con Pareja”, “Sin Pareja”	1 1 1 1 1 1…
Trabaja	Factor	w/2 levels	“Si”, “No”	1 1 1 1 1 2…
Ingreso	Factor	w/5 levels	“Más de 1200”, …	2 2 2 2 3 4…
Act_fisica	Factor	w/2 levels	“Si”, “No”	1 1 1 2 1 1…
Act_sexual	Factor	w/2 levels	“Algunas Veces”, “No”	1 1 1 1 1 1…
Atención_fam	Factor	w/2 levels	“Si”, “No”	2 1 1 1 2 1…
Perdida_aut	Factor	w/2 levels	“No”, “Si”	2 1 1 1 2 1…
Pasatiempo	Factor	w/2 levels	“Si”, “No”	1 2 1 2 2 1…
Depresion	Factor	w/2 levels	“Leve”, “Moderada – severa”	1 1 1 1 2 1…

Fuente: Elaboración propia

Este resultado nos permite visualizar a las variables consideradas para el estudio con sus escalas.

2. Análisis unidimensional de variables

Tabla 3

Códigos y descripción de las variables estudiadas

Variables	Unidades/ Valores que toma/ Codificación	Descriptivo
Género	-Femenino = 0	60 (46.9%)
	-Masculino = 1	68 (53.1%)
Edad	Años	Media: 76.3
		Min-Max: 60 - 89
		Mediana: 76
Estado Civil	-Con pareja = 0	40 (31.3%)
	-Sin pareja = 1	88 (68.7%)
Trabaja	-Si = 0	26 (20.3%)
	-No = 1	102 (79.7%)
Ingreso	-Más de 1200	2 (1.6%)
	-800 – 1200	9 (7.0%)
	-400 – 800	2 (1.6%)
	-100 – 400	30 (23.4%)
	-No tiene ingresos	85 (66.4%)
Actividad física	-Si = 0	59 (46.0%)
	-No = 1	69 (54.0%)
Actividad coital	-Algunas veces = 0	44 (34.4%)
	-No = 1	84 (65.6%)
Atención familiar	-Si = 0	84 (65.6%)
	-No = 1	44 (34.4%)
Pérdida de autoridad	-No = 0	81 (63.3%)
familiar	-Si = 1	47 (36.7%)
Pasatiempo	-Si = 0	76 (59.4%)
	-No = 1	52 (40.6%)
Depresión	-Leve	84 (65.6%)
	-Moderada - severa	44 (34.4%)

Fuente: Elaboración propia

De la tabla 3 se puede observar que la mayoría de los encuestados son de género masculino (53.1%), asimismo la edad promedio fue de 77 años. El 68.5% no tenían pareja y el 79.7% no realiza ninguna actividad laboral.

Con respecto al ingreso se puede observar que el 66.4% no cuenta con ingreso económico mensual, el 54% no realiza actividades físicas y el 65.6% manifestaron no tener actividad coital.

Se puede observar que el 65.6% no cuenta con atención familiar, asimismo el 63.3% siente que ha perdido la autoridad en su hogar. El 59.4% no cuenta algún pasatiempo, y el 65.6% tiene depresión leve, de acuerdo al test Yesavage.

3. Análisis de variables cualitativas en R

A continuación, mostramos el resultado del análisis bidimensional, donde vemos las relaciones entre dos variables, en especial entre los niveles de depresión y otras variables:

Tabla 4
Relación del tipo de depresión según el género

	Género	
Depresión	Femenino	Masculino
Leve	48	36
Moderada – Severa	12	32
Pearson's Chi-squared test Data: Table		
X-squared = 10.346	df = 1	p-value = 0.001298

Fuente: Elaboración propia

Del resultado anterior se observó que, dado el test de Chi-Cuadrado existe asociación entre el género y los niveles de depresión.

En la siguiente tabla, mostramos la relación de todas las variables categóricas con los niveles de depresión.

Tabla 5

Relación de las variables estudiadas según el tipo de depresión.

	DEPRESIÓN				
	Leve	**Moderada – Severa**	**Total**	**Chi-Cuadrado(X^2)**	**P-valor**
Género					
Femenino	48(80%)	12(20%)	60(47%)	10.35	0.001
Masculino	36(53%)	32(47%)	68(53%)		
Estado civil					
Con pareja	27(68%)	13(32%)	40(31%)	0.09	0.763
Sin pareja	57(65%)	31(35%)	88(69%)		
Actividad laboral					
Si	15(58%)	11(42%)	26(20%)	0.91	0.34
No	69(68%)	33(32%)	102(80%)		
Ingreso					
Más de 1200(*)	9(82%)	2(18%)	11(9%)		
800 – 1200(*)				5.53	0.24
400 – 800(**)	19(59%)	13(41%)	32(25%)		
100 – 400(**)					
No tiene ingresos	56(66%)	29(34%)	85(66%)		
Actividad física					
Si	31(53%)	28(47%)	59(46%)	8.30	0.004
No	53(77%)	16(23%)	69(54%)		
Actividad coital					
Algunas veces	36(82%)	8(18%)	44(34%)	7.79	0.005
No tiene	48(57%)	36(33%)	84(66%)		
Atención familiar					
Si	76(90%)	8(10%)	84(66%)		
No	8(18%)	36(82%)	44(34%)	66.90	<0.001
Pérdida de autoridad familiar					
No	73(90%)	8(10%)	81(63%)	56.69	<0.001
Si	11(23%)	36(77%)	47(37%)		
Pasatiempo					
Si	66(87%)	10(13%)	76(59%)	37.33	<0.001
No	18(35%)	34(65%)	52(41%)		

p: Test de independencia de Chi-cuadrado. Nivel de significación 0.05.

Fuente: Elaboración propia

De la tabla 5, se concluye que con excepción de las variables estado civil, actividad laboral y el ingreso económico mensual (p>0.05), las demás variables se asocian con los niveles de depresión.

El género masculino (47%), no tener actividad física (23%), no tener actividad coital (33%), no contar con atención familiar (82%), creer que ha perdido autoridad (77%), no tener pasatiempo (65%) fueron las categorías que se correspondieron con una depresión moderada o severa.

a) Ilustración gráfica de prevalencia de depresión y variables independientes

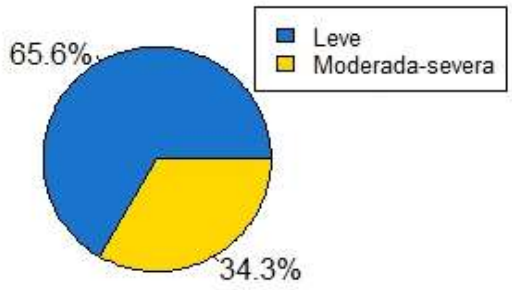

Figura 9. Prevalencia de depresión en toda la muestra. Elaboración propia

De la figura 9, se observó el porcentaje en los niveles de depresión en toda la muestra, siendo el 65.6% de adultos mayores con depresión leve.

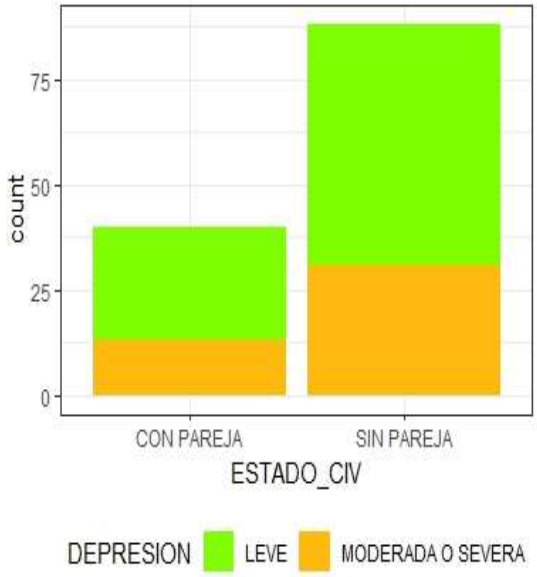

Figura 10. Prevalencia de depresión según el estado civil. Elaboración propia

De la figura 10, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue mayor en aquellos que no cuentan con pareja.

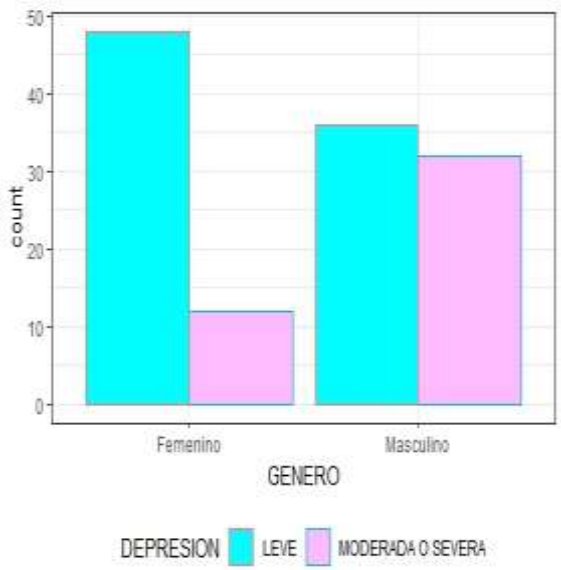

Figura 11. Prevalencia de depresión por género. Elaboración propia

De la figura 11, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue mayor en aquellos son del género masculino.

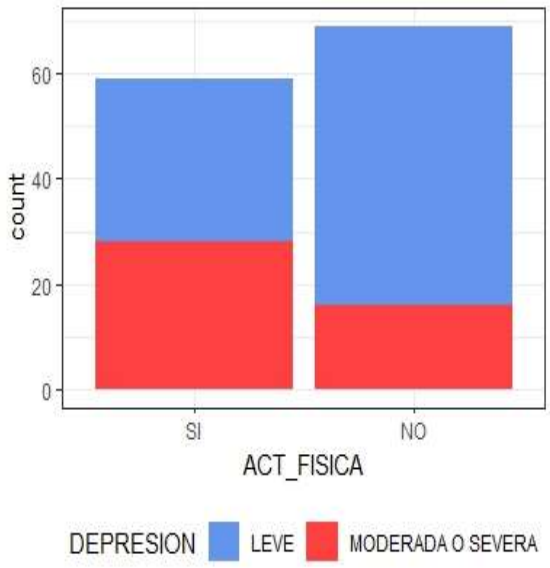

Figura 12. Prevalencia de depresión respecto a la actividad física. Elaboración propia

De la figura 12, se observó que la prevalencia de depresión moderada-severa en adultos mayores respecto a la actividad física fue mayor en aquellos que si realizan alguna actividad física.

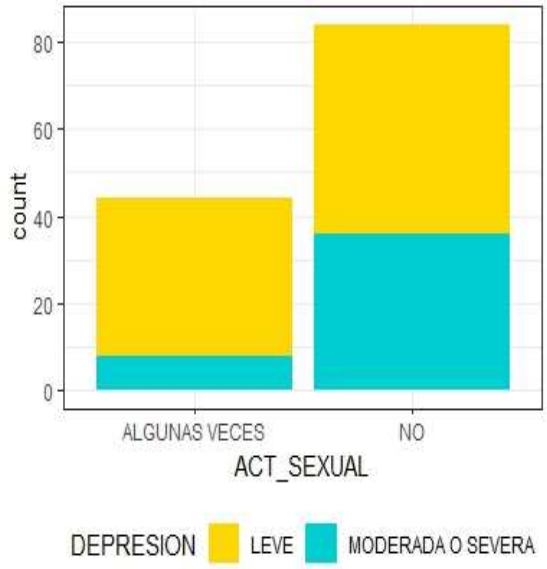

Figura 13. Prevalencia de depresión según la actividad sexual. Elaboración propia

De la figura 13, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue mayor en aquellos que no tenían relaciones coitales, respecto a la variable actividad sexual.

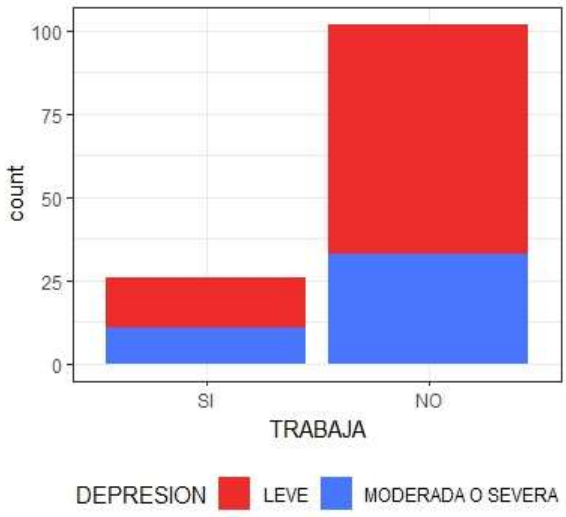

Figura 14. Prevalencia de depresión según la actividad laboral. Elaboración propia

De la figura 14, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue más resaltante en aquellos adultos mayores que no contaban con trabajo.

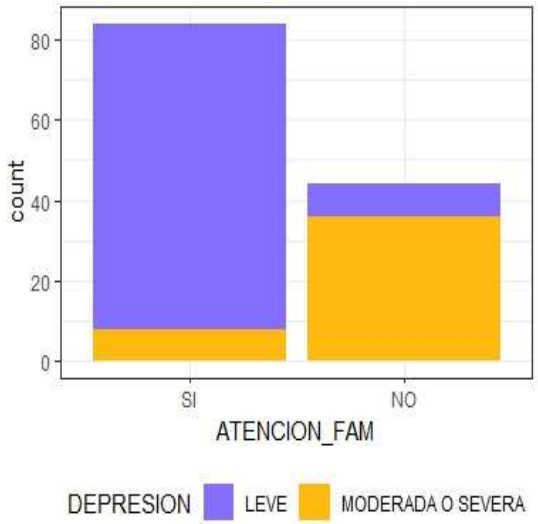

*Figura 1*5. Prevalencia de depresión respecto a la atención familiar. Elaboración propia

De la figura 15, se observó que la prevalencia de depresión moderada-severa en adultos mayores respecto a la variable atención familiar fue más evidente en aquellos adultos mayores que no cuentan con ninguna atención.

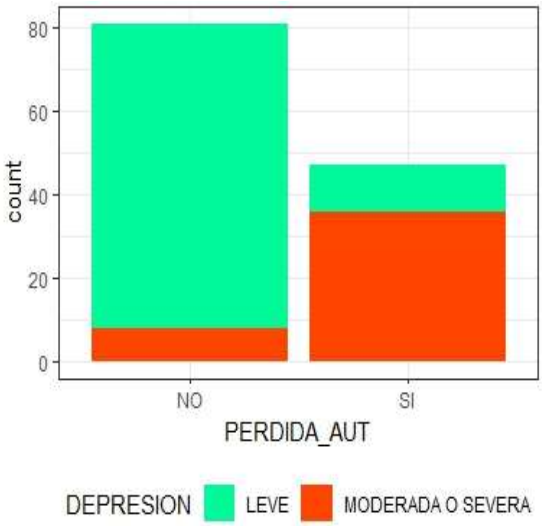

Figura 16. Prevalencia de depresión respecto a la pérdida de autoridad familiar. Elaboración propia

De la figura 16, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue mayor en aquellos que sentían que si habían perdido la autoridad en su familia.

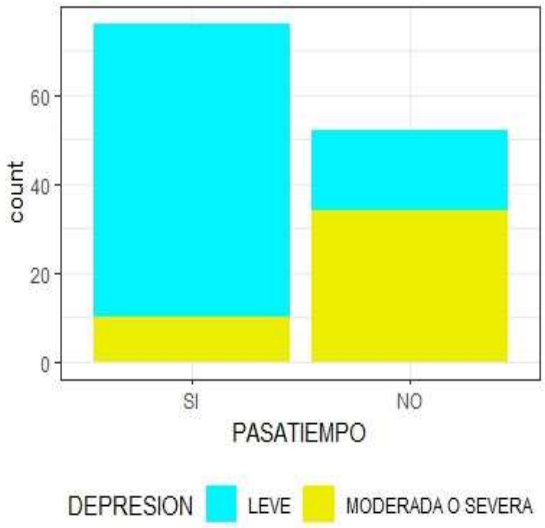

Figura 17. Prevalencia de depresión respecto al pasatiempo. Elaboración propia

De la figura 17, se observó que la prevalencia de depresión moderada-severa en adultos mayores fue más consistente en aquellos adultos mayores que no contaron con algún pasatiempo.

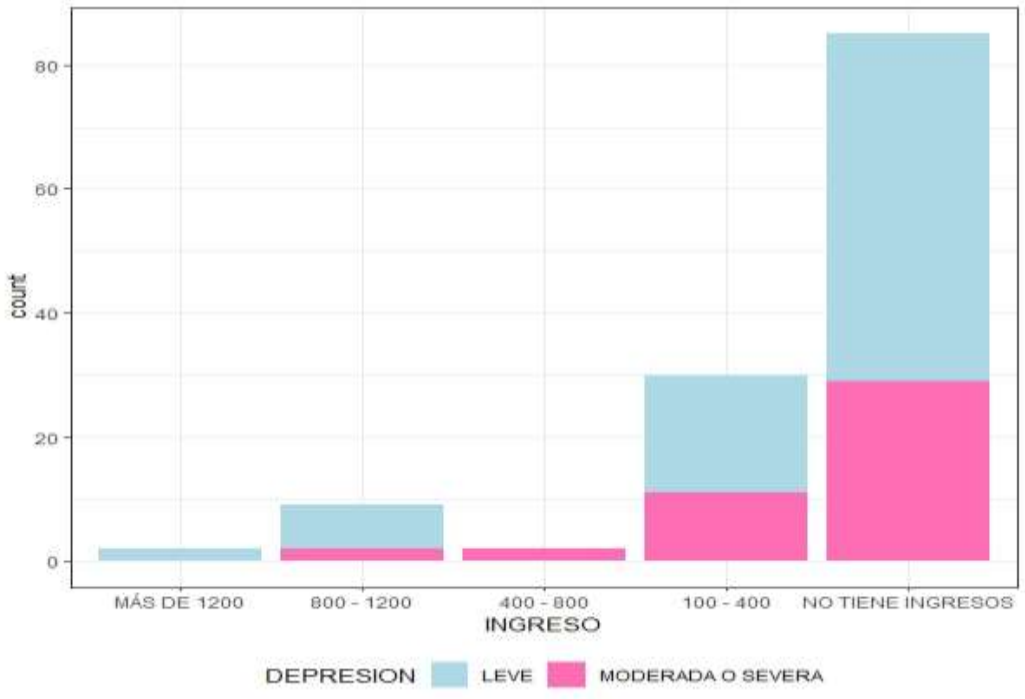

Figura 18. Prevalencia de depresión respecto al ingreso familiar. Elaboración propia

Los resultados anteriores de cada figura valida los resultados de la tabla 9, donde la prevalencia de Depresión Leve (65,6%) en adultos mayores de toda la muestra fue mayor significativamente al factor de riesgo de Depresión Moderada-

Severa (34,4%) con p<0,05. Asimismo, se puede visualizar un mayor porcentaje de Depresión Leve para cada una de las variables de las diferentes figuras.

4. Partición de datos en entrenamiento y prueba

La muestra de entrenamiento con la que se construirá los modelos y la muestra de prueba con la que se evaluaran los pronósticos, de acuerdo a los códigos del Rstudio que se indican en el anexo 3.

4.1. Verificación de la estructura de los datos particionados

Se obtuvo la siguiente salida:

Tabla 6
Pronóstico de los datos de entrenamiento

	Leve	**Moderada – Severa**
Entrenamiento	0.6538462	0.3461538
Prueba	0.6666667	0.3333333

Fuente: Elaboración propia

En la tabla 6, se observó el porcentaje de la división de datos de toda la muestra, considerando un 80% para los datos de entrenamiento y 20% para los datos de prueba.

5. Método Wrapper

5.1. Eliminación recursiva de variables

La eliminación recursiva es una estrategia que evita la búsqueda exhaustiva de combinaciones que a su vez genera un ranking de variables, mientras se va descartando las variables con menos puntaje.

Las variables seleccionadas serán comparadas con la importancia de los modelos de Regresión Logística y Red Neuronal.

Se obtuvo la siguiente salida:

Tabla 7
Variables óptimas mediante el método de remuestreo bootstraping

Recursive feature selection Outer resampling method: Bootstrapped (20 reps)				
Variables	**Accuracy**	**Kappa**	**AccuracySD**	**KappaSD**
1	0.8344	0.6448	0.06094	0.1449
2	0.8733	0.7221	0.06800	0.1482
3	0.8761	0.7258	0.04758	0.1196
4	0.8502	0.6720	0.05961	0.1430
5	0.8525	0.6808	0.07026	0.1595
6	0.8591	0.6909	0.06051	0.1413
7	0.8588	0.6913	0.06163	0.1424
8	0.8602	0.6919	0.05207	0.1215
9	0.8493	0.6716	0.06445	0.1481
10	0.8479	0.6683	0.06283	0.1464
11	0.8532	0.6787	0.62259	0.1479
12	0.8504	0.6715	0.06093	0.1434
13	0.8543	0.6799	0.05526	0.113
The top 3 variables (out of 3): **PERDIDA_AUTSI, ATENCION_FAMNO, PASATIEMPONO**				

Fuente: Elaboración propia

6. **Regresión logística:** Selección de variables a través de AIC para el modelo de Regresión Logística

El siguiente reporte permitió identificar las variables más importantes del mejor modelo de ajuste de regresión logística de acuerdo al coeficiente de información de Akaike (AIC).

Se obtuvo la siguiente salida**:**

Tabla 8
Identificación de las variables más importante según Akaike

Star AIC= 136.17 DEPRESION ~ 1			
	Df	Deviance	AIC
+ ATENCION_FAM	1	80.551	84.551
+ PERDIDA_AUT	1	82.156	86.156
+ PASATIEMPO	1	106.499	110.499
+ACT_SEXUAL	1	127.295	131.295
+ GENERO	1	128.662	132.662
+ ACT_FISICA	1	129.817	133.817
<none>		134.167	136.167
+ TRABAJA	1	133.687	137.687
+ ESTADO_CIV	1	134.089	138.089
+ EDAD	1	134.115	138.115
+ INGRESO	4	129.638	139.638
Star AIC= 84.55 DEPRESION ~ ATENCION_FAM			
	Df	Deviance	AIC
+ PASATIEMPO	1	70.826	76.826
+ PERDIDA_AUT	1	74.464	80.464
+ TRABAJA	1	77.279	83.279

+ ACT_FISICA	1	77.687	83.678
+ ACT_SEXUAL	1	78.077	84.077
<none>		80.551	84.551
+ GENERO	1	78.582	84.582
+ EDAD	1	80.526	86.526
+ ESTADO_CIV	1	80.538	86.538
+ INGRESO	4	75.792	87.792
- ATENCION_FAM	1	134.167	136.167

Star AIC= 76.83
DEPRESION ~ ATENCION_FAM + PASATIEMPO

	Df	Deviance	AIC
+ PERDIDA_AUT	1	67.432	75.432
+ ACT_FISICA	1	67.947	75.947
<none>		70.826	76.826
+ TRABAJA	1	69.726	77.726
+ ACT_SEXUAL	1	69.859	77.859
+ INGRESO	4	64.185	78.185
+ GENERO	1	70.291	78.291
+ EDAD	1	70.819	78.819
+ ESTADO_CIV	1	70.821	78.821
- PASATIEMPO	1	80.551	84.551
- ATENCION_FAM	1	106.499	110.499

Star AIC= 75.43
DEPRESION ~ ATENCION_FAM + PASATIEMPO + PERDIDA_AUT

	Df	Deviance	AIC

+ ACT_FISICA	1	64.237	74.237
<none>		67.432	75.432
+ TRABAJA	1	65.806	75.806
+ INGRESO	4	60.655	76.655
+ ACT_SEXUAL	1	66.770	76.770
-PERDIDA_AUT	1	70.826	76.826
+ GENERO	1	67.008	77.008
+ EDAD	1	67.294	77.294
+ ESTADO_CIV	1	67.368	77.368
- ATENCION_FAM	1	73.383	79.383
-PASATIEMPO	1	74.464	80.464

Star AIC= 74.24
DEPRESION ~ ATENCION_FAM + PASATIEMPO + PERDIDA_AUT + ACT_FISICA

	Df	Deviance	AIC
+ INGRESO	4	55.357	73.357
<none>		64.237	74.237
+ ACT_SEXUAL	1	62.838	74.838
-ACT_FISICA	1	67.432	75.432
+ GENERO	1	63.764	75.764
+ TRABAJA	1	63.933	75.933
-PERDIDA_AUT	1	67.947	75.947
+ ESTADO_CIV	1	64.133	76.133
+ EDAD	1	64.218	76.218
- ATENCION_FAM	1	69.691	77.691
-PASATIEMPO	1	71.569	79.569

Star AIC= 73.36
DEPRESION ~ ATENCION_FAM + PASATIEMPO + PERDIDA_AUT + ACT_FISICA + INGRESO

	Df	Deviance	AIC
+ TRABAJA	1	52.355	72.355
<none>		55.357	73.357
+ GENERO	1	54.194	74.194
-INGRESO	4	64.237	74.237
+ ACT_SEXUAL	1	54.625	74.625
+ ESTADO_CIV	1	54.784	74.784
-PERDIDA_AUT	1	59.304	75.304
+ EDAD	1	55.313	75.313
-ACT_FISICA	1	60.655	76.655
-ATENCION_FAM	1	62.454	78.454
-PASATIEMPO	1	65.453	81.453

Star AIC= 72.36
DEPRESION ~ ATENCION_FAM + PASATIEMPO + PERDIDA_AUT + ACT_FISICA + INGRESO + TRABAJA

	Df	Deviance	AIC
<none>		52.355	72.355
+ GENERO	1	50.715	72.15
-ACT_FISICA	1	54.932	72.932
+ ACT_SEXUAL	1	51.305	73.305
-TRABAJA	1	55.357	73.357
+ ESTADO_CIV	1	51.640	73.640
+ EDAD	1	51.710	73.710
-PERDIDA_AUT	1	56.637	74.637

-INGRESO	4	63.933	75.933
-PASATIEMPO	1	59.535	77.535
-ATENCION_ FAM	1	60.680	78.680

Fuente: Elaboración propia

Utilizando el método de pasos hacia adelante, y de acuerdo al coeficiente de determinación de Akaike cuyo menor valor fue de 72.36, el modelo de regresión logística tiene contiene las variables ATENCION_FAM, PASATIEMPO, PERDIDA_AUT, ACT_FISICA, INGRESO, TRABAJA

a) Ilustración gráfica

a.1. Variables más importantes a partir del valor absoluto del estadístico t para cada parámetro.

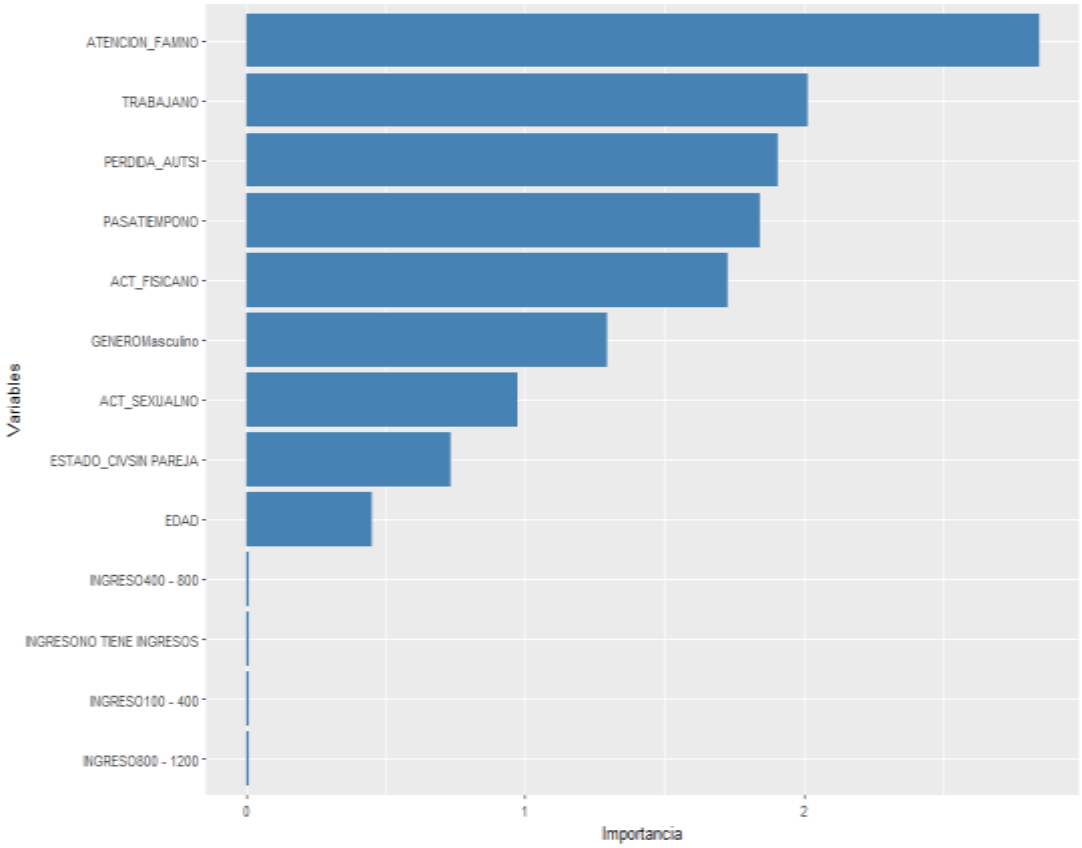

Figura 19. Orden de importancia de todas las variables. Elaboración propia

De la figura 19, se observó que la variable más importantes de acuerdo al estadístico t fue: No tener Atención Familiar y siendo la menos importante el ingreso económico.

b) Variables más importantes que contribuyen con la explicación de una mayor o menor depresión de acuerdo al valor absoluto T- student calculado, para el mejor modelo hallado de acuerdo al criterio de Akaike.

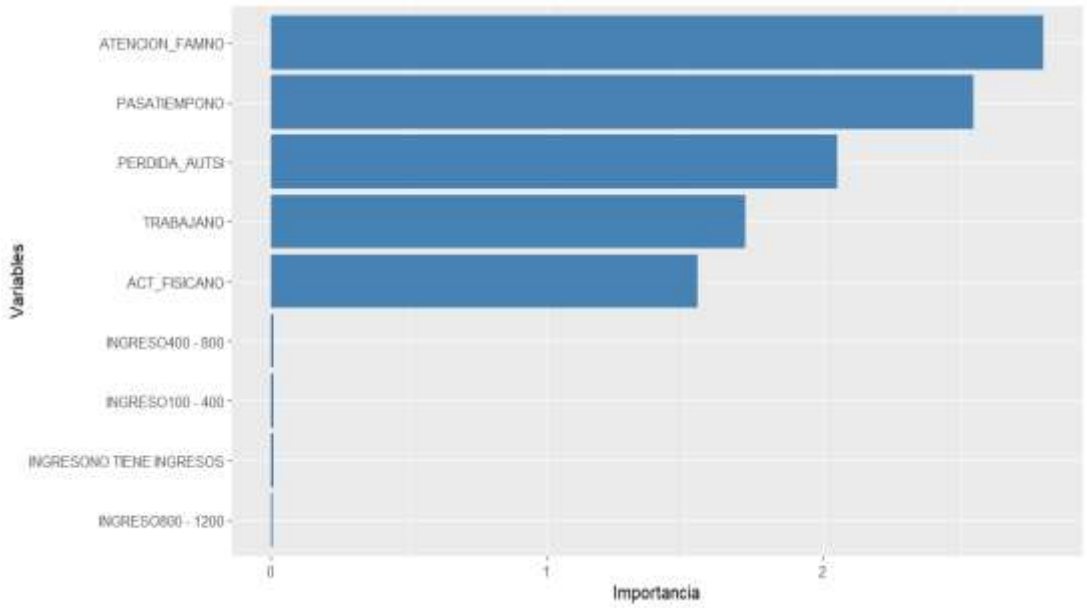

Figura 20. Orden de importancia de las 6 variables seleccionadas como las más influyentes de acuerdo al AIC. Elaboración propia

De la figura 20 se observa que las variables y sus categorías más importantes o influyentes en la variable depresión, categoría moderada – severa fueron: atención familiar: no, pasatiempo: no, pérdida de autoridad: si, actividad física: no, actividad laboral: no, ingreso económico mensual: 100-400, 400-800, 800-1200 y no tiene ingreso económico mensual. Éstas últimas categorías de ingreso como variables artificiales (dummy) aportan muy poca información para discriminar los deprimidos leves de los depresivo moderados- severos.

7. Significancia de las variables seleccionadas de Regresión Logística

a) Modelo de Regresión Logística de acuerdo a su significancia usando todas las variables de estudio (6 variables), se obtuvo la siguiente salida:

Tabla 9

Significancia de los coeficientes de las variables del mejor modelo según Akaike

Call:

glm(formula = Depresion ~ Atencion_fam + Pasatiempo + Perdida_aut + Act_fisica + Ingreso + Trabaja, family = binomial(link = logit), data = DATOS_E)

	Coefficients	Error	z value	Pr(>|z|)
(Intercep)	-23.3594	2399.5451	-0.010	0.9922
Atencion_famno	2.8416	1.0147	2.800	0.0051 **
Pasatiempono	2.0019	0.7858	2.547	0.0109 *
Perdida_autsi	1.9499	0.9492	2.054	0.0400 *
Act_fisicano	-1.3105	0.8471	-1.547	0.1219
Ingoreso800 – 1200	19.1557	2399.5452	0.008	0.9936
Ingreso400 – 800	33.1321	3393.4687	0.010	0.9922
Ingreso100 – 400	22.0239	2399.5451	0.009	0.9927
Ingresono tiene ingresos	21.8545	2399.5451	0.009	0.9927
Trabajano	-1.8174	1.0568	-1.720	0.0855

Signif. codes: 0 ‘*’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

Deviance Residuals:

Min	IQ	Median	3Q	Max
-2.1305	-0.3079	-0.1391	0.2474	2.5278

Fuente: Elaboración propia

Del resultado anterior se observó que solo tres coeficientes son significativos, de las 6 variables seleccionadas, que al ser comparados con las categorías más importantes de acuerdo AIC, coinciden con las tres más importantes, cuyas variables son: atención familiar, pasatiempo y pérdida de autoridad.

El 50% de las desviaciones de los residuales del modelo fueron menores o iguales a -0.1391, siendo el mayor residual igual a 2.5278.

b) Modelo de Regresión Logística excluyendo la variable Ingreso, de acuerdo a la significancia de los coeficientes, se obtuvo la siguiente salida:

Tabla 10

Significancia de los coeficientes excluyendo la variable ingreso

Call:

glm(formula = Depresion ~ Atencion_fam + Pasatiempo + Perdida_aut + Act_fisica + Trabaja, family = binomial(link = logit), data = DATOS_E)

	Coefficients	Error	z value	Pr(>|z|)
(Intercep)	-2.3493	0.7652	-3.070	0.00214 **
Atencion_famno	2.1721	0.9056	2.398	0.01647 *
Pasatiempono	1.6543	0.6840	2.419	0.01558 *
Perdida_autsi	1.7873	0.8892	2.010	0.04442 *
Act_fisicano	-0.9949	0.7432	-1.339	0.17067
Trabajano	-0.4868	0.8853	-0.550	0.58242

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Deviance Residuals:

Min	IQ	Median	3Q	Max
-2.5698	-0.4270	-0.2072	0.3474	2.4055

Fuente: Elaboración propia

Excluyendo la variable ingreso se obtuvo el mismo resultado, en cuanto a la significancia de los coeficientes de las tres variables más importantes, siendo estas: atención familiar, pasatiempo y pérdida de autoridad.

Los coeficientes de las variables: atención familiar, pasatiempo y pérdida de autoridad conservaron su no significancia ($p>0.05$).

El 50% de las desviaciones de los residuales del modelo fueron menores o iguales a -0.2027, siendo el mayor residual igual a 2.4055.

c) Modelo de Regresión Logística de acuerdo a su significancia, excepto las variables Ingreso y Trabaja, se obtuvo la siguiente salida:

Tabla 11

Significancia de los coeficientes excluyendo la variable ingreso y trabaja

Call:

glm(formula = Depresion ~ Atencion_fam + Pasatiempo + Perdida_aut + Act_fisica, family = binomial(link = logit), data = DATOS_E)

	Coefficients	**Error**	**z value**	**Pr(>|z|)**
(Intercep)	-2.6099	0.6143	-4.248	2.15e-05 **
Atencion_famno	2.0818	0.8853	2.351	0.01870 *
Pasatiempono	1.7548	0.6597	2.660	0.00781 *
Perdida_autsi	1.7408	0.8831	1.971	0.04870 **
Act_fisicano	-1.1693	0.6766	-1.728	0.08398

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Deviance Residuals:

Min	**IQ**	**Median**	**3Q**	**Max**
-2.4567	-0.3767	-0.2125	0.3167	2.3155

Fuente: Elaboración propia

Ahora excluyendo las variables ingreso y trabaja, los coeficientes de las tres variables más importantes conservaron su significancia estadística ($p<0.05$), con excepción de la variable actividad física cuyo p-valor fue mayor a 0.05 ($p=0.08398$).

El 50% de las desviaciones de los residuales del modelo fueron menores o iguales a -0.2125, siendo la mayor desviación residual igual a 2.3155

d) Modelo de Regresión Logística de acuerdo a su significancia, excepto la variable Ingreso, Trabaja y actividad Física, se obtuvo la siguiente salida:

Tabla 12

Significancia de los coeficientes excluyendo la variable ingreso, trabaja y actividad física

Call:

glm(formula = Depresion ~ Atencion_fam + Pasatiempo + Perdida_aut , family = binomial(link = logit), data = DATOS_E)

	Coefficients	**Error**	**z value**	**Pr(>|z|)**
(Intercep)	-3.1157	0.5721	-5.446	5.14e-08 ***
Atencion_famno	2.1265	0.38607	2.471	0.0135 *
Pasatiempono	1.6832	0.6416	2.624	0.0087 **
Perdida_autsi	1.6086	0.8526	1.887	0.0592 .

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Deviance Residuals:

Min	**IQ**	**Median**	**3Q**	**Max**
-2.1899	-0.2946	-0.2946	0.4366	2.5136

Fuente: Elaboración propia

Se observó que solo dos coeficientes son significativos correspondientes a las variables atención familiar y pasatiempo. El coeficiente de la variable pérdida de autoridad resultó ser no significativa al 95% de confiabilidad (p=0.0592). Aunque si lo fue al 94% de certeza.

El 50% de las desviaciones de los residuales del modelo fueron menores o iguales a -0.2946, siendo la mayor desviación residual igual a 2.5136.

Considerando el criterio de Parsimonia, el modelo de regresión logística resultante es el que se indica a continuación:

Tabla 13
Significancia de los coeficientes excluyendo la variable ingreso

Call: glm(formula = Depresion ~ Atencion_fam + Pasatiempo, family = binomial(link = logit), data = DATOS_E)		
Coefficients:		
(Intercep)	**Atencion_famno**	**Pasatiempono**
-3.004	3.279	1.886
Degrees of Freedom: 103 Total (i.e. Null); 101 Residual		
Null Deviance: 134.2		
Residual Deviance: 70.83 AIC: 76.83		

Fuente: Elaboración propia

8. Modelo final de Regresión Logística

A partir del resultado anterior se crea el modelo de Regresión Logística, obteniendo lo siguiente:

$$\mathbf{p} = 1/(1 + e^{-(-3.004 + 3.279 * \boldsymbol{ATENCION\_FAM} + 1.886 * \boldsymbol{PASATIEMPO})})$$

9. Evaluación del modelo

Para determinar si los predictores del modelo final contribuyen de forma significativa se emplea el test Wald, donde:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Obteniendo la siguiente salida:

Tabla 14
Evaluación del modelo mediando el test Wald

Analysis of Deviance Table (Type II tests) Response: DEPRESION			
	Df	Chisq	Pr(>Chisq)
Atencion_fam	1	34.007	5.298e-09 ***
Pasatiempo	1	11.193	0.0008209 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

Fuente: Elaboración propia

Del resultado anterior al aplicar la prueba de Wald en la prueba conjunta a los parámetros, se obtuvo que la probabilidad es menor que el nivel de significancia ($\alpha = 0.5$), por lo cual se rechaza la hipótesis nula de irrelevancia de los indicadores en forma conjunta.

10. Coeficiente de determinación R^2

El siguiente reporte nos permite visualizar la evaluación del modelo de ajuste a través del coeficiente de determinación de Nagelkerke.

Se obtuvo la siguiente salida:

Tabla 15
Resultados de la evaluación de modelos según Nagelkerke

Model: "glm, Depresion ~ Atencion_fam + Pasatiempo, binomial(link = logit), DATOS_E"
Null: "glm, DEPRESION ~ 1, binomial(link = logit), DATOS_E"

$Pseudo.R.squared.for.model.vs.null

	Pseudo. R. squared
McFadden	0.472107
Cox and Snell (ML)	0.456133
Nagelkerke (Cragg and Uhler)	**0.629368**

$Likelihood.ratio.test

Df.diff	**LogLik.diff**	**Chisq**	**p.value**
-2	-31.671	63.341	1.7605e-14

Fuente: Elaboración propia

El 62.94% de la variabilidad de la variable depresión es explicada por la variabilidad producida en las variables independientes atención familiar y pasatiempo, por lo que se concluye que el ajuste del modelo hacia los datos es bueno, de acuerdo al coeficiente de determinación de Nagelkerke.

11. Odds ratios e Intervalos de Confianza

Se obtuvo la siguiente salida:

Tabla 16
Riesgo relativo e intervalos de confianza para identificar factores de riesgo

	Coef	Pr(>|z|)	OR	2.5%	97.5%
(Intercep)	-3.0036	1.18e-07 ***	0.049607	0.0139821	0.133085
Atencion_famno	3.2790	1.29e-07 ***	26.548689	8.4106985	98.816189
Pasatiempono	1.8864	0.00248 **	6.595682	2.0115092	24.213120

Fuente: Elaboración propia

Del resultado anterior, se observó que todas las categorías de las variables consideradas de riesgo resultaron con ratios de riesgo mayores de 1, con intervalos de confianza al 95% que lo excluyen; lo que indica que los factores o variables atención familiar no y pasatiempo son considerados de riesgo de padecer depresión moderada – severa.

Cuando se interpreta los riesgos (OR) de cada variable, se asume que el resto de variables independientes se mantienen fijas. (Dueñas, 2006) Se interpretan a continuación las ratios de cada una de las variables independientes respecto a la categoría considerada como daño de la variable dependiente (y=1: depresión moderada-severa).

Los adultos mayores que no tienen atención familiar, tienen 26.549 veces más probabilidad de tener depresión moderada o severa respecto a aquellas que cuentan con atención familiar, resultando un factor de riesgo significativo al 95% de confiabilidad. Este resultado es respaldado por su intervalo de confianza [8.411, 98.816], ya que al no contener el 1, se considera un factor de riesgo.

Los adultos mayores que no realizan algún pasatiempo, tienen 6.596 veces más probabilidad de tener depresión moderada o severa respecto a aquellas que cuentan con algún pasatiempo, resultando un factor de riesgo significativo al 95% de confiabilidad, el intervalo de confianza del riesgo [2.011, 24.213] no contiene al 1.

12. Ajuste del modelo

Ho: No existe diferencia entre los valores observados y los valores pronosticados.

H_1: Existe diferencia entre los valores observados y los valores pronosticados.

Se obtuvo el siguiente resultado:

Tabla 17

Significancia del modelo según Hosmer Lemeshow

Hosmer and Lemeshow test (binary model)		
data: DATOS_E$DEPRESION, fitted(ModReg)		
X-squared = 0.38521	df = 1	p-value = 0.5348

Fuente: Elaboración propia

Se observó en los resultados que, no existe diferencia significativa entre los valores observados y esperados, en conclusión, el modelo ajusta bien a los datos.

13. Métricas de evaluación

13.1.Matriz de confusión en R usando datos de entrenamiento

Se obtuvo la siguiente salida:

Tabla 18

Matriz de confusión

	PRONOSTICADO		
OBSERVADO	Moderada - Severa	Leve	TOTAL
Moderada - Severa	29	7	36
Leve	7	61	68
TOTAL	36	68	104

Fuente: Elaboración propia

Precisión = (29+61) /104= **85.3%**

De la tabla 18, se observó que el porcentaje de clasificación correcta usando los datos de entrenamiento fue de 85.3%, es decir, los adultos mayores con depresión leve o moderada-severa serán clasificados correctamente en un 85.3% de los casos.

13.2. Matriz de confusión en R usando datos de prueba

El siguiente reporte permite calcular las diferentes métricas del modelo.

Tabla 19
Matriz de confusión

	PRONOSTICADO		
OBSERVADO	Moderada – Severa	Leve	TOTAL
Moderada - Severa	7	1	8
Leve	1	15	16
TOTAL	8	16	24

Fuente: Elaboración propia

Precisión = (7+15) /24= **91.67%**

De la tabla 19, se observó que el porcentaje de clasificación correcta usando los datos de entrenamiento fue de 91.67%, es decir, los adultos mayores con depresión leve o moderada-severa serán clasificados correctamente en un 91.67% de los casos

13.3. Métricas de evaluación en R

Se obtuvo los siguientes resultados a partir de la matriz de clasificación**:**

Tabla 20
Métricas del modelo de Regresión Logística

TASAS	VALOR
Precisión	91.67%
VPP	87.50%
VPN	93.75%
TVP (Sensibilidad)	87.50%
TVN (Especificidad)	93.75%
TFP	6.25%
TFN	12.50%
Error	8.33%
F1-Score	89.53%

Fuente: Elaboración propia

De la tabla 20, se observa que el porcentaje de clasificación correcta dada por el modelo es de 91.67%, lo que significa, que los adultos mayores pronosticados con depresión Leve o Moderada-severa serán clasificados correctamente como tales en el 91,67% de los casos. Si el test tiene resultado positivo (depresión moderada o severa) la probabilidad de que realmente sea así es de 87.50%; y si el test tiene resultado negativo (depresión leve) se tendrá la probabilidad (93.75%) de que así será.

La capacidad que tiene el modelo para captar a los adultos mayores con depresión moderada o severa es de 87.50%, lo que quiere decir, que de 100 verdaderos depresivos moderados o severos la prueba identifica como tal a 88 aproximadamente. La capacidad que tiene el modelo para identificar al paciente sin depresión moderada o severa es del 93.75%, lo que quiere decir, que de 100 pacientes con depresión leve la prueba detectará como tal a 94.

La probabilidad que tiene el modelo para diagnosticar un resultado positivo (depresión moderada o severa) estando realmente sano (depresión leve) es baja (6.25%); estos casos se les conoce como falsos positivos.

La probabilidad que tiene el modelo para diagnosticar personas con depresión leve teniendo realmente depresión moderada – severa es del 12.5%., estos casos son conocidos como falsos negativos

El rendimiento del modelo (F1score) para clasificar correctamente la depresión es de 89.53% y el error de pronóstico fue de 8.33%.

14. Curva ROC (Regresión Logística)

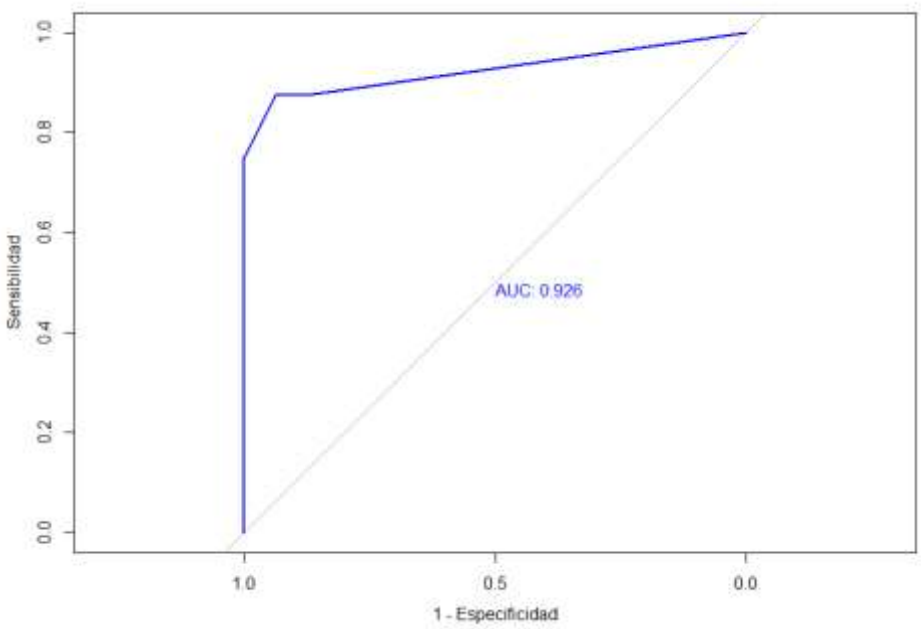

Figura 21. Área bajo la curva ROC para Depresión en adultos mayores. Elaboración propia

De la figura 21, se observó el coeficiente del área bajo la curva ROC fue de 0.926, es decir que de dos individuos uno con depresión leve y otro con depresión moderada-severa, la prueba (regresión logística) los detectará correctamente a ambos.

Tabla 21

Área bajo la curva. Evaluación de un método para determinar depresión

Área	Significación	95%IC
0.926	**0.00**	**0.7878 – 1**

Fuente: Elaboración propia

De la tabla 21 se observa que el área es de 0.926 con un intervalo de confianza que no contiene a 0.05 y significación 0.00, es decir estos aspectos indican que el modelo de regresión logística puede diagnosticar con alta precisión.

14.1. Resumen de resultados del modelo de Regresión Logística

Tabla 22
Resumen de resultados del mejor modelo de ajuste parsimonioso en comparación con los modelos analizados

	Coef	p-value	Precisión	R^2	AIC	Resid - Dev	ROC
Modelo 1 (2 variables)	3.2790	1.29e-07	91.667	0.629	76.83	70.83	92.6
	1.8864	0.00248					
Modelo 2 (3 variables)	2.1265	0.0135	91.667	0.653	75.43	67.43	92.2
	1.6832	0.0087					
	1.6086	0.0592					
Modelo 3 (6 Variables)	2.8416	0.0051	91.667	0.751	72.36	52.36	95.3
	2.0019	0.0109					
	1.9499	0.0400					
	-1.8174	0.0855					
	-1.3105	0.1219					
	19.1557	0.9936					
	33.1321	0.9922					
	22.0239	0.9927					
	21.8545	0.9927					

Nota: El modelo 3 de RL de seis variables cuyos indicadores son: actividad física, pérdida de autoridad, Ingreso económico, actividad laboral, pasatiempo y atención familiar; el modelo 2 de tres variables cuyos indicadores son: pérdida de autoridad, pasatiempo y atención familiar; El modelo 1 de dos variables cuyos indicadores son: pasatiempo y atención familiar.

Del resultado anterior se observa que los tres modelos resultantes obtuvieron muy idénticas precisiones, similares probabilidades bajo la curva ROC, parecidos coeficientes de determinación y Akaike, lo que indica que cualquiera de los tres

modelos de regresión logística podría ser utilizado para pronósticos de depresión en adultos mayores.

Sin embargo, los autores por parsimonia, presentan como modelo final de pronóstico de depresión en adultos mayores el modelo 1 que incluye como variables predictoras: pasatiempo y atención familiar.

15. **Red neuronal artificial:** Selección de variables a través del algoritmo de ponderaciones de pesos para el modelo de Red Neuronal

El siguiente reporte muestra las variables de mayor importancia que deben estar en el modelo de ajuste de red neuronal de acuerdo a los pesos mediante el algoritmo de ponderaciones.

Se obtuvo la siguiente salida:

Tabla 23
Pesos de las variables de estudio según el algoritmo de ponderaciones

	Variables	Importancia
1	TRABAJANO	-13.3913822
2	ACT_FISICANO	-6.9369079
3	EDAD	-0.4960565
4	INGRESO800 – 1200	0.3416099
5	ACT_SEXUALNO	4.1169468
6	GENEROMasculino	4.8187555
7	ESTADO_CIVSINPAREJA	5.1163534
8	INGRESO400- 800	5.8858370
9	PASATIEMPONO	7.0253400
10	INGRESO100- 400	13.3197998
11	PERDIDA_AUTSI	13.3948342
12	INGRESONO TIENE INGRESOS	17.2910643
13	ATENCION_FAMNO	19.0432241

Fuente: Elaboración propia

El algoritmo de ponderaciones calcula la importancia de las variables como el producto de los pesos de conexión de entrada oculta y salida oculta entre cada neurona de entrada y salida y suma el producto en todas las neuronas ocultas.

De acuerdo al algoritmo de ponderación de pesos para construir un modelo de red neuronal se encontraron como categorías de las variables con mayor peso negativo, es decir que protegen la ausencia de depresión moderada-severa, estas son TRABAJANO, ACT_FISICANO, EDAD y las de mayor peso positivo, es decir, favorece la depresión moderada-severa estas son ATENCION_FAMNO, INGRESONO TIENE INGRESO, INGRESO800-1200, PERDIDA_AUTSI,

INGRESO100-400, PASATIEMPONO, , INGRESO400-800, ESTADO_CIVSIN PAREJA, GENERO Masculino ,AC_SEXUALNO.

a) Ilustración gráfica de las variables más importantes a partir de la ponderación de pesos para Red Neuronal

De acuerdo al algoritmo de ponderaciones, se destaca la importancia de las variables en el modelo de ajuste hallado que se indican en la siguiente figura:

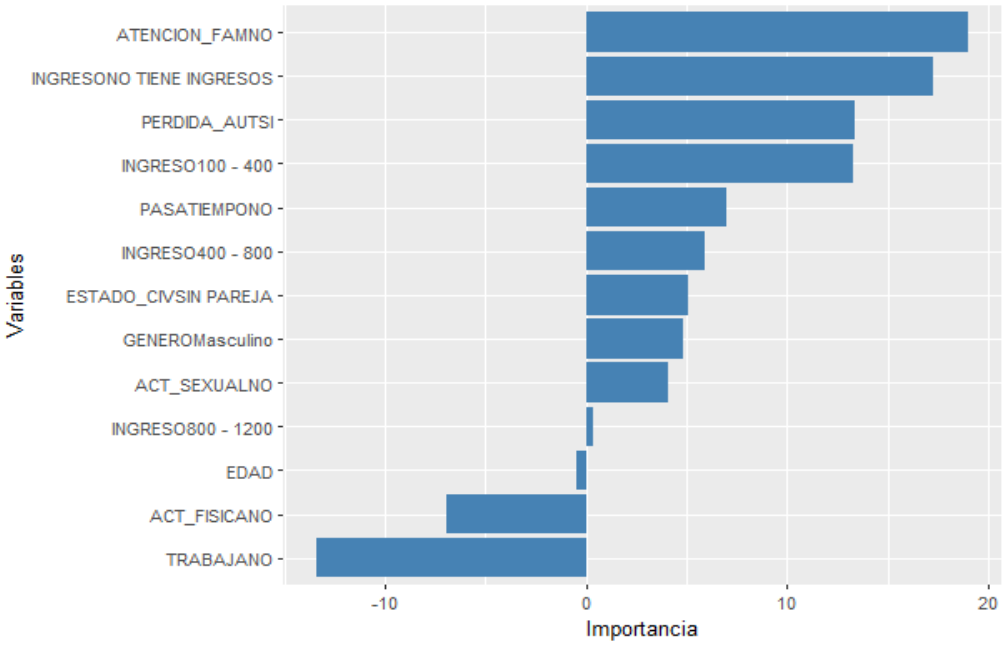

Figura 22. Importancia de variables independientes de una Red Neuronal. Elaboración propia

De la figura 22 se observa que las variables y sus categorías más importantes en la influencia de la variable Depresión son: atención familiar: No, pasatiempo: No, pérdida de autoridad: Si, actividad física: No, actividad laboral: No, ingreso económico mensual: 100-400, 400-800, 800-1200 y no tiene ingreso económico mensual.

16. Red Neuronal Artificial

El siguiente reporte permitió observar los diferentes coeficientes que pertenecen a la estructura del modelo de Red Neuronal.

Tabla 24

Valores de los coeficientes pertenecientes a la Red Neuronal

a 2-5-1 network with 21 weights		
options were - entropy fitting decay=0.05		
b->h1	i1->h1	i2->h1
-1.75	1.64	1.15
b->h2	i1->h2	i2->h2
0.89	-1.21	-0.64
b->h3	i1->h3	i2->h3
0.88	-1.20	-0.64
b->h4	i1->h4	i2->h4
-1.11	1.32	0.76
b->h5	i1->h5	i2->h5
0.89	-1.20	-0.64

b->o	h1->o	h2->o	h3->o	h4->o	h5->o
0.02	2.75	-1.76	-1.75	2.01	-1.75

Fuente: Elaboración propia

De la tabla 24, se observó los diferentes pesos que se conectan entre cada neurona, que a través de la red neuronal artificial construida con dos neuronas en la capa de entrada, 5 neuronas en la capa oculta y una neurona en la capa de salida

a) Modelo estructural de la Red Neuronal para diagnóstico de depresión

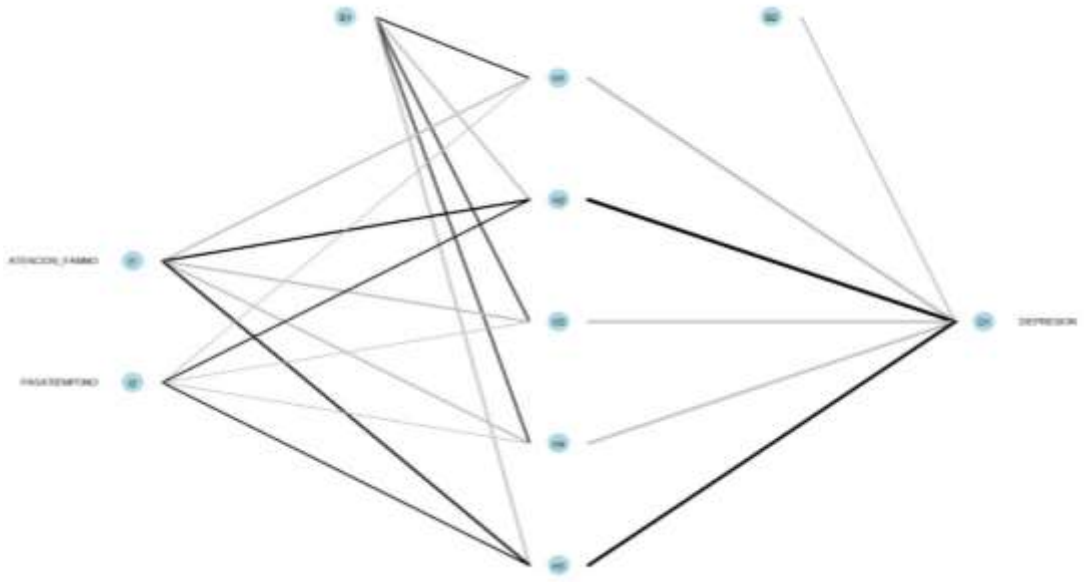

Figura 23. Modelo estructural de una Red Neuronal. Elaboración propia

De la figura 23, se puede observar la estructura de la red neuronal construida a partir de la selección de variables.

17. Modelo de Red Neuronal

A partir del resultado anterior se puede crear el modelo de Red Neuronal, obteniendo:

Valores de las neuronas de la capa oculta:

$H_1 = 1/(1 + e^{-(1{,}64*ATENCION\_FAM + 0{,}15*PASATIEMPO - 1.75)})$

$H_2 = 1/(1 + e^{-(-1{,}21*ATENCION\_FAM - 0{,}64*PASATIEMPO + 0{,}89)})$

$H_3 = 1/(1 + e^{-(-1.20*ATENCION\_FAM - 0{,}64*PASATIEMPO + 0{,}88)})$

$H_4 = 1/(1 + e^{-(1{,}32*ATENCION\_FAM + 0{,}76*PASATIEMPO - 1{,}11)})$

$H_5 = 1/(1 + e^{-(-1{,}20*ATENCION\_FAM - 0{,}64*PASATIEMPO + 0{,}89)})$

Valor de pronóstico:

$O_1 = 1/(1 + e^{-(2{,}75*H1 - 1{,}76*H2 - 1{,}75*H3 + 2{,}01*H4 - 1{,}75*H5 + 0{,}02)})$

Donde:

$y_i = \frac{1}{1 + e^{-(\sum \omega ij * xj + \theta i)}}$, con $y_i \in [0 - 1]$

18. Ajuste del modelo

Se obtuvo la siguiente salida:

Tabla 25
Resultados del ajuste del modelo según Hosmer and Lemeshow

Hosmer and Lemeshow test (binary model)		
data: DATOS_E$DEPRESION, fitted(ModRed)		
X-squared = 0.50183	df = 1	p-value = 0.4787

Se observó en los resultados que, no existe diferencia significativa entre los valores observados y esperados, en conclusión, el modelo ajusta bien a los datos.

19. Métricas de evaluación

19.1.Matriz de confusión en R usando datos de entrenamiento

Se obtuvo la siguiente salida:

Tabla 26
Matriz de confusión

	PRONOSTICADO		
OBSERVADO	Moderada - Severa	Leve	TOTAL
Moderada – Severa	29	7	36
Leve	7	61	68
TOTAL	36	68	104

Fuente: Elaboración propia

Precisión = (29+61) /104= **85.3%**

De la tabla 26, se observó que el porcentaje de clasificación correcta usando los datos de entrenamiento fue de 91.67%, es decir, los adultos mayores con depresión leve o moderada-severa serán clasificados correctamente en un 91.67% de los casos

19.2. Matriz de confusión en R usando datos de prueba

El siguiente reporte permitió calcular las diferentes métricas del modelo.

Tabla 27
Matriz de confusión

	PRONOSTICADO		
OBSERVADO	Moderada – Severa	Leve	TOTAL
Moderada - Severa	15	1	16
Leve	1	7	8
TOTAL	16	8	24

Fuente: Elaboración propia

Precisión = (15+7) /24= **91.67%**

De la tabla 27, se observó que el porcentaje de clasificación correcta usando los datos de entrenamiento fue de 91.67%, es decir, los adultos mayores con depresión leve o moderada-severa serán clasificados correctamente en un 91.67% de los casos

19.3. Métricas de evaluación en R

Se obtuvo los siguientes resultados**:**

Tabla 28
Métricas de una Red Neuronal

TASAS	VALOR
Precisión	91.67%
VPP	87.50%
VPN	93.75%
TVP (Sensibilidad)	87.50%
TVN (Especificidad)	93.75%
TFP	6.25%
TFN	12.50%
Error	8.33%
F1-Score	89.53%

Fuente: Elaboración propia

De la tabla 28, se observa que el porcentaje de clasificación correcta dada por el modelo es de 91.67%, lo que significa, que los adultos mayores pronosticados con depresión moderada o severa serán clasificados correctamente como tales en el 91,67% de los casos. Si el test tiene resultado positivo (depresión moderada o severa) la probabilidad de que realmente sea así es de 87.50%; y si el test tiene resultado negativo (depresión leve) se tendrá la certeza (93.75%) de que así será.

La capacidad que tiene el modelo para captar a los adultos mayores con depresión moderada o severa es de 87.50%, lo que quiere decir, que de 100 verdaderos depresivos moderados o severos la prueba identifica como tal a 88. La capacidad que tiene el modelo para identificar al paciente sin depresión moderada o severa es del 93.75%, lo que quiere decir, que de 100 pacientes con depresión leve la prueba detectará como tal a 94.

La capacidad que tiene el modelo de tener un resultado positivo (depresión moderada o severa) estando realmente sano (depresión leve) 6.25% y la capacidad que tiene el modelo para captar a personas sanas estando realmente enfermo es 12.5%.

El rendimiento del modelo (F1score) para clasificar correctamente la depresión es de 89.53% y el error de pronóstico fue de 8.33%.

20. Curva ROC (Red Neuronal)

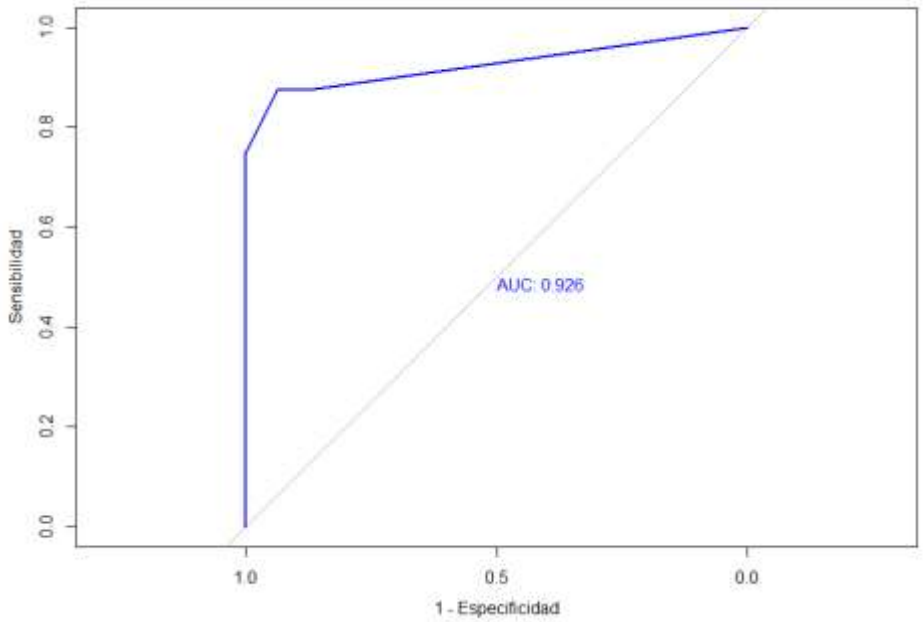

Figura 24. Área bajo la curva ROC para Depresión en adultos mayores. Elaboración propia

De la figura 24, se observó el coeficiente del área bajo la curva ROC fue de 0.926, es decir que de dos individuos uno con depresión leve y otro con depresión moderada-severa, la prueba (red neuronal) los detectará correctamente a ambos.

Tabla 29
Área bajo la curva. Evaluación de un método para determinar depresión

Área	Significación	95%IC
0.926	0.00	0.7878 – 1

Fuente: Elaboración propia

De la tabla 29 se observa que el área es de 0.926 con un intervalo de confianza que es mayor a 0.5 y significación 0.00, es decir estos aspectos indican que el modelo de red neuronal puede diagnosticar con alta precisión considerando sólo las variables pasatiempo y atención familiar.

20.1. Resumen de resultados del modelo de Red Neuronal

Tabla 30
Resumen de resultados del mejor modelo de ajuste parsimonioso en comparación con los modelos analizados

	N° neuronas en capa oculta	Decay	Precisión	Rang	ROC
Modelo 1 (2 variables)	5	5e-2	91.667	0.7	92.6
Modelo 2 (3 variables)	8	7e-2	91.667	0.7	92.2
Modelo 3 (6 Variables)	8	7e-2	91.667	0.7	93.8

Nota: El modelo 3 de RN de seis variables cuyos indicadores son: actividad física, pérdida de autoridad, Ingreso económico, actividad laboral, pasatiempo y atención familiar; el modelo 2 de tres variables cuyos indicadores son: pérdida de autoridad, pasatiempo y atención familiar; El modelo 1 de dos variables cuyos indicadores son: pasatiempo y atención familiar.

Los tres modelos resultantes obtuvieron idénticas precisiones y parecidas probabilidades bajo la curva ROC, lo que indica que cualquiera de los tres modelos

de red neuronal podría ser utilizado para pronósticos de depresión en adultos mayores.

Sin embargo, los autores por parsimonia, presentan como modelo final de pronóstico de depresión en adultos mayores el modelo 1 que incluye como variables predictoras: pasatiempo y atención familiar.

21. Comparación de modelos

21.1. Comparación de área bajo la curva ROC

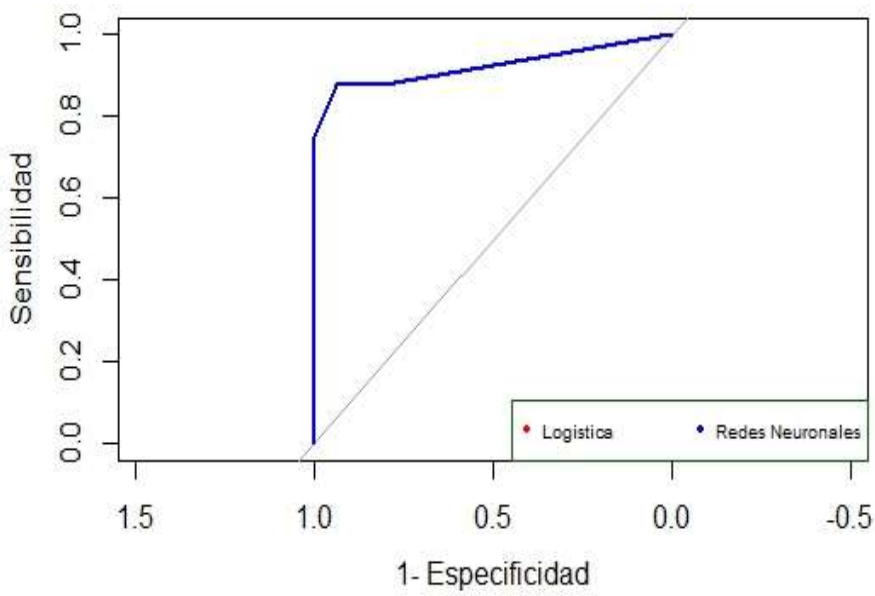

Figura 25. Comparación de área bajo la curva ROC para depresión de adultos mayores. Elaboración propia

De la figura 25, se observó que el coeficiente del área bajo la curva ROC para ambos modelos Regresión Logística y Red Neuronal se sobreponen, lo que indica que ambos modelos clasifican correctamente en 92.6% de los casos, es decir que de dos individuos uno con depresión leve y otro con depresión moderada-severa, la prueba los detectará correctamente a ambos.

21.2. Comparación de resultados

Tabla 31
Comparación de pruebas y métricas de los mejores modelos de Regresión Logística y Red Neuronal.

Métricas	Regresión Logística	Red Neuronal
Precisión (Entrenamiento)	0.8530	0.8530
Hosmer – Lemeshow	p =0.5348	p =0.4787
Precisión (Prueba)	0.9167	0.9167
Sensibilidad (Recall)	0.8750	0.8750
Especificidad	0.9375	0.9378
F-Score	0.8953	0.8953
Error	0.0833	0.0833
AUC	0.9260	0.9260
95% IC	[0.7878 - 1]	[0.7878 - 1]
R2 Nagelkerke	0.6294	

Fuente: Elaboración propia

De la tabla 31 se concluye que tanto el modelo de regresión logística como el modelo red neuronal que consideran sólo las variables predictoras Atención familiar y Pasatiempo, tienen similares valoraciones: idéntica precisión tanto con la muestra de entrenamiento y de prueba, el mismo resultado de acuerdo a Hosmer-Lemeshow, aunque mayor probabilidad de significancia la tiene el modelo de regresión logística, idénticas capacidades para detectar a los verdaderos pacientes con depresión(sensibilidad), muy parecidas capacidades de los modelos para identificar a los adultos mayores sin depresión(especificidad), iguales puntajes f1-score como capacidad de pronostico en general e iguales probabilidades bajo la curva roc (AUC) e iguales errores de pronóstico.

Por otro lado el 62.4% de las variaciones en el pronóstico de la depresión en adultos mayores son explicados por el modelo de regresión logística, sin reportar este resultado para el caso del modelo de red neuronal.

DISCUSIÓN

Este estudio realizado utilizando este tipo de modelos busca un conocimiento de las variables asociadas para poder lograr una prevención sobre la depresión es por eso que uno de los objetivos principales fue construir un modelo predictivo para cada caso utilizando las variables predictoras: edad, género, estado civil, ingreso económico, actividad física, actividad laboral, actividad sexual, atención familiar, autoridad familiar y recreación. Presentado en primer lugar los pasos básicos y fundamentales para la construcción de un modelo de regresión logística y red neuronal.

En la presente investigación, atención familiar, pérdida de autoridad y pasatiempo son los factores fuertemente asociados a la prevalencia depresión en adultos mayores con $x^2 = 66.90$, p<0.001, $x^2 = 56.69$, p<0.001 y $x^2 = 37.33$, p<0.001 respectivamente. Ávila (2015) en su estudio evidencia a la disfuncionalidad familiar como el factor fuertemente asociado a la prevalencia de depresión ($x^2 = 18.17$, p<0.001). Aldana y Pedraza (2012) en su investigación, observa una disminución en la prevalencia de depresión relacionada con el aumento del nivel educativo, mientras que Carmona y De los Santos (2012) reporta que en hombres el tener trabajo y ayuda económica reducel el riesgo de depresión (0.191 y 0.504) respectivamente.

En relación a los factores de riesgo del modelo de regresión logística, se evidencia que, el no contar con atención familiar y no tener ningún pasatiempo se relaciona con la depresión moderada-severa en adultos mayores con OR = 26.54, IC95%: 8.41-98.81 y OR = 6.59, IC95%: 2.01-24.21 respectivamente, en tanto Ävila (2015) en su estudio encuentra a la disfunción famililar como factor de riesgo con (OR = 26.54, IC95%: 8.41-98.81), mientras que en la investigación realizada por Ramírez, Bedoya, Correa, y Villada (2015) evidencian a las enfermedades neurológicas (OR=1,94, IC95% 1,14 – 3,31), a funcionalidad dependiente (OR=1,97, IC95% 1,30 - 2,98), la discapacidad auditiva (OR=2,21, IC95% 1,34 - 3,63) y el deterioro cognitivo (OR=1,54, IC95% 1,01 -2,35) como factores de riesgo del modelo de regresión logística con p<0,5. Francia (2010) en su estudio, reporta a discriminación (OR=13, IC95% 1,1 -151,9) y sentimiento de soledad (OR=112,7, IC95% 7,5 1701,8) como factores de riesgo con p<0.5.

Por otro lado, el modelo de red neuronal tiene una capacidad predictiva para clasificar correctamente de un 91.67%, considerando las variables: no tener atención famililar y no contar con ninún pasatiempo, mientras tanto Fumero y Navarrete (2014) en su investigación, evidencian que el modelo de red neuronal tiene una capacidad predictiva del 86.51% con un error de 0.51% condierando las variables: edad, neuroticismo, activación antes situaciones de estrés, trastorno esquizoide de la personalidad, trastorno límite de la personalidad, control externo y afecto negativo.

CAPITULO IV: CONCLUSIONES

1. Tanto la técnica multivariante de regresión logística dicotómica como la red neuronal con función sigmoidal proporcionan un modelo para pronóstico de depresión en adultos mayores con características muy similares.

2. Por parsimonia los autores reportan como mejor modelo de pronóstico de depresión en adultos mayores que acuden a los centros bajo estudio Hospital Provincial Docente "Belén" y el Centro Integral del Adulto Mayor (CIAM) de Lambayeque, al modelo de regresión logística que se indica a continuación:

$$\mathbf{p} = 1/\left(1 + e^{-(-3.004 + 3.279 * \boldsymbol{ATENCION\_FAM} + 1.886 * \boldsymbol{PASATIEMPO})}\right)$$

3. Las principales características de este modelo de regresión logística es que tiene una alta precisión (91.67%) y puntaje f1-score (89.53%), alta sensibilidad (87.50%), especificidad (93.75%) y bajo error (8.33%).

CAPITULO V: RECOMENDACIONES

1. Se propone a los servicios de geriatría de los diferentes servicios de salud de la región Lambayeque utilizar el modelo de regresión logística propuesto para pronóstico de depresión en adultos mayores.

2. Dado el valor de Nagelkerke (0.6293) se recomienda seguir investigando sobre otros factores que influyen en la depresión del adulto mayor dado que el 37.07% de la variabilidad del diagnóstico es explicado por otros factores.

BIBLIOGRAFIA

Academic. (2010). Obtenido de http://www.esacademic.com/dic.nsf/eswiki/508554

Aldana, R., & Pedraza, J. (2012). *Análisis de la depresión en el adulto mayor en la encuesta nacional de demografía y salud 2010.* Universidad del Rosario, Colombia.

Argibay, J. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. *Revistas UCES*, 15-33.

Ávila, S. (2015). *Determinantes sociales relacionados a la depresión del Adulto mayor en el centro de salud de la parroquia san juan Cantón Gualaceo provincia del Azuay 2015.* Universidad de Cuenca, Cuenca, Ecuador.

Bacca, A., Gonzáles, A., & Uribe, A. (2004). Validación de la Escala de Depresión de Yesavage (versión reducida) en adultos mayores colombianos. *Revista Javeriana*, 53 - 63.

Barón, J., & Téllez, F. (2004). *Apuntes de Bioestadistica: Identificación de factores de riesgo.* Málaga: Universidad de Málaga.

Bustos, E., Fernández, J., & Astudillo, C. (2017). Autopercepción de la salud, presencia de comorbilidades y depresión en adultos mayores mexicanos. *Scielo*, 11.

Carmona, S., & De los Santos, P. (2012). *Prevalencia de depresión en hombres y mujeres mayores en México y factores de riesgo.* Mèxico. Obtenido de http://www.scielo.sa.cr/pdf/psm/v15n2/1659-0201-psm-15-02-95.pdf

Dueñas, M. (2006). *Modelos de respuesta discreta en R y aplicación con datos reales.* Granada: Universidad de Granada.

Fernández, O., Marrero, M., Mesa, Y., Satiesteban, N., & Rojas, J. (2011). Depresión post–ictus: frecuencia y factores determinantes. *Dialnet*, 16.

Ferrando, J., & Anguiano, C. (2010). El análisis factorial como técnica de investigación en psicología. *Red de Revistas Científicas de América Latina, el Caribe, España y Portuga*, 18-33.

Flórez, R., & Fernández, J. (2008). *Las Redes Neuronales Artificiales Fundamentos Técnicos y Aplicaciones Prácticas.* La Coruña: Netbiblo S. L.

Francia, K. (2010). *Factores biopsicosociales que influyen en los niveles de depresión de los adultos mayores del C.S. Materno Infantil Tablada de Lurín, 2010.* Universidad Nacional Mayor de San Marcos, Lima - Perú.

Fumero, A., & Navarrete, G. (2014). Personalidad y Malestar Psicológico: Aplicación de un Modelo de Redes Neuronales. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 10.

Gómez, S., & Palacios, D. (2013). *Modelación logística multinomial para clasificar los hogares de El Salvador.* El Salvador: Universidad de El Salvador.

Grill, P. (14 de Mayo de 2014). *StackExchange*. Obtenido de https://tex.stackexchange.com/questions/176101/plotting-the-graph-of-hyperbolic-tangent

Ibrahim, O. (2013). A comparison of methods for assessing the relative importance of input variables in artificial neural networks. *Journal of Applied Sciences Research*, 5692-5700.

Lara, F. (sin fecha). *Fundamentos de redes neuronales artificiales.* México: Laboratorio de Cibernética aplicada Centro de nstrumentos.

Long, S. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Londres: SAGE Publications Inc.

López, J., & García, J. (2011). Eventos por Variable en Regresión Logística y Redes Bayesianas para Predecir Actitudes Emprendedoras. *Revista Electrónica de Metodología Aplicada*, 13-34.

Marín, J. (2012). *Introducción a las Redes Neuronales Aplicadas.* Madrid.

Mayorga, C. (2013). *Modelo de regresión logística para la identificación de factores de riesgo en pacientes con cáncer gástrico, del Hospital Belen de Trujillo.* Trujillo: Dirección de Sistemas de Informatica y Comunicación.

Morales, I. (2010). *Comparación teórico práctica entre modelos estadísticos y el perceptrón multicapa.* Valaparaíso.

Palmer, A., & Montaño, J. (2002). Redes neuronales artificiales aplicadas al análisis de supervivencia. En *un estudio comparativo con el modelo de regresión de Cox en su aspecto predictivo* (págs. 630-636). España : Psicothema.

Prieto, G., & Delgado, A. (2010). Fiabilidad y validez. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 67-74.

Ramírez, V., Bedoya, L., Correa, E., & Villada, J. (2015). *Riesgo de depresión y factores asociados en adultos mayores institucionalizados en la red de asistencia social al adulto Mayor.* Universidad CES, Medellin .

Rodas, E. (2009). *Factores asociados de riesgo para que una persona muera o sea diagnosticada con el virus A (H1N1) mediante el modelo de regresión logística, en el departamento del Cusco.* Cusco: Universidad Nacional de San Antonio Abad del Cusco.

Rojo, J. (2007). *Regresión con variable dependiente cualitativa.* Madrid .

Salcedo, C. (2002). *Estimación de la ocurrencia de incidencias en las declaraciones de pólizas de importación.* Lima: Universidad Mayor de San Marcos.

Sendra, J., Asensio, I., & Vargas, M. (2017). *Características y factores asociados a la depresión en el anciano en España desde una perspectiva de género.* España.

Serrano, A., Soria, E., & Martín, J. (2010). *Redes Neuronales Artificiales.* Valencia: Universidad Politécnica de Cartagena.

ANEXOS

Anexo 1. Test de Yesavage

ESCALA DE DEPRESIÓN DE YESAVAGE

Items	SI	NO
¿Está satisfecho con su vida?		
¿Ha abandonado muchas de sus actividades e intereses?		
¿Nota que su vida está vacía?		
¿Se encuentra a menudo aburrido?		
¿La mayor parte del tiempo está de buen humor?		
¿Tiene miedo de que le pase algo malo?		
¿Se siente feliz la mayor parte de tiempo?		
¿Se siente a menudo abandonado(a)?		
¿Prefiere quedarse en casa en vez de salir y hacer cosas?		
¿Cree que tiene más problemas de memoria que la mayoría de la gente?		
¿Cree que vivir es maravilloso?		
¿Le es difícil poner en marcha proyectos nuevos?		
¿se encuentra lleno de energía?		
¿Cree que su situación es desesterada?		
¿Cree que los otros están mejor que usted?		
Total		

Anexo 2. Encuesta

INSTRUMENTO DE EVALUACION FICHA DE IDENTIFICACION

Marcar con una tacha la respuesta que refiera el encuestado a cada pregunta, donde se encuentren respuestas de opción, y escribir el dato referido por el paciente en las preguntas abiertas.

Marque según sea su respuesta	
1. ¿Cuál es su género?	a) Masculino b) Femenino
2. ¿Cuál es su edad?	
3. Estado civil	a) Con pareja b) Sin pareja
4. ¿Cuál es su actividad laboral?	a)Si trabaja b)So trabaja
5. ¿Cuánto es el monto de sus ingresos?	a) No tiene ingresos b) de 100 a 400 soles c) de 400 a 800 soles d) de 800 a 1200 soles e) más de 1200 mensuales
6. ¿Usted realiza actividad física?	a) Si realiza b) No realiza
7. Actividad sexual: ¿En qué promedio usted realiza actividades coitales?	a) No b)Algunas veces
8. ¿Cuenta con la atención de un familiar¡	a) Si b) No
9. ¿Cree usted que ha perdido la autoridad en su familia?	a) Si b) No
10. Tiene pasatiempo favorito	a) Si b) No

Anexo 3. Procesamiento y análisis de datos en r

```
### CONFIABILIDAD Y ANÁLISIS FACTORIAL EN R ###

### Análisis de confiabilidad (kuder richardson)

> library(foreign)

> Yesavage <-read.spss("YESAVAGE.sav" , header = TRUE)

> YSV<- as.data.frame(Yesavage)

> KR<- as.matrix(YSV)

> cronbachs.alpha <-

                    function(X){

                     X <- data.matrix(X)

                     n <- ncol(X) # Number of items

                     k <- nrow(X) # Number of examinees

# Cronbachs alpha

          alpha <- (n/(n - 1))*(1 - sum(apply(X, 2,   var))/var(rowSums(X)))

          return(list("Crombach's alpha" = alpha,

          "Number of items" = n,

          "Number of examinees" = k))

  }

> dump("cronbachs.alpha", file = "cronbachs.alpha.R")

> cronbachs.alpha(KR) # compute cronbachs alpha
```

MODELO DE REGRESIÓN LOGÍSTICA Y RED NEURONAL

Lectura de datos en r

El siguiente código nos permite leer y mostrar la estructura de las variables con sus respectivas observaciones, las variables cualitativas fueron codificadas numéricamente sobre todo las variables con dos categorías, que se codificaron con 0 y 1.

```
> library(caret)

> library(foreign)

> setwd ("C:/Ubicación de la base de datos")

> DATOS=read.spss("DEPRESION.sav", to.data.frame = TRUE)

> str(DATOS)
```

Análisis de variables cualitativas en R

Para visualizar la asociación entre el género y los niveles de depresión, usamos la siguiente función **chisq.test()** aplicada a un objeto de tabla compara estos dos porcentajes a través de la prueba de independencia chi-cuadrado, donde la opción "**correct** =FALSE" en la función chisq.test desactiva la corrección de Yates para la prueba de chi-cuadrado (que se usa con tamaños de muestra pequeños), y proporciona el estadístico de prueba estándar de chi-cuadrado:

```
> Table <- xtabs(~DEPRESION+GENERO, data=DATOS)

> Test <- chisq.test(Table, correct=FALSE)

> Table;Test
```

ILUSTRACIÓN GRÁFICA DE VARIABLES

Para visualizar dicha relación de variables independientes respecto a la depresión se utilizó la función **ggplot().** Asimismo, para visualizar cada variable solo se reemplazó en la subfuncion **"x = "** y colores dentro de **"values = "** en las siguientes líneas de códigos:

```
# Gráfico circular

> pie(c(88,44),c("65.6%","34.3%"), main = "Niveles de Depresión", col = c("dodgerblue3", "gold1"))

> legend("topright", c("Leve","Moderada-severa"), cex = 0.8, fill = c("dodgerblue3", "gold1"))

# Gráfico de barras

> ggplot(DATOS, aes(x = ESTADO_CIV, y = ..count.., fill = DEPRESION)) + geom_bar()
                + scale_fill_manual(values = c("chartreuse", "darkgoldenrod1")) +
                theme_bw() + theme(legend.position = "bottom")

# Partición de datos en entrenamiento y prueba
```

Se utilizaron las siguientes líneas de códigos, donde **set.set()** ayuda a reutilizar el mismo conjunto de variables aleatorias, la función **createDataPartition()** nos permite dividir los datos en entrenamiento 80% y prueba 20% donde la opción "**list=FALSE**" se usa para evitar que el resultado sea una lista y "**times=1**" es el número de particiones que se realizarán:

```
> set.seed(123)

> train <- createDataPartition(y = DATOS$DEPRESION, p = 0.8, list = FALSE,  times= 1)

> DATOS_E <- DATOS[train, ]

> DATOS_P  <- DATOS[-train, ]
```

Verificación de la estructura de los datos particionados

Se obtuvo las frecuencias de las variables para este caso la variable Depresión, se utilizó la siguiente sentencia donde la función **prop.table()** visualiza el porcentaje de los datos que se dividieron:

```
> prop.table(table(DATOS_E$DEPRESION))

> prop.table(table(DATOS_P$DEPRESION))
```

MÉTODO WRAPPER

Eliminación recursiva de variables

```
> library(doMC)

> registerDoMC(cores = 4)

> subsets <- c(1:13) # Número de resamples para el proceso de bootstrapping

> repeticiones <- 20

> set.seed(342)

> ctrl_rfe <- rfeControl(functions = rfFuncs, method = "boot", number = repeticiones,
                         returnResamp = "all", allowParallel = TRUE, verbose = FALSE)

> rf_rfe <- rfe(DEPRESION ~ ., data = DATOS_E, sizes = subsets, metric = "Accuracy", #
                El accuracy es la proporción de clasificaciones correctas

                rfeControl = ctrl_rfe, ntree = 500)

> rf_rfe
```

REGRESIÓN LOGÍSTICA: Selección de variables en R

Selección de variables a través de AIC para el modelo de regresión logística

Las siguientes líneas de código permiten seleccionar variables a través de AIC, inicialmente se crea un modelo con todas las variables usando la función **glm()** y un modelo solo con la variable dependiente con la función **glm()** dado que la función **stepAIC()** irá eligiendo las variables a partir del modelo con todas las variables y colocando las selecciones en el modelo que contiene solo la variable dependiente hasta que el valor de AIC no disminuya.

```
> library(MASS)
```

```
> MODELOM1<-glm(DEPRESION~EDAD + GENERO + ESTADO_CIV +  TRABAJA
                + INGRESO +ACT_FISICA + ACT_SEXUAL + ATENCION_FAM +
                PERDIDA_AUT + PASATIEMPO, data = DATOS_E, family = binomial)

> modelo.inicial <-glm(DEPRESION ~ 1, data = DATOS_E, family = binomial)

> modelo.stp <- stepAIC(modelo.inicial, scope = list(upper = MODELOM1), direction =
                "both")
```

Ilustración gráfica de la importancia de variables a partir de la selección de aic para regresión logística

Para la importancia de variables del modelo de regresión logística se utilizaron las siguientes líneas, donde la función **varImp()** selecciona las variables del modelo, **rownames()** ubica los nombres y los coloca en fila, **ggplot()** permite visualizar la importancia de las variables a través de un gráfico.

```
> impor=varImp(modelo.stp)

> g=data.frame(impor)

> n=rownames(g, do.NULL = TRUE, prefix = "row")

> d1=g$Overall

> impor1<-data.frame(n,d1)

> impor1<- transform(data.frame(n,d1), n = reorder(n, d1))

> ggplot(data = impor1, aes(x = n, y = d1)) +
        geom_bar(stat = 'identity', fill = "steelblue") +
        xlab("Variables") +
        ylab("Importancia") +
        coord_flip()+
        theme_gray(base_size = 10)
```

Regresión logística y su significancia para cada modelo

Las siguientes líneas de códigos nos permiten crear el modelo de regresión logística a través de la función **glm()** y visualizar los valores de los coeficientes para cada variable del modelo:

a) Modelo Regresión Logística con 6 variables: Atención familiar, Pasatiempo, Pérdida de autoridad, Actividad física, Trabaja e Ingreso

```
> ModReg6<-glm(DEPRESION ~ ATENCION_FAM + PASATIEMPO +
                PERDIDA_AUT + ACT_FISICA + INGRESO + TRABAJA, family =
                binomial(link = logit),data=DATOS_E)

> summary(ModLog
```

b) Modelo Regresión Logística con 3 variables: Atención familiar, Pasatiempo y Pérdida de autoridad

```
> ModReg3<-glm(DEPRESION ~ ATENCION_FAM + PASATIEMPO +
                PERDIDA_AUT, family = binomial(link = logit),data=DATOS_E)

> summary(ModLog)
```

c) Modelo final de Regresión Logística con 2 variables: Atención familiar y Pasatiempo

```
> ModReg<-glm(DEPRESION ~ ATENCION_FAM + PASATIEMPO family =
                binomial(link = logit),data=DATOS_E)

> summary(ModLog)
```

Evaluación del modelo

Para la evaluación del modelo se utilizó el test Wald mediante las siguientes líneas de códigos:

```
> library(car)

> Anova(ModReg, type="II", test="Wald")
```

Coeficiente de determinación r2

Las siguientes líneas de códigos nos permiten ver el ajuste del modelo:

```
> library(rcompanion)

> nagelkerke(ModReg)
```

Odds ratios e intervalos de confianza

Para el cálculo de los odds ratio e intervalos de confianza se aplicó la siguiente línea de código, donde se calcula los exponenciales de los coeficientes con la función **exp()**, se crea un vector con la función **cbind(),** se extraen los coeficientes del modelo con la opción **coef()** y se muestra los intervalos de confianza al 95% con la opción **confint()** ya que el modelo se interpretará en base a éstas :

```
> exp(cbind(OR = coef(ModReg), confint(ModReg, level=0.95)))
```

Ajuste del modelo

Se utilizó las siguientes líneas de códigos para la prueba de Hosmer Lemeshow para estudiar la bondad de ajuste del modelo, comparando los valores esperados con los valores observados mediante la función **logitgof()** donde la opción **fitted()** devuelve los valores ajustados por el modelo:

```
> library(generalhoslem)

> logitgof(DATOS_E$DEPRESION, fitted(ModReg))
```

MATRIZ DE CONFUSIÓN

El porcentaje correcto de una clasificación se evalúa mediante el cálculo de números de ejemplos de la clase correctamente reconocidos (verdaderos positivos), el número de ejemplos correctamente reconocidos, pero no pertenecen a la clase (verdaderos negativos), los ejemplos que fueron incorrectamente asignados a la clase (falsos positivos) y los ejemplos que no fueron reconocidos a la clase (falsos negativos). Estas cuatro posibilidades constituyen una matriz de confusión.

		Predicción	
		Positivos	**Negativos**
Observación	**Positivos**	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 26. Matriz de confusión

MÉTRICAS

Para la clase individual, la evaluación está definida por VP, FN, FP, VN, Precisión, Valor de predicción positiva (VPP), Valor de predicción negativa (VPN) Sensibilidad (Recall), Especificidad, Tasa de falsos positivos (TFP), Tasa de falsos negativos (TFN), Error, F-score.

Precisión: Es la proporción entre el número de predicciones correctas (tanto positivas como negativas) y el total de predicciones.

$$\frac{VP+VN}{N} * 100$$

VPP: Es la probabilidad de que el test ha dado positivo, el individuo tiene realmente el daño.

$$\frac{VP}{VP+FP} * 100$$

VPN: Es la probabilidad de que el test ha dado negativo, el individuo no tenga la enfermedad.

$$\frac{VN}{VN+FN} * 100$$

Sensibilidad: Es la probabilidad de que un individuo realmente tenga el daño, la prueba lo detecte.

$$\frac{VP}{VP+FN} * 100$$

Especificidad: Es la probabilidad de que un individuo no tenga el daño, la prueba llegue a la misma conclusión.

$$\frac{VN}{VN+FP} * 100$$

TFP: Es la proporción de casos negativos que fueron erróneamente clasificados como positivos.

$$\frac{FP}{FP+VN} * 100$$

TFN: Es la proporción de casos positivos incorrectamente clasificados.

$$\frac{FN}{FN+VP} * 100$$

Error: Es la proporción de clasificaciones incorrectas.

$$100 - \frac{VP+VN}{N} * 100$$

F-score: Es el promedio ponderado de la Precisión y el Recall, en el cual el F1-Score alcanza su mejor valor en 1 y su peor valor en 0.

$$2 * \frac{Precisión * Recall}{Precisión + Recall}$$

Matriz de confusión en r usando datos de entrenamiento

Las siguientes líneas de códigos permitieron crear una matriz de confusión, donde la función **predict()** extrae las pronósticos del modelo con los datos de entrenamiento (DATOS_E) teniendo como respuesta probabilidades usando la opción "**type = response**":

```
> proba.pred=predict(ModReg,type="response")
```

```
> proba.pred

> clase.pred <- ifelse(proba.pred >= 0.5, "MODERADA - SEVERA" , "LEVE")

> clase.pred

> finaldata = cbind(DATOS_E, proba.pred,clase.pred)

> View(finaldata)
```

Matriz de confusión en r usando datos de prueba

Las siguientes líneas de códigos permitieron crear una matriz de confusión, donde la función **predict()** extrae las pronósticos del modelo con los datos nuevos (DATOS_P) teniendo como respuesta probabilidades usando la opción "**type=response**":

```
> Predicciones_probRL <- predict(ModReg, DATOS_P, type = "response")

> Pred=rep("MODERADA - SEVERA", length(Predicciones_probRL))

> Pred[Predicciones_probRL<=0.5]= "LEVE"

> Tabla_Conf=table(DATOS_P$DEPRESION, Pred, dnn=c("Observado","Pronosticado"))

> addmargins(Tabla_Conf)
```

Métricas de evaluación en R

Las siguientes líneas de código permitieron calcular la precisión, especificidad, sensibilidad y las diferentes métricas del modelo.

```
> VP = Tabla_Conf[2,2]; FN = Tabla_Conf[2,1];
> FP = Tabla_Conf[1,2]; VN = Tabla_Conf[1,1];
> N = VP+FN+FP+VN
> Precisión = ((VP+VN)/N)*100
> VPP = (VP/(VP+FP))*100
> VPN = (VN/(FN+VN))*100
> TVP = (VP/(VP+FN))*100
> TVN = (VN/(FP+VN))*100
> TFP = (FP/(FP+VN))*100
> TFN = (FN/(VP+FN))*100
> Error = 100 - Precisión
> F1score= 2*(Precisión*TVP)/(Precisión+TVP)
```

Curva roc (regresión logística)

Las siguientes líneas de códigos nos permiten calcular la sensibilidad, especificidad, intervalo de confianza así como el área bajo la curva.

```
> library(pROC)
> prediccionesRL <- predict(object = ModLog,  newdata = DATOS_P, type = "response")
> curva_roc <- roc(response = DATOS_P$DEPRESION,  predictor = prediccionesRL)
> auc(curva_roc)
> plot.roc(curva_roc, col="blue", print.auc=TRUE)
> ci.auc(curva_roc, conf.level = 0.95)
> ValorRL <- coords(curva_roc,"b",ret=c("specificity","sensitivity"),
                    best.method="closest.topleft")
```

```
> ValorRL
```

RED NEURONAL ARTIFICIAL: Selección de variables en R

Selección de variables a través del algoritmo de ponderaciones de pesos para el modelo de red neuronal

Para visualizar la importancia de variable de una red neuronal a través de los pesos que reporta el Rstudio, primero se crea un modelo con todas las variables usando la función **nnet()** y como siguiente paso usar la función **olden()**, creando 8 nodos en la capa oculta con la opción **"size=8"**, decaimiento de 0.07 con la opción **"decay=7e-2"** para evitar el sobre ajuste y la opción **"maxit=250"** que es el máximo de iteraciones que realizara el modelo.

```
> library(nnet)

> library(NeuralNetTools)

> set.seed(123)

> modeloRed<-nnet(DEPRESION ~ .,data = DATOS_E, size=8, decay=7e-2,rang=0.7,
                  maxit=250)

> impor<-olden(modeloRed)

> n=impor[["data"]][["x_names"]]

> d1=impor[["data"]][["importance"]]

> impor<-data.frame(n,d1)

> impor
```

Importancia de variables a partir del algoritmo de ponderaciones para red neuronal

Para visualizar de forma gráfica la importancia de variables se utilizó la función **ggplot().**

```
> impor1<- transform(data.frame(n,d1),n=reorder(n,d1))

> ggplot(data = impor1, aes(x = n, y = d1))+ geom_bar(stat = 'identity', fill = "steelblue") +
        xlab("Variables") + ylab("Importancia") +

        coord_flip()+theme_gray(base_size = 15)
```

Red neuronal artificial

Las siguientes líneas de códigos permitieron crear y visualizar los coeficientes del modelo de red neuronal, creando 8 nodos en la capa oculta con la opción "**size=8**", decaimiento de 0.07 con la opción "**decay=7e-2**" para evitar el sobre ajuste y la opción "**maxit=250**" que es el máximo de iteraciones que realizara el modelo:

a) Modelo Red Neuronal con 6 variables: Atención familiar, Pasatiempo, Pérdida de autoridad, Actividad física, Trabaja e Ingreso

```
> library(nnet)

> set.seed(1014)

> ModRed6<-nnet(DEPRESION ~ ATENCION_FAM + PASATIEMPO +
                PERDIDA_AUT + ACT_FISICA + INGRESO+ TRABAJA, data =
                DATOS_E, size=8, decay=7e-2, rang = 0.7, maxit=250)

> summary(ModRed)
```

b) Modelo Red Neuronal con 3 variables: Atención familiar, Pasatiempo y Pérdida de autoridad

```
> set.seed(1014)

> ModRed3<-nnet(DEPRESION ~ ATENCION_FAM + PASATIEMPO +
                PERDIDA_AUT, data = DATOS_E, size=8, decay=7e-2, rang = 0.7,
                maxit=250)

> summary(ModRed)
```

c) Modelo final de Red Neuronal con 2 variables: Atención familiar y Pasatiempo

```
> set.seed(1014)

> ModRed<-nnet(DEPRESION ~ ATENCION_FAM + PASATIEMPO, data =
                DATOS_E, size=5, decay=5e-2, rang = 0.7, maxit=250)

> summary(ModRed)
```

Modelo estructural de una red neuronal

Para visualizar la estructura de una red neuronal, se crearon las siguientes líneas de códigos:

```
> library(NeuralNetTools)
> plotnet(ModRed)
```

Ajuste del modelo

Para el ajuste del modelo utilizamos la prueba de Hosmer Lemeshow que compara los valores esperados por el modelo con los valores observados.

```
> library(generalhoslem)
> logitgof(DATOS_E$DEPRESION, fitted(ModRed))
```

Matriz de confusión en r usando datos de entrenamiento

```
> Predicciones_probRN <- predict(ModRed, DATOS_E, = "raw")
> Pred=rep("MODERADA - SEVERA ", length(Predicciones_probRN))
> Pred[Predicciones_probRN<=0.5]= "LEVE"
> Tabla_Conf=table(DATOS_E$DEPRESION, Pred, dnn=c("Observado","Pronosticado"))
> addmargins(Tabla_Conf)
```

Matriz de confusión en r usando datos de prueba

Las siguientes líneas de códigos permitieron clasificar los valores observados y los valores pronosticados en una matriz de confusión.

```
> Predicciones_probRN <- predict(ModRed, DATOS_P, = "raw")
> Pred=rep("MODERADA - SEVERA ", length(Predicciones_probRN))
> Pred[Predicciones_probRN<=0.5]= "LEVE"
> Tabla_Conf=table(DATOS_P$DEPRESION, Pred, dnn=c("Observado","Pronosticado"))
> addmargins(Tabla_Conf)
```

Métricas de evaluación en R

Las siguientes líneas de código permitieron calcular la precisión, especificidad, sensibilidad y las diferentes medidas de concordancia.

```
> VP = Tabla_Conf[2,2]; FN = Tabla_Conf[2,1];
> FP = Tabla_Conf[1,2]; VN = Tabla_Conf[1,1];
> N = VP+FN+FP+VN
> Precisión = ((VP+VN)/N)*100
> VPP = (VP/(VP+FP))*100
> VPN = (VN/(FN+VN))*100
> TVP = (VP/(VP+FN))*100
> TVN = (VN/(FP+VN))*100
> TFP = (FP/(FP+VN))*100
> TFN = (FN/(VP+FN))*100
> Error = 100-Precisión
> F1score= 2*(Precisión*TVP)/(Precisión+TVP)
```

Curva roc (red neuronal)

```
> library(pROC)
> prediccionesRN <- as.numeric(predict(object = ModRed,newdata = DATOS_P,
                        type = "raw"))
> curva_roc <- roc(response = DATOS_P$DEPRESION, predictor = prediccionesRN)
> auc(curva_roc)
> plot.roc(curva_roc, col="blue", print.auc=TRUE)
> ci.auc(curva_roc, conf.level = 0.95)
> ValorRN <- coords(curva_roc,"b",ret=c("specificity","sensitivity"),
              best.method="closest.topleft")
> ValorRN
```