Obesidad mórbida en adultos mayores de 60 años a partir de un modelo de lógica difusa

Maria Alejandra Cruz Suta

UNIVERSIDAD SANTO TOMAS
FACULTAD DE ESTADISTICA
PREGRADO EN ESTADISTICA
2022

Obesidad mórbida en adultos mayores de 60 años a partir de un modelo de lógica difusa

Maria Alejandra Cruz Suta

Trabajo final para optar por el titulo de:
Profesional en Estadística

Director: Profesor Gil Robert Romero

UNIVERSIDAD SANTO TOMAS
FACULTAD DE ESTADISTICA
PREGRADO EN ESTADISTICA
2022

A mis padres y a mi hermano.

Agradecimientos

A mis padres, por todo su apoyo incondicional, paciencia y constantes palabras de aliento.
A mi hermano Jeisson, por ser guía.
A la Universidad Santo Tomas por todo su apoyo.
A mi director Robert Romero por su acompañamiento, paciencia y guía constante.

Resumen

La obesidad mórbida es una enfermedad que ha venido en crecimiento a través de los años, convirtiéndose en un problema de clase mundial. En los adultos mayores de 60 años, que por su ritmo y calidad de vida sufren de enfermedades crónicas, el sobrepeso puede evitar que su tratamiento se maneje conforme a lo planeado, es por esto que, a través de un modelo de difusión lógica se desarrollan reglas que permiten evaluar factores orgánicos que aumenten la probabilidad de que una persona mayor de 60 años sufra de síndrome metabólico

Palabras claves:

Estadística, Lógica Difusa, Machine Learning, Arboles de decisión, Algoritmos

Abstract:

Morbid obesity is a disease that has been growing over the years, becoming a world-class problem. In adults over 60 years old, who due to their pace and quality of life suffer from chronic diseases, overweight can prevent their treatment to be managed as planned, that is why, through a logical diffusion model, rules are developed to evaluate organic factors that increase the probability that a person over 60 years old suffers from metabolic syndrome.

Keywords:

Statistics, Machine Learning, Fuzzy Logic

Contents

1 Introducción

A través de múltiples estudios realizados a lo largo de la historia, se han evidenciado las consecuencias en la salud que puede presentar una persona que sufre de obesidad. A pesar del amplio conocimiento que se tiene de esta enfermedad metabólica y los diferentes tratamientos que existen para la misma, es común encontrar un aumento en la prevalencia de obesidad en adultos de 60 años o más en el territorio colombiano, "por lo cual las personas con exceso de peso en el país llegan al $56,4\%$ del total de la población, lo que convierte a la obesidad en un problema de salud pública en el territorio nacional" (Cadena, E 2021)

Considerando que, debido a la obesidad mórbida se eleva significativamente el riesgo de sufrir enfermedades crónicas como hipertensión arterial, enfermedad isquemia coronaria, diabetes tipo 2, entre otras, además de provocar el incremento en la probabilidad de agudizar otras importantes enfermedades de base, vistas con más frecuencia en población de la tercera edad, conlleva a un aumento en la mortalidad prematura en el territorio nacional, como es explicado por Penny-Montenegro, 2017.

Uno de los tratamientos aplicados para la obesidad mórbida es la cirugía, con el fin de conseguir una reducción en el perímetro abdominal del paciente y con esto disminuir el riesgo de comorbilidades además de prevenir otras enfermedades de carácter psicológico, ya que "...los pacientes diagnosticados con obesidad mórbida cuentan con antecedentes psicopatológicos como lo son la ansiedad y la depresión, además de una deficiente interacción social" (Cofré Lizama, Floody Delgado, Mansilla Saldivia, & Jerez Mayorga, 2017), se realiza un procedimiento sin evaluar en muchas ocasiones otro tipo de escenarios menos invasivos, debido al alto precio que esto supone para el estado, pues cabe mencionar que el tratamiento de una persona con obesidad mórbida diferente de la cirugía requiere de profesionales especializados en diferentes áreas de la salud por un largo periodo de tiempo, .

Se ha encontrado que una de las consecuencias que tiene esta clase de cirugías en sus pacientes es la posibilidad de desarrollar síndrome metabólico y el incremento a la sensibilidad de la insulina (Shiordia Puente , Ugalde Velázquez , Cerón Rodríguez , & Vázquez García, 2012),síndrome que se caracteriza por la alteración del nivel de azúcar en sangre, tensión arterial elevada, triglicéridos plasmáticos altos, entre otros factores de riesgo, que pueden aparecen de forma secuencial o simultánea en un mismo paciente y no se trata de una sola enfermedad sino de múltiples problemas en la salud, causando una asociación importante con la disminución en la supervivencia, debido al aumento en la mortalidad cardiovascular. Lo que trae como consecuencia de dicha cirugía que una gran cantidad de individuos que son sometidos a este tratamiento para adelgazar se convierten en pacientes con enfermedades de base importantes.

Sabiendo que en Colombia el sobrepeso es un problema frecuente en los adultos mayores de 60 años y que esto afecta principalmente los riesgos de contraer enfermedades crónicas no transmisibles, lo que conlleva un aumento en las limitaciones y dependencia de los adultos, así como una disminución en su expectativa y calidad de vida, además de convertirlos en principales focos

de mortalidad de enfermedades de base, se deben evaluar y tener presentes las condiciones físicas y la forma de vida de los adultos con el fin de generar reglas que faciliten la identificación de la obesidad mórbida.

2 Problema

Las personas a medida que envejecen sufren en su cuerpo cambios con respecto a la cantidad de anticuerpos que producen o al funcionamiento de sus órganos, razón por la cual, se convierten en individuos más propensos a contraer enfermedades importantes. Con base a esto y sabiendo las complicaciones graves que trae para estas enfermedades crónicas el que un paciente sufra de obesidad, es importante desarrollar pautas que relacionen las causas orgánicas y la obesidad en los pacientes, esto con el fin de regular estos sucesos y mantenerlos dentro del rango de la normalidad y así evitar que los pacientes puedan generar nuevas complicaciones en sus afecciones.

3 Pregunta problema

¿Qué relación existe entre los factores físicos y biológicos de un adulto mayor y la probabilidad de sufrir de obesidad mórbida?

4 Objetivos

4.1 Objetivo general

- Generar reglas que ayuden a identificar la obesidad mórbida en individuos mayores de 60 años.

4.2 Objetivos especificos

- Determinar la función de pertenencia que optimice el modelo de lógica difusa.
- Hallar las variables lingüísticas que mejor representen el problema.
- Interpretar las reglas difusas para la identificación del estado de obesidad mórbida.

5 Justificación

La obesidad mórbida según la OMS consiste en la condición de existencia de un nivel mayor de tejido adiposo, lo que se considera como perjudicial para la salud. Esta obesidad se calcula con ayuda al IMC (índice de masa corporal), que es la razón entre el peso corporal de una persona en kilogramos y su altura en metros cuadrados. Una persona con obesidad mórbida tiene como IMC igual o superior a 30. “El IMC es la medida más precisa para el cálculo de la obesidad y la obesidad mórbida en adultos, ya que tiene la facilidad de encontrar los mismos valores para los dos sexos.” (Salud, 2021) “Con ayuda del IMC y teniendo

en cuenta otros factores es posible determinar los riesgos que tiene una persona de desarrollar enfermedades relacionadas con el peso" (Humanos, 2018). Cabe resaltar que el IMC no solo mide la grasa existente en el cuerpo, por lo cual no se considera un cálculo preciso, ya que al solo realizar el cálculo basándose en el peso y en la altura del individuo no es posible realizar la diferencia entre tejido adiposo y tejido muscular, por ende, puede resultar un cálculo sesgado.

La Federación Internacional de la Diabetes(IDF)[1] en octubre 2016 a falta de estudios propios del área de centro y Suramérica, sugiere que Colombia se acoja a las medidas del Sudoeste Asiático para determinar obesidad abdominal, la cual se considera una medida de fácil toma, sin costo y reproducible para evaluar de manera indirecta la grasa visceral. El objetivo de esta propuesta como medida independiente al IMC es ayudar a identificar la obesidad. Los puntos de corte para determinar la obesidad abdominal para mujeres son si el perímetro de la cintura es superior a 80 cm y para hombre si es superior o igual a 90cm. En este mismo sentido la ATP III[2] considera que el punto de corte para el perímetro abdominal para identificar obesidad es: superior a 102 cm para hombres y superior a 88cm para mujeres En cuanto al Grupo Europeo de Resistencia a la Insulina (EGIR)[3] propone que los puntos de corte para identificar la obesidad por medio del perímetro de la cintura son: para hombres superior a 94 cm y para mujeres superior a 80 cm. Sin embargo, cabe resaltar que la manera precisa es determinar el nivel de grasa visceral.(Social, 2016)

"Las personas de la tercera edad son mucho más propensas a sufrir de un aumento significativo en su peso, dado en su mayoría al sedentarismo en el que viven, acompañado de la mala alimentación y la falta de ejercicio físico, lo que tiene como consecuencia una limitación en su calidad de vida" (Penny-Montenegro, 2017). Así como una menor expectativa de vida en la población de la tercera edad con obesidad, debido a lo limitante que se convierte la realización que actividades diarias y la dependencia que genera.

6 Marco Teórico

6.1 Conceptos Fundamentales

6.2 Estadística

"Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de datos generados en una investigación sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusiones precisas o estimaciones futuras" (Salazar P. & del Castillo G., 2018)

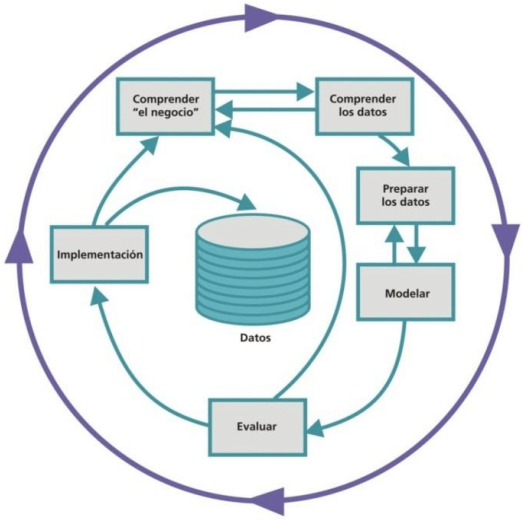
[1] Es una organización con más de 240 asociaciones nacionales de diabetes en 168 países y territorios

[2] Son las guías de tratamiento para disminuir lípidos y riesgos de enfermedad coronaria diseñadas por el National Cholesterol Education Program (NCEP)

[3] Grupo que se formó hace varios años interesado en el fenómeno de la resistencia a la insulina

6.3 Metodología CRISP-DM

Figure 1: Representación Metodología CRIPS-DM



Fuente: Bellini Saibene, Y., Volpaccio , M., Banchero, S., & Mezher Romina. (2014). Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA. Congreso Argentino.

Cross industry Standard Process for Data Mining Es una de las guías más utilizadas para el desarrollo de proyectos de minería de datos, en el cual se especifican las tareas a realizar en cada fase del proyecto, creando pequeñas tareas concretas, con un resultado esperado en cada una de ellas.
La metodología CRISP-DM se compone de las siguientes fases:

- **Fase de comprensión del negocio**: Es de las tareas más importantes, ya que contiene la comprensión y desarrollo de los objetivos del proyecto. Dentro de esta fase recae la responsabilidad de recoger la información adecuada para el alcance del proyecto y de que los mismos permitan una correcta interpretación de los resultados. (Rodríguez)

  Algunas de las tareas que se tienen en esta fase son: - Determinar el problema que se desea resolver - Evaluación de la situación - Determinación de los objetivos

- **Comprensión de los datos**:Abarca desde la recolección inicial de los datos, con el objetivo de realizar un análisis exploratorio del problema, permitiendo el desarrollo de las primeras pruebas de hipótesis.

  Las principales tareas a desarrollar en esta fase son:

  - Recolección de datos

  - Descripción de los datos

- Exploración de los datos

- Verificación de la calidad de los datos

- **Preparación de los datos**: Después de concluida la fase de recolección de datos, se procede a preparar los mismos para el método de Minería de Datos a realizar. Esta fase está muy relacionada con la fase de modelado, ya que teniendo en cuenta la técnica a utilizar los datos se procesan de diferentes formas (Rodríguez).

  Dentro de esta fase se encuentran las siguientes tareas: - Selección de datos

  - Limpieza de datos

  - Estructuración de los datos

  - Integración de los datos

  - Formateo de los datos

- **Modelado**: En esta fase se seleccionan las técnicas de modelamiento más apropiadas para el proyecto a realizar, estas técnicas a utilizar se eligen con base en 1. Su ajuste con el problema, 2. Datos adecuados para la técnica, 3. Cumplimiento de los requisitos del problema, 4. Tiempo, 5. Conocimiento del modelo.

  Cabe resaltar que antes del modelado de los datos se debe determinar un método de evaluación para los modelos con el fin de establecer un grado de bondad de ajuste. (Rodríguez)

  Las tareas a realizar en esta fase son:

  - Selección de la técnica

  - Generación del plan de prueba

  - Construcción del modelo

  - Evaluación del modelo

- **Evaluación**: En esta fase se evalúa el modelo, teniendo como criterio el cumplimiento de éxito del problema

  Las tareas involucradas en esta fase son:

  - Evaluación de los resultados

  - Proceso de revisión

  - Determinación de futuras fases

- **Implementación**: Una vez el modelo construido y evaluado, se transforman los resultados obtenidos en decisiones dentro del marco del proyecto.

  Las tareas a realizar en esta fase son:

  - Plan de implementación

  - Informe Final

6.4 Obesidad mórbida

La obesidad mórbida, según los *National Institutes of Health* (Institutos nacionales de salud) de los Estados Unidos, lo define como un sobrepeso de 50 al 100% superior al peso corporal normal o con un IMC superior a 40 y la IDF lo define tomando como referencia la circunferencia abdominal, donde es considerado obesidad abdominal para las mujeres si tienen un perímetro superior a 80 cm y para los hombres con un perímetro superior a 90cm.

El término mórbido es utilizado para enfatizar su relación con una enfermedad, generalmente síndrome metabólico. Enfermedad crónica multifactorial que se caracteriza por la acumulación excesiva de tejido adiposo, que se genera en condiciones normales cuando la ingesta de calorías es superior al gasto energético, lo que conlleva a un desequilibrio que se ve reflejado en el peso. (Rodrigo Cano, Soriano del Castillo, & Merino Torres, 2017)

Se ha reconocido esta enfermedad como un factor de riesgo para algunas enfermedades, como lo son: aterosclerosis y enfermedad cardiovascular subsecuente. Es un componente de un grupo de estados de riesgos cardiovasculares donde se encuentran incluidas la hipertensión, la resistencia a la insulina, las cuales combinadas forman lo que se denomina síndrome metabólico.

6.5 Encuesta SABE:

Es la encuesta realizada con el fin de obtener la situación real de los adultos mayores en el territorio colombiano, esto lo realizan teniendo en cuenta diferentes enfoques interdisciplinarios de la vejez y el envejecimiento. Cuenta con un corte transversal y representativo de la población de estudio.

Específicamente, desde una aproximación de trayectorias de vida y teniendo en cuenta las características sociales que conllevan a inequidades en los servicios de salud, además se evaluaron las diferentes dimensiones del marco de la política de los determinantes del envejecimiento activo, así lo menciona (Ortega Lenis & Méndez, 2019)

Es una encuesta de corte transversal con diseño metodológico cuantitativo y cualitativo, representando a los hombres y mujeres mayores de 60 años o más en el país.

Dentro del alcance de la encuesta se manifiesta que, "se espera que los resultados obtenidos en esta encuesta sobre la población de estudio sean en términos de determinantes de la vejez activa, esto quiere decir, conocer las condiciones socio demográficas en las que viven los individuos encuestados, sus hábitos, su actividad física, su estado nutricional, autocuidado y salud bucal." (Demografía, 2018).

Dentro de la encuesta se pueden encontrar las siguientes aplicaciones:

- Medias de funcionalidad, que las componen su velocidad al caminar, el equilibrio, fuerza de agarre y el tiempo que se tarda en levantarse de una

silla.

- Medidas antropométricas de los individuos encuestados, que contienen además del peso, altura, las medidas del perímetro abdominal, circunferencia de pantorrilla, brazo y rodilla.
- Medición de los niveles de azúcar en sangre, perfil lipídico y hemoglobina
- Toma de la presión arterial

  Cabe mencionar que la parte cualitativa se desarrolló con respecto a un enfoque interpretativo y de interacción simbólico, teniendo en cuenta las interacciones, dinamismo en las actividades sociales y las acciones que desempeñan los entrevistados.

La población de estudio está compuesta por los residentes de hogares urbanos y rurales de todas las regiones del país, teniendo como referencia la conformación del Ministerio de Salud y Protección Social; estos residentes debían ser mayores de 60 años, no institucionalizados, con predominio del lenguaje hispano. (Ortega Lenis & Méndez, 2019)

6.5.1 Antecedentes

Encuesta que esta precedida por dos encuestas, por una parte, la encuesta SABE realizada en 2011 por la Universidad Javeriana contando con el acompañamiento de Colciencias y los desarrollos de los protocolos de la encuesta SABE por la Universidad del Valle, con una relación contractual con el Ministerio de Salud y Protección Social.

Los resultados de esta encuesta le permiten al país recolectar información verídica sobre la salud, el bienestar, envejecimiento y la calidad de vida de los individuos mayores de 60 años, razón por la cual es insumo para determinar nuevas políticas de salud y planificar planes de acción estatal cuando se presenten eventos asociados al crecimiento demográfico del país.

6.5.2 Diseño Metodológico

Esta encuesta se construye con la necesidad de la mano con el Sistema Integral de Estudios y Encuestas Poblacionales para la Salud, regida por el ministerio de salud, con el objetivo de obtener toda la información posible acerca de la salud en Colombia.

Los indicadores diseñados para esta encuesta son en su mayoría estimaciones porcentuales y totales de individuos que cumplen con cierta característica. Es una encuesta realizada por medio de un muestreo probabilístico por conglomerados polietápicos tomando como cobertura todas las regiones del país. Se calculo el tamaño de muestra teniendo en cuenta la desagregación regional y la inclusión forzosa de grandes ciudades como Bogotá, Medellín, Cali y Barranquilla, además de parámetros como Proporción mínima establecida de 0.03, un

Deff de 1.2, error estándar de 0.05 y porcentaje de no respuesta del 20%.

Encuesta conformada por 405 preguntas divididas en 13 secciones de la siguiente manera:

- Sección 1: Identificación - Filtro - Etnia
- Sección 2: Aspectos socioeconómicos
- Sección 3: Medio ambiente físico
- Sección 4: Medio ambiente social
- Sección 5: Conducta
- Sección 6: Cognición y efecto
- Sección 7: Funcionalidad
- Sección 8: Condiciones médicas de salud
- Sección 9: Uso y acceso a servicios de salud
- Sección 10: Antropometría y valoración funcional
- Sección 11: Enlace submuestras estudio
- Sección 12: Registros submuestra estudio SABE persona mayor
- Sección 13: Datos de control

Cómo capacitación para optimizar el trabajo de campo de los encuestadores, se diseñarán manuales para las anteriores secciones que requieren de un apoyo adicional por estar consideradas de riesgo.

6.5.3 Recolección de la información

Fue llevado a cabo en el 2015 por el CNC bajo la supervisión de la UT SABE, la interventoría del ministerio de Salud y Colciencias. Este trabajo de campo consto de dos actividades principales, la segmentación y la aplicación de las encuestas y la toma de muestras. La encuesta se realizó de manera de entrevista entre el encuestador y el encuestado visitando en los hogares de cada adulto mayor. Para la transformación de datos se cuenta con una plataforma para él envió almacenamiento de la información recolectada.

6.5.4 Análisis de resultados

Para el análisis de la información se tuvieron en cuenta métodos descriptivos e inferenciales, teniendo en cuenta el tipo de variable, haciendo la estimación de promedios y proporciones. Estas estimaciones tuvieron en cuenta factores de expansión para dar cuenta de los efectos relacionados con el muestreo multietapico.

6.6 Lógica Difusa

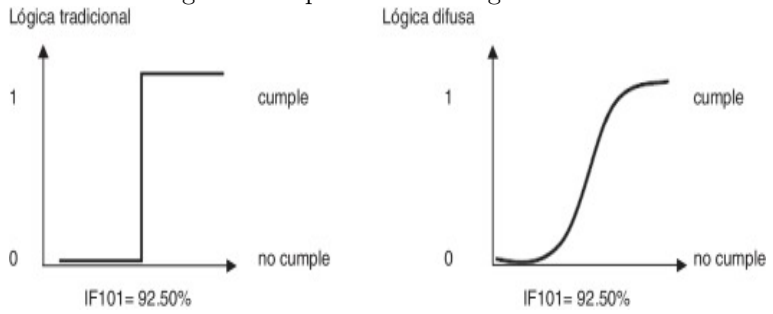
Se basa en la manera en que los seres humanos tomamos decisiones, en donde se tienen en cuenta muchos factores que junto con la experiencia se toma una decisión, que no necesariamente siempre es la misma, es decir, pueden existir variaciones (Hurtado Palacio, 2014)

Una de las ventajas que tiene el uso de la lógica difusa es que las reglas se establecen las reglas en un lenguaje natural, es decir, esta metodología permite tener un resultado a partir de información imprecisa, con ruido o incompleta.

Este concepto fue introducido por **Lofti A. Zaded**, quien, con cierto disgusto hacia los conjuntos clásicos, que evalúan dos opciones pertenencia o no pertenencia, la concibió como una forma de leer información con la posibilidad de que un individuo pertenezca parcialmente a un conjunto. (D'Negri & Eduardo, 2006).

Entonces la lógica difusa es desarrollada con la intención de mantener el concepto de vaguedad de los seres humanos que ha transcurrido a lo largo de la historia y así mismo demostrado por las afirmaciones de Zadeh, cuentan en sus artículos la necesidad de tener una exactitud en los puntos intermedios de los extremos exactos.

Figure 2: Representación Lógica Difusa



Fuente: Díaz Córdova, J. F., Coba Molina, F., & Navarrete, P. (2017). Lógica difusa y el riesgo financiero. Una propuesta de clasificación de riesgo financiero al sector cooperativo. ScienceDirect.

6.6.1 Conjuntos Difusos

"Un conjunto difuso es una pieza de información con incertidumbre que está especificada mediante una función que mapea el grado de pertenencia o certeza con la que se asocia cierto elemento a las diferentes categorías, atributos, estados de cosas o conjuntos que conforman el dominio" (Chivatá Cárdenas, 2008). Es decir, es el conjunto que puede contener elementos que no son totalmente ciertos o inciertos, denotando la diferencia entre la lógica binaria que solo permite tener elementos considerados como verdaderos o falsos. Permite la cuantificación de las expresiones que contienen ambigüedad, lo que facilita determinar analogías

entre individuos que no son estrictamente iguales. A diferencia de la teoría clásica, los conjuntos difusos se basan en el principio de que un elemento forma parte de un conjunto con un grado de pertenencia.

"Utiliza las expresiones que no se pueden afirmar como totalmente mentes correctas o totalmente incorrectas, lo que quiere decir que puede tomar cualquier valor de veracidad dentro de un conjunto de valores que oscilan entre los dos extremos. "(Pueyo). Lo que quiere decir que permite trabajar con datos imprecisos, por lo cual dentro del análisis se tiene en cuenta la presencia de imprecisiones o datos inexactos.

Permite la toma de decisiones teniendo en cuenta los grados intermedios de cumplimientos de una afirmación. Lo realiza a través de dos números aleatorios, pero con relación entre sí.

6.6.2 Operaciones de Conjuntos Difusos

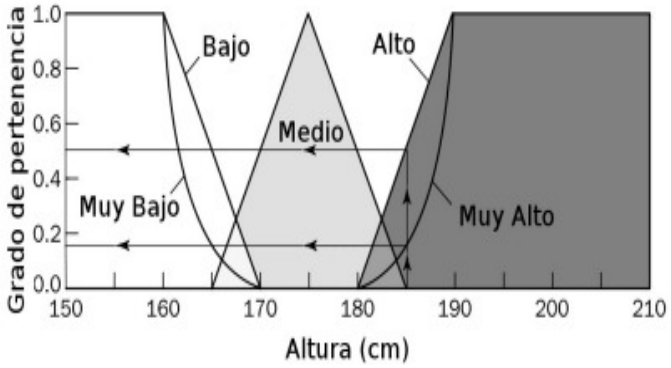
- La unión de dos conjuntos difusos es A y B es $A \cup B$
- La intersección entre dos conjuntos difusos A y B con función de pertenencia $\mu_{A \cap B}(x) = min\{\mu_A(x), \mu_B(x)\}$
- El complemento de un conjunto difuso A es el conjunto difuso $\bar{A}$ cuya función de pertenencia es $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$
- La igualdad de dos conjuntos difusos A y B se define como: $A = B \Leftrightarrow \mu_A(x) = \mu_B, \forall_x \in X$
- La inclusión de un conjunto difuso A en B se define como: $A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x), \forall \in X$
- El α - corte se define como: $A_\alpha = \{x \mid \mu_A(x) = \alpha, x \in X\}$

Estas operaciones cumplen las siguientes propiedades.

- Conmutativa: $A \cup B = B \cup A$; $A \cap B = B \cap A$
- Asociativa: $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$
- Distributiva: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Idempotencia: $A \cup A = A$; $A \cap A = A$
- Involución: $\neg(\neg A) = A$
- Transitiva: $A \subset B$ y $B \subset C$, implica $A \subset C$

6.6.3 Representación de Conjuntos Difusos

Figure 3: Representación conjuntos difusos



Fuente: González Morcillo, C. (s.f.). Lógica difusa. Una introducción practica

Para definir un conjunto difuso es importante definir su función de pertenencia. Lo más habitual es preguntarle un experto y presentarle diferentes funciones, siendo las más usuales las triangulares y trapezoidales. Para representar dicho conjunto difuso es necesario expresar esa función de pertenencia y mapear los elementos de conjunto con su grado de pertenencia. (Gonzáles Morcillo)

La siguiente figura corresponde con el conjunto difuso *Medio*, donde para la altura 165 se asocia el grado de pertenencia 0, o a la altura 175 el grado de pertenencia 1, y de nuevo a la altura el grado de pertenencia 0

6.6.4 Variables Lingüísticas

Para realizar la representación es necesario utilizar variables lingüísticas, la cual es aquella cuyos valores son sentencias en un lenguaje natural o artificial. De esta forma sirve para representar cualquier elemento que sea impreciso. (Gonzáles Morcillo)

Los símbolos terminales de las gramáticas

- Términos primarios: "bajo", "alto", ...
- Modificadores: "Muy", "más", "menos", "cerca de", ...
- Conectores lógicos: Usualmente NOT, AND y OR

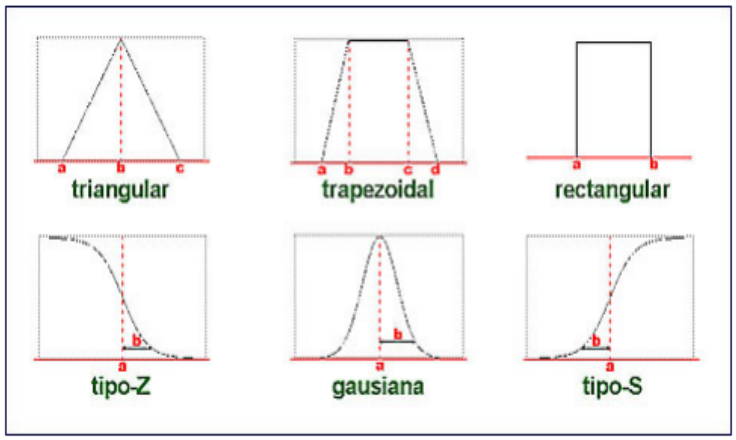
Las variables lingüísticas son usadas usualmente las reglas difusas, las cuales son aquellas reglas que indican un resultado basado de las condiciones dadas durante la fusificación y desfusicación.

6.6.5 Funciones de Pertenencia

Los conjuntos difusos pueden ser representados gráficamente cuando el conjunto universal es discreto.

Tipos de Funciones de Pertenencia

Figure 4: Funciones de Pertenencia



Fuente: Aguero Fernandez, C., Fernandez Cuesta, I., & Moya Collados , D. (2008). Valoración de inmuebles mediante técnicas de lógica difusa. Universidad Complutense de Madrid.

- Triangular: Esta definido por sus límites a y b y el valor de la moda m, tal que $a < m < b$
- Función Trapezoidal: Definida por sus límites inferior a y superior d, donde los límites de su soporte, b y c, siendo inferior y superior respectivamente.
- Rectangular: Definida por su límite superior a y el valor $k > 0$
- Función Z: Definida por su media m y el valor $k > 1$
- Función Gaussiana: Definida por su media μ y el valor de $k > 0$
- Función S: Definida por sus límites a y b, y su valor m o punto de inflexión, tal que $a < m < b$

6.6.6 Reglas de Difusión

Es un conjunto de reglas con formato SI-ENTONCES. Dónde la sentencia Si permite una acción que ha sido predeterminada con antelación, la cual se cumple si la condición tiene como resultado verdadero. (Hurtado Palacio, 2014)

Todo sistema de conjuntos difusos está conformado por una serie de reglas con bases de diferentes factores, esto quiere decir que se convierte en una regla

multi antecedente y multiconsecuente, las cuales siempre podrán ser convertidas en un sistema de reglas base con característica multi antecedente pero un solo consecuente.

Existe dos formas de obtener reglas difusas

- Permitir que sean los datos los que determinen los conjuntos difusos
- Los conjuntos difusos pueden ser predefinidos para antecedentes y consecuencias y luego si usarlo con los datos.

6.7 Árboles de Decisión

Se determinan árboles de clasificación cuando la variable dependiente es de tipo cualitativa. Básicamente se tiene como finalidad realizar un esquema de múltiples bifurcaciones, que permita que al final de cada rama se obtenga una predicción de tipo pertenencia.

Es el resultado de hacer una secuencia de preguntas ordenadas, y el tipo de pregunta que se hace en cada rama dependerá de las respuestas a las preguntas anteriores de la secuencia. El punto de partida de un árbol es denominado nodo raíz y consta de toda la información de aprendizaje en la parte superior del árbol. Un nodo es el subconjunto del conjunto de variables, y este puede ser de tipo terminal o no terminar, donde, un nodo no terminal es aquel que se divide en dos nodos, los cuales están condicionados por una variable booleana. Por el contrario, un nodo terminal, aquel que no se divide, se le asigna una etiqueta de clase. (Izenman, 2008)

Un método para escoger la mejor división es el error de clasificación, el índice de Gini o la Entropía, donde, el índice de Gini mide el grado de pureza de un nodo, mide la probabilidad de no sacar dos registros de la misma clase del nodo, a mayor índice de Gini menor pureza. La entropía es la medida que cuantifica un sistema, es decir si un nodo es puro su valor de entropía es 0 y solo tiene observaciones de una clase, pero si es igual a 1, existe la misma frecuencia para cada una de las clases.

Para evitar el sobreajuste del modelo se poda el árbol, este método consiste en eliminar divisiones de un nivel inferior que no son contribuyen significativamente a la precisión del modelo, esto también mejora la interpretación.

6.7.1 Creación del modelo

Para construir un modelo de clasificación por este método, principalmente es necesario cuestionarse qué condiciones booleanas separarán un nodo y que características debe cumplir el mismo para separar sus siguientes ramas, se debe cuestionar que debe cumplir un nodo para convertirse en un nodo terminal y que condiciones debe tener un grupo para ser asignada dentro de un nodo. (Izenman, 2008)

6.7.2 Estrategias de división

Para dividir un nodo en sus siguientes ramas, es necesario que el algoritmo de decisión decida cuál es la mejor variable para dividir, para esto es imprescindible considerar todas las posibles divisiones entre todas las variables para el nodo. Sin embargo, para una óptima división es necesario distinguir nuestras variables según su naturaleza.

- **Variables continuas y ordinarias:** Para estas variables el número posible de divisiones que se pueden hacer son menos que el número de variables observadas
- **Variables nominales y categóricas:** El número posible de divisiones para una variable con M distintas categorías son $2^{M-1} - 1$

Para determinar la mejor división entre todas las variables, primero se debe elegir la mejor división para una variable, para esto se determina para cada nodo su función de "impureza". Una de estas funciones es la *Entropía* denotada así:

$$i(\tau) = -\sum_{k=1}^{K} p(k|\tau) logp(k|\tau)$$

Donde $p(k|\tau)$ es una estimación de $P(\mathbf{X} \in \prod_k |\tau)$, la probabilidad condicional de que una observación $\mathbf{X}$ esta es $\prod_k$ dado que ca en el nodo τ.

La bondad de la división viene dada por la reducción de la impureza obtenida de dividir un nodo en sus siguientes ramas.

6.7.3 Podar el árbol

El desarrollo del algoritmo recae en hacer crecer el árbol para seguido podarlo de abajo hacia arriba hasta conseguir el tamaño optimo, es decir que un árbol podado es simplemente un subárbol del original (Izenman, 2008), para determinar la mejor manera de podar un árbol se considera el siguiente algoritmo:

- Conseguir un árbol grande T_{max}, en donde seguimos dividiendo el árbol hasta que cada uno de los nodos contenga n_{min} observaciones.
- Calcular una estimación de la tasa de clasificación errónea en cada nodo $\tau \in Tmax$
- Podar el árbol hacía su raíz, de modo que cada etapa de poda, la estimación de la tasa de clasificación errónea se minimiza.

6.8 Validación Cruzada (*Cross Validation*)

Se ha convertido en una herramienta esencial para el análisis y una característica importante al momento de ajustar modelo. Es una técnica que se utiliza para evaluar los resultados y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Esta técnica consiste en el cálculo de la media obtenida de las medidas de evaluación sobre diferentes particiones. Es usado con frecuencia en entornos donde se tiene como objetivo la predicción y se quiere estimar cuan preciso es el modelo.(Damian, 2012)

Cada uno de los modelos es entrenado y evaluado utilizando otra fuente de datos, estos datos son incluidos en el modelo y son bastante limitados a los datos que no están dentro de la base de entrenamiento.

Esto quiere decir que es un modelo que permite tener una capacidad predictiva en los modelos cuando son aplicados a observaciones que no están dentro de la base de datos, es decir, valiéndose solo de la información recogida con los datos de entrenamiento.

Consiste en separar de forma aleatoria los datos D en v subgrupos de aproximadamente el mismo tamaño, así que $D = \cup_{v=1}^{V} D_v$, donde $D_v \cap D_{v'} = \emptyset$, $v \neq v'$. Este proceso es iterativo y se realiza k veces, cambiando en cada iteración el grupo de validación. Con el fin de validar que tan bueno es el ajuste del modelo, es decir que tan cerca están los datos observados con los predichos del modelo, es importante calcular el error de entrenamiento, el cual puede ser calculado como la raíz cuadrada de una varianza.

6.9 Matriz de Confusión

Es una tabla de contingencia que tiene como finalidad el análisis de observaciones de forma emparejada. Como nos menciona *Comber*(2012) es adoptada como un estándar para informar sobre la exactitud temática de cualquier producto.

Esto quiere decir que la matriz de confusión es una herramienta que permite visualizar los resultados de los algoritmos de aprendizaje supervisado. Donde cada columna representa el número de predicciones y cada fila representa que tantos aciertos y errores está teniendo el modelo a la hora de pasar del aprendizaje con los datos.

Figure 5: Matriz de Confusión

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

Zelada, C. (10 de mayo de 2017). Rpubs. Obtenido de Evaluación de modelos de clasificación: https://rpubs.com/chzelada/275494

6.10 Curva ROC

- Sensibilidad: Probabilidad de obtener un resultado positivo cuando el individuo cuenta con la característica.

$$sensibilidad = \frac{Positivos}{Total} = \frac{VP}{VP + FN}$$

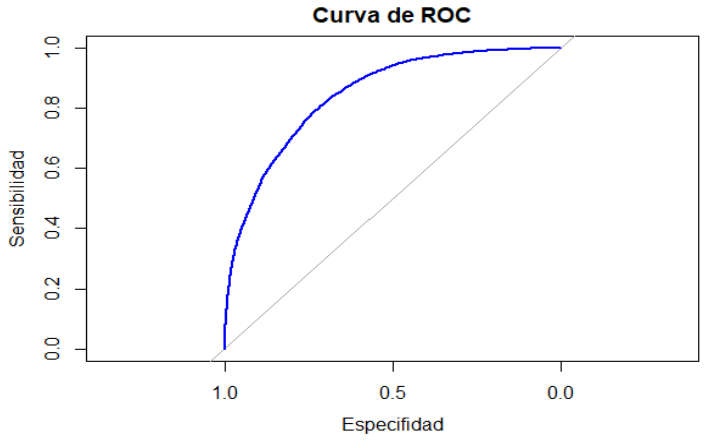
Siendo VP verdadero positivo y FN falso negativo

- Especificidad: Probabilidad de obtener un resultado negativo siendo el individuo negativo

$$Especificidad = \frac{Negativos}{Total} = \frac{VN}{VN + FP}$$

Siendo, VN verdadero negativo y FP falso positivo

Figure 6: Representación Curva ROC



Fuente: Juan Luis. (27 de abril de 2020). stackoverflow. Obtenido de ¿Cómo mejorar la gráfica de la curva ROC en R?: https://es.stackoverflow.com/questions/349954/c%C3%B3mo-mejorar-la-gr%C3%A1fica-de-la-curva-de-roc-en-r

La curva ROC es el gráfico en donde se pueden visualizar todos los pares de sensibilidad y especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados. Cada punto de la curva representa un par correspondiente a un nivel de confianza ya determinado.

En el eje y se puede encontrar la proporción de datos con sensibilidad positiva y en el eje x la proporción de datos con especificidad negativa, el área bajo esta curva clasifica estadísticamente bajo hipótesis el mejor modelo que se desarrolle en el aprendizaje y entrenamiento, para facilitar la interpretación se ha definido que:

- 0.5 a 0.6 no hay discriminación
- 0.61 a 0.7 Pobre
- 0.71 a 0.8 Aceptable
- 0.81 a 0.9 Bueno
- 0.91 - 1.00 Excelente

La interpretación de la curva ROC entrega herramientas con el fin de escoger los modelos óptimos y descartar aquellos que no lo son. La curva ROC es independiente de la distribución de las clases sobre las cuales está decidiendo.

Las tablas de contingencia pueden generar diferentes resultados, pero para el grafico del área bajo la curva ROC, solo es necesario tener los valores verdaderos positivos y los falsos positivos. El primer indicador mide hasta qué punto un clasificador es eficiente para clasificar los casos positivos de forma correcta, y los falsos positivos nos dice cuántos resultados están incorrectos.

7 Diseño presentación metodológica

Por medio de los resultados obtenidos de la aplicación de la encuesta SABE realizada a los adultos mayores de 60 años en Colombia. [4]

Se ejecutó una depuración y análisis exploratorio de la base, con el fin de determinar cálculos básicos de la población de estudio y empezar nuestro segundo paso de la metodología CRISP-DM, en la cual se necesita entender los datos.

Se hizo un modelo de difusión lógica con el fin de generar unas reglas que nos permitan reconocer los factores que aumentan la probabilidad de que una persona mayor a 60 años sufra de obesidad y esta información, compararla con la información que cuenta la bibliografía ahorita mismo, con el fin de comparar y concluir.

Se efectúa una predicción con un modelo de clasificación supervisado para comparar con el modelo anteriormente mencionado.

8 Resultado

8.1 Análisis Exploratorio

En la siguiente tabla se encuentra las principales medidas descriptivas para las variables numéricas de nuestro conjunto de datos, en donde ya podemos establecer ciertos factores que nos puedes ayudar a determinar qué pasos seguir en el modelo.

[4] Estos resultados se obtendrán de la página abierta del Ministerio de Salud.

P1005PESO_2	P1007CINTURA	P1009RODILLA	P1006TALLA	COLESTEROL_LDL	TRIGLICERIDOS	GLUCOSA	COLESTEROL_TOTAL	HEMOGLOBINA	IMC
Min. : 32.0	Min. : 66.00	Min. :30.00	Min. :133	Min. : 32.0	Min. : 42.0	Min. : 50.0	Min. : 71.0	Min. : 8.0	Min. :12.04
1st Qu.: 57.0	1st Qu.: 85.00	1st Qu.:45.00	1st Qu.:149	1st Qu.:102.0	1st Qu.: 106.0	1st Qu.: 86.0	1st Qu.:166.0	1st Qu.:123.0	1st Qu.:23.73
Median : 65.0	Median : 92.00	Median :48.00	Median :155	Median :125.0	Median : 143.0	Median : 94.0	Median :193.0	Median :136.0	Median :26.57
Mean : 65.7	Mean : 92.75	Mean :47.78	Mean :156	Mean :126.6	Mean : 164.7	Mean :101.2	Mean :194.8	Mean :126.3	Mean :27.01
3rd Qu.: 73.0	3rd Qu.:100.00	3rd Qu.:50.00	3rd Qu.:162	3rd Qu.:149.0	3rd Qu.: 195.2	3rd Qu.:103.0	3rd Qu.:221.0	3rd Qu.:148.0	3rd Qu.:29.97
Max. :146.0	Max. :125.00	Max. :62.00	Max. :185	Max. :328.0	Max. :2062.0	Max. :543.0	Max. :419.0	Max. :241.0	Max. :51.14

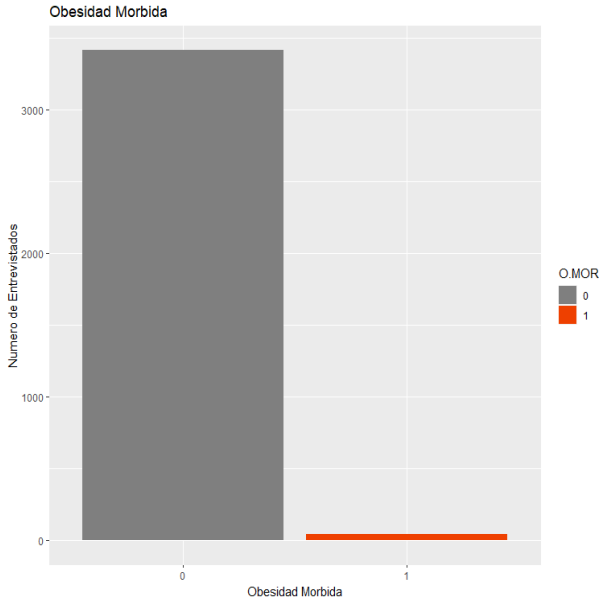
En un principio ya es posible observar comportamientos de las variables que posiblemente indiquen obesidad mórbida, como es el caso del peso máximo para los individuos que está en $146kg$, lo que corresponde a un dato bastante elevado para el peso corporal de una persona y la media del perímetro abdominal se sitúa sobre los 90 cm, lo que ya permite decir que en promedio los individuos de estudio sufren de obesidad

Dentro de los individuos de estudio en 44% tiene el colesterol HDL o también llamado colesterol malo elevado, el 45% tiene los triglicéridos altos y el 29% de los individuos cuenta con el nivel de azúcar en sangre alto, esto ya nos indica que muchos de nuestros pacientes tienen factores orgánicos de riesgo, que pueden ser consecuencia de enfermedades de base.

Se crea la variable de obesidad mórbida teniendo en cuenta el índice de Masa Corporal, que nos sugiere que para hombres un índice mayor a 40 y para mujeres un índice mayor a 35 determina obesidad mórbida.

Se puede evidenciar en el siguiente gráfico como nuestra información se encuentra desbalanceada, es decir, nos muestra muchos más individuos que no sufren obesidad mórbida en comparación con los que si la padecen, esto nos propone que la base debe ser balanceada para el clasificador, es decir en el momento de separar las bases de entrenamiento y de validación, estas deben tener un criterio que permita un mejor ajuste del modelo

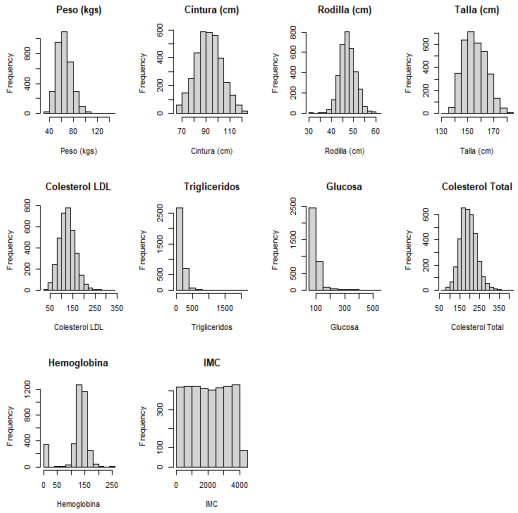
Figure 7: Histograma de la Variable Obesidad Mórbida



Fuente: Elaboración propia

En la siguiente figura podemos evidenciar el comportamiento frecuentista que tienen estas variables, aquí observamos como el peso sitúa su punto más alto por encima de los 60kgs, pero en la talla el punto más alto se encuentra sobre los 160 cm, esto para una persona con una estructura o sea normal, ya puede tener indicios de obesidad. Se evidencia que el comportamiento de los triglicéridos tiene un sesgo a la derecha y las demás variables tienen comportamiento parecido al de la campana de gauss.

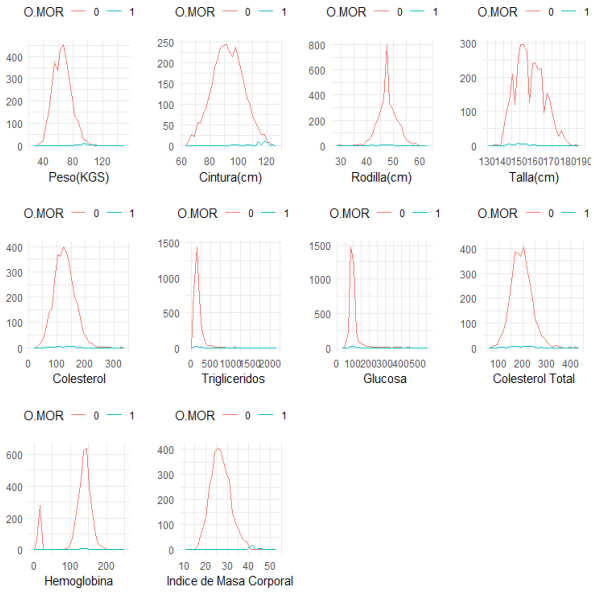
Figure 8: Histograma



Fuente: Elaboración propia

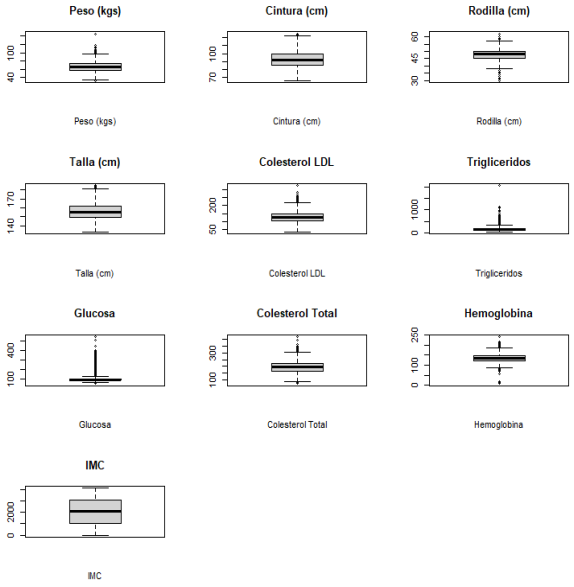
En la figura 9 se puede observar la densidad que tiene cada una de las variables, para así darnos una idea de la posible función de membresía que se utilizará para crear el modelo. Confirmamos que muchas de las variables tienen un comportamiento unimodal, con ligera similitud a la campaña de Gauss.

Figure 9: Grafico de Densidad de las Variables



En la figura 10 se observa que todas las variables tienen presentan datos atípicos, sobre todo las variables glucosa y triglicéridos, como se había mencionado antes, se puede observar como las variables parecen tener comportamientos simétricos entre sí y una gran dispersión entre ellas.

Figure 10: Boxplot



Fuente: Elaboración propia

8.2 Modelos

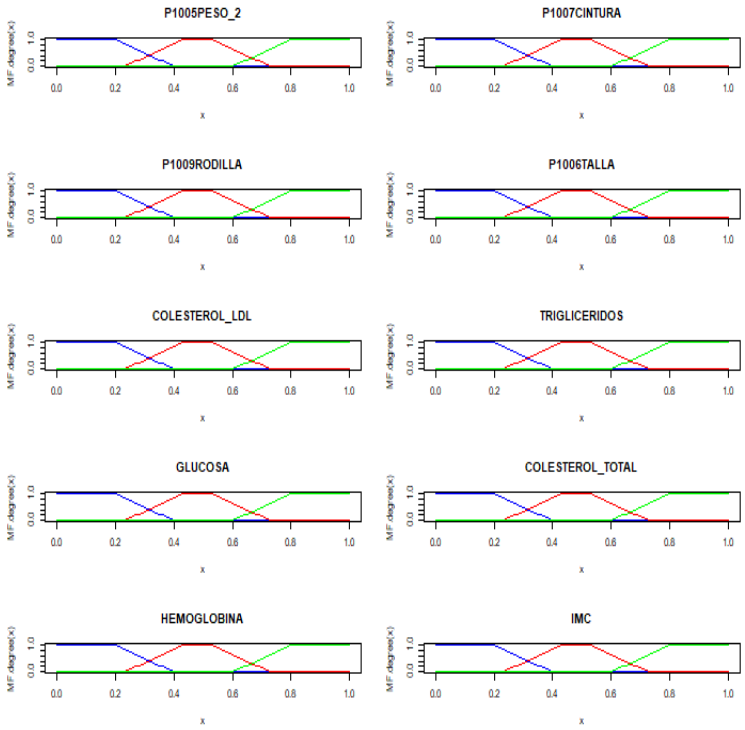
8.2.1 Lógica Difusa

Se procede a construir el modelo de lógica difusa, usando como algoritmo de aprendizaje el FRBCS.CHI la cual es la función interna que implementa un sistema de clasificación basado en reglas difusas utilizando el método CHI, es comúnmente utilizado para resolver problemas de clasificación, con una función de pertenencia de tipo GAUSSIANA.

Obteniendo las siguientes funciones de pertenencia.

Estas funciones de pertenencia nos indican el grado de membresía que tiene un valor numérico teniendo en cuenta un término lingüístico, en este caso, estas son las funciones de membresía que corresponder a las variables lingüísticas small, medium y large. Lo que quiere decir que nuestras reglas van a estar construidas de acuerdo a estas variables lingüísticas

Figure 11: Representación Funciones de Pertenencia



Fuente: Elaboración propia

Se encuentran las siguientes reglas:

1	2	3	4	5	6	7	8	9
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	large	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	large	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	medium	and
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	large	and
IF	P1005PESO_2	is	medium	and	P1007CINTURA	is	large	and

1	2	3	4	5	6	7	8
P1009RODILLA	is	medium	and	P1006TALLA	is	medium	and
P1009RODILLA	is	medium	and	P1006TALLA	is	small	and
P1009RODILLA	is	large	and	P1006TALLA	is	medium	and
P1009RODILLA	is	medium	and	P1006TALLA	is	medium	and
P1009RODILLA	is	medium	and	P1006TALLA	is	small	and
P1009RODILLA	is	medium	and	P1006TALLA	is	small	and
P1009RODILLA	is	medium	and	P1006TALLA	is	medium	and
P1009RODILLA	is	medium	and	P1006TALLA	is	medium	and

1	2	3	4	5	6	7	8
COLESTEROL_LDL	is	small	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	medium	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	medium	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	small	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	medium	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	small	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	small	and	TRIGLICERIDOS	is	small	and
COLESTEROL_LDL	is	small	and	TRIGLICERIDOS	is	small	and

1	2	3	4	5	6	7	8
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	small	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	medium	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	medium	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	small	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	medium	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	small	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	medium	and
GLUCOSA	is	small	and	COLESTEROL_TOTAL	is	medium	and

1	2	3	4	5	6	7	8	9	10	11
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	small	and	IMC	is	large	THEN	O.MOR	is	2
HEMOGLOBINA	is	medium	and	IMC	is	large	THEN	O.MOR	is	2

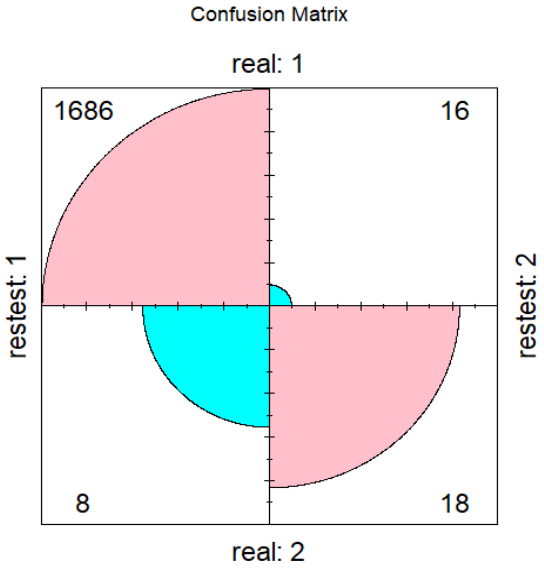
Al realizar la predicción obtenemos la siguiente matriz de confusión:

Donde se puede decir que estima correctamente que: a) hay 1686 individuos que no tienen obesidad mórbida y b) 18 individuos que, si la tienen, pero estima incorrectamente que hay: a) 16 individuos que no tienen obesidad mórbida cuando si la tienen y b)8 individuos que clasifica como si no tuviera obesidad

mórbida, pero si la tienen.

Cuenta con una exactitud del 98% que podemos clasificar como bueno, cuenta con una precisión de 30% que nos dice que tan cerca está el resultado de una predicción del valor verdadera, resultado que puede estar sesgado dado a lo raro que es que una persona tenga obesidad mórbida. Se obtiene una sensibilidad, es decir una tasa de verdaderos positivos del 69.2% y una especificidad (tasa de verdaderos negativos) del 97.5%

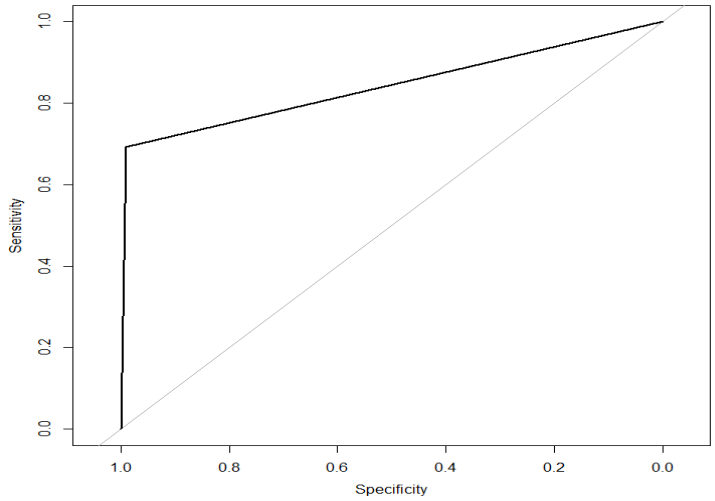
Figure 12: Matriz de Confusión para el modelo de Lógica Difusa



Fuente: Elaboración propia

Tenemos una curva valor de AUC de 0.8415, que como se había explicado anteriormente es un valor BUENO.

Figure 13: Curva Roc para Modelo de Difusión Lógica



Fuente: Elaboración propia

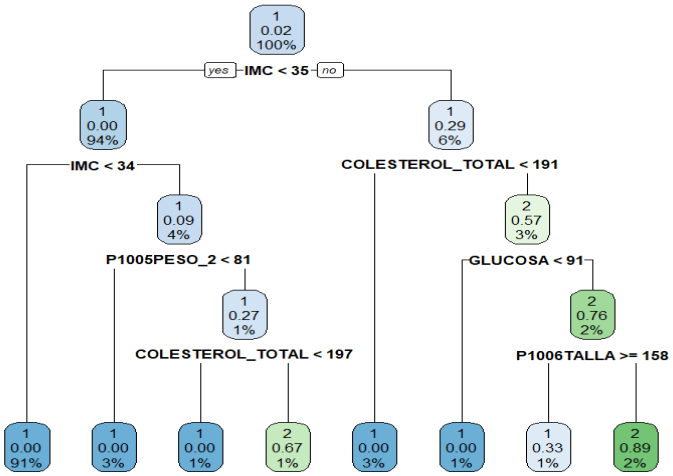
8.2.2 Arboles de Decisión

Se realiza el entrenamiento del modelo, partiendo de dividir la base 50 y 50 debido al desbalanceo de la misma.

El esquema de los resultados arrojado por el modelo de clasificación está representado en la figura 14, en donde se obtiene 15 nodos, cada uno relleno con un color que representa la categoría que ha predicho el modelo para dicho grupo, donde el verde representa a los individuos que tienen obesidad mórbida. Dentro de cada uno de los nodos nos arroja la proporción de encuestados que han sido clasificados allí.

Una de las reglas que podemos desprender de este árbol de decisión está relacionada con las variables IMC, nivel de colesterol total, glucosa, peso y talla. Se puede inferior que, si una persona tiene un índice de masa corporal superior a 35, un nivel de colesterol total menor a 191, un nivel de azúcar en sangre mayor a 91 y una talla superior o igual a 158 cm, puede ser que tenga la condición de obesidad mórbida.

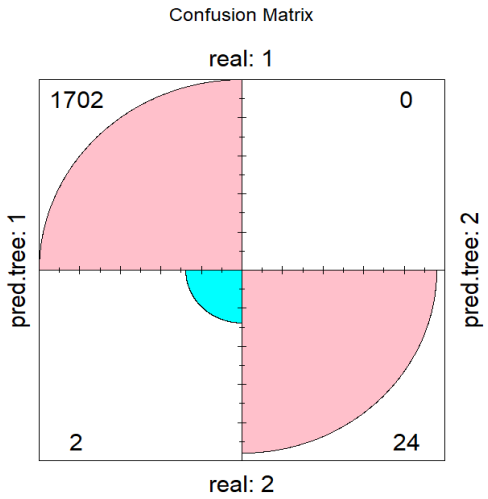
Figure 14: Modelo Árbol de Decisión



Fuente: Elaboración propia

Para evaluar la eficiencia de nuestro modelo, generamos las predicciones y obtenemos la siguiente matriz de confusión. En la cual evidencia como la predicción del modelo es mejor, ya que predice correctamente 1702 personas que no tienen obesidad mórbida y 24 que, si lo tienen, y predice incorrectamente 2 individuos que no presentan obesidad mórbida cuando si lo tienen.

Figure 15: Matriz de confusión para el modelo de árbol de decisión



Fuente: Elaboración propia

Se evidencia que la curva ROC mejora obteniendo un área bajo la curva más grande, lo que se convierte en una mejor predicción, obteniendo un AUC de 0.9615 que corresponde a una estimación EXCELENTE

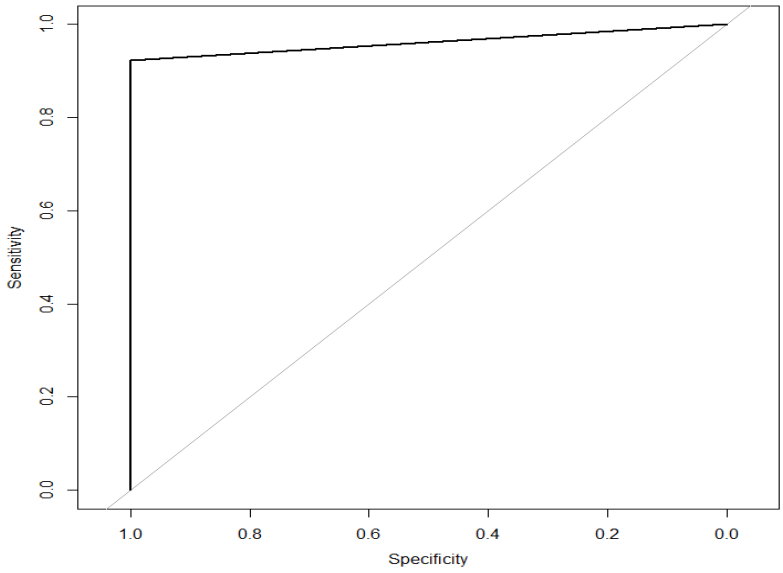


Figure 16: Curva ROC para Algoritmo de Arboles de Decisión

9 Conclusiones

- Se evidencio que la función de pertenencia que mejor ajustaba las variables, mayormente debido a su comportamiento fue la gaussiana, que asoció 3 variables lingüísticas a variables cuantitativas que permitieran una identificación de una persona que sufra de obesidad mórbida de manera más eficiente.
- Se determino que 3 variables lingüísticas eran los que ayudaban a mejorar el ajuste del modelo, debido a que cuando se realizaba el modelo teniendo en cuenta 4 variables lingüísticas, este se sobre ajustaba, provocando que clasificara más individuos verdaderos para obesidad mórbida, cuando no lo eran.
- En muchas ocasiones se estigma a las personas que sufren de síndrome metabólico a que son culpables ellos mismos de los kilos que puedan tener de más, lo que no se tiene en muchas ocasiones en cuenta es que hay causas orgánicas que pueden producir sobrepeso, como es el caso de la glucosa o los triglicéridos, que a un nivel determinado pueden ayudar al aumento del peso, como podemos evidenciar en las reglas mencionadas en este documento

List of Figures

References

[1] Cadena, E. (2021, MARZO 04). Obesidad, un factor de riesgo en el covid-19. Retrieved from https://www.minsalud.gov.co/Paginas/Obesidad-un-factor-de-riesgo-en-el-covid-19.aspx

[2] Chivatá Cárdenas, I. (2008). Estimación de la susceptibilidad ante deslizamientos: aplicación de conjuntos difusos y las teorías de la posibilidad y de la evidencia. Scielo

[3] Cofré Lizama , A., Floody Delgado, P. A., Mansilla Saldivia , C., & Jerez Mayorga, D. (2017).

[4] Diagnostico integral en pacientes obesos mórbidos candidatos a cirugía bariatrica y sugerencias para su tratamiento preoperatorio. Salud Pública , 528.

[5] Humanos, D. d. (2018, Agosto 8. Eunice Kennedy Shriver National Institute of Child Health and Human Development. Retrieved from https://espanol.nichd.nih.gov/salud/temas/obesity/informacion/diagnostican

[6] Penny-Montenegro, E. (2017, 07). Obesidad en la tercera edad. Retrieved from http://dx.doi.org/10.15381/anales.v78i2.13220

[7] Pueyo, R. P. (n.d.). Descripción general de las técncas de lógica difusa. In R. P. Pueyo.

[8] Salud, O. M. (2021, Junio 9). Obesidad y Sobrepreso. Retrieved from https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight

[9] Shiordia Puente , J., Ugalde Velázquez , F., Cerón Rodríguez , F., & Vázquez Garcia, A. (2012). Obesidad mórbida, síndrome metabólico y cirugía bariátrica: Revisión de la literatura. . Cirugía endoscopica , 85.

[10] Rodriguez, O. (s.f.). Metodología para el Desarrollo de Proyectos en Mineria de Datos CRISP-DM

[11] Salazar P. , C., & del Castillo G., S. (2018). Salazar P. , C., & del Castillo G., S. (2018). FUNDAMENTOS BASICOS DE ESTADISTICA.

[12] Shiordia PJ, Ugalde VF, Cerón RF, et al. Obesidad mórbida, síndrome metabólico y cirugía bariátrica: Revisión de la literatura. Rev Mex Cir Endoscop. 2012;13(2):85-94.

[13] Rodrigo Cano, S., Soriano del Castillo , J., & Merino Torres, J. (2017). Causas y tratamiento de la obesidad. Nutricion clinica y dietetica hospitalaria, 6.

[14] Ortega Lenis, D., & Mendez, F. (2019). Encuesta de salud, bienestar y envejecimiento sabe colombia 2015: reporte tecnico. Colombia Medica.

[15] Hurtado Palacio, J. P. (2014). Logica Difusa: Perspectiva y Aplicaciones. 39.

[16] Gonzales Morcillo, C. (s.f.). Logica Difusa. 29.

[17] Damian. (s.f.). VAlidación Cruzada. Scrib.

[18] Bouza, C., & Agustin, S. (2012). La Mineria de Datos: Arboles de decisión y su aplicación en estudios médicos. Universidad Autonoma de Guerrero , 15.

[19] Federation, I. D. (2021, 07 26). International Diabetes Federation. Retrieved from https://idf.org/who-we-are.html

[20] Social, M. d. (2016, 10). Guía Práctica Clínica. Colombia.

[21] Izenman, A. (2008). Modern Multivariate Statistical Techniques. Philadelphia: Springer.