

Universidad Católica de Santa María

Facultad de Ciencias e Ingenierías Físicas y Formales

Escuela Profesional de Ingeniería de Sistemas



SISTEMA PARA LA PREDICCIÓN DE OBESIDAD EN LA ADOLESCENCIA UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS

Tesis presentada por el Bachiller:

Ticona Sucari, Marco Antonio

Para optar el Título Profesional de

Ingeniero de Sistemas

Especialidad en Ingeniería de Software

Asesor: Dr. Sulla Torres, Jose Alfredo

Arequipa- Perú

2018



FACULTAD DE CIENCIAS E INGENIERIAS FISICAS Y FORMALES
ESCUELA PROFESIONAL DE INGENIERIA DE SISTEMAS

INFORME DICTAMEN DE BORRADOR TESIS

VISTO

El Borrador de TESIS titulado:

SISTEMA PARA LA PREDICCIÓN DE OBESIDAD EN
LA ADOLESCENCIA UTILIZANDO TÉCNICAS DE
MINERÍA DE DATOS.

Presentado por (el) (la) (los) Bachiller (es):

MARCO ANTONIO TICANA SUCARI

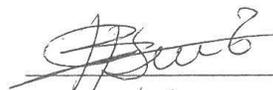
Nuestro dictamen es:

APROBADA

OBSERVACIONES:

Arequipa, _____ de _____ de _____


_____ 1748


1635
Jose A. Sullca S.

PRESENTACION

Sr. Director de la Escuela Profesional de Ingeniería de Sistemas.

Sres. Miembros del Jurado.

De conformidad con las disposiciones del Reglamento de Grados y Títulos de la Escuela Profesional de Ingeniería de Sistemas, pongo a vuestra consideración el presente trabajo de investigación titulado: **“SISTEMA PARA LA PREDICCIÓN DE OBESIDAD EN LA ADOLESCENCIA UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS”**, el mismo que de ser aprobado me permitirá optar por el Título Profesional de Ingeniero de Sistemas.

TICONA SUCARI MARCO ANTONIO.

DEDICATORIA

Este trabajo va dedicado a Dios, a mi familia, a mi enamorada y a mis amigos, que son personas muy especiales en mi vida, que en todo momento estuvieron brindándome su apoyo. Los padres son las primeras personas que conoces en el mundo, de ellos depende mucho el desarrollo de una persona, yo puedo decir que tuve los mejores padres, sin ellos no hubiese podido alcanzar esta meta, este trabajo va dedicado a ellos, que desde muy pequeño me inculcaron valores para ser una persona correcta en la vida. También dedico este trabajo a mi hermana, una persona muy alegre que siempre supo sacarme una sonrisa, y a mi enamorada, que durante todo el trayecto de este proyecto estuvo apoyándome y alentándome a seguir.

AGRADECIMIENTO

Agradezco al Ing. José Sulla Torres, al Dr. Marco Cossio y al grupo de personas que participaron en el proyecto de investigación en Perú. A ellos mi más amplio agradecimiento, por su confianza, paciencia y valioso apoyo durante todo este proceso.

Mi mayor reconocimiento y gratitud hacia ustedes.



RESUMEN

El proyecto consiste en analizar, diseñar e implementar un software que utilice un modelo de minería de datos, hecho en base a información recolectada de diversos colegios en el Perú, el software utiliza un algoritmo de árboles de decisión para predecir un posible resultado, de acuerdo a los datos o atributos que deben ser ingresados por el usuario (adolescente). Para la construcción del modelo de minería de datos se utilizó el Proceso de Descubrimiento de Conocimiento en Bases de Datos, se realizó detenidamente la fase de análisis y selección del algoritmo para deducir los patrones y tendencias que existen entre los datos.

Los algoritmos evaluados fueron J48, BayestNet, Multilayer Perceptron, ForestPA y NaiveBayes, obteniendo como mejor el algoritmo J48, con un porcentaje de precisión de 94.39%, y demostrando ser superior en otros indicadores. El algoritmo obtenido de las pruebas y comparaciones realizadas, fue implementado en una herramienta de software, con el objetivo de automatizar el proceso y evaluar a más personas para futuras investigaciones.

Palabras Clave

Aprendizaje Automático, Inteligencia Artificial, Minería de datos, Obesidad, Predicción.

ABSTRACT

The project consists of analyzing, designing and implementing a software that uses a data mining model, based on information collected from various schools in Peru, the software uses a decision tree algorithm to predict a possible outcome, in accordance to the data or attributes that must be entered by the user (adolescent). For the construction of the data mining model the Knowledge Discovery Process Databases was used, the analysis and selection phase of the algorithm was carefully carried out to deduce the patterns and trends that exist between the data.

The algorithms evaluated were J48, BayestNet, Multilayer Perceptron, ForestPA y NaiveBayes, obtaining as best the J48 algorithm, with a precision percentage of 94.39%, and proving to be superior in other indicators. The algorithm obtained from the tests and comparisons made was implemented in a software tool, with the aim of automating the process and evaluating more people for future research.

Keywords

Artificial Intelligence, Data mining, Machine Learning, Obesity, Prediction.

INTRODUCCION

Uno de los principales problemas que afronta la humanidad actualmente en el mundo y en el Perú, es la enfermedad llamada obesidad, que puede ser la base de otras muchas enfermedades peligrosas. La obesidad simboliza un problema que viene creciendo en las últimas décadas, especialmente en grupos de personas menores de edad, ocasionando cambios en el estilo de vida de la población. Las causas implicadas en la aparición del problema pueden ser varias, desde una mala alimentación o falta de ejercicio, hasta factores genéticos y orgánicos.

La forma más utilizada para diagnosticar si un paciente padece de obesidad es calculando su índice de masa corporal, luego por medio de este obtener el nivel de grasa corporal y de acuerdo a los resultados, determinar cuáles son los probables riesgos para la salud.

En ese contexto, se planteó el objetivo de saber si es viable desarrollar un sistema basado en algoritmos de minería de datos, que pueda diagnosticar o predecir si una persona de 5 a 17 años de edad, se encuentra en riesgo de padecer obesidad. Para lograr dicho objetivo se debe estudiar los pasos previos relacionados a la predicción de obesidad, realizar una comparación y evaluación de algoritmos de minería de datos, para identificar cuál es el algoritmo que mejor se adapta al conjunto de datos, y que mejora la precisión en la predicción de casos de obesidad. Posteriormente, implementar el algoritmo seleccionado en una aplicación web, que sea accesible por toda persona con acceso a internet. Y por qué no contribuir a la reducción de casos de obesidad en el Perú, previniendo y detectando a tiempo dicha enfermedad.

En el presente trabajo se presenta la documentación del proceso de análisis, desarrollo e implementación de una aplicación web para la predicción de casos de obesidad en el país, se encuentra organizado de la siguiente forma: El capítulo 1 describe el planteamiento teórico

concerniente a la investigación. En el capítulo 2 se expone el marco teórico respectivo, en el capítulo 3 se expone el análisis y definición de los requerimientos para el desarrollo del aplicativo, en el capítulo 4 se presenta el proceso de construcción y evaluación de los algoritmos de clasificación, en el capítulo 5 se realiza el diseño y desarrollo de la aplicación web, se presentan los resultados, conclusiones y las recomendaciones correspondientes.



INDICE

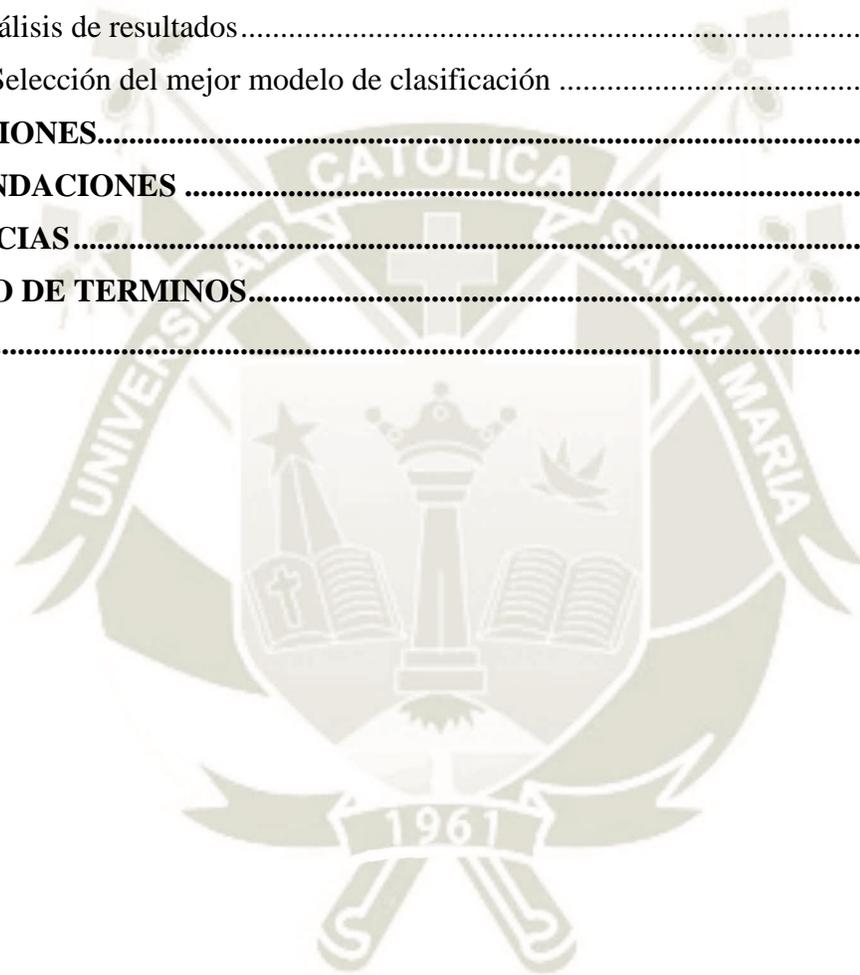
CAPITULO 1: PLANTEAMIENTO TEÓRICO	1
1.1 Título del Proyecto	1
1.2 Descripción del problema.....	1
1.3 Delimitaciones y definición del problema	2
1.3.1 Delimitaciones	2
1.3.2 Definición del Problema.....	3
1.4 Formulación del Problema	3
1.4.1 Problema Principal	3
1.5 Objetivos de la Investigación	3
1.5.1 Objetivo General.....	3
1.5.2 Objetivos Específicos	4
1.6 Viabilidad de la Investigación.....	4
1.6.1 Económica	4
1.6.2 Técnica.....	4
1.6.3 Operativa	4
1.7 Justificación e Importancia de la Investigación	5
1.7.1 Justificación	5
1.7.2 Importancia.....	6
1.8 Limitaciones de la Investigación.....	6
1.9 Área, línea, tipo y nivel de la investigación	7
1.9.1 Área de investigación	7
1.9.2 Línea de investigación	7
1.9.3 Tipo de investigación.....	7
1.9.4 Nivel de investigación	7
1.10 Método y diseño de investigación.....	7
1.10.1 Método de investigación.....	7
1.10.2 Forma de tratamiento de los datos.....	7
1.11 Cobertura del Estudio.....	9
1.11.1 Universo.....	9
1.11.2 Muestra	9
CAPITULO 2: MARCO TEÓRICO.....	10
2.1 Estado del Arte.....	10

2.2	Ingeniería de Software	17
2.3	Inteligencia Artificial	18
2.4	Aprendizaje Automático	19
2.5	Minería de datos	20
2.6	Proceso de Descubrimiento de Conocimiento en Bases de Datos	21
2.7	Técnicas de Minería de Datos	23
2.8	Asociación	23
2.8.1	Agrupamiento (Clustering).....	23
2.8.2	Clasificación	23
2.9	Modelo Cliente – Servidor	25
2.9.1	Lógica de Presentación	26
2.9.2	Lógica de Negocio	26
2.9.3	Lógica de Datos	27
2.9.4	Componentes	28
2.9.5	Ventajas	30
2.9.6	Desventajas	30
2.10	Lenguaje de programación PHP.....	30
2.10.1	Ventajas del lenguaje PHP	30
2.10.2	Desventajas del lenguaje PHP:.....	31
2.11	Metodología Ágil	31
2.11.1	Ventajas	32
2.11.2	Fases	33
2.11.3	Metodología Scrum	34
2.12	La guía SWEBOK.....	38
2.12.1	Objetivos SWEBOK.....	40
2.13	Obesidad.....	40
2.13.1	Causas.....	41
2.13.2	Consecuencias	41
2.13.3	Obesidad en Adolescentes	42
CAPITULO 3: ANÁLISIS Y DEFINICION DE REQUERIMIENTOS.....		43
3.1	Análisis de la situación actual	43
3.2	Definición de requerimientos del sistema	44

3.2.1	Adquisición de requerimientos	45
3.2.2	Identificación de actores	45
3.3	Catalogación de requerimientos	46
3.3.1	Requerimientos funcionales	46
3.3.2	Requerimientos no funcionales	50
3.4	Análisis de factibilidad para implementar la solución	52
3.5	Análisis de Casos de Usos.....	53
3.5.1	Diagrama de casos de uso.....	53
3.5.2	Elaboración del Caso de Uso Iniciar Sesión.....	54
3.5.3	Elaboración del Caso de Uso Registrar nuevo Usuario.....	55
3.5.4	Elaboración del Caso de Uso Registrar Cuestionario Obesidad.....	56
3.5.5	Elaboración del Caso de Uso Registrar Evaluación de IMC.....	57
3.5.6	Elaboración del Caso de Uso Consultar Evaluación de IMC.....	58
3.5.7	Elaboración del Caso de Uso Consultar Cuestionario Obesidad.....	59
3.5.8	Elaboración del Caso de Uso Generar Reporte individual	60
3.5.9	Elaboración del Caso de Uso Generar Reporte general Evaluaciones	61
3.5.10	Elaboración del Caso de Uso Generar Reporte general Cuestionarios	62
3.6	Análisis de diagramas de secuencia	63
3.6.1	Elaboración del Diagrama de secuencia Iniciar Sesión.....	63
3.6.2	Elaboración del Diagrama de secuencia Registrar nuevo Usuario.....	64
3.6.3	Elaboración del Diagrama de secuencia Registrar Cuestionario Obesidad.....	64
3.6.4	Elaboración del Diagrama de secuencia Registrar Evaluación de IMC	65
3.6.5	Elaboración del Diagrama de secuencia Consultar Evaluación de IMC	65
3.6.6	Elaboración del Diagrama de secuencia Consultar Cuestionario Obesidad.....	66
3.6.7	Elaboración del Diagrama de secuencia Generar reporte individual	66
3.6.8	Elaboración del Diagrama de secuencia Generar reporte general Evaluaciones.....	67
3.6.9	Elaboración del Diagrama de secuencia Generar reporte general Cuestionarios	67
3.6.10	Diagrama de clases	68
CAPITULO 4: CONSTRUCCION Y EVALUACION DE MODELOS DE CLASIFICACION		69
1.1.	Análisis previo de base de datos	69
1.2.	Construcción de los modelos de clasificación.....	71
1.2.1.	Métodos y herramientas.....	72

4.2.	Algoritmos de clasificación.....	73
4.2.1.	Clasificador J48 o C4.5	73
4.2.2.	Clasificador BayesNet	74
4.2.3.	Clasificador Multilayer Perceptron (MLP).....	74
4.2.4.	Clasificador ForestPA (Forest By Penalizing Attributes)	75
4.2.5.	Clasificador NaiveBayes	76
4.3.	Pruebas en la herramienta Weka	77
CAPITULO 5: DISEÑO, DESARROLLO E IMPLEMENTACION DEL SISTEMA DE PREDICCION.....		80
5.1	Arquitectura de Sistema	80
5.1.1	Arquitectura Lógica del Sistema	80
5.1.2	Arquitectura Física del Sistema.....	86
5.2	Diseño de la Base de Datos	87
5.2.1	Esquema conceptual de la base datos	87
5.2.2	Esquema de Implementación de la Base de Datos	88
5.3	Diseño de las Interfaces.....	88
5.3.1	Diseño de Interfaz de Inicio de Sesión para Usuario final	89
5.3.2	Diseño de Interfaz de Inicio de Sesión para Usuario administrador	90
5.3.3.	Diseño de Interfaz Menú principal de Usuario administrador	90
5.3.4.	Diseño de Interfaz Nivel Usuario final.....	98
5.4.	Desarrollo del Sistema	99
5.4.1.1.	Desarrollo de la Base de Datos	99
5.4.2.	Desarrollo de la Aplicación	100
5.5.	Implementación del Sistema – Migración a la nube	104
5.5.1.	Copia de seguridad de base de datos	104
5.5.2.	Servidor Web.....	105
5.5.3.	Configuración del Servidor Web.....	106
5.5.4.	Restauración de base datos en hosting	108
5.5.5.	Carga de aplicación a servidor web.....	109
5.6.	Evaluación del sistema.....	109
5.6.1.	Funcionalidad.	109
5.6.2.	Confiabledad.	110
5.6.3.	Facilidad de uso.....	112

5.6.4. Eficiencia.....	114
5.6.5. Facilidad de mantenimiento.....	117
5.6.6. Portabilidad.....	119
CAPITULO 6: RESULTADOS.....	122
6.1. Resultados de los algoritmos de clasificación.....	122
6.2. Evaluación y comparación de los modelos construidos.....	125
6.3. Análisis de resultados.....	127
6.3.1. Selección del mejor modelo de clasificación	127
CONCLUSIONES.....	131
RECOMENDACIONES	133
REFERENCIAS.....	134
GLOSARIO DE TERMINOS.....	141
ANEXOS.....	143



INDICE DE FIGURAS

- Figura 1 El modelo en cascada.
- Figura 2. Taxonomía de la Minería de Datos.
- Figura 3. El proceso de Descubrimiento de Conocimiento en Bases de Datos.
- Figura 4. El modelo Cliente/Servidor.
- Figura 5. El modelo Cliente/Servidor de 2 capas.
- Figura 6. Arquitectura cliente servidor para una filmoteca.
- Figura 7. Ciclo de desarrollo ágil.
- Figura 8. El proceso de Scrum.
- Figura 9. Las 5 primeras áreas de conocimiento.
- Figura 10. Esquema de los Requerimientos Funcionales.
- Figura 11. Esquema de los Requerimientos No Funcionales.
- Figura 12. Diagrama de casos de uso.
- Figura 13. Diagrama de secuencia Iniciar sesión
- Figura 14. Diagrama de secuencia Registrar nuevo usuario.
- Figura 15. Diagrama de secuencia Registrar cuestionario obesidad.
- Figura 16. Diagrama de secuencia Registrar Evaluación IMC.
- Figura 17. Diagrama de secuencia Consultar evaluación de IMC.
- Figura 18. Diagrama de secuencia Consultar cuestionario obesidad.
- Figura 19. Diagrama de secuencia Generar reporte individual.
- Figura 20. Diagrama de secuencia Generar reporte general Evaluaciones.
- Figura 21. Diagrama de secuencia Generar reporte general Cuestionarios.
- Figura 22. Diagrama de clases - SisPrediccion.
- Figura 23. Grafo del clasificador Naive Bayes obtenido en Weka
- Figura 24. Grafo de Perceptrón Multicapa
- Figura 25. Entorno inicial – Preprocesamiento en Weka.
- Figura 26. Lista de clasificadores de Weka.
- Figura 27. Resultados de análisis de clasificación en Weka.

- Figura 28. Resultados de visualización entre los atributos IMC y Edad.
- Figura 29. Patrón de diseño en capas.
- Figura 30. Arquitectura del sistema.
- Figura 31. Arquitectura del sistema de predicción.
- Figura 32. Módulos capa de presentación.
- Figura 33. Diagrama de servicios básicos.
- Figura 34. Diagrama de módulos servicios de negocio.
- Figura 35. Diagrama de servicios de Infraestructura.
- Figura 36. Diagrama de Arquitectura Física del sistema.
- Figura 37. Diagrama de Despliegue de la Arquitectura Física del sistema.
- Figura 38. Diagrama de Entidad – Relación del Sistema Predicción de Obesidad.
- Figura 39. Modelo Físico del Sistema Predicción de Obesidad.
- Figura 40. Interfaz de Inicio de Sesión de usuario final.
- Figura 41. Interfaz de Inicio de Sesión de usuario administrador.
- Figura 42. Interfaz de Menú Principal Usuario Administrador.
- Figura 43. Lista de opciones del menú Aplicaciones.
- Figura 44. Lista de opciones del menú Consultas.
- Figura 45. Cuestionario de Actitud hacia la obesidad.
- Figura 46. Resultado Cuestionario de Actitud hacia la obesidad
- Figura 47. Evaluación IMC.
- Figura 48. Resultado Evaluación IMC.
- Figura 49. Consulta Evaluación IMC.
- Figura 50. Consulta Cuestionario actitud hacia la obesidad.
- Figura 51. Reporte individual de usuario.
- Figura 52. Lista de opciones Reportes usuario administrador.
- Figura 53. Reporte general evaluaciones IMC.
- Figura 54. Formulario ingreso sugerencias.
- Figura 55. Menú de opciones reportes usuario final.

- Figura 56. Panel de control XAMPP.
- Figura 57. Esquema de base de datos implementado en Phpmyadmin.
- Figura 58. Estructura del proyecto sisPrediccion en Sublime.
- Figura 59. Ejemplo de uso AJAX. Función obtener preguntas.
- Figura 60. Clase conexión Database.
- Figura 61. Formulario de exportación de base de datos Phpmyadmin.
- Figura 62. Plan de servicios Hostinger Premium.
- Figura 63. Acceso al panel de control del servidor web.
- Figura 64. Propiedades de acceso FTP al servidor web.
- Figura 65. Restauración de base de datos db_sis_prediccion.
- Figura 66. Interfaz SmartFTP.
- Figura 67. Panel de control Thread Group.
- Figura 68. Panel de control de HTTP Request Defaults.
- Figura 69. Panel de control de HTTP Request.
- Figura 70. Gráfico del plan de pruebas.
- Figura 71. Navegador Mozilla - Sisprediccion.
- Figura 72. Navegador Internet Explorer - Sisprediccion.
- Figura 73. Navegador Opera - Sisprediccion.
- Figura 74. Navegador Safari - Sisprediccion.
- Figura 75. Resultados de Ratios de Verdaderos Positivos.
- Figura 76. Resultados de Ratios de Falsos Positivos.
- Figura 77. Resultados de Ratios de Valores de Precisión.
- Figura 78. Resultados de Ratios de Área bajo la curva ROC.
- Figura 79. Árbol J48 generado en Weka.

INDICE DE TABLAS

Tabla 1 Áreas de conocimiento del SWEBOK.

Tabla 2 Privilegios de autores.

Tabla 3 Total de atributos del conjunto de datos.

Tabla 4 Características antropométricas del conjunto de datos.

Tabla 5 Parámetros de predicción.

Tabla 6 Validación con usuarios.

Tabla 7 Indicadores para evaluar la complejidad del módulo.

Tabla 8 Resultados del clasificador Árbol de decisión J48.

Tabla 9 Resultados del clasificador BayesNet.

Tabla 10 Resultados del clasificador MultilayerPerceptron.

Tabla 11 Resultados del clasificador ForestPA.

Tabla 12 Resultados del clasificador NaiveBayes.

Tabla 13 Resultados de clasificación y comparación.

CAPITULO 1: PLANTEAMIENTO TEÓRICO

1.1 Título del Proyecto

SISTEMA PARA LA PREDICCIÓN DE OBESIDAD EN LA ADOLESCENCIA
UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS.

1.2 Descripción del problema

Según la Organización Mundial de la Salud (OMS), la obesidad y el sobrepeso han alcanzado caracteres de epidemia a nivel mundial. En 2016, más de 1900 millones de adultos de 18 o más años tenían sobrepeso, de los cuales, más de 650 millones eran obesos, 41 millones de niños menores de cinco años tenían sobrepeso o eran obesos, había más de 340 millones de niños y adolescentes (de 5 a 19 años) con sobrepeso u obesidad. El problema es mundial y está afectando progresivamente a muchos países de bajos y medianos recursos, en especial en el medio urbano. (Organización Mundial de la Salud [OMS], 2017:2)

La obesidad es un problema muy conocido en la actualidad. El número de personas obesas ha incrementado en los últimos años y se ha convertido en una epidemia mundial, existen pruebas de que la obesidad en los infantes o en los adolescentes, persiste en la adultez. En la actualidad este fenómeno es un problema cada vez más común en adolescentes modernos, es así que se debe tomar atención en los correlatos psicosociales relacionados con la estigmatización a la obesidad entre los adolescentes. (Gómez et al., 2015)

En ese sentido, los prejuicios y la discriminación contra las personas con sobrepeso y obesidad actualmente están muy extendidos dentro de nuestra sociedad, de hecho, varios estudios demostraron que el exceso de peso genera rechazo y problemas de discriminación en diferentes ámbitos, por ejemplo, en la asistencia sanitaria, en el lugar de trabajo, en la escuela, en las relaciones interpersonales, entre otros ámbitos. (Gómez et al., 2015).

Por lo tanto, predecir la obesidad a temprana edad resultaría beneficioso para la sociedad, y es que no se tiene un sistema que automatice dicha predicción, que muestre de manera inmediata las probabilidades que existen, y así poder evitar que deriven en posibles enfermedades cardiovasculares o en el peor de los casos la pérdida de vida. Si se logra detectar a tiempo, se podría actuar debidamente y brindar un tratamiento adecuado desde temprana edad.

Existen muchas formas de evitar y prevenir que se presente obesidad en la adultez, como adoptar buenos hábitos alimenticios o realizar actividad física. Lo importante es detectar la enfermedad con la debida anticipación.

1.3 Delimitaciones y definición del problema

1.3.1 Delimitaciones

a. Delimitación espacial

El actual trabajo de desarrollo del sistema de predicción, es llevado a cabo en la ciudad de Arequipa como proyecto de la Universidad Católica de Santa María.

b. Delimitación Temporal

El trabajo se inicia en noviembre de 2017 y culminará en setiembre de 2018.

c. Delimitación Social

Se encuentra enfocado a automatizar la predicción de obesidad en personas de 5 a 17 años, capturar datos mediante una herramienta de medición, con la cual se puede registrar: el índice de masa corporal, masa grasa, masa magra, la actitud hacia la obesidad en adolescentes y obtener una posible predicción sobre el estado de peso del usuario, por lo que su orientación social involucra capturar datos reales de adolescentes

en escuelas del Perú, con el uso de encuestas y evaluaciones antropométricas, de tal forma que se evalúen características relevantes para medir la obesidad.

d. Delimitación conceptual

Diseño y desarrollo de un sistema de predicción utilizando técnicas de minería de datos.

1.3.2 Definición del Problema

Actualmente en Arequipa no se cuenta con un sistema que automatice la predicción el proceso de minería de datos, para detectar patrones y detectar si el adolescente tiene una alta probabilidad de padecer de obesidad.

1.4 Formulación del Problema

1.4.1 Problema Principal

No se tiene un sistema que, de acuerdo a los síntomas y datos de un paciente, nos informe con cierta probabilidad en qué casos prevenir y tomar las acciones necesarias para evitar futuras enfermedades relacionadas a la obesidad y/o el sobrepeso.

1.5 Objetivos de la Investigación

1.5.1 Objetivo General

El objetivo del proyecto es analizar, diseñar e implementar un software, con el uso de técnicas de minería de datos, que pueda predecir si un adolescente a determinada edad, va a padecer de obesidad en el futuro, es decir en la adultez y que diagnostique cual es el estado actual del paciente. Para ello intervienen diversas variables socio-culturales, familiares, genéticas y antropométricas.

1.5.2 Objetivos Específicos

1. Realizar una revisión detallada de los antecedentes del proyecto.
2. Consolidar la información recolectada para realizar el análisis.
3. Seleccionar las variables de entrada que participarán en el modelo.
4. Comparar técnicas de minería de datos con el fin de determinar cuál es la mejor para el caso.
5. Interpretar el análisis respectivo de los resultados de la comparación.
6. Desarrollar una herramienta de software que utilice el modelo seleccionado con mejores resultados.

1.6 Viabilidad de la Investigación

1.6.1 Económica

Los recursos económicos necesarios son asumidos por la Universidad Católica de Santa María, como parte de los fondos que brinda para el apoyo de proyectos de investigación. Sin embargo, no se puede definir un valor monetario hasta un análisis de mercado bien realizado.

1.6.2 Técnica

Para el aspecto técnico se tiene los conocimientos teóricos y capacidad necesarios para resolver el problema planteado.

1.6.3 Operativa

Se cuenta con las herramientas bibliográficas, internet, servicio de bibliotecas de base de datos online, laboratorios y centro de estudios necesarios para la realización del proyecto.

Asimismo, la universidad Católica de Santa María, brinda acceso a la plataforma online Web of Science, que contiene bases de datos de información bibliográfica y recursos de análisis de la información que permiten evaluar y analizar el rendimiento de la investigación.

1.7 Justificación e Importancia de la Investigación

1.7.1 Justificación

Actualmente en nuestro país existe un alto índice de obesidad y sobrepeso en los jóvenes, causando a temprana edad enfermedades crónicas en los diferentes sistemas del cuerpo humano, por la falta de recursos económicos o accesibilidad a la atención médica; Por otro lado se sabe que la tecnología forma parte del desarrollo de la población y es de rápido, económico y fácil acceso; es por ello que en este proyecto de investigación, se desarrollará una aplicación web para la predicción de la obesidad, que diagnostique de forma inmediata, si un paciente tiene grandes riesgos de padecerla, con el fin de tomar las medidas respectivas a tiempo.

En el presente trabajo de investigación se busca desarrollar un sistema para la predicción de obesidad en la adolescencia utilizando técnicas de minería de datos, de tal manera facilitar el diagnóstico de obesidad a los usuarios.

Es de pertinencia a la escuela de ingeniería de sistemas, porque se tienen conocimientos necesarios para comprender técnicas y algoritmos de minería de datos, y utilizarlos para el desarrollo de una aplicación que realice la predicción de la obesidad.

Es de utilidad porque hoy en día gran parte de la población cuenta con un dispositivo tecnológico con acceso a internet, que permite el uso de esta aplicación de una forma rápida

y sencilla, prediciendo la obesidad para una atención temprana por especialistas de la salud y así evitar el deterioro físico, psicológico y social del usuario.

Es factible porque la minería de datos se encuentra ampliamente abordada en artículos extranjeros, que se tomaron en cuenta para la realización de esta investigación.

Es de aporte científico ya que tiene como objetivos favorecer la predicción temprana de la obesidad en adolescentes y hacer uso de tecnologías actuales como minería de datos, aprendizaje automático entre otras. Por otro lado, servirá para futuras investigaciones en nuestro país, logrando mejorar la precisión de predicción de la obesidad.

Asimismo, esta investigación forma parte de los fondos que la Universidad Católica de Santa María utiliza para el apoyo de proyectos.

1.7.2 Importancia

Este proyecto contribuirá con la comunidad educativa, en el desarrollo de un software que permita la predicción de la obesidad, no sólo en centros escolares sino también en centros de salud y hospitales de la región.

En ese contexto los aportes de este proyecto no solo se enmarcarán en los beneficios de la comunidad educativa, sino también estos resultados servirán para el uso entre profesionales de la salud, además servirá como referencia para comparar con otros estudios internacionales.

1.8 Limitaciones de la Investigación

Según investigaciones anteriores, existen ya sistemas de predicción de obesidad en personas, en las cuales se utiliza minería de datos, por lo tanto, el presente proyecto tiene

antecedentes que pueden servir como base para el desarrollo. Además, la realización del proyecto cuenta con el apoyo de conocimientos y experiencia de especialistas en la materia.

1.9 Área, línea, tipo y nivel de la investigación

1.9.1 Área de investigación

El área de investigación es la Ingeniería de Software.

1.9.2 Línea de investigación

La línea de investigación es la Inteligencia Artificial.

1.9.3 Tipo de investigación

Aplicada.

1.9.4 Nivel de investigación

Experimental, debido a que se requiere predecir posibilidades de enfermedad de obesidad en adolescentes y se pretende probar el uso de 3 técnicas de minería de datos y seleccionar la más adecuada.

1.10 Método y diseño de investigación

1.10.1 Método de investigación

Investigación predictiva, empleando la metodología ágil SCRUM.

1.10.2 Forma de tratamiento de los datos.

Para el cálculo de los IMC, se capturan los datos en tablas, se realizan los cálculos respectivos y en una columna resultado, se muestra el valor del IMC. Para el procesamiento de los datos recolectados en las encuestas, se tiene una tabla de valorización, en donde cada respuesta tiene un valor correspondiente, del 1 al 3. El

objetivo es transcribir las respuestas a números, para obtener un resultado o puntaje medible.

a) Tabla de valorización de respuestas

Los datos ingresados al sistema, son ordenados en una tabla, en la que cada pregunta tiene un valor determinado, este valor nos sirve para realizar el procesamiento de la información.

b) Conjunto de datos Entrenamientos

Se cargan los datos anteriormente recopilados en encuestas físicas, en colegios de Perú, esto con el objetivo de establecer las reglas para el árbol de decisión.

c) Conjunto de datos TEST

Se cargan los datos de los usuarios que utilicen el sistema, y que respondan las preguntas adecuadamente, esto con el objetivo de predecir que clientes tienden a tener obesidad.

d) Reportes generales

Se crean reportes en donde se pueda visualizar por edades y/o percentiles, qué porcentaje de esa población tiene un alto o bajo índice de personas propensas a ser obesas.

e) Clasificación de resultados

Los grupos de clasificación pueden ser: Bajo de peso, Rango normal, Sobrepeso y Obesidad.

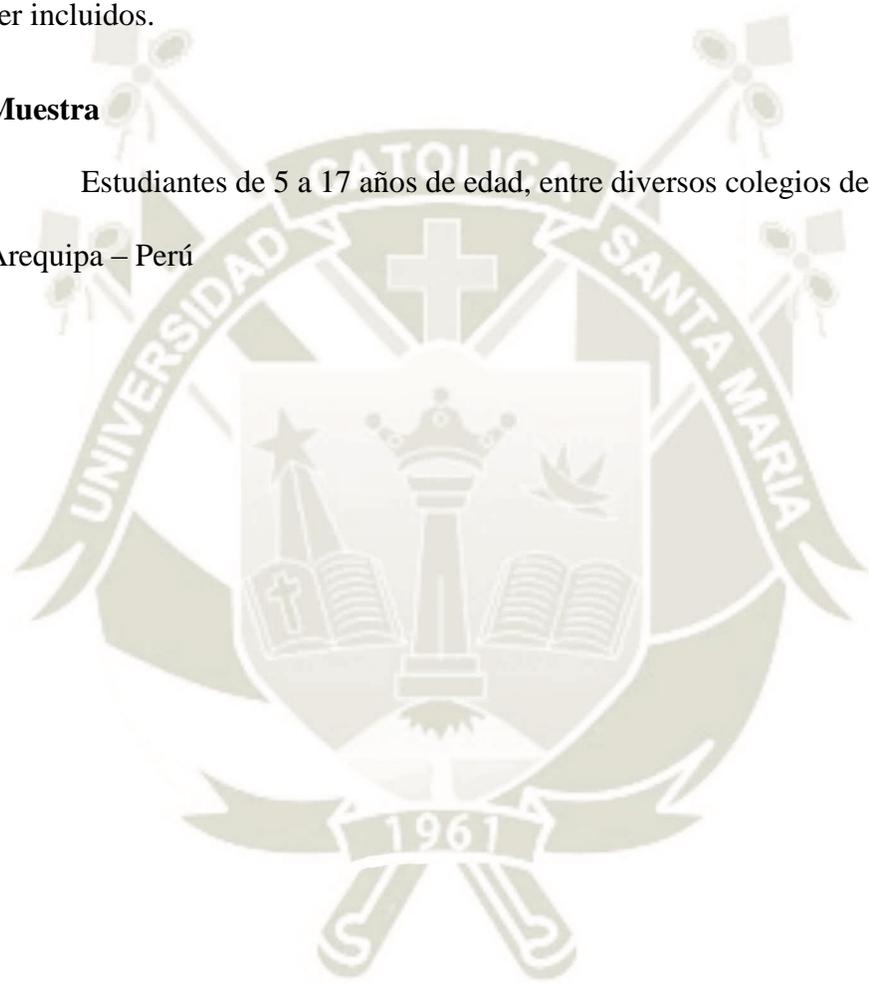
1.11 Cobertura del Estudio

1.11.1 Universo

Estudiantes, profesores de educación física y padres, entre diversos colegios de la ciudad de Arequipa – Perú. Especialistas en el campo de la salud también deben ser incluidos.

1.11.2 Muestra

Estudiantes de 5 a 17 años de edad, entre diversos colegios de la ciudad de Arequipa – Perú



CAPITULO 2: MARCO TEÓRICO

2.1 Estado del Arte

La minería de datos aplicada a la predicción de obesidad dirigido a una población de adolescentes, es un tema que no ha sido muy abordado en nuestro país, se sabe que a nivel mundial existen muchas investigaciones, utilizando minería de datos o aprendizaje automático, pero estas investigaciones no se ajustan del todo para la población peruana, es así que se crea la necesidad de aplicar dichas investigaciones en datos de nuestra localidad y automatizarlo en un sistema de Inteligencia artificial.

El 30% de la población mundial está afectada por problemas de sobrepeso y obesidad, y un número creciente de ellas presenta problemas de salud relacionados con estas causas. Tanto así que se ha evidenciado una tendencia creciente en las curvas de mortalidad como consecuencia de estos factores. (Fernández, 2017).

Alrededor de 360 millones de personas en Latinoamérica tienen un peso mayor al recomendado en función de la altura de la persona y otros indicadores. Esta cifra representa un 58% de la totalidad de habitantes de la región, de acuerdo a una investigación conjunta realizada por la Organización Panamericana de la Salud (OPS) y la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) que acaba de hacerse pública (BBC Mundo, 2017).

Según a la “Organización de las Naciones Unidas para la Alimentación y la Agricultura (2017)”, la obesidad y el sobrepeso han aumentado en el Perú y a lo largo de toda América Latina y el Caribe, con un impacto mayor en las mujeres y una tendencia

al alza en niños y niñas. El Perú se encuentra en el promedio regional: 6 de cada 10 hombres y mujeres adultos padecen sobrepeso en el país, exactamente el 58.2%.

La obesidad, por su parte, afecta a 140 millones de personas, el 23% de la población regional y las mayores prevalencias se pueden observar todas en países del Caribe: Bahamas (36,2%) Barbados (31,3%), Trinidad y Tobago (31,1%) y Antigua y Barbuda (30,9%). En Perú, también sobre el promedio regional, cerca del 20% de la población es obesa. ("Organización de las Naciones Unidas para la Alimentación y la Agricultura", 2017).

De acuerdo con Cossio-Bolaños, de Arruda, Sulla Torres, Urra Albornoz y Gómez Campos (2017), la determinación de la masa grasa y la masa libre de grasa (masa magra), es de considerable importancia en la evaluación del estado nutricional de niños, adolescentes y adultos. El aumento de grasa corporal se encuentra asociado al incremento de riesgo en enfermedades, como la obesidad, enfermedades cardiovasculares, diabetes tipo 2, hipertensión, entre otras. El objetivo de su investigación fue desarrollar ecuaciones de regresión para predecir la masa grasa, utilizando como método de referencia la absorciometría de rayos X de doble energía. Los resultados obtenidos fueron ecuaciones que mostraron aceptabilidad para determinar la masa grasa y además propusieron valores referenciales para evaluar la masa grasa en función de la edad cronológica y biológica.

Zheng y Ruggiero (2017) utilizaron cuatro modelos mejorados de aprendizaje automático para predecir la obesidad en estudiantes de secundaria al enfocarse en factores de riesgo y de protección: regresión logística binaria; árbol de decisión mejorado (IDT); vecino ponderado k más cercano (KNN); y red neuronal artificial (ANN). Utilizaron

como entradas modelo, nueve comportamientos relacionados con la salud del Sistema de Vigilancia del Comportamiento de Riesgo Juvenil 2015 (YRBSS) para el estado de Tennessee. Los resultados muestran que, en comparación con el modelo de regresión logística que logró una precisión del 56.02% y una especificidad del 54.77%, IDT, KNN ponderado y ANN tuvieron un desempeño significativamente mejor. El modelo IDT obtuvo una precisión del 80,23% y una especificidad del 90,74%, mientras que el modelo KNN ponderado obtuvo una precisión del 88,82% y una especificidad del 93,44%. El modelo ANN logró un 84.22% de precisión y un 99.46% de especificidad.

En un artículo desarrollado por Alliyarova et al., (2016), la aplicación de datos se utilizó para detectar factores clave que afectan el peso corporal de acuerdo con los registros del cuestionario tomados de adultos que viven en las regiones del norte y sur de Kazajstán. Además, se presentaron los modelos predictivos que pueden aceptarse como un método alternativo para definir el IMC entre adultos para ciudadanos de Kazajstán. Por lo tanto, aseguran que las tecnologías de minería de datos y aprendizaje automático se están convirtiendo en una herramienta popular que ayuda a analizar y recuperar valiosos conocimientos de datos brutos.

En un estudio realizado por Suca et al., (2016), se analizaron los pasos previos relacionados a la predicción de obesidad, es así que se hizo una comparación de algoritmos de clasificación de minería de datos para obtener el clasificador más adecuado que mejore la precisión de predicción en casos de obesidad. Los algoritmos clasificadores utilizados fueron árboles de decisión (J48), Naive Bayes, SVM y redes neuronales. Como resultados finales de la comparación, se demostró que el clasificador basado en árboles

de decisión (J48), es el más adecuado para realizar la predicción del tipo de obesidad con una tasa de precisión de 97.23%.

En la actualidad existen diversas causas que ocasionan la obesidad infantil, una de las más influyentes es el tabaquismo durante el embarazo. Es bien sabido que el tabaquismo materno durante el embarazo puede provocar una disminución del peso al nacer y de la estatura. Los recién nacidos de bajo peso a menudo muestran un mayor crecimiento de recuperación que los bebés de peso normal y posteriormente se vuelven obesos en la infancia. Artículos recientes de revisión han sugerido que el tabaquismo materno es un factor de riesgo significativo para la obesidad en la vida posterior. (Ino, 2010).

Así como existen diversas causas para la obesidad infantil, también lo hay para los adolescentes como el abandono de hábitos alimenticios saludables, el cambio hacia un ritmo de vida acelerado (que lleva a consumir comida rápida o comida basura) y la escasa práctica de ejercicio físico. (Sevilla Martínez, 2017).

Muchos métodos y técnicas como la evaluación de la composición corporal, las técnicas de minería de datos y la regresión logística se han aplicado para predecir la obesidad, pero sólo unos pocos lograron producir resultados precisos. El número de esfuerzos en la predicción de la obesidad debe aumentarse y las técnicas utilizadas deben ser improvisadas. La etapa inicial debe consistir en la recolección de datos de fuentes primarias. Luego, identificar factores de riesgo como la educación de los padres, el estilo de vida, los hábitos y las influencias ambientales, y proponer un marco de predicción de

la obesidad utilizando algún método. (Al Mamun, Cramb, O'Callaghan, Williams & Najman, 2009).

Las técnicas de Minería de datos han sido aplicadas a muchas áreas en el mundo de los negocios y en nuestra vida diaria, incluidos los servicios de la salud, y clínicos. Uno de los problemas más vistos es la obesidad y el sobrepeso, especialmente en niños y adolescentes. Para empezar a aplicar estas técnicas, primero se debe encontrar los factores de riesgo de estilo de vida más importantes asociados con el sobrepeso y la obesidad entre los estudiantes. En un artículo realizado por (Pochini, Wu & Hu, 2014), se recolectó dichos factores en una escuela de secundaria en Estados Unidos, mediante una encuesta, los datos de estilo de vida de la Encuesta Nacional de Comportamiento del Riesgo de la Juventud 2011 (YRBS, por sus siglas en inglés) se usaron con los estados de peso corporal de los estudiantes, sobrepeso u obesidad, considerados como dos variables objetivo. Se crearon modelos de regresión logística y modelos de árbol de decisión para cada variable objetivo. Tanto el método de regresión logística como el del árbol de decisión muestran que frecuentemente hacer actividad física y desayunar son todos factores de protección contra el sobrepeso u obesidad. Fumar y beber bebidas azucaradas con frecuencia se asoció con un mayor riesgo de ser obeso.

Zhang et al., (2009), realizaron una comparación de regresión logística con seis técnicas de extracción de datos, mejoró la precisión de la predicción: la predicción a 8 meses de precisión se mejora muy ligeramente, en ese caso mediante el uso de redes neuronales, mientras que para la predicción a los 2 años se mejora en más del 10% con el uso de métodos Bayesianos. También demuestra que la incorporación de interacciones

no lineales podría ser importante en la predicción epidemiológica y que las técnicas de minería de datos están suficientemente establecidas para ofrecer a la comunidad de investigación médica una alternativa válida a la regresión logística.

Por otro lado, se realizaron pruebas para comparar el rendimiento de regresión logística, redes neuronales y modelos de árboles de decisión para predecir la diabetes o prediabetes usando factores de riesgo comunes. En esta investigación los participantes vinieron de dos comunidades en Guangzhou, China; 735 pacientes confirmados con diabetes o prediabetes y 752 controles normales fueron reclutados. Se administró un cuestionario estándar para obtener información sobre características demográficas, historial de diabetes familiar, mediciones antropométricas y factores de riesgo de estilo de vida. Luego se desarrolló tres modelos predictivos utilizando 12 variables de entrada y una variable de salida a partir de la información del cuestionario; se evaluó los tres modelos en términos de precisión, sensibilidad y especificidad. El modelo de regresión logística logró una precisión de clasificación de 76.13% con una sensibilidad de 79.59% y una especificidad de 72.74%. El modelo redes neuronales alcanzó una precisión de clasificación de 73.23% con una sensibilidad de 82.18% y una especificidad de 64.49%; y el árbol de decisión (C5.0) logró una precisión de clasificación del 77.87% con una sensibilidad del 80.68% y una especificidad del 75.13%. El modelo de árbol de decisión tuvo la mejor precisión de clasificación, seguido del modelo de regresión logística, y la ANN dio la menor precisión. (Meng, Huang, Rao, Zhang & Liu, 2013).

Hoy en día, el problema de la obesidad entre los niños es uno de los temas comúnmente explorados usando técnicas de minería de datos. En este trabajo, se discute

la clasificación de la obesidad infantil entre los escolares del sexto año de dos distritos en Terengganu, Malasia. Los datos fueron recolectados de dos fuentes principales; a Estándar Kecergasan Fizikal Kebangsaan para Murid Sekolah Malasia / Estándar Nacional de Aptitud Física para el Programa de Evaluación de Niños Escolares de Malasia (SEGAK) y un conjunto de cuestionarios distribuidos. A partir de los datos recopilados, se analizaron prontamente 4,245 conjuntos de datos completos. El pre procesamiento de datos y la selección de características se implementaron en los conjuntos de datos. Las técnicas de clasificación, Redes Bayesianas, Árboles de Decisión, Redes Neuronales y Máquinas de Vectores de Soporte se implementaron y se compararon en los conjuntos de datos. (Abdullah et al., 2016) presenta la evaluación de varios métodos de selección de características basados en diferentes clasificadores.

Venkatasubramaniam, Wolfson, Mitchell, Barnes, JaKa y French (2017), aseguran que tanto los algoritmos CART como CTree identifican subgrupos homogéneos de población y ofrecen una precisión de predicción mejorada en relación con los enfoques basados en la regresión cuando los subgrupos están realmente presentes en los datos. Una distinción importante entre CART y CTree es que este último usa un marco formal de prueba de hipótesis estadísticas para construir árboles de decisión, lo que simplifica el proceso de identificación e interpretación del modelo de árbol final. Finalmente comprobaron que los árboles de decisión son una herramienta útil para identificar subgrupos homogéneos definidos por combinaciones de características individuales.

2.2 Ingeniería de Software

La ingeniería de software es un enfoque sistemático para la producción de software que toma en cuenta los temas prácticos de costo, fecha y confiabilidad, así como las necesidades de clientes y fabricantes de software. Como este enfoque sistemático realmente implementado varía de manera drástica dependiendo de la organización que desarrolla el software, el tipo de software y los individuos que intervienen en el proceso de desarrollo, no existen métodos y técnicas universales de ingeniería de software que sean adecuados para todos los sistemas y las compañías. Más bien, durante los últimos 50 años evolucionó un conjunto de métodos y herramientas de ingeniería de software. (Sommerville, 2011).

Existen muchos diferentes procesos de software, pero todos deben incluir cuatro que son fundamentales para la ingeniería de software.

1. Especificación de software. Tienen que definirse tanto la funcionalidad del software como las restricciones de su operación.
2. Diseño e implementación del software. Debe desarrollarse el software para cumplir con las especificaciones.
3. Validación del software. Hay que validar el software para asegurarse de que cumple lo que el cliente quiere.
4. Evolución del software. El software tiene que evolucionar para satisfacer las necesidades cambiantes del cliente.

Un modelo de proceso de software es una representación simplificada de este proceso. Cada modelo del proceso representa a otro desde una particular perspectiva y,

por lo tanto, ofrece sólo información parcial acerca de dicho proceso. Por ejemplo, un modelo de actividad del proceso muestra las actividades y su secuencia, pero quizá sin presentar los roles de las personas que intervienen en esas actividades. Para la realización de este proyecto se utilizó el proceso de software llamado “El modelo en cascada”.

El modelo en cascada (Waterfall). Éste toma las actividades fundamentales del proceso de especificación, desarrollo, validación y evolución y, luego, los representa como fases separadas del proceso, tal como especificación de requerimientos, diseño de software, implementación, pruebas, etcétera. (Sommerville, 2011).

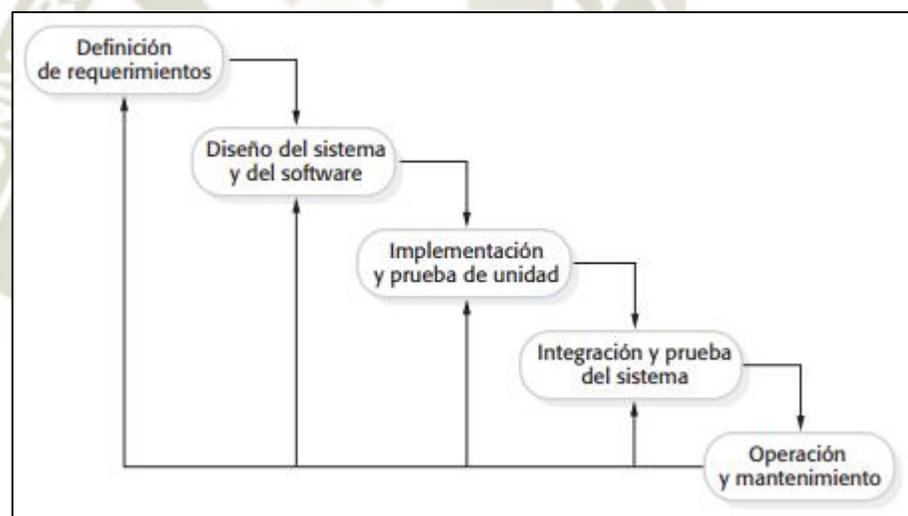


Figura 1. El modelo en cascada. Fuente: (Sommerville, 2011).

2.3 Inteligencia Artificial

La Inteligencia Artificial (IA) es una rama de las ciencias computacionales encargada de estudiar modelos de cómputo capaces de realizar actividades propias de los seres humanos en base a dos de sus características primordiales: el razonamiento y la

conducta. Existen distintas definiciones de IA de acuerdo a los distintos enfoques; algunas de estas definiciones se muestran a continuación. (Takeyas, 2007).

“La interesante tarea de lograr que las computadoras piensen ... máquinas con mente, en su amplio sentido literal.” (Haugeland, 1985)

“La automatización de actividades que vinculamos con procesos de pensamiento humano, actividades tales como la toma de decisiones, resolución de problemas, aprendizaje...” (Bellman, 1978)

“El estudio de las facultades mentales mediante el uso de modelos computacionales.” (Charniak y McDermott, 1985)

2.4 Aprendizaje Automático

El aprendizaje automático, es un campo de investigación el cual desarrolla y estudia algoritmos que pueden aprender y hacer predicciones sobre la data. También ofrece herramientas para resolver problemas del mundo real: clasificación, regresión, agrupamiento (clustering), aprendizaje en línea, aprendizaje semi supervisado, etc. (Christopher, M. B., 2016).

Desarrolla diferentes métodos para medir la efectividad de los procesos de aprendizaje, estudia la habilidad de un algoritmo de aprender con el objetivo de evitar la memorización y mejorar el rendimiento de generalización, que es la capacidad de aprender efectivamente el concepto seleccionado.

2.5 Minería de datos

La minería de datos es el núcleo del proceso Descubrimiento de Conocimiento en Bases de Datos, que implica inferir algoritmos que exploren datos, desarrollar un modelo y descubrir previamente patrones desconocidos. El modelo es usado para el entendimiento de fenómenos de los datos, análisis y predicción. En este proceso se utilizan métodos de inteligencia artificial, aprendizaje automático, estadística y bases de datos.

La accesibilidad y la abundancia de los datos de hoy hace que el Descubrimiento de Conocimiento y la Minería de Datos sean una cuestión de considerable importancia y necesidad. Dado el reciente crecimiento del campo, no es una sorpresa que una amplia variedad de métodos se encuentra actualmente disponible para investigadores y practicantes. (Maimon, O., & Browarnik, A., 2010).

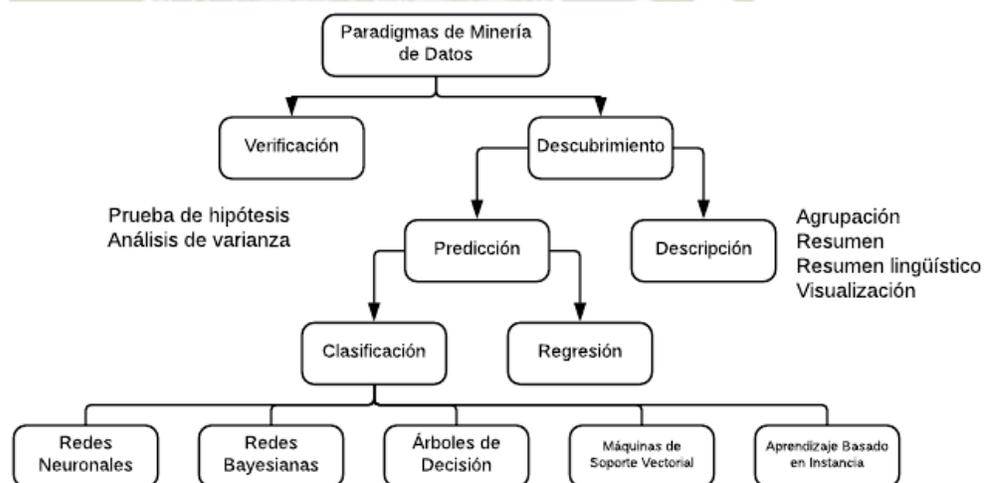


Figura 2. Taxonomía de la Minería de Datos. Fuente: (Maimon, O., & Browarnik, A., 2010).

2.6 Proceso de Descubrimiento de Conocimiento en Bases de Datos

El descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases) es un proceso organizado de identificar patrones válidos, novedosos, útiles y entendibles de conjuntos de datos grandes y complejos.

El proceso de descubrimiento de conocimiento es iterativo e interactivo, consiste en nueve pasos.

1. Desarrollar una comprensión del dominio de la aplicación. Es necesario preparar el escenario para el entendimiento de qué debería ser hecho con muchas decisiones sobre transformación, algoritmos, representación, etc. Además, es necesario entender y definir los objetivos de los usuarios finales y el ambiente en el cual el proceso de descubrimiento de conocimiento va a tomar lugar.
2. Seleccionar y crear un conjunto de datos en el que se realizará el proceso. Una vez definidos los objetivos, la información que será usada para el descubrimiento del conocimiento debe ser determinada. Esto incluye averiguar que data se tiene disponible, obtener data adicional y entonces integrar toda la información.
3. Pre procesamiento y limpieza. Esto incluye la limpieza de la información, el manejo de valores nulos, y la eliminación de valores erróneos.
4. Transformación de datos. Se prepara y se desarrolla mejor la información para el proceso de minería de datos.
5. Elegir la tarea apropiada de Minería de datos. Se decide el tipo de minería de datos a usar, por ejemplo, clasificación, regresión o agrupamiento. La decisión en la mayoría de los casos depende de los objetivos del proceso de Descubrimiento del Conocimiento y también de los pasos anteriores.

6. Elegir el algoritmo de Minería de datos. Se selecciona el método específico a ser usado para la búsqueda de patrones. Cada algoritmo tiene parámetros y tácticas de aprendizaje.
7. Empleando el algoritmo de minería de datos. Se realiza la implementación del algoritmo de minería de datos. En este paso se necesita emplear el algoritmo varias veces hasta obtener un resultado satisfactorio.
8. Evaluación. Se evalúa y se interpreta los patrones obtenidos (reglas, confiabilidad, etc.), con respecto a los objetivos definidos en el primer paso.
9. Uso el conocimiento descubierto. En este paso, se tiene todo lo necesario para implementar el conocimiento en otro sistema para dar mejor uso. (Maimon, O., & Browarnik, A., 2010).

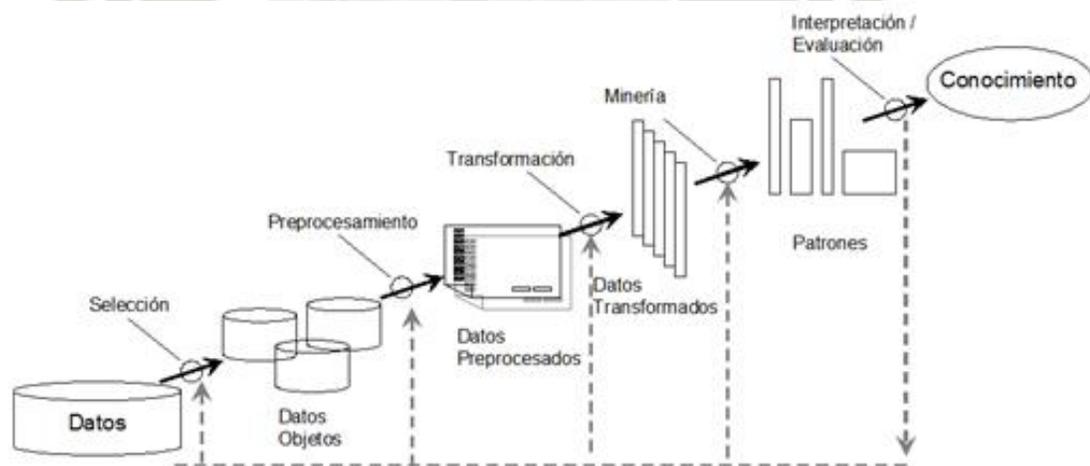


Figura 3. El proceso de Descubrimiento de Conocimiento en Bases de Datos. Fuente: (Maimon, O., & Browarnik, A., 2010).

2.7 Técnicas de Minería de Datos

Hay una variedad de técnicas importantes de minería de datos, que han sido desarrolladas y utilizadas en proyectos de minería de datos recientemente incluyendo asociación, clasificación, agrupamiento, predicción y patrones secuenciales, estos han sido usados para el descubrimiento de conocimiento en bases de datos.

2.8 Asociación

Asociación es una de las técnicas de minería de datos más conocidas. Esta técnica consiste en que un patrón es descubierto basado en una relación, de un elemento en particular con otros elementos de la misma transacción. (Raval, 2012).

2.8.1 Agrupamiento (Clustering)

Es el proceso de organizar objetos dentro de grupos de los cuales los miembros son similares de cierta forma. Por lo tanto, consiste en una colección de objetos que son “similares” entre ellos, y son “no similares” a los objetos que pertenecen a otros grupos o clusters. (Raval, 2012).

2.8.2 Clasificación

Es una técnica de minería de datos clásica, basada en aprendizaje automático. Básicamente, la clasificación es utilizada para clasificar cada ítem de un conjunto de datos dentro de un conjunto predefinido de clases o grupos. Los métodos de clasificación hacen uso de técnicas de matemáticas tales como árboles de decisión, programación lineal, redes neuronales y estadísticas. (Raval, 2012).

a) Redes neuronales

Redes Neuronales Artificiales, son los modelos computacionales inspirados por el cerebro humano. Muchos de los recientes avances han sido hechos en el campo de la inteligencia artificial, incluyendo reconocimiento de voz, reconocimiento de imágenes y robótica utilizando redes neuronales artificiales. Son simulaciones inspiradas biológicamente, realizadas en la computadora para realizar ciertas tareas específicas, como clustering, clasificación, reconocimiento de patrones, etc. (Kaur Gill, 2017).

b) Árboles de decisión

Los árboles de decisión son modelos secuenciales, los cuales lógicamente combinan una secuencia de simples pruebas; cada prueba compara un atributo numérico contra un valor límite o un valor nominal contra un conjunto de valores posibles. Dichos clasificadores simbólicos tienen una ventaja sobre los modelos de caja negra, como las redes neuronales, en términos de comprensibilidad. Las reglas lógicas seguidas por un árbol de decisión son mucho más fáciles de interpretar que los pesos numéricos de las conexiones entre los nodos de una red neuronal. (Kotsiantis, 2013).

c) Redes Bayesianas

Una red bayesiana, es un grafo acíclico dirigido (DAG, por sus siglas en inglés) en el cual los nodos representan las variables de sistema y los arcos simbolizan las dependencias o la relación causa-efecto entre las variables. Una red bayesiana es definida por un conjunto de nodos y un conjunto de arcos

dirigidos. Las aristas representan dependencias condicionales; los nodos que no se encuentran conectados representan variables que son independientes de las otras. (Weber, Medina-Oliva, Simon & Iung, 2012).

2.9 Modelo Cliente – Servidor

El modelo Cliente/Servidor es un modelo de aplicación distribuida en la que las tareas se reparten entre los proveedores de recursos o servicios, llamados servidores, y los demandantes, llamados clientes. Las aplicaciones Clientes realizan peticiones a una o varias aplicaciones Servidores, que deben encontrarse en ejecución para mantener dichas demandas.

El modelo Cliente/Servidor permite diversificar el trabajo que realiza cada aplicación, de forma que los clientes no se sobrecarguen, cosa que ocurriría si ellos mismos desempeñan las funciones que les son proporcionadas de forma directa y transparente. En esta arquitectura la capacidad de proceso está repartida entre los clientes y los servidores, aunque son más importantes las ventajas de tipo organizativo debidas a la centralización de la gestión de la información y la separación de responsabilidades, lo que facilita y clarifica el diseño del sistema. Tanto el cliente como el servidor son entidades abstractas que pueden residir en la misma máquina o en máquinas diferentes.

El modelo de servicios establece tres grandes conjuntos de funcionalidades, en cada uno de los cuales se encuadran las distintas tareas en las que se ve involucrado cualquier tipo de proyecto de desarrollo. Dicho modelo establece los siguientes conjuntos que comúnmente denominamos lógicas. (Marini, 2012).

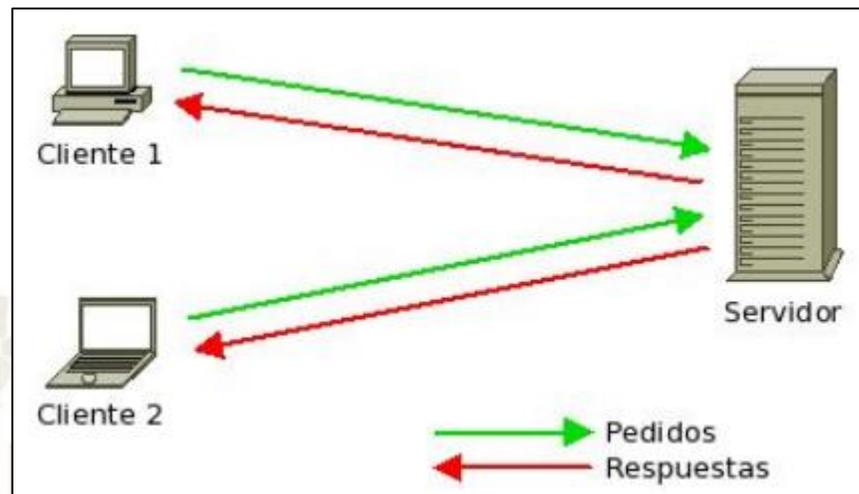


Figura 4. El modelo Cliente/Servidor. Fuente: (Marini, 2012).

2.9.1 Lógica de Presentación

Esta lógica es la responsable del control de todos los aspectos relacionados con la interacción entre el cliente y la aplicación. Para llevar a cabo esta tarea de control, es necesario conocer que tipos de usuarios utilizarán la aplicación, qué actividades tienen que realizar y, teniendo en cuenta estos datos, cuáles son los mejores estilos de interfaz para que esos usuarios realicen sus tareas. En esta lógica se engloban todas las tareas que deben ser realizadas por la parte del Cliente del modelo general. (Marini, 2012).

2.9.2 Lógica de Negocio

Es la lógica de la aplicación que controla la secuencia de acciones y fuerza el cumplimiento de las reglas de negocio propias de cada empresa; además asegura la integridad de las transacciones de las operaciones necesarias que haya que realizar para que se cumplan dichas reglas. La lógica de negocio también transforma una serie

de datos en información útil para el usuario mediante la aplicación de las reglas apropiadas.

El objetivo que debe de cumplir esta lógica es el de aislar las reglas de negocio, así como las transformaciones de datos de los consumidores (usuarios y otros componentes de esta misma capa) y de los sistemas de gestión de datos. (Marini, 2012).

2.9.3 Lógica de Datos

En este conjunto entran los procesos encargados de la gestión de los datos, es decir los procesos encargados del mantenimiento de los datos, de garantizar las reglas de integridad referencial establecidas, así como de la gestión de las transacciones. Estas tareas son realizadas, generalmente por un Sistema de Gestión de Base de datos Relacionales, como SQL Server, Oracle, MySQL, Informix, etc. (Marini, 2012).

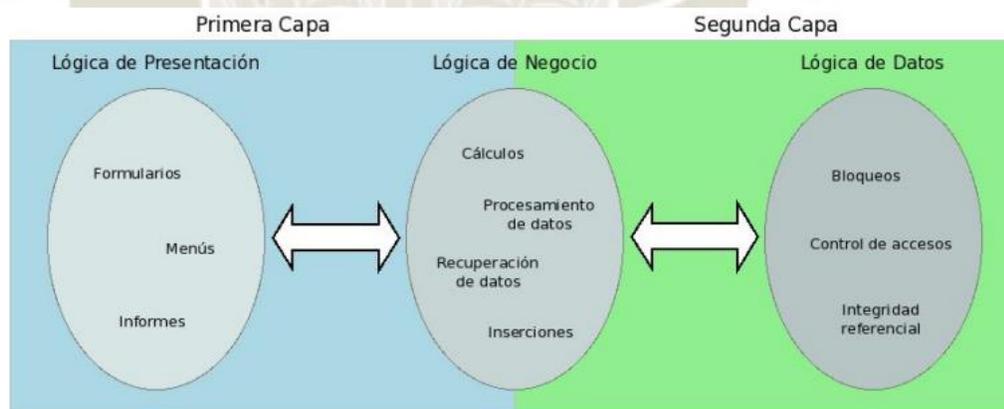


Figura 5. El modelo Cliente/Servidor de 2 capas. Fuente: (Marini, 2012).

2.9.4 Componentes

- **Servidor**

Ofrecen servicios a los otros componentes. Ejemplos de estos incluyen servidores de impresión; servidores de archivo que brindan servicios de administración de archivos, y un servidor compilador, que proporciona servicios de compilación de lenguaje de programación.

- **Clientes**

Este componente es el que solicita los servicios que ofrecen los servidores. Habrá usualmente varias instancias de un programa cliente que se ejecuten de manera concurrente en diferentes computadoras.

- **Red**

Es la red que permite a los clientes acceder a dichos servicios. La mayoría de los sistemas cliente-servidor se implementan como sistemas distribuidos conectados mediante protocolos de internet. (Sommerville, 2011).

La arquitectura cliente-servidor se consideran a menudo como arquitecturas de sistemas distribuidos; sin embargo, el modelo lógico de servicios independientes que opera en servidores separados puede implementarse en una sola computadora. Un beneficio importante es la separación e independencia. Los servicios y servidores pueden cambiar sin afectar otras partes del sistema.

En la figura 6, se muestra un ejemplo de un sistema que se basa en el modelo cliente-servidor. Se trata de un sistema multiusuario basado en la Web, para ofrecer un repertorio de películas y fotografías. En este sistema, varios servidores manejan y

despliegan los diferentes tipos de medios. Los cuadros de video necesitan transmitirse rápidamente y en sincronía, aunque a una resolución relativamente baja. Tal vez estén comprimidos en un almacén, de manera que el servidor de video puede manipular en diferentes formatos de compresión y descompresión del video. Sin embargo, las imágenes fijas deben conservarse en una resolución alta, por lo que es adecuado mantenerlas en un servidor independiente.

El catálogo debe manejar una variedad de consultas y ofrecer vínculos hacia el sistema de información Web, que incluye datos acerca de las películas y los videos, así como un sistema de comercio electrónico que soporte la venta de fotografías, películas y videos. El programa cliente es simplemente una interfaz integrada de usuario, construida mediante un navegador Web, para acceder a dichos servicios. (Sommerville, 2011).

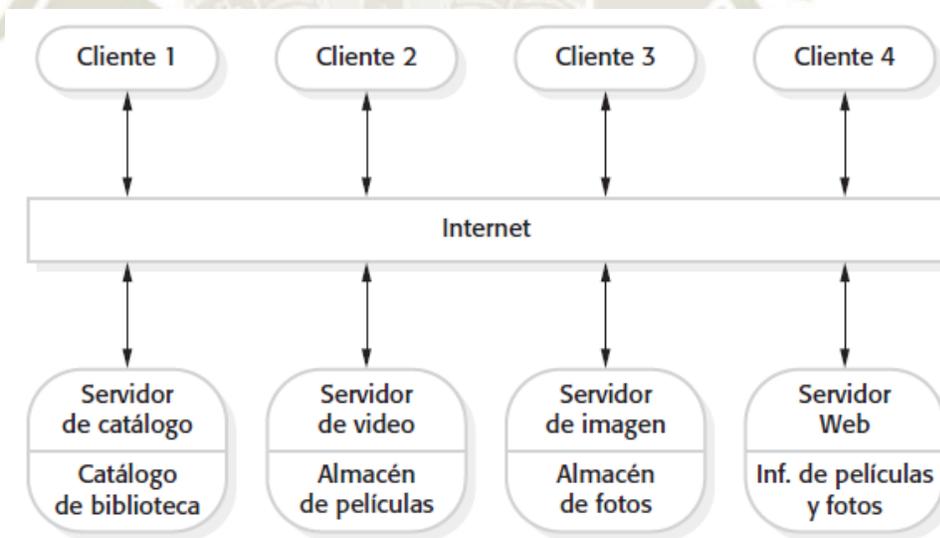


Figura 6. Arquitectura cliente servidor para una filmoteca. Fuente: (Sommerville, 2011).

2.9.5 Ventajas

Fácil de entender y soporta reutilización de transformación. El estilo del flujo de trabajo coincide con la estructura de muchos procesos empresariales. La evolución de agregar transformaciones es directa. Puede implementarse como un sistema secuencial o como un concurrente.

2.9.6 Desventajas

El formato para la transferencia de datos debe acordarse entre las transformaciones que se comunican. Cada transformación debe analizar sus entradas y sintetizar sus salidas al formato acordado. Esto aumenta la carga del sistema, y puede significar que sea imposible reutilizar transformaciones funcionales que usen estructuras de datos incompatibles. (Sommerville, 2011).

2.10 Lenguaje de programación PHP

PHP es el acrónimo de Hypertext Preprocesor. Es un lenguaje de programación del lado del servidor gratuito e independiente de plataforma, rápido, con una gran librería de funciones y mucha documentación (Beati, 2015).

2.10.1 Ventajas del lenguaje PHP

El lenguaje ofrece las siguientes ventajas:

- Una de sus características es la rapidez.
- Lo soportan la mayoría de las plataformas de alojamiento web.
- Tiene ciertas características de los lenguajes orientados a objetos como la utilización de clases y herencias.
- Puede mezclarse con código HTML, aunque esto dificulta su lectura.

- Puede manejar ficheros y conectarse a distintas bases de datos (MySQL, Oracle, SQL Server, Informix, PostgreSQL, etc.).
- El software que permite soportarlo en los servidores de hosting es libre y gratuito.
- Está en continuo desarrollo y soporta numerosas funcionalidades.
- Existe numerosa documentación sobre el lenguaje en Internet por lo que es relativamente sencillo resolver los problemas que nos puedan surgir durante el desarrollo de un sitio web (Beati, 2015).

2.10.2 Desventajas del lenguaje PHP:

- Para poder ver y testear las páginas que vayamos creando es necesario disponer de un servidor web que soporte PHP.
- Al ser interpretado en el servidor, es más fácil que se colapse cuando el número de peticiones de descarga de páginas aumenta.
- Parte del contenido de las páginas puede no ser accesible a los navegadores, dificultando el posicionamiento de las páginas. (Beati, 2015).

2.11 Metodología Ágil

Los métodos ágiles se apoyan universalmente en el enfoque incremental para la especificación, el desarrollo y la entrega de software. Son más adecuados para el diseño de aplicaciones en que los requerimientos del sistema cambian, por lo general rápidamente durante el desarrollo. Tienen la intención de entregar con prontitud el software operativo a los clientes, quienes entonces propondrán requerimientos nuevos y variados para incluir en posteriores iteraciones del sistema. Se dirigen a simplificar

el proceso burocrático al evitar trabajo con valor dudoso a largo plazo, y a eliminar documentación que quizá nunca se emplee (Sommerville, 2011).

La filosofía detrás de los métodos ágiles se refleja en el manifiesto ágil, que acordaron muchos de los desarrolladores líderes de estos métodos. Este manifiesto afirma:

Estamos descubriendo mejores formas para desarrollar software, al hacerlo y al ayudar a otros a hacerlo. Gracias a este trabajo llegamos a valorar:

A los individuos y las interacciones sobre los procesos y las herramientas

Al software operativo sobre la documentación exhaustiva

La colaboración con el cliente sobre la negociación del contrato

La respuesta al cambio sobre el seguimiento de un plan

Esto es, aunque exista valor en los objetos a la derecha, valoraremos más los de la izquierda (Sommerville, 2011).

2.11.1 Ventajas

En este contexto se podrían definir las siguientes ventajas:

- Gran capacidad de respuesta ante los cambios, los cuales no se entienden como un problema sino como algo necesario para que el producto sea mejor y satisfice al cliente. Los cambios formarán parte del proceso de desarrollo.
- Las entregas no se hacen al final, sino que se hacen pequeñas entregas. Estas entregas permiten al cliente valorar el producto además de ir trabajando con algunas funcionalidades.

- Los ciclos cortos de entrega ayudarán a disminuir los riesgos sobre todo al principio del proyecto.
- Se trabaja en equipo entre el cliente y los desarrolladores mediante una comunicación casi diaria para evitar errores y documentación innecesaria.
- Eliminar el trabajo que no es necesario y que realmente no aporta un valor al negocio.
- Buscar la mejor técnica y el mejor diseño para conseguir productos de calidad.
- Mejorar los procesos y al equipo que realiza el desarrollo. (Gallego, 2012).

2.11.2 Fases

Para entender el ciclo de desarrollo Scrum es necesario conocer las 5 fases que definen el ciclo de desarrollo ágil:

1. Concepto. Se define de forma general las características del producto y se asigna el equipo que se encargará de su desarrollo.
2. Especulación. En esta fase se hacen disposiciones con la información obtenida y se establecen los límites que marcarán el desarrollo del producto, tales como costes y agendas. Se construirá el producto a partir de las ideas principales y se comprueban las partes realizadas y su impacto en el entorno.

Esta fase se repite en cada iteración y consiste, en rasgos generales, en:

- Desarrollar y revisar los requisitos generales.
- Mantener la lista de las funcionalidades que se esperan.
- Plan de entrega. Se establecen las fechas de las versiones, hitos e iteraciones. Medirá el esfuerzo realizado en el proyecto.

3. Exploración. Se incrementa el producto en el que se añaden las funcionalidades de la fase de especulación.
 4. Revisión. El equipo revisa todo lo que se ha construido y se contrasta con el objetivo deseado.
 5. Cierre. Se entregará en la fecha acordada una versión del producto deseado.
- Al tratarse de una versión, el cierre no indica que se ha finalizado el proyecto, sino que seguirá habiendo cambios, denominados “mantenimiento”, que hará que el producto final se acerque al producto final deseado. (Gallego, 2012).



Figura 7. Ciclo de desarrollo ágil. Fuente: (Gallego, 2012).

2.11.3 Metodología Scrum

El enfoque Scrum es un método ágil general, su enfoque está en la administración iterativa del desarrollo, y no en enfoques técnicos específicos para la ingeniería de software ágil. La figura 9 representa un diagrama del proceso de

administración de Scrum. Este proceso no prescribe el uso de prácticas de programación, como la programación en pares y el desarrollo de primera prueba. Por lo tanto, puede usarse con enfoques ágiles más técnicos, como XP, para ofrecer al proyecto un marco administrativo (Sommerville, 2011).

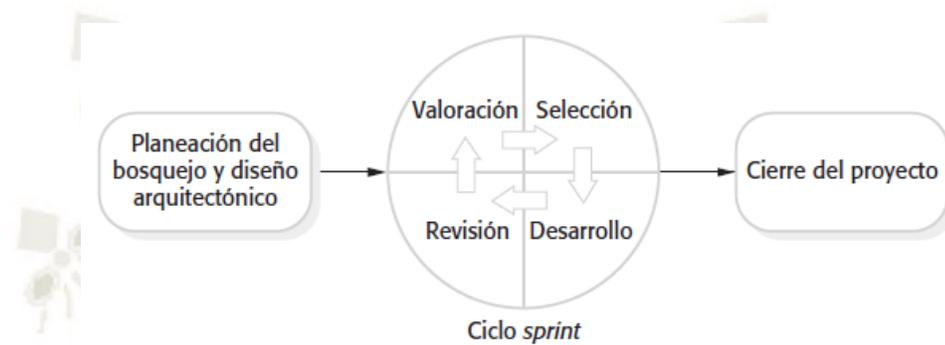


Figura 8. El proceso de Scrum. Fuente: (Sommerville, 2011).

Existen tres fases con Scrum. La primera es la planeación del bosquejo, donde se establecen los objetivos generales del proyecto y el diseño de la arquitectura de software. A esto le sigue una serie de ciclos sprint, donde cada ciclo desarrolla un incremento del sistema. Finalmente, la fase de cierre del proyecto concluye el proyecto, completa la documentación requerida, como los marcos de ayuda del sistema y los manuales del usuario, y valora las lecciones aprendidas en el proyecto. (Sommerville, 2011).

a) Componentes de Scrum.

Scrum se puede dividir de forma general en 3 fases, que podemos entender como reuniones. Las reuniones forman parte de los artefactos de esta metodología junto con los roles y los elementos (Gallego, 2012).

- **Las Reuniones.**

- Planificación del Backlog. Se definirá un documento en el que se reflejarán los requerimientos del sistema por prioridades. En esta fase se definirá también la planificación del Sprint 0, en la que se decidirá cuáles van a ser los objetivos y el trabajo que hay que realizar para esa iteración. Se obtendrá además en esta reunión un Sprint Backlog, que es la lista de tareas y que es el objetivo más importante del Sprint.
- Seguimiento del Sprint. En esta fase se hacen reuniones diarias en las que las 3 preguntas principales para evaluar el avance de las tareas serán:
 - ¿Qué trabajo se realizó desde la reunión anterior?
 - ¿Qué trabajo se hará hasta una nueva reunión?
 - Inconvenientes que han surgido y qué hay que solucionar para poder continuar
- Revisión del Sprint. Cuando se finaliza el Sprint se realizará una revisión del incremento que se ha generado. Se presentarán los resultados finales y una demo o versión, esto ayudará a mejorar el feedback con el cliente (Gallego, 2012).

- **Los Roles.**

Las personas que están comprometidas con el proyecto y el proceso de Scrum.

- Product Owner: Es la persona que toma las decisiones, y es la que realmente conoce el negocio del cliente y su visión del producto. Se encarga de escribir las ideas del cliente, las ordena por prioridad y las coloca en el Product Backlog.
- Scrum Master: Es el encargado de comprobar que el modelo y la metodología funciona. Eliminará todos los inconvenientes que hagan que el proceso no fluya e interactuará con el cliente y con los gestores.
- Equipo de Desarrollo: suele ser un equipo pequeño de unas 5-9 personas y tienen autoridad para organizar y tomar decisiones para conseguir su objetivo. Está involucrado en la estimación del esfuerzo de las tareas del Backlog.

A continuación, personajes que no son parte del proceso de Scrum, pero están involucrados en el proyecto.

- Usuarios: Es el destinatario final del producto.
- Stakeholders: Las personas a las que el proyecto les producirá un beneficio. Participan durante las revisiones del Sprint.
- Managers: Toma las decisiones finales participando en la selección de los objetivos y de los requerimientos (Gallego, 2012).

2.12 La guía SWEBOK

La guía del Cuerpo de Conocimiento de Ingeniería de Software (SWEBOK, por sus siglas en inglés) describe el conocimiento generalmente aceptado sobre ingeniería de software. Sus 15 áreas de conocimiento resumen conceptos básicos e incluyen una lista de referencias que apunta a información más detallada. La Guía SWEBOK también ha ganado reconocimiento internacional como ISO Technical Report 19759. En renovaciones futuras, Computer Society y sus voluntarios continuarán utilizando el proceso de consenso abierto y transparente que es una parte integral de SWEBOK. (Abran, Moore, Bourque, Dupuis, & Tripp, 2004)

A continuación, se muestran las áreas de conocimiento tratadas en la guía SWEBOK, cada una de estas áreas son tratadas como un capítulo en la guía:

Tabla 1

Áreas de conocimiento del SWEBOK. Fuente: (Abran et al., 2004).

Requerimientos del software
Diseño del software
Construcción del software
Pruebas del software
Mantenimiento del software
Gestión de la configuración del software
Gestión en la ingeniería del software
Métodos y herramientas de la ingeniería del software
Calidad del software

La guía utiliza una organización jerárquica para descomponer cada área de conocimiento en un conjunto de temas catalogados. Una descomposición en 2 o 3 niveles proporciona una manera razonable de encontrar los temas (Abran et al., 2004).

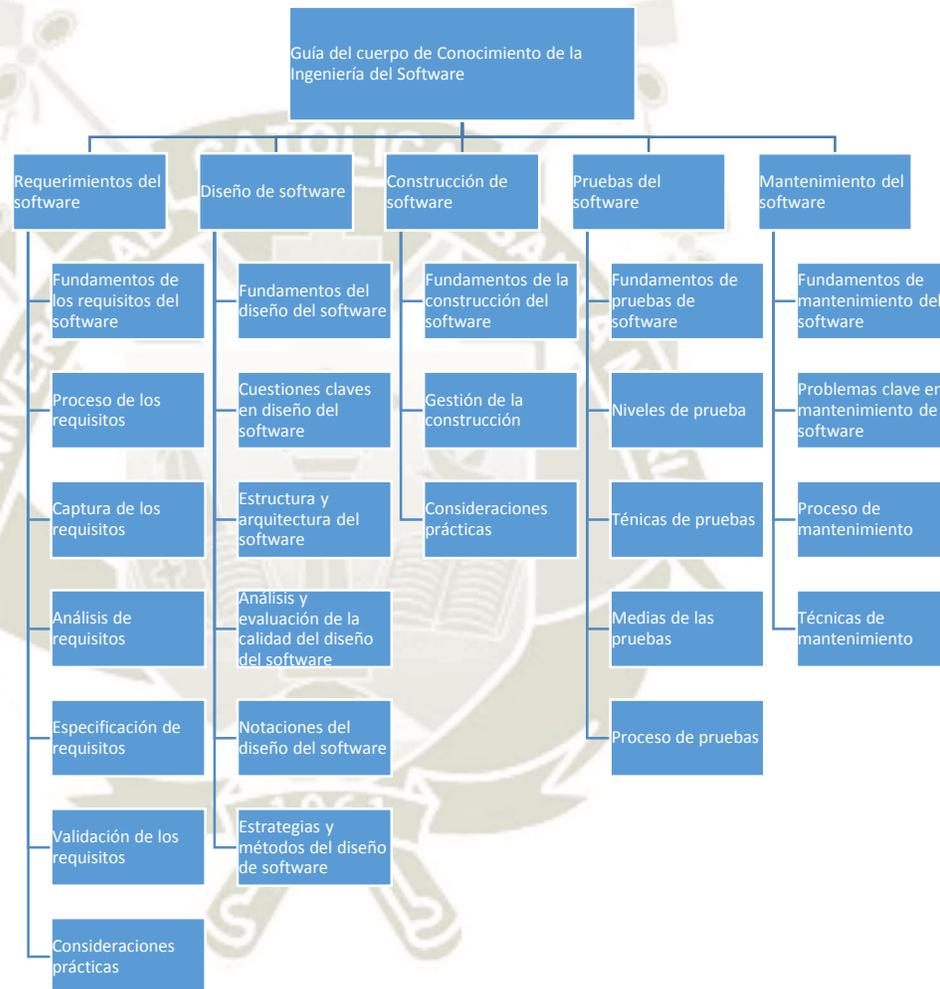


Figura 9. Las 5 primeras áreas de conocimiento. Fuente: (Abran et al., 2004).

La figura 10, describe las 5 primeras áreas de conocimiento siguiendo el tradicional ciclo de vida en cascada. Sin embargo, esto no implica que la guía adopta o fomenta el ciclo de vida en cascada o ningún otro (Abran et al., 2004).

2.12.1 Objetivos SWEBOK

La guía al cuerpo de conocimiento de ingeniería del software se estableció con los siguientes 5 objetivos:

1. Promover una visión consistente de la ingeniería del software en el mundo.
2. Clarificar la situación y definir fronteras de la ingeniería del software con respecto a otras disciplinas como la informática, gestión de proyectos, ingeniería informática y matemáticas.
3. Caracterizar los contenidos de la disciplina de la ingeniería del software.
4. Proporcionar al cuerpo de conocimiento de la ingeniería del software con los temas de interés.
5. Proporcionar una base para el desarrollo planes de estudio, certificaciones individuales y materiales para licencias. (Abran et al., 2004).

2.13 Obesidad

La obesidad es una enfermedad crónica, que corresponde a un exceso de masa grasa cuyas consecuencias para la salud son nefastas. Sus factores son múltiples e intrincados: densidad calórica de la alimentación, sedentarismo, factores psicológicos, genéticos, contexto socioeconómico. Su definición se basa en el cálculo del índice de masa corporal ($IMC = \text{peso}/\text{estatura}^2$: peso en kilogramos y estatura en metros). (Ciangura, Czernichow & Oppert, 2010).

2.13.1 Causas

La causa fundamental del sobrepeso y la obesidad es un desequilibrio energético entre calorías consumidas y gastadas. A nivel mundial ha ocurrido lo siguiente:

- Un aumento en la ingesta de alimentos de alto contenido calórico que son ricos en grasa; y
- Un descenso en la actividad física debido a la naturaleza cada vez más sedentaria de muchas formas de trabajo, los nuevos modos de transporte y la creciente urbanización.

A menudo los cambios en los hábitos alimentarios y de actividad física son consecuencia de cambios ambientales y sociales asociados al desarrollo y de la falta de políticas de apoyo en sectores como la salud; la agricultura; el transporte; la planificación urbana; el medio ambiente; el procesamiento, distribución y comercialización de alimentos, y la educación. (Organización Mundial de la Salud [OMS], 2017)

2.13.2 Consecuencias

Un IMC elevado es un importante factor de riesgo de enfermedades no transmisibles, como las siguientes:

- Las enfermedades cardiovasculares (principalmente las cardiopatías y los accidentes cerebrovasculares), que fueron la principal causa de muertes en 2012;

- Los trastornos del aparato locomotor (en especial la osteoartritis, una enfermedad degenerativa de las articulaciones muy incapacitante), y
- Algunos cánceres (endometrio, mama, ovarios, próstata, hígado, vesícula biliar, riñones y colon).

El riesgo de contraer estas enfermedades no transmisibles crece con el aumento del IMC. La obesidad infantil se asocia con una mayor probabilidad de obesidad, muerte prematura y discapacidad en la edad adulta. Sin embargo, además de estos mayores riesgos futuros, los niños obesos sufren dificultades respiratorias, mayor riesgo de fracturas e hipertensión, y presentan marcadores tempranos de enfermedades cardiovasculares, resistencia a la insulina y efectos psicológicos (Organización Mundial de la Salud [OMS], 2017).

2.13.3 Obesidad en Adolescentes

El número de niños y adolescentes de edades comprendidas entre los 5 y los 17 años que presentan obesidad se ha multiplicado por 10 en el mundo en las cuatro últimas décadas. Las conclusiones de un nuevo estudio dirigido por el Imperial College de Londres y la Organización Mundial de la Salud (OMS) indican que, si se mantienen las tendencias actuales, en 2022 habrá más población infantil y adolescente con obesidad que con insuficiencia ponderal moderada o grave.

Se afirma que, si se mantienen las tendencias observadas desde 2000, los niveles mundiales de obesidad en la población infantil y adolescente superarán en 2022 a los de la insuficiencia ponderal moderada o grave en el mismo grupo de edad. (Organización Mundial de la Salud [OMS], 2017).

CAPITULO 3: ANÁLISIS Y DEFINICION DE REQUERIMIENTOS

En este capítulo se procede a identificar y hacer la definición respectiva de cada uno de los requerimientos que debe cumplir el sistema, esto permite que se pueda llevar un plan del desarrollo del ciclo de vida del software. Para llevar a cabo dicha actividad se va a seguir el enfoque en “cascada”, definido anteriormente, este enfoque sigue básicamente los pasos de Análisis, Diseño, Desarrollo, Prueba e Implantación. Estos pasos serán realizados en una metodología SCRUM, realizando entregas parciales y regulares del producto final.

El proceso de obtención de requerimientos, que tiene como objetivo principal mostrar los requisitos del sistema, no solo es un proceso técnico, también es un proceso que relaciona diferentes personas, y por lo tanto tiene dificultades añadidas a su realización. La parte más difícil en la construcción de un sistema de software, es decidir precisamente qué construir. En esta situación, para tener un enfoque más realístico se debe iniciar realizando un análisis de la situación actual de la problemática que se intenta resolver.

Una vez realizado el análisis, se procede a plantear las soluciones en forma de funciones que realizará el sistema, y así establecer las especificaciones de los requerimientos, además se debe verificar lo realizable que es plasmar los requerimientos identificados, en el proyecto.

3.1 Análisis de la situación actual

En una escuela, los profesores encargados de la educación física de los alumnos, tienen un propósito central, que los adolescentes disfruten de la actividad física, los juegos, la iniciación deportiva y el deporte educativo como una forma de realización

personal. Adicionalmente se busca que los estudiantes asuman conductas de cooperación y se muestren tolerantes y serenos frente a la victoria o la derrota. Sin embargo, la evaluación en cualquier materia tiene un carácter formativo, que se caracteriza por la actividad de recoger información en diferentes momentos, con instrumentos variados para valorar las observaciones según los criterios establecidos.

Actualmente se captura la información de forma manual, es decir la primera base de datos que se obtuvo para este proyecto fue obtenida por medio de encuestas y evaluaciones escritas en documentos. Algo de lo que se pretende cambiar y automatizar, es la recepción de información de tal forma que los estudiantes realicen las encuestas en un dispositivo tecnológico, que bien puede ser una computadora, Tablet o celular y que esta información se almacene en una base de datos para ser procesada. Pero en un inicio se utilizó una base de datos manual y no automatizada, para realizar la evaluación de técnicas y modelos.

3.2 Definición de requerimientos del sistema

Para la definición de requerimientos del sistema, es necesario trabajar continuamente con los usuarios, en esta situación pueden ser los profesores y los alumnos de las escuelas. Realizando reuniones en las que se recopilará la información de los requerimientos que debe cumplir el sistema. Una vez concluida dicha actividad, se debe analizar la información obtenida dando una definición a cada uno de los requerimientos y prioridades, estas se añaden a un backlog, que es la lista de actividades a realizar, cada requerimiento se traduce en una función que debe realizar el sistema.

3.2.1 Adquisición de requerimientos

En la adquisición de requerimientos, es necesario recolectar la mayor cantidad de información verdadera posible, para lo cual es necesario desarrollar las siguientes actividades:

- **Identificación de usuarios especialistas (UE):** Se identificaron dos UE:
 - Dr. Marco Cossio – Universidad Católica del Maule - Chile
 - Ing. José Sulla Torres – Universidad Católica de Santa María
- **Primeras reuniones y evaluación de la viabilidad:** Estas reuniones son elaboradas junto a los usuarios UE, teniendo como propósito central determinar los requerimientos funcionales del sistema.
- **Extracción de información:** En este paso se repasó la información existente sobre el problema que se intenta resolver.
- **Educación de información a partir de los UE:** Se hicieron preguntas a los UE, con el propósito de desarrollar el sistema en base a sus conocimientos. En una parte inicial, se hicieron preguntas básicas y generales. Y ya luego se hicieron preguntas más detalladas, con ejemplos incluidos.

3.2.2 Identificación de actores

El Sistema se encuentra desarrollado principalmente para los usuarios que respondan las preguntas que se muestren y que realicen las evaluaciones. Sin embargo, debe haber dos tipos de **Usuarios**, uno con identificación y contraseñas únicas para el acceso al sistema **1) Usuario administrador**, y otro usuario que no posea identificación ni contraseña, es decir que sólo necesite ingresar un nombre **2)**

Usuario final. Este último usuario podrá hacer uso del sistema, sólo de ciertos reportes.

Tabla 2

Privilegios de autores. Fuente: Propia

	<i>Usuario administrador</i>	<i>Usuario final</i>
Ver historial de evaluaciones ingresadas	X	
Acceder la prueba de IMC	X	X
Acceder a cuestionario de obesidad	X	X
Reportes individuales	X	X
Reportes Generales	X	

3.3 Catalogación de requerimientos

Se catalogó lo requisitos funcionales y no funcionales que debe satisfacer el sistema, de acuerdo a la información recolectada, y a las sesiones que se tuvo con los UE.

3.3.1 Requerimientos funcionales

Los requerimientos funcionales, son las principales actividades que debe realizar el sistema, en otras palabras, es el comportamiento o función particular del sistema. En la siguiente figura se muestran los requerimientos funcionales del sistema.

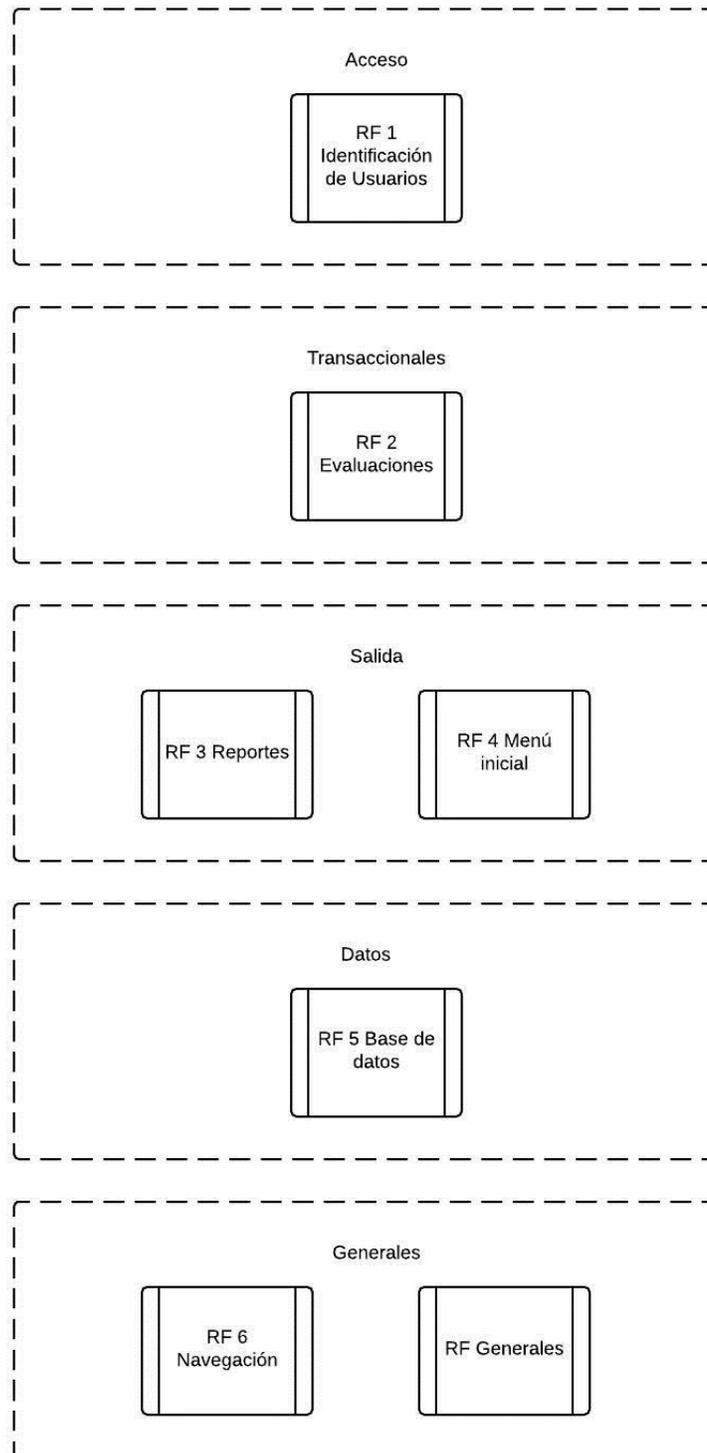


Figura 10. Esquema de los Requerimientos Funcionales.

RF1. Identificación de Usuario

RF1.1 El usuario que desee ingresar al sistema y acceder a las funcionalidades del sistema deberá escribir su nombre completo y presionar el botón continuar.

RF1.2 El usuario administrador debe ingresar sus credenciales, es decir un nombre de usuario y una contraseña.

RF1.3 Debe haber un link que permita al usuario cambiar de formularios, uno para el ingreso de usuario final y otro para el usuario administrador.

RF2. Evaluaciones

RF2.1 El usuario podrá seleccionar en el menú de opciones, la evaluación que desee rendir.

RF2.2 El usuario final podrá ingresar a la evaluación de imc, y luego responder todas las preguntas correspondientes, el sistema no permite terminar la evaluación, si no se respondieron todas las preguntas.

RF2.3 El usuario final podrá ingresar al cuestionario de actitud hacia la obesidad, y luego responder todas las preguntas correspondientes, el sistema no permite terminar la evaluación, si no se respondieron todas las preguntas.

RF3. Reportes

RF3.1 El sistema deberá permitir mostrar reportes.

RF3.2 El usuario podrá imprimir el reporte individual donde se deben mostrar los resultados de la evaluación imc y del cuestionario.

RF3.3 El sistema deberá permitir exportar cada reporte a Pdf.

RF3.4 El reporte general de cuestionario y evaluación imc, debe poder ser accedido solamente por usuarios administradores.

RF4. Consultas

RF4.1 El sistema permite acceder a una consulta rápida e individual por cada usuario.

RF4.2 El usuario podrá acceder a la consulta evaluación IMC, donde podrá ver todos los datos ingresados, y el resultado generado por el sistema.

RF4.3 El usuario podrá acceder a la consulta cuestionario, donde podrá ver todos los cuestionarios y datos ingresados, con el respectivo resultado, generado por el sistema

RF4. Menú inicial

RF4.1 El sistema mostrará al usuario identificado, todas las funciones que ofrece y a las que se permite acceder.

RF4.2 El sistema también deberá tener una opción de ayuda, en donde se permita descargar un manual de usuario.

RF5. Base de datos

RF5.1 El sistema debe poder almacenar cada una de las evaluaciones y/o cuestionarios de los usuarios autenticados.

RF6. Navegación

RF6.1 El sistema debe permitir al usuario desplazarse de un módulo a otro sin ningún inconveniente.

RF6.2 El sistema deberá tener una guía rápida en una de sus opciones, sobre cómo el usuario debe utilizar el sistema.

RF7. Generales

RF7.1 El sistema debe ser flexible, permitiendo a futuro modificar opciones, como las evaluaciones, reportes o privilegios de usuarios.

3.3.2 Requerimientos no funcionales

Son requisitos que imponen restricciones en el diseño o la implementación como restricciones en el diseño o estándares de calidad. Son propiedades o cualidades que el producto debe tener.

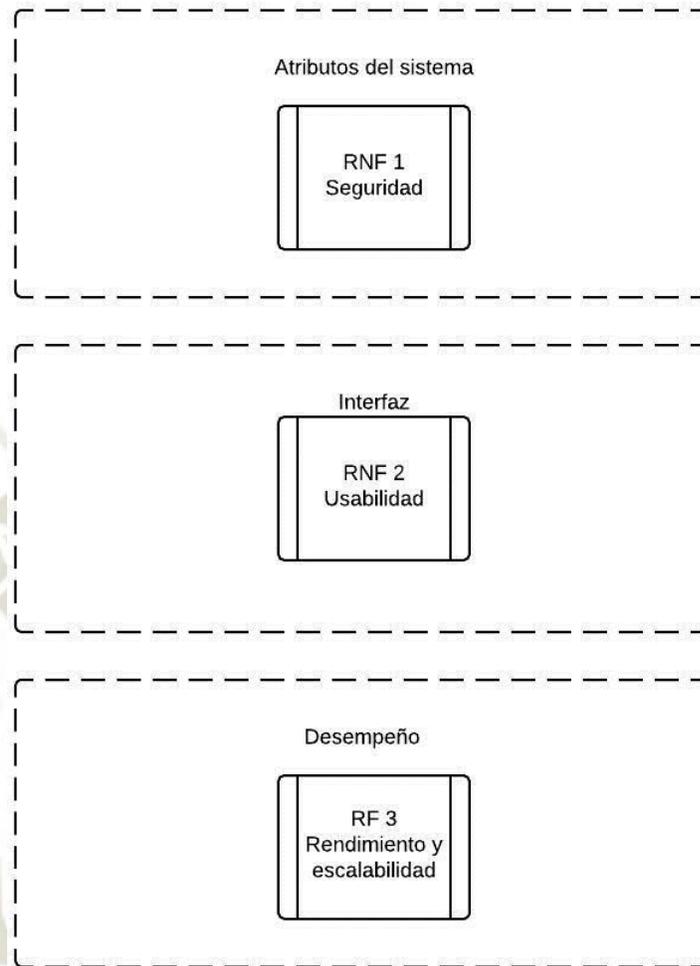


Figura 11. Esquema de los Requerimientos No Funcionales.

RNF1. Seguridad

RNF1.1 El usuario administrador debe tener acceso a una lista de opciones que un usuario final no tiene acceso.

RNF1.2 Los usuarios finales accederán a todas las evaluaciones y reportes individuales, excepto a los reportes generales.

RNF2. Usabilidad

RNF2.1 El usuario debe poder interactuar con el sistema de la forma más fácil, cómoda e intuitiva posible. Y la mejor forma de lograrlo es centrándonos en la simplicidad, y así diseñar una interfaz clara y sencilla.

RNF2.2 El sistema debe incluir en la lista de menú de opciones, una sección de ayuda al usuario, en donde se brinda una guía rápida de uso del sistema.

RNF3. Rendimiento y escalabilidad

RNF3.1 Las consultas y respuestas del sistema al servidor, no deben tomar demasiado tiempo.

RNF3.2 Debe tener la capacidad de manejar una creciente carga de trabajo, por ejemplo, un mayor número de conexiones de usuario.

3.4 Análisis de factibilidad para implementar la solución

Se considera que el sistema es factible técnicamente, puesto que ya existen sistemas que hacen uso de las tecnologías que se necesitan utilizar en el proyecto actual. El sistema involucra la implementación del algoritmo generado por WEKA con el lenguaje de programación PHP. Además, se debe utilizar un gestor de base de datos SQL para el almacenamiento de las evaluaciones de los usuarios.

3.5 Análisis de Casos de Usos

3.5.1 Diagrama de casos de uso

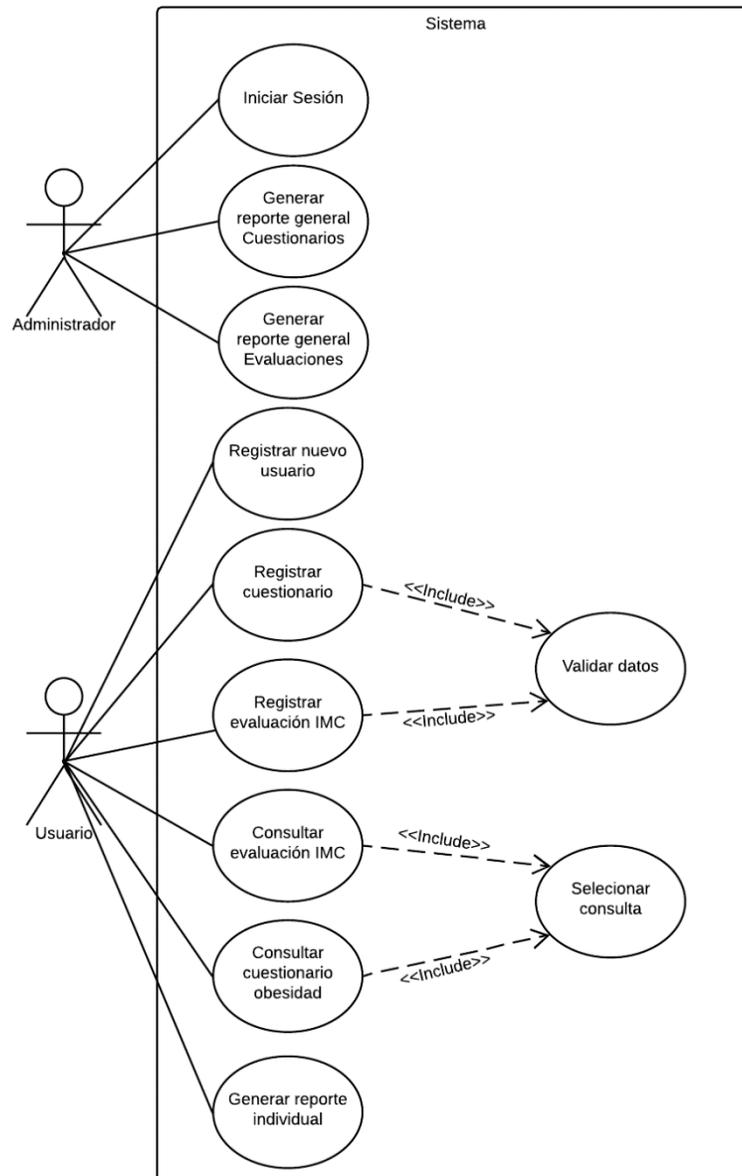


Figura 12. Diagrama de casos de uso.

3.5.2 Elaboración del Caso de Uso Iniciar Sesión

Descripción: El sistema permite al usuario administrador, iniciar sesión y acceder a todos los módulos del sistema, incluyendo evaluaciones, reportes individuales y generales.

Actores: Usuario administrador.

Propósito: Identificar y autenticar el usuario administrador, anteriormente registrado en la base de datos.

Precondiciones: El usuario debe tener conocimiento de una cuenta administrador.

Flujo de Eventos Principal:

1. Usuario ingresa a la aplicación
2. Usuario ingresa sus credenciales
3. Usuario presiona botón ingresar
4. El sistema valida las credenciales y permite el acceso

Flujo alternativo:

1. Usuario ingresa a la aplicación
2. Usuario ingresa sus credenciales
3. Sistema valida credenciales
4. Sistema notifica que las credenciales son incorrectas y pide ingresarlas nuevamente

Excepciones: Si los datos ingresados son incorrectos, se envía un mensaje al usuario, informando sobre el hecho, y se pide que intente nuevamente.

Post condiciones: Ingreso al menú principal del sistema.

3.5.3 Elaboración del Caso de Uso Registrar nuevo Usuario

Descripción: El sistema registra un nuevo usuario, permitiendo el acceso, sólo a los módulos de aplicaciones, consultar, reportes individuales y sugerencias.

Actores: Usuario final.

Propósito: Registrarse e ingresar a la aplicación web, para acceder a sus funciones.

Precondiciones: El usuario debe ingresar un nombre y presionar continuar.

Flujo de Eventos Principal:

1. Usuario ingresa la aplicación
2. Usuario ingresa nombre de usuario
3. Usuario presiona botón continuar
4. Sistema registra al usuario
5. Sistema permite el acceso

Flujo alternativo:

1. Usuario ingresa la aplicación
2. Usuario no ingresa un nombre
3. Usuario presiona botón continuar
4. Sistema notifica que debe ingresar obligatoriamente un nombre de usuario

Excepciones: La ocurrencia de un problema de red, y que el sistema no logre registrar correctamente al usuario.

Post condiciones: Ingreso al menú principal del sistema, sólo a los módulos de usuarios permitidos. Registro del usuario.

3.5.4 Elaboración del Caso de Uso Registrar Cuestionario Obesidad

Descripción: El sistema obtiene todas las preguntas pertenecientes al cuestionario obesidad, para luego registrar las respuestas del usuario. Finalmente muestra los resultados, de forma gráfica y textual, de acuerdo a los datos registrados.

Actores: Usuario final y Usuario administrador.

Propósito: Responder las preguntas del cuestionario obesidad, para posteriormente obtener un resultado.

Precondiciones: Ingresar correctamente al sistema.

Flujo de Eventos Principal:

1. Usuario elige Realizar cuestionario obesidad
2. Usuario ingresa datos personales y responde todas las preguntas
3. Usuario presionar guardar
4. El sistema valida los datos
5. El sistema muestra los resultados del cuestionario

Flujo alternativo:

1. Usuario elige Realizar cuestionario obesidad
2. Usuario ingresa datos personales
3. Usuario no completa todas las preguntas
4. Usuario presiona guardar

5. Sistema notifica que debe completar todas las preguntas

Excepciones: El sistema debe informar que aún faltan preguntas por responder.

Post condiciones: Cuestionario finalizado, se registra en la base de datos, se muestra inmediatamente el resultado del cuestionario, con un gráfico representativo, indicando cuál es la posición del usuario respecto a las tablas de medición.

3.5.5 Elaboración del Caso de Uso Registrar Evaluación de IMC

Descripción: El sistema muestra el formulario de evaluación IMC, luego permite registrar las respuestas del usuario. Finalmente muestra los resultados, de forma gráfica y textual, de acuerdo a los datos registrados.

Actores: Usuario final y administrador.

Propósito: Responder las preguntas de la evaluación, para así obtener un resultado, respecto al estado físico del usuario.

Precondiciones: Ingreso correcto al sistema.

- Flujo de Eventos Principal:**
1. Usuario elige Evaluación IMC
 2. Usuario ingresa los datos antropométricos
 3. Usuario responde las preguntas
 4. Usuario presionar guardar
 5. El sistema muestra los resultados de la evaluación

Flujo alternativo:

1. Usuario elige Evaluación IMC
2. Usuario ingresa sus datos antropométricos
3. Usuario no completa todas las preguntas
4. Usuario presionar guardar
5. Sistema notifica que se debe completar todas las preguntas

Excepciones: El sistema debe informar que faltan datos por ingresar.

Post condiciones: Evaluación finalizada, se registra en la base de datos, se muestra el resultado inmediatamente, y se muestra un gráfico representativo, del resultado obtenido.

3.5.6 Elaboración del Caso de Uso Consultar Evaluación de IMC

Descripción: El sistema obtiene todas las evaluaciones registradas por el usuario, luego permite ver los datos ingresados en la evaluación IMC.

Actores: Usuario final y administrador.

Propósito: Visualizar los datos y resultados ingresados en la evaluación IMC.

Precondiciones: Realizar al menos una evaluación de IMC.

Flujo de Eventos Principal:

1. Usuario elige Consultar Evaluación IMC
2. El sistema muestra los datos y resultados

Flujo alternativo:

1. Usuario elige Consultar Evaluación IMC

2. Sistema informa que no se registraron evaluaciones IMC

Excepciones: El sistema debe informar que el usuario no tiene registrado evaluaciones.

Post condiciones: Se muestra detalle de la evaluación realizada, con el resultado emitido por el sistema.

3.5.7 Elaboración del Caso de Uso Consultar Cuestionario Obesidad

Descripción: El sistema obtiene los datos registrados en el cuestionario obesidad por el usuario. Luego muestra una vista con los respectivos resultados obtenidos.

Actores: Usuario final y administrador.

Propósito: Visualizar los datos y resultados ingresados en el cuestionario obesidad.

Precondiciones: Haber realizado al menos un cuestionario.

Flujo de Eventos Principal:

1. Usuario elige Consultar cuestionario obesidad

2. El sistema muestra los datos y resultados

Flujo alternativo:

1. Usuario elige Consultar cuestionario obesidad

2. Sistema informa que no se registraron cuestionarios por el usuario

Flujo alterno: No haber realizado algún cuestionario de actitud hacia la obesidad.

Excepciones: El sistema debe informar que el usuario no tiene registrado cuestionarios.

Post condiciones: Se muestra detalle del cuestionario realizado, con el resultado emitido por el sistema.

3.5.8 Elaboración del Caso de Uso Generar Reporte individual

Descripción: El sistema obtiene todos los registros de evaluaciones y cuestionarios del usuario, para generar un reporte en formato pdf.

Actores: Usuario final y administrador.

Propósito: Generar un reporte individual de todas las evaluaciones y/o cuestionarios resueltos por el usuario.

Precondiciones: El usuario debe haber realizado alguna evaluación y/o cuestionario en el sistema.

Flujo de Eventos Principal:

1. Usuario elige Generar reporte individual
2. El sistema genera reporte individual, con las evaluaciones y/o cuestionarios realizados

Flujo alternativo:

1. Usuario elige Generar reporte individual

2. El sistema informa que el usuario no tiene evaluaciones o cuestionarios realizados

Excepciones: Se muestra el reporte en blanco, informando que aún no se han rendido evaluaciones.

Post condiciones: Reporte generado en formato pdf.

3.5.9 Elaboración del Caso de Uso Generar Reporte general Evaluaciones

Descripción: El sistema obtiene todos los registros de evaluaciones realizadas por todos los usuarios hasta la fecha, para luego generar reporte en formato Excel y descargarlo en el computador.

Actores: Usuario administrador.

Propósito: Generar un reporte general de todas las evaluaciones de todos los usuarios.

Precondiciones: El usuario debe haber ingresado correctamente al sistema en modo administrador.

- Flujo de Eventos Principal:**
1. Usuario elige Exportar evaluaciones
 2. Sistema genera reporte
 3. Sistema descarga el reporte en formato Excel

- Flujo alternativo:**
1. Usuario elige Exportar evaluaciones
 2. El sistema genera reporte

3. El sistema intenta descargar reporte
4. El navegador no permite las descargas de ciertos archivos

Excepciones: Se muestra un mensaje emitido por el navegador, para este caso es recomendable acceder a las configuraciones del navegador mismo y habilitar las descargas de archivos.

Post condiciones: Reporte generado en formato xls.

3.5.10 Elaboración del Caso de Uso Generar Reporte general Cuestionarios

Descripción: El sistema obtiene todos los registros de cuestionarios realizados por todos los usuarios hasta la fecha, para luego generar reporte en formato Excel y descargarlo en el computador.

Actores: Usuario administrador.

Propósito: Generar un reporte general de todos los cuestionarios del total de usuarios registrados hasta la fecha.

Precondiciones: El usuario debe haber ingresado correctamente al sistema en modo administrador.

- Flujo de Eventos Principal:**
1. Usuario elige Exportar cuestionarios
 2. Sistema genera reporte
 3. El sistema descarga el reporte en formato Excel

Flujo alternativo:

1. Usuario elige Exportar cuestionarios
2. El sistema genera reporte
3. El sistema intenta descargar reporte
4. El navegador no permite las descargas de ciertos archivos

Excepciones: Se muestra un mensaje emitido por el navegador, informando que no es posible descargar determinados archivos.

Post condiciones: Reporte generado en formato xls.

3.6 Análisis de diagramas de secuencia

3.6.1 Elaboración del Diagrama de secuencia Iniciar Sesión

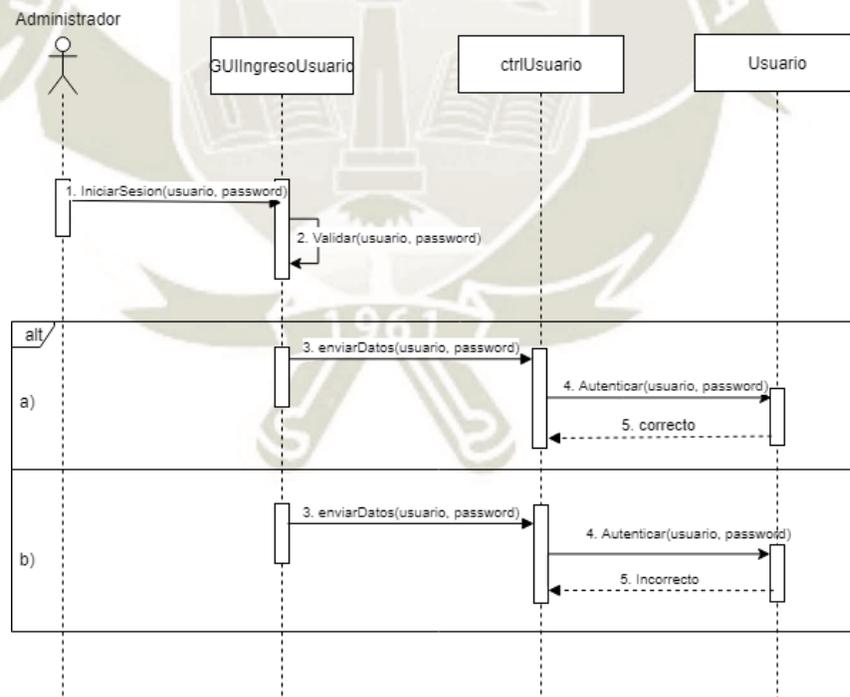


Figura 13. Diagrama de secuencia Iniciar sesión.

3.6.2 Elaboración del Diagrama de secuencia Registrar nuevo Usuario

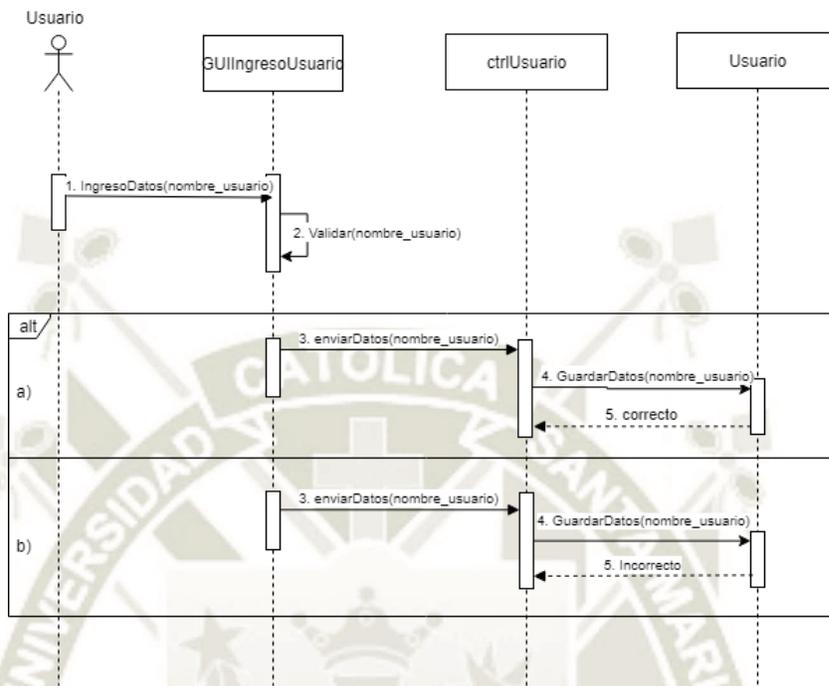


Figura 14. Diagrama de secuencia Registrar nuevo usuario.

3.6.3 Elaboración del Diagrama de secuencia Registrar Cuestionario Obesidad

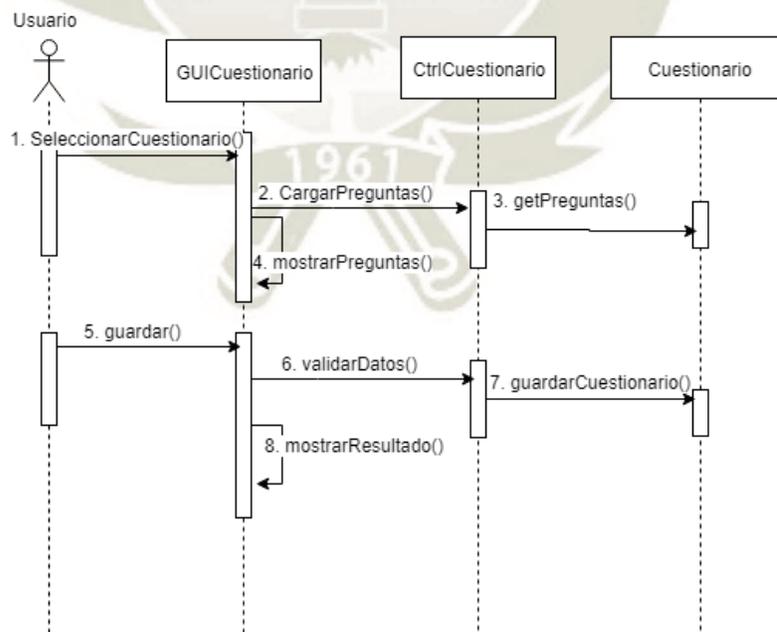


Figura 15. Diagrama de secuencia Registrar cuestionario obesidad.

3.6.4 Elaboración del Diagrama de secuencia Registrar Evaluación de IMC

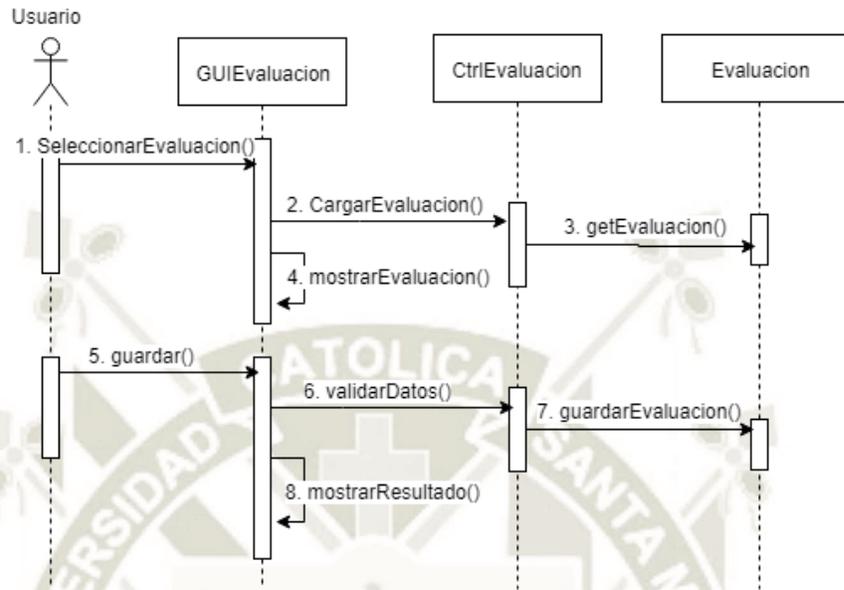


Figura 16. Diagrama de secuencia Registrar Evaluación IMC.

3.6.5 Elaboración del Diagrama de secuencia Consultar Evaluación de IMC

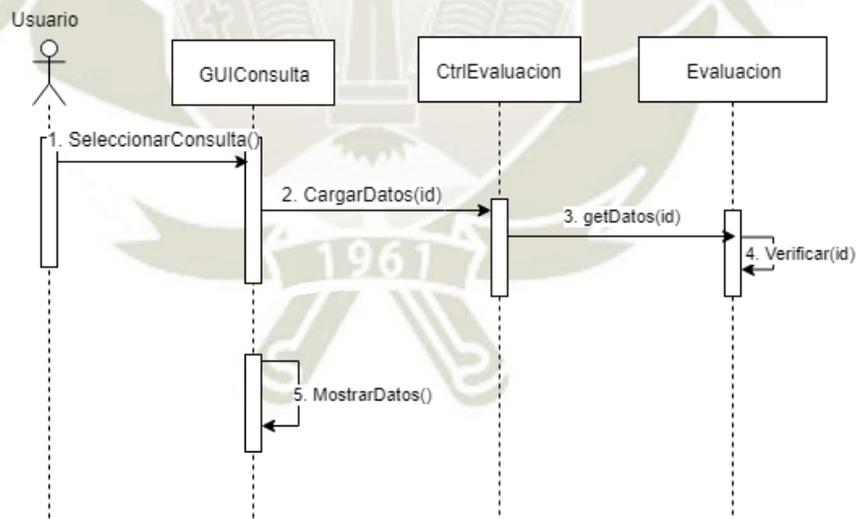


Figura 17. Diagrama de secuencia Consultar evaluación de IMC.

3.6.6 Elaboración del Diagrama de secuencia Consultar Cuestionario Obesidad

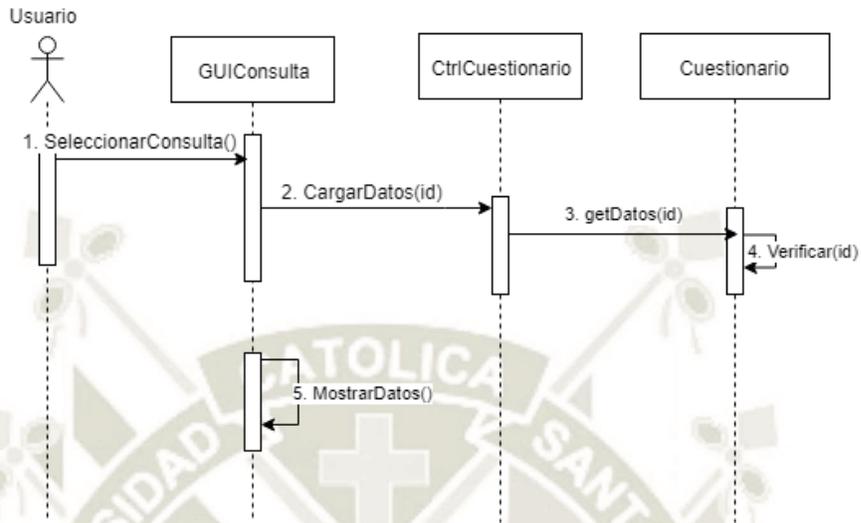


Figura 18. Diagrama de secuencia Consultar cuestionario obesidad.

3.6.7 Elaboración del Diagrama de secuencia Generar reporte individual

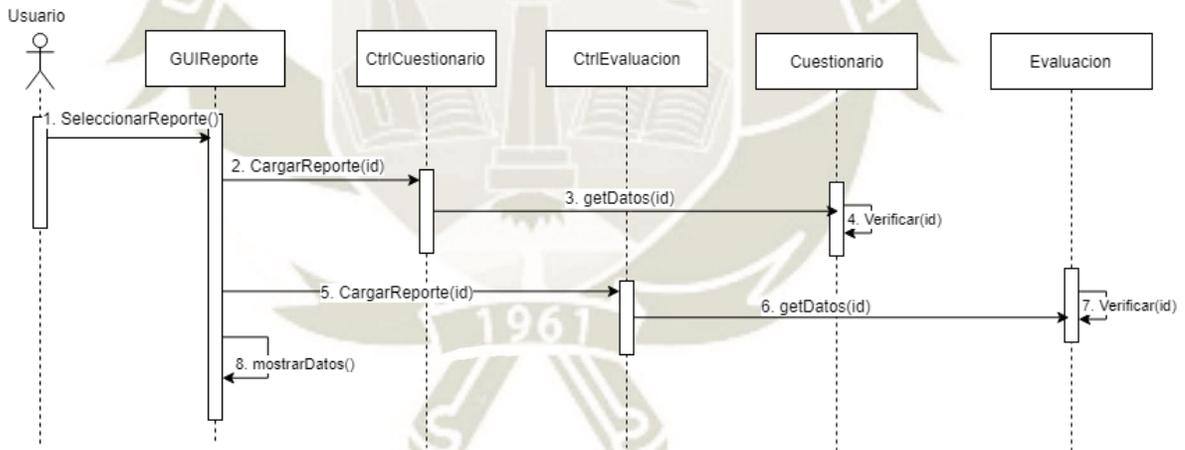


Figura 19. Diagrama de secuencia Generar reporte individual.

3.6.8 Elaboración del Diagrama de secuencia Generar reporte general Evaluaciones

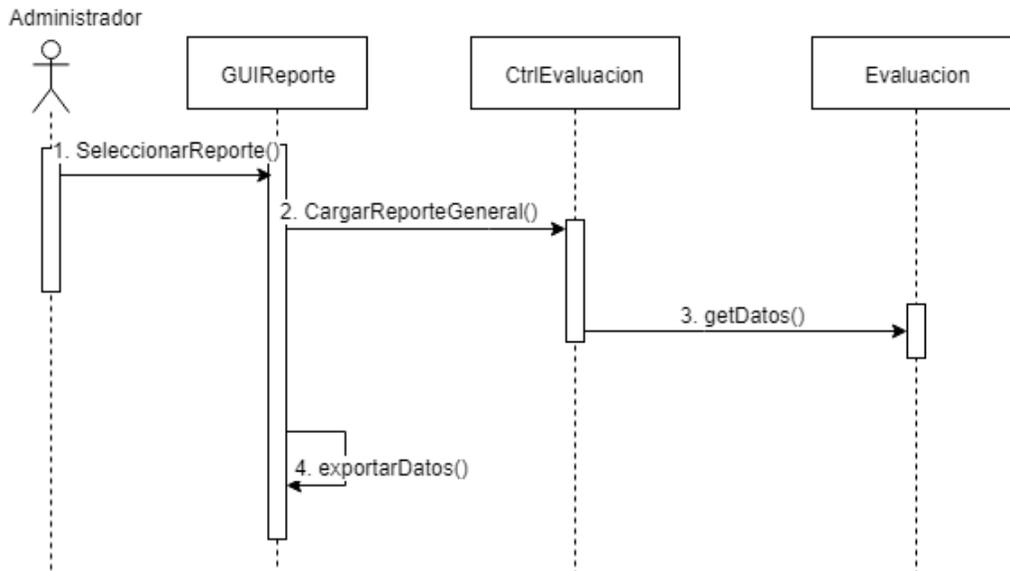


Figura 20. Diagrama de secuencia Generar reporte general Evaluaciones.

3.6.9 Elaboración del Diagrama de secuencia Generar reporte general Cuestionarios

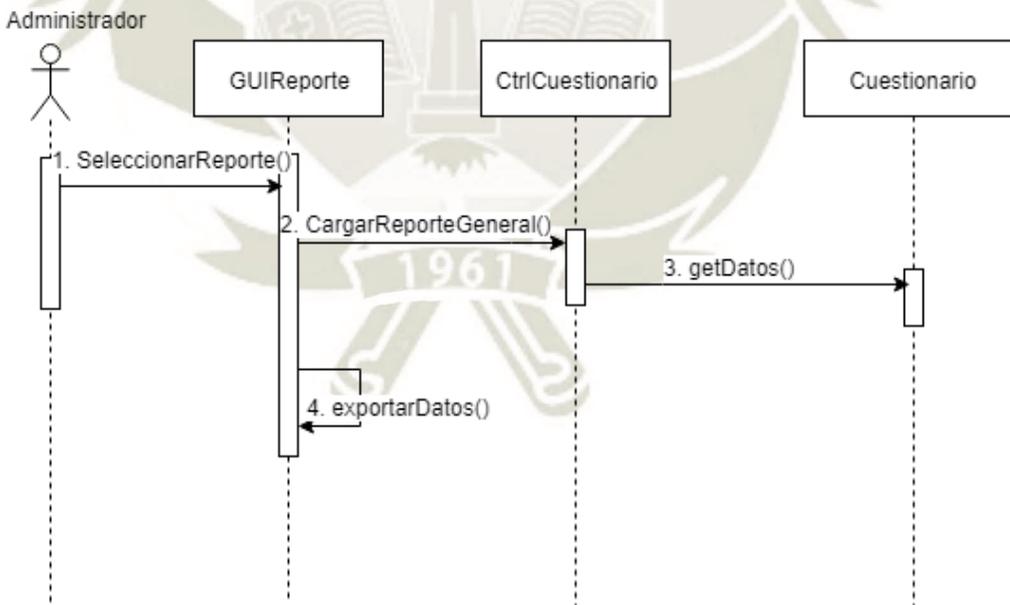


Figura 21. Diagrama de secuencia Generar reporte general Cuestionarios.

3.6.10 Diagrama de clases

En el siguiente diagrama de clases se muestra la estructura del sistema en concreto, modelando las clases, atributos, operaciones y relaciones entre objetos.

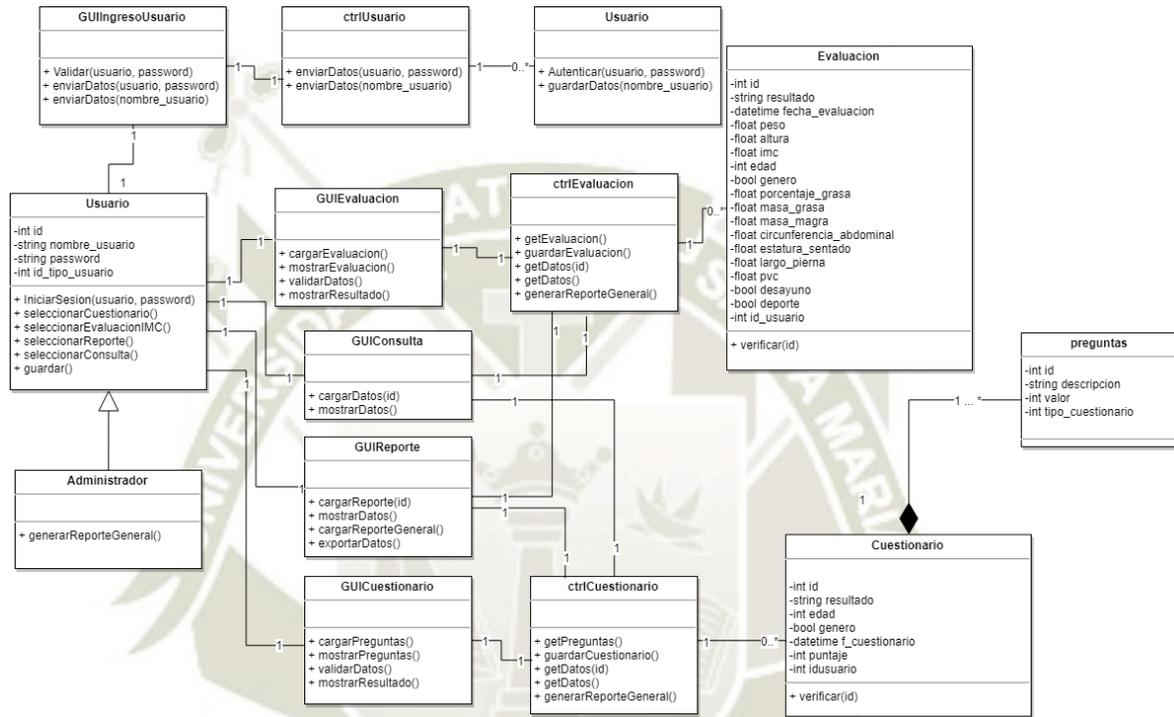


Figura 22. Diagrama de clases - SisPrediccion.

CAPITULO 4: CONSTRUCCION Y EVALUACION DE MODELOS DE CLASIFICACION

En este capítulo se desarrolla el proceso de minería de datos del sistema, empezando con la construcción, la comparación y finalmente seleccionar cuál es el modelo de clasificación que se ajusta de mejor forma a los requerimientos.

1.1. Análisis previo de base de datos

Uno de los pasos más importantes antes de empezar la construcción de modelos de clasificación, es realizar un análisis detallado de la fuente de datos a utilizar, cada técnica trabaja con una fuente de datos en particular, actualmente hay dos tipos, datos nominales categóricos y numéricos que pueden ser transformados si es que fuese necesario, para que se adapten a la técnica que será utilizada.

En total, la base de datos constó de 680 registros de estudiantes, recolectados de diversas escuelas de la ciudad de Arequipa, Perú. El rango de edad de los estudiantes es de 5 a 17 años. Los parámetros en total son los siguientes:

Tabla 3

Total, de atributos del conjunto de datos. Fuente: Propia

Número de categoría	Parámetros
1	Nombre
2	Fecha de nacimiento
3	Género
4	¿Hace Deporte?
5	¿Fuma?
6	¿Desayuna todos los días?
7	Estatura (cm)
8	Peso
9	Estatura parado (m)
10	Estatura sentado
11	Circunferencia abdominal (cm)
12	Codo (cm)
13	Muñeca (cm)
14	Rodilla (cm)
15	Tobillo (cm)
16	Antebrazo derecho (cm)
17	Pierna (cm)

De la muestra total recolectada se obtuvieron un total de 680 registros (374 mujeres y 306 varones).

Tabla 4

Características antropométricas del conjunto de datos. Fuente: Propia

Variables	Varones (n = 306)		Mujeres (n = 374)	
	X	DE	X	DE
Edad	14.63	2.79	14.68	2.16
Estatura (cm)	160.8	32.08	166.66	238.99
Peso	58.55	13.05	54.26	9.55
Estatura parado (m)	1.65	0.19	1.68	2.39
Estatura sentado (cm)	84.61	5.85	82.96	3.70
Circunferencia abdominal	77.38	9.97	74.07	8.65
Codo	5.84	0.75	5.71	3.70
Muñeca	5.01	0.80	4.90	3.25
Rodilla	8.95	1.06	8.64	4.76
Tobillo	6.25	0.78	6.00	2.15
Antebrazo derecho	25.41	3.78	24.24	1.84
Pierna	47.86	23.74	44.38	3.72

Donde X: promedio; DE: desviación estándar.

1.2. Construcción de los modelos de clasificación

Para la construcción de los modelos de clasificación se debe tomar en cuenta 3 técnicas principales: Redes neuronales, Árboles de decisión y Redes Bayesianas, ya que según la bibliografía son las técnicas más utilizadas para este tipo de caso y ofrecen mayor rendimiento para conjuntos de datos de mediano tamaño en comparación con otras técnicas.

Luego de realizar una limpieza de datos, la fuente de datos resultó con un número de registros de 660, de personas entre un rango de edad de 5 a 17 años, estos registros fueron proporcionados por instituciones educativas del Perú.

Para la preparación de datos, se deben almacenar todos los registros en un archivo en formato csv en Weka; del total de parámetros mostrados, los seleccionados son los siguientes: Los campos de tipo nominal son: Género, ¿Hace deporte?, ¿Fuma?, ¿Desayuna todos los días? y OUT_OBESITY; y los campos numéricos son: Edad, Peso, Estatura parado, Estatura sentado, Circunferencia abdominal, PVC, Porcentaje de grasa, Masa grasa y Masa magra.

Los parámetros seleccionados para realizar la predicción de información se muestran en la siguiente tabla.

Tabla 5

Parámetros de predicción. Fuente: Propia

Número de categoría	Parámetros
1	Edad
2	Sexo
3	¿Hace Deporte?
4	¿Fuma?
5	¿Desayuna todos los días?
6	Peso
7	Estatura parado
8	Estatura sentado
9	Circunferencia abdominal
10	Pico de Velocidad de Crecimiento (PVC)
11	Porcentaje grasa (%)
12	Masa grasa
13	Masa libre de grasa (Masa magra)

1.2.1. Métodos y herramientas

Antes de trabajar o seleccionar la herramienta o técnica de aprendizaje, se debe enfocar en la fuente de datos, tanto datos nominales categóricos como numéricos pueden ser transformados si es que es necesario, para que así sean compatibles con la técnica escogida. Con la herramienta Weka es posible convertir y aplicar filtros que pueden transformar nuestra data, de ser el caso.

Según Cruz & Wishart, (2006), se hizo una evaluación, dirigida a la detección de cáncer, luego de un adecuado análisis de las técnicas de aprendizaje aplicadas, realizando las comparaciones necesarias, se obtuvo que Redes neuronales fue la técnica más apropiada para ese caso. En nuestro caso (Obesidad en la adolescencia), también existen artículos que comparan métodos de máquinas de aprendizaje, es así que para nuestro caso cada técnica evaluada tendrá que ser la mejor versión posible, que se irá obteniendo de acuerdo a las pruebas.

Para realizar la comparación de las 3 técnicas, se basó en las métricas que pueden comparar el desempeño de métodos de clasificación de acuerdo a (Witten, Ian & Frank, 2005). Entre las 3 técnicas se vio la precisión como parte de que datos son bien clasificados, otros como sensibilidad, especificidad y la curva ROC que nos ayudan a ver cuál es la mejor técnica para nuestro caso.

4.2. Algoritmos de clasificación

4.2.1. Clasificador J48 o C4.5

J48 es un algoritmo de inducción que genera una estructura de reglas o árbol a partir de subconjuntos de casos extraídos de conjunto total de datos de “entrenamiento”. En ese contexto, la forma en que procesa los datos es similar al algoritmo ID3. El algoritmo genera una estructura de reglas y evalúa su “peso”, usando criterios que miden la precisión en la clasificación de casos, como resultado obtiene un árbol binario.

El enfoque de este árbol es más útil en problemas de clasificación, con esta técnica se construye un árbol para modelar el proceso de clasificación. Al construir un árbol J48, se ignoran los valores nulos o faltantes, es decir el valor de ese elemento, se puede predecir en función de lo que se sabe sobre los valores del atributo para los otros registros. La idea básica es dividir los datos en un rango basado en los valores de atributo para ese elemento que se encuentran en la muestra de entrenamiento. J48 permite la clasificación a través de árboles de decisión o reglas generadas a partir de ellos. (Patil & Sherekar, 2013).

4.2.2. Clasificador BayesNet

Las redes bayesianas modelan un conjunto de nodos y relaciones de dependencias entre estos nodos. De acuerdo a ello, es posible obtener inferencia bayesiana, con el fin de estimar la probabilidad posterior de las variables no conocidas, según las variables conocidas.

El clasificador BayesNet es considerado como un clasificador adecuado en el área de la medicina, además posee alta precisión cuando se aplica a grandes cantidades de datos, y es de gran utilidad en diagnósticos médicos (Adnan & Husain, 2012).

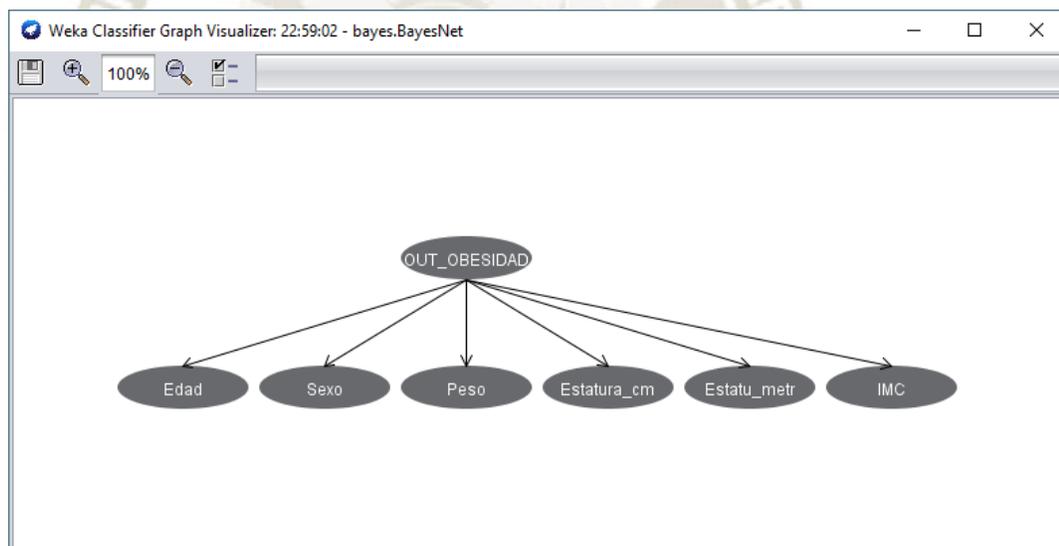


Figura 23. Grafo del clasificador Naive Bayes obtenido de Weka.

4.2.3. Clasificador Multilayer Perceptron (MLP)

Un Perceptrón Multicapa, es una red neuronal profunda y artificial, que se encuentra compuesta por más de un perceptrón, por una capa de entrada para recibir

uno o varios datos, una capa de salida que toma una decisión o predicción sobre la entrada, y entre estos dos, un número arbitrario de capas ocultas que son el verdadero cálculo del MLP.

MLP es a menudo aplicado a problemas de aprendizaje supervisado ellos entrenan sobre un conjunto de pares de entrada - salida, y aprenden a modelar la correlación entre esas entradas y salidas. El entrenamiento involucra ajustar los parámetros o pesos del modelo con el fin de minimizar el error. (Hassanien, Moftah, Azar, & Shoman, 2014).

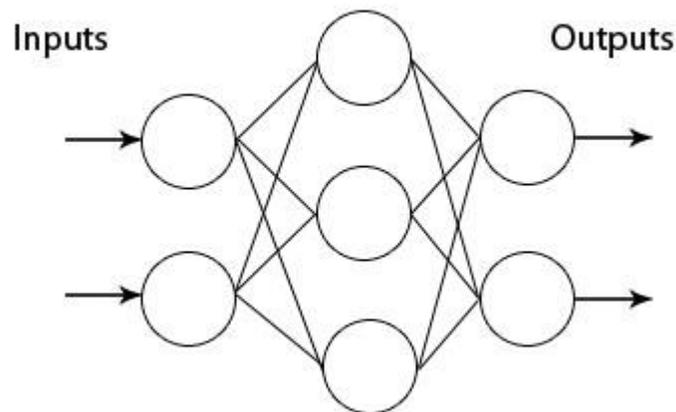


Figura 24. Grafo de Perceptrón Multicapa.

4.2.4. Clasificador ForestPA (Forest By Penalizing Attributes)

Forest by Penalizing Attributes (Forest PA) impone sanciones a los atributos de forma sistemática, de modo que un atributo probado en un nivel inferior (como en el nodo raíz) recibe una penalización más alta (menos peso) que un atributo comprobado en un nivel superior. La razón es que un atributo probado en un nivel inferior puede influir en más reglas que un atributo comprobado en un nivel superior.

Por lo tanto, para descubrir un conjunto diverso de reglas, los atributos probados en niveles inferiores deben evitarse en un árbol futuro más que aquellos atributos que se prueban en niveles más altos. Además, para aumentar la posibilidad de tener diferentes pesos entre los atributos en el mismo nivel, Forest PA selecciona aleatoriamente el peso de un atributo del rango de peso asignado para el nivel del atributo (Adnan & Islam, 2017).

4.2.5. Clasificador Naive Bayes

Es una técnica de clasificación basada en el teorema de Bayes con un supuesto de independencia entre los predictores. En términos simples, un clasificador de Naive Bayes supone que la presencia de una característica particular en una clase no está relacionada con la presencia de ninguna otra característica. Por ejemplo, una fruta puede considerarse una manzana si es roja, redonda y de aproximadamente 3 pulgadas de diámetro. Incluso si estas características dependen unas de otras o de la existencia de otras características, todas estas propiedades contribuyen de forma independiente a la probabilidad de que esta fruta sea una manzana y es por eso que se la conoce como "Naive".

El modelo Naive Bayes es fácil de construir y muy útil para grandes conjuntos de datos. Junto con la simplicidad, se sabe que Naive Bayes supera incluso a los métodos de clasificación altamente sofisticados. (Dulhare, 2018)

4.3. Pruebas en la herramienta Weka

Se vio por necesario utilizar Weka para evaluar la performance de los métodos de clasificación, también será posible utilizar la herramienta para predecir nuevos valores en cuanto a la obesidad infantil y obesidad en adolescencia.

Weka tiene un conjunto de algoritmos disponibles para diversos tipos de análisis de datos y modelado predictivo, además posee herramientas para visualizar datos, provee una interfaz gráfica que centraliza todo el proceso de pruebas a una sola herramienta (Suca et al., 2016).

En primera instancia se debe convertir la data a un formato aceptado por la herramienta, en este caso se convirtió a CSV (comma separated value). Se procede a abrir la data desde la interfaz de explorer, de la siguiente forma.

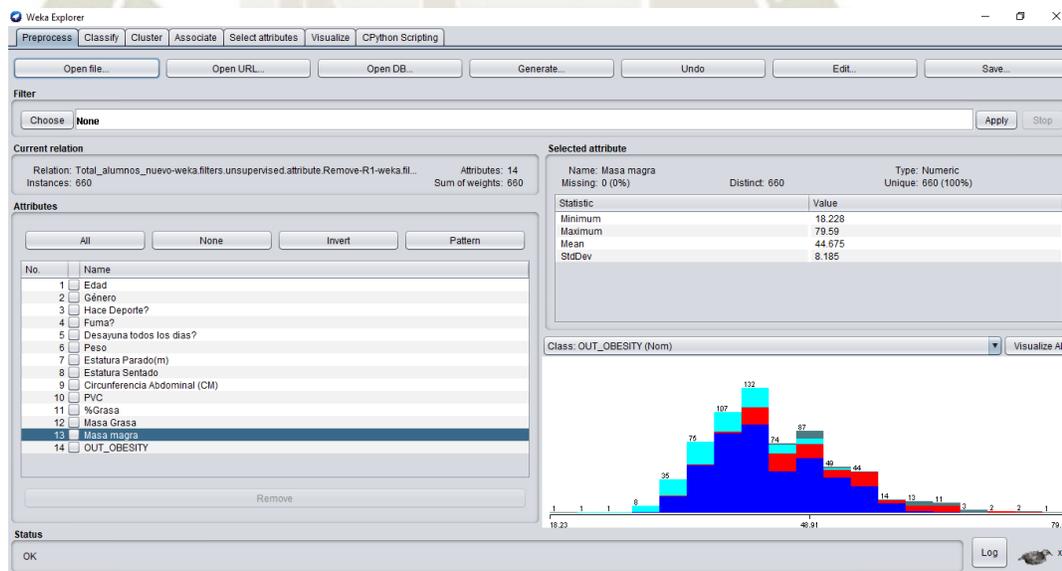


Figura 25. Entorno inicial – Pre procesamiento en Weka.

En el entorno de pre procesamiento de información, es posible hacer uso de los filtros que nos brinda Weka, con el objetivo de obtener data limpia y lista para pasar a la

fase de clasificación, en nuestro caso la data se encuentra preparada y no será necesario aplicar los filtros mencionados.

Dentro de las opciones que se observan en la herramienta, está la de escoger alguna técnica de clasificación, que se encuentran agrupadas según el tipo de método. Como se observa en la figura 26, a continuación.

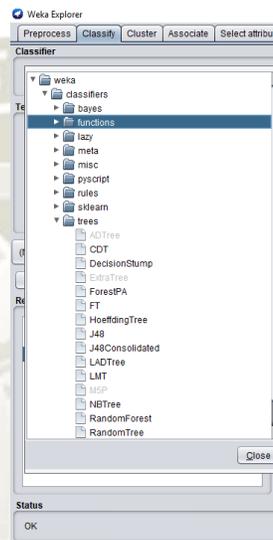


Figura 26. Lista de clasificadores de Weka.

Luego de haber seleccionado la técnica de clasificación, se debe iniciar para poder observar los resultados generales, y posteriormente realizar un análisis detallado de los resultados como el costo del modelo de datos de entropía o curva de ROC.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      623          94.3939 %
Incorrectly Classified Instances    37           5.6061 %
Kappa statistic                    0.8938
Mean absolute error                 0.0335
Root mean squared error             0.1625
Relative absolute error             12.5256 %
Root relative squared error         44.4615 %
Total Number of Instances          660

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.969   0.094   0.949   0.969   0.959   0.884   0.949   0.950   NORMAL
0.898   0.014   0.917   0.898   0.907   0.891   0.964   0.905   SOBREPESO
0.881   0.011   0.941   0.881   0.910   0.894   0.950   0.891   PESO BAJO
0.964   0.002   0.964   0.964   0.964   0.963   0.981   0.931   OBESO
Weighted Avg.   0.944   0.064   0.944   0.944   0.944   0.890   0.953   0.933

=== Confusion Matrix ===

 a  b  c  d  <-- classified as
412 7  6  0 | a = NORMAL
 9 88 0  1 | b = SOBREPESO
13 0 96 0 | c = PESO BAJO
 0 1 0 27 | d = OBESO
    
```

Figura 27. Resultados de análisis de clasificación en Weka.

Como una opción adicional a la clasificación se puede visualizar para observar la relación de atributos.

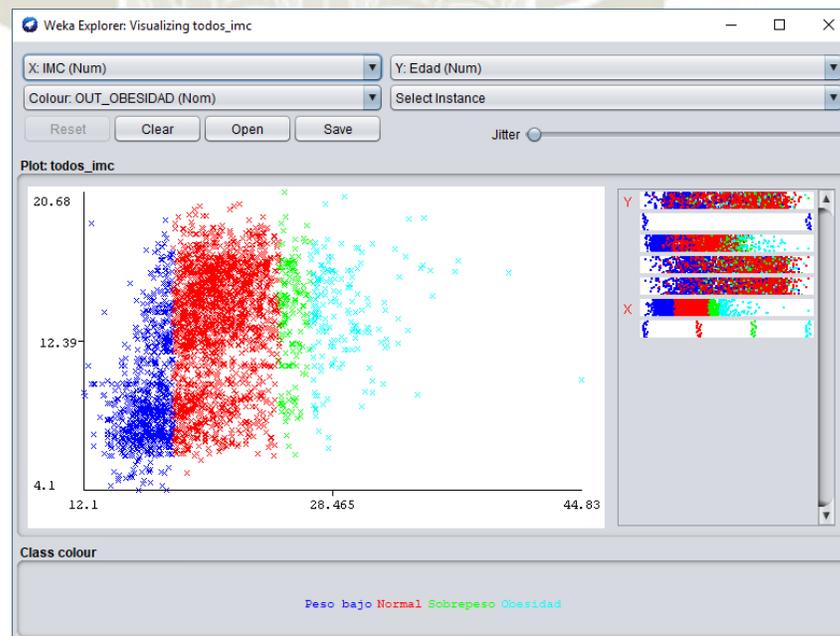


Figura 28. Resultados de visualización entre los atributos IMC y Edad.

CAPITULO 5: DISEÑO, DESARROLLO E IMPLEMENTACION DEL SISTEMA DE PREDICCIÓN

5.1 Arquitectura de Sistema

5.1.1 Arquitectura Lógica del Sistema

El patrón de arquitectura que se va a utilizar para el proyecto actual es el denominado 3 o más capas. Este patrón se usa en la gran mayoría de sistemas, el sistema gestiona datos y los guarda en una base de datos, además se tiene una interfaz visual que será la interacción con los usuarios y por último una parte se encargará de procesar y gestionar datos. La arquitectura en tres capas divide el sistema en 3 partes diferenciadas de tal forma que cada capa sólo se comunique con la inferior, esas 3 capas se llaman: Capa de presentación, Capa lógica de negocio y Capa de datos.

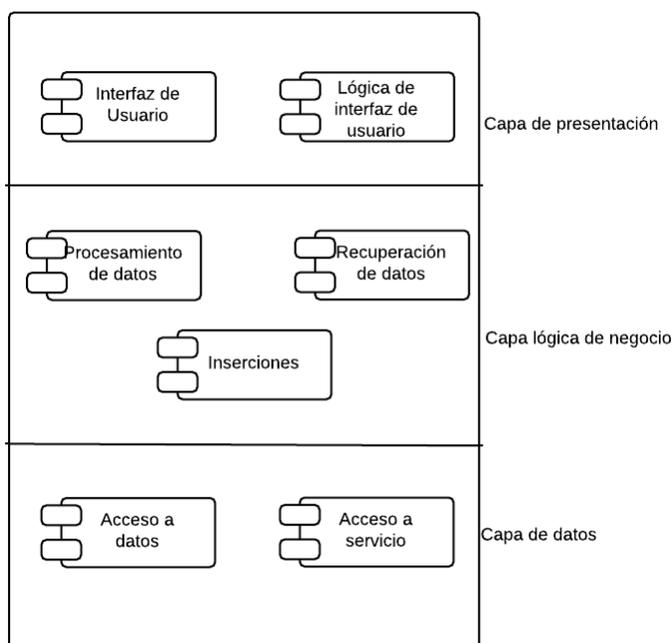


Figura 29. Patrón de diseño en capas. Fuente: Propia.

La capa de presentación es la que se encarga de que el sistema interactúe con el usuario y viceversa, muestra el sistema al usuario, presenta y obtiene la información que el usuario requiere, recibiendo los eventos accionados por los usuarios a través de la interfaz. Esta capa se comunica únicamente con la capa lógica de negocio.

La capa lógica de negocio es donde residen las funciones que se ejecutan, se reciben las peticiones del usuario, se procesa la información y se envían las respuestas tras el proceso. Esta capa se comunica con la capa de presentación, para recibir solicitudes y presentar los resultados, y con la capa de acceso a datos, para realizar solicitudes al gestor de base de datos, ya sea de recuperación o almacenamiento.

El sistema web empleará el estilo arquitectónico de capas y será organizado en tres capas: la capa de presentación, la capa lógica de negocio y la capa de datos. La capa de presentación contendrá la interfaz gráfica del usuario, la principal responsabilidad es mostrar la información al usuario, interpretar los comandos de este y realizar algunas validaciones simples de ingreso de datos. En la capa de lógica de negocio, se realizarán todos los cálculos basados en la información ingresada por el usuario, datos almacenados y validaciones. Por último, la capa de datos, que se encargará de guardar y recuperar datos requeridos por el sistema.

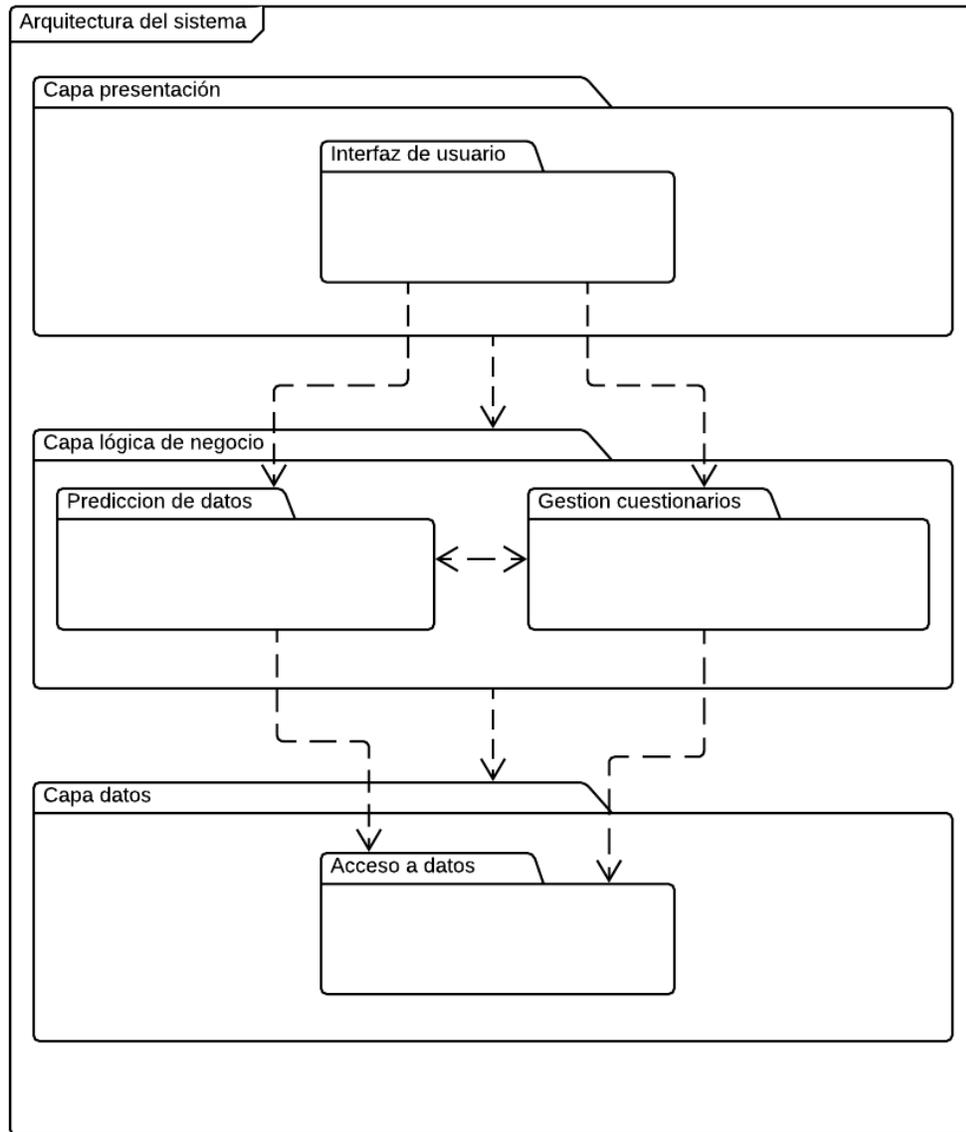


Figura 30. Arquitectura del sistema. Fuente: Propia.

La arquitectura propuesta para el sistema de predicción de obesidad es presentada en la siguiente figura:

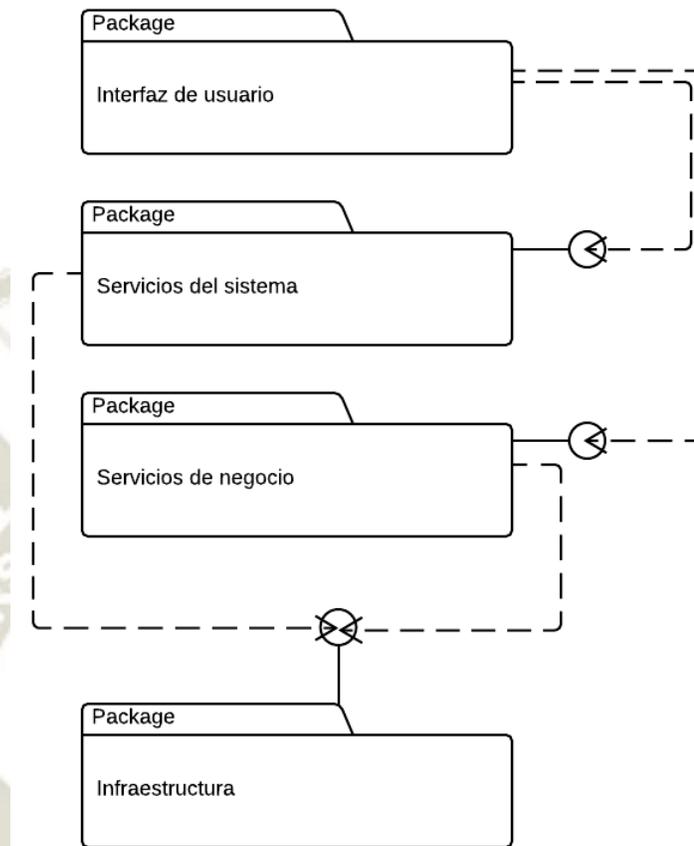


Figura 31. Arquitectura del sistema de predicción. Fuente: Propia.

La capa de Interfaz de usuario se encarga del manejo de la lógica del usuario, Los módulos identificados para esta capa son los siguientes:

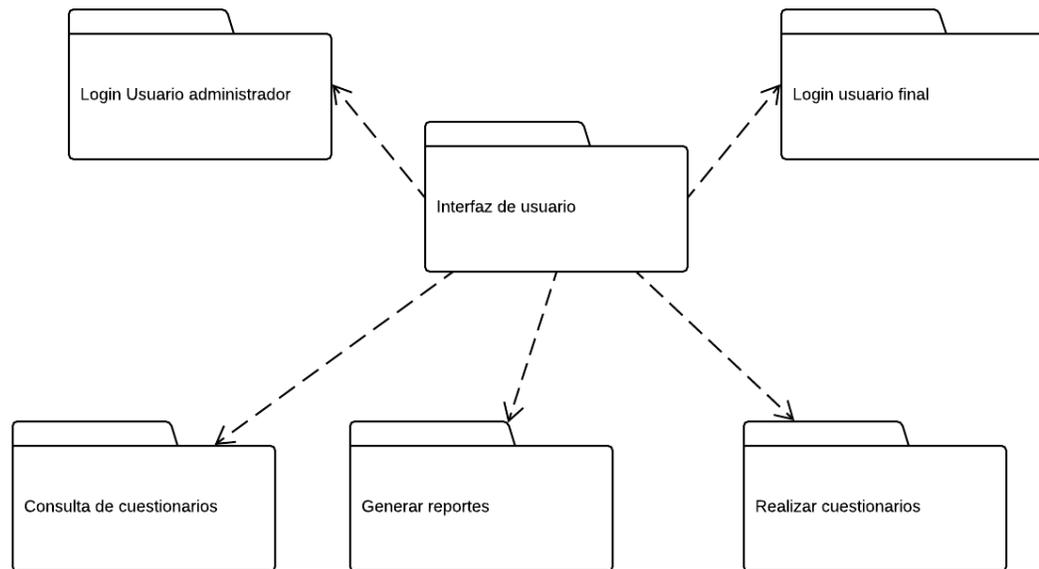


Figura 32. Módulos capa de presentación. Fuente: Propia.

En la capa de servicios del sistema se tiene a los servicios que se utilizan para los módulos de la capa superior, es decir la capa de presentación. En esta capa son definidas las clases controladoras encargadas de manejar la lógica de los casos de uso de los servicios básicos del sistema. En la siguiente figura se muestra el servicio básico de autenticación.

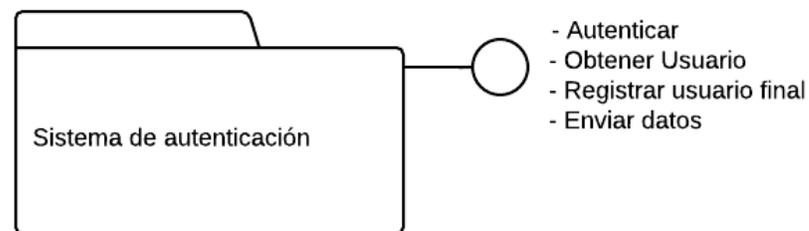


Figura 33. Diagrama de servicios básicos. Fuente: Propia.

La capa de servicios de negocio del sistema de predicción de obesidad, mantiene los módulos que representan los servicios para el manejo de información del negocio. Los módulos de esta capa poseen una única interfaz con los servicios que permiten que las operaciones de las capas superiores sean realizadas. A continuación, el diagrama de módulos.

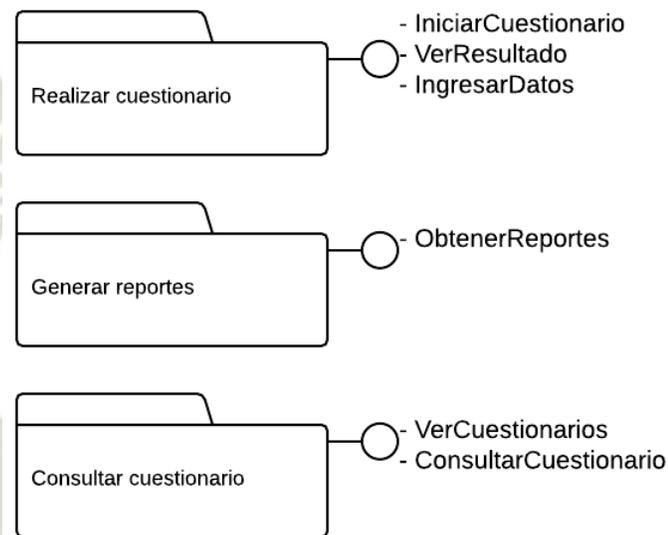


Figura 34. Diagrama de módulos servicios de negocio. Fuente: Propia.

La capa de Infraestructura se tiene principalmente a el módulo de acceso a datos, que se encarga de ingresar y recuperar datos almacenados en la base de datos. A continuación, los servicios de la capa.

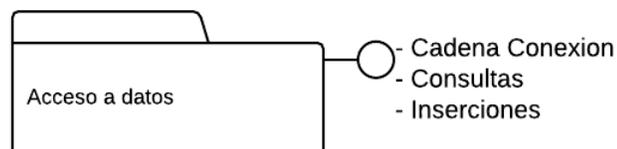


Figura 35. Diagrama de servicios de Infraestructura. Fuente: Propia.

5.1.2 Arquitectura Física del Sistema

En este punto se describe la topología del sistema, es decir cómo será asignado en forma física la aplicación web (software) a los diversos equipos de cómputo (hardware). Para el proyecto, la arquitectura física será representada de la siguiente forma:

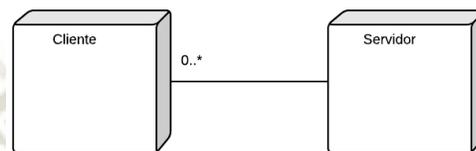


Figura 36. Diagrama de Arquitectura Física del sistema. Fuente: Propia.

Se consideran dos tipos de nodos: Cliente y Servidor, el primer nodo representa todas las estaciones de trabajo de los usuarios finales y usuarios administradores, de donde harán uso del sistema. El segundo nodo representa el computador donde se encuentra instalada la aplicación y además el alojamiento de la base de datos.

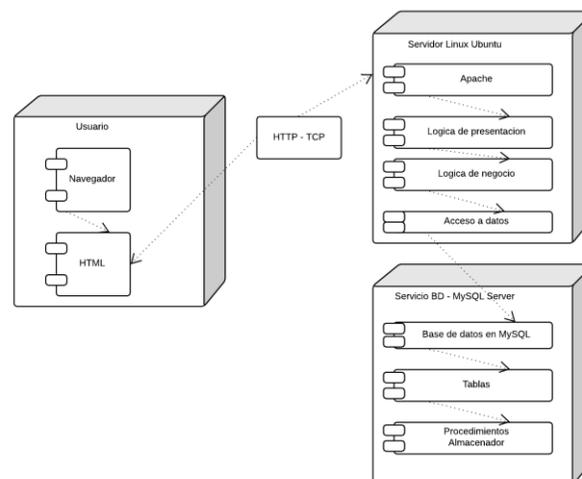


Figura 37. Diagrama de Despliegue de la Arquitectura Física del sistema. Fuente: Propia.

5.2 Diseño de la Base de Datos

Para el diseño de la base de datos, se tienen dos esquemas principales, el primero es el esquema conceptual de la base de datos, que es donde se hace un diagrama entidad relación y el segundo es donde se hace un esquema implementable, es decir que podrá ser ingresado directamente al gestor de base de datos.

5.2.1 Esquema conceptual de la base datos

En este diagrama se describe el flujo de cómo las “entidades”, como personas, objetos o conceptos, se relacionan entre si dentro del sistema. Este modelo es utilizado con el objetivo de diseñar la base de datos que será implementada en el gestor. Algo importante sobre este diagrama es que es de fácil entendimiento para el Usuario.

En la siguiente figura, se muestra el diagrama entidad relación del sistema.

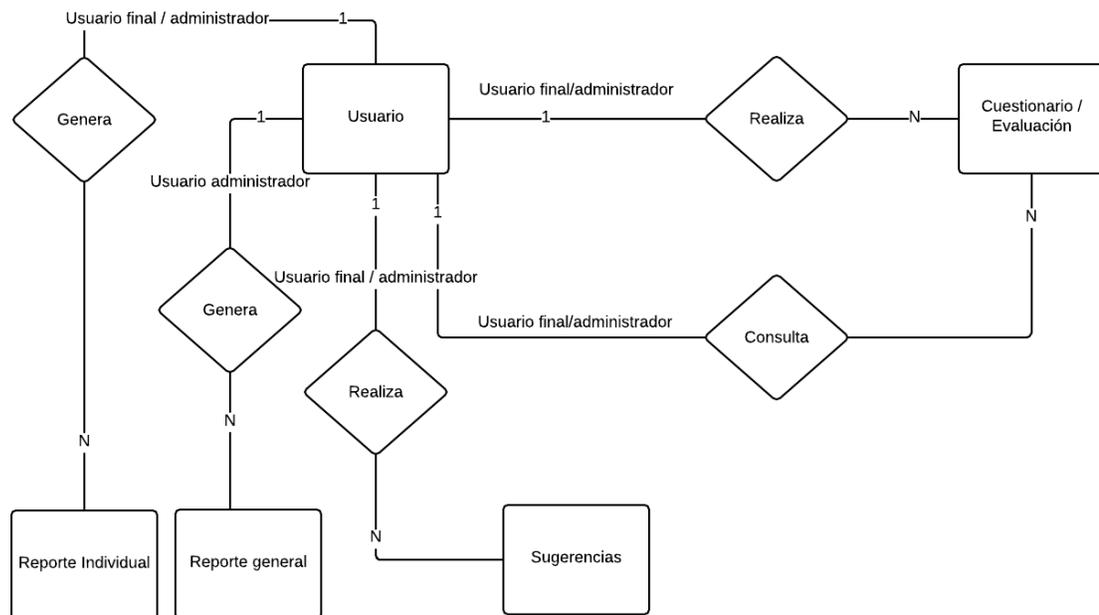


Figura 38. Diagrama de Entidad – Relación del Sistema Predicción de Obesidad. Fuente: Propia.

5.2.2 Esquema de Implementación de la Base de Datos

En este punto se define el esquema a implementar en el gestor de base de datos, también se incluyen detalles de la implementación de acuerdo al gestor a usar. En la siguiente figura se muestra el esquema implementable de la base de datos del sistema de predicción de obesidad en una vista de modelo relacional.

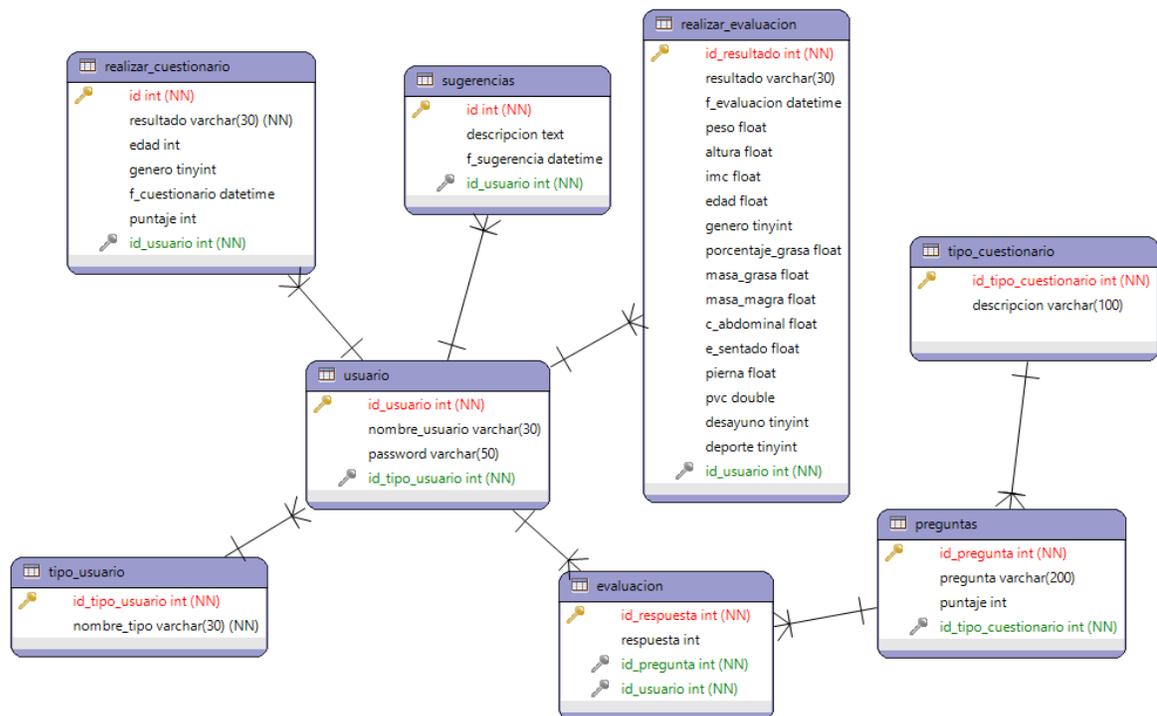


Figura 39. Modelo Físico del Sistema Predicción de Obesidad. Fuente: Propia.

5.3 Diseño de las Interfaces

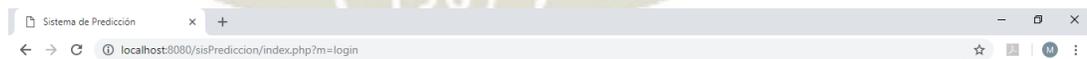
La interfaz web son los elementos gráficos que permitirán al usuario acceder a los contenidos, navegar e interactuar con el sistema de predicción de obesidad. Es por ello que es importante lograr un diseño atractivo para que el usuario tenga facilidad de acceso a los contenidos, interactúe con eficacia con todos los componentes y así sentirse cómodo

con el sistema. El lenguaje de programación a utilizarse será PHP en combinación con otras tecnologías como Javascript, CSS, JQUERY, AJAX y Bootstrap. Algo importante que resaltar es el uso del lenguaje Javascript, que hará de nuestro sistema algo más dinámico y nos permitirá reducir las constantes llamadas al servidor.

Por otro lado, el uso de CSS nos brindará ventajas como la separación de la forma y el contenido, es decir que se evitará duplicación de código, el mantenimiento del sistema será más simple y además reducirá en gran medida los tiempos de carga de páginas.

5.3.1 Diseño de Interfaz de Inicio de Sesión para Usuario final

En la siguiente figura, se muestra la primera página con la que interactúa el usuario, el inicio de sesión. Es en este formulario donde el usuario debe ingresar un nombre de usuario, en la casilla en blanco, con el motivo de que pueda registrar el nombre en la base de datos. Además, se tiene una opción de “Ingresar en modo administrador”, que nos direcciona a otro formulario de ingreso, dirigido a usuarios administradores.



Ingresar nombre de usuario

[Ingresar en modo administrador](#)
[Ingresar en modo anónimo](#)
[Descargar manual de usuario](#)

Figura 40. Interfaz de Inicio de Sesión de usuario final. Fuente: Propia.

5.3.2 Diseño de Interfaz de Inicio de Sesión para Usuario administrador

Para ingresar en modo administrador, además de ingresar un nombre de usuario, será necesario ingresar una contraseña, sólo los usuarios administradores pueden realizar reportes generales. Debajo de los dos casilleros en blanco se tiene un link “Ingresar en modo usuario”, que nos direcciona a la página de logueo del usuario final.

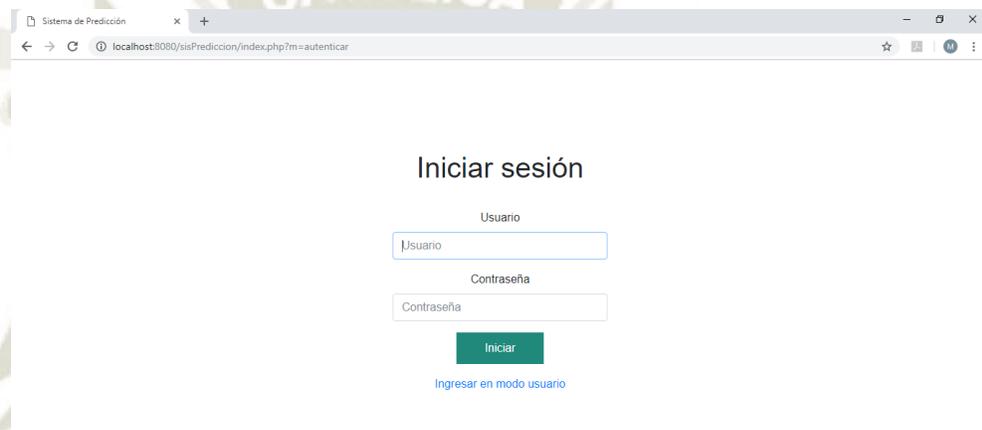
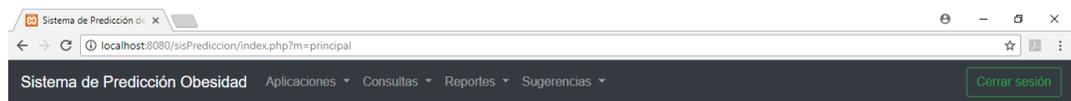


Figura 41. Interfaz de Inicio de Sesión de usuario administrador. Fuente: Propia.

5.3.3. Diseño de Interfaz Menú principal de Usuario administrador

En la figura 42, se muestra el formulario principal una vez autenticado el usuario, listando las diferentes opciones a las que puede acceder y navegar en el sistema.



Bienvenido

Sistema de Predicción Obesidad

Cuestionario

Figura 42. Interfaz de Menú Principal Usuario Administrador. Fuente: Propia.

El menú principal de un Usuario Administrador cuenta con las opciones de Aplicaciones, Consultas, Reportes y Sugerencias. Todas ellas se encuentran ubicadas en la barra ubicada en la parte superior de la página.

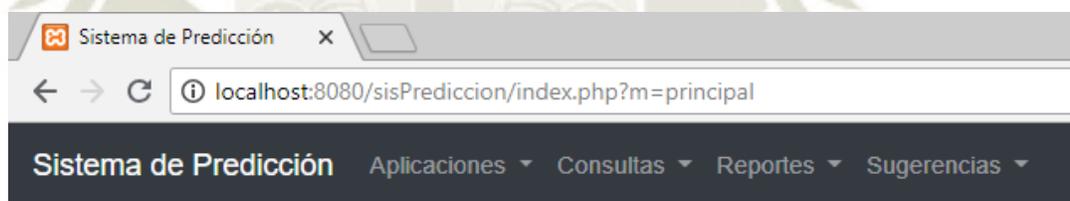


Figura 43. Lista de opciones del menú Aplicaciones. Fuente: Propia.

La lista de opciones del menú Aplicaciones es la siguiente: Cuestionarios, que tiene la opción cuestionario de Actitud hacia la Obesidad, nos direcciona a una página donde se muestra un formulario tipo cuestionario, con preguntas y opciones para marcar. La opción Realizar evaluación IMC, dirige a un pequeño formulario donde se piden datos que se ingresan en casillas de texto.

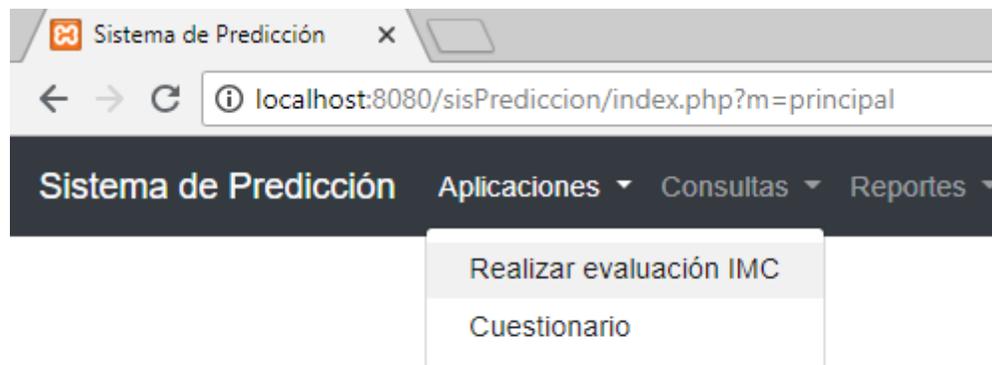


Figura 44. Lista de opciones del menú Consultas. Fuente: Propia.

La lista de opciones del menú Consultas, corresponde a los datos ingresados en los cuestionarios anteriormente mencionados en el menú de aplicaciones. Sirven para ver las preguntas que se respondieron y cuál es el resultado del cuestionario y/o evaluación.

El menú reporte cuenta con las opciones Reporte individual y Reporte general, en la primera opción se podrá ver el resultado final de las 3 (como máximo número de intentos) evaluaciones y/o cuestionarios, en la segunda opción se podrá realizar un reporte general de todos los usuarios que se ingresaron hasta la fecha. El menú de sugerencias es de libre acceso, tanto para usuarios administradores y finales, el propósito del menú de sugerencias, es mejorar el sistema, es decir que el usuario pueda sugerir nuevos cambios y mejoras al sistema.

CUESTIONARIO ACTITUD HACIA LA OBESIDAD

Fecha de nacimiento: Género: Hombre Mujer

T.A = Totalmente de acuerdo; NA-ND = Ni acuerdo, ni desacuerdo; T.D. = Totalmente desacuerdo.

Nro	Pregunta	T.A.	NA-ND	T.D.
1	Con frecuencia prefiero quedarme en casa, que salir a jugar al parque.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	A menudo no me llama la atención mucho hacer ejercicios físicos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Me agrada salir a reuniones sociales con mis amigos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Me siento a gusto comiendo en la calle que en mi casa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Cuando tengo tiempo libre, prefiero mirar televisión en casa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Cuando estoy en casa no me gusta ayudar en los quehaceres del hogar (limpiar, barrer, planchar, cortar pasto, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Me agrada más la idea de comer comida rápida que las frutas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	Cuando estoy en el colegio me gusta sentarme para conversar con mis amigos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	Te sentirías preocupado en el caso de que subas de peso repentinamente?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	Alguna vez te has sentido intranquilo por tu barriga?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11	Me preocupa mucho el hecho de no poder mantener mi peso ideal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

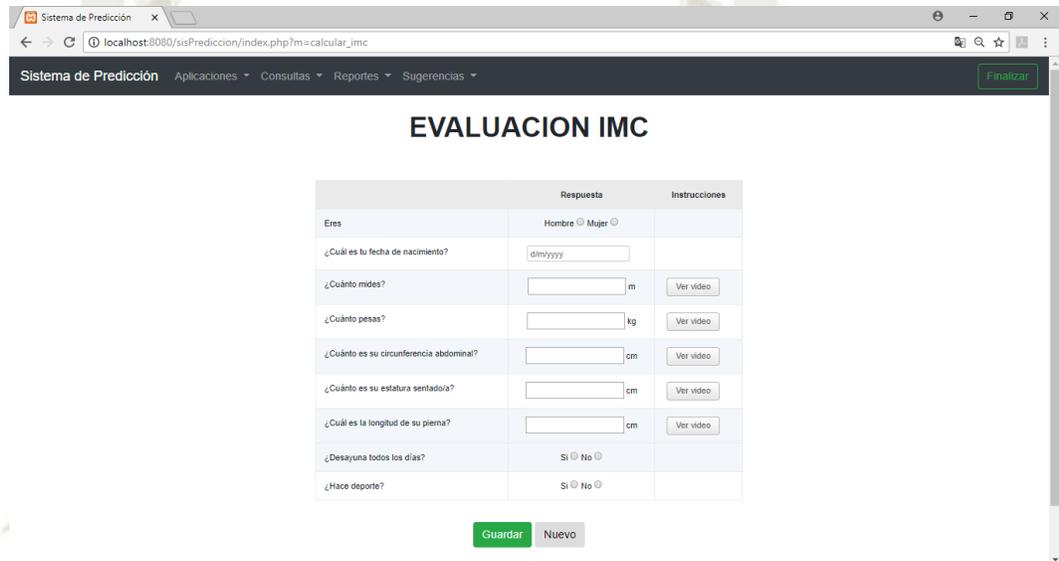
Figura 45. Cuestionario de Actitud hacia la obesidad. Fuente: Propia.

En el cuestionario de Actitud hacia la obesidad, se deben responder todas las preguntas, de lo contrario el sistema no permitirá guardar cambios, además se cuenta con una validación de fecha de nacimiento, como se mencionó anteriormente en el informe este cuestionario va dirigido a una población en un determinado rango de edad. Para enviar los cambios se debe presionar el botón “Guardar”.



Figura 46. Resultado Cuestionario de Actitud hacia la obesidad. Fuente: Propia.

Una vez finalizado satisfactoriamente el cuestionario, se enviarán los resultados generados por el sistema en formato texto, sobre el nivel de actitud hacia la obesidad, y además en un gráfico indicando la ubicación del usuario de acuerdo a la tabla de percentiles generada una investigación realizada por Gómez et al., (2015).



	Respuesta	Instrucciones
Eres	Hombre <input type="radio"/> Mujer <input type="radio"/>	
¿Cuál es tu fecha de nacimiento?	<input type="text" value="d/m/yyyy"/>	
¿Cuánto mides?	<input type="text"/> m	<input type="button" value="Ver video"/>
¿Cuánto pesas?	<input type="text"/> kg	<input type="button" value="Ver video"/>
¿Cuánto es su circunferencia abdominal?	<input type="text"/> cm	<input type="button" value="Ver video"/>
¿Cuánto es su estatura sentado/a?	<input type="text"/> cm	<input type="button" value="Ver video"/>
¿Cuál es la longitud de su pierna?	<input type="text"/> cm	<input type="button" value="Ver video"/>
¿Desayuna todos los días?	Si <input type="radio"/> No <input type="radio"/>	
¿Hace deporte?	Si <input type="radio"/> No <input type="radio"/>	

Figura 47. Evaluación IMC. Fuente: Propia.

En la evaluación IMC, se deben ingresar los atributos requeridos para realizar la predicción del estado de peso del usuario, los atributos se presentan en formato de interrogante, y son: ¿Eres? (Género), ¿Cuál es su fecha de nacimiento?, ¿Cuánto mide?, ¿Cuánto pesa?, ¿Cuánto es su circunferencia abdominal?, ¿Cuánto es su estatura sentado/a?, ¿Cuál es la longitud de su pierna?, ¿Desayuna todos los días? y ¿Hace deporte?, estas preguntas, de ser respondidas correctamente, permiten al sistema realizar el proceso de predicción para saber el estado de peso del usuario evaluado.



Figura 48. Resultado Evaluación IMC. Fuente: Propia.

Una vez ingresados correctamente los datos del usuario, el sistema deberá mostrar los resultados correspondientes a la evaluación (la predicción del algoritmo), y adicionalmente mostrará un gráfico de percentiles, detallando la ubicación actual del usuario.

Resultado	Peso (kg)	Altura (m)	IMC	Edad	Género	Porcentaje grasa	Masa grasa	Masa magra	Circunferencia abdominal(cm)	Estatura sentado(cm)	Pierna(cm)
1 PESO BAJO	42	1.45	19.98	0	Hombre	0%	0	0	50	56	42
2 NORMAL	41	1.43	20.05	14.58	Hombre	11.21%	4.6	36.4	51	53	48

Figura 49. Consulta Evaluación IMC. Fuente: Propia.

En el módulo de consultas, el usuario tiene la opción de ver las respuestas ingresadas, y también los resultados obtenidos por el sistema.

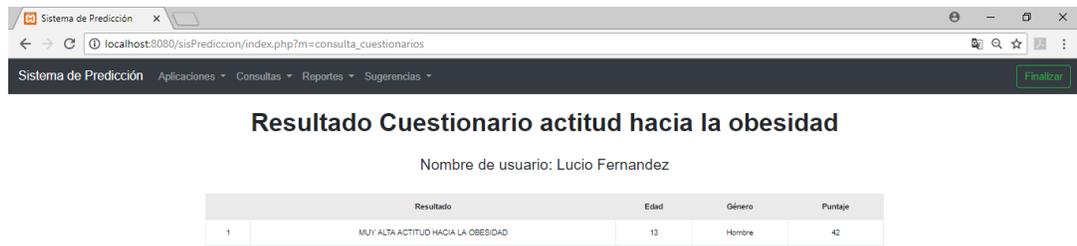


Figura 50. Consulta Cuestionario actitud hacia la obesidad. Fuente: Propia.

De la misma forma, el usuario puede ver datos ingresados en el cuestionario de actitud hacia la obesidad.

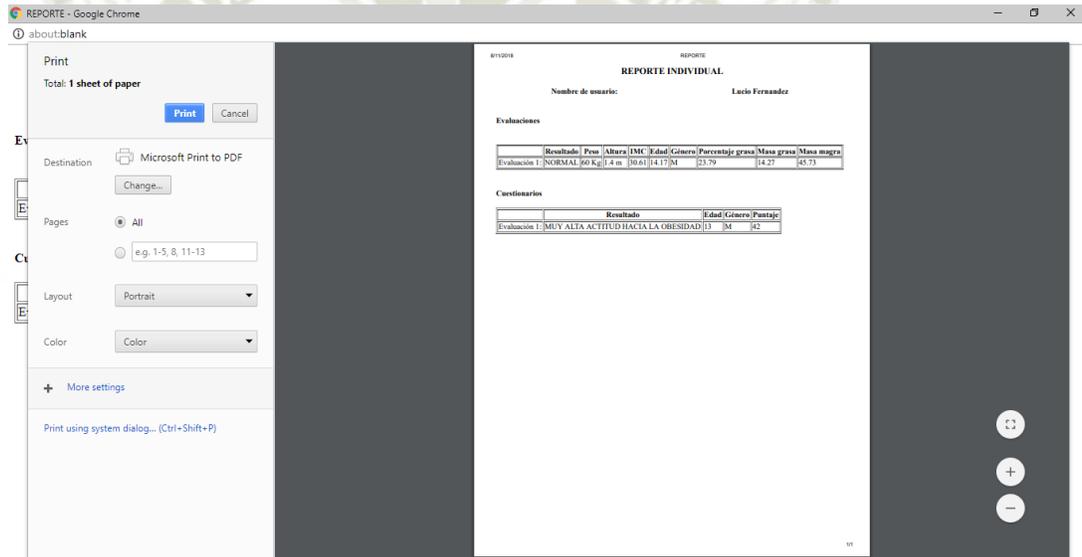


Figura 51. Reporte individual de usuario. Fuente: Propia.

Reporte individual, es una opción que tienen tanto usuarios finales como administradores, en este reporte se muestran todas las evaluaciones y cuestionarios que realizó el usuario, como se mencionó anteriormente, el máximo número de evaluaciones y/o cuestionario es de 3.

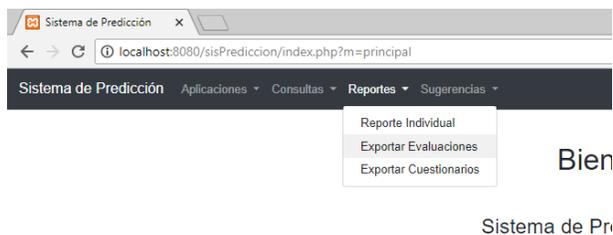


Figura 52. Lista de opciones Reportes usuario administrador. Fuente: Propia.

Un usuario administrador puede generar un reporte general, de todos los usuarios evaluados por el sistema, las opciones para realizar estas acciones son: “Exportar Evaluaciones” y “Exportar Cuestionarios”.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2	NOMBRE (RESULTADO PESO(kg)	ALTURA(m	IMC	EDAD	GENERO	PORCENT/MASA GR/	MASA MA	CIRCUNFE	ESTATURA	PIERNA(cm	PVC	DESAYUNI	DEPORTE	FECHA DE EVALUACION							
3	Ricardo HI SOBREPES	60	1.64	22.31	24.25	1	16.15	9.69	50.31	50	68	40	0	0	0	24/07/2018					
4	Ricardo HI NORMAL	65	1.72	21.97	15.33	1	13.69	8.9	56.1	40	45	30	0	0	1	24/07/2018					
5	Manuel S SOBREPES	70	1.74	23.12	25.33	1	17.37	12.16	57.84	50	50	30	0	1	1	24/07/2018					
6	Margarita SOBREPES	89	1.78	28.09	19.08	0	32.7	29.1	59.9	80	70	40	0	1	1	25/07/2018					
7	Juan Lope NORMAL	60	1.75	19.59	24.5	1	12.94	7.76	52.24	40	50	40	0	1	1	26/07/2018					
8	Marco Tic SOBREPES	80	1.76	25.83	23.5	1	20.2	16.16	63.84	60	60	20	0	1	0	26/07/2018					
9	Maria Can NORMAL	60	1.7	20.76	26.67	1	14.85	8.91	51.09	60	42	43	-1.75898	1	1	29/07/2018					
10	Maria Can SOBREPES	70	1.69	24.51	25.17	1	19	13.3	56.7	66	62	46	1.823836	1	1	29/07/2018					
11	Maria Can SOBREPES	70	1.66	25.4	24.33	0	30.68	21.48	48.52	65	60	30	-0.01822	0	1	29/07/2018					
12	Maria Campos	68	1.75	22.2	19.08	0	25.63	17.43	50.57	55	60	39	-1.63714	1	1	29/07/2018					
13	Maria Can NORMAL	55	1.55	22.89	19.75	1	15.81	8.7	46.3	60	40	56	-3.95076	0	1	29/07/2018					
14	Maria Can SOBREPES	76	1.69	26.61	26.58	0	32.65	24.81	51.19	90	65	70	4.477591	1	0	29/07/2018					
15	Maria Can SOBREPES	78	1.7	26.99	26.58	0	33.1	25.82	52.18	80	70	80	6.012161	1	0	29/07/2018					
16	Maria Can SOBREPES	70	1.69	24.51	18.58	1	17.49	12.24	57.76	80	79	82	1.531146	1	0	29/07/2018					
17	Maria Can NORMAL	50	1.63	18.82	22.33	1	11.52	5.76	44.24	60	40	40	-3.13497	1	1	29/07/2018					
18	Maria Can NORMAL	59	1.66	21.41	19.42	1	13.96	8.24	50.76	80	78	68	1.751867	1	1	29/07/2018					
19	Gilberto S NORMAL	60	1.56	24.65	19.58	1	17.88	10.73	49.27	60	70	49	0.874529	1	1	30/07/2018					
20	Gilberto S NORMAL	60	1.65	22.04	14.67	0	24.42	14.65	45.35	40	50	45	-4.1777	1	1	31/07/2018					
21	Humbert NORMAL	65	1.65	23.88	18.67	1	16.75	10.89	54.11	60	70	40	0.616349	1	0	1/8/2018					
22	Humbert SOBREPES	80	1.7	27.68	17.92	0	31.94	25.55	54.45	60	60	40	-2.3134	0	0	1/8/2018					
23	Humbert SOBREPES	85	1.75	27.76	17.33	1	21.1	17.94	67.06	60	80	40	1.60099	0	1	1/8/2018					

Figura 53. Reporte general evaluaciones IMC. Fuente: Propia.

En la figura 53 se muestra un ejemplo de un reporte general obtenido en formato xls, este reporte puede ser de mucha utilidad para futuras investigaciones, donde se pueden comprobar los resultados producidos por el sistema de actitud hacia la obesidad y/o evaluación IMC.

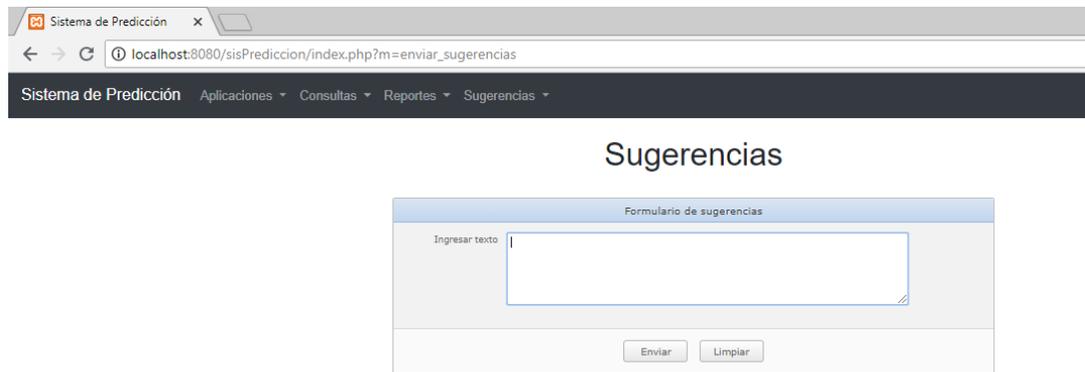


Figura 54. Formulario ingreso sugerencias. Fuente: Propia.

5.3.4. Diseño de Interfaz Nivel Usuario final

El usuario final cuenta con todas las opciones de un usuario administrador, excepto por los reportes generales, que son exclusivamente para usuarios administrador, por lo demás un usuario final posee las mismas condiciones.

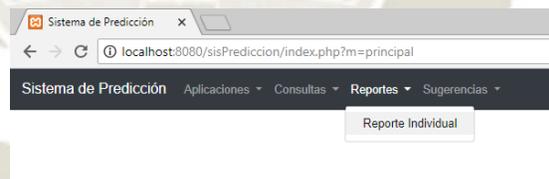


Figura 55. Menú de opciones reportes usuario final. Fuente: Propia.

En la figura 55, se puede ver que un usuario final, sólo cuenta con la opción Reporte individual.

5.4. Desarrollo del Sistema

En este punto se expondrá las partes principales del desarrollo del sistema, empezando con las herramientas que fueron utilizadas y explicando cada uno de los puntos de desarrollo fundamentales para el correcto funcionamiento del sistema.

5.4.1.1. Desarrollo de la Base de Datos

El primer paso para el desarrollo de la base de datos y del sistema web consiste en la instalación de XAMPP, paquete de software libre que contiene principalmente el sistema de gestión de base de datos MySQL, el servidor Apache y los intérpretes para lenguajes script PHP y Perl. Esta instalación facilita la configuración del servidor Apache y también permite realizar pruebas en el ordenador de desarrollo.

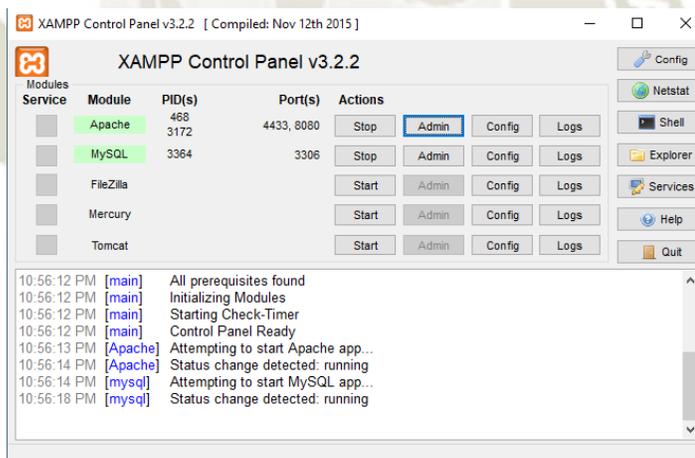


Figura 56. Panel de control XAMPP. Fuente: Propia.

Una vez realizada la conexión local, se dirige a crear la base de datos, tablas, vistas y procedimientos almacenados si fueran necesarios. En la figura que será mostrada a continuación se observa la representación del diagrama de base de datos, ya

implementado en MySQL, con sus respectivos elementos que son tablas, enlaces o relaciones, claves primarias y claves secundarias.

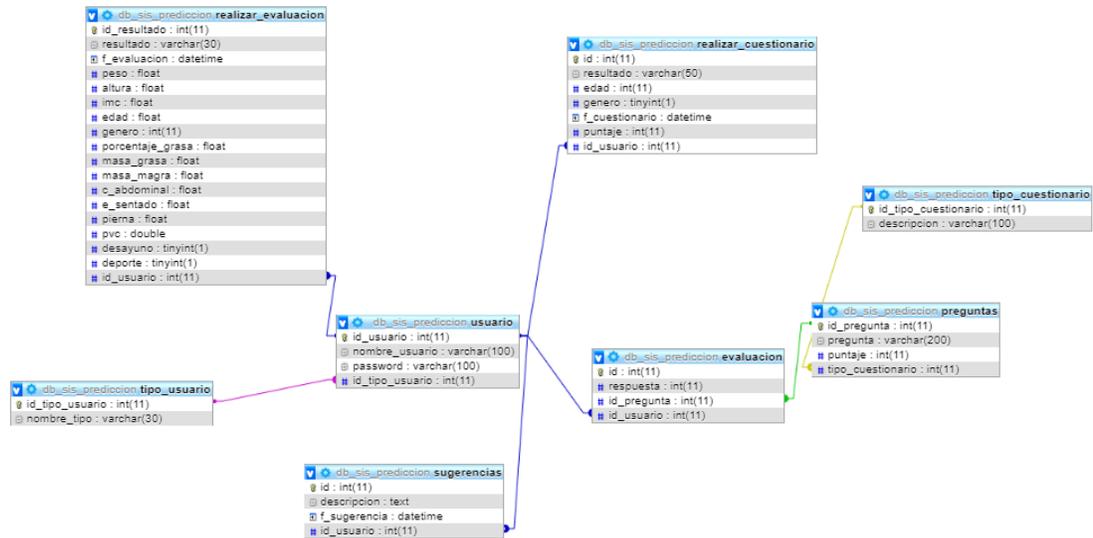


Figura 57. Esquema de base de datos implementado en Phpmyadmin. Fuente: Propia.

5.4.2. Desarrollo de la Aplicación

Para la codificación del sistema se utilizó el conocido editor de texto Sublime Text en su versión 3.1.1 trial. El nombre del proyecto es “sisPrediccion” programado en el lenguaje PHP.



Figura 58. Estructura del proyecto sisPrediccion en Sublime. Fuente: Propia.

Las carpetas creadas para organizar el proyecto son: bd, controlador, img, libraries, media, modelo, php, style y vistas. Estas nos proporcionan un cierto nivel de orden para el proyecto, además ayudan mantener la estructura MVC, modelo vista controlador.

Esta estructura mejora y contribuye al requerimiento no funcional de mantenibilidad del proyecto y evita la redundancia innecesaria de código, para futuras mejoras e implementaciones. De esta forma se hace uso de una buena metodología de programación Web.

El código Javascript facilita de gran forma el manejo de la interfaz de usuario, haciéndola más interactiva, permitiendo validar de forma inmediata si los datos que ingresa el usuario están permitidos dentro de los parámetros del sistema. Incluso esta tecnología nos permite mediante el uso de Ajax (Asynchronous Javascript and XML), que un servidor y un navegador intercambien información en XML, de una forma asíncrona. La gran diferencia y ventaja que nos brinda AJAX es realizar conexiones

e intercambios de información con el servidor sin necesidad de recargar toda la página.

```
function getPreguntas(tipo){
    var packet_message={ tipo:tipo, op:'getPreguntas'};
    $.ajax({
        url: 'php/webservices_get.php',
        type: 'GET',
        data: { packet_message: JSON.stringify(packet_message) },
        success: function(data)
        {
            if(data.success)
            {
                length_cuestionario = data.preguntas.length;
                editarTabla(data.respuestas, data.preguntas);
            }
            else{
                alert("Error");
            }
        },
        error: function(){
        }
    });
}

function editarTabla(array_respuestas, array_preguntas){
    var table = $("#tabla1 tbody");
    table.empty();
    var id=0;
    $.each(array_preguntas, function(idx, elem){
        array_preguntas_id.push(elem.id_pregunta);
        id++;
        table.append("<tr><td>"+id+"</td><td>"+elem.pregunta+"</td><td><input type='radio'
        name='radioGroup"+id+"' value="+elem.puntaje.substring(0, 1)+"
        onclick=""</td><td><input type='radio' name='radioGroup"+id+"'
        value="+elem.puntaje.substring(1, 2)+" onclick=""</td><td><input type='radio'
        name='radioGroup"+id+"' value="+elem.puntaje.substring(2, 3)+" onclick=""</td></tr>");
    });
}
```

Figura 59. Ejemplo de uso AJAX. Función obtener preguntas. Fuente: Propia.

El sistema trabaja directamente con una base de datos, por lo tanto, es necesario contar con una clase Database, que se encargue de realizar las gestiones y

conversaciones con la base de datos. Básicamente esta clase posee los datos requeridos para la conexión, como el nombre de base de datos, nombre de servidor, nombre de usuario y contraseña.

```
<?php
class Database
{
    private static $dbName = 'db_sis_prediccion' ;
    private static $dbHost = 'localhost' ;
    private static $dbUsername = 'root';
    private static $dbUserPassword = '';

    private static $cont = null;

    public function __construct() {
        die('Init function is not allowed');
    }

    public static function connect()
    {
        if ( null == self::$cont )
        {
            try
            {
                self::$cont = new PDO( "mysql:host=".self::$dbHost.";dbname=".self::$dbName,
                self::$dbUsername, self::$dbUserPassword);
            }
            catch(PDOException $e)
            {
                die($e->getMessage());
            }
        }
        return self::$cont;
    }

    public static function disconnect()
    {
        self::$cont = null;
    }
}
?>
```

Figura 60. Clase conexión Database. Fuente: Propia.

5.5. Implementación del Sistema – Migración a la nube

Se requiere implementar el sistema en servidor web o hosting web, para de una forma hacerlo accesible por una gran cantidad usuarios con conexión a internet. En esta situación se hará uso de los servicios de Hostinger, que renta espacios dentro de un servidor que se encuentra en línea 24/7, de esta forma el sistema jamás estará sin presencia en la red.

5.5.1. Copia de seguridad de base de datos

Será necesario restaurar completamente toda la base de datos en el servidor, por ello, se debe obtener un backup completo de la base de datos local, es decir tablas, vistas, procedimientos almacenados y/o disparadores.

En la siguiente figura, se puede ver la página de exportación de base de datos en Phpmyadmin, este método de obtener la copia de seguridad exporta un archivo en formato sql, para nuestro caso lleva el nombre de: db_sis_prediccion, que puede ser ejecutado en cualquier otro servidor apache dependiendo de la versión.

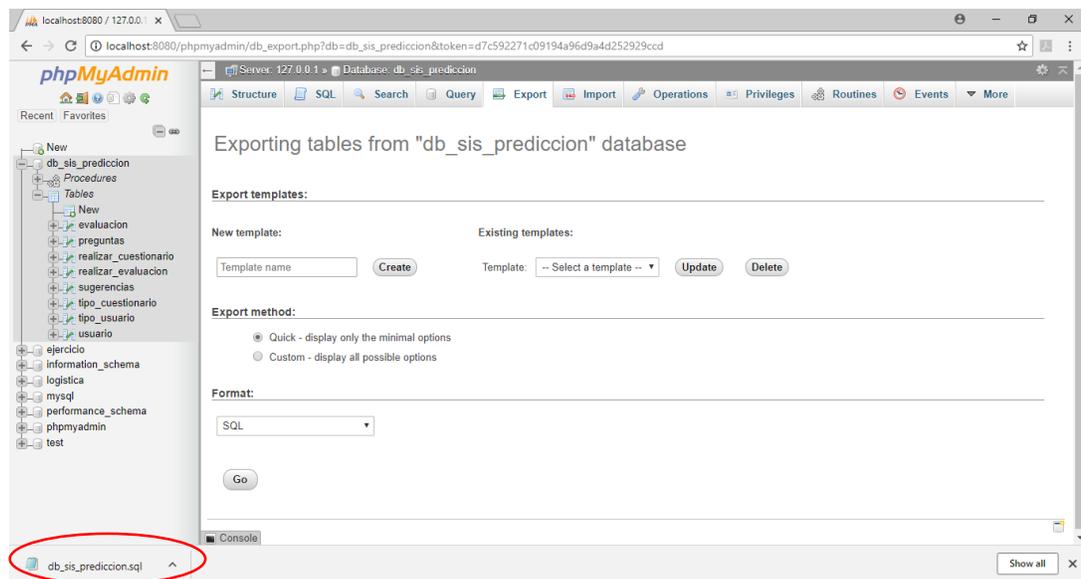


Figura 61. Formulario de exportación de base de datos Phpmysadmin. Fuente: Propia.

5.5.2. Servidor Web

El plan ofrecido por Hostinger, incluye múltiples versiones PHP, PHPMyadmin, Acceso FTP (File Transfer Protocol), copias de seguridad semanales, entre otros. El plan que se contrató es el llamado Hosting Premium, con el objetivo de lograr el correcto funcionamiento del sistema de predicción de obesidad.

Cuenta	
DOMINIO	reidebihu.net
SSL	Activar
DIRECCIÓN IP	185.28.20.193
USUARIO	u471842103
ACTIVADO	19/06/2015
CUENTAS DE CORREO	0
SUBDOMINIOS	1
DOMINIOS APARCADOS	0
CUENTAS FTP	1
BASES DE DATOS MYSQL	2

Figura 62. Plan de servicios Hostinger Premium. Fuente: Propia.

5.5.3. Configuración del Servidor Web

La configuración necesaria para migrar la aplicación a los servidores Hostinger, es relativamente sencilla, teniendo los datos necesarios. Principalmente se debe contar con las credenciales de acceso al servidor y a la base de datos, actualizar tales credenciales en el sistema y posteriormente restaurar la copia de seguridad de base de datos que se obtuvo en puntos anteriores del presente trabajo.

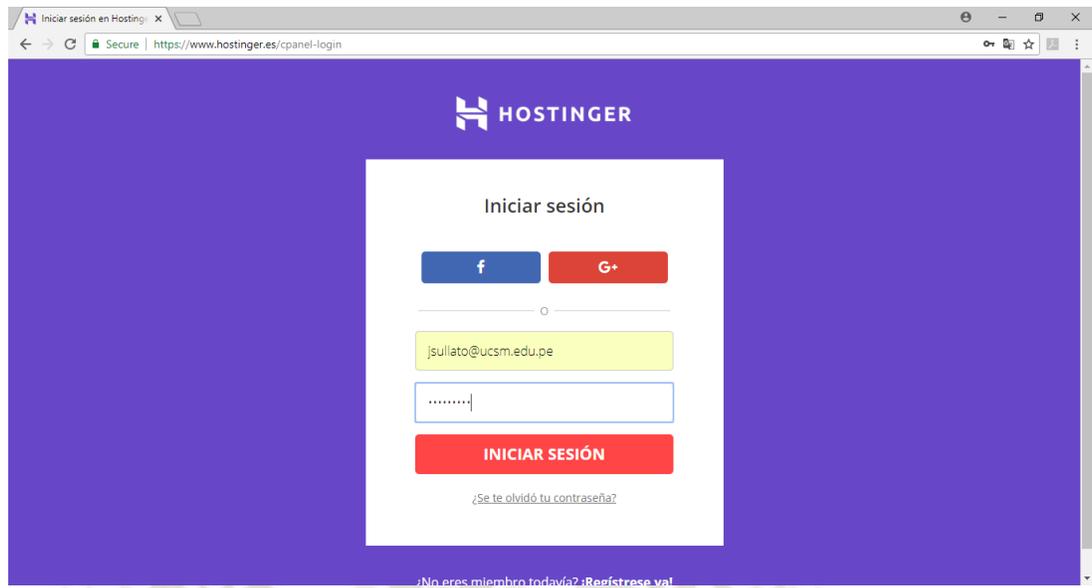


Figura 63. Acceso al panel de control del servidor web. Fuente: Propia.

Los datos más relevantes a tomar en cuenta son:

- Nombre del servidor: server45.hostinger.es
- Dominio: reidebihu.net
- Versión PHP: 5.5
- Versión Apache: Apache/2.4

Para realizar la carga de archivos de una forma más rápida se utilizó el protocolo FTP, conectándose al servidor por medio de la url files45.hostinger.es, mayores detalles en la siguiente figura.

⚙️ Acceso FTP	
Host FTP	files45.hostinger.es
IP FTP	185.28.20.193
Puerto FTP	21
Usuario FTP	u471842103
Contraseña FTP
Carpeta dónde Cargar Archivos	public_html
¿Olvidaste tu contraseña FTP?	Cambiar contraseña de la cuenta
Clientes FTP recomendados	SmartFTP ↗ FileZilla ↗

⚙️ Selección	
Refrescar	
▶ public_html	

⚙️ Crear Nueva Cuenta FTP	
DIRECTORIO	/public_html
USUARIO	<input type="text"/>
INGRESA CONTRASEÑA	<input type="password"/> Generar

Figura 64. Propiedades de acceso FTP al servidor web. Fuente: Propia.

5.5.4. Restauración de base datos en hosting

Para restaurar la base de datos en el servidor web, basta con ejecutar el script generado db_sis_prediccion, en la siguiente imagen se puede ver una de las formas en que es posible llevar a cabo dicha acción.

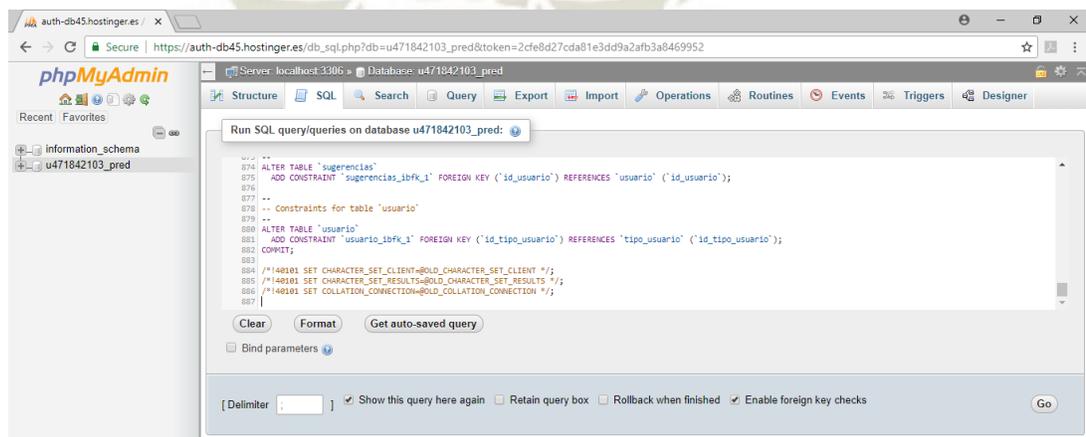


Figura 65. Restauración de base de datos db_sis_prediccion. Fuente: Propia.

5.5.5. Carga de aplicación a servidor web

Como se mencionó anteriormente en el presente trabajo, se utilizó el protocolo FTP, para efectuar la carga de archivos al servidor web. Asimismo, se hizo uso del programa recomendado por Hostinger, SmartFTP a continuación, en la siguiente imagen se puede observar la interfaz del programa, con los archivos subidos al servidor.

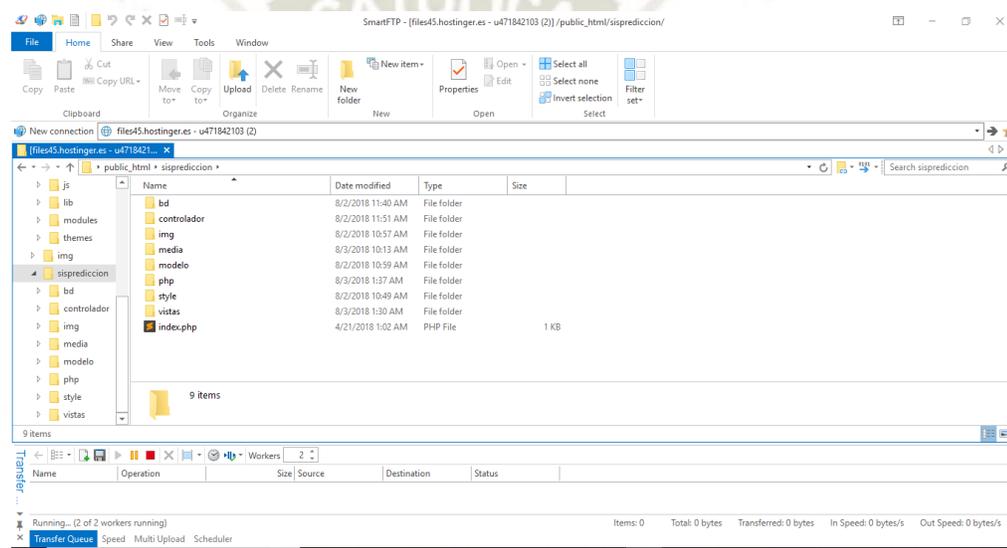


Figura 66. Interfaz SmartFTP. Fuente: Propia.

5.6. Evaluación del sistema

A fin de identificar los atributos de calidad del software, se deberá utilizar el estándar ISO 9126. Aplicando tal estándar, se identificará básicamente seis atributos claves de la calidad de software, que son los siguientes:

5.6.1. Funcionalidad.

Es el grado en que la aplicación satisface las necesidades que indican los siguientes sub atributos: idoneidad, exactitud, interoperabilidad, cumplimiento y seguridad.

En anteriores puntos se realizó la evaluación de la funcionalidad del software, cumpliendo con todos los requisitos funcionales planteados en la etapa de análisis y definición de requerimientos.

5.6.2. Confiabilidad.

Es el tiempo en que la aplicación se encuentra disponible para ser usada según los siguientes sub atributos: madurez, tolerancia a fallas, y facilidad de recuperación.

Medir la confiabilidad de una aplicación podría ser un problema difícil porque no se tiene un buen entendimiento sobre la naturaleza del software. Si no se puede medir la confiabilidad directamente, se puede medir algo que refleje las características relacionadas a la confiabilidad.

A continuación, algunas métricas de confiabilidad que pueden ser utilizadas para cuantificar el producto:

- Tiempo promedio de falla (Por sus siglas en inglés MTTF)
- Tiempo promedio para reparar (Por sus siglas en inglés MTTR)

Para obtener el indicador de confiabilidad Tiempo promedio entre fallas (Por sus siglas en inglés MTBF), es posible combinar las métricas mencionadas, obteniendo la siguiente fórmula.

$$MTBF = MTTF + MTTR \quad (1)$$

Para hallar tales variables tenemos las siguientes 2 fórmulas:

$$MTTF = \frac{\text{Tiempo total de funcionamiento}}{\text{Número de fallas}} \quad (2)$$

$$MTTR = \frac{\text{Tiempo total de inactividad}}{\text{Número de fallas}} \quad (3)$$

Reemplazando, según los datos requeridos para el tiempo promedio de falla se tiene lo siguiente:

$$MTTF = \frac{814}{2}$$
$$MTTF = 407$$

Reemplazando, según los datos requeridos para el tiempo promedio para reparar se tiene lo siguiente:

$$MTTR = \frac{2}{2}$$
$$MTTR = 1$$

Finalmente, para obtener nuestro indicador de confiabilidad, obtendremos el tiempo promedio entre fallas:

$$MTBF = 407 - 1$$
$$MTBF = 406$$

Por lo tanto, luego de obtener un MTBF de 406, este dato nos indica que una vez que ocurre una falla, se espera que ocurra la siguiente falla después de 406 horas. En este caso, las mediciones de tiempo son en tiempo real y no el tiempo de ejecución como el MTTF.

5.6.3. Facilidad de uso.

Es el nivel de facilidad con la cual se utiliza la aplicación de acuerdo a los siguientes sub atributos: facilidad de comprensión, facilidad de aprendizaje y operabilidad.

Según ISO/IEC 9126-4, recomienda que las métricas de usabilidad deberían incluir:

- Eficacia: La precisión y la integridad con la que los usuarios logran objetivos específicos.

Eficacia, puede ser calculada midiendo la tasa de finalización. Conocida como la métrica de usabilidad fundamental, la tasa de finalización se calcula asignando un valor binario de '1' si el participante de la prueba logra completar una tarea y '0' si no lo hace. Debido a su simplicidad la tasa de finalización es una métrica que es fácil de entender, la efectividad se puede representar como un porcentaje usando esta simple ecuación:

$$Eficacia = \frac{\text{Número de tareas completadas satisfactoriamente}}{\text{Total número de tareas realizadas}} \times 100\% \quad (4)$$

Tomando en cuenta ecuación mostrada, se plantea el siguiente caso para la evaluación del sistema de predicción:

Diez usuarios deben hacer uso de una funcionalidad del sistema, “Evaluación del IMC”. Luego de realizar la tarea, sólo 8 de los 10 usuario lograron completar satisfactoriamente el objetivo.

Aplicando la ecuación:

$$Eficacia = \frac{8}{10} \times 100\%$$

Como resultado para la evaluación de la funcionalidad del sistema se tiene un 80% de eficacia.

Validación con usuarios:

Para esta actividad se evaluaron un total de 10 usuarios no expertos. La tarea consiste en registrarse correctamente en el sistema, por medio del formulario inicial; posteriormente acceder a la tarea principal “Registrar evaluación”. Completar correctamente el formulario, con los datos: Género, Fecha de nacimiento, altura, peso, circunferencia abdominal, estatura sentado, longitud de pierna, si desayuno todos los días y si hace deporte. Los resultados obtenidos son los siguientes:

Tabla 6

Validación con usuarios. Fuente: Propia

Usuarios	Eficacia			
	¿Ingresó datos?	¿Accedió al sistema?	¿Completó la evaluación?	Errores cometidos
Karen Ticona	Si	Si	Si	0
Ruby Sucari	No	No	No	3
Joselyn Perlacios	Si	Si	Si	0
Ada Sucari	Si	Si	Si	0
Katty Givera	Si	Si	Si	0
Deyby Bedoya	Si	Si	Si	0
Steven Zegarra	Si	Si	Si	0
Christian Rojas	Si	Si	Si	0
Ramiro Álvarez	Si	Si	Si	0
Albert Pacco	Si	Si	No	1

5.6.4. Eficiencia.

Es el nivel con el que la aplicación emplea en forma óptima los recursos del sistema, de acuerdo a los siguientes sub atributos: comportamiento en el tiempo, comportamiento de recursos.

En este caso, para realizar la evaluación de la eficiencia, se utilizará la herramienta JMeter, la cual ofrece beneficios en el testing de rendimiento, incluye las siguientes evaluaciones: pruebas de carga y pruebas de estrés. Con esta herramienta podremos descubrir el máximo número de usuarios concurrentes que el sitio web puede soportar y nos proveerá una variedad de análisis gráfico de reportes de rendimiento.

Una vez instalada la herramienta, será necesario añadir un ThreadGroup, una vez abierta esta nueva ventana se procederá a ingresar las propiedades como en la siguiente imagen:

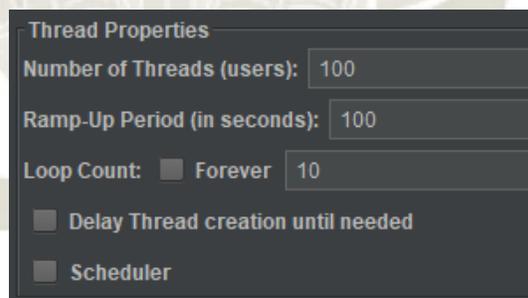
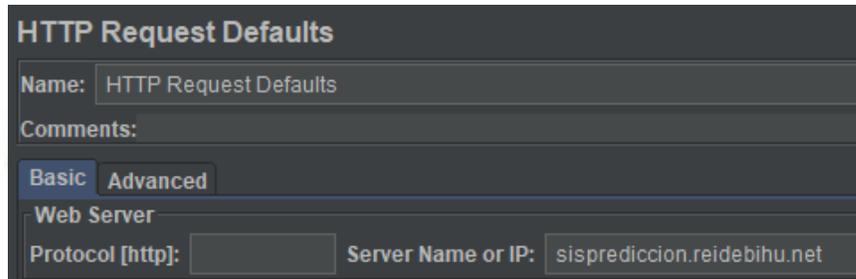


Figura 67. Panel de control Thread Group.

- Number of Threads: Número de usuario conectados al sitio web objetivo 100.
- Loop Count: Número de veces para ejecutar las pruebas.
- Ramp-Up Period: 100.

Posteriormente pasaremos a añadir un HTTP request default, configurando los parámetros de la siguiente forma:



HTTP Request Defaults

Name: HTTP Request Defaults

Comments:

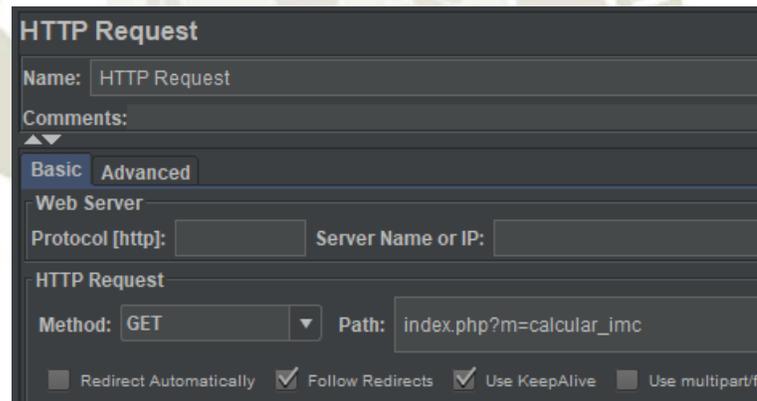
Basic Advanced

Web Server

Protocol [http]: Server Name or IP: sisprediccion.reidebihu.net

Figura 68. Panel de control de HTTP Request Defaults.

Luego, se agrega un HTTP Request, configuramos la propiedad Path de la siguiente forma:



HTTP Request

Name: HTTP Request

Comments:

Basic Advanced

Web Server

Protocol [http]: Server Name or IP:

HTTP Request

Method: GET Path: index.php?m=calcular_imc

Redirect Automatically Follow Redirects Use KeepAlive Use multipart/form-data

Figura 69. Panel de control de HTTP Request.

Finalmente, para visualizar los resultados de una forma gráfica añadiremos un Graph result. Esto nos ayudará a realizar análisis y conclusiones de los resultados.

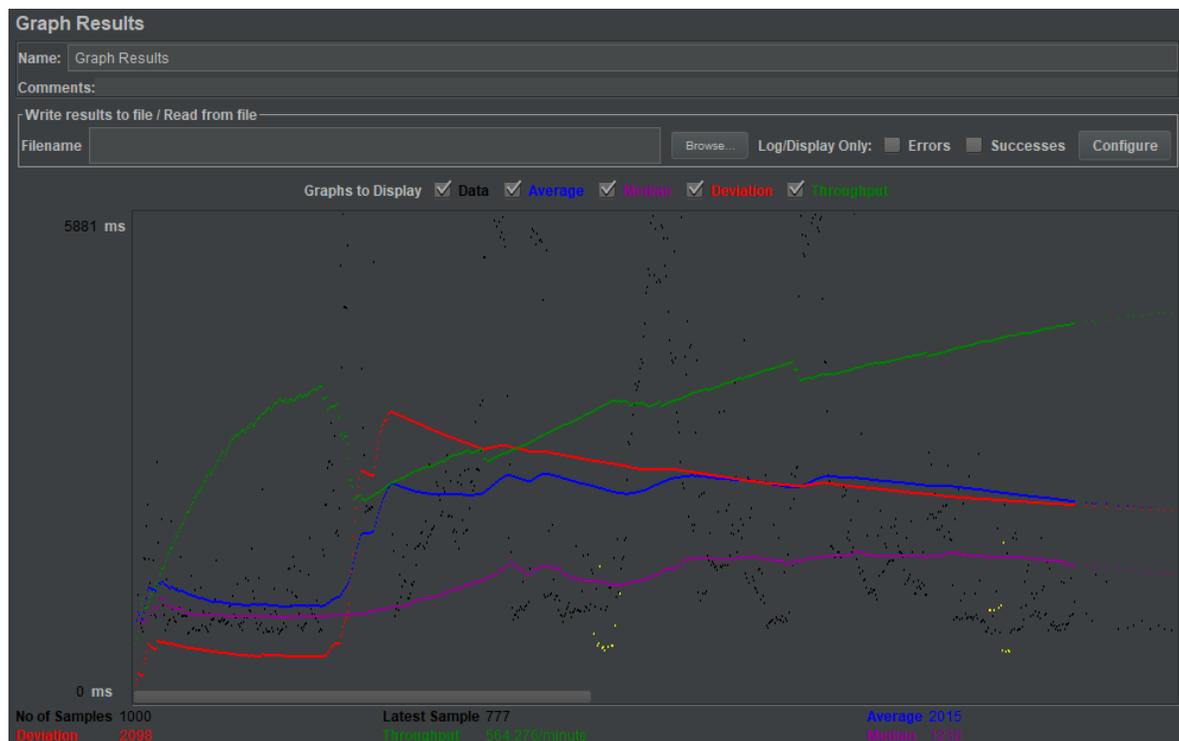


Figura 70. Gráfico del plan de pruebas.

Para analizar el rendimiento, del sitio web, se enfocará en 2 parámetros: Throughput y Deviation (Rendimiento y Desviación respectivamente).

El rendimiento es el parámetro más importante, representa la habilidad del servidor para manejar carga pesada. Mientras más alto, mejor. En este caso, como valor resultado se obtuvo 564.276/minuto. Esto significa que nuestro servidor Hostinger, puede manejar 564.276 pedidos por minuto. Este valor es relativamente medio. Así que se puede concluir que el servidor tiene un rendimiento medio.

La desviación se muestra de color rojo, en el gráfico, e indica la desviación desde el promedio. Mientras más bajo, mejor. En esta prueba se obtuvo una desviación de 2098, así que podemos determinar que el rendimiento del sitio web es medio.

5.6.5. Facilidad de mantenimiento.

Es el nivel de facilidad con que se recompone la aplicación de acuerdo a los siguientes sub atributos: facilidad de análisis, facilidad de cambio, estabilidad, y facilidad de testeo.

No existe una sola métrica para indicar si una aplicación es más fácil de mantener que la otra y no existe una sola herramienta que pueda analizar el depósito de código y proporcionarle una respuesta precisa. No hay sustituto para un revisor humano, pero incluso los humanos no pueden analizar los repositorios de códigos completos para dar una respuesta definitiva. Se necesita cierta cantidad de automatización.

Las métricas más importantes a ser utilizadas en el testing de facilidad de mantenimiento son:

- Capacidad de prueba: Qué parte de la aplicación se está probando.
- Comprensibilidad: La legibilidad del código.
- Modificabilidad: Simplicidad estructural y de diseño.
- Requisito para el mapeo de implementación: ¿qué tan fácil es decir "qué" debe hacer la aplicación y correlacionarla con "cómo" se está haciendo?

No existe una medida única que pueda capturar con precisión la noción de mantenimiento de una aplicación. Existen métricas compuestas como índice de mantenimiento (MI) que ayudan a predecir la mantenibilidad de la aplicación utilizando el volumen de Halstead, la complejidad ciclomática, el SLOC total (líneas de código de origen) y la relación de comentarios:

$$MI = \max\left[0, 100x^{\frac{171-5.2 \ln V - 0.23G - 16.2 \ln L + 50 \sin \sqrt{(2.4C)}}{171}}\right] \quad (5)$$

Dónde:

- V es el volumen medio de Halstead por módulo. Para nuestro caso: 462.27
- G es la complejidad ciclomática promedio por módulo. Para nuestro caso: 11
- L es el número promedio de líneas de código fuente (SLOC) por módulo. Para nuestro caso: 290.5
- C es el número promedio de líneas de comentario por módulo. Para nuestro caso: 23

Antes de hallar el volumen de Halstead, se deben obtener ciertos indicadores que son los que se muestran en la siguiente tabla:

Tabla 7

Indicadores para evaluar la complejidad del módulo.

Parámetro	Significado	Valores
n1	Número de un solo operador	17
n2	Número de un solo operando	26
N1	Número total de casos de operadores	139
N2	Número total de casos de operandos	143

$$\begin{aligned} \text{Volume: } V &= N \log_2(n) & (6) \\ V &= 283 \log_2(43) \end{aligned}$$

$$V = 462.27$$

Reemplazamos los valores correspondientes en la fórmula 5:

$$MI = \max\left[0, 100x \frac{171 - 5.2 \ln 462.27 - 0.23 * 11 - 16.2 \ln * 290.5 + 50 \sin \sqrt{(2.4 * 23)}}{171}\right]$$
$$MI = 52.77$$

Luego realizar los cálculos y operaciones, se obtiene un índice de facilidad de mantenimiento igual a 52.77, en una evaluación en un rango del 1 al 100, la evaluación es relativamente media, y se puede decir que los costos para realizar una mantención del código son medios.

5.6.6. Portabilidad.

Nivel de facilidad con que se lleva la aplicación de un entorno a otro, de acuerdo a los siguientes sub atributos: adaptabilidad, facilidad para ser instalado, cumplimiento, y facilidad para ser reemplazado.

Ya que la aplicación se encuentra desarrollada en una versión web, de por sí ya cuenta con varios de los sub atributos que requiere la portabilidad, para nuestro caso se realizará una prueba, que consiste en determinar la vista que muestra la aplicación en varios navegadores tales como Mozilla, Internet Explorer, Opera y Safari.

- Mozilla.

The screenshot shows a Mozilla browser window displaying the 'Sistema de Predicción' interface. The main content is a form titled 'EVALUACION IMC'. At the top, it asks for 'Eres:' with radio buttons for 'Hombre' and 'Mujer'. Below this are several questions with input fields and 'Ver video' buttons: '¿Cuál es su fecha de nacimiento?' (with a date picker), '¿Cuánto mide?' (with a height input), '¿Cuánto pesa?' (with a weight input), '¿Cuánto es su circunferencia abdominal?' (with a waist circumference input), '¿Cuánto es su estatura ventral?' (with a chest height input), '¿Cuál es la longitud de su pierna?' (with a leg length input), '¿Desayuna todos los días?' (with radio buttons for 'Si' and 'No'), and '¿Hace deporte?' (with radio buttons for 'Si' and 'No').

Figura 71. Navegador Mozilla - Sisprediccion.

- Internet Explorer.

The screenshot shows an Internet Explorer browser window displaying the 'Sistema de Predicción' interface. The main content is a form titled 'CUESTIONARIO ACTITUD HACIA LA OBESIDAD'. It includes a date picker for 'Fecha de nacimiento' and radio buttons for 'Género: Hombre' and 'Mujer'. Below this, there is a legend: 'T.A. = Totalmente de acuerdo; NA-ND = Ni acuerdo, ni desacuerdo; T.D. = Totalmente desacuerdo.' The form contains a table with 4 rows of questions and 3 columns of radio buttons for 'T.A.', 'NA-ND', and 'T.D.':

Nro	Pregunta	T.A.	NA-ND	T.D.
1	Con frecuencia prefiero quedarme en casa, que salir a jugar al parque.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	A menudo no me llama la atención mucho hacer ejercicios físicos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Me agrada salir a reuniones sociales con mis amigos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Más siento a mucha molestia en la noche más en mi casa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

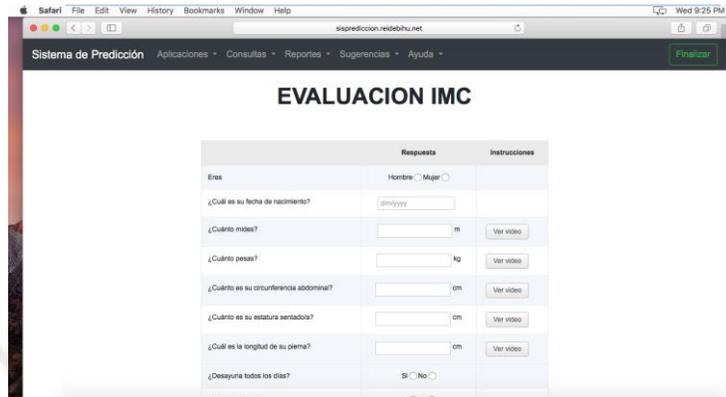
Figura 72. Navegador Internet Explorer - Sisprediccion.

- Opera

The screenshot shows an Opera browser window displaying the 'Sistema de Predicción' interface. The main content is a form titled 'EVALUACION IMC'. At the top, it asks for 'Eres:' with radio buttons for 'Hombre' and 'Mujer'. Below this are several questions with input fields and 'Ver video' buttons: '¿Cuál es su fecha de nacimiento?' (with a date picker), '¿Cuánto mide?' (with a height input), '¿Cuánto pesa?' (with a weight input), '¿Cuánto es su circunferencia abdominal?' (with a waist circumference input), '¿Cuánto es su estatura ventral?' (with a chest height input), '¿Cuál es la longitud de su pierna?' (with a leg length input), '¿Desayuna todos los días?' (with radio buttons for 'Si' and 'No'), and '¿Hace deporte?' (with radio buttons for 'Si' and 'No').

Figura 73. Navegador Opera - Sisprediccion.

- Safari



	Respuesta	Instrucciones
Eres	Hombre <input type="radio"/> Mujer <input type="radio"/>	
¿Cuál es su fecha de nacimiento?	<input type="text" value="dd/yyyy"/>	
¿Cuánto mides?	<input type="text"/> m	Ver video
¿Cuánto pesas?	<input type="text"/> kg	Ver video
¿Cuánto es su circunferencia abdominal?	<input type="text"/> cm	Ver video
¿Cuánto es su estatura sentada?	<input type="text"/> cm	Ver video
¿Cuál es la longitud de su pierna?	<input type="text"/> cm	Ver video
¿Desayuna todos los días?	Si <input type="checkbox"/> No <input type="checkbox"/>	

Figura 74. Navegador Safari - Sisprediccion.

De 4 navegadores utilizados para realizar las pruebas, en los 4 la aplicación se adaptó correctamente, este es un buen indicador que nos demuestra que la aplicación se puede adaptar en diversos ambientes, cumpliendo con todas sus funcionalidades.

CAPITULO 6: RESULTADOS

Es necesario recapitular a los objetivos planteados al inicio del proyecto para exponer los resultados, como objetivo principal se tiene el crear una aplicación que haga uso de técnicas de minería de datos para así poder predecir el estado de peso de un usuario (en este caso adolescente), además diagnosticar su estado de peso actual. Los atributos que se determinaron relevantes en los modelos son: Edad, Sexo, ¿Hace Deporte?, ¿Desayuna todos los días?, Peso, Estatura parado, Estatura sentado, Circunferencia abdominal, Pico de Velocidad de Crecimiento (PVC), Porcentaje grasa (%), Masa grasa y Masa libre de grasa (Masa magra).

6.1.Resultados de los algoritmos de clasificación

En este punto se muestran los resultados obtenidos de los algoritmos de clasificación de J48, BayesNet, MultilayerPerceptron, ForestPA y NaiveBayes para que posteriormente sean evaluados y saber cuál es el más adecuado para el presente caso.

En cuanto a las cantidades de data analizadas y clasificadas, 660 registros de entrenamiento fueron usados; 10% fueron utilizados para la fase de entrenamiento y 90% para la fase de prueba, con la herramienta Weka, y el tipo de prueba Cross Validation.

Es necesario saber que la cantidad de verdaderos positivos (TP) es equivalente a la cantidad de ejemplos que son verdaderos positivos y los falsos negativos (FP) son el número de ejemplos falsos negativos encontrados.

Para la evaluación se utilizarán los criterios de precisión y Recall, que pueden ser hallados con la siguiente fórmula.

$$Precision = \frac{TP}{TP + FP} \quad (Cruz \ \& \ Wishart, \ 2006) \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \text{ (Zhang et al., 2009)} \quad (8)$$

Donde FN es la cantidad de ejemplos que son falsos negativos detectados. A continuación, los resultados de cada algoritmo serán mostrados en las siguientes tablas.

a) Árbol de decisión J48

Tabla 8

Resultados del clasificador Árbol de decisión J48. Fuente: Propia

Clase	Tasa TP	Tasa FP	Precisión	Área ROC
Peso bajo	0.881	0.011	0.941	0.950
Normal	0.969	0.094	0.949	0.949
Sobrepeso	0.898	0.014	0.917	0.964
Obesidad	0.964	0.002	0.964	0.981
Instancias correctamente clasificadas			623	94.39%
Instancias incorrectamente clasificadas			37	5.60%

b) Clasificador BayesNet

Tabla 9

Resultados del clasificador BayesNet. Fuente: Propia

Clase	Tasa TP	Tasa FP	Precisión	Área ROC
Peso bajo	0.780	0.082	0.654	0.951
Normal	0.838	0.179	0.894	0.921
Sobrepeso	0.765	0.060	0.688	0.956
Obesidad	0.643	0.008	0.783	0.979
Instancias correctamente clasificadas			534	80.90%
Instancias incorrectamente clasificadas			126	19.09%

c) Clasificador MultilayerPerceptron

Tabla 10

Resultados del clasificador MultilayerPerceptron. Fuente: Propia

Clase	Tasa TP	Tasa FP	Precisión	Área ROC
Peso bajo	0.890	0.024	0.882	0.989
Normal	0.955	0.085	0.953	0.985
Sobrepeso	0.898	0.020	0.889	0.984
Obesidad	0.857	0.002	0.960	0.998
Instancias correctamente clasificadas			615	93.18%
Instancias incorrectamente clasificadas			45	6.81%

d) Clasificador ForestPA

Tabla 11

Resultados del clasificador ForestPA. Fuente: Propia

Clase	Tasa TP	Tasa FP	Precisión	Área ROC
Peso bajo	0.862	0.007	0.959	0.979
Normal	0.979	0.106	0.943	0.977
Sobrepeso	0.888	0.016	0.906	0.987
Obesidad	0.857	0.002	0.960	0.999
Instancias correctamente clasificadas			621	94.09%
Instancias incorrectamente clasificadas			39	5.9%

e) Clasificador NaiveBayes

Tabla 12

Resultados del clasificador NaiveBayes. Fuente: Propia

Clase	Tasa TP	Tasa FP	Precisión	Área ROC
Peso bajo	0.734	0.083	0.635	0.946
Normal	0.835	0.204	0.881	0.914
Sobrepeso	0.735	0.062	0.673	0.955
Obesidad	0.607	0.011	0.708	0.987
Instancias correctamente clasificadas			524	79.39%
Instancias incorrectamente clasificadas			136	20.6%

6.2.Evaluación y comparación de los modelos construidos

En este punto se realiza una comparación de los 5 clasificadores seleccionados, en base a las métricas (Ratios de Verdaderos Positivos, Ratios de Falsos Positivos, Precisión y Área bajo curva ROC). (Suca et al., 2016).

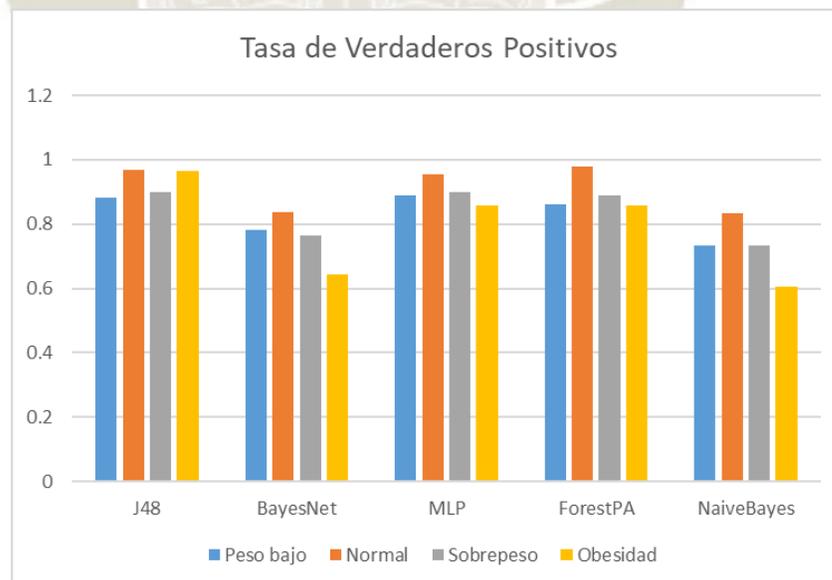


Figura 75. Resultados de Ratios de Verdaderos Positivos.

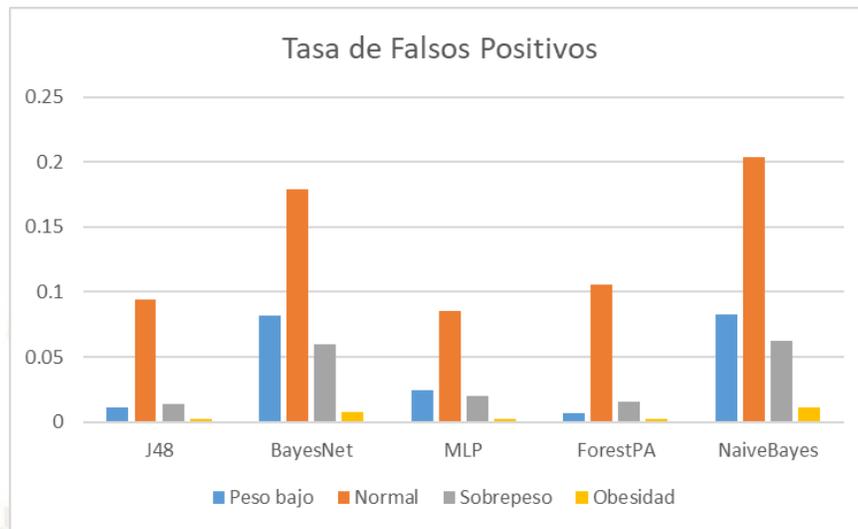


Figura 76. Resultados de Ratios de Falsos Positivos.

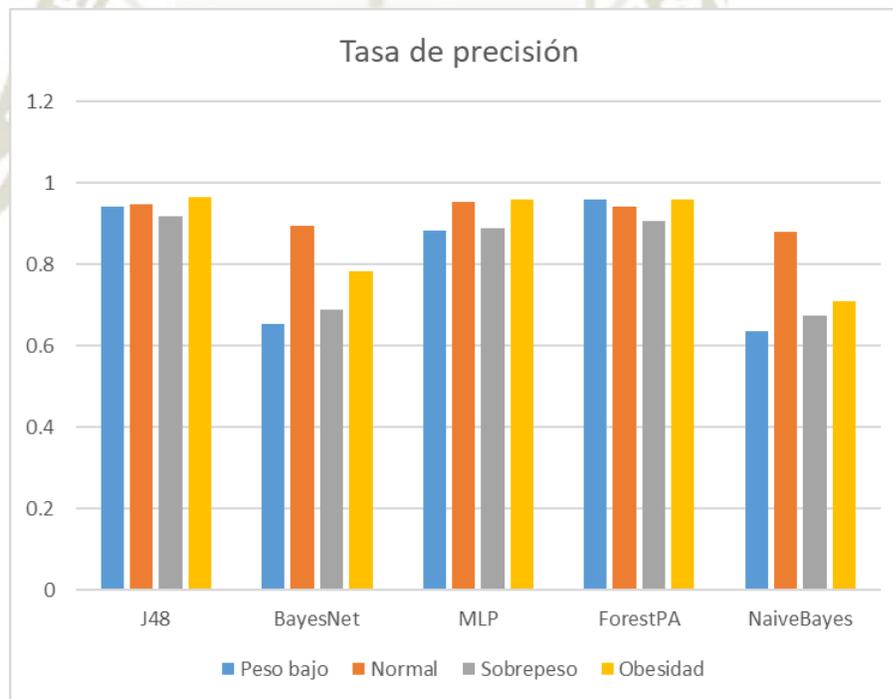


Figura 77. Resultados de Ratios de Valores de Precisión.

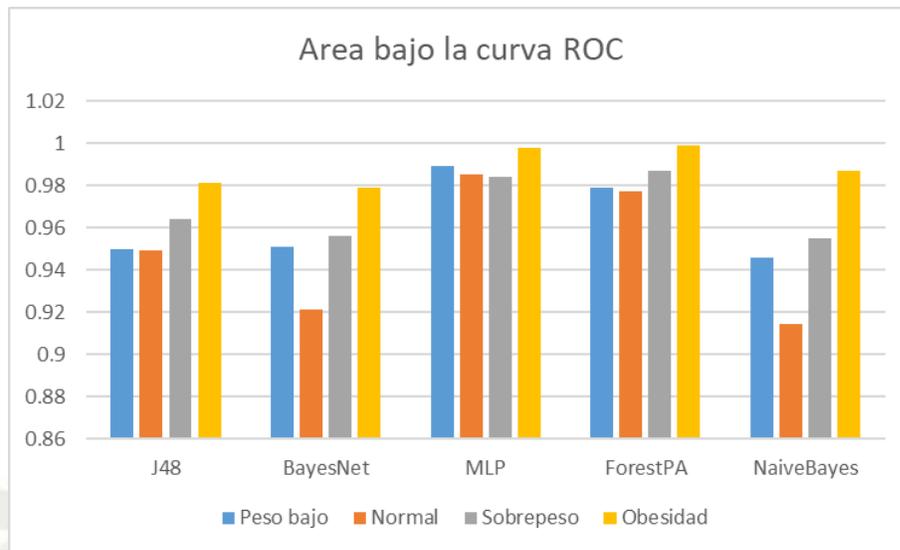


Figura 78. Resultados de Ratios de Área bajo la curva ROC.

6.3. Análisis de resultados

Como se puede observar en los gráficos, la mayor parte de los resultados obtenidos nos dirigen a la elección entre 2 clasificadores de árboles de decisión, J48 y ForestPA. A pesar de que en otras investigaciones BayesNet y NaiveBayes se consideran como clasificadores muy aceptados para diagnósticos médicos, para nuestro caso obtienen bajos niveles en cuanto métricas, lo más probable es que haya sido por la dependencia de variables que requiere, por tal motivo no obtuvo los mejores resultados. Por otro lado, el algoritmo MultilayerPerceptron obtuvo valores muy acertados en cuanto a la realidad, en ciertos indicadores obtuvo mejor puntaje que J48.

6.3.1. Selección del mejor modelo de clasificación

De acuerdo a los datos y resultados obtenidos de las evaluaciones realizadas para cada algoritmo, la técnica que debe ser seleccionada y que obtuvo los más altos valores en base a las métricas seleccionadas es J48.

J48 pruned tree

```

-----
Masa Grasa <= 17.488602
| %Grasa <= 9.358596: PESO BAJO (58.0)
| %Grasa > 9.358596
| | Peso <= 46.8
| | | %Grasa <= 20.248711
| | | | Género = F
| | | | | %Grasa <= 20.047507: PESO BAJO (37.0)
| | | | | %Grasa > 20.047507
| | | | | | Edad <= 14.90411: NORMAL (2.0)
| | | | | | Edad > 14.90411: PESO BAJO (2.0)
| | | | | Género = M: NORMAL (6.0)
| | | | %Grasa > 20.248711: NORMAL (33.0)
| | Peso > 46.8
| | | Peso <= 66.1
| | | | %Grasa <= 28.150971
| | | | | %Grasa <= 9.903285
| | | | | | Edad <= 15.334247: NORMAL (4.0)
| | | | | | Edad > 15.334247: PESO BAJO (6.0/1.0)
| | | | | %Grasa > 9.903285
| | | | | | Circunferencia Abdominal (CM) <= 68
| | | | | | | Estatura Parado(m) <= 1.63: NORMAL (30.0)
| | | | | | | Estatura Parado(m) > 1.63
| | | | | | | Género = F
| | | | | | | | Circunferencia Abdominal (CM) <= 66: NORMAL (2.0)
| | | | | | | | Circunferencia Abdominal (CM) > 66: PESO BAJO (5.0)
| | | | | | | Género = M: NORMAL (5.0)
| | | | | | | | Circunferencia Abdominal (CM) > 68: NORMAL (328.0/6.0)
| | | | | %Grasa > 28.150971: SOBREPESO (10.0/2.0)
| | | | Peso > 66.1
| | | | | %Grasa <= 17.206416: NORMAL (13.0)
| | | | | %Grasa > 17.206416: SOBREPESO (27.0/1.0)
Masa Grasa > 17.488602
| %Grasa <= 33.95511
| | Género = F
| | | %Grasa <= 28.065086: NORMAL (5.0/1.0)
| | | %Grasa > 28.065086: SOBREPESO (51.0)
| | Género = M
| | | %Grasa <= 24.066204: SOBREPESO (8.0)
| | | %Grasa > 24.066204: OBESO (15.0)
| %Grasa > 33.95511: OBESO (13.0)

```

Number of Leaves : 21

Size of the tree : 41

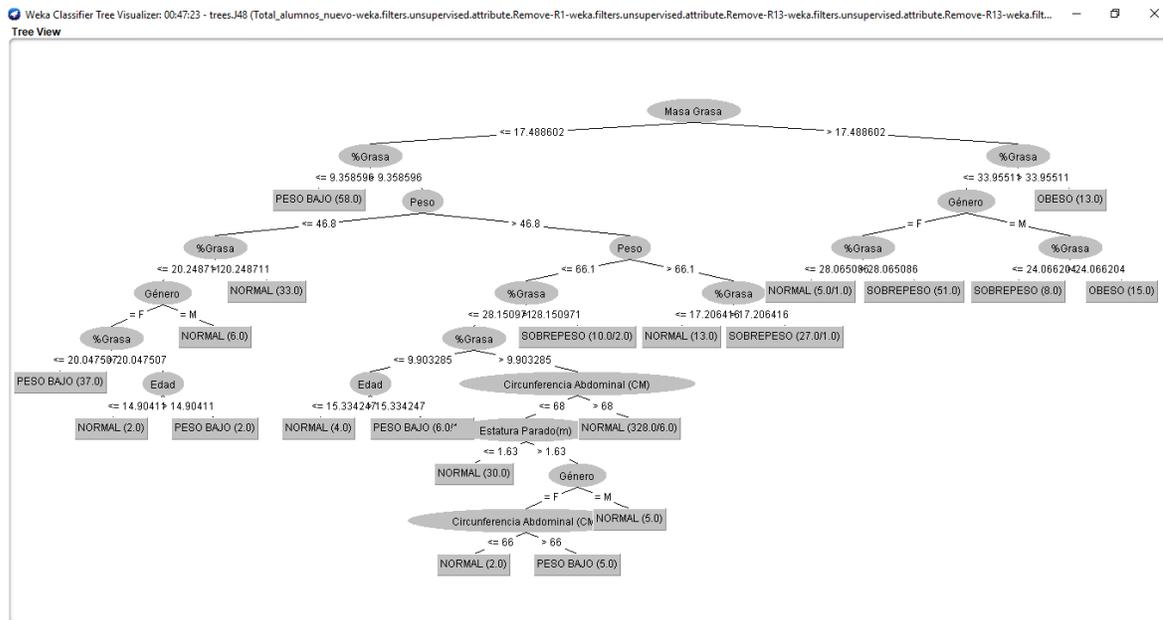


Figura 79. Árbol J48 generado en Weka.

En cuanto a valores de predicción positivos, J48 obtuvo una media de 0.928 comparada con los valores que obtuvieron BayesNet, MultilayerPerceptron, ForestPA y NaiveBayes, es superior.

En la figura 79, se muestra el árbol generado por el algoritmo seleccionado J48, este es el árbol implementado en el sistema de predicción de obesidad.

Se hizo la comparación de algoritmos de minería de datos y aprendizaje automático, obteniendo resultados que en su momento fueron analizados para darles la interpretación necesaria y ser implementados en el desarrollo de la herramienta de software. Como mejor algoritmo se obtuvo a J48 que pertenece a la técnica de árboles de decisiones, con una precisión de 94.39%.

Un punto importante es que se debe tener en consideración, es la cantidad de información que se utilizó para la elaboración de los modelos de predicción, en el presente trabajo se utilizó 660 registros de personas en un rango de edad de 5 a 17 años, en la fase de entrenamiento del modelo; sin embargo, se considera que esta cantidad podría ser ampliada en un futuro para volver a entrenar los modelos y ver si se consigue resultados más óptimos. Una vez seleccionado el algoritmo J48, se implementó en la herramienta de software. Otro punto importante que resaltar, es la implementación del cuestionario de actitud hacia la obesidad (Gómez et al., 2015), este contribuye a brindar mayor información al usuario sobre su estado actual y nivel de actitud hacia la obesidad.

Se lograron conseguir todos los objetivos planteados, desde la revisión de los antecedentes relacionados al proyecto, selección del mejor algoritmo hasta el completo desarrollo e implementación de la herramienta de software.

Tabla 13

Resultados de clasificación y comparación. Fuente: Propia

Clasificador	Instancias	Atributos	Resultado
J48 (Mca, 2010).	3414	39	99.95%
J48 (Tello, Eslava, & Tobías, 2013).	1000	10	51.3%
J48 (Kumar, 2012).	600	12	89.33%
J48 (Presente investigación)	660	12	94.39%

CONCLUSIONES

PRIMERA:

Se hizo la revisión necesaria de artículos e investigaciones relacionadas al tema, obteniendo la información requerida para la correcta realización del proyecto. La documentación revisada proviene de fuentes confiables, plataformas como Web of Science, IEEE y Springer.

SEGUNDA:

La información recolectada de colegios de Arequipa fue consolidada, preparada, limpiada y procesada con el fin de utilizarla para el entrenamiento de los modelos de predicción.

TERCERA:

Se logró obtener las variables más influyentes en el modelo de predicción, luego de la realización de numerosas pruebas en la herramienta WEKA. Al seleccionar las variables o atributos, la precisión del algoritmo mejoró, y así fue quedando el que mejor se adaptaba al objetivo.

CUARTA:

Se compararon las técnicas de minería de datos, en total fueron tres técnicas: Árboles de decisión, Redes Bayesianas y Redes Neuronales, en cuanto a clasificadores se compararon los algoritmos: J48, BayesNet, Multilayer Perceptron, ForestPA y NaiveBayes. La técnica que obtuvo mayor precisión y obtuvo mejores resultados en cuanto a las ratios evaluados fue Árboles de decisión. La selección previa de las variables y/o atributos fue crucial para la obtención de este objetivo.

QUINTA:

El mejor modelo obtenido J48(94.39%) fue entrenado a partir del análisis de la comparación con otros modelos de minería de datos. El modelo MultilayerPerceptron obtuvo un puntaje de 93.18% de precisión, Redes Neuronales, presenta un buen nivel de precisión y predicción, quizás con el análisis de mayores registros de información, habría obtenido mejores resultados; Se esperaba que la técnica BayesNet y NaiveBayes obtenga mejores resultados, de acuerdo a la documentación revisada este se acomoda mejor en cuanto a temas de salud.

SEXTA:

Se logró implementar el algoritmo con mejor resultado en una herramienta de software, al alcance de toda persona con conexión a internet. Por otro lado, se desarrollaron funciones adicionales al sistema, que complementan el objetivo general del proyecto.

SEPTIMA:

Se hizo la evaluación de calidad del software, tomando como referencia el estándar ISO 9126, que se compone por cuatro atributos principales: Funcionalidad, Eficiencia, Confiabilidad, Facilidad de uso, Facilidad de mantenimiento y Portabilidad

OCTAVA:

De acuerdo a la comparación con los resultados obtenidos en otras investigaciones, utilizando técnicas de minería de datos con el árbol de decisión J48, el nivel de instancias correctamente clasificadas obtenido por el presente trabajo, es aceptable y puede brindar conocimiento en la clasificación de nuevas instancias.

RECOMENDACIONES

Para mejorar los resultados, en futuras investigaciones se debe contar de ante mano con una cantidad mayor de datos, suficiente para demostrar y observar cambios en el entrenamiento de modelos.

En futuros trabajos se puede probar con otros algoritmos y/o técnicas de minería de datos y aprendizaje automático. Y así contrastar los resultados con los obtenidos en este proyecto.

Si se hacen pruebas con un mayor número de muestras que las utilizadas en este proyecto, resultaría más ventajoso probar variando los tamaños de los conjuntos de entrenamiento.

Se podría construir e implementar nuevas funcionalidades a la herramienta de software, el desarrollo de software es un proceso que debe estar en mejoras continuas, y no quedar en un determinado punto.

REFERENCIAS

- Abdullah, F. S., Manan, N. S. A., Ahmad, A., Wafa, S. W., Shahril, M. R., Zulaily, N., ... & Ahmed, A. (2016, August). Data Mining Techniques for Classification of Childhood Obesity Among Year 6 School Children. In International Conference on Soft Computing and Data Mining (pp. 465-474). Springer, Cham.
- Adnan, Md Nasim & Islam, Md. (2017). Forest PA: Constructing a Decision Forest by Penalizing Attributes used in Previous Trees. Expert Systems with Applications. 89. 10.1016/j.eswa.2017.08.002.
- Adnan, M. H. M., & Husain, W. (2012). Hybrid approaches using decision tree, naïve bayes, means and euclidean distances for childhood obesity prediction. International Journal of Software Engineering and Its Applications, 6(3), 99-106.
- Abran, A., Moore, J. W., Bourque, P., Dupuis, R., & Tripp, L. L. (2004). Software engineering body of knowledge. IEEE Computer Society, Angela Burgess.
- Alliyarova ST, Kainarbayeva MS, Khassenova G, Chuyenbekova AB, Ushanskaya YY, Bykybayeva SA, Kozhakhmetova AN y Bakirova MA (2016). Selection Of Obesity Prediction Attributes Among Adults Via Data Mining Application. 7(2), 2011-2019.
- Al Mamun, A., Cramb, S., O'Callaghan, M., Williams, G., & Najman, J. (2009). Childhood Overweight Status Predicts Diabetes at Age 21 Years: A Follow-up Study. Obesity. <http://dx.doi.org/10.1038/oby.2008.660>

- BBC Mundo (2017). El impresionante aumento del sobrepeso y la obesidad en América Latina. BBC. Recuperado el 30 de enero 2018 de <http://www.bbc.com/mundo/noticias-38693438>
- Beati, H. (2015). PHP-Creación de páginas Web dinámicas 2a edición. Alfaomega Grupo Editor.
- Bellman, R. (1978). An introduction to artificial intelligence: Can computers think?. Thomson Course Technology.
- Charniak, E., & McDermott, D. (1985). Introduction to AI. Reading (Mass.): Addison.
- Christopher, M. B. (2016). Pattern Recognition and Machine Learning. Springer-Verlag New York.
- Ciangura, C., Czernichow, S., & Oppert, J. (2010). Obesidad. EMC - Tratado De Medicina, 14(1), 1-9. [http://dx.doi.org/10.1016/s1636-5410\(10\)70517-1](http://dx.doi.org/10.1016/s1636-5410(10)70517-1)
- Cossio-Bolaños, Marco & Vasquez, Pablo & Luarte, Cristian & Sulla-Torres, Jose & Gomez-Campos, Rossana. (2016). Evaluación de la autopercepción de la aptitud física y propuesta de normativas en adolescentes escolares chilenos: estudio EAPAF. Archivos argentinos de pediatría. 114. 319-328.
- Cossio-Bolaños M, de Arruda M, Sulla Torres J, Urra Alborno C & Gómez Campos R (2017). Desarrollo de ecuaciones y propuesta de valores referenciales para estimar la masa grasa de niños y adolescentes chilenos. Arch Argent Pediatr. 115(5), 453-461. <http://dx.doi.org/10.5546/aap.2017.453>

Cruz, J., & Wishart, D. (2006). Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*, 2, 117693510600200. doi: 10.1177/117693510600200030

Fernández, C. (2017). El 30 por ciento de la población mundial tiene exceso de peso. El Tiempo. Recuperado el 4 de octubre 2017, de <http://www.eltiempo.com/vida/salud/personas-con-obesidad-y-sobrepeso-en-el-mundo-segun-informe-97900>

Gallego, M. T. (2012). Metodología Scrum. Universitat Oberta de Catalunya.

Gómez-Campos, Rossana, Bersano Benavides, Gonzalo, Luarte-Rocha, Cristian, Nayer Tumi Figueroa, Ernesto, Urra Albornoz, Camilo, Sulla-Torres, Jose, & Cossio-Bolaños, Marco Antonio. (2016). Validación y propuesta normativa para la valoración de la preocupación por la alimentación de adolescentes chilenos: estudio EPACH. *Nutrición Hospitalaria*, 33(2), 337-344. <https://dx.doi.org/10.20960/nh.114>

Gómez C, Rossana, de Arruda, Miguel, Sulla T, Jose, Alvear, Fernando, Urra A, Camilo, González P, Carolina, Luarte R, Cristian, & Cossio B, Marco. (2015). Construcción de un instrumento para valorar la actitud a la obesidad en adolescentes. *Revista chilena de nutrición*, 42(3), 277-283. <https://dx.doi.org/10.4067/S0717-75182015000300008>

Dulhare, U. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Biomedical Research*, 29(12). doi: 10.4066/biomedicalresearch.29-18-620

- Hassanien, A. E., Moftah, H. M., Azar, A. T., & Shoman, M. (2014). MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. *Applied Soft Computing*, 14, 62-71.
- Haugeland, J. (1985). *Artificial intelligence: The very idea*. MIT press.
- Ino, T. (2010). Maternal smoking during pregnancy and offspring obesity: Meta-analysis. *Pediatrics International*, 52(1), 94-99. <http://dx.doi.org/10.1111/j.1442-200x.2009.02883.x>
- Kaur Gill, J. (2017). *Overview and Applications of Artificial Neural Networks*. xenonstack. Recuperado de www.xenonstack.com/blog/data-science/overview-of-artificial-neural-networks-and-its-applications.
- Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4), 261-283.
- Kumar, K. (2012). Knowledge Extraction From Trained Neural Networks. *International Journal of Information and Network Security*, 1(4), 282.
- Maimon, O., & Browarnik, A. (2010). Introduction to Knowledge Discovery and Data Mining. In *Data mining and knowledge discovery handbook* (pp. 1-18). Springer, Boston, MA.
- Marini, E. (2012). El Modelo Cliente/Servidor, 5. <https://www.linuxito.com/docs/el-modelo-cliente-servidor.pdf>

Mca, Judith. (2010). diagnóstico de diabetes utilizando los algoritmos apriori y j48. Renia.

Meng, X., Huang, Y., Rao, D., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung Journal Of Medical Sciences*, 29(2), 93-99. <http://dx.doi.org/10.1016/j.kjms.2012.08.016>

Organización de las Naciones Unidas para la Alimentación y la Agricultura FAO (2017). Perú: sobrepeso afecta a más de la mitad de la población (Informe de FAO/OPS). Recuperado el 30 de enero 2018 de <https://observatoriosocioeclesial.pe/peru-sobrepeso-afecta-a-mas-de-la-mitad-de-la-poblacion-informe-de-faoopa/>

Organización Mundial de la Salud. (2017). Nota descriptiva sobre Obesidad y sobrepeso. Autor. Recuperado el 2 de Noviembre 2017, de <http://www.who.int/mediacentre/factsheets/fs311/es/>

Patil, T. R., & Sherekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications*, 6(2), 256-261.

Pochini, A., Wu, Y., & Hu, G. (2014). Data Mining for Lifestyle Risk Factors Associated with Overweight and Obesity among Adolescents. 2014 IIAI 3Rd International Conference On Advanced Applied Informatics. <http://dx.doi.org/10.1109/iiiai-aai.2014.175>

Raval, K. M. (2012). Data Mining Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(10).

Sevilla Martínez, M. (2017). *Obesidad: causas y soluciones en los adolescentes*. CuidatePlus.

Recuperado el 4 de Octubre 2017, de
<http://www.cuidateplus.com/familia/adolescencia/2016/11/12/dia-mundial-obesidad-causas-soluciones-adolescentes-134635.html>

Sommerville, I. (2011). *Ingeniería del software*. Pearson Educación.

Suca, C., Córdova, A., Condori, A., Cayra, J., & Sulla, J. (2016). *Comparación De Algoritmos De Clasificación Para La Predicción De Casos De Obesidad Infantil*. Universidad Nacional de San Agustín.

Takeyas, B. (2007). *Introduccion a la Inteligencia Artificial*. Instituto Tecnológico de Nuevo Laredo. Nuevo Laredo, Tamps, México. Recuperado el 3 de febrero 2018 de:
<http://www.itnuevolaredo.edu.mx/takeyas/Articulos/Inteligencia%20Artificial/ARTICULO%20Introduccion%20a%20la%20Inteligencia%20Artificial.pdf>

Tello, M. L., Eslava, H. J., & Tobías, L. B. (2013). *Análisis y evaluación del nivel de riesgo en el otorgamiento de créditos financieros utilizando técnicas de minería de datos*. *Visión electrónica*, (1), 13-26.

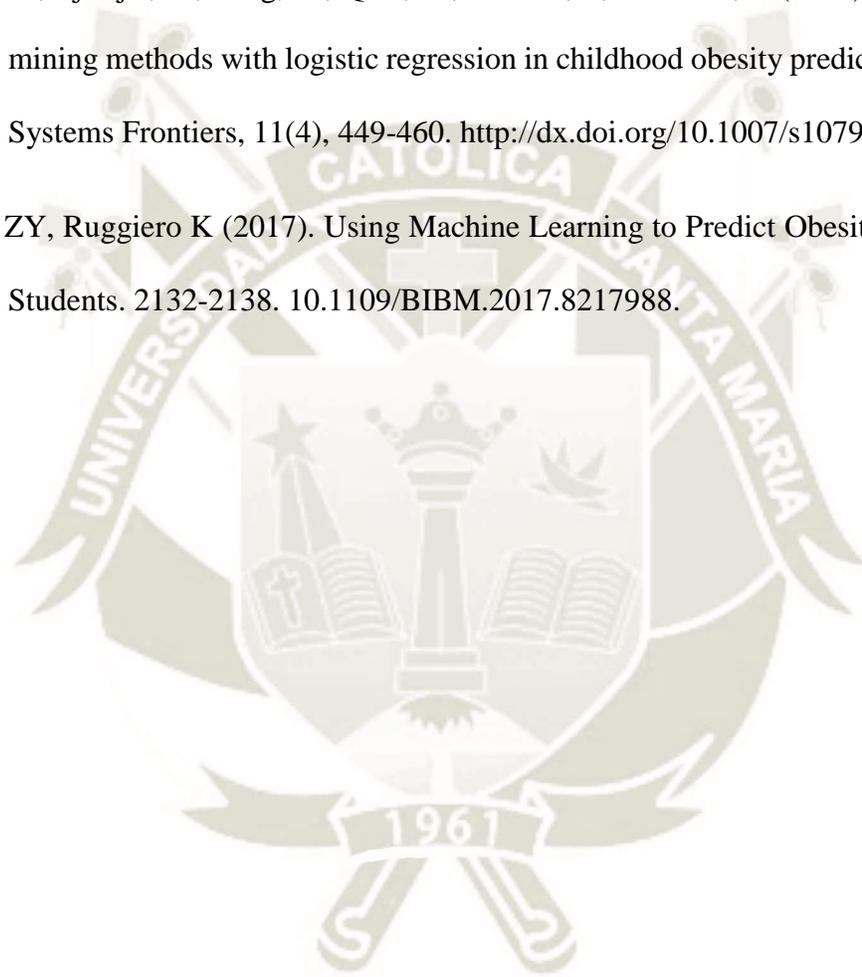
Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S (2017). *Decision trees in epidemiological research*. 14(11). 10.1186/s12982-017-0064-4.

Weber, P., Medina-Oliva, G., Simon, C., & Iung, B. (2012). *Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas*. *Engineering Applications Of Artificial Intelligence*, 25(4), 671-682.
<http://dx.doi.org/10.1016/j.engappai.2010.06.002>

Witten, Ian & Frank, Eibe. (2005). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems).

Zhang, S., Tjortjis, C., Zeng, X., Qiao, H., Buchan, I., & Keane, J. (2009). Comparing data mining methods with logistic regression in childhood obesity prediction. Information Systems Frontiers, 11(4), 449-460. <http://dx.doi.org/10.1007/s10796-009-9157-0>

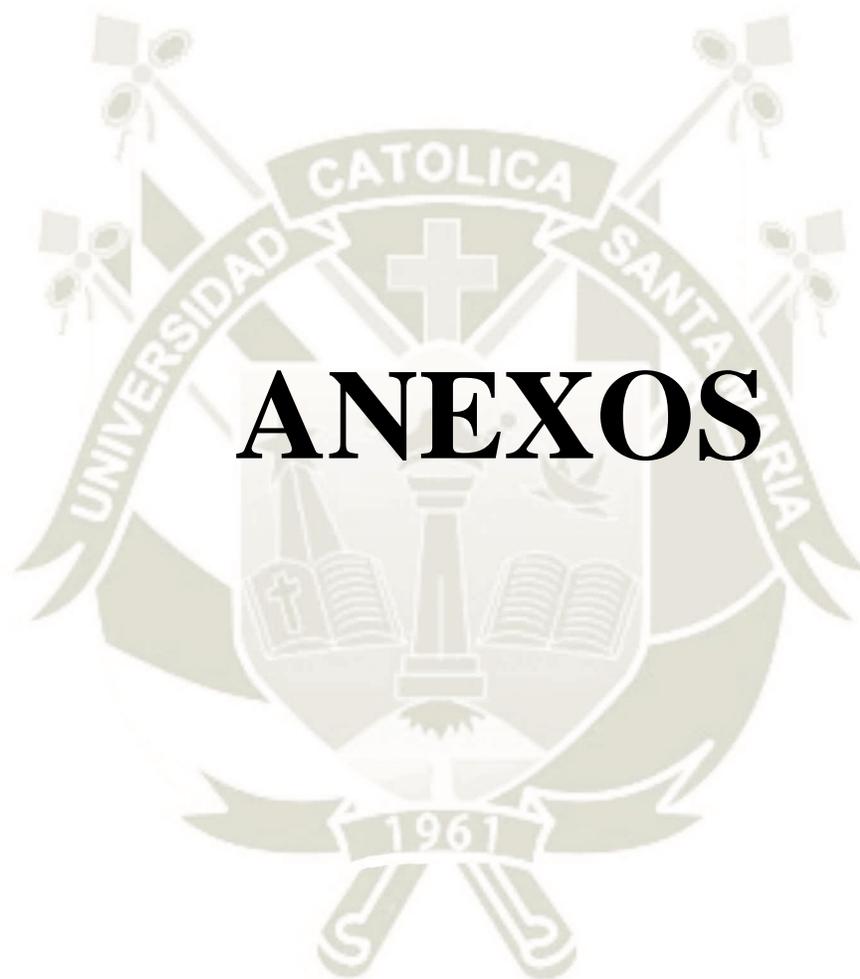
Zheng ZY, Ruggiero K (2017). Using Machine Learning to Predict Obesity in High School Students. 2132-2138. 10.1109/BIBM.2017.8217988.



GLOSARIO DE TERMINOS

TERMINOS	CONCEPTO/DEFINICION
Aprendizaje automático	Es un tipo de inteligencia artificial (AI) que proporciona a las computadoras la capacidad de aprender, sin ser programadas explícitamente. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden cambiar cuando se exponen a nuevos datos.
Sobrepeso	Es un estado premórbido de la obesidad y al igual que ésta se caracteriza por un aumento del peso corporal y se acompaña a una acumulación de grasa en el cuerpo, esto se produce por un desequilibrio entre la cantidad de calorías que se consumen en la dieta y la cantidad de energía (en forma de calorías) que se gasta durante las actividades físicas.
Antropometría	Es una ciencia muy antigua que trata de las medidas del cuerpo humano, principalmente las que se refieren a su tamaño; al tamaño de sus segmentos, formas, fuerza y capacidad de trabajo y es una de las bases fundamentales de la ergonomía.
SCRUM	Es un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto.
Metodología ágil	Es un conjunto de metodologías para el desarrollo de proyectos que precisan de rapidez y flexibilidad para adaptarse a condiciones cambiantes del sector o mercado, aprovechando dichos cambios para proporcionar ventaja competitiva.
Percentiles	Es una de las llamadas medidas de posición no central (cuartiles, deciles, quintiles, percentiles, etc) que se puede describir como una forma de comparación de resultados, por ello es un concepto ampliamente utilizado en campos como la estadística o el análisis de datos.
Prototipo	Es un modelo del comportamiento del sistema que puede ser usado para entenderlo completamente o ciertos aspectos de él y así clarificar los requerimientos. Un prototipo es una representación de un sistema, aunque no es un sistema completo, posee las características del sistema final o parte de ellas.
Organización Mundial de la Salud (OMS)	Es el organismo internacional del sistema de las Naciones Unidas responsable de la salud. Constituido por un grupo de expertos elaboran directrices y normas sanitarias, y ayudan a los países a abordar las cuestiones de salud pública, apoyar y promover las investigaciones sanitarias.

Organización Panamericana de la Salud (OPS)	Es el organismo especializado de salud del sistema interamericano, encabezado por la OEA, y también está afiliada a la OMS, desde 1949, de manera que forma parte igualmente del sistema de las Naciones Unidas.
Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO)	La FAO es la agencia de las Naciones Unidas que lidera el esfuerzo internacional para poner fin al hambre. Su objetivo es lograr la seguridad alimentaria para todos, y al mismo tiempo garantizar el acceso regular a alimentos suficientes y de buena calidad para llevar una vida activa y sana.
Sistema de Vigilancia del Comportamiento de Riesgo Juvenil 2015 (YRBSS)	Es una encuesta bienal estadounidense sobre el riesgo de salud de los adolescentes y las conductas protectoras de la salud, como el tabaquismo, el consumo de alcohol, el consumo de drogas, la dieta y la actividad física realizada por los Centros para el Control y la Prevención de Enfermedades. Es una de las mejores fuentes de información sobre estos comportamientos de riesgo, y las agencias federales lo usan para rastrear el consumo de drogas, el comportamiento sexual y otros comportamientos de riesgo.
Regresión Logística	Es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras.
Diabetes	Es una enfermedad producida por la presencia de altos niveles de azúcares en la sangre. Básicamente la diabetes se reproduce en el organismo a causa de la falta de una hormona segregada por el páncreas llamada Insulina, también por la resistencia que pueda prestar el cuerpo a dicha hormona.
Cross Validation	La validación cruzada o cross-validation es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones.
AJAX	AJAX es el acrónimo de Asynchronous Javascript and XML, es decir: Javascript y XML Asíncrono. Se define como una técnica para el desarrollo de páginas (sitios) web que implementan aplicaciones interactivas.



ANEXOS

ANEXO I
MANUAL USUARIO/ADMINISTRADOR

**SISTEMA DE PREDICCIÓN OBESIDAD EN
ADOLESCENTES – UCSM**

INTRODUCCIÓN

El sistema de predicción de obesidad en adolescentes fue desarrollado para permitir al usuario realizar evaluaciones y cuestionarios, con el fin de que el sistema haga una predicción acerca de su estado físico, también fue diseñado de tal forma que los usuarios administradores puedan exportar toda la información recolectada por el sistema, teniendo los privilegios para acceder a los reportes generales de evaluación IMC y cuestionario de actitud hacia la obesidad.

INICIO DE SESIÓN - ADMINISTRADOR

El sistema de autenticación para usuario administradores, solicita el ingreso de un nombre de usuario y una contraseña, estos datos deben ser otorgados por los desarrolladores del sistema, en caso contrario se debe solicitar tales credenciales, para así poder acceder con ese nivel de usuario.

En la página de inicio de sesión, se deben ingresar las credenciales de usuario administrador otorgadas, y seguidamente presionar en el botón de “Iniciar”.



Iniciar sesión



Usuario

Contraseña

Iniciar

[Ingresar en modo usuario](#)

INICIO DE SESIÓN - USUARIO

El usuario que desee ingresar al sistema de predicción, solamente deberá ingresar su propio nombre, luego presionar en el botón “Continuar”. Desde el formulario inicial, el usuario tiene la opción de ingresar de forma anónima, en caso sea así lo requiera, además también puede descargar el manual de usuario. Los accesos a estas últimas dos opciones se encuentran debajo del botón “Continuar”.



Ingresar nombre de usuario

Continuar

[Ingresar en modo administrador](#)

[Ingresar en modo anónimo](#)

[Descargar manual de usuario](#)

- **Cerrar sesión**

Una vez autenticado, el sistema muestra una barra de menú en la parte superior, en esta se tiene un botón para cerrar sesión en el momento que lo desee.

La barra se encuentra ubicada exactamente debajo de la barra de texto url, para cerrar sesión el usuario tendrá que presionar en el botón “Finalizar”.



MODULO DE APLICACIONES

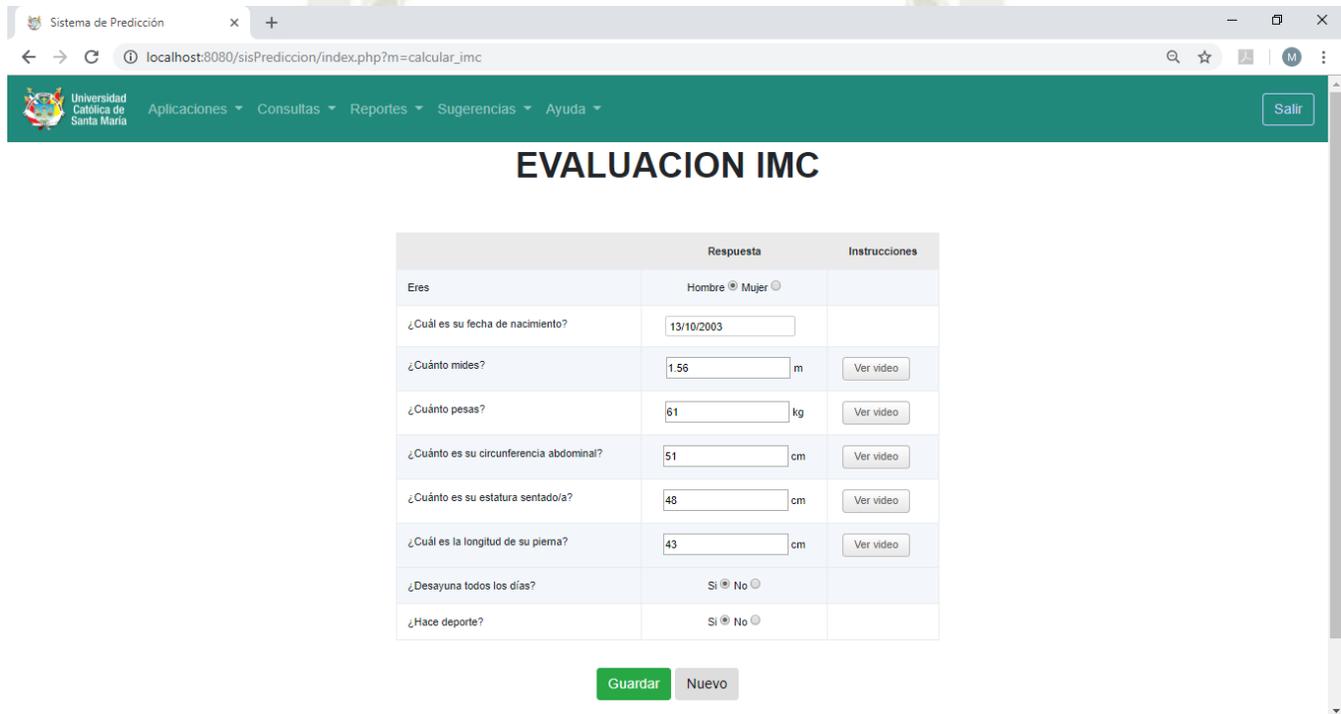
El módulo de aplicaciones contiene hasta el momento sólo dos aplicaciones, Evaluación IMC y Cuestionario de Actitud hacia la obesidad. En ambas aplicaciones el usuario tiene un número máximo de oportunidades de 3, una vez excedido tal número el usuario emitirá un mensaje de notificación, indicando que ya es posible realizar más intentos.

- **Realizar evaluación IMC**

1. Clic en la opción “Realizar Evaluación IMC”, en la barra oscura ubicada en la parte superior.



2. En el siguiente formulario es donde se deben ingresar los datos correspondientes a la evaluación, seguidamente se debe presionar en el botón “Guardar”.



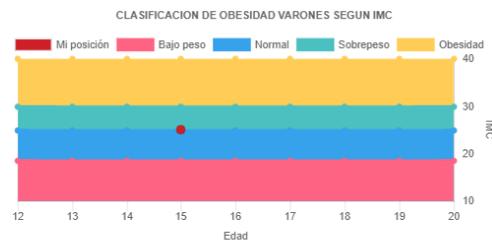
	Respuesta	Instrucciones
Eres	Hombre <input checked="" type="radio"/> Mujer <input type="radio"/>	
¿Cuál es su fecha de nacimiento?	<input type="text" value="13/10/2003"/>	
¿Cuánto mides?	<input type="text" value="1.56"/> m	<input type="button" value="Ver video"/>
¿Cuánto pesas?	<input type="text" value="61"/> kg	<input type="button" value="Ver video"/>
¿Cuánto es su circunferencia abdominal?	<input type="text" value="51"/> cm	<input type="button" value="Ver video"/>
¿Cuánto es su estatura sentado/a?	<input type="text" value="48"/> cm	<input type="button" value="Ver video"/>
¿Cuál es la longitud de su pierna?	<input type="text" value="43"/> cm	<input type="button" value="Ver video"/>
¿Desayuna todos los días?	Si <input checked="" type="radio"/> No <input type="radio"/>	
¿Hace deporte?	Si <input checked="" type="radio"/> No <input type="radio"/>	

3. Una vez ingresada la evaluación, sistema emitirá una notificación avisando que los datos fueron ingresados correctamente, luego se emitirá un texto y un diagrama con los resultados de predicción.



EVALUACION IMC

Resultado: Probabilidad de que su estado de peso sea NORMAL



En el resultado se ve el posible estado de peso del usuario, más abajo se ve un gráfico donde se muestra el estado de peso actual, de acuerdo a la edad y el IMC.

- **Realizar Cuestionario Actitud hacia la Obesidad**

1. Para acceder a este cuestionario, se debe ir a “Cuestionario” opción listada en Aplicaciones.



2. En este cuestionario se ingresan los siguientes datos: Fecha de nacimiento, Género y una lista de 20 preguntas, en conjunto estos datos determinarán cuál es el nivel del usuario en cuanto a la actitud hacia la obesidad.

Sistema de Predicción

localhost:8080/sisPrediccion/index.php?m=cargarCuestionario&op=5

Universidad Católica de Santa María

Aplicaciones Consultas Reportes Sugerencias Ayuda

Salir

CUESTIONARIO ACTITUD HACIA LA OBESIDAD

Fecha de nacimiento: Género: Hombre Mujer

T.A.= Totalmente de acuerdo; NA-ND = Ni acuerdo, ni desacuerdo; T.D. = Totalmente desacuerdo.

Nro	Pregunta	T.A.	NA-ND	T.D.
1	Con frecuencia prefiero quedarme en casa, que salir a jugar al parque.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	A menudo no me llama la atención mucho hacer ejercicios físicos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Me agrada salir a reuniones sociales con mis amigos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Me siento a gusto comiendo en la calle que en mi casa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Cuando tengo tiempo libre, prefiero mirar televisión en casa.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Cuando estoy en casa no me gusta ayudar en los quehaceres del hogar (limpiar, barrer, planchar, cortar pasto, etc).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Me agrada más la idea de comer comida rápida que las frutas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	Cuando estoy en el colegio me gusta sentarme para conversar con mis amigos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	Te sentirías preocupado en el caso de que subas de peso repentinamente?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Luego de responder todas las preguntas e ingresar los datos solicitados, se debe hacer clic en el botón “Guardar”, ubicado al final del cuestionario.

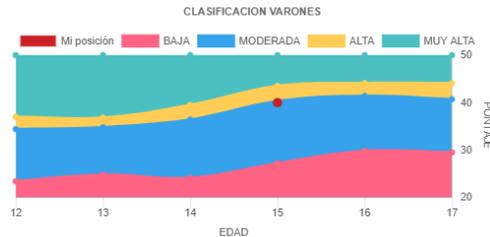
19	Creo que puedo subir de peso si me dejo vencer por la flojera.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20	Considero que los ejercicios físicos son importantes para mantener mi peso corporal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- Si los datos fueron ingresados correctamente, el sistema emitirá un mensaje informando sobre la acción, posteriormente se mostrarán los resultados, un resultado tipo texto y otro resultado tipo gráfico.



CUESTIONARIO ACTITUD HACIA LA OBESIDAD

Resultado: MODERADA ACTITUD hacia la obesidad, el puntaje acumulado es 40 y la edad es 15



En el resultado tipo texto como es de entenderse, se muestra el nivel de actitud hacia la obesidad, detectado por el sistema, también se muestran el puntaje acumulado y la edad.

En el resultado tipo gráfico, se muestra la ubicación del usuario en cuanto a su resultado obtenido.

5. Si se excede el número de intentos, el sistema enviará el siguiente mensaje:



CUESTIONARIO ACTITUD HACIA LA OBESIDAD

Usted ha excedido el número de veces en realizar este cuestionario

MODULO DE CONSULTAS

En el módulo de consultas, el usuario tiene la opción de ver los datos y resultado de cada evaluación y/o cuestionario realizado.

- **Consultar evaluación IMC**

1. Se debe ir a la opción “Evaluación IMC”, de la lista de opciones en el menú consultas.



- En esta opción se pueden ver los datos ingresados en la evaluación IMC, siempre cuando se haya realizado al menos una, de lo contrario el sistema emitirá un mensaje indicando que no se realizó evaluaciones y que para hacerlo se debe dirigir al módulo de “Aplicaciones”.

Sistema de Predicción

localhost:8080/sisPrediccion/index.php?m=consulta_imc

Universidad Católica de Santa María

Aplicaciones Consultas Reportes Sugerencias Ayuda

Salir

Resultado Evaluación IMC

Nombre de usuario: administrator

Resultado	Peso (kg)	Altura (m)	IMC	Edad	Género	Porcentaje grasa	Masa grasa	Masa magra	Circunferencia abdominal(cm)	Estatura sentado(cm)	Pierna(cm)	
1	NORMAL	51	1.56	20.96	12.25	Hombre	11.77%	6	45	40	42	39

- Consultar Cuestionario Actitud hacia la Obesidad**

- Para consultar los resultados del cuestionario actitud hacia la obesidad, es necesario ir al módulo de consultas y hacer clic en la opción “Cuestionario”.

Sistema de Predicción

localhost:8080/sisPrediccion/index.php?m=principal

Universidad Católica de Santa María

Aplicaciones Consultas Reportes Sugerencias Ayuda

Salir

Evaluaciones

Cuestionarios

Universidad Católica de Santa María

- En seguida el sistema proporcionará los datos del cuestionario realizado, en una tabla con los campos: Resultado, Edad, Género y Puntaje.

Sistema de Predicción

localhost:8080/sisPrediccion/index.php?m=consulta_cuestionarios

Universidad Católica de Santa María

Aplicaciones Consultas Reportes Sugerencias Ayuda

Salir

Resultado Cuestionario actitud hacia la obesidad

Nombre de usuario: administrator

Resultado	Edad	Género	Puntaje	
1	MUY ALTA ACTITUD HACIA LA OBESIDAD	15	Mujer	50
2	MODERADA ACTITUD HACIA LA OBESIDAD	16	Hombre	36
3	MODERADA ACTITUD HACIA LA OBESIDAD	15	Mujer	36

MODULO DE REPORTE

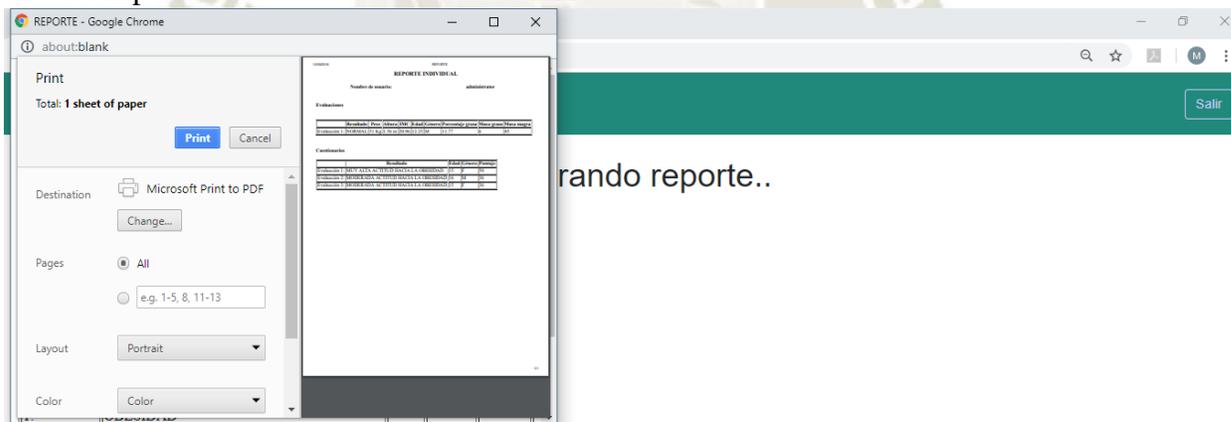
En el módulo de reportes, el usuario tiene la opción de generar documentos en base a los datos ingresados y emitidos por el sistema. Los tipos de reportes que se pueden generar son, el reporte individual y reporte general.

- **Reporte individual**

1. Para generar el reporte individual, el usuario administrador se debe dirigir al menú reportes, y hacer clic en la opción “Reporte individual”.



2. El sistema tardará unos segundos para generar este reporte, pero luego de ese tiempo aparecerá una ventana flotante, donde se muestran tanto las evaluaciones y/o cuestionarios realizados por el usuario, en formato de documento pdf, preparado para realizar una impresión o si se desea guardar en su computador.



Este documento se compone de dos partes principales, Evaluaciones y Cuestionarios. En cada una se muestran tablas, donde se listan las evaluaciones y/o cuestionarios realizados por el usuario, cabe mencionar que en la parte superior se puede observar el nombre del usuario.

- **Reporte Exportar evaluaciones**

1. Este reporte es exclusivo de los usuarios administradores, consiste en exportar todas las evaluaciones imc realizadas hasta la fecha en un archivo Excel. Para generar tal reporte se debe acceder al módulo de reportes, y en la lista de opciones hacer clic a “Exportar Evaluaciones”.



2. Posteriormente se descargará un archivo en formato Excel, para ver la descarga, se debe observar la parte inferior del navegador.

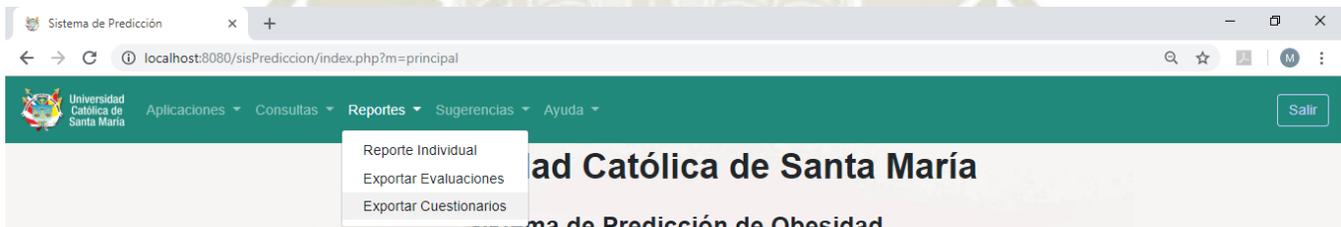


3. Abrir el archivo para ver el reporte generado.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2		NOMBRE (RESULTAD	PESO(kg)	ALTURA(m	IMC	EDAD	GENERO	PORCENT	MASA GR	MASA MA	CIRCUNFE	ESTATURA	PIERNA	(cr	PVC	DESAYUN	DEPORTE	FECHA DE EVALUACION			
3		Maria Can	NORMAL	60	1.7	20.76	26.67	1	14.85	8.91	51.09	60	42	43	-1.75898	1	1	29/07/2018			
4		Maria Can	SOBREPES	70	1.69	24.51	25.17	1	19	13.3	56.7	66	62	46	1.823836	1	1	29/07/2018			
5		Maria Can	SOBREPES	70	1.66	25.4	24.33	0	30.68	21.48	48.52	65	60	30	-0.01822	0	1	29/07/2018			
6		Maria Campos		68	1.75	22.2	19.08	0	25.63	17.43	50.57	55	60	39	-1.63714	1	1	29/07/2018			
7		Maria Can	NORMAL	55	1.55	22.89	19.75	1	15.81	8.7	46.3	60	40	56	-3.95076	0	1	29/07/2018			
8		Maria Can	SOBREPES	76	1.69	26.61	26.58	0	32.65	24.81	51.19	90	65	70	4.477591	1	0	29/07/2018			
9		Maria Can	SOBREPES	78	1.7	26.99	26.58	0	33.1	25.82	52.18	80	70	80	6.012161	1	0	29/07/2018			
10		Maria Can	SOBREPES	70	1.69	24.51	18.58	1	17.49	12.24	57.76	80	79	82	1.531146	1	0	29/07/2018			
11		Maria Can	NORMAL	50	1.63	18.82	22.33	1	11.52	5.76	44.24	60	40	40	-3.13497	1	1	29/07/2018			
12		Maria Can	NORMAL	59	1.66	21.41	19.42	1	13.96	8.24	50.76	80	78	68	1.751867	1	1	29/07/2018			
13		Gilberto S	NORMAL	60	1.56	24.65	19.58	1	17.88	10.73	49.27	60	70	49	0.874529	1	1	30/07/2018			
14		Gilberto S	NORMAL	60	1.65	22.04	14.67	0	24.42	14.65	45.35	40	50	45	-4.1777	1	1	31/07/2018			
15		Humbertc	NORMAL	65	1.65	23.88	18.67	1	16.75	10.89	54.11	60	70	40	0.616349	1	0	1/8/2018			
16		Humbertc	SOBREPES	80	1.7	27.68	17.92	0	31.94	25.55	54.45	60	60	40	-2.3134	0	0	1/8/2018			
17		Humbertc	SOBREPES	85	1.75	27.76	17.33	1	21.1	17.94	67.06	60	80	40	1.60099	0	1	1/8/2018			
18		Gerardo L	SOBREPES	70	1.6	27.34	14.75	1	20	14	56	65	60	40	-2.17434	1	1	1/8/2018			
19		Pedro Pec	NORMAL	65	1.69	22.76	14.5	1	14.45	9.39	55.61	65	67	50	-1.63861	1	1	2/8/2018			
20		Pedro Pec	NORMAL	70	1.75	22.86	18.58	1	15.51	10.86	59.14	60	60	50	-1.00216	1	1	2/8/2018			
21		Rodrigo Soza		68	1.78	21.46	1819.5	1	428.04	291.07	-223.07	40	50	42	521.6031	1	1	3/8/2018			
22		Rodrigo S	NORMAL	68	1.72	22.99	18.58	1	15.66	10.65	57.35	60	50	40	-2.31556	1	1	3/8/2018			
23		Enrique G	NORMAL	70	1.81	21.37	18.67	1	13.74	9.62	60.38	60	50	50	-2.48579	1	1	3/8/2018			

- **Reporte Exportar cuestionarios**

1. Este es otro de los reportes sólo accesibles por usuarios administradores, para generarlo el usuario debe ir al módulo de reportes, y hacer clic en la opción “Exportar Cuestionarios”.



2. A continuación, se debe descargar automáticamente un archivo Excel en la parte inferior del navegador.



3. Hacer clic en el cuestionario para ver la información generada en el reporte.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2		NOMBRE	RESULTAD	EDAD	GENERO	PUNTAJE	FECHA DE EVALUACION												
3		Gilberto S	MUY ALTA	12	1	39	31/07/2018												
4		Gilberto S	MODERAC	15	1	38	31/07/2018												
5		Gilberto S	MODERAC	15	1	35	31/07/2018												
6		Gilberto S	MUY ALTA	13	1	38	31/07/2018												
7		Martin Hernandez		0	1	36	31/07/2018												
8		Humberto Palomino		19	1	36	1/8/2018												
9		Humberto	MODERAC	15	1	35	1/8/2018												
10		Humberto Palomino		11	1	36	1/8/2018												
11		Pedro Pec	MODERAC	14	1	36	2/8/2018												
12		Pedro Pec	ALTA ACTI	13	1	36	2/8/2018												
13		Pedro Pec	MODERAC	14	1	36	3/8/2018												
14		Fabio Mar	MODERAC	14	1	36	3/8/2018												
15		Fabio Mar	ALTA ACTI	12	1	36	3/8/2018												
16		Marco Tici	MODERAC	14	1	36	3/8/2018												
17		Diego Lop	ALTA ACTI	13	1	36	4/8/2018												
18		Diego Lop	ALTA ACTI	12	1	36	4/8/2018												
19		Diego Lop	MODERAC	14	1	36	4/8/2018												
20		Joselyn Pe	MODERAC	14	1	36	5/8/2018												
21		Joselyn Pe	MODERAC	16	1	36	5/8/2018												
22		Joselyn Pe	MUY ALTA	16	1	46	5/8/2018												
23		Karen Ticc	MODERAC	14	1	36	5/8/2018												

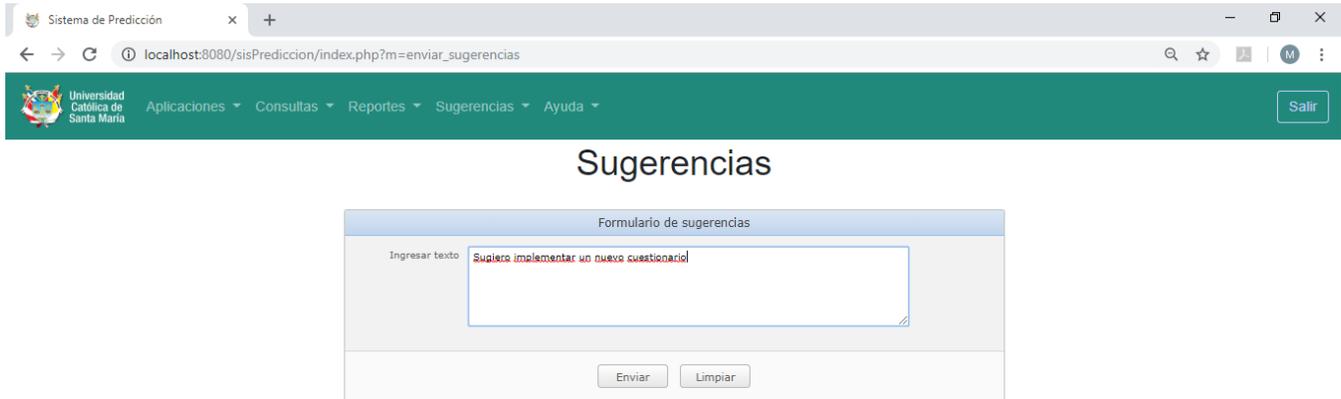
MODULO DE SUGERENCIAS

En este módulo, tanto usuario final como usuario administrador, pueden enviar sugerencias. Estas sugerencias serán tomadas en cuenta para futuras posibles mejoras y/o actualizaciones en el sistema.

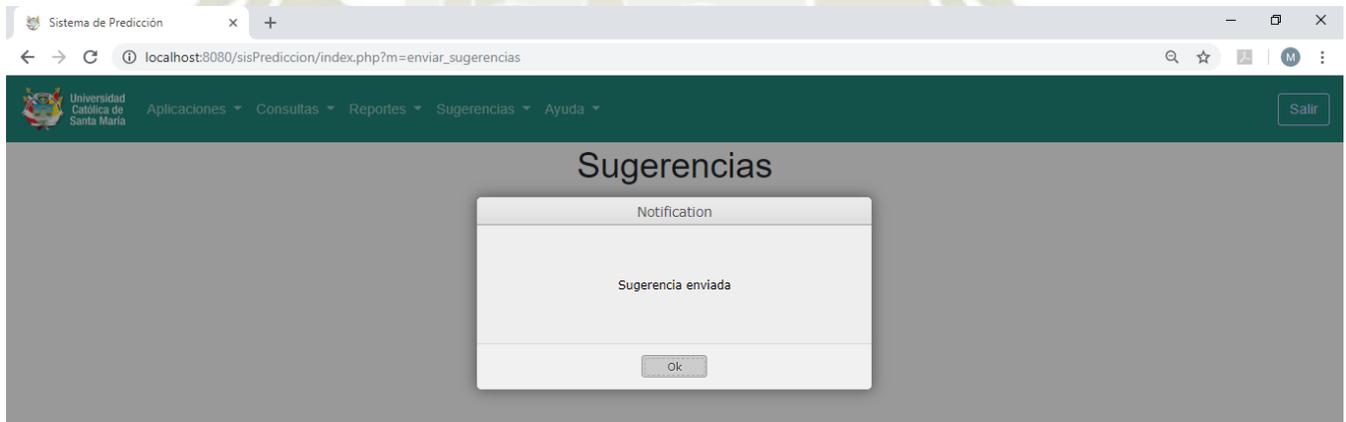
1. Para acceder, se debe ir al menú sugerencias, y hacer clic en el botón “Enviar Sugerencia”.



2. Seguidamente el sistema dirigirá al usuario a un formulario, donde el usuario podrá explayarse y enviar la sugerencia que desee.



3. Una vez que se ingrese el texto, se debe presionar en el botón “Enviar”, luego el sistema notificará al usuario si la sugerencia fue enviada satisfactoriamente o no.





ANEXO II
PLAN DE TESIS

Id	Activo	Nombre	Duración	Comienzo	Fin
1	Sí	Obtener conceptos básicos	6 días	6 noviembre 2017	11 noviembre 2017
2	Sí	Definición del problema	3 días	13 noviembre 2017	15 noviembre 2017
3	Sí	Dominio del problema	3 días	16 noviembre 2017	18 noviembre 2017
4	Sí	Preparación de datos	5 días	20 noviembre 2017	24 noviembre 2017
5	Sí	Exploración de datos	5 días	25 noviembre 2017	30 noviembre 2017
6	Sí	Generación de modelos	5 días	1 diciembre 2017	6 diciembre 2017
7	Sí	Exploración y validación de modelos	6 días	7 diciembre 2017	13 diciembre 2017
8	Sí	Implementación y actualización de modelos	8 días	14 diciembre 2017	20 diciembre 2017
9	Sí	Inicio del desarrollo del sistema	1 día	21 diciembre 2017	22 diciembre 2017
10	Sí	Identificar Requerimientos	3 días	23 diciembre 2017	26 diciembre 2017
11	Sí	Levantamiento de los requerimientos	12 días	27 diciembre 2017	9 enero 2018
12	Sí	Listar los requerimientos	6 días	10 enero 2018	16 enero 2018
13	Sí	Clasificar RF y RNF	2 días	17 enero 2018	18 enero 2018
14	Sí	Analizar los requerimientos clasificados	5 días	19 enero 2018	24 enero 2018
15	Sí	Elaborar Diagramas	10 días	25 enero 2018	7 febrero 2018
16	Sí	Casos de Uso y Diagrama de Actividades	4 días	8 febrero 2018	12 febrero 2018
17	Sí	Corrección de Diagramas	5 días	13 febrero 2018	17 febrero 2018
18	Sí	Diseñar Interfaces	16 días	19 febrero 2018	8 marzo 2018
19	Sí	Prototipo interfaz principal	7 días	9 marzo 2018	16 marzo 2018
20	Sí	Prototipo interfaces de módulos	9 días	17 marzo 2018	27 marzo 2018
21	Sí	Conexión con base de datos	2 días	28 marzo 2018	29 marzo 2018
22	Sí	Definir modelo Cliente Servidor	4 días	30 marzo 2018	3 abril 2018
23	Sí	Desarrollo del Sistema	45 días	4 abril 2018	25 mayo 2018
24	Sí	Módulo principal de ingreso de datos	12 días	26 mayo 2018	8 junio 2018
25	Sí	Módulo de consultas	10 días	9 junio 2018	20 junio 2018
26	Sí	Módulo de Generador de reportes	10 días	21 junio 2018	2 julio 2018
27	Sí	Presentación del sistema a usuario final	2 días	3 julio 2018	4 julio 2018
28	Sí	Evaluación de la funcionalidad	2 días	5 julio 2018	6 julio 2018
29	Sí	Ingresar información recopilada	2 días	7 julio 2018	9 julio 2018
30	Sí	Integración de datos	4 días	10 julio 2018	13 julio 2018
31	Sí	Minería de datos	5 días	14 julio 2018	19 julio 2018
32	Sí	Evaluación de Patrones	8 días	20 julio 2018	28 julio 2018
33	Sí	Presentación del conocimiento	5 días	30 julio 2018	3 agosto 2018
34	Sí	Pruebas del sistema	7 días	4 agosto 2018	10 agosto 2018
35	Sí	Desarrollo de informe de Tesis	10 días	11 agosto 2018	20 agosto 2018
36	Sí	Elaborar informe y revisar formato APA	7 días	21 agosto 2018	27 agosto 2018
37	Sí	Elaborar anexos del sistema	8 días	28 agosto 2018	4 setiembre 2018
38	Sí	Presentar borradores de tesis	7 días	5 setiembre 2018	11 setiembre 2018
39	Sí	Corrección de observaciones	30 días	12 setiembre 2018	12 octubre 2018
40	Sí	Sustentación de Tesis	1 día	25 octubre 2018	25 octubre 2018