



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE INGENIERÍA DE SISTEMAS

**DISEÑO Y DESARROLLO DE UNA APLICACIÓN INTELIGENTE PARA APOYAR LA
EVALUACIÓN E INTERVENCIÓN PSICOLÓGICA DE NIÑOS DE 6 A 10 AÑOS BAJO
RIESGO SOCIAL**

Trabajo de titulación previo a la obtención del
título de Ingeniero de Sistemas

AUTOR: ADRIÁN ALFREDO POVEDA PULLA

TUTOR: ING. VLADIMIR ESPARTACO ROBLES BYKBAEV, PhD.

Cuenca - Ecuador

2023

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Adrián Alfredo Poveda Pulla con documento de identificación N° 0104079082, manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 25 de abril del 2023

Atentamente,



Adrián Alfredo Poveda Pulla

0104079082

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Adrián Alfredo Poveda Pulla con documento de identificación N° 0104079082, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto técnico: “Diseño y desarrollo de una aplicación inteligente para apoyar la evaluación e intervención psicológica de niños de 6 a 10 años bajo riesgo social”, el cual ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 25 de abril del 2023

Atentamente,



Adrián Alfredo Poveda Pulla

0104079082

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Vladimir Espartaco Robles Bykbaev con documento de identificación N° 0300991817, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **DISEÑO Y DESARROLLO DE UNA APLICACIÓN INTELIGENTE PARA APOYAR LA EVALUACION E INTERVENCION PSICOLOGICA DE NINOS DE 6 A 10 ANOS BAJO RIESGO SOCIAL**, realizado por Adrián Alfredo Poveda Pulla con documento de identificación N° 0104079082, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 25 de abril del 2023

Atentamente,



Ing. Vladimir Espartaco Robles Bykbaev, PhD.

0300991817

Agradecimientos

Mi agradecimiento inicial es hacia a mis padres, por haberme brindado la oportunidad de estudiar en la universidad de mi preferencia, con todo el esfuerzo que esto conlleva, además de ofrecerme su ayuda a lo largo de mi período académico y dentro de mi vida personal. Hago un gran énfasis, dando el más sincero agradecimiento a mi tutor de proyecto, PhD. Vladimir Robles; quien, con su sabia y constante guía, han hecho posible lograr esta meta, y finalmente agradecer a todas las personas que colaboraron previamente en este proyecto y fueron un soporte durante las diferentes etapas de desarrollo de este trabajo.

Dedicatoria

Este logro va dedicado de manera especial a mis padres, quienes me han brindado su apoyo incondicional y continuo, además a mis familiares y amigos más cercanos, quienes me han ofrecido una mano y brindado sus palabras en los momentos oportunos.

Resumen

En este trabajo se presenta un ecosistema inteligente que se apoya en herramientas TIC y sistemas inteligentes para dar soporte en diversas actividades que realizan el personal y psicólogos que laboran en instituciones de protección. El ecosistema proporciona funcionalidades para apoyar la evaluación e intervención psicológica en niños, utilizando para ello aplicaciones móviles basadas en “micromundos” y juegos lúdicos. El sistema fue puesto a prueba con 102 niños y alcanzó una precisión de 81 % en la etapa de evaluación del paciente [1], dentro de este proyecto se incluye una fase de experimentación, con el apoyo de técnicas de machine learning con el dataset obtenido de los diagnósticos en el plan de investigación, dentro de la primera fase se utiliza la herramienta Weka, y en la segunda fase se desarrolla un módulo aplicando los algoritmos que presentaron mejor precisión en la primera fase, alcanzando una precisión de 85.71 %.

Palabras Claves: Intervención psicológica, herramientas TIC, ecosistema inteligente, juegos lúdicos.

Abstract

Recent studies provided by the United Nations Children's Fund (UNICEF) claim that 275 million children are victims of violence in the world. An important percentage of these children must be separated from their families to be placed in infant protection institutions, where, it is estimated, about 8 million children live nowadays. However, in developing countries the most of these institutions do not have enough personnel, technological resources and appropriate structures to provide adequate psychological intervention services. In the same way, several institutions require informatics health systems that support the children's monitoring as well their progress. On those grounds, in this paper we present an intelligent ecosystem that relies on IT tools and intelligent systems to provide support in several activities that are done by personnel and psychologists that work in protection institutions. The ecosystem provides functionalities to support the psychological evaluation and intervention on children, using for it mobile applications based on ludic micro and nano worlds, and ludic games. The system was put to test with 102 children and has reached 81 % of precision in the patient's evaluation stage [1]. In this project, an intelligent module is developed with machine learning techniques that reaches a precision of 85.71 %

Keywords: Psychological intervention, IT tools, Intelligent ecosystem, Ludic games.

Índice general

Índice de figuras	XI
Índice de cuadros	XIV
1. Introducción	1
2. Antecedentes	4
3. Objetivos	6
3.1. General	6
3.2. Específicos	6
4. Fundamentos teóricos	8
4.1. Trabajo relacionado	8
4.2. Marco Teórico	15
4.2.1. Perfil psicológico	15
4.2.1.1. El acogimiento institucional	15
4.2.1.2. Perfiles psicológicos	17

<i>ÍNDICE GENERAL</i>	x
4.3. Técnicas de análisis de datos	20
5. Marco metodológico	30
5.1. Diseño de interfaz de aplicación móvil	30
5.2. Diseño de módulo inteligente y preprocesamiento de datos	44
6. Módulo de experimentación y resultados	61
6.1. Levantamiento de requerimientos	61
6.2. Desarrollo del módulo	62
6.3. Actualización de experimentación	84
7. Cronograma	90
8. Presupuesto	94
9. Conclusiones	96
10. Trabajo Futuro	98
11. GlosarioTerminos	100
Bibliografía	102

Índice de figuras

2.1. Arquitectura MicroLudi	5
4.1. Etapas Clustering	24
5.1. Pantalla inicio	31
5.2. Inicio sesión	32
5.3. Registro usuario	33
5.4. Usuario creado	34
5.5. Menú principal	35
5.6. Registro niño	35
5.7. Listado niños	36
5.8. Selección sexo	37
5.9. Categoría partes cuerpo	38
5.10. Subcategorías partes masculino	39
5.11. Subcategorías partes femenino	39
5.12. Subcategoría rostro	40

5.13. Subcategoría ojos masculino	41
5.14. Subcategoría ojos femenino	41
5.15. Subcategoría torso masculino	42
5.16. Subcategoría torso femenino	43
5.17. Avatar completado	44
5.18. Frecuencias Boca	49
5.19. Frecuencias Cuello	50
5.20. Frecuencias Cabeza	50
5.21. Frecuencias Ojos	51
5.22. Frecuencias Orejas	51
5.23. Frecuencias Nariz	52
5.24. Frecuencias Torso	52
5.25. Frecuencias Brazos	53
5.26. Frecuencias Manos	53
5.27. Frecuencias Piernas	54
5.28. Frecuencias Pies	54
6.1. Carga de dataset en herramienta Weka	65
6.2. Análisis de atributos aplicando Cross Val 5 Folds	66
6.3. Análisis de atributos aplicando Cross Val 10 Folds	67
6.4. Análisis de atributos aplicando Full Training Set	68
6.5. Aplicación algoritmo EM 8 atributos	69

6.6. Aplicación algoritmo COBWEB 8 atributos	71
6.7. Aplicación algoritmo K-Means 8 atributos	72
6.8. Aplicación algoritmo FarthestFirst 8 atributos	74
6.9. LoadDatasetMethod	85
6.10. NaibeBayesMethod	85
6.11. NaibeBayesPrecision	86
6.12. MultiLayerMethod	86
6.13. MultiLayerTest	87
6.14. LogisticRegresionMethod	87
6.15. LogisticRegresionTest	88
6.16. decisionTreeMethod	88
6.17. decisionTreeTest	89
7.1. Cronograma	93
8.1. Presupuesto	95

Índice de cuadros

4.1. Técnicas de análisis de datos [2]	21
5.1. Partes del cuerpo Avatar	45
5.2. Resumen de conteo de selecciones en partes del cuerpo inst. acogida	46
5.3. Resumen de conteo de selecciones en partes del cuerpo inst. regular	47
5.4. Moda de selecciones en partes del cuerpo	48
5.5. Resumen comparativo de frecuencias	56
5.6. Rango medio de selecciones en partes del cuerpo	57
5.7. Desviaciones estándar del cuadro 5.6 de rangos medios	58
5.8. Valoración normalizada de elementos de partes del cuerpo	59
6.1. Resultados precisión archivo normalizado 12 partes	77
6.2. Resultados precisión archivo normalizado 12 partes orden aleatorio	78
6.3. Partes del cuerpo 8 atributos	79
6.4. Resultados precisión archivo normalizado 8 partes	79
6.5. Resultados precisión archivo normalizado 8 partes orden aleatorio	80

6.6. Partes del cuerpo 7 atributos	80
6.7. Resultados precisión archivo normalizado 7 partes orden aleatorio	81
6.8. Partes del cuerpo 6 atributos	81
6.9. Resultados precisión archivo normalizado 6 partes	82
6.10. Resultados precisión archivo normalizado 6 partes orden aleatorio	82
6.11. Partes del cuerpo 5 atributos	83
6.12. Resultados precisión archivo normalizado 5 partes orden aleatorio	83
6.13. Métodos a implementarse	84
6.14. Resultados precisión experimentación personalizada	89

Capítulo 1

Introducción

Según el estudio presentado por [3], dentro de un análisis aplicado a 670 alumnos escolares, 28 tienen altas capacidades, basándose en 5 pruebas diferentes para medir distintas capacidades, quienes lograron tener un puntaje mayor a 75/100 en 3 pruebas, población dentro de la que se logra analizar las características similares entre sí. Dentro del experimento se dividen en tres grupos a los niños con altas capacidades, el primer grupo obtuvo un alto puntaje en 5 de las 5 pruebas realizadas, un segundo grupo con altas calificaciones en 4 pruebas, y el tercer grupo, con puntaje sobre el 75/100 en 3 de las pruebas. Se tomará como referencia, al primer grupo. Dentro de su desarrollo en el ambiente familiar 7 de 10 son primogénitos, siendo la primera característica identificada. Se consulta a los padres acerca de rasgos emocionales, la respuesta de mayor frecuencia, indica que son sensibles, siendo relacionada en cómo manejan la intensidad y complejidad de los sentimientos. Además 7 niños realizan actividades extracurriculares, ya sea de deportes o clases extra. A las madres de los niños, se trata de medir la influencia de su edad, pero no lleva a una respuesta congruente, con lo que no se toma como una característica en su desarrollo, se descarta también su ocupación. Se nota una similitud en relación al tiempo que emplean su ayuda con las tareas con sus hijos emplean entre 1 y 5 horas para cumplir con las tareas escolares, sin importar su

nivel académico. Dentro del análisis a los padres, 7 dedican entre 1 y 5 horas para ayudar en las tareas a los niños y 3 no realizan esta actividad, apoyando a la característica de mayor porcentaje en los primogénitos, ya que los padres brindan mayor tiempo, atención y oportunidades de desarrollo a los hijos mayores, confirmando que la figura paterna en la infancia tiene un alto nivel de influencia en el comportamiento afectivo e intelectual de los niños. Dentro de las características del núcleo familiar siete son casados, dos en unión libre y un divorciado, percibimos claramente que la unión de los padres influye de manera positiva en el desarrollo intelectual de los niños. Con ello, se concluye que existen características identificables en el desarrollo de los niños, para que puedan obtener un buen desempeño en el área escolar.

Por contraparte, los menores ingresan a instituciones de acogida u hogares temporales, debido a diversas situaciones asociadas con la incapacidad de los padres para cuidar de sus hijos, como violencia doméstica, privación de libertad, procesos migratorios, la pobreza y la muerte. Sin embargo, estas entidades generalmente enfrentan condiciones complicadas que hacen que la asistencia sea difícil; la escasez de recursos, y la guía general recibida por su personal, puede conducir a procedimientos inadecuados que pueden perpetuar la violencia dentro de la propia institución, al tener una visión deficiente de la interacción de los niños que comúnmente reproduce la violencia vivida en su estilo de vida anterior [4], según estimaciones de [5] alrededor de 8 millones de niños a nivel mundial, viven esta situación hoy en día.

Para la experimentación, se partirá de un conjunto de datos obtenido en la ciudad de Cuenca-Ecuador, en asociación con psicólogas colaboradoras de la Universidad del Azuay, en la que se tuvo una población total de 102 niños, 71 estudiantes regulares y 31 pacientes clínicos, en los que se almacenó los datos de selección en la aplicación móvil del proyecto MicroLudi. Se estudiarán y analizarán diferentes técnicas de análisis de datos para

implementarlo en un módulo inteligente que clasifique entre estos dos grupos con la mayor precisión.

Capítulo 2

Antecedentes

Actualmente, las aplicaciones informáticas han sido desarrolladas con soporte de inteligencia de computador, ya que proveen una gran ayuda en el diagnóstico en varias áreas de la psicología, como el beneficio de identificar de manera automatizada las diferentes características de los rasgos psicológicos, utilizados para evaluar similitudes con perfiles previamente analizados. Dentro de este proyecto se aplicarán análisis en relevancia a la personalidad de los niños.

La aplicación MicroLudi, desarrollada como tema de titulación en la Universidad Politécnica Salesiana, incluye aspectos como el de sistematizar pruebas de proyección psicológica individual y grupal; en la primera fase se desarrolla el test de la persona, la misma que consiste en que una persona proyecte la percepción de sí misma, a través de un dibujo, lo que sirve para interpretación de cada una de las características psicológicas de su autor (niño). Además, de incluir otra prueba individual, la ubicación de la persona en un área, ya que su ubicación y tamaño en el dibujo, son indicadores de ciertos aspectos psicológicos. Por otra parte, se incluye un test proyectivo grupal, el de la familia, que consiste en elaborar cada uno de los integrantes de la familia. Dentro del cual existen aspectos que se toman en cuenta, como el orden en que se elaboran, el tamaño de cada personaje, los gestos de cada uno,

etc., indicadores con su respectiva interpretación. Todas las interacciones en la aplicación son almacenadas, para posteriormente realizar un diagnóstico del perfil del participante. La aplicación está diseñada para niños en un rango de edad de 6 a 10 años. La arquitectura usada en la aplicación MicroLudi, se presenta en la figura 2.1 [1].

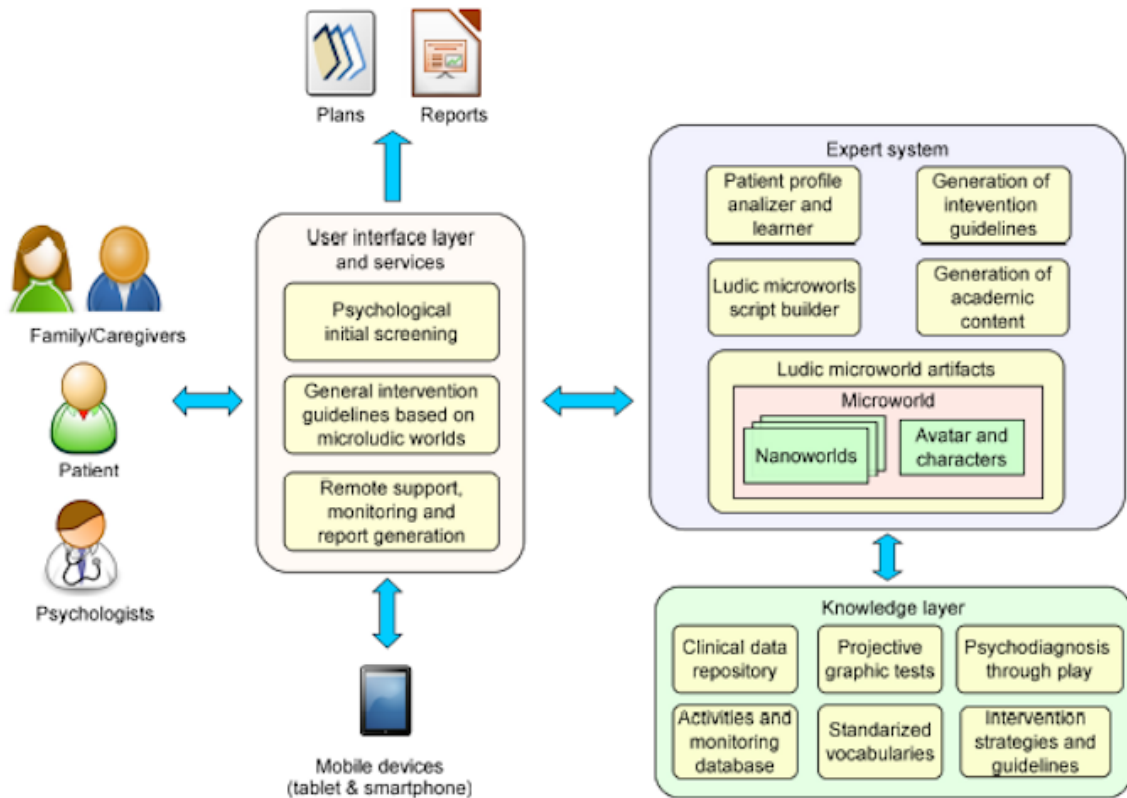


Figura 2.1 Arquitectura general del sistema MicroLudi [1]

Los resultados presentados en el análisis de los perfiles psicológicos de los participantes, alcanzó una precisión de 81 %, se busca implementar un módulo de entrenamiento y experimentación de algoritmos de machine learning, para elevar la precisión del diagnóstico que ofrece la herramienta, facilitando un diagnóstico automatizado más preciso al personal psicológico.

Capítulo 3

Objetivos

3.1. General

Diseñar y desarrollar una aplicación inteligente para apoyar la evaluación e intervención psicológica de niños de 6 a 10 años bajo riesgo social.

3.2. Específicos

- **OE1.** Estudiar y conocer las diferentes características de los perfiles psicológicos de niños y niñas en situación de riesgo social.
- **OE2.** Diseñar y desarrollar una aplicación móvil prototipo para determinar los rasgos psicológicos de los niños en base al test de la personalidad (Draw-A-Person Test).
- **OE3.** Diseñar y desarrollar un módulo inteligente que permita realizar clasificación de niños en función de su perfil psicológico, empleando para ello técnicas de clustering e inteligencia artificial.

- **OE4.** Diseñar y ejecutar un plan de experimentación que permita validar el módulo inteligente y la aplicación móvil.

Capítulo 4

Fundamentos teóricos

4.1. Trabajo relacionado

Dentro de aplicaciones interactivas destinadas a desarrollar diferentes habilidades y conocimientos, existen dos tipos principales: los juegos epistémicos o "serios", y los juegos lúdicos. Los juegos epistémicos, se definen como la participación dentro de una plataforma de simulación que da acceso al marco epistémico; siendo un conjunto de modelos de fases, jerarquías dentro de sistemas dinámicos [6]. Su principal objetivo no es proporcionar diversión, sino tienen un interés en la identificación, la comprensión y la práctica del conocimiento. Este tipo de juegos requiere mucha práctica, ya que a menudo contienen análisis matemático, asociación lógica, toma de decisiones y otras tareas de razonamiento. Por otra parte, se encuentran los juegos lúdicos o de diversión, que de una manera no invasiva el conocimiento es adquirido a través de la navegación en el juego. A diferencia de los "juegos serios" que mantienen una estructura más robusta. Según [7], los juegos serios se distinguen de otros, porque su prioridad se basa en objetivos educativos, en lugar de solo entretenimiento; así que la principal diferencia entre ellos y los juegos lúdicos, es que los primeros se consideran más adecuados para los niños. Hoy en día, las aplicaciones informáticas se han desarrollado con

el apoyo de la inteligencia computacional, brindando una gran ayuda en el diagnóstico dentro de varias áreas de la psicología. Esta coordinación se puede encontrar en la identificación automatizada de ciertas características que corresponden a criterios que evalúan ciertos perfiles analizados, en este caso la personalidad del niño.

Dentro de este ámbito, se cuenta con aplicaciones como GWAP, propuesta por [8], que cuenta con un entorno SortIT, que presenta juegos lúdicos como rompecabezas, en los que el jugador se enfrenta a una batalla, pudiendo defender o atacar el/ella mismo/a. El jugador interactúa con el desarrollo del juego a través de múltiples consultas, y dependiendo de sus respuestas, el nivel de vida de los personajes disminuye o se mantiene. Utiliza intervalos de probabilidad o expresiones verbales que sirven como herramienta para las tareas de obtención de conocimiento, para la identificación futura de ciertas características de los perfiles de personalidad.

Dentro del área de la inteligencia artificial, destacan diferentes sistemas como Copernicus, propuesto por [9], el cual brinda apoyo al área de la psicología clínica, basándose en diagnósticos clínicos para el análisis de perfiles de los pacientes. Consta de una base de conocimiento, basada en reglas de datamining, al ingresar un nuevo caso permite seleccionar el tipo de técnica a aplicarse para su clasificación.

El modelo (*GOM*) (Game Object Model) Versión II, contiene un espacio de visualización, en el que presenta diferentes interfaces para la interacción con módulos destinados a promover el pensamiento crítico, los descubrimientos, las habilidades, a través de un orden lineal de la historia. Su módulo principal, llamado Khozi, es un juego de aventura con fines educativos relacionados con enfermedades en África (VIH, SIDA, malaria, tuberculosis y cáncer). Se lleva a cabo en un micromundo en 3D, en el que el jugador explora el entorno para resolver varios rompecabezas interrelacionados para recrear el hilo de la historia. Presenta un rendimiento notable en el diseño de juegos con aprendizaje auténtico [10].

Dentro del área lúdica, la aplicación web “Lúdo kids”, presentada por [11], bajo un enfoque para niños de primer grado para la asignatura de "gramática", donde con el apoyo del instructor, se verifica que la ejecución del juego se lo realice de una manera efectiva. Para definir su interfaz, con la mayor aceptación entre los niños, realizó una encuesta, resultando la de mayor aceptación, un fondo de mar. Se manejan diferentes ambientes con imágenes y sonidos, en los que se puede seleccionar entre un listado de juegos como:

- Juguemos con las vocales: se aprenderá a identificar las vocales
- Armemos el rompecabezas playero: se armará palabras con “Q” en “que” y “qui”
- Reconoce los nombres comunes: se clasificará los nombres comunes.
- El cofre de las burbujas: donde aprenderá a identificar y asociar las consonantes b y d
- Vamos al parque: se unirá palabras inversas y mixtas
- Formemos oraciones en la playa: aprenderá a formar oraciones correctamente
- Repaso final: puede repasar todas las demás actividades en una sola

Donde al realizar acciones correctas en cualquier juego, se visualiza una estrella de color amarillo, y al cometer una equivocación se visualiza una estrella de color rojo.

Dentro de Ecuador, destaca el trabajo desarrollado por [12] el "sistema lúdico para fortalecer valores en niños de 6 a 8 años", en el que se presenta una animación de video con relatos basados en la religión, se presentan pequeñas historias del actuar de Jesús con diferentes personajes. Estos relatos se desenvuelven en un ambiente adaptado a la arquitectura y época, para una mejor proyección de los niños hacia las historias. Al finalizar las historias, se propone una secuencia de preguntas, en modo de trivia para almacenar su nivel de percepción y comprensión.

Dentro de la categoría de los juegos epistémicos o serios, se puede encontrar a "¿Cuánto suman los ángulos interiores de un triángulo?", en donde [13] expone que es un proyecto sobre evaluación y desarrollo de competencias matemáticas, que estaría aplicado a los profesores de primaria y facilitaría la efectivización del análisis de conocimiento con la actividad matemática. Se hace frente a la problemática que expone [14] "la noción de conocimiento nos parece una y evidente, pero en el momento en que se le interroga, estalla, se diversifica, se multiplica en nociones innumerables, planteando cada una de ellas una nueva interrogante". Con lo que se debe adoptar modelos explícitos con elementos y sus variedades, si se desea el desarrollo de tales conocimientos. Dentro de la evaluación, siguiendo diferentes guías se llega a la conclusión de formular preguntas que cumplan con los siguientes objetivos: describir una experiencia de formación en la que articule su formación matemática y presentar una "guía" para el reconocimiento de objetos y procesos necesarios para una cognición matemática. Con lo que se formula en un desarrollo bastante breve, ya que consta del desarrollo de las siguientes preguntas.

- ¿Cuánto suman los ángulos interiores de cualquier triángulo? Justifica la respuesta.
- ¿Qué conocimientos se ponen en juego en la resolución de este problema?

La primera pregunta evalúa conocimiento común, mientras que la segunda, requiere un conocimiento especializado, la ser necesario hacer explícitos los conocimientos matemáticos con conocimientos didácticos, dentro de un contexto introductorio a geometría.

Dentro del análisis de las respuestas, para la primera pregunta se encuentra que reafirman su respuesta con distintos argumentos, también con justificaciones parciales, incorrectas o no pertinentes. Lo que lleva a la conclusión de no tener claro el concepto de justificar. Para la segunda pregunta, la mayoría presenta conceptos sin interpretación, por ejemplo, se utilizan propiedades de ángulos y triángulos, sin describirlas, con lo que se llega a la conclusión de que no describen los procedimientos enunciados.

Como indica [15], “el diseñar e implementar situaciones didácticas para la formación de profesores cuyo objetivo central sea el metaanálisis.” Dentro de este proyecto, desarrollado por [13], nos dice que la actividad trata de determinar un componente clave de la enseñanza, quizás inicialmente sea conflictiva por suponer usualmente un cierto nivel de actividad metacognitiva a la que no están habituados.

Otro juego serio, que se puede destacar es el presentado por [16], que tiene un enfoque hacia niños con discapacidades, en el que se define a la discapacidad intelectual como “una afección que produce un funcionamiento intelectual general por debajo del promedio y una carencia de las destrezas necesarias para la vida diaria.”, lo que implica que dado el nivel motriz de estos niños para el desempeño en los juegos, haya sido desarrollado con el sensor Kinect, que de manera inalámbrica detecta gestos, comando de voz, objetos e imágenes. Este juego está compuesto por dos actividades.

Actividad 1: Se visualiza el cuerpo humano dividido en tres o seis cortes, según se escoja inicialmente. Dispuestos de la siguiente manera:

- Cuerpo humano con tres cortes
 - Cabeza
 - Torso
 - Piernas

- Cuerpo humano con seis cortes
 - Cabeza
 - Torso (mano derecha)
 - Torso (mano izquierda)
 - Torso solo

- Pierna derecha
- Pierna izquierda

Al momento que el niño selecciona una parte del cuerpo, se reproduce un sonido indicando a qué parte del cuerpo corresponde, entonces debe colocarla en el cuerpo, en el que recibe estímulos al realizarlo correctamente, y en caso de un error, la parte del cuerpo vuelve a la posición inicial con una cara triste.

Actividad 2: El niño debe asociar objetos a las partes de su cuerpo. Inicialmente se visualiza el cuerpo completo de un niño, con diferentes elementos a su alrededor que se usan para la salud e higiene. Dispuestos de la siguiente manera:

- Cepillo de dientes - Boca
- Jabón - Manos, pies y cara
- Shampoo - Pelo
- Ducha - Cabeza
- Perfume - Cuello
- Talco - Pies
- Peine - Pelo

El niño debe ubicar cada objeto con la parte del cuerpo que corresponda, de igual manera, si lo hace correcto recibe estímulos de éxito, además de escuchar una explicación de su uso, y si falla, el objeto regresa a su posición y se muestra una cara triste.

Se puede mencionar la aplicación del autor [17], desarrollada bajo multiplataforma, en el lenguaje Dart con el Framework Flutter, la que consiste en una serie de juegos serios para

niños y adolescentes con diversidad cognitiva. Para acceder al juego, el niño necesita un registro de usuario con el cual podrá visualizar la sección de estadísticas. Que almacena datos como tiempo de juego, datos referentes a sus últimas 4 partidas, suma general de aciertos, fallos y retos conseguidos, además de su récord personal en puntaje de cada juego. Dentro de la sección de juegos, se cuenta con algunos de diferentes modalidades cognitivas, como calculadora, memoria, memoria progresiva, razonamiento y orientación. Al iniciar cada juego, verifica el nivel almacenado para el niño, si es que no existe un registro presenta una pantalla con las instrucciones del juego correspondiente y al terminarlo, sea de manera satisfactoria o incorrecta, se dirige a la pantalla de resultados con la posibilidad de jugarlo nuevamente, mientras vaya superando los retos presentados en los juegos la siguiente partida será con un nivel de dificultad superior adaptándose así a las diversidades cognitivas.

Otro proyecto que se puede destacar, dentro de los juegos serios, fue desarrollado por [18] bautizado como RAICES, y tiene el propósito didáctico para la enseñanza, el autor añade que "posibilita a los jugadores obtener un conjunto de conocimientos sobre las culturas aborígenes en forma natural y distendida, despertando curiosidad y deseos por aprender.^{E1} enfoque de este juego, es que los niños además de socializar, cooperar y divertirse en las redes sociales, puedan adquirir conocimientos. Colaborando en un proceso de concientización sobre los valores de las culturas originarias y sus problemáticas. Fue desarrollado en Argentina, con lo que, dentro del juego se tomaría la información de sus pueblos originarios. Nos indica [18] que dentro de RAICES "se intenta generar un estado emocional que propicie procesos de aprendizaje más efectivos que los métodos tradicionales de enseñanza." Dentro de su diseño, el esquema adoptado fue definir la temática a transmitir y posteriormente escoger el género del juego y sus mecánicas para mejorar la transmisión de la temática. El juego es una aplicación web, con arquitectura cliente-servidor que se integrará con la red social Facebook, en la máquina cliente se ejecuta el motor de juego y las entidades en su desarrollo, que son integradas al navegador web. El juego consiste en que el personaje avance en un espacio en

el que puede interactuar con otros distintos, al iniciar el nivel, los diálogos son cargados por procesamiento de archivos XML, creando una estructura de datos, la que será recorrida con la interacción del jugador con distintos personajes. Los conocimientos que intervienen en la aplicación son alimentación, armas, arte, tradiciones, historia, etc.

4.2. Marco Teórico

4.2.1. Perfil psicológico

4.2.1.1. El acogimiento institucional

El art. 232 del Código de la Niñez y Adolescencia, indica que el acogimiento institucional es “una medida transitoria de protección dispuesta por la autoridad judicial, en los casos en que no sea posible el acogimiento familiar para aquellos niños, niñas y adolescentes que se encuentran privados del medio familiar. Este es el último recurso, y se cumple únicamente en aquellas entidades de atención debidamente autorizadas” [19].

Dentro de nuestro país, [20] indica que, durante la ejecución de esta medida, la entidad responsable tiene la obligación de preservar, mejorar, fortalecer o restituir los vínculos familiares, prevenir el abandono, procurar la reinserción familiar del niño, niña o adolescente en su familia biológica o procurar su adopción.

El MIES, señala que, en el mes de junio del 2018, ha brindado atención a una población total de 2552 entre niños, niñas y adolescentes atendidos en las diferentes entidades con convenio, directas y privadas. Las causas principales de su atención, fueron negligencia (29.90%), maltrato (26.84%) y abandono (17.99%), al clasificar esta población por género, se obtiene el (57.48%) correspondiente al género femenino y el (42.52%) al masculino [20].

Dentro de las principales causas por las que se institucionalizan a los menores, se encuentra el maltrato infantil, que [21] define como “los abusos y la desatención de que son objeto los menores de 18 años, e incluye todos los tipos de maltrato físico o psicológico, abuso sexual, desatención, negligencia y explotación comercial o de otro tipo que causen o puedan causar un daño a la salud, desarrollo o dignidad del niño, o poner en peligro su supervivencia, en el contexto de una relación de responsabilidad, confianza o poder. La exposición a la violencia de pareja también se incluye a veces entre las formas de maltrato infantil”.

En [21] se señala, que el maltrato infantil es un problema mundial con graves consecuencias, puede llegar a durar toda la vida. Los diagnósticos actuales de detección, son muy variables, dependiendo del país y del método de investigación utilizado. A pesar de ello, estudios internacionales presentan que, una cuarta parte de los adultos han sufrido maltratos físicos de niños. Otra de las consecuencias, es el que genera sufrimiento en los infantes y en sus familias, causa estrés y se asocia a trastornos de desarrollo cerebral temprano. Incluso dentro de la etapa de gestación, se pueden presentar trastornos en el proceso de neurodesarrollo, motivo por el cual se incluye el maltrato prenatal dentro de los tipos de maltrato infantil [22].

Los adultos pueden presentar mayor riesgo de sufrir problemas conductuales, físicos y mentales, como [21]:

- Actos de violencia (como víctimas o perpetradores).
- Depresión
- Diferentes adicciones
- Comportamientos sexuales de alto riesgo
- Embarazos no deseados

Lo que conlleva a un impacto económico, que englobaría costos de hospitalización por tratamientos por motivos de salud mental, servicios sociales para los infantes [21].

En estudios realizados por [23], a niños en hogares de acogida, se determina que en su perfil psicológico existe alta incidencia en trastornos de tipo externalizante (de manifestación o proyección exterior), definida a nivel conductual con rasgos como la agresividad, trastornos de conducta o los cuadros oposicionista y negativista desafiante que pueden alcanzar hasta un porcentaje del 80 %, y teniendo un promedio del 42 %. Por contraparte, la influencia en la salud mental interna, los trastornos internalizantes (malestares de carácter interno), definida a nivel cognitivo con rasgos como la retraimiento o inhibición, ansiedad, depresión, tienen un menor grado de afección que los externalizantes, pero bastante superior a la población infanto-juvenil general [24].

4.2.1.2. Perfiles psicológicos

Dentro de la investigación desarrollada por [25], realizada en una población de 284 estudiantes escolares en Huancavelica, Perú, gestionado dentro del Proyecto Curricular Institucional (PCI), en el que sus participantes son estudiantes que provienen de familias disfuncionales, en su mayoría constan de antecedentes de maltrato psicológico o físico por parte de sus padres o adultos a cargo. Esto influye en un nivel bajo de autoestima, con actitudes reprimidas e incluso con rasgos de agresividad con sus similares, influyendo en sus habilidades sociales. Las habilidades sociales definidas como “las conductas o destrezas sociales específicas requeridas para ejecutar competentemente una tarea de índole interpersonal. Es decir, son un conjunto de comportamientos interpersonales complejos que se ponen en juego en la interacción con otras personas” [26].

Dentro de las definiciones de habilidades sociales, se pueden encontrar los siguientes tipos: [27]:

- Aceptación de sus iguales: “Usan índices de aceptación de los iguales o popularidad. Entran en estos índices los niños socialmente hábiles que son populares o aceptados en su comunidad. La desventaja de este tipo es que no identifican los comportamientos específicos relacionados con su aceptación” [27].
- Conductual: “Es la adquisición de habilidades interpersonales específicas que permiten experimentar relaciones con otros, que son personal o mutuamente satisfactorias, también tiene influencia en decrecer la probabilidad de castigo bajo ciertas acciones. Tienen la ventaja que se pueden identificar, especificar sus antecedentes y consecuencias de los comportamientos sociales particulares con fines de evaluación e intervención” [27].
- Validación social: “Se incluye dentro de un conjunto de competencias conductuales que posibilitan que el niño mantenga relaciones sociales positivas, afrontando de manera efectiva, y adaptativa, las demandas de su entorno social. Aspectos que contribuyen significativamente a la aceptación por sus compañeros, con un adecuado ajuste y adaptación social” [27]. Se puede referenciar dentro de este tipo, la popularidad, influencias de juicios de los demás niños, hacia sí mismo.

Como indica [28], dentro del rasgo psicológico de Habilidades Sociales en los niños, podrían aplicarse las siguientes evaluaciones para determinar los rasgos en el perfil psicológico del niño:

- Categorías observacionales del Código de la Observación de la Interacción Social (COIS)
- Programa de Enseñanza de Habilidades de Interacción Social (PEHIS)
- Cuestionario de Habilidades de Interacción Social (CHIS): Es un cuestionario que cuenta con 60 ítems donde se analizan diferentes factores, como habilidades básicas de interacción social, habilidades para hacer amigos, habilidades sociales relacionadas

a sentimientos y emociones, teniendo un enfoque en la facilidad de expresar y recibir emociones, ya sea defender opiniones o expresar desacuerdos, y su facilidad de relacionarse con adultos [26].

En [29] se especifica que la aceptación social en la infancia depende en gran medida de sus habilidades sociales, es decir, el grado con que un niño es aceptado y querido por sus similares, midiendo su nivel de adaptación se puede predecir su aceptación futura. Cuando un niño es diagnosticado con un perfil psicológico, bajo en habilidades sociales, es muy importante su modificación de conducta por la relación entre el niño y sus relaciones interpersonales.

En cuanto a los estímulos brindados por sus padres, el presentarse poco afectivos o distantes, lo receptan con poca aceptación presentando inseguridades, posiblemente potenciando características de ansiedad, por contraparte en una actuación de sobreprotección, proyectan una idea muy peligrosa de su ambiente, haciendo que el niño evite interactuar o experimentar acciones en espacios externos, reforzando la dependencia hacia los demás [30].

Dentro de las características internalizantes, se identifica a la inhibición, que [31] la describe como “una variable temperamental que genera timidez, retraimiento, evitación social, inquietud o malestar social, además de miedo ante personas, objetos y/o situaciones desconocidas”, está presente dentro de nuestros primeros meses, cuando se rechaza una interacción con nuevas personas, que no sean de nuestro núcleo familiar, aunque es parte natural del desarrollo. Dentro de la niñez puede ser motivo para dejar de jugar o interactuar con sus similares, posiblemente presente signos de sufrimiento por lo que tratan de estar próximos a sus padres, por la seguridad que sienten. Al no sentirse cómodos en actividades grupales, presentan una respuesta de estrés a las actividades escolares, haciendo incluso que se expresen con pocas palabras, influyendo de manera negativa en su comunicación con niños y adultos, por lo que habitualmente prefieren interactuar con un sólo amigo en su hogar [32].

Dentro de las características externalizantes, la agresividad o violencia escolar se considera un fenómeno mundial, que en los últimos años los actos de violencia y conflictos han ido en aumento sin importar fronteras ni niveles de desarrollo, con lo que se ha convertido en parte de la cotidianidad [33]. Siendo posible suponer que las formas de expresarlo están relacionadas con las condiciones sociales de cada centro educativo y su entorno social, es decir; que varían las formas de expresarlo y su intensidad. La violencia en los niños es únicamente la replicación de acciones observadas en su entorno, siendo un patrón el que los niños maltratados presenten estas acciones, generalmente los niños más violentados, son los más violentos en la escuela, por lo que es necesaria una guía externa en su desarrollo e interacción con similares, ya que los diferentes niveles de agresividad no son correctamente discernidos [34].

4.3. Técnicas de análisis de datos

Una técnica constituye el enfoque conceptual para extraer la información de los datos, y es implementada por varios algoritmos. Cada algoritmo representa la manera de desarrollar una determinada técnica paso a paso, de forma que es preciso un entendimiento de alto nivel de los algoritmos, para identificar cuál es la técnica más apropiada para cada problema. Otro de los factores importantes es el conocer los parámetros y las características de los algoritmos para preparar los datos a analizar [35].

Técnicas de análisis de datos:

Cuadro 4.1 Técnicas de análisis de datos [2]

Técnicas de análisis de datos		
No supervisadas	Clustering	Numérico
		Conceptual
		Probabilístico
	Asociación	A Priori
Supervisadas	Predicción	Regresión
		Árboles de Predicción
		Estimador de Núcleos
	Clasificación	Tabla de decisión
		Árboles de Decisión
		Inducción de Reglas
		Bayesiana
		Basado en ejemplares
		Redes Neuronales
		Razonamiento impreciso
Técnicas Genéticas		

El aprendizaje no supervisado, realiza el proceso de aprendizaje sin la ayuda de un supervisor, trata de ordenar los ejemplos en una jerarquía según las regularidades en la distribución de los pares atributo-valor sin una guía, éste es el proceder de los sistemas que realizan clustering. Otra posibilidad contemplada para estos sistemas es la de sintetizar conocimiento cualitativo o cuantitativo, objetivo de los sistemas que llevan a cabo tareas de descubrimiento [2].

El aprendizaje supervisado usa un atributo especial, normalmente denominado clase, que especifica si el ejemplo pertenece o no a un cierto concepto (que será el objetivo del aprendizaje). El atributo clase, normalmente toma los valores “+” y “-”, que significan la pertenencia o no del ejemplo al concepto que se trata de aprender. El objetivo del aprendizaje supervisado es: a partir de un conjunto de ejemplos, denominados de entrenamiento, de un cierto dominio D de ellos, construir criterios para determinar el valor del atributo clase en un ejemplo cualquiera del dominio. Esos criterios están basados en los valores de uno o varios de los otros pares (atributo; valor) que intervienen en la definición de los ejemplos [2].

Dentro de las bondades del aprendizaje no supervisado, se pueden manejar grandes cantidades de datos, ya que no se necesita tener los registros etiquetados. En el punto de análisis de datos, se debe utilizarlo para tener una visión general de sus características y estructura. Recomienda aplicar un aprendizaje por supervisión solamente en los grupos de salida [36].

Otra característica elemental, es que no se conoce la información de salida, con lo que al aplicar los diferentes algoritmos, se debe definir una cantidad de iteraciones para su funcionamiento máximo o la convergencia del algoritmo.

Dentro de los algoritmos de *clustering*, [36] expone que puede considerarse el más utilizado dentro del aprendizaje no supervisado, consiste en observaciones o vectores de características similares o clusters, en las que se aplica una de las características del aprendizaje no supervisado, que sus valores de respuesta o salida no están definidos; su información de entrada es un grupo de objetos sin etiqueta, y su resultado es un conjunto de grupos en los que sus elementos tienen características similares, es decir, son más similares que los demás elementos del total de datos.

Las fases del proceso de agrupamiento o clustering son las siguientes [37]:

- Preproceso: En algunos casos los datos requieren un paso previo, el preprocesar los datos para facilitar la agrupación o por su alta cantidad de registros. Es importante realizarlo de manera correcta, ya que tiene alta influencia en los datos de salida.
 - Extracción de características: se encarga de minimizar la cantidad de datos de entrada, extrayendo los datos más relevantes.
 - Selección de características: escoge un subconjunto de datos de entrada con valoración predictiva.
- Diseño del algoritmo
 - Medida de similitud o distancia: bajo la selección de una métrica de similitud o distancia, se determina cuánto se asemejan o diferencian los objetos. La técnica afecta directamente a los *clusters* de salida.
 - Criterio para el agrupamiento: se analiza qué criterios deben tomarse en cuenta, dependiendo el tipo de grupos a encontrarse como parámetros de salida.
 - Selección del algoritmo: tomando en cuenta los requisitos del sistema, se debe escoger el algoritmo que mejor se adapte.
- Validación de resultados: al obtener los datos de salida, es necesario verificar que las agrupaciones sean correctas, se valida si el resultado es el esperado, sino se modifican las herramientas de fases previas.
- Interpretación de resultados: en esta fase se analiza el resultado del procesamiento, obteniendo información de los grupos obtenidos o dando soporte al desarrollo de modelos.

Las fases son expuestas en la figura 4.1 para una mejor comprensión.

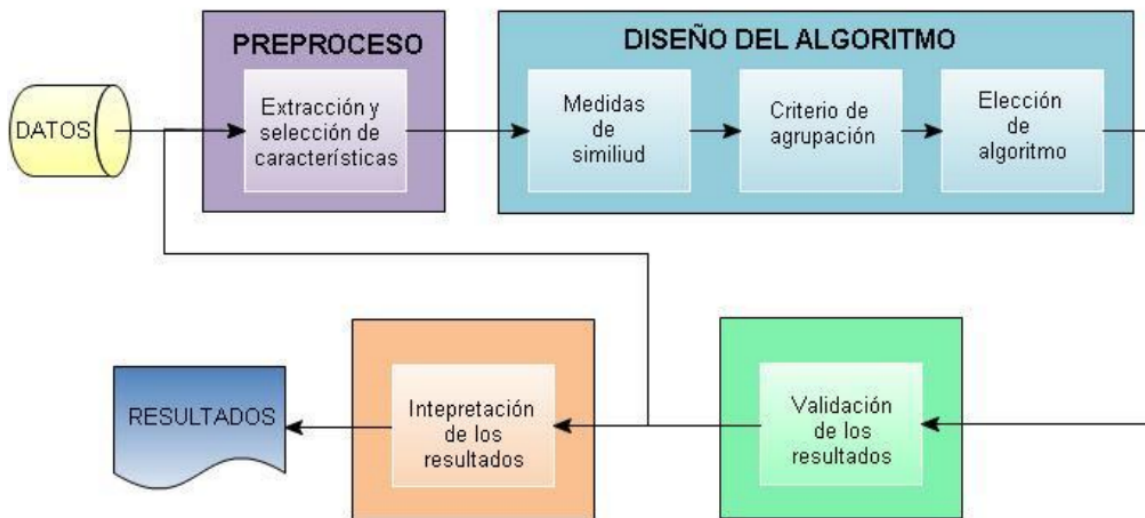


Figura 4.1 Etapas proceso agrupamiento en clustering [38]

Dentro del *clustering* se tienen varias formas de agrupación, una de ellas es la numérica, [39] las relaciona con objetos cuantitativos, pueden ser numéricos como estatura, edad, peso, etc. y no numéricos, como valores booleanos. También se puede encontrar la técnica probabilística, que [38] indica que son objetos generados por distintas distribuciones de probabilidad, en donde se aplican ciertas funciones y modelos, adecuando sus parámetros, su principal desventaja es su complejidad en la asignación de valores de entrada, haciendo que tenga una baja tasa de convergencia. Dentro del tipo Conceptual, [40] indica que parten de la idea básica de un conjunto, planteando que los conjuntos no se crean sino se determinan, aportando información sobre el sentido de pertenencia en los objetos hacia un mismo grupo, de tal manera que, se le puede brindar información extensional con los elementos de cada grupo y por otra parte brindando las características o conceptos para esas agrupaciones.

Dentro de la *asociación*, [41] indica que su objetivo es encontrar asociaciones o correlaciones entre sus elementos, para ello se debe determinar pares que tienen las descripciones atributo-valor, y se les llama *item*, al conjunto de pares *item-sets*. Para encontrar las reglas a aplicarse en la asociación, se debe considerar combinaciones de pares atributo-valor. Algunos *item-sets*, producen más de una regla y otros ninguna, un parámetro para las reglas es la

precisión mínima o *confianza* que dará valores mínimos tolerados. Las reglas de asociación apriori, debe cumplir los pasos dentro de la asociación, con la condición de que todas las reglas satisfacen los niveles mínimos de *confianza*, para obtener los mejores resultados de la asociación.

Los algoritmos supervisados, se dividen en predicción y clasificación, [38] expone dentro de la predicción que “la información de salida es el valor real a estimar (información cuantitativa) y de clasificación, los valores deseados corresponden a las etiquetas de cada caso (información cualitativa)”. Tomando un ejemplo para cada subcategoría, se puede predecir el rendimiento de un atleta, en el que se necesitan datos de peso, estatura, tiempos de duración, etc., y de forma cualitativa para la clasificación, se puede tratar de determinar cuál de las categorías discretas se deben asociar un grupo de caracteres.

Las fases del aprendizaje supervisado son las siguientes [38]:

- Obtención de datos de entrada: sirve para extraer los datos que recibirá nuestro módulo inteligente, debe contar con la cantidad suficiente y cierta diversidad, para que el algoritmo responda de manera adecuada ante el mayor tipo de registros, en ciertos casos debe recibir un pre-procesamiento para facilitar la creación de un modelo aprendido con mayor precisión.
- Aprendizaje: se aplican los distintos algoritmos de aprendizaje y para la obtención de un patrón o la creación de un modelo.
 - Separación de los datos de entrada: consiste en dividir en dos partes los datos de entrada, datos de entrenamiento para entrenar al módulo y los datos de validación o prueba, para probar el modelo generado.
 - Selección de algoritmo de aprendizaje: bajo un análisis de los parámetros disponibles y resultados esperados, se debe escoger el algoritmo con mayor adaptación.

- Entrenamiento: se ejecuta de manera iterativa el algoritmo con los parámetros seleccionados, tomando la diferencia entre el error de salida y el esperado. En cada iteración el modelo se ajustará en función del error, dando como resultado el modelo aprendido ajustado a los datos de entrenamiento.
- Validación de resultados: se realiza una comparación del modelo utilizando los datos de validación, contra los de entrenamiento, obteniendo el error real del modelo. Si el error es inferior al planteado, el proceso de aprendizaje concluye, caso contrario se podrá volver a alguna de las fases iniciales para reajustar el proceso global.

Dentro de la Predicción se encuentran los algoritmos de Regresión, [42] indica que permiten evaluar la relación entre una variable (dependiente) respecto a otras variables en conjunto (independientes), es decir; analiza el impacto de ciertos cambios en variables respecto a sus datos similares. Para evaluar esta relación, se requiere calcular el coeficiente de correlación y el análisis de la varianza, obteniendo el porcentaje de la varianza de la variable dependiente.

Los árboles de predicción, [43] señala que son creados bajo una secuencia de preguntas que dividen recursivamente los datos, generalmente en dos ramas; hasta que una hoja terminal cumpla un criterio de detención o convergencia (por ejemplo, un tamaño mínimo de hoja). La región que limita los datos en el nodo hoja terminal, representa una submuestra específica de los datos completos con propiedades similares. Dentro de esta hoja, se aplica un modelo para obtener la predicción, se puede aplicar especialmente en situaciones donde existen muchas variables que interactúan de manera no lineal.

Dentro de la clasificación expuesta, [44] indica que los estimadores de núcleos logran funciones que se construyen en cada punto del eje real (objetivo del modelo) acorde a los valores muestrales más cercanos, que constituyen un entorno *ventana*. Por ejemplo, que los vecinos más cercanos tengan mayor peso que los más alejados dentro de una *ventana* de datos.

Se pueden utilizar diversas funciones de ponderación (llamadas K o “Kernel”) que están bajo el criterio del analista, ya que debe satisfacer ciertos errores cuadráticos como Error cuadrático integrado esperado (o promedio) (MISE) y Error cuadrático integrado esperado asintótico (AMISE).

Dentro de las técnicas supervisadas, se tiene la categoría de Clasificación, en la que se encuentran los siguientes algoritmos:

En las tablas de decisión, [45] indica que permiten agrupar todas las combinaciones de condiciones y todas las posibilidades lógicas con análisis y comprensión simples, controla que no se omita ninguna alternativa y cubierto todas las posibilidades. Con un lenguaje de tablas normalizado puede utilizarse como un instrumento de programación, por facilitar las tareas del programador, al simplificar la transformación de las condiciones y acciones de las tablas.

Dentro de la técnica de árboles de decisión, [38] señala que son utilizados generalmente en operaciones de búsqueda que requieran un análisis de decisión, ayudan a identificar la estrategia con mayor probabilidad de alcanzar el objetivo planteado en el modelo. Desde su nodo raíz, se consulta el valor de una propiedad particular del patrón, dependiendo de este valor; se dirige al enlace hacia su nodo hijo. Los enlaces son mutuamente distintos, al último nivel de nodo hijo, se lo llama *nodo hoja*, cada uno de ellos, tiene su etiqueta que se presenta como la respuesta de la técnica. Es de fácil interpretación, pero su limitación es la complejidad por el tamaño del conjunto de datos. Además [43] señala que está diseñado para clasificar o predecir datos discretos.

En la Inducción de Reglas, [46] indica que “representan funciones que establecen una relación entre los ejemplos (descritos mediante un conjunto de rasgos) y las clases de decisión. Se expresan de la forma *If P then Q*, donde P es la parte condicional formada usualmente por una conjunción de condiciones elementales (p_1 and p_2 and ... p_k), y Q es la parte

de decisión que asigna un valor de decisión (clase) a un objeto que cumpla la condición”. “Las reglas constituyen patrones que establecen una dependencia entre los valores de los atributos de condición en P y el valor de decisión Q” [46]. Se puede ir agregando reglas P incrementalmente hasta que se cumpla con la condición dada.

La clasificación Bayesiana, [47] define como “una representación gráfica de dependencias para razonamiento probabilístico, en la cual los nodos representan variables aleatorias y los arcos representan relaciones de dependencia directa entre las variables”. La topología de la red representa las dependencias probabilísticas entre las variables, además de independencias condicionales de una variable (o conjunto de variables) dada(s) otra(s) variable(s).

Razonamiento basado en casos, [48] señala que su objetivo se fundamenta en tratar de resolver problemas usando experiencias pasadas. [49] la define como “generación de soluciones a problemas actuales sobre la base de planes pasados, los cuales se almacenan como casos en una memoria, el caso cuyas características se asemejen más a la situación actual, y aplicándole estrategias que lo adapten a las necesidades actuales. Para esta selección se hace necesaria, por lo tanto, una métrica de similitud”.

Las redes neuronales, [38] señala que es una técnica que está inspirada en el funcionamiento del sistema nervioso de los animales, un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Dentro de la informática, son sistemas paralelos de cálculo con gran número de procesadores simples (neuronas o nodos) con muchas interconexiones. Una neurona puede recibir varias entradas, pero devolver una salida.

Las redes neuronales están formadas por los tres siguientes niveles de capas [38]:

- Capa de entrada: una única capa formada por neuronas de entrada, puede tener m neuronas.
- Capa oculta: puede tener varias o varias capas, cada una con n neuronas.

- Capa de salida: una única capa formada por neuronas de salida, puede tener c neuronas.

La salida de una red neuronal es representada por $f(x)$, la complejidad de esta función dependerá del número de nodos ocultos en la red. Una desventaja de este algoritmo es que generalmente sufre un sobre-ajuste (*overfitting*), que en la fase de entrenamiento ha aprendido el modelo, pero no responde de manera adecuada en la fase de validación, lo que conlleva a varios ajustes en el diseño de la red, [38] recomienda para evitarlo, cubrir el mayor número de casos en el entrenamiento o usar la técnica de *cross-validation* que realiza el análisis de datos de entrenamiento y su validación.

El razonamiento impreciso es una lógica que tiene la capacidad de tratar la incertidumbre, es decir; puede representar, medir y procesar información incierta o que los datos no vayan acorde a la común lógica o sean datos ambiguos, por lo tanto, llevan un amplio margen de error y requerirán conocimiento no explícito. Esta técnica tiene la capacidad de procesar datos cuantitativos de la incertidumbre. [50] [51]

Las técnicas genéticas, son descritas por [52] como “algoritmos de búsqueda basados en la mecánica de selección natural y de la genética natural. Combinan la supervivencia del más apto entre estructuras de secuencias con un intercambio de información estructurado, aunque aleatorizado, para constituir así un algoritmo de búsqueda que tenga algo de las genialidades de las búsquedas humanas”. Dentro del análisis de sus variables, las que presenten mejores resultados tendrán mayor capacidad reproductiva; cumpliendo el que a mayor reproducción, perdurarán y crecerán de generación en generación. Se puede introducir elementos aleatorios para la modificación de las variables (mutaciones), iteración a iteración las soluciones “malas” se irán descartando [35].

Capítulo 5

Marco metodológico

5.1. Diseño de interfaz de aplicación móvil

La aplicación está enfocada al diagnóstico ágil de niños entre 5 y 10 años, permitiendo la proyección del niño con partes gráficas del cuerpo humano, las cuales previamente fueron diseñadas con el apoyo de psicólogas y diseñador gráfico, para intentar asociar con el nivel mayor de expresión cada parte del cuerpo a rasgos psicológicos de la personalidad del niño. La aplicación será una herramienta de soporte al personal psicológico, para realizar diagnósticos ágiles. Estará desarrollada para el sistema operativo Android, que es el de mayor alcance.

La pantalla inicial se presenta en la figura 5.1.

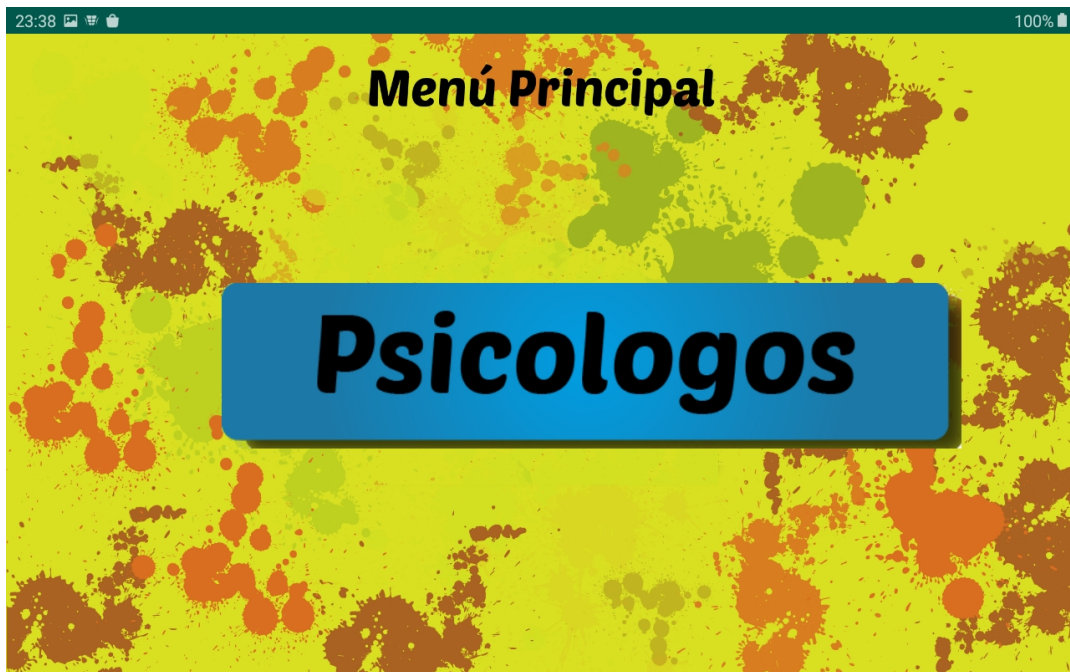
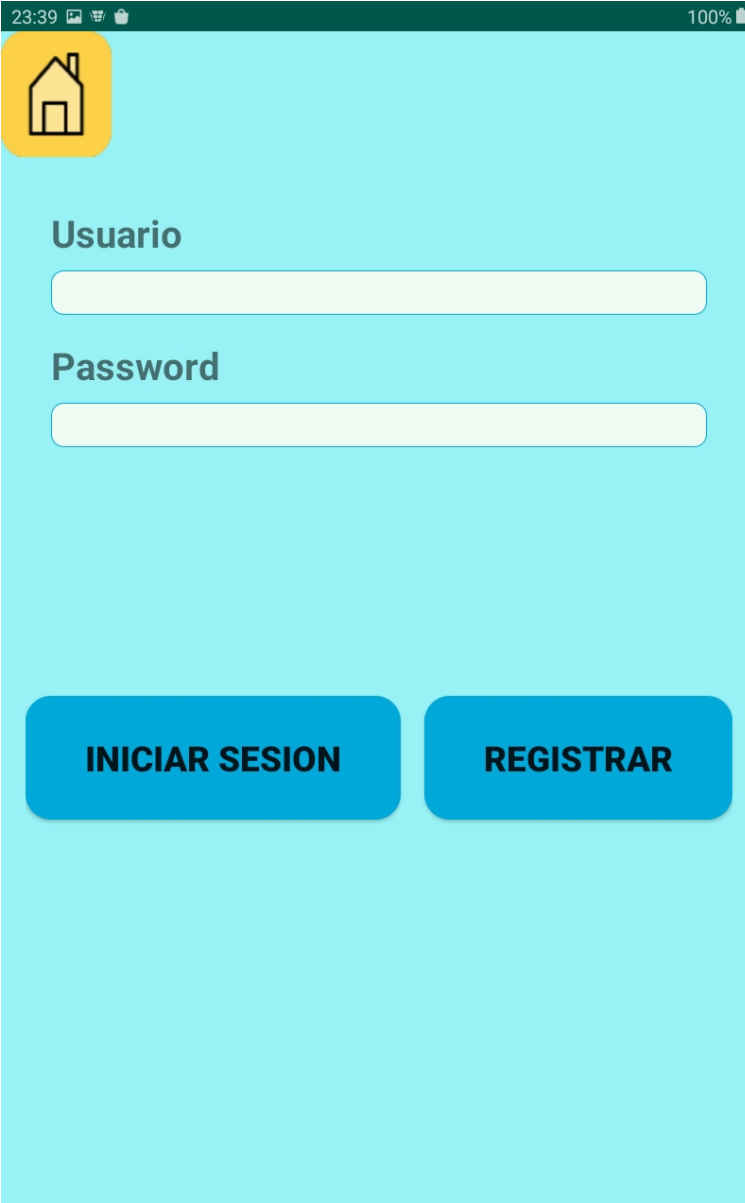


Figura 5.1 Pantalla inicio

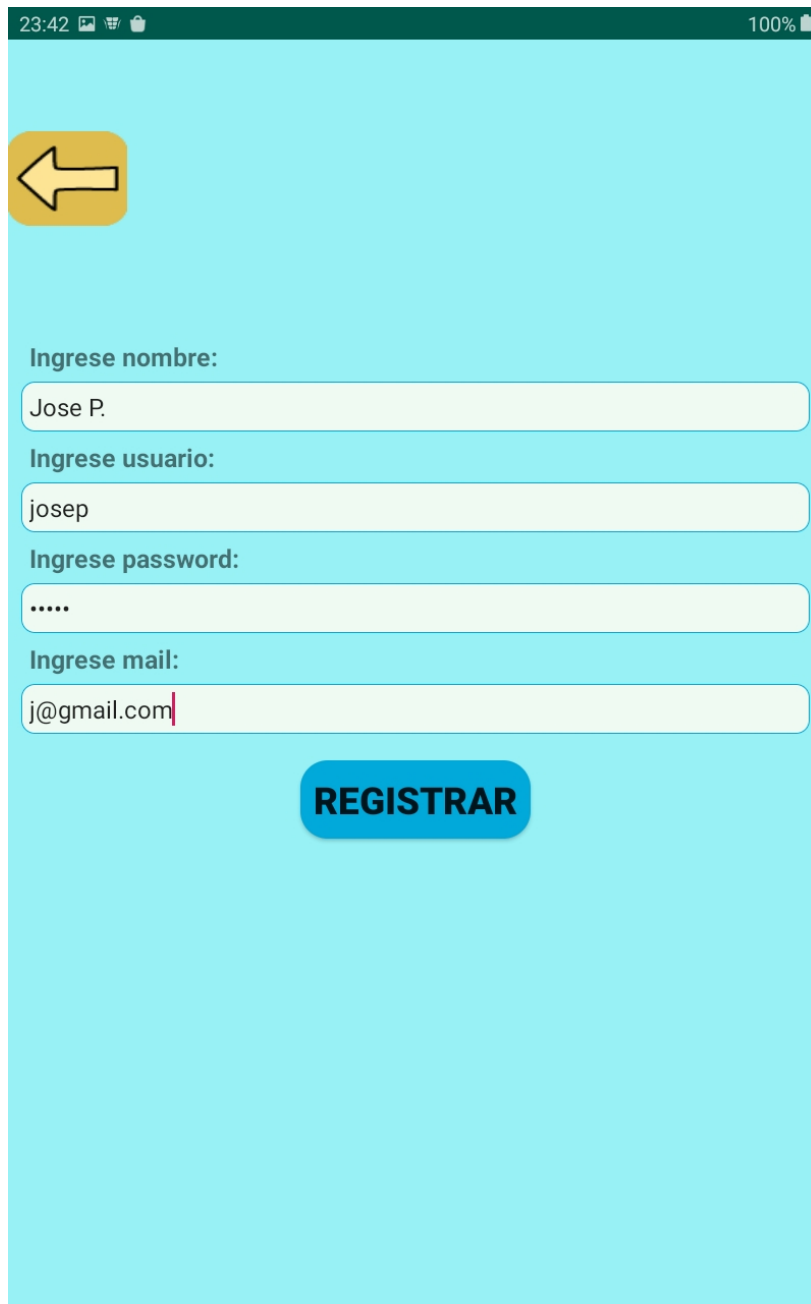
Una vez iniciada la aplicación, al presionar el botón “Psicologos”, es necesario registrar el usuario de la persona que va a desarrollar los test, con lo que es obligatorio el registro del psicólogo. Por ello la siguiente pantalla es la de inicio de sesión y registro de usuario, el que se visualiza en la figura 5.2.



The image shows a mobile application interface for user login. The background is a light blue gradient. At the top, there is a dark green status bar with the time '23:39' and battery level '100%'. Below the status bar is a yellow home icon. The main content area contains two input fields: 'Usuario' and 'Password'. Below the input fields are two blue buttons: 'INICIAR SESION' and 'REGISTRAR'.

Figura 5.2 Inicio sesión

En el que acceden los usuarios previamente registrados, ingresando su usuario y contraseña, seguido de presionar el botón “Iniciar sesión”. En caso contrario, de requerir el ingreso de un nuevo usuario, se debe presionar el botón “Registrar”. El cual nos llevará a la pantalla de registro de usuario, expuesta en la figura 5.3.



23:42 100%

←

Ingrese nombre:
Jose P.

Ingrese usuario:
josep

Ingrese password:
.....

Ingrese mail:
j@gmail.com

REGISTRAR

Figura 5.3 Registro usuario

Este usuario va a registrarse en la base de datos de la aplicación, con un almacenamiento satisfactorio del usuario, se presenta el mensaje indicado en la figura 5.4.



Figura 5.4 Usuario creado

Una vez registrado el usuario, se debe volver al menú de Inicio de sesión con la flecha en la parte superior izquierda, acceder con el usuario, desplegándose el Menú principal. Como se puede observar en la figura 5.5 en el que se presenta, tres opciones de interacción, “Nuevo

niño” en la que se registran en la base de datos de la aplicación, los datos requeridos del nuevo niño que va a participar en el test, son los presentados en la figura 5.6.

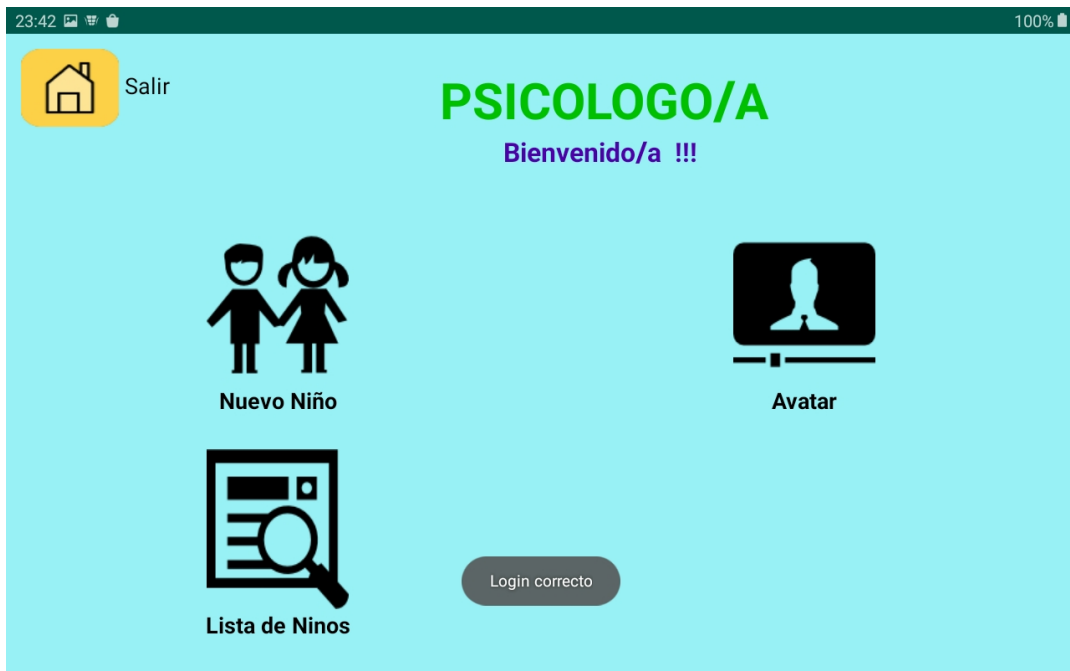


Figura 5.5 Menú principal



Figura 5.6 Registro niño

La segunda opción es la de obtener el listado de niños registrados, se accede desde la opción “Lista de niños”, el formato se lo presenta en la figura 5.7 y se regresa al menú principal con el icono de la casa amarilla en la parte inferior.

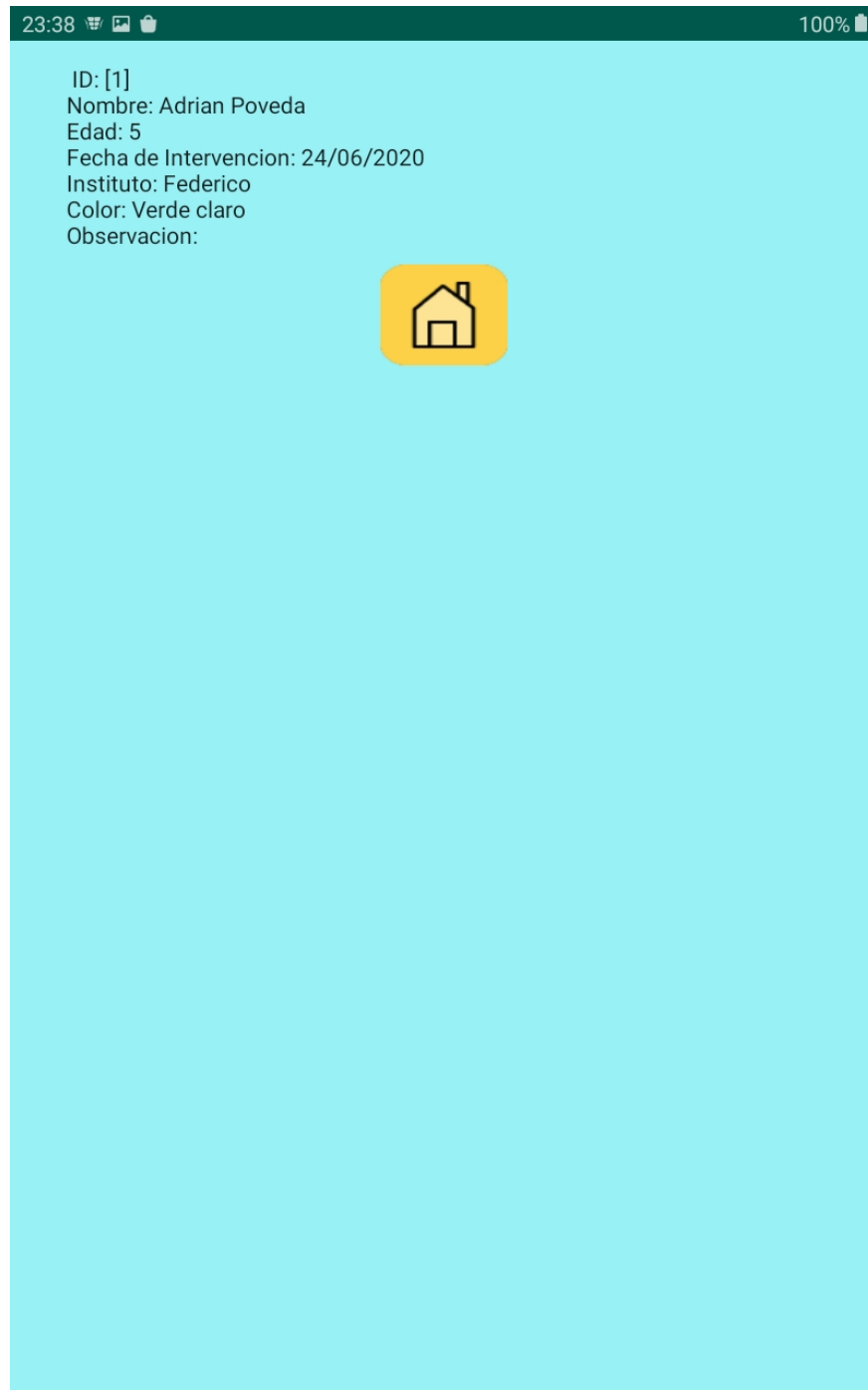


Figura 5.7 Listado niños

La tercera opción es la de ingresar al ambiente de construcción del avatar, desde el botón “Avatar”, donde el niño será el autor de esta actividad. En su primera pantalla, presentada en la figura 5.8, en el que ingresan su nombre y selecciona el sexo con el que se siente identificado, presionando sobre el cuerpo seleccionado.

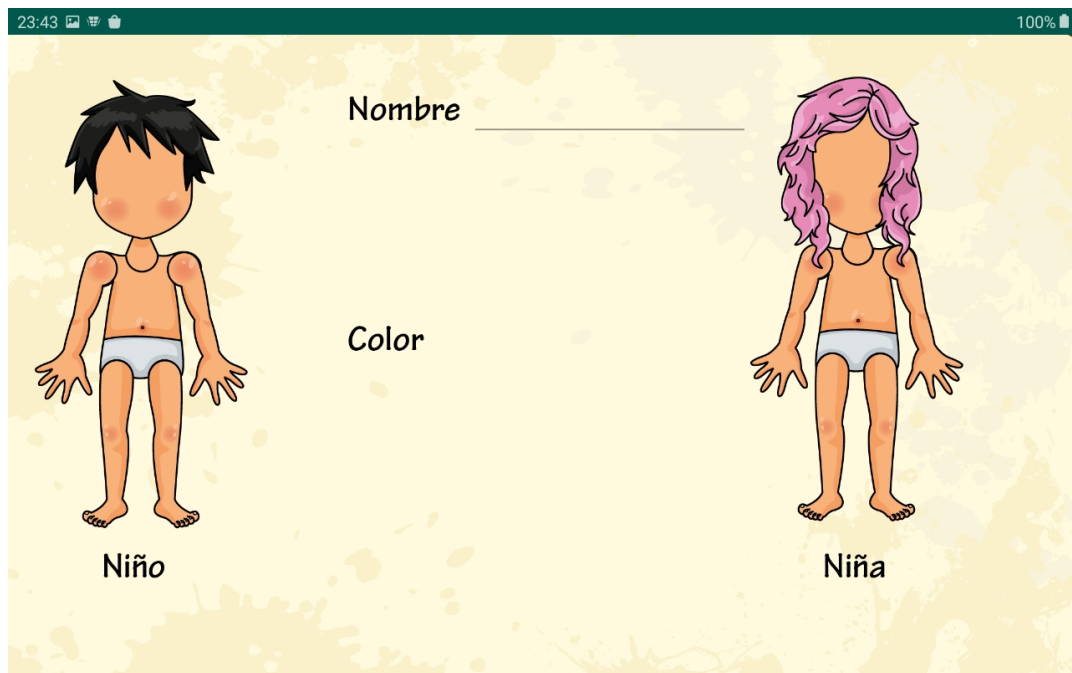


Figura 5.8 Selección sexo

Las partes del cuerpo, están agrupadas en dos categorías Cabeza y, Tronco con extremidades, como se observa en la figura 5.9.

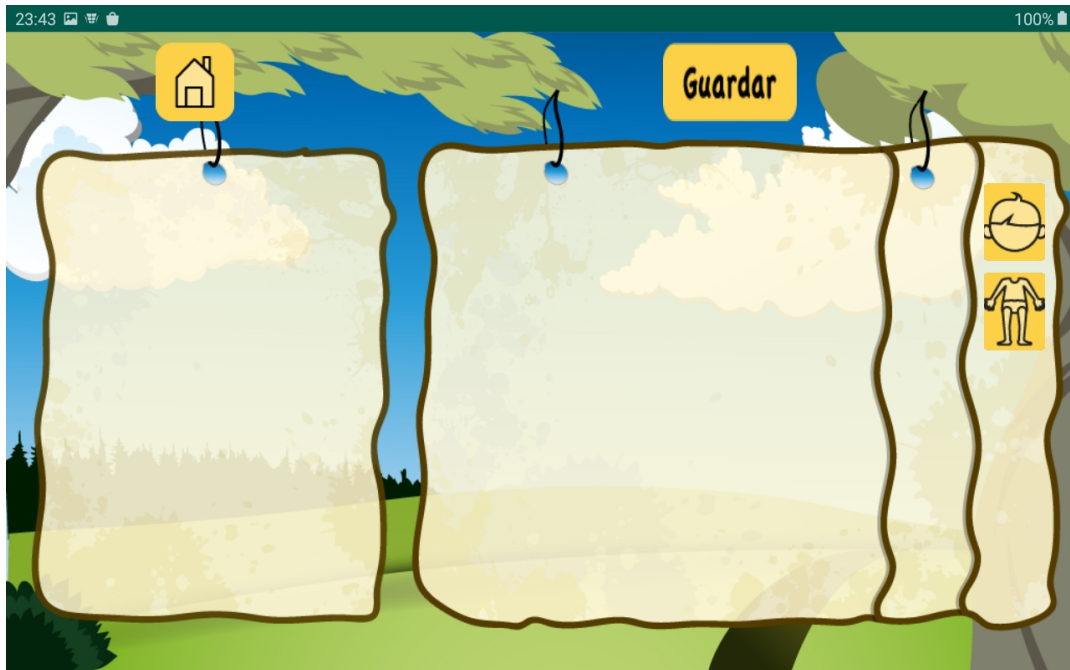


Figura 5.9 Categoría partes cuerpo

Al seleccionar la categoría de la cabeza, se despliegan las subcategorías relacionadas, que formarán el rostro del niño. Se observan, en la figura 5.10, al seleccionarse el sexo masculino, caso contrario, con una selección de sexo femenino, se tiene las subcategorías presentes en la figura 5.11.

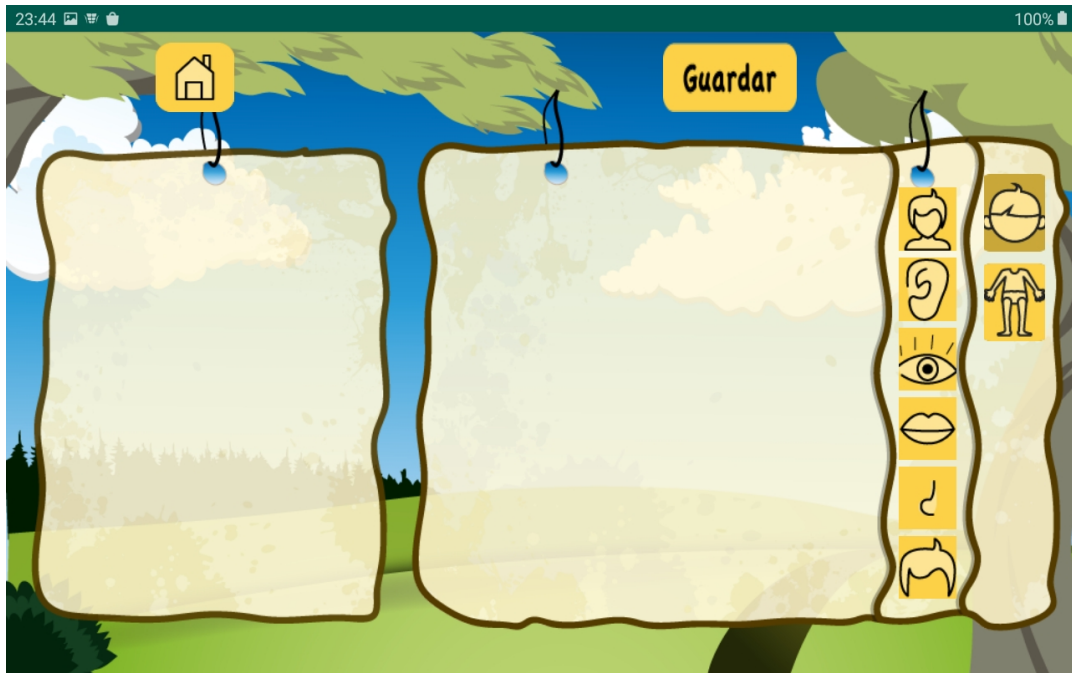


Figura 5.10 Subcategorías partes masculino

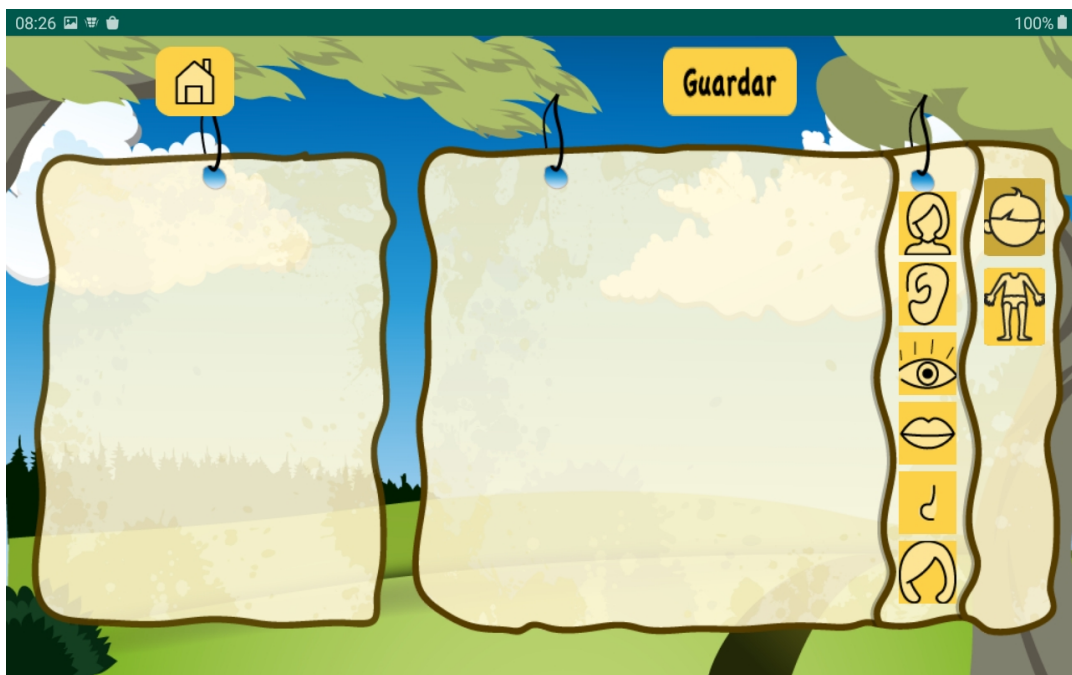


Figura 5.11 Subcategorías partes femenino

Al seleccionar una subcategoría, se visualizan a su izquierda, en un panel, la cantidad de alternativas relacionadas. Las opciones son presentadas de manera aleatoria, para evitar una asociación por orden de visualización del niño. Dentro de las subcategorías, existen partes similares en ambos sexos y partes que son específicas para cada uno, para obtener una mejor proyección para ambos sexos. Las subcategorías que tienen las mismas partes como el rostro, se puede visualizar en la figura 5.12.

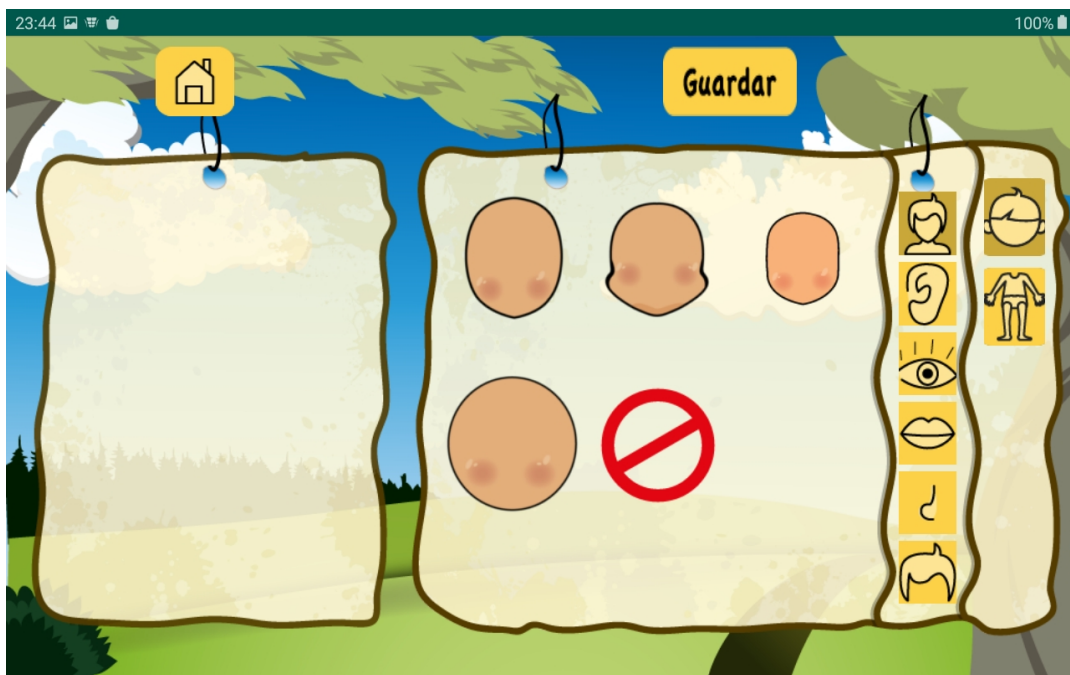


Figura 5.12 Subcategoría rostro

Las subcategorías en las que se tiene partes específicas para cada sexo, son cabello, ojos y torso. Dentro de la subcategoría de los ojos, en la figura 5.13 se visualiza las alternativas para el sexo masculino, y en la figura 5.14 las alternativas del sexo femenino.

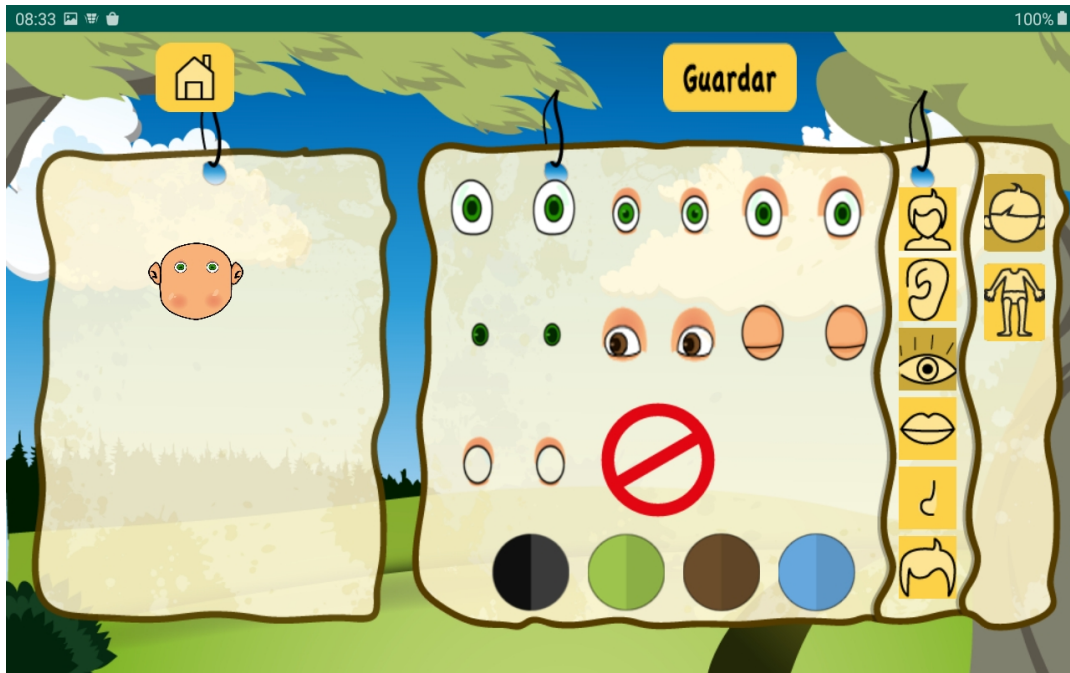


Figura 5.13 Subcategoría ojos masculino

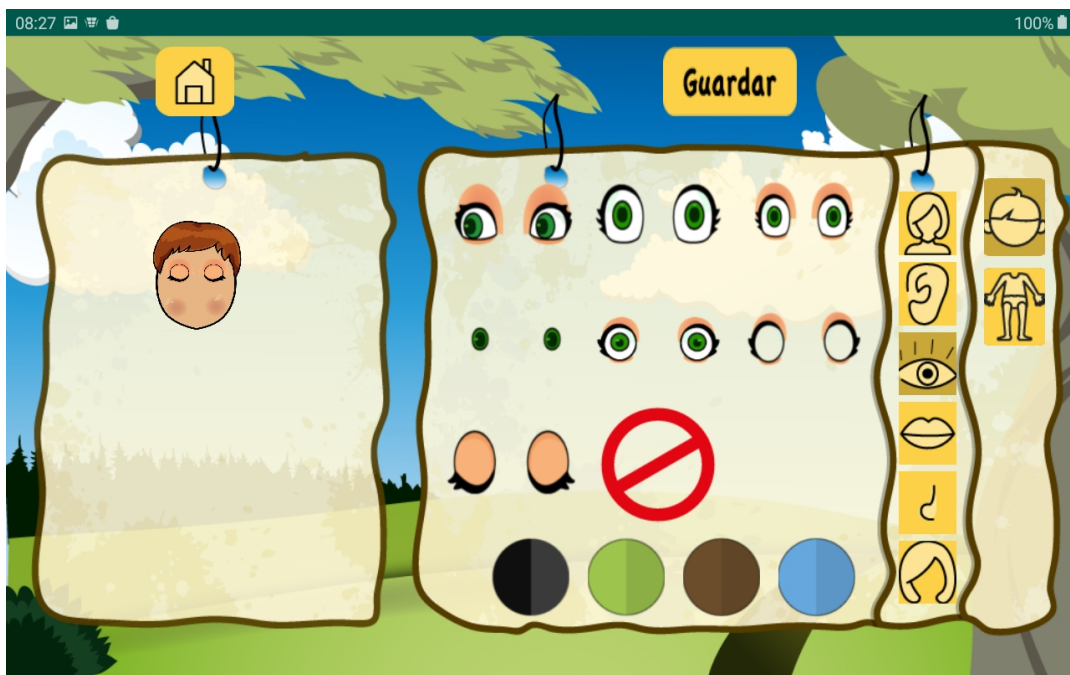


Figura 5.14 Subcategoría ojos femenino

Otra parte en la que existe diferenciación entre sexos, es la subcategoría del torso. En la figura 5.13 se visualizan las alternativas para el sexo masculino, y en la figura 5.14 las alternativas del sexo femenino.

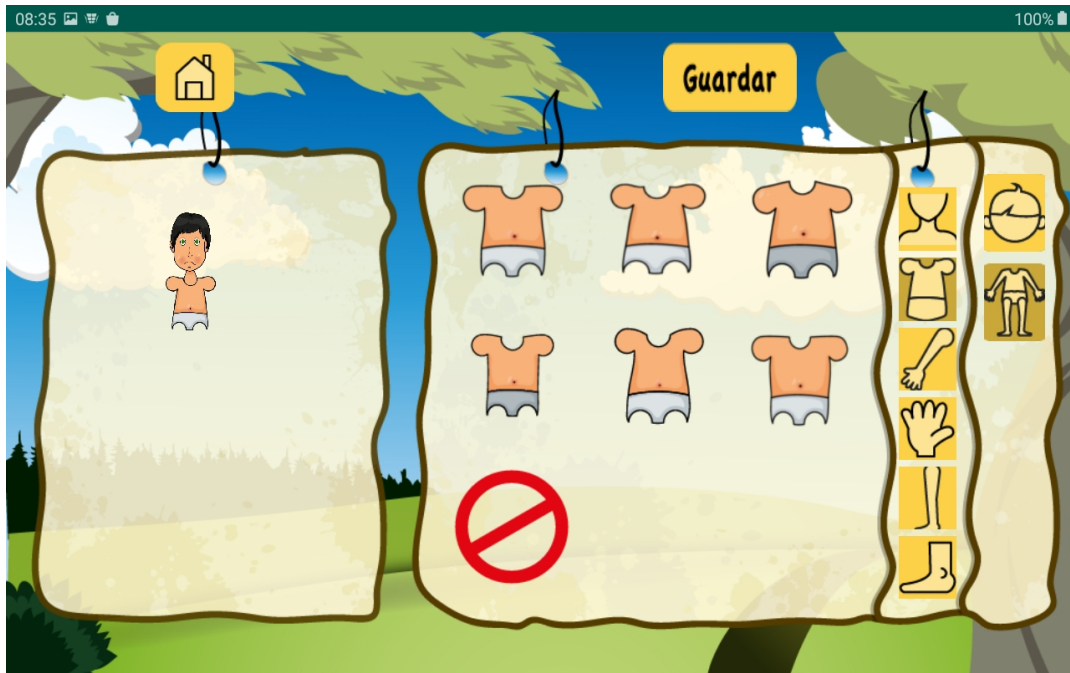


Figura 5.15 Subcategoría torso masculino

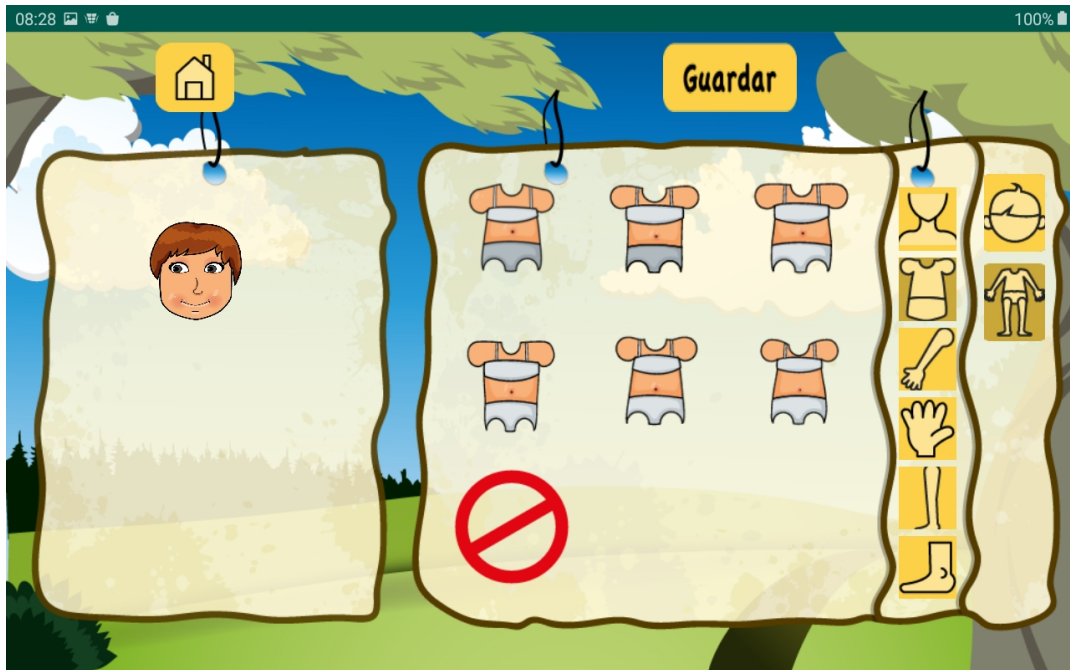


Figura 5.16 Subcategoría torso femenino

Al finalizar la construcción del avatar se visualiza el cuerpo completo, presentado en la figura 5.17. También se le puede incluir vestimentas al avatar, de varias alternativas de prendas.

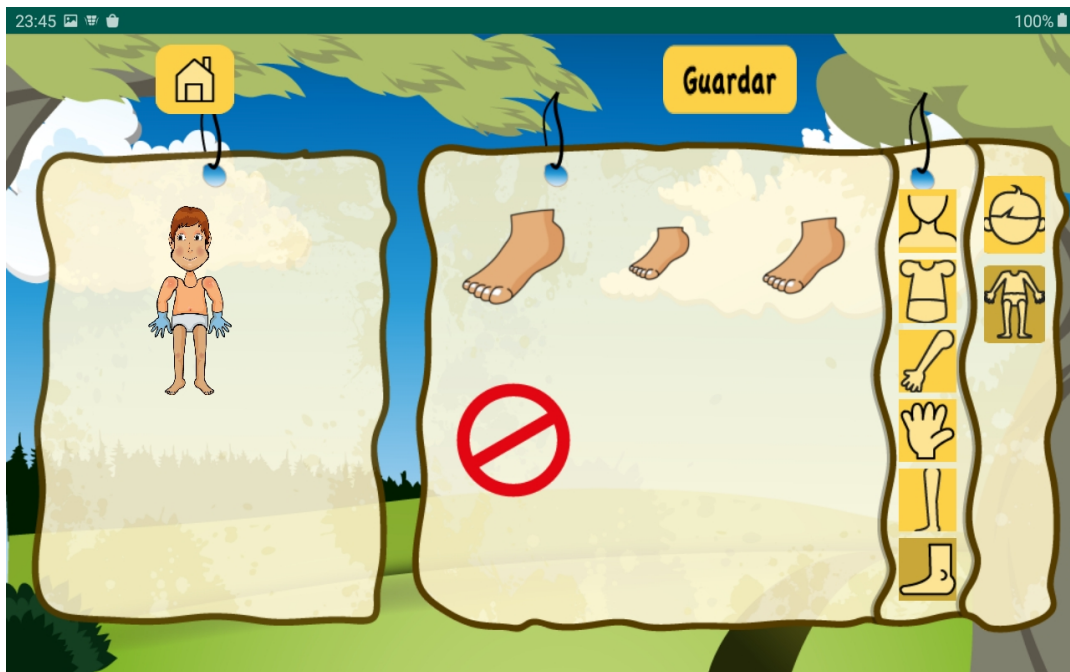


Figura 5.17 Avatar completado

Al finalizar, se debe presionar el botón "Guardar", para almacenar la imagen del niño con el nombre ingresado, para una posterior revisión por parte de los psicólogos. Además, se almacena las selecciones por código de cada parte seleccionada, para su posterior análisis y poderlo utilizar como origen de datos, para el desarrollo del módulo inteligente.

5.2. Diseño de módulo inteligente y preprocesamiento de datos

Dentro del proyecto MicroLudi, desarrollado por la Universidad Politécnica Salesiana en colaboración de la Universidad del Azuay, se obtuvo una base de datos con los registros almacenados de la realización del test de la persona, en que el niño proyecta ciertos rasgos psicológicos expresados en las diferentes partes del cuerpo con el proyecto MicroLudi [1]. La población de los niños, fue de 31 en instituciones de acogida y 71 en instituciones regulares,

en un rango de edad entre 6 y 10 años. Partiendo de esta base de datos, se realizarán análisis estadísticos, para obtener una base de datos con información relevante y estandarizada, para usarse en la experimentación del módulo inteligente.

Dentro del test de la “persona” en el proyecto MicroLudi, es necesario seleccionar 14 partes para proyectar el cuerpo. Las que son enlistadas a continuación:

Cuadro 5.1 Partes del cuerpo Avatar

Partes cuerpo
Boca
Cuello
Cabeza
Ojos
Color ojos
Orejas
Cabello
Color Cabello
Nariz
Torso
Brazos
Manos
Piernas
Pies

Para la mayoría de las partes, existen 7 alternativas de selección, con lo que los cuadros serán creados con esta cantidad para realizar los cálculos y análisis de manera general a todas las partes, en las partes que no se disponga con opciones de selección se asignará el valor de 0.

Se inicia el proceso con un conteo de selecciones por código de las partes, el cual es explicado en los siguientes cuadros. No se han tomado en cuenta en los análisis las opciones que tengan selección de colores, ya que, por sugerencia del equipo de psicólogas, no se pueden asociar los colores a los rasgos de personalidad. Esto aplica, en las opciones de color de cabello y ojos, por lo que posteriormente no se los incluirá en el cuerpo de datos para la experimentación en el capítulo VI.

Cuadro 5.2 Resumen de conteo de selecciones en partes del cuerpo inst. acogida

Parte	1	2	3	4	5	6	7
Boca	1	6	2	10	1	11	0
Cuello	7	16	3	5	0	0	0
Cabeza	7	20	3	1	0	0	0
Ojos	3	4	7	15	5	0	0
Color ojos	0	0	0	0	0	0	0
Orejas	8	6	13	2	0	0	0
Cabello	9	5	7	4	3	3	0
Color Cabello	0	0	0	0	0	0	0
Nariz	12	2	8	3	4	2	0
Torso	10	5	7	5	2	2	0
Brazos	5	13	3	2	3	5	0
Manos	7	12	4	2	5	0	1
Piernas	1	9	3	18	0	0	0
Pies	16	10	5	0	0	0	0

Cuadro 5.3 Resumen de conteo de selecciones en partes del cuerpo inst. regular

Parte	1	2	3	4	5	6	7
Boca	4	13	9	25	1	10	0
Cuello	19	25	11	5	0	0	0
Cabeza	27	22	6	7	0	0	0
Ojos	9	5	16	18	12	0	0
Color ojos	0	0	0	0	0	0	0
Orejas	25	15	17	4	0	0	0
Cabello	17	9	21	7	5	5	0
Color Cabello	0	0	0	0	0	0	0
Nariz	23	7	9	12	2	10	0
Torso	15	17	8	4	9	11	0
Brazos	10	29	7	7	4	5	0
Manos	9	15	13	9	6	0	12
Piernas	4	19	11	28	0	0	0
Pies	34	20	5	0	0	0	0

Existen diferencias en la cantidad de selecciones en las partes del avatar entre los niños de cada tipo de institución.

El cálculo realizado, es la moda, es decir, qué código de cada parte fue la más elegida, tanto para niños de instituciones de acogida como regulares. Los resultados se presentan en el cuadro 5.4.

Cuadro 5.4 Moda de selecciones en partes del cuerpo

Parte	Elem. Inst. Regular	Elem. Inst. Acogida
Boca	4	6
Cuello	2	2
Cabeza	1	2
Ojos	4	4
Color ojos		
Orejas	1	3
Cabello	3	1
Color Cabello		
Nariz	1	1
Torso	2	1
Brazos	2	2
Manos	2	2
Piernas	4	4
Pies	1	1

Otro cálculo aplicado, son las frecuencias, es decir; qué porcentaje de veces han sido seleccionados respecto a los demás niños. Para realizarlo, se inicia desde los cuadros de conteo 5.2 y 5.3, y se obtiene su porcentaje de selección. Para esto, se divide la suma de selecciones por cada código, para la cantidad de niños que realizan la prueba, seguido multiplicarlo por 100 para obtener su porcentaje. En el caso del cuadro 5.2, se cuenta con 31 participantes y en el cuadro 5.3 con 71.

Para una mejor percepción, se realizaron gráficas comparativas de las frecuencias de cada tipo de institución para cada parte del cuerpo.

El análisis de frecuencias en las selecciones de la boca, se expone en la figura 5.18.

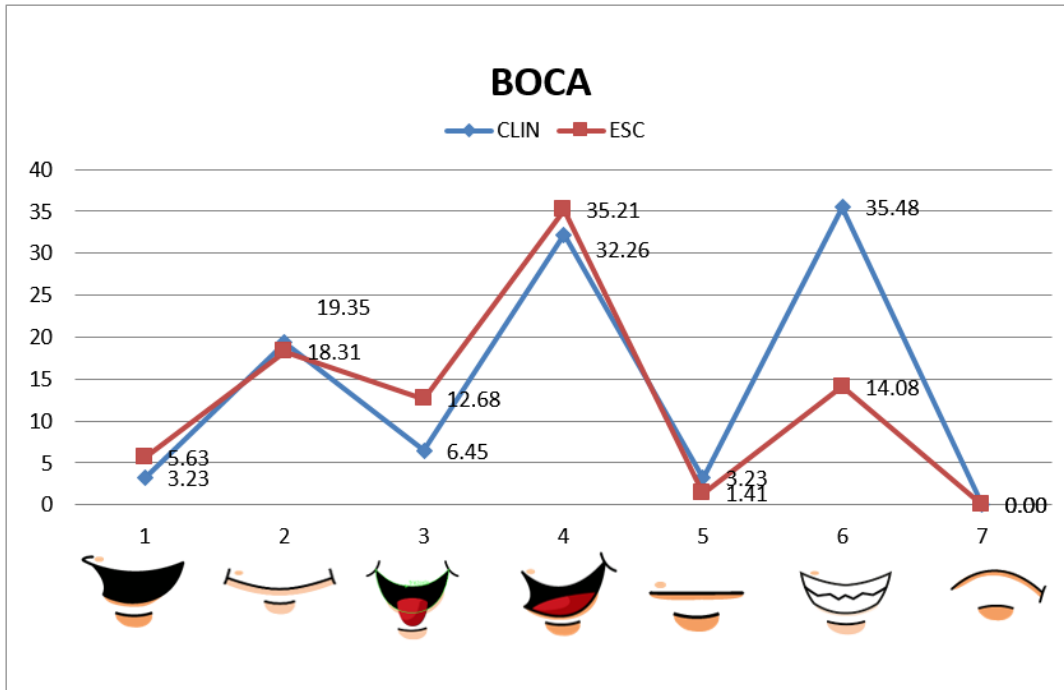


Figura 5.18 Frecuencias Boca

El análisis de frecuencias en las selecciones del cuello, se expone en la figura 5.19.

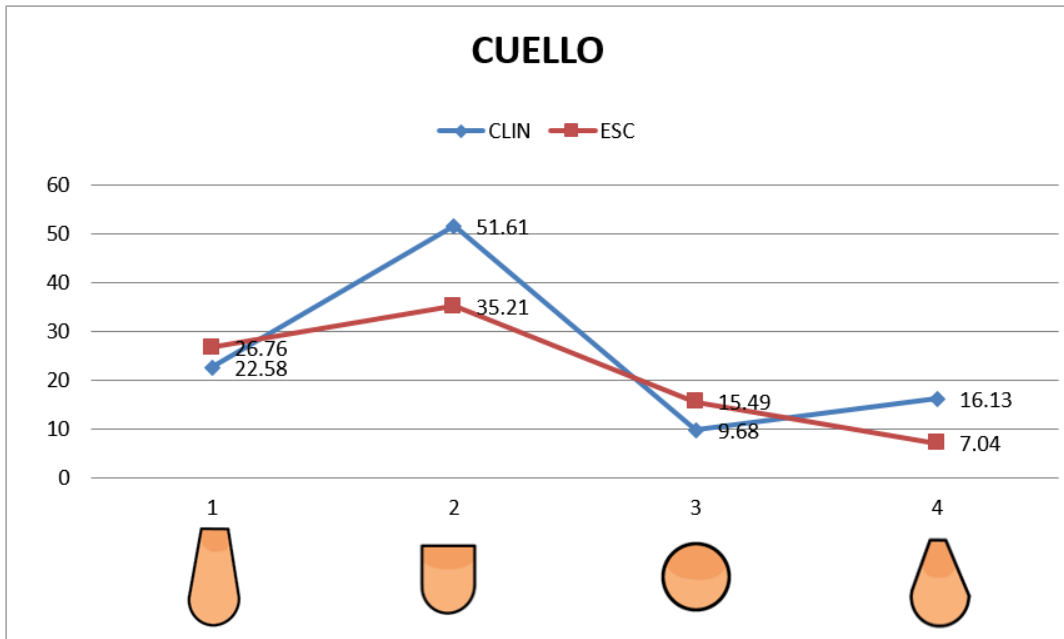


Figura 5.19 Frecuencias Cuello

El análisis de frecuencias en las selecciones de la cabeza, se expone en la figura 5.20.

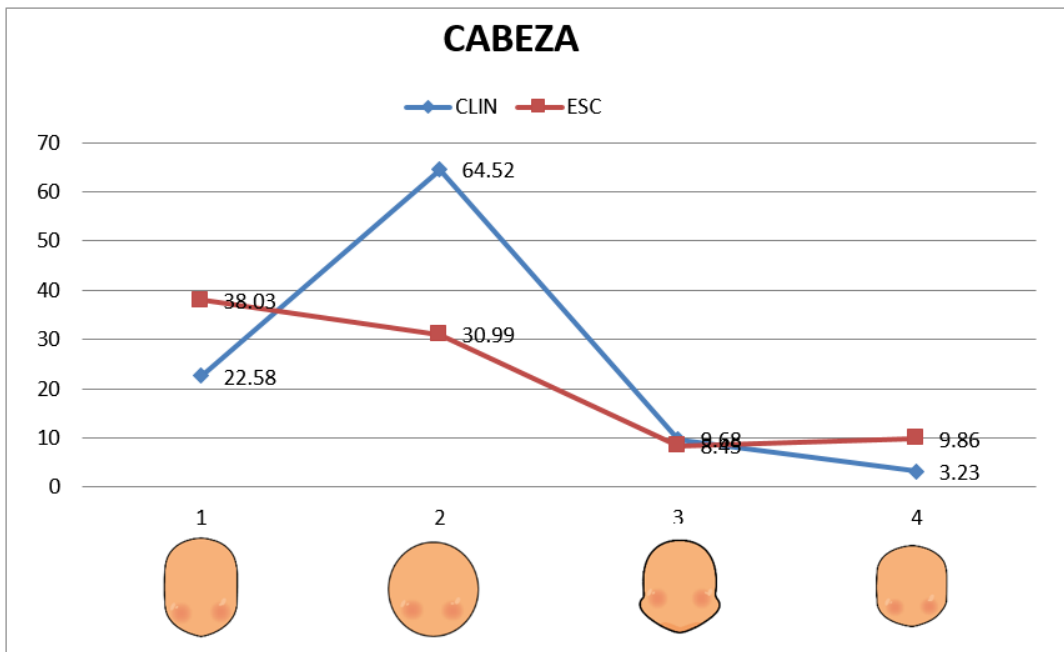


Figura 5.20 Frecuencias Cabeza

El análisis de frecuencias en las selecciones de los ojos, se expone en la figura 5.21.

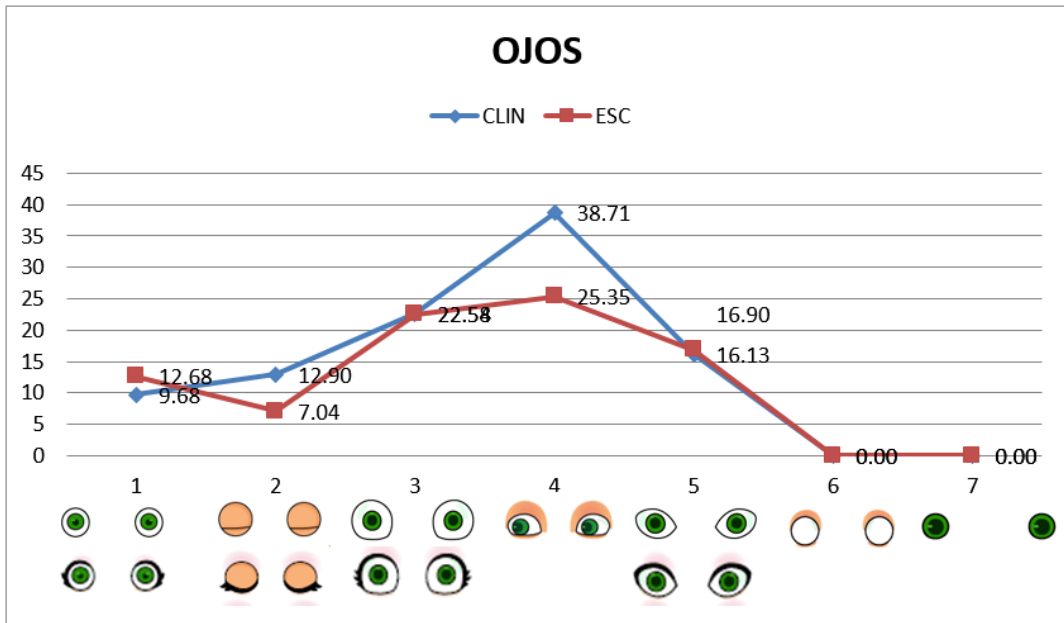


Figura 5.21 Frecuencias Ojos

El análisis de frecuencias en las selecciones de las orejas, se expone en la figura 5.22.

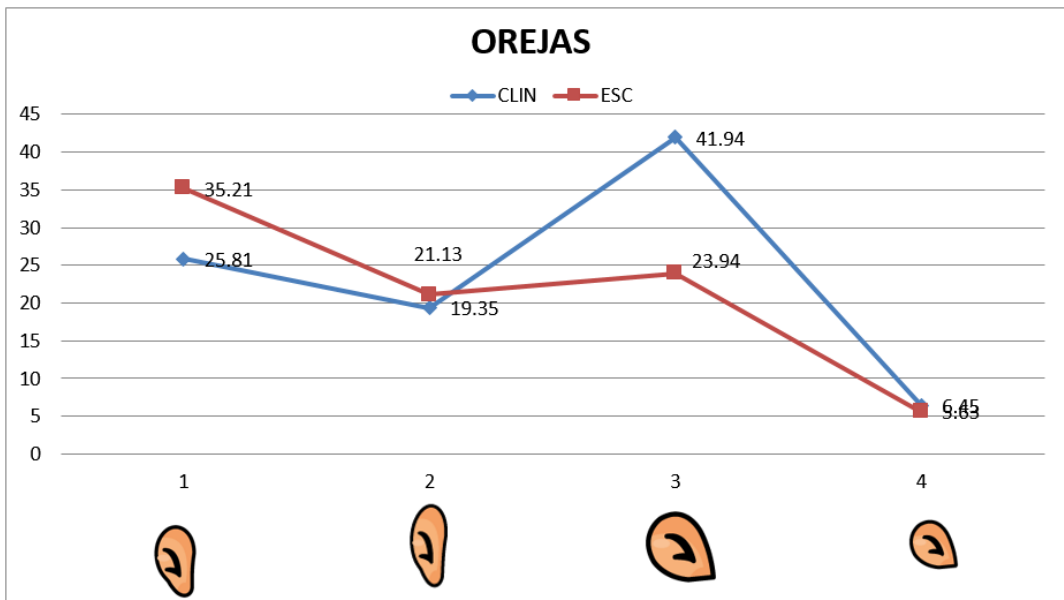


Figura 5.22 Frecuencias Orejas

El análisis de frecuencias en las selecciones de la nariz, se expone en la figura 5.23.

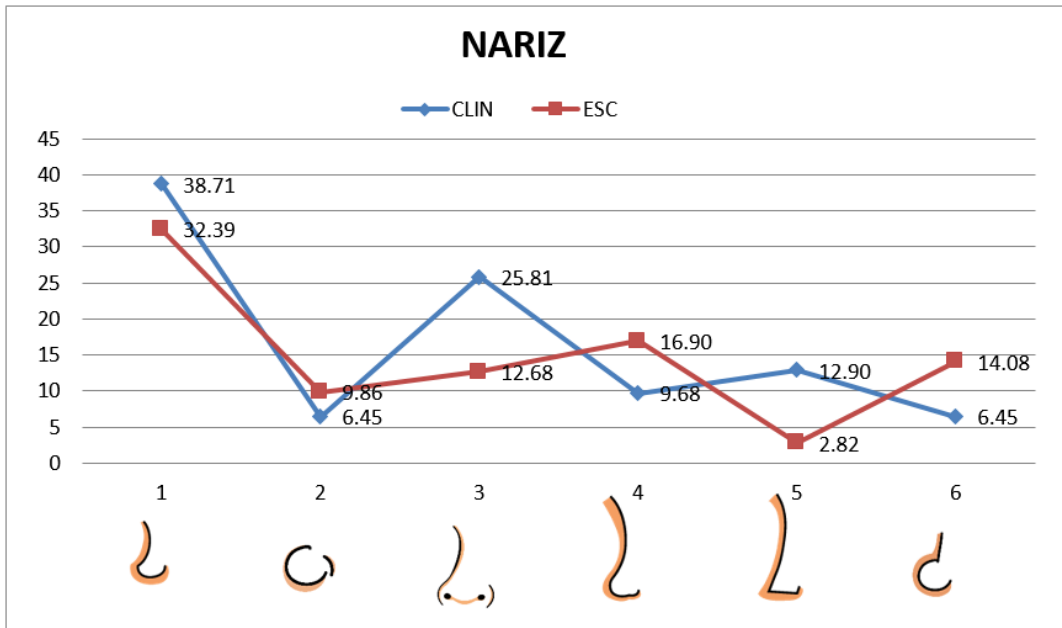


Figura 5.23 Frecuencias Nariz

El análisis de frecuencias en las selecciones del torso, se expone en la figura 5.24.

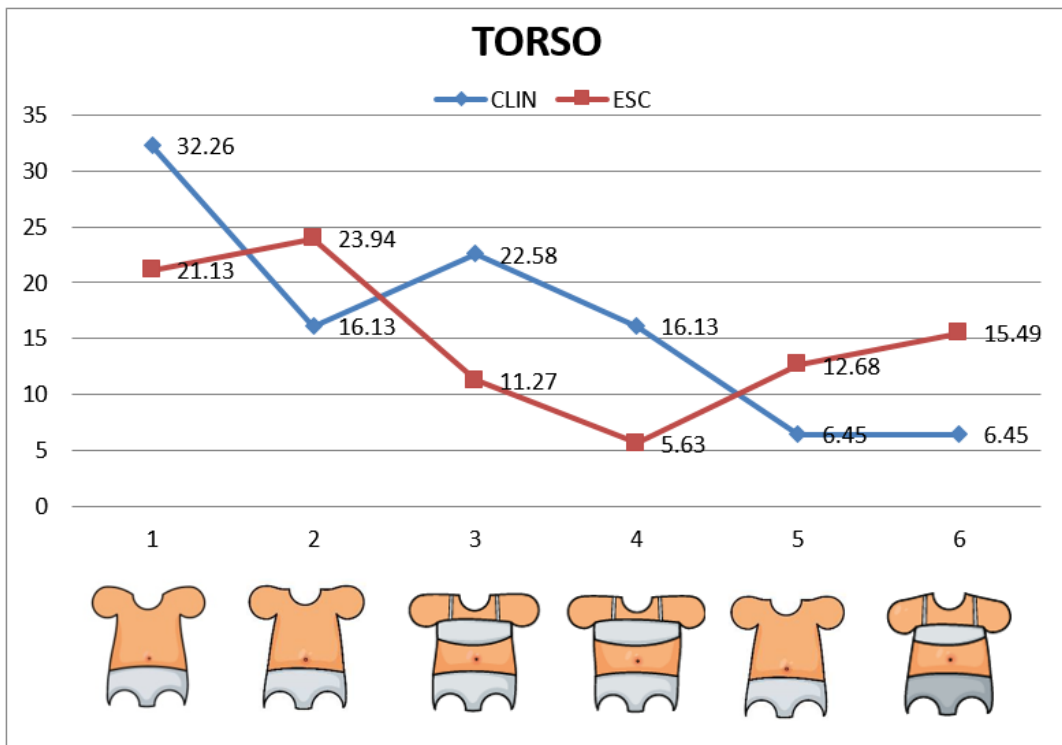


Figura 5.24 Frecuencias Torso

El análisis de frecuencias en las selecciones de los brazos, se expone en la figura 5.25.

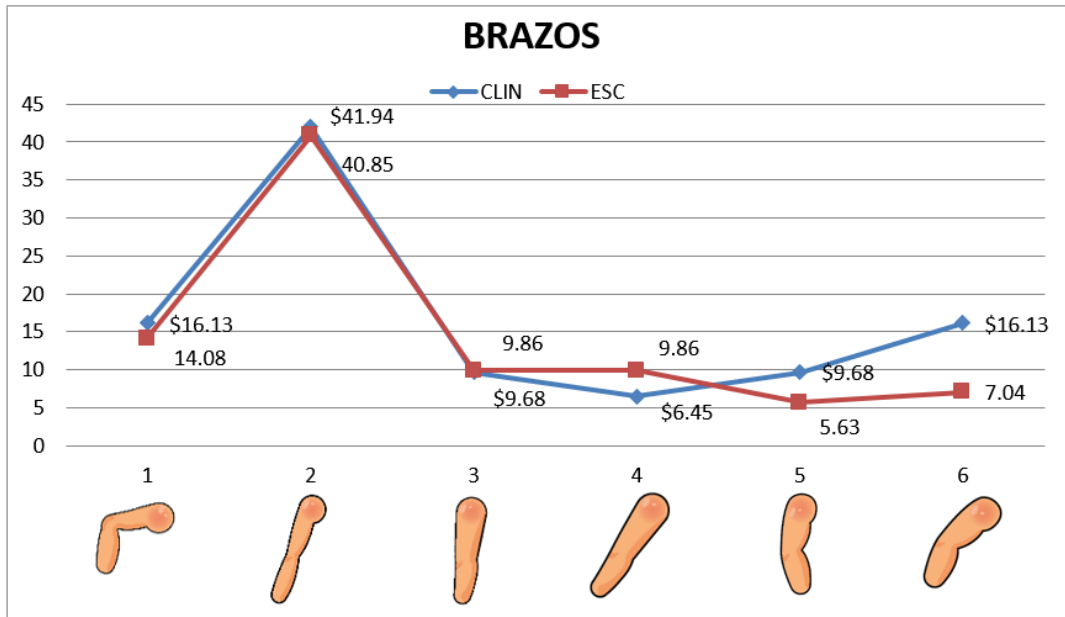


Figura 5.25 Frecuencias Brazos

El análisis de frecuencias en las selecciones de las manos, se expone en la figura 5.26.

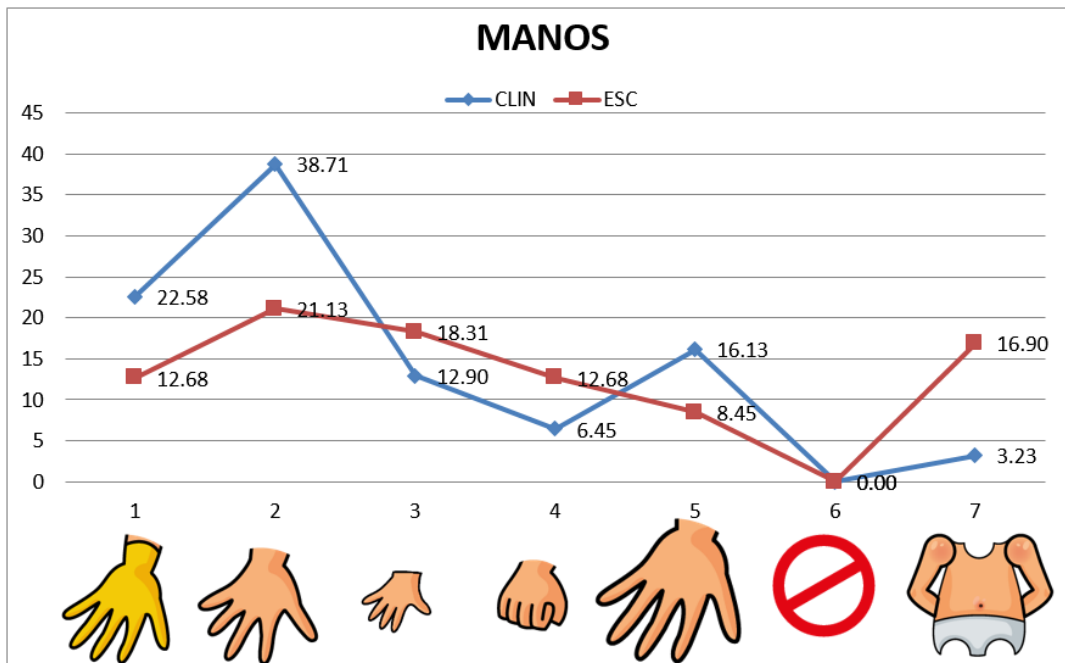


Figura 5.26 Frecuencias Manos

El análisis de frecuencias en las selecciones de las piernas, se expone en la figura 5.27.

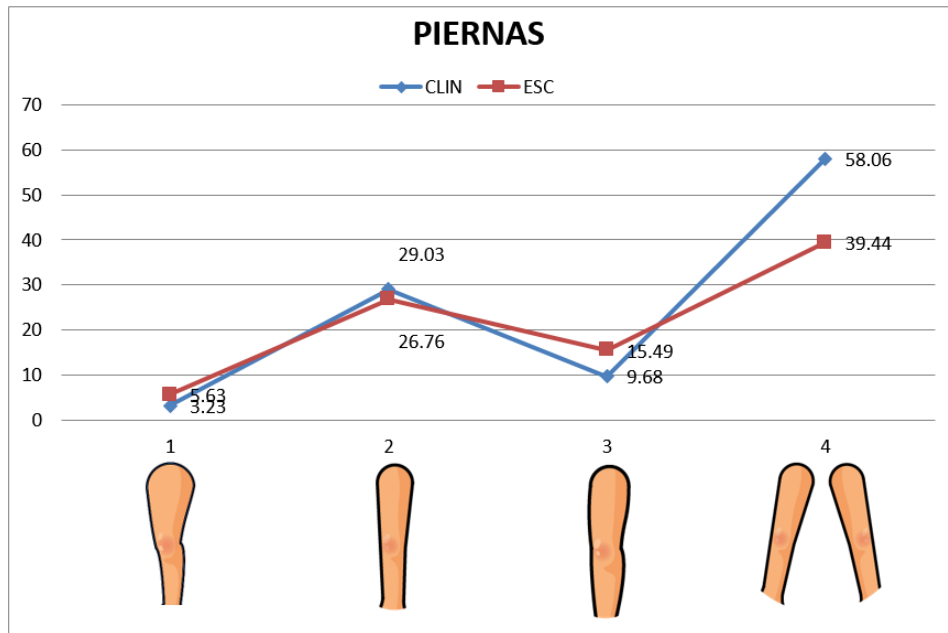


Figura 5.27 Frecuencias Piernas

El análisis de frecuencias en las selecciones de los pies, se expone en la figura 5.28.

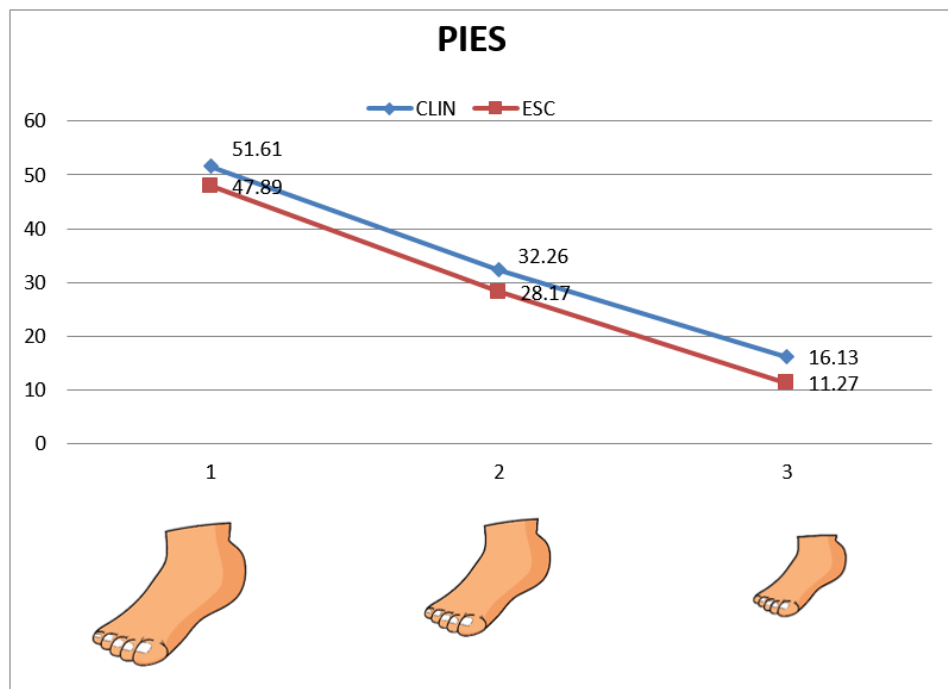


Figura 5.28 Frecuencias Pies

En modo de resumen, en el cuadro 5.5 se presenta una comparativa de las frecuencias para cada parte del cuerpo tomando en cuenta el tipo de institución.

Cuadro 5.5 Resumen comparativo de frecuencias

Parte	Tipo de Institución	1	2	3	4	5	6	7
Boca	Acogida	3.2	19	6.5	32	3.2	35	0
	Regular	5.6	18	13	35	1.4	14	0
Cuello	Acogida	23	52	9.7	16	0	0	0
	Regular	27	35	15	7	0	0	0
Cabeza	Acogida	23	65	9.7	3.2	0	0	0
	Regular	38	31	8.5	9.9	0	0	0
Ojos	Acogida	9.7	13	23	39	16	0	0
	Regular	13	7	23	25	17	0	0
Orejas	Acogida	26	19	42	6.5	0	0	0
	Regular	35	21	24	5.6	0	0	0
Cabello	Acogida	29	16	23	13	9.7	9.7	0
	Regular	24	13	30	9.9	7	7	0
Nariz	Acogida	39	6.5	26	9.7	13	6.5	0
	Regular	32	9.9	13	17	2.8	14	0
Torso	Acogida	32	16	23	16	6.5	6.5	0
	Regular	21	24	11	5.6	13	15	0
Brazos	Acogida	16	42	9.7	6.5	9.7	16	0
	Regular	14	41	9.9	9.9	5.6	7	0
Manos	Acogida	23	39	13	6.5	16	0	3.2
	Regular	13	21	18	13	8.5	0	17
Piernas	Acogida	3.2	29	9.7	58	0	0	0
	Regular	5.6	27	15	39	0	0	0
Pies	Acogida	52	32	16	0	0	0	0
	Regular	48	28	11	0	0	0	0

Para entender cómo se extienden los datos, se encontrará el rango medio entre las frecuencias de cada tipo de institución. Para calcularlo, se resta el valor mínimo del máximo. Aplicando el cálculo a los valores del cuadro 5.5, obteniendo los siguientes resultados:

Cuadro 5.6 Rango medio de selecciones en partes del cuerpo

Parte	1	2	3	4	5	6	7
Boca	2.41	1.04	6.22	2.95	1.82	21.4	0
Cuello	4.18	16.4	5.82	9.09	0	0	0
Cabeza	15.45	33.53	1.23	6.63	0	0	0
Ojos	3	5.86	0.05	13.36	0.77	0	0
Orejas	9.4	1.77	17.99	0.82	0	0	0
Cabello	5.09	3.45	7	3.04	2.64	2.64	0
Nariz	6.32	3.41	13.13	7.22	10.09	7.63	0
Torso	11.13	7.81	11.31	10.5	6.22	9.04	0
Brazos	2.04	1.09	0.18	3.41	4.04	9.09	0
Manos	9.9	17.58	5.41	6.22	7.68	0	13.68
Piernas	2.41	2.27	5.82	18.63	0	0	0
Pies	3.73	4.09	4.86	0	0	0	0

A los valores del cuadro 5.6, se aplicará el cálculo de la desviación estándar. [53] indica que “es una medida de variación de todos los valores con respecto a la *media*, si son cercanos producirán una desviación estándar pequeña, mientras que si son muy dispersos producirán una desviación estándar más grande.” Se representa con s , y se calcula con la fórmula 5.1 [53]:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (5.1)$$

Donde \bar{x} es la media de la distribución, N es la cantidad total de muestras, x_i es el valor en la posición i de la muestra.

Calculando la desviación estándar de cada valor del cuadro 5.6, se obtiene los valores dispuestos en el cuadro 5.7 de rangos medios.

Cuadro 5.7 Desviaciones estándar del cuadro 5.6 de rangos medios

Parte	Desviación Estándar
Boca	7.4
Cuello	6.1
Cabeza	13
Ojos	4.9
Orejas	6.9
Cabello	2.2
Nariz	4.3
Torso	4
Brazos	3.2
Manos	5.8
Piernas	6.7
Pies	2.3

Con los valores de la desviación estándar presentados en el cuadro anterior, aplicando una técnica de normalización sugerida por el área psicológica que trabajó en el proyecto, se podría dar valores de 0 a 4, en función del valor de la desviación estándar correspondiente a la parte que se esté analizando. El valor de 0 en las partes, se mantiene hasta superar el valor de la aplicación de las siguientes fórmulas incrementalmente:

Donde ds , es la desviación estándar de la parte que se analice.

$$1 = 1 * ds \tag{5.2}$$

$$2 = 2 * ds \tag{5.3}$$

$$3 = 3 * ds \tag{5.4}$$

$$4 = 4 * ds \tag{5.5}$$

Utilizando la técnica de normalización expuesta, resulta el siguiente cuadro:

Cuadro 5.8 Valoración normalizada de elementos de partes del cuerpo

Parte	1	2	3	4	5	6	7
Boca	0	2	0	0	0	1	2
Cuello	0	2	1	1	0	0	0
Cabeza	1	2	0	0	0	0	0
Ojos	0	1	0	2	0	0	0
Orejas	1	0	2	0	0	0	0
Cabello	2	1	3	1	1	1	0
Nariz	1	0	3	1	2	1	0
Torso	3	2	3	2	1	2	0
Brazos	0	0	0	1	1	3	0
Manos	1	3	1	1	1	0	2
Piernas	0	0	0	2	0	0	0
Pies	1	1	2	0	0	0	0

Con el cuadro presentado 5.8, se evidencia la importancia en cada código de las partes, el cual se sustituye en la base de datos de selecciones de los niños y posteriormente añadirá la etiqueta del tipo de institución del participante en cada registro, resultando el corpus que se utilizará para la experimentación con las diferentes técnicas de machine learning. El que será explicado con mayor detalle en el capítulo VI (Experimentación y resultados).

Capítulo 6

Módulo de experimentación y resultados

6.1. Levantamiento de requerimientos

Las funcionalidades que deberá satisfacer este módulo, son las de recibir el conjunto de datos obtenido en el capítulo V como su origen de datos, seguido se debe realizar la normalización de sus registros, con la aplicación de cálculos y técnicas estadísticas, para facilitar el procesamiento de los diferentes algoritmos en base a medidas de similitud detectadas, con el conjunto de datos obtenido se debe seleccionar los algoritmos aplicables a los registros utilizando varias técnicas de clasificación, finalmente se debe realizar la interpretación de los resultados y alcanzar una mejor precisión que el proyecto inicial. La definición de la arquitectura y el proceso de construcción de este módulo, se realizará en base a la figura 4.1 (Etapas proceso agrupamiento en clustering), que incluye las siguientes fases:

- Preproceso
 - Extracción de característica
 - Selección de característica

- Diseño del algoritmo
 - Medida de similitud o distancia
 - Criterio para el agrupamiento
 - Selección del algoritmo
- Validación de resultados
- Interpretación de resultados

Al tener un dataset en el que se plantean varios atributos tomados en cuenta por los colaboradores del área psicológica para realizar un análisis de rasgos psicológicos, y no afirmar de manera precisa que las medidas de similitud son las efectivas, por la influencia de varios parámetros o circunstancias presentadas en el entorno de desarrollo de las pruebas aplicadas a los niños, se realizará un proceso reiterativo en la selección de medidas de similitud y elección del algoritmo, hasta alcanzar en la interpretación de resultados la más alta precisión, superando el ratio de 81 % de la fase inicial del proyecto. La fase de selección de algoritmo, se realizará en base a los grupos disponibles para experimentación en la aplicación de machine learning Weka. Al tener dos tipos de clasificación, supervisada y no supervisada, se modificarán los atributos de dataset en función de las técnicas a aplicarse, seleccionando los algoritmos con mayor precisión al dataset optimizado, para realizar una comparación de los resultados obtenidos con evaluación de medidas de similitud y algoritmos.

6.2. Desarrollo del módulo

El dataset normalizado debe especificar valores entre 0 y 1, para brindar una mejor información a los diferentes algoritmos de machine learning, debido a la cantidad de atributos (partes del cuerpo) que son 12, se iniciará con un análisis de estos para obtener las medidas de

similitud, y revelar los atributos con mayor aporte hacia los algoritmos y aplicarlos de manera más eficiente. Para esto se utiliza la herramienta Weka, desarrollada por la Universidad de Waikato en Nueva Zelanda, que en su página web la definen como “un software de aprendizaje automático de código abierto al que se puede acceder a través de una interfaz gráfica de usuario, aplicaciones de terminal estándar o una API de Java. Es usado ampliamente para la enseñanza, la investigación y las aplicaciones industriales, contiene una gran cantidad de herramientas integradas para tareas de aprendizaje automático estándar.”[54]

Weka utiliza para su origen de datos, el formato *.arff* de Attribute-RelationFile Format, el cual consta de una estructura dividida en 3 partes, la cabecera, que contiene el nombre de la relación, la segunda, con la declaración de atributos según su tipo de dato, para nuestro proyecto se utilizan los tipos *Real*, que representan valores numéricos y *Nominal*, que consta del valor de la etiqueta de identificación del grupo al que pertenece el registro, que será útil para una posterior predicción con la aplicación de técnicas supervisadas. La tercera parte, contiene la sección de datos. [55]

Al conocer previamente el tipo de institución de los registros (escolar o acogida), se podrá brindar una etiqueta a cada registro para realizar el entrenamiento del algoritmo, con esto se hará énfasis al uso de técnicas de clasificación supervisada. Dando al atributo del tipo de institución el tipo nominal, para realizar la comparación entre el valor real y la clasificación del algoritmo “entrenado”.

Dentro de la herramienta existen tres módulos, [56] señala que el módulo *Explorer*, es la interfaz más popular para procesamiento de datos. [55] añade que consta de algoritmos de machine learning para varias tareas o *conceptos*, como clasificación, asociación, clustering y predicción numérica, los cuales se pueden aplicar a un solo archivo. Otro de los módulos disponibles, [56] señala que son *Knowledge Flow* o flujo de conocimiento, que los usuarios diseñan y conectan widgets representando componentes WEKA y *Experimenter*, que permite la comparación a gran escala de rendimiento predictivo de algoritmos de aprendizaje.

Dentro del módulo *Explorer*, existen las siguientes pestañas [56]:

- **Preprocess**: donde se carga el conjunto de datos y se aplican los filtros necesarios.
- **Classify**: permite aplicar diferentes algoritmos y técnicas de clasificación, entrenar modelos y evaluar su precisión.
- **Cluster**: aplicación de algoritmos de clustering o agrupación.
- **Associate**: aplicación de algoritmos de búsqueda usando reglas de asociación.
- **Select Attributes**: búsqueda supervisada de atributos representativos.
- **Visualize**: herramienta interactiva de presentación gráfica.

El dataset con los datos normalizados y en formato para procesamiento con Weka, tendrá el nombre de “NormalizadoFull.arff”, el cual será cargado en la pestaña *Preprocess*. Con lo que en base a nuestro atributo nominal se constata la cantidad de participantes en el proyecto en los diferentes tipos de instituciones, 31 en instituciones de acogida y 71 en instituciones regulares. Se puede comprobar en la figura 6.1.

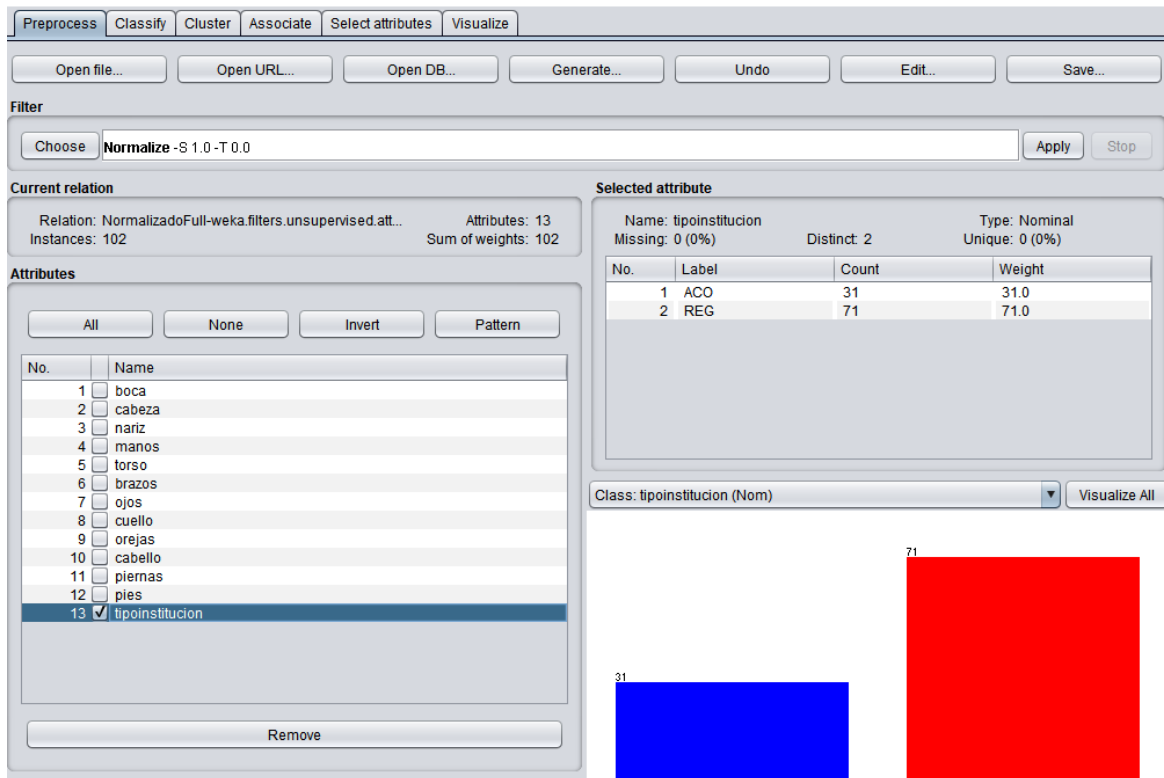


Figura 6.1 Carga de dataset en herramienta Weka

El dataset consta de 12 atributos, que ciertas partes tenían una amplia similitud respecto a la selección en los dos tipos de institución con lo que dentro del módulo de *Select Attributes* se analizaron los atributos. Se utilizó el evaluador *CorrelationAttributeEval* para medir la correlación de los atributos individuales con la clase (valor nominal), en combinación con el método de búsqueda *Ranker* que clasifica los atributos por sus evaluaciones individuales.

Se usa tres configuraciones en el modo de selección del atributo, inicialmente se aplica la técnica de Cross-Validation, que [57] señala que su forma de aplicación se basa en el número de *Folds* k , dividiendo a la totalidad de registros del conjunto de datos en la cantidad k , seguido se construye el clasificador con las $k-1$ partes restantes y se realiza pruebas con esa.

En la figura 6.2, se presenta el resultado de la aplicación de cross-validation con 5 folds.

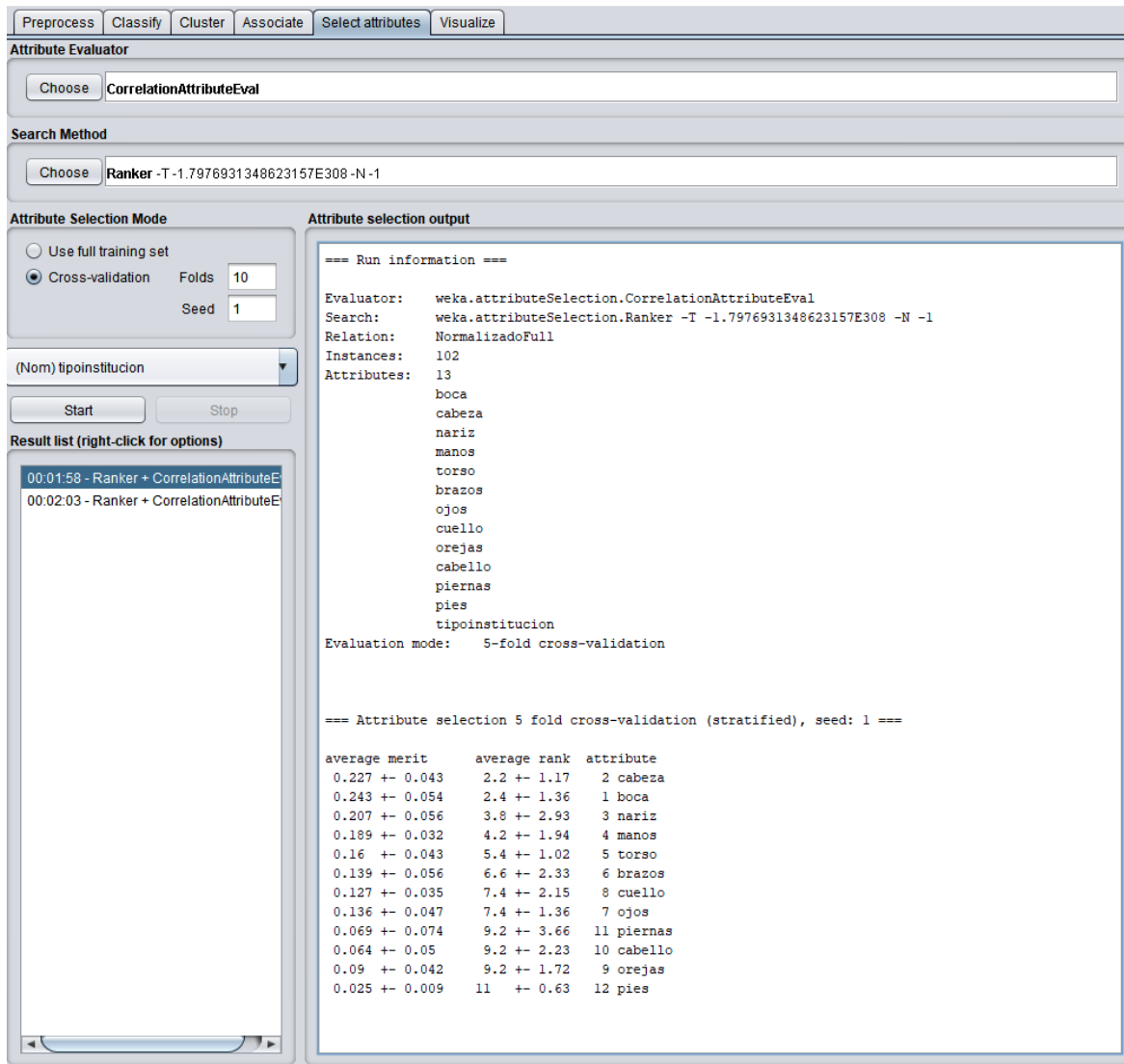


Figura 6.2 Análisis de atributos aplicando Cross Val 5 Folds

En la figura 6.3, se presenta el resultado de la aplicación de cross-validation con 10 folds.

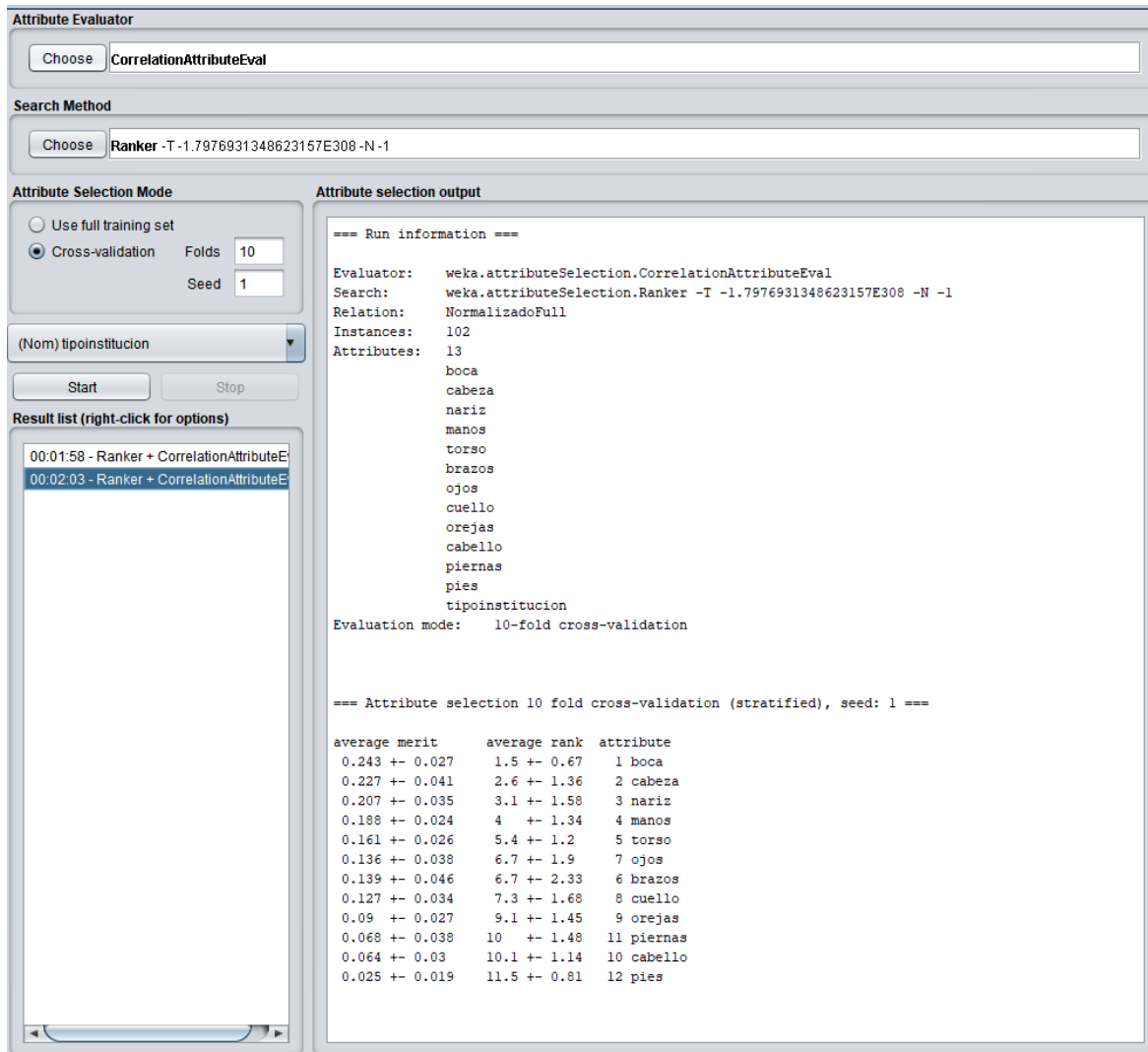


Figura 6.3 Análisis de atributos aplicando Cross Val 10 Folds

Otra configuración utilizada es *Use Full Training Set*, que directamente se realizará el análisis con todos los registros en el dataset. En la figura 6.4, se presenta el resultado de la aplicación.

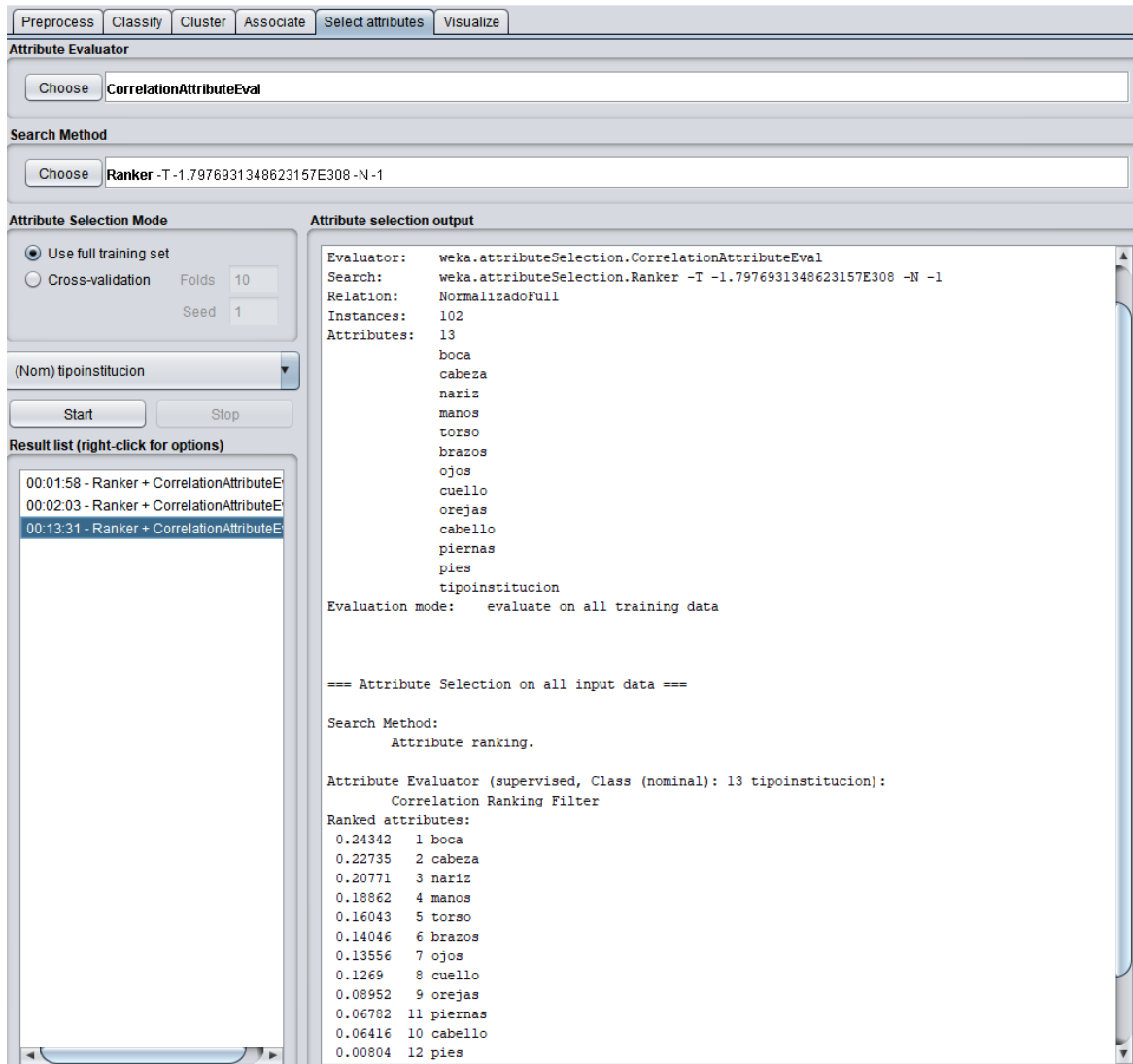


Figura 6.4 Análisis de atributos aplicando Full Training Set

Comparando los 3 análisis de atributos realizados con las técnicas aplicadas, existen pequeñas variaciones en los valores de correlación. En todos, los 4 atributos que tienen menor influencia son Orejas, Pies, Piernas y Cabello, que estarían debajo del 0.1 en los valores del cálculo, con lo que se retirarán del dataset. El nuevo archivo con 8 atributos, tendrá el nombre de “NormalizadoSeleccion8.arff”.

Existen dos tipos de técnicas de machine learning, Supervisado (módulo *Classify*) y No supervisado (módulo *Cluster*).

Se inicia con la aplicación de técnicas de clustering, al utilizar un aprendizaje no supervisado, el dataset necesita que no se especifique una etiqueta o valor nominal. Se indica en [58] que los algoritmos de agrupamiento buscan grupos de instancias con características similares en base a los atributos reales, según un criterio de comparación. Con lo que se crea el archivo con esta especificación con el nombre de “NormalizadoSeleccion8SinLbl.arff”

El algoritmo EM, en que [58] indica que "permite buscar el número de grupos más apropiado, asigna a cada instancia una distribución de probabilidad de pertenencia a cada cluster. El algoritmo puede decidir cuántos clusters crear, basado en Cross-Validation o se le puede especificar a priori cuantos debe generar". Con lo que se lo utilizará para determinar el número de clusters recomendados en nuestro dataset.

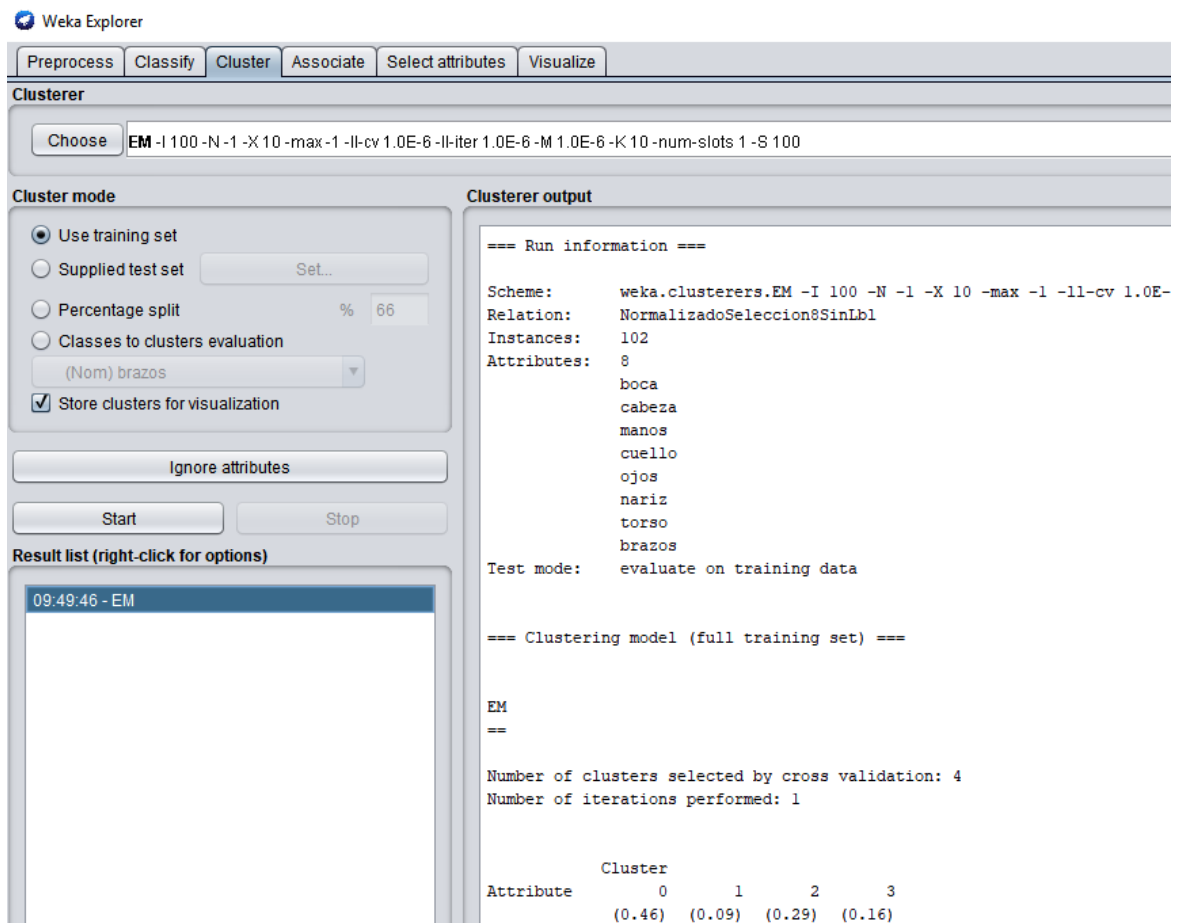


Figura 6.5 Aplicación algoritmo EM 8 atributos

Se observa en la figura 6.5, que lo recomendable es utilizar 4 clusters, con los grupos de pertenencia detectados en los valores del dataset. Con lo que no sería factible aplicarlo, ya que se dispone de una clasificación de los registros en dos tipos de instituciones.

Otro tipo de análisis en el clustering, es el agrupamiento simbólico, en el que [58] señala que se realiza un análisis cualitativo que se estructura jerárquicamente para organizar los datos. Con estas categorías se forma un criterio probabilístico de “utilidad” alcanzando homogeneidad de los valores en los atributos. Logrando separarlos en cada categoría, formando un árbol de conceptos. Dentro de esta categoría se tiene el algoritmo COBWEB, que realiza un *clustering jerárquico* al utilizar un aprendizaje incremental, realizando agrupaciones instancia a instancia, que durante su ejecución va formando un árbol de clasificación, donde las hojas representan segmentos y el nodo raíz engloba por completo el conjunto de datos de entrada.

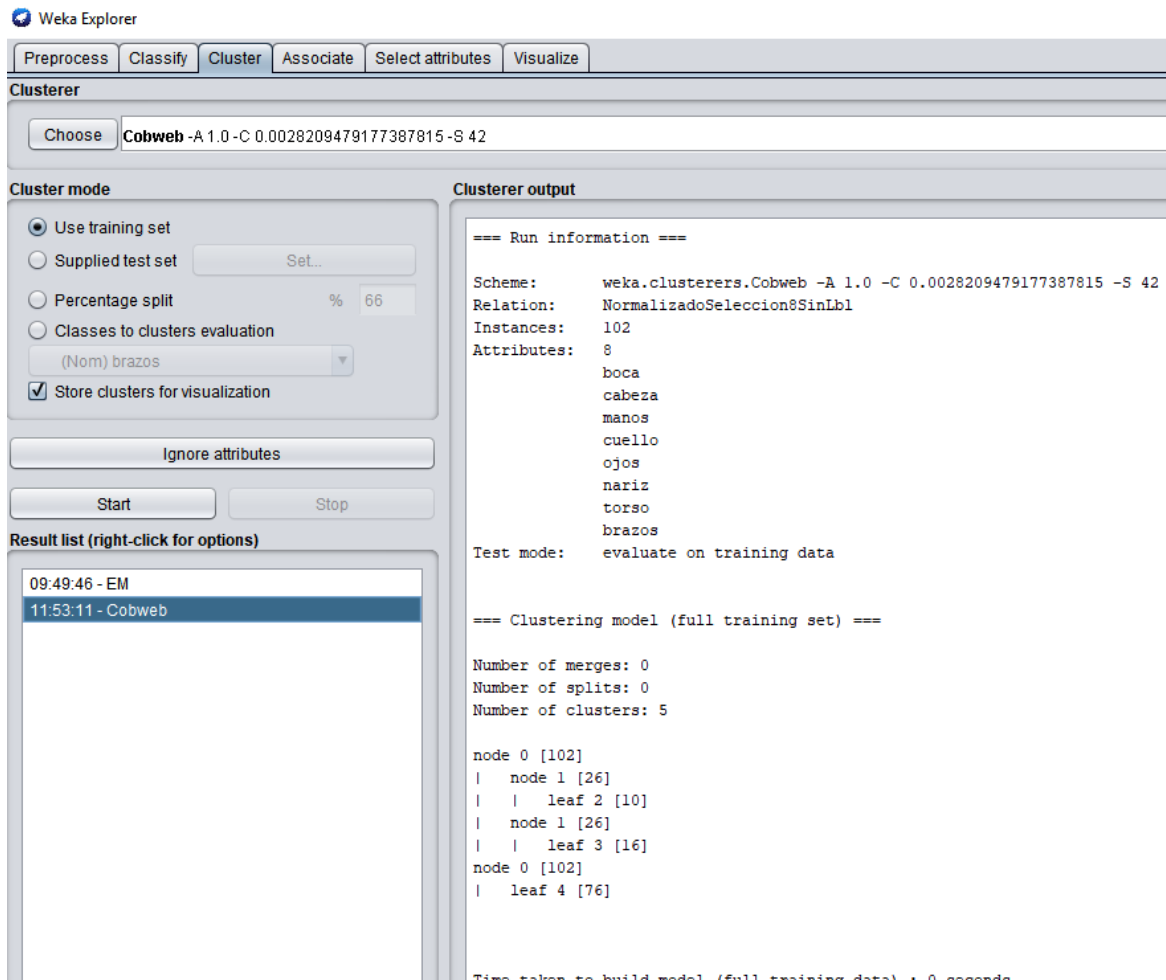


Figura 6.6 Aplicación algoritmo COBWEB 8 atributos

Se observa en la figura 6.6, que lo recomendable es utilizar 5 clusters con su estructura identificada, en los grupos de pertenencia detectados en los valores del dataset. Con lo que tampoco sería factible aplicarlo, ya que están clasificados los registros en dos tipos.

Se aplicará el método de agrupamiento numérico Simple K-means, [58] indica que “es hasta ahora el más utilizado en aplicaciones científicas e industriales, representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. Usa atributos únicamente numéricos, su función es representada por la suma de los cuadrados de los errores entre los puntos y sus centroides respectivos.” Dentro de las configuraciones por

defecto se encuentra especificada la cantidad de 2 clusters, que es la cantidad de grupos en nuestro dataset, así que se aplica el algoritmo con las configuraciones por defecto.

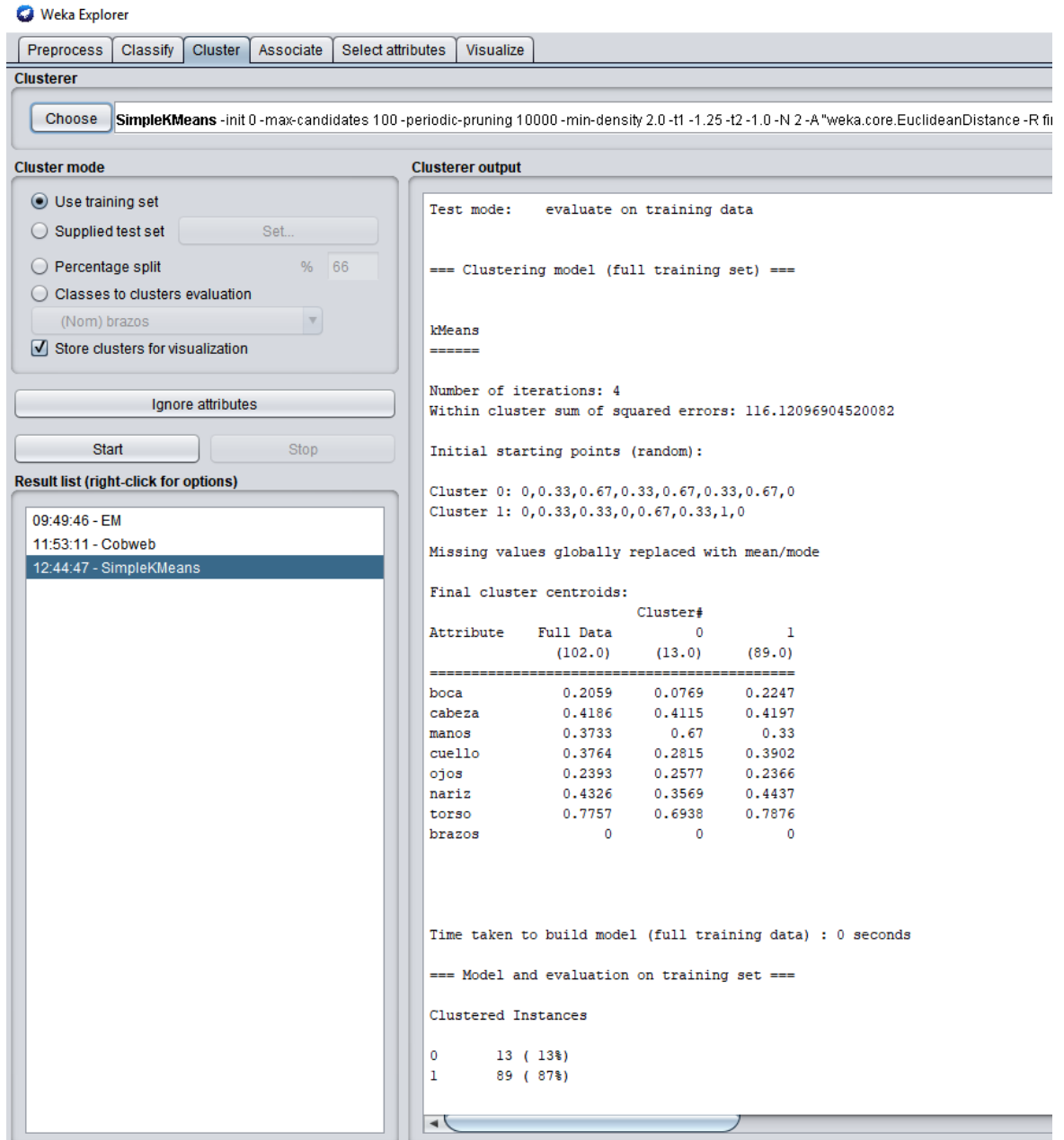


Figura 6.7 Aplicación algoritmo K-Means 8 atributos

Se observa en la figura 6.7, que su precisión respecto al grupo de niños en instituciones es realmente baja, ya que agrupa correctamente 13 registros de una población de 31, significando el 42 % de precisión.

También se aplica el algoritmo FarthestFirst, que [59] indica que es una variante del algoritmo K-means que ubica cada centro de cluster, en el punto más alejado de los centros de cluster existentes. Este punto debe estar dentro del área de datos. Esto acelera la agrupación porque en la mayoría de casos, se necesita menos reasignación y ajuste.

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer

Choose FarthestFirst -N 2 -S 1

Cluster mode

Use training set
 Supplied test set Set...
 Percentage split % 66
 Classes to clusters evaluation (Nom) brazos
 Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

09:49:46 - EM
 11:53:11 - Cobweb
 12:44:47 - SimpleKMeans
 12:58:04 - HierarchicalClusterer
 12:58:30 - FarthestFirst

Clusterer output

```

=== Run information ===

Scheme:      weka.clusterers.FarthestFirst -N 2 -S 1
Relation:    NormalizadoSeleccion8SinLbl
Instances:   102
Attributes:  8
             boca
             cabeza
             manos
             cuello
             ojos
             nariz
             torso
             brazos

Test mode:   evaluate on training data

=== Clustering model (full training set) ===

FarthestFirst
=====

Cluster centroids:

Cluster 0
    0.0 0.33 0.33 0.33 0.67 0.33 0.67 0
Cluster 1
    1.0 0.67 0.33 0.0 0.0 0.33 1.0 0.33

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      75 ( 74%)
1      27 ( 26%)

```

Figura 6.8 Aplicación algoritmo FarthestFirst 8 atributos

Se observa en la figura 6.8, que la cantidad de los dos clusters se aproxima a la disposición de nuestro dataset, con 31 participantes en el grupo de Acogida y 71 en el regular, con los resultados de 75 y 27 respectivamente, pero luego de exportar los resultados de la agrupación y comparar con las etiquetas, se detectó que 23 de 31 estuvieron incorrectamente separados del cluster1 al cluster0, con lo que representa el 26% de precisión.

Evidenciando una baja precisión con los algoritmos de clustering, se procederá con un plan de experimentación con técnicas de clasificación supervisada, como se explica en el capítulo V (Marco teórico).

Para aplicar este tipo de aprendizaje, bajo el *concepto* Classify. El plan de experimentación se desarrollará con algoritmos de las categorías más representativas. Dentro de la categoría Bayes, se dispone del algoritmo NaiveBayes, que [60] define como “un método intuitivo que utiliza las probabilidades condicionales de cada atributo que pertenece a cada clase para hacer una predicción. Utiliza el Teorema de Bayes, una fórmula que calcula la probabilidad luego que la evidencia sea analizada, contando la frecuencia de los valores y sus combinaciones en los datos históricos.” Una importante acotación nos aporta [61], en que requiere pequeños datos de entrenamiento, que se podría adaptar a la cantidad de registros del dataset con el que se está trabajando.

Dentro de la categoría Functions, existe el algoritmo MultilayerPerceptron en la que [60] señala que es una red neuronal que es entrenada utilizando *back propagation*, consta de tres capas: una capa de entrada con los datos de entrada para cada atributo, una capa oculta en la que están conectados todos los nodos de entrada y una capa de salida, con los datos resultantes. Dentro del algoritmo perceptron, se dispone de VotedPerceptron, que [60] indica es una modificación del algoritmo perceptrón básico, permitiendo usar un segundo multiplicador para usar actualizaciones multiplicativas, que lo hace más flexible. Otra función es la Logística, que [60] señala es “es una implementación alternativa para construir y usar un modelo de regresión logística multinomial.”

Dentro de la categoría Rules se dispone del algoritmo OneR [61], indica que su nombre es una abreviatura de “One Rule”, es un algoritmo preciso para ser simple. Genera una regla para cada predictor en los datos, y luego selecciona la regla con el error total más pequeño como su “única regla”, como ejemplo de una regla, se puede encontrar la clase más frecuente

y se asigne ese valor a los predictores, calcular el error de cada predictor y usar el algoritmo para encontrar el predictor que posea el error total más pequeño.

Dentro de la categoría Trees, se tiene a J48 del que [58] señala que es una implementación del algoritmo C4.5, en que [62] indica es uno de los algoritmos de minería de datos más utilizado, que tiene las características de particionar el valor del atributo continuo en un conjunto discreto de intervalos que se conoce ampliamente como “discretización” y usa valores de probabilidad, para valores perdidos o no asignados en lugar de asignar los valores más comunes existentes de ese atributo. Estos valores de probabilidad se calculan a partir de las frecuencias observadas en esa instancia. Se acota en [58] que “supone una mejora moderada en las prestaciones, y podrá conseguir una probabilidad de acierto ligeramente superior.”

Con una breve explicación de las técnicas a aplicarse en nuestro plan de experimentación, se realiza un cuadro comparativo de los resultados de precisión de las técnicas expuestas para la fase de entrenamiento de los diferentes algoritmos. Utilizando la técnica de división del dataset en dos partes, la mayor parte de registros para entrenamiento y los registros restantes para realizar pruebas de la clasificación. Las medidas para la división de nuestro dataset son 70 % para entrenamiento con 30 % para pruebas y 80 % para entrenamiento y 20 % para pruebas. Además se aplicará el método de Validación Cruzada o Cross Validation, con 5 y 10 Folds para un análisis específico de cada dataset.

Los resultados iniciales, se obtendrán de los experimentos con el dataset “Normalizado-Full.arff”, fuente explicada al inicio del capítulo. Los registros están ordenados por tipo de institución, en la parte inicial con instituciones de acogida seguido de instituciones regulares. La precisión de las diferentes técnicas, se las observa en el cuadro 6.1. Además de recopilar las precisiones de las diferentes técnicas, se normalizará la calidad del dataset, ya que si se tienen mejores datos para un entrenamiento supervisado, se analizará la técnica con mayor

efectividad en el dataset con mayor efectividad. Para poder medir la efectividad del dataset, se extrae un promedio de las precisiones en cada cuadro de resultados.

Los resultados iniciales, se obtendrán de los experimentos con el archivo “Normalizado-Full.arff”, el cuál contiene las selecciones de todas las partes del cuerpo, es decir; no se le ha aplicado un análisis de atributos. Los registros están ordenados por tipo de instituciones de los niños, los de instituciones de acogida en la parte inicial seguido de los de instituciones regulares. La precisión de las diferentes técnicas, se las observa en el cuadro 6.1.

Cuadro 6.1 Resultados precisión archivo normalizado 12 partes

Entrenamiento	Bayes		Functions		Rules	Trees
	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	65.68	60.78	69.6	66.66	65.68	68.62
Cross V. 5 F.	65.68	58.82	72.54	69.6	63.72	68.62
Training 80 %	70	65	70	75	75	70
Training 70 %	70.96	67.74	67.74	74.19	70.96	77.41

El promedio de resultados en el cuadro 6.1, es de 68.75.

Se modifica el dataset y se lo ordena de manera aleatoria, para que el algoritmo se entrene de mejor manera con los valores de las instancias y no influya el orden como un patrón del análisis. Este nuevo dataset se llama “NormalizadoFullShuffler.arff”. Los resultado se presentan en el cuadro 6.2 y tiene un promedio de resultados de 64.08.

Cuadro 6.2 Resultados precisión archivo normalizado 12 partes orden aleatorio

	Bayes		Functions		Rules	Trees
Entrenamiento	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	67.64	60.78	66.66	64.7	61.76	45.16
Cross V. 5 F.	63.72	65.68	62.74	67.64	62.74	70
Training 80 %	55	70	65	65	65	65.68
Training 70 %	58.06	70.96	70.96	67.74	64.51	60.78

Con la comparación entre los cuadros 6.1 y 6.2, se puede evidenciar que en la mayoría de los resultados, la precisión disminuye debido a la influencia del orden de las instancias, siendo el orden aleatorio el más real en la precisión, ya que en la vida real cuando se recolecta la información, no se la tiene organizada. El nivel de precisión aún es muy baja, por lo que se debe mejorar el dataset.

Para reducir los atributos se usará el análisis presentado en la figura 6.2 y figura 6.3, sus valores más bajos lo tienen las partes de pies, cabello, piernas y orejas. Con lo que se mantienen 8 atributos reales en nuestro nuevo dataset, los que serán listados en el cuadro 6.3 para una mejor referencia en el plan de experimentación.

Cuadro 6.3 Partes del cuerpo 8 atributos

Partes cuerpo
Boca
Cuello
Cabeza
Ojos
Nariz
Manos
Torso
Brazos

Los valores de las partes listadas en el cuadro 6.3 estarán en el dataset con nombre “NormalizadoSeleccion8.arff” y con un orden aleatorio con nombre “NormalizadoSeleccion8Shuffler.arff” Los resultados con los algoritmos especificados, dentro del nuevo dataset son presentados en el cuadro 6.4.

Cuadro 6.4 Resultados precisión archivo normalizado 8 partes

Entrenamiento	Bayes	Functions			Rules	Trees
	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	71.56	74.19	73.52	69.6	67.64	68.62
Cross V. 5 F.	71.56	70	70.58	68.62	65.68	72.54
Training 80 %	75	61.76	80	75	75	80
Training 70 %	74.19	66.66	70.96	77.41	70.96	74.19

Cuadro 6.5 Resultados precisión archivo normalizado 8 partes orden aleatorio

Entrenamiento	Bayes	Functions			Rules	Trees
	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	71.56	70.58	68.62	71.56	62.74	63.72
Cross V. 5 F.	70.58	69.6	70	71.56	66.66	68.62
Training 80 %	75	60	70	70	60	60
Training 70 %	67.74	48.38	61.29	61.29	61.29	51.61

El promedio de resultados del cuadro 6.4 es de 71.93 y del cuadro 6.5 es de 65.52, se evidencia que aún existe una diferencia amplia entre los resultados con datos ordenados frente a los aleatorios, incluso mayor que en la iteración anterior de nuestro plan, y se ha mejorado los resultados de precisión lo que indica que existe una mejora en nuestro dataset.

Se procede a reducir un atributo más en base a nuestro análisis de atributos, se reducirá la parte del cuello. Para resumir la cantidad de datos se aplicará al dataset un orden aleatorio, que presenta una información más real. El nuevo dataset tendrá el nombre “Normalizado-Seleccion7Shuffler.arff”. El listado de partes se presenta en el cuadro 6.6 para una mejor referencia en el plan de experimentación.

Cuadro 6.6 Partes del cuerpo 7 atributos

Partes cuerpo

Boca

Cabeza

Manos

Ojos

Nariz

Torso

Brazos

Cuadro 6.7 Resultados precisión archivo normalizado 7 partes orden aleatorio

	Bayes	Functions			Rules	Trees
Entrenamiento	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	71.56	66.66	72.54	66.66	66.66	67.64
Cross V. 5 F.	68.62	73.52	68.62	68.62	60.78	68.62
Training 80 %	70	50	70	65	70	70
Training 70 %	61.29	61.29	58.06	61.29	58.06	58.06

Como se percibe en el cuadro 6.7, los valores son ligeramente mayores al de la iteración anterior, alcanza un promedio de resultados de 65.56, con lo que se avanza a la siguiente iteración, para mejorar nuestro dataset. Se restará la parte de los brazos, según nuestro análisis de atributos. Con lo que las partes con las que se trabajará, se presentan en el cuadro 6.8

Cuadro 6.8 Partes del cuerpo 6 atributos

Partes cuerpo

Boca

Cabeza

Manos

Ojos

Nariz

Torso

Ya que existe una amplia mejora, en base a los resultados obtenidos en los dos tipos de ordenamiento en nuestro dataset.

Cuadro 6.9 Resultados precisión archivo normalizado 6 partes

Entrenamiento	Bayes	Functions			Rules	Trees
	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	70.58	64.7	67.64	70.58	67.64	69.60
Cross V. 5 F.	71.56	67.64	71.56	71.56	65.68	72.54
Training 80 %	70	75	75	75	75	80
Training 70 %	74.19	70.96	70.96	77.41	70.96	77.41

Cuadro 6.10 Resultados precisión archivo normalizado 6 partes orden aleatorio

Entrenamiento	Bayes	Functions			Rules	Trees
	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	70.58	63.72	71.56	71.56	66.66	72.54
Cross V. 5 F.	71.56	65.68	73.52	74.5	64.7	75.49
Training 80 %	70	70	75	80	65	65
Training 70 %	80.64	80.64	80.64	80.64	70.96	70.96

El promedio de resultados del cuadro 6.9 es de 71.8 y del cuadro 6.10 es de 72.15, en ambos casos superan los resultado de la iteración anterior, alcanzando mejores resultados en el dataset con orden aleatorio. Hay que resaltar que en este análisis se obtuvo la mayor precisión con el método Cross Validation, con la función logística se obtuvo el valor de 74.5 y bajo un entrenamiento del 70% del dataset, se obtuvo un valor de precisión de 80.64 en 4 funciones, NaibeBayes, Multilayer, Voted Perceptron y Logistic. Lo que indica que el dataset, contiene datos de valor y hasta el momento la mejor precisión.

Se procede a reducir un atributo más en base a nuestro análisis de atributos, para experimentar con los resultados, con lo que se reducirá la parte de los ojos. Para resumir la cantidad de datos se utiliza únicamente el dataset en orden aleatorio, el nuevo dataset tendrá el nombre “NormalizadoSeleccion5Shuffler.arff”. El listado de partes se presenta en el cuadro 6.11.

Cuadro 6.11 Partes del cuerpo 5 atributos

Partes cuerpo
Boca
Cabeza
Manos
Nariz
Torso

Cuadro 6.12 Resultados precisión archivo normalizado 5 partes orden aleatorio

	Bayes	Functions			Rules	Trees
Entrenamiento	NaiveBayes	Multilayer Percep	Voted Percep	Logistic	OneR	J28
Cross V. 10 F.	71.56	66.66	68.62	70.58	64.7	68.62
Cross V. 5 F.	69.6	68.62	69.6	69.6	64.7	69.6
Training 80 %	75	65	70	80	65	55
Training 70 %	77.41	58.06	70.96	77.41	70.96	61.29

Como se percibe en el cuadro 6.12, los valores han disminuido ligeramente, siendo menores al de la iteración anterior. Alcanza un promedio de resultados de 68.69, con lo que se finaliza el plan de experimentación. Ya que al reducir atributos, se cuenta con menor cantidad de criterios para agrupar los registros, presentando una menor precisión con las diferentes técnicas.

Dentro del plan de experimentación se alcanzó una precisión que satisface un valor aceptable de clasificación, al ser mayor a 80% de precisión.

6.3. Actualización de experimentación

En la primera fase de experimentación, se obtiene como resultado la tabla 6.10 en la que se presenta la mayor precisión de 80.64 % obtenido con entrenamiento de división del 70 % del dataset.

Se vio necesaria una optimización para alcanzar una precisión mayor, con lo que se requiere una nueva fase de experimentación implementada con mayor personalización para la cual se implementará en el lenguaje de desarrollo Python, para utilizar las librerías disponibles en los paquetes de machine learning para aplicar una comparación estimada a las técnicas aplicadas en la fase inicial, además de un manejo más glanuar de las configuraciones en los algoritmos, asegurando de mejor manera su personalización y validez.

Con los datos de precisión obtenidos en la fase 1 y el dataset normalizado, se plantea utilizar los siguientes métodos de machine learning:

Cuadro 6.13 Métodos a implementarse

Métodos a implementarse en segunda etapa de experimentación
Naibe Bayes
Perceptron Multicapa
Regresion Logística
Árbol de decisión

Como evidencia de la implementación en el lenguaje Python, se adjuntará una captura del código fuente implementado de cada método y su precisión alcanzada.

Para iniciar el proceso de experimentación se carga en la aplicación desarrollada, el dataset obtenido, se adjunta la captura del código fuente:

```
def loadDataSet(source_dataset, fileName):
    filePath = source_dataset + fileName
    dataSetLoaded = pd.read_csv(filePath, engine='python')
    print(f'===== DataSet {fileName} CARGADO CORRECTAMENTE =====')
    print("=====")
    return dataSetLoaded
```

Figura 6.9 Carga de dataset obtenido en fase de experimentación en fase 1

En este método se carga el archivo para ser procesado, recibe como parámetro el *path* del archivo y su *nombre*.

El plan de experimentación para cada algoritmo de machine learning, consta de una separación del dataset del 80 % para entrenamiento y 20 % para prueba, una segunda división de 70 % para entrenamiento y 30 % para prueba, seguido de una validación cruzada con 10-Folds y 5-Folds.

La implementación inicia con el método Naive Bayes, en la figura 6.10 se encuentra el código fuente y los resultados en la figura 6.11.

```
def naiveBayesMethod(dataSet, percentTest):
    # Division de data en entradas y salidas
    x = dataSet.iloc[:, 0:5].values
    y = dataSet.iloc[:, 6].values

    # Asignacion de tamaño de datos de prueba
    X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=percentTest, stratify=y)

    # escalar los datos de entrada
    sc_X = StandardScaler()
    X_train = sc_X.fit_transform(X_train)
    X_test = sc_X.fit_transform(X_test)

    classifier = BernoulliNB() #BernoulliNB() GaussianNB() MultinomialNB()

    classifier.fit(X_train, y_train)
    y_pred = classifier.predict(X_test)

    return accuracy_score(y_test, y_pred)
```

Figura 6.10 Implementación método Naive Bayes

```

===== MÉTODO NAIBY BAYES =====
===== ENTRENAMIENTOS A DATA SET =====
===== División de 30.0 % de entrenamiento =====
Precisión: 0.8387096774193549
===== División de 20.0 % de entrenamiento =====
Precisión: 0.7619047619047619
===== CROSS-VALIDATION =====
##### 10 FOLDS #####
Prec Avg: 0.6945454545454546
Precisión: [0.81818182 0.72727273 0.7      0.8      0.6      0.6
0.7      0.8      0.7      0.5      ]
##### 5 FOLDS #####
Prec Avg: 0.7052380952380952
Precisión: [0.80952381 0.66666667 0.6      0.75      0.7      ]
=====

```

Figura 6.11 Precisión método Naibe Bayes

Dentro del método Perceptrón Multicapa, se adjunta la implementación en la figura 6.12 y sus resultados en la figura 6.13.

```

def multiLayerPerceptron(dataSet, percentTest):
    x = dataSet.iloc[:, 0:5].values
    y = dataSet.iloc[:, 6].values

    # assign test data size
    X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=percentTest, stratify=y, random_state=0)

    #hidden layer NCapas, y numero de nodos
    clf = MLPClassifier(max_iter=500, hidden_layer_sizes=(400, 150, ), warm_start=True, activation="relu", solver="adam").fit(X_train, y_train)

    precision = clf.score(x, y)

    return precision

```

Figura 6.12 Implementación método Naibe Bayes

```

=====
===== MÉTODO MULTILAYER PERCEPTRON =====
===== ENTRENAMIENTOS A DATA SET =====
===== División de 30.0 % de entrenamiento =====
Precisión: 0.7941176470588235
===== División de 20.0 % de entrenamiento =====
Precisión: 0.8137254901960784
===== CROSS-VALIDATION =====
##### 10 FOLDS #####
Precisión: 0.6745454545454546
##### 5 FOLDS #####
Precisión: 0.5966666666666667
=====

```

Figura 6.13 Precisión método Perceptrón Multicapa

Dentro del método Regresión Logística, se adjunta la implementación en la figura 6.14 y sus resultados en la figura 6.13.

```

def logisticRegression(dataSet, percentTest):
    # División de data en entradas y salidas
    x = dataSet.iloc[:, 0:5].values
    y = dataSet.iloc[:, 6].values

    # Asignacion de tamaño de datos de prueba
    X_train, X_test, y_train, y_test = train_test_split(x, y, stratify=y, test_size=percentTest)

    clf = LogisticRegression().fit(X_train, y_train) #random_state=0, max_iter=50
    precision = clf.score(X_test, y_test)

    return precision

```

Figura 6.14 Implementación método Regresión Logística

```

=====
===== MÉTODO LOGISTIC REGRESSION =====
===== ENTRENAMIENTOS A DATA SET =====
===== División de 30.0 % de entrenamiento =====
Precisión: 0.7741935483870968
===== División de 20.0 % de entrenamiento =====
Precisión: 0.8571428571428571
===== CROSS-VALIDATION =====
##### 5 FOLDS #####
Precisión: 0.7552380952380953
##### 10 FOLDS #####
Precisión: 0.7454545454545454
=====

```

Figura 6.15 Precisión método Regresión Logística

Finalmente la implementación del método árbol de decisión, se adjunta código fuente en la figura 6.14 y sus resultados en la figura 6.13.

```

def logisticRegression(dataSet, percentTest):
    # División de data en entradas y salidas
    x = dataSet.iloc[:, 0:5].values
    y = dataSet.iloc[:, 6].values

    # Asignacion de tamaño de datos de prueba
    X_train, X_test, y_train, y_test = train_test_split(x, y, stratify=y, test_size=percentTest)

    clf = LogisticRegression().fit(X_train, y_train) #random_state=0, max_iter=50
    precision = clf.score(X_test, y_test)

    return precision

```

Figura 6.16 Implementación método Árbol de decisión

```

=====
===== MÉTODO LOGISTIC REGRESSION =====
===== ENTRENAMIENTOS A DATA SET =====
===== División de 30.0 % de entrenamiento =====
Precisión: 0.7741935483870968
===== División de 20.0 % de entrenamiento =====
Precisión: 0.8571428571428571
===== CROSS-VALIDATION =====
##### 5 FOLDS #####
Precisión: 0.7552380952380953
##### 10 FOLDS #####
Precisión: 0.7454545454545454
=====
    
```

Figura 6.17 Precisión método Árbol de decisión

Para resumir los resultados de esta fase de experimentación, se los especificará en la tabla 6.14

Cuadro 6.14 Resultados precisión experimentación personalizada

	Bayes	Functions		Trees
Entrenamiento	NaiveBayes	Multilayer Percep	Logistic	DecisionTree
Cross V. 10 F.	69.45	67.45	74.54	66.63
Cross V. 5 F.	70.52	59.66	75.52	62.57
Training 80 %	76.19	81.37	85.71	76.19
Training 70 %	83.87	79.41	77.41	70.96

Con la finalización de esta fase de experimentación, se obtuvieron las mayores precisiones con el método de regresión logística dentro del método de separación 80% entrenamiento y 20% de prueba con una precisión de **85.71 %** y dentro de la validación cruzada con **5-Folds** se alcanzó la precisión de **75.52 %** y **10-Folds** con **74.54 %**.

Capítulo 7

Cronograma

Total, Horas: 460 h

Fecha de Inicio: lunes, 22 de octubre del 2018

Fecha de Finalización: domingo, 30 de octubre del 2022

CARRERA DE INGENIERÍA DE SISTEMAS

FORMATO DE SEGUIMIENTO DE TUTOR – TRABAJO DE TITULACIÓN

No.	Actividad desarrollada	Recursos/Materiales/ Conocimientos Requeridos	Fecha	Horas requeridas
1	Analizar y estudiar los aspectos fundamentales para la alteración psicológica en niños.	Computador, internet, inducción personal psicológico en colaboración	22/10/2018	Tutor: 0 Estudiante: 12
2	Identificar las principales alteraciones psicológicas en la población infantil	Computador, internet	26/10/2018	Tutor: 2 Estudiante: 15
3	Investigar concepto y diagnóstico de cada uno de los perfiles identificados	Computador, internet, inducción personal psicológico en colaboración	29/10/2018	Tutor: 0 Estudiante: 15
4	Validar requerimientos en las actividades en la aplicación móvil con las psicólogas inmersas en el proyecto	Computador, internet, inducción personal psicológico en colaboración	05/11/2018	Tutor: 2 Estudiante: 10
5	Rediseño de interfaz de la aplicación en Android	Computador, tableta, conocimiento desarrollo en Android	08/11/2018	Tutor: 5 Estudiante: 20
6	Seleccionar e implementar medios multimedia, facilitados por diseñador del proyecto	Computador, tableta, conocimiento desarrollo en Android	15/11/2018	Tutor: 0 Estudiante: 30
7	Desarrollar aplicación en plataforma Android	Computador, tableta, conocimiento desarrollo en Android	27/11/2018	Tutor: 6 Estudiante: 32
8	Documentar código fuente y funcionamiento	Computador	07/12/2018	Tutor: 0 Estudiante: 10
9	Analizar variables relevantes de información obtenida por aplicación móvil	Computador, conocimientos en procesamiento de datos	12/12/2018	Tutor: 4 Estudiante: 8

10	Analizar y estudiar técnicas de clustering e inteligencia artificial aplicables desde Weka	Computador, internet	04/12/2020	Tutor: 3 Estudiante: 30
11	Identificar la técnica más adecuada para los datos obtenidos	Computador, internet, dataset obtenido	18/12/2020	Tutor: 2 Estudiante: 20
12	Diseñar módulo inteligente en Weka	Computador, internet, dataset obtenido	27/12/2020	Tutor: 0 Estudiante: 20
13	Desarrollo de módulo inteligente Weka	Computador, conocimientos machine learning	08/01/2020	Tutor: 2 Estudiante: 30
14	Realizar pruebas de funcionamiento del algoritmo	Computador	22/01/2020	Tutor: 0 Estudiante: 16
15	Identificar variables de entrada y salida a incluir en plan de experimentación	Computador	29/01/2020	Tutor: 2 Estudiante: 15
14	Diseñar el plan de experimentación en Weka	Computador, conocimientos machine learning	04/02/2020	Tutor: 0 Estudiante: 20
15	Analizar los resultados obtenidos y ajustar plan de experimentación	Computador, conocimientos machine learning	13/02/2020	Tutor: 0 Estudiante: 32
16	Documentar los resultados obtenidos	Computador	01/04/2020	Tutor: 0 Estudiante: 15
17	Desarrollo de documentación de memoria técnica	Computador	28/05/2020	Tutor: 4 Estudiante: 50
18	Definición de FASE 2 DE EXPERIMENTACIÓN	Computador	03/10/2022	Tutor: 4 Estudiante: 4
19	Investigación de uso y sintaxis de lenguaje de programación Python	Computador, internet	05/10/2022	Tutor: 0 Estudiante: 16
20	Identificación e investigación de implementación de métodos de machine learning	Computador, internet, dataset obtenido	10/10/2022	Tutor: 2 Estudiante: 10
21	Diseño de plan de experimentación	Computador	13/10/2022	Tutor: 0 Estudiante: 4
22	Desarrollo de plan de experimentación	Computador	17/10/2022	Tutor: 0 Estudiante: 10

23	Ejecución de plan de entrenamiento	Computador, internet, dataset obtenido	21/10/2022	Tutor: 0 Estudiante: 6
24	Análisis de resultados obtenidos y ajuste de plan de experimentación	Computador, internet, dataset obtenido	24/10/2022	Tutor: 2 Estudiante: 6
25	Documentar los resultados obtenidos	Computador	30/10/2022	Tutor: 0 Estudiante: 4
Total de horas de trabajo: 500				Tutor: 40 Estudiante: 460

Figura 7.1 Cronograma

Capítulo 8

Presupuesto

DENOMINACIÓN	CANTIDAD (unidades)	COSTO UNITARIO (dólares)	COSTO TOTAL (dólares)
1. Bienes			
Resma Papel Bond A-4 de 80 gr.	1	30,00	30,00
Copias	100	0,05	5,00
Materiales de escritorio	1	20,00	20,00
Impresiones	100	0,10	10,00
Flash memory	1	12,00	12,00
2. Personal			
Estudiante Investigador	1	1200,00	1200,00
Asesoría especializada (UDA)	1	600,00	600,00
3. Servicios			
Servicio de transporte	1	30,00	30,00
Servicios de Internet	4	30,00	120,00
Material Bibliográfico	10	23,00	230,00
4. Tecnológicos			
Equipo de computo	1	250,00	450,00
5. Otros			
Imprevistos	1	50,00	50,00
TOTAL			2707,00

Figura 8.1 Presupuesto

Capítulo 9

Conclusiones

Se planteó inicialmente la posibilidad de un nivel mayor de rechazo en la realización del test en niños de instituciones de acogida, pero al explicarse que se lo realizará con un enfoque lúdico, se consiguió gran acogida. El mismo escenario se presentó con los niños de instituciones regulares. Con lo que no se necesitó mayor influencia de los psicólogos de las instituciones. Fue importante que la parte psicológica de este proyecto, haya llevado a cabo el registro con los niños, ya que se maneja de mejor forma la interacción.

Dentro del plan de experimentación en el módulo inteligente, como regla fundamental debe aplicarse la fase de diseño en el plan de experimentación, que se ejecutará con las diferentes técnicas de machine learning, dentro del desarrollo de este plan, en este trabajo se evidenció importantes mejoras en la precisión con el uso de técnicas previas que nos permitían actualizar nuestro dataset con atributos de mayor influencia, de tal manera que la aplicación de las técnicas específicas presenten mayor beneficio en la precisión resultante.

Dentro de la aplicación de técnicas de machine learning, no se vio factible extenderse en la aplicación de técnicas de clustering, debido a la cantidad reducida de la población y los valores dispersos en los atributos, ya que existían muchas selecciones similares entre

los grupos de los participantes. Por lo que se obtuvo mejores resultados con técnicas de entrenamiento supervisado, con el pre-entrenamiento que los caracteriza.

Se puede concluir que en el diagnóstico psicológico de los niños que participaron en el test, presentan rasgos similares a los descritos en otros países. Se ha comprobado en la documentación de este trabajo, la dificultad de detección y su atención posterior pueden presentar diversas dificultades, en nuestra sociedad. Siendo la principal, el no contar con una herramienta que tome en cuenta todos los factores influyentes, ya que es complejo medir aspectos como los sociales que son cualitativos y poderlos analizar cuantitativamente. Conllevando a una terapia sin la dirección totalmente correcta.

La herramienta presenta una aceleración en el diagnóstico, cuando se lo realiza en instituciones, ya que, debido a la cantidad de niños, utilizando una aplicación móvil, se optimizan tiempos de aplicación del test y registros elaborados manualmente. En las conversaciones con el personal psicológico, fue expresada la utilidad de la herramienta para diagnósticos ágiles.

Dentro de la etapa de actualización en la experimentación, se obtuvo una precisión bastante superior donde se evidenció que en base al conocimiento aplicado a una mayor cantidad de variables, se refleja una mayor precisión utilizando el mismo dataset de la etapa previa. Dentro del beneficio de las técnicas de aprendizaje supervisado nos permite realizar un contraste con los resultados de los algoritmos y los obtenidos dentro del proyecto, alcanzando un diagnóstico eficaz de alrededor de 8 diagnósticos correctos de 10 aplicados, por medio de técnicas machine learning.

Capítulo 10

Trabajo Futuro

A lo largo del desarrollo de la aplicación móvil, surgieron ciertas ideas que se podrían mejorar su aplicación o funcionalidades, las que se recomienda podrían aplicarse como trabajo futuro. Se recomienda el experimentar e investigar a un nivel más amplio la influencia psicológica implícita en el uso de las imágenes que se utilizaron para cada parte del cuerpo de los niños. Esto, con el fin de lograr obtener una identificación más específica de los perfiles psicológicos de los niños, en la interacción con los niños.

Actualmente se recopiló de manera externa una base de datos con las interacciones de los niños en el proyecto, para un posterior análisis con el módulo inteligente. Podría mejorarse con un envío de datos directamente desde la aplicación elaborando una interfaz entre los datos recopilados por la aplicación y la ejecución de la clasificación. Una vez se obtenga el algoritmo entrenado resultante de la aplicación del plan de experimentación. Es decir, se puede comunicar directamente con una aplicación cliente, que permita analizar los resultados al finalizar la interacción con el niño. Para poder enfocar los trabajos posteriores, en función del perfil diagnosticado de manera inmediata.

Dentro de la aplicación móvil, podría desarrollarse tanto en multiplataforma, como para un tamaño de pantalla adaptable (responsive), permitiendo una diversificación mayor en la cantidad de dispositivos soportados, con esto los psicólogos que deseen utilizar la herramienta, no tendrían limitaciones técnicas de uso del resultado de este proyecto.

Capítulo 11

Glosario

- Algoritmos: Instrucciones organizadas para resolver un problema.
- Aplicación móvil: Software diseñado para ser ejecutado en un dispositivo móvil.
- Atributo: Variable del conjunto de datos a utilizarse como fuente de datos para la aplicación de técnicas de machine learning.
- Clustering: Agrupación de elementos sin una categorización en grupos con atributos similares por medio de un sistema de aprendizaje automático.
- Cluster: Colección de elementos relacionados por atributos similares.
- Dato: Representación de un atributo o valor que va a ser gestionado por un sistema informático.
- Fase de entrenamiento: Fase de experimentación con un enfoque de aprendizaje con la aplicación de un algoritmo de machine learning.
- Inteligencia artificial: Es la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana [63].

- Lúdico: Actividad con fines de entretenimiento.
- Modelo: Es un esquema organizado de relaciones y/o operaciones entre los datos [64].
- Patrón: Es un conjunto de comportamientos que se repiten de manera predecible.
- Predicción: La predicción es el proceso de estimar el resultado de una situación basada en datos pasados y presentes.
- Personalidad: Son rasgos de comportamiento formados a partir de experiencias y circunstancias, la herencia genética, la cultura, la educación y otros factores, estos rasgos pueden cambiar con el tiempo, en respuesta a la experiencia, el aprendizaje y otros factores.
- TIC: Tecnologías de la información y comunicación.

Bibliografía

- [1] A. Pacurucu-Pacurucu, F. Pesántez-Avilés, S. Escobar-Cobos, A. Poveda-Pulla, E. Borck-Vintimilla, V. C. L. Wong, V. Robles-Bykbaev, and M. Cobos-Cali, “An intelligent ecosystem to support psychological evaluation and intervention of children under social risk: towards graphics design validation and new findings,” in *IEEE CACIDI 2016-IEEE Conference on Computer Sciences*. IEEE, 2016, pp. 1–6.
- [2] S. M. Weiss and C. A. Kulikowski, *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems*. Morgan Kaufmann Publishers Inc., 1991.
- [3] B. I. C. Soto and G. A. Tomasini, “Niños con altas capacidades: Análisis de las variables familiares implicadas en el desarrollo del potencial,” *Electronic Journal of Research in Education Psychology*, vol. 16, no. 45, pp. 273–300, 2018.
- [4] P. S. Pinheiro, “Informe del experto independiente para el estudio de la violencia contra los niños, de las naciones unidas,” 2011.
- [5] Unicef *et al.*, *ESTADO MUNDIAL DE LA INFANCIA 2007: La mujer y la infancia El doble dividendo de la igualdad de género*. UNICEF, 2006.
- [6] A. Collins and W. Ferguson, “Epistemic forms and epistemic games: Structures and strategies to guide inquiry,” *Educational psychologist*, vol. 28, no. 1, pp. 25–42, 1993.
- [7] D. R. Michael and S. L. Chen, *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.
- [8] Y. Cao, *Ludic elicitation: Using games for knowledge elicitation*. The Pennsylvania State University, 2014.
- [9] K. Pancerz, O. Mich, A. Burda, and J. Gomuła, “A tool for computer-aided diagnosis of psychological disorders based on the mmpi test: An overview,” in *Applications of Computational Intelligence in Biomedical Technology*. Springer, 2016, pp. 201–213.
- [10] A. Amory, “Game object model version ii: a theoretical framework for educational game development,” *Educational Technology Research and Development*, vol. 55, no. 1, pp. 51–77, 2007.
- [11] W. H. Organization *et al.*, “Global estimates of health consequences due to violence against children,” *Geneva: WHO*, 2006.

- [12] D. M. Feria Naranjo and M. L. Pomaquero Yuquilema, "Creación de un sistema lúdico para fortalecer valores en niños y niñas de 6 a 8 años," B.S. thesis, 2013.
- [13] J. D. Godino, M. Gonzato, and T. Fernández, "¿ cuánto suman los ángulos interiores de un triángulo? conocimientos puestos en juego en la realización de una tarea metemática," 2010.
- [14] E. Morin, "El método i; la naturaleza de la naturaleza (1986)," *Madrid: Cátedra*, 1977.
- [15] B. Jaworski, "Tools and tasks for learning and meta-learning," 2005.
- [16] M. E. Castagnola, M. A. Bosio, and G. A. Chiodi, "Juegos serios aplicados a niños con discapacidades," in *II Simposio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 44 (Rosario, 2015)*, 2015.
- [17] J. González Díez *et al.*, "Diseño y desarrollo de una aplicación móvil de juegos serios para niños y adolescentes con diversidad cognitiva," 2019.
- [18] L. J. Nomdedeu, "Raíces: un juego serio social para revalorizar las culturas originarias," Ph.D. dissertation, Universidad Nacional de La Plata, 2015.
- [19] D. pública del Ecuador, "Código de la niñez y adolescencia. Última reforma: Edición constitucional del registro oficial 1, 31-v-2017," 2017. [Online]. Available: <https://biblioteca.defensoria.gob.ec/handle/37000/2112?mode=full>
- [20] MIES, "Informe acogimiento institucional reporte de datos al 30 de junio de 2018," 2018. [Online]. Available: <https://www.inclusion.gob.ec/wp-content/uploads/2018/08/Informe-Junio..pdf>
- [21] O. M. de la Salud, "Maltrato infantil," 2016. [Online]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/child-maltreatment>
- [22] M. D. Cornelius, N. M. De Genna, L. Goldschmidt, C. Larkby, and N. L. Day, "Prenatal alcohol and other early childhood adverse exposures: direct and indirect pathways to adolescent drinking," *Neurotoxicology and teratology*, vol. 55, pp. 8–15, 2016.
- [23] V. Keil and J. M. Price, "Externalizing behavior disorders in child welfare settings: Definition, prevalence, and implications for assessment and treatment," *Children and youth services review*, vol. 28, no. 7, pp. 761–779, 2006.
- [24] H. Meltzer, R. Gatward, T. Corbin, R. Goodman, and T. Ford, "The mental health of young people looked after by local authorities in England," *London: The Stationery Office*, 2003.
- [25] M. M. Carassa De La Sota, C. De La Sota, M. Melissa, F. Damiano Vega, and F. R. Flores Laban, "Aplicación de las técnicas del pehis de inés monjas en el área de comunicación para incrementar los niveles de habilidades sociales en los estudiantes de cuarto grado de educación primaria de la institución educativa parroquial virgen de la candelaria-villa maría del triunfo, ugel-01," 2017.
- [26] M. Monjas, B. Arias, and M. Verdugo, "Desarrollo de un código de observación para evaluar la interacción social en alumnos de primaria.(cois)," in *III Congreso de Evaluación Psicológica*, 1991.

- [27] F. M. Gresham and K. L. Lemanek, "Social skills: A review of cognitive-behavioral training procedures with children," *Journal of Applied Developmental Psychology*, vol. 4, no. 3, pp. 239–261, 1983.
- [28] P. De Miguel, "Enseñanza de habilidades de interacción social en niños con riesgo de exclusión," *Revista de Psicología Clínica con Niños y Adolescentes*, vol. 1, no. 1, pp. 17–26, 2014.
- [29] J. Rodríguez and F. Carrillo, *Técnicas de modificación de conducta*. Biblioteca Nueva, 2008.
- [30] R. M. Rapee *et al.*, "The development of generalized anxiety," *The developmental psychopathology of anxiety*, pp. 481–503, 2001.
- [31] C. G. Coll, J. Kagan, and J. S. Reznick, "Behavioral inhibition in young children," *Child development*, pp. 1005–1019, 1984.
- [32] A. Ordóñez-Ortega, L. Espinosa-Fernández, L.-J. García-López, and J.-A. Muela-Martínez, "Inhibición conductual y su relación con los trastornos de ansiedad infantil," *terapia psicológica*, vol. 31, no. 3, pp. 355–362, 2013.
- [33] B. A. Clavijo, "Violencia escolar, un fenómeno mundial," *Magazín Aula Urbana*, no. 40, pp. 20–20, 2003.
- [34] M. Valdés, "La violencia ronda la escuela," *Educación y cultura*, vol. 9, no. 14, p. 15, 1991.
- [35] P. V. Britos, "Procesos de explotación de información basados en sistemas inteligentes," Ph.D. dissertation, Universidad Nacional de La Plata, 2008.
- [36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.
- [37] S. Theodoridis, "et k. koutroumbas (2006)," *Pattern Recognition, Third Edition*. Orlando, FL, USA: Academic Press, Inc.
- [38] M. Gallardo Campos, "Aplicación de técnicas de clustering para la mejora del aprendizaje," Master's thesis, 2009.
- [39] R. M. Valdovinos Rosas *et al.*, *Técnicas de submuestreo, Toma de decisiones y Análisis de diversidad en aprendizaje supervisado con Sistemas Múltiples de Clasificación*. Universitat Jaume I, 2016.
- [40] Y. R. González, A. C. Arceo, N. M. Sánchez, and A. H. Domínguez, "Agrupamiento conceptual lógico combinatorio: una alternativa para la toma de decisiones," *Inteligencia Artificial*, vol. 19, no. 57, pp. 82–96, 2016.
- [41] E. Morales and J. Gonzales, "Reglas de asociación," 2013.
- [42] I. M. Peláez, "Modelos de regresión: lineal simple y regresión logística," *Revista Seden*, 2016.

- [43] M. Carrasco Kind and R. J. Brunner, “Tpz: photometric redshift pdfs and ancillary information by using prediction trees and random forests,” *Monthly Notices of the Royal Astronomical Society*, vol. 432, no. 2, pp. 1483–1501, 2013.
- [44] E. Servy, C. B. Cuesta, and G. P. Marí, “Comparación del comportamiento de diversos estimadores basados en núcleos,” 2004.
- [45] C. J. M. ALESSANDRO, “Tablas de decisión,” 2005.
- [46] Y. Filiberto, R. BELLO, Y. CABALLERO, and M. FRÍAS, “Algoritmo para el aprendizaje de reglas de clasificación basado en la teoría de los conjuntos aproximados extendida,” *Dyna*, vol. 78, no. 169, pp. 62–70, 2011.
- [47] L. E. Sucar, “Redes bayesianas,” *BS Araujo, Aprendizaje Automático: conceptos básicos y avanzados*, pp. 77–100, 2006.
- [48] B. López, “Case-based reasoning: a concise introduction,” *Synthesis lectures on artificial intelligence and machine learning*, vol. 7, no. 1, pp. 1–103, 2013.
- [49] P. S. Lagos, “Inteligencia artificial distribuida y razonamiento basado en casos en la arquitectura de un sistema basado en el conocimiento para la educación a distancia (sbced).” *Chile: Departamento de Metodología de la Investigación e Informática Educativa, Facultad de Educación, Univ. de Concepción*, 2003.
- [50] R. O. D’Aquila, “Desarrollo e implementación de un motor de inferencias basado en lógica borrosa para el razonamiento inexacto incierto e impreciso en sistemas expertos basados en reglas,” Ph.D. dissertation, Informatica, 1997.
- [51] R. A. E. Andrade, E. F. González, and E. G. Caballero, “Un sistema lógico para el razonamiento y la toma de decisiones: la lógica difusa compensatoria basada en la media geométrica,” *Investigación operacional*, vol. 32, no. 3, pp. 230–245, 2014.
- [52] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1989.
- [53] M. F. Triola, *Estadística Noveda Edición*. Pearson Educación de México, 2004.
- [54] U. de Waikato, “Weka the workbench for machine learning,” 2020. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>
- [55] D. G. Morate, “Manual de weka,” *Disponível através do e-mail diego. garcia. morate@mail.com*, 2008.
- [56] T. university of Waikato, “Weka in the ecosystem for scientific computing,” accessed: 15-05-2020. [Online]. Available: https://www.cs.waikato.ac.nz/~eibe/WEKA_Ecosystem.pdf
- [57] C. L. Corso and C. Lorena, “Aplicación de algoritmos de clasificación supervisada usando weka,” *Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba*, 2009.
- [58] M. G. Jiménez and A. Álvarez, “Análisis de datos en weka—pruebas de selectividad,” *línea] disponible en http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf*, 2010.

- [59] N. Sharma, A. Bajpai, and M. R. Litoriya, “Comparison the various clustering algorithms of weka tools,” *facilities*, vol. 4, no. 7, pp. 78–80, 2012.
- [60] F. R. Alam and S. Pachauri, “Comparative study of j 48 , naive bayes and oner classification technique for credit card fraud detection using weka,” 2017.
- [61] V. Bhuyar, “Comparative analysis of classification techniques on soil data to predict fertility rate for aurangabad district,” *Int. J. Emerg. Trends Technol. Comput. Sci*, vol. 3, no. 2, pp. 200–203, 2014.
- [62] M. M. Mazid, S. Ali, K. S. Tickle *et al.*, “Improved c4. 5 algorithm for rule based classification,” in *Proceedings of the 9th WSEAS international conference on Artificial intelligence, knowledge engineering and data bases*. World Scientific and Engineering Academy and Society (WSEAS), 2010, pp. 296–301.
- [63] L. Rouhiainen, “Inteligencia artificial,” *Madrid: Alienta Editorial*, 2018.
- [64] M. Marqués, *Bases de datos*. Castelló de la Plana: Publicacions de la Universitat Jaume I. Servei de . . . , 2011.