



UNIVERSIDAD NACIONAL DE COLOMBIA

Modelo computacional para evaluación de discapacidad intelectual usando datos de funcionamiento cognitivo

NELSON FABIAN LEIVA RUIZ

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial

Bogotá, Colombia

2021

Modelo computacional para evaluación de discapacidad intelectual usando datos de funcionamiento cognitivo

NELSON FABIAN LEIVA RUIZ

Tesis presentada como requisito parcial para optar al título de:
Magister en Ingeniería de Sistemas y Computación

Director:
Luis Fernando Niño Vásquez, Ph.D.

Codirectora:
Aura Nidia Herrera Rojas, Ph.D.

Grupos de Investigación:
Laboratorio de Investigación en Sistemas Inteligentes (LISI)
Laboratorio de Psicometría

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia

2021

Dedicada a mis padres.

Agradecimientos

A los profesores Luis Fernando Niño Vásquez, director del Grupo de Investigación LISI, y Aura Nidia Herrera Rojas, directora del Laboratorio de Psicometría, quienes durante el desarrollo de este proyecto brindaron el mejor soporte técnico y humano; así como a Jazmine Escobar Pérez, profesora del departamento de Psicología, quien amablemente revisó la tesis e hizo recomendaciones pertinentes. También a Henry Delgado, coordinador del área de evaluación del SAP, quien facilitó muy amablemente el insumo de datos para esta investigación. Así como a los compañeros de los grupos de investigación mencionados, quienes con sus aportes ayudaron a estructurar esta tesis.

Resumen

A lo largo de esta investigación, se llevó a cabo un proceso de descubrimiento de conocimiento (KDD, por sus siglas en inglés), en el marco de ciencia de datos, aplicado a datos de evaluación de capacidad cognitiva aportados por el Servicio de Atención Psicológica (SAP) de la Universidad Nacional de Colombia. Se realizó el preprocesamiento y tratamiento de datos en concordancia con el objetivo de la investigación: desarrollar un modelo computacional que permita determinar las agrupaciones de variables, asociadas al Coeficiente intelectual, que predican el diagnóstico de discapacidad intelectual. Se obtuvo un total de 18 variables cognitivas más informativas, con las cuales se implementaron Árbol de decisión y Regresión logística como modelos predictivos e interpretables.

Palabras clave: Ciencia de datos, Discapacidad intelectual, Evaluación psicológica, Inteligencia y KDD

Abstract

Throughout this research, a knowledge discovery process (KDD) was carried out, within the framework of data science, applied to cognitive capacity assessment data, provided by the Psychological Attention Service (SAP) of the National University of Colombia. Data pre-processing and treatment was carried out in accordance with the objective of the research: to develop a computational model that allows determining the groupings of variables, associated with the IQ, that predict the diagnosis of intellectual disability. A total of 18 more informative cognitive variables were obtained, with which Decision Tree and Logistic Regression were implemented as predictive and interpretable models.

Keywords: Data science, Intellectual disability, Psychological assessment, Intelligence and KDD

Esta tesis de maestría se sustentó el 04 de junio de 2021 a las 08:00 am,
y fue evaluada por los siguientes jurados:

Jazmine Escobar Pérez (Msc)

Afiliación: Universidad Nacional de Colombia, Departamento de Psicología

Jose Ismael Peña Reyes (Ph.D.)

Afiliación: Universidad Nacional de Colombia, Departamento de Ingeniería de Sistemas e
Industrial

Tabla de contenido

1. Marco teórico.....	2
1.1 Discapacidad intelectual	2
1.1.1 Funcionamiento cognitivo: definición y medición de inteligencia	3
1.1.2 Medida de inteligencia: Coeficiente Intelectual	5
1.1.3 WISC-IV: Escala de inteligencia Wechsler para niños	5
▪ Limitaciones de la prueba WISC-IV	8
1.2 Ciencia de datos	9
1.3 Descubrimiento de Conocimiento en Bases de Datos KDD	10
1.3.1 Estudio del dominio de aplicación	10
1.3.2 Creación de un conjunto de datos objetivo	10
1.3.3 Limpieza y preprocesamiento de datos.....	10
1.3.4 Reducción y transformación de datos	11
▪ Eliminación recursiva de características	11
1.3.5 Compaginar los objetivos del KDD con los métodos de minería	12
1.3.6 Seleccionar los métodos de minería de datos.....	12
▪ Aprendizaje automático	13
1.3.7 Aplicar los métodos de minería.....	13
▪ Árboles de decisión	14
▪ Bosque aleatorio	15
▪ Regresión logística	15
1.3.8 Evaluación e interpretación de resultados.....	16
▪ Medidas de desempeño.	16
▪ Curvas ROC	17
1.3.9 Presentación o documentación del conocimiento descubierto	18
1.4 Herramientas computacionales en salud mental	19
1.4.1 Aplicaciones computacionales en la evaluación de información psicológica ..	21
2. Planteamiento del problema y metodología.....	25
2.1 Descripción del problema	25
2.2 Objetivos	26
2.2.1 Objetivo general.....	26
2.2.2 Objetivos específicos	26
2.3 Metodología	26
▪ Etapa 1: Selección y preprocesamiento de datos	27
▪ Etapa 2: Análisis Exploratorio de datos	27
▪ Etapa 3: Transformación y modelamiento de datos	27
▪ Etapa 4: Evaluación del modelo computacional	27
3. Preprocesamiento y análisis descriptivo de los datos.....	28
3.1 Organización del conjunto de datos	28
3.1.1 Descripción datos socioeconómicos	29
3.1.2 Descripción datos de consulta	30
3.1.3 Descripción datos WISC-IV.....	31
3.2 Limpieza y adecuación del conjunto de datos	33
3.2.1 Limpieza de variables categóricas	33
3.2.2 Limpieza de fechas.....	37
3.2.3 Dicotomización del diagnóstico y balanceo de clases	38

3.2.4	Limpieza de datos WISC-IV	38
▪	Diseño con cubos	39
▪	Semejanzas	39
▪	Retención de dígitos	39
▪	Conceptos con dibujos	39
▪	Vocabulario	39
▪	Sucesión de números y letras	40
▪	Matrices	40
▪	Comprensión	40
3.3	Análisis exploratorio	40
4.	Selección de características y generación del modelo predictivo	48
4.1	Selección de características, primera aproximación: Importancia de características de Bosque aleatorio	48
4.1.1	Limitación de la primera aproximación	52
4.2	Selección de características, segunda aproximación: Eliminación recursiva de características	53
4.3	Generación de modelo predictivo	56
4.3.1	Implementación del modelo de Árbol de decisión	56
4.3.2	Implementación del modelo de Regresión logística	61
5.	Resultados y discusión	67
6.	Conclusiones, recomendaciones y trabajo futuro	72
7.	Bibliografía	75

Lista de figuras

Figura 1-1: Seudocódigo del algoritmo RFE	12
Figura 1-2: Modelo de Regresión logística	16
Figura 1-3: Ejemplo de curva ROC	18
Figura 3-1: Recuento de hombres y mujeres para la variable Género	41
Figura 3-2: Distribución de la edad de los evaluados	42
Figura 3-3: Distribución de estrato de los evaluados	42
Figura 3-4: Distribución de Tipo de colegio de los evaluados	43
Figura 3-5: Distribución de Motivo de consulta de los evaluados	44
Figura 3-6: Distribución del diagnóstico	44
Figura 3-7: Distribución de diagnósticos con discapacidad intelectual	45
Figura 3-8: Descriptivos subpruebas WISC-IV	46
Figura 3-9: Boxplot de subpruebas WISC-IV, agrupados por diagnóstico de DI	46
Figura 4-1: Método de selección de características por Bosque Aleatorio	49
Figura 4-2: Salida métricas de rendimiento para Bosque aleatorio	50
Figura 4-3: Top 10 de características más importantes para RandomForestClassifier	51
Figura 4-4: Método de selección de características por RFECV	53
Figura 4-5: Ajuste automático de selección de características con RFECV	54
Figura 4-6: Esquema de implementación del Árbol de decisión	57
Figura 4-7: Salida métricas de rendimiento para Árbol de decisión	58
Figura 4-8: Curva ROC para el Árbol de decisión	59
Figura 4-9: Niveles de importancia de características para el Árbol de decisión	60
Figura 4-10: Modelo de clasificación final del Árbol de decisión	60
Figura 4-11: Esquema de implementación del modelo de Regresión logística	61
Figura 4-12: Salida métricas de rendimiento para Regresión logística	62

Contenido	10
Figura 4-13: Curva ROC para el modelo de Regresión logística	64
Figura 4-14: Coeficientes del modelo de Regresión logística	65

Lista de tablas

Tabla 1-1: matriz de confusión	17
Tabla 1-2: resumen técnicas de aprendizaje automático en salud mental	24
Tabla 3-1: Descripción de los campos de la prueba WISC-IV	31
Tabla 3-2: Descripción de datos de diagnóstico	32
Tabla 3-3: Descripción de los campos de la prueba WISC-IV	33
Tabla 4-1: Intercepto y coeficientes del modelo de Regresión logística	64

Introducción

El proceso de evaluación y diagnóstico de pacientes con discapacidad intelectual está ligado al uso de pruebas psicométricas de funcionamiento intelectual y comportamiento adaptativo; así como la experticia del profesional en salud mental. Dicho proceso, de manera convencional, puede demandar bastante tiempo, así como recursos de diversa índole. Por otro lado, pese a la validación y confiabilidad de las pruebas psicométricas usadas tanto para el diagnóstico de discapacidad intelectual, como para otros tipos de desórdenes psicológicos, no dejan de tener sus limitaciones en su uso diagnóstico, particularmente, en casos donde el evaluado presenta desórdenes psicológicos profundos que limitan la comprensión y expresión del lenguaje.

Ante este panorama, las herramientas computacionales y la inteligencia artificial ofrecen el potencial de revolucionar el diagnóstico y tratamiento de enfermedades, al ser capaz de hacer análisis y clasificaciones que involucran una gran cantidad de datos muy difíciles o imposibles de manejar para los humanos. De este modo, las herramientas computacionales pueden coadyuvar, junto con la experticia de los profesionales en salud mental, a mejorar la efectividad del diagnóstico, así como dar recomendaciones o señalar cursos de acción, de manera que se tomen mejores decisiones y sea eficiente el uso de recursos, tanto en la evaluación como en el tratamiento patologías o malestar psicológico.

1. Marco teórico

En este capítulo se hará una revisión de los conceptos asociados al problema que se ha desarrollado a lo largo de la investigación, tanto en el ámbito psicológico como en el ingenieril, así como las investigaciones previas que abordan esta temática.

1.1 Discapacidad intelectual

Por cuestionamientos que pueden ir desde lo idiosincrático o lo ético, hasta una mejor o más actualizada definición, el concepto de discapacidad intelectual (DI) se usa cada vez más en lugar de retraso mental [1], pese a que este último permanece en las definiciones de las Clasificación Internacional de Enfermedades (CIE-10). La Asociación Americana de Discapacidades Intelectuales y del Desarrollo (AAIDD) [2] define la DI como una discapacidad caracterizada por limitaciones en el funcionamiento intelectual y el comportamiento adaptativo, que abarca habilidades prácticas y sociales del diario vivir.

El funcionamiento intelectual hace referencia a la capacidad mental general, como el aprendizaje, razonamiento, resolución de problemas, etc. [2]. El comportamiento adaptativo se refiere a una colección de habilidades conceptuales, sociales y prácticas que son aprendidas y desarrolladas por las personas en su diario vivir [2].

Según la Clasificación Internacional de Enfermedades, décima revisión (CIE-10) [3], la discapacidad intelectual configura un estado mental incompleto, caracterizado por un deterioro de las capacidades que contribuyen al nivel global de inteligencia, y se manifiesta durante la fase de desarrollo. La discapacidad intelectual, según CIE-10 se codifica en el rango F70-F79, dando paso a los siguientes niveles de diagnóstico:

- F70 Retraso mental leve. Coeficiente intelectual aproximado de 50 a 69.
- F71 Retraso mental moderado. Coeficiente intelectual aproximado de 35 a 49.

- F72 Retraso mental grave. Coeficiente intelectual aproximadamente de 20 a 34.
- F73 Retraso mental profundo. Coeficiente intelectual inferior a 20
- F78 Otros tipos de retraso mental
- F79 Retraso mental, no especificado

También se usan las siguientes subdivisiones, en un cuarto carácter, con las categorías F70-F79, para codificar el grado de deterioro de la conducta [3]:

- 0. Deterioro del comportamiento nulo o mínimo
- 1. Deterioro del comportamiento significativo, que requiere atención o tratamiento
- 8. Otros deterioros del comportamiento
- 9. Deterioro del comportamiento de grado no especificado

1.1.1 Funcionamiento cognitivo: definición y medición de inteligencia

La evaluación de la inteligencia tanto en humanos como en sistemas ha acarreado un reto desde tiempo atrás, sin llegar a una solución unificada aún ahora, lo que se debe a que existen diferentes definiciones y/o interpretaciones de la inteligencia [4]. Las teorías de inteligencia pretenden identificar la mayor cantidad de componentes independientes del comportamiento inteligente [4]. En este sentido, algunas aproximaciones intentan definir una “inteligencia universal” [5], que sea capaz de medir cualquier tipo de inteligencia (biológica o computacional) y derivada de una precisa formalización con principios computacionales para cualquier contexto o entorno.

La definición de inteligencia no se refiere a un constructo simple; en su lugar, el término genérico se refiere a una red nomológica de diferentes constructos tales como habilidad mental, capacidades cognitivas y conocimiento aculturado [6]. Las habilidades mentales son definidas en general como las fuentes de varianza en la realización de una tarea, que requiere procesamiento mental, comprensión y manipular información, que se deriva en varias habilidades específicas (razonamiento cuantitativo, percepción visoespacial,

velocidad cognitiva); mientras que la habilidad cognitiva es general, y cubre todas aquellas habilidades mentales específicas, y fue formalmente propuesta por Charles Spearman en 1904 como Factor "g", el cual refleja todas las diferencias individuales en inteligencia [6], [21]. Por lo tanto, la habilidad cognitiva no es unidimensional, como ha sugerido mucha evidencia empírica basada en análisis factorial, y se compone de una amplia variedad de habilidades cognitivas altamente inter-correlacionadas (como habilidad numérica, habilidad verbal y habilidad espacial, entre otras) [7].

Raymond Cattell, entre 1991 y 1993, dividió el factor g de Spearman en dos factores generales: inteligencia cristalizada, o habilidades que son adquiridas por experiencia; e inteligencia fluida, que refleja el potencial natural y biológico del cerebro para el razonamiento lógico sin información previa [7], [8]. Muchas de las pruebas para medir coeficiente intelectual (CI) han sido derivadas de esta teoría factorial de la inteligencia, y algunos de ellos se refieren a un factor general de inteligencia, tales como la prueba de matrices progresivas de Raven y las escalas Wechsler (WPPSI, WAIS, WISC) [8], [9].

Por otro lado, otras teorías han sido propuestas en cuanto a la medición psicológica se refiere. Los dos principales modelos de medida psicológica con la Teoría Clásica de las Pruebas (CTT en inglés), que refiere al concepto de puntuación verdadera como valor esperado de la puntuación; y la Teoría de Respuesta al Ítem (IRT en inglés) que refiere a la puntuación de una variable latente unidimensional [8].

Existen también otros enfoques para medir la inteligencia en diferentes grados y cualidades, los cuales asumen que la inteligencia puede ser reconocida, identificada y medida por observación, los cuales son: observación y entrevista, rendimiento en tareas específicas, pruebas sistemáticas, enfoque de teoría de la información algorítmica y simulación de operaciones mentales (modelos y sistemas cognitivos) [10].

1.1.2 Medida de inteligencia: Coeficiente Intelectual

El CI (o cociente intelectual) es, en sí mismo, una medida del funcionamiento intelectual o inteligencia, y se deriva a partir de una serie de pruebas. Existen muchos tipos de pruebas cuyo objetivo es medir tanto habilidades generales como específicas: lectura, aritmética, vocabulario, memoria, conocimiento general, razonamiento abstracto, habilidades visuales y habilidades verbales, entre otros [1].

En la actualidad, los resultados de las pruebas se estandarizan en una muestra representativa de la población y, por tanto, las puntuaciones del CI de una persona son relativas a personas de la misma edad [1]. El resultado medio se define como 100 y una desviación estándar es de 15 puntos. Por lo tanto, el 95.4% de la población tiene puntuaciones dentro de dos desviaciones estándar de la media (es decir, dentro de un rango de CI de 70 a 130).

1.1.3 WISC-IV: Escala de inteligencia Wechsler para niños

Esta escala, de la cual se extrajeron los datos para la investigación, es un instrumento clínico de aplicación individual para evaluación de capacidad cognitiva en niños desde los 6 años hasta los 16 años con 11 meses de edad, que proporciona subpruebas y puntuaciones compuestas que representan el funcionamiento intelectual en dominios cognitivos específicos y una puntuación compuesta que representa la capacidad intelectual general o CI [11].

El WISC-IV está constituido por 15 subpruebas que conforman los 5 índices o puntuaciones compuestas, como son: CI total (CIT), que describe la habilidad intelectual general; Índice de Comprensión Verbal (ICV), que mide la habilidad verbal; Índice de Razonamiento Perceptual (IRP), el cual implica la manipulación de objetos o procesamiento de estímulos visuales para resolver problemas de manera no verbal; Índice de Memoria de Trabajo (IMT), que mide la memoria a corto plazo; e Índice de Velocidad de Procesamiento (IVP), el cual mide la eficiencia del procesamiento cognitivo; y toda la prueba tiene un tiempo estimado de aplicación de entre 65 a 80 minutos [11], [12].

En cuanto a la agrupación de las subpruebas principales entre los índices compuestos (ICV, IRP, IMT, IVP), se tiene tres para Comprensión verbal, tres para Razonamiento perceptual, dos para Memoria de trabajo y dos para Velocidad de procesamiento, las cuales se agrupan como sigue [11], [12]:

- Comprensión Verbal (ICV): Semejanzas, Vocabulario y Comprensión.
- Razonamiento Perceptual (IRP): Diseño con cubos, Conceptos con dibujos y Matrices.
- Memoria de Trabajo (IMT): Retención de dígitos y, Sucesión de letras y números.
- Velocidad de Procesamiento (IVP): Claves y Búsqueda de símbolos.

Las subpruebas que conforman el WISC-IV se dividen en 10 principales, como son: Diseño con cubos, Semejanzas, Retención de dígitos, Conceptos con dibujos, Claves, Vocabulario, Sucesión de números y letras, Matrices, Comprensión y Búsqueda de símbolos; así como 5 subpruebas suplementarias, como son: Figuras incompletas, Registros, Información, Aritmética y Palabras en contexto [11], [12]. A continuación, se transcribe la descripción dada por [12] para las 8 subpruebas que se usaron para esta investigación:

- *Diseño con cubos*: “Se pide al examinado que repita un conjunto de patrones geométricos bidimensionales en maqueta o impreso usando cubos rojos con blanco, dentro de un límite de tiempo específico.”
- *Semejanzas*: “Se pide al examinado que describa cómo dos palabras que representan objetos o conceptos comunes son similares.”
- *Retención de dígitos*: “En Retención de dígitos directo se pide al examinado que repita los números tal y como los dijo el examinador. En Retención de dígitos de orden inverso, se pide al examinado que repita números en el orden inverso a como los dijo el examinador.”

- *Conceptos con dibujos*: “Se requiere que el examinado elija un dibujo, de entre dos o tres hileras de dibujos presentados, para formar un grupo con una característica común.”
- *Vocabulario*: “Se pide al examinado que nombre dibujos o proporcione definiciones para las palabras.”
- *Sucesión de números y letras*: “Se lee al examinado una sucesión de números y letras y se le pide que repita los números en orden ascendente y las letras en orden alfabético.”
- *Matrices*: “Se pide al examinado que complete la parte faltante de una matriz de dibujos seleccionando una de cinco opciones de respuesta.”
- *Comprensión*: “Se pide al examinado que responda a una serie de preguntas basadas en su comprensión de principios generales y situaciones sociales.”

A partir de la puntuación del Cociente Intelectual Total (CIT) del WISC-IV, se generan unos intervalos sobre los cuales se mide los niveles de inteligencia; es decir, lo que se considera inteligencia normal o promedio puntúa entre 80 y 130 puntos del CIT, inteligencia superior tiene una puntuación de CIT mayor a 130; mientras que inteligencia límite puntúa entre 70 y 79 del CIT y discapacidad Intelectual puntúa por debajo de 69 puntos del CIT [3], [11], [12]. Además de eso, la DI también contempla diferentes niveles de profundidad, que también se miden en diferentes rangos del CIT por debajo de 69 puntos [1], [3].

Usando una validación de estructura alternativa de WISC-IV, basada la teoría de Catell-Horn-Carroll (CHC) [7], que estratifica los constructos de inteligencia fluida y cristalizada, el WISC-IV puede reagruparse en cinco factores que miden: Inteligencia cristalizada (Gc), que refiere a la amplitud y profundidad del conocimiento adquirido por una persona acerca de una cultura y su aplicación efectiva; el Procesamiento visual (Gv), que es la habilidad de generar, percibir, analizar, sintetizar almacenar, recuperar, manipular, transformar y pensar con patrones y estímulos visuales; Razonamiento fluido (Gf), que refiere a las operaciones mentales que utiliza un individuo cuando se enfrenta a una tarea relativamente nueva; Memoria a corto plazo (Gsm), que es la habilidad de aprehender y mantener información en la consciencia inmediata y después usarla en cuestión de poco tiempo; y

Velocidad de procesamiento (Gs), que es la habilidad de realizar tareas cognitivas de manera fluida y automática, especialmente bajo presión para mantener la atención y la concentración enfocada [12], las cuales agrupan las subpruebas principales del WISC-IV como sigue:

- Procesamiento visual (Gv): Diseño con cubos, Matrices y, Búsqueda de símbolos
- Razonamiento fluido (Gf): Conceptos con dibujos y Matrices
- Memoria a corto plazo (Gsm): Retención de dígitos y Sucesión de letras y números
- Velocidad de procesamiento (Gs): Claves y Búsqueda de símbolos
- Inteligencia cristalizada (Gc): Semejanzas, Vocabulario y Comprensión.

▪ **Limitaciones de la prueba WISC-IV**

Una limitación de la prueba surge cuando dicha puntuación se ubica muy cerca del límite entre dos categorías diagnósticas. Por ejemplo, si un niño que es evaluado presenta una puntuación de 69 o 70 en el CIT, no se podría definir con certeza si su condición es de inteligencia límite o discapacidad intelectual, aun teniendo en cuenta el intervalo de confianza calculado para dicha puntuación [12], lo que eventualmente requerirá que se apliquen otra serie de pruebas que midan el funcionamiento intelectual para definir la condición actual del niño a nivel cognoscitivo, y así mismo establecer la mejor ruta de intervención, como también otra serie de decisiones a nivel educativo o médico que afectarán a largo plazo al niño evaluado.

Otra limitación que se puede presentar con la aplicación del WISC-IV es que los índices o puntuaciones compuestas (ICV, IRP, IMT, IVP) den valores muy lejanos o dispersos entre ellos. En este caso, es necesario considerar los cuatro Índices del WISC-IV, restar el Índice más bajo del más alto y verificar si la magnitud de la diferencia de la puntuación estándar es menor de 23 puntos (menor de 1.5 desviaciones estándar) [12]. Del mismo modo, si la diferencia entre puntuaciones escalares de subpruebas que pertenecen al mismo índice es inusualmente grande [12], si la diferencia es mayor, automáticamente da como

resultado que la prueba se considere como “no interpretable”, puesto que no se puede resumir la habilidad intelectual global en una sola puntuación (CIT) o los índices no son unitarios y, como consecuencia, acarrea los mismos inconvenientes de la limitación anterior [11], [12].

1.2 Ciencia de datos

Como campo de acción, la ciencia de datos se refiere a una emergente área de trabajo que engloba la recolección, preparación, análisis, visualización y preservación de una gran cantidad de información [13]. Por otro lado, desde una perspectiva disciplinar, la ciencia de datos es un nuevo campo interdisciplinario que se construye a partir otros campos como estadística, informática, computación, comunicación, administración y sociología para estudiar los datos y su contexto, con el fin de transformar dichos datos en nuevo conocimiento que soporte la toma de decisiones [14].

La ciencia de datos distingue un conjunto de actividades clave para lograr transformar los datos en conocimiento o decisiones, las cuales son: adquirir u obtener los datos necesarios; preparar los datos con tareas de adecuación, limpieza y organización; analizar con el fin de encontrar conexiones o patrones en los datos y; actuar o tomar acción en la solución de problemas, toma de decisiones o adquirir nuevo conocimiento [15]. Un modelo de estados más completo y maduro de las actividades en ciencia de datos se usa como un marco común para describir el progreso en la ganancia de conocimiento o el producto de datos final. Dichos estados son: recolectar bases de datos internas y externas; describir o aplicar estadística descriptiva a los datos; descubrir o identificar patrones o relaciones ocultas; predecir utilizando observaciones pasadas para revelar futuras observaciones y, finalmente, informar o asesorar sobre posibles decisiones, optimizaciones o acciones a seguir [15].

1.3 Descubrimiento de Conocimiento en Bases de Datos KDD

En el marco de Ciencia de datos, el término KDD (Descubrimiento de conocimiento en bases de datos, en español) se emplea para describir todo el proceso de extracción de conocimiento a partir de datos y, en este contexto, conocimiento significa relaciones y patrones entre elementos de datos [16]. El proceso de descubrimiento de conocimiento es iterativo e interactivo, y consta de nueve pasos: 1) Estudio del dominio de aplicación; 2) Creación del conjunto de datos objetivo; 3) Limpieza y preprocesamiento de datos; 4) reducción y transformación de datos; 5) Compaginar los objetivos del KDD con los métodos de minería; 6) Seleccionar los métodos de minería de datos; 7) aplicar los métodos de minería; 8) Evaluación e interpretación de resultados; 9) Presentación o documentación del conocimiento descubierto [16], [17].

1.3.1 Estudio del dominio de aplicación

Este primer paso, en términos del desarrollo de la tesis, consiste en la identificación del problema y el entendimiento del conocimiento relevante que lo componen [17], de modo que haya sinergia entre el objetivo de la tesis con el proceso de KDD.

1.3.2 Creación de un conjunto de datos objetivo

Este segundo paso consiste en seleccionar un conjunto de datos o, si el conjunto de datos está establecido, enfocarse en un subconjunto de variables o muestras que va a ser utilizadas para el descubrimiento de conocimiento [17]. En términos de Ciencia de datos, se habla de la recolección o el acceso a los datos, ya sea que pertenezcan a una organización o sean de acceso público [15].

1.3.3 Limpieza y preprocesamiento de datos

Este tercer paso incluye todas aquellas operaciones encaminadas a manipular los datos y prepararlos de acuerdo con las necesidades de análisis [15] u objetivos de la investigación.

Estas operaciones incluyen remover ruido, tratamiento de datos perdidos o remoción de datos redundantes [17].

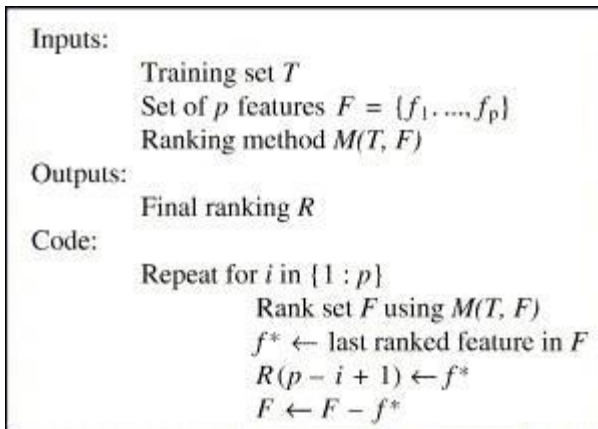
1.3.4 Reducción y transformación de datos

El cuarto paso hace referencia a encontrar aquellas características que mejor representen el conjunto de datos [17] de acuerdo con los objetivos de la investigación. Estas características pueden ser seleccionadas a partir de técnicas de reducción de dimensionalidad [17] o teniendo en cuenta los criterios del investigador, conforme al dominio de conocimiento en el problema identificado. Por otro lado, si es necesario, se puede acudir a estrategias de transformación de datos, como normalización (los datos deben estar en un rango específico), conversión de formato (p. ej., binario a decimal) o transformación de dominio (p. ej., tiempo a frecuencia) [15]. A continuación, se describe el método usado para la selección de características más relevantes en el desarrollo de esta tesis.

- **Eliminación recursiva de características**

En la Eliminación recursiva de características (RFE, acrónimo en inglés), dado un estimador que asigna pesos o importancias a cada característica (Modelo de Bosque aleatorio para esta investigación), el objetivo de RFE es ir eliminando la característica menos relevante en cada iteración del algoritmo, a la vez que se vuelve a medir la importancia de las características con el estimador de pesos, como se muestra en la Figura 1-1 [18], [19].

Figura 1-1: Seudocódigo del algoritmo RFE



Tomado de [19]

En cada iteración, RFE va considerando recursivamente un set cada vez más pequeño de características, hasta encontrar el número deseado de ellas [18]. Esta recursividad es necesaria con respecto a la medición de la importancia relativa de cada característica, dado que dicha importancia puede cambiar sustancialmente en cada subconjunto remanente de características, sobremanera para características altamente correlacionadas [19].

1.3.5 Compaginar los objetivos del KDD con los métodos de minería

Una vez realizados los pasos anteriores, se puede retomar el objetivo inicial, de forma que se analicen los métodos apropiados para cumplir con dicho objetivo [15], teniendo en cuenta las características del conjunto de datos. Por ejemplo, clasificación, regresión, agrupación, entre otros [15], [17].

1.3.6 Seleccionar los métodos de minería de datos

En este punto se deben seleccionar específicamente los algoritmos de minería de datos para buscar patrones en los datos. Esto incluye decidir qué modelos y parámetros pueden

ser apropiados [17]. Dado que los modelos que se usaron para esta tesis están basados en aprendizaje automático, a continuación, se describirá con más detalle este concepto:

- **Aprendizaje automático**

El aprendizaje automático (en inglés, machine learning) se refiere a cómo los sistemas aprenden de los datos a partir de métodos o modelos de entrenamiento [20]. Existen dos clasificaciones básicas para los modelos de aprendizaje automático: aprendizaje supervisado y no supervisado. El supervisado tiene dos tipos: clasificación, consiste en aprender de datos que están etiquetados para un conjunto de atributos dado; y regresión, en el que se predice un valor numérico a partir de variables de características. El aprendizaje no supervisado consiste en el proceso de encontrar los patrones o conexiones en los datos que no están etiquetados [20].

Otras clasificaciones de los modelos de aprendizaje automático están relacionadas con la naturaleza del problema a resolver. En este sentido existen dos tipos fundamentales: clasificación y regresión. Los métodos de clasificación consisten en predecir la clase o grupo, dados unos datos de entrada; mientras que los métodos de regresión consisten en predecir un valor futuro a partir de los datos de entrada [15]. Algunos ejemplos de algoritmos de clasificación son: bosque aleatorio, redes neuronales, árbol de decisión y regresión logística; mientras que algunos ejemplos de algoritmos de regresión son: regresión lineal y regresión paso a paso (en inglés, stepwise regression).

1.3.7 Aplicar los métodos de minería

En este paso se implementan los métodos de minería seleccionados, en la búsqueda de patrones de interés, ya sea en una representación particular o un conjunto de representaciones [17]. A continuación, se profundizará en aquellos métodos que fueron de utilidad para la investigación:

▪ Árboles de decisión

Como forma de representación, el árbol de decisión es un diagrama de flujo con estructura de árbol, donde los nodos representan una prueba condicional de un atributo particular, mientras que cada rama representa el resultado de dicha prueba y, al final, cada hoja (o nodo terminal) representa la etiqueta de clase [21], [22].

Como método de aprendizaje automático, los árboles de decisión son un método de aprendizaje supervisado, no paramétrico, usado para clasificación y regresión [23]. El objetivo de este método es predecir el valor de la variable objetivo, a partir del mapeo de observaciones sobre características de los datos, por medio del aprendizaje de reglas simples [21], [23].

Cada nodo, conforme se divide a partir de la prueba condicional, debería resultar en un subconjunto de datos homogéneo asociado a una clase particular; sin embargo, esto es casi imposible de obtener en un conjunto de datos de la vida real [22]. En este sentido, la noción de impureza fue introducida como criterio de división en cada nodo para medir este valor de impureza; varios índices se han propuesto, incluyendo el GINI y la ganancia de información, entre otros [22].

Algunas ventajas de los árboles de decisión son [23]:

- Simple de entender e interpretar, así como de visualizar.
- Requiere poca preparación de los datos.
- Puede trabajar tanto con datos numéricos como categóricos.
- Puede manejar problemas multiclase.
- Capacidad de medir la importancia de las características a partir de los índices de impureza [22].

▪ **Bosque aleatorio**

Bosque aleatorio es un método de aprendizaje de conjunto (Ensemble learning method, en inglés) usado para clasificación y regresión, el cual combina múltiples modelos para resolver el mismo problema, así como ganar mayor precisión [24]. Este método combina el método de muestreo de Bootstrap y la selección aleatoria de características, en orden que construir una colección de árboles de decisión con variaciones controladas [19], [24].

Cada árbol construido actúa como un clasificador de base para determinar la etiqueta de clase (en el caso de clasificación), por tanto, cada árbol lanza una etiqueta de clase predicha y, al final, la etiqueta de clase se determina por mayoría de votos [19], [24].

Algunas de las ventajas del Bosque aleatorio son [24]:

- Por generalización, evita el sobreajuste (overfitting) y mejora la precisión.
- Es simple y fácilmente paralelizado.
- Adopta las ventajas del árbol de decisión y reduce sus limitaciones.
- Capacidad de medir la importancia de las características a partir del error de predicción [19].

▪ **Regresión logística**

La regresión logística es un método que intenta modelar una variable binaria (0 o 1) en función de una o más variables independientes, por medio de un modelo de probabilidad basado en la función logística o logit [20]. En este sentido, hay una relación con los modelos de clasificación, de modo que se intenta relacionar observaciones con categorías de salida, como es el caso de esta investigación, al relacionar los puntajes observados de la prueba WISC-IV y el diagnóstico de discapacidad intelectual.

El modelo logístico se expresa como muestra la Figura 1-2:

Figura 1-2: Modelo de Regresión logística:

$$y = \frac{e^{f(x)}}{1 + e^{f(x)}}, \quad f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Tomado de [20]

En este modelo de regresión logística, Y está restringida al rango $(0, 1)$ y es la variable categórica u objetivo, $x = \{x_1, x_2, x_3, \dots, x_K\}$ representan las observaciones o variables independientes y, $\beta = \{\beta_0, \beta_1, \beta_2, \dots, \beta_K\}$ son los coeficientes o pesos que se ajustan durante el entrenamiento del modelo [20].

Una vez ajustados los coeficientes del modelo, se puede calcular la variable Y usando un nuevo conjunto de observaciones [20]. Para esta investigación, se calcula el valor del diagnóstico (0 o 1) usando las observaciones de las puntuaciones en los reactivos de la prueba de inteligencia. Dado que el resultado no es exactamente 0 o 1, sino la probabilidad de diagnóstico positivo, se adopta un punto de corte, generalmente 0.5, de modo que si el valor de probabilidad es superior a 0.5, entonces la predicción será un diagnóstico positivo, y viceversa.

1.3.8 Evaluación e interpretación de resultados

El octavo paso del KDD refiere a la interpretación de los patrones minados y, dados los resultados, requerirá volver a cualquiera de los 7 pasos anteriores [17]. En este paso se desarrollan las visualizaciones de los patrones extraídos y el desempeño de modelos implementados [17].

- **Medidas de desempeño.**

En muchos casos en los que un diagnóstico de salud es realizado, se distingue entre dos grupos que reflejan la presencia o ausencia de una condición médica relevante [25], como es el caso del tema de investigación, en el cual, el modelo computacional refleja la presencia o ausencia de discapacidad intelectual. Sin embargo, como el modelo no es efectivo siempre, habrá casos en los que la predicción arroje lo opuesto a la condición real

del paciente. Por tanto, los resultados del modelo, tanto aciertos como fallos en el diagnóstico, pueden resumirse en lo que se conoce como matriz de confusión [25].

Tabla 1-1: matriz de confusión

		Test result	
		0	1
True situation	0	True negative (TN)	False positive (FP)
	1	False negative (FN)	True positive (TP)

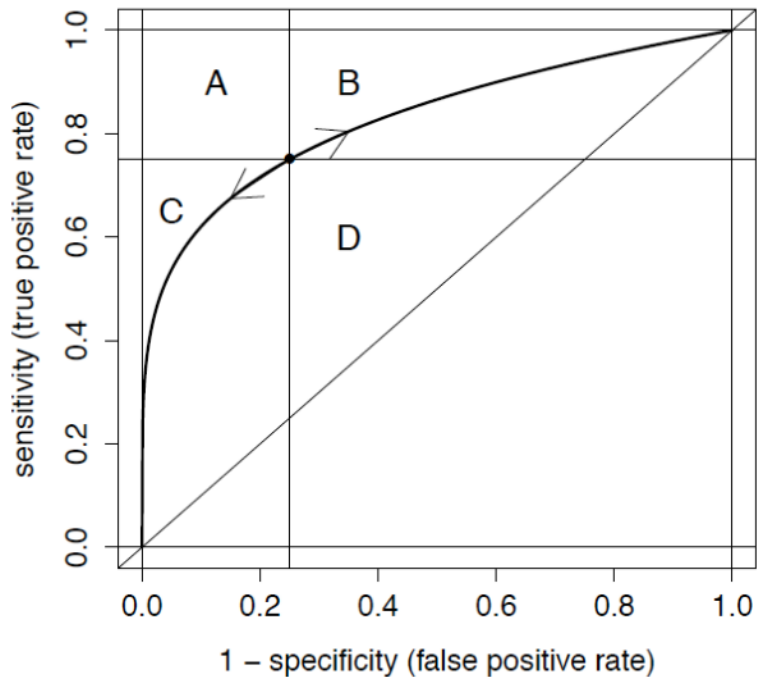
Tomado de [25]

La matriz de confusión está compuesta por los verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). A partir de ella, se puede determinar el desempeño de la predicción, utilizando medidas como exactitud, sensibilidad, especificidad y precisión.

La *Exactitud* (o prevalencia), que es la medida más general de rendimiento o porcentaje de predicciones correctas, se define como $(TP+TN)/(TP+TN+FP+FN)$; la *Sensibilidad*, que es el porcentaje de predicciones positivas correctas para el total de pacientes con presencia de la condición médica, se define como $TP/(TP+FN)$; la *Especificidad*, definida como el porcentaje predicciones negativas correctas para pacientes con ausencia de la condición médica, se define como $TN/(TN+FP)$; y la *Precisión* como el porcentaje de predicciones positivas correctas $TP/(TP+FP)$ [25].

▪ Curvas ROC

Adicionalmente a las métricas descritas a partir de la matriz de confusión, se puede obtener una representación gráfica del rendimiento del modelo a partir de los valores de sensibilidad y especificidad, en las curvas ROC (acrónimo de Receiver Operating Characteristic, en inglés) [25]. En esta curva, como lo muestra la Figura 1-3, para cada umbral en los valores del clasificador se obtiene un valor de sensibilidad y especificidad, de modo que decrementando el umbral se incrementará la sensibilidad y decrementa la especificidad, mientras que decrementando el umbral sucederá lo contrario [25].

Figura 1-3: Ejemplo de curva ROC

Tomado de [25]

Una forma de resumir la curva ROC es mediante el área bajo la curva o AUC (Area under the ROC curve, en inglés), donde un AUC sobre 0.5 indica que el clasificador no es bueno en predecir las clases o categorías objetivo, mientras que valores superiores indican un buen nivel de clasificación [25]. Por el contrario, si el valor de AUC está por debajo de 0.5 indica que las clases están fuera de lugar o intercambiadas; por tanto, un valor mayor a 0.5 para AUC en el modelo de clasificación es informativo para los resultados [25].

1.3.9 Presentación o documentación del conocimiento descubierto

Por último, este paso involucra el uso directo del conocimiento obtenido a lo largo del proceso de KDD, ya sea con la implementación de un sistema de acción posterior o la

documentación de resultados [17]. Este paso también puede involucrar la toma de decisiones, en el caso de las organizaciones, que pueden llegar a tener un impacto económico, social o político [15], como también puede implicar la revaluación de creencias previas o conflictos con el conocimiento extraído [17].

1.4 Herramientas computacionales en salud mental

Ciencia de datos, como se mencionó antes, consiste en la extracción de conocimiento a partir de un gran volumen de datos, usando habilidades en computación, estadística y conocimiento de expertos, que en el ámbito de la salud puede revolucionar el monitoreo, diagnóstico, tratamiento y prevención de la enfermedad [26]. En salud mental, los desórdenes mentales representan la mayor carga “oculta” de mala salud en la sociedad a largo plazo, dada la complejidad que puede representar el diagnóstico, por lo que la analítica avanzada de la ciencia de datos sobre los vastos conjuntos de datos longitudinales (incluyendo registros electrónicos de prescripciones, educación, bienestar, socio-demográfico, laboratorios y monitoreo en tiempo real) ofrece una gran oportunidad para diagnósticos más robustos, predicción de consecuencias y respuestas a tratamiento, así como las preferencias de los pacientes para la intervención [26].

En este mismo sentido, Big Data, que ha estado de moda en todos los campos, ofrece la posibilidad de la rápida integración de enormes cantidades de información de diferentes fuentes, incluyendo la amplia disponibilidad de tecnologías móviles y medios sociales (con sensores que monitorean actividad física, patrones de sueño, ubicación, biometría y patrones de comportamiento), y datos biomédicos (registros electrónicos médicos y administrativos, genética, biomarcadores, neuroimagen, etc.) [27], que junto con procesos de minería de datos, podría ofrecer conclusiones innovadoras en todas las especialidades de las ciencias de la salud y del comportamiento [28].

Otro aporte del Big Data, enfocado principalmente en el uso de teléfonos móviles y redes sociales, es definir metodologías que sirvan tanto a la psicología como a la psiquiatría,

para mediciones confiables de emoción, cognición y comportamiento, que junto con otras fuentes de información, promueven avances teóricos en dominios psicológicos como la personalidad y el bienestar, con miras a mejorar el diagnóstico, tratamiento clínico y prevención de la enfermedades mentales como la depresión y las adicciones [29], [30]. Como es el caso de esta investigación, donde usando información existente sobre funcionamiento cognitivo, se puede hacer más eficiente el diagnóstico de DI, permitiendo hacer un mejor uso de recursos diagnósticos, así como tomar mejores decisiones en cuanto a la intervención educativa de niños con este tipo de discapacidad.

Una vez integrados los datos de grandes fuentes de información, entran a actuar las técnicas o métodos para modelar los datos y así extraer nuevo conocimiento, los cuales han sido derivadores de los avances en la inteligencia artificial (IA). Entre las técnicas principales se encuentra el aprendizaje de máquina y las redes neuronales artificiales, cuyo objetivo es dar a los computadores la habilidad de aprender sin estar explícitamente programados, a través de la exposición a los datos y experiencia [31]. El aprendizaje de máquina tiene muchos subcampos y aplicaciones, incluyendo métodos de aprendizaje estadístico, redes neuronales, aprendizaje basado en instancias, algoritmos genéticos, minería de datos, reconocimiento de imágenes y procesamiento de lenguaje natural, entre otros [31].

Entre los beneficios de uso de las técnicas basadas en IA para el cuidado de la salud mental, se encuentra la integración y personalización del cuidado, donde máquinas inteligentes que proporcionan cuidados son programadas con diversos enfoques basados en evidencia, para tener la capacidad de sensibilidad y adaptación a aspectos específicos del paciente, y entonces predecir la terapia más apropiada [31]; así como mejorar las capacidades de los profesionales de la salud, a partir de la integración y análisis de toda la información del mundo real de los pacientes, en tiempo real, para la toma de decisiones clínicas, usando computación cognitiva [32].

En cuanto al modelamiento predictivo de la información en salud mental, se encuentran varios enfoques o tipos, entre ellos: modelo de evaluación del riesgo, que se enfoca en la

predicción de problemas de salud mental usando técnicas de minería en los datos epidemiológicos existentes, para identificar factores de riesgo de enfermedad que no son fácilmente detectables con técnicas estadísticas tradicionales; modelo de predicciones a corto plazo durante el tratamiento, enfocado en la predicción de los estados de evolución del paciente durante la intervención, con el fin de mejorar la adherencia al tratamiento y los resultados, proveyendo ciclos de retroalimentación; modelo de predicción del resultado del tratamiento, que se enfoca en predecir el resultado del proceso completo de tratamiento, incluyendo el abandono del mismo, mientras éste se encuentre activo; y modelo para predicción del riesgo de recaída, que es la recurrencia de los síntomas en el largo plazo y determina la cantidad de postcuidado requerido [33].

Dentro de las técnicas de modelamiento predictivo usadas en salud mental se encuentran: los árboles de decisión, enfoques bayesianos, máquinas de soporte vectorial, algoritmos genéticos, análisis discriminante lineal, reglas de asociación y redes neuronales artificiales [33], [34]. Encontrando, por ejemplo, que en estudios de Alzheimer se han aplicado: redes neuronales artificiales, clasificador Bayesiano, bosque aleatorio, máquinas de soporte vectorial, reglas de asociación, JRIP y árbol de decisión; en demencia se han aplicados técnicas de minería como: redes neuronales artificiales, clasificador Bayesiano, bosque aleatorio, máquinas de soporte vectorial, árbol de decisión y regresión logística; mientras que en depresión, se han aplicado técnicas de redes neuronales artificiales, clasificador Bayesiano, bosque aleatorio, máquinas de soporte vectorial, árbol de decisión, regresión logística, k vecinos más cercanos y regresión lineal [34].

1.4.1 Aplicaciones computacionales en la evaluación de información psicológica

En cuanto a las aplicaciones computacionales y técnicas de minería de datos aplicadas en salud mental, se encontraron dos referencias al diagnóstico de discapacidad cognitiva: la primera, de 2007, se propone la creación de una herramienta computacional para la correcta y rápida clasificación de nivel de retraso mental, apoyándose en técnicas computacionales como lógica difusa y algoritmos genéticos [35]. La segunda es una

aplicación de un sistema experto, basado en lógica difusa y sus hibridaciones, que evalúa datos recolectados de pacientes afectados por diferentes grados de discapacidad intelectual y que están bajo tratamiento, y que muestra buenos resultados en el diagnóstico automático debido al manejo de numerosas fuentes de incertidumbre [36].

Se encontraron varias aplicaciones computacionales enfocadas al diagnóstico de dislexia, dado que esta discapacidad del aprendizaje conlleva, como parte del diagnóstico, la evaluación de capacidad intelectual, como son: el uso de técnicas de clasificación usando test screening basado en computador para la detección temprana de dislexia, usando clasificador bayesiano, con un resultado de 97.6% de precisión diagnóstica [37]; diagnóstico de la dislexia mediante análisis computacional usando clasificadores como k-means, redes neuronales artificiales y lógica difusa para el modelo analítico propuesto [38]; y un estudio de la dislexia utilizando diferentes algoritmos de aprendizaje automático como árbol de decisión, máquina de soporte vectorial, clasificador bayesiano y red neuronal, con técnicas de extracción de datos, basándose en investigación sobre Coeficiente Intelectual (IQ en inglés) y Coeficiente de Inteligencia Emocional (EQ en inglés) [39].

Aplicado a diagnósticos múltiples, un estudio de 2011 tiene por objetivo el uso de la computadora para simular los modos de pensamiento del cerebro humano, y establecer un sistema experto de inteligencia artificial para el diagnóstico y tratamiento de trastornos de salud mental infantil basados en redes neuronales artificiales y sistemas expertos, involucrando psiquiatría infantil, psicología infantil, medición psicológica, terapia psicológica y ciencias de la computación; y combinando datos de diagnóstico estándar de CIE-10, DSM-IV, CCMD-2, así como datos epidemiológicos y clínicos [40]; una aplicación de Big Data para estudio y archivo de datos de salud mental, utilizando MongoDB, para diferentes tipos de desórdenes mentales [41]. Así como un enfoque de la minería de datos con árboles de decisión para analizar los resultados vocacionales de los individuos con discapacidades psiquiátricas que recibieron servicios estatales de rehabilitación vocacional [42].

Otras aplicaciones han sido conducidas para evaluar la salud mental de estudiantes, como por ejemplo el uso de técnicas de minería de datos en un conjunto de datos de salud mental de estudiantes universitarios, usando reglas de aprendizaje supervisado, para encontrar perfiles de información útiles en los problemas de los estudiantes [43]; un estudio sobre la aplicación del algoritmo del árbol de decisión en la información psicológica de estudiantes universitarios vocacionales, analiza la insuficiencia del sistema de manejo psicológico existente y proponiendo la tendencia de desarrollo de la educación en salud psicológica en las universidades [44]; un diseño de un sistema de administración y minería de datos para salud mental de estudiantes universitarios, basado en redes neuronales y usando la herramienta MATLAB 2014a [45]; y la estimación del bienestar psicológico de los adolescentes basada en un algoritmo de minería de datos [46].

Otras implementaciones buscan el modelamiento general de la salud mental, por ejemplo: ajuste de riesgo para la salud mental con categorías clínicas y aprendizaje automático [47]; y modelado conductual para la salud mental utilizando algoritmos de aprendizaje automático [48]. Por último, se encontraron aplicaciones enfocadas en la evaluación de los avances en las aplicaciones de inteligencia artificial general, como, por ejemplo: medición del rendimiento de un sistema de inteligencia artificial en una prueba de coeficiente intelectual verbal para niños pequeños [49]; un enfoque para evaluar los sistemas de razonamiento de sentido común [50]; un análisis computacional de pruebas de inteligencia general para evaluar desarrollo cognitivo [51]; y la medición del razonamiento abstracto en redes neuronales [52].

A continuación, a modo de resumen, se muestra en la tabla 1-2 las técnicas de aprendizaje automático más utilizadas según el enfoque de diagnóstico en salud mental:

Tabla 1-2: resumen técnicas de aprendizaje automático en salud mental

Discapacidad Intelectual	Dislexia	Diagnósticos múltiples
Lógica Difusa	Clasificador Bayesiano	Redes Neuronales
Algoritmos Genéticos	K-means	Sistemas Expertos
Sistemas Expertos	Redes Neuronales	Arboles de Decisión
	Lógica Difusa	
	Maquina de Soporte Vectorial (SVM)	
	Arbol de Decisión	

Elaboración propia

2.Planteamiento del problema y metodología

En este capítulo se desarrolla una descripción del problema que se abordó en esta tesis y los objetivos propuestos. Posteriormente, se presenta la metodología implementada para llevarla a término.

2.1 Descripción del problema

La revisión de antecedentes muestra que se han hecho algunos acercamientos en cuanto al análisis de información diagnóstica en pacientes con DI, usando principalmente lógica difusa, con el fin de lograr crear un sistema capaz de hacer diagnóstico automático efectivo [25], [26]. Pero la literatura muestra un vacío en cuanto a la investigación usando métodos computacionales exploratorios o analíticos, que dé cuenta de los componentes asociados al CI, que aporten más información o claridad al diagnóstico de DI.

En este sentido, y dadas las limitaciones descritas de la prueba de inteligencia WISC IV en el primer capítulo, esta tesis presenta el desarrollo un modelo computacional que determina aquellas variables asociadas al Coeficiente Intelectual que mejor predicen el diagnóstico de Discapacidad Intelectual, e implementa un modelo predictivo con dichas variables. De esta forma, se crea un modelo reducido de la prueba de inteligencia, aplicando aprendizaje automático, que permitirá a los psicólogos hacer una evaluación rápida y efectiva en los casos donde los evaluados presentan sospecha de discapacidad intelectual. Además, aporta conocimiento sobre los dominios del funcionamiento intelectual que mejor predicen dicho diagnóstico.

2.2 Objetivos

2.2.1 Objetivo general

Desarrollar un modelo computacional para determinar agrupaciones de variables, dimensiones o componentes asociados al Coeficiente Intelectual (CI) que predican el diagnóstico de Discapacidad Intelectual (CIE 10 F70-79).

2.2.2 Objetivos específicos

- Identificar patrones sobre el conjunto de datos de datos de Coeficiente Intelectual a través de la selección y aplicación de técnicas de minería de datos para análisis exploratorio, en el contexto de evaluación psicológica del Servicio de Atención Psicológica (SAP) de la Universidad Nacional de Colombia.
- Predecir casos de discapacidad cognitiva a través de la selección e implementación de modelos de aprendizaje automático sobre el conjunto de datos, teniendo en cuenta métricas de desempeño de los modelos y los criterios de diagnóstico clínico.
- Desarrollar un modelo computacional que permita identificar las características que aportan más información, así como predecir la ocurrencia de casos de discapacidad intelectual.

2.3 Metodología

Con el fin de llevar esta investigación a buen término, se implementó una metodología estructurada en cuatro etapas, las cuales dieron lugar a la consecución de los objetivos propuestos. Como el objetivo principal del proyecto fue desarrollar un modelo computacional analítico, se usaron métodos y herramientas en el marco de Ciencia de datos, vinculando procesos de KDD y técnicas de aprendizaje automático, los cuales no solo ayudaron a generar un modelo predictivo e interpretable, sino que dio lugar a nuevo conocimiento que ayudará al diagnóstico diferencial de discapacidad intelectual. A continuación, se describen las etapas seguidas en el desarrollo de la investigación:

- **Etapa 1: Selección y preprocesamiento de datos**

En esta etapa se realizó la recopilación y preparación de los datos, los cuales fueron gestionados por el Servicio de Atención Psicológica (SAP) de la Universidad Nacional, y contienen tanto información socioeconómica de los evaluados, como de las puntuaciones obtenidas en la prueba de evaluación de funcionamiento WISC-IV. La preparación de los datos incluyó tratamiento de datos perdidos e imputación de valores, dados los errores de digitación con los que se encontraba la base inicialmente, así como eliminación de variables con alto porcentaje de datos perdidos.

- **Etapa 2: Análisis Exploratorio de datos**

Una vez el conjunto de datos fue limpiado, se efectuó una exploración general del conjunto de datos usando estadística descriptiva, con el fin de describir la información socioeconómica que comprende el conjunto de datos y hallar patrones en la población.

- **Etapa 3: Transformación y modelamiento de datos**

En esta etapa se filtró y dicotomizó la variable de diagnóstico final, de modo que solo tuviera en cuenta si los evaluados eran positivos o negativos para discapacidad intelectual. También se efectuó balanceo de clases, dada la diferencia de proporción entre las dos etiquetas del diagnóstico. Posteriormente se implementaron los métodos de selección de características y de aprendizaje automático necesarios.

- **Etapa 4: Evaluación del modelo computacional**

En esta etapa se evaluó la calidad del modelo en términos de las métricas de rendimiento, y se analizan los resultados a la luz de los conocimientos previos sobre inteligencia y discapacidad intelectual.

3. Preprocesamiento y análisis descriptivo de los datos

En este capítulo se describe las estrategias implementadas para la limpieza y preparación de datos

3.1 Organización del conjunto de datos

El conjunto de datos obtenido cuenta con una base de datos de 349 columnas y 502 registros, y fue obtenido a partir de la transcripción manual que miembros del Servicio de Atención Psicológica (SAP), de la Universidad Nacional, han hecho de ciertos datos de las historias clínicas que han considerado relevantes, a un archivo en formato Excel, y que conformó la base de datos necesaria para el desarrollo de esta investigación. Por tal motivo, la base de datos no fue construida a partir de un diseño de modelo de datos. El modelo de datos considerado para esta base tiene un solo sentido, que relaciona el ID de la persona evaluada con una serie de objetos o atributos que la conforman, en tres categorías: datos socioeconómicos, datos de consulta y puntuaciones en la prueba WISC-IV, como se observa en la siguiente lista:

1. ID (Identificación del evaluado)
 - a. Datos socioeconómicos
 - i. Género
 - ii. Fecha de nacimiento
 - iii. Fecha de aplicación
 - iv. Escolaridad
 - v. Tipo de colegio
 - vi. Escolaridad padre
 - vii. Escolaridad madre

- viii. Estrato
- ix. Procedencia
- b. Datos de consulta
 - i. Motivo de consulta
 - ii. Diagnóstico
- c. Datos WISC-IV (Puntajes en la prueba de Inteligencia)
 - i. Diseño con cubos: 14 campos
 - ii. Semejanzas: 23 campos
 - iii. Retención de dígitos: 16 campos
 - iv. Conceptos con dibujos: 28 campos
 - v. Vocabulario: 36 campos
 - vi. Sucesión de números y letras: 10 campos
 - vii. Matrices: 35 campos
 - viii. Comprensión: 21 campos
 - ix. Figuras incompletas: 38 campos
 - x. Información: 33 campos
 - xi. Aritmética: 34 campos
 - xii. Palabras en contexto: 24 campos

3.1.1 Descripción datos socioeconómicos

Los datos socio económicos comprenden datos básicos, tanto del niño evaluado, como de los padres. Estos datos están constituidos por 9 campos, como son: género, fecha de nacimiento, fecha de aplicación, escolaridad, tipo de colegio, escolaridad padre, escolaridad madre, estrato y procedencia. A continuación, en la tabla 2-1, se desglosan las categorías de cada campo:

Tabla 3-1: Descripción de los campos de la prueba WISC-IV

Género	Escolaridad	Tipo de colegio	Escolaridad del padre	Escolaridad de la madre	Procedencia
			0. No se tienen datos	0. No se tienen datos	
1. Mujer	1. Primaria y preescolar	1. Publico	1. Sin escolaridad	1. Sin escolaridad	1. Bogotá
2. Hombre	2. Secundaria	2. Privado	2. Primaria	2. Primaria	2. Cundinamarca
			3. Secundaria	3. Secundaria	3. Boyaca
			4. Profesional	4. Profesional	4. Tolima
			5. Postgrado	5. Postgrado	5. Nariño
			6. Tecnico	6. Tecnico	6. Antioquia
			7. Tecnológico	7. Tecnológico	7. Huila
					8. Santander
					9. Valle
					10. Casanare
					11. Otro

Elaboración propia

Estas categorías de los campos socioeconómicos fueron definidas desde el SAP, y son una forma de estandarizar y limitar el número de opciones posibles para cada campo.

3.1.2 Descripción datos de consulta

Los datos de consulta comprenden 2 campos: Motivo de consulta y Diagnóstico. El motivo de consulta hace referencia a la razón principal por la cual el niño o evaluado fue remitido al SAP para valoración. El diagnóstico es, en síntesis, el resultado final de todo el proceso de evaluación, que da cuenta de la afección o enfermedad (en términos de salud mental) que aqueja al paciente. A continuación, se listan las categorías para el campo de motivo de consulta:

- A. Evaluación Cognoscitiva
- B. Habilidades
- C. interés propio
- D. Problema de atención
- E. Problemas de aprendizaje
- F. Bajo rendimiento académico
- G. Valoración desarrollo
- H. Orientación vocacional sencilla

- I. Reingreso y traslado
- J. Problemas emocionales
- K. Retraso mental
- L. Discapacidad cognitiva

En cuanto al diagnóstico, se establece por medio de la codificación CIE-10, que es la Clasificación internacional de enfermedades, en su décima edición, y determina la clasificación y codificación de enfermedades. En la tabla 3-2, se describen los capítulos generales del CIE-10 sobre los cuales se encuentran los datos de diagnóstico de la base de datos.

Tabla 3-2: Descripción de datos de diagnóstico

DIAGNÓSTICOS CIE - 10		
CAPÍTULO	RANGO DE CODIFICACIÓN	DESCRIPCIÓN
5	F01-F99	TRASTORNOS MENTALES, DEL COMPORTAMIENTO Y DEL DESARROLLO NEUROLÓGICO
18	R00-R99	SÍNTOMAS, SIGNOS Y RESULTADOS ANORMALES DE PRUEBAS COMPLEMENTARIAS, NO CLASIFICADOS BAJO OTRO CONCEPTO
19	S00-T88	LESIONES TRAUMÁTICAS, ENVENENAMIENTOS Y OTRAS CONSECUENCIAS DE CAUSAS EXTERNAS
21	Z00-Z99	FACTORES QUE INFLUYEN EN EL ESTADO DE SALUD Y CONTACTO CON LOS SERVICIOS SANITARIOS

Elaboración propia

3.1.3 Descripción datos WISC-IV

La estructura interna de los datos de la prueba de inteligencia WISC IV, se descompone en 8 subpruebas principales (Diseño con cubos, Semejanzas, Retención de dígitos, Conceptos con dibujos, Vocabulario, Sucesión de números y letras, Matrices y Comprensión) y 4 complementarias (Figuras incompletas, Información, Aritmética y Palabras en contexto). Cada ítem o reactivo puede presentar una escala de puntuación diferente, de forma que algunos ítems solo puntúan 0 o 1, mientras que otros pueden llegar a una puntuación de 7 puntos. Esta caracterización de la puntuación fue clave en el proceso de limpieza de la base. En la tabla 2-3, se describe los valores de puntuación de cada ítem:

Tabla 3-3: Descripción de los campos de la prueba WISC-IV

Descripción de los campos de la prueba WISC-IV	
Subprueba	Puntuación (valores enteros)
Diseño con cubos	Reactivos 1-3: 0, 1 o 2 puntos
	Reactivos 4-8: 0 o 4 puntos
	Reactivos 9-14: 0, 4, 5, 6 o 7 puntos
Semejanzas	Reactivos 1-2: 0 o 1 puntos
	Reactivos 3-23: 0, 1 o 2
Retención de dígitos	Puntuación de 0, 1 o 2 para cada reactivo, en 16 reactivos
Conceptos con dibujos	Puntuación de 0 o 1 para cada reactivo, en 28 reactivos
Vocabulario	Reactivos 1-4: 0 o 1 punto
	Reactivos 5-36: 0, 1 o 2 puntos
Sucesión de números y letras	Puntuación de 0, 1, 2 o 3 para cada reactivo, en 10 reactivos
Matrices	Puntuación de 0 o 1 para cada reactivo, en 35 reactivos
Comprensión	Puntuación de 0, 1 o 2 puntos para cada reactivo, en 21 reactivos
Figuras incompletas*	Puntuación de 0 o 1 punto para cada reactivo, en 38 reactivos
Información*	Puntuación de 0 o 1 punto para cada reactivo, en 33 reactivos
Aritmética*	Puntuación de 0 o 1 punto para cada reactivo, en 34 reactivos
Palabras en contexto*	Puntuación de 0 o 1 punto para cada reactivo, en 24 reactivos

Elaboración propia

En cuanto a los nombres de los reactivos en la base de datos, se usaron las primeras letras o acrónimos, junto con un número al final que da la posición del reactivo dentro de cada subprueba. A continuación, se listan los nombres usados para los reactivos usados para esta investigación en cada subprueba:

- *Diseño con cubos*: 'cubos1', 'cub2', 'cub3', 'cub4', 'cub5', 'cub6', 'cub7', 'cub8', 'cub9', 'cub10', 'cub11', 'cub12', 'cub13', 'cub14'.
- *Semejanzas*: 'semejanzas1', 'sem2', 'sem3', 'sem4', 'sem5', 'sem6', 'sem7', 'sem8', 'sem9', 'sem10', 'sem11', 'sem12', 'sem13', 'sem14', 'sem15', 'sem16', 'sem17', 'sem18', 'sem19', 'sem20', 'sem21', 'sem22', 'sem23'.
- *Retención de dígitos*: 'digitos 1 progresivo', 'dig2p', 'dig3p', 'dig4p', 'dig5p', 'dig6p', 'dig7p', 'dig8p', 'digitos 1 inverso', 'dig2i', 'dig3i', 'dig4i', 'dig5i', 'dig6i', 'dig7i', 'dig8i'.
- *Conceptos con dibujos*: 'conceptos con dibujos 1', 'condib2', 'condib3', 'condib4', 'condib5', 'condib6', 'condib7', 'condib8', 'condib9', 'condib10', 'condib11', 'condib12', 'condib13', 'condib14', 'condib15', 'condib16', 'condib17', 'condib18',

'condib19', 'condib20', 'condib21', 'condib22', 'condib23', 'condib24', 'condib25', 'condib26', 'condib27', 'condib28'.

- *Vocabulario*: 'vocabulario1', 'voc2', 'voc3', 'voc4', 'voc5', 'voc6', 'voc7', 'voc8', 'voc9', 'voc10', 'voc11', 'voc12', 'voc13', 'voc14', 'voc15', 'voc16', 'voc17', 'voc18', 'voc19', 'voc20', 'voc21', 'voc22', 'voc23', 'voc24', 'voc25', 'voc26', 'voc27', 'voc28', 'voc29', 'voc30', 'voc31', 'voc32', 'voc33', 'voc34', 'voc35', 'voc36'.
- *Sucesión de números y letras*: 'sucesión de números y letras1', 'suletr2', 'suletr3', 'suletr4', 'suletr5', 'suletr6', 'suletr7', 'suletr8', 'suletr9', 'suletr10'.
- *Matrices*: 'matrices 1', 'matri2', 'matri3', 'matri4', 'matri5', 'matri6', 'matri7', 'matri8', 'matri9', 'matri10', 'matri11', 'matri12', 'matri13', 'matri14', 'matri15', 'matri16', 'matri17', 'matri18', 'matri19', 'matri20', 'matri21', 'matri22', 'matri23', 'matri24', 'matri25', 'matri26', 'matri27', 'matri28', 'matri29', 'matri30', 'matri31', 'matri32', 'matri33', 'matri34', 'matri35'.
- *Comprensión*: 'comprensión 1', 'comp2', 'comp3', 'comp4', 'comp5', 'comp6', 'comp7', 'comp8', 'comp9', 'comp10', 'comp11', 'comp12', 'comp13', 'comp14', 'comp15', 'comp16', 'comp17', 'comp18', 'comp19', 'comp20', 'comp21'.

3.2 Limpieza y adecuación del conjunto de datos

Como se mencionó al inicio de este capítulo, la elaboración de esta base de datos se desarrolló de manera manual, por lo que se encontraron tanto datos faltantes como errores de digitación, los cuales se solventaron con las estrategias descritas a continuación.

3.2.1 Limpieza de variables categóricas

Para esta base de datos, las variables categóricas las conforman los 11 campos de los datos socioeconómicos y los datos de consulta, ya que son campos que describen una condición particular del evaluado y toma un valor definido. Estas variables son: Género, Fecha de nacimiento, Fecha de aplicación, Escolaridad, Tipo de colegio, Escolaridad padre, Escolaridad madre, Estrato, Procedencia, Datos de consulta, Motivo de consulta y Diagnóstico.

En primer lugar, en todos los campos categóricos, se convirtió el tipo de variable a carácter, o 'str' en Python, con el fin de evitar errores al aplicar funciones o métodos de imputación de datos. Una vez convertidos a carácter, se cambiaron a mayúsculas y se eliminaron los espacios intermedios, con el fin de reducir el número de opciones.

En el campo Género, las opciones de valores que quedaron fueron ['2', '1', 'M', 'FEMENINO', 'MASCULINO', 'F', '2.HOMBRE', '1.MUJER'], las cuales se reemplazaron de la siguiente forma:

- '1', 'F', 'FEMENINO' y '1.MUJER'; reemplazados por 'MUJER'
- '2', 'M', 'MASCULINO' y '2.HOMBRE'; reemplazados por 'HOMBRE'

En el campo Escolaridad, los valores resultantes fueron ['1', 'PRIMARIA', 'PREESCOLAR', '2', 'BACHILLERATO', '3', '4', '11', 'NAN', 'ACELERACION', '5', '50M', '0', 'TERCERO', 'SÉPTIMO', 'SEGUNDO', 'PRIMERO', 'QUINTO', 'CUARTO', '-', 'NOESTABLECIDA', '1.PRIMARIA', '1.PRIMARIAYPREESCOLAR', '2.SECUNDARIA'], los cuales se reemplazaron como sigue:

- '0', '1', '3', '4', '5', 'PRIMARIA', 'PREESCOLAR', 'ACELERACION', 'TERCERO', 'SEGUNDO', 'PRIMERO', 'QUINTO', 'CUARTO', '1.PRIMARIA' y '1.PRIMARIAYPREESCOLAR'; reemplazados por "PRIMARIA Y PREESCOLAR"
- '2', '11', 'BACHILLERATO', 'SÉPTIMO' y '2.SECUNDARIA'; reemplazados por 'SECUNDARIA'
- 'NOESTABLECIDA', '50M', '-' y 'NAN'; reemplazados por 'NO ESTABLECIDA'

En el campo Tipo de colegio, los valores resultantes fueron ['2', '1', 'NAN', 'PRIVADO', 'PUBLICO', '0', 'PÚBLICO', 'NE', 'HC2019', '-', 'NOESTABLECIDA', '1.PÚBLICO', '2.PRIVADO']; los cuales se reemplazaron como sigue:

- '1', 'PUBLICO' y '1.PÚBLICO'; reemplazados por 'PÚBLICO'
- '2' y '2.PRIVADO'; reemplazados por 'PRIVADO'

- 'NAN', '0', 'NE', 'HC2019', '-' y 'NOESTABLECIDA'; reemplazados por 'NO ESTABLECIDO'

En los campos Escolaridad padres y Escolaridad Madre, los valores resultantes fueron ['3', '4', '2', 'BACHILLERATO', 'TECNICO', 'PRIMARIA', '6', '0', 'UNIVERSITARIO', '5', 'NA', 'NAN', '1', 'NE', 'NOESTABLECIDA', 'POSGRADO', 'TÉCNICO', '7', 'POSTGRADO', 'NOREGISTRA', 'HC2019', '-', 'SININFORMACIÓN', '...', '3.SECUNDARIA', '4.PROFESIONAL', '7.TECNOLÓGICO', '0.NOSETIENENDATOS']; ['6', '4', '3', 'TECNICO', 'BACHILLERATO', 'UNIVERSITARIO', '2', 'POSTGRADO', 'NAN', '5', '0', 'TÉCNICO', '1', '7', 'NE', 'PRIMARIA', 'POSGRADO', 'NOESTABLECIDA', ' ', 'BACHILLER', 'UNIVERSITARIA', 'TÉCNICA', 'NOREGISTRA', 'HC2019', '-', 'SININFORMACIÓN', 'TECNOLÓGICO', '3.SECUNDARIA', '4.PROFESIONAL', '6.TÉCNICO', '0.NOSETIENENDATOS']; los cuales se reemplazaron como sigue:

- '0', '1', '2', '3', '4', '5', '6' y '7'; reemplazados respectivamente por 'SIN DATOS', 'SIN ESCOLARIDAD', 'PRIMARIA', 'SECUNDARIA', 'PROFESIONAL', 'POSTGRADO', 'TÉCNICO' y 'TECNOLÓGICO'
- ' ', 'NAN', 'NA', 'NE', 'NOESTABLECIDA', 'NOREGISTRA', 'HC2019', '-', 'SININFORMACIÓN', '...', '0.NOSETIENENDATOS', 'NO ESTABLECIDA', 'NO REGISTRA', 'HC 2019', 'SIN INFORMACIÓN' y '0.NO SE TIENEN DATOS'; reemplazados por 'SIN DATOS'
- 'BACHILLERATO', '3.SECUNDARIA' y 'BACHILLER'; reemplazados por 'SECUNDARIA'
- 'TECNICO', 'TÉCNICA' y '6.TÉCNICO'; reemplazados por 'TÉCNICO'
- '7.TECNOLÓGICO'; reemplazado por 'TECNOLÓGICO'
- 'UNIVERSITARIO', 'UNIVERSITARIA' y '4.PROFESIONAL'; reemplazados por 'PROFESIONAL'
- 'POSGRADO'; reemplazado por 'POSTGRADO'

En el caso de estrato, los valores resultantes fueron ['1', '2', '3', '5', '4', 'II', 'I', 'III', 'NAN', 'HC2019', '-']; los cuales se reemplazaron como sigue:

- 'I', 'II' y 'III' reemplazados respectivamente por '1', '2' y '3'
- 'NAN', 'HC 2019', 'HC2019' y '-'; reemplazados por 'SIN DATOS'

Para el campo Procedencia, los valores resultantes fueron ['1', 'BOGOTA', 'BOGOTAD.C', 'CHÍA', 'NAN', 'BOGOTÁ', '11', 'CHAPINERO', 'SOACHA', 'ENGATIVÁ', '2', '6', 'TOCANCIPÁ', '8', '4', '7', 'USME', 'CIUDADBOLIVAR', 'USAQUEN', 'KENNEDY', 'SANJOSÉDELGUAVIARE', 'SUBA', 'MEDELLÍN', 'HC2019', '-', '2.CUNDINAMARCA', '1.BOGOTÁ', '10']; los cuales se reemplazaron como sigue:

- '1', '2', '3', '4', '5', '6', '7', '8', '9', '10' y '11'; reemplazados respectivamente por 'BOGOTÁ', 'CUNDINAMARCA', 'BOYACÁ', 'TOLIMA', 'NARIÑO', 'ANTIOQUIA', 'HUILA', 'SANTANDER', 'VALLE', 'CASANARE' y 'OTRO'
- 'BOGOTAD.C', 'BOGOTA', 'CHAPINERO', 'ENGATIVÁ', 'KENNEDY', 'CIUDADBOLIVAR', 'USME', 'USAQUEN', 'SUBA' y '1.BOGOTÁ'; reemplazados por 'BOGOTÁ'
- 'CHÍA', 'SOACHA', 'TOCANCIPÁ' y '2.CUNDINAMARCA'; reemplazados por 'CUNDINAMARCA'
- 'SANJOSÉDELGUAVIARE'; reemplazado por 'OTRO'
- 'HC2019', '-' y 'NAN'; reemplazados por 'SIN DATOS'

En el campo de Consulta, los valores resultantes fueron ['0', '5', '3', 'NAN', '4', 'EVAL', '6', '2', '11', '1', '9', 'DESEMPEÑO', '10', '2Y3', 'HC2019', '-', '0/4', 'DESEMPEÑOESCOLAR', '3.PROBLEMADEATENCIÓN' y '0.EVALUACIÓNCOGNOSCITIVA']; los cuales se reemplazaron como sigue:

- '0', '1', '2', '3', '4', '5', '6', '7', '8', '9' y '10'; reemplazados respectivamente por 'EVALUACIÓN COGNOSCITIVA', 'HABILIDADES', 'INTERÉS PROPIO',

'PROBLEMA DE ATENCIÓN', 'PROBLEMAS DE APRENDIZAJE', 'BAJO RENDIMIENTO ACADÉMICO', 'VALORACIÓN DESARROLLO', 'ORIENTACIÓN VOCACIONAL SENCILLA', 'REINGRESO Y TRASLADO', 'PROBLEMAS EMOCIONALES' y 'DISCAPACIDAD COGNITIVA'

- 'EVAL' y '0.EVALUACIÓNCOGNOSCITIVA'; reemplazados por 'EVALUACIÓN COGNOSCITIVA'
- 'NAN', '2Y3', 'HC2019', '-' y '0/4'; reemplazados por 'SIN DATOS'
- 'DESEMPEÑO' y 'DESEMPEÑOESCOLAR'; reemplazados por 'BAJO RENDIMIENTO ACADÉMICO'
- '3.PROBLEMADEATENCIÓN'; reemplazado por 'PROBLEMA DE ATENCIÓN'

Por último, frente al campo de Diagnóstico, la ventaja particular es que los códigos CIE-10 estaban descritos en los primeros 4 caracteres de cada dato, por lo cual solo fue necesario segmentar la cadena de caracteres y obtener los 4 primeros.

3.2.2 Limpieza de fechas

La base de datos aporta dos campos de fechas: Fecha de nacimiento y Fecha de aplicación. La primera es del nacimiento del evaluado, y la segunda es la fecha de evaluación del evaluado. Con estos dos datos se obtiene un campo adicional, como es la edad del evaluado al momento de la evaluación. Para este nuevo campo, lo esperado es que la edad coincida con el rango de edad de aplicación del WISC-IV, que es desde los 6 años hasta los 16 años.

El trabajo de limpieza para estos campos se limitó a la adecuación del formato de fecha, de modo que Python pueda reconocerlos como variables de tipo fecha, permitiendo operarlos para hallar los valores de edad de los evaluados.

3.2.3 Dicotomización del diagnóstico y balanceo de clases

Dado que esta investigación está enfocada en la predicción de la discapacidad intelectual, el campo de diagnóstico se convierte en la variable objetivo, de modo que se redujeron el número de clases objetivo a dos, una para positivo y otra para negativo como diagnóstico final discapacidad intelectual. Para tal fin, el diagnóstico fue dicotomizado conforme a la siguiente estrategia, de acuerdo con lo visto en el primer capítulo sobre la codificación CIE-10 de la discapacidad intelectual:

- Si los primeros dos caracteres del valor de diagnóstico corresponden a 'F7', se recodifica con la clase 1 (positivo para discapacidad intelectual).
- En cualquier otro caso, se recodifica con la clase 0 (negativo para discapacidad intelectual).

Una vez dicotomizado el diagnóstico, se usó la librería 'resample' en Python para efectuar el balanceo de clases, de modo que se tomaron los casos positivos para discapacidad intelectual y se replicaron aleatoriamente, hasta igualar el número de casos negativos. De este modo, la muestra total fue aumentada a 804 casos, distribuidos equitativamente en 402 casos negativos y 402 casos positivos.

3.2.4 Limpieza de datos WISC-IV

Para esta sección de datos, dado que pertenecen a la prueba de inteligencia y son puntuaciones numéricas, se optó, en primera instancia, por la siguiente secuencia de pasos como estrategia de limpieza, teniendo siempre en cuenta las reglas de puntuación para cada subprueba, descritas en el manual de aplicación del WISC-IV, y la supervisión de los miembros del laboratorio de Psicometría:

1. Reemplazar todo valor que no sea numérico (letras o palabras) por un valor tipo NaN (Not a Number).
2. Rellenar los valores perdidos sobre el eje de las columnas, en sentido derecha-izquierda, con el primer dato válido.

3. Rellenar los datos perdidos que aún se mantengan con una puntuación de 0.
4. Convertir el tipo de variable a entero

Una vez culminada esta primera estrategia, se toman las puntuaciones descritas para cada subprueba (ver Tabla 3-3), y con ellas, se buscan en la base aquellos registros que tengan puntuaciones por fuera de los valores permitidos y se ajustan conforme a la regla. Este proceso se siguió de manera independiente para cada subprueba, de modo que no hubiera errores debido a la generalización de la estrategia. A continuación, se desglosa las reglas usadas para la limpieza de datos en cada subprueba:

- **Diseño con cubos**

1. Reactivos 1-3: Si la puntuación es mayor a 2, se imputa el valor de 2
2. Reactivos 4-8: Si la puntuación es diferente de 0 y 4, se imputa el valor de 4
3. Reactivos 9-14: Si la puntuación es mayor a 0 y menor a 4, se imputa el valor de 4
4. Reactivos 9-14: Si la puntuación es mayor a 7, se imputa el valor de 7

- **Semejanzas**

1. Reactivos 1-2: Si la puntuación es mayor a 1, se imputa el valor de 1
2. Reactivos 3-23: Si la puntuación es mayor a 2, se imputa el valor de 2

- **Retención de dígitos**

1. Todos los reactivos: Si la puntuación es mayor a 2, se imputa el valor de 2

- **Conceptos con dibujos**

1. Reactivos 1-28: Si la puntuación es mayor a 1, se imputa el valor de 1

- **Vocabulario**

1. Reactivos 1-4: Si la puntuación es mayor a 1, se imputa el valor de 1

2. Reactivos 5-36: Si la puntuación es mayor a 2, se imputa el valor de 2

- **Sucesión de números y letras**

1. Reactivos 1-10: Si la puntuación es mayor a 3, se imputa el valor de 3

- **Matrices**

1. Reactivos 1-35: Si la puntuación es mayor a 1, se imputa el valor de 1

- **Comprensión**

1. Reactivos 1-21: Si la puntuación es mayor a 2, se imputa el valor de 2

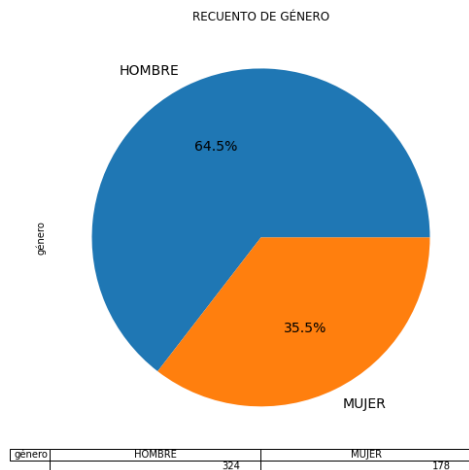
Por último, dado que las subpruebas complementarias (Figuras incompletas, Información, Aritmética y Palabras en contexto) tienen un porcentaje de datos perdidos superior al 50%, se decidió eliminarlas de la base de datos. Por tanto, el número de variables de la base de datos se redujo a 183.

3.3 Análisis exploratorio

En esta sección, se realizaron algunos análisis descriptivos para develar las distribuciones, tendencias o patrones en las variables categóricas.

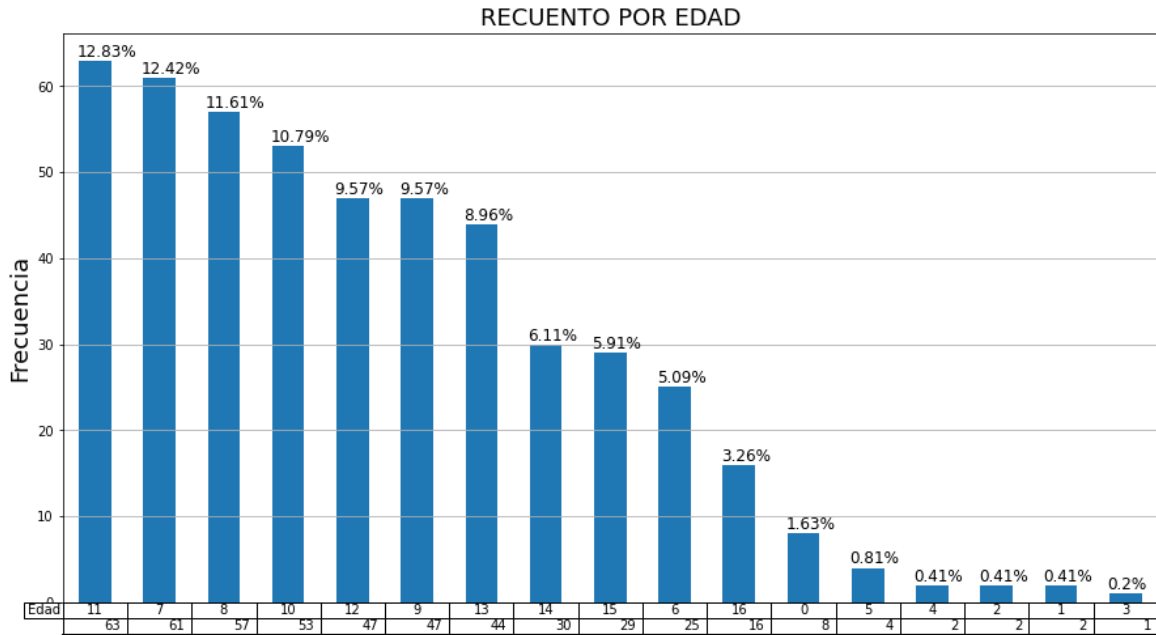
En la distribución de la variable Género, en la población de la base de datos (ver Figura 3-1), se encontraron 324 hombres y 178 mujeres, lo que corresponde al 64.5% y 35.5% respectivamente, dando como mayoría a la categoría de los hombres.

Figura 3-1: Recuento de hombres y mujeres para la variable Género



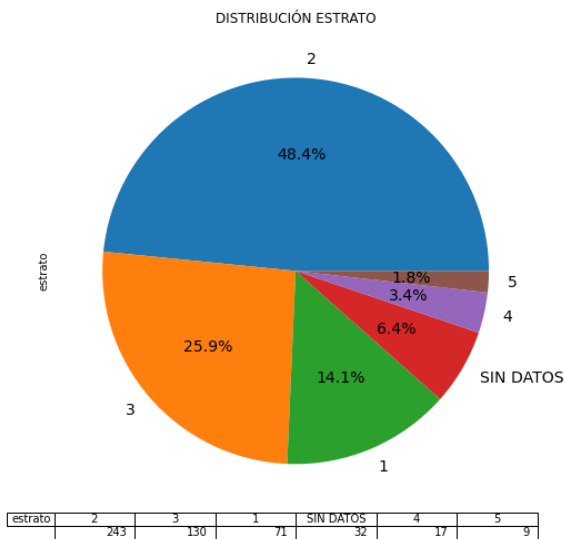
La edad de los evaluados fue generada como un campo numérico a partir de los campos de fechas de nacimiento y fecha de aplicación. El WISC-IV está diseñado para evaluar niños desde los 6 años hasta los 16 años con 11 meses de edad, pero la distribución muestra registros de niños menores de 6 años, así como campos para los cuales no fue posible el cálculo de la edad, lo que se debe posiblemente a errores de digitación o del formato de fecha. El conteo de estos datos atípicos es de 19 sobre 502 registros, lo que representa un 3,8% de la muestra. La mayoría de los evaluados corresponde a una edad de 11 años (ver Figura 3-2).

Figura 3-2: Distribución de la edad de los evaluados



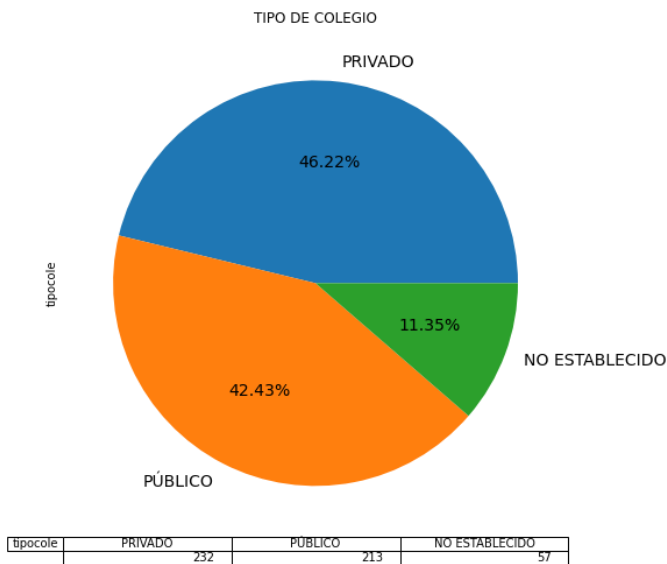
La distribución por estratos (ver figura 3-3) da como mayoría al estrato 2, con casi la mitad de la muestra, seguidos por los estratos 3 y 1 que juntos suman un 40%, y una mínima representación de los estratos 4 y 5 que juntos suman 5,2%. El restante 6,4% pertenece a datos perdidos.

Figura 3-3: Distribución de estrato de los evaluados



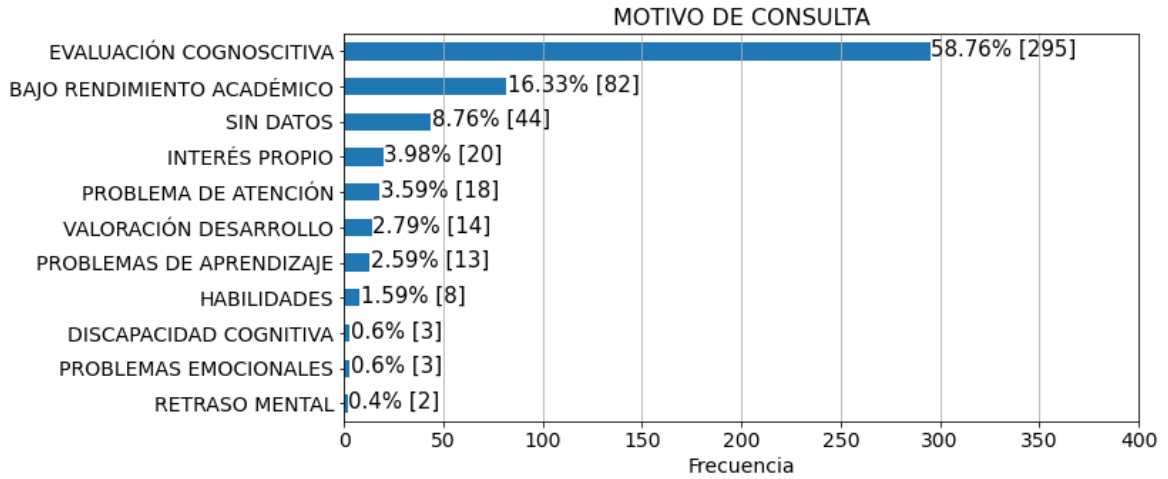
En el tipo de colegio de los evaluados (ver Figura 3-4), se distribuyen de forma casi equitativa, con un 46,2% de los evaluados en colegio privado y 42,4% en colegio público. El restante 11,3% pertenece a aquellos para quienes no se pudo establecer el tipo de colegio o datos perdidos.

Figura 3-4: Distribución de Tipo de colegio de los evaluados



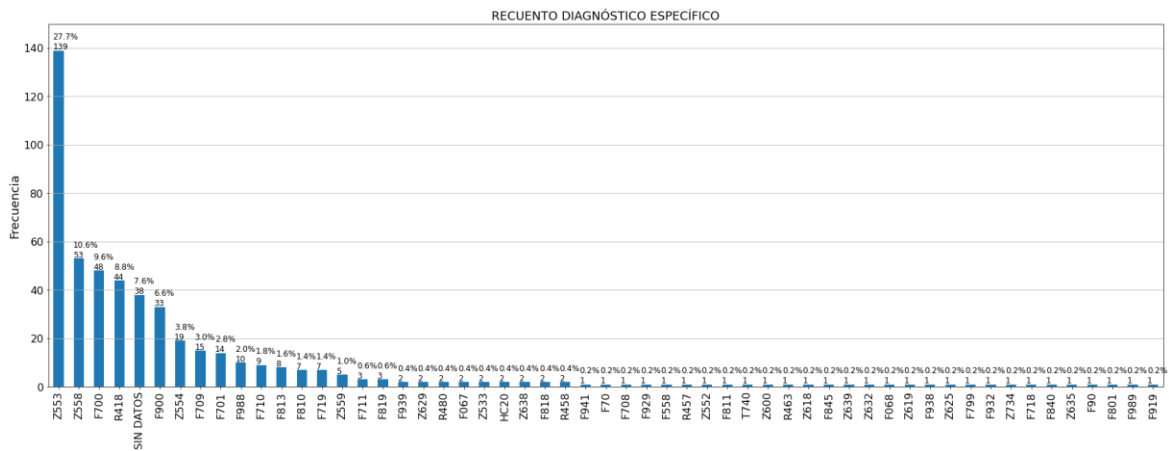
En cuanto al motivo de consulta (ver Figura 3-5), el motivo por el que más asisten a evaluación psicológica, cerca del 60%, es por Evaluación cognoscitiva, seguido de un 16,3% que van por bajo rendimiento académico. Los datos perdidos representan un 8,76% de la muestra. Los demás motivos de consulta tienen una baja representación.

Figura 3-5: Distribución de Motivo de consulta de los evaluados



En lo referente al diagnóstico, que está codificado según la Clasificación Internacional de Enfermedades (CIE-10), los diagnósticos más frecuentes están relacionados con problemas de bajo rendimiento escolar (Z553-8) sumando un 38,3%; seguidos en tercer lugar por Retraso Mental Leve (F700) con 9,6%. Un 7,6% de la muestra pertenece a datos perdidos; es decir, 38 casos de 502 en total.

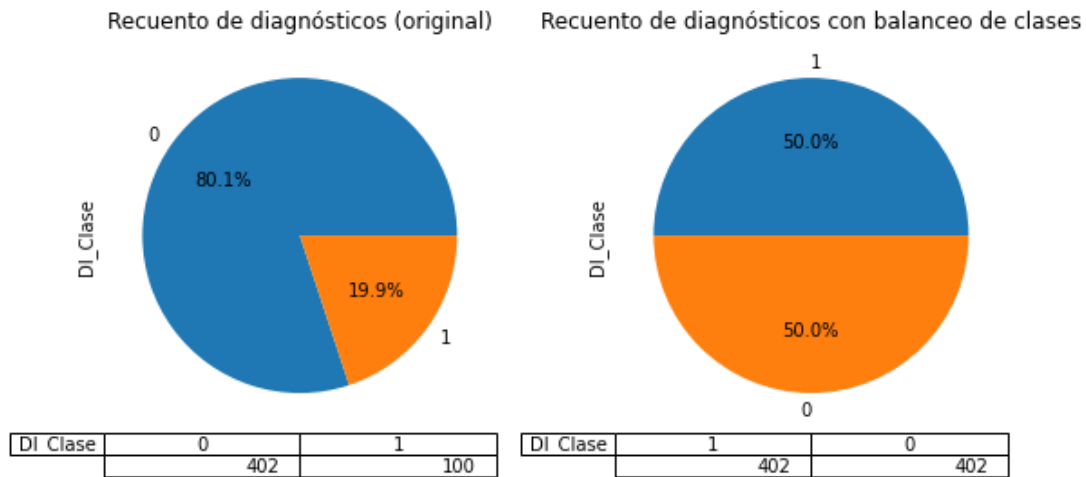
Figura 3-6: Distribución del diagnóstico



Por último, una vez realizada la dicotomización del diagnóstico, se puede ver en la Figura 3-7 (lado izquierdo) que los casos positivos representan un 20%; mientras que los negativos conforman el 80%, habiendo una diferencia de proporción de 4 a 1 entre las

clases objetivo. Una vez realizado el balanceo de clases (ver Figura 3-6, lado derecho), los diagnósticos se equilibraron en proporción, quedando un 50% de la muestra para cada clase.

Figura 3-7: Distribución de diagnósticos con discapacidad intelectual



En cuanto a los puntajes de la prueba WISC-IV, ya que son 183 ítems, la forma de resumirlos y explorar su comportamiento conjunto es mediante las puntuaciones naturales, las cuales se hallan a partir de la sumatoria de todos los reactivos en cada subprueba. La figura 3-8 muestra un resumen descriptivo de las 8 subpruebas que conforman la base de datos.

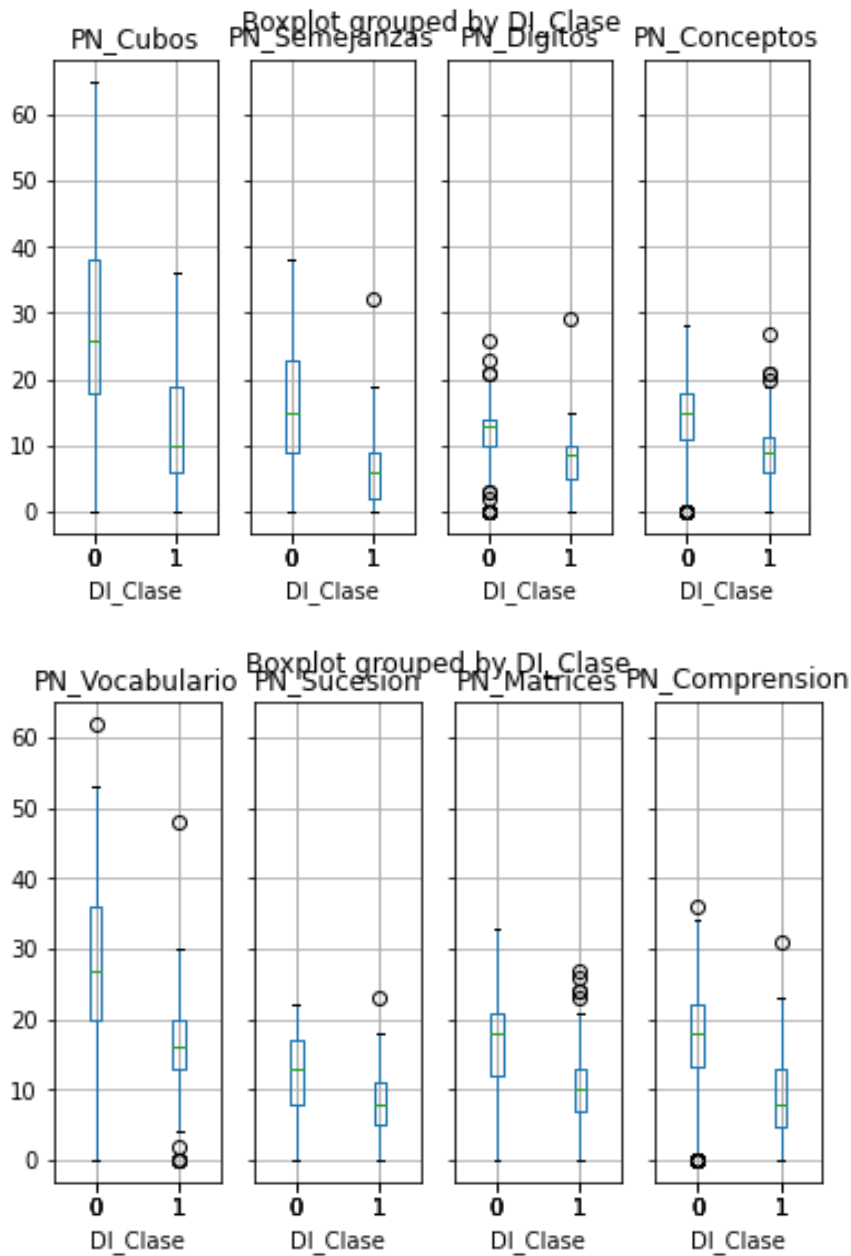
Figura 3-8: Descriptivos subpruebas WISC-IV

	PN_Cubos	PN_Semejanzas	PN_Digitos	PN_Conceptos	PN_Vocabulario	PN_Sucesion	PN_Matrices	PN_Comprension
count	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000	502.000000
mean	25.262948	13.687251	11.362550	13.356574	25.348606	11.017928	15.717131	15.687251
std	14.347087	9.149447	3.993512	5.477804	11.311591	5.808803	6.637737	7.861818
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	14.000000	6.000000	9.000000	10.000000	18.000000	7.000000	11.000000	10.000000
50%	25.000000	13.000000	12.000000	14.000000	24.000000	12.000000	16.000000	16.000000
75%	35.000000	21.000000	14.000000	17.000000	34.000000	16.000000	21.000000	21.000000
max	65.000000	38.000000	29.000000	28.000000	62.000000	23.000000	33.000000	36.000000

Estos descriptivos muestran que la subprueba con la puntuación máxima más alta es Diseño con cubos, con 65 puntos; mientras que la puntuación máxima más baja es la de Sucesión de letras y números, con 23 puntos. El mismo comportamiento se puede ver en las puntuaciones medias. En todas las subpruebas, hubo personas evaluadas que obtuvieron puntuación mínima de 0 puntos. Cabe anotar que todas las subpruebas tienen una puntuación máxima diferente, por lo que no son equiparables entre sí sus puntuaciones naturales.

Al agrupar los datos de las subpruebas por la variable objetivo (diagnóstico positivo o negativo de DI), y hacer diagramas de caja y bigotes, se puede ver que en todas las 8 subpruebas existe una diferencia en el comportamiento, mostrando puntuaciones más bajas en evaluados con discapacidad intelectual (clase 1), al comparar la mediana y los rangos intercuartílicos en cada grupo.

Figura 3-9: Boxplot de subpruebas WISC-IV, agrupados por diagnóstico de DI



Esto es un buen indicio inicial, ya que las variables independientes (los reactivos de las subpruebas), muestran un potencial global latente para clasificar el diagnóstico de discapacidad intelectual, mediante el ajuste de un modelo predictivo, con aquellas variables independientes que más aporten información.

4. Selección de características y generación del modelo predictivo

En este capítulo se describirán los métodos usados para la selección de características, sobre el conjunto de datos que abarca todos los reactivos de la prueba de inteligencia WISC-IV, que aportan más información a la predicción de Discapacidad intelectual. Con este conjunto de datos reducido, se implementó el modelo predictivo para detectar futuros casos de Discapacidad intelectual. La implementación se elaboró en lenguaje Python, sobre el entorno de Google Colaboratory.

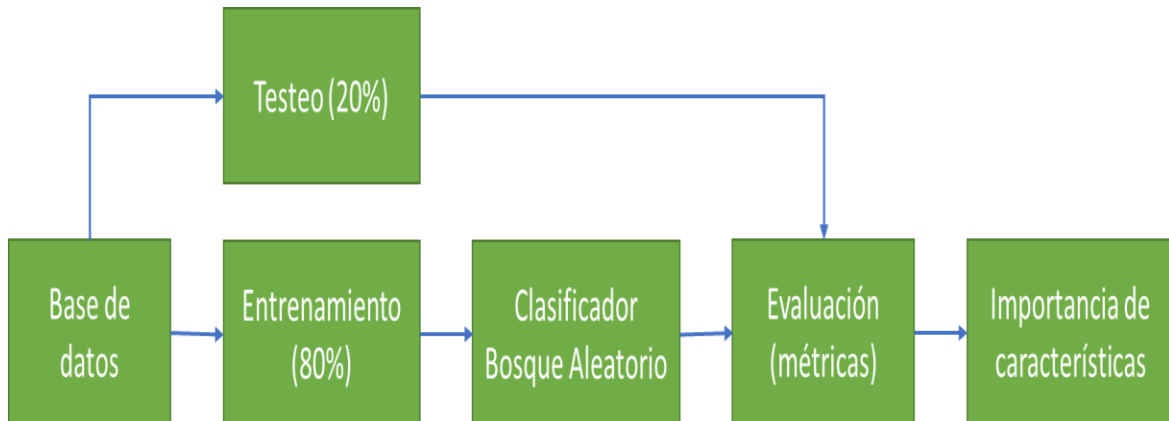
4.1 Selección de características, primera aproximación: Importancia de características de Bosque aleatorio

Para esta primera aproximación de selección de características, se optó por usar el clasificador Bosque Aleatorio (RandomForestClassifier), de la librería sklearn en Python, que es un método de aprendizaje supervisado no paramétrico, basado en árboles de decisión, lo cual lo hace útil para este proyecto dada la naturaleza discreta de todas las variables, así como la poca preparación de datos que requiere, al no ser necesaria la normalización. Otra ventaja de este modelo es que es relativamente fácil de implementar, al basarse en reglas sencillas de decisión.

Para usar este método, se usaron todas las variables del WISC-IV que se extrajeron después de la limpieza de datos. La variable objetivo, como se mencionó en el capítulo anterior, es la variable de diagnóstico dicotomizada para discapacidad intelectual, con la clase 1 para positivo y clase 0 para negativo. A continuación, basado en la Figura 4-1, se describirá el método seguido para implementar esta primera aproximación.

Inicialmente, se realiza la segmentación de los datos entre entrenamiento y testeo, se decide dejar arbitrariamente 80% para entrenamiento y 20% para testeo, ya que es una proporción de uso común en estos casos, teniendo en cuenta que se cuentan con pocos registros.

Figura 4-1: Método de selección de características por Bosque Aleatorio



Como otros clasificadores, Bosque Aleatorio requiere de dos parámetros: una matriz de entrenamiento (X_{train}), de tamaño $[n_{muestras}, n_{características}]$; y un vector de entrenamiento (Y_{train}), de tamaño $[n_{muestras}]$, que contiene los valores objetivo (etiquetas de clase), que en este caso es el diagnóstico binario. Con estos parámetros, se procede a crear y entrenar el modelo de Bosque Aleatorio, empleando el clasificador `RandomForestClassifier`.

Una vez entrenado el clasificador, se usa el segmento de testeo (X_{test}) para calcular un vector de valores predichos ($Y_{predict}$), los cuales se van a contrastar con los valores del vector de testeo (Y_{test}), por medio de las métricas de rendimiento halladas a partir de la matriz de confusión.

Como se muestra en la Figura 4-2, la matriz de confusión muestra: 79 verdaderos negativos, 3 falsos positivos, 2 falsos negativos y 77 verdaderos positivos. A partir de estos

valores se calculan valores como precisión, recall y f1-score, que permiten un análisis más detallado del rendimiento del modelo.

Figura 4-2: Salida métricas de rendimiento para Bosque aleatorio

```
=== Confusion Matrix ===
[[79  3]
 [ 2 77]]

=== Classification Report ===
              precision    recall  f1-score   support

     0       0.98      0.96      0.97         82
     1       0.96      0.97      0.97         79

 accuracy          0.97
 macro avg         0.97
 weighted avg      0.97

=== All AUC Scores ===
[0.99634146 0.99146341 0.99695122 0.98841463 0.9803125 0.9921875
 0.996875 0.991875 0.999375 0.98625 ]

=== Mean AUC Score ===
Mean AUC Score - Random Forest: 0.9920045731707317

=== Accuracy Score ===
Accuracy Score - Random Forest: 0.968944099378882
```

Para obtener el valor de la precisión, dividimos el número total de casos positivos clasificados correctamente por el número total de casos positivos previstos. Alta precisión indica que un caso etiquetado como positivo es efectivamente positivo (pequeño número de Falsos Positivos), el cual para este modelo es de 0.98 para un diagnóstico negativo (0) de DI y 0.96 para un diagnóstico positivo (1) de DI.

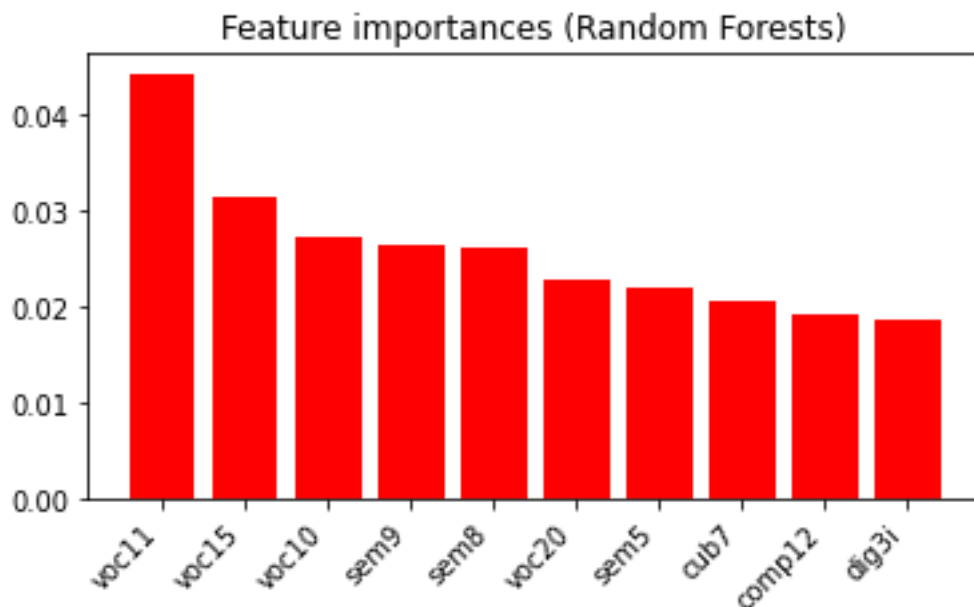
Recall puede definirse como la relación entre el número total de casos positivos clasificados correctamente y el número total de casos positivos. Un recall alto indica que la clase se reconoce correctamente (número pequeño de Falsos Negativos), el cual para este modelo es de 0.96 para un diagnóstico negativo (0) de DI y 0.97 para un diagnóstico positivo (1) de DI.

Dado que tenemos dos medidas (Precisión y Recall), es útil tener una medida que represente a ambas. Esta medida se calcula con F1, que usa la media armónica en lugar de la media aritmética, ya que castiga más los valores extremos. F1 siempre estará más cerca del valor más pequeño de Precisión o Recall. Para este modelo es de 0.97 para un diagnóstico negativo (0) de DI y 0.97 para un diagnóstico positivo (1) de DI.

AUC (área bajo la curva para curvas ROC) proporciona una medida agregada del rendimiento en todos los umbrales de clasificación posibles, y mide la capacidad del modelo para distinguir entre clases. Una forma de interpretar el AUC es la probabilidad de que el modelo clasifique un caso positivo aleatorio más que un caso negativo aleatorio. AUC varía de 0 a 1, de modo que un modelo cuyas predicciones son 100% incorrectas tiene un AUC de 0.0; uno cuyas predicciones son correctas al 100% tiene un AUC de 1.0. Para este modelo, se presenta un valor medio de AUC del 0.99.

El paso final consiste en hallar la importancia de cada característica, o el nivel relativo de información que aporta cada uno de los 182 ítems del WISC-IV, con respecto a la predicción del diagnóstico, dispuestos en el modelo de clasificación RandomForestClassifier, lo cual se puede ver en la Figura 4-3, para las 10 características con mayor valor de importancia:

Figura 4-3: Top 10 de características más importantes para RandomForestClassifier



Para este modelo, la suma de las importancias de las 182 características es 1, por lo que la característica con mayor importancia apenas supera el umbral de 0.04. A partir de ordenar las características con base en su nivel de importancia, se puede tomar deliberadamente las 10 primeras como muestra la Figura 4-3, y podrían seleccionarse más o menos a conveniencia de la investigación.

4.1.1 Limitación de la primera aproximación

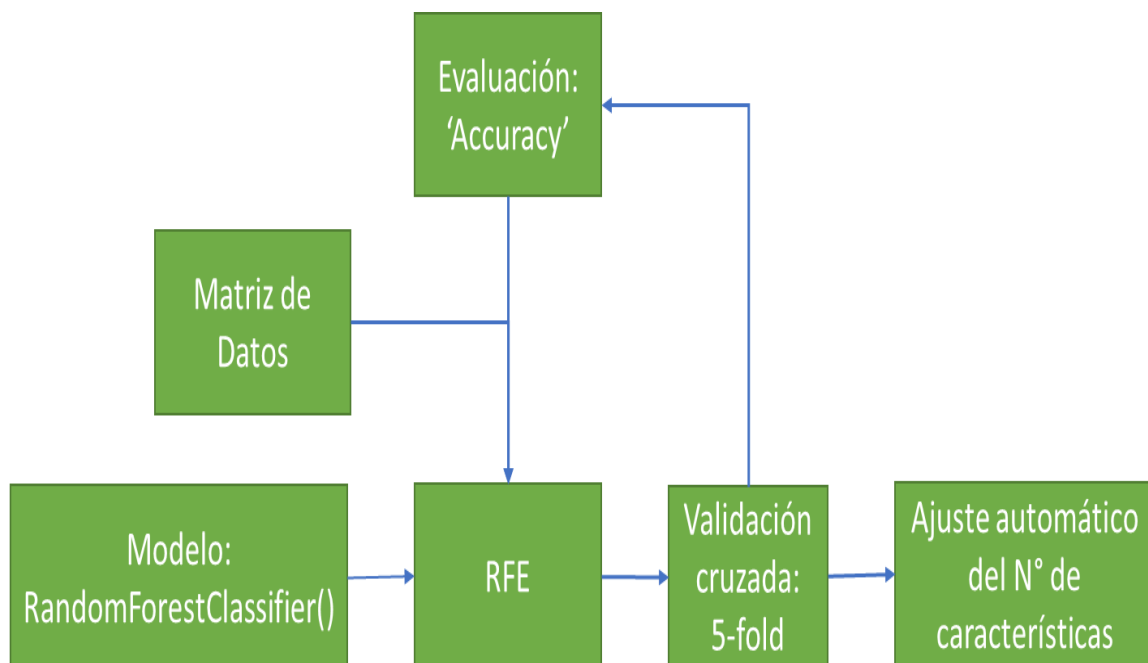
En primera instancia, cabe resaltar que el modelo de Bosque aleatorio funciona muy bien con respecto a sus métricas, lo cual permite tener un alto nivel de confianza en la selección de las características que más aportan información al diagnóstico. Sin embargo, usado por sí solo como método de selección de características, no permite distinguir un criterio o un umbral para decidir en número de características que más aportan información.

4.2 Selección de características, segunda aproximación: Eliminación recursiva de características

Para esta segunda aproximación, se optó por usar la eliminación recursiva de características (Recursive feature elimination, RFE, en inglés), el cual usa el modelo de Bosque aleatorio para asignar los pesos a las características, con el fin de seleccionarlas recursivamente, considerando en cada iteración un grupo más pequeño, eliminando las que menos importancia tienen, hasta encontrar el número deseado de características.

Existe, además, dentro de la librería sklearn, una versión especial de RFE que implementa un bucle de validación cruzada, el cual permite encontrar el número óptimo de características. Esta versión, denominada RFECV, fue especialmente útil para el desarrollo de la investigación, dado que supera la limitación de la primera aproximación. En la figura 4-4 se describe el proceso de RFECV en su implementación.

Figura 4-4: Método de selección de características por RFECV

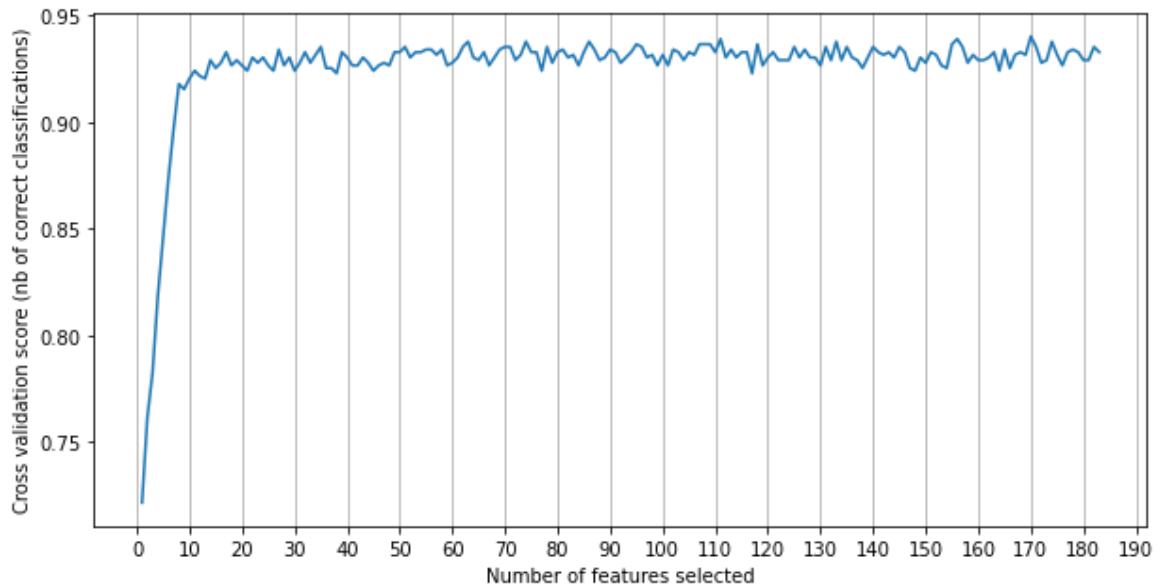


Como se mencionó en la primera aproximación, el bosque aleatorio tuvo muy buenos resultados en cuanto a sus métricas de predicción; en consecuencia, se adoptó como modelo de asignación de pesos de las características para esta estrategia (ver Figura 4-4), usando el atributo de 'feature importances' para asignar dichos pesos.

El esquema de validación cruzada, en el método RFECV, usa por defecto un k-fold con $k=5$, lo que deja también una distribución del 80% para datos de entrenamiento y 20% para datos de testeo en cada iteración. La métrica usada para evaluar el clasificador (Bosque aleatorio), en cada subconjunto de características remanente fue "accuracy", dado que las clases están balanceadas, lo que permite esta métrica como medida global del rendimiento.

El método RFECV elimina recursivamente una característica a la vez, en cada paso, empezando desde en total de características hasta que quede una sola, como muestra la Figura 4-5. Esto permite evaluar, por un lado, la característica que menos aporta información y eliminarla del siguiente subconjunto de datos y, por otro lado, el rendimiento del modelo con cada subconjunto de datos. Por consiguiente, este método evalúa tantos subconjuntos de datos como número de características exista; es decir, como para esta investigación se tienen 183 características, el método evalúa 183 subconjuntos de datos. De este modo, se puede seleccionar aquel subconjunto que obtenga el mejor rendimiento del modelo, dada la métrica escogida.

Figura 4-5: Ajuste automático de selección de características con RFECV



Como se observa en la Figura 4-5, con un número mayor a 10 características, el rendimiento del modelo supera el 90% de clasificaciones correctas. Esto significa que con 10 características o más, la eficacia del diagnóstico no varía significativamente y, por tanto, con un porcentaje mínimo de características se puede obtener el mismo poder de diagnóstico que con el conjunto completo.

Sin embargo, este método tiene una debilidad, no permite seleccionar un número específico de características, sino que selecciona el número de características conforme encuentra el mejor rendimiento posible del modelo. Por tanto, el número de características seleccionadas puede cambiar cada vez que se ejecute el algoritmo.

En este sentido, se implementa un segundo paso de esta aproximación, en la cual se usa el método RFE sin validación cruzada, el cual permite, dentro de sus parámetros, ingresar el número de características a seleccionar. Dado que con más de 10 características se obtienen un buen rendimiento, y el total de características son 183, se decidió seleccionar un 10% del conjunto original; es decir, 18 características finales, las cuales se listan a continuación:

- Reactivo 9 de la subprueba Diseño con cubos
- Reactivos 5, 7, 8 y 9 de la subprueba Semejanzas
- Reactivo 3 directo y reactivo 3 inverso de la subprueba Retención de dígitos
- Reactivos 10, 11, 12, 13, 15 y 20 de la subprueba Vocabulario
- Reactivos 3 y 4 de la subprueba Sucesión de letras y números
- Reactivo 21 de la subprueba Matrices
- Reactivos 10 y 12 de la subprueba Comprensión

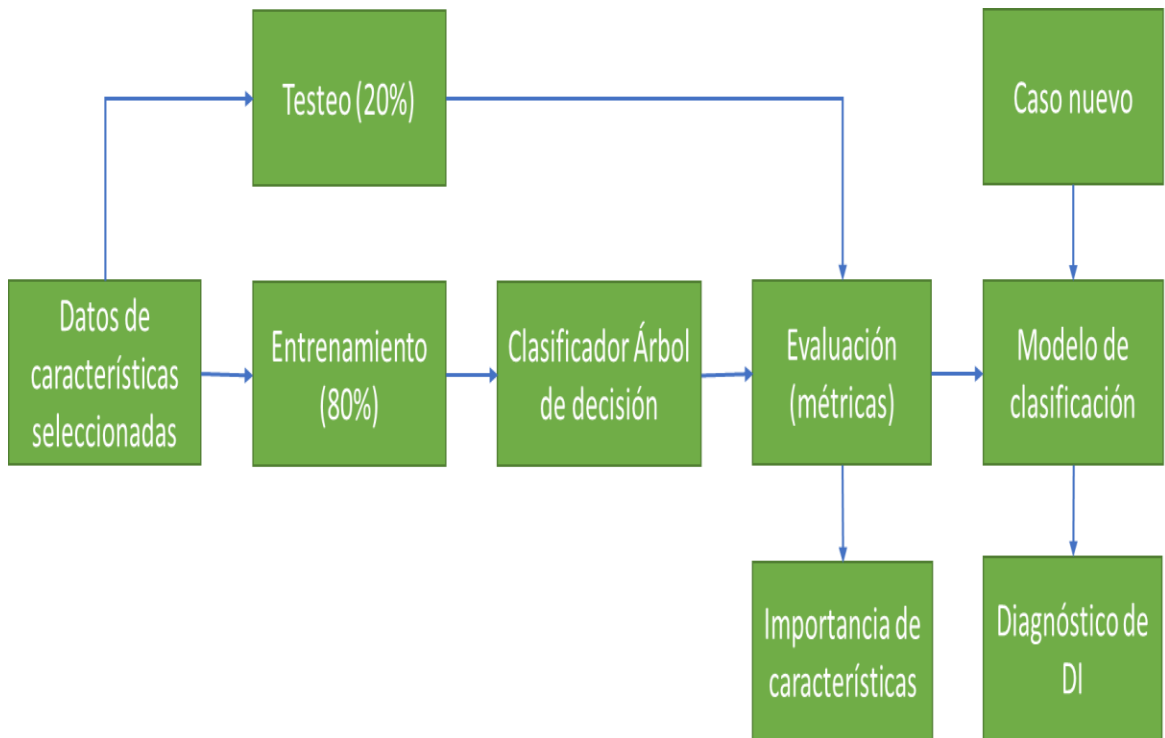
4.3 Generación de modelo predictivo

Para la generación del modelo predictivo, con las 18 características seleccionadas, se optó por implementar dos opciones y contrastarlas entre sí: Árbol de decisión y Regresión logística. Estos dos modelos tienen dos ventajas en particular, su fácil implementación y su nivel de interpretabilidad, lo que permite analizar mejor los resultados y generar nuevo conocimiento.

4.3.1 Implementación del modelo de Árbol de decisión

Retomando lo mencionado en el marco teórico, el árbol de decisión es una técnica de aprendizaje supervisado no paramétrico, usado en esta investigación como clasificador, el cual aprende a predecir el diagnóstico de discapacidad intelectual a partir de reglas de decisión simples inferidas de las 18 características cognitivas seleccionadas de la prueba WISC IV.

Como se observa en la Figura 4-6, la implementación del modelo de clasificación con Árbol de decisión sigue el mismo esquema de implementación del Bosque aleatorio para selección de características, con la diferencia de que el producto es el modelo final de clasificación, para predecir casos futuros.

Figura 4-6: Esquema de implementación del Árbol de decisión

Por consiguiente, se mantienen los mismos criterios frente a la proporción de datos de entrenamiento y de prueba, que son 80/20 respectivamente, y el clasificador empleado de la librería sklearn es DecisionTreeClassifier. Conforme el resultado de la selección de características, el número de características que ingresa son las 18 seleccionadas, y el número de registros son 502.

Un parámetro que se ajustó de este clasificador fue la máxima profundidad (`max_depth`), el cual, si se deja por defecto, los nodos se expandirán hasta que todas las hojas sean puras; sin embargo, esto haría el árbol mucho más grande y difícil de visualizar. Adicionalmente, se hicieron pruebas variando la profundidad y evaluando las métricas, de modo que, con una profundidad de 7 niveles, se obtiene un buen equilibrio entre rendimiento y complejidad.

En cuanto a las métricas de rendimiento para el Árbol de decisión, como se muestra en la Figura 4-7, la matriz de confusión muestra: 66 verdaderos negativos, 10 falsos positivos, 2 falsos negativos y 83 verdaderos positivos, a partir de los cuales se calcularon las métricas para su evaluación.

Figura 4-7: Salida métricas de rendimiento para Árbol de decisión

```
=== Confusion Matrix ===
[[66 10]
 [ 2 83]]

=== Classification Report ===
              precision    recall  f1-score   support

     0       0.97         0.87         0.92         76
     1       0.89         0.98         0.93         85

 accuracy          0.93         0.93         0.93         161
 macro avg         0.93         0.92         0.92         161
 weighted avg      0.93         0.93         0.93         161

=== All AUC Scores ===
[0.90335366 0.89969512 0.94603659 0.90792683 0.8103125 0.9546875
 0.9515625 0.84125 0.956875 0.93375 ]

=== Mean AUC Score ===
Mean AUC Score - Tree Decision: 0.9105449695121951

=== Accuracy Score ===
Accuracy Score - Tree Decision: 0.9254658385093167
```

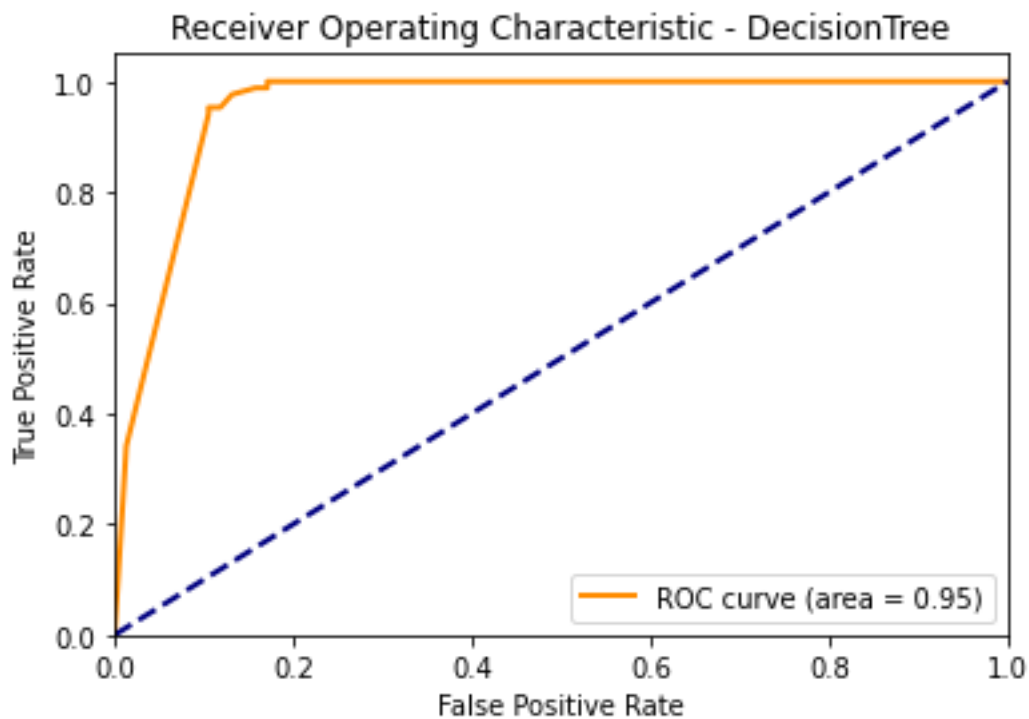
El valor de la precisión (ver Figura 4-7), para este modelo es de 0.97 para un diagnóstico negativo (0) de DI y 0.89 para un diagnóstico positivo (1) de DI, siendo más preciso el modelo para los casos negativos.

El valor de Recall (ver Figura 4-7), para este modelo es de 0.87 para un diagnóstico negativo (0) de DI y 0.98 para un diagnóstico positivo (1) de DI, lo cual indica que el modelo es mejor para reconocer la clase positiva correctamente

La medida de F1 (ver Figura 4-7), que es una media armónica entre Precisión y Recall, para este modelo es de 0.92 para un diagnóstico negativo (0) de DI, y 0.93 para un diagnóstico positivo (1) de DI, siendo un indicador alto para ambas clases. Por último, la medida de AUC para este modelo, tiene un valor medio de 0.91.

Otra métrica que se puede usar para medir la calidad del clasificador es la curva ROC, la cual, como se observa en la Figura 4-8, en el eje Y expresa la tasa de verdaderos positivos (sensibilidad), y la tasa de falsos positivos (1-especificidad) en el eje X.

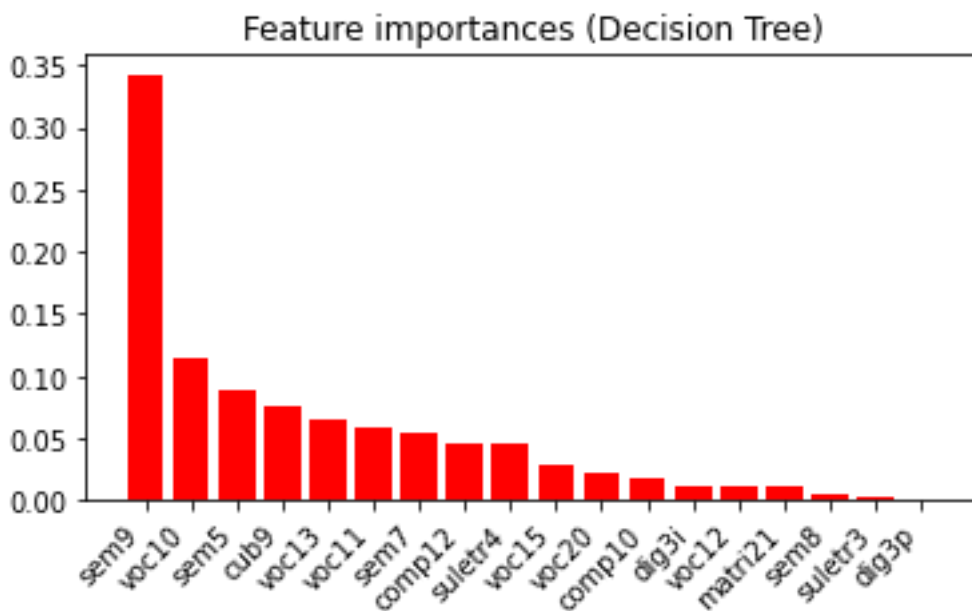
Figura 4-8: Curva ROC para el Árbol de decisión



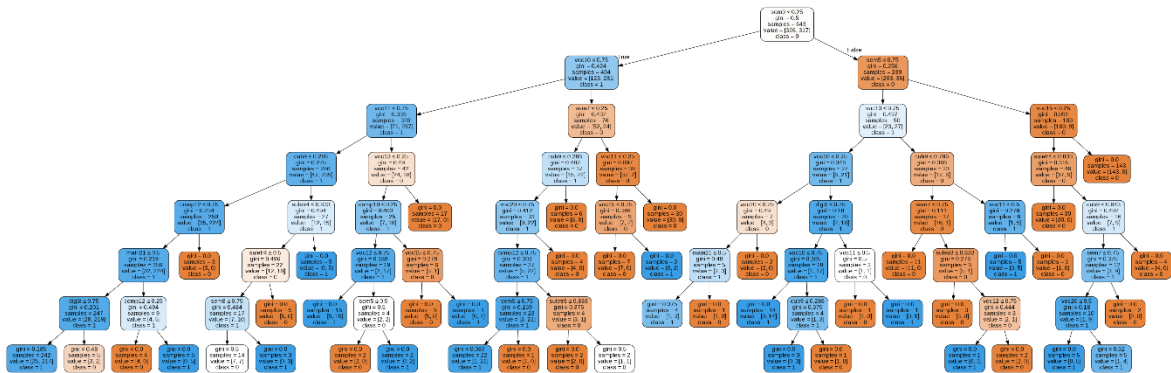
Para la curva ROC, lo esperado es que su punto de mayor inflexión se aproxime tanto como sea posible al punto superior izquierdo de la gráfica, de modo que la tasa de verdaderos positivos esté cercana a 1, y la tasa de falsos positivos tenga un valor cercano a cero. Cerca de este punto esperado, el área bajo la curva también se aproxima a 1. Para este clasificador de Árbol de decisión, el área bajo la curva es de 0.95.

El árbol de decisión, al igual que el bosque aleatorio, tiene la capacidad de medir la importancia o el nivel de información que aporta cada característica al momento de realizar la clasificación de clases. Como se observa en la Figura 4-9, las características están ordenadas de mayor a menor, de manera que la primera característica (reactivo 9 de la subprueba Semejanzas), es la más “pura” o la que mejor discrimina las clases objetivo; y va disminuyendo la importancia de cada característica hasta la última (reactivo 3 de la subprueba Retención de dígitos), la cual aporta la menor información dentro del modelo.

Figura 4-9: Niveles de importancia de características para el Árbol de decisión



Finalmente, la visualización del árbol de decisión (ver Figura 4-10) muestra en detalle el conjunto de reglas simples que conforman el modelo de clasificación final. La forma correcta de seguir el algoritmo para evaluar un caso nuevo es de arriba hacia abajo, de modo que la primera decisión se toma en la raíz, y se va bifurcando conforme a la decisión que tome en cada nodo, hasta llegar a alguna de las hojas en la parte inferior, donde declara si el diagnóstico es positivo (clase 1) o negativo (clase 0).

Figura 4-10: Modelo de clasificación final del Árbol de decisión

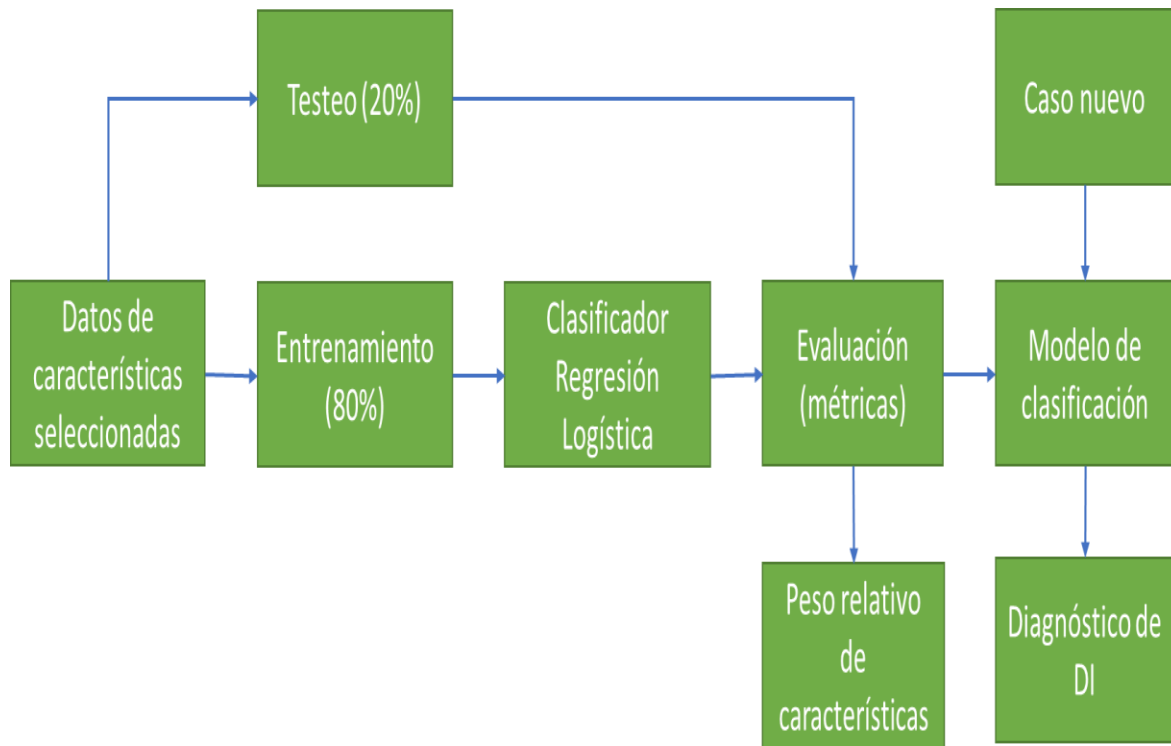
Si se compara la gráfica del árbol con respecto a los niveles de importancia de las características (ver Figuras 4-9 y 4-10), se puede observar que las características más importantes se ubican cerca a la raíz del árbol (parte superior), mientras que las menos importantes se ubican cerca de las hojas (parte inferior).

4.3.2 Implementación del modelo de Regresión logística

Para esta implementación, la regresión logística tiene un sentido particular, y es evaluar diferencialmente las características que aportan información a cada una de las clases, a saber, la clase positiva y la clase negativa del diagnóstico de discapacidad intelectual; así como medir su poder predictivo en contraste con el modelo de árbol de decisión.

Como se observa en la Figura 4-11, la implementación del clasificador por regresión logística sigue el mismo esquema que el clasificador por árbol de decisión. Por tanto, se usan los mismos datos de entrenamiento y testeo, proporción 80/20 respectivamente. Así como los productos serán el modelo de clasificación final y los pesos de las características para cada una de las clases.

Figura 4-11: Esquema de implementación del modelo de Regresión logística



Con respecto al modelo de clasificador implementado, se empleó LogisticRegression de la librería sklearn. Para favorecer la interpretabilidad del modelo, no se incluyó ningún tipo de regularización en el modelo, con el parámetro `penalty='none'`, de modo que no se restringe la estimación de los coeficientes del modelo, pero se corre el riesgo de sobreajuste del modelo a los datos de entrenamiento (overfitting). Una vez entrenado el modelo, se observan las métricas de rendimiento en la Figura 4-12.

Figura 4-12: Salida métricas de rendimiento para Regresión logística

```
=== Confusion Matrix ===
[[62 14]
 [19 66]]

=== Classification Report ===
              precision    recall  f1-score   support

     0       0.77       0.82       0.79         76
     1       0.82       0.78       0.80         85

 accuracy          0.80         0.80         161
 macro avg         0.80         0.80         161
 weighted avg      0.80         0.80         161

=== All AUC Scores ===
[0.85243902 0.88506098 0.91829268 0.81768293 0.828125  0.8453125
 0.93      0.8325      0.945      0.86625  ]

=== Mean AUC Score ===
Mean AUC Score - Logistic Regression: 0.8720663109756099

=== Accuracy Score ===
Accuracy Score - Logistic Regression: 0.7950310559006211
```

El valor de la precisión (ver Figura 4-12), para este modelo es de 0.77 para un diagnóstico negativo (0) de DI y 0.82 para un diagnóstico positivo (1) de DI, siendo más preciso el modelo para los casos negativos.

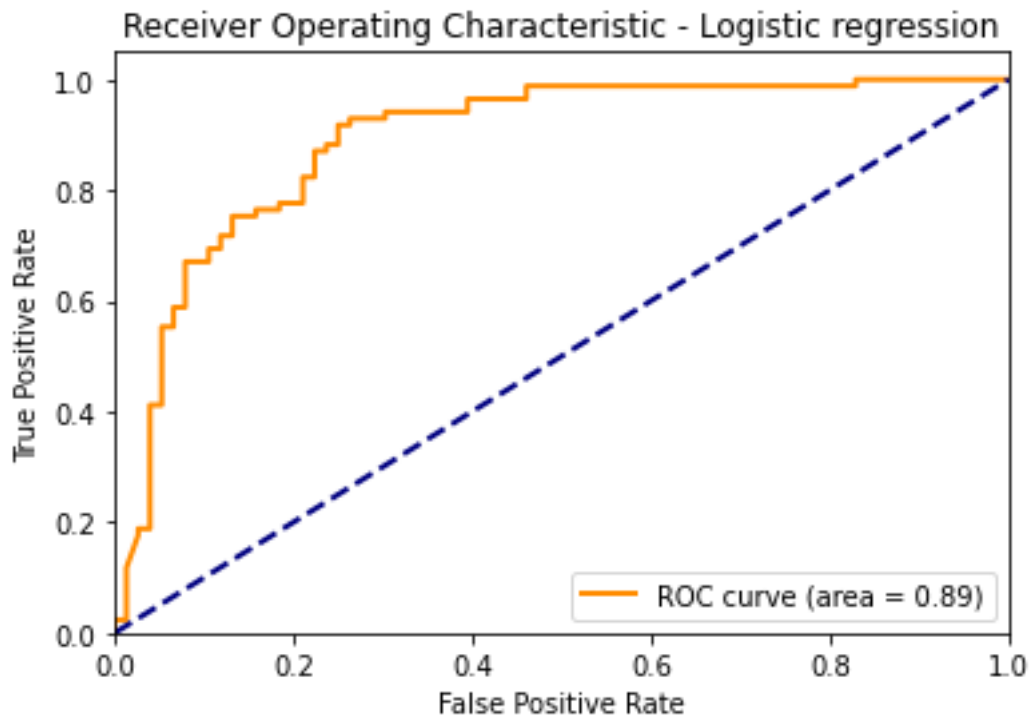
El valor de Recall (ver Figura 4-12), para este modelo es de 0.82 para un diagnóstico negativo (0) de DI y 0.78 para un diagnóstico positivo (1) de DI, lo cual indica que el modelo es mejor para reconocer la clase negativa correctamente

La medida de F1 (ver Figura 4-12), que es una media armónica entre Precisión y Recall, para este modelo es de 0.79 para un diagnóstico negativo (0) de DI, y 0.80 para un diagnóstico positivo (1) de DI, siendo un indicador bueno para ambas clases. Por último, la medida de AUC para este modelo, tiene un valor medio de 0.79.

En cuanto a la curva ROC, lo esperado es que su punto de mayor inflexión se aproxime tanto como sea posible al punto superior izquierdo de la gráfica, de modo que la tasa de

verdaderos positivos esté cercana a 1, y la tasa de falsos positivos tenga un valor cercano a cero. Cerca de este punto esperado, el área bajo la curva también se aproxima a 1. Para este clasificador de Regresión logística (ver Figura 4-13), la forma de la curva es irregular y el área bajo la curva es de 0.89.

Figura 4-13: Curva ROC para el modelo de Regresión logística



Como resultado del modelo, se obtuvieron los coeficientes mostrados en la Tabla 4-1, incluyendo el intercepto del modelo logístico.

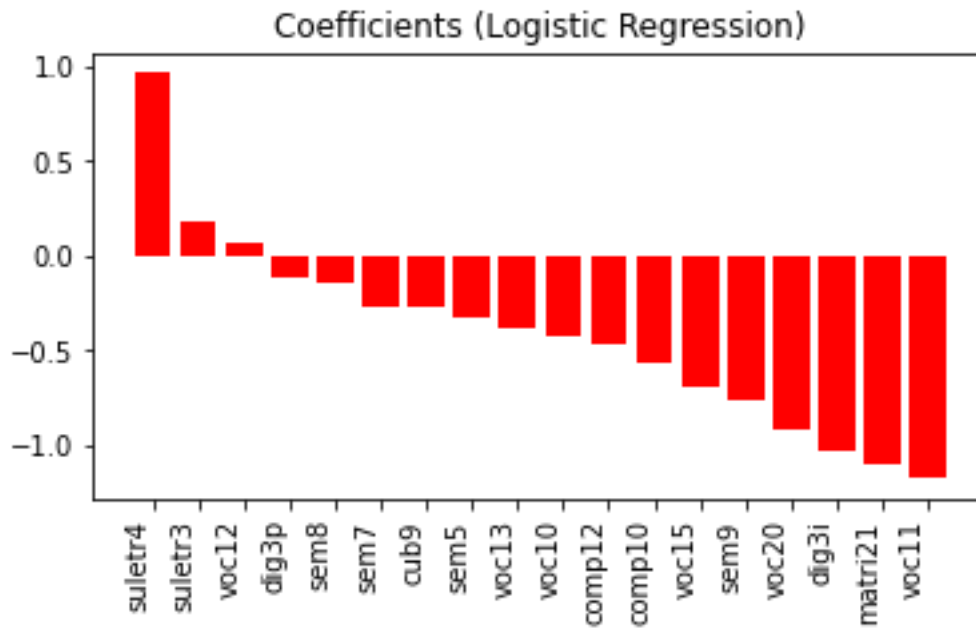
Tabla 4-1: Intercepto y coeficientes del modelo de Regresión logística

Característica / Intercepto	Coefficientes / Pesos relativos
Intercepto	2.41592762
Reactivo 9 de Diseño con cubos (cub9)	-0.27440250250745724
Reactivo 5 de Semejanzas (sem5)	-0.3236656379060563

Reactivo 7 de Semejanzas (sem7)	-0.26718347358360967
Reactivo 8 de Semejanzas (sem8)	-0.1410824623000159
Reactivo 9 de Semejanzas (sem9)	-0.7712324454297979
Reactivo 3 de Retención de dígitos directo (dig3p)	-0.11217498929538951
Reactivo 3 de Retención de dígitos inverso (dig3i)	-1.0397908574551231
Reactivo 10 de Vocabulario (voc10)	-0.43187488696776966
Reactivo 11 de Vocabulario (voc11)	-1.176340074959952
Reactivo 12 de Vocabulario (voc12)	0.06331911157951593
Reactivo 13 de Vocabulario (voc13)	-0.3834347091541495
Reactivo 15 de Vocabulario (voc15)	-0.6926474973228878
Reactivo 20 de Vocabulario (voc20)	-0.9257661974533463
Reactivo 3 de Sucesión de letras y números (suletr3)	0.17859609607511082
Reactivo 4 de Sucesión de letras y números (suletr4)	0.9667285399335227
Reactivo 21 de Matrices (matri21)	-1.1085540147272608
Reactivo 10 de Comprensión (comp10)	-0.5629575501580905
Reactivo 12 de Comprensión (comp12)	-0.47409637175293623

Estos valores de coeficientes, al igual que en el modelo de Árbol de decisión, se comportan como pesos o importancia de las características; sin embargo, se diferencian en que tienen valores positivos o negativos, lo cual indica si las características aportan información a la predicción del diagnóstico positivo o negativo, respectivamente.

Figura 4-14: Coeficientes del modelo de Regresión logística



En este sentido, la Figura 4-14 muestra tanto la magnitud de los coeficientes como el sentido (Positivo/Negativo). Por tanto, se tienen tres características que aportan información al diagnóstico positivo, mientras que las restantes 15 aportan información al diagnóstico negativo de discapacidad intelectual.

5. Resultados y discusión

Como resultado de todo el proceso de investigación, y cumpliendo con el objetivo principal de hallar la agrupación de variables asociadas al Coeficiente Intelectual (CI) que predican el diagnóstico de Discapacidad Intelectual (CIE 10 F70-79), se obtuvieron 18 reactivos agrupados por las subpruebas a las que pertenecen:

- Diseño con cubos: Reactivo 9
- Semejanzas: Reactivos 5, 7, 8 y 9
- Retención de dígitos: Reactivo 3 directo y reactivo 3 inverso
- Vocabulario: Reactivos 10, 11, 12, 13, 15 y 20
- Sucesión de letras y números: Reactivos 3 y 4
- Matrices: Reactivo 21
- Comprensión: Reactivos 10 y 12

Si se reagrupan estos 18 reactivos resultantes a partir de los índices compuestos de la prueba WISC-IV, se tienen las siguientes proporciones:

- Comprensión Verbal (ICV): Semejanzas, Vocabulario y Comprensión = 12 reactivos (66.6% del total de 18 reactivos)
- Razonamiento Perceptual (IRP): Diseño con cubos y Matrices = 2 reactivos (11.1% del total de 18 reactivos)
- Memoria de Trabajo (IMT): Retención de dígitos y Sucesión de letras y números = 4 reactivos (22.2% del total de 18 reactivos)

Tomando las anteriores proporciones como ponderaciones del aporte de información de cada uno de los índices del Coeficiente intelectual, en primer lugar, el mayor aporte a la evaluación de Discapacidad intelectual lo da el índice de Comprensión verbal (ICV), con

un 66.6% de los reactivos, mostrando que el razonamiento verbal, la formación de conceptos, el bagaje de conocimientos, la comprensión verbal y, en general, el desarrollo de las capacidades de comprensión y expresión verbal para manejar información práctica [12], son más relevantes en la evaluación de la discapacidad intelectual.

En segundo lugar, el índice de Memoria de trabajo (IMT) es el siguiente en el aporte de información al diagnóstico, con un 22.2% de los reactivos, mostrando que la capacidad de formar y seguir secuencias, la concentración y atención, la flexibilidad cognitiva y, en general, la capacidad de procesamiento de información mentalmente y la memoria a corto plazo [12] coadyuvan en la evaluación de la discapacidad intelectual.

En último lugar, con un 11.1% de los reactivos, el índice de Razonamiento perceptual (IRP) también aporta al diagnóstico, a partir de medir la formación de conceptos no verbales, la organización y percepción visual y, en general, el procesamiento de información visual e inteligencia fluida [12].

Usando el mismo procedimiento, al agrupar los 18 reactivos ahora por los factores CHC, se tienen las siguientes proporciones:

- Procesamiento visual (Gv): Diseño con cubos y Matrices = 2 reactivos (11.1% del total de 18 reactivos)
- Razonamiento fluido (Gf): Matrices = 1 reactivo (5.5% del total de 18 reactivos)
- Memoria a corto plazo (Gsm): Retención de dígitos y Sucesión de letras y números = 4 reactivos (22.2% del total de 18 reactivos)
- Inteligencia cristalizada (Gc): Semejanzas, Vocabulario y Comprensión = 12 reactivos (66.6% del total de 18 reactivos)

De igual manera, tomando las proporciones como ponderaciones del aporte de información de cada uno de estos factores CHC, se obtiene, en primer lugar, con 66.6% de los reactivos, el mayor aporte al diagnóstico por parte de la Inteligencia cristalizada (Gc), la cual refiere al almacenamiento de conocimiento verbal o basado en el lenguaje, que se

adquiere a través de las experiencias de aprendizaje [12] y, en general, se relaciona con la comprensión del lenguaje y el conocimiento general [7].

El segundo lugar en la ponderación de aporte de información, con 22.2% de los reactivos, se tiene la Memoria a corto plazo (Gsm), la cual es un sistema limitado en capacidad, capaz de retener 7 segmentos de información (más o menos 2) en la mayoría de los individuos, y esta información se retiene por unos segundos antes de desaparecer [12]. Este sistema se relaciona con la aprehensión y mantenimiento consciente de información, que es útil en resolución de problemas en varios pasos [7].

En tercer lugar, con 11.1% de los reactivos, se tiene al Procesamiento visual (Gv), el cual refiere a habilidades de percepción, discriminación y manipulación de siluetas y formas visuales, así como al uso de imaginación visual para resolver problemas [7], [12].

En último lugar, con mínima proporción, se tiene el Razonamiento fluido (Gf), que refiere al control deliberado y flexible de la atención a fin de resolver problemas novedosos y, contrario a la inteligencia cristalizada, no requiere conocimiento previo para resolver el problema [7]. En el caso de la subprueba Matrices, se hace uso de razonamiento secuencial general o razonamiento deductivo [12].

En cuanto a los resultados del modelo de Árbol de decisión, respecto a la importancia relativa de las 18 características seleccionadas, muestra que el reactivo 9 de semejanzas tiene la mayor importancia relativa (ver figura 4-9), muy por encima de las demás características, por lo cual se ubica en la raíz de árbol y vale la pena revisar el contenido de dicho reactivo. En el reactivo 9 de Semejanzas, se le presentan al evaluado dos conceptos comunes y tiene que describir en qué son similares, para lo cual se le lee la siguiente pregunta: ¿En qué se parecen el ENOJO y la ALEGRÍA? A lo cual tiene que responder mediante el habla [11]

Según la respuesta al reactivo 9 de Semejanzas, se le da al evaluado entre 0 y 2 puntos; dando 2 puntos si responde que son emociones o respuestas emocionales, sentimientos,

estados de ánimo o estados emocionales; se da 1 punto si responde que son expresiones, formas de sentirse, respuestas, reacciones, actitudes o expresiones faciales; y 0 puntos si responde acciones, comportamientos, señas o rostros [11].

Como se mencionó anteriormente, la subprueba de Semejanzas hace parte del índice de Comprensión verbal y de Inteligencia cristalizada, los grupos con mayor número de reactivos. Por tanto, requiere que el evaluado escuche y comprenda la instrucción verbal que se le está dando; así como usar el conocimiento previo para asociar conceptos que son abstractos, hallar las similitudes y expresar una respuesta verbal con otro concepto más general, que dé cuenta de dichas similitudes (emociones o estados de ánimo). Así mismo, cuanto más concreta sea la respuesta o esté basada en la descripción de comportamiento, menor será la puntuación. Cabe anotar que este reactivo también requiere la comprensión de emociones o sentimientos, lo cual podría agregar una dificultad adicional que le permita al reactivo discriminar mejor entre personas con y sin discapacidad intelectual.

Por otro lado, el modelo de Regresión logística da cuenta de 3 reactivos que aportan información al diagnóstico positivo de discapacidad intelectual, los cuales son el reactivo 12 de la subprueba Vocabulario, y los reactivos 3 y 4 de la subprueba Sucesión de letras y números, siendo este último el que mayor peso tiene, muy por encima de los otros dos (ver Figura 4-12), por lo que también vale la pena revisar su contenido. En el reactivo 4 de Sucesión de letras y números, se lee al evaluado una combinación de letras y números, y se le pide que recuerde y reproduzca verbalmente los números en orden ascendente y las letras en orden alfabético [11]. En este caso, el reactivo tiene tres ensayos, que se listan a continuación:

- Ensayo 1: D-2-9, a lo cual debe responder 2-9-D o D-2-9
- Ensayo 2: R-5-B, a lo cual debe responder 5-B-R o B-R-5 (este ensayo tiene una indicación adicional en caso de respuesta errada)
- Ensayo 3: H-9-K, a lo cual debe responder 9-H-K o H-K-9

A diferencia de los reactivos anteriores 1, 2 y 3 de la subprueba, los cuales solo comprenden una letra y un número de 1 dígito (reactivos 1 y 2), o una letra y dos números de 1 dígito (reactivo 3); el reactivo 4 tiene dos letras y un número de un dígito (en los ensayos 2 y 3); lo cual, en principio, marcaría la diferencia para que este reactivo sea más informativo para el diagnóstico positivo.

Como se mencionó anteriormente, la subprueba de Sucesión de letras y números se encuentra dentro del índice de Memoria de trabajo y la Memoria a corto plazo. Por tanto, mide la formación de secuencias, el manejo de información mentalmente, memoria auditiva a corto plazo, atención y concentración [12]. En este sentido, dado que la subprueba está diseñada para aumentar la dificultad conforme avanza cada reactivo, la diferencia en el reactivo 4 radica en el manejo del alfabeto como una secuencia organizada, dado que el evaluado tiene que recordar el orden de 27 letras (desde A hasta Z) y comparar la posición relativa de dos letras en este orden, a la vez que recuerda un número; mientras que en el reactivo 3, compara la posición relativa de dos números con respecto a la secuencia ordenada de 10 dígitos (desde 0 hasta 9), mientras recuerda una letra. Esta diferencia de dificultad en el reactivo 4 es lo que le permitiría aportar información al diagnóstico positivo.

6. Conclusiones, recomendaciones y trabajo futuro

En este trabajo se desarrolló un modelo computacional que permite seleccionar las características cognitivas que más aportan información al diagnóstico de discapacidad intelectual, mediante la implementación de eliminación recursiva de características (RFE, por sus siglas en inglés). También se implementó un árbol de decisión y una regresión logística como modelos predictivos e interpretables, los cuales permiten predecir casos futuros de discapacidad intelectual usando datos de funcionamiento cognitivo. Los resultados muestran un total de 18 características cognitivas más informativas; como también que el modelo de árbol de decisión presenta mejores métricas en la predicción de casos futuros de discapacidad intelectual.

En cuanto al aporte de los modelos de predicción desarrollados, tanto el Árbol de decisión como la Regresión logística, son modelos susceptibles de ser implementados por parte del Servicio de Atención Psicológica (SAP) de la Universidad Nacional, como pruebas de tamizaje para la detección de discapacidad intelectual, ante la sospecha por parte del psicólogo evaluador al inicio de un proceso de evaluación cognoscitivo-emocional. Su implementación como prueba de tamizaje sería importante dado que, en contraste, la aplicación completa del WISC-IV consume bastante tiempo y requiere personal calificado, lo cual implica mayores costos económicos. También aportaría, a nivel educativo, en la rápida detección de niños con dificultades de aprendizaje, para quienes se podría enfocar una evaluación especial.

Así mismo, una vez aplicado el WISC-IV, y a pesar de que las puntuaciones no sean interpretables o se ubiquen en el borde entre inteligencia límite y discapacidad intelectual, se podrá aplicar el modelo de predicción como herramienta que ayude a enfocar efectivamente los recursos de evaluación, al descartar o detectar desde el inicio la

condición de discapacidad. Cabe resaltar que se recomendará el uso del modelo de Árbol de decisión sobre el de Regresión logística, dada la superioridad mostrada por el Árbol de decisión en las métricas de rendimiento (ver capítulo 4). Se podrían aplicar ambos modelos en un mismo caso particular, en la medida en que el psicólogo evaluador requiera una mayor seguridad para descartar un posible caso de discapacidad intelectual.

Una etapa en el desarrollo de esta tesis que demandó gran parte de esfuerzo y tiempo fue la de preprocesamiento de datos, dada la falta de estandarización en los valores de las características de la base de datos, y los posibles errores de digitación, por lo que se recomienda a futuro, en la creación de la base de datos, unificar o restringir el conjunto de valores que puede tener cada característica, de manera que cada dato copie fielmente la información de las historias clínicas.

Para futuras investigaciones, se recomienda trabajar con datos que abarquen medidas tanto de coeficiente intelectual como de comportamiento adaptativo, de modo que se cubra gran parte del dominio de evaluación de la discapacidad intelectual. También se recomienda trabajar con un conjunto de datos más grande, de tal forma que no sean necesaria la implementación de técnicas de aumento de datos o balanceo de clases.

En este sentido, se recomienda la aplicación de este modelo de evaluación para descartar posibles casos de discapacidad intelectual, dadas las limitaciones en el insumo de datos que, por un lado, no abarcan el componente de comportamiento adaptativo, y por otro, tiene una proporción pequeña de casos con diagnósticos positivos de discapacidad intelectual.

Otra recomendación para trabajos futuros es la aplicación del modelo computacional en muestras de poblaciones diferentes, dado que la muestra del SAP es una población particular de niños y adolescentes que van a consulta, en muchos casos, motivados por dificultades o malestar psicológico que afecta su rendimiento académico o su vida cotidiana. Por tanto, no se puede generalizar los resultados de este modelo computacional en su aplicabilidad a la población general.

Dado que la evaluación psicológica cobija un gran número de diagnósticos diferentes, relacionados con salud mental, como lo mostró la descripción de la variable de diagnósticos en esta investigación, una posible evolución de este trabajo sería un modelo de predicción multiclase o multi-diagnóstico, que se entrene con datos de variadas pruebas psicométricas, las cuales evalúan diversos constructos psicológicos o comportamentales relacionados con dichos diagnósticos.

También, como trabajo futuro, se podrían implementar otras técnicas, tanto para selección de características como para modelos de predicción, de modo que se puedan contrastar tanto en sus métricas como en sus bondades interpretativas. Así mismo, se podría generar un modelo predictivo basado en la combinación o paralelización de distintos algoritmos de aprendizaje automático.

7. Bibliografía

- [1] X. Ke and J. Liu, *Discapacidad intelectual*. Ginebra: Asociación Internacional de Psiquiatría del Niño y el Adolescente y Profesiones Afines: IACAPAP, 2017.
- [2] American Association on Intellectual and Developmental Disabilities (AAIDD), "Definition of Intellectual Disability." <https://aaid.org/intellectual-disability/definition> (accessed May 08, 2019).
- [3] O. P. de la salud OPS and O. M. de la S. OMS, "CIE-10 clasificación estadística internacional de enfermedades y problemas relacionados con la salud," *554*, vol. 3, no. 554, p. 758, 2008, Accessed: Feb. 01, 2021. [Online]. Available: <http://ais.paho.org/classifications/Chapters/pdf/Volume3.pdf>.
- [4] A. Srivastava, "A Vector Measure for the Intelligence of a Question-Answering (Q-A) System," *IEEE Trans. Syst. Man. Cybern.*, vol. 25, no. 5, pp. 814–823, 1995, doi: 10.1109/21.376494.
- [5] J. Hernández-Orallo and D. L. Dowe, "Measuring universal intelligence: Towards an anytime intelligence test," *Artif. Intell.*, vol. 174, pp. 1508–1539, 2010, doi: 10.1016/j.artint.2010.09.006.
- [6] C. L. Reeve, C. Scherbaum, and H. Goldstein, "Manifestations of intelligence: Expanding the measurement space to reconsider specific cognitive abilities," *Hum. Resour. Manag. Rev.*, vol. 25, pp. 28–37, 2015, doi: 10.1016/j.hrmr.2014.09.005.
- [7] W. J. Schneider and D. A. Newman, "Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities," 2015, doi: 10.1016/j.hrmr.2014.09.004.
- [8] J. Hernández-Orallo, D. L. Dowe, and M. Victoria Hernández-Lloreda, "Universal psychometrics: Measuring cognitive abilities in the machine kingdom," 2014, doi: 10.1016/j.cogsys.2013.06.001.
- [9] S. R. Vrana and D. T. Vrana, "Can a computer administer a Wechsler intelligence test?," *Prof. Psychol. Res. Pract.*, vol. 48, no. 3, pp. 191–198, 2017, doi: 10.1037/pro0000128.
- [10] J. Hernández-Orallo, F. Martínez-Plumed, U. Schmid, M. Siebers, and D. L. Dowe, "Computer models solving intelligence test problems: Progress and implications," *Artif. Intell.*, vol. 230, pp. 74–107, 2016, doi: 10.1016/j.artint.2015.09.011.
- [11] D. Wechsler, *WISC-IV: escala Wechsler de inteligencia para niños-IV: manual técnico*. México: Manual Moderno, 2005.
- [12] D. Flanagan and A. Kaufman, *Claves para la evaluación con WISC-IV*, 2nd ed. El Manual Moderno, 2012.
- [13] J. Stanton, *VERSION 3: AN INTRODUCTION TO DATA SCIENCE*, Third. iTunes Open Source eBook, 2012.
- [14] L. Cao, "Data science: A comprehensive overview," *ACM Comput. Surv.*, vol. 50, no. 43, 2017, doi: 10.1145/3076253.
- [15] M. Herman *et al.*, *The field guide to data science*. Booz Allen Hamilton, 2013.
- [16] R. A. Pazminño-Maji, F. J. García-Peñalvo, and M. A. Conde-González, *Statistical implicative analysis*

- approximation to KDD and Data Mining: A systematic and mapping review in Knowledge Discovery Database framework*, no. c. 2017.
- [17] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases) (© AAAI)," Mar. 1996. doi: 10.1609/AIMAG.V17I3.1230.
- [18] "1.13. Feature selection — scikit-learn 0.24.1 documentation." https://scikit-learn.org/stable/modules/feature_selection.html#rfe (accessed Feb. 24, 2021).
- [19] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemom. Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, Sep. 2006, doi: 10.1016/j.chemolab.2006.01.007.
- [20] S. R. Das, *Data Science: Theories, Models, Algorithms and Analytics, a web book*. Das, Sanjiv Ranjan, 2013.
- [21] S. Kumar and H. Sharma, "A Survey on Decision Tree Algorithms of Classification in Data Mining," 2016. Accessed: Jan. 27, 2021. [Online]. Available: www.ijsr.net.
- [22] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*," *Artic. Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020, doi: 10.14569/IJACSA.2020.0110277.
- [23] "1.10. Decision Trees — scikit-learn 0.24.1 documentation." <https://scikit-learn.org/stable/modules/tree.html> (accessed Jan. 27, 2021).
- [24] K. Fawagreh, M. Medhat Gaber, E. Elyan, and M. M. Gaber, "Random forests: from early developments to recent advancements," *Syst. Sci. Control Eng. An Open Access J.*, vol. 2, no. 1, pp. 602–609, 2014, doi: 10.1080/21642583.2014.956265.
- [25] M. Kohl, "Performance Measures in Binary Classification," *Int. J. Stat. Med. Res.*, vol. 1, no. 1, pp. 79–81, 2012, doi: 10.1016/j.ipm.2009.03.002.
- [26] A. M. McIntosh *et al.*, "Data science for mental health: a UK perspective on a global challenge," *The Lancet Psychiatry*, vol. 3, no. 10, pp. 993–998, 2016, doi: 10.1016/S2215-0366(16)30089-X.
- [27] R. Stewart and K. Davis, "'Big data' in mental health research: current status and emerging possibilities," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 51, no. 8, pp. 1055–1072, 2016, doi: 10.1007/s00127-016-1266-8.
- [28] D. Hidalgo-Mazzei, A. Murru, M. Reinares, E. Vieta, and F. Colom, "Big Data in mental health: A challenging fragmented future," *World Psychiatry*, vol. 15, no. 2, pp. 186–187, 2016, doi: 10.1002/wps.20307.
- [29] M. Kosinski and T. Behrend, "Editorial overview: Big data in the behavioral sciences," *COBEHA*, vol. 18, pp. iv–vi, 2017, doi: 10.1016/j.cobeha.2017.11.007.
- [30] A. Markowitz, K. Błaszczewicz, C. Montag, C. Switala, and T. E. Schlaepfer, "Psycho-Informatics: Big Data shaping modern psychometrics," *Med. Hypotheses*, vol. 82, pp. 405–411, 2014, doi: 10.1016/j.mehy.2013.11.030.
- [31] D. D. Luxton, "An Introduction to Artificial Intelligence in Behavioral and Mental Health Care," in *Artificial Intelligence in Behavioral and Mental Health Care*, Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States, 2015, pp. 1–26.

- [32] C. C. Bennett and T. W. Doub, "Expert Systems in Mental Health Care: AI Applications in Decision-Making and Consultation," in *Artificial Intelligence in Behavioral and Mental Health Care*, School of Informatics and Computing, Indiana University, Bloomington, IN, United States, 2015, pp. 27–51.
- [33] D. Becker, W. van Breda, B. Funk, M. Hoogendoorn, J. Ruwaard, and H. Riper, "Predictive modeling in e-mental health: A common language framework," *Internet Interv.*, vol. 12, pp. 57–67, 2018, doi: 10.1016/j.invent.2018.03.002.
- [34] S. G. Alonso *et al.*, "Data Mining Algorithms and Techniques in Mental Health: A Systematic Review," *J. Med. Syst.*, vol. 42, no. 9, 2018, doi: 10.1007/s10916-018-1018-2.
- [35] A. G. Di Nuovo, V. Catania, S. Di Nuovo, and S. Buono, "Psychology with soft computing: An integrated approach and its applications," *Appl. Soft Comput. J.*, vol. 8, no. 1, pp. 829–837, 2007, doi: 10.1016/j.asoc.2007.03.001.
- [36] A. Di Nuovo, S. Di Nuovo, S. Buono, and V. Cutello, "Benefits of fuzzy logic in the assessment of intellectual disability," in *IEEE International Conference on Fuzzy Systems*, 2014, pp. 1843–1850, doi: 10.1109/FUZZ-IEEE.2014.6891834.
- [37] S. Nor, W. Shamsuddin, N. Siti, F. Nik, W. Malini, and W. Isa, "Classification Techniques for Early Detection of Dyslexia Using Computer-Based Screening Test," *World Appl. Sci. J.*, vol. 35, no. 10, pp. 2108–2112, 2017, doi: 10.5829/idosi.wasj.2017.2108.2112.
- [38] H. M. Al-Barhamtoshy and D. M. Eldeen Motaweh, "Diagnosis of Dyslexia using Computing Analysis," 2017. Accessed: May 10, 2019. [Online]. Available: <http://www.joetsite.com/wp-content/uploads/2017/07/Vol.-62-37-2017.pdf>.
- [39] H. Selvi and M. S. Saravanan, "A Study of dyslexia using different machine learning algorithm with data mining techniques," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 4, Research and Development Centre Science, Bharathiar University, India, pp. 3406–3411, 2018.
- [40] B.-M. Chen, X.-P. Fan, Z.-M. Zhou, and X.-R. Li, "Application of computer system based on artificial neural network and artificial intelligence in diagnosing child mental health disorders," *J. Clin. Rehabil. Tissue Eng. Res.*, vol. 15, no. 13, pp. 2467–2470, 2011, doi: 10.3969/j.issn.1673-8225.2011.13.044.
- [41] P. Dhaka and R. Johari, "Big data application: Study and archival of mental health data, using MongoDB," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, 2016, pp. 3228–3232, doi: 10.1109/ICEEOT.2016.7755300.
- [42] D. A. Rosenthal, J. A. Dalton, and R. Gervery, "Analyzing vocational outcomes of individuals with psychiatric disabilities who received state vocational rehabilitation services: A data mining approach," *Int. J. Soc. Psychiatry*, vol. 53, no. 4, pp. 357–368, 2007, doi: 10.1177/0020764006074555.
- [43] C. Yuan, "Data mining techniques with its application to the dataset of mental health of college students," in *Proceedings - 2014 IEEE Workshop on Advanced Research and Technology in Industry Applications, WARTIA 2014*, 2014, pp. 391–393, doi: 10.1109/WARTIA.2014.6976277.
- [44] D. Cheng, T. Li, and L. Niu, "A Study on the Application of the Decision Tree Algorithm in Psychological Information of Vocational College Students," in *MATEC Web of Conferences*, 2015, vol. 22, doi: 10.1051/mateconf/20152201044.
- [45] J. QingHua, "Data mining and management system design and application for college student mental health," in *Proceedings - 2016 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2016*, 2017, pp. 410–413, doi: 10.1109/ICITBS.2016.96.

- [46] S. V Tyulyupo, A. A. Andrakhanov, B. A. Dashieva, and A. V Tyryshkin, "Adolescents psychological well-being estimation based on a data mining algorithm," in *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 - Proceedings*, 2018, vol. 1, pp. 475–478, doi: 10.1109/STC-CSIT.2018.8526628.
- [47] A. Shrestha, S. Bergquist, E. Montz, and S. Rose, "Mental Health Risk Adjustment with Clinical Categories and Machine Learning," *Health Serv. Res.*, vol. 53, pp. 3189–3206, 2018, doi: 10.1111/1475-6773.12818.
- [48] M. Srividya, S. Mohanavalli, and N. Bhalaji, "Behavioral Modeling for Mental Health using Machine Learning Algorithms," *J. Med. Syst.*, vol. 42, no. 5, 2018, doi: 10.1007/s10916-018-0934-5.
- [49] S. Ohlsson, R. H. Sloan, G. Turán, and A. Urasky, "Measuring an artificial intelligence system's performance on a Verbal IQ test for young children*," *J. Exp. Theor. Artif. Intell.*, vol. 29, no. 4, pp. 679–693, 2017, doi: 10.1080/0952813X.2016.1213060.
- [50] S. Ohlsson, R. H. Sloan, G. Turán, D. Uber, and A. Urasky, "An approach to evaluate AI commonsense reasoning systems," in *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25*, 2012, pp. 371–374, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84864991190&partnerID=40&md5=1f74a6af3d3d81c4a3bb261335f63ef2>.
- [51] F. Martínez-Plumed, C. Ferri, J. Hernández-Orallo, and M. J. Ramírez-Quintana, "A computational analysis of general intelligence tests for evaluating cognitive development," *Cogn. Syst. Res.*, vol. 43, pp. 100–118, 2017, doi: 10.1016/j.cogsys.2017.01.006.
- [52] D. G. T. Barrett, F. Hill, A. Santoro, A. S. Morcos, and T. Lillicrap, "Measuring abstract reasoning in neural networks," in *35th International Conference on Machine Learning, ICML 2018*, 2018, vol. 10, pp. 7118–7127, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85057325020&partnerID=40&md5=e585112ce962e18dc3b0143e9dc7f36e>.