

**ESTUDIO COMPARATIVO DE ALGORITMOS DE INTELIGENCIA ARTIFICIAL Y  
MINERÍA DE DATOS ENFOCADOS A LA TOMA DE DECISIONES  
EMPRESARIALES DE SELECCIÓN DE PERSONAL**

**Cesar Alejandro Romero Romero**

**UNIVERSIDAD DE CUNDINAMARCA**

**FACULTAD DE INGENIERÍA**

**INGIENERIA DE SISTEMAS**

**FACATATIVÁ**

**2018**

**ESTUDIO COMPARATIVO DE ALGORITMOS DE INTELIGENCIA ARTIFICIAL Y  
MINERÍA DE DATOS ENFOCADOS A LA TOMA DE DECISIONES  
EMPRESARIALES DE SELECCIÓN DE PERSONAL**

**Cesar Alejandro Romero Romero**

**DIRECTOR:**

**Alexander Espinosa**

**CO-DIRECTOR:**

**Cesar Yesid Barahona**

**UNIVERSIDAD DE CUNDINAMARCA**

**FACULTAD DE INGENIERÍA**

**INGIENERIA DE SISTEMAS**

**FACATATIVÁ**

**2018**

**Notas de aceptación:**

---

---

---

---

---

---

---

**Firma del Jurado:**

---

**Firma del Jurado:**

---

## **AGRADECIMIENTOS**

La presente Tesis se encuentra dedicada a todas aquellas personas que siempre se encontraron presentes en esta etapa, aportando continuamente sus conocimientos y apoyo. En primer lugar, a Alexander Espinosa Director del presente trabajo, siendo el docente que siempre guio el camino de sus alumnos, viendo más allá de un simple conocimiento a convertir una carrera en una pasión. A mi familia que siempre estuvo presente en cada uno de los peldaños de este camino y a todos los que conforman la comunidad de la universidad de Cundinamarca.

**Cesar Alejandro Romero Romero**

ESTUDIO COMPARATIVO DE ALGORITMOS DE INTELIGENCIA ARTIFICIAL Y  
MINERÍA DE DATOS ENFOCADOS A LA TOMA DE DECISIONES  
EMPRESARIALES DE SELECCIÓN DE PERSONAL.

**LÍNEA DE INVESTIGACIÓN:**

Software, sistemas emergentes y nuevas tecnologías.

## TABLA DE CONTENIDO

RESUMEN.....	11
1. INTRODUCCIÓN.....	12
2. GENERALIDADES.....	14
2.1. Impacto del Proyecto.....	14
2.2. Planteamiento del problema.....	14
2.3. Objetivos.....	15
2.3.1. Objetivo General.....	15
2.3.2. Objetivos Específicos .....	15
2.4. Metodología .....	16
3. INFORME DE INVESTIGACIÓN.....	17
3.1 Marcos de referencia .....	17
3.1.1 Marco Histórico.....	17
3.1.2 Marco Teórico.....	19
3.1.3. Marco Legal .....	22
4. ESTADO DEL ARTE .....	28
5. ARQUITECTURA Y MODELADO .....	58
5.1 Diseño de base de datos .....	58
5.1.1 Modelo Entidad Relación .....	58
5.2. Dimensión Estática (Estructura) .....	61
5.2.1. Diagrama de casos de uso .....	61
5.3. Dimensión Dinámica (Interacción).....	66
5.3.1. Diagrama de Secuencias.....	66
5.4. Dimensión Funcional (comportamiento).....	67
5.4.1. Diagrama de Actividades .....	67
5.4.2. Diagrama de Clases.....	69
5.5 . Dimensión Organizacional (Implementación) .....	70
5.5.1. Diagrama de Despliegue .....	70
5. RESULTADOS .....	72
5.1. Modelo propuesto.....	72
5.2. Comparación de algoritmos e machine learning.....	75
5.3. Secuencia de actividades para el postulado .....	81

<b>5.4. Secuencia de actividades para el personal de selección.....</b>	<b>111</b>
<b>5.5 Análisis de resultados .....</b>	<b>118</b>
<b>5.6 Análisis de algoritmo de distancias.....</b>	<b>130</b>
<b>6. CONCLUSIONES Y RECOMENDACIONES .....</b>	<b>134</b>
<b>BIBLIOGRAFÍA.....</b>	<b>135</b>

## LISTA DE ILUSTRACIONES

Ilustración 1 Modelo Entidad Relación .....	59
Ilustración 2 Casos de Uso- Postulado .....	62
Ilustración 3 Casos de Uso- Selección .....	63
Ilustración 4 Casos de Uso- Administrador .....	65
Ilustración 5 Diagrama de Secuencias.....	66
Ilustración 6 Diagrama de Actividades.....	68
Ilustración 7 Diagrama de Clases .....	69
Ilustración 8 Diagrama de Despliegue.....	71
Ilustración 9. Hojas de vida .....	82
Ilustración 10. Subir hoja de vida .....	83
Ilustración 11. Crear hoja de vida .....	83
Ilustración 12. Datos básicos.....	84
Ilustración 13. Test de perfilación .....	85
Ilustración 14. Gráficos resultados de test.....	110
Ilustración 15. Interface para personal de selección .....	111
Ilustración 16. Hojas de vida sin procesar.....	112
Ilustración 17 Hojas de vida tratadas.....	113
Ilustración 18. Filtros por personalidad. ....	113
Ilustración 19 Perfiles organizados por porcentaje de compatibilidad .....	114
Ilustración 20 Perfiles organizados por porcentaje de compatibilidad graficas .....	114
Ilustración 21 Perfiles Algoritmos de machine learning evaluados .....	115
Ilustración 22 Perfiles Creación del perfil mas idóneo para el cargo.....	116
Ilustración 23 Resultados del algoritmo de distancias por porcentaje de compatibilidad con el perfil meta.....	116
Ilustración 24 Conclusiones.....	117
Ilustración 25 Resultado de búsqueda perfil 1.....	118
Ilustración 26 Resultado de test 1.....	119
Ilustración 27 Resultado de búsqueda perfil 2.....	120
Ilustración 28 Resultado test personalidad 2.....	121
Ilustración 29 Resultado de búsqueda perfil 3.....	122
Ilustración 30 Resultado test personalidad 3.....	123
Ilustración 31 Resultado de búsqueda perfil 4.....	124
Ilustración 32 Resultado test personalidad 4.....	125
Ilustración 33 Resultado de búsqueda perfil 5.....	126
Ilustración 34 Resultado test personalidad 5.....	127
Ilustración 35 Resultado de búsqueda perfil 6.....	128
Ilustración 36 Resultado test personalidad 6.....	129
Ilustración 37 Distancia con el perfil meta .....	130



Ilustración 38 Distancia con el perfil meta 3 dimensiones.....	131
Ilustración 39 Grupos encontrados .....	132
Ilustración 40 Características compatibles con el perfil idóneo.....	133

## LISTA DE TABLAS

Tabla 1 Modelo Entidad Relación.....	61
Tabla 2 Casos de Uso- Postulado .....	63
Tabla 3 Casos de Uso- Proyecto .....	64
Tabla 4 Casos de Uso- Administrador .....	65
Tabla 5 Diagrama de Secuencias .....	67
Tabla 6 Diagrama de Actividades- .....	69
Tabla 7 Diagrama de Clases.....	70
Tabla 8 Diagrama de Despliegue .....	71

## RESUMEN

El presente trabajo de tesis cuyo objetivo es la toma de decisiones empresariales mediante técnicas de machine learning, responde a la importancia que ejercen estas decisiones dentro de las organizaciones dado que en este proceso se aíslan todas las posibles rutas de solución en el momento que se define un plan de acción. Usualmente la totalidad de variables o procesos internos que se verán involucrados en dicha solución no son conocidos, de hecho, el porcentaje usual de conocimiento sobre la situación específica es del 40% al 70%, esta brecha es la recomendable para tomar la decisión dado que obtener el 100% de la información implica un retardo significativo para que esta sea efectiva o ejerza un impacto notorio (Colin Powel). Desde esta perspectiva, la abstención de decisiones para generar ideas, también cuenta como “la que más afecta las organizaciones” no permitiendo que se adquiera experiencia y por lo tanto, cuando se deba enfrentar un nuevo reto terminara de la forma errónea.

Machine learning es un conjunto de técnicas algorítmicas que se aplican a la descripción de un problema con base en el aprendizaje autónomo, siendo la mejor opción para problemas específicos y de crecimiento exponencial en la mayoría de organizaciones, como la selección de personal. La selección de personal se realiza actualmente de forma manual aplicando una serie de reglas ordenadas empleando herramientas informáticas en las cuales se repite el proceso una y otra vez hasta conseguir un resultado esperado, pero de esta forma se ignora una serie de variables que pueden influir fuertemente en la perfilación correcta.

El estudio exploratorio de los algoritmos trae consigo una clasificación de estos, el enfoque en que mejor se desenvuelven y aún más importante los resultados en cuanto a características tales como fortalezas, debilidades, tiempo de ejecución, facilidad de implementación, etc. Se dará lugar al diseño de un modelo de sistema inteligente que permita una demostración sobre la toma de decisiones en contratación de personal, mediante la aplicación de los algoritmos de clasificación más adecuados según el estudio previo. El diseño estará sustentado en el estudio comparativo, finalmente, se construirá un software a partir del diseño generado para la toma de decisiones dentro de una organización que permitirá realizar la tarea de forma eficiente y mejorada respecto a los métodos tradicionales usados en la actualidad por las distintas áreas de la organización especialmente la selección de personal.

## 1. INTRODUCCIÓN

El cambio constante que se presenta en las organizaciones debido al entorno competitivo, ocasiona una evolución de los procesos,(Collins & Porras, 1996). Para adaptarse a las grandes cantidades de datos y variables nuevas (Big Data), se ha empleado tecnología informática como herramienta fundamental, cambiando los modelos e implicando que la estrategia ya no puede plantearse exclusivamente en las tres dimensiones espaciales, sino que se debe enfocar adicionalmente en una variable crítica: el tiempo.(Manyika et al., 2011)

Las organizaciones deben elegir continuamente entre varios caminos para solucionar conflictos o afrontar nuevos retos, muchos de estos son desechados sin la posibilidad de evaluar los pros y contras, incluso desde el momento en que se toma una decisión se da por sentado que las demás son erróneas y no se evidencian patrones que pueden ser útiles para retroalimentación de la organización. La integración de la toma de decisiones mediante machine learning brinda un análisis de los datos para posteriormente extraer un conjunto de características e identificar patrones relevantes entre ellas.

Para comprender el concepto de aprendizaje autónomo se debe partir de la concepción de inteligencia artificial. Esta nació a partir de la idea de que las maquinas podían desarrollar tareas inteligentes simulando procesos de inteligencia humana. Estos procesos incluyen el aprendizaje (la adquisición de información y reglas para el uso de la información), el razonamiento (usando las reglas para llegar a conclusiones aproximadas o definitivas) y la autocorrección.(Autónoma De Occidente, n.d.) El segundo concepto es el manejo de grandes cantidades de datos (Big Data), que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, describiendo grandes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costosos de cargar a un base de datos relacional para su análisis. De tal manera que, el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.(Mayer-Schonberger & Cukier, 2013)

En el presente documento se aborda el estado del arte que contiene una búsqueda intensiva del tema en cuestión, hallando las principales técnicas de machine learning, evaluando posteriormente sus características y desempeño ante el problema en particular de toma de decisiones empresariales enfocadas en la selección de personal y comparando estos resultados con las investigaciones actuales llegando a un modelo para la construcción de un software de selección de personal. Se debe tener en cuenta que este proceso es fundamental para las organizaciones y el modelo que se obtenga debe por ultimo pasar una etapa de reevaluación y retroalimentación de resultados para así llegar a conclusiones que registren el resultado final. En paralelo se toma la selección de personal desde el punto de vista de las personas que la ejercen; usualmente psicólogos, conociendo a profundidad todos los distintos test usados en este entorno (American Psychological Association, n.d.) comparando las características principales para seleccionar el test que arroje los resultados más óptimos y llevarlo a software unificándolo con el tratamiento de las hojas de vida y los algoritmos anteriormente evaluados.(American Psychological Association, n.d.) (CATTELL & DREYDAHL, 1955).

## **2. GENERALIDADES**

### **2.1. Impacto del Proyecto**

Con el software para la toma de decisiones empresariales especialmente enfocadas a la selección de personal dará un giro bastante interesante en cuanto a la forma en que se evalúan los postulados a un cargo específico, los métodos que se usaban para la clasificación de hojas de vida y perfiles. El beneficio principal entregado por el software, es la reducción considerable de datos que se presenta al personal de selección reduciendo los tiempos considerablemente, la visualización más clara de perfiles y sus características. La automatización de test de personalidad da mayor claridad y veracidad sobre los resultados.

### **2.2. Planteamiento del problema**

Desde el punto en que la humanidad crea un sistema social definido como: sociedad, depende para su continuación y supervivencia de insumos con los sistemas ambientales para generar una estabilidad en relación al cambio, se crean subgrupos de población con tareas y características totalmente distintas, y es allí cuando se puede dar la primera concepción de selección de personal. (Trillas, 1977)

La división en subgrupos en la sociedad, implica una especialización de los individuos en tareas particulares, brindando un aporte y recibiendo una contraprestación. De esta forma aquel que brindaba un servicio específico tal como defensa era alimentado. (Barsotti, 1981) Estos sistemas sociales fueron evolucionando y con ello la cantidad de tareas creció significativamente, así mismo la especialización en cada una de ellas aumento exigiendo un mayor nivel de selección en el momento de desempeñar estas funciones.(Universidad del Zulia. Facultad de Ciencias Económicas y Sociales, Gerardo J.; Canet Giner, 2012)

El éxito en las empresas se ha visto directamente relacionado con la capacidad que tienen estas para definir esquemas claros y concisos de cada una de las tareas que realizan. La delimitación de las funciones de cada uno de los cargos ocasiona procesos más ágiles con metas bien establecidas y prioridades

preestablecidas, pero siguen teniendo el problema más relevante en esta segmentación de procesos,(Erro carmen, 2003) el cual consiste en que se crean constantemente nuevos procesos, incluso, se cambian los ya existentes, pero la selección del personal que desempeñara estas labores se ha desarrollado de forma arbitraria y sin conocimiento mínimo de un perfil específico para responder a las características de este cargo.(Aramburu Goya & Rivera, n.d.)

Sobre la base de las consideraciones anteriores las organizaciones se vieron en una encrucijada por llevar la creación y modificación de procesos a la par con la correcta selección y perfilación del personal a desarrollarla. Por lo tanto emplearon tecnología informática como herramienta fundamental para acompañar estos procesos,(Hamidian, Soto, & Poriet, n.d.) con programas bastante básicos y de uso común llegando a formalizar más este proceso,(“El impacto del uso efectivo de las TIC sobre la eficiencia técnica de las empresas españolas,” 2007) pero aun así surge de nuevo el problema de la limitación de estas herramientas en una perfilación profunda que recolecte y analice los factores decisivos para la correcta toma de decisiones enfocada a la selección de personal.

## **2.3. Objetivos**

### **2.3.1. Objetivo General**

Realizar un análisis comparativo y software demostrativo acerca de algoritmos de inteligencia artificial y minería de datos aplicables en la toma de decisiones empresariales enfocadas al proceso de selección de personal.

### **2.3.2. Objetivos Específicos**

- Identificar los principales algoritmos en minería de datos e inteligencia artificial enfocados a la toma de decisiones empresariales.
- Hacer un estudio comparativo sobre los algoritmos identificados con el fin de determinar un modelo adecuado para un sistema inteligente de selección de personal.

- Adecuación de los modelos generados por el estudio previo para la creación del esquema general sobre el que se construirá el software.
- Realizar Software para la recolección de datos y posterior alimentación de los algoritmos, llegando a la representación de los resultados comparativos entre ellos.

## **2.4. Metodología**

Considerando el estudio previo, el cual se evidencia en el estado del arte, se ha realizado un estudio comparativo de los principales algoritmos de machine learning que más se acoplan al problema específico de selección de personal en organizaciones, llegando de esta forma al planteamiento de un modelo que aplique a este entorno específico.

Para esta investigación y posterior realización de la herramienta de estimación se aplicará la metodología de desarrollo ágil SCRUM, que permitirá avances temporales en un tiempo determinado, ayudando a una mejor organización, un mayor compromiso por parte de los integrantes de este estudio, un mejor trabajo en equipo y solución de errores a comienzo de cada iteración del cronograma que será planteado.

- En el despliegue del proyecto se plantean varias etapas las cuales se irán desarrollando en el transcurso de 6 meses a partir de la fecha. La primera fase del proyecto se basa en Levantamiento de datos sobre las hojas de vida y perfiles.
- En segundo lugar se realizara el script para subir los datos de pdf a la base de datos, para su posterior normalización, cuando se tenga unos datos claros se realizara una interface gráfica para poder visualizarlos.
- El estudio comparativo de los algoritmos de machine learning permite elegir de forma clara el mejor algoritmo para el modelo más adecuado en la selección de personal y así realizar una comparación con métodos tradicionales.



### 3. INFORME DE INVESTIGACIÓN

#### 3.1 Marcos de referencia

##### 3.1.1 Marco Histórico

###### Inteligencia artificial

En 1943, Warren McCulloch y Walter Pitts propusieron un modelo de neurona del cerebro humano y animal. Estas neuronas nerviosas abstractas son conceptualmente la representación simbólica de actividad cerebral. Luego de ello Norbert Wiener unificó estas ideas, dentro del mismo campo, que se llamó "cibernética", de esta forma nacería en los años 50 la inteligencia artificial. Los primeros investigadores de esta innovadora ciencia, tomaron como base la neurona formalizada de McCulloch quien postuló que: " El cerebro es un solucionador inteligente de problemas, de modo que imitemos al cerebro". Pero si consideramos la enorme complejidad del mismo esto es ya prácticamente imposible, ni que mencionar que el hardware de la época ni el software estaban a la altura para realizar semejantes proyectos.(DW Rolston, AP Gama, n.d.) La comunidad científica quedó profundamente cautivada al publicarse un artículo del inglés Alan Turing llamado "Maquinaria computacional e Inteligencia, Turing basó su artículo en una pregunta elemental ¿pueden las máquinas pensar? Con una serie de argumentos bastante fuertes él sostenía su postura a favor de la posibilidad de una inteligencia artificial en las máquinas, se inició un intenso debate integrando la IA con la psicología centrándose en el análisis de una serie de problemas implicados en la aplicación de términos mentalistas en la computación a lo cual Turing sostenía que esa barrera entre inteligencia artificial e inteligencia natural.(Turing, n.d.) El trabajo de Turing, quien falleció prematuramente, fue continuado en los Estados Unidos por John Von Neumann durante la década de los cincuenta. Su contribución central fue la idea de que las computadoras deberían diseñarse tomando como modelo al cerebro humano. Von Neumann fue el primero en "antropomorfizar" el lenguaje y la concepción de la computación al hablar de la "memoria", los "sensores", etc., de las computadoras. Construyó una serie de máquinas utilizando lo que a principios de los cincuenta se conocía sobre el cerebro humano, y diseñó los primeros programas almacenados en la memoria de una computadora. (Bruckmann & Weber, 1971)(ESTUDIOS. filosofía-historia-letras & Otoño 1987, n.d.)

## Minería de datos

Aunque desde un punto de vista académico el término data mining es una etapa dentro de un proceso mayor llamado extracción de conocimiento en bases de datos, en el entorno comercial, así como en este trabajo, ambos términos se usan de manera indistinta. Lo que en verdad hace el data mining es reunir las ventajas de varias áreas como la Estadística, la Inteligencia Artificial, la Computación Gráfica, las Bases de Datos y el Procesamiento Masivo, principalmente usando como materia prima las bases de datos. Una definición tradicional es la siguiente:

Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Desde el punto de vista empresarial, se define como: La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión (Ozzie, 2006)

La idea de data mining no es nueva. Ya desde los años sesenta los estadísticos manejaban términos como data fishing, data mining o data archaeology con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido. A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, empezaron a consolidar los términos de data mining y KDD. A finales de los años ochenta sólo existían un par de empresas dedicadas a esta tecnología; en 2002 existieron más de 100 empresas en el mundo que ofrecen alrededor de 300 soluciones. (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004)

Las listas de discusión sobre este tema las forman investigadores de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios. El data mining es una tecnología compuesta por etapas que integra varias áreas y que no se debe confundir con un gran software. Durante el desarrollo de un proyecto de este tipo se usan diferentes aplicaciones software en cada etapa que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente. (Pérez López & Santín González, 2007)

Actualmente existen aplicaciones o herramientas comerciales de data mining muy poderosas que contienen un sinnúmero de utilidades que facilitan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta. La data mining es la etapa de descubrimiento en el proceso de KDD: Paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos preprocesados (Fayyad et al., 1996) Aunque se suelen usar indistintamente los términos KDD y Minería de Datos.

### **3.1.2 Marco Teórico**

Dado que el presente documento contendrá un estudio comparativo sobre distintos algoritmos para la toma de decisiones empresariales enfocadas a la selección de personal, resulta fundamental empezar por la definición de toma de decisiones en organizaciones: La toma de decisiones y las alternativas se encuentran en problemas como en nuevos retos, valores cuantitativos y no cuantitativos, en la mayoría de ocasiones la toma de decisiones busca la implementación de metodologías cuantitativas que brinda la administración. En concreto se podría definir como el proceso para identificar y solucionar un curso de acción para resolver un problema específico.

El objetivo principal del machine learning es desarrollar técnicas para poder emular el aprendizaje del ser humano en las máquinas y aunque tal vez estas tecnologías nos rodean a diario, desde Facebook con filtros de nuestros gustos personales hasta google y distintas herramientas para el filtrado de spam, pero en este punto surge la pregunta primordial ¿Cómo aprenden los ordenadores? Para lograr entender este funcionamiento se debe tener en cuenta que la base de este aprendizaje es los datos que ingresan para ser analizados, posteriormente a estos se les extraen sus características para que un algoritmo pueda identificar patrones que estén a fin con el resultado esperado, es claro que para el ordenador estos patrones no significan nada en contexto todo esto toma forma hasta el punto de pasar por un modelo matemático, este modelo puede adaptarse constantemente a los cambios presentados es decir, el modelo matemático se adapta a nuevas metas y preferencias.

## - **Herramientas de software**

### **PHP**

Este lenguaje cuyo acrónimo es Hipertext Preprocessor posee una sintaxis similar a la que se trabaja en C++ o Java, lo que caracteriza a PHP de estos es la generación dinámica de páginas web en donde es más popular. Además, suele ir introducido en páginas HTML o XHTML siendo el servidor web quien lo ejecuta.

Asimismo, es un lenguaje de tipo libre disponible para muchos sistemas (Windows, Linux, etc.). Asimismo, una de sus tantas ventajas es el hecho de poseer muchas extensiones para conectar a base de datos, manejo de sockets, generación de PDF, etc. Y tal vez algo a considerar es que si ocurre algún problema como por ejemplo virus el sistema se podrá seguir usando desde que el navegador web tenga conexión al servidor. (Duarte, 2015)

Igualmente, según el conocimiento adquirido se conoce que PHP es un lenguaje de desarrollo orientado a la programación WEB y este consta de dos etapas llamadas BackEnd y FrontEnd.

BackEnd se conoce como la etapa de acceso de datos y es la encargada de todo el procesamiento lógico que trae una página web para que esta funcione correctamente. Trabaja directamente con el servidor, recibe peticiones desde la fase web y como retroalimentación envía código HTML como resultado, cabe aclarar que todo el procesamiento que se realiza en esta etapa queda oculto a la vista del cliente.

FrontEnd es la etapa de diseño de la página web, esta consta de instrucciones HTML que procesará el navegador web para visualizarse al cliente el software programado. En esta fase el programador se encarga de diseñar la estructura y organización visual del software implementando estilos en CSS, colores, fondos, animaciones y demás parte gráfica. (Cobo, Gómez, Pérez, & Rocha, 2005)

### **MYSQL**

Es un sistema de gestión de base de datos relacional de código abierto operado por ORACLE Corporation, Es una de las bases de datos más utilizadas en el mundo y su fuerte es el entorno web, aunque en ocasiones es utilizada para aplicaciones de prueba de escritorio.

MySQL cuenta con varias interfaces de usuario para ser utilizada de manera más practica por los desarrolladores, entre las más conocidas se encuentran SQLyog, MySql Workbench, PHPMysqlAdmin.

PHPMyAdmin es la interfaz más conocida por los desarrolladores PHP, ya que muchos de los servidores web que trabajan este lenguaje de programación cuentan con esta herramienta y permiten administrar la base de datos del sistema en su mismo servidor.

MySQL le garantiza al usuario la integridad de sus datos en donde todas las transacciones realizadas serán atómicas, es decir, si existe una falla en el sistema no la deja a la mitad y la cancela. (Cobo et al., 2005)

## - Metodología de desarrollo

### SCRUM

Es una metodología de proceso ágil para el desarrollo de proyectos de software se caracteriza por la elaboración incremental en vez de la planificación y ejecución completa del proyecto final que es como trabajan otras metodologías y además en cada fase de desarrollo se realizan las tareas una tras otras de forma secuencia.

SCRUM fue identificado por Ikujiro Nonaka e Hirotaka Takeuchi a principios de los 80 al observar como las principales empresas manufactureras desarrollaban los nuevos productos; pero no fue hasta 1995 que *Ken Schwaber* presento "Scrum Development Process" en OOPSLA 95 (Object Oriented Programming Systems & Applications conference) con una serie de reglas para desarrollar software que había empleado en la elaboración de Delphi y *Jeff Sutherland* en su empresa Easel Corporation que es una compañía que en los macrojuegos de compras y fusiones, se integraría en VMARK, y luego en Informix y finalmente en Ascential Software Corporation). (Palacio & Ruata, 2011)

En Scrum existen tres roles:

- **El dueño del producto (Product owner):** Es quien trata de maximizar el trabajo en equipo y el valor del producto, asimismo es el responsable de realizar la lista de trabajo que incluye ordenación de elementos para cumplir con los objetivos de la mejor manera, optimización de trabajo por parte del equipo de desarrollo, asegurarse de que la lista sea clara y entendible para todos.
- **El equipo de desarrollo (Development Team):** son los que desempeñan el trabajo y se encargan de entregar un incremento terminado para ponerlo en producción al finalizar cada sprint (tiempo

donde se entregará un avance de una parte del proyecto que se esté realizando).

- **El Scrum Master:** es el responsable de cerciorarse de que la metodología es entendido y cumplido. Además, es un líder que está pendiente del equipo de desarrollo asegurándose del cumplimiento de roles y sus debidas responsabilidades. (Schwaber & Sutherland, 2013)

### **Tipo de investigación:**

La investigación experimental consiste en la manipulación de una variable experimental no comprobada, en condiciones rigurosamente controladas, con el fin de describir de qué modo o por qué causa se produce una situación o acontecimiento en particular.

Se trata de un experimento porque precisamente el investigador provoca una situación para introducir determinadas variables de estudio manipuladas por él, para controlar el aumento o disminución de esa variable, y su efecto en las conductas observadas. El investigador maneja deliberadamente la variable experimental y luego observa lo que sucede en situaciones controladas.

### **3.1.3. Marco Legal**

Este proyecto se desarrollará basado en software libre, no se tendrán que adquirir licencias para trabajar con él.

Software libre se refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, cambiar, y mejorar el software. De modo más preciso, se refiere a cuatro libertades de los usuarios del software libre:

- La libertad de usar el programa, con cualquier propósito (Libertad 0).
- La libertad de estudiar cómo funciona el programa, y adaptarlo a sus necesidades. El acceso al código fuente es una condición previa a esto.
- La libertad de distribuir copias, con lo que pueda ayudar a otras personas.

- La libertad de mejorar el programa y hacer públicas las mejoras a los demás, de modo que toda la comunidad se beneficie. (Libertad 3). El acceso al código fuente es un requisito previo a esto.

Para que estas libertades sean reales, deben ser irrevocables mientras no se haga nada incorrecto; si el desarrollador del software tiene poder de revocar la licencia aunque no se le haya dado motivos, el software no es libre.

El marco legal de esta tesis está regido por las siguientes estipulaciones.

El derecho de autor es una especie dentro de la institución de la propiedad intelectual, en virtud de la cual se otorga protección a las creaciones expresadas a través de los géneros literario o artístico, los derechos de autor nacen desde el momento de la creación de la obra, es decir, desde que se expresa la idea del autor. A esto se le llama principio de protección automática: en cuanto existe obra, existe protección, aun si la obra no ha sido publicada. Para que exista obra no basta con solo la idea, sino que ella debe ser exteriorizada.

La constitución política de Colombia de 1991, en el artículo 61, expresa “El estado protegerá la propiedad intelectual por el tiempo y mediante las formalidades que establezca la ley” (Dirección nacional de derechos de autor).

Ley 23 de 1982

En el artículo 1 define: Los autores de obra literarias, científicas y artísticas gozaran de protección para sus obras en la forma prescrita por la presente ley y en cuanto duere compatible con ella, por el derecho común.

En el artículo 2 define: Los derechos de autor recaen sobre las obras científicas literarias y artísticas las cuales comprenden toda la creación del espíritu en el campo científico, literario y artístico, cualquiera que sea el modo o forma de expresión y cualquiera que sea su distribución, tales como libros, folletos u otros escritos las conferencias, alocuciones, sermones y otras obras de la misma naturaleza.

Decreto 460 de 1995

Artículo 1. El Registro Nacional del Derecho de Autor es competencia de la Unidad Administrativa Especial - Dirección Nacional del Derecho de Autor, con carácter único para todo el territorio nacional.

Artículo 2. Para los efectos del artículo 3 de la Ley 44 de 1993, el Registro Nacional del Derecho de Autor es un servicio que presta el Estado a través de la Unidad Administrativa Especial - Dirección Nacional del Derecho de Autor, cuya finalidad es la de brindarle a los titulares del derecho autor y derechos conexos un medio de prueba y de publicidad a sus derechos así como a los actos y contratos que transfieran o cambien ese dominio amparado por la ley, y garantía de autenticidad y seguridad a los títulos de derecho de autor y de derechos conexos y a los actos y documentos que a ellos se refiere.

Artículo 3. La protección que se brinda a las obras literarias y artísticas, así como a las interpretaciones y demás producciones salvaguardadas por el derecho conexo, no estará subordinada a ningún tipo de formalidad, y en consecuencia el registro que aquí se reglamenta será para otorgar mayor seguridad jurídica a los autores y titulares.

Artículo 4. Los datos consignados en el Registro Nacional del Derecho de Autor se presumirán ciertos, hasta tanto se demuestre lo contrario.

Artículo 5. Las inscripciones realizadas en el Registro Nacional del Derecho de Autor son de carácter público y, en consecuencia, pueden ser consultadas en virtud del derecho de petición y conforme a sus principios reguladores.

La reproducción de las obras editadas o inéditas y la consulta de las obras inéditas inscritas sólo se podrá realizar por los autores de ellas, por sus derechohabientes que acrediten tal calidad y por las autoridades judiciales o por quienes ellos dictaminen.

Artículo 6. El Jefe de la Oficina de Registro de la Dirección Nacional del Derecho de Autor podrá, de oficio o a solicitud de parte, corregir los simples errores mecanográficos o numéricos cometidos al realizar una inscripción, atendiendo lo dispuesto sobre el particular en el Régimen de Instrumentos Públicos.



Las cancelaciones, adiciones o modificaciones de las inscripciones efectuadas en el Registro Nacional del Derecho de Autor, solo procederán a solicitud del autor y de los derechohabientes que demuestren tal calidad, quienes deberán allegar la documentación que soporte su petición, o en virtud de orden judicial.

Artículo 7. Para todos los efectos, el Registro Nacional del Derecho de Autor deberá ajustarse en lo posible, a la forma y términos prescritos en el Estatuto de Registro de Instrumentos Públicos.

## OBRAS IMPRESAS

a) Impreso de carácter monográfico: publicación completa en una sola parte o que se piensa completar con un número determinado de partes, publicadas por separado y que no pertenece a una serie. Los impresos de carácter monográfico abarcan: libros, folletos, pliegos sueltos.

Libro: Reunión de muchas hojas de papel, vitela, u otras, ordinariamente impresas, que se han cosido o encuadernado juntas con cubierta de papel, cartón, pergamino u otra piel, y que forman un volumen.

Folleto: Obra impresa, no periódica, que no consta de bastantes hojas para formar un libro.

Pliego: Pieza suelta de papel impresa por uno o ambos lados;

b) Publicación seriada: publicación que aparece en partes sucesivas, a intervalos regulares o irregulares, cada una de las cuales presenta designaciones numéricas o cronológicas y que pretende continuarse indefinidamente. Las publicaciones seriadas incluyen periódicos o diarios; anuarios, revistas, memorias, actas, entre otros, de entes corporativos;

c) Material cartográfico: cualquier material que presente la totalidad o una parte de la tierra o de cualquier cuerpo celeste. Los materiales cartográficos abarcan: mapas

o planos en dos o tres dimensiones; cartas aeronáuticas, de navegación o celestes; atlas; globos; diagramas en bloque; fotografías aéreas con fines cartográficos; vistas a ojo de pájaro; croquis, grabados topográficos; imágenes aéreas, espaciales y terrestres; modelos de relieve; entre otros;

d) Música: Serie de pentagramas en donde están impresas todas las partes instrumentales y/o vocales de una obra musical, colocados uno debajo de otro en forma vertical, de modo que las partes puedan leerse simultáneamente. Así mismo, los pentagramas para una de las voces o instrumentos que participan en una obra musical. Incluye: partituras abreviadas, partituras cortas, partituras de bolsillo, partes de piano del director, partituras vocales, partituras para piano, partituras corales, partituras y partes en general.

## SOFTWARE Y BASE DE DATOS

g) Archivo de datos legibles por máquina: cuerpo de información codificado por métodos que requieren el uso de una máquina (típicamente una computadora) para el procesamiento. Pertenecen a esta categoría: archivos almacenados en cinta magnética, módulos de disco, tarjetas de marca sensible, documentos fuente en caracteres de reconocimiento óptico;

El término de datos legibles por máquina, se refiere tanto a los datos almacenados en forma legible por máquina como a los programas usados para procesar esos datos;

h) Material gráfico: representación en dos dimensiones, puede ser opaca o destinada a ser vista o proyectada, sin movimiento, por medio de un aparato óptico. Los materiales gráficos abarcan: carteles, diagramas, diapositivas, dibujos técnicos, estampas, estereografías, fotobandas, fotografías, reproducciones de obras de arte, tarjetas nemotécnicas, tarjetas postales y transparencias.

i) Microforma: Término genérico para cualquier medio, ya sea transparente u opaco, que contenga microimágenes, como las microfichas, microfilmes, microopacos, etc.

Artículo 24. La Biblioteca Nacional de Colombia ser la entidad responsable del Depósito Legal.

Artículo 25. El Depósito Legal se deberá efectuar observando lo siguiente:

a) Tratándose de obras impresas de carácter monográfico, publicaciones seriadas, material cartográfico, material gráfico, microformas, soporte lógico (software), música o archivo de datos legible por máquina, entre otros, el editor deberá entregar dos (2) ejemplares a la Biblioteca Nacional de Colombia, un (1) ejemplar a la Biblioteca del Congreso y un (1) ejemplar a la Biblioteca de la Universidad Nacional de Colombia.

Si la obra ha sido editada en lugar diferente al Departamento de Cundinamarca, deberá además entregarse otro ejemplar a la biblioteca departamental donde tenga asiento principal el editor.

### **Habeas Data**

El Habeas Data es el derecho que tiene toda persona para conocer, actualizar y rectificar toda aquella información que se relacione con ella y que se recopile o almacene en centrales de información.

Este derecho está regulado por la ley 1266 de 2008.

Información positiva:

Las centrales de riesgo están legalmente facultadas Para conservar en sus archivos toda la información histórica de una persona, esto es, tanto los datos positivos como los negativos.

En relación el aspecto relacionado con la permanencia de la información en las base de datos el artículo 13 de la Ley 1266 de 2008 dispone que la información de carácter Positivo permanecerá indefinidamente en los bancos de datos de los operadores de información.

Información negativa:

Los datos cuyo contenido hacen referencia al tiempo de mora, tipo de cobro, estado de la cartera y en general, aquellos datos referentes a una situación de incumplimiento de obligaciones, se regirán por un término máximo de cuatro (4) años, a partir de la fecha en que sean pagadas las cuotas vencidas, o que se extinga la obligación por cualquier modo. Sin embargo, la Corte Constitucional estableció un plazo de permanencia del doble de la mora para los deudores que presentaron mora inferior a dos (2) años y asumieron voluntariamente el pago de la obligación y plazo de permanencia será de (4) años cuando la mora máxima alcanzada fue superior a los (2) años.

#### 4. ESTADO DEL ARTE

En las empresas se deben tomar decisiones las cuales se encuentran basadas en opciones, estas opciones no siempre son realizadas conociendo todos las consecuencias o resultados derivados de cada una de ellas e incluso sin las bases suficientes en cuanto a conocimientos específicos de aquella necesidad.(Ruigrok, Peck, & Keller, 2006)

Sobre la base de las consideraciones anteriores los marcos de referencia sobre las decisiones se han transformado en distintos estándares y metodologías que brindan mayor seguridad a la base de manejo de los procesos. (Simon, n.d.) Este conocimiento, que no se tenía, sobre la forma en que se crean y evolucionan los procesos dentro de una organización presentaba la mayor dificultad para los administrativos, surgiendo el software como solución a esta carencia, siendo estos procesos la fuente de información necesaria para posteriormente categorizarlos para su estandarización. (REILLY, 2014)

La selección de personal es una de las decisiones más difíciles para una empresa debido a que en esta etapa es donde se ingresan personas ajenas a la organización acarreando vulnerabilidades, demoras en tiempos de aprendizaje, adaptación a un nuevo entorno y en casos muy particulares la reubicación del personal interno en otra dirección con el fin de poder explotar sus capacidades o puntos de vista distintos. (PAGUAY VARGAS, 2016) Por ello en el transcurso de las organizaciones se han creado técnicas de clarificación la mayoría con una incertidumbre total sobre los perfiles necesarios a tener en cuenta para un cargo, ocasionando de esta forma la selección inadecuada. La buena selección en una organización puede definir perfectamente que tan productiva llega a ser, por eso es de vital importancia la forma en la cual se tienen estos parámetros, normas o técnicas, estando compuesta en la mayoría de casos por 6 etapas:

- 1- Elaboración del perfil de la vacante
- 2- Reclutamiento
- 3- Selección y evaluación
- 4- Contratación y presentación

## 5- Seguimiento

La fase restante se encuentra entre el reclutamiento y la contratación siendo está en la que se definen si pasaran a la siguiente etapa de contratación. (Sánchez Chávez, 2017)

Las técnicas usadas en selección de personal basadas en perfiles han evolucionado a medida que aumentan los postulados a vacantes y el nivel de especialización requerido en ellas, llegando así a la integración con las técnicas de minería de datos basadas en arboles de decisión y reglas de asociación, que en conjunto generan una relación entre la información del personal con su desempeño laboral y la retención dentro de la misma organización. (Chien & Chen, 2008)

Dadas las condiciones que anteceden se observa claramente que los métodos basados en una solución ideal sesgando o dejando atrás a posibles soluciones que brindan puntos fuertes en otras áreas son descartadas por el alto nivel de subjetividad en los conjuntos de reglas empleadas en la evaluación, para ello se han integrado métodos mucho más versátiles que se incorporan a la evaluación la distancia entre las distintas alternativas disponibles de candidatos generando un incentivo a los más calificados según un conjunto de reglas de tomadores de decisiones (TOPSIS). (Kelemenis & Askounis, 2010)

La cantidad de datos que crecen constantemente dentro de una organización, pero ¿Cómo se adapta la toma de decisiones a estos grandes volúmenes de datos?, La minería de datos (Data Mining) como respuesta a ello busca patrones ocultos en los datos para ser usados posteriormente en la predicción de comportamientos a futuro convirtiendo así el conocimiento en proactivo.(Bhargavi, Jyothi, Jyothi, & Sekar, 2008) (Han, Pei, & Kamber, 2011)

Los mecanismos y métodos de minería de datos personales se emplean para identificar la información relevante que de otra manera probablemente permanecería sin descubrir. Los usuarios suministran datos personales que pueden ser analizados conjuntamente con datos asociados con otros usuarios para proporcionar información útil que puede mejorar las operaciones comerciales y / o la calidad de vida. Los datos personales pueden ser extraídos individualmente o en

conjunto con datos de terceros para identificar correlaciones entre los datos de los usuarios asociados. Las aplicaciones o servicios pueden interactuar con dichos datos y presentarlos a los usuarios en una gran variedad de modalidades, por ejemplo, las notificaciones de oportunidades.

(Mansuri & Sarawagi, 2006) (Ozzie, 2006)

La necesidad de convertir los procesos en componentes más pequeños, requiere un esfuerzo mayor y demasiadas horas para el equipo (Aplicado a equipos de reclutamiento), además si estos procesos se extienden más, tanto a la organización como a los postulados les acarrea consecuencias negativas. La automatización cumple con el propósito de realizar varios (En ocasiones todos) los pasos por defecto, poder gestionar con facilidad y eficacia los datos y pruebas (Test, Entrevistas) de los postulados, reduciendo constantemente el periodo entre la postulación y la etapa final de la selección, posterior contratación. (Saidi Mehrabad & Fathian Brojeny, 2007)

(Faerber, Weitzel, Keim, & Färber, 2003) (L, n.d.)(Universidad del Zulia. Departamento de Ciencias Humanas., Universidad del Zulia. Facultad Experimental de Ciencias., MORENO, & MARTÍN, 2002)(Beddoe & Petrovic, 2006)(Storey Hooper, Galvin, Kilmer, & Liebowitz, 1998)(Young, 1965) (C mercer - J Dunbar, n.d.)

En este punto y con las bases anteriormente expuestas, se integra el machine learning. El aprendizaje autónomo es un fenómeno multifacético dado que incluye la adquisición de nuevos conocimientos, la reorganización de conocimientos previos, para poder aplicar técnicas mediante la experimentación y la observación. Desde los años 50 se ha presentado el fenómeno del aprendizaje autónomo el cual consiste en implantar estas capacidades de aprendizaje en los equipos de cómputo. El estudio y la modelización informática de los procesos de aprendizaje en sus múltiples manifestaciones constituye el tema del aprendizaje automático. En la actualidad, el campo del aprendizaje automático está organizado en torno a tres tipos de estudio:

- Estudios orientados a tareas: el desarrollo y análisis de sistemas de aprendizaje para mejorar el rendimiento en un conjunto predeterminado de tareas (también conocido como "enfoque de ingeniería") (Pablo Felgaer, n.d.)(Piessens, Doncker-Kapenga, Überhuber, & Kahaner, 1983)
- Simulación Cognitiva: la investigación y la simulación por computadora de

procesos de aprendizaje humano.

- Análisis teórico: la exploración teórica del espacio de posibles métodos de aprendizaje y algoritmos independientes del dominio de la aplicación, Aunque muchos esfuerzos de investigación se dedican principalmente hacia uno de estos objetivos, el progreso hacia un objetivo a menudo conduce al progreso hacia otro. Por ejemplo, para investigar el espacio de posibles métodos de aprendizaje.

Considerando el campo del aprendizaje autónomo anteriormente expuesto, se ha generado una investigación bastante amplia sobre los métodos, arrojando las siguientes técnicas:

- Árboles de decisiones
- Algoritmos genéticos
- Redes neuronales artificiales
- Reglas de asociación
- Máquinas de vectores de soporte
- Algoritmos de agrupamiento
- Redes bayesianas
- Análisis de discriminantes lineales
- Regresión en procesos Gaussianos
- k-vecinos más próximos

(Render & Heizer, 2004)(Gutiérrez Martínez & Bello Pérez Andrés Tellería Rodríguez, 2002)

(Estévez & Resumen, n.d.)(Isasi Viñuela & Galván León, 2004)(Valero Orea, Salvador Vargas, & García Alonso, n.d.)(Universidad de Murcia. Facultad de Filosofía y Ciencias de la Educación. Sección de Psicología., Universidad de Murcia. Secretariado de Publicaciones e Intercambio Científico., & SPARC (Organization), 1984)

A continuación, se presenta una comparación y breve historia de cada una de estas técnicas:

#### 1) Árboles de decisiones:

Esta técnica nació de la necesidad de elegir, dado los cursos de acción entre distintas alternativas posibles llegando al denominado proceso decisorio. En muchas oportunidades, los efectos o consecuencias emergentes de una decisión, tomada en el momento actual, desencadenan la necesidad de tomar una nueva decisión, de manera inmediata o en un horizonte temporal cercano. La nueva decisión estará condicionada por los resultados de la alternativa elegida en el presente. Esta técnica en específico se encuentra basada en poder analizar decisiones lineales enfocadas a el uso de resultados y probabilidad asociados, estos se usan en sistemas expertos, arboles de juegos y búsquedas binarias, las principales ventajas son:

- Es de fácil interpretación cuando se llega a una decisión.
- Se reduce considerablemente el número de variables independientes.
- Puede llegar a explicar con detalle el comportamiento ante una sola decisión.
- Es una de las mejores herramientas para el control de gestión empresarial.

Generalmente estos árboles son binarios, esto hace referencia al hecho de que cuenta con 2 opciones, claro está que existen arboles de muchas más opciones tres o más.

#### Árbol binario:

En este método se puede evidenciar que el árbol se encuentra partido en 2 partes y así se subdivide hasta llegar a la respuesta final. (“Árboles de Decisión,” n.d.)

#### Arboles de juego:

Este tipo de árboles son una aplicación de los árboles de decisión y se usan en sistemas expertos o de aprendizaje dado que son más precisos que las decisiones que pudiese a tomar un ser humano desarrollando diagnostico en algo en particular, ya que el hombre puede llegar a dejar detalles en la evaluación, caso contrario del sistema que mediante el árbol explora todas las soluciones y aunque esto puede llegar a ser mucho más lento, la exploración de todas las soluciones compensa esta deficiencia.



Un árbol de decisión también puede usarse para ayudar a crear modelos predictivos automatizados, que puedan emplearse en el aprendizaje automático, la minería de datos y las estadísticas. Conocido como "el aprendizaje basado en árboles de decisión", este método toma en consideración las observaciones sobre un elemento para predecir su valor.(Lucid, n.d.)

En árboles de decisión, los nodos representan datos en lugar de decisiones. Este tipo de árbol también se conoce como "árbol de clasificación". Cada ramificación contiene un conjunto de atributos o reglas de clasificación asociadas a una etiqueta de clase específica, que se halla al final de la ramificación.

Estas reglas, también conocidas como "reglas de decisión", se pueden expresar en una cláusula "Si... entonces...". Cada valor de datos o decisión forma una cláusula, de tal manera que, por ejemplo, "si las condiciones 1, 2 y 3 se cumplen, entonces el resultado X será el resultado definitivo con certeza Y".

Cada dato adicional ayuda a que el modelo prediga de forma más precisa a qué conjunto finito de valores pertenece el asunto en cuestión. Esa información se puede usar posteriormente como una entrada en un modelo más grande de toma de decisiones.(Valero Orea, Salvador Vargas, & García Alonso, n.d.)

A veces la variable predicha será un número real, como un precio. Los árboles de decisión con resultados posibles, infinitos y continuos se llaman "árboles de regresión". Para una mayor precisión, se suelen combinar múltiples árboles con métodos de ensamblado:

Bagging (Ensamblado) crea múltiples árboles de decisión haciendo un remuestreo de los datos de origen, luego hace que los árboles voten para llegar a un consenso.

Un clasificador de selva aleatoria consiste en múltiples árboles diseñados para aumentar la tasa de clasificación.

Los árboles ampliados se pueden usar en árboles de clasificación y de regresión.

Los árboles en una Selva aleatoria (Rotation Forest) se entrenan mediante el uso del análisis de componentes principales (ACP) en una porción aleatoria de los datos.

Un árbol de decisión se considera ideal cuando representa la mayor cantidad de datos con el menor número de niveles o preguntas. Los algoritmos diseñados para crear árboles de decisión optimizados incluyen CART, ASSISTANT, CLS y ID3/4/5. Un árbol de decisión también se puede generar mediante la creación de reglas de asociación, ubicando la variable objetivo a la derecha.

Cada método debe determinar cuál es la mejor forma de dividir los datos en cada nivel. Los métodos comunes para hacerlo incluyen la medición de la impureza de Gini, la obtención de información y la reducción de variaciones.

Emplear los árboles de decisión en el aprendizaje automático tiene numerosas ventajas:

- El costo del uso del árbol para predecir los datos disminuye con cada punto de datos adicional.
- Funciona para los datos numéricos o categóricos.
- Puede modelar problemas con múltiples resultados.
- Usa un modelo de caja blanca (lo que hace que los resultados sean fáciles de explicar).
- La fiabilidad de un árbol se puede cuantificar y poner a prueba.
- Tiende a ser preciso independientemente de si viola las suposiciones de los datos de origen.

Pero también tienen algunas desventajas:

- Cuando se presentan datos categóricos con múltiples niveles, la información obtenida se inclina a favor de los atributos con mayoría de niveles.
- Los cálculos pueden volverse complejos al lidiar con la falta de certezas y numerosos resultados relacionados.
- Las conjunciones entre nodos se limitan a AND, mientras que los gráficos de decisión admiten nódulos relacionados mediante OR.

(Lucid, n.d.)(Frías Blanco, Ortiz Díaz, Ramos Jimenez, Morales Bueno, & Caballero Mota, 2010)(“Árbol de decisión,” n.d.)

## 2) Algoritmos genéticos

Los Algoritmos Genéticos son métodos adaptativos, generalmente usados en problemas de búsqueda y optimización de parámetros, basados en la reproducción sexual y en el principio de supervivencia del más apto (Fogel, IEEE Neural Networks Council., & John Wiley & Sons., 2006). Más formalmente, y siguiendo la definición dada por Goldberg, “los Algoritmos Genéticos son algoritmos de búsqueda basados en la mecánica de selección natural y de la genética natural. Combinan la supervivencia del más apto entre estructuras de secuencias con un intercambio de información estructurado, aunque aleatorizado, para constituir así un algoritmo de búsqueda que tenga algo de las genialidades de las búsquedas humanas” (Jofuku, Schipper, & Goldberg, 1989).

Cualquier solución potencial a un problema puede ser presentada dando valores a una serie de parámetros. Suele hacerse por valores binarios designando una cantidad de bits a cada parámetro luego de ello se realiza una discretización de la variable, según el ajuste y la precisión que se desee obtener se asignan la cantidad de bits, no todas las variables deben ser de la misma cantidad de bits claro está dado que depende es de la codificación que se realice.

### Algoritmo Principal

Los Algoritmos Genéticos trabajan sobre una población de individuos.

Cada uno de ellos representa una posible solución al problema que se desea resolver.

Todo individuo tiene asociado un ajuste de acuerdo a la bondad con respecto al problema de la solución que representa (en la naturaleza el equivalente sería una medida de la eficiencia del individuo en la lucha por los recursos). Una generación se obtiene a partir de la anterior por medio de los operadores de reproducción. Existen 2 tipos:

- Cruce. Se trata de una reproducción de tipo sexual. Se genera una descendencia a partir del mismo número de individuos (generalmente 2)

de la generación anterior. Existen varios tipos que se detallarán en un punto posterior. (Estévez & Resumen, n.d.)

- Copia. Se trata de una reproducción de tipo asexual. Un determinado número de individuos pasa sin sufrir ninguna variación directamente a la siguiente generación.

El funcionamiento genérico de un Algoritmo Genético puede apreciarse

en el pseudocódigo, Si desea optarse por una estrategia elitista, los mejores individuos de cada generación se copian siempre en la población temporal, para evitar su pérdida.

A continuación, comienza a generarse la nueva población en base a la aplicación de los operadores genéticos de cruce y/o copia. Una vez generados los nuevos individuos se realiza la mutación con una probabilidad  $P_m$ . La probabilidad de mutación suele ser muy baja, por lo general entre el 0.5% y el 2%. Se sale de este proceso cuando se alcanza alguno de los criterios de parada fijados. Los más usuales suelen ser:

- Los mejores individuos de la población representan soluciones suficientemente

buenas para el problema que se desea resolver.

- La población ha convergido. Un gen ha convergido cuando el 95% de la población tiene el mismo valor para él, en el caso de trabajar con codificaciones binarias, o valores dentro de un rango especificado en el caso de trabajar con otro tipo de codificaciones. Una vez que todos los genes alcanzan la convergencia se dice que la población ha convergido. Cuando esto ocurre la media de bondad de la población se aproxima a la bondad del mejor individuo.

- Se ha alcanzado el número de generaciones máximo especificado. Sobre este algoritmo inicialmente propuesto por Holland se han definido numerosas variantes. Quizás una de las más extendidas consiste en prescindir de la población temporal de manera que los operadores genéticos de cruce y mutación se aplican directamente sobre la población genética. Con esta variante el proceso de cruces varía ligeramente. Ahora no basta, en el caso de que el cruce se produzca, con insertar directamente la descendencia en la población. Puesto que el número de individuos de la población se ha de mantener constante, antes de insertar la descendencia en la población se le ha de hacer sitio. Es decir, para ubicar a los descendientes generados previamente se han de eliminar otros individuos de la población genética. Existen para ello diversas opciones, que se comentarán con más detalle en un punto posterior. Evidentemente, trabajando con una única población no se puede identificar las características de la generación cuando se llene la población, pues siempre está llena. En este caso el paso a la siguiente generación se producirá una vez que se hayan alcanzado cierto número de cruces y mutaciones. Este número dependerá de la tasa de cruces y mutaciones especificadas por el usuario y del tamaño de la población. Así, con una tasa de cruces del 90%, una tasa de 18 Introducción a los Algoritmos Genéticos y la Programación Genética mutaciones del 2% y trabajando con 100 individuos se pasará a la siguiente generación cuando se alcanzasen 45 cruces (cada cruce genera 2 individuos con lo que se habrían insertado en la población 90 individuos, esto es el 90%) o 2 mutaciones. Otra variación común consiste en la modificación del esquema de selección de los individuos que serán mutados. En el esquema mostrado, sólo los descendientes originados a partir de un cruce son mutados (proceso que imita los errores de transcripción del ADN que tienen lugar en la naturaleza); otra opción habitual es la selección aleatoria del individuo a mutar entre todos los que forman parte de la población genética. (Marcos Gestal- Daniel Rivero - Alejandro Pazos, n.d.)(Yolis, Yolis, Britos, Perichisky, & García-Martínez, 2003)(Morales & Casas, n.d.) (Estévez & Resumen, n.d.)

### 3) Redes neuronales artificiales

Una Red Neuronal Artificial es un modelo matemático vagamente inspirado en el comportamiento biológico de las neuronas y en la estructura del cerebro, y que es utilizada para resolver un amplio rango de problemas. Debido a su flexividad, una única red neuronal es capaz de realizar diversas tareas. (Tablada & Torres, 2009)

Las Redes Neuronales Artificiales, ANN (Artificial Neural Networks) están inspiradas en las redes neuronales biológicas del cerebro humano. Están constituidas por elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes. Estos elementos están organizados de una forma parecida a la que presenta el cerebro humano. Las ANN al margen de "parecerse" al cerebro presentan una serie de características propias del cerebro. Por ejemplo las ANN aprenden de la experiencia, generalizan de ejemplos previos a ejemplos nuevos y abstraen las características principales de una serie de datos.

**Aprender:** adquirir el conocimiento de una cosa por medio del estudio, ejercicio o experiencia. Las ANN pueden cambiar su comportamiento en función del entorno. Se les muestra un conjunto de entradas y ellas mismas se ajustan para producir unas salidas consistentes.

**Generalizar:** extender o ampliar una cosa. Las ANN generalizan automáticamente debido a su propia estructura y naturaleza. Estas redes pueden ofrecer, dentro de un margen, respuestas correctas a entradas que presentan pequeñas variaciones debido a los efectos de ruido o distorsión. **Abstraer:** aislar mentalmente o considerar por separado las cualidades de un objeto. Algunas ANN son capaces de abstraer la esencia de un conjunto de entradas que aparentemente no presentan aspectos comunes o relativos. (Pablo Felgaer, n.d.) (Tablada & Torres, 2009). En las Redes Neuronales Artificiales, ANN, la unidad análoga a la neurona biológica es el elemento procesador, E (process element). Un elemento procesador tiene varias entradas y las combina, normalmente con una suma básica. La suma de las entradas es modificada por una función de transferencia y el valor de la salida de esta función de transferencia se pasa directamente a la salida del elemento procesador. La salida del PE se puede conectar a las entradas de otras neuronas artificiales (PE) mediante conexiones ponderadas correspondientes a la eficacia de la sinapsis de las conexiones neuronales. (Izaurieta & Saavedra, n.d.) Una red neuronal consiste en un conjunto de unidades elementales PE conectadas de una forma concreta. El interés de las ANN no reside solamente en el modelo del elemento PE sino en las formas en que se conectan estos elementos procesadores. Generalmente los elementos PE están organizados en grupos llamados niveles o capas. Una red típica consiste en una secuencia de capas con conexiones entre capas adyacentes consecutivas.

Existen dos capas con conexiones con el mundo exterior. Una capa de entrada, buffer de entrada, donde se presentan los datos a la red, y una capa buffer de salida que mantiene la respuesta de la red a una entrada. El resto de las capas reciben el nombre de capas ocultas.

Recientemente diversos trabajos (Pitarque, Roy, & Ruiz, 1998) han analizado las relaciones entre redes neurales y técnicas estadísticas convencionales. Aunque los resultados no son concluyentes, en general las redes neurales han mostrado una capacidad clasificatoria igual o superior que las técnicas estadísticas, con la ventaja de poderse utilizar independientemente del cumplimiento de los supuestos teóricos relativos a estas técnicas (de ahí que se haya hablado de ellas como de "técnicas no paramétricas"). Se presentan una serie de simulaciones y aplicaciones sobre datos reales que apoyan esta idea. En concreto redes neurales del tipo perceptrón multi-capas contra modelos de regresión logística y análisis discriminante en tareas de clasificación, manipulando los patrones de correlación existentes entre los predictores (o variables de entrada) por un lado, y entre predictores con el criterio (variable de salida) por otro. Los resultados muestran que las redes neurales clasifican mejor que las técnicas estadísticas, incluso en la condición a priori más favorable a éstas, es decir, cuando existen altas correlaciones entre los predictores con el criterio pero bajas correlaciones entre los predictores. Un patrón de resultados similares aparece en tareas de clasificación con datos reales. Se discuten los resultados en el marco de la polémica redes neurales vs modelos estadísticos convencionales (ver p.e. Sarle, 1999), valorando si vale la pena ganar en capacidad clasificatoria a costa de incrementar el costo computacional y de recursos involucrados en el entrenamiento de las redes neurales. (Pitarque, Ruiz, & Roy, 2000)(Hilera González & Martínez Hernando, 2000)("Utilización de las Redes Neuronales en la toma de decisiones: aplicación en un problema de secuenciación," n.d.)

#### 4) Reglas de asociación

Las reglas de asociación son usadas para descubrir hechos que ocurren de manera común dentro de un conjunto de datos, se ha investigado muy ampliamente métodos para aprendizaje de reglas de asociación relevando asociaciones entre conjuntos de datos

(Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004).

Una regla de asociación consta de un conjunto de ítems, el cuerpo de regla, que conducen a otro ítem, la cabecera de regla. La regla de asociación relaciona el

cuerpo de regla con la cabecera de regla. Una regla de asociación puede contener las características siguientes:

- Información estadística sobre la frecuencia de aparición
- Fiabilidad
- Importancia de esta relación

En el ejemplo siguiente, los bañadores y las toallas de playa representan el cuerpo de regla. Las gafas de sol representan la cabecera de regla. Soporte y Confianza son medidas de la fiabilidad y la frecuencia de aparición.

Ejemplo

[bañadores] + [toallas de playa] ==> [gafas de sol] Soporte=24%

Confianza=60% Elevación=2.0

La regla del ejemplo anterior significa que, si los clientes compran bañadores y toallas de playa, en el 60% de los casos también compran gafas de sol. La combinación compuesta por bañadores, toallas de playa y gafas de sol aparece en el 24% de todas las transacciones.

Éstas son las características de una regla de asociación:

- Soporte

Una transacción da soporte a una regla de asociación si la transacción contiene el cuerpo de regla y la cabecera de regla. El soporte de regla es la proporción de transacciones que dan soporte a la regla de asociación con respecto al número total de transacciones de la base de datos de transacciones.

En el ejemplo anterior, puede haber 24 transacciones de 100 que den soporte a la regla de asociación. Esto significa que 24 transacciones de 100 contienen, como mínimo, el ítem bañadores, toallas de playa o gafas de sol.

- Soporte absoluto

El soporte absoluto de regla es el número de transacciones que dan soporte a la regla de asociación en la base de datos de transacciones.

- Confianza

La confianza de una regla de asociación es su firmeza o fiabilidad. La confianza se define como el porcentaje de transacciones que dan soporte a la regla con respecto

a todas las transacciones que dan soporte al cuerpo de regla. Una transacción da soporte al cuerpo de regla si contiene todos los ítems del cuerpo de regla.

En el ejemplo anterior, la confianza es de un 60%. Esto significa que el 60% de todas las transacciones que contienen bañadores y toallas de playa también contiene gafas de sol.

#### - Elevación

El valor de elevación de una regla de asociación es el factor por el cual la confianza sobrepasa la confianza esperada. Se determina dividiendo la confianza de la regla por el soporte de la cabecera de regla.

Por ejemplo, si la regla de asociación [bañadores] + [toallas de playa] => [gafas de sol] tiene 2,0 como factor de elevación, los bañadores y las toallas de playa tienen un efecto muy positivo en la compra de gafas de sol puesto que un factor de elevación alto indica una firme asociación entre los ítems.

#### - Soporte multiplicado por Confianza

La medida soporte de regla multiplicado por confianza de regla le ayuda a identificar reglas que pueden ser importantes. Tiene en cuenta el valor de confianza y el valor de soporte. Si el valor de confianza y el valor de soporte son altos, la medida soporte de regla multiplicado por confianza de regla también es alta.

En la Vista Reglas, puede ordenar las reglas por los valores correspondientes a la medida soporte de regla multiplicado por confianza de regla.

#### - Grupo de reglas

Los grupos de reglas le ayudan a distinguir conjuntos de reglas que no tienen una relación directa o indirecta.

Si dos reglas, por ejemplo R1 y R2, comparten al menos un ítem (independientemente de si el ítem aparece en la cabecera de regla o en el cuerpo de regla), estas reglas están directamente relacionadas ( $R1 \sim R2$ ) y por lo tanto pertenecen al mismo grupo de reglas.

El grupo de reglas de una determinada regla R1 consta de todas las reglas del modelo que están directa o indirectamente relacionadas con R1. Esto significa que un grupo de reglas contiene todas las reglas que están conectadas mediante una cadena de reglas directamente relacionadas en el modelo.

(González & Morales, n.d.)(Delgado, Ruiz, & Sánchez, 2008)



## 5) Máquinas de vectores de soporte

Una Máquina de Soporte Vectorial (SVM) aprende la superficie de decisión de dos clases distintas de los puntos de entrada. Como un clasificador de una sola clase, la descripción dada por los datos de los vectores de soporte es capaz de formar una frontera de decisión alrededor del dominio de los datos de aprendizaje con muy poco o ningún conocimiento de los datos fuera de esta frontera. Los datos son mapeados por medio de un kernel Gaussiano u otro tipo de kernel a un espacio de características en un espacio dimensional más alto, donde se busca la máxima separación entre clases. Esta función de frontera, cuando es traída de regreso al espacio de entrada, puede separar los datos en todas las clases distintas, cada una formando un agrupamiento. (Universidad Tecnológica de Pereira., 2005) Las máquinas de vectores soporte (SVM, del inglés Support Vector Machines) tienen su origen en los trabajos sobre la teoría del aprendizaje estadístico y fueron introducidas en los años 90 por Vapnik y sus colaboradores (Cortes & Vapnik, 1995). Aunque originariamente las SVMs fueron pensadas para resolver problemas de clasificación binaria, actualmente se utilizan para resolver otros tipos de problemas (regresión, agrupamiento, multclasificación). También son diversos los campos en los que han sido utilizadas con éxito, tales como visión artificial, reconocimiento de caracteres, categorización de texto e hipertexto, clasificación de proteínas, procesamiento de lenguaje natural, análisis de series temporales. De hecho, desde su introducción, han ido ganando un merecido reconocimiento gracias a sus sólidos fundamentos teóricos. Dentro de la tarea de clasificación, las SVMs pertenecen a la categoría de los clasificadores lineales, puesto que inducen separadores lineales o hiperplanos, ya sea en el espacio original de los ejemplos de entrada, si éstos son separables o cuasi-separables (ruido), o en un espacio transformado (espacio de características), si los ejemplos no son separables linealmente en el espacio original. Como se verá más adelante, la búsqueda del hiperplano de separación en 1 estos espacios transformados, normalmente de muy alta dimensión, se hará de forma implícita utilizando las denominadas funciones kernel. Mientras la mayoría de los métodos de aprendizaje se centran en minimizar los errores cometidos por el modelo generado a partir de los ejemplos de entrenamiento (error empírico), el sesgo inductivo asociado a las SVMs radica en la minimización del denominado riesgo estructural.

La idea es seleccionar un hiperplano de separación que equidista de los ejemplos más cercanos de cada clase para, de esta forma, conseguir lo que se denomina un margen máximo a cada lado del hiperplano. Además, a la hora de definir el hiperplano, sólo se consideran los ejemplos de entrenamiento de cada clase que caen justo en la frontera de dichos márgenes. Estos ejemplos reciben el nombre de vectores soporte. Desde un punto de vista práctico, el hiperplano separador de margen máximo ha demostrado tener una buena capacidad de generalización, evitando en gran medida el problema del sobreajuste a los ejemplos de

entrenamiento. Desde un punto de vista algorítmico, el problema de optimización del margen geométrico representa un problema de optimización cuadrático con restricciones lineales que puede ser resuelto mediante técnicas estándar de programación cuadrática. Las propiedades de convexidad exigida para su resolución garantizan una solución única, en contraste con la no unicidad de la solución producida por una red neuronal artificial entrenada con un mismo conjunto de ejemplos.(Fernando et al., 2011)(Argañaraz & Entraigas, n.d.)

## 6) Algoritmos de agrupamiento

Debido al desarrollo alcanzado con los microprocesadores se pueden manejar grandes volúmenes de datos como puntos en espacios de una alta dimensionalidad, éstos aparecen en varias esferas de la vida real, en bases de datos de corporaciones financieras, telecomunicaciones, medicina, imágenes de satélite, etc., numerosas son las aplicaciones en las que se requiere del manejo de bases de datos espaciales con el objetivo de conocer acerca de la identificación de grupos, para descubrir importantes distribuciones del espacio en estudio, lo cual puede ser resuelto con el empleo de algún algoritmo de agrupamiento conveniente, por lo que el estudio, aplicación y creación de nuevos algoritmos constituye un desafío importante en la actualidad.

El Reconocimiento de Formas constituye un amplio conjunto de técnicas para el tratamiento de datos entre las que se puede mencionar: la selección y extracción de características, la clasificación de un objeto en un grupo dado y la división de los datos en grupos (agrupamiento). Uno de los enfoques en el Reconocimiento de Formas, en función del tipo de espacio de representación utilizado es el Reconocimiento Estadístico de Formas que se apoya en la Teoría de Decisión, éste asume que el espacio de representación tiene una estructura de espacio vectorial y/o métrico y no se supone ninguna relación estructural entre las distintas características. Dentro de este enfoque, se distingue entre las aproximaciones paramétrica y no paramétrica. En el primer caso, se asume un conocimiento a priori sobre la forma funcional de las distribuciones de probabilidad de cada clase sobre el espacio de representación, las cuales vendrán determinadas por un conjunto finito y normalmente fijo de parámetros, las fronteras de decisión estarán definidas por dichas distribuciones de clases. La aproximación no paramétrica no supone ninguna forma de las distribuciones de probabilidad sobre el espacio de representación, de modo que el único conocimiento a priori será el correspondiente a la información inducida a partir de un conjunto de muestras, las fronteras de decisión estarán determinadas por las muestras del conjunto de entrenamiento. En este trabajo nos interesa el Reconocimiento Estadístico de Formas no paramétrico, pero haremos

una pequeña alusión a la aproximación paramétrica con el estudio de las mixturas finitas. El problema del agrupamiento puede definirse como sigue: dados  $n$  puntos en un espacio  $n$ -dimensional particionar los mismos en  $k$  grupos tales que los puntos dentro de un grupo son más similares que cada uno a los de los otros grupos, dicha similaridad se mide atendiendo a alguna función distancia (función de disimilaridad) o alguna función de similaridad. También es importante para el mejor funcionamiento de los algoritmos de agrupamiento la detección de ruido para evitar la influencia negativa de éstos en la formación de los grupos, así como la estimación del número correcto de grupos a determinar. (Pla Bañón José Salvador Sánchez Garreta, n.d.)a

Los métodos de agrupamiento no paramétricos pueden dividirse en tres grupos fundamentales: jerárquicos, particionales y basados en densidad. (Pascual, Pla, & Sánchez, n.d.)(Pla Bañón José Salvador Sánchez Garreta, n.d.)(J. Contreras, Member IEEE, n.d.)

## 7) Redes bayesianas

Un criterio que se ha utilizado para describir la estructura de una red bayesiana atiende al grado en que sus elementos son “visibles” para el usuario (Edwards & Fasolo, 2001). En este sentido una red bayesiana sería una estructura compuesta por cuatro niveles. En el nivel superior, una red bayesiana sería un conjunto de variables representadas por nodos y un conjunto de flechas que relacionan estas variables en términos de influencia. En un nivel inferior estarían los niveles o estados, también conocidos como espacio de estados (“Linking Bayesian networks and PLS path modeling for causal analysis,” 2010), que pueden asumir cada una de las variables del modelo. En tercer lugar, tendríamos un conjunto de funciones de probabilidad condicional, una para cada nodo, donde se representaría la probabilidad de ocurrencia de cada estado de la variable condicionado a los posibles valores de las variables que determinan el valor de la variable. Por último, en el nivel más subordinado estarían un conjunto de algoritmos que permitirían que la red recalculase las probabilidades asignadas a cada uno de sus niveles cuando conocemos alguna evidencia sobre el modelo. Sin embargo, la descripción más frecuente de una red bayesiana se basa en dos elementos, una dimensión cualitativa y otra cuantitativa (García García et al., n.d.) Los fundamentos teóricos de estos dos elementos o dimensiones se sustentan en dos grandes pilares de la modelización matemática, la teoría de grafos y la teoría de la probabilidad. (Sucar, n.d.)(Tuya, Ramos Román, & Dolado Cosín, 2007)(Jorge López Puga<sup>1\*\*</sup>, Juan García García<sup>1\*\*</sup> & SPARC (Organization), 1984)

Las redes bayesianas híbridas son un caso particular de las redes bayesianas en el que las variables aleatorias no toman únicamente valores discretos o continuos, sino que existen variables de ambos tipos. De esta manera se puede trabajar conjuntamente dentro del modelo con datos cuantitativos y cualitativos. Por lo general, no es posible realizar inferencia exacta cuando se usa cualquier tipo de distribuciones de probabilidad condicional, siendo necesario recurrir a algún tipo de aproximación. Una primera opción para paliar este problema es discretizar los valores continuos y trabajar con una red bayesiana en la que todos sus estados son discretos. Otra opción, propuesta por Lauritzen para los modelos híbridos, es asumir que la distribución condicional de las variables continuas dadas las discretas sea una Gaussiana multivalente.

Esta última aproximación es la que se describe en este capítulo y la que se usa en esta memoria. (Jaime Regueró Alvarez Directora Julia Díaz García, n.d.) (Jorge López Puga<sup>1\*\*</sup>, Juan García García<sup>1\*\*</sup> & SPARC (Organization), 1984)

## 8) Análisis de discriminantes lineales

Supongamos que un conjunto de objetos se clasifica en una serie de grupos; el Análisis Discriminante equivale a un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y Las variables independientes son continuas y determinan a qué grupos pertenecen los objetos. Se trata de encontrar relaciones lineales entre las variables continuas que mejor discriminen en los grupos dados a los objetos. Además, se trata de definir una regla de decisión que asigne un objeto nuevo, que no sabemos clasificar previamente, a uno de los grupos prefijados.

Se presentan una serie de restricciones o supuestos:

- (i) Se tiene una variable categórica y el resto de variables son de intervalo o de razón y son independientes respecto de ella.
- (ii) Es necesario que existan al menos dos grupos y para cada grupo se necesitan dos o más casos.
- (iii) El número de variables discriminantes debe ser menor que el número de objetos menos dos:  $X_1, \dots, X_p$ , donde  $p < (n - 2)$  y  $n$  es el número de objetos.
- (iv) Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.

(v) El número máximo de funciones discriminantes es igual al mínimo entre el número de variables y el número de grupos menos 1 (con  $q$  grupos,  $(q - 1)$  funciones discriminantes).

(vi) Las matrices de covarianzas dentro de cada grupo deben ser aproximadamente iguales.

(vii) Las variables continuas deben seguir una distribución normal multivariante

(SPARC (Organization), PATRICIA; TREJOS, ALVARO; SOTO MEJÍA & Universidad Tecnológica de Pereira., 1995)(Arellano, Gil, & Martínez, 2014)(Peña, 2002)

## 9) Regresión en procesos Gaussianos

- Estas distribuciones nos cuantifican nuestra incertidumbre en los modelos.
- Los procesos gaussianos se definen como una distribución de probabilidad sobre funciones aleatorias.
- De hecho son sobre colecciones infinitas de variables (funciones), tal que cualquier subconjunto de variables aleatoria finita tiene una distribución gaussiana  $n$ -multivariable.
- Los procesos gaussianos se empezaron a estudiar para predicción de series de tiempo en los 40's (Foster, 1961)
- Mas adelante (70's) se usaron en geo estadística y meteorología, en lo que se llamó kriging usando 2 o 3 dimensiones.
- Mas adelante en estadística espacial
- También en regresión: (Ver Hoef & Cressie, 1993)
- Se realizaron experimentos en computadoras (sin ruido) a finales de los 80's (Sacks, Welch, Mitchell, & Wynn, n.d.)
- Se popularizaron recientemente en aprendizaje computacional por Williams y Rasmussen (1996) y Neal (1996) para resolver problemas de regresión.  $n$  (Guevara & Vargas, n.d.)(Isaza, Rios, & Mosquera, 2008)(Baumgartner, Rivero, & Pucheta, n.d.)

## 10) k-vecinos más próximos

El algoritmo K-NN fue diseñado y desarrollado por Fix y Hodges en 1951, como método de clasificación supervisada, el cual realiza un aprendizaje supervisado basado en un conjunto de entrenamiento y prototipos. Es uno de los algoritmos más simples de los existentes y muestra la esencia del aprendizaje basado en distancias.

El método K Nearest Neighbors ( KNN ) tiene como objetivo categorizar los puntos de consulta cuya clase es desconocida dado sus respectivas distancias a puntos en un conjunto de aprendizaje (es decir, cuya clase se conoce a priori). Es una de las herramientas de aprendizaje automático supervisadas más populares. Una versión simple de KNN se puede considerar como una extensión del método vecino más cercano (el método NN es un caso especial de KNN,  $k = 1$ ).

El enfoque de clasificación KNN supone que cada ejemplo en el conjunto de aprendizaje es un vector aleatorio en  $R^n$ . Cada punto se describe como  $x = \langle a_1(x), a_2(x), a_3(x), \dots, a_n(x) \rangle$  donde  $a_r(x)$  denota el valor  $r$  del atributo  $r$ th.  $a_r(x)$  puede ser una variable cuantitativa o cualitativa.

Para determinar la clase del punto de consulta  $X_q$ , cada uno de los  $k$  puntos más cercanos  $X_1, \dots, X_k$  a  $X_q$  procede a votar. La clase de  $X_q$  corresponde a la clase de la mayoría.

Distancias: se pueden usar varias medidas de distancia para calcular similitudes en el algoritmo de K Nearest Neighbors . Las opciones varían de acuerdo con el tipo de variables que caracterizan las observaciones (cualitativas o cuantitativas).

Distancias disponibles para datos cuantitativos (métricas): Euclidiana que es un caso de Minkowski, Manhattan, Tchebychev, Canberra

Distancias disponibles para datos cuantitativos (kernels): lineales, sigmoideas, logarítmicas, potencia, gaussianas, laplacianas

Distancias disponibles para datos cualitativos: Superposición métrica (OM), métrica de diferencia de valor (VDM)(Tan, Chen, Zhou, & Zhang, 2005)

(Cover & Hart, 1967)

De la misma forma se ha llevado a cabo un estudio de las principales técnicas usadas para la selección de personal, llegando a tener como resultado los principales test usados y la evaluación de cada uno de ellos para seleccionar el que más se adapta a el problema específico planteado.

Usualmente, la cantidad de postulados a un cargo específico, aunque puede parecer alta en número, en realidad los perfiles son muy distantes al cargo siendo necesario que la persona que se encuentre en el área de selección tenga una gran capacidad para identificar aquellos que más puedan aportar. Cuando se realiza la entrevista se deben tener en cuenta varios aspectos dentro de la misma, incluso la manera de realizar las preguntas a el postulado interfiere con las respuestas que pueda llegar a dar el mismo y para una óptima entrevista se tienen en cuenta puntos fundamentales sobre las preguntas: las preguntas que sean siniestradas deben entenderse de la manera más fácil posible, solo se puede realizar una pregunta una sola vez, evitar que las preguntas condicionen las respuestas, las preguntas do deben inducir a eludir una respuesta o a adoptar una actitud negativa.(Ma Jesús Montes Alonso y Pablo González Rodríguez, n.d.) Existen distintos tipos de preguntas para este proceso:

- Preguntas de sondeo

Son preguntas cortas y bastante sencillas tales como: ¿por qué?, ¿Cuál fue la causa?, etcétera.

- Preguntas cerradas

Estas preguntas pueden ser contestadas por una sola palabra, por lo general, sí no. De esta respuesta suele derivar otra pregunta.

- Preguntas hipotéticas

En estas preguntas se le presentan situaciones hipotéticas a el postulado, ¿Qué haría usted si...?, ¿Cómo manejaría usted...?. Estas preguntas son relativamente peligrosas esto indica que se debe prestar bastante intención al momento de formularlas, aunque en casos específicos suele ser de mayor utilidad un ejemplo claro de ello es cuando se desea evaluar conocimientos específicos de un cargo procesos, etc.

- Preguntas malintencionadas

Obligan a el postulado a tomar uno de los dos caminos y se tiene claro desde el inicio que los dos caminos son malas opciones. Estas suelen ser poco recomendadas.

- Preguntas provocadoras

El propósito de estas preguntas es desafiar a el postulado e forma tal que se pueda evaluar su reacción, en ellas como factor fundamental tiene que ir la sorpresa es decir se formulan en el momento menos propicio.

- Preguntas que sugieren la respuesta esperada

Aquellas donde queda clara la respuesta que el postulado dara.

- Preguntas abiertas

Inducen a el entrevistado a explayarse sobre el tema y permiten obtener mucha información y evaluar otros aspectos de su desempeño, como lo son expresión natural o uso del lenguaje. (Alles, 2003)

En las ciencias fácticas se miden las características de los objetos que estudian.

La psicología es una ciencia fáctica y para medir los atributos o características psicológicas del ser humano utiliza como instrumentos a los tests. (Guzmán Evangelista, 2017) Estos tests pueden ser proyectivos o psicométricos, los psicométricos se usan explícitamente para medición un ejemplo claro de ello es el primer test psicométrico creado el cual fue la escala métrica de la inteligencia creada por 2 franceses llamados Binet y Simon en el año de (1905). (Antonio, Mérida, Luis, & Jorge, n.d.) con la que se introdujo en psicología el concepto de edad mental. Uniendo este concepto con el de edad cronológica, el psicólogo alemán William Stern (1912) creó el concepto de Cociente Intelectual o CI.(Pardon & Stern, 2001) Por otro lado, el Cuestionario de Datos de Personalidad de Robert S. Woodsworth .(Woodworth & Thorndike, 1901) es reconocido como el primer test de personalidad.

Los tests de personalidad son herramientas que permiten evaluar rasgos psicológicos y de la personalidad de un individuo. Conocidos porque se utilizan en el ámbito de la psicología clínica, también se utilizan en los procesos de selección para contrastar si la información aportada en la entrevista es acorde al test.(Parra et al., 2005) Cada test de personalidad mide unos determinados rasgos. Un mismo rasgo puede ser nombrado de forma diferente según el test. Algunos de los más usuales son liderazgo, trabajo en equipo, autonomía, orientación a resultados, orientación al logro o flexibilidad. Las preguntas pueden estar planteadas de múltiples maneras y en diferentes escalas, desde preguntas con dos opciones de



respuesta, “verdadero – falso”, a preguntas que tienen una escala que oscila entre “muy frecuentemente – nunca”. Dos de los tests que se utilizan habitualmente en selección de personal son el KOSTICK PAPI o 16 PF.(Ferrando, 1989)

Dado lo anterior se han tomado los 3 principales test usados en el área de selección de personal para realizar un estudio comparativo dentro de ellos.

#### 1) 16 PF.

Esta es una prueba usada por investigadores alrededor del mundo, como instrumento de medida y comprensión de la personalidad. Este test es el resultado de más de 40 años de investigación por el psicólogo Raymond B. Cattell, quien redujo 4504 rasgos de personalidad a 16 rasgos básicos, mediante análisis factorial.(AMerican psychological Association, n.d.) La prueba se encuentra basada en varios factores psicológicos medidos independientemente, hoy en día existen varias formas (A,B,C,D & E) esto facilitando la aplicación de la prueba según la necesidad específica.

La validez de esta prueba en particular depende directamente de una evidencia previa, dado que esto apoya de sobremanera las inferencias que arroje basado en puntuaciones obtenidas de los factores básicos los cuales son:

El Factor A mide el carácter gregario del individuo; el grado en que la persona busca establecer contacto con otras personas porque encuentra satisfactorio y gratificante el relacionarse con éstas. Las personas que obtienen puntuaciones altas (A+) tienen una mayor disposición hacia el afecto, tienden a ser más cariñosos, expresivos, dispuestos a cooperar, generosos, activos, y no temen a las críticas que puedan hacerse de su persona. Los A+ prefieren los proyectos grupales en vez de la competencia a nivel individual, y disfrutan de empleos que enfatizan la interacción social como las ventas, el trabajo social o la enseñanza. Los A- tienden a ser más reservados, formales, impersonales y escépticos. Prefieren trabajar solos y son rígidos y precisos al hacer sus cosas; pueden ser, en ocasiones, altamente críticos y rudos.

El Factor B mide inteligencia en base al predominio del pensamiento abstracto o del pensamiento concreto; considerando el predominio del abstracto como característico de una persona de inteligencia mayor y el concreto como indicador de una inteligencia menor. Una persona con puntuaciones altas (B+) demuestra tener un pensamiento abstracto y se le percibe como muy inteligente. Puede captar, analizar y comprender rápidamente y con facilidad las ideas o conceptos que se le

presenten; y tienden a ser muy alertas. Los que obtienen puntuaciones bajas (B-) tienden a interpretar la mayoría de las cosas de manera literal y concreta. Tienen dificultades para comprender conceptos y para el aprendizaje en general. Se les describe como lentos al reaccionar y de baja inteligencia.

El Factor C está relacionado a la estabilidad emocional de la persona y a la manera en que se adapta al ambiente que le rodea; determina específicamente la fortaleza de ego. Puntuaciones altas (C+) son características de personas realistas y estables emocionalmente. Se les considera maduros, con una alta fortaleza de ego; y se les adjudica una capacidad para mantener unida la moral de un grupo. Los que obtienen puntuaciones bajas (C-) son por lo general personas que se frustran rápidamente bajo condiciones no-satisfactorias, tienden a evadir la realidad y tienen una fortaleza de ego muy baja.

Estas personas se ven afectadas fácilmente por los sentimientos, son neuróticos, la mayor parte del tiempo se encuentran insatisfechos; y tienden a padecer de fobias, dificultades al dormir y problemas de tipo psicossomático.

El Factor E mide el grado de control que tiende a poseer la persona en sus relaciones con otros seres humanos; se determina en términos de si es dominante o es sumiso. Puntuaciones altas (E+) indican que la persona es muy dominante. A este tipo de persona le resulta muy agradable y atractivo el estar en posiciones de poder para controlar y criticar a otros. Son agresivos, competitivos, tercos, asertivos y muy seguros de sí mismos. Tienden a ser muy autoritarios con otros y no se someten a la autoridad. Las personas que obtienen puntuaciones bajas (E-) tienden a ser sumisos, humildes y dóciles. Se dejan llevar fácilmente por otros, son conformistas, pasivos y considerados. Debido a que les interesa evitar los conflictos en sus relaciones interpersonales, se esfuerzan en complacer y en ganarse la aprobación de los demás.

El Factor F está relacionado al nivel de entusiasmo evidente en contextos sociales. Las personas con puntuaciones altas (F+) tienden a ser altamente entusiastas, espontáneas, expresivas y alegres. Estos individuos son muy francos, impulsivos y mercuriales. Con frecuencia salen electos como líderes. Puntuaciones bajas (F-) son características de personas más sobrias, prudentes, serias y taciturnas. Estas personas son introspectivas, restringidas y, por lo general, pesimistas.

El Factor G mide la internalización de los valores morales; estructuralmente se explora el superego según descrito por Freud. Altas puntuaciones en este factor (G+) son representativas de personas altamente moralistas, conformistas, responsables y concienzudas que tienden a actuar siempre de acuerdo a las reglas. Los que obtienen puntuaciones bajas (G-) son personas que no se comportan de acuerdo a las reglas, ni se someten por completo a las normas de la sociedad o de su cultura. Su necesidad de logro se percibe como baja o ninguna, pero esto no implica que no sean productivos. El que la persona salga bajo (G-) en este factor no indica que ésta no experimente la presencia de su superego como una fuerza o agencia interna de control; sino que responde a una serie de valores distintos a los que establece la sociedad.

El Factor H mide la reactividad del sistema nervioso en base a la tendencia en la persona de un dominio parasimpático o simpático. En las personas que obtienen puntuaciones altas en este factor (H+) resulta dominante el sistema parasimpático. Estas personas son capaces de funcionar bajo altos niveles de estrés, ignoran las señales que indiquen o presagien peligros externos, les encanta correr riesgos y disfrutan del éxtasis que les produce el ser aventureros. Puntuaciones bajas en este factor (H-) son características de personas bajo el dominio de su sistema simpático. Estas personas tienden a reaccionar de manera exagerada a cualquier percepción de posible amenaza. Los H- se limitan a lo seguro, predecible y estable para evitar situaciones o estímulos que puedan alterar su delicada homeostasis interna.

El Factor I se utiliza para medir el predominio, ya sea de los sentimientos o del pensamiento racional, en la persona en su toma de decisiones al conducirse en su diario vivir. Los que obtienen puntuaciones altas (I+) funcionan bajo el dominio de sus sentimientos. Estas personas tienden a ser muy emotivas y de una sensibilidad extrema. Se les puede describir como distraídos, soñadores, intuitivos, impacientes, temperamentales y, por lo general, no son muy realistas. Los (I-) se rigen por su pensamiento racional, siendo muy prácticos y realistas. Estos individuos son independientes, responsables, escépticos y, en ocasiones, pueden resultar cínicos y rudos.(Howarth, E., & Browne, n.d.)

El Factor L explora la identidad social del individuo; específicamente mide en qué grado la persona se siente identificado o unido a la raza humana en general. Los que puntúan alto (L+) poseen unas fronteras personales tan marcadas que se desconectan del resto de la humanidad. Estas personas tienden a desconfiar de los demás, y se ha encontrado que su comportamiento tiende a ser paranoico. Por estas razones, las relaciones interpersonales de los L+ son generalmente problemáticas, deteriorándose por el exceso de celos, sospechas y el escepticismo

de estos individuos. Las personas que obtienen bajas puntuaciones (L-) se caracterizan primordialmente por sentirse uno con los demás. Estas personas no se perciben como un mundo aparte, sino como parte de un mundo compuesto por toda la humanidad. Los L- son personas que confían en los demás, se adaptan fácilmente, se preocupan por sus compañeros, son abiertos, tolerantes y muy poco competitivos. (American Psychological Association, n.d.)

El Factor M se basa en que los humanos pueden percibir de dos modos. La primera manera de percibir se nutre del contacto directo entre los cinco sentidos y el ambiente. La otra forma se compone mayormente de un diseño interno de conexión subliminal de pensamientos y especulaciones que van organizando la información. Las personas con altas puntuaciones (M+) se caracterizan por una intensa vida interna. Estas personas van por el mundo sumergidas en sus pensamientos, distraídas e inatentas a lo que sucede a su alrededor. Son seres con una gran imaginación, muy creativos, poco convencionales e interesados sólo en la esencia de las cosas. Los que salen bajos (M-) responden al mundo externo en vez de al interno. Estas personas son muy realistas y prácticas; valoran lo concreto y lo obvio. En términos de creatividad, los M- poseen niveles muy bajos o inexistentes. En casos de emergencia, tienden a mantener la calma y son capaces de resolver la situación.

El Factor N está relacionado a las máscaras sociales; describe en que grado las personas se ocultan, mostrando sólo aquellos rasgos que generen las respuestas que desean obtener de los demás. Las personas que obtienen puntuaciones altas (N+) tienden a ser calculadoras, refinadas, diplomáticas y muy conscientes socialmente. Los N+ se pueden describir como utilitaristas; usan sus destrezas sociales para relacionarse con personas a las que les puedan sacar provecho para realizar sus planes. Los que puntúan bajo (N-) son personas genuinas, abiertas, directas y sinceras que no se esfuerzan por impresionar a otros. Estos individuos son muy espontáneos y auténticos; si quieren algo, lo piden sin incurrir en planes elaborados de interacciones humanas.

El Factor O explora la auto-estima de las personas en base a tendencias a experimentar culpa o inseguridades. Este factor no pretende categorizar a las personas entre altas y bajas auto-estimas ya que el nivel al momento de la prueba puede ser uno de carácter transitorio, influenciado por eventos recientes. Altas puntuaciones (O+) son obtenidas por personas cuya vida interna se rige por el sufrimiento.

Estas personas tienen expectativas personales muy altas, se preocupan demasiado, experimentan muchos sentimientos de culpa, son inseguros y no se sienten aceptados en situaciones grupales. Los que puntúan bajo (O-) tienen una visión muy positiva de su persona, son seguros de sí mismos y no están propensos a experimentar culpa. Estas personas se sienten tan satisfechas con lo que son que, en ocasiones, tienden a ser insensibles hacia los sentimientos y necesidades de los demás, pues creen merecerlo todo.

El Factor Q1 explora la orientación psicológica hacia el cambio. Las personas que la prueba define como Q1+ están abiertas y dispuestas al cambio. Los Q1+ se sienten menos atados a su pasado que el resto de las personas, tienden a ser muy liberales y rechazan lo tradicional y convencional. Estos individuos son, por lo general, intelectuales y escépticos que se preocupan por estar bien informados y están menos inclinados a moralizar y más propensos a experimentar en la vida. Al otro lado del continuo se encuentran los Q1-, muy conservadores y tradicionales. Estas personas aceptan lo establecido sin cuestionarlo, no les interesa el pensamiento intelectual o analítico y demuestran una marcada resistencia al cambio.

El Factor Q2 mide el grado de dependencia de la persona. Los que obtienen puntuaciones altas en este factor (Q2+) son individuos autosuficientes que acostumbran tomar decisiones sin preocuparse por las opiniones ajenas, prefieren estar solos la mayor parte del tiempo y hacen sus cosas sin pedir ayuda a los demás. Los que sí demuestran un alto grado de dependencia son los que obtienen puntuaciones bajas en este factor. Los Q2- demuestran una preferencia por estar en grupo la mayor parte del tiempo y toman sus decisiones en base a lo que piensan otros y lo que establece la sociedad, en vez de utilizar su propio juicio. Estas personas necesitan sentir que pertenecen a un grupo donde son aceptados y queridos.

El Factor Q3 explora los esfuerzos del individuo por mantener una congruencia entre su yo ideal y su yo real; moldeándose de acuerdo a patrones establecidos y aprobados por la sociedad. Las personas que obtienen puntuaciones altas (Q3+) se esfuerzan por igualar su conducta a la imagen ideal y socialmente aceptable que se han creado. Estas personas tienden a controlar sus emociones, son muy autoconscientes, compulsivos y perfeccionistas. Por otro lado, los Q3- no se esfuerzan por controlarse y disciplinarse para lograr igualarse a los ideales de conducta, y no le dan importancia alguna a las reglas que establece la sociedad. Estas personas llevan una vida más relajada y menos estresante que la de los Q3+, pero tienden a ser menos exitosos y reconocidos.

El Factor Q4 mide las sensaciones desagradables que tienden a acompañar la excitación del sistema nervioso autónomo; comúnmente conocida como tensión nerviosa. Las personas que puntúan alto (Q4+) experimentan niveles extremos de tensión nerviosa. Estos individuos padecen de una incomodidad subjetiva constante, son impacientes y se distinguen por su incapacidad para mantenerse inactivos. Se les describe además como frustrados, pues su conducta se interpreta como un exceso de impulsos que se expresan inadecuadamente. Al otro extremo del continuo se encuentran los Q4-, quienes se caracterizan primordialmente por la ausencia de tensión nerviosa. Los Q4- llevan una existencia tranquila y relajada, regida por la calma, la paciencia y un alto grado de satisfacción que podría conducir a la vagancia y al conformismo.(Howarth, E., & Browne, n.d.)(Furnham, 1990)(Gorsuch & Cattell, 1967) (Universidad de Murcia. Facultad de Filosofía y Ciencias de la Educación. Sección de Psicología. et al., 1984) (Machover & Gutiérrez, n.d.)

(Corral, 2004)

## 2) Assessment center

Con el paso del tiempo se ha ido afinando la forma de llevar a cabo esta metodología y cómo entrenar a los profesionales que lo realizan, como sucedió en el año 1979 con un grupo de expertos quienes implementaron las directrices, consideraciones éticas y guías de entrenamiento. Hacia 1989, un segundo comité conformado por Douglas Bray y George Thornton III, miembros del Development Dimensions Internacional y Colorado State University, respectivamente, sugirieron la siguiente definición del AC:

Es una evaluación estandarizada de la conducta basada en múltiples datos, varios observadores entrenados (sombras) y diversas técnicas, donde los juicios hechos acerca de la conducta son en su mayor parte provenientes de simulaciones específicamente desarrolladas para la evaluación. Esta información se conjunta y sintetiza en reuniones que tienen los observadores (sombras) o por procesos de integración estadística (Bray & Campbell, 1968). Específicamente, los criterios para definir un AC son:

- Inclusión de dos o más métodos durante su procedimiento, en donde por lo menos uno de ellos sea individual y el otro grupal.
- Uno o más asesores deben observar el comportamiento de los evaluados, en al menos uno de los ejercicios de simulación.

- Los métodos para realizar la evaluación del desempeño deben ser integrados por un procedimiento estadístico, en una Calificación Global de Evaluación.
- El procedimiento debe tener una duración de al menos dos horas (Hermelin, Lievens, & Robertson, 2007). El AC, es en realidad un método que aplica una serie de ejercicios, con el fin de que cada candidato evaluado tenga la oportunidad de demostrar sus habilidades y competencias a los observadores que lo están evaluando (Grados, 2004).

Para hacer más comprensivo este concepto, Lievens y Thornton III (2005) explicaron lo que no es un AC, los métodos que erróneamente se están utilizando y cuyos elementos esenciales no se ajustan al mismo.

No componen un AC:

- Los métodos que sólo aplican pruebas de papel y lápiz.
- Los que involucran sólo un asesor, contrariamente a lo que manifiesta Hermelin, et al (2007) en los criterios que definen a los AC, mencionados anteriormente.
- Los métodos que no implican la observación de conducta abierta.
- Los ejercicios computarizados en donde el participante debe elegir entre un conjunto de comportamientos alternativos predefinidos.

## Test de personalidad

La medición de la personalidad, es de los aspectos más difíciles de evaluar, puesto que se puede considerar como una invasión a la privacidad de la persona. Adicionalmente, el evaluado puede realizar distorsiones en las respuestas debido a la llamada deseabilidad social, respondiendo sobre la base del deber ser de sus actuaciones o también por conocimiento previo de los test. En un meta-análisis realizado por Goffin, Rothstein, y Johnston (1996) sobre la validez incremental de las pruebas de personalidad y de los AC en la gestión de selección, tomando como criterio el desempeño y promoción del trabajador, hallaron que las pruebas de personalidad eran al menos equivalentes a los AC, en cuanto a la predicción del rendimiento.

Por lo tanto, es probable que al combinar las dos técnicas antes mencionadas, se puede llegar a predecir significativamente el rendimiento de una persona en el trabajo, a través de su personalidad. “Varios estudios sugieren que el uso de pruebas diseñadas específicamente para medir la personalidad en contextos de trabajo aumenta el criterio relacionado con la validez” (Sanz et al, 2008, 47). Prueba de ello, son las investigaciones realizadas con uno de los test más utilizados para evaluar la personalidad: el Inventario de los Cinco Grandes Factores, en razón a

que ha demostrado estar relacionado con comportamientos de rendimiento en el trabajo. En un estudio realizado por Heinsman et al (2007), se halló que la conciencia es un predictor válido para todos los puestos de trabajo; los empleados con altos puntajes en este rasgo favorecen la planificación, son organizados y responsables. Además, en este metaanálisis Barrick y Mount (1991), encontraron que el rasgo de apertura a la experiencia, es un predictor válido para la formación de la aptitud. Por lo tanto, conciencia y apertura a la experiencia, son competencias relacionadas principalmente con la dimensión de pensamiento.

Una vez analizados los ejercicios que comúnmente se utilizan en un AC, es importante tener en cuenta la pertinencia de estos con la dimensión que se desea medir. Un sólo ejercicio tiene la cualidad de evaluar diferentes dimensiones de comportamiento, de tal manera que no es recomendable hacer una serie de ejercicios que finalmente afectarán la validez convergente del AC.(Hermelin et al., 2007)(Cascio & Silbey, 1979)(MACAN, AVEDON, PAESE, & SMITH, 1994)

### 3) Test de matrices progresivas de Raven

El test de matrices progresivas de Raven es una prueba más que conocida y utilizada en el ámbito psicológico y psicopedagógico. Este test diseñado en 1938 por el psicólogo inglés John C. Raven, tenía el objetivo de calcular el factor “G” de inteligencia y su administración se restringió a los oficiales de la armada de Estados Unidos.

El factor “G” de inteligencia hace referencia a la inteligencia general que condiciona cualquier ejecución o resolución de problemas, y es común a todas las habilidades que requieran de un componente intelectual. Este factor evidencia la habilidad de una persona a la hora de realizar un trabajo intelectual.(Arthur & Day, 1994)

La característica principal de esta prueba es la de incentivar el razonamiento análogo, la percepción y la capacidad de abstracción. Además, al ser un test no verbal se sirve de la comparación entre formas y el razonamiento por analogías, sin la necesidad de que la persona precise de una cultura o conocimientos previos.

En la actualidad existen diferentes versiones de esta prueba, las cuales se administran en función de la edad y las habilidades de la persona a evaluar. Estas tres versiones son: La Escala general para personas de entre 12 y 65 años



Matrices progresivas en color para niños de entre 3 y 8 años con algún tipo de diversidad funcional intelectual matrices avanzadas para la evaluación de personas con capacidades superiores al promedio características de la prueba existen una serie de características distintivas que han convertido esta prueba en una de las más utilizadas. Estas características se dan tanto a nivel de administración, como de objetivos y fiabilidad. (Ivanovic Marincovich, Rodolfo; Forno Sparosvich, Hernán; Durán Santana, María Cristina; Hazbún Game, Julia; Castro Gómez, n.d.)

### 1. Objetivo

Otro de los objetivos del Test de matrices progresivas de Raven es el de medir la capacidad eductiva de la persona, la cual explicaremos más adelante, mediante la comparación de formas y la utilización del razonamiento por analogía; todo esto de manera independiente a los conocimientos previamente adquiridos por el sujeto.

### 2. Material

Se trata de una prueba que utiliza series de figuras geométricas abstractas e incompletas que se presentan a la persona de manera gradual y con dificultad ascendente. La prueba puede administrarse mediante fichas impresas o también de forma virtual.

### 3. Administración

Otra de las ventajas de esta prueba es que esta es susceptible de ser autoadministrada, así como de administrarse tanto de manera individual como colectiva.

El tiempo de aplicación de esta prueba es de entre 30 y 60 minutos, no obstante este suele completado a los 45 minutos de su inicio.

### 4. Fiabilidad y validez

Finalmente, en cuanto a la fiabilidad y validez de esta prueba esta presenta un 0.87-0.81 de fiabilidad, mientras que en validez se obtuvo un índice de 0.86. Estos datos fueron obtenidos con las fórmulas de Kuder-Richardson y con los criterios de Terman Merrill.

(Miguel De Bortoli<sup>1</sup>, Patricia Barrios y Rosana Azpiroz (Universidad Nacional de San Luis, 2004)(Rossi Casé et al., 2013)

## 5. ARQUITECTURA Y MODELADO

En este capítulo se encontrará toda la arquitectura del software y la realización del modelado.

Estos modelos UML se dividen en tres grupos definidos como:

- **Dimensión Estática:** En este grupo se encuentra el modelo relacionado a la estructura del software como el diagrama de Casos de Uso.
- **Dimensión Dinámica:** En este grupo se encuentran los modelos relacionados a la interacción de objetos en el software como lo son el diagrama de Secuencias y el diagrama de Colaboración.
- **Dimensión Funcional:** En este grupo se encuentran los modelos relacionados al comportamiento del software como lo son el Diagrama de Actividades y el Diagrama de Clases.

### 5.1 Diseño de base de datos

#### 5.1.1 Modelo Entidad Relación

El Modelo Entidad relación (MER) describe el diseño de la base de datos del software SoftArbitrium en MYSQL en donde se almacenará toda la información, la base de datos del aplicativo web consiste de 11 tablas conectadas identificadas cada una por un nombre una llave primaria o Primary Key (PK) y están relacionadas entre sí. (*Observar Ilustración 1 Modelo Entidad Relación*)

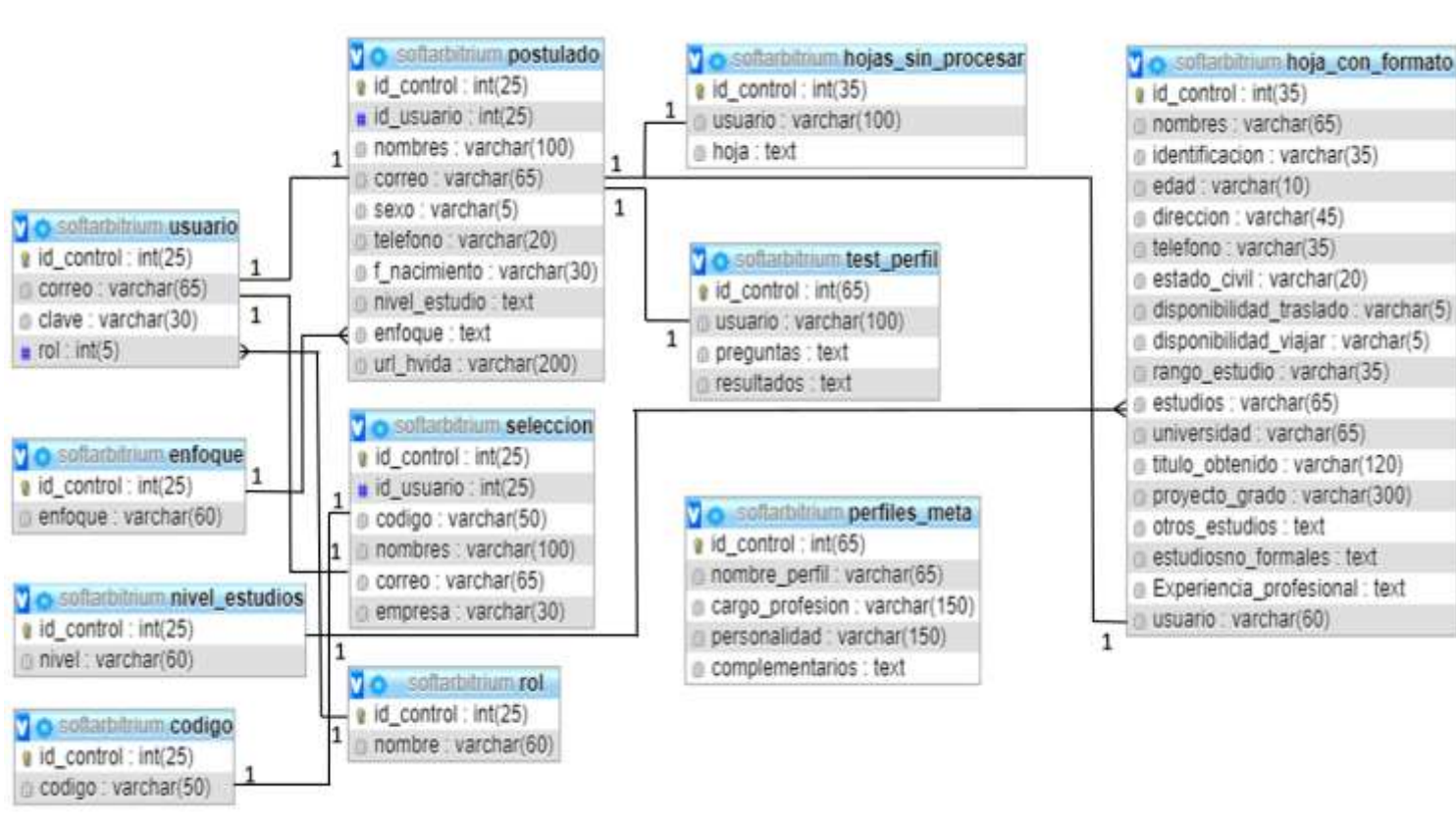


Ilustración 1 Modelo Entidad Relación

- **Descripción del Modelo Entidad Relación**

En la *Tabla 1 Modelo Entidad Relación* se observará una breve descripción de cada una de las tablas existentes en el MER.

<b>Tabla</b>	<b>Descripción</b>
<b>Usuario</b>	En esta tabla se almacena los datos necesarios para la entrada de los usuarios como lo son: correo y clave, también el rol que tendrá dentro de la plataforma.
<b>Rol</b>	En esta tabla se almacenarán los tres tipos de roles: Postulado, selección y administrador.
<b>Enfoque</b>	Esta tabla almacenará la descripción del enfoque de cada postulado y se identificará por el id_control.
<b>Nivel_estudios</b>	Esta tabla almacenará el nivel de estudios cursados por el postulado.
<b>Codigo</b>	En esta tabla se encuentran los códigos de validación para el registro del personal de selección.
<b>Postulado</b>	En esta tabla se almacenan los datos básicos del postulado y se crea el enlace con la tabla de hoja de vida
<b>Selección</b>	En esta tabla se almacenan los datos básicos del personal de selección.
<b>Hojas_sin_procesar</b>	Esta hoja contendrá la información de la hoja de vida que el usuario ha subido previamente en formato pdf, en una cadena de texto.
<b>Test_perfil</b>	En esta tabla se almacena pregunta a pregunta del test de perfilación del usuario y sus respectivas respuestas

<b>Hoja_con_formato</b>	Esta tabla contendrá un modelo de hoja de vida, como se debería diligenciar y campos requeridos
<b>Perfiles_meta</b>	Esta tabla será la fundamental para la realización de la selección de personal dado que contendrá uno a uno los perfiles meta y basados en ellos la selección de los postulados.

*Tabla 1 Modelo Entidad Relación*

## **5.2. Dimensión Estática (Estructura)**

### **5.2.1. Diagrama de casos de uso**

El Diagrama de Casos de uso contiene dos elementos esenciales para su descripción como lo son los actores que intervienen en el software y las funcionalidades que realiza cada uno de ellos. En SoftArbitrium se dispone de tres roles (Administrador del sistema, selección, postulado) y cada uno de ellos tendrá diferentes funcionalidades explicados en los siguientes diagramas.

#### **5.2.1.1. Casos de Uso- Postulado**

En el diagrama de casos de uso expuesto en la *Ilustración 2 Casos de Uso-* se mostrarán las funcionalidades que tiene el rol de postulado.

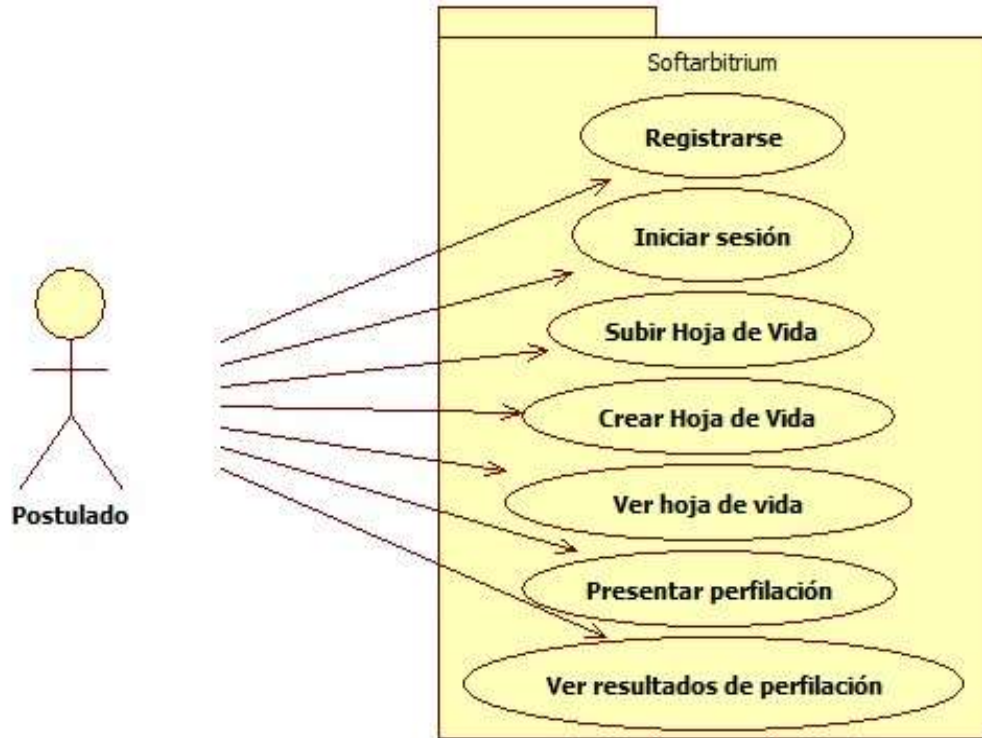


Ilustración 2 Casos de Uso- Postulado

**- Descripción Diagrama de Casos de Uso – Postulado**

En la *Tabla 2 Casos de Uso- postulado* se describirán las funcionalidades mostradas en el diagrama de casos de uso.

Caso de Uso	Descripción
<b>Registrarse</b>	El postulado se registra en el sistema Nombres y Apellidos Correo Sexo Teléfono Clave
<b>Iniciar sesión</b>	Login de ingreso a la plataforma Correo Sexo
<b>Subir hoja de vida</b>	Funcionalidad de subir la hoja de vida en formato pdf

<b>Crear hoja de vida</b>	Crea la hoja de vida en un formulario con un formato estándar
<b>Ver hoja de vida</b>	Puede ver la hoja de vida que fue creada anteriormente
<b>Presentar perfilacion</b>	Test de perfil con los 16 factores de personalidad
<b>Ver resultados de perfilacion</b>	Graficas de los resultados del perfil

Tabla 2 Casos de Uso- Postulado

### 5.2.1.2 Casos de Uso- Selección

En el diagrama de casos de uso expuesto en la *Ilustración 3 Casos de Uso-* se mostrarán las funcionalidades que tiene el rol selección.

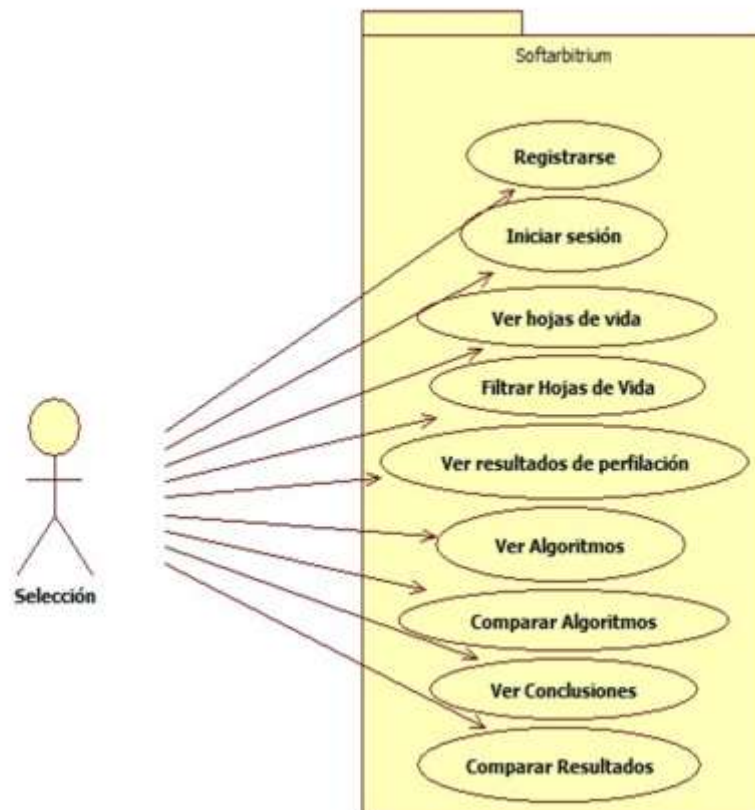


Ilustración 3 Casos de Uso- Selección

### - Descripción del diagrama de Casos de uso – Selección

En la *Tabla 3 Casos de Uso- Proyecto* se describirán todas las funcionalidades mostradas en el diagrama de casos de uso.

Caso de uso	Descripción
<b>Registrarse</b>	Registro de personal de selección
<b>Iniciar sesión</b>	.inicio de sesión login
<b>Ver hojas de vida</b>	Se listan las hojas de vida por 2 factores tratadas y no tratadas
<b>Filtrar hojas de vida</b>	Realiza un filtrado de las hojas de vida por palabras clave o por usuario
<b>Ver resultados de perfilación</b>	Resultados con los 16 tipos de perfiles
<b>Ver algoritmos</b>	Listado de los algoritmos evaluados
<b>Comparar algoritmos</b>	Comparación de algoritmo con resultados esperados
<b>Ver conclusiones</b>	Conclusiones de la investigación
<b>Comparar resultados</b>	Resultados de los perfiles meta

*Tabla 3 Casos de Uso- Proyecto*

#### **5.2.1.3. Casos de Uso- Administrador**

En el diagrama de casos de uso expuesto en la *Ilustración 4 Casos de Uso-* se mostrarán las funcionalidades que tiene el rol administrador.



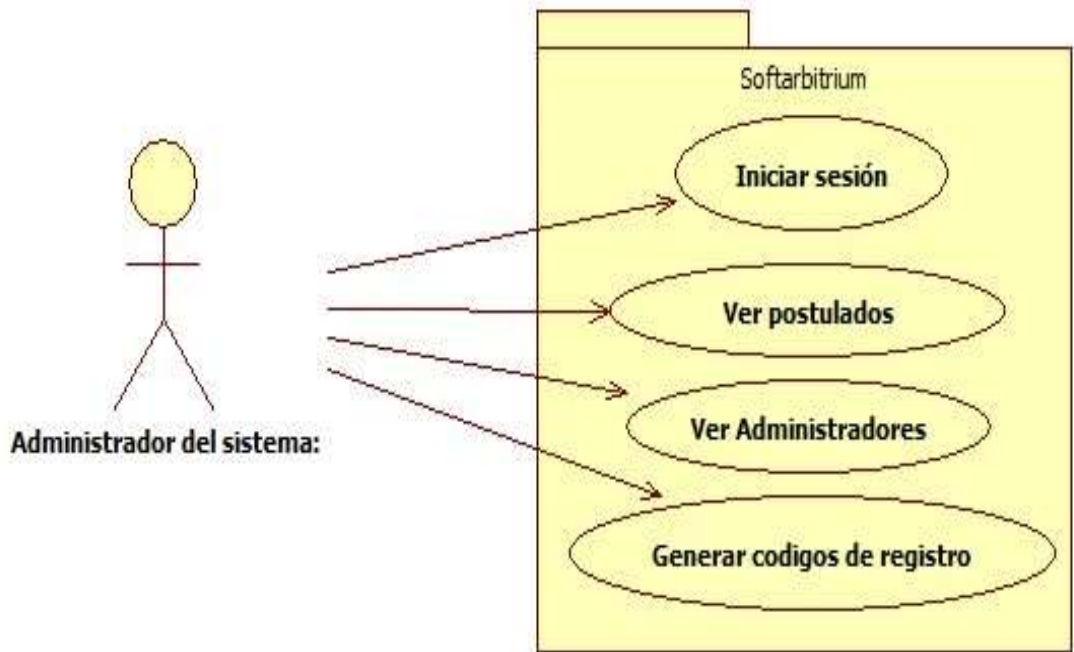


Ilustración 4 Casos de Uso- Administrador

**- Descripción Diagrama de Casos de uso – Administrador**

En la *Tabla 4 Casos de Uso-* se describirán todas las funcionalidades mostradas en el diagrama de casos de uso.

Caso de Uso	Descripción
<b>Iniciar sesión</b>	Inicio de sesión login
<b>Ver postulados</b>	Listado de los postulados en plataforma
<b>Ver administradores</b>	Listado del personal de selección.
<b>Generar códigos de registro</b>	Genera códigos para la inscripción de el personal de selección

Tabla 4 Casos de Uso- Administrador

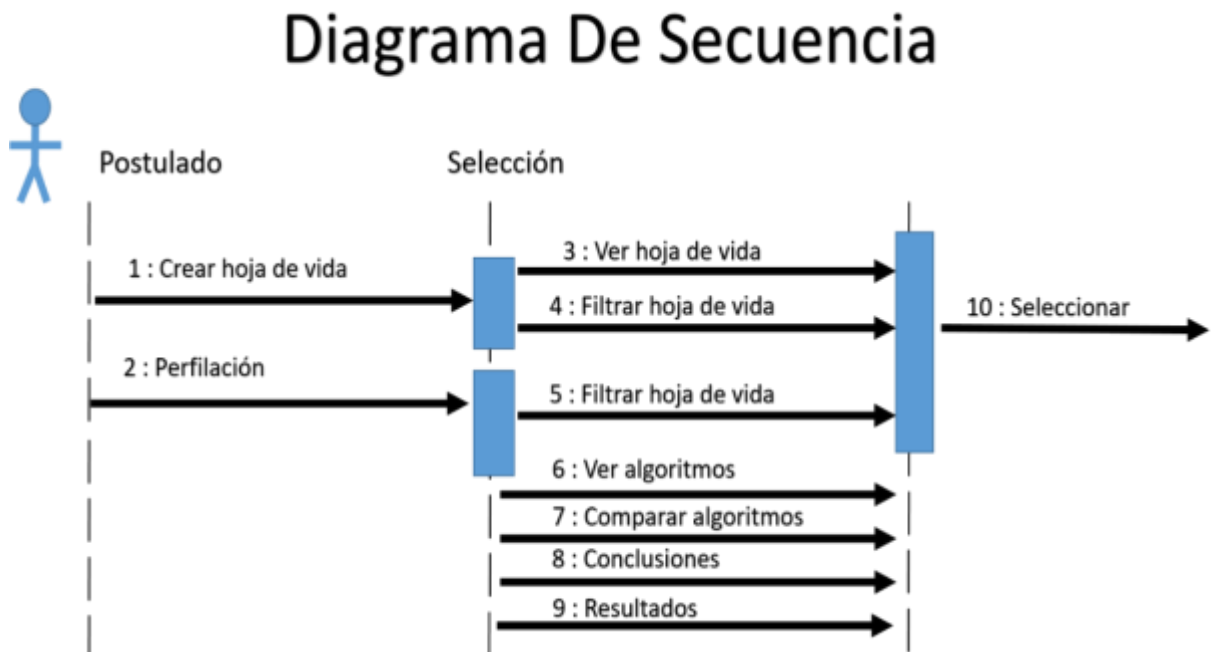
### 5.3. Dimensión Dinámica (Interacción)

#### 5.3.1. Diagrama de Secuencias

El diagrama de secuencias muestra la interacción de objetos realizada en el sistema y explica detalladamente el modelo de casos de uso en la secuencia que se aplican las funcionalidades.

##### 5.3.1.1. Diagrama de Secuencias

En la *Ilustración 5 Diagrama de Secuencias* se mostrarán todas las funcionalidades según la secuencia de funcionalidades permitidas según su rol.



*Ilustración 5 Diagrama de Secuencias*

#### - Descripción del Diagrama de Secuencias

En la *Tabla 5 Diagrama de Secuencias* se describirán cada uno de los objetos del diagrama de secuencias.

Objeto	Descripción
<b>Postulado</b>	El Postulado puede realizar 2 acciones iniciales, la primera es crear la hoja de vida y la segunda realizar el test de perfilacion.
<b>Selección</b>	El personal de selección puede ver y filtrar tanto hojas de vida como perfiles, luego de esto se ejecuta un método de machine learning para seleccionar el perfil más idóneo .

Tabla 5 Diagrama de Secuencias

## 5.4. Dimensión Funcional (comportamiento)

### 5.4.1. Diagrama de Actividades

El diagrama de actividades hace parte de la dimensión funcional que se encarga de explicar todo el comportamiento de los roles en su interacción con el software. Este diagrama muestra el flujo de información con el que usuario tendrá acceso a la herramienta.

#### 5.4.1.1. Diagrama de Actividades- Administrador del Sistema

En la *Ilustración 6 Diagrama de Actividades-* se encuentra el diagrama de actividades.

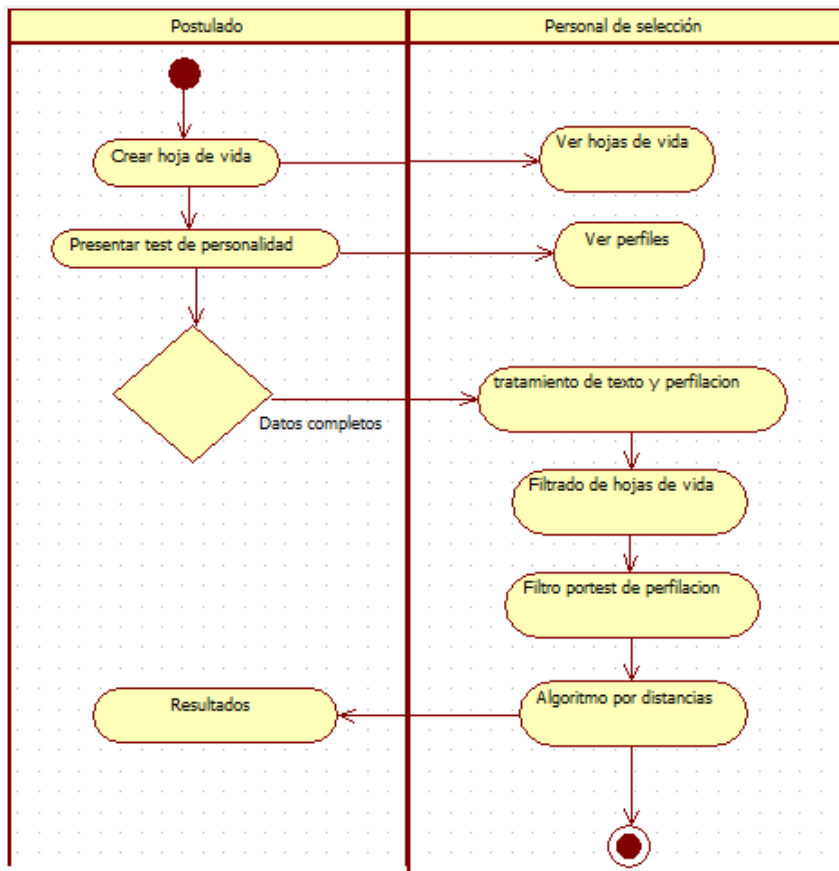


Ilustración 6 Diagrama de Actividades-

### - Descripción del diagrama de actividades

En la *Tabla 6 Diagrama de Actividades-* se describirán todos los procesos mostrados en el diagrama de actividades.

Actividad	Descripción
<b>Crear hoja de vida</b>	El postulado sube su hoja de vida en formato pdf o puede crarla con un formato pre establecido
<b>Presentar test de personalidad</b>	Se presenta el test 16PF
<b>Ver hojas de vida</b>	Puede ver la hoja anteriormente creada
<b>Ver test de perfilacion</b>	Se ven las graficas de resultados del test 16PF
<b>Tratamiento de texto</b>	Se realiza un tratamiento de texto mediante cadenas para saber la distancia de un termino a otro
<b>Filtrado de la hoja de vida</b>	Se pueden filtrar hojas de vida por términos o por usuario

<b>Filtrado test de perfilacion</b>	Se ven los resultados separados en graficas del test 16PF
<b>Algoritmo por distancias</b>	Se realiza la unificación de hojas de vida con perfiles apra evaluar por distancias
<b>Resultados</b>	Resultados de distancias a el perfil meta

Tabla 6 Diagrama de Actividades-

## 5.4.2. Diagrama de Clases

El diagrama de clases describe la estructura del sistema mostrando las clases del mismo, así como sus funciones y atributos, además muestra la interacción entre clases. En la *Ilustración 7 Diagrama de Clases*.

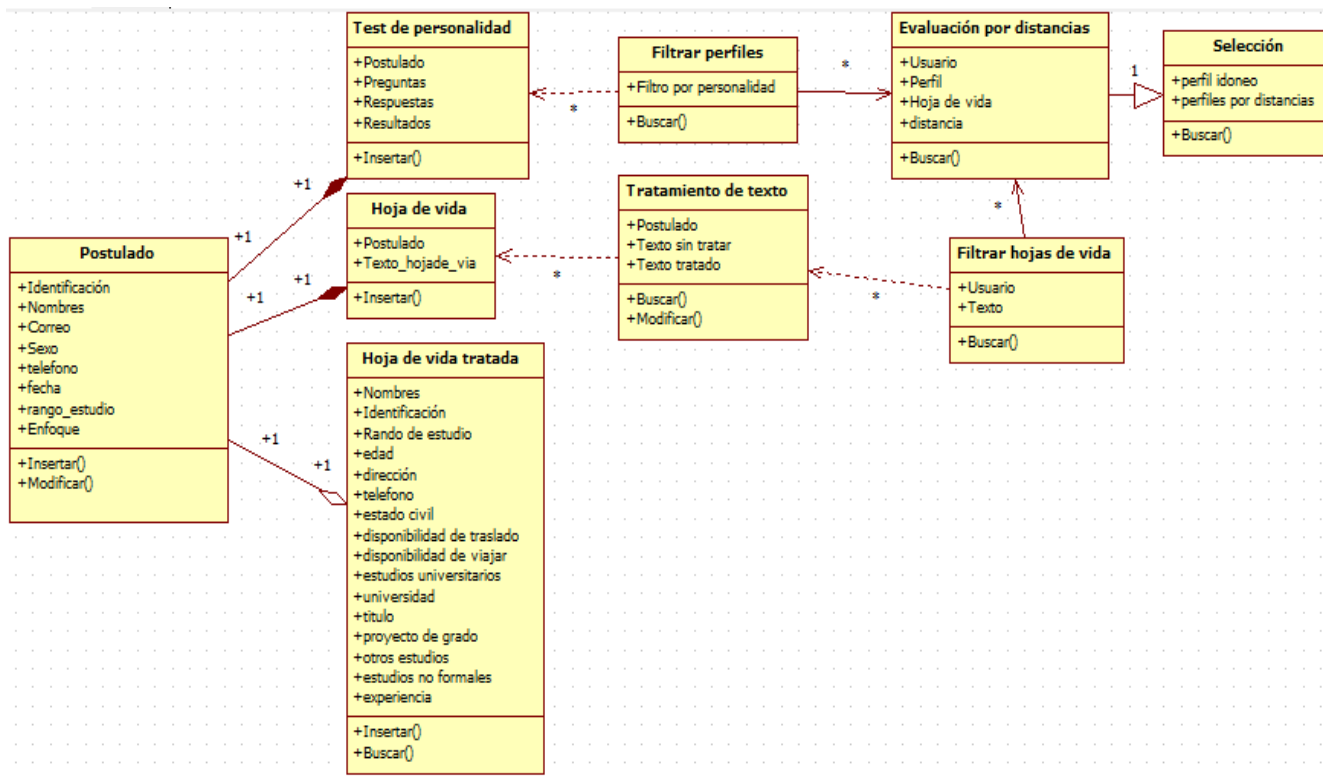


Ilustración 7 Diagrama de Clases

## Descripción Diagrama de Clases

En la *Tabla 7 Diagrama de Clases* se describirán los atributos, funciones e interacciones de cada una de las clases del diagrama.

Clase	Descripción
<b>Postulado</b>	Contiene toda la información relevante en cuanto al postulado para su posterior clasificación
<b>Test personalidad</b>	Este test se encuentra basado en el test 16 pf y arroja los 16 factores de personalidad
<b>Hoja de vida</b>	Toma la hoja de vida que sube el postulado en formato pdf y toma el texto para almacenarlo
<b>Hoja de vida tratada</b>	Hoja de vida bajo los parámetros y formato dados
<b>Filtrar perfiles</b>	Se filtran los perfiles por la personalidad que se requiera
<b>Tratamiento de texto</b>	El texto se toma con distancias entre términos para hacerlo inclusivo
<b>Tratamiento de hojas de vida</b>	Se trata el texto para su posterior manejo eliminando posibles errores
<b>Algoritmo por distancias</b>	Unifica test de personalidad con hojas de vida para comparar distancias a el perfil meta
<b>Selección</b>	Puede tanto crear el perfil idóneo o meta como traer los mas opcionados

*Tabla 7 Diagrama de Clases*

## 5.5. Dimensión Organizacional (Implementación)

### 5.5.1. Diagrama de Despliegue

Este Diagrama denominado de despliegue muestra los componentes de hardware y software usados para el funcionamiento de la herramienta web-(ver *Ilustración 8 Diagrama de Despliegue* )

# Diagrama De Distribución o Despliegue

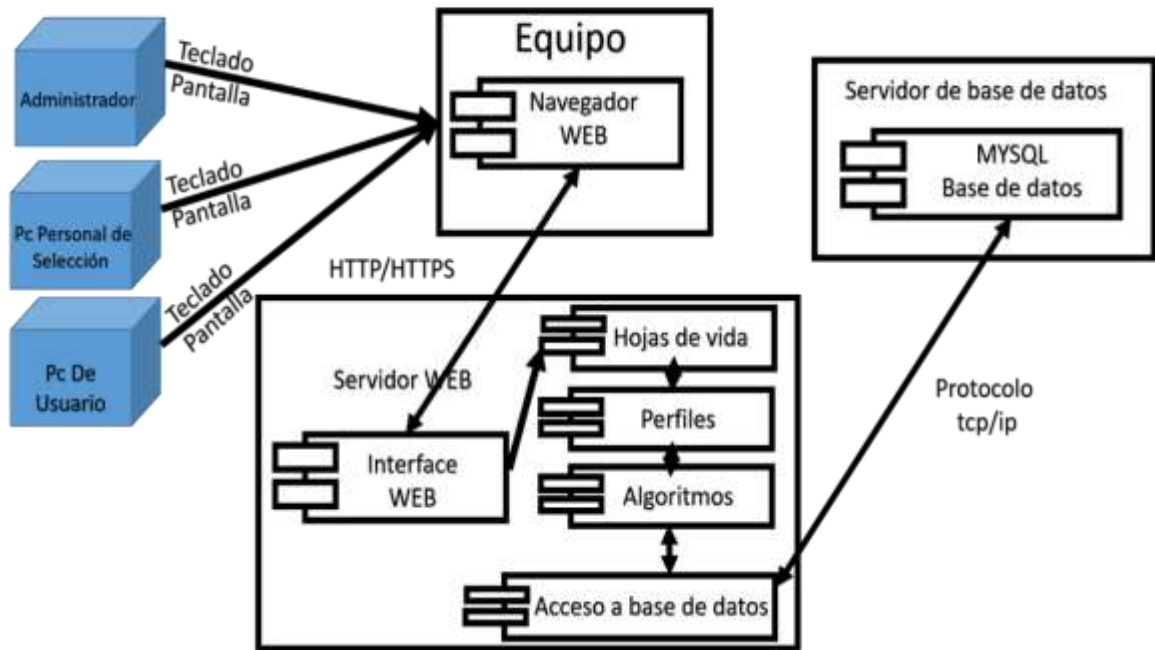


Ilustración 8 Diagrama de Despliegue

## - Descripción Diagrama de Despliegue

En la *Tabla 8 Diagrama de Despliegue* se describirán todos los componentes incluidos en el diagrama de despliegue.

Componente	Descripción
<b>Cliente</b>	Es aquel componente por el cual el cliente tiene comunicación con la herramienta, al ser un aplicativo web es por medio de un navegador web.
<b>Servidor Web</b>	Es aquel componente en donde está alojada la herramienta web.
<b>SGBD (Sistema Gestor de base de datos)</b>	Es aquel componente donde están almacenados los datos de la herramienta web, el gestor de base de datos es MYSQL.

Tabla 8 Diagrama de Despliegue

## 5. RESULTADOS

### 5.1. Modelo propuesto

La siguiente propuesta dada al problema específico de toma de decisiones empresariales enfocadas a la selección de personal, se ha fundamentado fuertemente en los resultados dados por el estado del arte, el proceso se ha dividido en 4 fases fundamentales.

1) Tratamiento de hojas de vida: Dado el número de variables que se encuentran en una hoja de vida y la cantidad de postulados a un cargo específico se ha tomado este tratamiento en 4 etapas.

- a) Se realiza un procesamiento de los documentos ingresados al sistema (Hojas de vida), usando el reconocimiento de cadenas de texto y de esta forma llevarlo a la base de datos en forma de texto plano para su posterior procesamiento.
- b) Se realiza una normalización de la base de datos, dada la redundancia de los datos o algunos datos que se encuentran planteados de forma errónea por el postulado.
- c) Se implementa una adaptación del algoritmo de Levenshtein, el cual evalúa las distancias o los reemplazos necesarios dentro de una cadena dada de la hoja de vida y las variables que se tienen en cuenta durante la selección, haciendo del software más incluyente y más flexible en cuanto a errores resultantes de fallas de digitación o creación de la hoja de vida como un ejemplo veremos:

Ingeniero = Ingeniera

O el siguiente error de digitación

psicólogo = sicologo



- d) Se realiza una segmentación por cuartiles de los resultados del texto para hallar datos atípicos, lo cual será repetido posteriormente en la organización de los perfiles.

2) La segunda fase consiste en un correcto análisis de factores de personalidad de cada postulado, para lo cual se ha implementado una adaptación del test 16PF que es ampliamente usado en este campo específico, la prueba arroja varios factores psicológicos medidos de forma independiente, de este test se obtienen resultados numéricos que posteriormente serán analizados mediante técnicas algorítmicas de distancias, como ejemplo de los resultados tenemos: 12,19,18,17,8,5,7,12,15,13,14,16,12,12,20,16 siendo (A,B,C,E,F,H,I,L,M,N,O,Q1,Q2,Q3,Q4)

Las características de cada uno de los perfiles son:

A = abierto, participativo, uso fácil de la comunicación

B = pensamiento abstracto, Rápido para comprender

C = emocionalmente estable, tranquilo, maduro

E = independiente, agresivo, competitivo obstinado

F = entusiasta, alegre, tendencia a ser siempre el líder

G = escrupuloso, perseverante, con un gran sentido del deber

H = bastante emprendedor, socialmente atrevido, espontáneo

I = idealista y soñador, a veces solicita ayuda de otros

L = suspicaz, difícil de engañar, desconfiado, se complica

M = imaginativo, centrado en necesidades íntimas, abstraído

N = astuto, calculador, sutil, previsor, puede ser “cabeza dura”

O = depresivo, ansioso, ante dificultades se presenta infantil

Q1 = analítico, crítico, de ideas libres y pensamiento abierto

Q2 = independiente, prefiere sus propias decisiones, actúa por su cuenta.

Q3 = prudente, refleja mucho control en sus emociones y su conducta.

Q4 = impaciente, tenso, frustrado e impaciente, incapaz de permanecer inactivo.

3) La tercera parte y pilar fundamental en el proceso de selección, es definir correctamente el perfil meta al que se quiere llegar, esto quiere decir el cargo que se ocupara y está compuesto este perfil por los siguientes factores:

- Nombre para el perfil
- Cargo o profesión
- Cargo o profesión opcional
- Filtro de perfil:
- Características complementarias

Cada una de las características anteriormente mencionadas tiene unos respectivos grados de relevancia, estos definirán el peso que darán sobre esa característica a la hoja de vida los cuales son:

- Cargo o profesión = Muy baja, Baja, Media, Alta y Muy alta.
- Cargo o profesión opcional = Muy baja, Baja, Media, Alta y Muy alta.
- Filtro de perfil: Incluyente, Excluyente. (Esto determina si se quiere discriminar alguna característica en particular del perfil)
- Características complementarias = Muy baja, Baja, Media, Alta y Muy alta.

4) Por último se realiza una unificación de los test de personalidad con el texto procesado de la hoja de vida y a todo ello se realiza una comparación por distancias para tomar esa distancia y convertirla en un porcentaje de compatibilidad del postulado, el algoritmo de distancias se encuentra inspirado en la técnica de K-means por las siguientes ventajas:

- El coste del aprendizaje es nulo.

- No necesita hacer ninguna suposición sobre los conceptos a aprender.
- Puede aprender conceptos complejos usando funciones sencillas como aproximaciones locales.
- Puede extender el mecanismo para predecir un valor continuo (regresión).
- Es muy tolerante al ruido.

## 5.2. Comparación de algoritmos e machine learning

Dado que el machine learning tiene una gran cantidad de técnicas y cada técnica se desempeña mejor en problemas específicos, se ha realizado una comparación entre las técnicas que mejor se adaptan a la toma de decisiones empresariales enfocadas a la selección de personal, la siguiente lista se encuentra compuesta de 3 parámetros que son: Nombre, fortalezas y debilidades.

Nombre: Árboles de decisiones

Fortalezas:

- Plantean el problema para que todas las opciones sean analizadas.
- Permiten analizar totalmente las posibles consecuencias de tomar una decisión.
- Proveen un esquema para cuantificar el costo de un resultado y la probabilidad de que suceda.
- Ayuda a realizar las mejores decisiones sobre la base de la información existente y de las mejores suposiciones.
- Provee una estructura sumamente efectiva dentro de la cual se puede estimar cuales son las opciones e investigar las posibles consecuencias de seleccionar cada una de ellas
- Nos ayuda a realizar las mejores decisiones sobre la base de la información existente y de las mejores suposiciones

Debilidades:

- Sólo es recomendable para cuando el número de acciones es pequeño y no son posibles todas las combinaciones.
- En la elección de un modelo, existe una cantidad muy limitada y dificulta para elegir el árbol óptimo.
- Presenta inconvenientes cuando la cantidad de alternativas es grande y cuanto

las decisiones no son racionales.

- Al no tener claridad de objetivos, es difícil de organizar las ideas.

Nombre: Algoritmos genéticos

Fortalezas:

- No requieren conocimientos específicos del problema para llevar a cabo la búsqueda

- Usan operadores aleatorios en vez de operadores determinísticos, lo que hace que la convergencia de la técnica varíe con respecto al tiempo.

- Operan de forma simultánea con varias soluciones, tomando información de varios puntos del espacio de búsqueda como guía.

- Resultan menos afectados por los máximos locales que las técnicas de búsqueda tradicionales.

Debilidades:

- El lenguaje que se debe utilizar debe tener la capacidad de tolerar cambios aleatorios; que no lleguen a producir resultados sin sentido o errores fatales. Una solución posible será definir a los individuos por listas de números donde cada uno de estos números representa algún aspecto de la solución que se tenga.

- Puede demorarse bastante en converger o no en absoluto, esto depende de cierto modo en los parámetros que se estén utilizando, por ejemplo el tamaño de la población, número de generaciones, etc.

- Pueden converger prematuramente debido a una serie de problemas. Si un individuo que es más apto que la mayoría de sus competidores emerge muy pronto en el curso de la ejecución, se puede reproducir de tal forma que reduce la diversidad de la población muy pronto, haciendo que el algoritmo converja hacia el óptimo local. Este problema se presenta en poblaciones pequeñas, donde una variación aleatoria en el ritmo de reproducción provoca que un genotipo se haga dominante sobre los otros

Nombre: Redes neuronales artificiales

#### Fortalezas:

- Aprendizaje: Las RNA tienen la habilidad de aprender mediante una etapa que se llama etapa de aprendizaje. Esta consiste en proporcionar a la RNA datos como entrada a su vez que se le indica cuál es la salida (respuesta) esperada.
- Auto organización: Una RNA crea su propia representación de la información en su interior, descargando al usuario de esto.
- Tolerancia a fallos: Debido a que una RNA almacena la información de forma redundante, ésta puede seguir respondiendo de manera aceptable aun si se daña parcialmente.
- Flexibilidad: Una RNA puede manejar cambios no importantes en la información de entrada, como señales con ruido u otros cambios en la entrada (ej. si la información de entrada es la imagen de un objeto, la respuesta correspondiente no sufre cambios si la imagen cambia un poco su brillo o el objeto cambia ligeramente)
- Tiempo real: La estructura de una RNA es paralela, por lo cuál si esto es implementado con computadoras o en dispositivos electrónicos especiales, se pueden obtener respuestas en tiempo real.

#### Debilidades:

- Complejidad de aprendizaje para grandes tareas, cuanto más cosas se necesiten que aprenda una red, mas complicado será enseñarle.
- Tiempo de aprendizaje elevado. Esto depende de dos factores: primero si se incrementa la cantidad de patrones a identificar o clasificar y segundo si se requiere mayor flexibilidad o capacidad de adaptación de la red neuronal para reconocer patrones que sean sumamente parecidos, se deberá invertir mas tiempo en lograr que la red converja a valores de pesos que representen lo que se quiera enseñar.
- No permite interpretar lo que se ha aprendido, la red por si sola proporciona una salida, un número, que no puede ser interpretado por ella misma, sino que se requiere de la intervención del programador y de la aplicación en si para encontrarle un significado a la salida proporcionada.
- Elevada cantidad de datos para el entrenamiento, cuanto más flexible se requiera que sea la red neuronal, más información tendrá que enseñarle para que realice de forma adecuada la identificación.

Nombre: Reglas de asociación

Fortalezas:

- muy ampliamente métodos para aprendizaje de reglas de asociación relevando asociaciones entre conjuntos de datos.
- Fácil implementación inicial.

Debilidades:

- El procesamiento se torna lento en grandes conjuntos de datos
- Se debe realizar un previo tratamiento a las variables para que apliquen las reglas de formas más óptima.

Nombre: Máquinas de vectores de soporte

Fortalezas:

- Óptimas para reconocimiento de patrones
- Ampliamente utilizadas en series de tiempo
- El desempeño de las SVM no depende del tamaño de la muestra que se va utilizar para el problema, por lo que puede ser utilizado para una cantidad limitada de datos en contraste con otras metodologías que presentan mejor desempeño cuando el tamaño de la muestra es grande.
  
- El algoritmo detrás de las SVM se puede ajustar a problemas no lineales y la solución se realiza bajo programación cuadrática, lo cual hace que su solución sea única y generalizable.

Debilidades:

- El mapeo suele ser lento
- La representación de los hiperplanos suele ser costoso.

Nombre: Algoritmos de agrupamiento

Fortalezas:

- Útil cuando no se conoce de antemano el número de clases del problema (número de clusters desconocido).

- Encuentra agrupamientos de tal forma que los objetos de un grupo sean similares entre sí y diferentes de los objetos de otros grupos.
- La definición de la métrica de similitud/distancia será distinta en función del tipo de dato y de la interpretación semántica que nosotros hagamos.

Debilidades:

- Dependencia del orden de presentación (comportamiento sesgado por el orden de presentación de los patrones).
- Presupone agrupamientos compactos separados claramente de los demás (puede no funcionar adecuadamente en presencia de ruido).

Nombre: Redes bayesianas

Fortalezas:

- Permiten representar al unísono la dimensión cualitativa y cuantitativa de un problema en un entorno gráfico inteligible.
- Pueden trabajar con datos perdidos de una manera eficiente, algo que en la práctica es deseable.
- Permiten reducir el sobre ajuste de los datos.
- Posibilitan el descubrimiento de la estructura causal subyacente en un conjunto de datos.
- Representan toda la información en un único formato (probabilístico y gráfico) lo que hace sencillas las interpretaciones, permite retractarse de conclusiones obtenidas con anterioridad ya que no son razonables a la luz de nuevas evidencias, proporcionan una visión general del problema, generan un conjunto de alternativas ordenadas y facilita la explicación de las conclusiones.
- Permiten realizar inferencias bidireccionales; esto es, desde los efectos a las causas y desde las causas a los efectos, etc.

Debilidades:

- Toma evidencia de mucho atributos lo que requiere un procesamiento aún mayor.
- Poco efectiva en escalabilidad cuando no se conoce la cantidad de variables a

manejar.

Nombre: Análisis de discriminantes lineales

Fortalezas:

- Simplicidad del método.
- Las probabilidades a posteriori son fáciles de obtener.
- Disponible en muchos paquetes de programas estadísticos.

Debilidades:

- Las suposiciones de Normalidad e igualdad de varianza no siempre se cumplen.
- La clasificación de nuevas observaciones no es muy eficiente a medida que se incrementa el número de variables en el modelo. Se acostumbra a hacer selección de variables antes de aplicar el método.

Nombre: Regresión en procesos Gaussianos

Fortalezas:

- Estas distribuciones nos cuantifican nuestra incertidumbre en los modelos.
- Los procesos gaussianos se definen como una distribución de probabilidad sobre funciones aleatorias.
- Son sobre colecciones infinitas de variables (funciones), tal que cualquier subconjunto de variables aleatoria finita tiene una distribución gaussiana multivariable.

Debilidades:

- Cuando se usan en predicción el tiempo es mayor y aumenta el grado de incertidumbre
- Alto grado de trabajo computacional
- En contraste con OLS la función objetivo es no diferenciable en el origen y por consiguiente no puede darse una solución cerrada
- No se cuenta aún con un buen desarrollo de la teoría asintótica
- Una gran cantidad de autores están trabajando en este punto

Nombre: k-vecinos más próximos



Fortalezas:

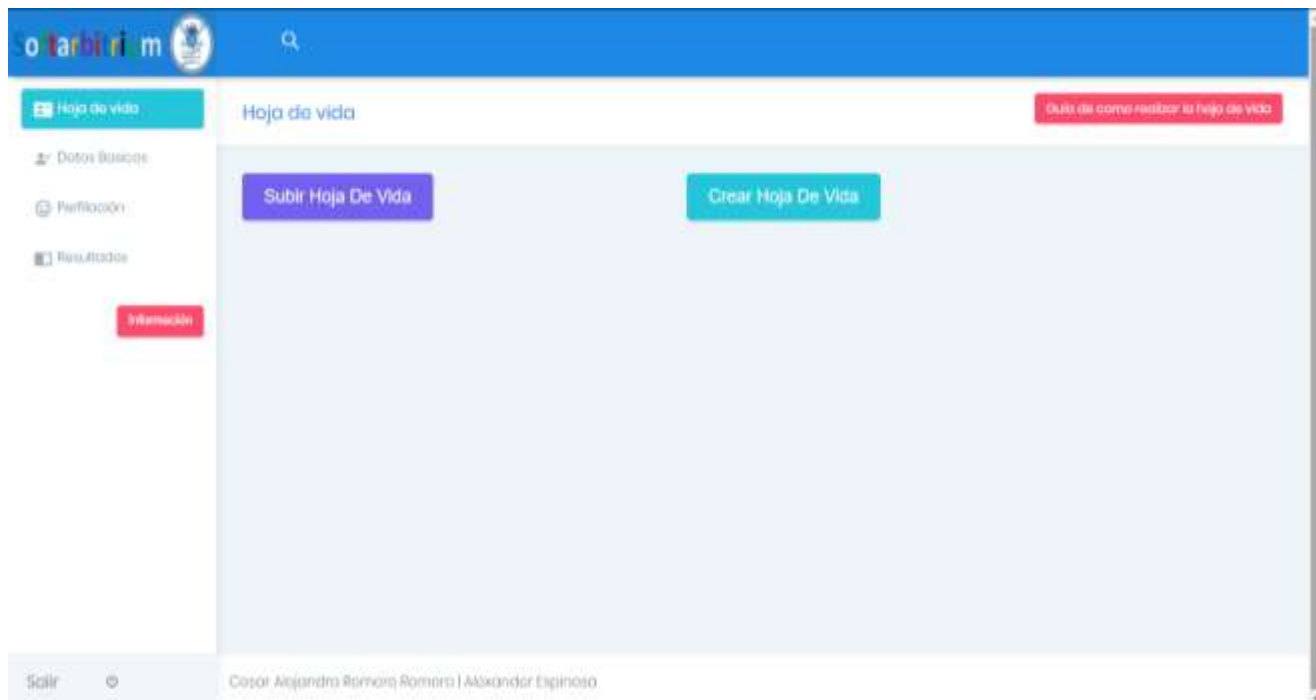
- no es necesario construir un modelo explícito.
- se puede adaptar constantemente a nuevos parámetros
- Fácil escalabilidad

Debilidades:

- lentitud en la clasificación cuando hay muchos datos
- dependencia de la función de distancia
- la influencia nociva de atributos irrelevantes

### **5.3. Secuencia de actividades para el postulado**

Se dará un recorrido paso a paso dentro de la plataforma para que el postulado presente de forma correcta el proceso de hoja de vida y perfilaciónn.



*Ilustración 9. Hojas de vida*

En esta interface gráfica se tendrán 2 opciones, tanto subir hoja de vida en formato pdf para su posterior tratamiento como la creación de hoja de vida con un formato ya establecido.

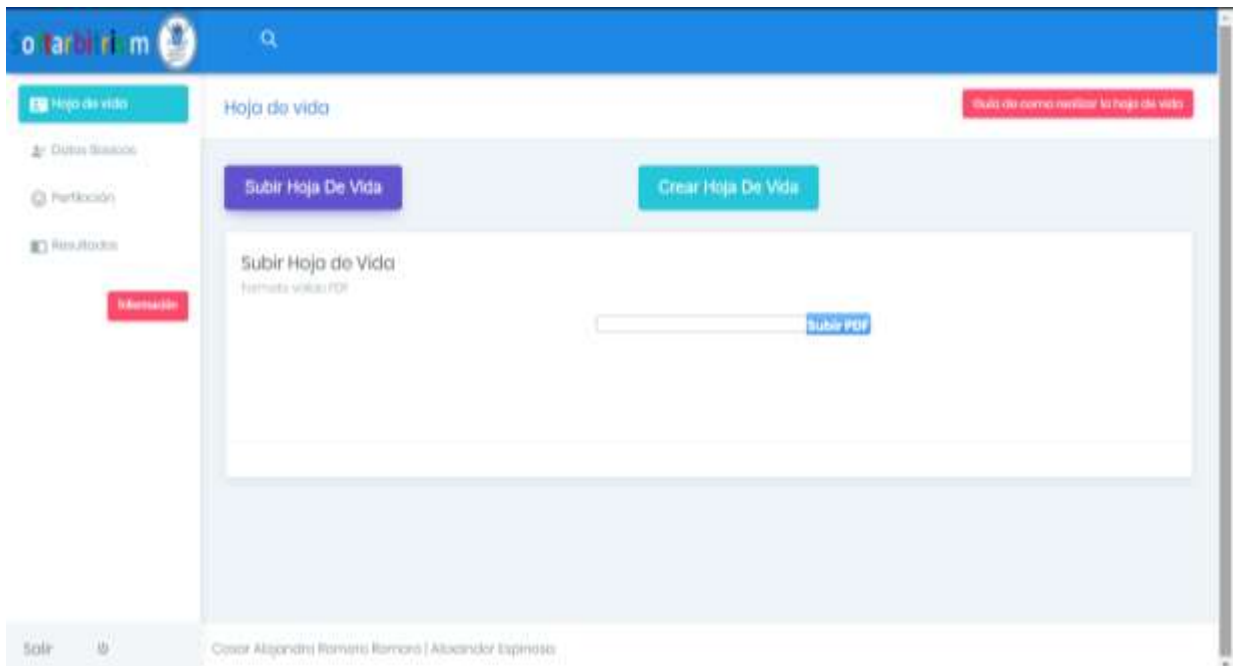


Ilustración 10. Subir hoja de vida

Se tiene la opción de subir pdf de una ubicación del equipo, si el formato es válido y sube correctamente tendrá una confirmación el usuario.

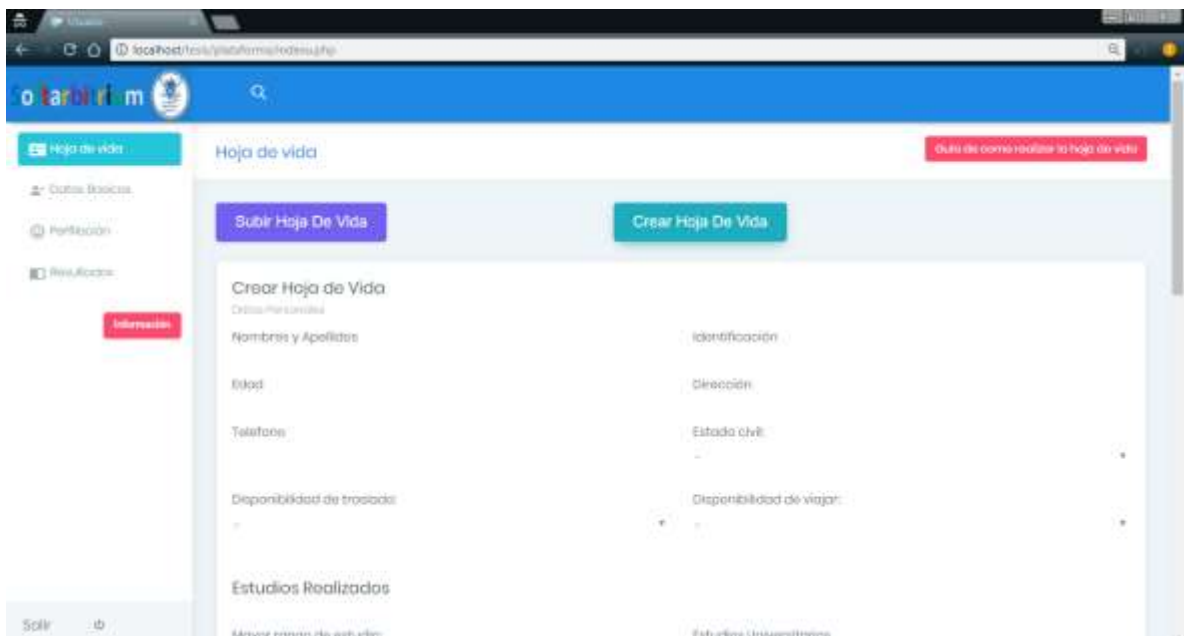


Ilustración 11. Crear hoja de vida

En este formulario tendrá la opción el postulado de crear la hoja de vida con los parámetros datos, también contiene una ayuda.

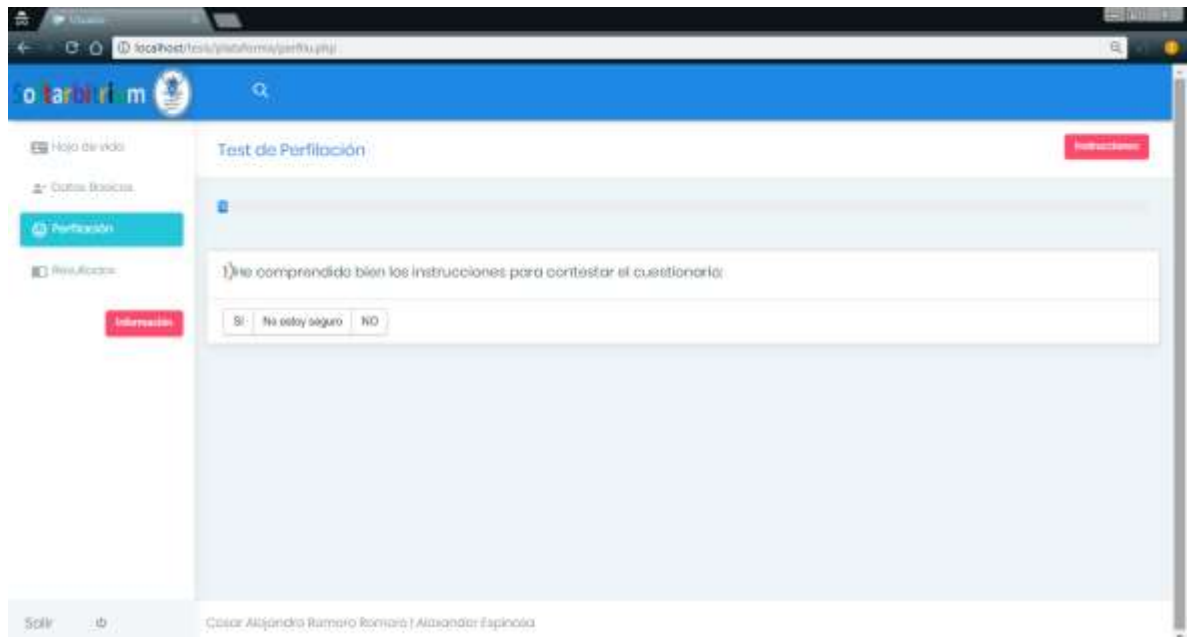
The screenshot shows a web interface for 'o tarbi m'. The main content area is titled 'Datos Básicos De Usuario'. A message at the top of the form states: 'Debe llenar todos los campos para una correcta perfilación'. The form contains the following fields and labels:

- Identificación: 0
- Correo: usuariopruebas@pruebas.com
- Teléfono: 3213333335
- Mayor rango de estudio: -
- Nombres: usuario de pruebas
- Sexo: M
- Fecha de nacimiento: dd/mm/aaaa
- Describa su enfoque o especialidad: -

At the bottom of the form, there is a red 'Guardar' button and a blue 'Cambiar clave' button. On the left sidebar, there are menu items: 'Hoja de vida', 'Datos Básicos', 'Perfilación', 'Resultados', and 'Información' (highlighted in red). At the bottom left of the sidebar is a 'Salir' link.

Ilustración 12. Datos básicos

Datos básicos del usuario.



Como se observa en la ilustración 13 se presenta un test de perfilación que se compone de una *Ilustración 13. Test de perfilación*

serie de preguntas que se verán a continuación, para obtener los resultados del test de personalidad “16PF”.

1. He comprendido bien las instrucciones para contestar el cuestionario:

- A. Sí
- B. No estoy seguro
- C. No

2. Estoy dispuesto a contestar todas las cuestiones con sinceridad:

- A. Sí
- B. No estoy seguro
- C. No

3. ¿Cuál de las siguientes palabras es diferente de las otras dos?

- A. Algo
- B. Nada
- C. Mucho

4. Poseo suficiente energía para enfrentarme a todos mis problemas:

- A. Siempre
  - B. Frecuentemente
  - C. Raras veces
5. Evito criticar a la gente y sus ideas:
- A. Sí
  - B. Algunas veces
  - C. No
6. Hago agudas y sarcásticas observaciones a la gente si creo que las merece:
- A. Generalmente
  - B. Algunas veces
  - C. Nunca
7. Me gusta más la música semiclásica que las canciones populares:
- A. Verdadero
  - B. No estoy seguro
  - C. Falso
8. Si veo peleándose a los niños de mis vecinos:
- A. Les dejo solucionar sus problemas
  - B. No estoy seguro
  - C. Razono con ellos la solución
9. En situaciones sociales:
- A. Fácilmente soy de los que toman la iniciativa.
  - B. Intervengo algunas veces
  - C. Prefiero quedarme tranquilamente a distancia
10. Sería más interesante ser:
- A. Ingeniero de la construcción
  - B. No estoy seguro entre las dos
  - C. Escritor de teatro
11. Generalmente puedo tolerar a la gente presuntuosa, aunque fanfarronee o piense demasiado bien de ella misma:
- A. Sí
  - B. Término medio
  - C. No
12. Cuando una persona no es honrada, casi siempre se le puede notar en la cara:

- A. Verdadero
- B. Término medio
- C. Falso

13. Aceptaría mejor el riesgo de un trabajo donde pudiera tener ganancias mayores, aunque eventuales, que otro con sueldo pequeño, pero seguro:

- A. Sí
- B. No estoy seguro
- C. No

14. De vez en cuando siento un vago temor o un repentino miedo, sin poder comprender las razones:

- A. Sí
- B. Término medio
- C. No

15. Cuando me critican duramente por algo que no he hecho:

- A. No me siento culpable
- B. Término medio
- C. Todavía me siento un poco culpable

16. Casi todo se puede comprar con dinero:

- A. Sí
- B. No estoy seguro
- C. No

17. La mayoría de las personas serían más felices si convivieran más con la gente de su nivel e hicieran las cosas como los demás:

- A. Sí
- B. Término medio
- C. No

18. En ocasiones, mirándome en el espejo, me entran dudas sobre lo que es mi derecha o izquierda:

- A. Verdadero
- B. No estoy seguro
- C. Falso

19. Cuando algo realmente me pone furioso, suelo calmarme muy pronto:

- A. Sí
- B. Término medio
- C. No

20. Prefiero tener una casa:

- A. En un barrio con vida social
- B. Término medio
- C. Aislada en el bosque

21. Con el mismo horario y sueldo, sería más interesante ser:

- A. El cocinero de un buen restaurante
- B. No estoy seguro entre ambos
- C. El que sirve las mesa en el restaurante

22. <<Casado>> es a <<trabajar>> como <<orgullosa>> es a:

- A. Sonreír
- B. Tener éxito
- C. Ser feliz

23. Me pongo algo nervioso ante animales salvajes, incluso cuando están encerrados en fuertes jaulas:

- A. Sí
- B. No estoy seguro
- C. No

24. Una ley anticuada debería cambiarse:

- A. Sólo después de muchas discusiones
- B. Término medio
- C. Inmediatamente

25. La mayor parte de las personas me consideran un interlocutor agradable:

- A. Sí
- B. No estoy seguro
- C. No

26. Me gusta salir a divertirme o ir a un espectáculo:

- A. Más de una vez por semana (más de lo corriente)
- B. Alrededor de una vez por semana (lo corriente)
- C. Menos de una vez por semana (menos de lo corriente)

27. Cuando veo gente desaliñada y sucia:

- A. Lo acepto simplemente
- B. Término medio
- C. Me disgusta y me fastidia



28. Estando en un grupo social me siento un poco turbado si de pronto paso a ser el foco de atención:

- A. Sí
- B. Término medio
- C. No

29. Cuando voy por la calle prefiero detenerme antes a ver a un artista pintando que a escuchar a la gente discutir:

- A. Verdadero
- B. No estoy seguro
- C. Falso

30. Cuando me ponen al frente de algo, insisto en que se sigan mis instrucciones; en caso contrario, renuncio:

- A. Sí
- B. Algunas veces
- C. No

31. Sería mejor que las vacaciones fueran más largas y obligatorias para todas las personas:

- A. De acuerdo
- B. No estoy seguro
- C. En desacuerdo

32. Hablo acerca de mis sentimientos:

- A. Sólo si es necesario
- B. Término medio
- C. Fácilmente, siempre que tengo ocasión

33. Me siento muy abatido cuando la gente me critica en un grupo:

- A. Verdadero
- B. Término medio
- C. Falso

34. Si mi jefe (profesor) me llama a su despacho:

- A. Aprovecho la ocasión para pedirle algo que deseo
- B. Término medio
- C. Temo haber hecho algo malo

35. Mis decisiones se apoyan más en:

- A. El corazón
- B. Los sentimientos y la razón por igual
- C. La cabeza

36. En mi adolescencia pertenecía a equipo deportivos:

- A. Algunas veces
- B. A menudo
- C. La mayoría de las veces

37. Cuando hablo con alguien, me gusta:

- A Decir las cosas tal como se me ocurren
- B Término medio
- C Organizar antes mis ideas

38. A veces me pongo en estado de tensión y agitación cuando pienso en los sucesos del día:

- A. Sí
- B. Término medio
- C. No

39. He sido elegido para hacer algo:

- A. Sólo en pocas ocasiones
- B. Varias veces
- C. Muchas veces

40. ¿Cuál de las siguientes cosas es diferente de las otras dos?

- A. Vela
- B. Luna
- C. Luz eléctrica

41. <<Sorpresa>> es a <<extraño>> como <<miedo>> es a:

- A. Valeroso
- B. Ansioso
- C. Terrible

42. A veces no puedo dormirme porque tengo una idea que me da vueltas en la cabeza:

- A. Verdadero
- B. Término medio
- C. Falso

43. Me siento desasosegado cuando trabajo en un proyecto que requiere una acción rápida que afecta a los demás:

A. Verdadero

B. Término medio

C. Falso

44. Indudablemente tengo menos amigos que la mayoría de las personas:

A. Sí

B. Término medio

C. No

45. Aborrecería tener que estar en un lugar donde hubiera poca gente con quien hablar:

A. Verdadero

B. No estoy seguro

C. Falso

46. Creo que es más importante mucha libertad que buena educación y respeto por la ley:

A. Verdadero

B. No estoy seguro

C. Falso

47. Siempre me alegra formar parte de un grupo grande, como una reunión, un baile o una asamblea:

A. Sí

B. Término medio

C. No

48. En mi época de estudiante me gustaba (me gusta):

A. La música

B. No estoy seguro

C. La actividad de tipo manual

49. Si alguien se enfada conmigo:

A. Intento calmarle

B. No estoy seguro

C. Me irrito con él

50. Para los padres es más importante

A. Ayudar a sus hijos a desarrollarse afectivamente

B. Término medio

C. Enseñarles a controlar sus emociones

51. Siento de vez en cuando las necesidades de ocuparme de una actividad física enérgica:

- A. Sí
- B. Término medio
- C. No

52. Hay veces en que no me siento con humor para ver a alguien:

- A. Muy raramente
- B. Término medio
- C. Muy a menudo

53. A veces los demás me advierten que yo muestro mi excitación demasiado claramente en la voz o en los modales:

- A. Sí
- B. Término medio
- C. No

54. Lo que el mundo necesita es:

- A. Ciudadanos más sensatos y constantes
- B. No estoy seguro
- C. Más "idealistas" con proyectos para un mundo mejor

55. Preferiría tener un negocio propio, no compartido con otra persona:

- A. Sí
- B. No estoy seguro
- C. No

56. Tengo mi habitación organizada de un modo inteligente y estético, con las cosas colocadas casi siempre en lugares conocidos:

- A. Sí
- B. Término medio
- C. No

57. En ocasiones dudo si la gente con quien estoy hablando se interesa realmente por lo que digo:

- A. Sí
- B. Término medio
- C. No

58. Si tuviera que escoger, preferiría ser:

- A. Guarda forestal
- B. No estoy seguro
- C. Profesor de enseñanza media

59. ¿Cuál de las siguientes fracciones es diferente de las otras dos?

- A.  $\frac{3}{7}$
- B.  $\frac{3}{9}$
- C.  $\frac{3}{11}$

60. <<Tamaño>> es a <<longitud>> como <<delito>> es a:

- A. Prisión
- B. Castigo
- C. Robo

61. En mi vida personal consigo casi siempre todos mis propósitos:

- A. Verdadero
- B. No estoy seguro
- C. Falso

62. Tengo algunas características en las que me siento claramente superior a la mayor parte de la gente:

- A. Sí
- B. No estoy seguro
- C. No

63. Sólo asisto a actos sociales cuando estoy obligado, y me mantengo aparte en las demás ocasiones:

- A. Sí
- B. No estoy seguro
- C. No

64. Es mejor ser cauto y esperar poco que optimista y esperar siempre el éxito:

- A. Verdadero
- B. No estoy seguro
- C. Falso

65. Algunas veces la gente dice que soy descuidado, aunque me considera una persona agradable:

- A. Sí
- B. Término medio
- C. No

66. Suelo permanecer callado delante de personas mayores (con mucha más experiencia, edad o jerarquía):

- A. Sí
- B. Término medio
- C. No

67. Tengo un buen sentido de la orientación (sitúo fácilmente los puntos cardinales), cuando me encuentro en un lugar desconocido:

- A. Sí
- B. Término medio
- C. No

68. Cuando leo en una revista un artículo tendencioso o injusto, me inclino más a olvidarlo que a replicar o <<devolver el golpe>>:

- A. Verdadero
- B. No estoy seguro
- C. Falso

69. En tareas de grupo, preferiría:

- A. Intentar mejorar los preparativos
- B. Término medio
- C. Llevar las actas o registros y procurar que se cumplan las normas

70. Me gustaría más andar con personas corteses que con individuos rebeldes y toscos:

- A. Sí
- B. Término medio
- C. No

71. Sí mis conocidos me tratan mal o muestran que yo les disgusto:

- A. No me importa nada
- B. Término medio
- C. Me siento abatido

72. Siempre estoy alerta ante los intentos de propaganda en las cosas que leo:

- A. Sí
- B. No estoy seguro
- C. No

73. Me gustaría más gozar de la vida tranquilamente y a mi modo que ser admirado por mis resultados:

- A. Verdadero
- B. No estoy seguro
- C. Falso

74. Para estar informado, prefiero:

- A. Discutir los acontecimientos con la gente
- B. Término medio
- C. Apoyarme en las informaciones periodísticas de actualidad

75. Me encuentro formado (maduro) para la mayor parte de las cosas:

- A. Verdadero
- B. No estoy seguro
- C. Falso

76. Me encuentro más abatido que ayudado por el tipo de crítica que la gente suele hacer:

- A. A menudo
- B. Ocasionalmente
- C. Nunca

77. En las fiestas de cumpleaños:

- A. Me gusta hacer regalos personales
- B. No estoy seguro
- C. Pienso que comprar regalos es un poco latoso

78. <<AB>> es a <<de>> como <<SR>> es:

- A. qp
- B. pq
- C. tu

79. <<Mejor>> es a <<pésimo>> como <<menor>> es a:

- A. Mayor
- B. Optimo
- C. Máximo

80. Mis amigos me han fallado:

- A. Muy rara vez
- B. Ocasionalmente
- C. Muchas veces

81. Cuando me siento abatido hago grandes esfuerzos por ocultar mis sentimientos a los demás:

- A. Verdadero
- B. Término medio
- C. Falso

82. Gasto gran parte de mi tiempo libre hablando con los amigos sobre situaciones sociales agradables vividas en el pasado:

- A. Sí
- B. Término medio
- C. No

83. Pensando en las dificultades de mi trabajo:

- A. Intento organizarme antes de que aparezcan
- B. Término medio
- C. Doy por supuesto que puedo dominarlas cuando vengan

84. Me cuesta bastante hablar o dirigir la palabra a un grupo numeroso:

- A. Sí
- B. Término medio
- C. No

85. He experimentado en varias situaciones sociales el llamado “nerviosismo del orador”:

- A. Muy frecuentemente
- B. No estoy seguro
- C. No

86. Prefiero leer:

- A. Una narración realista de contiendas militares o políticas
- B. No estoy seguro
- C. Una novela imaginativa y delicada

87. Cuando la gente autoritaria trata de dominarme, hago justamente lo contrario de lo que quiere:

- A. Sí
- B. Término medio
- C. No



88. Suelo olvidar muchas cosas triviales y sin importancia, tales como los nombres de las calles y tiendas de la ciudad:

- A. Sí
- B. Término medio
- C. No

89. Me gustaría la profesión de veterinario, ocupado con las enfermedades y curaciones de los animales:

- A. Sí
- B. Término medio
- C. No

90. Me resulta embarazoso que me dediquen elogios o cumplidos:

- A. Sí
- B. Término medio
- C. No

91. Siendo adolescente, cuando mi opinión era distinta de la de mis padres, normalmente:

- A. Mantenía mi opinión
- B. Término medio
- C. Aceptaba su autoridad

92. Me gusta formar parte activa en las tareas sociales, trabajos de comité, etc.:

- A. Sí
- B. Término medio
- C. No

93. Al llevar a cabo una tarea, no estoy satisfecho hasta que se ha considerado con toda atención al menor detalle:

- A. Verdadero
- B. Término medio
- C. Falso

94. Tengo ocasiones en que me es difícil alejar un sentimiento de compasión hacia mí mismo:

- A. A menudo
- B. Algunas veces
- C. Nunca

95. Siempre soy capaz de controlar perfectamente la expresión de mis sentimientos:

- A. Sí B. Término medio C. No

96. Ante un nuevo invento utilitario, me gustaría:

- A. Trabajar sobre él en el laboratorio
- B. No estoy seguro
- C. Venderlo a la gente

97. La siguiente serie de letras XOOOOXOOOXXX continúa con el grupo:

- A. OXXX
- B. OOX
- C. XOOO

98. Algunas personas parecen ignorarme o evitarme, aunque no sé por qué:

- A. Verdadero
- B. No estoy seguro
- C. Falso

99. La gente me trata menos razonablemente de lo que merecen mis buenas intenciones:

- A. A menudo
- B. Ocasionalmente
- C. Nunca

100 Aunque no sea en un grupo de mujeres y hombre, me disgusta que se use en lenguaje obsceno:

- A. Sí
- B. Término medio
- C. No

101 Me gusta hacer cosas atrevidas y temerarias sólo por el placer de divertirme:

- A. Sí
- B. Término medio
- C. No

102 Me resulta molesta la vista de una habitación muy sucia:

- A. Sí
- B. Término medio
- C. No

103 Cuando estoy en un grupo pequeño, me agrada, quedarme en un segundo término y dejar que otros lleven el peso de la conversación

- A. Sí
- B. Término medio
- C. No

104 Me resulta fácil mezclarme con la gente en una reunión social:

- A. Verdadero
- B. No estoy seguro
- C. Falso

105 Sería más interesante ser:

- A. Orientador vocacional para ayudar a los jóvenes en la búsqueda de su profesión
- B. No estoy seguro
- C. Directivo de una empresa industrial

106 Por regla general, mis jefes y mi familia me encuentran defectos sólo cuando realmente existen:

- A. Verdadero
- B. Término medio
- C. Falso

107 Me disgusta el modo con que algunas personas se fijan en otras en la calle o en las tiendas:

- A. Sí
- B. Término medio
- C. No

108 Como los alimentos con gusto y placer, aunque no siempre tan cuidadosa y adecuadamente como otras personas:

- A. Verdadero
- B. No estoy seguro
- C. Falso

109 Temo algún castigo incluso cuando no he hecho nada malo:

- A. A menudo
- B. Ocasionalmente
- C. Nunca

110 Me gustaría tener más un trabajo con:

- A. Un determinado sueldo fijo
- B. Término medio
- C. Un sueldo más alto, pero siempre que demuestre a los demás que lo merezco

111 Me molesta que la gente piense que mi comportamiento es demasiado raro o fuera de lo corriente:

- A. Mucho
- B. Algo
- C. Nada en absoluto

112 A veces dejo que mis sentimientos de envidia o celos influyan en mis acciones:

- A. Sí
- B. Término medio
- C. No

113 En ocasiones, contrariedades muy pequeñas me irritan mucho:

- A. Sí
- B. Término medio
- C. No

114 Siempre duermo bien, nunca hablo en sueños ni me levanto sonámbulo:

- A. Sí
- B. Término medio
- C. No

115 Me resulta más interesante trabajar en una empresa:

- A. Atendiendo a los clientes
- B. Término medio
- C. Llevando las cuentas o los archivos

116 <<Azada>> es a <<Cavar>> como <<cuchillo>> es a:

- A. Cortar
- B. Afilar
- C. Picar

117 Cuando la gente no es razonable, yo normalmente:

- A. Me quedo tranquilo
- B. Término medio
- C. El menosprecio

118 Si los demás hablan en voz alta cuando estoy escuchando música:

- A. puedo concentrarme en ella sin que me moleste
- B. Término medio
- C. Eso me impide disfrutar de ella y me incomoda

119 Creo que se me describe mejor como:

- A. Comedido y reposado
- B. Término medio
- C. Enérgico

120 Preferiría vestirme con sencillez y corrección que con un estilo personal y llamativo

- A. Verdadero
- B. No estoy seguro
- C. Falso

121 Me niego a admitir sugerencias bien intencionadas de los demás, aunque sé que no debería hacerlo

- A. Algunas veces
- B. Casi nunca
- C. Nunca

122 Cuando es necesario que alguien emplee un poco de diplomacia y persuasión para conseguir que la gente actúe, generalmente sólo me lo encargan a mí:

- A. Sí
- B. Término medio
- C. No

123 Me considero a mí mismo como una persona muy abierta y sociable:

- A. Sí
- B. Término medio
- C. No

124 Me gusta la música:

- A. Ligera, movida y animada
- B. Término medio
- C. Emotiva y sentimental

125 Si estoy completamente seguro de que una persona es injusta o se comporta egoístamente, se lo digo. Incluso si esto me causa problemas

- A. Sí
- B. Término medio
- C. No

126 En un viaje largo, preferiría:

- A. Leer algo profundo pero interesante
- B. No estoy seguro
- C. Pasar el tiempo charlando sobre cualquier cosa con un compañero de viaje

127 En una situación que puede llegar a ser peligrosa, creo que es mejor alborotar o hablar alto, aun cuando se pierda la calma y la cortesía:

- A. Sí
- B. Término medio
- C. No

128 Es muy exagerada la idea de que la enfermedad proviene tanto de causas mentales como físicas:

- A. Sí
- B. Término medio
- C. No

129 En cualquier gran ceremonia oficial debería mantenerse la pompa y el esplendor:

- A. Sí
- B. Término medio
- C. No

130 Cuando hay que hacer algo, me gustaría más trabajar:

- A. En equipo
- B. No estoy seguro
- C. Yo sólo

131 Creo firmemente que "tal vez el jefe no tenga siempre la razón, pero siempre tiene la razón por ser el jefe"

- A. Sí
- B. No estoy seguro
- C. No

132 Suelo enfadarme con las personas demasiado pronto:

- A. Sí
- B. Término medio
- C. No

133 Siempre puedo cambiar viejos hábitos sin dificultad y sin volver a ellos:

- A. Sí
- B. Término medio
- C. No

134 Si el sueldo fuera el mismo, preferiría ser:

- A. Abogado
- B. No estoy seguro entre ambos
- C. Navegante o piloto

135 <<Llama>> es a <<calor>> como <<rosa>> es a:

- A. Espina
- B. Pétalo
- C. Aroma

136 Cuando se acerca el momento de algo que he planeado y esperado, en ocasiones pierdo la ilusión por ello:

- A. Verdadero
- B. Término medio
- C. Falso

137 Puedo trabajar cuidadosamente en la mayor parte de las cosas sin que me molesten las personas que hacen mucho ruido a mi alrededor:

- A. Sí
- B. Término medio
- C. No

138 En ocasiones hablo a desconocidos sobre cosas que considero importantes, aunque no me pregunten sobre ellas:

- A. Sí
- B. Término medio
- C. No

139 Me atrae más pasar una tarde ocupado en una tarea tranquila a la que tenga afición que estar en una reunión animada:

- A. Sí
- B. Término medio
- C. No

140 Cuando debo decir algo, tengo siempre presente las reglas básicas de lo justo de lo injusto:

- A. Sí
- B. Término medio
- C. No

141 En el trato social:

- A. Muestro mis emociones tal como las siento
- B. Término medio
- C. Guardo mis emociones para mis adentros

142 Admiro más la belleza de un poeta que la de un arma de fuego bien construida:

- A. Sí
- B. No estoy seguro
- C. No

143 A veces digo en broma disparates, sólo para sorprender a la gente y ver que responden:

- A. Sí
- B. Término medio
- C. No

144 Me agradaría ser un periodista que escribe sobre teatro, conciertos, opera, etc.:

- A. Sí
- B. No estoy seguro
- C. No

145 Nunca siento la necesidad de garabatear, dibujar o moverme cuando estoy sentado en una reunión:

- A. Verdadero
- B. No estoy seguro
- C. Falso

146 Si alguien me dice algo que yo se que no es cierto, suelo pensar:

- A. Es un mentiroso
- B. Término medio
- C. Evidentemente no está bien informado

147 La gente me considera con justicia una persona activa pero con éxito solo mediano:

- A. Sí
- B. No estoy seguro
- C. No

148 Si se suscitara una controversia violenta entre otros miembros de un grupo de discusión:



- A. Me gustaría ver quién es el ganador
- B. Término medio
- C. Desearía que se suavizara de nuevo la situación

149 Me gusta planear mis cosas solo, sin interrupciones y sugerencias de otros:

- A. Sí
- B. Término medio
- C. No

150 Me gusta seguir mis propios caminos, en vez de actuar según normas establecidas:

- A. Verdadero
- B. No estoy seguro
- C. Falso

151 Me pongo nervioso (tenso) cuando pienso en todas las cosas que tengo que hacer:

- A. Sí
- B. Algunas veces
- C. No

152 No me perturba que la gente me a haga alguna sugerencia cuando estoy jugando:

- A. Verdadero
- B. No estoy seguro
- C. Falso

153 Me parece más interesante ser:

- A. Artista
- B. No estoy seguro
- C. Secretario de un club social

154 ¿Cuál de las siguientes palabras es diferente de las otras dos?

- A. Ancho
- B. Zigzag
- C. Recto

155 He tenido sueños tan intensos que no me han dejado dormir bien:

- A. A menudo
- B. Ocasionalmente
- C. Prácticamente nunca

156 Aunque tenga pocas posibilidades de éxito, creo que todavía me merece la pena correr el riesgo:

- A. Sí
- B. Término medio
- C. No

157 Cuando yo sé muy bien lo que el grupo tiene que hacer, me gusta ser el único en dar las ordenes:

- A. Sí
- B. Término medio
- C. No

158 Me consideran una persona muy entusiasta:

- A. Sí
- B. Término medio
- C. No

159 Soy una persona bastante estricta, he insistido siempre en hacer las cosas tan correctamente como sea posible:

- A. Verdadero
- B. No estoy seguro
- C. Falso

160 Me disgusta un poco que la gente me esté mirando cuando trabajo:

- A. Sí
- B. Término medio
- C. No

161 Como no siempre es posible conseguir las cosas utilizando gradualmente métodos razonables, a veces es necesario emplear la fuerza

- A. Verdadero
- B. Término medio
- C. Falso

162 Si se pasa por alto una buena observación mía:

- A. La dejo pasar
- B. Término medio
- C. Doy a la gente la oportunidad de volver a escucharla

163 Me gustaría hacer el trabajo de un oficial en cargo de los casos de delincuentes bajo fianza:

- A. Sí
- B. Término medio
- C. No

164 Hay que ser prudente antes de mezclarse con cualquier desconocido, puesto que hay peligros de infección de otro tipo:

- A. Sí
- B. No estoy seguro
- C. No

165 En un viaje al extranjero preferiría ir en un grupo organizado, con un experto, que planear yo mismo los lugares que deseo visitar:

- A. Sí
- B. No estoy seguro
- C. No

166 Si la gente se aprovecha de mi amistad, no me quedo resentido y lo olvido pronto:

- A. Verdadero
- B. Término medio
- C. Falso

167 Creo que la sociedad debería aceptar nuevas costumbres, de acuerdo con la razón, y no olvidar los viejos usos y tradiciones:

- A. Sí
- B. Término medio
- C. No

168 Aprendo mejor:

- A. Leyendo un libro bien escrito
- B. Término medio
- C. Participando en un grupo de discusión

169 Me gusta esperar a estar seguro de que lo que voy a decir es correcto, antes de exponer más ideas:

- A. Siempre
- B. Generalmente
- C. Sólo si es posible

170 Algunas veces “me sacan de quicio” de un modo insoportable pequeñas cosas, aunque reconozca que son triviales:

- A. Sí
- B. Término medio
- C. No

171 No suelo decir, sin pensarlas, cosas que luego lamento mucho:

- A. Verdadero
- B. No estoy seguro
- C. Falso

172 Si se me pudiera colocar en una compañía caritativa:

- A. Aceptaría
- B. No estoy seguro
- C. Diría cortésmente que estoy muy ocupado

173 <<Pronto>> es << nunca>> como <<cerca>> es a:

- A. En ningún sitio
- B. Lejos
- C. En otro sitio

174 Si cometo una falta social desagradable, puedo olvidarla pronto:

- A. Sí
- B. No estoy seguro
- C. No

175 Se me considera un “hombre de ideas” que casi siempre puede apuntar una solución aun problema:

- A. Sí
- B. Término medio
- C. No

176 Creo que se me da mejor mostrar:

- A. Aplomo en las pugnas y discusiones de una reunión
- B. No estoy seguro
- C. Tolerancia en los deseos de los demás

177 Me gusta un trabajo que presente cambios, variedad y viajes, aunque implique – algún peligro:

- A. Sí
- B. Término medio
- C. No

178 Me gusta un trabajo que requiera dotes de atención y exactitud:

- A. Sí
- B. Término medio
- C. No

179 Soy de ese tipo de personas con tanta energía que siempre están ocupadas:

- A. Sí
- B. No estoy seguro
- C. No

180 En mi época de estudiante, prefería (prefiero):

- A. Lengua o literatura
- B. No estoy seguro
- C. Matemáticas o aritmética

181 Algunas veces me a turbado el que la gente diga a mi espalda cosas desagradables de mí sin fundamento:

- A. Sí
- B. No estoy seguro
- C. No

182 Hablar con personas corrientes, convencionales y rutinarias:

- A. Es a menudo muy interesante e instructivo
- B. Término medio
- C. Me fastidia por que no hay profundidad o se trata de chismes y cosas sin importancia.

183 Algunas cosas me irritan tanto que creo que entonces lo mejor es no hablar:

- A. Sí
- B. Término medio
- C. No

184 En la formación del niño es más importante:

- A. Darle bastante afecto
- B. Término medio
- C. Procurar que aprenda hábitos y actitudes deseables

185 Los demás me consideran una persona firme e imperturbable, impasible ante los vaivenes de las circunstancias:

- A. Sí
- B. Término medio
- C. No

186 Creo que en el mundo actual es más importante resolver:

- A. El problema de la intensión moral
- B. No estoy seguro
- C. Los problemas políticos

187 Creo que no me he saltado ninguna cuestión y he contestado a todas de modo apropiado:

- A. Sí
- B. No estoy seguro
- C. No



Ilustración 14. Gráficos resultados de test

Y finalmente, se podrá visualizar los distintos tipos de graficos sobre los resultados del test con una explicación de cada uno de los factores de personalidad.

#### 5.4. Secuencia de actividades para el personal de selección

El personal de selección realizara tanto el filtrado de hojas de vida y perfiles, como la creacion de perfiles meta para poder evaluar las distancias de los postulados.

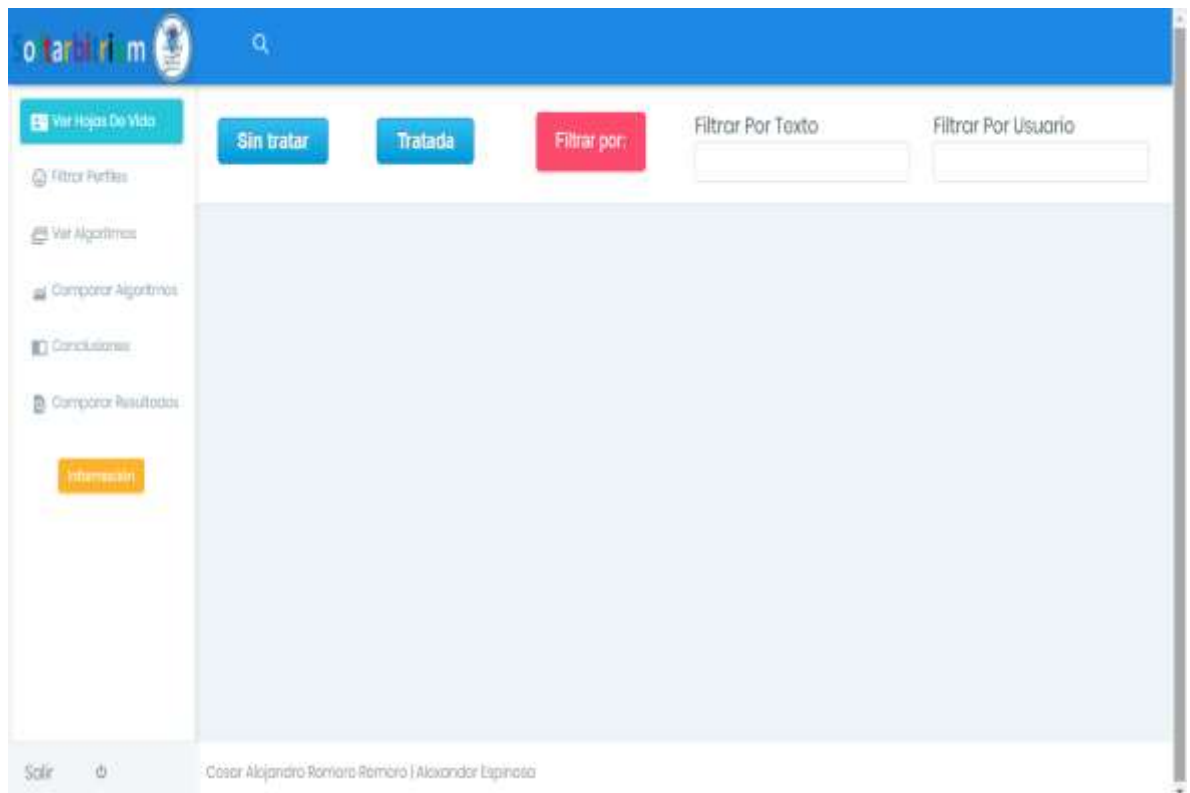


Ilustración 15. Interface para personal de selección

En la figura 16. Se puede evidenciar las hojas de vida sin procesar y sus respectivos filtrados, como parte fundamental de este proceso el personal de selección puede dar click en una hoja de vida especifica para traer la informacion total del postulado.

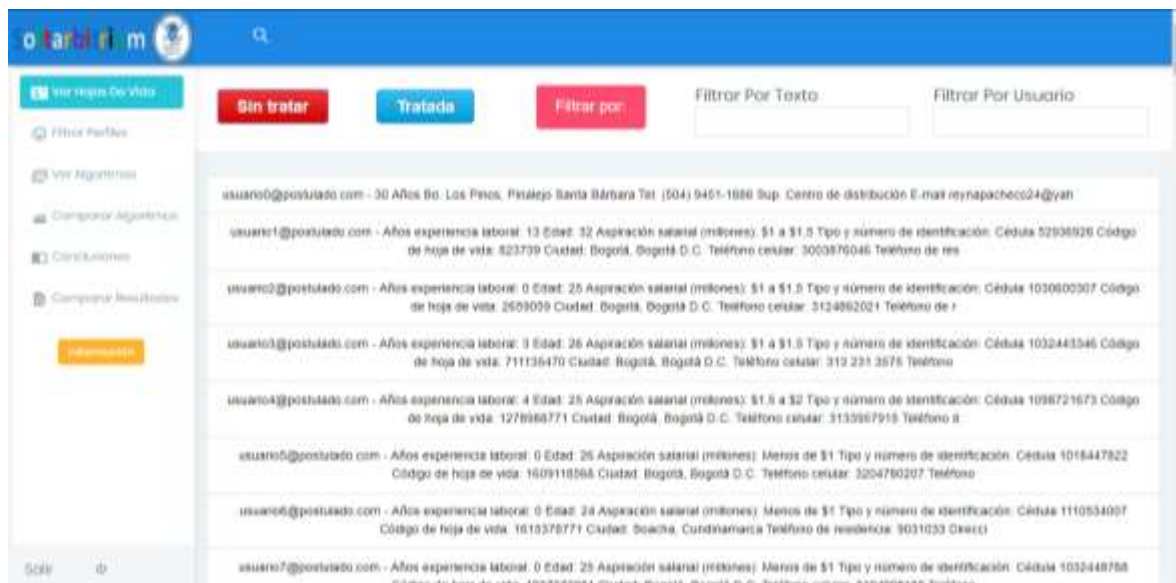


Ilustración 16. Hojas de vida sin procesar

En la figura 17 se puede evidenciar las hojas de vida ya con formato pre establecido.



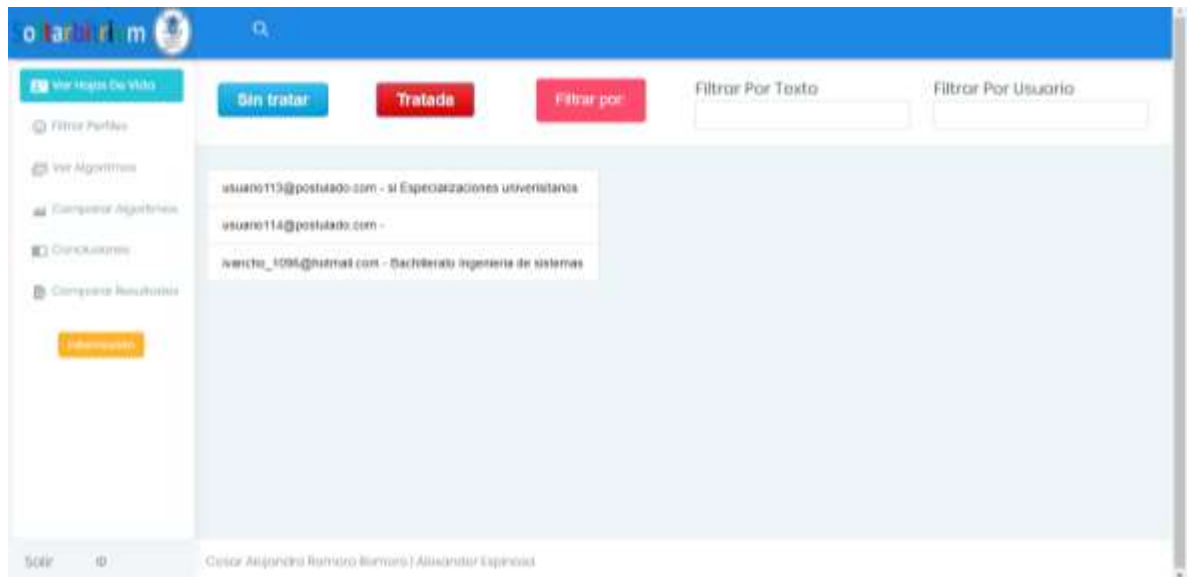


Ilustración 17 Hojas de vida tratadas

En la figura N. 18 se puede ver los filtros de personalidad por los cuales el personal de selección realiza una clasificación de forma descendente de el porcentaje de compatibilidad.

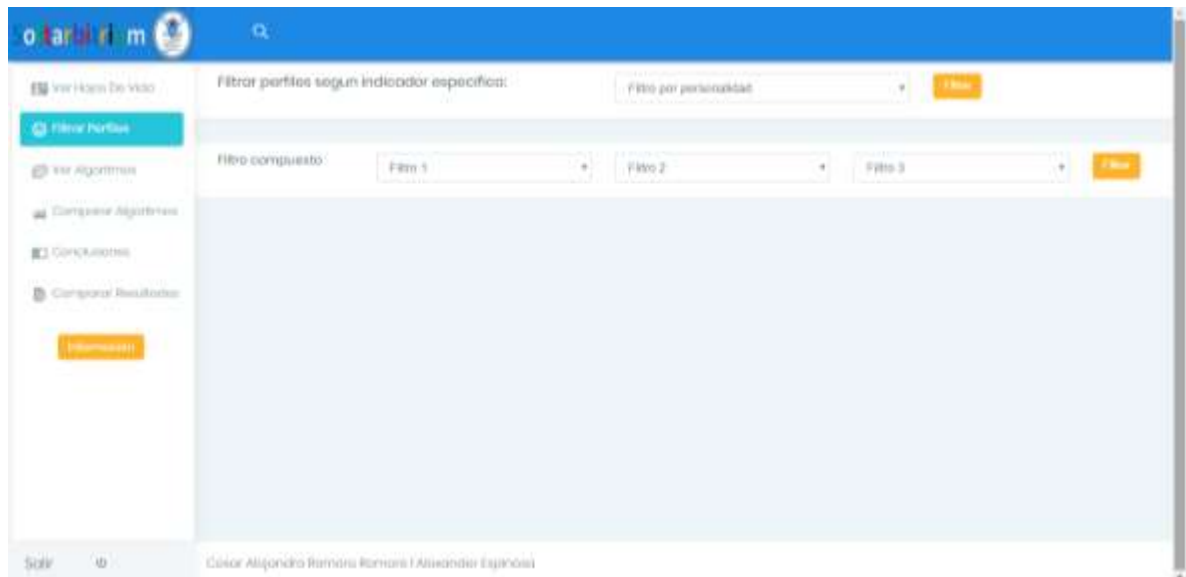


Ilustración 18. Filtros por personalidad.



Como se evidencia en la figura N 19 se organizan los perfiles por un porcentaje de igualdad y posteriormente al hacer click sobre cualquier perfil se podrán ver los resultados factor por factor y su respectiva explicación Figura N. 20.

En la figura N. 21 se listan los algoritmos de machine learning que se investigaron para la implementación en el modelo propuesto.



*Ilustración 21 Perfiles Algoritmos de machine learning evaluados*

Comparar algoritmos es la base fundamental de la demostración en la cual se puede crear un perfil con las características específicas del seleccionador para su posterior búsqueda y organización.

Ilustración 22 Perfiles Creación del perfil mas idóneo para el cargo

Como se puede evidenciar en la figura N. 22. Se crea el perfil meta y se le asignan los respectivos grados de relevancia para posteriormente aplicar ese perfil a la selección.

Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario22@postulado.com	36	4	6	66 %	0	Ver	Ver
usuario27@postulado.com	36	3	3	85 %	0	Ver	Ver
usuario3@postulado.com	35	3	3	82 %	0	Ver	Ver
usuario71@postulado.com	34	1	6	67 %	0	Ver	Ver
usuario18@postulado.com	32	0	1	76 %	2	Ver	Ver

Ilustración 23 Resultados del algoritmo de distancias por porcentaje de compatibilidad con el perfil meta

The screenshot shows a web application interface. At the top left, there is a logo with the text 'o. tarul. i. m.' and a search icon. Below the logo is a navigation menu with the following items: 'Ver Hojas De vida', 'Filtros Perfiles', 'Ver Algoritmos', 'Compara Algoritmos', 'Conclusiones' (highlighted in blue), and 'Compara Resultados'. Below the menu is an orange button labeled 'Iniciar Sesión'. The main content area on the right displays a list of five numbered conclusions. At the bottom left, there is a 'Salir' button. At the bottom right, the text 'César Aljondra Ramos Romero | Aljondra Esquivel' is visible.

1) En la investigación se aprecia como en cada uno de los pasos que se llevan a cabo en la selección de personal se presentan errores, dada la forma en que se evalúan los perfiles dejando datos que pueden ser de gran relevancia para la organización.

2) Se identifica que cada tipo de test realizado a los postulados a un cargo específico cumple con una serie de características según el perfil, siendo totalmente diferente en perfiles altos o en cargos básicos.

3) Se presenta una adaptación del test 16PF ampliamente utilizado en este campo, para llegar a evaluar el perfil óptimo.

4) El tratamiento de las hojas de vida es susceptible a errores de digitación y de planteamiento, llevando a usar análisis de cadenas de texto mediante una adaptación del algoritmo de Levenshtein para lograr hacer las búsquedas más incluyentes y exactas.

5) La unificación de los tests de personalidad y el análisis de las hojas de vida es más óptimo llevarlo a cabo por métodos de distancias, como K-NN siendo el que se evaluó en el presente documento.

Ilustración 24 Conclusiones

## 5.5 Análisis de resultados

Se presentarán a continuación los perfiles idóneos para el cargo que se plantearon y la posterior respuesta del sistema.

### 1) Nombre para el perfil: Ingeniero Electromecánico

Cargo o profesión: Ingeniero electromecánico - Relevancia Muy alta

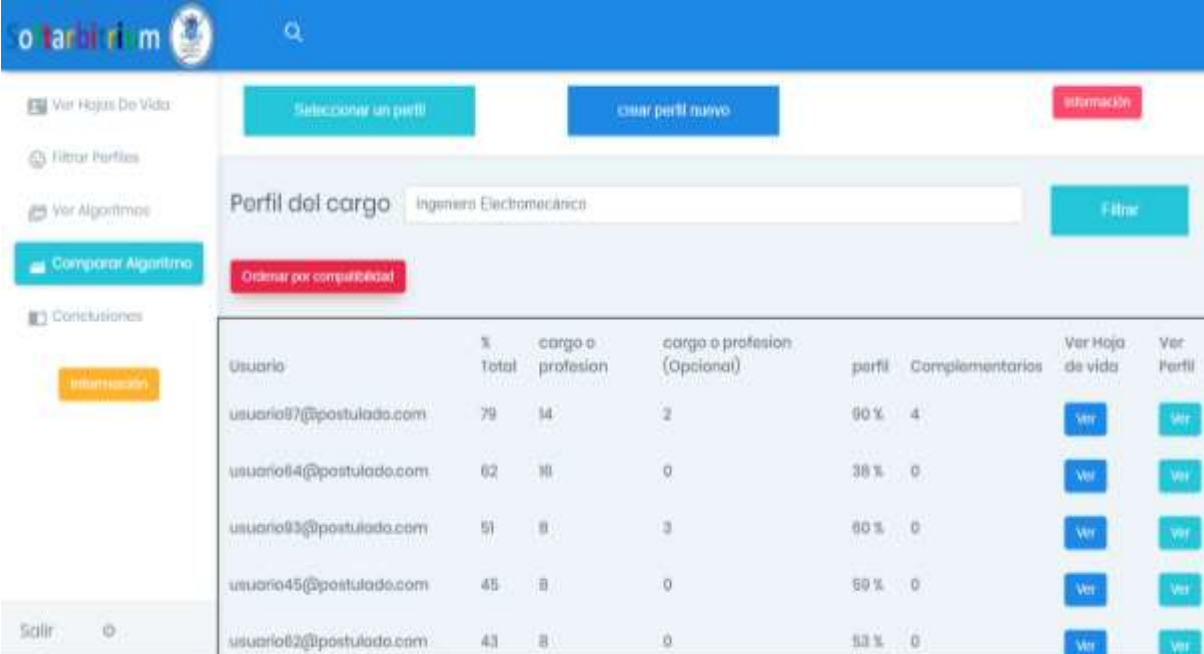
Cargo o profesión opcional: conocimientos hidráulicos - Relevancia Alta

Filtro de perfil: G – Incluyente, C – Incluyente, N – Excluyente

complementarios:

ingles, mantenimiento mecánico, soldadura - Relevancia Media

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados Figura 25. En donde se puede ver el porcentaje de compatibilidad con el perfil idóneo.



Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario17@postulado.com	79	14	2	80 %	4	Ver	Ver
usuario64@postulado.com	62	10	0	38 %	0	Ver	Ver
usuario63@postulado.com	51	8	3	60 %	0	Ver	Ver
usuario45@postulado.com	45	8	0	59 %	0	Ver	Ver
usuario82@postulado.com	43	8	0	53 %	0	Ver	Ver

Ilustración 25 Resultado de búsqueda perfil 1

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo como el 79% a el perfil que más se acoplaba y midiendo las distancias a partir de él.

Como se puede ver en la Ilustración N. 25 el perfil que más se aproximó a el perfil idóneo fue: [usuario97@postulado.com](mailto:usuario97@postulado.com).

Analizando el primer perfil [usuario97@postulado.com](mailto:usuario97@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 79% de compatibilidad con el perfil meta
- 14 coincidencias en cargo o profesión
- 2 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 90%
- 4 compatibilidades con habilidades complementarias.



Ilustración 26 Resultado de test 1

En la Ilustración 26 muestra los resultados del test de personalidad, con los cuales se puede deducir que es una persona bastante emprendedor y obstinado, pero también imaginativo.

- 2) Nombre para el perfil: Encargado de diagnóstico infraestructura  
 Cargo o profesión: arquitecto – Relevancia Muy alta  
 Cargo o profesión opcional: diagnostico a la infraestructura –Relevancia alta  
 Filtro de perfil: B – Incluyente, I - Excluyente ,Q3 – Incluyente  
 complementarios:  
 atención de desastre, ISO 2001, planes de sostenibilidad  
 Relevancia - Media

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados  
 Figura 27. En donde se puede ver el porcentaje de compatibilidad con el  
 perfil idóneo.

Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario53@postulado.com	78	10	7	41%	0	Ver	Ver
usuario65@postulado.com	57	12	0	50%	0	Ver	Ver
usuario88@postulado.com	57	13	2	38%	0	Ver	Ver
usuario10@postulado.com	57	12	0	50%	0	Ver	Ver
usuario42@postulado.com	49	11	1	38%	0	Ver	Ver
usuario36@postulado.com	43	11	0	26%	0	Ver	Ver
usuario6@postulado.com	41	10	0	34%	0	Ver	Ver

Ilustración 27 Resultado de búsqueda perfil 2

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo como el 78% a el perfil que más se acoplaba y midiendo las distancias a partir de él.



Como se puede ver en la Ilustración N. 27 el perfil que más se aproximó a el perfil idóneo fue: [usuario53@postulado.com](mailto:usuario53@postulado.com).

Analizando el primer perfil [usuario53@postulado.com](mailto:usuario53@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 78% de compatibilidad con el perfil meta
- 16 coincidencias en cargo o profesión
- 7 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 41%
- 0 compatibilidades con habilidades complementarias



Ilustración 28 Resultado test personalidad 2

En la Ilustración 28 muestra los resultados del test de personalidad, con los cuales se puede evidenciar que es una persona emocionalmente estable e independiente, imaginativa.

- 3) Nombre para el perfil: Asistente - secretaria  
 Cargo o profesión: Gestión administrativa – Relevancia Media  
 Cargo o profesión opcional: Estudios contables –Relevancia Media  
 Filtro de perfil: I – Incluyente, A - Incluyente ,G – Incluyente  
 complementarios:  
 Word, Excel, ingles  
 Relevancia - Alta

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados  
 Figura 29. En donde se puede ver el porcentaje de compatibilidad con el  
 perfil idóneo.

Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario4@postulada.com	31	0	0	87%	0	Ver	Ver
usuario72@postulada.com	31	4	1	76%	2	Ver	Ver
usuariopruebas@pruebas.com	31	0	2	80%	2	Ver	Ver
usuario35@postulada.com	30	0	0	87%	0	Ver	Ver
usuario54@postulada.com	30	0	0	70%	4	Ver	Ver

Ilustración 29 Resultado de búsqueda perfil 3

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo como el 31% a el perfil que más se acoplaba y midiendo las distancias a partir de él.

Como se puede ver en la Ilustración N. 29 el perfil que más se aproximó a el perfil idóneo fue: [usuario72@postulado.com](mailto:usuario72@postulado.com).

Analizando el primer perfil [usuario72@postulado.com](mailto:usuario72@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 31% de compatibilidad con el perfil meta
- 4 coincidencias en cargo o profesión
- 1 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 76%
- 2 compatibilidades con habilidades complementarias



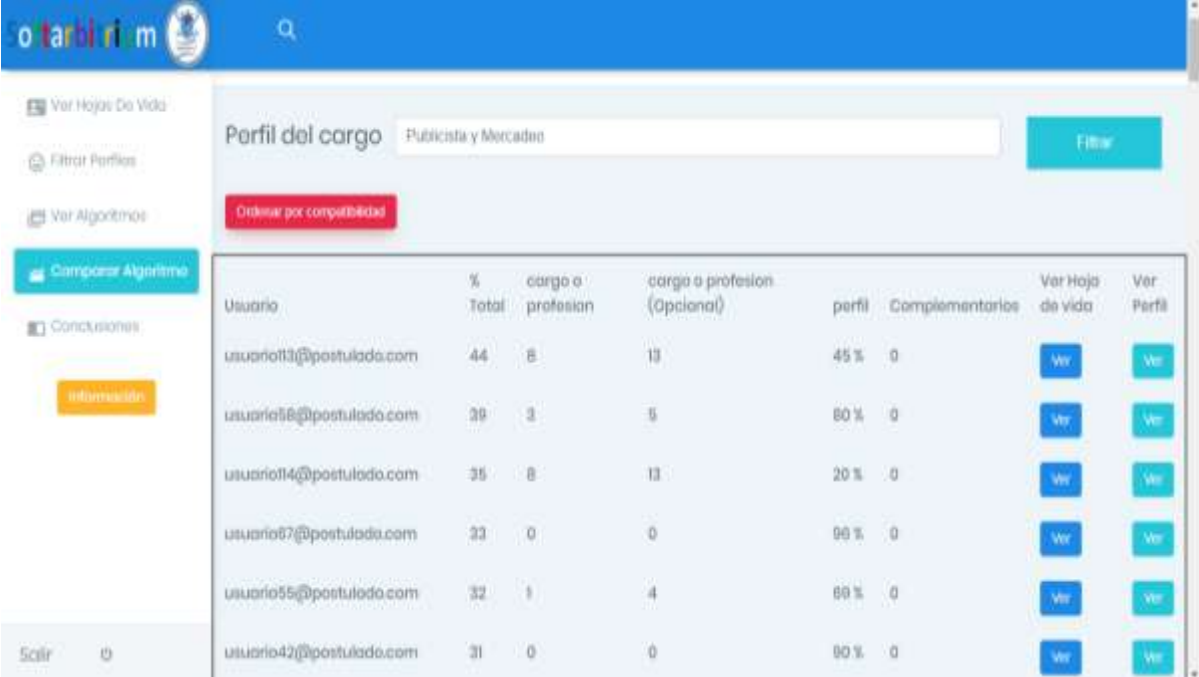
Ilustración 30 Resultado test personalidad 3

En la Ilustración 30 muestra los resultados del test de personalidad, con los cuales se puede evidenciar que es una persona entusiasta, alegre, jovial y socialmente atrevido, bastante espontáneo.

- 4) Nombre para el perfil: Publicista y Mercadeo  
Cargo o profesión: Publicidad – Relevancia Media  
Cargo o profesión opcional: Mercadeo –Relevancia Alta

Filtro de perfil: I – Incluyente, C - Incluyente ,A – Incluyente complementarios:  
 Inglés, logística  
 Relevancia - Media

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados Figura 31. En donde se puede ver el porcentaje de compatibilidad con el perfil idóneo.



Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario113@postulado.com	44	8	13	45%	0	Ver	Ver
usuario56@postulado.com	29	3	5	80%	0	Ver	Ver
usuario114@postulado.com	35	8	13	20%	0	Ver	Ver
usuario67@postulado.com	22	0	0	99%	0	Ver	Ver
usuario55@postulado.com	32	1	4	88%	0	Ver	Ver
usuario42@postulado.com	31	0	0	80%	0	Ver	Ver

Ilustración 31 Resultado de búsqueda perfil 4

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo como el 44% a el perfil que más se acoplaba y midiendo las distancias a partir de él.

Como se puede ver en la Ilustración N. 31 el perfil que más se aproximó a el perfil idóneo fue: [usuario113@postulado.com](mailto:usuario113@postulado.com).

Analizando el primer perfil [usuario113@postulado.com](mailto:usuario113@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 44% de compatibilidad con el perfil meta
- 8 coincidencias en cargo o profesión
- 13 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 45%
- 0 compatibilidades con habilidades complementarias



Ilustración 32 Resultado test personalidad 4

En la Ilustración 32 muestra los resultados del test de personalidad, con los cuales se puede evidenciar que es una persona emocionalmente estable, tranquilo, maduro y abstraído, despreocupado de los asuntos prácticos.

- 5) Nombre para el perfil: tratamiento de aguas residuales  
Cargo o profesión: Ingeniería Ambiental – Relevancia Media  
Cargo o profesión opcional: tratamiento aguas residuales  
–Relevancia Muy Alta  
Filtro de perfil: C – Incluyente, F - Incluyente ,A – excluyente  
complementarios:

impacto ambiental, sistemas, ingles  
Relevancia - Media

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados  
Figura 33. En donde se puede ver el porcentaje de compatibilidad con el  
perfil idóneo.

Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarios	Ver Hoja de vida	Ver Perfil
usuario23@postulado.com	36	4	0	66%	0	Ver	Ver
usuario27@postulado.com	36	3	3	85%	0	Ver	Ver
usuario3@postulado.com	35	3	3	82%	0	Ver	Ver
usuario71@postulado.com	34	1	5	67%	0	Ver	Ver
usuario9@postulado.com	32	0	1	78%	2	Ver	Ver
usuario18@postulado.com	31	0	2	71%	0	Ver	Ver
usuario55@postulado.com	31	2	0	80%	0	Ver	Ver

Ilustración 33 Resultado de búsqueda perfil 5

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo como el 36% a el perfil que más se acoplaba y midiendo las distancias a partir de él.

Como se puede ver en la Ilustración N. 33 el perfil que más se aproximó a el perfil idóneo fue: [usuario23@postulado.com](mailto:usuario23@postulado.com).

Analizando el primer perfil [usuario23@postulado.com](mailto:usuario23@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 36% de compatibilidad con el perfil meta
- 4 coincidencias en cargo o profesión
- 6 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 66%
- 0 compatibilidades con habilidades complementarias



Ilustración 34 Resultado test personalidad 5

En la Ilustración 34 muestra los resultados del test de personalidad, con los cuales se puede evidenciar que es una persona socialmente atrevido, bastante espontáneo, la mayoría de sus respuestas son emocionales y motivado por sí mismo.

- 6) Nombre para el perfil: Jefe de tratamiento a residuos  
 Cargo o profesión: Ingeniería Ambiental – Relevancia Alta  
 Cargo o profesión opcional: tratamiento de residuos  
 –Relevancia Muy Alta  
 Filtro de perfil: G – Incluyente, N - Incluyente, B – Incluyente  
 complementarios:  
 gestion del riesgo, sanitaria, manejo residuos solidos, ingles  
 Relevancia - Alta

Basado en el perfil meta o idóneo se obtuvieron los siguientes resultados  
 Figura 35. En donde se puede ver el porcentaje de compatibilidad con el  
 perfil idóneo.

Usuario	% Total	cargo o profesion	cargo o profesion (Opcional)	perfil	Complementarias	Ver Hoja de vida	Ver Perfil
usuario12@postulado.com	93	27	0	28 %	1	Ver	Ver
usuario68@postulado.com	70	14	0	79 %	0	Ver	Ver
usuario21@postulado.com	66	14	0	69 %	0	Ver	Ver
usuario19@postulado.com	63	11	3	55 %	3	Ver	Ver
usuario18@postulado.com	60	9	8	47 %	0	Ver	Ver
usuario88@postulado.com	52	9	0	71 %	0	Ver	Ver

Ilustración 35 Resultado de búsqueda perfil 6

Los postulados se organizan de mayor a menor porcentaje de compatibilidad con el perfil propuesto por el personal de selección, teniendo



como el 93% a el perfil que más se acoplaba y midiendo las distancias a partir de él.

Como se puede ver en la Ilustración N. 35 el perfil que más se aproximó a el perfil idóneo fue: [usuario12@postulado.com](mailto:usuario12@postulado.com).

Analizando el primer perfil [usuario12@postulado.com](mailto:usuario12@postulado.com) se evidenciaron las siguientes características que demuestran la compatibilidad con el perfil

- 93% de compatibilidad con el perfil meta
- 27 coincidencias en cargo o profesión
- 0 en cargo o profesión opcional
- compatibilidad con el test de personalidad esperado de 28%
- 1 compatibilidades con habilidades complementarias



Ilustración 36 Resultado test personalidad 6

En la Ilustración 36 muestra los resultados del test de personalidad, con los cuales se puede evidenciar que es una persona socialmente atrevido, bastante espontáneo, la mayoría de sus respuestas son emocionales y motivado por sí mismo.

## 5.6 Análisis de algoritmo de distancias

En los resultados se pudo evidenciar como los perfiles de postulados se dan en un punto espacial con relación al perfil meta como se puede ver en la Ilustración N. 37.

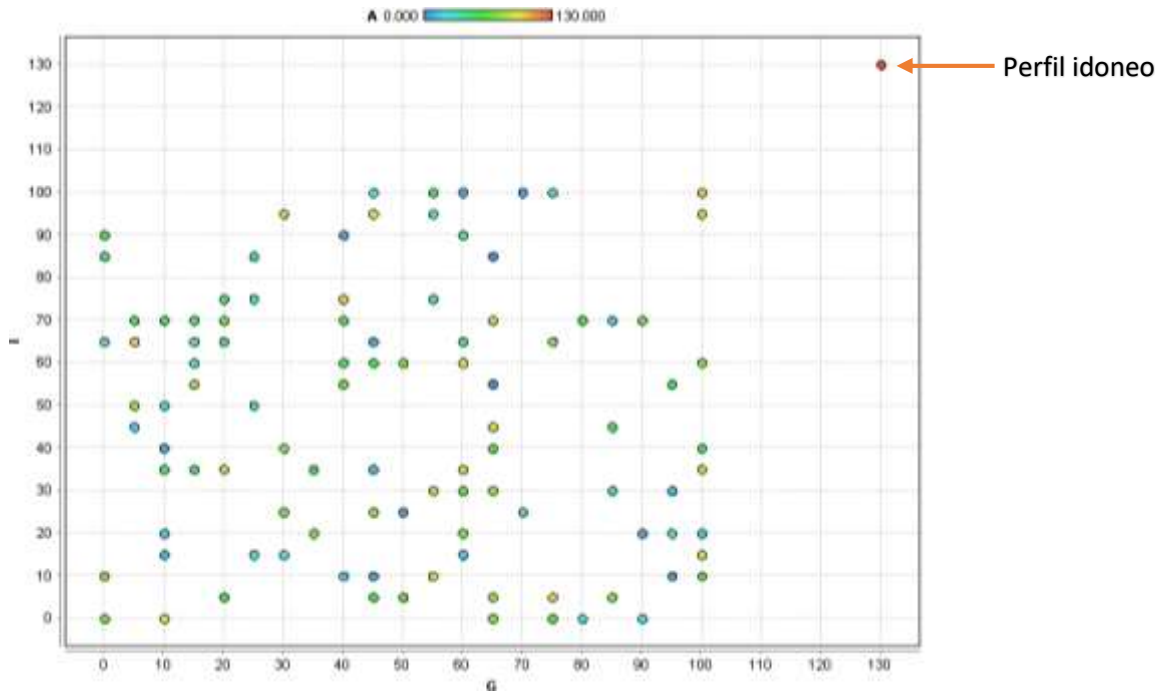
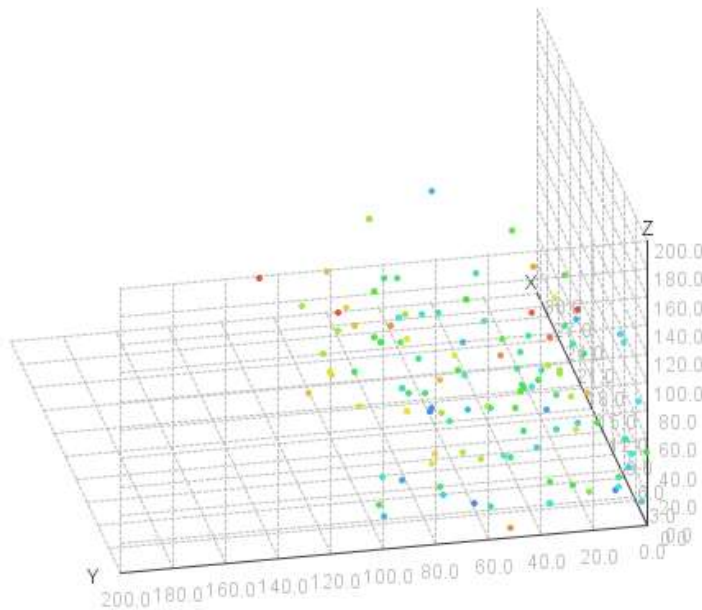


Ilustración 37 Distancia con el perfil meta

Dada la complejidad del problema específico, la cantidad de variables tan grandes a evaluar la dimensionalidad no se puede representar de forma clara, ya que son 17 dimensiones, en la figura N. 38 se puede ver la posición de los perfiles respecto a 3 dimensiones: factor de personalidad X, Distancia de términos de hoja de vida Y, Factores complementarios a el perfil meta Z.



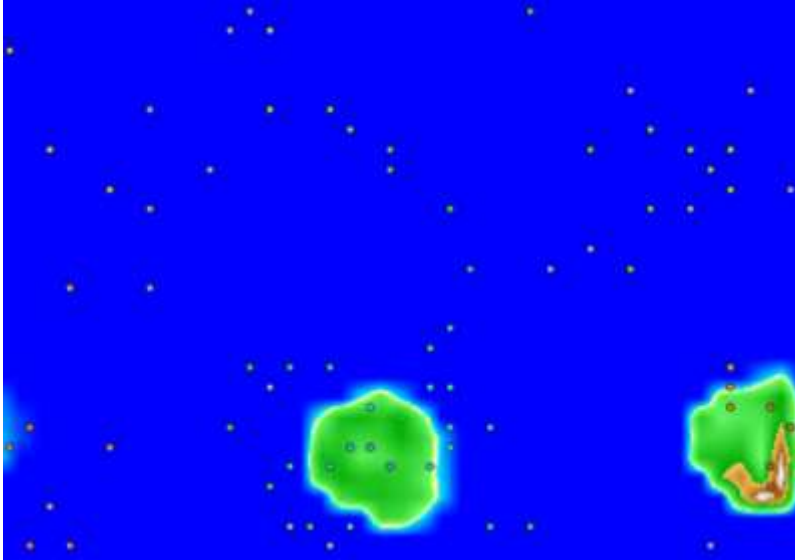
*Ilustración 38 Distancia con el perfil meta 3 dimensiones*

Cuando se crea un perfil con características bastante definidas en el conjunto de datos se generan grupos Como se puede ver en la ilustración N. 39, en este caso específico se han dado como características:

- 1) Ingeniero de sistemas
- 2) Conocimiento en bases de datos

### 3) Conocimiento en php

Los dos grupos se han formado dado que tanto como ingenieros como técnicos cumplían con los conocimientos en base de datos como en php.



*Ilustración 39 Grupos encontrados*

El perfil de cada postulado tiene una serie de características, tanto fortalezas como debilidades estas se evidencian en el grafico 40, se puede apreciar los postulados que se aproximan a el perfil idóneo definido por el personal de selección siendo el círculo rojo el perfil idóneo o meta.

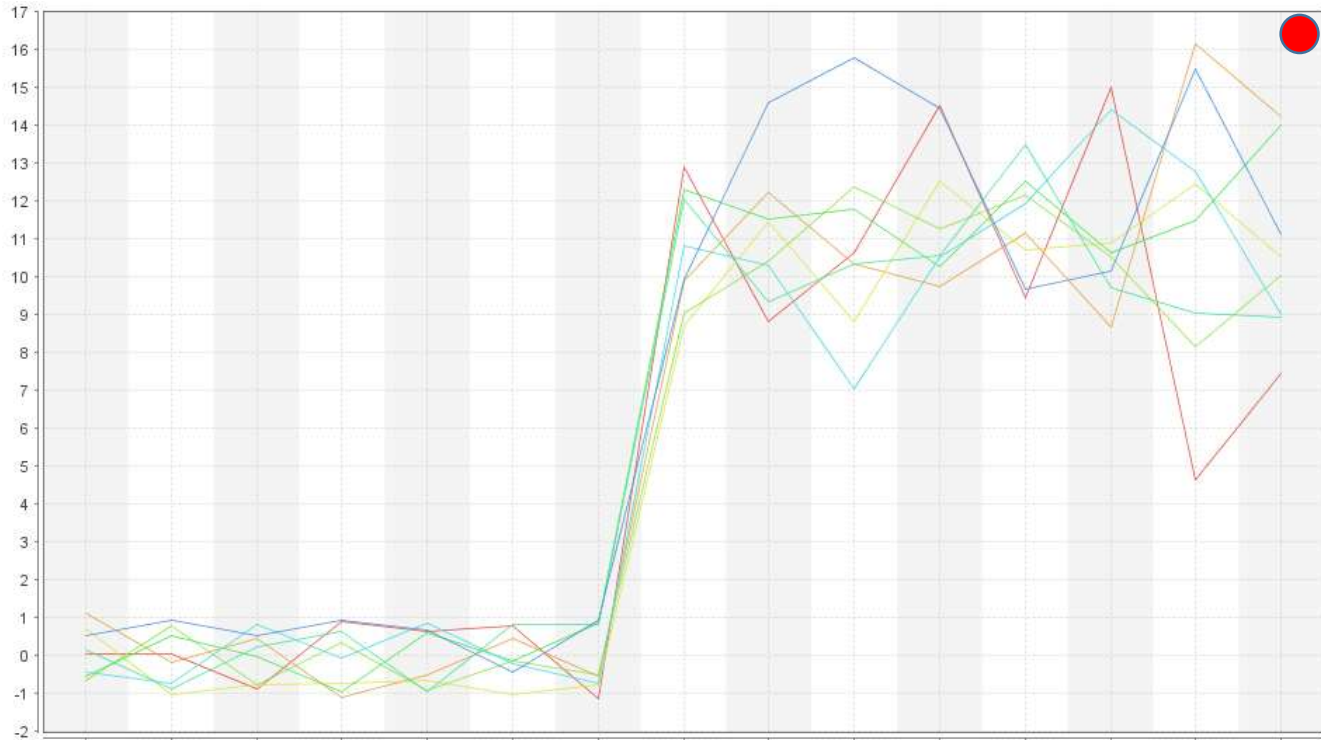


Ilustración 40 Características compatibles con el perfil idóneo

## 6. CONCLUSIONES Y RECOMENDACIONES

### CONCLUSIONES

- A lo largo del desarrollo del proyecto se observa que los métodos tradicionales de selección de personal implican errores dados las fallas de digitación o planteamiento inicial de las hojas de vida, por tanto, se hace necesario un procesamiento de cadenas de texto haciendo que estos errores tengan una menor influencia en los resultados, volviendo el modelo incluyente.
- Los métodos para medir la distancia de los postulados a un perfil idóneo planteado por el personal de selección, se comportan mejor entre más parámetros se definan como meta, de lo contrario los resultados se vuelven vagos y ambiguos, seleccionando perfiles no compatibles con el perfil meta.
- Para que el sistema funcione de manera óptima debe tener una retroalimentación constante de las decisiones que ha tomado el personal de selección, de esta forma se depuran errores que son difíciles de inferir para el modelo como lo son los perfiles sobre calificados en un perfil bajo.
- Los perfiles que tienen características de estudio o factores complementarios con un nivel medio como lo son los técnicos, se ven opacados por perfiles más altos, por lo que se debe realizar una media de los perfiles y tomar los datos mayores o menores como datos atípicos y ver de forma clara las divisiones entre perfiles altos medios y bajos, para una selección más clara .
- Se identifica que cada tipo de test realizado a los postulados a un cargo específico cumple con una serie de características según el perfil, siendo totalmente diferente en perfiles altos o en cargos básicos y llegando a obtener inferencias para posteriores perfiles de la misma organización.

## RECOMENDACIONES

- Se recomienda unificar los formatos de hoja de vida bajo el estándar propuesto, esto llevara a que en posteriores versiones del software se pueda realizar una evaluación mucho más exacta.
- Se recomienda que el sistema tenga un método de retroalimentación en el cual se premie o se castigue los pesos que se han usado en la selección por distancias de un perfil específico, de esta forma el sistema se depura y los errores se reducen progresivamente.
- Se recomienda guardar un histórico de las decisiones del personal de selección, es decir: El personal que fue contratado, tiempos y desempeño en el cargo específico en la organización poyado en el Software, de esta forma se podrán hallar patrones dentro de este conjunto de datos para retroalimentar el sistema y hacerlo mucho más exacto.
- Se recomienda el uso de cuartiles para la comprensión de las distribuciones en los datos, de esta forma visualizar los datos atípicos de una mejor forma.

## BIBLIOGRAFÍA

- Alles, M. A. (2003). *Elija al mejor : cómo entrevistar por competencias, las preguntas necesarias para una buena selección de personal*. Granica.
- American Psychological Association. (n.d.). The 16 P.F. and basic personality structure: A reply to Eysenck. Retrieved January 9, 2018, from <http://psycnet.apa.org/record/1973-20044-001>
- American Psychological Association. (n.d.). A guide to the clinical use of the 16 PF. Retrieved January 9, 2018, from <http://psycnet.apa.org/record/1977-20113-000>
- Antonio, J., Mérida, M., Luis, M., & Jorge, M. (n.d.). La Escala de Inteligencia de Binet y Simon (1905) su recepción por la Psicología posterior. Retrieved from <https://s3.amazonaws.com/academia.edu.documents/33983610/BINET.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1515712726&Signature=oCev0XwmjqU4LWsD0mi8YqoHL8w%3D&response-content-disposition=inline%3B filename%3DBINET.pdf>
- Aramburu Goya, N., & Rivera, O. (n.d.). *Organización de empresas*. Retrieved from <https://books.google.es/books?hl=es&lr=&id=jKuNuxFZU7YC&oi=fnd&pg=PA25&dq=organización+en+empresas&ots=8P58nxT96I&sig=M9NnaHmo9GrqUPww88UeCRgP4Nc#v=onepage&q=organización en empresas&f=false>
- Árbol de decisión. (n.d.). Retrieved from <http://www.utm.mx/~jahdezp/archivos/estructuras/DESICION.pdf>
- Arellano, A. S., Gil, J. A., & Martínez, A. H. (2014). El Análisis Discriminante en la Previsión de la Insolvencia en las Empresas de Seguros de no Vida. <http://dx.doi.org/10.1080/02102412.2003.10779487>. <https://doi.org/10.1080/02102412.2003.10779487>
- Argañaraz, J. P., & Entraigas, I. (n.d.). Análisis comparativo entre las máquinas de vectores soporte y el clasificador de máxima probabilidad para la discriminación de cubiertas de suelo. Retrieved from [http://www.aet.org.es/revistas/revista36/Numero36\\_03.pdf](http://www.aet.org.es/revistas/revista36/Numero36_03.pdf)
- Arthur, W., & Day, D. V. (1994). Development of a Short form for the Raven



- Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, 54(2), 394–403. <https://doi.org/10.1177/0013164494054002013>
- Autónoma De Occidente, U. (n.d.). El hombre y la máquina. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/940408/47803014.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1508861006&Signature=3Rv2TSQiTUZSPr9%2Bt1dcDgUH9Mc%3D&response-content-disposition=inline%3Bfilename%3D2008\\_-In\\_Spanish-\\_Review\\_to\\_Alejandro\\_P.pdf](https://s3.amazonaws.com/academia.edu.documents/940408/47803014.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1508861006&Signature=3Rv2TSQiTUZSPr9%2Bt1dcDgUH9Mc%3D&response-content-disposition=inline%3Bfilename%3D2008_-In_Spanish-_Review_to_Alejandro_P.pdf)
- Barsotti, C. A. (1981). LA ORGANIZACIÓN SOCIAL DE LA REPRODUCCIÓN DE LOS AGENTES SOCIALES, LAS UNIDADES FAMILIARES Y SUS ESTRATEGIAS. *Demografía Y economía*. El Colegio de Mexico. <https://doi.org/10.2307/40602275>
- Baumgartner, J., Rivero, C. R., & Pucheta, J. (n.d.). PRONÓSTICO DE LLUVIA EN UN PUNTO DESDE DIVERSOS PUNTOS GEOGRÁFICOS DE OBSERVACIÓN MEDIANTE PROCESOS GAUSSIANOS. Retrieved from [https://www.researchgate.net/profile/Julian\\_Pucheta/publication/283308087\\_Pronostico\\_de\\_lluvia\\_en\\_un\\_punto\\_desde\\_diversos\\_puntos\\_geograficos\\_de\\_observacion\\_mediante\\_Procesos\\_Gaussianos/links/5640c1a508aebaaea1f69978.pdf](https://www.researchgate.net/profile/Julian_Pucheta/publication/283308087_Pronostico_de_lluvia_en_un_punto_desde_diversos_puntos_geograficos_de_observacion_mediante_Procesos_Gaussianos/links/5640c1a508aebaaea1f69978.pdf)
- Beddoe, G. R., & Petrovic, S. (2006). Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering. *European Journal of Operational Research*, 175(2), 649–671. <https://doi.org/10.1016/j.ejor.2004.12.028>
- Bhargavi, P., Jyothi, B., Jyothi, S., & Sekar, K. (2008). Knowledge Extraction Using Rule Based Decision Tree Approach. *IJCSNS International Journal of Computer Science and Network Security*, 8(7). Retrieved from [http://paper.ijcsns.org/07\\_book/200807/20080744.pdf](http://paper.ijcsns.org/07_book/200807/20080744.pdf)
- Bray, D. W., & Campbell, R. J. (1968). Selection of salesmen by means of an assessment center. *Journal of Applied Psychology*, 52(1, Pt.1), 36–41. <https://doi.org/10.1037/h0025353>
- Bruckmann, G., & Weber, W. (Eds.). (1971). *Contributions to the Von Neumann Growth Model*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-24667-2>
- C mercer - J Dunbar. (n.d.). A conceptual model for recruitment and retention: Allied health workforce enhancement in western victoria, Australia.
- Cascio, W. F., & Silbey, V. (1979). Utility of the assessment center as a selection device. *Journal of Applied Psychology*, 64(2), 107–118. <https://doi.org/10.1037/0021-9010.64.2.107>
- CATTELL, R. B., & DREVDAHL, J. E. (1955). A COMPARISON OF

PERSONALITY PROFILE (16 P.F.) OF EMINENT RESEARCHERS WITH THAT OF EMINENT TEACHERS AND ADMINISTRATORS, AND OF GENERAL POPULATION. *British Journal of Psychology*, 46(4), 248–261. <https://doi.org/10.1111/j.2044-8295.1955.tb00547.x>

Chien, C.-F., & Chen, L.-F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications*, 34(1), 280–290. <https://doi.org/10.1016/j.eswa.2006.09.003>

Collins, J. C., & Porras, J. I. (1996). Building Your Company ' S Vision Making Differences Matter : a New Paradigm for Managing Diversity Breaking the Functional Mind-Set in the Pitfalls of Parenting Mature Companies. *Harvard Business Review*, 74(5), 65. <https://doi.org/Article>

Corral, S. (2004). *TPT : test de personalidad de Tea : manual*. TEA. Retrieved from [https://books.google.es/books?hl=es&lr=&id=7P\\_LpUQYb5AC&oi=fnd&pg=PA5&dq=test+de+personalidad&ots=yjpvRouRmc&sig=-Sp0N4A4kpk3Pd7Bg-0W4e\\_Bw#v=onepage&q=test+de+personalidad&f=false](https://books.google.es/books?hl=es&lr=&id=7P_LpUQYb5AC&oi=fnd&pg=PA5&dq=test+de+personalidad&ots=yjpvRouRmc&sig=-Sp0N4A4kpk3Pd7Bg-0W4e_Bw#v=onepage&q=test+de+personalidad&f=false)

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>

Delgado, M., Ruiz, M. D., & Sánchez, D. (2008). Reglas de Asociación Difusas: Nuevos Retos. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/32257345/2008\\_estilf-pres-nuevos-retos.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1510530496&Signature=8OAr3cspcHedZ4bsWiM%2BGzKjTBc%3D&response-content-disposition=inline%3B+filename%3D2008\\_estilf-pres-nuevos-retos.pdf](https://s3.amazonaws.com/academia.edu.documents/32257345/2008_estilf-pres-nuevos-retos.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1510530496&Signature=8OAr3cspcHedZ4bsWiM%2BGzKjTBc%3D&response-content-disposition=inline%3B+filename%3D2008_estilf-pres-nuevos-retos.pdf)

DW Rolston, AP Gama, I. Z.-1990. (n.d.). Principios de inteligencia artificial y sistemas expertos.

Edwards, W., & Fasolo, B. (2001). Decision Technology. *Annual Review of Psychology*, 52(1), 581–606. <https://doi.org/10.1146/annurev.psych.52.1.581>

El impacto del uso efectivo de las TIC sobre la eficiencia técnica de las empresas españolas. (2007). *Estudios Gerenciales*, 23(103), 65–84. [https://doi.org/10.1016/S0123-5923\(07\)70010-4](https://doi.org/10.1016/S0123-5923(07)70010-4)

Erro, C. (2003). *Historia empresarial : pasado, presente y retos de futuro*. Ariel. Retrieved from <http://centaur.reading.ac.uk/21424/>

Estévez, P., & Resumen, V. \*. (n.d.). OPTIMIZACIÓN MEDIANTE ALGORITMOS

- GENÉTICOS. Retrieved from  
[https://www.researchgate.net/profile/Pablo\\_Estevez/publication/228708779\\_Optimizacion\\_Mediante\\_Algoritmos\\_Geneticos/links/0912f51111f82b2a61000000.pdf](https://www.researchgate.net/profile/Pablo_Estevez/publication/228708779_Optimizacion_Mediante_Algoritmos_Geneticos/links/0912f51111f82b2a61000000.pdf)
- ESTUDIOS. filosofía-historia-letras, & Otoño 1987. (n.d.). Breve historia de la Inteligencia Artificial. Retrieved October 24, 2017, from  
[https://biblioteca.itam.mx/estudios/estudio/estudio10/sec\\_16.html](https://biblioteca.itam.mx/estudios/estudio/estudio10/sec_16.html)
- Faerber, F., Weitzel, T., Keim, T., & Färber, F. (2003). Association for Information Systems AIS Electronic Library (AISeL) An Automated Recommendation Approach to Selection in Personnel Recruitment AN AUTOMATED RECOMMENDATION APPROACH TO SELECTION IN PERSONNEL RECRUITMENT. Retrieved from  
<http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1768&context=amcis2003>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37.  
<https://doi.org/10.1609/AIMAG.V17I3.1230>
- Fernando, J., Gutiérrez, M., Fernando, L., Velandia, M., De, B., & República, L. (2011). Pronóstico de incumplimientos de pago mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito. Retrieved from  
[https://www.researchgate.net/profile/Luis\\_Melo8/publication/254392485\\_Pronostico\\_de\\_incumplimientos\\_de\\_pago\\_mediante\\_maquinas\\_de\\_vectores\\_de\\_soporte\\_una\\_aproximacion\\_inicial\\_a\\_la\\_gestion\\_del\\_riesgo\\_de\\_credito/links/541904820cf203f155adb8b.pdf](https://www.researchgate.net/profile/Luis_Melo8/publication/254392485_Pronostico_de_incumplimientos_de_pago_mediante_maquinas_de_vectores_de_soporte_una_aproximacion_inicial_a_la_gestion_del_riesgo_de_credito/links/541904820cf203f155adb8b.pdf)
- Ferrando, P. J. C. E. T. J. M. (1989). *Psicothema*. *Psicothema* (Vol. 14). Universidad de Oviedo. Retrieved from  
<http://www.redalyc.org/html/727/72714325/>
- Fogel, D. B., IEEE Neural Networks Council., & John Wiley & Sons. (2006). *Evolutionary computation : toward a new philosophy of machine intelligence*. John Wiley & Sons.
- Foster, M. (1961). An Application of the Wiener-Kolmogorov Smoothing Theory to Matrix Inversion. *Journal of the Society for Industrial and Applied Mathematics*, 9(3), 387–392. <https://doi.org/10.1137/0109031>
- Frías Blanco, I. I., Ortiz Díaz, A. A., Ramos Jimenez, G., Morales Bueno, R., & Caballero Mota, Y. (2010). CLASIFIers and ensemble of classifiers based on decision trees that handle concept drift. *INTELIGENCIA ARTIFICIAL*, 14(45).  
<https://doi.org/10.4114/ia.v14i45.1080>
- Furnham, A. (1990). The fakeability of the 16 PF, Myers-Briggs and FIRO-B personality measures. *Personality and Individual Differences*, 11(7), 711–716.

[https://doi.org/10.1016/0191-8869\(90\)90256-Q](https://doi.org/10.1016/0191-8869(90)90256-Q)

- García García, J., Puga, J. L., Jesús, C., Guillén, C., Belén, A., Segura, G., ... Sánchez, F. (n.d.). APLICACIÓN DE LAS REDES BAYESIANAS AL MODELADO DE LAS ACTITUDES EMPRENDEDORAS. Retrieved from [http://jpuga.es/Publicaciones/Congresos/RB\\_MODEL\\_EMP.pdf](http://jpuga.es/Publicaciones/Congresos/RB_MODEL_EMP.pdf)
- González, J., & Morales, E. (n.d.). REGLAS DE ASOCIACIÓN. Retrieved from <https://ccc.inaoep.mx/~jagonzalez/ML/ReglasA.pdf>
- Gorsuch, R. L., & Cattell, R. B. (1967). Second Stratum Personality Factors Defined In The Questionnaire Realm By The 16 P.F. *Multivariate Behavioral Research*, 2(2), 211–223. [https://doi.org/10.1207/s15327906mbr0202\\_5](https://doi.org/10.1207/s15327906mbr0202_5)
- Guevara, R. D., & Vargas, A. (n.d.). Intervalos de confianza para los índices de capacidad Cpm y Cpmk en procesos estacionarios gaussianos. Retrieved from [http://www.emis.ams.org/journals/RCE/ingles/V29/V29\\_2\\_153GuevaraVargas.pdf](http://www.emis.ams.org/journals/RCE/ingles/V29/V29_2_153GuevaraVargas.pdf)
- Gutiérrez Martínez, I., & Bello Pérez Andrés Tellería Rodríguez, R. E. (2002). UN SISTEMA BASADO EN CASOS PARA LA TOMA DE DECISIONES EN CONDICIONES DE INCERTIDUMBRE. *REVISTA INVESTIGACION OPERACIONAL*, 23(2). Retrieved from [https://s3.amazonaws.com/academia.edu.documents/46787722/IO-23202-1\\_1.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1510464052&Signature=qQJQ9alz4XUyKISSsaYT7A%2FUrDo%3D&response-content-disposition=inline%3Bfilename%3DUN\\_SISTEMA\\_BASADO\\_EN\\_CASOS\\_PARA\\_LA\\_TOMA.pdf](https://s3.amazonaws.com/academia.edu.documents/46787722/IO-23202-1_1.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1510464052&Signature=qQJQ9alz4XUyKISSsaYT7A%2FUrDo%3D&response-content-disposition=inline%3Bfilename%3DUN_SISTEMA_BASADO_EN_CASOS_PARA_LA_TOMA.pdf)
- Guzmán Evangelista, F. S. (2017). Propiedades Psicométricas del Cuestionario de Madurez Psicológica – PSYMAS en adolescentes de colegios públicos de Otuzco. *Universidad César Vallejo*. Retrieved from <http://181.224.246.201/handle/UCV/678>
- Hamidian, L. B., Soto, L. G., & Poriet, L. Y. (n.d.). PLATAFORMAS VIRTUALES DE APRENDIZAJE: UNA ESTRATEGIA INNOVADORA EN PROCESOS EDUCATIVOS DE RECURSOS HUMANOS. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/32537167/266.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1515489455&Signature=XS%2BzYyfaYwajaW7OpEusSmWVvoU%3D&response-content-disposition=inline%3Bfilename%3DPLATAFORMAS\\_VIRTUALES\\_DE\\_APRENDIZAJE\\_UNA.pdf](https://s3.amazonaws.com/academia.edu.documents/32537167/266.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1515489455&Signature=XS%2BzYyfaYwajaW7OpEusSmWVvoU%3D&response-content-disposition=inline%3Bfilename%3DPLATAFORMAS_VIRTUALES_DE_APRENDIZAJE_UNA.pdf)
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. San Mateo, CA: Morgan Kaufmann Publishers) Helmi, A., & de Zeeuw, P. T. 2000, *MNRAS* (Vol. 319657). Elsevier/Morgan Kaufmann. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-)

ANIE9823>3.3.CO;2-C

- Hermelin, E., Lievens, F., & Robertson, I. T. (2007). The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A meta-analysis. *International Journal of Selection and Assessment*, 15(4), 405–411. <https://doi.org/10.1111/j.1468-2389.2007.00399.x>
- Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). *Introducción a la minería de datos*. Pearson Prentice Hall. Retrieved from <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=UCC.xis&method=post&formato=2&cantidad=1&expresion=mfn=064234>
- Hilera González, J. R., & Martínez Hernando, V. J. (2000). *Redes neuronales artificiales :fundamentos, modelos y aplicaciones /*. Retrieved from <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=AGRIUAN.xis&method=post&formato=2&cantidad=1&expresion=mfn=014157>
- Howarth, E., & Browne, J. A. (1971). (n.d.). An item-factor-analysis of the 16 PF. Retrieved January 11, 2018, from <http://psycnet.apa.org/record/1972-06813-001>
- Isasi Viñuela, P., & Galván León, I. M. (2004). *Redes de neuronas artificiales : un enfoque práctico*. Prentice Hall. Retrieved from <http://go.galegroup.com/ps/anonymous?id=GALE%7CA146892307&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=16666046&p=AONE&sw=w&authCount=1&isAnonymousEntry=true>
- Isaza, C., Rios, W., & Mosquera, J. (2008). Modelo de maduración de fruto de banano empleando procesos gaussianos de regresión y filtros acusto-ópticos. *Scientia et Technica*, 1(38). <https://doi.org/10.22517/23447214.3785>
- Ivanovic Marincovich, Rodolfo; Forno Sparosvich, Hernán; Durán Santana, María Cristina; Hazbún Game, Julia; Castro Gómez, C. I. M. (n.d.). Estudio de la capacidad intelectual (test de Matrices Progresivas de Raven). Retrieved January 13, 2018, from <http://europa.sim.ucm.es/compludoc/AA?articulold=97596>
- Izaurieta, F., & Saavedra, C. (n.d.). Redes Neuronales Artificiales. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/36957207/Redes\\_neuronales.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1508448349&Signature=FZyj7vLpAlvl%2BsXA6zvgxa0st8g%3D&response-content-disposition=inline%3B filename%3DRedes\\_Neuronales\\_Artificiales.pdf](https://s3.amazonaws.com/academia.edu.documents/36957207/Redes_neuronales.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1508448349&Signature=FZyj7vLpAlvl%2BsXA6zvgxa0st8g%3D&response-content-disposition=inline%3B filename%3DRedes_Neuronales_Artificiales.pdf)
- J. Contreras, Member IEEE, R. M. y L. U. (n.d.). Algoritmos para Identificación de Modelos Difusos Interpretables. Retrieved from [https://s3.amazonaws.com/academia.edu.documents/44027403/Algoritmos\\_p](https://s3.amazonaws.com/academia.edu.documents/44027403/Algoritmos_p)

ara\_Identificacin\_de\_Modelos20160323-25553-1yqq4pg.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1510537802&Signature=vR%2F4ru9cmUYHy94%2FhDG8Q%2BXhQIM%3D&response-content-disposition=

- Jaime Regueró Alvarez Directora Julia Díaz García, A. (n.d.). Aplicación de las redes bayesianas dinámicas a la predicción de series de datos y a la detección de anomalías. Retrieved from [http://afrodita.ii.uam.es/nueva\\_web/intranet/ga/tfdm/trabajos/Reguero\\_Alvarez\\_Jaime.pdf](http://afrodita.ii.uam.es/nueva_web/intranet/ga/tfdm/trabajos/Reguero_Alvarez_Jaime.pdf)
- Jofuku, K. D., Schipper, R. D., & Goldberg, R. B. (1989). A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *The Plant Cell*, 1(4), 427–35. <https://doi.org/10.1105/tpc.1.4.427>
- Jorge López Puga<sup>1\*\*</sup>, Juan García García<sup>1\*\*</sup>, L. de la F. S. de M. S. de P. e I. C., & SPARC (Organization). (1984). *Las redes bayesianas como herramientas de modelado en psicología\**. *Anales de Psicología* (Vol. 23). Secretariado de Publicaciones e Intercambio Científico. Retrieved from <http://www.redalyc.org/html/167/16723218/>
- Kelemenis, A., & Askounis, D. (2010). A new TOPSIS-based multi-criteria approach to personnel selection. *Expert Systems with Applications*, 37(7), 4999–5008. <https://doi.org/10.1016/j.eswa.2009.12.013>
- L, U. del Z. D. de C. H. F. U. del Z. F. E. de C. (n.d.). Universidad del Zulia. Departamento de Ciencias Humanas. F Universidad del Zulia. Facultad Experimental de Ciencias. L. Retrieved from <http://pendientedemigracion.ucm.es/info/pslogica/rodolfo2.pdf>
- Linking Bayesian networks and PLS path modeling for causal analysis. (2010). *Expert Systems with Applications*, 37(1), 134–139. <https://doi.org/10.1016/J.ESWA.2009.05.021>
- Lucid. (n.d.). Qué es un diagrama de árbol de decisión | Lucidchart. Retrieved December 12, 2017, from <https://www.lucidchart.com/pages/es/qué-es-un-diagrama-de-árbol-de-decisión>
- Ma Jesús Montes Alonso y Pablo González Rodríguez. (n.d.). Selección de Personal - “Ma Jesús Montes Alonso y Pablo González Rodríguez” - Google Libros. Retrieved January 11, 2018, from <https://books.google.es/books?hl=es&lr=&id=0jkELJ2nfVwC&oi=fnd&pg=PA1&dq=selección+de+personal&ots=cH8ZZrC7aX&sig=YfQUdMJ1jcS8cwPtB18zPxXU9Eo#v=onepage&q=selección+de+personal&f=false>
- MACAN, T. H., AVEDON, M. J., PAESE, M., & SMITH, D. E. (1994). THE EFFECTS OF APPLICANTS' REACTIONS TO COGNITIVE ABILITY TESTS AND AN ASSESSMENT CENTER. *Personnel Psychology*, 47(4), 715–738.

<https://doi.org/10.1111/j.1744-6570.1994.tb01573.x>

- Machover, K., & Gutiérrez, J. M. ,tr. (n.d.). Proyección de la personalidad en el dibujo de la figura humana :un método de investigación de la personalidad /. S.n,. Retrieved from <http://www.sidalc.net/cgi-bin/wxis.exe/?IisScript=UCC.xis&method=post&formato=2&cantidad=1&expresion=mfn=122569>
- Mansuri, I. R., & Sarawagi, S. (2006). Integrating Unstructured Data into Relational Databases. In *22nd International Conference on Data Engineering (ICDE'06)* (pp. 29–29). IEEE. <https://doi.org/10.1109/ICDE.2006.83>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity.
- Marcos Gestal- Daniel Rivero - Alejandro Pazos. (n.d.). Introducción a los algoritmos Genéticos y la programación Genética. Retrieved from <http://www.galeon.com/dantethedestroyer/algoritmos.pdf>
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data : la revolución de los datos masivos*. Turner. Retrieved from [https://books.google.es/books?hl=es&lr=&id=uO9FbEcaMpkC&oi=fnd&pg=PA11&dq=que+es+big+data&ots=VZx\\_7jqFD\\_&sig=afynnrB9cKYh7w7rJb8jvh87T\\_U#v=onepage&q=que es big data&f=false](https://books.google.es/books?hl=es&lr=&id=uO9FbEcaMpkC&oi=fnd&pg=PA11&dq=que+es+big+data&ots=VZx_7jqFD_&sig=afynnrB9cKYh7w7rJb8jvh87T_U#v=onepage&q=que es big data&f=false)
- Miguel De Bortoli1, Patricia Barrios y Rosana Azpiroz (Universidad Nacional de San Luis, A. (2004). *Relaciones entre los potenciales evocados cognitivos auditivos y el Test de Matrices Progresivas de Raven*. [publisher not identified]. Retrieved from <http://www.redalyc.org/html/337/33720207/>
- Morales, Á. K., & Casas, J. G. (n.d.). ALGORITMOS GENÉTICOS. Retrieved from [https://b35e8169-a-62cb3a1a-sites.googlegroups.com/site/unamfcienciascomputoevolutivo/materials/Kuri-Galaviz-AG.pdf?attachauth=ANoY7co83RwZXBFPBycLT6RYgBJRsznFidT2BWGUUnemdfIsMGLXnBkJwbGZEVDTwK3M3sxxg4LUUAscN2pAi8wRbDtUdzxdI8QT9ikot-hjdUthkJEgH2Hj0kUa5ZKeSRffvYITyejgud1v1A4guCGWZuB4AQSxIPMiUXATluXZ3VbPbOhcT96gJXIJ2BORrRzr1aUwVetxhZXXC5mcPWw458z5H\\_MY2Ra9T9VEIF0x0NVwQ0FKBR5N8z4Gb557fyWFQ\\_\\_vxqN&attredirects=0](https://b35e8169-a-62cb3a1a-sites.googlegroups.com/site/unamfcienciascomputoevolutivo/materials/Kuri-Galaviz-AG.pdf?attachauth=ANoY7co83RwZXBFPBycLT6RYgBJRsznFidT2BWGUUnemdfIsMGLXnBkJwbGZEVDTwK3M3sxxg4LUUAscN2pAi8wRbDtUdzxdI8QT9ikot-hjdUthkJEgH2Hj0kUa5ZKeSRffvYITyejgud1v1A4guCGWZuB4AQSxIPMiUXATluXZ3VbPbOhcT96gJXIJ2BORrRzr1aUwVetxhZXXC5mcPWw458z5H_MY2Ra9T9VEIF0x0NVwQ0FKBR5N8z4Gb557fyWFQ__vxqN&attredirects=0)
- Ozzie, R. E. (2006). Personal data mining. Retrieved from <https://www.google.com/patents/US7930197>
- Pablo Felgaer. (n.d.). Optimización de Redes Bayesianas basado en Técnicas de Aprendizaje por Inducción. Retrieved from <https://pdfs.semanticscholar.org/71a9/59860bca578976393a224ed71f30c7916610.pdf>

- PAGUAY VARGAS, E. N. (2016). ELABORACIÓN Y SOCIALIZACIÓN DE UN MANUAL DE SELECCIÓN DE PERSONAL BASADO EN UN PERFIL DE COMPETENCIAS PARA DISMINUIR EL ALTO ÍNDICE DE ROTACIÓN DE PERSONAL EN LA EMPRESA LABORATORIO DENTAL EMIDENT UBICADA EN EL SECTOR NORTE DEL DISTRITO METROPOLITANO DE QUITO 2016. Retrieved from <http://www.dspace.cordillera.edu.ec/handle/123456789/2190>
- Pardon, W., & Stern, M. (2001). Pure Hodge structure on the L2 -cohomology of varieties with isolated singularities. *Journal Für Die Reine Und Angewandte Mathematik (Crelles Journal)*, 2001(533), 55–80. <https://doi.org/10.1515/crll.2001.026>
- Parra, D. M., Blasco, C., Nuria, C., Buixo, M., Ponz, E., Joan, C., ... Parra, M. (2005). THE INFLUENCE OF PERSONALITY TRAITS IN THE CHOICE OF DIALYSIS TECHNIQUE Influencia de los rasgos de personalidad en la elección de la técnica de diálisis. *REVISTA Rev Soc Esp Enferm Nefrol*, 8(81), 13–17. Retrieved from <http://scielo.isciii.es/pdf/nefro/v8n1/art02.pdf>
- Pascual, D., Pla, F., & Sánchez, S. (n.d.). Algoritmos de agrupamiento. Retrieved from [http://marmota.dlsi.uji.es/WebBIB/papers/2007/1\\_Pascual-MIA-2007.pdf](http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf)
- Peña, D. (2002). ANÁLISIS DE DATOS MULTIVARIANTES. Retrieved from [https://www.researchgate.net/profile/Daniel\\_Pena4/publication/40944325\\_Analisis\\_de\\_Datos\\_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf](https://www.researchgate.net/profile/Daniel_Pena4/publication/40944325_Analisis_de_Datos_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf)
- Piessens, R., Doncker-Kapenga, E., Überhuber, C. W., & Kahaner, D. K. (1983). *Quadpack : a Subroutine Package for Automatic Integration*. Springer Berlin Heidelberg.
- Pitarque, A., Roy, J. F., & Ruiz, J. C. (1998). Redes neurales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. *Psicológica*, 19, 387–400. Retrieved from <https://www.uv.es/psicologica/articulos3.98/pitarque.pdf>
- Pitarque, A., Ruiz, J. C., & Roy, J. F. (2000). Las redes neuronales como herramientas estadísticas no paramétricas de clasificación. *Psicothema*, 12(Suplemento), 459–463. Retrieved from <https://www.uniovi.es/reunido/index.php/PST/article/view/7730>
- Pla Bañón José Salvador Sánchez Garreta, F. (n.d.). Algoritmos de Agrupamiento basados en densidad y Validación de clusters. Retrieved from <http://www.cerpamid.co.cu/sitio/files/DamarisTesis.pdf>
- REILLY, R. T. C. & T. (2014). Making Hard Decisions with DecisionTools - Robert T. Clemen, Terence Reilly - Google Libros. Retrieved August 28, 2017, from <https://books.google.es/books?hl=es&lr=&id=CNMbBQAAQBAJ&oi=fnd&pg=P>



P1&dq=+Making+decisions+in+companies&ots=WRIw-ojmFw&sig=HBmjCeBbpdfW2ChdbRFq\_4zfc6Y#v=onepage&q=Making decisions in companies&f=false

- Render, B., & Heizer, J. (2004). *Principios de administración de operaciones*. Pearson Educación. Retrieved from <https://books.google.es/books?hl=es&lr=&id=jVlwSsVHUfAC&oi=fnd&pg=PA1&dq=+Árboles+de+decisiones+++&ots=FpA4aT5o-B&sig=T23qIDjDL19Q1PAfLC84iKaLRrs#v=onepage&q=Árboles de decisiones&f=false>
- Rossi Casé, L. E., Neer, R. H., Lopetegui, M. S., Doná, S. M., Biganzoli, B., Farinon, E., & Garzaniti, R. (2013). Test de Raven. Retrieved from <http://sedici.unlp.edu.ar/handle/10915/45565>
- Ruigrok, W., Peck, S. I., & Keller, H. (2006). Board characteristics and involvement in strategic decision making: Evidence from Swiss companies. *Journal of Management Studies*, 43(5), 1201–1226. <https://doi.org/10.1111/j.1467-6486.2006.00634.x>
- Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (n.d.). Design and Analysis of Computer Experiments. *Statistical Science*. Institute of Mathematical Statistics. <https://doi.org/10.2307/2245858>
- Saidi Mehrabad, M., & Fathian Brojeny, M. (2007). The development of an expert system for effective selection and appointment of the jobs applicants in human resource management. *Computers & Industrial Engineering*, 53(2), 306–312. <https://doi.org/10.1016/j.cie.2007.06.023>
- Sánchez Chávez, J. A. (2017). Gestión de Recursos Humanos : Desarrollo de los Recursos Humanos en las Organizaciones. Retrieved from <http://repositorio.unan.edu.ni/3930/>
- Simon, H. A. (n.d.). Theories of Decision-Making in Economics and Behavioral Science. *The American Economic Review*. American Economic Association. <https://doi.org/10.2307/1809901>
- SPARC (Organization), PATRICIA; TREJOS, ALVARO; SOTO MEJÍA, J., & Universidad Tecnológica de Pereira. (1995). *APLICACIÓN DEL ANALISIS DISCRIMINANTE PARA EXPLORAR LA RELACIÓN ENTRE EL EXAMEN DE ICFES Y EL RENDIMIENTO EN ALGEBRA LINEAL DE LOS ESTUDIANTES DE INGENIERÍA DE LA UTP EN EL PERÍODO 2001-2003*. *Scientia Et Technica* (Vol. X). Universidad Tecnológica de Pereira. Retrieved from <http://www.redalyc.org/html/849/84911685034/>
- Storey Hooper, R., Galvin, T. P., Kilmer, R. A., & Liebowitz, J. (1998). Use of an expert system in a personnel selection process<sup>1</sup>The views expressed in this article are those of the authors and do not necessarily reflect the official policy

- or position of the Department of the Army, Department of Defense or the U.S. Government.1. *Expert Systems with Applications*, 14(4), 425–432.  
[https://doi.org/10.1016/S0957-4174\(98\)00002-5](https://doi.org/10.1016/S0957-4174(98)00002-5)
- Sucar, L. E. (n.d.). Redes Bayesianas. Retrieved from  
<https://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>
- Tablada, C. J., & Torres, G. A. (2009). Redes Neuronales Artificiales. *Revista de Educación Matemática*, 24(3). Retrieved from  
<https://revistas.unc.edu.ar/index.php/REM/article/view/10280>
- Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2005). Recognizing Partially Occluded, Expression Variant Faces From Single Training Image per Person With SOM and Soft<math>\text{&gt;}\\$k\\$&lt;/math>-NN Ensemble. *IEEE Transactions on Neural Networks*, 16(4), 875–886.  
<https://doi.org/10.1109/TNN.2005.849817>
- Trillas, E. (1977). El sistema de las sociedades modernas TALCOTT PARSONS, 13–41. Retrieved from  
[https://s3.amazonaws.com/academia.edu.documents/32510959/talcott-parsons-el-sistema-de-las-sociedades-modernas.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1509175896&Signature=%2F3wpVWBUE8wLYdQaEI9nUdwy%2BBc%3D&response-content-disposition=inline%3Bfilename%3DEI\\_sistema\\_de\\_las\\_sociedades\\_modernas\\_TA.pdf](https://s3.amazonaws.com/academia.edu.documents/32510959/talcott-parsons-el-sistema-de-las-sociedades-modernas.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1509175896&Signature=%2F3wpVWBUE8wLYdQaEI9nUdwy%2BBc%3D&response-content-disposition=inline%3Bfilename%3DEI_sistema_de_las_sociedades_modernas_TA.pdf)
- Turing, A. M. (n.d.). Computing Machinery and Intelligence. *Mind*. Oxford University PressMind Association. <https://doi.org/10.2307/2251299>
- Tuya, J., Ramos Román, I., & Dolado Cosín, J. (2007). *Técnicas cuantitativas para la gestión en la ingeniería del software*. Netbiblo. Retrieved from  
<https://books.google.es/books?hl=es&lr=&id=PZQoZ9KTNaEC&oi=fnd&pg=PR15&dq=Redes+bayesianas&ots=P0VD1AfjIV&sig=1zo20pZepva2f1dnoScLzOQXgdQ#v=onepage&q=Redes bayesianas&f=false>
- Universidad de Murcia. Facultad de Filosofía y Ciencias de la Educación. Sección de Psicología., J. G. G. J. F. S. L. de la; de la F. S. E. I., Universidad de Murcia. Secretariado de Publicaciones e Intercambio Científico., & SPARC (Organization). (1984). *Anales de psicología : revista de la Facultad de Filosofía y Ciencias de la Educación, (Sección de Psicología), Universidad de Murcia. Anales de Psicología (Vol. 23)*. Secretariado de Publicaciones e Intercambio Científico. Retrieved from  
<http://www.redalyc.org/html/167/16723218/>
- Universidad del Zulia. Departamento de Ciencias Humanas., F., Universidad del Zulia. Facultad Experimental de Ciencias., L., MORENO, M., & MARTÍN, P. (2002). *Opción. Revista Colombiana de Computación (Vol. 3)*. Universidad del Zulia, Facultad Experimental de Ciencias, Oficina de Publicaciones

- Científicas. Retrieved from <http://www.redalyc.org/articulo.oa?id=31005208>
- Universidad del Zulia. Facultad de Ciencias Económicas y Sociales, Gerardo J.; Canet Giner, M. T. (2012). *Tipos de trabajo y la formación de la especialización de tareas en la organización*. *Revista de Ciencias Sociales (Ve)* (Vol. XVIII). Universidad del Zulia. Retrieved from <http://www.redalyc.org/html/280/28022785005/>
- Universidad Tecnológica de Pereira., G. A. (2005). *Scientia et technica*. *Scientia et technica* (Vol. 1). Retrieved from <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/6895/4139>
- Utilización de las Redes Neuronales en la toma de decisiones: aplicación en un problema de secuenciación. (n.d.). Retrieved from [https://www.researchgate.net/profile/David\\_Fuente/publication/28124933\\_Utilizacion\\_de\\_las\\_Red\\_Neuronales\\_en\\_la\\_toma\\_de\\_decisiones\\_aplicacion\\_en\\_un\\_problema\\_de\\_secuenciacion/links/00b4953803a3118e9b000000/Utilizacion-de-las-Redes-Neuronales-en-la-toma-de](https://www.researchgate.net/profile/David_Fuente/publication/28124933_Utilizacion_de_las_Red_Neuronales_en_la_toma_de_decisiones_aplicacion_en_un_problema_de_secuenciacion/links/00b4953803a3118e9b000000/Utilizacion-de-las-Redes-Neuronales-en-la-toma-de)
- Valero Orea, S., Salvador Vargas, A., & García Alonso, M. (n.d.). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Retrieved from <https://pdfs.semanticscholar.org/0d9b/8031e1ce7d89e518e1f875873bb5436503b7.pdf>
- Ver Hoef, J. M., & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, 25(2), 219–240. <https://doi.org/10.1007/BF00893273>
- Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, 8(3), 247–261. <https://doi.org/10.1037/h0074898>
- Yolis, E., Yolis, E., Britos, P., Perichisky, G., & García-Martínez, R. (2003). Algoritmos Genéticos Aplicados a la Categorización Automática de Documentos. *Revista Eletrônica de Sistemas de Informação* <BR>ISSN 1677-3071 doi:10.21529/RESI, 2(2). <https://doi.org/10.5329/133>
- Young, A. (1965). MODELS FOR PLANNING RECRUITMENT AND PROMOTION OF STAFF. *British Journal of Industrial Relations*, 3(3), 301–310. <https://doi.org/10.1111/j.1467-8543.1965.tb00908.x>