

## Selección de personal utilizando técnicas de minería de datos

Ing. Ivan Alarcón Catalán<sup>1</sup>, MTI. Rafael Hernández Reina<sup>2</sup>, MTI. Eloy Cadena Mendoza<sup>3</sup>, MTI. Juan Miguel Hernández Bravo<sup>4</sup>, MTI. Jorge Carranza Gómez<sup>5</sup>, Ing. Armando Montes Torres.<sup>6</sup>

**Resumen**— El presente trabajo se enfoca en la aplicación de técnicas de minería de datos para la selección de personal, centrándose en personas interesadas a ingresar al Sistema Nacional de Bachillerato (SNB) para ocupar un puesto administrativo en el estado de Guerrero. Propuesta que analiza resolver el problema usando este tipo de técnicas, debido a que se deben tomar en cuenta las pruebas básicas de recursos humanos para cubrir los perfiles de los puestos, los cuales ya se encuentran definidos por la Dirección General de Educación Tecnológica Industrial (DGETI). Aplicando las herramientas y técnicas adecuadas es posible establecer parámetros para evaluarlos y así poder identificar a los más calificados de acuerdo a lo que se esté buscando y a las necesidades a cubrir en las instituciones de la entidad.

**Palabras clave**— Minería de datos, selección de personal, perfiles, toma de decisiones.

### Introducción

El objetivo de este trabajo, es el análisis para el desarrollo de un sistema que aplique minería de datos y que ayude a seleccionar de entre miles de registros a aquellos que cumplan cierto criterio pre-establecido de búsqueda sobre la base de datos de candidatos a ingresar a la DGETI en el estado de Guerrero específicamente al CETis No. 116.

En la actualidad no se cuenta con ninguna estrategia de contratación en el medio, es por ello que surge la necesidad de elaborar una herramienta que en condiciones de incertidumbre nos permita clasificar y minimizar los riesgos al momento de elegir un candidato. Al mismo tiempo el sistema nos brinde información de la forma en que el candidato se relacionará con los futuros compañeros laborales.

Por tal motivo, crear la herramienta de selección de candidatos a ingresar al sistema educativo, ayudará a mejorar el proceso dirigido a la cobertura de necesidades que puedan existir en el plantel educativo haciendo más ágil el proceso de selección. Dicha herramienta nos permitirá incorporar información o atributos ambiguos o subjetivos y aunque sea imprecisa esta información, se podrá utilizar como manera determinística en el proceso de selección.

Para ello, a la hora de valorar los talentos o aptitudes del individuo será necesario no solo el perfil del puesto y el compromiso que este conlleve en sus tareas a desarrollar, será necesario también tomar en cuenta las cualidades de los candidatos al igual que las distintas personalidades de cada uno de ellos.

Esto se logra, realizando test de aptitudes, cuestionarios de personalidad, psicométricos, pruebas situacionales y entrevistas, esto independientemente de los datos que se sustraen de una encuesta de trabajo. En este trabajo, se ha determinado el uso del algoritmo determinado C4.5, el cual nos provee de una clasificación mediante valores previamente definidos y que de acuerdo a la experiencia que ya se tiene de contrataciones pasadas es que nos ayuda a determinar los candidatos factibles a contratación.

Algunos elementos que sirven de base para el análisis son los siguientes:

---

<sup>1</sup> Ing. Ivan Alarcón Catalán, estudiante de la Maestría en Sistemas Computacionales, [ivan\\_alarcon@hotmail.com](mailto:ivan_alarcon@hotmail.com), teléfono: 744-2213996, Domicilio: Col. Granjas del Márquez, CP. 39890, Acapulco Gro.

<sup>2</sup> Docente en Instituto Tecnológico de Acapulco, Depto. de Sistemas y Computación, [rhernan7@yahoo.com](mailto:rhernan7@yahoo.com), Tel. 744-4099427, Acapulco, Gro.

<sup>3</sup> Docente en Instituto Tecnológico de Acapulco, Depto. de Sistemas y Computación, [Eloy\\_cadena@yahoo.com](mailto:Eloy_cadena@yahoo.com), Acapulco, Gro.

<sup>4</sup> Docente en Instituto Tecnológico de Acapulco, Depto. de Sistemas y Computación, [jmhernan@yahoo.com](mailto:jmhernan@yahoo.com), Acapulco, Gro.

<sup>5</sup> Docente en Instituto Tecnológico de Acapulco, Depto. De Sistemas y Computación, [jcarranzamx@hotmail.com](mailto:jcarranzamx@hotmail.com), Acapulco Gro.

<sup>6</sup> Ing. Armando Montes Torres, estudiante de la Maestría en sistemas computacionales, [ing\\_armando\\_mtorres@hotmail.com](mailto:ing_armando_mtorres@hotmail.com), Acapulco Gro.

### *Aprendizaje supervisado*

Este tipo de herramientas es usado en minería de datos, el aprendizaje que se usa es una técnica para deducir una función a partir de datos de entrenamiento, es decir, se entrena o se provee de los datos previos de entrada para que estos a la vez puedan ser comparados con los nuevos datos ingresados y predecir o determinar las salidas o resultados que se estén buscando.

### *Sistemas expertos.*

Un sistema experto, es una aplicación informática que sobre una base de conocimientos que posee información de uno o más expertos para solucionar un conjunto de problemas en un área específica nos brindan soporte para poder clasificar o catalogar. Este tipo de sistemas aplicados a la herramienta que se diseña nos brindara ayuda para la clasificación de la información obtenida mediante las encuestas aplicadas a los candidatos. [4]

### *Base de conocimiento.*

El sistema cuenta con un algoritmo supervisado de clasificación, esto para que sea más ágil a la hora de tomar decisiones, ya que este tipo de algoritmos posee los datos de cotejo para la determinación de los valores que a su vez determinan el perfil del individuo contra el perfil del puesto. Ya que al contar con el algoritmo supervisado C4.5 de clasificación, se logra el “cotejo” entre perfil de candidato y perfil de puesto. [4]

### *Lógica difusa.*

La minería de datos es un método o modelo matemático que pretende emular la habilidad que tienen las personas para la toma de decisiones acertadas a partir de datos vagos o imprecisos y que se encuentra expresado lingüísticamente, es decir, en lenguaje natural. Por lo tanto, permite tratar dicha información como conjuntos difusos que se combinan en reglas para poder definir acciones, por ejemplo, “si se encuentra nublado, entonces lloverá”. Esta técnica aplicada a los cuestionarios de aptitud, psicológicos y de adaptación social, nos brinda ciertas ponderaciones para poder hacer una pre-selección de candidatos. [3]

### *Función de pertenencia.*

La forma de aplicar dicha forma depende de los criterios a aplicarse en la resolución del problema, sin embargo, para poder ser aplicada en los algoritmos de minería de datos deben de considerarse los valores entre 0 y 1. Una de las fórmulas más utilizadas en este tipo de problemas es la triangular: definida por los límites (a, 0) y (c, 0) y por el punto central (b, 1), tal que:

$$\mu_F(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in (a, b] \\ \frac{c-x}{c-b} & \text{si } x \in [b, c) \\ 1 & \text{si } x \geq c \end{cases}$$

Donde, x= variable de entrada; a= punto en el eje x que señala el primer vértice del triángulo; b= punto en el eje x que señala el segundo vértice del triángulo; c= punto en el eje que señala el tercer vértice del triángulo. Esto es: por ejemplo, si tomáramos el atributo edad: se puede determinar si es un candidato a ser contratado o no (siempre y cuando este fuera uno de los atributos principales al puesto), valores difusos: “joven”, “adulto” o “viejo”. Si este fuera el caso tenemos entonces que evaluar y poner valores en el rango de edades de acuerdo a las necesidades y/o características del puesto. [2]

### Antecedentes del problema

El proceso de selección de candidatos que se lleva a cabo en la actualidad es de forma subjetiva, esto es, que de acuerdo a criterios propios del departamento de recursos humanos, determinan de acuerdo a experiencias si es o no un candidato a ser contratado, esto sin mencionar factores externos que no son relacionados al mismo departamento, además de las malas prácticas de selección de candidatos a ingresar al área administrativa ha perjudicado gravemente la estructura y el ambiente laboral por la falta de análisis de las actitudes del individuo y su forma de relacionarse con otros en equipo.

### Metodología de desarrollo

La metodología a usarse será la de prototipos, ya que es la que nos apoya a ir avanzando paso a paso cada una de los procesos he ir interactuando con las personas que harían uso del sistema. Además de que es un método flexible en su uso y no necesariamente necesita de todos los datos o atributos para ir verificando los avances en él. Como se ha mencionado en el presente artículo, el trabajo se centra en el análisis de las entradas y lo que se espera de salida llevando entre sí una comparativa entre perfil del individuo y del puesto, sin embargo, como nunca ha existido un proceso de reclutamiento no se cuenta con datos específicos de la toma de decisión. Véase figura 1. [7]

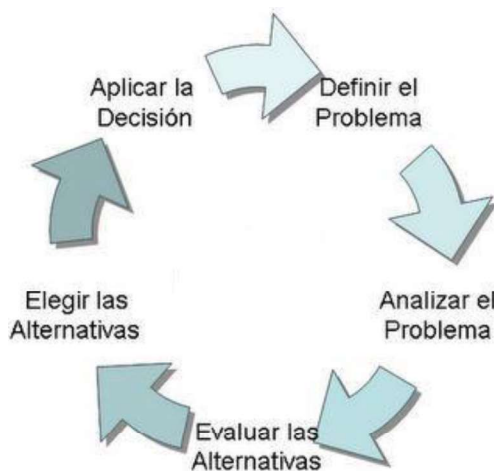


Figura 1. Esquema que ejemplifica la toma de decisión.

La selección de personal en una empresa y/o organización está constituida por distintos procesos, herramientas y decisiones que ofrecen información relevante de los candidatos a ingresar, todo esto a través de pruebas psicométricas, entrevistas y perfiles, lo que ayuda a medir las aptitudes, competencias, rasgos de personalidad, valores y aspiraciones. Todo este tipo de pruebas brindará datos que pueden o no ser objetivos para determinar si es realmente el candidato que se está buscando, ya que todas las pruebas están aún evaluadas por una persona. Sin embargo, si todo este proceso fue realizado bajo criterios ya establecidos, es decir, trasladados a procesos y funciones matemáticos. Con todo esto es posible determinar los candidatos en base a los datos arrojados por dichas herramientas. [3]

La metodología a seguir, es el de analizar la estructura de la base de datos que contendrán los datos de todas las personas que deseen ingresar al sistema educativo en el estado.

Dichas bases de datos tendrán:

- Datos personales.
- Resultados de encuestas.
- Datos académicos. (solo para puestos específicos)

Los datos personales, son todos aquellos que puedan identificar al individuo de todos los demás, por ejemplo: Nombre completo, edad, experiencia, lugar de nacimiento.

Se realizará una serie de encuestas de admisión, pruebas que ya están establecidas normalmente en toda empresa para reclutamiento por parte de recursos humanos, mismas que ya tienen ciertos puntos de ponderación y que pueden ser medibles.

Los datos académicos, por la naturaleza de los puestos administrativos que se están analizando podrían en un momento no ser relevantes para el proceso de selección, ya que solo para algunos puestos administrativos es necesaria. Sin embargo, lo que sí se debe de tomar en cuenta es la experiencia que pueda tener cada uno de los individuos. Existen ciertos perfiles dentro de la estructura que, si requieren niveles de estudios a nivel licenciatura y otros a nivel técnico, para estos casos también es necesario el conocerlos, pero como se menciona, es solo para casos muy específicos y que en realidad esos puestos forman parte de niveles altos en jerarquía. Vea la tabla 3 como ejemplo de jerarquías en la estructura.

Y que todo esto junto con los perfiles académicos (dependiendo del puesto) se almacenará en la base de datos en una tabla de resultados, lo que dará ciertos valores para cada individuo y de donde se obtendrán los candidatos que serán considerados para el puesto vacante.

Una vez teniendo dicha tabla, se somete al algoritmo de minería de datos de clasificación, ya que, por la cantidad de registros, el ir verificando y validando uno a uno los registros, será un proceso demasiado tardado. Es por eso que es necesaria este tipo de técnicas o herramientas para la optimización de tiempos de respuesta en los resultados.

*“La decisión es la acción que define al candidato que habrá de ocupar un puesto de trabajo, por lo tanto, debe ser considerado un proceso que se encuentra insertado en el flujo de procedimiento para la selección de personal”.* [1]

De acuerdo a lo antes expuesto, se puede determinar que es posible asignar valores a cada uno de los exámenes o pruebas de recursos humanos, al igual que a la edad y experiencia o según sea el perfil del puesto a ocupar, es por ello, es necesario considerar los planteamientos de valores que se asignaran a cada uno de ellos y que grado de pertenencia tendrán de manera que deben de quedar bien específicos para que estos sean objetivos al seleccionar el personal, por ejemplo, siguiendo con el ejemplo pasado si es de considerarse la edad y se le asignan los valores de acuerdo a que edad es la mínima y máxima para ser contratada, este tipo de limitaciones están aplicadas a todas las empresas. En la tabla 1 y 2 se muestra un ejemplo de como se le asignan valores en este caso a las edades de los individuos para ser contratados y su experiencia respectivamente; y de cómo se aplica minería de datos a estos:

Edad	
Valores asignados	Vértice
20 - 25	Joven
26 - 30	Adulto
31 - 35	Viejo

**Tabla 1:** valores asignados a la edad permitida.

Experiencia	
Años	Vértice
00 – 02	Poca
03 – 04	Regular
05 – 06	Mucha

**Tabla 2:** valores de Experiencia laboral.

Se debe de recordar que la única condición que debe de cumplir una función de pertenencia (la cual se usa en lógica difusa) es que debe de tomar valores entre 0 y 1 con continuidad. La tabla solo representa un ejemplo de este tipo de modelos, no necesariamente deben ser estos los valores a aplicarse. [2]

Se puede entonces decir, que lo restante sería la “defusificación” o interpretar los datos. Esto es un método por el cual se obtienen los valores de salida del conjunto difuso, que es la agregación de todas las reglas y valores asignados a las distintas pruebas. Entonces podemos utilizar uno de los siguientes casos:

- Método del máximo: se elige como valor de salida aquel para el cual la función del algoritmo aplicado brinde un valor máximo. Este método no es un método óptimo, pues el valor puede ser alcanzado por varias salidas.
- Método del centroide: utiliza como valores de salida el centro de gravedad de la función característica de salida. Con este método se obtiene un valor único.
- Método de la altura: primero se calculan los centros de gravedad para cada regla del conjunto difuso de salida y después la media ponderada.

Y de todos los valores arrojados se obtendrá una lista de los posibles candidatos a ocupar la plaza, ya que estos serán a los que se les aplicará el examen de ingreso al sistema, el cual es aplicado por el departamento de recursos humanos del plantel y de la comisión dictaminadora, que es la que normalmente designa las “plazas” una vez que se han definido los candidatos. Los valores se pueden ver en la figura 2, donde se muestran los vértices de cada una de los atributos y la interacción entre cada uno de ellos.

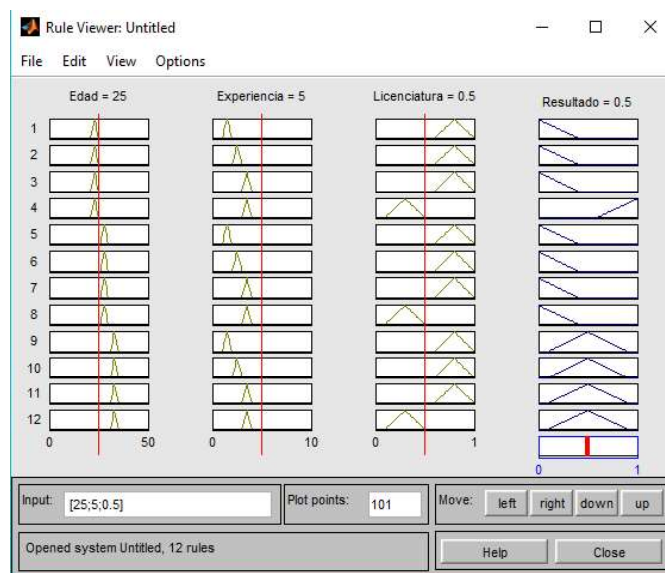


Figura 2: Valores arrojados por MatLab.

Cabe mencionar que las plazas que se analizan son las administrativas, en específico las “XA”, como se muestra en la tabla 3.

CLAVE	CATEGORIA	NIVEL
XS05008	LINOTIPISTA	3
XS06002	INTENDENTE	3
XS06006	AUXILIAR DE INTENDENCIA	3
XS07008	AUXILIAR DE MANTENIMIENTO	3
XS08011	TECNICO EN MANTENIMIENTO	4
XS09007	ELECTRICISTA	4
XS12030	JARDINERO	4
XS13008	CHOFER	4
XS14003	VIGILANTE	4
XS05006	IMPRESOR	5

XS05015	OPERADOR DE MAQUINAS DE REPRODUCCION	5
XS12004	PINTOR	5
XS12008	CARPINTERO	5
XS12012	PLOMERO	5

**Tabla 3:** Categorías administrativas de personal de apoyo y de servicios.

### Conclusiones

El proceso de selección para la contratación de personal administrativo será realizada más rápida y eficientemente a través de una base de datos que almacene la información seleccionada de los interesados a ingresar al Sistema Nacional de Bachillerato ya que como se mencionó, actualmente no hay algún sistema o proceso que ayude a analizar la información de los participantes de una forma estándar, evitando la buena o mala interpretación del o de los reclutadores, ya que este sistema puede analizar los datos, compararlos con los parámetros que se desean como óptimos y resolver aún con pequeñas discrepancias entre los datos de los participantes. Esto significa un 90 % más rápido la selección.

### Recomendaciones

Este proceso se realiza para el plantel CETis No. 116, sin embargo, forma parte de un plan general que puede aplicarse en toda institución o empresa, pública o privada ya que, por su naturaleza de aplicar las principales técnicas de recursos humanos, es que, la hace factible el poder hacerlo.

### Referencias

- [1] Andrés, Paolo, Castaño, Vélez (2011), Optimización de selección de proveedores integrando un árbol de decisión a un proceso de negocio, Artículo.
- [2] Carlos, A., Díaz, Contreras (2012), Minería de datos vs. Modelo de regresión múltiple para la selección de personal, Artículo.
- [3] Francisco, Javier, Ruvalcaba, Coyaso (2015), Minería de datos para la toma de decisiones y la selección de personal, Artículo.
- [4] Héctor, A., Tabares-Ospina (2012), Modelo de sistema experto para la selección de personal docente universitario, Artículo.
- [5] José, Hernández, Orallo (2006), Curso de doctorado extracción automática de conocimiento en bases de datos e ingeniería del software, Artículo.
- [6] López, González, E, (1996), La selección del personal con un algoritmo genético borroso, Artículo.
- [7] Sommerville, Ian (2011), Ingeniería de software, Libro.