

DOI: 10.20983/culcyt.2020.2.2.1

Integración de grupos de trabajo con apoyo de C4.5

Integration of working groups with the support of C4.5

M. Soltero-Romero¹, J. Rodas-Osollo¹

¹ Universidad Autónoma de Ciudad Juárez

RESUMEN

El perfil de una persona respecto a su desempeño al realizar diversas actividades en áreas diferentes, en lo individual o trabajando en equipo, es de suma importancia para un negocio u organización, ya que conociendo los perfiles de las personas se pueden integrar grupos de trabajo más eficientes. De las pruebas de perfilado de personalidad, la DISC es una de las más utilizadas por los departamentos de recursos humanos. Sin embargo, todos los test de perfilado son altamente subjetivos, por lo que con frecuencia un perfil puede ser indeterminado. En este artículo se comparte el resultado de utilizar el C4.5, un algoritmo de clasificación supervisada, como apoyo a los departamentos de recursos humanos y se determina, de acuerdo al test de personalidad DISC, la más adecuada integración.

PALABRAS CLAVE: DISC; C4.5; modelo de clasificación; reglas de decisión.

ABSTRACT

The profile of a person regarding their performance when carrying out various activities in different areas, individually or working as a team, has a great importance for a business or organization, since knowing the profiles of the people more efficient work groups can be integrated. Among the personality profiling tests, the DISC test is one of the most used by human resources departments. However, all profiling tests are highly subjective, so a profile can often be indeterminate. This article shares the result of using the C4.5, a supervised classification algorithm, to support the human resources department and determine, according to the DISC personality test, the most appropriate integration.

KEYWORDS: DISC; C4.5; classification model; decision rules.

Correspondencia:

DESTINATARIO: Marlene Soltero Romero
INSTITUCIÓN: Universidad Autónoma de Ciudad Juárez,
Instituto de Ingeniería y Tecnología
DIRECCIÓN: Av. del Charro no. 450 norte, col. Partido
Romero, Ciudad Juárez, Chihuahua, México, C. P. 32310
CORREO ELECTRÓNICO: al194545@alumnos.uacj.mx

Fecha de recepción: 10 de febrero de 2020. **Fecha de aceptación:** 3 de noviembre de 2020. **Fecha de publicación:** 15 de noviembre de 2020.



I. INTRODUCCIÓN

La personalidad es un factor clave para determinar el cómo una persona se va a desempeñar en una actividad en su entorno. Diferentes tipos de personalidades indican que personas con diferentes características se adaptan mejor a otras que complementan esas características.

El cuestionario DISC abarca una serie de preguntas que determinan a qué ritmo la persona suele llevar las relaciones, qué prioridades tiene respecto a estas, cómo suele actuar en situaciones de presión o riesgo, cómo influye en las personas, a qué le teme y el estilo de personalidad al que mejor se adecua, lográndose así la más acorde al individuo evaluado. Esta técnica para obtener las personalidades es muy acertada y permite determinar varios factores que son indiscutibles a la hora de obtener resultados.

Por esta razón, en este trabajo se creó un árbol de clasificación para manejar estas personalidades en cuanto a reglas de decisión, y la herramienta de uso libre Weka dio la posibilidad de elegir varios algoritmos para hacer el entrenamiento de los datos. En este caso se optó por utilizar el algoritmo C4.5 y con la ayuda de un archivo con todos los posibles puntajes de la prueba DISC y unas adecuaciones para que esta herramienta pudiera interpretarlo, se creó el árbol de clasificación con un 99.3336% de efectividad, con el cual se desarrollaron varias pruebas con distintos cuestionarios ya contestados y con los puntajes de los posibles resultados.

El clasificar correctamente una personalidad da beneficios en diversos ámbitos, por ejemplo, en la mayoría de las empresas se crean grupos de trabajos que resuelven necesidades conforme a sus habilidades y al giro del negocio. Estos equipos son formados generalmente por personas que contribuyen con estrategias a un proyecto que, al complementarlas, ayudan al desarrollo y éxito de este ^[1].

Situaciones problemáticas derivadas de no tomar en cuenta las características principales de la personalidad de cada individuo pueden impactar tanto en la calidad del resultado del proyecto, el tiempo en que se desarrolla este o incluso propicia conflictos entre los participantes ^[1]. De hecho, un problema recurrente en las empresas es la rotación o despido de personal que no cumple con los requisitos de los proyectos o de la natu-

raleza de las actividades, generando pérdidas de tiempo y recursos ^[2]. Sin embargo, el conocer la personalidad de los miembros de una empresa brinda la oportunidad de saber cómo incentivar a los empleados a desempeñar sus labores más asertivamente, de manera que no afecte su entorno laboral.

En Hunt ^[3] se realizó un estudio que destaca la importancia de las características de la personalidad de quienes participan en la toma de decisiones para efectuar cambios de carácter legal. Como resultado, se obtuvo una correcta clasificación de la personalidad que apoya en la elección de adecuados perfiles para tal encomienda, de lo contrario, las personas encargadas de estas labores con seguridad pueden tomar decisiones erróneas con impactos adversos a la comunidad.

Existen varias herramientas en forma de cuestionarios psicométricos, como DISC ^[4] o Emergenetics ^[5], que, a partir de una serie de preguntas o enunciados, clasifican la personalidad del usuario detallando varios aspectos importantes que una persona o los departamentos de recursos humanos desearían saber de su personalidad y sus fortalezas o habilidades más destacadas.

El utilizar un cuestionario para obtener el perfilado de una persona puede arrojar resultados no deseables, que incluyen desde obtener ligeras discrepancias en la personalidad del individuo hasta la indeterminación de una personalidad. En situaciones así es común recurrir a más de una prueba, siendo necesario invertir un mayor tiempo de valoración e interpretación de resultados. Por lo anterior, hoy en día es frecuente que este tipo de procesos se apoyen en herramientas del Aprendizaje Automático para facilitar la obtención de resultados.

En el trabajo de Tarrago *et al.* ^[6] se habla de las distintas técnicas que se pueden emplear para tratar información similar y cómo algunas son más acertadas que otras, sin embargo, el uso de algoritmos de clasificación como el C4.5 es eficiente al momento de obtener reglas de clasificación si la información es distribuida en forma de clases con cada uno de sus atributos. Por ejemplo, en Apolaya *et al.* ^[7] se utilizan para determinar el nivel académico de estudiantes que cursan distintas materias, es decir, qué tan alto es su nivel de aprobación y si pueden cursar o no esa materia el semestre siguiente. La efectividad de su trabajo fue del 90%, por lo que esta técnica resultó eficaz para los datos manejados por estos autores en el dominio académico.

Como se ha mencionado anteriormente, los árboles de decisión no son el único camino para lograr la clasificación de la personalidad. En un ejercicio de predicción de personalidad [8], se utilizó como fuente de información audios de entrevistas y percepciones de las mismas. Se aplicaron funciones de predicción lineales con cinco personalidades dominantes para ir dando puntaje de semejanza hacia cada uno de los audios. Una vez obtenidos estos puntajes, se emplearon las técnicas de máquinas de soporte de vectores, algoritmo del vecino más cercano y perceptrón multicapa para clasificar cada uno de los audios con su personalidad más baja y alta.

En este ejercicio, la técnica del vecino más cercano obtuvo resultados más favorables, con una predicción del 90%. En el mismo sentido, se realizó un trabajo similar pero con la técnica de Rough Sets, obteniendo muy resultados semejantes [9].

Por otra parte, en el trabajo de Bhardwaj *et al.* [10] se adoptó un modelo de red neuronal para medir el nivel de compatibilidad entre nuevos trabajadores y el gerente, con el fin de clasificar cuáles son más compatibles y cuáles no en relación a sus personalidades. Se utilizaron los perfiles DISC de estos nuevos trabajadores y el gerente y se entrenó un clasificador para predecir si el empleado era compatible con las necesidades. Como resultado, se observó que es factible el empleo de esta técnica para predecir el comportamiento de una persona, pues de obtuvo un nivel alto de precisión en la personalidad predicha, disminuyendo la indeterminación de cuando se utiliza exclusivamente la prueba DISC.

Existe una gran cantidad de referencias de casos de estudio relacionados con el tema abordado en este artículo, aparte de los señalados en párrafos anteriores, en las cuales las técnicas de árboles de decisión son muy utilizadas cuando se parte de conocimiento previo a lo que se desea clasificar como etiquetas de clasificación que indican la correspondencia de los datos.

Al clasificar las distintas personalidades que arroja una prueba como DISC, es fácil de implementar este método y, al igual que en [11], el presente artículo reporta el trabajo con el algoritmo de clasificación C4.5 que crea reglas de clasificación fáciles de verificar y determina objetos que no se pueden clasificar e incluso el nivel de error que tiene el árbol de decisión.

II. METODOLOGÍA

Se utilizó un procedimiento para llevar a cabo la clasificación inspirado en los pasos realizados para el entrenamiento de algoritmos de aprendizaje automático (Figura 1), estableciendo las etapas necesarias para obtener reglas de clasificación y, tal como se muestra en la investigación de Apolaya, haciendo uso de Knowledge Discovery in DataBases [7]. Además, se utilizó la metodología Action Design Research [12], la cual consta de las cuatro etapas descritas a continuación.

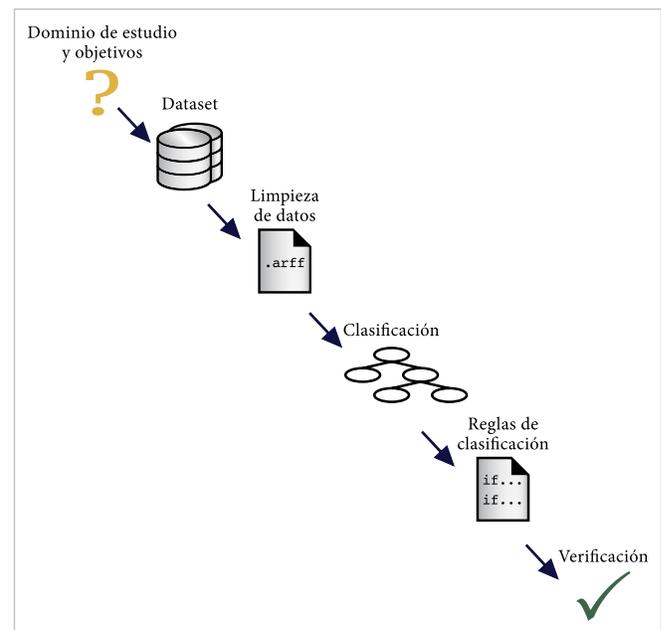


Figura 1. En la imagen se aprecia el procedimiento para llevar a cabo la clasificación inspirado en los pasos para entrenamiento de algoritmos de aprendizaje automático. Fuente: los autores.

La metodología tiene las etapas siguientes: *formulación de problemas*, donde se establecieron los objetivos que se esperaba cumplir; *construcción, intervención y evaluación*, en la que se clasificaron los datos posibles a obtener en la prueba DISC; *reflexión y aprendizaje*, en la que se utilizó el algoritmo C4.5, y *formalización del aprendizaje*, donde, una vez obtenidas estas reglas que el algoritmo proporcionó, se realizó una serie de ejercicios para verificar que estas reglas fueran correctas.

Etapa 1. Formulación de problemas

El dominio de este estudio es la prueba DISC, la cual se basa en cuatro personalidades predominantes y de ahí deriva sus resultados dependiendo de un puntaje, el

cual se obtiene de las diferentes respuestas que fue dando la persona entrevistada. Los objetivos de este trabajo son los siguientes:

- Crear una serie de reglas de clasificación para poder predecir la personalidad dominante de una persona por medio de los puntajes posibles a obtener, realizando el cuestionario de clasificación DISC, que cuenten con el rasgo de personalidad.
- Comprobar que el algoritmo C4.5 funciona para crear árboles de clasificación (clasificador) a partir de un conjunto de datos funcionales.
- Comprobar que las reglas de clasificación obtenidas son óptimas para predecir la personalidad.
- Concluir qué tan confiables son estas reglas.

Etapa 2. Construcción, intervención y evaluación

El DISC cuenta con un conjunto de puntajes que se asocian con un tipo de personalidad y de esta forma se hace la predicción. De estos datos se evaluaron y obtuvieron los puntajes posibles de la prueba DISC con sus respectivas personalidades, dejando de lado los parámetros con los que se obtienen tales puntajes y clasificándolos en las diferentes personalidades resultantes, como se muestra en la Tabla 1.

TABLA 1
ALGUNOS DE LOS PUNTAJES CON SUS RESPECTIVAS PERSONALIDADES

PUNTAJE	PERSONALIDAD
7664	Alentador
7657	Superactivo
7647	Creativo
7622	Alentador
7611	Alentador
7573	Realizador
7572	Realizador
7557	Superactivo
7554	Orientado a resultados
7545	Creativo
7542	Orientado a resultados
7476	Investigador
7475	Investigador
7474	Realizador

Fuente: los autores.

Una vez obtenidos todos los puntajes, se construyó el modelo de los datos que haría la clasificación conforme

a las cuatro personalidades sobresalientes (Tabla 2), separando por columna cada una de las personalidades y la definición del puntaje.

TABLA 2
CLASIFICACIÓN DIVIDIDA EN CADA RASGO DE LA PERSONALIDAD DE LOS PUNTAJES ANTERIORES

D	I	S	C	PERSONALIDAD
7	7	3	3	Alentador
7	7	2	4	Alentador
7	7	2	6	Evaluador
7	6	7	7	Superactivo
7	6	6	5	Superactivo
7	5	5	4	Orientado a resultados
7	6	7	3	Realizador
7	5	4	7	Creativo
7	5	4	4	Orientado a resultados
7	4	7	6	Investigador
7	4	5	1	Realizador
7	4	3	2	Orientado a resultados
7	3	2	2	Resolutivo

Fuente: los autores.

A partir de los puntajes ya clasificados en sus respectivas personalidades, se realizó una intervención de los datos a través de la limpieza de los puntajes que no arrojan una personalidad o de los cuales no se determinan rasgos característicos, es decir, cuando la personalidad está sin definición. Para lograr esto se eliminaron los datos (puntajes) que, una vez contestada la prueba, no mostraron una predicción de la personalidad. Cuando se obtuvo una total limpieza de los datos, se procedió a migrarlos a una base de datos con archivos en formato ARFF, donde se establecieron las clases con las diferentes personalidades posibles y además se agregaron como atributos las cuatro personalidades sobresalientes: Dominancia, Influencia, Estabilidad y Conformidad (Figura 2), con el fin de adaptarse a la herramienta a utilizar para la clasificación.

```

%1. Tilt: DISC
%
%2. Creator:
%
@RELATION disc

@ATTRIBUTE DOMINANCIA NUMERIC
@ATTRIBUTE INFLUENCIA NUMERIC
@ATTRIBUTE ESTABILIDAD NUMERIC
@ATTRIBUTE CONFORMIDAD NUMERIC
    
```

Figura 2. Atributos sobresalientes y las clases que se derivan y construyen el encabezado. Fuente: los autores.

Etapa 3. Reflexión y aprendizaje

El aprendizaje del clasificador se llevó a cabo mediante el algoritmo C4.5, llamado J48 en la herramienta Weka, la cual es un software de uso libre desarrollado en Java que permite utilizar distintos algoritmos de clasificación con la finalidad de realizar varios experimentos, como la generación de árboles de decisión, que verifiquen el funcionamiento de estos como técnicas para clasificar clases o conjuntos basados en características. Consta de una interfaz gráfica que permite visualizar los datos en diferentes modelos, desde algoritmos como gráficos de barras; hace validaciones de cada uno de los algoritmos que utiliza sin necesidad de hacer cálculos manualmente y, al obtener los resultados, también se consiguen datos como el error cuadrático medio, la precisión o el rendi-

miento. En general, ofrece soluciones para poder clasificar o predecir dado algún conjunto de atributos [13].

La ejecución de Weka se llevó a cabo en una máquina con sistema operativo MacOS, donde se eligió la base de datos antes creada y el algoritmo de clasificación C4.5 para obtener después las reglas de clasificación.

Con el árbol de clasificación que se obtuvo se puede deducir una serie de reglas que determinan el comportamiento de los puntajes y a qué clase de personalidad apuntarán (Figura 3), donde la personalidad predominante es Conformidad y de ahí se divide en Dominante y Estable para encontrar coincidencias que puedan describir la personalidad, según el puntaje a partir de las reglas de clasificación que ahí se muestran.

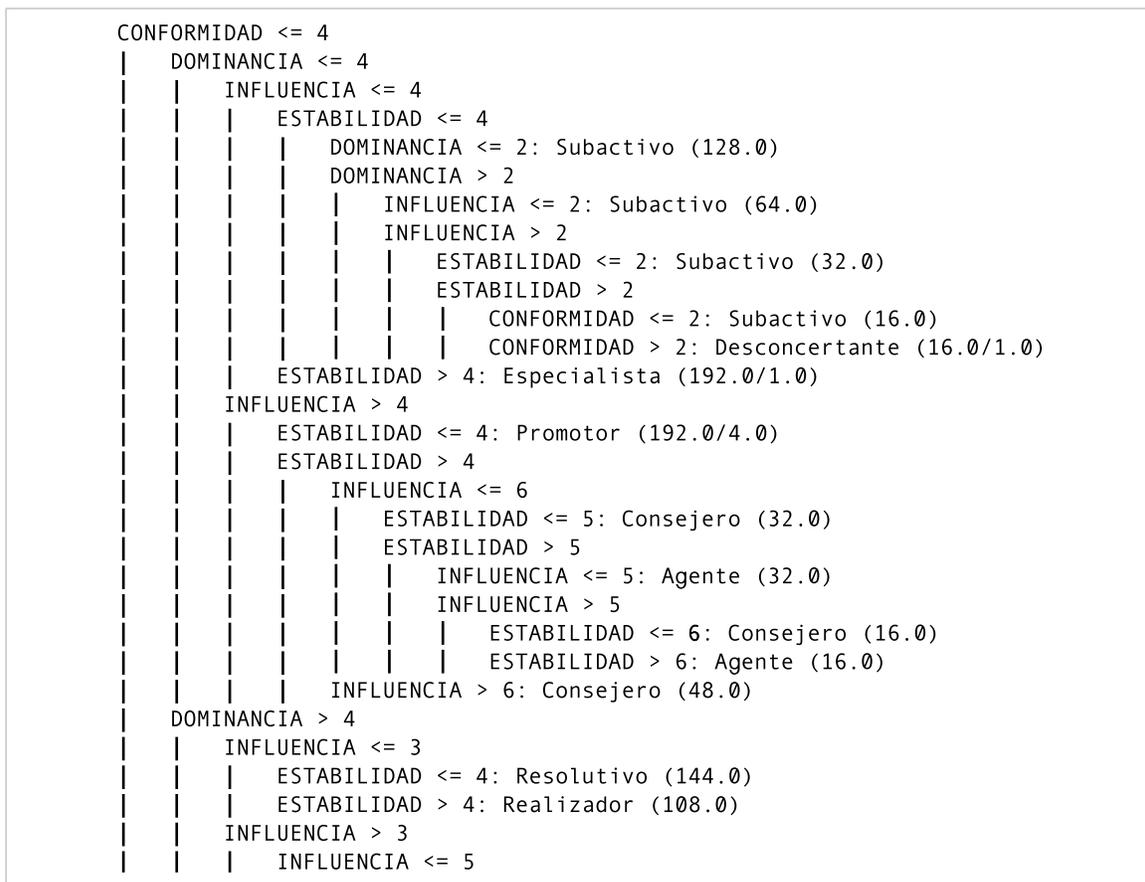


Figura 3. Algunas reglas de decisión obtenidas a partir de la clasificación. Fuente: los autores.

Etapa 4. Formalización del aprendizaje

Una vez obtenido el árbol de clasificación, se pueden realizar comprobaciones o incluso aplicar mejoras en los datos que fueron utilizados para crearlo. En ocasio-

nes resulta más factible omitir algún conjunto de datos o agregar más para obtener un modelo de mayor efectividad. En este caso, con los datos utilizados se obtuvieron buenos resultados.

Las reglas de clasificación obtenidas fueron fácilmente validadas mediante los puntajes de los perfiles de la prueba DISC. Se comprobaron estas reglas observando el puntaje que arroja un perfil versus la regla de clasificación que apunta a ese resultado.

En la Figura 4 se muestra cómo efectivamente se puede comprobar, a través de las reglas de clasificación obtenidas, que los puntajes corresponden tanto para el árbol de clasificación como para un puntaje ya predicho, mostrado en la Figura 5.



Figura 4. Puntajes del DISC en cuanto a personalidades dominantes del perfil "Creativo". Fuente: los autores.

```

CONFORMIDAD > 4
| ESTABILIDAD <= 4
| | INFLUENCIA <= 4
| | | DOMINANCIA <= 4: Objetivo (192.0/1.0)
| | | DOMINANCIA > 4: Creativo (144.0)
    
```

Figura 5. Regla de decisión que indica que el perfil anterior es "Creativo". Fuente: los autores.

III. RESULTADOS

El porcentaje de error obtenido tras la utilización del algoritmo C4.5 en su versión J48 de la herramienta Weka fue del 0.6664% de los datos, cuya matriz de confusión se observa en la Figura 6.

Se puede verificar que de todas las personalidades clasificadas, solamente en cuatro de ellas se generó una confusión. Esto se deriva de un pre-procesamiento de los datos, limpieza y simplificación, obteniendo un conjunto de personalidades funcional al algoritmo y dejando de lado la información que no era de utilidad para la clasificación.

```

=== Confusion Matrix ===
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  o  r  <-- classified as
0 123 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  a = Superactivo
0 0 186 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  b = Alentador
0 0 0 180 0 0 0 0 0 0 0 0 0 0 0 0 0 0  c = Evaluador
0 4 0 0 160 0 0 0 0 0 0 0 0 0 0 0 0 0  d = Creativo
0 0 0 0 0 87 0 0 0 0 0 0 0 0 0 0 0 0  e = Realizador
0 0 0 0 0 0 108 0 0 0 0 0 0 0 0 0 0 0  f = Orientado a resultados
0 0 0 0 0 0 0 144 0 0 0 0 0 0 0 0 0 0  g = Investigador
0 0 0 0 0 0 0 0 111 0 0 0 0 0 3 0 0 0  h = Resolutivo
0 0 0 0 0 0 0 0 0 40 0 0 0 0 0 0 0 0  i = Consejero
0 0 0 0 0 0 0 0 0 0 60 0 0 0 0 0 0 0  j = Persuasivo
0 1 0 0 0 1 0 0 1 0 0 15 1 1 1 0 1 1  k = Agente
0 0 0 0 0 0 0 0 0 0 0 0 137 0 0 0 0 0  l = Desconcertante
0 0 0 0 0 0 0 0 0 0 0 0 0 188 0 0 0 0  m = Profesional
0 0 0 0 0 0 0 0 0 0 0 0 0 0 191 0 0 0  n = Promotor
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 240 0 0  o = Objetivo
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 191 0  p = Subactivo
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 143  q = Especialista
r = Perfeccionista
    
```

Figura 6. Matriz de confusión que muestra las personalidades en las cuales hubo discrepancias. Fuente: los autores.

IV. CONCLUSIONES

Derivado de este ejercicio se puede afirmar que para este tipo de dominio y los datos que le representan, la técnica de aprendizaje supervisado de árbol de decisión es adecuada, especialmente cuando se tienen datos precisos en cantidad considerable, ya que, en este

caso, el nivel de error varía muy poco debido a que todas las personalidades pertenecen a su correspondiente puntaje y los datos a clasificar no superan los 1300 registros.

Al trabajar con datos del dominio de la prueba de personalidad DISC que consideran personalidades indeter-

minadas —sin descripción de estas, sin clasificación—, estos quedaron excluidos para este ejercicio con el fin de hacer más eficiente el desempeño del C4.5. Aunque no es el tema de este artículo, se puede señalar que existen otras técnicas del *machine learning* que podrían ser consideradas para una clasificación que sí incluyera esta información faltante.

Finalmente, clasificar correctamente las personalidades es un gran apoyo a los procesos que llevan a cabo los departamentos de recursos humanos en las empresas u organizaciones que buscan mejorar el desempeño de sus empleados, especialmente al conformar grupos de trabajo o incluso en lo individual, para lograr la mayor eficiencia en la gestión de proyectos.

REFERENCIAS

- [1] I. Lykourantzou, A. Antoniou, Y. Naudet, y S. P. Dow, “Personality matters: Balancing for personality types leads to better outcomes for crowd teams”, *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW*, vol. 27, pp. 260-273, 2016, DOI: [10.1145/2818048.2819979](https://doi.org/10.1145/2818048.2819979)
- [2] A. M. Espín, F. Reyes-Pereira y L. F. Ciria, “Organizations should know their people: A behavioral economics approach”, *J. Behav. Econ. Policy*, vol. 1, no. esp., pp. 41-48, 2017. Disponible en: <https://sabeconomics.org/journal/RePEc/beh/JBEPv1/articles/JBEP-1-S-8.pdf>
- [3] S. J. Hunt, “Communication and Change: A Qualitative Study of Law Enforcement Team DISC Personality Traits”, disertación de doctorado, Grand Canyon University, Phoenix, Arizona, 2017. [En línea]. Disponible en: https://www.drsheilahunt.com/uploads/1/2/4/6/124627095/dr._hunt_descriptive_dissertation_pdf_version_2.18.19.pdf
- [4] DISC. *La introducción completa*, Axiom Internet Group Limited, Manchester, Reino Unido, 2020. Disponible en: <https://www.discusonline.com/es-mx/es-pe/features/free-guide-to-disc-personality-test.php>
- [5] R. Wendell Williams, “The Emergenetics Profile Technical Report”, Emergenetics LLC, 2018.
- [6] D. S. Tarrago, F. Herrera y R. Bello, “Dominio de competencia del C4.5 a través de métricas de la Teoría de los Conjuntos Aproximados”, en *XI National Congress of Mathematics and Computer Sciences COMPUMAT 2009*, en La Habana, Cuba, enero 2009.
- [7] C. Apolaya, A. Espinosa y A. Barrientos. (2015, junio). Predicción del Rendimiento Académico en carreras de Computación utilizando Árboles de Decisión. ResearchGate [En línea]. Disponible en: https://www.researchgate.net/publication/333582116_Prediccion_del_Rendimiento_Academico_en_carreras_de_Computacion_utilizando_Arboles_de_Decision
- [8] S. Jothilakshmi y R. Brindha, “Speaker trait prediction for automatic personality perception using frequency domain linear prediction features”, *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 2016, pp. 2129-2132. DOI: [10.1109/WiSPNET.2016.7566518](https://doi.org/10.1109/WiSPNET.2016.7566518)
- [9] U. Gupta y N. Chatterjee, “Personality Traits Identification Using Rough Sets Based Machine Learning”, *2013 International Symposium on Computational and Business Intelligence*, Nueva Delhi, 2013, pp. 182-185. DOI: [10.1109/ISCBI.2013.44](https://doi.org/10.1109/ISCBI.2013.44)
- [10] C. A. Bhardwaj, M. Mishra y S. Hemalatha, “An Automated Compatibility Prediction Engine using DISC Theory Based Classification and Neural Networks”, *Int. J. Eng. Technol. Sci. Res.*, vol. 4, no. 8, 2017.
- [11] H. Jantan, “Human Talent Prediction in HRM using C4.5.pdf”, *Int. J. Comput. Sci. Eng.*, vol. 02, no. 08, pp. 2526-2534, 2010.
- [12] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi y R. Lindgren, “Action Design Research”, *MIS Q.*, vol. 35, no. 1, pp. 37-56, marzo 2011. DOI: [10.2307/23043488](https://doi.org/10.2307/23043488)
- [13] E. Frank, M. Hall, L. Trigg, G. Holmes e I. H. Witten, “Data mining in bioinformatics using Weka”, *Bioinformatics*, vol. 20, no. 15, pp. 2479-2481, 12 octubre 2004. DOI: [10.1093/bioinformatics/bth261](https://doi.org/10.1093/bioinformatics/bth261)