

Tipo de artículo: Artículo original

## Método para el análisis lingüístico de estadísticas médica

### Method for linguistic analysis of medical statistics

Barbara Bron Fonseca <sup>1\*</sup>, <https://orcid.org/0000-0001-9463-8408>

Omar Mar Cornelio <sup>2</sup>, <https://orcid.org/0000-0002-0689-6341>

<sup>1</sup> Soluciones MarBro S.R.L. Cuba.

<sup>2</sup> Centro de Estudio de Matemática Computacional, Universidad de las Ciencias Informáticas. Cuba.

\*Autor para la correspondencia: [barbarabronfonseca@gmail.com](mailto:barbarabronfonseca@gmail.com)

---

## RESUMEN

El proceso de evaluación del desempeño de las personas que trabajan en proyectos de innovación tecnológica para la salud genera una gran cantidad de información, que condiciona la toma de decisiones oportuna en tiempo de ejecución y sirve como base de estudio para perfeccionar o corregir futuras estrategias de gestión. Cuando el volumen de información cuantitativa generado es muy alto, es difícil identificar la información verdaderamente importante para proyectos de este tipo. El objetivo de la investigación es diseñar un método para el análisis lingüístico de la evaluación de los recursos humanos de proyectos médicos. La investigación desarrolla mediante un enfoque de estudio mixto, de tipo retrospectivo y diseño experimental, de corte transversal. Se realizó un análisis documental para identificar las técnicas de Inteligencia Artificial abordadas por la ciencia para resolver problemas de análisis e interpretación de datos. Como resultado se obtuvo el diseño de un método basado en el resumen lingüístico de datos para la



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional**  
(CC BY 4.0)

construcción de resúmenes lingüísticos a partir de la generación de reglas de asociación. El método fue corrido sobre una base de datos que almacena indicadores evaluativos diarios, sobre el desempeño de los recursos humanos de proyectos médicos. La sumarización lingüística de datos es un método de softcomputing que apoya la toma de decisiones extrayendo el conocimiento de grandes volúmenes de datos, no interpretables para los humanos sin ayuda computacional. El método propuesto, permitió identificar resúmenes lingüísticos que pueden servir de base para tomar decisiones sobre el desempeño general de los recursos humanos de un proyecto.

**Palabras clave:** Resumen lingüístico de datos; recursos humanos; reglas de asociación; proyectos médicos

### **ABSTRACT**

The performance evaluation process of people working in technological innovation projects for health generates a large amount of information, which conditions timely decision-making at execution time and serves as a basis for study to improve or correct future management strategies. When the volume of quantitative information generated is very high, it is difficult to identify the truly important information for projects of this type. The objective of the research is to design a method for the linguistic analysis of the evaluation of human resources in medical projects. The research is developed through a mixed study approach, retrospective type and experimental design, cross-sectional. A documentary analysis was carried out to identify the Artificial Intelligence techniques addressed by science to solve data analysis and interpretation problems. As a result, a method based on the linguistic summary of data was designed for the construction of linguistic summaries from the generation of association rules. The method was run on a database that stores daily evaluative indicators on the performance of human resources in medical projects. Linguistic data summarization is a soft computing method that supports decision making by extracting knowledge from large volumes of data that are not interpretable by humans without computational assistance. The proposed method made it possible to identify linguistic summaries that can serve as a basis for making decisions about the overall performance of a project's human resources.

**Keywords:** Linguistic data summarization, human resources, association rules, medical projects

**Recibido:** 10/09/2024

**Aceptado:** 12/12/2024



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional**  
(CC BY 4.0)

**En línea: 01/01/2025**

---

## Introducción

Con el desarrollo cognoscitivo alcanzado por la sociedad, se ha generalizado una tendencia hacia la organización por proyectos (Torres-López et al., 2014). Los proyectos en su concepción requieren implementar una correcta planeación y control para su seguimiento y evaluación, surgiendo así la gestión de proyectos como una actividad profesional.

El ritmo acelerado de los procesos actuales, el creciente número de proyectos que se conceptualizan y se desarrollan, apoyados en los vertiginosos avances de las tecnologías genera un enorme cúmulo de información valiosa que necesita ser procesada y analizada pero que resulta imposible ser examinada en su totalidad en condiciones humanas. La capacidad para manejar grandes volúmenes de datos se perfila cada vez más necesaria en una sociedad que, sin duda alguna, está basada en el conocimiento.

Los Recursos Humanos (RR.HH) de los proyectos constituyen un área clave para garantizar el cumplimiento de los objetivos estratégicos. Realizar un control y evaluación sistemática del desempeño laboral de los RR.HH que constituyen la cédula fundamental de los proyectos de innovación tecnológica para la salud, ayuda a identificar los niveles de desempeños generales del proyecto, el cumplimiento de los planes pactados, el aprovechamiento de la jornada laboral, la calidad de los entregables entre otros criterios de impacto para los proyectos.

La evaluación del desempeño de los RR.HH surge como un instrumento evaluativo para cuantificar el rendimiento laboral de cada especialista. Esta evaluación se realiza en periodos específicos para cada proyecto y la secuencia varía en función de la estructura organizativa. En el sector de la salud, estos criterios tienen un impacto muy específico, ya que las fallas originadas durante el proceso pueden provocar la pérdida de vidas humanas. Las evaluaciones realizadas a cada trabajador son almacenadas en bases de datos para su posterior consulta y análisis, pues permite la identificación de patrones y comportamientos en determinadas temporadas que ayudan a prevenir retrasos productivos, fallas en la configuración de tecnologías médicas, y a la toma de decisiones de un proyecto.



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional**  
(CC BY 4.0)

El proceso de evaluación del desempeño, sea este exitoso o no, genera una gran cantidad de información que condiciona la toma de decisión oportuna en tiempo de ejecución y que sirve como base de estudio para perfeccionar o corregir estrategias de administración futura. Cuando el volumen de información generado es muy elevado, se dificulta la identificación de la información verdaderamente importante y surge la necesidad de introducir en el proceso herramientas de gestión que apoyen la toma de decisiones en el proceso de evaluación de los RR.HH (Marín Sánchez & Lugo García, 2016).

En entrevistas realizadas a directivos y especialistas de proyectos de innovación tecnológica para la salud, se demostró que es más efectivo el proceso de chequeo y evaluación de los RR.HH. cuando se cuenta con información resumida y mostrada de forma explícita, que cuando solo se cuenta con grandes volúmenes de datos almacenados de forma numérica. El manejo y acceso a las fuentes de información no garantizan que el conocimiento y las experiencias acumuladas sean interpretados de manera útil y real, por lo que realizar una extracción de información exhaustiva forma parte del perfeccionamiento de todos los procesos actuales.

Un aspecto de prioridad de esta investigación lo constituye la función de recomendación inteligente que pueda proporcionar información valiosa que apoye la toma de decisiones en el proceso de evaluación del desempeño de los RR.HH de un proyecto. En este contexto es imprescindible el uso de técnicas de descubrimiento del conocimiento como el Aprendizaje Automático (*Machine Learning*) y la Minería de Datos (*Data Mining*, DM). Existe una clasificación ampliamente aceptada por la comunidad investigadora que pretende establecer una tipología de tareas diferenciadas de entre todas las incluidas en el amplio concepto de minería de datos. Dicha tipología está compuesta por seis tareas fundamentales: la detección de anomalías, el aprendizaje de reglas de asociación, la agrupación o clustering, la clasificación, la regresión y el resumen.

En esta investigación el resumen será objeto de estudio y aplicación por la importancia que cobra en la toma de decisiones. El objetivo de la presente investigación es desarrollar un método que permita realizar la interpretación de datos numéricos en información textual haciendo uso de herramientas de Soft Computing como las etiquetas lingüísticas y las sentencias cuantificadas (Yager, 1982), (Fonseca et al., 2020), que permita apoyar la toma de decisiones en la evaluación el desempeño de los RR.HH de un proyecto.

La investigación estará estructurada en Resumen, Introducción, Método y Resultados. En la sección Método se describen los principales elementos que sustentan la investigación, se analizan los elementos



fundamentales que caracterizan la Sumarización Lingüística de Datos, las sentencias lingüísticas y las protoformas. Al finalizar la sesión se describen los elementos del método para la generación de resúmenes lingüísticos. En la sesión Resultados se implementa el método propuesto sobre un caso real en un proceso de evaluación del desempeño de los Recursos Humanos de un proyecto. Al finalizar se muestran las Conclusiones y Referencias bibliográficas.

## Métodos o Metodología Computacional

Los líderes de proyecto necesitan modelos de evaluación del desempeño que sean capaces de ofrecer resultados entendibles e interpretables por los responsables del Departamento de Recursos Humanos (Merigó et al., 2014). En esta sesión se describen los principales referentes teóricos que serán abordados en la investigación. Al finalizar se propone un método para la construcción de resúmenes lingüísticos.

### Sumarización Lingüística de Datos (SDL)

La extracción de información de un conjunto de datos sin procesar suele ser muy engorroso para aquellas personas que tienen la responsabilidad de tomar decisiones basadas en ese conocimiento. Los expertos en el área de la Gestión de proyectos insisten en la necesidad de integrar técnicas de sintetización del conocimiento almacenado, de manera más interpretable, ya que no siempre la información almacenada atendiendo a criterios numéricos ayuda a facilitar la interpretación de los resultados.

Una manera muy utilizada para resolver los problemas de análisis e interpretación de la información es implementando funciones de descubrimiento de conocimiento a través de la construcción de resúmenes lingüísticos ya que se favorece la presentación de información al usuario, y que por tanto facilita el manejo y entendimiento de dicha información por parte de este (Bello & Verdegay, 2010). La acción de resumir consiste en reducir una cierta cantidad de información a términos breves y precisos. No sólo es posible realizar resumen lingüístico de información representada textualmente, sino que también puede hacerse de información con otras representaciones como las numéricas.

Para comprender mejor la función principal de los resúmenes se muestra el siguiente ejemplo:

**Ejemplo 1:** Se cuenta con una Base de datos que almacena la evaluación diaria de los RR.HH de un proyecto de innovación tecnológica para la salud, en un intervalo de  $[0,1]$  puntos, donde:



RH: cada trabajador evaluado

- a*: cumplimiento de cronogramas de entrega.
- b*: Satisfacción de los requisitos del cliente.
- c*: Calidad de los artefactos generados.
- d*: Aprovechamiento de la jornada laboral.

**Tabla 1.** Evaluación diaria de los RR.HH de un proyecto.

RH	Lunes				Martes				Miércoles				Jueves				Viernes			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	0.1	0.25	0.39	0.2	0.35	0.55	0.60	0.58	0.74	1	1	1	0.35	0.55	0.50	0.48	0.12	0.20	0.31	0.35
2	0.3	0.38	0.1	0.12	0.36	0.47	0.5	0.65	0.79	1	0.95	0.93	0.36	0.47	0.5	0.55	0.12	0.23	0.38	0.19
3	0.2	0.32	0.21	0.25	0.5	0.34	0.65	0.59	1	0.92	0.87	1	0.5	0.43	0.55	0.49	0.2	0.32	0.11	0.55
4	0.3	0.37	0.1	0.15	0.57	0.65	0.58	0.45	0.89	1	1	0.73	0.57	0.65	0.58	0.42	0.23	0.3	0.17	2.25
5	0.1	0.2	0.34	0.12	0.43	0.45	0.65	3	0.98	0.86	1	1	0.43	0.45	0.65	0.50	0.15	0.42	0.3	0.12
6	0.2	0.1	0.35	0.13	0.59	0.52	0.57	0.40	1	0.86	1	0.93	0.54	0.52	0.53	0.54	0.22	0.24	0.65	0.48

Basado en la Tabla I y haciendo uso de la SDL, un resumen quedaría de la siguiente manera:

- “La evaluación de los RRHH es mala al comenzar la semana, alcanza su máxima evaluación a mitad de semana y vuelve a ser mala al concluir la semana.”
- “Los días cercanos a los fines de semana son en los que se obtiene más mala evaluación.”
- “Los lunes y viernes la evaluación es mala, los martes y jueves es regular y los miércoles es buena.”

Un componente imprescindible de estos resúmenes son las variables lingüísticas, las cuales brindan la oportunidad de modelar de una forma más entendible los datos almacenados en registros históricos. Las variables lingüísticas ofrecen mayor versatilidad para modelar conceptos. Por esta razón, son una herramienta muy ampliamente utilizada en diversos métodos de resumen cuando en el proceso aparece implicado el lenguaje natural (Boran et al., 2014), (Yager, 2015). Para comprender mejor la función de las variables lingüísticas se muestra el siguiente ejemplo:

**Ejemplo 2:** Se tiene la sentencia “Los RR.HH tienen una evaluación de 3.84; 3.87; 3.89; 3.92; 3.85”

Para que un administrador de proyectos, aunque entiende la interpretación de estos valores, es difícil utilizarlos o recordarlos. Haciendo uso de las variables lingüísticas esta información se puede mostrar de la siguiente manera:

- “Los RR.HH tienen una evaluación regular”

Partiendo de estas aclaraciones se tiene que:



Dado:

$$Y = \{y_1, \dots, y_n\} \quad (1)$$

donde:

Y: conjunto de elementos de una Base de datos, en este caso conjunto de RR.HH.

$$A = \{A_1, \dots, A_m\} \quad (2)$$

Donde:

A: conjunto de atributos (en este escenario Variables difusas) que caracterizan objetos de Y, en este caso A son los atributos de cada uno de los RR.HH, Y. Ejemplo de A sería: *Tiempo planificado, Tiempo trabajado, Calidad.*

Por tanto, si se tiene una base de datos D, de evaluación de RR.HH, entonces  $A_j(y_i)$  denota un valor de atributo  $A_j$  para el objeto  $y_i$ .

### Composición de una proposición o sentencia lingüística

Atendiendo a las investigaciones de un SDL cuenta con un resumidor S, un cuantificador Q y un grado de validez o verdad T donde (Yager, 1982):

**S:** Summarizer (S), definido en el dominio de atributo, se compone por un atributo junto con un valor lingüístico definido en el dominio del atributo  $A_j$ . Ejemplo “Buena” para el atributo “Calidad”.

**Q:** Quantifier (Q), cuantificador lingüístico que expresa la cantidad. Conjunto difuso con universo de discurso en el intervalo  $[0, 1]$  que expresa una cantidad. Ejemplo: “Más” “Algunos”, “La mayoría”, “La mitad”, “Aproximadamente el 80%”, “Solo 3”, “Aproximadamente la mitad”.

**T:** Truth (T), grado de verdad (validez) del resumen. Es un número del intervalo  $[0,1]$  que evalúa la verdad del resumen. Generalmente, solo los resúmenes con un alto valor de T sirven de apoyo efectivo a la toma de decisiones.

**R:** Es un cuantificador opcional, constituye otro atributo junto con un valor lingüístico definido en el dominio del atributo  $A_k$ , ejemplo “joven” para el atributo “edad”.

Por lo tanto, los resúmenes lingüísticos pueden ejemplificarse según (Zadeh, 1997) como muestran las ecuaciones 3 y 4:

$$Qy \text{ 's are } S \quad (3)$$



QRy's are S (4)

**Ejemplo 3:** Atendiendo a estos ejemplos, un resumen sería:

Un resumen para Qy's son S, sería: "La mayoría de los RR.HH tienen un desempeño bueno."

Un resumen para QRy's son S, sería: "La mayoría de los RR.HH con edad joven tienen un desempeño bueno."

En la implementación de esta investigación se calculará T a partir de la propuesta de (Zadeh, 1997). Según muestran las ecuaciones 5 y 6.

$$T(Qy's are S) = \mu_Q \left[ \frac{1}{n} \sum_{i=1}^n \mu_S(y_i) \right] \quad (5)$$

$$T(QRy's are S) = \mu_Q \left[ \frac{\sum_{i=1}^n (\mu_R(y_i) \wedge \mu_S(y_i))}{\sum_{i=1}^n \mu_R(y_i)} \right] \quad (6)$$

Además de usar T, se han propuesto otras medidas de validez para determinar la calidad de los resúmenes. En (Kacprzyk & Yager, 2001) se puede encontrar:

- (T1) grado de verdad: corresponde con el mencionado T.
- (T2) grado de imprecisión: solo depende de la forma del resumen y expresa la falta de claridad del resumen debido a su descripción S.
- (T3) Grado de cobertura: indica cuántos objetos en la base de datos que se corresponden con el cuantificador R, están cubiertos por el resumen, es decir, por la descripción particular S.
- (T4) Grado de conformidad: describe cómo es la característica de la base de datos particular resumen encontrado.
- (T5) Longitud del resumen: se define utilizando el número de los términos del resumen S.

En contribuciones como (Kacprzyk & Zadrozny, 2009) se usa el concepto de protoforma como esqueleto tipo o plantilla en la que se basa la construcción del mensaje final. El término protoforma fue introducido en el ámbito de la computación por Zadeh en (Zadeh, 2002). Los protoformas forman una jerarquía donde a más alto nivel, más abstractas, en este caso la protoforma más abstracta corresponde a (3) y (4). La Tabla 2 muestra diferentes tipos de protoformas. Para esta investigación se utilizará la protoforma 5.

**Tabla 2.** Diferentes tipos de protoformas.

Tipo	Protoforma	Dado	Busca
------	------------	------	-------





0	QRy's are S	Todo	Validez(T)
1	Qy's are S	S	Q
2	QRy's are S	S y R	Q
3	Qy's are S	Q y S <sup>estructura</sup>	S <sup>valor</sup>
4	QRy's are S	Q,R y S <sup>estructura</sup>	S <sup>valor</sup>
5	QRy's are S	Nada	S,R y Q

### Reglas de asociación

Las reglas de asociación (RA) representan las relaciones entre varias variables unas con otras, es decir, se debe considerar que una RA es una implicación de la forma  $X \rightarrow Y$ , donde X es un conjunto de ítems llamados antecedentes, e Y es el ítem consecuente. Al menos cinco parámetros deben ser considerados en visualización de las RA: el conjunto de ítems antecedentes, ítems consecuentes, las asociaciones entre antecedentes y consecuentes, el soporte de las reglas, y la confianza de las reglas.

La investigación sobre visualización de RA se puede categorizar en tres grupos principales basados en: tablas, matrices, y grafos. Las técnicas basadas en tablas son el enfoque más común y tradicional para representar RA (González & Cornelio, 2013). Las columnas de una tabla representan generalmente los ítems, antecedentes, consecuentes, el soporte y la confianza de las RA, y cada fila representa una RA.

### Método para la Sumarización Lingüística de Datos usando reglas de asociación

La sumarización lingüística de datos puede ser abordada usando un enfoque basado en reglas de asociación (Kacprzyk & Zadrozny, 2016). Con este enfoque se pueden generar resúmenes del tipo: Si a1 es S1 y b1 es S2 entonces c1 es S3, siendo a1, b1 y c1 atributos de los objetos, y S1, S2 y S3 los sumarizadores modelados por términos difusos, usados en el resumen.

### Ejemplo 4:

*“Si X es largo y Y es medio entonces Z es grande”.*

*“Si cumplimiento de cronogramas es Superior y satisfacción de los requisitos es Adecuado entonces la calidad de los artefactos es Adecuado”.*

Las reglas de asociación para la generación de resúmenes de datos lingüísticos tienen como desventaja que si se generan muchas reglas entonces no sería de ayuda a los decisores en la toma de decisiones sobre la evaluación de los RR.HH de un proyecto. En investigaciones como (Zhao et al., 2021), se proponen



algoritmos para la extracción de reglas de asociación. Partiendo de estas investigaciones se propone un método para la sumarización lingüística de datos usando reglas de asociación que consta de las siguientes actividades:

Actividad 1: Seleccionar los atributos de los objetos.

Actividad 2: Extraer reglas de asociación

Actividad 3: Definir los Sumarizadores (S)

Actividad 4: Definir los cuantificadores (Q)

Actividad 5: Transformar las reglas en resúmenes lingüísticos

A continuación se detallan cada una de las actividades propuestas:

### **Actividad 1. Seleccionar los atributos de los objetos**

En este paso se realiza la selección de los atributos que se tendrán en cuenta para la construcción de las reglas de asociación. No todos los atributos almacenados tienen que incluirse necesariamente en las reglas.

Si se tiene el conjunto de atributos,

$$ACOMPLETO = \{A1, A2, A3, A4\},$$

A partir de la selección de los atributos de interés, se puede generar un nuevo conjunto

$$ASELECCIONADO = \{A1, A2, A3\}, \text{ siendo } ASELECCIONADO \subseteq ACOMPLETO$$

El valor que posee cada uno de estos atributos, es un valor numérico, como se mostró en la tabla 1. Para convertir las variables numéricas a variables lingüísticas, se utilizará un modelo de representación de la información lingüística basado en 2-tuplas. Este modelo está fundamentado en el concepto de traslación simbólica y permite operar con etiquetas lingüísticas sin pérdida de información. Para establecer las preferencias se emplean un conjunto de términos lingüísticos propuestos tal como muestra la tabla 3.

**Tabla 3.** Conjunto de valores lingüísticos empleados.

<b>Término lingüístico</b>	<b>Valor</b>
Deficiente (D)	(0,0;0,33)
Adecuado (A)	(0,33;0,66)
Superior (S)	(0,66;1)

### **Actividad 2. Extraer reglas de asociación**

La extracción de reglas de asociación se hará usando algoritmos propuestos para la extracción de reglas de asociación (Pupo et al., 2020).



Dentro de los algoritmos que más destacan en este ámbito se encuentra el a priori el cual Busca elementos frecuentes usando generación de candidatos. Su nombre se debe a que usa conocimiento a priori para la generación de ítems frecuentes. Este algoritmo se resume en dos etapas:

1. Generación de todos los ítems que contienen un solo elemento, utilización de estos para generar elementos que contengan dos elementos, y así sucesivamente. Se toman todos los posibles pares de ítems que cumplen con las medidas mínimas de soporte inicialmente preestablecidas; esto permite ir eliminando posibles combinaciones: aquellas que no cumplan con los requerimientos de soporte no entrarán en el análisis.
2. Generación de las reglas revisando que cumplan con el criterio mínimo de confianza. Es interesante observar que si una conjunción de consecuentes de una regla cumple con los niveles mínimos de soporte y confianza, sus subconjuntos (consecuentes) también los cumplen; en el caso contrario, si algún ítem no los cumple no tiene caso considerar sus superconjuntos.

### **Actividad 3. Definir los Sumarizadores (S)**

Un sumarizador  $S$ , es un atributo junto con un valor lingüístico (predicado difuso) definido en el dominio del atributo (Kacprzyk & Yager, 2001).

**Ejemplo 5:** Para el atributo "cumplimiento de cronogramas", un sumarizador puede ser "cumplimiento de cronogramas Deficiente"

### **Actividad 4. Definir los cuantificadores (Q)**

Un cuantificador lingüístico es un conjunto difuso definido para un universo de discurso entre  $[0, 1]$  tal que:

$$\mu_Q(x) = \begin{cases} 1 & \text{para } x \geq 0.8 \\ 2x - 0.6 & \text{para } 0.3 < x < 0.8 \\ 0 & \text{para } x \leq 0.3 \end{cases} \quad (7)$$

El cuantificador puede ser absoluto o relativo. Ejemplo de proposiciones lingüísticamente cuantificadas sería:

### **Ejemplo 6:**

Para la ecuación (3):





lingüísticos que le puedan servir a un decisor a la hora de tomar decisiones en el proyecto o sobre la evaluación el desempeño de un RR.HH.

**Actividad 1:** Seleccionar los atributos de los objetos

Si se tiene el conjunto de atributos:

$$ACOMPLETO = \{A_1, \dots, A_m\},$$

donde:

$$ACOMPLETO = \{ \text{"Cumplimiento de cronogramas"}, \text{"Satisfacción de requisitos"}, \text{"Calidad de artefactos"}, \text{"Aprovechamiento de jornada"}, \text{"Eficiencia"} \}.$$

A partir de la selección, de los atributos de interés se genera el nuevo conjunto:

$$ASELECCIONADO = \{ \text{"cumplimiento de cronogramas"}, \text{"Satisfacción de requisitos"}, \text{"Calidad de artefactos"}, \text{"Aprovechamiento de jornada"} \}, \text{ siendo } ASELECCIONADO \subseteq ACOMPLETO$$

Atendiendo al conjunto de valores lingüísticos definidos en la tabla 2 se procede a la discretización de los datos donde los valores numéricos de la tabla 1 serán transformados en datos lingüísticos, tal como muestra la tabla 4.

**Tabla 4.** Datos con valores lingüísticos.

RH	Lunes				Martes				Miércoles				Jueves				Viernes			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	D	D	A	D	A	A	A	A	S	S	S	S	A	A	A	A	D	D	D	A
2	D	A	D	D	A	A	A	A	S	S	S	S	A	A	A	A	D	D	A	D
3	D	D	D	D	A	A	A	A	S	S	S	S	A	A	A	A	D	D	D	A
4	D	A	D	D	A	A	A	A	S	S	S	S	A	A	A	A	D	D	D	D
5	D	D	A	D	A	A	A	A	S	S	S	S	A	A	A	A	D	A	D	D
6	D	D	A	D	A	A	A	A	S	S	S	S	A	A	A	A	D	D	A	A

**Actividad 2: Extraer reglas de asociación**

La extracción de reglas de asociación se hará usando el algoritmo 2 LDSAssociationRules, propuesto por (Castillo Ortega, 2013), para el minado de reglas de asociación, que está inspirado en el algoritmo a priori y que se muestra a continuación:

**Algoritmo 2:** LDSAssociationRules



Esta obra está bajo una licencia *Creative Commons* de tipo **Atribución 4.0 Internacional** (CC BY 4.0)

**Entradas :**

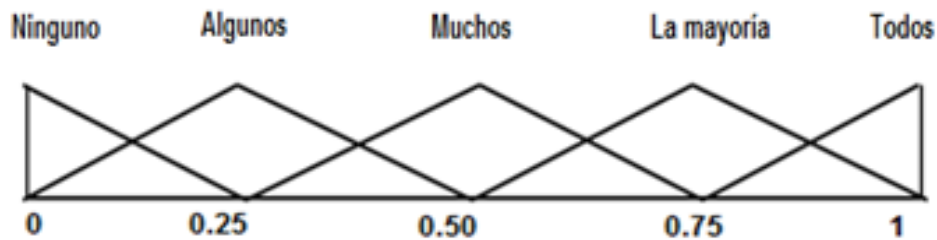
- D: conjunto de datos para el análisis compuesto por un conjunto de atributos A
- T: conjunto de métodos predefinidos para el cálculo de los valores de calidad T.
- CLV: variable lingüística de los cuantificadores de los resúmenes.
- Paso 1.** Construir la variable lingüística de cada uno de los  $A_i$  atributos
- Paso 2.**  $A_{LV} = \text{linguistic\_variables}(D)$
- Paso 3.** Transformación de los datos D en borrosos  $DB = \Delta(D, A_{LV})$
- Paso 4.** Establecer valores de lift, soporte y confianza.
- Paso 5.** Generar reglas de asociación  $\text{apriori\_rules} = \text{do\_apriori}(DB, \text{support}, \text{confidence}, \text{lift})$
- Paso 6.** Calcular los cuantificadores a partir del soporte y la confianza
- Paso 7.**  $\text{quantiffiers} = \text{calculate\_quantiffiers}(C_{LV}, \text{support}, \text{confidence})$
- Paso 8.** Generar resumen por cada regla  $\text{generate\_LDS}(\text{apriori\_rules}, \text{quantiffiers}, T)$

**Actividad 3: Definir los Sumarizadores (S)**

El par formado por el atributo  $A_i$  y el valor lingüístico definido en la tabla 2, conforma un sumador S para un atributo A.

**Actividad 4: Definir los cuantificadores (Q)**

Los cuantificadores definidos en esta investigación están dados según muestra la fig. 1.



**Fig. 1.** Variable lingüística de los cuantificadores (Q) de los resúmenes.

**Actividad 5: Transformar las reglas en resúmenes lingüísticos**



Las reglas de asociación obtenidas son transformadas en resúmenes lingüísticos. El cálculo del grado de validez (T) de los resúmenes obtenidos fue realizado según las ecuaciones (5) y (6) obteniendo en todos los casos un valor para  $T > 0.85$ , lo cual se considera adecuado según <sup>(3)</sup>. Algunos de los resúmenes obtenidos se muestran a continuación:

**Resumen:** *"Todos los RR.HH con cumplimiento de cronogramas Superior, Satisfacción de requisitos Superior, Calidad de artefactos Superior tienen Aprovechamiento de jornada Superior"*.

- Interpretación: Cuando un RR.HH cumple los planes pactados y los cronogramas de entrega con calidad, significa que tiene un buen aprovechamiento de la jornada laboral.
- Decisión a tomar: El evaluador debe generar incentivos de estímulo al RR.HH.

**Resumen:** *"Algunos RR.HH con cumplimiento de cronogramas Deficiente, Satisfacción de requisitos Deficiente, Calidad de artefactos Deficiente, tienen Aprovechamiento de jornada Adecuado"*

- Interpretación: No es posible que un RR.HH que no cumpla con ningún indicador del desempeño haya aprovechado adecuadamente la jornada laboral.
- Decisión a tomar: El evaluador debe identificar las fallas en la evaluación de indicadores del desempeño.

**Resumen:** *"Todos los RR.HH, los día miércoles tienen desempeño Superior"*.

- Interpretación: El miércoles es el día de mayor eficiencia y eficacia de los RR.HH del proyecto
- Decisión a tomar: Generar ese día las tareas de mayor impacto y complejidad para el proyecto.

**Resumen:** *"Todos los RR.HH, los día lunes tienen desempeño Deficiente"*.

- Interpretación: El desempeño laboral de los RR.HH no es bueno los lunes.
- Decisión a tomar: Identificar tareas que eleven la motivación de los RR.HH para que sean desempeñadas los lunes.

## Conclusiones



La sumarización lingüística de datos es un método de softcomputing que apoya la toma de decisiones extrayendo el conocimiento de grandes volúmenes de datos, no interpretables para los humanos sin ayuda computacional.

En este trabajo se implementa un método basado en sumarización lingüística de datos basado en la generación de reglas de asociación para apoyar la toma de decisiones en el proceso de evaluación de los RR.HH de un proyecto de innovación tecnológica para la salud. La implementación del método propuesto, permitió identificar resúmenes lingüísticos que pueden servir de base para tomar decisiones sobre el desempeño laboral de los RR.HH de un proyecto.

Como trabajos futuros se pretende utilizar el método propuesto en la evaluación de proyectos que son administrados con herramientas de gestión de proyectos.

## Referencias

- Bello, R., & Verdegay, J. L. (2010). Los conjuntos aproximados en el contexto de la Soft Computing. *Revista Cubana de Ciencias Informáticas*, 4(1-2).
- Boran, F. E., Akay, D., & Yager, R. R. (2014). A probabilistic framework for interval type-2 fuzzy linguistic summarization. *IEEE Transactions on Fuzzy Systems*, 22(6), 1640-1653.
- Castillo Ortega, R. M. (2013). *Resumen lingüístico de series de datos mediante técnicas de soft computing: una aplicación a los cubos olap con dimensión tiempo*. Universidad de Granada. <https://digibug.ugr.es/bitstream/handle/10481/24731/2146621x.pdf?sequence=1&isAllowed=y>
- Fonseca, B. B., Cornelio, O. M., & Marzo, F. R. R. (2020). Tratamiento de la incertidumbre en la evaluación del desempeño de los Recursos Humanos de un proyecto basado en conjuntos borrosos. *Serie Científica de la Universidad de las Ciencias Informáticas*, 13(6), 84-93. <https://dialnet.unirioja.es/download/articulo/8590282.pdf>
- González, J. L. G., & Cornelio, O. M. (2013). Propuesta de algoritmo de clasificación genética. *Revista Cubana de Ingeniería*, 4(2), 37-42. <https://biblat.unam.mx/hevila/Revistacubanadeingenieria/2013/vol4/no2/4.pdf>





- Kacprzyk, J., & Yager, R. R. (2001). Linguistic summaries of data using fuzzy logic. *International Journal of General System*, 30(2), 133-154.
- Kacprzyk, J., & Zadrozny, S. (2009). Protoforms of linguistic database summaries as a human consistent tool for using natural language in data mining. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, 1(1), 100-111.
- Kacprzyk, J., & Zadrozny, S. (2016). Queries with fuzzy linguistic quantifiers for data of variable quality using some extended OWA operators. Flexible Query Answering Systems 2015: Proceedings of the 11th International Conference FQAS 2015, Cracow, Poland, October 26-28, 2015,
- Marín Sánchez, J., & Lugo García, J. A. (2016). Control de proyectos de software: actualidad y retos para la industria cubana. *Ingeniare. Revista chilena de ingeniería*, 24(1), 102-112.  
<https://www.scielo.cl/pdf/ingeniare/v24n1/art10.pdf>
- Merigó, J. M., Casanovas, M., & Yang, J.-B. (2014). Group decision making with expertons and uncertain generalized probabilistic weighted aggregation operators. *European Journal of Operational Research*, 235(1), 215-224.
- Pupo, I. P., Vacacela, R. G., Pérez, P. P., Mahdi, G. S. S., & Peña, M. (2020). Experiencias en el uso de técnicas de softcomputing en la evaluación de proyectos de software. *Investigación Oper*, 41, 108-119. <https://rev-inv-ope.pantheonsorbonne.fr/sites/default/files/inline-files/41120-08.pdf>
- Torres-López, S., Lugo-García, J. A., Piñero-Pérez, P. Y., Torres-Quñones, K. M., Perdomo-Alonso, A., Cuza-García, B., & Aldana-Cuza, M. L. (2014). Técnicas formales y de inteligencia artificial para la gestión de recursos humanos en proyectos informáticos. *Revista Cubana de Ciencias Informáticas*, 8(3), 43-55. <http://scielo.sld.cu/pdf/rcci/v8n3/rcci04314.pdf>
- Yager, R. R. (1982). A new approach to the summarization of data. *Information Sciences*, 28(1), 69-86.
- Yager, R. R. (2015). Multicriteria decision making with ordinal/linguistic intuitionistic fuzzy sets for mobile apps. *IEEE Transactions on Fuzzy Systems*, 24(3), 590-599.
- Zadeh, L. A. (1997). The roles of soft computing and fuzzy logic in the conception, design and deployment of intelligent systems. Proceedings of 6th International Fuzzy Systems Conference,



- Zadeh, L. A. (2002). A prototype-centered approach to adding deduction capability to search engines-the concept of protoform. 2002 annual meeting of the North American fuzzy information processing society proceedings. NAFIPS-FLINT 2002 (Cat. No. 02TH8622),
- Zhao, Z., Jian, Z., Gaba, G. S., Alroobaea, R., Masud, M., & Rubaiee, S. (2021). An improved association rule mining algorithm for large data. *Journal of Intelligent Systems*, 30(1), 750-762.

### **Conflicto de interés**

Los autores autorizan la distribución y uso de su artículo.

### **Contribuciones de los autores**

1. Conceptualización: Barbara Bron Fonseca
2. Curación de datos: Barbara Bron Fonseca, Omar Mar Cornelio
3. Análisis formal: Barbara Bron Fonseca, Omar Mar Cornelio
4. Investigación: Barbara Bron Fonseca, Omar Mar Cornelio
5. Metodología: Barbara Bron Fonseca, Omar Mar Cornelio
6. Administración del proyecto: Barbara Bron Fonseca
7. Recursos: Barbara Bron Fonseca, Omar Mar Cornelio
8. Software: Barbara Bron Fonseca, Omar Mar Cornelio
9. Supervisión: Omar Mar Cornelio
10. Validación: Barbara Bron Fonseca, Omar Mar Cornelio
11. Visualización: Barbara Bron Fonseca, Omar Mar Cornelio
12. Redacción – borrador original: Barbara Bron Fonseca, Omar Mar Cornelio
13. Redacción – revisión y edición: Barbara Bron Fonseca, Omar Mar Cornelio

### **Financiación**

La investigación no requirió fuente de financiamiento.

