

# Caracterización de variables utilizando inteligencia computacional para identificar alteraciones en la salud de bovinos

## *Variables characterization by using computing intelligence to identify the cattle's health disorders*

Fecha de recepción: 18 de octubre de 2014  
Fecha de aceptación: 16 de diciembre de 2014

Edgar Leonardo Sarmiento-Pacanchique<sup>1</sup>,  
Oscar Iván Torres-Corredor<sup>2</sup>, Javier Antonio Ballesteros-Ricaurte<sup>3</sup>,  
Gustavo Cáceres-Castellanos<sup>4</sup>

### Resumen

La detección de alteraciones de la salud de los animales mediante pruebas de laboratorio es un proceso complejo que implica relacionar diversas variables y factores clínicos de los individuos; por ello, en esta investigación se evaluaron técnicas de inteligencia computacional que contribuyeron a la identificación de patrones de comportamiento de las alteraciones detectadas en las pruebas de hemograma aplicadas en bovinos. Aunque diversos algoritmos de inteligencia computacional son utilizados en la solución de problemas médicos, no se encontró registro de investigaciones en procesos médicos veterinarios. Una vez hecha una minuciosa caracterización de las variables y la evaluación de las técnicas de inteligencia computacional, se determinó que el algoritmo que mejor se ajusta al propósito de análisis de datos planteado es **FP-Growth**.

**Palabras clave:** Análisis de datos, Algoritmos, Sanidad animal, FP-Growth.

### Abstract

Detecting disorders in lab tests applied in animals is a complex process that implies linking different variables and clinical factors of the individuals. It is why during the development of the present research, some computing intelligence techniques were evaluated, which contributed to the behavior patterns identification of the most important disorders detected in CBC tests applied in cattle. Although several computing intelligence algorithms are used in medical troubleshooting, no record of researches in veterinary

---

1 Profesional Independiente.

2 Profesional Independiente.

3 M.Sc. Universidad Pedagógica y Tecnológica de Colombia (Tunja, Colombia). javier.ballesteros@uptc.edu.co.

4 M.Sc. Universidad Pedagógica y Tecnológica de Colombia (Tunja-Colombia). gustavo.caceres@uptc.edu.co.

medical processes was found. Once the thorough characterization of the variables and the evaluation of the computing intelligence techniques were made, it was determined that the algorithm that best fits to the purpose of the proposed data analysis is **FP-Growth**.

**Keywords:** Data analysis, Algorithms, Animal Health, FP-Growth.

## Introducción

El uso de herramientas computacionales ofrece una alternativa confiable a los investigadores para descubrir conocimiento que a simple vista no puede ser detectado. Los algoritmos de inteligencia computacional son aplicados en procesos médicos con el objetivo de identificar alteraciones en la salud, que coadyuvan a los profesionales de la salud en el diagnóstico de enfermedades.

El Grupo de Investigación en Medicina Veterinaria y Zootécnica de la Universidad Pedagógica y Tecnológica de Colombia (Uptc) realiza estudios epidemiológicos en diferentes municipios del departamento de Boyacá, donde registra variables que dan indicios del estado sanitario de las fincas y de sus animales. Los resultados obtenidos en este proceso ofrecen un marco de trabajo para los investigadores en la detección y prevención de anomalías presentadas por los individuos del área en estudio. Por tal razón, con el objetivo de apoyar estas actividades sanitarias se evaluaron ocho algoritmos de inteligencia computacional, de los que seleccionó uno, que al ser implementado permitió identificar alteraciones en los resultados de los estudios desarrollados por el grupo de investigación para el apoyo de diagnósticos.

## Contextualización

La identificación de estudios enmarcados en el apoyo de sistemas de información a procesos médicos proveyó información valiosa de los problemas presentados, de las técnicas aplicadas y de los resultados obtenidos; adicionalmente, ofreció una base de conocimiento sólida que soportó la realización de esta investigación.

**Sanidad animal:** la sanidad animal es el estado fisiológico óptimo en el organismo de un animal para mostrar su potencial de producción. Cuando este estado se encuentra afectado es porque el individuo presenta algún tipo de enfermedad que perturba su normal desenvolvimiento en su entorno (3).

Las enfermedades que afectan la sanidad animal pueden presentarse por diferentes agentes pató-

genos, como virus, bacterias, parásitos y hongos, o por alteraciones nutricionales o predisposición hereditaria.

Los problemas de salud animal también pueden convertirse en problemas de salud en los humanos, debido a que la mayoría de enfermedades que afectan a las personas tienen origen animal y son transmitidas por animales. La correcta estructuración y ejecución de planes de sanidad contribuirá a la mitigación de enfermedades en humanos causadas por la transmisión de vectores de origen animal.

Las fincas implementan medidas sanitarias para garantizar salubridad en los individuos que las conforman; estas medidas deben ser correctamente estructuradas y evaluadas para que se ajusten a las necesidades propias de cada población. Algunos aspectos importantes que contribuyen a mantener los individuos libres de enfermedades son (3):

- Diagnóstico de enfermedades
- Planes de vacunación
- Registros sanitarios
- Nutrición balanceada a los animales

**Inteligencia computacional:** en los últimos años ha aumentado la capacidad de procesamiento de los equipos de computación, pero, a pesar de ello, aún algunas tareas que implican grandes cálculos de información tardan mucho tiempo en ejecutarse. Los organismos biológicos realizan tareas similares todo el tiempo y todo los días, es por eso que los expertos en computación han querido simular dichos procesos para ser ejecutados en un computador. La Inteligencia Computacional (IC) se ocupa de la teoría, diseño, desarrollo y aplicación de paradigmas computacionales inspirados en procesos biológicos (1).

Los términos de inteligencia computacional e inteligencia artificial (IA) se suelen confundir, pero estas dos disciplinas tienen enfoques diferentes, algunos aspectos en los que difieren son (2):

- **Fuente de representación de la inteligencia:** la IA se basa en la representación del conocimiento de los expertos de cierto dominio,

mientras que la IC extrae el conocimiento de los datos disponibles.

- **Los mecanismos para el procesamiento de la inteligencia:** la IC utiliza métodos de razonamiento simbólico, y la IA está basada en métodos numéricos.
- **Las interacciones con el ambiente:** la IA ajusta el ambiente a las soluciones conocidas representadas por conocimiento estático, mientras que la IC trata de aprender con el ambiente.

**Inteligencia Computacional en la solución de problemas médicos:** el uso de técnicas de IC en investigaciones médicas es un proceso emergente aún, sin embargo, algunos estudios desarrollados en este campo han dado resultados significativos, tal como se describe a continuación:

- Los sistemas de información inteligentes no suplantán la labor de los profesionales de la salud, ofrecen conocimiento implícito de los datos que permite identificar alteraciones de la salud de los individuos analizados, tal como lo comentan Cabrera, Pardeni, Delgado y otros en su artículo "Aplicaciones médicas como ayuda al diagnóstico en la medicina" (4). Esta investigación propone la implementación de un sistema experto que contenga una lista de reglas con la información sobre medicamentos, plaguicidas, animales, plantas, drogas psicotrópicas, productos industriales y productos de hogar, que permitan diagnosticar problemas de toxicología en los pacientes analizados.
- La investigación "Extracción de reglas de asociación en una base de datos clínicos de pacientes con VIH/SIDA" (5), de los autores Chausa Fernández, Gómez Aguilera y Cáceres Taladriz, identificó las relaciones entre los diferentes atributos almacenados en una base de datos de historias clínicas de las personas infectadas con VIH/SIDA, que son tratadas en el Hospital Clinic de Barcelona, España. Los autores utilizaron la técnica de reglas de asociación para detectar patrones de comportamiento en los datos de los pacientes, y obtuvieron

como resultado asociaciones que los médicos ya conocían por la experiencia, pero además identificaron nuevos patrones, los cuales sirven como insumo para alimentar otras líneas de investigación médica (5).

- La investigación "Diseño de un algoritmo genético para la generación de conocimiento en el diagnóstico del Síndrome de Autista" (6), desarrollada por Yima Cárdenas, Luis Guerra, David Santos y Mauricio Sánchez, plantea el desarrollo de un algoritmo genético para diagnosticar parcialmente la presencia de los síntomas del autismo en una persona, teniendo como base una serie de características determinadas como hallazgos dentro de la medicina. Cada hallazgo es representado con los números del sistema binario, de acuerdo con el dictamen de la Organización Mundial de la Salud (OMS), el Instituto de investigación en autismo (ARI) y la Asociación Americana de Psiquiatría (APA).

## Metodología

La investigación se desarrolla en tres fases: la primera corresponde a la caracterización de las variables; la segunda, a la evaluación de los algoritmos, y la tercera a la aplicación del algoritmo, relacionando las variables para la búsqueda de alteraciones en la salud de los bovinos. A continuación se describen estos procesos.

### A. Caracterización de variables

La caracterización de las variables utilizadas por el grupo para registrar el estado sanitario de la población en estudio se enfocó en dos actores importantes que se describen a continuación:

**Variables de la finca.** Los datos requeridos para la finca se enfocan en información sanitaria, nutricional, epidemiológica y reproductiva. La Tabla I muestra el consolidado de variables identificadas.

**Tabla I.** Información general finca

Aspecto	Descripción
Información básica	Datos de identificación básica.
Ubicación	Incluye datos de ubicación geográfica de la finca dentro del departamento de Boyacá
Muestras	Corresponde a las muestras enviadas al laboratorio.
Síntoma	Son los principales síntomas presentados por los individuos de la finca en un periodo de un año.
Inventario	Inventario de animales presentes en la finca, especificando especie y clasificación.
Eventos	Los eventos corresponden a procedimientos sanitarios aplicados a grupos o la totalidad de los individuos.
Parámetros	Los parámetros describen variables relevantes para cada finca en procesos lácteos y reproductivos.

**Variables del individuo.** Los datos registrados para los individuos corresponden a información de su comportamiento lácteo, reproductivo y clínico. La Tabla II corresponde a la información caracterizada.

**Tabla II.** Información del individuo

Aspecto	Descripción
Identificación	Datos básicos de identificación
Producción láctea	Información del comportamiento lácteo
Parámetros reproductivos	Información del comportamiento reproductivo
Laboratorio	Resultados de laboratorio de las pruebas aplicadas

## B. Selección del algoritmo de inteligencia computacional

En esta fase del proyecto se caracterizaron y evaluaron ocho algoritmos de IC, para seleccionar uno e implementarlo en el proyecto. La caracterización de los algoritmos por evaluar se hizo identificando cuáles de estos habían incursionado en el campo de la medicina y habían sido aplicados en procesos médicos.

La metodología para evaluar los algoritmos se basó en una valoración de ponderaciones y gradaciones de parámetros y factores (7). Una vez caracterizados los algoritmos y definida la metodología de evaluación, se realizó la valoración para cada uno de estos. Las Tablas III y IV muestran los parámetros y factores utilizados en el proceso de valoración.

**Tabla III.** Parámetros de evaluación

Parámetro	Descripción
Eficiencia	Capacidad de obtener resultados óptimos utilizando la técnica de computación más sencilla.
Costo computacional	Cantidad de recursos consumidos durante la ejecución del algoritmo.
Rendimiento	Capacidad de arrojar resultados precisos en tiempos cortos.

**Tabla IV.** Factores de evaluación

Parámetro	Factor	Descripción
Eficiencia	Simplicidad algorítmica	Capacidad del algoritmo en procesar los datos de manera más simple que otros algoritmos, obteniendo resultados de significativos.
Costo computacional	Consumo de hardware	Cantidad de recursos de máquina que consume el algoritmo, teniendo en cuenta aspectos como el crecimiento en el número de datos por procesar y la capacidad de la máquina donde se ejecuta.
	Acceso a los datos	Forma en la que el algoritmo accede a los datos que van a ser procesados.
Rendimiento	Tiempo de ejecución	Tiempo utilizado por el algoritmo para ejecutarse y presentar los resultados.
	Precisión	Calidad de los resultados arrojados luego de ejecutar el algoritmo.

Una vez definidos los parámetros y factores, se asignó una valoración para cada uno de estos, de acuerdo con una adaptación de la escala Likert, donde 1 es bajo y 10 es alto. Las Tablas V y VI muestran la calificación asignada a cada parámetro y factor.

**Tabla V.** Valoración parámetros de evaluación

Parámetro	Ponderado
Eficiencia	10
Costo computacional	8
Rendimiento	8

**Tabla VI.** Valoración factores de evaluación

Factor	Ponderado
Simplicidad algorítmica	8
Consumo hardware	7
Acceso a los datos	7
Tiempo de ejecución	9
Precisión	8

## Análisis de los resultados

**Resultado de la evaluación de los algoritmos:** fue un proceso complejo que involucró la utilización de una metodología para garantizar confiabilidad

en los resultados obtenidos. El estado del arte de los algoritmos en la solución de problemas médicos fue el factor clave en la valoración de los parámetros y factores.

El algoritmo FP-Growth es uno de los más eficientes de las reglas de asociación, porque utiliza una estructura en memoria llamada FP-tree (árbol de patrones recurrentes), la cual permite reducir el número de consultas a la fuente de datos; este algoritmo está basado en un árbol de prefijos de una base de datos de transacciones, y utiliza un esquema de eliminación recursiva, en la que se van eliminando los conjuntos de datos no frecuentes (8); obtuvo la calificación final más alta con un puntaje de 4.34, lo que promete resultados significativos en la detección de patrones de comportamiento en los datos, apoyando de esta manera a los investigadores del Gidimevezt en la detección de enfermedades y anomalías médicas en los animales en estudio.

La utilización del algoritmo de IC FP-Growth para el análisis de datos permite identificar patrones de comportamiento en los resultados de las pruebas de hemograma. La Fig.1 muestra los patrones de comportamiento identificados en un estudio determinado.



Figura 1. Patrones de comportamiento identificados.

Con el fin de ofrecer un contexto amplio al análisis de los resultados de las pruebas de hemograma, el sistema permite graficar el comportamiento de las alteraciones y variables en estudio.

**Comportamiento de alteraciones:** la gráfica del comportamiento de las alteraciones mide la cantidad de individuos del proyecto que presentaron cierta alteración, tal como se presenta en la Fig.2.

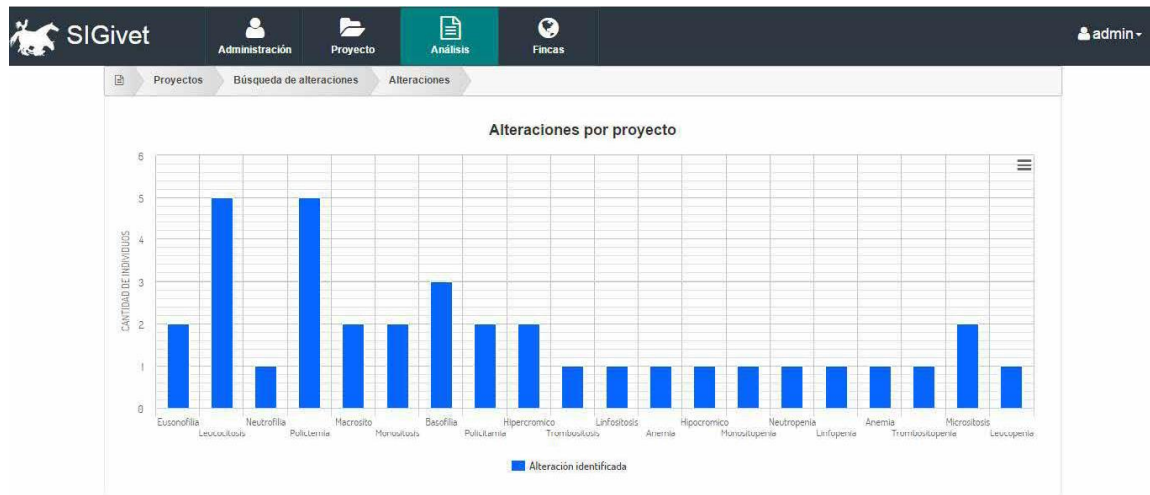


Figura 2. Comportamiento de las alteraciones.



**Comportamiento de variables:** comportamiento de variables cuenta el número de individuos que presentaron valores normales y valores alterados para cada variable de las pruebas de hemograma, tal como se visualiza en la Fig.3.



**Figura 3.** Comportamiento de las variables.

## Conclusiones

La cantidad de variables y actores involucrados en el desarrollo de proyectos de sanidad animal es cuantiosa; por tal razón se hizo una minuciosa caracterización de las variables que permitió identificar y corregir algunos problemas de integridad de los datos solicitados, obteniendo al final un nuevo formato para el registro de la información de las fincas.

La implementación del algoritmo FP-Growth en un sistema de información apoya a los investigadores del grupo en el diagnóstico de enfermedades, por medio de la identificación de patrones de comportamiento de las alteraciones detectadas en las variables de las pruebas de hemograma.

La administración de la información salubre en la aplicación desarrollada permite contar con datos actualizados del estado sanitario de los animales y vigilar el comportamiento de alteraciones identificadas, contribuyendo de esta forma a la disminución de enfermedades en las personas, debido a que cerca del 60% de las enfermedades en huma-

nos son de origen animal y, aproximadamente, el 75% son transmitidas por animales.

## Referencias

- (1) J. Muñoz Pérez, *Inteligencia computacional inspirada en la vida*, Universidad de Málaga, España, 2011.
- (2) J. L. Camargo-Cuervo, E. F. Suárez-Mendoza, and J. A. Ballesteros-Ricaurte, "Comparison between Oracle BPM and JBPM in the Admission Process Optimization", *Fac. Ing. UPTC*, vol. 22, no. 34, pp. 85-96, 2013.
- (3) A. R. Novoa, *Salud animal, manejo y administración de sistemas de producción de leche*. San José de Costa Rica, 1983.
- (4) M. Cabrera-Hernández, M. del C. Paderini-López, R. Hita Torres, A. Delgado-Ramos, M. A. Tardío-López, and D. Derivet-Thaureaux, "Aplicaciones médicas como ayuda al diagnóstico en la medicina. Experiencia SOfTEL-MINSAP", *Rev. Cuba. Informática Médica*, vol. 4, no. 2, pp. 199-212, 2012.



- (5) P. Chausa-Fernández, E. J. Gómez-Aguilera, C. Cáceres-Taladriz, F. García-Alcaide, and J. M. Gatell-Artigas, *Extracción de Reglas de Asociación en una Base de Datos Clínicos de Pacientes con Vih/Sida*, p. 4, 2006.
- (6) Y. Cárdenas, L. Guerra y D.Mauricio, *Diseño de un algoritmo genético para la generación de conocimiento en el diagnóstico del Síndrome Autista*. Facultad de Ingeniería de Sistemas e Informática Universidad Nacional Mayor de San Marcos, 2014.
- (7) R. C. Naranjo-Cuervo and L. M. Sierra-Martínez, "Software tool for analyzing the family shopping basket without candidate generation", *Ing. e Investig.* Vol. 29, No. 1, 2010.
- (8) J. S. González-Sanabria and G. Cáceres-Castellanos, "Comparison of GIS desktop tools-fordevelopment of SIGPOT," *Lat. Am. Trans. IEEE (Revista IEEE Am. Lat.)*, vol. 11, no. 1, pp. 86-90, 2013.