

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

<http://dx.doi.org/10.35381/r.k.v5i9.649>

Analítica de datos puros dentro del ámbito productivo y reproductivo de las ganaderías de leche

Pure data analytics within the productive and reproductive field of dairy farms

Santiago Morales-Cardoso

smorales@uce.edu.ec

Universidad Central del Ecuador, Quito
Ecuador

<https://orcid.org/0000-0002-3833-9654>

Mario Morales-Morales

mmoralesm@uce.edu.ec

Universidad Central del Ecuador, Quito
Ecuador

<https://orcid.org/0000-0002-7493-8072>

Alicia Andrade-Bazurto

aandrade@uce.edu.ec

Universidad Central del Ecuador, Quito
Ecuador

<https://orcid.org/0000-0002-7153-1875>

Laura Cevallos-Black

lgcevallos@uce.edu.ec

Universidad Central del Ecuador, Quito
Ecuador

<https://orcid.org/0000-0002-6886-1013>

Recibido: 1 de noviembre de 2019

Revisado: 20 de noviembre de 2019

Aprobado: 15 de diciembre de 2019

Publicado: 20 de enero de 2020

RESUMEN

La investigación presenta la metodología M3S como alternativa en el desarrollo de proyectos basados en la analítica de datos, por cuanto permite usar modelos matemáticos y explotar datos para generar conocimiento sin dejar de lado aspectos importantes para cualquier actividad como el nivel de madurez de la información y métodos que tienen que ver con Inteligencia Artificial como son los árboles de decisión

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

y el aprendizaje automático, el principal objetivo de la metodología es juntar todos estos aspectos que servirán de mucha ayuda dentro de los procesos de extracción, transformación y carga, asegurándonos que las decisiones que sean tomadas dentro de la organización sean las mejores. Mediante la implementación de una metodología de decisión se puede mejorar la explotación de información y obtener resultados más confiables, llegando a tomar decisiones más oportunas y eficaces en el tiempo.

Descriptor: Genética animal; nutrición animal; inteligencia artificial. (Palabras tomadas del Tesoro AGROVOC Multilingüal).

ABSTRACT

The research presents the M3S methodology as an alternative in the development of projects based on data analytics, because it allows to use mathematical models and exploit data to generate knowledge without neglecting important aspects for any activity such as the maturity level of the information and methods that have to do with Artificial Intelligence, especially, decision trees and machine learning. The main objective of the methodology is bringing all these aspects that will help a lot in the processes of extraction, transformation and loading, ensuring the best choice within the organization. By implementing a decision methodology, the exploitation of information can be improved and more reliable results may be obtained, making more timely and effective decisions over time.

Descriptors: Animal genetics; animal nutrition; artificial intelligence. (Words taken from the AGROVOC Multilingual Thesaurus).

INTRODUCCIÓN

En la actualidad el uso de la Analítica de Datos permite obtener información relevante y confiable la cual permite generar indicadores estratégicos para el negocio ya que se puede realizar acciones y tomar buenas decisiones que aporten con la productividad del mismo. Se habla mucho sobre la inteligencia de negocios y se dice que: “Se entiende por BI al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización”. (Díaz, 2010), (Vilchez, 2011); y otra como en la que se afirma que: “El objetivo primario de la Inteligencia de Negocios es contribuir a tomar decisiones que mejoren el desempeño de la empresa y promover su ventaja competitiva en el mercado” (Calzada & Abreu, 2009), en definitiva se analizan dos perspectivas: Tomar mejores decisiones más rápido y convertir datos en información.

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Uno de los problemas que se presentan durante este tipo de proyectos es el retraso significativo que existe durante el proceso; por lo que hemos conveniente usar algoritmo de decisión basado en ID3, con nodos que brinden más información, provenientes de atributos que se los trae de variadas fuentes de datos ya sean de bases estructuradas o no, los mismos que deben ser revisados por medio de sentencias sql para procesos de limpieza para que en el estudio los conozcamos como datos “puros”. Una vez con nodos puros estos deben ser clasificados en base a cálculos recursivos de entropía y ganancia de información para conseguir estructurar un árbol de decisión útil.

En la propuesta metodológica una de las definiciones importantes es la de *Árbol de Decisión*, del que en una de ellas dice: “Is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decisión”. (Peng, Chen, & Zhou, 2012).

Metodología M3S

M3S, es una metodología fácil y práctica con la que se pueden llevar a cabo proyectos de BI, y sus etapas están descritas literalmente a continuación.

Tabla 1

Ciclo de Vida M3S

A. JUSTIFICACIÓN

Evaluar negocio/ Constitución

B. PLANIFICACIÓN

Evaluar Infraestructura Técnica – No técnica /
Alcance / Cronograma

C. ANÁLISIS REQUISITOS DE NEGOCIOS

Requisitos / Prototipos

D. DISEÑO

Metodología Decisión

E. CONSTRUCCIÓN

Desarrollo Aplicativo

F. IMPLEMENTACIÓN

Puesta en marcha / Evaluar solución / Cierre Proyecto

Fuente: (Morales, Morales, & Rizo, 2019)

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Algoritmo de decisión

Para la extracción y transformación de datos usamos un algoritmo de decisión aplicando análisis de la entropía y ganancia de información (Quinlan, 1986) que como resultado describa la posición de cada Nodo (Atributo) en el árbol. Según lo descrito en la tabla 2 de la metodología de decisión se cuenta con las siguientes etapas:

Tabla 2

Metodología de Decisión

ETAPAS
<p>1. Análisis de Calidad de Datos.</p> <p>Dado A conjunto datos estructurados, y B No estructurados (redes sociales) ponderados (1..5) Si $X \in A$ ó $X \in B$</p> <p style="padding-left: 20px;">Sea $f : X \rightarrow Y$, en donde:</p> <p style="padding-left: 20px;">X casos posibles de entradas; Y espacio de salidas</p> <p>2. Hipótesis de Datos</p> <p>Serán todas las $h \in H$</p> <p>El objetivo encontrando una descripción del problema planteado.</p> <p>3. Generación de la tabla para Entrenamiento</p> <p>Si $f \in H$ entonces es Aprendizaje considerado Si $f \notin H$ entonces no será considerado.</p> <p>Producto final de datos limpios, formalización de reglas de negocio. A sí se genera el proceso matemático para la creación del mejor AD.</p> <p>4. Inferencia del Árbol de decisión</p> <p>Usar conocimiento a priori para asegurar que $H \subset f$ en donde la inferencia se lo realiza mediante:</p> <p>Cálculo de la entropía</p> <p>Dados:- Un problema con dos clases (+ ó -) - S, el conjunto de ejemplos.</p> $Entropia(S) = \sum_i^k p_i \log_2 \left(\frac{1}{p_i} \right) \text{ con } k \text{ clases} \quad (1)$ <ul style="list-style-type: none"> • Por la ganancia que se obtendrá

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

$$Ganancia(S, A) = Entropia(S) - \sum_{v \in val. A} \frac{|S_v|}{|S|} Entropia(S_v) \quad (2)$$

5. Pruebas de datos sobre el AD.

Usamos medidas evaluación que son:

$$promedio\ positivo\ verdadero = \frac{TP}{TP + FN} \quad (3)$$

$$promedio\ positivo\ falso = \frac{FP}{FP + TN} \quad (4)$$

$$presición = \frac{TP}{TP + FP} \quad (5)$$

$$corrección = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

Dónde: TP: Verdaderos positivos; TN: Verdaderos negativos.

FP: Falsos positivos; FN: Falsos negativos

Fuente: (Morales, Morales, & Rizo, 2019)

En el sector de la sierra ecuatoriana la preocupación de muchas haciendas productoras de leche es tener una buena rentabilidad para lo cual era necesario analizar de forma manual los datos que se genera día a día buscando soluciones y respuestas a varios problemas que se presentan en el negocio sin embargo actualmente la aplicación de Inteligencia de Negocios hace que todo este proceso sea más sencillo facilitándole la vida al personal encargado de tomar decisiones que afecten a la producción.

Mediante el uso de la Analítica de Datos apoyada en la metodología M3S se busca aprovechar la información histórica de la hacienda, mediante el uso de determinados indicadores productivos y reproductivos generar información que sea relevante para el personal técnico, para que de esta forma puedan tomar decisiones que aporten de forma positiva en la hacienda.

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Aplicación M3S

A. Justificación

La producción lechera es una de las actividades agropecuarias que está utilizando tecnologías modernas en su gestión productiva y reproductiva. Una de las alternativas para reducir el número de vacas que no aportan a una ganadería de leche es el descarte en base a indicadores técnicos por esto es pertinente aportar con estudios basados en el uso de la metodología M3S la cual nos ayudará a conseguir un proceso adecuado para llevar a cabo el descarte de ganado.

Objetivo General

Utilizar la metodología M3S para descartar ganado, el cual será dado de baja de acuerdo a determinadas características que indiquen que no aportan en la eficiencia productiva y reproductiva dentro de la hacienda.

Objetivos específicos

- Validar la Metodología M3S de inteligencia de negocios, aplicándola en el sector pecuario del país.
- Aplicar árboles de decisión aplicando reglas del negocio que nos entregue información oportuna y verídica para generar nuevo conocimiento.
- Justificar nuestra decisión basados en el algoritmo ID3, el cual tendrá una gran fortaleza con el apoyo de veterinarios como arquitectos de información.
- Delimitar secciones de la base de datos adecuadas que nos permitan aplicar árboles de decisión que en definitiva muestren la menor incertidumbre posible.
- Aplicar todas las fases de M3S, en el caso Tabacundo-Ecuador, Hda. Sta Gertrudiz de la cual se obtendrá información que será de mucha utilidad para el crecimiento de la misma.

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

B. Planificación

- Estrategia: Se ubico una hacienda en la zona de Tabacundo con ganado Holstein.
- Infraestructura: Registros almacenados desde hace 20 años en una base Visual FoxPro.
Equipo técnico informático (Ingenieros UCE).
- Cronograma: Proyecto completo 4 semanas.

C. Análisis y requisitos del negocio

Se tomará en cuenta algunos reportes extraídos del Sistema GANASI (sistema transaccional que ayuda en la parte operativa diaria en una hacienda de ganadería de leche), los cuales presentan información importante acerca del hato, así como índices reproductivos y productivos los cuales servirán de apoyo para el desarrollo del proyecto, como son:

Resumen Productivo – Reproductivo

En la tabla 3 podemos obtener información de la situación del ganado a febrero del 2019 como el número de cabezas que se encuentran preñadas y vacías (no preñadas), dentro de la categoría de vacías y preñadas tenemos el promedio de días abiertos que es el periodo comprendido entre el parto y la siguiente preñez de la vaca, con un período ideal de 85-90 días, se tiene también el promedio de los días de lactancia únicamente para las preñadas.

Tabla 3

Análisis Reproductivo-Productivo

		HATO		PROMEDIO			
		ADULTO		PRODUCCIÓN			
		112		0,00			
VACIOS		SERVIDOS		SIN SERVICIO		PROMEDIO	
No.	%	No.	%	No.	%	DIAS	

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

PREÑADAS		CONCEPCIÓN		ABIERTOS		ABIERTOS	
No.	%			PROMEDIO	PROMEDIO	DÍAS	DÍAS
74	66,07	49	66,21	25	33,78	293,72	
38	33,92	2,32		275,1		382,42	

Fuente: Sistema GANASI(2019)

Reporte nacimientos y bajas

La tabla 4 resume el número de nacimientos dentro de un determinado periodo de tiempo, cuantos animales han nacido vivos, muertos, machos y hembras; la sección de bajas de ganado indica cuantos animales han salido, cuantos son machos y hembras; todos con sus respectivos porcentajes.

Tabla 4

Reporte Nacimientos y Bajas

NACIMIENTOS			PORCENTAJE BAJAS		
Vivos	26	92,86%	Descarte	11	73,33%
Muertos	2	7,14%	Obsequio	1	6,67%
Machos	14	51,85%	Venta	3	20,00%
Hembras	14	48,15%	Total, Bajas	15	100,00%
Total, Nacimientos	28	100,00%			

Fuente: Sistema GANASI (2019).

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Índices Reproductivos

En la tabla 5 se observa algunos parámetros que son utilizados para medir la eficiencia reproductiva dentro del hato, los cuales tienen una meta que debe llegar a cumplirse para asegurar que el hato se encuentra en buenas condiciones reproductivas.

Tabla 5

Índices Reproductivos

INDICE	META	REAL
	12,5 - 13	
1.- promedio intervalo partos	Meses	1-1-1 A-M-D
2.- promedio días de lactancia		273,88Días
3.- promedio de preñez postparto		275,1Días
3.- i. Por concepción	1,8 - 3	2,45
4.- rango de aborto	<15%	7,41%
5.- % de vacas 3 o más veces inseminadas	<20%	28,57%
6.- % de vacas no preñadas	10 - 15%	24,64%

Fuente: Sistema GANASI (2019)

Tabla 6

Distribución según días de lactancia

DIAS DE LECHE EN VACAS	PREÑADAS	VACIAS
0 y 100 días	0	11
100 y 200 días	0	17
200 y 300 días	2	11
300 y 400 días	10	12
Mas de 400 días	17	18
DIAS EN LECHE	501	294

Fuente: Sistema GANASI (2019)

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Tabla 7

Distribución según número de partos

No. PARTOS VACAS	TOTAL	PORCENTAJE
1	32	32,65%
2	21	21,43%
3	20	20,41%
4	15	15,31%
5	6	6,12%
6	4	4,08%
Mas de 6 partos	0	0,00%
TOTAL, ANIMALES	98	100,00%

Fuente: Sistema GANASI (2019)

Tabla 8

Proyección de crecimiento del hato

MES	Vacas Lechando	Vacas por Parir	Vacas por Parir	Vacas a Secar	Vacas de Venta	Total, Vacas
JUNIO	89	8	0	4	2	91
JULIO	91	6	3	1	2	97
AGOSTO	97	3	2	3	2	97
SEPTIEMBRE	97	2	0	2	2	95
OCTUBRE	95	3	0	2	2	94
NOVIEMBRE	94	4	0	0	2	96
DICIEMBRE	96	2	0	0	2	96

Fuente: Sistema GANASI (2019)

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Dentro de la revisión bibliográfica se desprende que datos se deben tomar para generar mayor conocimiento en una ganadería de los cuales se va a elegir algunos indicadores como se refleja en la siguiente tabla:

Tabla 9

Indicadores de Producción

% de descarte	% Vacas Abiertas en Rejo
% de vacas secas	% Vacas Servidas en Rejo
Litros de leche por vaca	% Vacas Preñadas en Rejo (>45%)
Litros por potrero	% Preñez en Rejo y Seco (>50%)
	% Preñez en Hato (>60%)

Fuente: (Ecuador, 2018)

D. Aplicación de la Metodología de Decisión dentro de M3S

ANÁLISIS DE DATOS

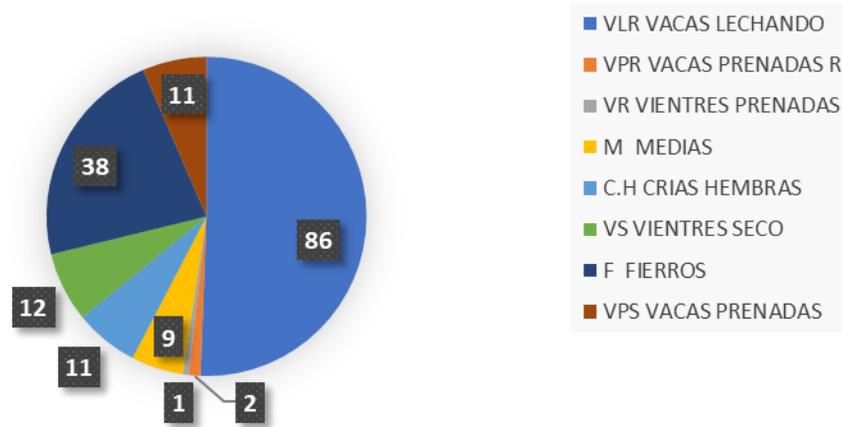
El caso propuesto revisa y contempla datos acerca del hato de la hacienda, correspondientes al periodo 2000 – 2017. El total de animales utilizados para el estudio es de 171, de los cuales nos dedicaremos solamente aquellos correspondientes al hato adulto como son Rejos y Secas (Hembras).

En gráfico 1, se cuenta con información acerca de la clasificación del hato según el tipo.

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Grafico 1

Hato según el tipo



Fuente: Autores, 2019.

HIPÓTESIS DE DATOS

Dentro de este caso de estudio se revisaron un gran número de atributos y en base al análisis realizado se identificaron aquellos que podrían generar mayor valor y a su vez tener posibilidad de dicotomía para la generación del árbol de decisión; por su importancia del estudio se han considerado exclusivamente animales que han alcanzado su edad adulta, cinco de ellos están a continuación están las siguientes reglas.

ABORTOS si:

≥ 2 en el caso que el número de abortos sea mayor o igual a 2.

PARTOS si:

>4 en el caso que el número de abortos sea mayor a 4.

PRODUCCION si:

≤ 3000 el caso que litros de leche producidos sean menores o iguales a 3000 litro.

DÍAS LACTANCIA si:

≥ 280 en el caso que los días de lactancia sean mayores o iguales a 280.

ENFERMEDAD si:

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

>= 3 en el caso que el número de enfermedades sea mayor o igual a 3.

GENERACIÓN DE LA TABLA DE ENTRENAMIENTO (TE)

Esta tabla que contiene los datos de mayor relevancia se genera en base a la información de expertos “veterinarios” que brindando su experiencia ayudan a la creación de las reglas del negocio para focalizar en este caso particular de optimizar el número de vacas que no aportan reproductivamente a la ganadería lechera.

Descarte = Si

- Edad < 7, Parto <=4, Aborto >=2, Enfermedad >=3, Días Lactancia >=280 o Días Lactancia <280, Producción <3000.
- Edad < 7, Parto >4, Aborto >=2, Enfermedad >=3, Días Lactancia >=280 o Días Lactancia <280, Producción <3000
 - SI en el campo Decisión.
 - NO, en todas las otras combinaciones

Tabla 9

Reglas de Negocio en base a indicadores relevantes

Edad	Partos	Abortos	Clínico	Días Lac	Producción	Descarte	
< 7	<= 4	>= 2	>= 3	>=280	>=3000	NO	
					<3000	SI	
				<280	>=3000	NO	
					<3000	SI	
				>=280	>=3000	NO	
					<3000	NO	
		< 2	>= 3	< 3	>=280	>=3000	NO
					<3000	NO	
	<280				>=3000	NO	
					<3000	NO	
	>=280				>=3000	NO	
					<3000	NO	

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

				>=280	>=3000	NO	
					<3000	SI	
			<280	>=3000	NO		
	> 4	>= 2	>= 3		<3000	SI	
				>=280	>=3000	NO	
					<3000	NO	
			<280	>=3000	NO		
			< 3		<3000	SI	
				>=280	>=3000	SI	
					<3000	SI	
			<280	>=3000	SI		
	>=7	<= 4	>= 2	>= 3		<3000	SI
				>=280	>=3000	NO	
					<3000	NO	
			<280	>=3000	NO		
			< 3		<3000	SI	

Fuente: Autores, 2019.

En el momento de aplicarlas condiciones propuestas por los veterinarios, las vacas que posean la característica de SI en la decisión, serán los seleccionados para ser eliminadas de la hacienda y en donde al momento ya se contaría con la TE que se presenta en la tabla 10, la cual contiene toda la información necesaria, para poder continuar al siguiente paso.

Tabla 10
Tabla de Entrenamiento

Arete	Aborto	Parto	Clínico	Litros Producidos	Días Lactancia	Edad	Descarte
177	<2	<=4	>=3	<3000	>=280	>=7	SI
183	<2	>4	>=3	>=3000	>=280	>=7	SI
38	<2	<=4	>=3	<3000	>=280	>=7	SI
185	<2	<=4	>=3	>=3000	<280	>=7	NO

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

188	<2	<=4	<3	<3000	>=280	>=7	SI
181	<2	<=4	>=3	>=3000	<280	>=7	NO
187	<2	<=4	>=3	<3000	>=280	>=7	SI
192	<2	<=4	>=3	>=3000	<280	>=7	NO
198	<2	<=4	>=3	>=3000	<280	>=7	NO
201	<2	<=4	>=3	>=3000	<280	>=7	NO
209	<2	<=4	>=3	<3000	>=280	>=7	SI
219	<2	<=4	>=3	<3000	>=280	>=7	SI
265	<2	<=4	>=3	>=3000	>=280	<7	NO

Fuente: Autores, 2019

Inferencia del árbol de decisión (USO ID3):

En el momento que se cuenta con la TE de datos se procede a calcular la entropía que nos permitirá decidir entre todos cuál de los atributos dará mayor ganancia de información tras clasificar los vectores de entrada. Esta elección la haremos basada en los pasos que se explican en la Tabla 2, los que se aplicarán con los atributos Edad, Partos, Abortos, Enfermedad, días de lactancia y Producción.

ITERACIÓN 1:

Primeramente, se calcula la entropía total (Nivel 1) de todos los atributos y en este caso se obtiene un valor de 0,27; dado esto en forma recursiva se obtiene como se muestra en la tabla 1, el siguiente resumen de datos:

Entropía Total (Nivel 1): 0,70

Tabla 11

Tabla de Cálculos 1er. Nivel Ganadería

ATRIBUTO	H ()	GANANCIA INFORMACIÓN
Edad	0,38	0,38024
Partos	0,63	0,08740
Abortos	0,69	0.02120

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Clínico	0,70	0,01933
Días lactancia	0.62	0.0956
Producción	0.24	0.4760

Fuente: Autores, 2019

De este primer barrido con todos los atributos, observamos que obtenemos la mejor ganancia de información con el nodo Producción, el mismo que quedará en la parte inicial del árbol de la siguiente forma:



ITERACIÓN 2:

Una vez con el primer nivel, repetimos el procedimiento para la rama ≥ 3000 , pero sin tomar en cuenta el nodo Producción, y se tiene lo que se muestra en la tabla 5:

Entropía Total (Nivel 2): 0,45

Tabla 12

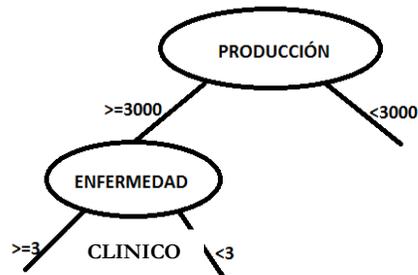
Tabla de Cálculos 2do. Nivel Ganadería

Edad	0,29	0,1546
Partos	0,16	0,2860
Abortos	0,45	0.0000
Clínico	0,13	0,3196
Días lactancia	0.36	0.0837

Fuente: Autores, 2019

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

En el segundo barrido con todos los atributos a excepción de Producción, se define que la mejor ganancia de información se tiene con el nodo Producción- Enfermedad para la rama ≥ 3 , el mismo ingresa al segundo nivel del árbol de la siguiente forma:



ITERACIÓN 3:

A continuación, reiteramos el mismo procedimiento en forma recursiva para la rama < 3000 , pero esta vez sin tomar en cuenta los dos nodos seleccionados, como se obtiene en la tabla 6, el siguiente resumen de datos:

Entropía Total (Nivel 2): 0,45

Tabla 13

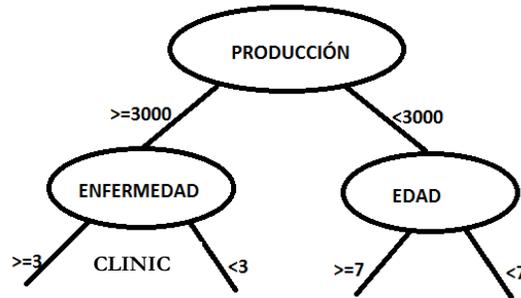
Tabla de Cálculos 2do. Nivel Ganadería

ATRIBUTO	H ()	GANANCIA INFORMACIÓN
Edad	0,10	0,7247
Partos	0,74	0,0884
Abortos	0,80	0,0285
Clínico	0,79	0,0298
Días lactancia	0,72	0.1130

Fuente: Autores, 2019

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Haciendo referencia al resultado obtenido, se definiría que el segundo nivel para la rama < 3 se forma por el nodo que correspondería a Edad.



ITERACIÓN 4:

En este instante se cuenta con dos niveles, se vuelve a repetir el proceso en forma recursiva pero ahora para la rama ≥ 3000 y ≥ 3 , y en esta ocasión sin contar con el nodo Producción y Enfermedad, y se obtendrá tal como se muestra en la tabla, el siguiente resumen de datos:

Entropía Total (Nivel 3): 0,49

Tabla 14

Tabla de Cálculos 3er. Nivel Ganadería

ATRIBUTO	H ()	GANANCIA INFORMACIÓN
Edad	0,33	0,1552
Partos	0,18	0,3039
Abortos	0,49	0,0000
Días lactancia	0,38	0,1055

Fuente: Autores, 2019

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Con el respectivo resultado ya se podría decir que el tercer nivel para la rama ≥ 3 está formado por el nodo que hace referencia a Partos.

ITERACIÓN 5:

Al momento se cuenta con dos niveles, reiteramos el mismo proceso de manera recursiva para la rama < 3000 y > 7 , en esta ocasión sin tomar en cuenta el nodo Producción y Edad, y se obtendrá tal y como se muestra en la tabla, el siguiente resumen de datos:

Entropía Total (Nivel 3): 0,13

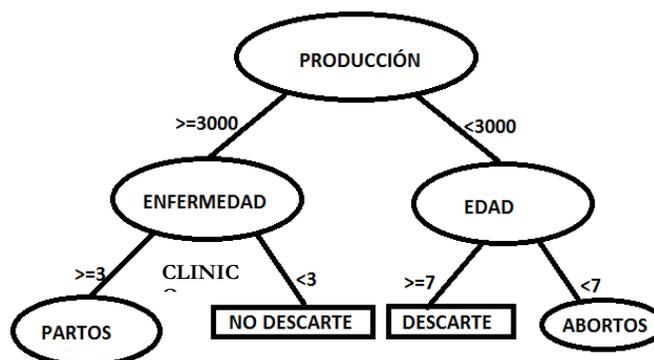
Tabla 15

Tabla de Cálculos 3er. Nivel Ganadería

ATRIBUTO	H ()	GANANCIA INFORMACIÓN
Clínico	0,13	0,0066
Partos	0,03	0,1096
Abortos	0,00	0,1371
Días lactancia	0,03	0,1099

Fuente: Autores, 2019

El resultado obtenido definiría que el tercer nivel para la rama < 7 estaría formado por el nodo que hace referencia a Abortos.



Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

ITERACIÓN 6:

En este instante tenemos tres niveles, reiteramos el proceso en forma recursiva para la rama ≥ 3000 , ≥ 3 y >4 , pero en esta ocasión sin tomar en cuenta los nodos Producción, Clínico y Partos, y lo que obtendremos será como se muestra en la tabla 7, el respectivo resumen de datos:

Entropía Total (Nivel 4): 0,99

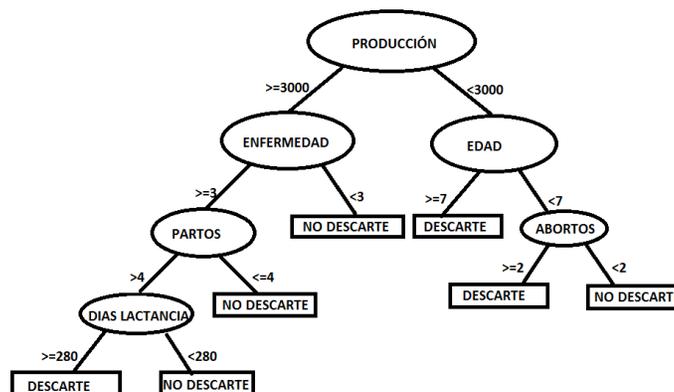
Tabla 16

Tabla de Cálculos 4to. Nivel Ganadería

ATRIBUTO	H ()	GANANCIA INFORMACIÓN
Edad	0,99	0,0000
Abortos	0,99	0,0000
Días lactancia	0,00	0,9852

Fuente: Autores, 2019

Con el siguiente resultado se obtendrá que el cuarto nivel para la rama >4 estaría formado por el nodo que hace referencia a Días de Lactancia. De esta manera la estructura total del árbol sería de la siguiente forma:



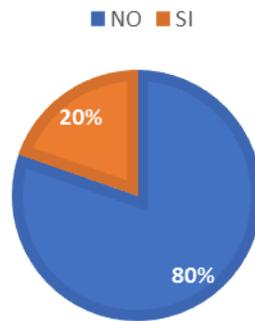
Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Test de Datos

Cuando ya se ha establecido la estructura total del árbol de decisión se procederá a realizar las respectivas pruebas sobre el mismo, y se llega a concluir que, de un total de 112 registros analizados, 22 de los resultados que representan al 20% de la muestra tomada deberá descartarse en el ganado lechero, para un mejor entendimiento se representa los datos mencionados en la siguiente gráfica.

Grafico 2

Test de datos



DECISIÓN	NÚMERO DE REGISTROS
NO	90
SI	22

Fuente: Autores,2019

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

CONCLUSIONES

M3S es una metodología que tiene por objetivo apoyar la construcción de un proyecto basado en BI simplificando tiempo en un 40% y costos en un 50% para llevar a cabo el mismo.

M3S puede ser probada en pequeñas y medianas empresas de cualquier área sea productivo o comercial.

Mediante la implementación de una metodología de decisión se puede mejorar la explotación de información y obtener resultados más confiables, llegando a tomar decisiones más oportunas y eficaces en el tiempo.

FINANCIAMIENTO

No monetarios

AGRADECIMIENTO

A la Universidad Central del Ecuador por el apoyo técnico y científico realizado a la presente investigación.

REFERENCIAS CONSULTADAS

ASINFO, A. I. (s.f.). Sistema GANASI. [GANASI system]. Quito, Ecuador.

Balarezo, L., Montenegro, F., & Mora, R. (2015). *Obtención de Parámetros Productivos, Reproductivos y Nutricionales en explotaciones Lecheras del Carchi*. [Obtaining Productive, Reproductive and Nutritional Parameters in Carchi Dairy farms].

Calzada, L., & Abreu, J. (09 de 2009). El Impacto de las herramientas de Inteligencia de Negocios. [The Impact of Business Intelligence tools]. *DAENA (International Journal of Good Conscience)*, 52. Obtenido de [http://www.spentamexico.org/v4-n2/4\(2\)/02016-52.pdf](http://www.spentamexico.org/v4-n2/4(2)/02016-52.pdf)

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

- Cedeño, D., & Vargas, B. (2004). *Efecto de la raza y el manejo sobre la vida Productiva del Bovino Lechero en Costa Rica*. [Breed and management effect on the Productive life of the Dairy Cattle in Costa Rica].
- Chamba, H., Armijos, R., & Vidal, P. (2017). *Estudio de los parámetros reproductivos de hatos ganaderos de la parroquia Valladolid- Palanda- Zamora- Chinchipe*. [Study of the reproductive parameters of cattle herds of the Valladolid- Palanda- Zamora and Chinchipe parish].
- Díaz, J. C. (2010). *Introduction Business Intelligence* [Introducción a la inteligencia de negocios]. (1era ed., Vol. Primera Edición). (E. C. SA, Ed.) Barcelona, ESPAÑA: UOC. Recuperado el 16 de 07 de 2016, de www.editorialuoc.com
- Ecuador, A. H. (2018). *Asociación Holstein Friesian del Ecuador*. [Holstein Friesian Association of Ecuador]. Obtenido de <http://www.holsteinecuador.com/>
- Infocarne, (2010). [Infocarne, (2010)]. Obtenido de <http://www.infocarne.com>
- ainudeen, M., & Hafez, E. (2000). Ciclos reproductivos y productivos en bovinos y búfalos. [Reproductive and productive cycles in cattle and buffalo]. En E. Hafez, & B. Hafez, *Reproducción e Inseminación artificial en animales* (págs. 163-167). México: Mc Graw Hill.
- Morales, S., Morales, M., & Rizo, R. (2019). *Entropía en búsqueda de atributos puros en Sistemas de Información y Redes Sociales. M3S Metodología basada en ID3. 2*. [Entropy in search of pure attributes in Information Systems and Social Networks. M3S, methodology based on ID3. 2].
- Peng, W., Chen, J., & Zhou. (2012). *An Implementation of ID3. TREE LEARNING ALGORITHM*. Sidney, Australia: University of New South Wales, School of Computer Science & Engineering. [Una implementación de ID3. ALGORITMO DE APRENDIZAJE BASADO EN ÁRBOLES. Sydney, Australia: Universidad de Nueva Gales del Sur, Escuela de Informática e Ingeniería]. Obtenido de <https://n9.cl/l61qp>
- Quinlan, J. (1986). *Induction of decision trees. Machine learning*. [Inducción de árboles de decisión. Aprendizaje automático]. Kluwer Academic Publishers. Obtenido de <https://doi.org/10.1007/BF00116251>

Santiago Morales-Cardoso; Mario Morales-Morales; Alicia Andrade-Bazurto; Laura Cevallos-Black

Vilchez, M. &. (2011). *Desarrollo de un modelo de inteligencia de negocios usando Data Mining para optimizar la toma de decisiones en los procesos de ventas en la empresa San Roque SA. Universidad Señor de Sipán. Facultad de ingeniería.* [Development of a business intelligence model using Data Mining to optimize sales processes decision-making in San Roque company Ltd. Señor de Sipán University. Faculty of Engineering]. Escuela Profesional Ingeniería de sistemas. (E. d. Sistemas, Ed.) Chiclayo, Perú: Universidad Señor de Sipán. Obtenido de <http://repositorio.uss.edu.pe/handle/us/387>

©2020 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).