



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

### Minería de Datos enfocada al Embarazo de Alto Riesgo

Sandra Munguía-Gutiérrez –Sergio Eduardo Ledesma-Orozco – Segovia-Domínguez Ignacio

#### Resumen

Se realizó un estudio para comprender los diferentes síntomas de un embarazo de riesgo y sus explicaciones médicas; así como la comprensión de selección de los factores de riesgo realizada por los doctores durante el periodo de gestación.

La información utilizada para realizar este estudio proviene de cuadros clínicos llamados Historias Clínicas Perinatales, comúnmente localizados en los archivos de cualquier hospital. La información recabada se analizó mediante diversas técnicas de Minería de Datos. En el presente trabajo se muestran los resultados obtenidos al emplear los métodos: 1) Mapas de Auto-organización y 2) K-means. Al término de las pruebas, ambos algoritmos presentaron un comportamiento similar. Sin embargo, se observó que la agrupación generada por los Mapas de Auto-organización (SOM) se asemeja más al diagnóstico proporcionado por el médico.

#### Abstract

A study was conducted to understand the different symptoms of pregnancy risk and their medical explanations. Also, this study intends to understand the selection process of risk factors during the gestation period.

The information used for this study comes from clinical reports named Perinatal Medical Records, which are usually located in the archives of any hospital. The collected information is analyzed by applying techniques of data mining. This paper presents several results obtained by utilizing two main methods: 1) self-organizing maps and 2) k-means. Upon completion of tests, both algorithms presented similar behavior. However, it is worthwhile to mention that the group found by the self-organizing maps (SOM) is closer to the diagnosis provided by the doctor.



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

Palabras Clave **Historia Clínica Perinatal, SOM, K-means**

### 1. Introducción

La muerte materna es el evento más dramático que puede sufrir un amplio sector de las mujeres mexicanas durante el embarazo. En este contexto, la inteligencia artificial ha tenido valiosas aportaciones dentro de la medicina; como es el caso de los sistemas de diagnóstico médico basados en la experiencia de casos previos, los cuales permiten realizar un diagnóstico médico interactivo a través de la Internet [1]. Un sistema de esta naturaleza puede ser usado por los mismos pacientes para diagnosticarse sin tener la necesidad de ir con el médico. Otra herramienta empleada en aplicaciones médicas son las Redes Neuronales; por ejemplo para el caso en diagnósticos de cáncer [2] y para tratamientos en infartos miocárdicos [3].

La minería de datos es una técnica de Inteligencia Artificial usada para el descubrimiento del conocimiento en grupos grandes. La manera en la que estas herramientas extraen conocimiento a partir de los datos puede ser empleada en diversas áreas de la medicina. En este sentido, Maojo [4] publicó un artículo sobre la utilización de tecnología de minería de datos para el tratamiento de datos biomédicos, donde menciona el uso de la posibilidad de la integración de datos clínicos y genómicos. Es tal el rango de posibles aplicaciones de la minería de datos que actualmente es empleado para el análisis de exploraciones de datos del sueño (Exploratory analysis of sleep data) [5]; en este contexto, se desarrollan aplicaciones usando reconocimiento estadístico de patrones para analizar las irregularidades del sueño en los seres humanos.

El objetivo fundamental de este trabajo de investigación es la eficaz atención en la prevención de la muerte materna en el periodo de gestación en el que se



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

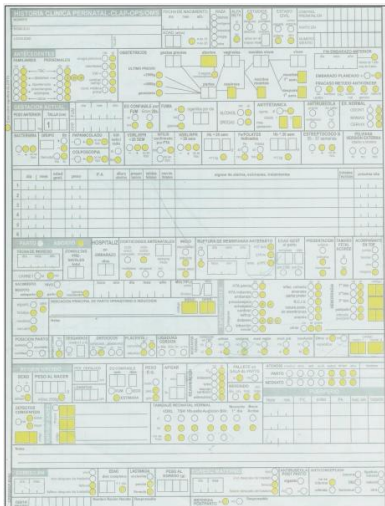
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

encuentra la mujer, a través de un análisis profundo del comportamiento de la población mediante técnicas de reconocimiento de patrones/minería de datos.

### 2. Metodología

La población está representada por las pacientes embarazadas que acudieron al Centro de Salud de Valle de Santiago a solicitar servicio médico para el control de su embarazo. Debe señalarse que los factores de riesgo son simplemente expresiones de probabilidad, esto es, no son términos absolutos, ni diagnósticos. Los factores de riesgo son características individuales que influyen sobre la probabilidad de que un individuo pueda sufrir una enfermedad particular o un grupo de enfermedades en un periodo de tiempo definido [6]. Los factores de riesgo son los siguientes: nivel socioeconómico, edad, anemia, tensión arterial, proteinuria, obesidad. Se consideraron más variables que no pertenecían a los factores de riesgo: edad gestacional, fecha del último parto, número de embarazos, ruptura de la membrana, líquido amniótico, sufrimiento fetal, trimestre infecciones, antecedentes de hipertensión, placenta previa y número de cesáreas. Los casos clínicos son expedientes de mujeres embarazadas que acuden a la institución con la finalidad de la revisión prenatal durante el periodo de embarazo y hasta la culminación del parto; tal y como se muestra en la figura 1 el formato que se llena en el centro de Salud.



**Figura 1 “Historia Clínica Perinatal-CLAPOPS/OMS”**

#### 2.1 Aplicación de la Minería de Datos

##### 2.1.1 Preparación de datos

Teniendo un total de 215 casos, lo primero es buscar los datos incompletos, esto es consultas prenatales y no culminaron con el parto en la



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

institución o viceversa, por lo tanto se eliminaron se procedió a normalizar los datos en formato numérico con un intervalo de 1 a -1.

### 2.1.2 Aplicación de las Técnicas.

#### 2.1.2.1 Self Organizing Map (SOM)

Es un tipo de red neuronal artificial entrenada utilizando el aprendizaje no supervisado para producir una representación de baja dimensión (normalmente de dos dimensiones), discretizada en el espacio de entrada de las muestras de entrenamiento, llamado mapa [7]. El objetivo de este aprendizaje es categorizar los datos que se introducen en la red [8].

#### 2.1.2.2 K-Means (K-medias).

En las estadísticas y de aprendizaje automático, k-means clustering es un método de análisis. Tiene por objeto las particiones  $n$  en grupos  $k$  en el que cada clúster pertenece a la agrupación más cercana con la media.

### 2.2 Selección de variables en las pruebas

Se analizaron solamente tres variables, siendo estas la edad de la paciente, la edad de la gestación y el líquido amniótico.

### 2.3 Normalización de las variables

Con el fin de facilitar el manejo de los datos obtenidos de las variables seleccionadas; se determinó expresar estos datos en valores relativos de por unidad (pu), que se define como la división del valor obtenido entre un valor base seleccionado.

#### 2.3.1 Variable Edad

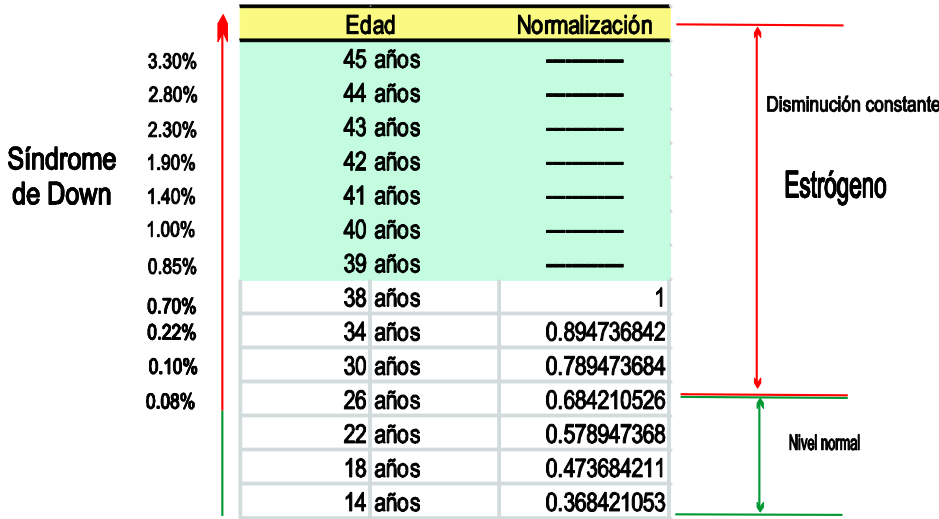
En la figura 2, se indican los datos recolectados asociados con su factor de normalización, así como también el nivel de estrógeno 1, que se mantiene constante hasta los 25 años, a partir de estos empieza a disminuir



**“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”**  
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

constantemente. Durante el embarazo la producción de estrógenos se incrementa, debido a su importancia para el mantenimiento de la gestación [9].

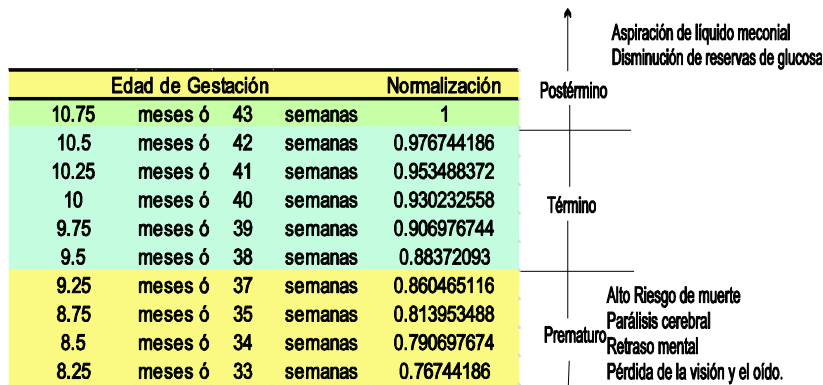


**Figura 2 Variable Edad normalizada**

**2.3.2 Variable Edad de Gestación**

En la figura 3 se muestra los datos normalizados de la variable edad de gestación. El recién nacido post maduro es el producto de una

gestación pos término o gestación prolongada que presenta signos clínicos de post madurez [11].



**Figura 3 Variable Edad de gestación**

**2.3.3 Variable Líquido Amniótico**

En la siguiente figura 4 se puede observar las características del líquido amniótico

normalizadas junto con el riesgo que pueden presentar.

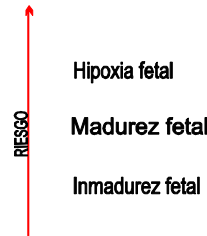


**“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”**  
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

Cuando el médico observa la presencia de meconio durante el parto, puede tratar a la madre con una amnio-infusión para intentar evitar estas complicaciones en el bebé [12].

Líquido Amniótico	Normalización
aumentado	1
normal	0.75
disminuido	0.5
transparente	0.25

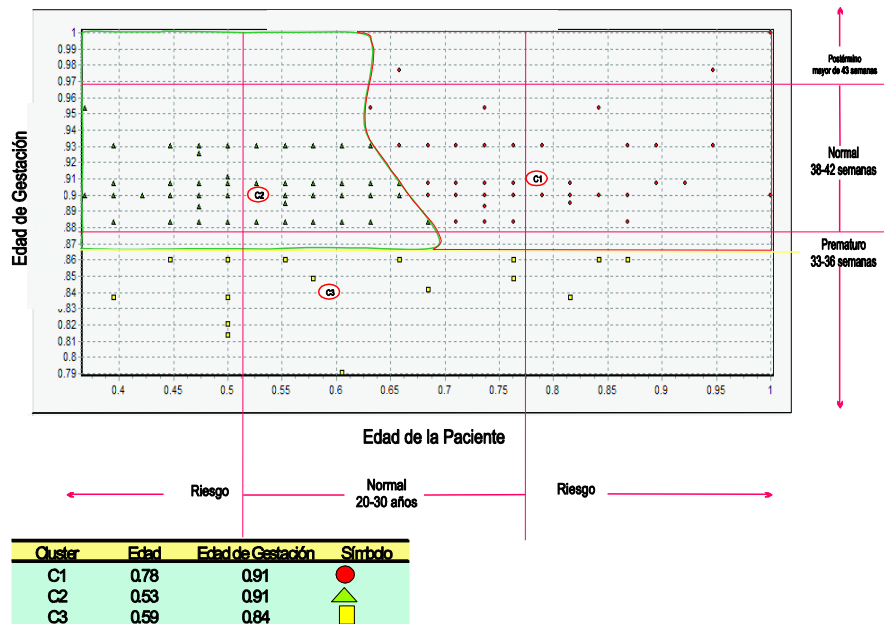


**Figura 4 Variable**  
**Líquido Amniótico**

**3. Resultados**

**3.1 Prueba Edad de la Paciente VS Edad de Gestación.**

En esta prueba se podrá observar la distribución de la población dentro de la gráfica, referida a la comparación entre los factores de las variables de la edad de la paciente contra la edad de gestación, es decir cómo evoluciona el feto de acuerdo a la edad de la madre. El resultado de k-means se muestra en la figura 5.



**Figura 5 Gráfica**  
**Resultado K-means**

Se puede apreciar que son tres clústeres los que se crean. El clúster denominado con la etiqueta C1 agrupa las pacientes con edades

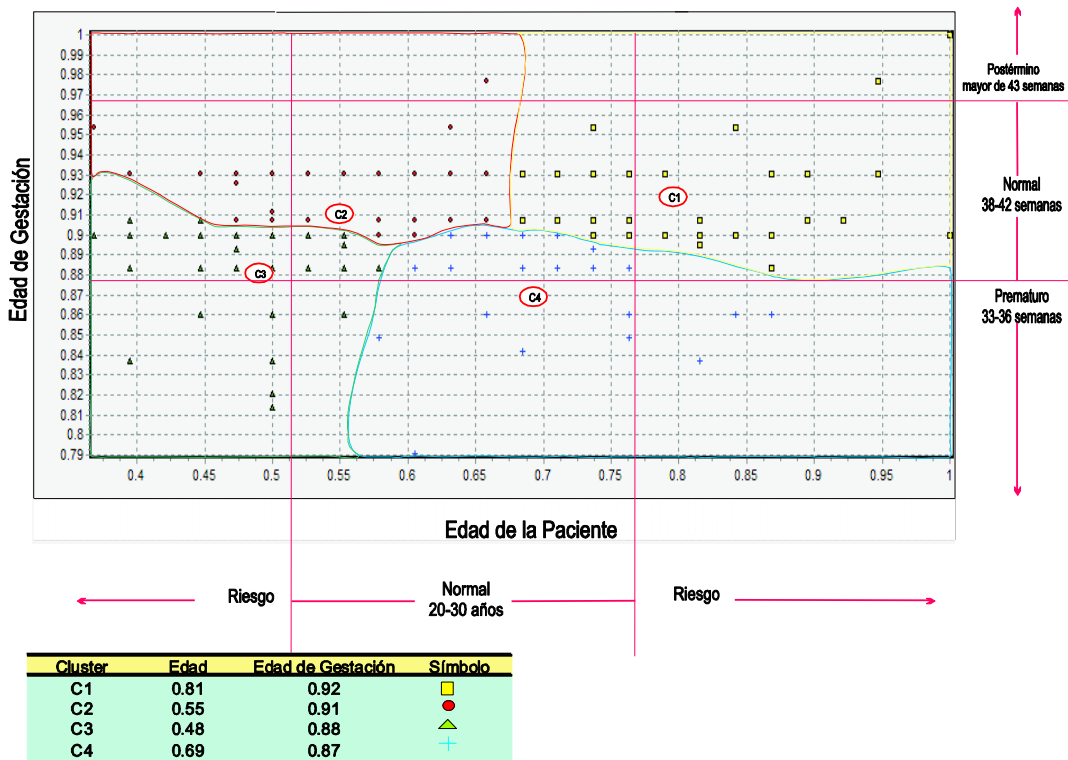
aproximadamente de 27 a 38 años es decir, su límite inferior es un poco más de la



**“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”**  
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

mitad del rango normal en la edad de la paciente y va incrementando. El clúster denominado con la etiqueta C2 agrupa las pacientes lo inverso de C1, su límite superior es la mitad del rango normal en la edad. Entre C1 y C2 se aprecia que agrupan los recién nacidos normales y pos términos. El clúster C3 agrupa a las pacientes de todas las edades con los recién nacidos prematuros. El resultado de SOM se muestra en la siguiente figura 6.



**Figura 6**  
**Gráfica**  
**Resultado**  
**de SOM**  
Se generan cuatro clústeres, se observa que el clúster C1 abarca las pacientes con edades de 27 años

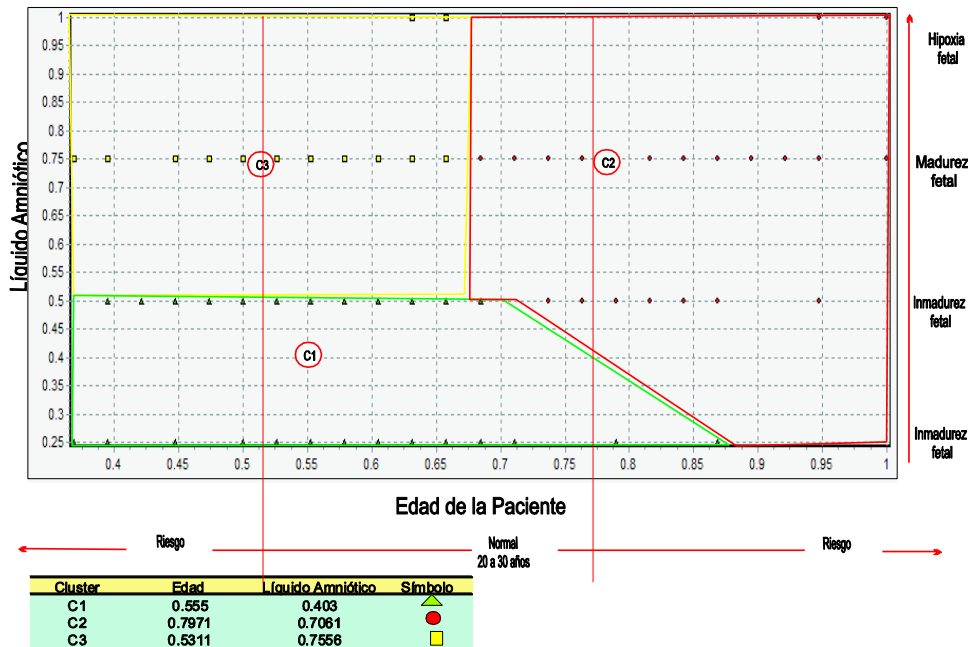
en adelante. Mientras que el clúster C2 abarca inversamente el clúster C1 en la variable edad, es decir su límite superior es de los 27 años y se va incrementando.

**“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”**  
Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

### 3.2 Prueba Líquido Amniótico vs Edad de la paciente.

Esta prueba está compuesta por las características que presenta el líquido amniótico durante el parto y edad de la paciente. El resultado de k-means se muestra en la figura 7. Se observa que se generaron tres clústeres. El clúster C1 agrupa las mujeres el rango de 14 a 33 años de edad, con el riesgo de inmadurez fetal que puede presentar su recién nacido de acuerdo a las características del líquido amniótico.



**Figura 7.**  
**Resultado del k-means.**

El clúster C2 abarca las pacientes con edad de 25 a 38 años de edad, teniendo el riesgo que sus recién nacidos puedan tener el riesgo de hipoxia fetal e inmadurez fetal. El clúster C3 agrupa a las mujeres de 14 a 33 años, con el riesgo que su recién nacido pueda presentar inmadurez fetal.

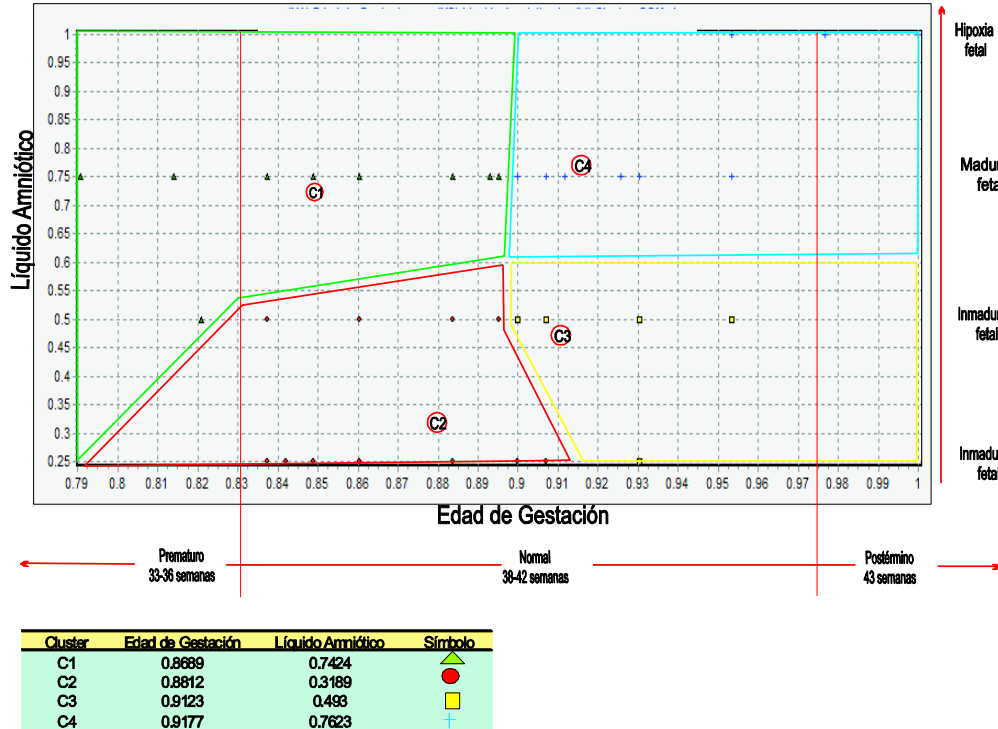
En la figura 8 se observa el resultado de SOM en esta prueba.

Se observan cuatro clústeres, el clúster C1 agrupa parte de las mujeres jóvenes hasta los 25 años y sus recién nacidos pueden presentar madurez fetal e hipoxia fetal por las características del líquido amniótico.





**“CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”**  
 Multidisciplinario  
 21 y 22 de abril de 2016, Cortazar, Guanajuato, México



**Figura 8**  
**Resultado de SOM**  
 El clúster C2 agrupa a las pacientes de 25 a 38 de edad de igual modo sus recién nacidos pueden presentar

madurez fetal e hipoxia fetal. El clúster C3 abarca a las mujeres con jóvenes hasta los 26 años con y sus recién nacidos pueden presentar inmadurez fetal por las características del líquido amniótico. El clúster C4 agrupa a las mujeres de 26 a 38 años y sus recién nacidos pueden presentar inmadurez fetal.

**4. Conclusiones**

En el desarrollo de las pruebas se pudo constatar que SOM genera cuatro clústeres y K-means genera solo tres, esto es indicativo de la eficacia de SOM para poder agrupar, ya que son menos datos que le corresponden a cada clúster y la distancia entre los clústeres es menor.

SOM tiene como fundamento buscar la preservación de la unión de sus clústeres y K-means dependiendo del número de datos que se van a clasificar coloca sus clústeres. Es decir, SOM actúa de manera sistemática y K-means solo lo hace de



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

manera automática y calculando tan solo la cantidad de datos que se quieren clasificar. Por lo tanto, se concluye que la edad de la paciente es importante porque varía la edad de la gestación y está a su vez varía las características del líquido amniótico; es decir la edad de la paciente cambia el estado en que se encuentre el recién nacido. Asimismo es importante la opinión médica ya que no se pretende sustituir los conocimientos médicos, solo determinar si la mujer cuenta con una edad dentro del rango de 24 a 30 años, el producto contará con las características dentro de lo aceptable, pero si la edad de la mujer está fuera de este rango el producto tiene la posibilidad de estar en niveles altos de riesgo y de salud.

### 5. Bibliografía

- [1] Manickam, 2000 S. Manickam, S.S.R. Abidi (2000): Unsupervised case classification and retrieval using Kohonen self-organising feature maps in a case base reasoning system, Proc. Of the IEEE TENCON 2000, Kuala Lumpur, 2000, p.24-31.
- [2] S.A Karkanis, G.D Magoulas, D.K Iakovidis, D.A Karras, and D.E Maroulis, Evaluation of Textural Feature Extraction Schemes for Neural Network-based Interpretation of Regions in Medical Images, in Proc. ICIP 2001, Thessaloniki, Greece, pp. 281-284, 2001
- [3] Heden B, Ohlsson M, Edenbrandt L, Rittner R, Pahlm O, Peterson C. Artificial neural networks for recognition of electrocardiographic lead reversal. Am J Cardiol. 1995; 75:929-33.
- [4] Maojo, V., F. Martín-Sánchez, J. Crespo, and H. Billhardt. Theory, Abstraction and Design in Medical Informatics. Methods of Information in Medicine 2002b, 41: 44-50.
- [5] ANALYSIS OF SLEEP DATA Project Members Faculty: Carolina Ruiz, Sergio A. Alvarez (Boston College), Majaz Moonis (University of Massachusetts Medical School). (Graduate Student: Parameshvyas Laxminarayan).
- [6] González-Merlo, Jesús (2006). “Capítulo 38: Parto pretérmino”, Obstetricia, 5ta edición (en español), Elsevier España. ISBN 8445816101.
- [7] “Introducción a la SOM de Kohonen Teuvo”. SOM Toolbox. Retrieved el 2006-06-18.
- [8] Kohonen, T. Honkela, T. (2007). “Red de Kohonen”.
- [9] Paul S. Cooke; Afia Naaz (2004). “Role of Estrogens in Adipocyte Development and Function”. Experimental Biology and Medicine (229): pp. 1127-1135.



## “CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2016”

Multidisciplinario

21 y 22 de abril de 2016, Cortazar, Guanajuato, México

[10] Botero Uribe, Jaime; Alfonso Júbiz Hazbún y Guillermo Henao (2004). “Capítulo 38: Parto pretérmino”, Obstetricia y Ginecología, 7ma edición (en español), Corporación para Investigaciones Biológicas, pp. 349. ISBN 9583358339.

[11] Manual Merck. Postmature Infant. [www.merck.com](http://www.merck.com), section 19 ch.260

[12] Ross, M.G., Brace, R.A. y participantes del taller del NIH. National Institute of Child Health and Development Conference Summary: Amniotic Fluid Biology - Basic and Clinical Aspects. Journal of Maternal-Fetal Medicine, volumen 10, febrero de 2001, Págs. 2-19.