

## Minería de datos: una aplicación para determinar cuáles factores socio-económicos influyen en el embarazo adolescente

Martín A. Quezada, Alejandro Tobón-Rivera y Omar D. Castrillón-Gómez

Universidad Nacional de Colombia, Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Industrial, Campus La Nubia Bloque Q piso 2, Manizales - Colombia. (correo-e: [mquezada@unal.edu.co](mailto:mquezada@unal.edu.co), [altobonr@unal.edu.co](mailto:altobonr@unal.edu.co), [odcastrillong@unal.edu.co](mailto:odcastrillong@unal.edu.co))

*Recibido Abr. 29, 2020; Aceptado Jun. 26, 2020; Versión final Jul. 20, 2020, Publicado Dic. 2020*

---

### Resumen

El objetivo de este trabajo es establecer los factores socio-económicos más influyentes en el embarazo adolescente por medio de técnicas de minería de datos. La muestra consiste de 700 adolescentes. Se analiza una variable dependiente (tiene hijos) y 11 variables independientes. Se seleccionaron estadísticamente aquellas variables más influyentes que fueron analizadas con el algoritmo J48 desde la plataforma Weka. Los resultados muestran que, con una efectividad superior al 97%, los factores socio-económicos más influyentes en el embarazo de las adolescentes son: tiene pareja actualmente, entidad promotora de salud (EPS) ha instruido métodos anticonceptivos, estudia actualmente, empezó a tener relaciones a los quince años o más, planifica cuando tiene relaciones y tipo de educación. Se concluye que esta investigación permite transformar una metodología tradicionalmente subjetiva en una metodología objetiva, basada en técnicas inteligentes, lo cual permitirán a las instituciones públicas redefinir las políticas para el control del embarazo en adolescentes.

*Palabras clave:* embarazo adolescente; minería de datos; Bayesiano; técnicas inteligentes

## Data mining: an application to determine what socio-economic factors influence teen pregnancy

### Abstract:

The objective of this research study is to determine the most influential socio-economic factors in teenage pregnancy by using data mining techniques. The sample consists of 700 teenagers. A dependent variable (to have children) and 11 independent variables are examined. The most influential variables were selected statistically and were analyzed by using the algorithm J48 from the platform Weka. The results show with over a 97% accuracy that the most influential factors in teenage pregnancy are: to have a partner, to be instructed on contraceptive methods by a health insurance company, to be studying, to have become sexually active at age 15 or older, to be using contraceptive methods, and the level of formal education. It is concluded that the present study transformed a subjective methodology into an objective methodology, based on intelligence techniques, which will allow public institutions to redefine their policies to better manage pregnancy in teenagers.

*Keywords:* teenage pregnancy; data mining; Bayesian; intelligence techniques

## INTRODUCCIÓN

Según la Sociedad Colombiana de Pediatría, el embarazo en la adolescencia, se ha convertido en un problema de salud pública creciente e importante, hasta el punto que se ha pedido asumir su prevención como una política de Estado. Este problema, tiene marcadas consecuencias de tipo social y económico, las cuales aumentan la imposibilidad de que las personas más pobres puedan abandonar lastradas como: Pobreza, mala de calidad de vida, mala educación, bajos logros escolares, necesidades básicas insatisfechas y la falta de oportunidades, entre otro gran número de consecuencias negativas no solo para las adolescentes sino también para sus familias como se propone en Quintero y Rojas, (2015). Autores como Flórez y Soto, (2006) muestran que la salud física, mental y social es afectada por el embarazo en la adolescencia constituyéndose este en un factor importante de mortalidad materno infantil. Este problema, puede ser consecuencia de varios aspectos como: inequidad social, ausencia de oportunidades, entre una gran variedad de aspectos que afectan las personas más vulnerables de una sociedad como lo escriben Reyes y González, (2014) y sus valores éticos definidos en Osuna y Luna, (2011) y Royo et al., (2015).

En Girma y Paton, (2015) se propone la educación como la mejor forma de prevenir el embarazo en adolescentes, acompañada de una serie de políticas como promoción de la anticoncepción reversible, factores ligados a mejoras en rendimiento educativo, disminución del uso del alcohol y drogas. Por su parte en Sorhaindo et al., (2016) se muestra un estudio de prevención de embarazo adolescente, identificando las mujeres jóvenes que están en riesgo de embarazo. En Hadley et al., (2016) se proponen estrategias para evitar el embarazo en adolescentes, dado que este es problema que genera gran desigualdad, afectando el bienestar y salud de las personas. Con este programa se logra una reducción hasta del 51% en la tasa de embarazos adolescentes. Por su parte en Brindis C., (2017), se proponen iniciativas comunitarias para la prevención del embarazo adolescente. En Summers et al., (2017) se analizan los factores contribuyentes al embarazo entre mujeres afroamericanas, con el fin proporcionar sugerencias a los programas de prevención. Así mismo en Vafai et al., (2020), se analiza la relación entre los episodios depresivos y el embarazo adolescente, encontrándose una relación directa entre estos.

En el mismo sentido del párrafo anterior, autores como Jalanko et al., (2020) analizan el riesgo de la morbilidad psiquiátrica en adolescente inducidos a abortar, encontrándose un mayor riesgo de morbilidad psiquiátrica en el grupo de abortos en adolescente. En Ohene y García, (2020), se realiza una recopilación de las experiencias de las adolescentes embarazadas con el objetivo de entender como sus principios básicos, explican su participación escolar y rendimiento académico. Por su parte en González-Andrade y Saeteros-Cordero, (2020) se realiza un estudio sobre el embarazo en la adolescencia y la morbilidad en neonatales encontrando en la falta chequeos prenatales un factor influyente en la morbilidad. Así mismo en Wong et al., (2020) se analiza la relación entre el estado socioeconómico, la salud mental y el uso de sustancias con el embarazo en adolescentes. En Fortin-Langelier et al., (2019) se encuentra una relación entre la infancia, el abuso sexual y el embarazo adolescente. En Sukhato et al., (2015) se realiza un análisis de la intervención para que los niños de las adolescentes embarazadas nazcan con mejor peso. En Lohan et al., (2014) se realiza una intervención educativa basada en películas para evitar el embarazo en adolescentes.

De otro lado autores como Zangiacomi y Leite da Roza, (2020) emplean un modelo bayesiano para describir los patrones espacio-temporales en los porcentajes de nacimientos vivos en madres adolescentes y sus asociaciones con los índices de desarrollo Humano en las regiones donde residen, encontrándose una relación entre el embarazo adolescente y el medio ambiente. Por su parte en López-Martínez et al., (2019) se desarrolla una red neuronal para predecir la sepsis neonatal logrando una precisión de 83.1%. En Mani et al., (1997) se diseña un modelo bayesiano para predecir el retardo en los recién nacidos. Igualmente, en Rawashdeha et al., (2020) se diseña un sistema inteligente basado en técnicas de minería de datos para la predicción de nacimientos prematuros en mujeres con cerclaje cervical. Un factor importante para disminuir las complicaciones asociadas a los partos prematuros. Otros estudios relacionados con esta introducción pueden ser encontrados en: Alzate et al., (2020), Ellis-Sloan (2019), Tsakiridis et al., (2020), Scott et al., (2019), Thomas A. (2019), Karataşlı et al., (2019).

Las diferentes revisiones bibliográficas muestran que la mayoría de estudios realizados sobre el embarazo en adolescentes, solo se identifican las causas de este problema desde diferentes enfoques como: Social, comunitario, poblacional, psicológico y familiar, entre otros muchos aspectos, encontrándose muy pocos trabajos que emplean las diferentes técnicas inteligentes en el análisis de este problema. Sin embargo, es importante resaltar que si bien, en las revisiones literarias realizadas no existen un análisis de causas del embarazo en adolescentes, basado en técnicas inteligentes, si se analizan problemas relacionados por medio de técnicas como: Redes neuronales, algoritmos bayesianos, técnicas matemáticas y estadísticas, entre muchas otras. Lo anterior constituyen razones de peso, para el análisis de este problema, más aún cuando el mismo, es considerado un problema de salud pública, el cual puede traer repercusiones negativas para las adolescentes, sus familias y toda la sociedad en general.

Con base en la problemática descrita en el párrafo anterior, el objetivo fundamental de este artículo es determinar los factores socio-económicos más influyentes en el embarazo adolescente. Esta predicción está basada en el algoritmo de clasificación J48, el cual es ejecutado por medio de la plataforma de aprendizaje automático y minería de datos denominado Weka (<https://www.cs.waikato.ac.nz/ml/weka/>). El aprendizaje automático permite, a partir de técnicas inteligentes como la minería de datos, entrenar un sistema previamente definido, con el fin de que este sistema pueda predecir el comportamiento de una variable dependiente deseada Tiene hijos (*TH*) después de una fase previa de entrenamiento. Una vez establecido el comportamiento de esta variable dependiente, será posible identificar las principales causas socio económicas que generan esta conducta. Es importante resaltar que una buena fase de entrenamiento se logra con una cantidad de suficiente de datos, no obstante, estos clasificadores ofrecen muy buenos resultados con pocos datos como se encuentra en: Valencia et al., (2015) y Castrillón et al., (2020).

Finalmente, este artículo ha sido organizado de la siguiente forma: Primero se presenta la metodología empleada en el análisis y resolución de este problema. Posteriormente, en la sección de resultados se establecen las principales causas socio económicas que explican el embarazo adolescente. En la discusión de resultados, se realiza una comparación de esta investigación con otras investigaciones referenciadas. Posteriormente, se realizan las conclusiones, consecuencia directa de los análisis y resultados obtenidos en esta investigación. Posteriormente, se incluye una sección de agradecimiento y en la última sección se ilustran las referencias empleadas en esta investigación.

## MATERIALES Y MÉTODOS

Para la estructuración y desarrollo de la metodología, se empleó el algoritmo de clasificación J48, ejecutado mediante la plataforma de aprendizaje automático y minería de datos denominada Weka (<https://www.cs.waikato.ac.nz/ml/weka/>). Este algoritmo permite predecir el comportamiento de una variable dependiente deseada Tiene hijos (*TH*). Una vez se establece el comportamiento de esta variable dependiente, es posible identificar las principales causas que generan este comportamiento. La metodología propuesta en esta investigación, es estructurada en cinco pasos, así: (1) Construcción de la base de datos; (2) Análisis de correlación; (3) Diseño del archivo arff (4) Generación del árbol de clasificación; (5) Principales factores socio-económicos influyentes en el embarazo adolescente.

*Paso 1. Construcción de la base de datos:* A partir de la encuesta nacional de demografía y salud (ENDS) realizada en el año 2015, se tomaron los principales factores que muestran una correlación con la maternidad en adolescentes y la importancia de cada una de las variables que influyen en esta problemática. Para el estudio se tomaron las adolescentes entre 15 y 19 años y las que tienen hijos. Para la toma de los datos fue necesario remitirse al programa DHS (Demographic and Health Survey) y solicitar su respectiva autorización para la toma de los datos ya que, aunque hacen parte de una encuesta, se requiere su autorización (vía electrónica) para poder descargar la información y poderla procesar con fines académicos

Con base en lo anterior, se tomaron las siguientes variables: a) Variable dependiente: Tiene Hijos (*TH*), b) Variables independientes: Tiene pareja actualmente (*TPA*), estrato social (*ES*), responsable de hogar trabaja (*RT*), Supervisión padres (*SP*), EPS pertenece ha instruido métodos anticonceptivos (*EI*), estudia actualmente (*EA*), todos sus hijos son del mismo padre (*HMP*), empezó a tener relaciones a los quince años o más (*RQ*), planifica cuando tiene relaciones (*PR*), tipo de educación– primaria – secundaria – Universidad (*TE*), edad primer hijo madre (*EPHM*), edad primer hijo padre (*EPHP*). Las anteriores variables fueron seleccionadas entre 1500 disponibles. Igualmente se considera el estudio denominado *Determinantes del embarazo en adolescentes en Colombia*, ilustrado en: Royo et al., (2015). Se tomaron 750 registros por cada variable. Estas variables se seleccionaron tomando como referencia análisis empíricos previos, los cuales mostraban una mayor incidencia de estas variables en el contexto colombiano. Constituyéndose, esta investigación en una oportunidad formal para validar la efectividad de las variables seleccionadas.

*Paso 2. Análisis de Correlación:* Una vez elaborada la base de datos del Paso 1, la correlación entre cada una de las variables independientes y la variable dependiente es calculada, lo anterior permite seleccionar las variables con una incidencia moderada sobre la variable dependiente. En este punto solo se seleccionarán las variables con una correlación mayor a 0.2 y menor a 0.8. Es decir, las variables con una incidencia insignificante o muy alta son descartadas. Esto último con el fin de evitar variables que no aporten nada o variables que dominen completamente la variable dependiente.

*Paso 3. Diseño del archivo .Arff :* Tomando como referencia la base de datos construida en el Paso 1, y las variables seleccionadas en el paso 2, se construye el respectivo archivo para ser interpretado por medio del clasificador J48, desde la plataforma de aprendizaje automático y minería de datos denominada Weka. El diseño de este archivo se estructura en dos partes encabezado y datos.

*Paso 4. Generación del árbol de clasificación:* El archivo elaborado en el paso 3, se interpreta por medio de la plataforma Weka, empleando el algoritmo de clasificación J48, mediante la técnica de validación cruzada.

Esta interpretación permite generar el respectivo árbol de decisión. Comparativamente los algoritmos de clasificación bayesiana permiten obtener buenos resultados con pocos datos (Valencia et al., 2015).

*Paso 5. Principales factores socio-económicos influyentes en el embarazo adolescente.:* Con base en el árbol, generado en el Paso 4, se determinan la influencia de los principales factores socio – económicos sobre la variable dependiente Tiene hijos (TH). Los nodos en este árbol representarían las causas (factores socio - económicos), mientras que las hojas representarían la variable dependiente objeto de análisis en esta investigación.

## RESULTADOS

*Paso 1. Construcción de la base de datos:* Como resultado de la aplicación del Paso 1 de la metodología, se obtuvo la siguiente base de datos, la cual es ilustrada en la Tabla 1. Por razones de espacio no se muestra el archivo completo, pero este puede ser obtenido de los autores, previa autorización *The Demographic and Health Surveys (DHS) Program*.

Tabla 1: Bases de datos construida

TPA	ES	RT	SP	EI	EA	HMP	RQ	PR	TE	EPHM	EPHP	TH
1	4	0	1	1	1	1	2	2	3	15	24	0
1	3	0	1	2	1	2	2	2	3	32	30	0
2	3	1	1	1	1	2	1	1	2	17	20	1
1	4	1	1	2	1	1	2	2	2	19	20	0
1	3	1	1	2	1	1	2	2	2	17	22	0
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
1	4	1	1	2	1	1	2	2	3	20	25	0
1	4	0	1	1	1	1	2	2	3	15	24	0
1	3	0	1	2	1	2	2	2	3	32	30	0

*Paso 2. Análisis de Correlación:* Las correlaciones entre cada una de las variables independientes y la variable dependiente son ilustradas en la Tabla 2. Esta tabla, muestra una correlación moderada entre cada una de las variables independientes y la variable dependiente, razón por la cual se seleccionan todas las variables para continuar el análisis en esta investigación.

Tabla 2: Correlación variables independientes vs variable dependiente.

Vd/Vi	TPA	ES	RT	SP	EI	EA	HMP	RQ	PR	TE	EPHM	EPHP
TH	0,77	0,25	0,21	0,44	0,31	0,47	0,25	0,53	0,49	0,26	0,36	0,27

*Paso 3. Diseño del archivo .Arff:* Con el fin de analizar los datos desde la plataforma de aprendizaje automático y minería de datos denominada Weka (<https://www.cs.waikato.ac.nz/ml/weka/>), se tomaron todas las variables analizadas en el paso 2, posteriormente, se estructura el respectivo archivo. arff. Este archivo, es ilustrado en las Tabla 3 y 4. La Tabla 3 muestra el encabezado y la Tabla 4 el cuerpo del archivo. Por razones de espacio no se muestra el archivo completo, pero este puede ser obtenido de los autores, previa autorización *The Demographic and Health Surveys (DHS) Program*.

Tabla 3: Encabezado archivo .Arff

Variable	Tipo de Datos
@attribute TPA_No=1_Si=2	{1,2}
@attribute ES=1_2_3_4_5	Numeric
@attribute RT_No=0_Si=1	{0,1}
@attribute SP_Alta1_Baja2	{1,2}
@attribute EI_No=2_Si=1	{1,2}
@attribute EA_No=2_Si=1	{1,2}
@attribute HMP_No=2_Si=1	{1,2}
@attribute RQ_No=2_Si=1	{1,2}
@attribute PR_No=1_Si=2	{1,2}
@attribute TE Primaria=1_Secundaria=2_Universidad=3	{1,2,3}
@attribute EPHM	Numeric
@attribute EPHP	Numeric
@attribute TH_No=0_Si=1	{0,1}

Tabla 4: Cuerpo archivo .Arff

1	4	0	1	1	1	1	2	2	3	15	24	0
1	3	0	1	2	1	2	2	2	3	32	30	0
2	3	1	1	1	1	2	1	1	2	17	20	1
1	4	1	1	2	1	1	2	2	2	19	20	0
1	3	1	1	2	1	1	2	2	2	17	22	0
1	3	1	1	2	2	1	2	2	2	19	23	0
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
1	5	0	1	1	1	1	2	2	3	25	18	0
1	5	0	1	1	2	1	2	2	2	26	29	0
1	4	0	1	2	1	1	2	2	2	15	24	0
1	3	0	1	2	2	1	2	2	3	27	26	0
1	3	0	1	2	2	1	2	2	2	19	31	0
1	4	1	1	2	1	1	2	2	3	20	25	0

Paso 4. Generación del árbol de clasificación: Como resultado de aplicar la metodología en este paso, se obtiene el árbol de clasificación de la Figura 1, el cual presenta una efectividad superior al 97%, por medio de una validación cruzada. No obstante, si se emplea el mismo set de datos para entrenar el sistema y validarlo, se puede obtener una clasificación superior al 98% como se ilustra en la Tabla 6. Las Tablas 7 y 8 muestran la matriz de precisión para cada una de estas dos formas de clasificación respectivamente, mientras que en la Tabla 9 se compara el algoritmo J48 frente a otros algoritmos.

Tabla 5: Efectividad por validación cruzada.

Correctly Classified Instances	733	97.86 %
Incorrectly Classified Instances	16	2.14 %
Kappa statistic	0.94	
Mean absolute error	0.03	
Root mean squared error	0.14	
Relative absolute error	9.29 %	
Root relative squared error	33.14 %	
Total Number of Instances	749	

Tabla 6: Efectividad con el mismo set de datos

Correctly Classified Instances	737	98.40%
Incorrectly Classified Instances	12	1.60%
Kappa statistic	0.96	
Mean absolute error	0.03	
Root mean squared error	0.12	
Relative absolute error	7.95%	
Root relative squared error	28.21%	
Total Number of Instances	749	

Tabla 7. Matriz de precisión. Validación cruzada

Positivos verdaderos	Falsos – Positivos	Precisión	Recalculo	Medida – F	Área ROC	Clase
0.993	0.066	0.979	0.993	0.986	0.971	0
0.934	0.007	0.977	0.934	0.955	0.971	1
0.979	0.052	0.979	0.979	0.978	0.971	←Prom

Tabla 8: Matriz de precisión. Empleando el mismo set de datos

Positivos verdaderos	Falsos – Positivos	Precision	Recalculo	Medida – F	Área ROC	Clase
0.998	0.060	0.981	0.998	0.99	0.987	0
0.94	0.002	0.994	0.940	0.966	0.987	1
0.984	0.046	0.984	0.984	0.984	0.987	← Prom

Tabla 9: Comparación del algoritmo J48

Técnica	Efectividad	Referencia
Adtree	95.19	<a href="https://weka.sourceforge.io/doc.stable/weka/classifiers/trees/ADTree.html">https://weka.sourceforge.io/doc.stable/weka/classifiers/trees/ADTree.html</a>
BFTree	97.32	<a href="https://weka.sourceforge.io/doc.packages/bestFirstTree/weka/classifiers/trees/BFTree.html">https://weka.sourceforge.io/doc.packages/bestFirstTree/weka/classifiers/trees/BFTree.html</a>
DescionStump	91.58	<a href="https://weka.sourceforge.io/doc.stable-3-8/weka/classifiers/trees/DecisionStump.html">https://weka.sourceforge.io/doc.stable-3-8/weka/classifiers/trees/DecisionStump.html</a>
FT	95.99	<a href="https://weka.sourceforge.io/doc.packages/functionalTrees/weka/classifiers/trees/FT.html">https://weka.sourceforge.io/doc.packages/functionalTrees/weka/classifiers/trees/FT.html</a>
J48Graft	97.86	<a href="https://weka.sourceforge.io/doc.packages/J48graft/weka/classifiers/trees/J48graft.html">https://weka.sourceforge.io/doc.packages/J48graft/weka/classifiers/trees/J48graft.html</a>
LADTree	95.72	<a href="https://weka.sourceforge.io/doc.stable/weka/classifiers/trees/LADTree.html">https://weka.sourceforge.io/doc.stable/weka/classifiers/trees/LADTree.html</a>
LMT	95.99	<a href="https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/LMT.html">https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/LMT.html</a>
NBtree	95.59	<a href="https://weka.sourceforge.io/doc.packages/naiveBayesTree/weka/classifiers/trees/NBTree.html">https://weka.sourceforge.io/doc.packages/naiveBayesTree/weka/classifiers/trees/NBTree.html</a>
RandomForest	96.12	<a href="https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/RandomForest.html">https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/RandomForest.html</a>

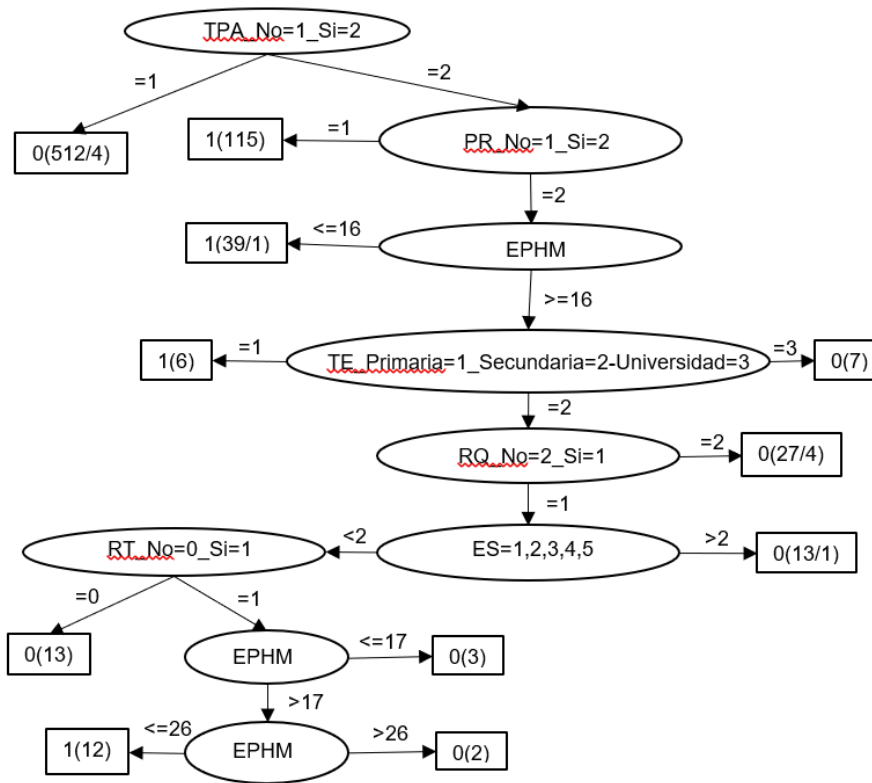


Fig. 1: Árbol de clasificación. Weka

Paso 5. Principales factores socio-económicos influyentes en el embarazo adolescente.

En el árbol de la Figura 1, se muestran las variables más influyentes en el embarazo adolescente: Tiene Pareja Actualmente (TPA), Panifica cuando tiene relaciones (PR), Edad del Primer Hijo Mujer (EPHM), Tipo de educación (TE) y Estrato socio – Económico (ES). Esta figura presenta las siguientes interrelaciones de las causas (variables independientes), influyentes en el embarazo adolescente (ver tabla 10):

Tabla 10: Interrelaciones variables independientes vs variable dependiente

Nodo 1	Nodo 2	Nodo 3	Nodo 4	Hoja
ES <= 2 →	RT = 1(si) →	EPHM >17 →		1 (Tiene hijos)
TPA=2(si) →	PR = 2 (si) →	EPHM >16 →	TE = 1 (Primaria) →	1(Tiene hijos)
TPA=2(si) →	PR = 2 (si) →	EPHM < 16 →		1 (Tiene hijos)
TPA=2(si) →	PR = 1 (no) →	1 (Tiene hijos)		

DISCUSIÓN

Un análisis de los resultados ilustrados en la sección anterior muestra el comportamiento de las causas socio-económicas más influyentes en el embarazo adolescente. Entre estas variables se encuentran: a) Tiene pareja actualmente (TPA), siendo esta la variable de más influyente en el árbol anterior. b) Estrato socio económico (ES), las adolescentes de los estratos sociales 1 y 2 (ES <=2), donde el responsable del hogar trabaja (RT = 1) son más susceptibles de presentar un embarazo adolescente, talvez por la falta de

capacitación y oportunidades carentes por quienes viven en estos estratos, así como la ausencia más marcada del jefe del hogar. c) El tipo de educación (*TE*), tiene una influencia importante sobre la prevención del embarazo adolescente. Entre más alto es el tipo de educación la posibilidad de un embarazo adolescente es menor. Como lo muestra la Figura 1. En este árbol cuando el tipo de educación es superior ( $TE=3$ ) no existen embarazos adolescentes. Si el tipo de educación es primaria ( $TE=1$ ) se presenta un porcentaje mayor de embarazos. d) Los métodos de planificación (*PR*), no son suficientes para prevenir un embarazo estudiantil, siempre será necesario un mayor nivel de educación. No obstante, no emplear métodos de planificación familiar ( $PR = 1$ ) necesariamente, conduce a un embarazo adolescente, como era de esperarse. e) quienes tuvieron un hijo muy joven (antes de los 16 años) están en riesgo de volver tener otro hijo aun siendo adolescentes ( $EPHM < 16$ ).

Cuando se revisa la literatura referenciada en esta investigación en citas como: Quintero y Rojas, (2015); Flórez y Soto, (2006); Reyes y González, (2014); Ohene y García (2020), etc. Se encuentra que este problema siempre ha sido analizado con herramientas basadas en áreas como la Psicología, Sociología, etc, como se ilustró en la introducción. Este documento constituye, una de las primeras investigaciones donde este problema es analizado con herramientas basadas en técnicas inteligentes, como minería de datos. La técnica empleada en este trabajo, permite ilustrar de una forma objetiva el comportamiento de las variables socio-económicas más influyentes en la prevención del embarazo adolescente, transformando un método subjetivo en un método objetivo.

## CONCLUSIONES

En este artículo se identifica de una forma objetiva, con un nivel de confiabilidad superior al 97%, las principales causas del embarazo adolescente: Tiene pareja Actualmente (*TPA*), Planifica cuando tiene relaciones (*PR*), Edad primer hijo mujer (*EPHM*), Tipo de educación (*TE*), Estrato socio económico (*ES*), responsable del hogar trabaja (*RT*). Se resalta que esta investigación permite transformar una metodología tradicionalmente subjetiva en una metodología objetiva, basada en técnicas inteligentes. Los anteriores resultados, les permitirán a las instituciones públicas, redefinir las políticas para el control de este problema. Finalmente, se expresa que es necesario avanzar hacia nuevos estudios los cuales contemplen un número más amplio de variables, así como probar diferentes técnicas inteligentes.

## AGRADECIMIENTOS

Se agradece la colaboración a la Universidad Nacional de Colombia Sede Manizales y en especial al Departamento de Ingeniería Industrial de esta Sede. Igualmente se agradece la colaboración prestada a *The Demographic and Health Surveys (DHS) Program*.

## REFERENCIAS

- Alzate, M., Dongarwar, D., y otros 2 autores, *The Effect of Sexual Literacy on Adolescent Pregnancy in Colombia*, <https://doi.org/10.1016/j.jpaga.2019.09.005>, Journal of pediatric & adolescent Gynecology, 33 (1), 72-82 (2020)
- Brindis, C., *Advancing the Field of Teenage Pregnancy Prevention Through Community-Wide Pregnancy Prevention Initiatives*, DOI: <https://doi.org/10.1016/j.jadohealth.2016.11.027>, J. adolescent health, 60 (3), Supplement S1-S2 (2017)
- Castrillon, O., Sarache, W., y Ruiz, H., *Predicción del Rendimiento Académico por Medio de Técnicas de Inteligencia Artificial*, <http://dx.doi.org/10.4067/S0718-50062020000100093>, Formación Universitaria, 13(1), 93 -102 (2020)
- Ellis-Sloan, K., *Teenage Mothers in Later Life: Time for a Second Look*, <https://doi.org/10.1016/j.adolescence.2019.10.007>, Journal of Adolescence, (77), 98–107 (2019)
- Flórez, C., y Soto, V., *Fecundidad Adolescente y Desigualdad en Colombia y la Región de América Latina y el Caribe*, Reunión de Expertos sobre Población y Pobreza en América Latina y el Caribe, Santiago de Chile, (2006)
- Fortin-Langelier, E., Daigneault, I., y Otros 4 autores, *A Matched Cohort Study of the Association Between Childhood Sexual Abuse and Teenage Pregnancy*, <https://doi.org/10.1016/j.jadohealth.2019.03.001>, Journal of Adolescent Health, (65), 384 – 389 (2019)
- Girma, S., y Paton, D., *Is Education the Best Contraception: The Case of Teenage Pregnancy in England*, <http://dx.doi.org/10.1016/j.socscimed.2015.02.040>, Social Science & Medicine, (131), 1 -9 (2015)
- González-Andrade, F., y Saeteros, X., *Pregnancy in Adolescence and Adverse Neonatal Outcomes in Ecuadorian Mestizo Newborns*, <https://doi.org/10.1016/j.pedneo.2019.11.004>, Pediatrics and Neonatology, (61), 216-223 (2020)
- Hadley, A., Chandra-Mouli, V., y Ingham, R., *Implementing the United Kingdom Government's 10-Year Teenage Pregnancy Strategy for England (1999 -2010): Applicable Lessons for Other Countries*, <http://dx.doi.org/10.1016/j.jadohealth.2016.03.023>, Journal of Adolescent Health, (59) , 68 -74 (2016)

- Jalanko, E., Leppälahti, S., y Otros 2 autores, *The Risk of Psychiatric Morbidity Following Teenage Induced Abortion and Childbirth - A Longitudinal Study From Finland*, <https://doi.org/10.1016/j.jadohealth.2019.08.027>, Journal of Adolescent Health, (66), 345 -351 (2020)
- Karataşlı, V., Göksel, A., y Otros 3 autores, *Maternal and Neonatal Outcomes of Adolescent Pregnancy*, <http://dx.doi.org/10.1016/j.jogoh.2019.02.011>, Journal of Gynecology Obstetrics and Human Reproduction, (48), 347–350 (2019)
- Lohan, M., Aventin, A., y Otros 4 autores, *Feasibility Trial of a Film-Based Educational Intervention for Increasing Boys' and Girls' Intentions to Avoid Teenage Pregnancy: Study Protocol*, <http://dx.doi.org/10.1016/j.ijer.2014.08.003>, International Journal of Educational Research, (68), 35–45 (2014)
- López-Martínez, F., Rolando, E., y Otros 2 autores, *A Neural Network Approach to Predict Early Neonatal Sepsis*, <https://doi.org/10.1016/j.compeleceng.2019.04.015>, Computers and Electrical Engineering, (76), 379–388 (2019)
- Mani, S., McDermott, S., y Valtorta, M., *MENTOR: A Bayesian Model for Prediction of Mental Retardation in Newborns*, D.O.I: 10.1016/S0891-4222(97)00012-7, Research in Developmental Disabilities, 18 (5), 303-318 (1997)
- Royo, M., Forero, L., y Otros 9 autores, *Determinantes del Embarazo en Adolescentes en Colombia: Explicando las Causas de las Causas. Un Estudio a Profundidad a Partir de la Encuesta Nacional de Demografía y Salud – ENDS 2015*, Ministerio de Salud y Social - Profamilia, ISBN: 978-958-8164-51-9, (2015) Recuperado de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/informe-determinantes-sociales-embarazo-en-adolescente.pdf>, (2020)
- Ohene, S., y Garcia, A., *Narratives of Women's Retrospective Experiences of Teen Pregnancy, Motherhood, and School Engagement while Placed in Foster Care*, <https://doi.org/10.1016/j.childyouth.2019.104563>, Children and Youth Services Review, (108), 104563 (2020)
- Osuna, C., y Luna, E., *Valores Éticos en la Formación Universitaria de las Áreas de Ciencias Naturales e Ingeniería y Tecnología, en el Contexto de la Sociedad del Conocimiento*, <http://dx.doi.org/10.4067/S0718-50062011000500005>, Formación Universitaria, 4(5), 29-36 (2011)
- Quintero, P., y Rojas, H., *El Embarazo a Temprana Edad, un Análisis Desde la Perspectiva de Madres Adolescentes*, Revista Virtual Universidad Católica del Norte, ISSN: 0124-5821, (44), 222-237 (2015)
- Rawashdeha, h., Awawdehb, s., y Otros 5 autores, *Intelligent System Based on Data Mining Techniques for Prediction of Preterm Birth for Women with Cervical Cerclage*, <https://doi.org/10.1016/j.compbiolchem.2020.107233>, Computational Biology and Chemistry, (85), 107233 (2020)
- Reyes, D., y González, E., *Elementos Teóricos para el Análisis del Embarazo Adolescente*, Revista Latinoamericana de Sexualidad, Salud y Sociedad, ISSN 1984-6487, (17), 98-123 (2014)
- Scott, R., Wellings, K., y Lindberg, L., *Adolescent Sexual Activity, Contraceptive Use, and Pregnancy in Britain and the U.S.: A Multidecade Comparison*, <https://doi.org/10.1016/j.jadohealth.2019.11.310>, Journal of Adolescent Health, 66(5), 582-588 (2020)
- Sorhaindo, A., Bonell, C., y Otros 4 autores, *Being Targeted: Young Women's Experience of Being Identified for a Teenage Pregnancy Prevention Programme*, <http://dx.doi.org/10.1016/j.adolescence.2016.03.013>, Journal of Adolescence, (49), 181 – 190 (2016)
- Sukhato, K., Wongrathanandha, C., y Otros 4 autores, *Efficacy of Additional Psychosocial Intervention in Reducing Low Birth Weight and Preterm Birth in Teenage Pregnancy: A Systematic Review and Meta-Analysis*, <http://dx.doi.org/10.1016/j.adolescence.2015.07.013>, Journal of Adolescence, (44), 106 -116 (2015)
- Summers, L., Lee, Y., y Lee, H., *Contributing Factors of Teenage Pregnancy Among African-American Females Living in Economically Disadvantaged Communities*, <http://dx.doi.org/10.1016/j.apnr.2017.07.006>, Applied Nursing Research, (37), 44–49 (2017)
- Thomas, A., *Teenage Pregnancy and Long-Acting Contraception*, <https://doi.org/10.1016/j.jadohealth.2019.07.015>, Journal of Adolescent Health, (65), 438 – 439 (2019)
- Tsakiridis, I., Mamopoulos, A. y Otros 2 autores, *Trends in Induced Abortions in Greece: A Cross-sectional Study*, Journal of pediatric & adolescent Gynecology, 33(2), 149-152 (2020)
- Vafai, Y., Thoma, M., y Steinberg, J., *Association Between First Depressive Episode in the Same Year as Sexual Debut and Teenage Pregnancy*, <https://doi.org/10.1016/j.jadohealth.2020.02.001>, J. Adolescent Health, abril, 1 – 6 (2020)
- Valencia, M., Correa, J., y Díaz, F., *Métodos Estadísticos Clásicos y Bayesianos para el Pronóstico de Demanda. Un Análisis Comparativo*, <https://doi.org/10.15446/rev.fac.cienc.v4n1.49775>, Revista Facultad de Ciencias Universidad Nacional de Colombia, 4(1), 52 -67 (2015)
- Wong, S., Twynstra, J., y Otros 3 autores, *Risk Factors and Birth Outcomes Associated with Teenage Pregnancy: A Canadian Sample*, <https://doi.org/10.1016/j.jpag.2019.10.006>, Journal of pediatric & adolescent Gynecology, 33 (2), 153-159, (2020)
- Zangiacomí, E., y Leite da Roza, D., *Ecological Analysis of Adolescent Birth Rates in Brazil: Association with Human Development Index*, <https://doi.org/10.1016/j.wombi.2019.04.002>, Women and Birth, (33), e191–e198 (2020)