

## Metodología para la Predicción del Grado de Riesgo Psicosocial en Docentes de Colegios Colombianos utilizando Técnicas de Minería de Datos

Rodolfo Mosquera<sup>(1)</sup>, Liliana Parra-Osorio<sup>(2)</sup> y Omar D. Castrillón<sup>(1)</sup>

(1) Universidad Nacional de Colombia, Facultad de Ingeniería y Arquitectura, Bloque Q, Campus La Nubia, Manizales - Colombia. (e-mail: rmosqueran@unal.edu.co, odcastrillong@unal.edu.co)

(2) Universidad Libre, Seccional Bogotá, Facultad de Derecho, Maestría en Seguridad Social y Derecho Laboral, Colombia. (e-mail: lilianaparraosorio@hotmail.com)

Recibido Abr. 4, 2016; Aceptado May. 26, 2016; Versión final Jun. 20, 2016, Publicado Dic. 2016

---

### Resumen

En este artículo se presenta la aplicación de la minería de datos en diagnósticos de evaluaciones psicosociales para la identificación del grado de riesgo psicosocial en docentes de educación básica primaria y media secundaria en Colombia. Se tomó una muestra de datos de 1100 registros, correspondientes a evaluaciones personales. Se realizó la revisión y limpieza de los datos para poder utilizarlos como entradas en la herramienta de minería de datos WEKA. Se comparan los resultados obtenidos al aplicar las técnicas de minería de datos, árboles de clasificación J48 y Naive Bayes. Finalmente se obtiene como resultado las características psicosociales predictivas con una efectividad del 91% comparado con el diagnóstico clínico. Se concluye que la herramienta puede ser usada como un instrumento para la prevención de la aparición de factores de riesgo psicosocial en docentes de colegios públicos colombianos.

*Palabras clave:* clasificación de riesgo psicosocial; J48; Naive Bayes; docentes colombianos; jornada básica primaria y secundaria; minería de datos.

## Methodology for Predicting the Psychosocial Risk Level on Colombian Teachers using Data Mining Techniques

### Abstract

This paper proposed the application of data mining to identify the psychosocial risk level in teaching elementary, middle y high school education in Colombia. A sample of 1100 data records corresponding to individual assessments were analyzed and processed. The samples were used as input data to the data mining tool called WEKA. Results were compared to assess the performance when applying data mining techniques and classification trees J48 and Naive Bayes. Finally, the application of this predictive tool allows the accuracy 91% compared to the clinic diagnostic. It is concluded that the method can be used tool for preventing the occurrence of psychosocial risk factors in Colombian public school teachers.

*Keywords:* Psychosocial risk factors classification; WEKA J48; Colombian teachers; teaching elementary data mining.

## INTRODUCCIÓN

Actualmente se han propuesto metodologías y estándares nacionales para la identificación de los factores de riesgo psicosocial presente en los ambientes de trabajo, como la guía británica de estandarización y la serie nacional en la salud psicológica y la seguridad en el lugar de trabajo en Canadá (BSI, 2011, PRIMA-EF 2011, BNQ, CSA Group and MHCC, 2013), se sigue insistiendo en la necesidad de desarrollar nuevas herramientas y metodologías que contribuyan a la identificación, prevención y control de modo flexible, dependiendo el tipo de organización a analizar (Leka et al., 2013), por ejemplo, la guía electrónica para la gestión del estrés y los riesgos psicosociales que facilita la Agencia Europea para la Seguridad y la Salud en el Trabajo (EU-OSHA 2015). La prevención e identificación de los riesgos psicosociales que afectan el ambiente de trabajo se consideran como difíciles de manejar, ya que considerar todos los factores que afectan a los trabajadores y que producen problemas originados por el estrés del trabajo derivan de las relaciones entre personas y organizaciones, los cuales pueden ser por diversos factores (Agencia Europea para la Seguridad y la Salud en el Trabajo EU-OSHA 2012; Leka et al., 2015).

Uno de los problemas más comúnmente encontrados en una amplia variedad de situaciones organizacionales es, cómo, incorporar la participación de los miembros pertenecientes a la organización (Hernández et al., 2016). A pesar de que algunos métodos de estudio pueden tener ventajas significativas, a menudo, estos métodos no proporcionan suficiente evidencia de las relaciones causales y las inferencias estadísticas entre un gran número de factores de entrada y respuestas (Ricardo 2016). Para hacer frente a este problema, la minería de datos integrada (Blanco 2014; Mamani 2015) es una técnica que puede ser un procedimiento alternativo para la evaluación del grado de riesgo psicosocial, que puede reducir la dimensionalidad de forma sistemática.

Un aspecto muy importante sobre el sistema de gestión de la seguridad y la salud en el trabajo es el seguimiento y control de procesos que tengan relación directa con éste (Arias 2014; Saizarbitoria et al., 2014). Por lo tanto, se hace necesario el monitoreo de procesos organizacionales. Estos procesos sintetizan las diferentes etapas del proceso de organización del trabajo, razón por la cual las variables contempladas en cada uno deben tener relación directa con el objeto mismo del trabajo docente (Huerta et al., 2016), así como su análisis y seguimiento, el cual se puede dar a través de las técnicas de minería de datos que muestran en diversos estudios haber sido efectivas como técnicas de inteligencia artificial (Castro et al., 2014; Matos et al., 2006; Castrillón 2014; Coca et al., 2013).

En docentes de colegios públicos se ha establecido que los niveles de riesgo con síntomas de estrés son altos; el 62% presenta riesgo muy alto y alto, los síntomas de estrés y los factores de riesgo psicosocial intralaborales tienen una asociación moderada pero significativa con el Liderazgo y relaciones sociales en el trabajo, con el control sobre el trabajo y con las demandas del trabajo y las recompensas, factores que se deben priorizar para formular programas o estrategias de intervención, se ha concluido que, es urgente la intervención en los docentes desde un enfoque de vigilancia epidemiológica para prevenir las posibles afectaciones en la salud mental y controlar la exposición a los diferentes factores de riesgo psicosocial (Espinosa 2015).

Para el desarrollo de la investigación se consideraron 114 variables, con valores cuantitativos y cualitativos, y 1100 registros de muestra, correspondientes a la aplicación de la batería de riesgos psicosocial (Ministerio de la Protección Social, 2010). El estudio comenzó identificando las variables que estarán directamente implicadas y la clase de referencia que guiará a la información de acuerdo con el conjunto de datos con el que se cuenta para determinar el perfil del docente.

Se empleó como apoyo tecnológico la herramienta de minería de datos WEKA (Waikato Environment for Knowledge Analysis V.3.6.13) (Mark et al., 2009) por medio de la cual se llevó a cabo el pre-procesamiento de los datos mediante la técnica de selección de atributos; posterior a esto, se utilizó como proceso de clasificación los algoritmos de minería Naive Bayes y el algoritmo C4.5 (J48) para ser comparados en la eficiencia en cuanto a la clasificación de las variables de riesgo psicosocial, lo que finalmente, da como mejor herramienta para la clasificación a futuro del riesgo psicosocial en docentes de colegio, a los árboles de decisión basados en el algoritmo C4.5 (J48) (Mark et al., 2009).

## MATERIALES Y METODOS

Se busca establecer un sistema de clasificación basada en un algoritmo de inteligencia artificial que fusione aquellos elementos importantes de la minería de datos y los factores de riesgo psicosociales, de tal manera que sus resultados puedan ser utilizados para aplicar posteriormente sobre la misma situación, la prevención del riesgo psicosocial en docentes como se muestra en la figura 1. Luego se muestran a nivel general los pasos que integran la metodología:

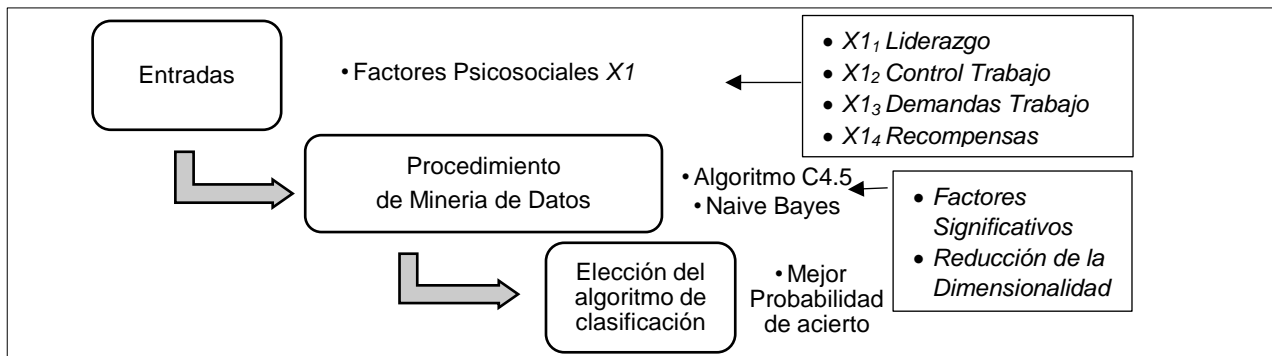


Fig. 1: Esquema general de la metodología a aplicar para la clasificación del grado de riesgo psicosocial en docentes de educación básica primaria y media secundaria en Colombia. Fuente: Elaboración propia.

#### Paso 1. Determinación del conjunto de datos a intervenir

El conjunto de datos que se analizó proviene de la información que se recolectó de evaluaciones realizada a docentes de colegios públicos de una ciudad colombiana. Dicha información está conformada por datos básicos de los docentes y por una clasificación o categorización propia del magisterio colombiano.

#### Paso 2. Limpieza de datos y pre-procesamiento

Las grandes cantidades de información que contiene la base de datos requieren una eficiente presentación, no solo que reduzcan la dimensionalidad, sino también que preserven la información relevante para una clasificación eficiente, para lo cual se verificaron los campos para eliminar los que no contenían información que permitieran realizar el proceso de predicción.

#### Paso 3. Reducción de las variables

En el proceso de reducción de las variables debe identificarse el tipo de información que estas transmiten, dicha información puede ser de tres tipos: 1 Redundante: información repetitiva o predecible; 2 Irrelevante: información que no aporta al proceso de descubrimiento de la información; y 3 Básica: la relevante, la que se constituye como parte importante en un proceso de predicción o descubrimiento de información (Caamaño et al., 2015). La importancia de la reducción de los datos radica en el mejoramiento de los datos de entrada para que los algoritmos puedan clasificar eficientemente la relación entre variables.

#### Paso 4. Filtros de atributos

WEKA (Mark et al., 2009) permite realizar manipulaciones sobre los datos aplicando filtros. Se pueden aplicar en dos niveles, atributos e instancias. Decidimos aplicar un refinamiento al modelo y poder conseguir una probabilidad de acierto ligeramente superior. Las operaciones de filtrado tienen la opción de aplicarse “en cascada”, de manera que cada filtro toma como entrada el conjunto de datos resultante de haber aplicado un filtro anterior. En el modelo utilizamos el método de búsqueda exhaustiva, el cual se puede expresar como una tupla (Anon 2016, p.3):

$$(X_1, X_2, \dots, X_n) \quad (1)$$

Satisfaciendo unas restricciones

$$P(X_1, X_2, \dots, X_n) \quad (2)$$

Optimizando una cierta función objetivo. En cada momento, el algoritmo se encontrará en un cierto nivel  $k$ , con una solución parcial

$$(X_1, \dots, X_k) \quad (3)$$

Cada conjunto de posibles valores de la tupla representa un nodo del árbol de soluciones. Se sigue hasta que la solución parcial sea una solución completa del problema, o hasta que no queden más posibilidades por probar.”

### Paso 5. Evaluación de las técnicas de minería de datos

Para el desarrollo de esta investigación se utilizaron las técnicas de clasificación y predicción en la construcción de modelos a partir de los datos para ser comparados y definir que algoritmo es mejor para clasificar la predicción del grado de riesgo psicosocial en docentes colombianos. Como clasificador inicial utilizamos el clasificador bayesiano (Naive Bayes) y en segunda instancia los árboles de decisión C4.5 (J48). Para clasificar los datos se ha utilizado la herramienta de minería de datos llamada WEKA y para medir la efectividad del algoritmo de clasificación se ha comparado la clase predicha con la clase real de las instancias. Existen diversos modos para llevar a cabo la evaluación, en este caso se empleó "use training set" que permite utilizar la misma muestra para entrenar y probar (Hepner 1990). Entre los algoritmos que proporciona WEKA, se analizaron los siguientes:

#### Algoritmo Naive Bayes

Un clasificador Bayesiano, es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales (Agarwal et al., 2015). El clasificador Bayes combina este modelo con una regla de decisión. La primera regla en común, es para recoger la hipótesis del más probable, también conocido como el máximo a posteriori o MAP. El clasificador Bayer (la función Classify) (Larrañaga et al., 1997, p.6) se define como:

$$p(X_1 = x_1, \dots, X_n = x_n / C = c) = \prod_{i=1}^n p(X_i = x_i / C = c) \quad (4)$$

Donde: En el caso de que las  $n$  variables predictoras  $X_1, \dots, X_n$  sean continuas, se tiene que el paradigma Naive Bayes se convierte en buscar el valor de la variable  $C$ , que denotamos por  $c$ , que maximiza la probabilidad a posteriori de la variable  $C$ , dada la evidencia expresada como una instanciación de las variables  $X_1, \dots, X_n$ , esto es,  $X = (X_1, \dots, X_n)$ . Por tanto, en el paradigma naïve Bayes, la búsqueda del diagnóstico más probable,  $c^*$ , una vez conocidos los síntomas  $(X_1, \dots, X_n)$  de un determinado paciente, se reduce a:

$$c^* = \arg_c \max p(C = c | X_1 = x_1, \dots, X_n = x_n) \quad (5)$$

$$= \arg_c \max p(C = c) \prod_{i=1}^n p(X_i = x_i / C = c) \quad (6)$$

#### Algoritmo C4.5 (J48)

El algoritmo C4.5 construye árboles de decisión de un sistema de datos del entrenamiento de la misma forma que el algoritmo ID3, que usa el concepto de entropía de la información. Los datos del entrenamiento son un sistema  $S = s_1, s_2, \dots$  de muestras ya clasificados. Cada ejemplo  $s_j = \{x_1, x_2, \dots\}$  es un vector donde  $x_1, x_2, \dots$  representan los atributos o características del ejemplo. Los datos de entrenamiento son aumentados con un vector  $C = \{c_1, c_2, \dots\}$  donde  $c_1, c_2, \dots$  representan la clase a la que pertenece cada muestra. C4.5 es una extensión del algoritmo ID3 desarrollado anteriormente por Quinlan. Los árboles de decisión generador por C4.5 pueden ser usados para clasificación, y por esta razón, C4.5 está casi siempre referido como un clasificador estadístico. (Quinlan, J. R. 1993).

### Paso 6. Definición de la técnica de minería de datos

La evaluación de los datos obtenidos a partir de la aplicación de los dos métodos, se efectúa utilizando como variables de comparación las siguientes: Instancias correctas, error Absoluto, matriz de confusión y facilidad para la interpretación de los datos.

## RESULTADOS Y DISCUSIÓN

### Paso 1. Determinación del conjunto de datos a intervenir

Aplicamos el cuestionario de factores de riesgo psicosocial intralaboral de la batería de instrumentos para la evaluación de factores de riesgo psicosocial, desarrollada por el Ministerio de Trabajo de Colombia (2008) para obtener los resultados cualitativos del cuestionario, donde se tiene en cuenta las siguientes variables: De los dominios: Liderazgo y relaciones sociales en el trabajo, Control sobre el trabajo, Demandas del trabajo y Recompensas. Las variables son 114, como se muestra en la tabla 1.

Tabla 1. Variables de los dominios de factores de riesgo psicosocial intralaboral.

| No. | Dominios             | Variable  |
|-----|----------------------|---|
| x1  | Demandas del Trabajo | El ruido en el lugar donde trabajo es molesto   |
| x2  |                      | En el lugar donde trabajo hace mucho frío   |
| x3  |                      | En el lugar donde trabajo hace mucho calor  |
| x4  |                      | El aire en el lugar donde trabajo es fresco y agradable                                       |
| x5  |                      | La luz del sitio donde trabajo es agradable   |
| x6  |                      | El espacio donde trabajo es cómodo  |
| x7  |                      | En mi trabajo me preocupa estar expuesto a sustancias químicas que afecten mi salud           |
| x8  |                      | Mi trabajo me exige hacer mucho esfuerzo físico   |
| x9  |                      | Los equipos o herramientas con los que trabajo son cómodos                                    |
| x10 |                      | En mi trabajo me preocupa estar expuesto a microbios, animales o plantas que afecten mi salud |
| x11 |                      | Me preocupa accidentarme en mi trabajo  |
| x12 |                      | El lugar donde trabajo es limpio y ordenado   |
| x13 |                      | Por la cantidad de trabajo que tengo debo quedarme tiempo adicional                           |
| x14 |                      | Me alcanza el tiempo de trabajo para tener al día mis deberes                                 |
| x15 |                      | Por la cantidad de trabajo que tengo debo trabajar sin parar                                  |
| x16 |                      | Mi trabajo me exige hacer mucho esfuerzo mental   |
| x17 |                      | Mi trabajo me exige estar muy concentrado   |
| x18 |                      | Mi trabajo me exige memorizar mucha información   |
| x19 |                      | En mi trabajo tengo que tomar decisiones difíciles muy rápido                                 |
| x20 |                      | Mi trabajo me exige atender a muchos asuntos al mismo tiempo                                  |
| x21 |                      | Mi trabajo requiere que me fije en pequeños detalles  |
| x22 |                      | En mi trabajo respondo por cosas de mucho valor   |
| x23 |                      | En mi trabajo respondo por dinero de la empresa   |
| x24 |                      | Como parte de mis funciones debo responder por la seguridad de otros                          |
| x25 |                      | Respondo ante mi jefe por los resultados de toda mi área de trabajo                           |
| x26 |                      | Mi trabajo me exige cuidar la salud de otras personas   |
| x27 |                      | En el trabajo me dan órdenes contradictorias  |
| x28 |                      | En mi trabajo me piden hacer cosas innecesarias   |
| x29 |                      | En mi trabajo se presentan situaciones en las que debo pasar por alto normas o procedimientos |
| x30 |                      | En mi trabajo tengo que hacer cosas que se podrían hacer de una forma más práctica            |
| x31 |                      | Trabajo en horario de noche   |
| x32 |                      | En mi trabajo es posible tomar pausas para descansar  |
| x33 |                      | Mi trabajo me exige laborar en días de descanso, festivos o fines de semana                   |
| x34 |                      | En mi trabajo puedo tomar fines de semana o días de descanso al mes                           |
| x35 |                      | Cuando estoy en casa sigo pensando en el trabajo  |
| x36 |                      | Discuto con mi familia o amigos por causa de mi trabajo                                       |
| x37 |                      | Debo atender asuntos de trabajo cuando estoy en casa  |
| x38 |                      | Por mi trabajo el tiempo que paso con mi familia y amigos es muy poco                         |
| x39 |                      | Puedo tomar pausas cuando las necesito  |
| x40 |                      | Puedo parar un momento mi trabajo para atender algún asunto personal                          |
| x41 |                      | Los cambios que se presentan en mi trabajo dificultan mi labor                                |
| x42 |                      | Atiendo clientes o usuarios muy enojados  |
| No. | Dominios             | Variable  |
| x43 | Demandas del Trabajo | Atiendo clientes o usuarios muy preocupados   |
| x44 |                      | Atiendo clientes o usuarios muy tristes   |
| x45 |                      | Mi trabajo me exige atender personas muy enfermas   |
| x46 |                      | Mi trabajo me exige atender personas muy necesitadas de ayuda                                 |
| x47 |                      | Atiendo clientes o usuarios que me maltratan  |
| x48 |                      | Para hacer mi trabajo debo demostrar sentimientos distintos a los míos                        |
| x49 |                      | Mi trabajo me exige atender situaciones de violencia  |
| x50 |                      | Mi trabajo me exige atender situaciones muy tristes o dolorosas                               |

|     |   |   |
|-----|---|---|
| x51 | Control sobre el Trabajo                      | Mi trabajo me permite desarrollar mis habilidades                                     |
| x52 |   | Mi trabajo me permite aplicar mis conocimientos                                       |
| x53 |   | Mi trabajo me permite aprender nuevas cosas   |
| x54 |   | Me asignan el trabajo teniendo en cuenta mis capacidades                              |
| x55 |   | Puedo decidir cuánto trabajo hago en el día   |
| x56 |   | Puedo decidir la velocidad a la que trabajo   |
| x57 |   | Puedo cambiar el orden de las actividades en mi trabajo                               |
| x58 |   | Los cambios en mi trabajo han sido beneficiosos                                       |
| x59 |   | Me explican claramente los cambios que ocurren en mi trabajo                          |
| x60 |   | Puedo dar sugerencias sobre los cambios que ocurren en mi trabajo                     |
| x61 |   | Cuando se presentan cambios en mi trabajo se tienen en cuenta mis ideas y sugerencias |
| x62 |   | Me informan con claridad cuáles son mis funciones                                     |
| x63 |   | Me informan cuáles son las decisiones que puedo tomar en mi trabajo                   |
| x64 |   | Me explican claramente los resultados que debo lograr en mi trabajo                   |
| x65 |   | Me explican claramente el efecto de mi trabajo en la empresa                          |
| x66 |   | Me explican claramente los objetivos de mi trabajo                                    |
| x67 |   | Me informan claramente quien me puede orientar para hacer mi trabajo                  |
| x68 |   | Me informan claramente con quien puedo resolver los asuntos de trabajo                |
| x69 |   | La empresa me permite asistir a capacitaciones relacionadas con mi trabajo            |
| x70 |   | Recibo capacitación útil para hacer mi trabajo  |
| x71 |   | Recibo capacitación que me ayuda a hacer mejor mi trabajo                             |
| x72 | Liderazgo y Relaciones Sociales en el Trabajo | Mi jefe me da instrucciones claras  |
| x73 |   | Mi jefe ayuda a organizar mejor el trabajo  |
| x74 |   | Mi jefe tiene en cuenta mis puntos de vista y opiniones                               |
| x75 |   | Mi jefe me anima para hacer mejor mi trabajo  |
| x76 |   | Mi jefe distribuye las tareas de forma que me facilita el trabajo                     |
| x77 |   | Mi jefe me comunica a tiempo la información relacionada con el trabajo                |
| x78 |   | La orientación que me da mi jefe me ayuda a hacer mejor el trabajo                    |
| x79 |   | Mi jefe me ayuda a progresar en el trabajo  |
| x80 |   | Mi jefe me ayuda a sentirme bien en el trabajo  |
| x81 |   | Mi jefe ayuda a solucionar los problemas que se presentan en el trabajo               |
| x82 |   | Siento que puedo confiar en mi jefe   |
| x83 |   | Mi jefe me escucha cuando tengo problemas de trabajo                                  |
| x84 |   | Mi jefe me brinda su apoyo cuando lo necesito   |
| x85 |   | Me agrada el ambiente de mi grupo de trabajo  |
| x86 |   | En mi grupo de trabajo me tratan de forma respetuosa                                  |
| x87 |   | Siento que puedo confiar en mis compañeros de trabajo                                 |
| x88 |   | Me siento a gusto con mis compañeros de trabajo                                       |
| x89 |   | En mi grupo de trabajo algunas personas me maltratan                                  |
| x90 |   | Entre compañeros solucionamos los problemas de forma respetuosa                       |
| x91 |   | Hay integración en mi grupo de trabajo  |
| x92 |   | Mi grupo de trabajo es muy unido  |
| x93 |   | Las personas en mi trabajo me hacen sentir parte del grupo                            |

| No.  | Dominios                                      | Variable  |
|------|---|---|
| x94  | Liderazgo y Relaciones Sociales en el Trabajo | Cuando tenemos que realizar trabajo de grupo los compañeros colaboran |
| x95  |   | Es fácil poner de acuerdo al grupo para hacer el trabajo              |
| x96  |   | Mis compañeros de trabajo me ayudan cuando tengo dificultades         |
| x97  |   | En mi trabajo las personas nos apoyamos unos a otros                  |
| x98  |   | Algunos compañeros de trabajo me escuchan cuando tengo problemas      |
| x99  |   | Me informan sobre lo que hago bien en mi trabajo                      |
| x100 |   | Me informan sobre lo que debo mejorar en mi trabajo                   |
| x101 |   | La información que recibo sobre mi rendimiento en el trabajo es clara |

|      |             |   |
|------|-------------|---|
| x102 |             | La forma como evalúan mi trabajo en la empresa me ayuda a mejorar     |
| x103 |             | Me informan a tiempo sobre lo que debo mejorar en el trabajo          |
| x104 | Recompensas | En la empresa confían en mi trabajo                                   |
| x105 |             | En la empresa me pagan a tiempo mi salario                            |
| x106 |             | El pago que recibo es el que me ofreció la empresa                    |
| x107 |             | El pago que recibo es el que merezco por el trabajo que realizo       |
| x108 |             | En mi trabajo tengo posibilidades de progresar                        |
| x109 |             | Las personas que hacen bien el trabajo pueden progresar en la empresa |
| x110 |             | La empresa se preocupa por el bienestar de los trabajadores           |
| x111 |             | Mi trabajo en la empresa es estable                                   |
| x112 |             | El trabajo que hago me hace sentir bien                               |
| x113 |             | Siento orgullo de trabajar en esta empresa                            |
| x114 |             | Hablo bien de la empresa con otras personas                           |

Cada una de ellos con un resultado final donde aparece el grado de riesgo al cual está expuesto el docente, asociado a un conjunto de datos de 1082, después del proceso de limpieza y pre-procesamiento que fue estudiado, el cual tiene los siguientes atributos: Edad, Sexo, casado, antigüedad en el cargo, Grado de riesgo psicosocial, sR: sin riesgo, bajo, medio, alto y muy alto.

Tabla 2. Ficha técnica de las evaluaciones realizadas.

|                           |  |
|---------------------------|--|
| Diseño muestral           | Probabilístico estratificado.  |
| Población objetivo        | Hombres y mujeres mayores de 18 años.  |
| Universo representado     | 115 colegios públicos de la ciudad de Cali.                                    |
| Técnica                   | Entrevista cara a cara, en colegios de estrato 1 a 3.                          |
| Tamaño de muestra         | 1100 en 32 colegios de la ciudad   |
| Momento estadístico       | 12 de junio a 12 agosto de 2014  |
| Financiación              | Recursos propios   |
| Margen de error observado | 1.6% total (Error estándar relativo máximo, con un nivel de confianza del 95%) |

#### *Probabilístico.*

Esto significa que las unidades de selección tienen una probabilidad conocida y distinta de cero de entrar en la muestra para cada miembro de la población, lo cual, permite conocer la precisión de los resultados muestrales.

#### *Estratificado.*

Porque las unidades de selección son agrupadas por características similares de tipo geográficas y Socioeconómicas. (Colegios públicos de la ciudad de Santiago de Cali, ubicados en estrato 2 y 3, en el segundo cuatrimestre del año 2014)

#### *Paso 1. Datos de la población a intervenir*

La caracterización demográfica mostró que la población tiene una edad mínima de 23 años y máxima de 62 años con una edad media de 36 años, el 48% (529) son de sexo Femenino y el 51% (553) de sexo masculino, de los cuales el 42% (462) reporta estar casado, existen docentes con tres (2) años de antigüedad en el cargo y docentes con veinte (20) años de trabajo, como se puede ver en la tabla 3.

#### *Paso 2. Limpieza de datos y pre-procesamiento*

En el conjunto de datos que se estudió se encuentran datos continuos y categóricos, por lo tanto, es necesario convertir los continuos a categóricos. Se obtiene el número de segmentos y rangos en que pueden agruparse los segmentos de tiempo, conservando la información relevante (García 2007), Este conjunto de datos se sometió a etapas de limpieza y pre-procesamiento con el fin de lograr la reducción. De los 1100 datos iniciales obtenidos de las evaluaciones personales, al tabular los datos se encontraron campos sin información, lo que conlleva a eliminar estos registros y tener 1082 datos.

#### *Paso 3. Reducción de las variables*

De acuerdo con los tres tipos de información definidos antes y a partir de los datos que son objeto de estudio en este artículo, se eliminaron las variables: Unión libre y número de personas que dependen económicamente de la persona, por no aportar mucha información para el estudio.

Tabla 3. Distribuciones de la población laboral según variables socio demográficas y laborales en docentes de colegios públicos de una ciudad de Colombia.

| <i>Características</i> | <i>Descriptivas</i> | <i>%</i> |
|------------------------|---------------------|----------|
| Edad n (Años)          | 1082                |          |
| Mínimo                 | 23                  |          |
| Máximo                 | 62                  |          |
| Media                  | 36                  |          |
| Desviación estándar    | 8                   |          |
| <i>Características</i> | <i>Descriptivas</i> | <i>%</i> |
| Sexo                   |                     | 100      |
| F                      | 529                 | 48.89%   |
| M                      | 553                 | 51.10%   |
| Casado                 | 1082                | 100      |
| Sí                     | 462                 | 42.69%   |
| No                     | 620                 | 57.30%   |
| Antigüedad             |                     |          |
| Mínimo                 | 2                   |          |
| Máximo                 | 20                  |          |
| Media                  | 12                  |          |
| Riesgo Total           | 1082                | 100      |
| Sin Riesgo             | 102                 | 9.42%    |
| Bajo                   | 258                 | 23.84%   |
| Medio                  | 279                 | 25.78%   |
| Alto                   | 162                 | 14.97%   |
| Muy Alto               | 281                 | 25.97%   |

#### Paso 4. Filtros de atributos

Después de realizar la categorización de los datos, se trabaja con el conjunto de datos modificado y filtrado, compuesto por 45 columnas y 1082 registros. El resultado obtenido conformado por cuarenta y cuatro campos que contienen la información relevante para el proceso de la información se muestra en la estructura de la tabla 4.

$$(X_1, X_2, \dots, X_{114}) \quad (5)$$

Satisfaciendo unas restricciones, las que están determinadas por el funcionamiento del algoritmo.

$$P(X_1, X_2, \dots, X_{44}) \quad (6)$$

Donde  $x_i = (0, 1)$ . (7)

Los datos agrupados en la tabla hacen referencia a todas las variables analizadas en el proceso de minería de datos, agrupado en los principales dominios: Liderazgo y relaciones sociales en el trabajo, Control sobre el trabajo, Demandas del Trabajo y Recompensas.

Tabla 4. Atributos incluidos en el proceso de minería de datos

| <i>No.</i> | <i>Dominios</i>                               | <i>Variable</i>                                | <i>Abreviatura</i> |
|------------|---|--|--------------------|
| 1          |   | Edad   |                    |
| 2          | Liderazgo y relaciones sociales en el trabajo | Características del liderazgo                  | CLD                |
| 3          |   | relaciones sociales en el trabajo              | RST                |
| <i>No.</i> | <i>Dominios</i>                               | <i>Variable</i>                                | <i>Abreviatura</i> |
| 4          | Liderazgo y relaciones sociales en el trabajo | retroalimentación del desempeño                | RAD                |
| 5          |   | relación con los colaboradores                 | RCC                |
| 6          |   | organización del trabajo por parte del jefe    | OTR                |
| 7          |   | puntos de vista y opiniones por parte del jefe | PVO                |
| 8          |   | comunicación asertiva del trabajo con el jefe  | JCT                |
| 9          |   | confianza en el jefe                           | CNJ                |
| 10         |   | integración en el grupo de trabajo             | INTG               |
| 11         |   | colaboración entre compañeros                  | COLB               |



|    |                          |   |        |
|----|--------------------------|---|--------|
| 12 |                          | apoyo laboral entre compañeros  | APY    |
| 13 |                          | información sobre el rendimiento en el trabajo                          | IRTC   |
| 14 |                          | realimentación sobre puntos a mejorar en el trabajo                     | LIDZ   |
| 15 | Control sobre el trabajo | Claridad de rol   | CR     |
| 16 |                          | capacitación  | CAP    |
| 17 |                          | participación y manejo del cambio                                       | PMC    |
| 18 |                          | oportunidades para el uso y desarrollo de habilidades y conocimientos   | ODHC   |
| 19 |                          |   |        |
| 20 |                          | control y autonomía sobre el trabajo                                    | CAST   |
| 21 | Demandas del Trabajo     | Demandas ambientales y de esfuerzo físico                               | DAEF   |
| 22 |                          | demandas emocionales  | DE     |
| 23 |                          | demandas cuantitativas  | DC     |
| 24 |                          | influencia del trabajo sobre el entorno extra laboral                   | ITEE   |
| 25 |                          | exigencias de responsabilidad del cargo                                 | ERC    |
| 26 |                          | demandas de carga mental  | DCM    |
| 27 |                          | consistencia de rol   | CNR    |
| 28 |                          | demandas de la jornada de trabajo                                       | DJT    |
| 29 |                          | exposición a discomfort térmico   | CALOR  |
| 30 |                          | esfuerzo mental   | ESFM   |
| 31 |                          | altos niveles de concentración  | CONC   |
| 32 |                          | memorización excesiva de información                                    | EMEM   |
| 33 |                          | seguridad física de otras personas                                      | SGO    |
| 34 |                          | pensamiento laboral en casa   | TRC    |
| 35 |                          | atención a asuntos del trabajo en casa                                  | AATC   |
| 36 |                          | atención a clientes o usuarios muy enojados                             | ACE    |
| 37 |                          | atención a clientes o usuarios muy preocupados                          | ACUP   |
| 38 |                          | atención a personas enfermas  | APE    |
| 39 |                          | atención a clientes o usuarios que me maltratan                         | ACUM   |
| 40 | Recompensas              | Recompensas derivadas de la pertenencia a la organización y del trabajo | RD POT |
| 41 |                          |   |        |
| 42 |                          | reconocimiento y compensación   | RYC    |
| 43 |                          | pago acorde a la labor  | PAGT   |
| 44 |                          | desarrollo profesional en la empresa                                    | ASCN   |

#### Paso 5. Evaluación de las técnicas de minería de datos

*Comparación entre los resultados del algoritmo Naive Bayes y C4.5 (J48):* Luego de aplicar los algoritmos Naive Bayes y J48 al conjunto de datos de entrenamiento, se obtuvieron los resultados que se muestran en la tabla 5, la cual presenta de manera comparativa las instancias correctas generadas por cada algoritmo.

Tabla 5. Comparación entre algoritmos de clasificación

| % Entrenamiento | % Validación    | % Clasificación J48 | % Clasificación Naive Bayes |
|-----------------|-----------------|---------------------|-----------------------------|
| 10              | 90              | 89                  | 89                          |
| 20              | 80              | 91                  | 90                          |
| 40              | 60              | 92                  | 90                          |
| 50              | 50              | 92                  | 90                          |
| 70              | 30              | 92                  | 90                          |
| 80              | 20              | 92                  | 90                          |
|                 | Promedio Simple | 91                  | 89                          |

En cuanto a los datos de entrenamiento con el 10% de los datos el algoritmo J48 presenta un 89% de clasificación correcta, mientras que el algoritmo Naive Bayes presenta igual porcentaje, la diferencia radica en que a mayor número de datos de entrenamiento, el algoritmo J48 clasifica mejor, con un 91% de eficiencia frente a un 89% del algoritmo Naive Bayes.

#### Comparación de matrices de confusión

Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases (Corso et al., 2009). A continuación se muestran las matrices de confusión generadas por cada uno de los algoritmos, aplicados al mismo conjunto de datos. En la tabla 6 podemos observar que los valores de la diagonal son los aciertos y el resto los errores. Para el Algoritmo Naive Bayes se observa que de los 260 usuarios con perfil b, 240 fueron bien clasificados y 20 presentaron errores, del perfil c, 149 fueron bien clasificados y 131 presentaron error, del perfil d, 120 fueron bien clasificados y 40 presentaron error y con perfil e, los 280 fueron bien clasificados. Para el Algoritmo J48 en la Tabla 7 se observa que de los 260 usuarios con perfil a, 240 fueron bien clasificados y 20 presentaron errores y de los 160 usuarios con perfil d, 120 fueron bien clasificados y 40 presentaron errores.

Tabla 6. Matriz de confusión – Adaptado de: weka.classifiers.bayes.NaiveBayes (Mark et al., 2009)

| Matriz de Confusión |     |     |     |     |                    |
|---------------------|-----|-----|-----|-----|--------------------|
| a                   | b   | c   | d   | e   | ← Clasificado como |
| 100                 | 0   | 0   | 0   | 0   | a = sR             |
| 0                   | 240 | 20  | 0   | 0   | b = Bajo           |
| 0                   | 80  | 149 | 51  | 0   | c = Medio          |
| 0                   | 0   | 0   | 120 | 40  | d = Alto           |
| 0                   | 0   | 0   | 0   | 280 | e = MAlto          |

Tabla 7. Matriz de confusión – Adaptado de: weka.classifiers.trees.J48 -C 0.25 -M 2 (Mark et al., 2009)

| Matriz de Confusión |     |     |     |     |                    |
|---------------------|-----|-----|-----|-----|--------------------|
| a                   | b   | c   | d   | e   | ← Clasificado como |
| 100                 | 0   | 0   | 0   | 0   | a = sR             |
| 0                   | 240 | 20  | 0   | 0   | b = Bajo           |
| 0                   | 0   | 280 | 0   | 0   | c = Medio          |
| 0                   | 0   | 0   | 120 | 40  | d = Alto           |
| 0                   | 0   | 0   | 0   | 280 | e = MAlto          |

Por lo tanto, observando los resultados obtenidos en el conjunto de datos de entrenamiento y en la comparación de las matrices de confusión, se decide trabajar con el algoritmo J48 para realizar la clasificación y la predicción de los factores de riesgo psicosocial.

#### Árbol de decisión generado

El algoritmo C4.5 (J48), que es un clasificador de tipo árbol de decisión que genera gráficos tipo árbol, el cual permite tomar decisiones respecto a las reglas generadas. El árbol de decisión que muestra la figura 2, se acerca más a la realidad del entorno social de los docentes colombianos, pues facilita el identificar y poder prevenir posibles apariciones de factores de riesgo psicosocial en la labor docente primaria y secundaria, es de resaltar, como la variable información sobre el rendimiento en el trabajo juega un papel fundamental en el estado de salud mental de los docentes evaluados, muy ligada a características del liderazgo, atención a clientes o usuarios que me maltratan y colaboración entre compañeros. También, es de notar que los docentes con edades entre 56 - 60 años y más presentan sin riesgo o riesgo bajo y la característica organización del trabajo por parte del jefe, pago acorde al salario, demandas cuantitativas del trabajo y relación con los compañeros, presentan un riesgo muy alto, se identifica qué, la variable PVO, puntos de vista y opiniones por parte del jefe, está relacionada con el grado de riesgo alto, con docentes que lleven más de 13 años en su cargo y que la variable OTR, organización del trabajo por parte del jefe, ocasiona en docentes del sexo femenino un riesgo muy alto, estos resultados, posiblemente facilitarán la aplicación de la clasificación en otros colegios de la ciudad y del país, ayudando de esta manera a la detección de factores generadores de riesgo psicosocial y a la prevención de enfermedades derivadas por el estrés en el trabajo.

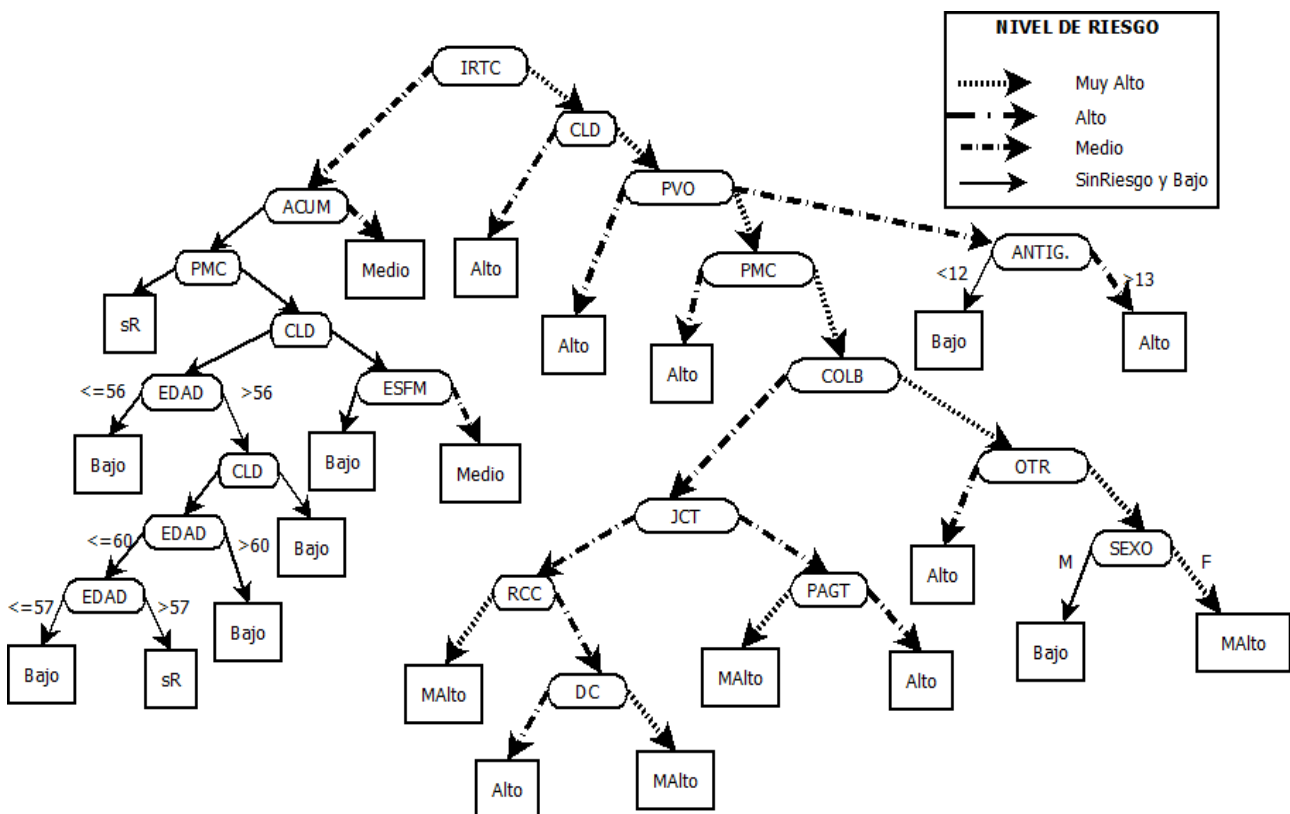


Fig. 2: Árbol de decisión con datos generales generado por el algoritmo J48 para un porcentaje de clasificación del 91% en promedio – Adaptado de WEKA (Mark et al., 2009)

A pesar de que en la literatura evaluada (BSI, 2011, PRIMA-EF, BNQ, CSA Group and MHCC, 2013; EU-OSHA 2012; Leka et al., 2015) y sus antecedentes, los procesos de investigación de los factores de riesgo psicosocial utilizan técnicas fundamentadas en la sociología, la psicología laboral, la organización y administración, la organización del trabajo, no se identifica que se utilicen técnicas de inteligencia artificial para la identificación y clasificación de riesgo psicosocial en la seguridad y salud en el trabajo. Asimismo, el problema identificado es de interés relevante debido a que se orienta a consolidar la eficiencia y la competitividad de las organizaciones educativas públicas.

Las técnicas de inteligencia artificial para la clasificación de patrones como las Redes Neuronales Artificiales (ANN) y las técnicas de lógica difusa, pueden ser de gran utilidad para la clasificación y reconocimiento de patrones en el análisis de los factores relativos al trabajo que puedan generar condiciones laborales que conlleven la aparición de factores psicosociales, que pueden ser agentes de estrés ocupacional y de enfermedad laboral en cualquier ámbito académico.

A partir de este estudio, se puede vislumbrar futuros estudios a realizar en el campo de la Inteligencia Artificial, es interesante, la posibilidad que ofrecen los sistemas inteligentes de construir y evaluar un clasificador aplicando técnicas inteligentes a la identificación de riesgo psicosocial como herramienta para la gestión y control gerencial de los sistemas de seguridad y salud en el trabajo, que sirva como elemento para fortalecer la toma de decisiones en la política de la alta dirección empresarial. También, se deja claro la necesidad de realizar mejoramiento continuo a las herramientas que tiene el país para la identificación, evaluación, control y seguimiento que se hace a los factores de riesgo psicosocial, tanto en empresas públicas como privadas.

Finalmente, se resalta que en la revisión literaria no se encuentra una metodología para la predicción del grado de riesgo psicosocial utilizando clasificadores, que tenga como objetivo identificar los factores de riesgo psicosocial para prevenir enfermedades mentales y físicas derivadas de patologías asociadas al estrés laboral. Por lo general, las investigaciones existentes centran el problema en el tratamiento de la sintomatología de los docentes, desvinculando el proceso de promoción y prevención en la salud. Por tanto, se pueden generar nuevos modelos que además de permitir la predicción de resultados con una alta eficiencia, puedan abordar variables que permiten obtener un mayor beneficio de las condiciones laborales desde la teoría de los Entornos laborales Saludables o programas como el Total Worker Health (Schill et al., 2013)

## CONCLUSIONES

Los resultados permiten obtener las siguientes conclusiones: 1) La metodología definida presenta ventajas importantes sobre otras metodologías, ya que se obtienen características apropiadas de factores de riesgo psicosocial (precisión, capacidad de clasificar diversas características en la población final de respuesta y asimismo, capacidad para identificar diversidad de frentes que pueden ser tratados). 2) Las características más influyentes en la clasificación son: Edad, sexo, antigüedad, Información sobre el rendimiento en el trabajo, Características del liderazgo, Atención a clientes o usuarios que me maltratan, Puntos de vista y opiniones por parte del jefe, Participación y manejo del cambio, Esfuerzo mental, Colaboración entre compañeros, Organización del trabajo por parte del jefe, Comunicación asertiva del trabajo con el jefe, Pago acorde a la labor, Relación con los colaboradores Y Demandas cuantitativas. 2) La descripción detallada de la metodología, permite que sea replicada por parte de las personas encargadas de la seguridad y la salud en el trabajo. 3) Al analizar las tendencias organizacionales (relaciones en el trabajo, bienestar, calidad de vida), las rectorías de los colegios, pueden seleccionar diferentes programas de promoción y prevención. 4) El tamaño de clasificación debe analizarse de forma integral con otras variables, las cuales pueden depender en mayor o menor grado del comportamiento del entorno de trabajo y de su interpretación por parte de los equipos de seguridad y salud en el trabajo de los colegios públicos; de ésta manera, en estudios posteriores se podría robustecer la metodología incluyendo factores que provengan de la dinámica organizacional y de sistemas de vigilancia epidemiológica, al igual que incluyendo elementos tomados de las técnicas inteligentes. 5) La aplicación de la metodología predictiva con árboles de decisión J48, a los factores de riesgo psicosocial de instituciones educativa públicas, muestra que, se abre una línea de trabajo y de investigación que permite apoyar la labor de los grupos encargados de la seguridad y salud en el trabajo y facilita la labor de la identificación, prevención y control de los riesgos psicosociales en los entornos de trabajo. 6) La metodología propuesta está diseñada de acuerdo a las características de los colegios públicos en Colombia, lo cual permite que pueda ser aplicada en diferentes instituciones educativas cuya jornada sea continua (mañana, tarde), empleando diversas técnicas inteligentes, con el fin de obtener un mejor análisis sobre su verdadero impacto en la disminución en las tasas morbilidad y de ausentismo de los docentes.

## AGRADECIMIENTOS

Se expresan los más sinceros agradecimientos a la Universidad Nacional de Colombia –sede Manizales–, por su apoyo en la realización del proyecto que actualmente adelanta el autor del artículo, bajo la reglamentación definida por el programa de Doctorado en Ingeniería; Industria y Organizaciones perteneciente a la Facultad de Ingeniería y Arquitectura, cuyos resultados parciales se ilustran en el presente trabajo.

## REFERENCIAS

- Agarwal, B., y Mittal, N. Text classification using machine learning methods-a survey. In *Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012* (pp. 701-709). Springer India (2014)
- Anon, D. Búsqueda exhaustiva (En línea: <http://dis.um.es/~domingo/apuntes/AlgBio/1213/exhaustiva.pdf>, acceso: 2 de Agosto 2015), Universidad de Murcia, España (2016)
- Arias-González, M. Integración de los Sistemas de Gestión de Calidad, el Medio Ambiente y la Seguridad y Salud del Trabajo. *Ciencias Holguín*, 20(2), 39-48. (2014)
- Blanco Amado, R. M. Base de datos y minería de datos en contornos empresariales (En línea: <http://docplayer.es/1910803-Base-de-datos-y-mineria-de-datos-en-contornos-empresariales.html>, acceso: 5 de Agosto de 2015) Universidad de la Coruña, España (2014)
- BNQ, CSA Group and MHCC. Psychological health and safety in the workplace – prevention, promotion, and guidance to staged implementation (CAN/CSAZ1003-13/BNQ 9700-803/2013). (En línea: <http://shop.csa.ca/en/restofworld/occupational-health-and-safety-management/canca-z1003-13bnq-9700-8032013/inv/z10032013>, acceso: 2 de Agosto 2015), Ottawa, Ontario: Standards Council of Canada (2013)
- British Standards Institution (BSI). 2011. PAS1010: guidance on the management of psychosocial risks in the workplace (En línea: <http://shop.bsigroup.com/en/Navigate-by/PAS/PAS-10102010/?rdt=wmt>, acceso 5 de agosto 2015), BSI, London, UK (2011)

- Caamaño, A. J., Echeverría, M. M., Retamal, V. O., Navarro, C. T., y Espinosa, F. T. Modelo predictivo de fuga de clientes utilizando minería de datos para una empresa de telecomunicaciones en Chile. *Universidad Ciencia y Tecnología*, 18(72) (2015)
- Castro, A., Sifuentes, E., González, S., y Rascón, L. H. Uso de Minería de Datos en el manejo de Información Geográfica. *Información tecnológica*, 25(5), 95-102. (2014)
- Castrillón, O.D. Combinación entre Algoritmos Genéticos y Aleatorios para la Programación de Horarios de Clases basado en Ritmos Cognitivos. *Información tecnológica*, 25(4) (2014)
- Coca, G.A, Castrillón, Omar D, y Ruiz, Santiago. Programación Multiobjetivo en un Sistema de Fabricación Tipo "Contra Existencia" (JOB SHOP). *Información tecnológica*, 24(6), 79-92 (2013)
- Corso, C. L. Alternativa de herramienta libre para la implementación de aprendizaje automático. (En línea: [http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos\\_labsis/cynthia/Alternativa\\_de\\_herramienta\\_para\\_Mineria\\_Datos\\_CNEISI\\_2009.pdf](http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/Alternativa_de_herramienta_para_Mineria_Datos_CNEISI_2009.pdf) , acceso: 10 de Agosto de 2015), Argentina (2009)
- D. A. García, Algoritmo de discretización de series de tiempo basado en entropía y su aplicación en datos colposcópicos. Tesis de Maestría en Inteligencia Artificial. Universidad Veracruzana. México. (En línea: <http://cdigital.uv.mx/bitstream/123456789/32352/1/garcialopezdaniel.pdf>, acceso: 10 de Agosto de 2015 (2007)
- EU-OSHA – European Agency for Safety and Health at Work, 'Drivers and Barriers for Psychosocial Risk Management: An analysis of findings of the European survey of enterprises on new and emerging risks', (En línea: <https://osha.europa.eu/en/tools-and-publications/publications/reports/drivers-barriers-psycho-social-risk-management-esener> , acceso: 15 de Agosto de 2015 ), Publications Office of the European Union, Luxembourg (2012)
- Espinoza Cubillos, L. M. Prevalencia de riesgo psicosocial en un grupo de docentes y directivos del Distrito Capital. Tesis de Maestría. Universidad del Rosario, Bogotá, Colombia (2015)
- Eurofound, E. O. Psychosocial Risks in Europe: Prevalence and Strategies for Prevention (En línea: <https://osha.europa.eu/en/publications/reports/psychosocial-risks-eu-prevalence-strategies-prevention/view> , acceso: 5 de Enero de 2016), Publications Office of the European Union, Luxembourg, (2014)
- Frawley, W., Piatetsky, G.S., Matheus, C. Knowledge discovery in databases: an overview. *AI Magazine Fall*, 213e228, (1992)
- Hernández, O. V., y Miramontes, B. L. M. The relation of empowerment with organizational commitment influenced by human capital and high involvement practices of employees in automotive assembly plants. In *Global Conference on Business & Finance Proceedings* (Vol. 11, No. 1, p. 855). Institute for Business & Finance Research, (2016)
- Hepner, G. F. Artificial neural network classification using a minimal training set. Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, 56(4), 469-473 (1990)
- Jiménez, M. G., y Álvarez, A. Análisis de datos en WEKA–pruebas de selectividad. (En línea: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf> , acceso: 5 de Enero de 2016), Universidad Carlos III, España (2010).
- Karasek Jr, R. A. Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly*, 285-308 . (1979)
- Larrañaga, P., Inza, I., y Moujahid, A. Tema 6. Clasificadores Bayesianos. Departamento de Ciencias de la Computación e Inteligencia Artificial (En línea: <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t6bayesianos.pdf> acceso: 9 de Enero de 2016), Universidad del País Vasco-Euskal Herriko Unibertsitatea, España (1997)
- Leka, S., Jain, A., Cox, T., y Kortum, E. The development of the European framework for psychosocial risk management: PRIMA-EF. *Journal of occupational health*, 53(2), 137-143 (2011)
- Leka, S., Jain, A., Hassard, J. and Cox, T. Managing psychosocial risks: Drivers and barriers Occupational Health & Safety Wikipedia, European Agency for Safety & Health at Work. (En línea: [https://oshwiki.eu/wiki/Managing\\_psychosocial\\_risks:\\_Drivers\\_and\\_barriers](https://oshwiki.eu/wiki/Managing_psychosocial_risks:_Drivers_and_barriers) acceso: 10 de Enero de 2016) (2013)

- Leka, S., Van Wassenhove, W., y Jain, A. Is psychosocial risk prevention possible? Deconstructing common presumptions. *Safety science*, 71, 61-67 (2015)
- Huerta, L. Ruiz, J., Cabrera, N., Montiel, L., Benítez, F., y Ramírez, V. Data Mining: Impact of Daily Activities on Student Performance. *International Journal of Innovation and Applied Studies*, 14(4), 927-935 (2016)
- Mark Hall y otros 5 autores; *The WEKA Data Mining Software: An Update*; SIGKDD Explorations, Volume 11, Issue 1 (2009)
- Mamani Rodríguez, Z. E. Aplicación de la minería de datos distribuida usando algoritmo de clustering k-means para mejorar la calidad de servicios de las organizaciones modernas caso: Poder judicial. Tesis de Maestría. Universidad Nacional Mayor de San Marcos. Programa Cybertesis PERÚ (En línea: <http://cybertesis.unmsm.edu.pe/handle/cybertesis/4472> acceso: 12 de Enero de 2016). Lima, Perú (2015)
- Matos, G., Chalmeta, R., y Coltell, O. Metodología para la extracción del conocimiento empresarial a partir de los datos. *Información tecnológica*, 17(2), 81-88. (2006)
- Ministerio de la Protección Social, Resolución 2646 Julio 2008 (En línea: <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=31607> acceso: 12 de Enero de 2016), República de Colombia, (2008)
- Ricardo, L. La participación del educador en el cambio educativo. *InterCambios: Dilemas y transiciones de la Educación Superior*, 2(2), (2016)
- Saizarbitoria, I. H., Vilamitjana, M. B., y Fa, M. C. La integración de sistemas de gestión basados en estándares internacionales: resultados de un estudio empírico realizado en la Capv. *Revista de dirección y administración de empresas*, 1(14), 155-174. (2014).
- Schill, A. L., y Chosewood, L. C. The NIOSH Total Worker Health™ program: an overview. *Journal of Occupational and Environmental Medicine*, 55, S8-S11. (2013)
- Siegrist, J. Adverse health effects of high-effort/low-reward conditions. *Journal of occupational health psychology*, 1(1), 27 (1996)
- Theorell, T. The demand-control-support model for studying health in relation to the work environment: an interactive model. *Behavioral medicine approaches to cardiovascular disease prevention*, 69-85 (1996)
- Villalobos G. Diseño de un sistema de vigilancia epidemiológica de factores de riesgo psicosocial en el trabajo. Tesis doctoral, Doctorado en Ciencias de la Salud, Escuela Nacional de Salud Pública, La Habana, Cuba (2005)
- Quinlan, J. R. C4. 5: programs for machine learning. Elsevier, Burlington, Canada. (1993)