



Minería sobre streams de datos: una nueva estrategia para afianzar las relaciones de confianza en el mundo empresarial

Data stream mining: a new strategy to strengthen relationships of trust in the business world

Sonia JARAMILLO-VALBUENA [1](#); Esperanza ESPITIA [2](#) Sergio Augusto CARDONA [3](#)

Recibido: 17/11/2017 • Aprobado: 18/12/2017

Contenido

[1. Introducción](#)

[2. Las relaciones de confianza en el mundo empresarial](#)

[3. Algoritmos de minería sobre streams de datos](#)

[4. Técnicas de minería sobre streams de datos para afianzar las relaciones de confianza en el mundo empresarial](#)

[5. Conclusiones](#)

[Referencias bibliográficas](#)

RESUMEN:

Las tecnologías de la información y el *Big Data*, están generando una inmensa transformación en el mundo empresarial, dando lugar lo que se conoce como Nueva Economía o Economía Digital. El Big Data se define en torno a 4 dimensiones, a saber: volumen, velocidad, variedad y veracidad, y puede verse como un conjunto de datos de gran volumen, que es dinámico, complejo y en donde la información puede ser obtenida desde diversas fuentes y en variedad de formatos. Las fuentes del Big Data incluyen, entre otros, datos provenientes de aplicaciones empresariales. El procesamiento de esta información requiere la aplicación de técnicas que permitan descubrir o extraer conocimiento, para que pueda ser utilizado en la toma de decisiones con el objetivo de hallar nuevas oportunidades de negocio (al identificar tendencias de los clientes), aumentar las ganancias, mejorar el resultado de campañas publicitarias, realizar operaciones en un menor tiempo y lograr clientes más satisfechos. El objetivo de este artículo es revisar el estado del arte en lo referente a algoritmos de minería sobre streams de datos. Se describen las técnicas más representativas y se analiza cómo mediante su uso, es posible afianzar las relaciones efectivas de confianza en el mundo empresarial.

Palabras-Clave: clasificación, clustering, grandes volúmenes de datos, patrones, análisis de sentimientos

ABSTRACT:

Information technologies and Big Data are generating an immense transformation in the business world, giving rise to what is known as New Economy or Digital Economy. Big Data is defined around 4 dimensions, namely: volume, speed, variety and veracity, and can be seen as a set of high volume data, which is dynamic, complex and where information can be obtained from various sources and in a variety of formats. The sources of Big Data include, among others, data from business applications. The processing of this information requires the application of techniques that allow discovering or extracting knowledge, so that it can be used in decision-making in order to find new business opportunities (when identifying customer trends), increase profits, improve the result of advertising campaigns, perform operations in a shorter time and achieve more satisfied customers. The objective of this article is to review the state of the art in relation to data mining algorithms on streams. The most representative techniques are described and it is analyzed how through its use, it is possible to strengthen the effective relationships of trust in the business world.

Keywords: Big Data, classification, clustering, patterns, Sentiment Analysis

1. Introducción

En (Moss, 2005) se define la confianza como expectativa de éxito que produce motivación, voluntad, esfuerzo y energía, y que está respaldada por un sistema que genera responsabilidad, colaboración e iniciativa. Es un lazo fuerte y perdurable que mantiene acoplada a la organización y que se fundamenta en la credibilidad, la imparcialidad y el respeto. La confianza, en el mundo empresarial, es un activo que cimienta el crecimiento

económico y la estabilidad social.

Las empresas, para generar confianza, deben responder a las necesidades cambiantes de la sociedad del conocimiento; una sociedad que emerge a partir de gran cantidad de innovaciones tecnológicas, en lo referente a telecomunicaciones, informática y medios de comunicación. En la actualidad, las empresas se enfrentan a un exceso de información nunca antes visto y por ende, deben asumir un rol de liderazgo al demostrar su capacidad para producir y utilizar la información (Cebrián, 2009). Este exceso de información se debe en gran medida a que la tecnología permite producir fácilmente contenido en Internet (Gabelas, 2002).

Procesar apropiadamente grandes enormes volúmenes de datos le permite a la empresa conocer a sus empleados y clientes, identificar gustos y tendencias. Lo anterior redundará en un aumento significativo de su productividad, al reducir el margen de error al momento de ofrecer servicios, productos o implementar nuevas políticas y permite llevar la delantera a la hora de identificar mercados potenciales.

El objetivo de este artículo es mostrar como el Big Data, y en especial, la aplicación de técnicas de minería sobre streams de datos puede ser usado por la empresa para fortalecer las relaciones de confianza. En la sección 1 se describe la importancia de las relaciones de confianza y credibilidad en el mundo empresarial. La sección 2 describe el estado del arte en lo referente a algoritmos de minería sobre streams de datos. La sección 3 muestra como la aplicación de técnicas de Minería sobre streams de datos puede utilizarse para generar relaciones de confianza. Finalmente, la sección 4 presenta las conclusiones.

2. Las relaciones de confianza en el mundo empresarial

El medio competitivo actual exige que las organizaciones se comprometan con el modelamiento del entorno, que expresen con claridad (a través de obrar), su capacidad para intervenir y transformar. De aquí surge la necesidad de construir un nuevo modelo de gestión fundamentado en generación de reputación y confianza.

La confianza es el resultado de la responsabilidad y de generar valor agregado. Por ello la organización debe alinear su planeación estratégica, el liderazgo de los ejecutivos y la congruencia filosófica para generar confianza. Esta es una práctica única en cada empresa, que debe incorporarse a la forma de actuar de cada uno de los integrantes de la organización. La confianza debe ser percibida por los empleados, los clientes y cualquier otro grupo organizado que tenga relación con la empresa.

La confianza aporta a la toma de decisiones y es parte integral del clima organizacional. Además, propicia espacios de autonomía para que el empleado sienta que puede tomar la iniciativa y logre desarrollar de forma responsable su trabajo, esto supone el apoyo para operar y modelar el resto de la organización.

Una organización que tiene una buena reputación, y por ende genera confianza, se mueve en varias dimensiones: el buen gobierno, el liderazgo, los productos, los servicios, la innovación, el entorno de trabajo, los resultados, la población y el medio ambiente.

La sociedad y el mercado responden efectivamente ante aquellas empresas, que generan confianza y esto se logra mediante la implementación de políticas, operaciones, prácticas y productos socialmente responsables (Ministerio de Educación de Colombia, 2006).

3. Algoritmos de minería sobre streams de datos

Los dos principales modelos para trabajar sobre streams de datos son el evolutivo y el de ventana deslizante. El primero ellos considera todo el histórico de los datos. El segundo, y más utilizado, hace uso de ventanas deslizantes, que permite considerar los datos más recientes. Una ventana deslizante, se comporta de forma similar a una cola (Primero en entrar primero en salir), almacenando y procesando únicamente el sub string definido por el tamaño de la ventana (Jaramillo, Cardona & Fernandez, 2015). El Algoritmo 1 presenta el código básico para la construcción de una ventana deslizante.

Algoritmo básico de ventana deslizante

Entrada:

S: Un stream de ejemplos

W: ventana de ejemplos

Salida:

C: un clasificador construido en los datos de la ventana **W**

1. Inicializar la ventana W;

2. Para todos los ejemplos $x_i \in S$ hacer

3. WB $W \cup \{ x_i \}$;

4. Si es necesario remover ejemplos desactualizados desde W

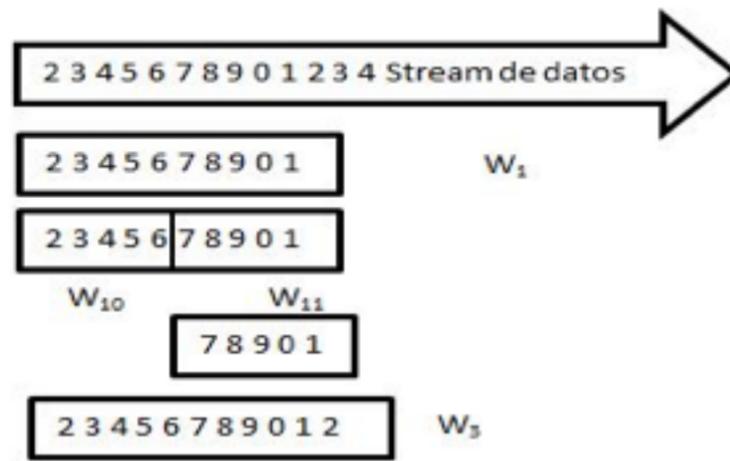
5. reconstruir/actualizar C usando W

Las ventanas de tamaño fijo en algunos casos pueden no ser óptimas para la exactitud del modelo, ya que no necesariamente el concept-drift se va a alinear con los límites de la ventana (Gehrke, Korn y Srivastava, 2001).

Algunas aproximaciones proponen que el tamaño de la ventana puede ser ajustado dinámicamente, ver Figura 2, creciendo si el modelo gana exactitud o decreciendo cuando la pierde. ADWIN (Bifet y Gavaldá, 2009) es una aproximación de este tipo, que hace uso del límite de Hoeffding. ADWIN es una estrategia de ventana deslizante adaptativa libre de parámetros que compara todas las sub-ventanas adyacentes a una ventana de partición que contiene todos los datos. Este método es mucho más preciso pero es más costoso a nivel de ejecución en grandes *streams* de datos. En ADWIN se mantienen una ventana con todas las instancias que llegan desde que se detecta el cambio.

Figura 1

Ventana Deslizante adaptativa (Finlay, Pears y Connor, 2014)



En lo referente a ajustar los algoritmos de minería para trabajar sobre streams de datos hay varias aproximaciones. En la Tabla 1 se muestran algunos de los algoritmos de clasificación existente para este tipo de ambientes.

Tabla 1

Algoritmos de clasificación en ambiente *streaming*
(The University of Waikato, 2014) (Jaramillo, Cardona & Fernandez, 2015).

Algoritmo	Descripción	Evolutivo	ADWIN
Very Fast Decision Trees - VFDT	Implementación del árbol de Hoeffding. Este algoritmo incremental induce árboles de decisión para <i>streams</i> de datos. La idea tras este algoritmo es que es posible elegir un atributo de particionamiento óptimo solo usando un pequeño número de ejemplos. El árbol producido es prácticamente idéntico al producido por la versión batch de este algoritmo (Hulten, Spencer y Domingos, 2001).	-	-
Concept- adapting Very Fast Decision Trees -CVFDT	Genera un árbol de decisión desde <i>streams</i> de datos que cambian en el tiempo a grandes velocidades. Este método de clasificación trabaja el concept-drift, pero no funciona bien con atributos (Hulten, Spencer y Domingos, 2001)..	-	-
HoeffdingAdaptiveTree	Este algoritmo puede tratar el concept drift. Se basa en el uso de detectores de cambio y módulos estimadores. No requiere conjeturas previas sobre la rapidez o frecuencia en la que el <i>stream</i> presentará drift (Bifet y Gavaldá, 2009).	X	X
NaiveBayes	Este clasificador predice para cada instancia no etiquetada I la clase C a la que pertenece con gran precisión. Dicha predicción es realizada asumiendo entradas independientes (The University of Waikato, 2014).	-	-
OzaBag (bagging)	Versión Incremental de bagging para <i>streams</i> de datos. El mecanismo de bagging para <i>streams</i> genera conjuntos de clasificadores a partir de una evidencia. En este método se	-	-

	forma aleatoriamente y con reemplazamiento una muestra de m ejemplos de entrenamiento del conjunto inicial de entrenamiento compuesto por m instancias (Orallo, Ramírez y Ferri, 2005). La probabilidad de que un ejemplo sea elegido para replicación tiende a la distribución de Poisson. Cada subconjunto de entrenamiento aprende un modelo. Para clasificar un ejemplo se predice la clase de ese ejemplo para cada clasificador y se clasifica la clase con voto mayoritario (Oza y Russell, 2001)		
OzaBoost	<p>Versión incremental de boosting para <i>streams</i> de datos. El mecanismo de boosting asigna un peso a cada conjunto de entrenamiento. Cada vez que se realiza una iteración, el modelo aprendido minimiza la suma de los pesos de aquellos ejemplos clasificados incorrectamente. Partiendo de los errores obtenidos en cada iteración se actualizan los pesos del conjunto de entrenamiento, dándole más peso a los mal clasificados, que a aquellos clasificados correctamente (Oza y Russell, 2001)</p> <p>En OzaBoost, el procedimiento de pesado divide el peso total del ejemplo en dos partes – una mitad peso se asigna a los ejemplos correctamente clasificados, y la otra mitad a los ejemplos clasificados incorrectamente. Este método utiliza la distribución de Poisson para decidir la probabilidad aleatoria de que un ejemplo sea usado para el entrenamiento.</p>	-	-
Adaptive-Size Hoeffding Tree (ASHT) con Bagging-OzaBagASHT	Bagging utilizando árboles de diferentes tamaños. El ASHT se deriva mediante el uso de Árboles de Hoeffding (Bifet et al., 2009).	X	-
OzaBagADWIN	Bagging que utiliza ADWIN como detector de cambio (Bifet y Gavalda, 2009).	X	X
Perceptron	Clasificador perceptron que ejecuta incremental aprendizaje multiclase.	X	-
SGD	Implementa gradiente descendiente para los modelos de aprendizaje: <i>Máquinas de soporte vectorial</i> para clasificación binaria, <i>regresión lineal</i> y <i>logística binaria</i> .	X	-

En relación a métodos de clustering sobre *streams* de datos, es decir, a agrupación de un conjunto de datos en subconjuntos en los que los miembros de un mismo clúster comparten propiedades comunes, la Tabla 2 muestra algunos de los algoritmos más relevantes.

Tabla 2
Clustering en *streams* (The university of Waikato, 2014) (Jaramillo, Cardona & Fernandez, 2015)

Algoritmo	Técnica utilizada	Clúster de forma arbitraria	Valores atípicos	drift
StreamKM++	Usa k-means++ algoritmo como una técnica semilla para seleccionar los primeros valores del clúster. Para calcular una muestra pequeña utiliza una estructura de datos llamada coresets tree (Ackermann et al., 2010).	-	-	-
CluStream	Mediante el uso de micro-clústeres se mantiene información estadística de los datos. Los micro clústeres, extensiones temporales de vectores de características de clústeres, son almacenados como	-	X	-

	captura instantánea en el tiempo, siguiendo un patrón piramidal que permite recordar las estadísticas de resumen de diferentes horizontes de tiempo (Aggarwal et al., 2003).			
ClusTree	Utiliza una estructura indexada compacta y auto adaptativa para mantener resúmenes de datos. Se basa en densidad (Kranen et al., 2011).	-	X	X
DenStream	Utiliza 2 estructuras para resumir los clústeres y manejar la información de valores atípicos, a saber: micro-clústeres densos y micro-clúster de valores atípicos (Cao et al., 2006)	X	X	-
D-Stream	Este método mapea cada entrada en una grid y calcula la densidad de la grid. Las grids son agrupadas basadas en la densidad. Este algoritmo adopta la técnica de descomposición de densidad para capturar los cambios dinámicos de un flujo de datos (Chen y Tu, 2007).	X	X	X
CobWeb	Usa un árbol de clasificación. Cada nodo en el árbol representa un concepto y se etiqueta haciendo uso de medidas probabilísticas que resumen las distribuciones atributo-valor de los objetos clasificados bajo dicho nodo.	-	-	-

La Tabla 3 presenta la clasificación de algunos de los trabajos realizados con base en el modelo de cómputo y los algoritmos de extracción de conocimiento.

Tabla 3
Clasificación de algunos de los trabajos realizados sobre minería de datos en redes sociales (Jaramillo, Cardona & Fernandez, 2015)

Referencia	Contexto	Contribución	Modelo de computo	Algoritmo de minería/ extracción	Conclusión
(Bifet y Frank, 2010)	Análisis de sentimientos de streams provenientes de Twitter.	Presenta Kappa como una medida para performance del clasificador en casos en los cuales se hay desbalanceo de clases	Ventana deslizante de tamaño 1000	-Multinomial Nàive Bayes, -Stochastic gradient descent, -Hoeffding tree	De los 3 algoritmos evaluados recomiendan el algoritmo de Gradiente descendiente estocástico (SGD). Además, se indica que Kappa debe ser la medida utilizada, en la ventana deslizante, para medir el performance del clasificador
(Balasubramanyan, Routledge y Smith, 2010)	Correlacionan las observaciones expresadas en encuestas de opinión pública	---	Una media móvil a través de una ventana de los últimos días	Dentro de mensajes tópicos, se cuentan los mensajes que contienen	Se detectan altas correlaciones (aproximadamente 80%, dependiendo del dataset) y se capturan

	con los resultados del análisis de los sentimientos efectuado sobre textos de Twitter		k, que permite realizar análisis de cambio opinión en el tiempo	palabras positivas y negativas en el contexto de una palabra clave	tendencias a gran escala.
(Aston et al. 2014)	Análisis de sentimientos. Se evalúa la exactitud del clasificador	Habilidad para detectar dinámicamente características importantes para análisis de sentimientos	Intervalos de tiempo de 100 publicaciones secuenciales	-Winnow Balanceado Modificado (MBW) en conjunción con MOA	La selección de características contribuye de forma independiente a la precisión de la predicción sentimiento.
(AT&T Labs, 2013)	Un motor de búsqueda visual en línea para minería de texto sobre streams en tiempo real	-Posibilidad ver gráficamente el stream de tweets. -Uso de métodos de clustering y análisis semántico para resumir y categorizar los tweets	Ventana dinámica	-Modelo de <i>Asignación de Dirichlet latente</i> (LDA) -Clustering modular	-Clasificación automática de tweets. - Identificación de los tópicos más populares y rastreo de la discusión de dichos tópicos de interés en el pasado. -Preservación de la coherencia espacio-temporal durante el ajuste de la ventana
(Wood, Zheludev y Treleven, 2014)	Minería sobre datos sociales, permite identificación de tendencias de cambio, sentimientos globales y propagación de historias	Una plataforma capaz de monitorear noticias y opiniones desde las redes más populares	-	-	Plataforma aún en desarrollo. Se espera que soporte análisis de opinión, integración con ATRADE para poder evaluar simultáneamente datos financieros y provenientes de redes sociales.
(Popovici, Weiler y Grossniklaus, 2014)	Clustering on-line para detección de tópicos en tiempo real en redes sociales	Nuevo algoritmo de clustering en línea capaz de procesar de forma rápida e incremental los streams y de detectar los cambios en los modelos de clustering	Modelo de ventana	-DenStream extendido como técnica de poda -DBSCAN en fase de macroclustering	Detección de tópicos y asociación de imágenes relacionadas con el tópico.

(Bifet et al., 2011)	Detección de cambios de sentimiento en streams de Tweets	Una extensión de MOA llamada TweetReader	ADWIN (ADaptive sliding WINDOW)	-Hoeffding Tree	Un software que ayuda a comprender lo que ocurre en un momento dado en el mundo, mediante el análisis de sentimientos
-----------------------------	--	--	---------------------------------	-----------------	---

4. Técnicas de minería sobre streams de datos para afianzar las relaciones de confianza en el mundo empresarial

La aplicación de técnicas de Big Data se consolida como una importante alternativa para generar confianza en el ámbito empresarial. La aplicación de técnicas para descubrir o extraer conocimiento de forma semiautomática a partir de los datos objeto de análisis de la empresa, puede ser utilizada en la toma de decisiones, y por ende, también para generar relaciones de confianza (Alasadi & Bhaya, 2017).

La Minería de Datos abre nuevas oportunidades de negocios, al contribuir a la toma de decisiones tácticas y estratégicas. Esto se debe, a que los modelos de Minería de Datos pueden predecir valores, buscar relaciones, dependencias y generar resúmenes de datos (Larose, 2005). El tipo de conocimiento a extraer puede ser predictivo o descriptivo (Valcárcel, 2004). El conocimiento descriptivo hace referencia a relaciones entre las variables. Por su parte el conocimiento predictivo permite predecir sucesos futuros partiendo del modelo que define al sistema (Shi-Nash & Hardoon, 2017).

La minería de datos permite clasificar o agrupar (empleados o clientes) de acuerdo a sus características, para brindar asesoría o seguimiento, realizar adaptaciones y personalizaciones (ofrecer productos de acuerdo a las necesidades del cliente o del empleado), monitorizar la realización de actividades (para lo cual podrían tenerse en cuenta por ejemplo datos de interacción como ofertas, clics, quejas, reclamos), analizar datos actitudinales (opiniones, necesidades y preferencias), buscar patrones de comportamiento generales e inusuales (detectar riesgo psicosocial, fraude y comportamientos de compra, entre otros). La detección de patrones y descubrimiento, se refiere a una colección de enfoques cuyo objetivo es identificar estructuras inusuales o relaciones en los datos. Algunos ejemplos al respecto son búsqueda de objetos de baja probabilidad, búsqueda de formas en series de tiempo, estructuras en secuencias y reglas de asociación. El análisis de estas últimas se utiliza en escenarios tales como la cesta de mercado y filtrado colaborativo.

Otra importante aplicación de la minería de datos para generar relaciones de confianza es el análisis de sentimientos. Mediante el análisis de sentimientos es posible comprender la actitud que tiene la persona que escribe sobre una temática o la polaridad contextual (positiva, negativa o neutra) que expresa. Se considera sentimiento a un estado de ánimo junto con una emoción conceptualizada, es decir, un subtipo de actitud duradera, una impresión afectiva, preferencia o predisposición, que provoca determinado objeto, persona o situación (Scherer & Ceschi, 1997) (Castilla, 2000). En (Maas, et al., 2011) (Potts, 2011) se afirma que las teorías cognitivas de la emoción son una importante herramienta para identificar semejanzas o contraste entre los sentimientos expresados en un texto, esto contribuye a identificar opiniones opuestas y similares en este tipo de sistemas. La Figura 2 muestra las 8 emociones básicas expuestas en (Robert Plutchik, 2002) y su relación con los sentimientos.

Figura 2

En a) se observa el contraste de emociones según (Plutchik, 1980).
En b) El contraste de sentimientos (Dubiau & Ale, 2013)

a.	Emoción Básica	Emoción Opuesta	b.	Emociones Básicas	Sentimiento	Sentimiento Opuesto
	Alegría	Tristeza		Anticipación + Alegría	Optimismo	Desaprobación
	Confianza	Disgusto		Alegría + Confianza	Amor	Remordimiento
	Miedo	Ira		Confianza + Miedo	Sumisión	Desprecio
	Sorpresa	Anticipación		Miedo + Sorpresa	Temor	Agresividad
				Sorpresa + Tristeza	Desaprobación	Optimismo
				Tristeza + Disgusto	Remordimiento	Amor
				Disgusto + Ira	Desprecio	Sumisión
				Ira + Anticipación	Agresividad	Temor

La identificación de sentimientos es una tarea compleja, puesto que requiere analizar aspectos entre los que se destacan: morfología, sinonimia, ortografía, el tratamiento de la negación, el análisis a nivel de aspecto, la adaptación al dominio y además, analizar textos en los que se expresa ironía y sarcasmo (Liu, 2012) (Pang & Lee, 2008). Además, según (Manning & Jurafsky, 2012) para el análisis computacional de sentimientos se debe considerar: detectar las emociones, determinar quién las genera, determinar la razón o el motivo generador de la emoción, determinar la polaridad de la emoción o su tipo (me encanta, lo odio, me gusta) y hallar las sentencias en las que se encuentran dichas emociones.

Entre la información que puede obtenerse empleando Análisis de sentimientos se tiene: nivel de fidelización de clientes, polaridad de sentimientos (en lo referente a críticas sobre productos, servicios, entre otros), tendencias de mercado y predicciones sobre comportamientos tanto de clientes como empleados.

El análisis de sentimientos puede ser aplicado en las empresas, por ejemplo para analizar el riesgo psicosocial de sus empleados. Esto puede ayudar a generar recomendaciones que permitan evitar problemas referentes a la salud mental de los empleados, ver Tabla 4, y disminuir el estrés laboral.

Tabla 4
Síntomas que puede determinar que el empleado está en riesgo psicosocial (Cano, 2002) (Duque, 2016).

A nivel cognitivo-subjetivo	A nivel fisiológico	A nivel motor u observable
Preocupación, temor, inseguridad, dificultad para decidir, miedo, pensamientos negativos sobre uno mismo, pensamientos negativos como se actúa, temor a la pérdida del control, dificultades para pensar, estudiar, o concentrarse	Sudoración, tensión muscular, palpitations, taquicardia, temblor, molestias en el estómago, molestias gástricas, dificultades respiratorias, sequedad de boca, dificultades para tragar, dolores de cabeza, mareo, náuseas, tiritar	Evitar situaciones temidas, fumar, comer o beber en exceso, intranquilidad motora (movimientos repetitivos, rascarse, tocarse, etc.), ir de un lado para otro sin una finalidad concreta, tartamudear, llorar, quedarse paralizado.

El hecho de que un empleado sienta confianza redonda en una mayor productividad y un mejor clima organizacional. También puede ser utilizado, para identificar si nuevos productos son de interés o no para el cliente.

5. Conclusiones

La Red ha hecho que las relaciones de proximidad, al igual que las identidades de los individuos cambien. La empresa debe adaptarse a esta nueva circunstancia y valerse de las técnicas de Big Data para valorar y sopesar la información con la que cuenta, para poder tomar decisiones acertadas que beneficien a toda la empresa. La aplicación de técnicas de minería de datos aporta conocimiento para la toma de decisiones tácticas y estratégicas, al permitir realizar predicciones, identificar relaciones y detectar dependencias. La empresa tiene la obligación moral de anticipar el impacto de las acciones que realice con la información que recibe y manipula. La ética siempre debe estar presente durante el tratamiento de la información.

Lograr que se den relaciones efectivas de confianza en el mundo empresarial es un proceso que requiere la participación de todas las instancias de actuación, y la incorporación y apropiación de una nueva sociedad de la información. Además, implica la construcción de estrategias desde lo público y lo privado. Un mercado eficiente que vaya de la mano con la confianza reduce el costo del capital y contribuye al crecimiento económico.

Referencias bibliográficas

- ALASADI, S., & BHAYA, W. (2017). **Review of data preprocessing techniques in data mining**. Journal of Engineering and Applied Sciences, 12(16), 4102-4107. Obtenido de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029819648&doi=10.3923%2fjeasci.2017.4102.4107&partnerID=40&md5=f0d04fb63c552259aeb4f1a3a5a95bba>
- CASTILLA, C. (2000). **Teoría de los sentimientos**. Tusquets Editores.
- CEBRIÁN, J. L. (31 de mayo de 2009). **Gacetilleros, gansos y embaucadores**. Diario El País.
- DUBIAU, L., & ALE, J. M. (2013). **Análisis de Sentimientos sobre un Corpus en Español**. 14th Argentine Symposium on Artificial Intelligence, ASAI 2013.
- DUQUE MARÍN, NORMA CONSTANZA (2016). **Análisis de riesgos psicosociales y estrés en trabajadores de selección y operativos de Digitex de Manizales**. Proyecto de desarrollo para optar al título de psicólogo,

UNAD.

GABELAS, J. A. (2002). **Las TIC en la educación. Una perspectiva desmitificadora y práctica sobre los entornos de aprendizaje generados por las nuevas tecnologías.** Obtenido de <http://www.uoc.edu/web/esp/art/uoc/gabelas0102/gabelas0102.html>

JARAMILLO, CARDONA & FERNANDEZ (2015). **Minería de datos sobre *streams* de redes sociales, una herramienta al servicio de la Bibliotecología.** Información, cultura y sociedad.

LAROSE, D. (2005). **Discovering Knowledge in Data: An Introduction to Data Mining.** Obtenido de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84983721794&doi=10.1002%2f0471687545&partnerID=40&md5=2f3c404fe092f5c39cf4d09801e18a3b>

LIU, B. (2012). **Sentiment Analysis and Opinion Mining.** Morgan & Claypool Publishers.

MAAS, A. L., DALY, R. E., PHAM, P. T., HUANG, D., NG, A. Y., & POTTS, C. (2011). **Learning Word Vectors for Sentiment Analysis.** (págs. 142-150). Stroudsburg, PA, USA: Association for Computational Linguistics. Obtenido de <http://dl.acm.org/citation.cfm?id=2002472.2002491>

MANNING, C., & JURAFSKY, D. (2012). **NLP Course at Stanford from Chris Manning and Dan Jurafsky.** Obtenido de <https://aclweb.org/portal/content/free-online-nlp-course-stanford-chris-manning-and-dan-jurafsky>

MINISTERIO DE EDUCACIÓN DE COLOMBIA. (2006). **Responsabilidad social empresarial.** Obtenido de <http://www.mineducacion.gov.co/cvn/1665/article-93439.html>

MOSS, R. (2005). **Harvard Business Review.** Harvard Business Review.

PANG, B., & LEE, L. (2008). **Opinion Mining and Sentiment Analysis.** Found. Trends Inf. Retr., 2(1-2), 1-135. Obtenido de <http://dx.doi.org/10.1561/1500000011>

PLUTCHIK, R. (1980). **Emotion: A Psychoevolutionary Synthesis.** New York. Harper & Row.

POTTS, C. (2011). **Sentiment Symposium Tutorial.** Obtenido de <http://sentiment.christopherpotts.net/>

RENGIFO, J. (2012). **Clasificación minería de datos, de acuerdo al conocimiento adquirido.** IBM, Colombia.

SCHERER, K. R., & CESCHI, G. (Sep de 1997). **Lost Luggage: A Field Study of Emotion--Antecedent Appraisal.** Motivation and Emotion, 21(3), 211-235. Obtenido de <https://doi.org/10.1023/A:1024498629430>

SHI-NASH, A., & HARDOON, D. (2017). **Data analytics and predictive analytics in the era of big data.** Obtenido de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019876521&doi=10.1002%2f9781119173601.ch19&partnerID=40&md5=f1ab8df553b2f61d9d919649635f7cd0>

VALCÁRCEL, V. (2004). **Data mining y el descubrimiento del conocimiento.** Industrial Data Revista de Investigación .

1. Doctora en Ingeniería. Computer Engineering Dept., Universidad del Quindío, UQ. Armenia, (Colombia), sjaramillo@uniquindio.edu.co

2. Magister en Ingeniería. Universidad del Quindío, UQ, Armenia, (Colombia), eespitia@uniquindio.edu.co

3. Doctor en Ingeniería. Computer Engineering Dept., Universidad del Quindío, UQ, Armenia, (Colombia), sergio_cardona@uniquindio.edu.co

Revista ESPACIOS. ISSN 0798 1015
Vol. 39 (Nº 13) Año 2018

[Index]

[En caso de encontrar un error en esta página notificar a [webmaster](#)]

©2018. revistaESPACIOS.com • @Derechos Reservados