

Análisis de las alternativas del diseño de un sistema difuso para la predicción del precio del Bitcoin (BTC).

Analysis of alternatives for the design of a fuzzy system for predicting the price of Bitcoin (BTC).

Jeison Ivan Roa Mora 

Universidad Distrital Francisco José de Caldas, Colombia

Open Access

Recibido:

10 de abril de 2018

Aceptado:

18 julio de 2018

Publicado:

11 diciembre de 2018

Correspondencia:

jroam@correo.udistrital.edu.co

DOI:

<https://doi.org/10.17081/invinno.7.1.2996>



© Copyright: Investigación e Innovación en Ingenierías

Resumen

Objetivo: Presentar tres alternativas para el diseño de un sistema difuso, con la intención de predecir la señal precio del Bitcoin (BTC) en dólares (USD). **Metodología:** Se realizó un análisis matemático y gráfico de la señal del precio del Bitcoin [BTC]. Los datos de entrada se han seleccionado como precios de cierre para cada periodo, donde el periodo es constante y de 30 minutos. Con esto se busca identificar posibles características de interés, patrones periódicos o presencia de ruido. **Resultados:** En la primera aproximación a la solución del problema, se construyó un sistema difuso con motor de inferencia Mamdani en el que los parámetros de la máquina se ajustan manualmente, basándose en los patrones encontrados en el estudio previo de la señal y en el conocimiento de un experto en el área. La segunda solución se generó mediante el algoritmo ANFIS, que realiza el ajuste automático de los parámetros empleando algoritmos de gradiente descendente; y la tercera solución se diseñó a través de la implementación de un algoritmo bioinspirado, conocido como algoritmo genético simple. **Conclusiones:** En los tres casos, son cinco las entradas del sistema, cada una de las cuales corresponde a muestras de la señal del precio; la salida, por su parte, es la predicción del precio para el periodo siguiente.

Palabras claves: Bitcoin, pronóstico, lógica difusa, sistema difuso, ANFIS, algoritmo evolutivo, algoritmo genético simple, Mamdani, Sugeno.

Abstract

Objective: To present three alternatives for the design of a diffuse system that has, as a main objective, the forecast of the Bitcoin (BTC) Price in USA dollars. **Methodology:** To perform a mathematical analysis and graphs of Bitcoin prices [BTC]. The entry prices have been selected as the closing prices for each period, where the period is a constant of 30 minutes, seeking to identify possible characteristics of interest, periodic patterns, and/or the presence of noise. **Results:** In the first approach to the problem's solution, a diffuse system is built with a Mamdani motor of inference where the machine parameters are manually adjusted, based on the patterns found in the previous study of the signal, and in the knowledge of an expert in the area. The second solution is made by using an ANFIS algorithm that carries out an automatic adjustment of parameters using diminishing gradient algorithms, and the third solution is designed through the implementation of a bio-inspired algorithm known as a simple genetic algorithm. **Conclusions:** In all three cases, the entries to the system are five, and correspond to price signals and the closing is the price forecast for the following period.

Keywords: Bitcoin, forecasting, fuzzy logic, fuzzy system, ANFIS, evolutionary algorithm, simple genetic algorithm, Mamdani, Sugeno.

Como citar (IEEE): J.Roa - Mora, "Análisis de las Alternativas del Diseño de un Sistema Difuso para la Predicción del Precio del Bitcoin (BTC)", *Revista Investigación e Innovación en Ingenierías*, vol. 7, n°. 1, 2019. DOI: <https://doi.org/10.17081/invinno.7.1.2996>

Introducción

El bitcoin es una criptomoneda lanzada e implementada a principios del año 2009. Su red está basada en el protocolo P2P, que garantiza su descentralización. El sistema de la criptomoneda hace uso de algoritmos de prueba de trabajo para confirmar las transacciones e impedir el doble gasto. Debido a la carencia de controles estatales o jurídicos en dicho mercado, el bitcoin presenta un comportamiento altamente volátil, y permite obtener importantes rentabilidades a corto plazo, incluso más sustanciales que las que se obtendrían con una estrategia de Buy & Hold, siempre que se logre hacer una predicción correcta de su tendencia y punto de inflexión. Esta moneda, como cualquier otra, tiene una equivalencia con las principales divisas: USD, EUR, COP, entre otras. Actualmente, la mayoría de compras y ventas de BTC se realizan en dólares [USD] [1]. En este contexto, el estudio de este problema se centra en comprender la señal como una serie de tiempo, lo que nos llevó finalmente a utilizar un Modelo Auto-Regresivo de Media Móvil (ARMA) para su predicción.

En cuanto a los sistemas difusos respecta, estos se entienden como una clase de objetos con continuos grados de pertenencia, que oscilan entre cero y uno. En tales conjuntos, como en cualquier otro, se presentan las nociones de inclusión, unión, intersección, complemento, relación, convexidad, etc., así como varias propiedades de estas nociones [2]. Por otra parte, un sistema difuso se basa en reglas heurísticas de la forma SI (antecedente) ENTONCES (consecuente), donde el antecedente y el consecuente son elementos pertenecientes a conjuntos difusos, ya sea puros, o como resultado de operar con ellos. En este documento se exponen los resultados de algunos sistemas difusos diseñados con el fin de predecir el estado de una señal determinada.

Marco teórico

Series de tiempo

Una serie de tiempo es una secuencia de un conjunto de datos, medida en un intervalo establecido de tiempo y ordenada en forma cronológica. Matemáticamente, se define como un conjunto de vectores. [3]

$$x(t); t = 0, 1, 2, \dots \quad (1)$$

Sistema difuso

En un sistema basado en lógica difusa, se transforman los datos o valores generalmente numéricos de la entrada al dominio de reglas intuitivas y lingüísticas, que permitan difuminar la naturaleza de la información. En dicho dominio se realizan las operaciones respectivas entre conjuntos, según una base de conocimiento previamente estructurada. Luego de esto

se realiza un proceso de defusificación en el que se devuelven los resultados del proceso de inferencia al dominio numérico o al que corresponda para hacerlos interpretables. Un sistema difuso se puede expresar en términos de sus conjuntos difusos de entrada y salida, número de las mismas, base de reglas y tipo de motor de inferencia.

ANFIS (Adaptative Network based Fuzzy Inference System)

ANFIS es un tipo de red neuronal artificial que se basa en el sistema de inferencia difusa Takagi-Sugeno, desarrollado a principios de la década de 1990. Dado que integra las redes neuronales y los principios de lógica difusa, tiene el potencial de capturar los beneficios de ambos en un marco único. Su sistema de inferencia corresponde a un conjunto de reglas difusas SI-ENTONCES, que tienen capacidad de aprendizaje para aproximar funciones no lineales. Por lo tanto, ANFIS se considera un estimador universal [4].

Algoritmo genético simple

El algoritmo genético simple se fundamenta en la teoría de la evolución biológica por selección natural definida por Charles Darwin. En términos muy generales, la selección natural es la manera en que una especie se adapta a su medio ambiente. Este proceso lleva al cambio evolucionario cuando individuos con ciertas características poseen una tasa de supervivencia o reproducción más alta que los otros individuos de la población, y pasan estas características genéticas heredables a su descendencia mediante la reproducción [5].

Con el transcurrir del tiempo, las poblaciones evolucionan adquiriendo características específicas en respuesta al medio ambiente, de este modo consiguen una mejor adaptación y, por ende, una mayor probabilidad de supervivencia.

El algoritmo genético simple propicia la evolución de una población de individuos, sometiéndola a acciones aleatorias análogas a las que actúan en la evolución biológica, tales como mutaciones y recombinaciones genéticas. Este algoritmo también somete a los individuos a una evaluación de acuerdo con una función objetivo ("Fitness Function") determinada por el problema a tratar, y con base en la cual se decide cuáles son los individuos más adaptados, que sobreviven y tienen una mayor tasa de reproducción, y cuáles los menos aptos, que son descartados. A este método de selección se le conoce como *Roulette Wheel*.

Estadísticos de evaluación de desempeño

A continuación se exponen los indicadores estadísticos seleccionados para realizar la evaluación del desempeño de los predictores. Se debe realizar una cuidadosa selección de los estadísticos (E), pues estos son los encargados de juzgar la predicción.

E1. MSE (Mean Square Error): El Error Cuadrático Medio (MSE) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. Está definido como:

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_{real}(i) - Y_{predicho}(i))^2 \quad (2)$$

E2. MAPE (Mean Average Percentage): El Error Porcentual Absoluto Medio (MAPE) es un indicador del desempeño del Pronóstico de Demanda, que mide el tamaño del error absoluto en términos porcentuales. Se define como:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{Y_{real}(i) - Y_{predicho}(i)}{Y_{predicho}(i)} \right| \quad (3)$$

E3. NMSE (Normalized Mean Square Error): El Error cuadrático medio normalizado (NMSE) es un estimador de las desviaciones generales entre los valores pronosticados y medidos.

$$NMSE = \frac{\sum_{i=1}^N Y_{real}(i) - Y_{predicho}(i)}{\sum_{i=1}^N |Y_{predicho}(i) - Y_{predicho}(i-1)|} \quad (4)$$

Metodología

Análisis del problema

Para llevar a cabo este proceso, en primer lugar, se realiza un análisis matemático y gráfico de la señal del precio del Bitcoin [BTC]. Los datos de entrada fueron seleccionados como los precios de cierre para cada periodo, donde el periodo es constante y de 30 minutos. De esta manera, se buscó identificar posibles características de interés, patrones periódicos o presencia de ruido.

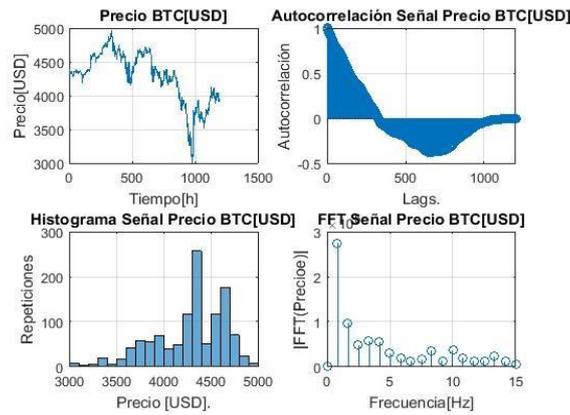
Dicho análisis se realiza empleando herramientas matemáticas como la auto-correlación, el histograma de la señal y la Transformada Rápida de Fourier (FFT). Con base en los resultados obtenidos, se observó que la señal cuenta con un nivel DC considerable, por ello una de las primeras modificaciones que se realizó consistió en eliminar dicho nivel, convirtiéndola en bipolar.

La auto-correlación permite identificar si existe una relación lineal directa entre el estado de la señal en el presente y en el pasado. En la gráfica correspondiente se observa una periodicidad de baja frecuencia enmascarada bajo el ruido. Con base en el análisis global, se confirma, por tanto, que la utilización de un modelo auto-regresivo de medias móviles (ARMA) puede ser apropiada.

Al aplicar la FFT se observa que no posee componentes de alta frecuencia importantes y que su energía está concentrada sobre todo en las bajas frecuencias. Además, se confirma la presencia de un armónico con alta energía ubicado en la región de baja frecuencia. Se propuso eliminar dicha componente para no sobrecargar el sistema de predicción, ya que su periodicidad permite caracterizarla fácilmente. Al hacer esto se logró afinar la solución hacia la predicción de la componente no periódica de la señal, es decir, su componente aleatoria.

Al finalizar el proceso de predicción, se añade a dicha respuesta la componente periódica conocida y el nivel DC restado. De esta manera, se obtuvo el precio en términos reales. A continuación se exponen las gráficas resultantes al aplicar los tratamientos a la serie de tiempo:

Figura 1. Análisis de Señal Precio BTC/USD~



Sistema difuso basado en la experienci

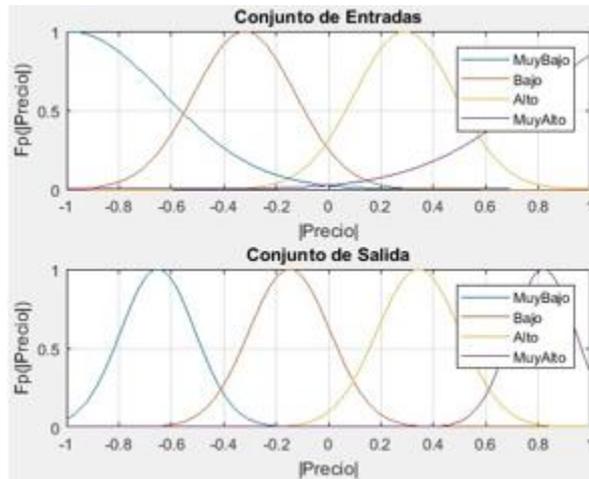
El primer sistema difuso propuesto para la solución del problema, se diseñó con la ayuda de un experto en *trading* de criptodivisas. Las entradas del sistema difuso se definen como los cinco estados anteriores de la señal y su estado presente $[P(k-4), P(k-3), P(k-2), P(k-1), P(k)]$. La salida, en cambio, es el valor en el periodo siguiente $P(k+1)$. Se define como 5 el número de entradas, ya que en estos rangos la señal de precio (periodo de muestreo de 30 minutos) evidenció cambios de tendencia importantes.

Para cada una de las 5 entradas se generan 4 funciones de pertenencia, correspondientes a descriptores lingüísticos que difuminan la entrada. Tales funciones se ajustan con base en el conocimiento del experto y se exponen en la Figura 2.

B1. Parámetros del sistema difuso:

- Funciones de pertenencia: Gaussiana.
- Normas T: Producto
- Normas S: Máximo
- Defusificador: Centroide

Figura 2. Funciones de pertenencia en Entradas y Salida de Sistema Difuso I Diseñado con la experiencia.



Ahora bien, una base de conocimiento se define con 30 reglas que relacionan las entradas entre sí. El motor de inferencia utiliza dicha base de conocimiento para realizar la predicción del precio para el instante (K+1). En la Tabla 1 se resumen 4 de las más importantes reglas de la base de reglas.

Tabla 1. Tabla con 4 Reglas del Sistema Difuso Diseñado con la experiencia

P(k-4)	P(k-3)	P(k-2)	P(k-1)	P(k)	P(k+1)
Bajo	Bajo	Bajo	Bajo	MuyBajo	MuyBajo
MuyBajo	MuyBajo	MuyAlto	MuyAlto	MuyAlto	Bajo
MuyAlto	MuyBajo	MuyBajo	MuyAlto	Bajo	Alto
Muy Alto	Muy Alto	Muy Bajo	Muy Alto	Muy Alto	MuyAlto

Por ejemplo, para la primera regla (fila) de la tabla se deduce que: Si P(k-4) es Bajo & P(k-3) es Bajo & P(k-2) es Bajo & P(k-1) es Bajo & P(k) es Muy Bajo ENTONCES P(k+1) será MuyBajo.

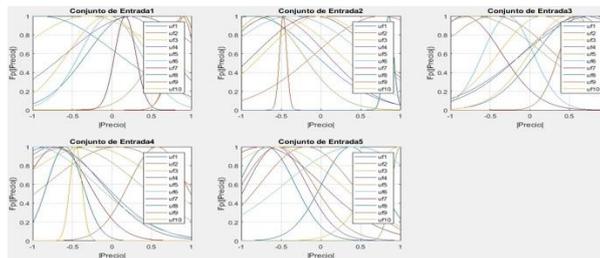
ANFIS

Para la implementación de este algoritmo se realizó una exploración local del paisaje de optimización, variando diferentes parámetros, como el número de reglas, el número de experimentos y el número de épocas. En este caso, para cada parámetro se especificaron 3 valores fijos y se realizaron todas las combinaciones posibles entre ellos. Se tomaron valores de 100, 200 y 300 para el número de experimentos, el número de épocas en 200, 400 y 600. Y el número de reglas toma valores de 5, 10 y 15.

Luego de correr y explorar las posibles combinaciones entre los diferentes parámetros, se obtuvo la mejor solución (menor error de predicción) y se analizaron sus conjuntos difusos y base de reglas. Además, se realizó una evaluación en detalle de dicha solución para posteriormente comparar las 3 alternativas de diseño planteadas.

El mejor sistema difuso fue encontrado tras realizar 100 experimentos de 400 épocas y con un total de 10 reglas. El sistema difuso diseñado por este método es de tipo Sugeno. En la Figura 3 se observan las funciones de pertenencia encontradas para las entradas.

Figura 3. Funciones de Pertenencia de Entradas del Sistema Difuso



Los parámetros de inicialización utilizados en el algoritmo ANFIS fueron los siguientes:

C1. Parámetros ANFIS:

-
- Épocas = Variable
- Tasa de aprendizaje = 1%.

Método de optimización = Híbrido (BackPropagation & Least-squares).

*El sistema difuso inicial se genera de forma aleatoria.

En este punto se evidencia que es complejo extraer información de las funciones de pertenencia y base de reglas resultantes; sin embargo, al comparar con el sistema difuso planteado en el apartado I, se observan algunas semejanzas en cuanto a los centros de las campanas y sus desviaciones.

Algoritmo genético simple

En la implementación de este algoritmo se generó la población inicial de manera aleatoria. Se realizó también una exploración en los parámetros de inicialización, variando entonces el número de generaciones y el número de experimentos. En la siguiente lista se encuentran los parámetros iniciales utilizados en este algoritmo evolutivo:

Parámetros algoritmo evolutivo

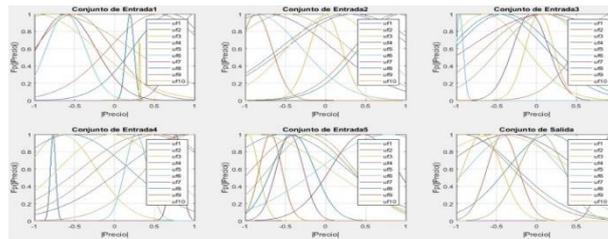
- Generaciones = Variable.
- Tamaño Población = 30.
- Presión Selectiva = 0.01.
- Probabilidad de Cruce = 70%.
- Probabilidad de Mutación = 5%.
- Experimentos = Variable.

La función objetivo se diseñó cuidadosamente teniendo en cuenta la naturaleza del problema y utilizando los 3 indicadores estadísticos anteriormente mencionados.

$$f_{obj} = MSE + (MAPE/100) * NMSE \quad (5)$$

Al terminar de correr el algoritmo se analizaron las diferentes poblaciones resultantes y se escogieron los mejores individuos para cada una. Dichos individuos volvieron a ser comparados para dejar finalmente solo una solución, que es la comparada con los anteriores sistemas difusos diseñados por otros métodos. La mejor solución se encontró con un total de 10 reglas. A continuación se exponen sus funciones de pertenencia resultantes.

Figura 4. Funciones de pertenencia de Entradas del Sistema Difuso III



Se volvieron a encontrar semejanzas con las funciones definidas en el sistema I, aunque de manera no tan marcada. Además, como el número de reglas aumentó, la interpretación de las mismas se hizo más compleja.

Resultados

Para realizar una evaluación del desempeño de cada uno de los sistemas difusos diseñados mediante los diferentes métodos, se utilizaron los estadísticos de error MSE, MAPE y NMSE. Enseguida se exponen los resultados de evaluación, además de las gráficas resultantes cuando los sistemas se someten a una prueba para predecir puntos contiguos, y su comparación con la serie real. Cabe aclarar que se utilizó un esquema de validación cruzada durante todos los procesos de evaluación y entrenamiento.

En la siguiente tabla, se concatenan los resultados de los indicadores de evaluación de desempeño de las mejores soluciones obtenidas para cada uno de los tres métodos de diseño.

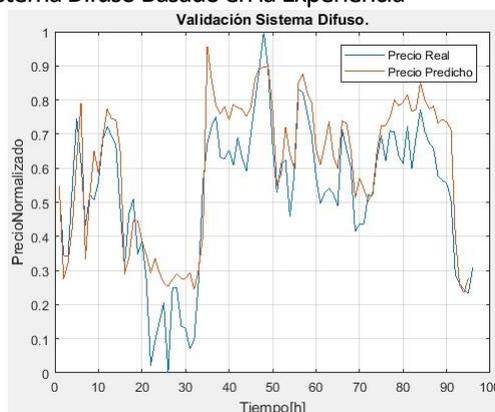
Tabla 2. Estadísticos de Evaluación de Desempeño

	MSE	MAPE	NMSE
Experiencia	0.025	154 %	1.829
ANFIS	0.25	114 %	0.80
Algoritmo Evolutivo	0.0619	478 %	20.52

Las siguientes figuras son el resultado obtenido para cada uno de los procesos de diseño, cuando los predictores son sometidos a una entrada de datos contiguos.

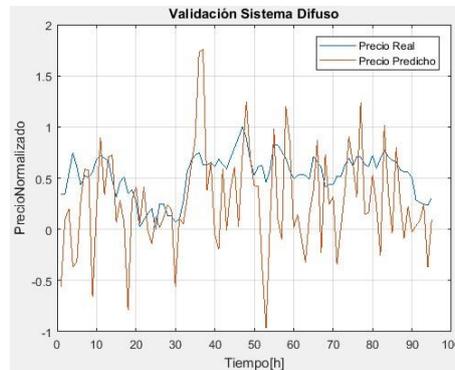
Diseño basado en la experiencia

Figura 5. Validación Sistema Difuso Basado en la Experiencia



Diseño utilizando ANFIS

Figura 6. Validación Sistema Difuso diseñado mediante ANFIS



Diseño mediante algoritmo genético simple

Figura 7. Validación Sistema Difuso Diseñado Mediante Algoritmo Genético Simple



Discusión

El menor error MSE se refleja en el sistema difuso diseñado con base en la experiencia, dado que un número tan elevado de reglas puede permitir una buena aproximación a los datos. Se evidenció que el indicador MAPE muestra las incongruencias en tendencia que presenta la señal predicha en los diferentes sistemas. El valor MAPE más alto se obtuvo en el sistema difuso III, diseñado mediante el algoritmo evolutivo simple. Por otro lado, si el indicador NMSE es mayor a 1, se infiere que una caminata aleatoria podría ser una buena alternativa de diseño. Esta condición solo se cumplió en el sistema diseñado mediante el algoritmo evolutivo simple.

Basándose en la experiencia, se diseñó un sistema difuso con una gran cantidad de reglas que aunque parece poco práctico, demostró tener resultados comparables con los otros modelos. La predicción de tendencias es bastante significativa, lo cual indica que se pueden obtener resultados positivos si se opera con base en este sistema.

El algoritmo ANFIS logra ajustar los parámetros del sistema difuso con muy buenos resultados. Sin embargo con él se puede caer fácilmente en un mínimo local, ya que

no realiza una exploración masiva del resto del panorama de optimización. Falencia mejorada en el algoritmo genético simple, que implementa mutaciones aleatorias para realizar dicha exploración.

Conclusiones

Se evidenció que el algoritmo ANFIS presenta un mejor comportamiento frente al problema según los indicadores. Esto puede deberse al método de propagación de error que utiliza en su estructura, que en este caso se acopló en gran medida a la naturaleza del problema.

El mal desempeño encontrado con el algoritmo genético simple, puede deberse a una selección no adecuada de la función objetivo, ya que de ella depende en gran medida el rumbo de la evolución de las poblaciones.

La unión hombre-máquina puede ser una buena alternativa de solución de problemas, ya que la maquina sola es incapaz de programarse a sí misma para la solución de un problema, y siempre requiere una inicialización. Por otro lado, el hombre posee una capacidad de entendimiento muy alta en comparación a las maquinas, sin embargo su capacidad de procesamiento de datos es reducida y lenta. Por ello, este acople resulta tan idóneo y poderoso para la aplicación en problemas de diversas áreas.

Referencias bibliográficas

1. D. T. Rice, The Past and Future of Bitcoins in Worldwide Commerce., ser. Business La Today. American Bar Association, 2013.
2. L. A. ZADEH, Fuzzy Sets. Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California, 1965.
3. R. Adhikari and R. K. Agrawal, An Introductory Study on Time Series Modeling and Forecasting. LAP Lambert Academic Publishing, Germany, 2013.
4. J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system. Dept. of Electr. Eng. Comput. Sci., California Univ., Berkeley, CA, USA, 1993.
5. J. Navarro, L. Perez., J. Molina., N. Blanco., M. Amador., E. Vasquez., C. Alvarino. Caracterización de la seguridad y salud en el trabajo en las empresas de Zona Franca y Via 40 de la ciudad de Barranquilla. En N. A. (Corporación U. L. León Castro (Ed.), Seguridad y salud en el trabajo: Análisis de las empresas de Barranquilla (Primera, pp. 51-90). Barranquilla: Corporación Universitaria Latinoamericana., 2018.