Plataforma de Big Data para la predicción de accidentes de salud y seguridad

Resumen

Propósito

En este documento se destaca el uso de tecnologías de Big Data para el análisis de riesgos de salud y seguridad en el dominio de la infraestructura energética con grandes conjuntos de datos de riesgos de salud y seguridad que suelen ser dispersos y ruidosos.

Diseño/metodología/enfoque

El estudio se centra en el uso de marcos de Big Data para diseñar una arquitectura robusta que permita gestionar y analizar (análisis exploratorio y predictivo) accidentes en infraestructuras eléctricas . La arquitectura diseñada se basa en un ciclo de vida de análisis de riesgos para la salud coherente . Se implementó en Java un prototipo de la arquitectura que interconectaba diversos artefactos tecnológicos para predecir la probabilidad de ocurrencia de riesgos para la salud.

Se realizó una evaluación preliminar de la arquitectura propuesta con un subconjunto de 17 datos objetivos, obtenidos de una empresa líder de infraestructura energética del Reino Unido que ofrece una amplia gama de servicios de infraestructura energética.

Hallazgos

La arquitectura propuesta permitió identificar variables relevantes y mejorar la precisión de las predicciones preliminares y la capacidad explicativa. También permitió extraer conclusiones sobre las causas de los riesgos para la salud. Los resultados representan una mejora significativa en la gestión de la información sobre accidentes en la construcción, especialmente en el ámbito de las infraestructuras eléctricas .

Originalidad/valor

Este estudio realiza una revisión bibliográfica exhaustiva para impulsar la gestión de riesgos de salud y seguridad en la construcción. También destaca la incapacidad de las tecnologías convencionales para gestionar conjuntos de datos no estructurados e incompletos para el procesamiento analítico en tiempo real . El estudio propone una técnica de Big Data para encontrar patrones complejos y establecer la cohesión estadística de patrones ocultos para una óptima toma de decisiones futuras .

Palabras clave: Análisis de Big Data, Aprendizaje automático, Análisis de riesgos para la salud, Salud y seguridad

1. Introducción

Los accidentes laborales son motivo de preocupación en la sociedad moderna, especialmente en las obras de construcción , donde se lleva a cabo un gran número de actividades de construcción (Zhu et al., 2016). El sector de suministro de infraestructura eléctrica , por ejemplo, tiene una alta incidencia de lesiones laborales no mortales , ya que los trabajadores que utilizan maquinaria pesada se enfrentan a riesgos para la salud como la radiación, el polvo, las temperaturas extremas y los productos químicos, entre otros (McDermott y Hayes, 2016). Según la Dirección de Salud y Seguridad del Reino Unido, en 2014/15 se invirtieron 4800 millones de libras en lesiones laborales (HSE, 2016). De igual manera, los costos de reparación de las líneas de comunicación enterradas son significativos cuando se interrumpen durante las excavaciones (McDermott y Hayes, 2016).

Se han utilizado diversas técnicas de aprendizaje automático para la predicción de riesgos de salud y seguridad en la construcción. Por ejemplo, los árboles de decisión (Cheng et al., 2011), el modelo lineal generalizado (Esmaeili et al., 2015) y el método neuronal difuso (Debnath et al., 2016) se han utilizado para analizar datos de incidentes y reducir las tasas de accidentes. Se han utilizado técnicas como la red bayesiana.

51 se utiliza para cuantificar las tasas de accidentes laborales (Papazoglou et al., 2015) y las redes bayesianas difusas 52 para el análisis de equipos dañados (Zhang et al., 2016). Otras opciones son la representación de pajarita 53 para la evaluación de riesgos laborales (Jacinto y Silva, 2010) y los modelos de Poisson 54 para la modelización del impacto de las lesiones laborales (Yorio et al., 2014).
55

56 Sin embargo, un problema importante asociado con estos modelos existentes es su capacidad limitada para 57 procesar datos brutos a gran escala, ya que se necesita un esfuerzo considerable para transformarlos en una 58 forma interna apropiada para lograr una alta precisión de predicción (Esmaeili et al. 2015).
Los datos sobre accidentes en la construcción suelen ser amplios, heterogéneos y dinámicos (Fenrick y Getachew , 2012), presentan relaciones no lineales entre las variables causales de los accidentes (Gholizadeh y Esmaeili , 2016), presentan datos desequilibrados y presentan valores faltantes considerables (Bohle et al., 2015). Además, estas técnicas simplifican algunos factores clave y prestan poca atención al análisis de las relaciones entre un fenómeno de seguridad y los datos de seguridad (Landset et al., 2015).
Con base en lo anterior, la tecnología de Big Data, gracias a su capacidad de procesamiento paralelo y a su capacidad para gestionar eficientemente datos de alta dimensión y ruido con relaciones no lineales, será beneficiosa para el análisis de riesgos de salud y seguridad en el ámbito de la infraestructura eléctrica. Además, esta tecnología revelará los posibles factores que contribuyen a los accidentes en este ámbito. Por lo tanto, los objetivos de este estudio son trazar las etapas del ciclo de vida del análisis de riesgos laborales y desarrollar una arquitectura de Big Data para la gestión de riesgos de salud y seguridad.
70

71 1.1. Big data para el análisis de riesgos de salud y seguridad
72

73 Big Data es una tecnología emergente que se refiere a conjuntos de datos que son muchos órdenes de magnitud más grandes que los archivos estándar transmitidos a través de Internet (Suthakar et al. 2016).
Existe un gran interés en utilizar la información de Big Data para diversos análisis (exploratorios, descriptivos, predictivos y prescriptivos) a fin de determinar sucesos futuros. Lo más importante es que las tecnologías de Big Data respaldan las técnicas analíticas para el análisis de riesgos de salud y seguridad ocupacional ; por lo tanto, el sistema propuesto en este estudio, denominado Plataforma de Predicción de Accidentes de Big Data (B-DAPP), ofrece oportunidades inigualables para minimizar los riesgos ocupacionales en las obras de construcción. La perfecta combinación de las siguientes tecnologías: Big Data , Salud y Seguridad, y Aprendizaje Automático es el resultado de una robusta herramienta de gestión de riesgos de salud y seguridad para ayudar a las partes interesadas a tomar decisiones adecuadas para minimizar los accidentes ocupacionales en proyectos de Infraestructura Eléctrica.

El análisis de riesgos para la salud y la seguridad depende de un cómputo de alto rendimiento y un almacenamiento de datos a gran escala, lo que requiere una gran cantidad de conjuntos de datos diversos sobre riesgos para la salud y la seguridad, así como conocimientos de aprendizaje automático para proporcionar con éxito las responsabilidades analíticas necesarias. Sin embargo, estos conjuntos de datos son poco fiables, no estructurados, incompletos y desequilibrados (Chen et al.).
88 2017). Por lo tanto, almacenar los conjuntos de datos mediante tecnologías convencionales y someterlos a procesamiento en tiempo real para análisis avanzados es un gran desafío. Una técnica robusta para encontrar patrones complejos y establecer la cohesión estadística de patrones ocultos en dichos conjuntos de datos.

91 Para una toma de decisiones óptima en el futuro es inevitable. Por lo tanto, se motiva el uso de tecnologías de Big Data 92 para abordar estos desafíos.

93

94 1.2. Justificación de la investigación

95 Existe una brecha tecnológica aparente en la literatura existente con respecto a la gestión de riesgos de salud y seguridad . En particular, hay investigación limitada sobre la aplicación de técnicas de Big Data para gestionar el riesgo de salud y seguridad en la infraestructura eléctrica. El desarrollo de un B -DAPP robusto para el riesgo de salud y seguridad es el objetivo del esfuerzo continuo de I+D. La herramienta propuesta proporcionará a las partes interesadas información bien informada y basada en datos para reducir los accidentes e incidentes en las obras de construcción. Por lo tanto, se propone una arquitectura de Big Data para gestionar los riesgos de salud y seguridad. Además, se realiza una presentación de los componentes y las tecnologías relevantes de la arquitectura propuesta necesarias para almacenar y analizar conjuntos de datos de riesgo de salud y seguridad para la exploración y predicción en tiempo real. El término "Arquitectura", como se usa en este texto, se refiere a las estructuras de alto nivel de un sistema de software. De manera similar, en el contexto de este estudio 105 , 'Accidente' es un evento no planificado y no premeditado causado por actos o condiciones inseguras 106 que resultan en lesiones, mientras que 'Incidente' es un evento que causa daño real a la propiedad (incluida la planta 107 o el equipo) u otra pérdida con potencial de causar lesiones.

108

El resto del documento se estructura de la siguiente manera: la Sección 2 analiza la metodología de investigación , el análisis de Big Data y el ecosistema de Big Data. La Sección 3 analiza el ciclo de vida del análisis de riesgos para la salud . La Sección 4 presenta la arquitectura de Big Data propuesta para la gestión de riesgos de salud y seguridad, mientras que la Sección 5 presenta los resultados preliminares. Las conclusiones y el trabajo futuro se presentan en la Sección 6.

114

115

116 2. Metodología 117

En esta sección, se analiza la metodología empleada en esta investigación. En primer lugar, se realiza una revisión exhaustiva de la literatura para avanzar en la gestión de riesgos de salud y seguridad con respecto a la arquitectura del sistema y el ciclo de vida del análisis de sistemas. Posteriormente, la arquitectura propuesta y el ciclo de vida del análisis de riesgos laborales se validan mediante un análisis preliminar de los datos relacionados con los riesgos de salud y seguridad. Para ofrecer una arquitectura holística de Big Data y un ciclo de vida del análisis de riesgos laborales , se realizó una revisión exhaustiva de la literatura existente sobre modelos de predicción de riesgos de salud y seguridad, Big Data y aprendizaje automático . En este sentido, se buscan artículos de investigación entre 2005 y 2017 en bases de datos en línea como Journal of Big Data, Big Data Research, Safety 126 Science, Journal of construction engineering, Journal of Decision Systems, Journal of Safety 127 Research, Journal of Construction Engineering and Management, Reliability Engineering y 128 System Safety. También se consideran revisiones recientes de investigaciones y libros sobre Big Data Analytics (Camann et al. 2011; Gandomi 130 y Haider 2015; Guo et al. 2016).

131

132 Ejemplos de palabras de búsqueda utilizadas incluyen: "gestión de riesgos de salud y seguridad", "estrategias de diseño 133 para riesgos laborales en la construcción", "modelos de predicción para riesgos de salud ocupacional", "Big 134 Data en la construcción", "Arquitectura de aplicaciones basada en Big Data" y "Análisis de Big Data". En 135 general, se seleccionaron 94 publicaciones a pesar de que la búsqueda bibliográfica no fue exhaustiva como resultado 136 de una gran cantidad de artículos publicados. Sin embargo, se cree que la búsqueda bibliográfica 137 ha capturado una muestra equilibrada y representativa de la investigación relacionada. Se incluyeron estudios en los que se utiliza Big 138 Data para desarrollar aplicaciones empresariales, y se excluyeron aquellos centrados en los peligros relacionados con el tráfico vial 139 y los peligros para la salud en dominios no relacionados con la construcción (por ejemplo, minería 140 y pesca). Este procedimiento de eliminación redujo aún más los artículos seleccionados a 141 66. Además, estos artículos se examinan para determinar su relevancia mediante la lectura de resúmenes, introducciones, 142 y conclusiones. Finalmente, los artículos se reducen a 50. La Tabla 1 muestra cómo estos 143 artículos seleccionados son relevantes y contribuyen al desarrollo de la arquitectura propuesta, que se basa esencialmente en tres conceptos, a saber, Big Data, Riesgo de salud y seguridad, y Aprendizaje automático . En este estudio, presentamos la arquitectura B-DAPP propuesta y el 146 etapas del ciclo de vida del análisis de peligros para la gestión de incidentes y accidentes.

147

148 2.1. Análisis de big data 149 El

big data consiste en conjuntos de datos grandes y complejos, a menudo difíciles de manipular con los métodos de procesamiento convencionales. Tiene seis atributos que lo definen (Gandomi y Haider, 2015): volumen , variedad, velocidad, veracidad, variabilidad y complejidad, y valor. El término «volumen» representa la magnitud de los datos (medida en terabytes, petabytes y más). La «variedad» es la heterogeneidad estructural de un conjunto de datos, mientras que la «velocidad» es la tasa de generación de datos.

La " Veracidad " se refiere a la falta de fiabilidad inherente a las fuentes de datos, mientras que la "Variabilidad" (complejidad) representa la variación en el flujo de datos. Finalmente, el "Valor" mide la información extraída de conjuntos de datos históricos de incidentes para optimizar las decisiones de control, mitigar los incidentes y reducir su impacto .

Estos atributos son evidentes en un conjunto típico de datos de salud y seguridad de infraestructura eléctrica, que suele ser extenso, heterogéneo y dinámico (Fenrick y Getachew, 2012). El análisis de big data es un concepto que inspecciona, depura, transforma y modela el big data para descubrir información útil que respalde la toma de decisiones (Power, 2014). El análisis de big data tiene una rica tradición intelectual y se nutre de una amplia variedad de campos relacionados, como la estadística, la minería de datos, el análisis de negocios, el descubrimiento de conocimiento a partir de datos (KDD) y la ciencia de datos. Las formas del análisis de big data son descriptivas (Schryver et al., 2012), predictivas (Esmaeili et al., 2015), prescriptivas (Delen y Demirkan, 2013) y causales (Schryver et al., 2012).

166

167 2.2. Big Data para la gestión de riesgos de seguridad

168 Existe una amplia variedad de tecnologías y arquitecturas heterogéneas disponibles para implementar aplicaciones de Big Data . Dado que este documento pretende desarrollar una arquitectura robusta de Big Data para el análisis de riesgos sanitarios , se presenta una breve descripción de las herramientas y plataformas de Big Data para facilitar la creación de una arquitectura compacta y mejorar la comprensión del concepto. Principalmente,

172 centrándose en el ecosistema Hadoop, un sistema diseñado para resolver problemas de Big Data.

173

174 Tabla 1: Resumen de los artículos revisados

175

# Artículo	Contribución a la arquitectura de análisis de riesgos de salud y seguridad		
	Riesgos para la salud y la seguridad de las máquinas	aprendiendo	Big data
1 Liu y Tsai (2012)			
2 Zhou y otros (2015)			
3 García-Herrero et al. (2012) 4 Groves et al. (2007)			
5 Le et al. (2016) 6 Soltanzadeh et al. (2016)			
7 Poder (2014)			
8 Yi y otros (2016)			
9 Cheng y otros (2011)			
10 Silva y otros (2016)			
11 Raviv y otros (2017)			
12 Liao y Perng (2008)			
13 Li y Bai (2008)			
14Törner y Pousette (2009)			
15 Pinto et al. (2011)			
16 Tixier y otros (2016) 17 Hallowell y Gambatese (2009)			
18 Pääkkönen y Pakkala (2015)			
19 Venturini y otros (2017)			
20 Suthakar y otros (2016)			
21 Najafabadi y otros (2015)			
22 Landset y otros (2015)			
23 Tsai y otros (2015)			
24 Zang y otros (2014)			
25 Jin y otros (2015)			
26 Rahman y Esmailpour (2016)			
27 Al-Jarrah y otros (2015)			
28 Zhang et al. (2016) 29 Love & Teo (2017)			
30 Rivas et al. (2011) 31 Guo et al. (2016) 32 Zou et al. (2007)			
33 Wu y otros (2010)			
34 Carbonari et al. (2011)			
35 Weng et al. (2013) 36 Naderpour et al. (2016) 37 Yoon et al. (2016)			
38 Favarò y Saleh (2016) 39 Jocelyn et al. (2017) 40 Papazoglou et al. (2017) 41 Papazoglou et al. (2015) 42 Fragiadakis et al. (2014) 43 Ciarapica y Giacchetta (2009) 44 Khakzad et al. (2015) 45 Galizzi y Tempestades (2015)			
46 Gürcanli y Müngena (2009) 47 Debnath et al. (2016) 48 Nanda et al. (2016) 49 Zeng et al. (2008)			
50 Guo et al. (2016)			

176

177

178 2.2.1. Ecosistema Hadoop

179 Hadoop es un motor de procesamiento MapReduce con sistemas de archivos distribuidos (White 2012).
180 Sin embargo, ha evolucionado hasta convertirse en una vasta red de proyectos (ecosistema Hadoop)
relacionados con cada paso 181 de un flujo de trabajo de Big Data. El concepto se conoce ahora como ecosistema Hadoop, que abarca proyectos y productos relacionados , desarrollados para complementar o reemplazar los componentes originales 183. A continuación, se analizarán ambos conceptos con más detalle para facilitar su comprensión.
184

185 El proyecto Hadoop consta de cuatro módulos (White 2012):

186 a) El sistema de archivos distribuido Hadoop (HDFS) es un sistema de archivos tolerante a fallos diseñado para almacenar
187 Datos masivos en múltiples nodos de hardware básico. Tiene un sistema maestro-esclavo.
188 Arquitectura compuesta por nodos de datos y nodos de nombre. Los nodos de datos almacenan bloques.
189 de los datos, recuperar datos a solicitud e informar al nodo de nombre con inventario. El
190 El nodo de nombre mantiene registros del inventario y dirige el tráfico a los nodos de datos.
191 solicitudes de clientes.
192 b) Motor de procesamiento de datos MapReduce. Un trabajo MapReduce consta de una fase de mapeo y una fase de
193 reducción. La fase de mapeo organiza los datos sin procesar en pares clave-valor, mientras que la fase de reducción...
194 La fase procesa datos en paralelo.
195 c) YARN (“Yet Another Resource Negotiator”) es un administrador de recursos de Hadoop
196 Proyecto introducido para abordar las limitaciones de MapReduce. Separa las infraestructuras de las
197 representaciones de programas.
198 d) Común es un conjunto de utilidades requeridas por los demás módulos de Hadoop. Estas incluyen
199 códecs de compresión, utilidades de E/S, detección de errores, autorización de usuarios proxy,
200 autenticación y confidencialidad de los datos.
201

El ecosistema Hadoop consta de varias herramientas desarrolladas sobre los módulos centrales de Hadoop descritos anteriormente para apoyar a investigadores y profesionales en todos los aspectos del análisis de datos. La estructura del ecosistema consta de las siguientes capas: almacenamiento, procesamiento y gestión. La Figura 1 muestra ejemplos de herramientas estándar utilizadas en aplicaciones de Big Data. La selección correcta requiere un conocimiento profundo de las características críticas de estas plataformas y las características del problema a resolver. En el caso del análisis de riesgos para la salud, la adaptación de las plataformas como resultado del aumento de la carga de trabajo supera el resto de los criterios de selección. En realidad, el ecosistema Hadoop está compuesto por más de 100 proyectos, y se recomienda a los lectores consultar White (2012) o el sitio web de Hadoop para obtener más información.

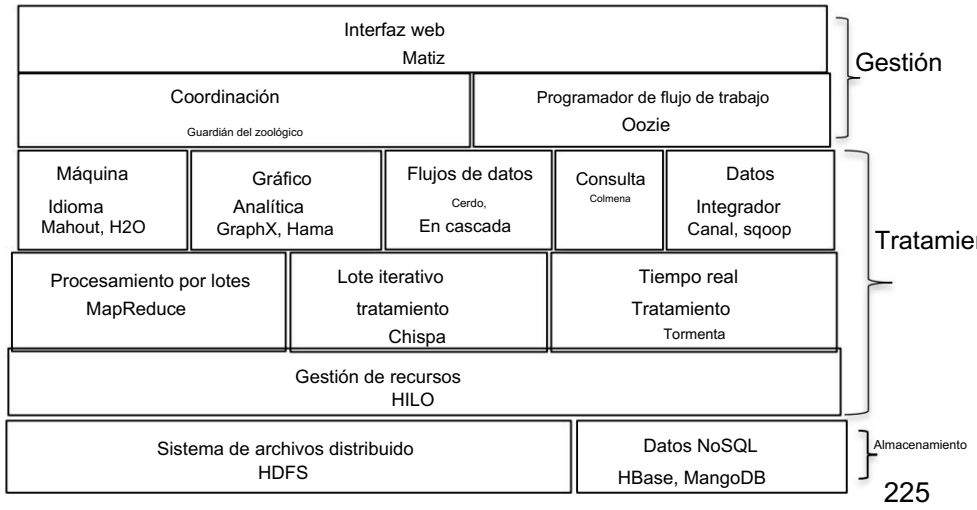


Figura 1. Ecosistema Hadoop

a) Capa de almacenamiento: esta capa incluye el HDFS descrito anteriormente y el no relacional.
bases de datos (NoSQL). Las bases de datos no relacionales son anidadas, semiestructuradas y
datos no estructurados que respaldan tareas de aprendizaje automático. Estas bases de datos utilizan
siguientes modelos de representación de datos: almacenes de clave-valor (es decir, Redis), almacenes de documentos
(es decir, MongoDB), datos orientados a columnas (es decir, HBase) y modelos basados en gráficos
(Neo4J) El modelo gráfico se considera más flexible que otros modelos.

b) Capa de procesamiento: Esta capa realiza el análisis real utilizando YARN, lo que permite
Uno o más motores de procesamiento para ejecutarse en un clúster Hadoop. Además, una capa tiene
marcos para la transferencia, agregación e interacción de datos. Algunos ejemplos son Flume,
Sqoop, Hive, Spark y Pig. Flume recopila, agrega y transfiere registros de datos en HDFS.
Kafka es un sistema de mensajería distribuida en HDFS y Sqoop transporta datos masivos
entre HDFS y bases de datos relacionales. Hive es un motor de consulta para consultar datos.
almacenados en las bases de datos HDFS y NoSQL. Spark admite el cálculo iterativo y
Mejora la velocidad y los recursos al utilizar computación en memoria. Finalmente,
Pig ofrece un marco de ejecución y un lenguaje de flujo de datos para respaldar las necesidades definidas por el usuario.
funciones escritas en Python, Java, JavaScript, etc. Los marcos de aprendizaje automático son
Se utilizan para realizar tareas de aprendizaje automático en Hadoop. Algunos ejemplos son Mahout, H2O, etc.
Mahout es una de las herramientas de aprendizaje automático más conocidas. Es conocida por tener una
Amplia selección de algoritmos robustos, pero con tiempos de ejecución ineficientes debido a la lentitud.
Motor MapReduce. H2O proporciona un motor de procesamiento paralelo, bibliotecas de análisis, matemáticas y
aprendizaje automático para el preprocesamiento y la evaluación de datos.

c) Capa de gestión: Esta capa tiene herramientas para la interacción del usuario y de alto nivel.
Organización. Realiza funciones como programación, supervisión y coordinación, entre otras. Ejemplos de herramientas
disponibles en esta capa son Oozie, Zookeeper y
Hue. Oozie es un programador de flujo de trabajo que administra trabajos para muchas de las herramientas en el

Capa de procesamiento. Zookeeper proporciona herramientas para gestionar la coordinación de datos y protocolos y puede gestionar fallos parciales de red. Incluye API para Java y C y También incluye enlaces para clientes Python y REST. Hue es una interfaz web para Hadoop. Proyectos con soporte para componentes del ecosistema Hadoop ampliamente utilizados.

3. Etapas propuestas de análisis de riesgos para la salud

Desarrollar una herramienta de análisis de riesgos para la salud para datos de riesgos de salud y seguridad es una tarea desafiante , ya que los datos suelen ser dinámicos (Fenrick y Getachew, 2012) y desequilibrados, con valores faltantes significativos (Bohle et al., 2015). Además, el modelado tradicional de causas de accidentes puede ignorar o simplificar algunos factores clave, así como asumir el mismo formato para los datos de entrada . Por lo tanto, una metodología eficiente para abordar estos desafíos requiere un proceso bien articulado para dividir la tarea en etapas más pequeñas y manejables para garantizar la preparación adecuada de diversos enfoques analíticos. En esta sección, se analiza el ciclo de vida de la arquitectura de Big Data propuesta para la herramienta de análisis de riesgos para la salud. El ciclo de vida consta de seis etapas (véase la Figura 2) que se ejecutan iterativamente para adaptarse a los requisitos de la propuesta.

Herramienta .

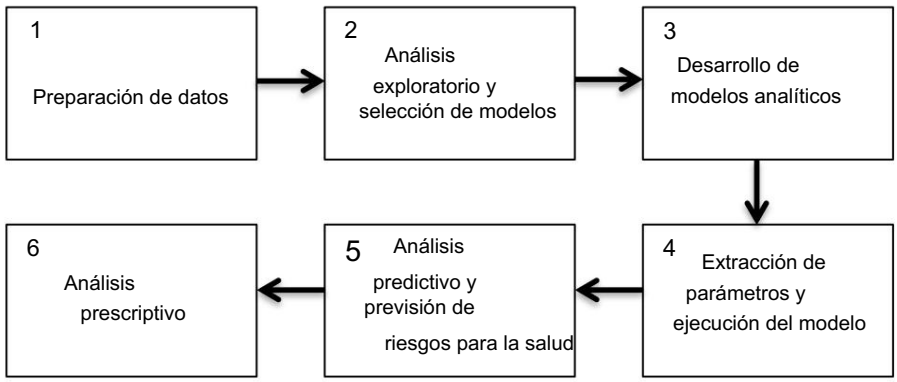


Figura 2. Etapas del análisis de riesgos para la salud

3.1. Preparación de datos

La preparación de datos es un procedimiento para detectar y corregir errores en el conjunto de datos. Para el análisis de riesgos para la salud , se requiere una calidad de datos suficiente para un análisis de alta calidad. Por lo tanto, se obtienen datos de diversas fuentes , se transforman y se cargan en el almacén de datos centralizado. Antes de...

En este contexto, los valores atípicos se eliminan inadvertidamente mediante técnicas como la imputación de media/moda, la transformación y la clasificación por intervalos. Los problemas de datos faltantes también deben resolverse mediante la tecnología adecuada . La imputación de k-vecinos más cercanos (kNN) y la imputación de media/moda son algunos ejemplos para eliminar el problema de los datos faltantes. Al parecer, las técnicas de aprendizaje automático también pueden aplicarse para filtrar rápidamente cientos de miles de narrativas (textos) y así recuperar y rastrear de forma precisa y consistente causas de lesiones de alta magnitud, alto riesgo y emergentes. La información recuperada se utiliza luego para orientar el desarrollo de intervenciones para prevenir incidentes futuros.

En caso de tener grandes cantidades de datos, pueden requerirse métodos para el movimiento de datos en paralelo, lo que puede requerir el uso del componente adecuado del ecosistema Hadoop. Los datos a menudo son

300 analizados para familiarizarse con el riesgo de salud y seguridad en lo que respecta al ámbito de la construcción. Para el análisis preliminar que se presenta aquí, los datos de salud y seguridad se proporcionan como archivos .csv almacenados en el clúster de Hadoop. Se consultan los archivos correspondientes para recuperar detalles específicos sobre riesgos para la salud y la seguridad, como lesiones corporales, lesiones por pérdida y equipos dañados , entre otros. Para ello, herramientas como Apache Flume son de gran importancia para capturar versiones actualizadas de los conjuntos de datos.
306

307 3.2. Análisis exploratorio y selección de modelos

308 Para la gestión de riesgos para la salud, el análisis comienza con análisis exploratorios y luego con análisis predictivos. Para cada actividad en la herramienta propuesta, es esencial tener un objetivo claro para seleccionar correctamente los enfoques analíticos (prescripción, exploratorio, predictivo, etc.) a ejecutar . La exploración de datos de los registros de salud y seguridad se realiza para comprender la relación entre las diferentes variables explicativas. Este análisis exploratorio de datos informa la selección de variables relevantes para construir un modelo robusto de predicción de riesgos para la salud. En este estudio, se utiliza una técnica de visualización para el análisis exploratorio de datos. En esta fase, el propósito del análisis es capturar predictores esenciales y variables independientes, eliminando las menos relevantes para la construcción del modelo. Los métodos de selección de variables incluyen la regresión de todas las variables posibles , la regresión gradual hacia adelante, la regresión del mejor subconjunto, etc. Estos métodos de selección suelen ser iterativos y requieren una serie de pasos para identificar las variables más útiles para el modelo dado. Herramientas como R Studio pueden utilizarse para construir estos modelos.
320

321 3.3. Desarrollo de modelos analíticos

En esta etapa, se crean modelos analíticos para la predicción de riesgos de salud y seguridad mediante técnicas robustas de análisis de Big Data . Los datos se dividen primero en conjuntos de entrenamiento y de prueba. Posteriormente, los modelos analíticos se ajustan a los datos de entrenamiento y se evalúan con los datos de prueba. Se seleccionan los modelos con una precisión óptima o mayor capacidad predictiva. A menudo, este paso puede implicar abordar ciertos problemas de optimización, como la multicolinealidad. Se selecciona e implementa el mejor modelo para predecir riesgos de salud y seguridad a partir de un gran volumen de datos. En muchas ocasiones, el entorno de producción puede requerir el ajuste y la reimplementación de modelos para dar soporte a situaciones más prácticas.
329 (Camann y otros, 2011).
330

331 3.4. Extracción de parámetros y ejecución del modelo

332 Aquí se extraen parámetros vitales para ejecutar los modelos predictivos. Se extraen parámetros como la tarea , el tipo de equipo, la complejidad del proyecto, etc., y se explora la relación entre un fenómeno de seguridad y los datos de seguridad para descubrir los posibles factores que contribuyen a la probabilidad de accidentes. Estas relaciones enfocan las posibles tendencias que podrían utilizarse para predecir el riesgo de salud y seguridad de un proyecto de infraestructura en ejecución. Se aplica una serie de transformaciones para facilitar el uso de la aplicación. Específicamente, se estandarizan los contenidos mediante la ontología ifcOWL (Chaudhuri y Dayal, 1997). Los datos se...

339 almacenados como formatos anotados en gráficos para soportar cálculos más amplios requeridos por la herramienta propuesta.
341

342 3.5. Análisis predictivo y previsión de riesgos para la salud

La predicción de riesgos para la salud proporciona la base necesaria para comprender las causas y los tipos de riesgos para la salud y la seguridad que surgen de un proyecto de construcción en ejecución. Por lo tanto, en esta etapa se emplean modelos predictivos generados mediante análisis de big data para analizar bases de datos de riesgos para la salud y la seguridad y advertir sobre la posible ocurrencia de un riesgo para la salud. De hecho, lo fundamental de esta evaluación es la precisión de los modelos de predicción de riesgos para la salud y la seguridad empleados.

El modelado tradicional de causas de accidentes presenta las siguientes limitaciones: puede ignorar o simplificar algunos factores clave, utiliza análisis cualitativo y se centra en el análisis de causalidad y las explicaciones de un accidente (Landset et al., 2015). Por lo tanto , estos métodos prestan poca atención al análisis de las relaciones entre un fenómeno de seguridad y los datos de seguridad. Tampoco pueden descubrir factores potenciales que contribuyen a la probabilidad de accidentes, como la frecuencia, la relevancia, la ubicación y la puntualidad.

355 El desarrollo de modelos robustos de predicción de riesgos para la salud es el objetivo final de este 356 ciclo de vida, y utilizando los modelos de predicción, se generan 357 pronósticos integrales de accidentes y daños a los equipos para que las organizaciones implementen estrategias y técnicas para mejorar 358 la seguridad de sus sitios de construcción.
359

360 3.6. Análisis prescriptivo

Esta fase optimiza diversas estrategias de seguridad basándose en una gran variedad de factores (la interacción entre las deficiencias en los equipos de trabajo, el lugar de trabajo, los equipos y materiales, las condiciones meteorológicas, etc.) para recomendar la mejor estrategia para una situación dada. Utiliza la simulación y la optimización para ofrecer la mejor estrategia para los diferentes riesgos de salud y seguridad. Como resultado, se genera una gran cantidad de planes de optimización alternativos que se convierten en recomendaciones fáciles de usar para las partes interesadas, lo que facilita la toma de decisiones basada en datos para minimizar los accidentes.
367

368 3.7. Análisis y resultados preliminares

La arquitectura propuesta se confirma y valida aún más con datos objetivos obtenidos de una empresa constructora líder del Reino Unido que ofrece una amplia gama de servicios de infraestructura eléctrica , incluyendo la construcción y renovación de líneas aéreas, subestaciones, cableado subterráneo, fibra óptica, etc. La empresa utiliza una base de datos relacional para almacenar los datos sobre riesgos para la salud y la seguridad , que consisten en un gran número de proyectos de infraestructura eléctrica construidos a lo largo de 13 años (2004 a 2016) en cinco regiones del Reino Unido. Cada vez que ocurre un incidente (o peligro), se crea un registro digital en la base de datos. Se incluyen detalles de algunas de las variables explicativas relevantes en el
En la Tabla 2 se muestran 376 bases de datos.
377

Se utilizó un subconjunto de 5000 proyectos seleccionados al azar de un total de 20000 proyectos para un evaluación y análisis preliminares presentados en este estudio. Los criterios para esta selección incluyen tipos de proyectos (es decir, líneas aéreas, cableado y subestaciones) y modo de construcción (es decir, nueva construcción, remodelación). La distribución de datos en las regiones del Reino Unido ayudará a generar visualizaciones avanzadas como el mapa de calor geográfico. Se accede a los datos de la base de datos relacional a través de la aplicación front-end y se exportan a archivos separados por comas (.csv). Claramente, los datos de riesgos laborales de 5000 proyectos no se etiquetarán como Big Data para justificar el uso de plataformas intensivas en datos para su análisis. Sin embargo, el enfoque adoptado en este estudio puede ser utilizado para analizar conjuntos más grandes de datos de riesgos de salud y seguridad. Se aplica análisis de datos exploratorios para comprender las tendencias subyacentes en los datos utilizando dimensiones geográficas y cronológicas . Por lo tanto, para la investigación de datos se utilizan diversas visualizaciones, como gráficos de barras, gráficos de caja y mapas de calor geográficos .

Tabla 2:Variables explicativas en la base de datos

Variable	Significado
Referencia de incidente	Identificación de un incidente determinado
Tipo de proyecto	El proyecto específico (línea aérea, cableado, offshore, etc.)
Contrato de proyecto	El proyecto de construcción natural en construcción (es decir, obra nueva, mantenimiento, remodelación)
Región	La región específica del sitio de construcción (Escocia, Norte, Sureste, Midlands, etc.)
Subregión	La subregión donde se encuentra el sitio, es decir, Yorkshire East, Midlands North, East England, Tyrone, etc.
Ciudad	Ciudades del Reino Unido donde se ubica la obra.
Ubicación	Un área o ubicación específica del sitio
Cliente	Una organización que utiliza los servicios de la empresa de infraestructura eléctrica.
Tipo de equipo	Especifica la maquinaria (por ejemplo, taladro, martillo, transportador, etc.) utilizada para una tarea.
Edad	La edad de la víctima en el momento del accidente.
Año	El año en que se produjo el riesgo para la salud.
Estación	Factores externos como el clima
Mes	El mes (1-12) en que ocurrió el incidente
Tiempo	El incidente ocurrió en el período (0-6-madrugada, 6-12-mañana, 12-18-tarde, 18-23-noche).
Día de la semana	Día (1-31) en que ocurrió el accidente.
Día de la semana	El día de la semana, es decir, lunes, martes, miércoles, etc.
Tarea	Tarea u operación específica a realizar (excavar, levantar, cortar, etc.).
Tipo de accidente	El tipo de accidente, por ejemplo, caída, tropezón, golpe, inhalación, atrapado, etc.
Tipo de lesión	La consecuencia física para la víctima, es decir, primeros auxilios, fatal, sin lesiones, etc.
Costo de gravedad	Costo financiero incurrido como resultado del accidente
Tipo de peligro	Formas de peligro para la salud, por ejemplo, enfermedad, lesión, pérdida o daño, etc.
Parte del cuerpo lesionada	La parte del cuerpo que está lesionada, es decir, dedos, hombro, cabeza, espalda, etc.
Costo total	El costo del proyecto
Equipo	Parte del equipo dañado durante el funcionamiento.

396 4. Arquitectura de big data propuesta para el análisis de riesgos para la salud

397 En esta sección se analiza la arquitectura de Big Data propuesta para el análisis de riesgos para la salud (véase la Figura 3). Los componentes de la arquitectura son la capa de aplicación, la capa de análisis y modelo funcional , la capa semántica y la capa de almacenamiento de datos, que se analizan posteriormente.
400 subsecciones.

401

402 4.1. Almacenamiento de datos

403 Esta capa es la fuente de datos (riesgos financieros y de salud y seguridad), que son necesarios para el funcionamiento eficiente de B-DAPP y el desarrollo de modelos analíticos (predictivos y prescriptivos).

Los datos financieros 405 incluyen información como el costo del proyecto, el margen, el costo de la mano de obra, el costo del material, etc.

406 Los datos de salud y seguridad contienen datos históricos de riesgos laborales, mientras que los datos multimedia 407 consisten en imágenes y vídeos que representan escenas de accidentes.

408 Como resultado de la naturaleza diversa de los datos que se almacenarán en esta capa, una base de datos NoSQL (es decir, 409 MongoDB, Neo4J, Oracle NoSQL) se utiliza para la implementación debido a su robusto almacenamiento 410 mecanismos y manejo eficiente de datos estructurados, semiestructurados y no estructurados (Leavitt 411 2010).

412

413 4.2. Capa semántica

414 Esta capa proporciona el formato de intercambio de datos y el aprovisionamiento de datos a la capa de aplicación.
415 El formato de intercambio de datos permite compartir un formato de datos común en todo el sistema.
416 El DDAXML se utiliza para compartir datos entre los diferentes módulos del sistema, ya que es un esquema con soporte industrial para compartir información. La funcionalidad de aprovisionamiento de datos proporciona a la capa de aplicación de la arquitectura acceso fluido a las bases de datos mediante el servicio web REST (Transferencia de Estado de Representación). Este enfoque de acceso a bases de datos se considera el más adecuado debido a la diferente naturaleza de los datos de riesgos de salud y seguridad.

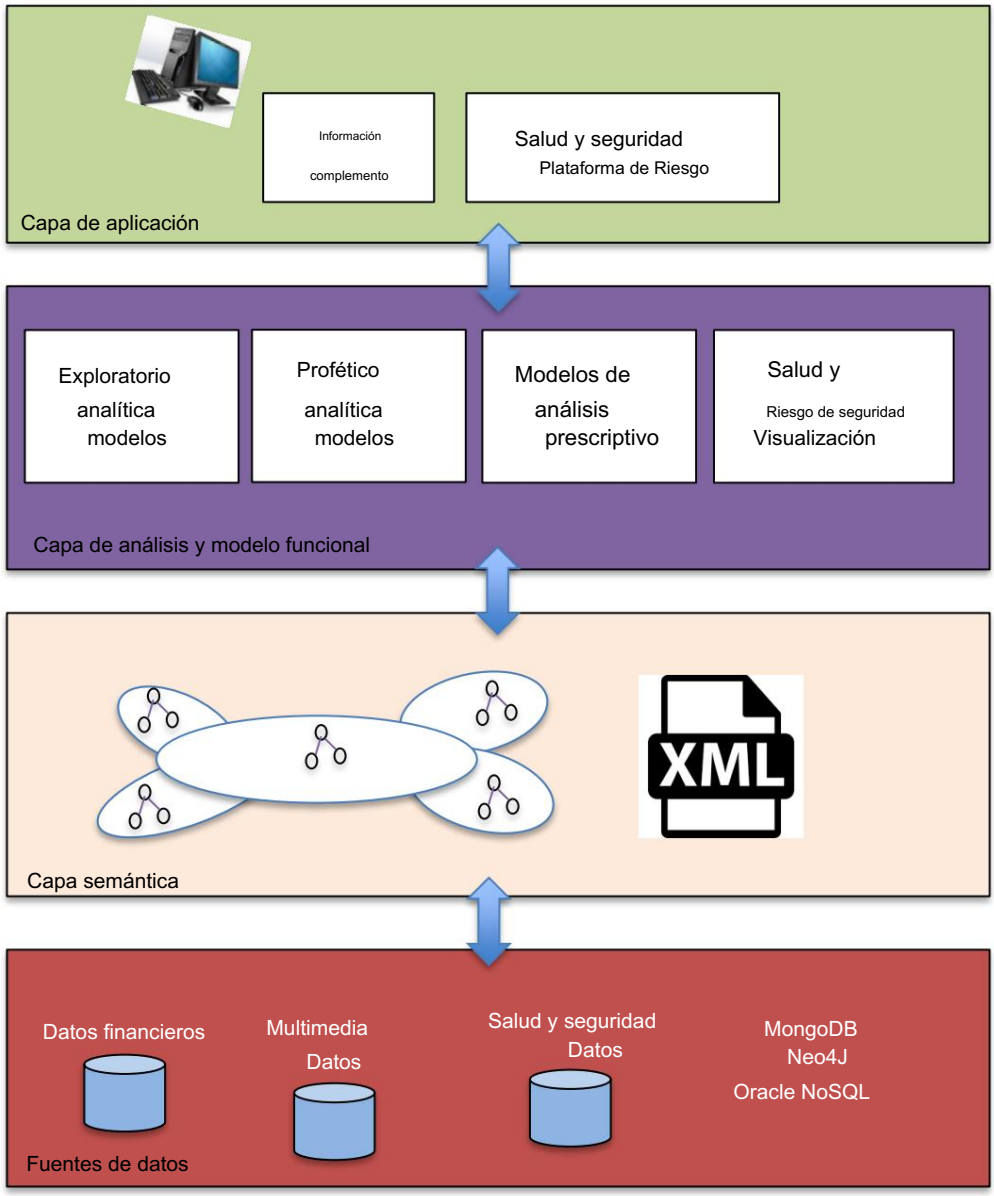
421

422 4.3. Analítica y capa de modelo funcional

La importancia de la herramienta de gestión de riesgos de salud y seguridad reside en su capacidad para analizar y actuar con prontitud sobre datos complejos y de gran volumen. La capa cuenta con un modelo funcional (visualización de Salud y Seguridad) y tres modelos analíticos (analizados anteriormente): análisis exploratorio , análisis predictivo y análisis prescriptivo. Como se mencionó anteriormente, la predicción y la gestión de riesgos para la salud se basan en datos y son muy intensivas. Por consiguiente, Apache Spark

Se eligió el motor 428 en lugar de MapReduce para desarrollar los análisis (predictivos y prescriptivos), 429 debido a su eficiente almacenamiento en memoria y computación (Ryza et al., 2015). Los procesos analíticos 430 para la gestión de riesgos para la salud se implementan utilizando SparkR, H2O y GraphX.

431
432



433
434 Figura 3. Arquitectura B-DAPP
435
436

437 Durante cada iteración del proceso analítico, se exploran y optimizan diferentes modelos predictivos de riesgos para la salud para lograr una precisión óptima.

El marco H2O se selecciona por su rica interfaz gráfica de usuario (GUI) y numerosas herramientas para desarrollar modelos de redes neuronales profundas. Además, ofrece un completo kit de herramientas de aprendizaje automático de código abierto , adecuado para big data (Landset et al., 2015). También proporciona herramientas para diversas tareas de aprendizaje automático, herramientas de optimización, preprocesamiento de datos y redes neuronales profundas . Además, ofrece una integración coherente con Java, Python, R y R Studio, así como con Sparkling Water para la integración con Spark y MLlib . Antes o durante la construcción de un proyecto de infraestructura, se predicen los riesgos para la salud y se difunden a las partes interesadas para ayudar a mitigar su impacto.

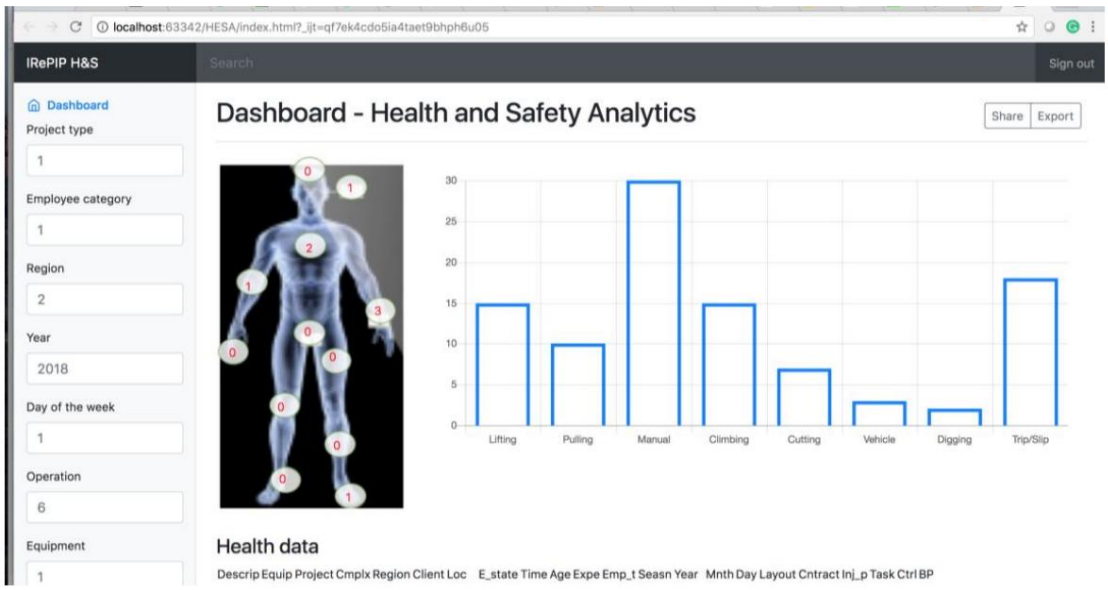
447 4.4. Capa de aplicación

Esta capa se construye aprovechando sus potentes programas API. Los usuarios finales de la herramienta son las partes interesadas (ingenieros, responsables de seguridad y salud, jefes de obra, directores de alto nivel, etc.) . Las variables explicativas de los proyectos de infraestructura bajo B-DAPP se capturan mediante la interfaz de usuario correspondiente y se cargan en el HDFS y, posteriormente, en Triplestore. Spark Streaming activa el flujo de análisis para predecir riesgos para la salud y sugiere información práctica para minimizarlos. Las predicciones y recomendaciones se comunican mediante el Lenguaje de Marcado de Modelos Predictivos (PMML). Se proporciona información a las partes interesadas para gestionar eficazmente los riesgos para la salud.

456
457
458 5. Resultados y discusiones

459 El prototipo de la arquitectura B-DAPP se implementa considerando e interconectando los diversos artefactos tecnológicos. 460 Captura de pantalla de ejemplo generada mediante la simulación del sistema B-DAPP.
461 es como se muestra en la Figura 4, donde el sistema predice la probabilidad y el número de lesiones en partes del cuerpo 462 después de la especificación de los parámetros de entrada (es decir, “Tipo de proyecto”, “Región”, “Operación”, etc.).
463 Informa a las partes interesadas sobre los riesgos probables y les permite prestar la atención adecuada a los factores de riesgo 464 al gestionar los riesgos laborales para lograr un entorno más seguro.
La arquitectura B- DAPP se evalúa mediante análisis exploratorio de datos y se presentan algunos resultados preliminares. El propósito de esta evaluación es comprobar la idoneidad de los componentes arquitectónicos de la B- DAPP y presentar algunos de estos resultados iniciales. Cabe destacar que los resultados obtenidos respaldan los hallazgos de la literatura. El objetivo futuro es realizar una evaluación más rigurosa mediante análisis predictivo, aprovechando los resultados preliminares del análisis presentados en este artículo.

471



472
473
474 Figura 4. Captura de pantalla del submódulo
475
476

477 5.1. Distribución de lesiones por partes del cuerpo

478 Dado que el conjunto de datos de Salud y Seguridad incluye la variable de tipo de operación, que describe el tipo de operación (levantar, tirar, cortar, etc.) con la herramienta (equipo) específica para la tarea dada.
480 Comprender la distribución de lesiones por partes del cuerpo puede resaltar las operaciones principales, por ejemplo , que resultan en accidentes en partes del cuerpo. Una herramienta estadística gráfica (gráfico circular) para explorar 482 esta información se muestra en la Figura 5, donde se observa que ciertas partes del cuerpo son propensas 483 a lesiones durante la construcción del proyecto de infraestructura eléctrica. La distribución de lesiones de las 484 5 partes del cuerpo principales, como se especifica en la base de datos, es la siguiente: Dedos (23 %), Mano (13 %), 485 Espalda/Glúteos (12 %) y Tobillo (8 %). Las cinco operaciones principales que resultan en estas lesiones son 486 tirar (encordar), levantar, cargar/descargar, manipulación manual y cortar porque estas partes 487 son esenciales para llevar a cabo estas operaciones (Chi y Han 2013). La observación de esto es 488 probablemente que la mayoría de los accidentes son el resultado de descuido, distracciones e indiferencia 489 por los procedimientos de seguridad. Los resultados del análisis exploratorio concuerdan con Fan et al. (2014).
490 Este conocimiento detallado no sólo es fundamental para el desarrollo de una gestión sólida de riesgos de salud y seguridad en la construcción, sino que también es fundamental para que las partes interesadas apliquen las mejores prácticas de seguridad para minimizar los accidentes.

493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509

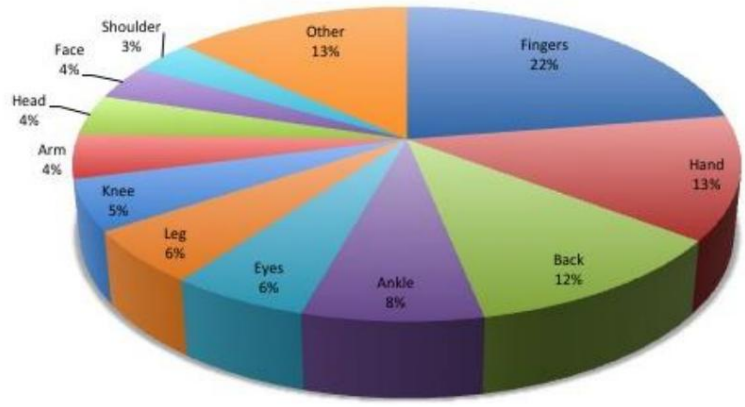


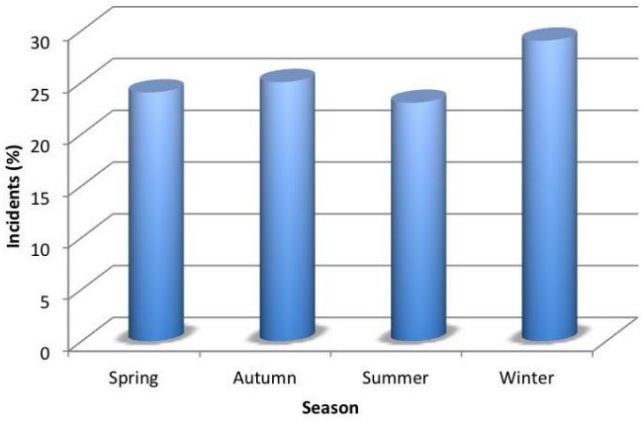
Figura 5. Distribución de lesiones por partes del cuerpo

510 5.2. Distribución de incidentes por temporada

La construcción de infraestructura eléctrica (por ejemplo, líneas aéreas) es principalmente una actividad al aire libre, y ciertos tipos de accidentes son más probables debido a las condiciones estacionales cambiantes (verano, invierno, otoño y primavera). La Figura 6 muestra que el invierno presenta el mayor porcentaje de incidentes (29%), seguido del otoño (25%), la primavera (24%) y el verano (23%). Escocia tiene un clima templado y oceánico , muy frío en invierno debido a las frecuentes e intensas lluvias de granizo y nieve.
516 Gales también tiene un clima templado y tiende a ser más húmedo que Inglaterra.
Los tropiezos, resbalones y caídas se encuentran entre los incidentes más comunes en estas regiones debido a la visibilidad reducida . Las temperaturas cercanas o inferiores al punto de congelación y los fuertes vientos también pueden provocar enfermedades y lesiones graves. Además, los accidentes de tráfico ocurren debido a los efectos del hielo y la nieve en el barro .
520 carreteras.

521 El uso de análisis de Big Data para la extracción y difusión automática de las condiciones climáticas de una región en tiempo real contribuirá en gran medida a mitigar los daños que son propios de esa región (ubicación).

524

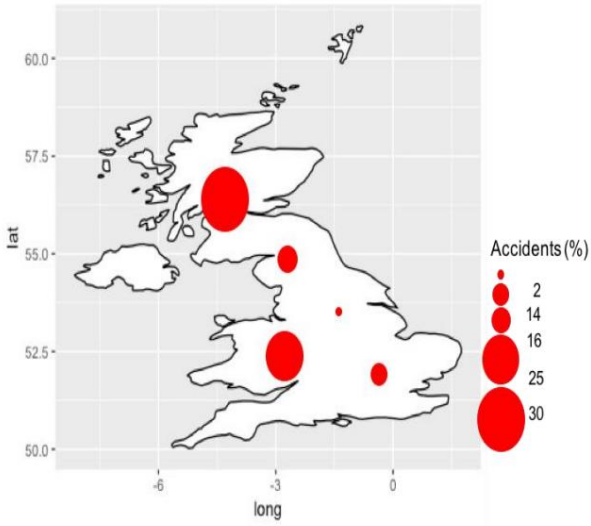


525
526 Figura 6. Distribución de incidentes por temporada
527
528 5.3. Distribución de accidentes mediante análisis espacial

529 A menudo, la alta dirección de una empresa constructora puede estar interesada en regiones con altas tasas de incidentes . Ofrecer este servicio proporcionará a los gerentes la información adecuada para reaccionar proactivamente a los desafíos de salud y seguridad en dichas regiones. Por lo tanto, el análisis espacial es de suma importancia en tales situaciones, ya que permite el análisis de incidentes en toda la distribución topológica y geográfica. En el conjunto de datos de salud y seguridad, la información de ubicación se captura en la columna "sitio". Para el análisis espacial, el conjunto de datos se preprocesa para extraer el código postal del Reino Unido de cada registro de incidente y se vincula con los datos de latitud y longitud correspondientes de Doogal (http://www.doogal.co.uk/UKPostcodes.php). El mapa de calor geográfico se utiliza para visualizar los datos resultantes. La Figura 7 muestra el resumen de esta distribución, donde el tamaño de las esferas representa la proporción de accidentes (calculada como porcentajes) en cada región. Escocia tiene el más alto (30%), seguido de Gales y Sur 540 Oeste (25%), Norte (16%), Sureste (14%) y Midlands (2%). Se observa que la frecuencia del clima severo 541 es la principal causa de accidentes en Escocia, así como en Gales y las regiones 542 Suroeste. El viento fuerte, por ejemplo, puede provocar la rotura de parabrisas de vehículos 543 y el derrumbe de una cerca o unidad. El clima helado puede provocar tropezones y resbalones. Además, la operación de maquinaria pesada 544 (es decir, excavación y corte de carreteras) es a menudo la causa de daños a los servicios públicos 545 (es decir, tuberías de gas, suministro de agua). Aunque las condiciones geológicas en diferentes ciudades 546 son complejas, los enfoques existentes de gestión de riesgos de salud y seguridad no consideran hacer 547 que esta información esté disponible para una prevención adecuada de riesgos de salud y seguridad. Para controlar eficazmente los riesgos para la salud y la seguridad en la obra, la incorporación de un módulo para calcular automáticamente el estado geológico e hidrológico de las obras en tiempo real mejorará el control óptimo de los riesgos laborales.

552
553
554

555



556

557 Figura 7. Análisis espacial de los accidentes

558

559

560 Además, el resultado de analizar las regiones en relación con la tasa de incidentes (o accidentes) puede refinarse aún más a ciudades y una ubicación específica. Merece la pena explorar más a fondo el impacto de la ubicación en los incidentes . Esta investigación se centra en futuras investigaciones sobre la arquitectura propuesta .

564

565 5.4. Modelado de la relación entre variables

Se han realizado enormes esfuerzos de I+D para reducir los impactos de los riesgos para la salud ocupacional . Uno de esos intentos consiste en modelar y analizar varias variables (es decir, determinar las relaciones entre los predictores (variables independientes) y la variable dependiente).

Se emplean técnicas de aprendizaje automático robustas y eficientes, como el aprendizaje profundo, las máquinas de potenciación de gradiente y la regresión lineal multivariante, para modelar las relaciones entre variables . En este artículo, se presenta una demostración de la técnica de regresión lineal debido a su simplicidad .

573

574 La regresión multivariante lineal, en este sentido, promueve métodos para analizar los riesgos para la salud 575 con respecto al coste del proyecto. Este concepto no solo permite el análisis exploratorio de lesiones 576 , sino que también permite el análisis predictivo de accidentes. El principio de la regresión multivariante lineal consiste en

577 predecir Y como una combinación lineal de las variables de entrada (1 , 2 , ,) más un término de error .

578 $= {}_{0} + {}_{1}{}_{1} + {}_{2}{}_{2} + + + ,$ [1,]

579 n es el número de datos de muestra, p el número de variables y 580 se puede un sesgo. Este modelo puede 0 escribir convenientemente como = + , dónde

581 $= ({}_{1,} ,) , = ({}_{1,} ,) , = ({}_{1,} ,) ,$ $y = (1 \; {}_{1} \; {}_{1} \; {}_{1} \; {}_{2} \; {}_{11} \; {}_{21} \; {}_{1} \; {}_{1} \; {}_{1})$

582

583 El valor predicho o ajustado es, por tanto, = , donde es la estimación de mínimos cuadrados de .

584 El modelo se puede utilizar, por ejemplo, para predecir la parte del cuerpo lesionada dado un conjunto de entradas tales como 585 el tipo de operación (tarea), equipo utilizado, tipo de proyecto de infraestructura eléctrica, la 586 complejidad del proyecto, tipo de contrato del proyecto, etc. Una ilustración práctica pero directa es 587 determinar la relación entre el costo del proyecto y los riesgos laborales (regresión lineal 588 con un predictor) se representa utilizando un gráfico de líneas (Figura 8). El eje x del gráfico 589 representa el costo del proyecto mientras que el eje y representa el riesgo de riesgos para la salud (incidentes y 590 accidentes). El gráfico de líneas muestra un aumento significativo en el número de riesgos para la salud (accidentes 591 e incidentes) a medida que aumenta el costo del proyecto. En consecuencia, el número de riesgos para la salud 592 ocupacional es proporcional al costo del proyecto. Este resultado es esperado ya que el costo del proyecto es un factor crucial 593 para determinar la complejidad de un proyecto. Así, cuanto más complejo es un proyecto, más 594 incidentes hay asociados a él.

595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621

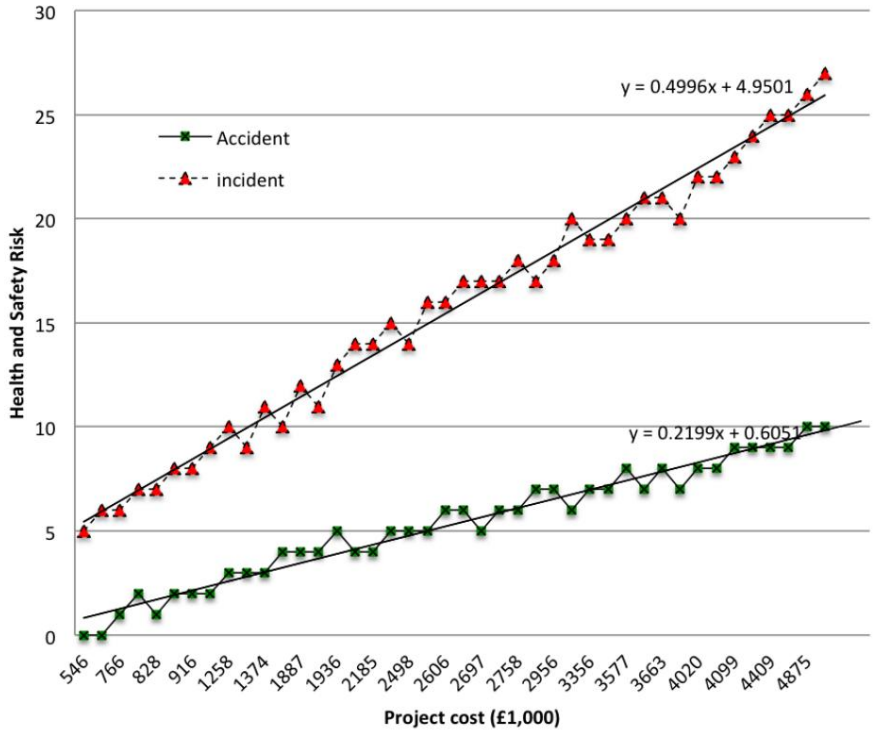


Figura 8. Relación entre variables

622 6. Conclusiones

Los análisis de riesgos de seguridad en la construcción son actualmente limitados debido a que las técnicas existentes pasan por alto la naturaleza compleja y dinámica de las obras. Además, ignoran o simplifican algunos factores clave y prestan poca atención al análisis de la relación entre un fenómeno de seguridad y los datos de seguridad. Hoy en día, se deben analizar datos grandes y dinámicos de diversos tipos. Al implementar la herramienta de gestión de riesgos para la salud, se propone la arquitectura de Big Data, basada en un ciclo de vida de análisis de riesgos para la salud coherente. La tecnología de Big Data fue seleccionada por su compatibilidad con datos masivos, de alta dimensión, heterogéneos, complejos , no estructurados, incompletos y con ruido.

Los resultados preliminares obtenidos en este estudio, utilizando diversos marcos de Big Data, nos han permitido diseñar una arquitectura robusta para gestionar y analizar datos de accidentes en infraestructuras eléctricas . La arquitectura propuesta puede identificar variables relevantes y mejorar la precisión de las predicciones preliminares y la capacidad explicativa. También ha permitido extraer conclusiones sobre las causas de los riesgos para la salud. Los resultados obtenidos en este estudio representan una mejora significativa en la gestión de la información sobre accidentes en la construcción, en particular para las empresas de infraestructuras eléctricas. Los resultados satisfactorios de la herramienta B-DAPP han indicado la fiabilidad e idoneidad de los componentes de Big Data seleccionados para

640 estudios sobre riesgos para la salud en la construcción y sus causas.

641

Las investigaciones futuras se centran en evaluar rigurosamente la precisión tanto de la predicción como de la prescripción del software implementado en tiempo real. Además, otros investigadores deberían centrarse en el diseño y la planificación de modelos más ambiciosos y a mayor escala para comprender mejor las causas de los accidentes en diversos sectores industriales.

647

648 Referencias

649 Al-Jarrah, OY et al., 2015. Aprendizaje automático eficiente para Big Data: una revisión. Big Data 650 651 Bohle, P. et Investigación, 2(3), pp.87–93.

al., 2015. Salud y bienestar de los trabajadores mayores: comparación de sus asociaciones 652 653 654 Camann, DE et al., Con desequilibrio esfuerzo-recompensa y presión, desorganización y fallos regulatorios. Trabajo y Estrés, 2011. 8373, pp. 1-14.

Ciencia de datos y análisis de Big Data, Nueva York: EMC.

655 Carbonari, A., Giretti, A. y Naticchia, B., 2011. Un sistema proactivo para la gestión de la seguridad en tiempo real en obras de construcción. Automatización en la Construcción, 20(6), pp. 686–698. 657 Disponible en: http://dx.doi.org/10.1016/j.autcon.2011.04.019.

658 Chaudhuri, S. y Dayal, U., 1997. Una descripción general del almacenamiento de datos y la tecnología OLAP. 659 ACM SIGMOD Rec, 26, págs. 65–74.

660 Chen, J., Qiu, J. y Ahn, C., 2017. Reconocimiento de posturas incómodas en trabajadores de la construcción mediante la descomposición tensorial de movimiento supervisada. Automatización en la Construcción, 77 , págs. 67–81.

662 Cheng, C. et al., 2011. Aplicación de técnicas de minería de datos para explorar los factores que contribuyen a las lesiones laborales en la industria de la construcción de Taiwán. Análisis y Prevención de Accidentes, 48, págs. 214-222 .

665 Chi, S. y Han, S., 2013. Análisis de la teoría de sistemas para la prevención de accidentes en la construcción, con referencia específica a los informes de accidentes de la OSHA. Revista Internacional de Gestión de Proyectos , 31 (7), págs. 1027-1041 .

668 Ciarapica, FE y Giacchetta, G., 2009. Clasificación y predicción de lesiones laborales 669 670 671 Debnath, J. et al., 2016. Riesgo mediante técnicas de soft computing: un estudio italiano. Safety Science, 47(1), pp.36–49. Disponible en: http://dx.doi.org/10.1016/j.ssci.2008.01.006.

Modelo de inferencia difusa para evaluar riesgos laborales en 672 sitios de construcción. Revista Internacional de Ergonomía Industrial, 55, págs. 114-128.

673 Delen, D. y Demirkan, H., 2013. Datos, información y análisis como servicios. Decision 674 Support Systems, 55(1), págs. 359–363.

675 Esmaeili, B., Hallowell, MR y Rajagopalan, B., 2015. Riesgo de seguridad basado en atributos 676 677 678 Fan, Evaluación. II: Predicción de resultados de seguridad mediante modelos lineales generalizados. Revista de Ingeniería y Gestión de la Construcción, 141(8), pp. 1-11.

ZJ et al., 2014. Asociación entre la combinación de fuerza manual y antebrazo 679 680 681 Favarò, FM y Saleh, JH, 2016. Postura e incidencia de epicondilitis lateral en la población laboral. Factores humanos, 56, págs. 151-165. Hacia

la evaluación de riesgos 2.0: Control de supervisión de seguridad 682 y monitoreo de peligros basado en modelos para intervenciones de seguridad informadas por el riesgo. Confiabilidad 683 Ingeniería y Seguridad del Sistema, 152, págs. 316–330. Disponible en: 684 http://dx.doi.org/10.1016/j.ress.2016.03.022.

685 Fenrick, L. y Getachew, S., 2012. Comparaciones de costos y confiabilidad de sistemas subterráneos y 686 687 Fragiadakis, N., líneas eléctricas aéreas. Política de servicios públicos, 20(1), págs. 31–37.

Tsoukalas, V. y Papazoglou, V., 2014. Una inferencia neurodifusa adaptativa 688 689 690 Galizzi, M. y Tempesti, T., 2015. Tolerancia al Modelo del sistema de información de riesgos laborales (ANFIS) para la evaluación de riesgos laborales en la industria de la construcción naval. Safety Science, 63, págs. 226-235.

riesgo de los trabajadores y lesiones ocupacionales. Análisis de Riesgo , 35(10), págs. 1858–1875.

692 Gandomi, M. y Haider, A., 2015. Más allá de la publicidad: conceptos, métodos y análisis de big data . Revista internacional de gestión de la información, 35(2), págs. 137-144.

694 García-Herrero, S. et al., 2012. Condiciones laborales, síntomas psicológicos/físicos y accidentes laborales. Modelos de redes bayesianas. Safety Science, 50 (9), pp. 1760–
696 1774.

697 Gholizadeh, P. y Esmaeili, B., 2016. Aplicación de árboles de clasificación para analizar accidentes de contratistas eléctricos . En el Congreso de Investigación de la Construcción. San Juan, Puerto Rico, págs. 2699–2708 .

Groves, W., Kecejovic, V. y Komljenovic, D., 2007. Análisis de fatalidades y lesiones que involucran 701 702 Guo, B., Yiu, T. y González, V., Equipos de minería. Revista de Investigación en Seguridad, 38(4), pp.461–470.

2016. Predicción del comportamiento de seguridad en la industria de la construcción: 703 Desarrollo y prueba de un modelo integrador. Safety Science, 84, págs. 1–11. Disponible en : http://dx.doi.org/10.1016/j.ssci.2015.11.020.

Guo, S. et al., 2016. Una plataforma basada en macrodatos del comportamiento laboral: Observaciones de campo. Análisis y Prevención de Accidentes, 93, pp. 299-309 . Disponible en: http://dx.doi.org/10.1016/j.aap.2015.09.024 .

708 Gürcanli, G. y Müngena, U., 2009. Un método de análisis de riesgos de seguridad laboral en 709 obras de construcción mediante conjuntos difusos. Revista Internacional de Ergonomía Industrial, 39(2), 710 , págs. 371–387.

Hallowell , MR y Gambatese, JA, 2009. Mitigación de riesgos de seguridad en la construcción. Revista de Ingeniería y Gestión de la Construcción, 135 (12), págs. 1316-1323. Disponible en: http://ascelibrary.org/doi/10.1061/%28ASCE%29CO.1943-7862.0000107 .

714 Haslam, RA et al., 2005. Factores contribuyentes en accidentes de construcción. Ergonomía Aplicada, 715 36(4), págs. 401–415.

716 HSE, 2016. Estadísticas resumidas sobre salud y seguridad en el trabajo para Gran Bretaña. Disponible en: http://www.hse.gov.uk/statistics/overall/hssh1516.pdf?pdf=hssh1516 [Consultado el 14 de abril de 2017].

719 Jacinto, C. y Silva, C., 2010. Una evaluación semicuantitativa de los riesgos laborales utilizando 720 721 Jin, X. et al., 2015. Representación de pajarita. Safety Science, 48, págs. 973–979.

Importancia y desafíos de la investigación de Big Data. Big Data Research, 722 2(2), págs. 59–64.

Jocelyn, S. et al., 2017. Aplicación del análisis lógico de datos a la prevención de accidentes relacionados con maquinaria con base en datos escasos. Ingeniería de Confiabilidad y Seguridad de Sistemas, 159 (mayo de 2016), pp. 223-236 . Disponible en: http://dx.doi.org/10.1016/j.ress.2016.11.015.

726 Khakzad, N., Khan, F. y Amyotte, P., 2015. Accidentes mayores (cisnes grises): Modelado de probabilidad 727 mediante precursores de accidentes y razonamiento aproximado. Análisis de riesgos, 35(7), 728 , págs. 1336-1347.

729 Landset, S. et al., 2015. Un estudio de herramientas de código abierto para aprendizaje automático con big data. 730 731 Leavitt, N., El ecosistema Hadoop. Journal of Big Data, 2(1), pp.1–36.

2010. ¿Cumplirán las bases de datos NoSQL su promesa? IEE Computer Journal, 732 43(2), pp. 12-14.

733 Li, H. et al., 2016. Modelo de secuencia de estados estocásticos para predecir estados de seguridad en obras de construcción. 734 735 mediante sistemas de localización en tiempo real. Safety Science, 84, págs. 78-87.

Li, Y. y Bai, Y., 2008. Comparación de características entre accidentes mortales y con lesiones en las zonas de construcción de carreteras. 736 Safety Science, 46(4), págs. 646–660.

737 Liao, C.-W. y Perng, Y.-H., 2008. Minería de datos para lesiones ocupacionales en la industria de la construcción de Taiwán. Safety Science, 46(7), págs. 1091–1102.

739 Liu, H. y Tsai, Y., 2012. Un enfoque de evaluación de riesgos difusos para riesgos laborales en la industria de la construcción. Safety Science, 50(4), pp. 1067–1078 . Disponible en: http://dx.doi.org/10.1016/j.ssci.2011.11.021 .

742 Love, PED y Teo, P., 2017. Análisis estadístico de frecuencias de lesiones y no conformidad 743 744 en construcción: Modelo de regresión binomial negativa. Revista de ingeniería y gestión de la construcción, 143(8), pp.1–9.

745 McDermott, V. y Hayes, J., 2016. “Seguimos chocando con cosas”: La eficacia de los procesos de terceros para la prevención de impactos en oleoductos. En Actas de la undécima conferencia internacional sobre oleoductos 747 (IPC 2016). Calgary, Alberta, Canadá, págs. 1-10.
748 Naderpour, M., Lu, J. y Zhang, G., 2016. Evaluación de un sistema de apoyo a la toma de decisiones crítico para la seguridad 749 mediante medidas de conocimiento de la situación y carga de trabajo. Ingeniería de Confiabilidad y Seguridad de Sistemas 750 , 150, págs. 147-159. Disponible en: http://dx.doi.org/10.1016/j.ress.2016.01.024.
751 Najafabadi, MM et al., 2015. Aplicaciones del aprendizaje profundo y desafíos en el análisis de big data. 752 753 Nanda, G. et al., 2016. Revista de Big Data, 2(1), pp.1–21.
Soporte de decisiones bayesiano para la codificación de datos de lesiones laborales. Journal of Safety Research, 57 , págs. 71-82. Disponible en: 755 http://dx.doi.org/10.1016/j.jsr.2016.03.001.

756 Pääkkönen, P. y Pakkala, D., 2015. Arquitectura de referencia y clasificación de tecnologías , productos y servicios para sistemas de big data. Big Data Research, 2(4), 758, pp. 166-186. Disponible en: http://dx.doi.org/10.1016/j.bdr.2015.01.001.

759 Papazoglou, I. et al., 2017. Modelo cuantitativo de riesgo ocupacional: Riesgo único. Confiabilidad 760 Ingeniería y Seguridad de Sistemas, 160, págs. 162-173.
761 Papazoglou, I. et al., 2015. Evaluación de la incertidumbre en la cuantificación de las tasas de riesgo de 762 763 Pinto, A., Accidentes de trabajo. Análisis de riesgos, 35(8), pp.1536–1561.
Nunes, I. y Ribeiro, R., 2011. Evaluación de riesgos laborales en la industria de la construcción 764 – Panorama general y reflexión. Safety Science, 49, págs. 616–624.
765 Power, D., 2014. Uso de “Big Data” para análisis y soporte de decisiones. Journal of Decision 766 767 Rahman, MN y Sistemas, 23(2), pp.222–228.
Esmailpour, A., 2016. Una arquitectura de centro de datos híbrido para Big Data. Big 768 769 Raviv, G., Shapira, A. y Fishbain, B., 2017. Investigación de datos, 3, págs.29–40.
Análisis basado en AHP del potencial de riesgo de incidentes de seguridad: Estudio de caso de grúas en la industria de la construcción. Safety Science, 91, 771 págs. 298–309 . Disponible en: http://dx.doi.org/10.1016/j.ssci.2016.08.027.

772 Rivas, T. et al., 2011. Explicación y predicción de accidentes laborales utilizando técnicas de minería de datos . Ingeniería de confiabilidad y seguridad de sistemas, 96(7), págs. 739–747 .
774 Ryza, 0] S. et al., 2015. Análisis avanzado con Spark, Cambridge: O'Reilly,.
775 Schryver, J., Shankar, M. y Xu, S., 2012. Pasando del análisis descriptivo al causal: Estudio de caso 776 sobre el descubrimiento de conocimiento del almacén de indicadores de salud de EE. UU. En el Taller sobre Informática de la Salud de la ACM SIGKDD 777. Pekín, China, págs. 1-8.
Silva, SA et al., 2016. Prácticas organizacionales para el aprendizaje sobre accidentes laborales a lo largo de su ciclo de información. Safety Science, en prensa.
Soltanzadeh , A. et al., 2016. Análisis de lesiones humanas causadas por accidentes laborales: Un estudio de caso en industrias y obras de construcción. Revista de Ingeniería Civil y Tecnología de la Construcción, 7(1), pp. 1-7. Disponible en : http://academicjournals.org/journal/JCECT/article-abstract/15EEFC357741 .

784 Suthakar, U. et al., 2016. Una estrategia eficiente para la recopilación y el almacenamiento de grandes volúmenes de datos para computación. Journal of Big Data, 3(1), pp. 1–17.
786 Tixier, A.. et al., 2016. Aplicación del aprendizaje automático a la predicción de lesiones en la construcción. 787 Automatización en la construcción, 69, págs. 102-114.
788 Törner, M. y Pousette, A., 2009. Seguridad en la construcción: una descripción completa de las características de los altos estándares de seguridad en la construcción, desde la perspectiva conjunta de supervisores y trabajadores experimentados. Journal of Safety Research, 40(6), 791 , págs. 399-409.

792 Tsai, CW et al., 2015. Análisis de big data: una encuesta. Journal of Big Data, 2(1), págs. 1–32.
793 Venturini, L., Baralis, E. y Garza, P., 2017. Escalado de la clasificación asociativa para conjuntos de datos muy grandes. Journal of Big Data, 4(1). Disponible en: https://doi.org/10.1186/s40537-017-
795 0107-2.
796 Weng, J., Meng, Q. y Wang, DZW, 2013. Enfoque de regresión logística basada en árboles para la evaluación del riesgo de accidentes en zonas de trabajo. Análisis de Riesgos, 33(3), págs. 493–504 .
798 White, T., 2012. Hadoop: La guía definitiva, Sebastopol, CA: O'Reilly Media, Inc.
799 Wu, W. et al., 2010. Hacia un sistema autónomo de seguimiento en tiempo real de cuasi accidentes en obras de construcción. Automatización en la Construcción, 19(2), pp. 134-141. 801 Disponible en: http://dx.doi.org/10.1016/j.autcon.2009.11.017.
802 Yi, W. et al., 2016. Desarrollo de un sistema de alerta temprana para trabajos en el sitio en climas cálidos y Entornos: Un estudio de caso. Automatización en la Construcción, 62, pp. 101-113. Disponible en: http://dx.doi.org/
húmedos 803 804 0.1016/j.autcon.2015.11.003.

Yoon , YS, Ham, DH y Yoon, WC, 2016. Aplicación de la teoría de la actividad al análisis de accidentes humanos: Método y casos prácticos. Ingeniería de Confiabilidad y Seguridad de Sistemas, 150 , pp. 22-34. Disponible en: http://dx.doi.org/10.1016/j.ress.2016.01.013.

Yorio, PL, Willmer, DR y Haight, JM, 2014. Interpretación de las citaciones de la MSHA desde la perspectiva de los sistemas de gestión de la seguridad y salud ocupacional: Investigación de su impacto en las lesiones y enfermedades mineras (2003-2010). Análisis de Riesgos, 34(8), págs. 1538-1553.

811 Zang, W. et al., 2014. Estudio comparativo entre aprendizaje incremental y aprendizaje por conjuntos en flujos de datos 812 : Caso práctico. Journal Of Big Data, pp. 1-16. Disponible en: 813 http://www.journalofbigdata.com/content/1/1/5/abstract.

Zeng , SX, Tam, VWY y Tam, CM, 2008. Hacia sistemas de salud y seguridad ocupacional en la industria de la construcción de China. Safety Science, 46, págs. 1155-1168 .

816 Zhang, L. et al., 2016. Hacia un enfoque basado en redes bayesianas difusas para el análisis de riesgos de seguridad de daños en tuberías inducidos por túneles. Análisis de riesgos, 36 (2), págs. 278–301.

818 Zhou, Z., Goh, Y. y Li, Q., 2015. Resumen y análisis de estudios de gestión de la seguridad en la construcción. 819 Industria de la construcción. Safety Science, 72, págs. 337–350.

820 Zhu, Z. et al., 2016. Predicción de movimientos de trabajadores en obra y equipos móviles para mejorar la seguridad en la construcción. Automatización en la Construcción, 68, págs. 95-101.

822 Zou, PXW, Zhang, G. y Wang, J., 2007. Comprensión de los riesgos clave en la construcción 823 824 Proyectos en China. Revista Internacional de Gestión de Proyectos, 25(6), pp.601–614.