Implementación de Data Science: Metodología

Data Science Implementation: Methodology

Liza Zarahi Domínguez Acosta[a]*, Benito Alan Ponce Rodríguez[b], Javier Molina Salazar[a]

[a]Departamento de Ingeniería Industrial y Manufactura, Universidad Autónoma de Ciudad Juárez, México
[b]Departamento de Ingeniería Eléctrica y Computación, Doctorado en Tecnología, Universidad Autónoma de Ciudad Juárez, México
* Autor de correspondencia. Correo: ing.liza.dominguez@gmail.com

No. de resumen
2CP21-59

Formato
Ponencia

Evento
2.º Coloquio de Posgrados IIT

Presentador
Liza Zarahi Domínguez Acosta

Tema
Cómputo Aplicado

Estatus
Estudio en curso

Fecha de la presentación
Noviembre 12, 2021

RESUMEN

En la actualidad, organizaciones industriales, comerciales y de gobierno generan una gran cantidad de datos a través de la comunicación M2M (máquina a máquina), por medio de sensores, cámaras y/o archivos de registros; por los humanos, consultando redes sociales, o realizando búsquedas; y mediante la organización, utilizando los sistemas de planificación de recursos empresariales. A su vez, estos son obtenidos de diferentes medios, como archivos de texto, CSV y archivos multimedia. Esta información no se puede procesar utilizando métodos tradicionales, por lo cual se utiliza el campo de Data Science para encontrar patrones a través del análisis y tomar decisiones a partir de grandes volúmenes de datos (Big Data). La presente investigación consiste en aplicar Data Science a un conjunto de datos procedentes de los equipos de prueba eléctrica dentro de un proceso de manufactura electrónica. Para ello, se definieron cuatro pasos metodológicos: 1) Obtención de los datos, 2) Extracción de los datos, 3) Limpieza y, por último, 4) el análisis exploratorio. El primer método consiste en leer cada uno de los conjuntos de datos que se encuentran almacenados en un servidor en extensión CSV. El segundo método se enfoca en transformar los datos en un formato estructurado utilizando arreglos de la misma dimensión. El tercer método elimina información que no es de utilidad. Y el cuarto método tiene como objetivo utilizar herramientas gráficas, medidores de desempeño y estadísticas para identificar posibles causas que originan un problema. Se diseñará una interfaz de visualización Dash.

Palabras clave: Data Science; Big Data; análisis de datos; manufactura electrónica.

ABSTRACT

Currently, industrial, commercial and government organizations generate a large amount of data through M2M (machine-to-machine) communication, through sensors, cameras and / or log files; by humans, consulting social networks, or performing searches; and through organization, using enterprise resource planning systems. In turn, these are obtained from different media, such as text files, CSV and multimedia files. This information cannot be processed using traditional methods, which is why the field of Data Science is used to find patterns through analysis and make decisions from large volumes of data (Big Data). This research consists of applying Data Science to a set of data from electrical test equipment within an electronic manufacturing process. To do this, four methodological steps were defined: 1) Data collection, 2) Data extraction, 3) Cleaning and, finally, 4) exploratory analysis. The first method is to read each of the data sets that are stored on a server in CSV extension. The second method focuses on transforming the data into a structured format using arrays of the same dimension. Method 3 deletes information that is not useful. Method 4 aims to use graphical tools, performance meters and statistics to identify possible causes that originate a problem. A Dash display interface will be designed.

Keywords: Data Science; Big Data; data analysis; electronic manufacturing.

Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

Financiamiento

Beca CONACYT, CVU: 387595.

Conflictos de interés

Los autores declaran que no existe conflicto de intereses.