

NUEVA METODOLOGÍA CSKT PARA MEJORAR LOS PROYECTOS DE IMPLEMENTACIÓN DE MACHINE LEARNING EN INGENIERÍA INDUSTRIAL EN UNA UNIVERSIDAD PÚBLICA

NEW CSKT METHODOLOGY TO IMPROVE MACHINE LEARNING IMPLEMENTATION PROJECTS IN INDUSTRIAL ENGINEERING AT A PUBLIC UNIVERSITY

Tipo de Publicación: Artículo Científico

Recibido: 15/02/2025

Aceptado: 16/03/2025

Publicado: 28/03/2025

Código Único AV: e461

Páginas: 1 (140-151)

DOI: <https://doi.org/10.5281/zenodo.15102636>

Autores:


José Antonio Ogosi Auqui

Ingeniero de Sistemas y Cómputo

Maestría en Ingeniería, mención Gestión de Tecnologías de Información

Maestría en Ingeniería mención Inteligencia Artificial

Doctor en Ingeniería de Sistemas

 <https://orcid.org/0000-0002-4708-610X>

E-mail: jogosi@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Lima – Perú


Jorge Lira Camargo

Ingeniero de Sistemas

Maestría en Ingeniería mención Ingeniería Industrial y

Gestión Empresarial

Doctor en Ingeniería de Sistemas

 <https://orcid.org/0000-0003-2364-5226>

E-mail: jlira@unfv.edu.pe


Afiliación: Universidad Nacional Federico Villarreal

País: Lima - Perú

Francisca Sonia Vera Tito

Ingeniero de Sistemas

Maestría en Administración

 <https://orcid.org/0000-0002-7258-2391>

E-mail: fvera@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Lima – Perú


César Gerardo León-Velarde

Licenciado en Educación mención Filosofía y Ciencias Sociales

Maestría en Educación mención Evaluación de la Calidad Educativa

Maestría en Educación mención Gestión de la Educación

Doctor en Educación

 <https://orcid.org/0000-0002-8273-1995>

E-mail: cleon@unfv.edu.pe

Afiliación: Universidad Nacional Federico Villarreal

País: Lima – Perú

Resumen

La investigación propone una metodología tomando las mejores partes de los enfoques de CRISP-DM, SEMMA, KDD y TDSP, para ello primero se realizó una revisión sistemática, se orientó a un enfoque empresarial, tomando en consideración la pautas de la minería de datos, en el proceso de la validación piloto se realizó en una universidad pública para evaluar que la satisfacción de la propuesta del modelo, obteniendo un 67%, lo cual implica que el modelo tiene muchas oportunidades de mejorar y madurar para lograr un modelo de referencia. A pesar de haber implementado dentro de la carrera de Ingeniería Industrial, se logró determinar que dicho modelo puede lograr los mismos resultados o mejores en una empresa sea pública o privada. El modelo permite mostrar las actividades a seguir con un enfoque empresarial y se logre ser referente para las implementaciones de Machine Learning.

Palabras Clave

Modelo de referencia, Machine Learning, Implementación, Metodología CSKT, Empresas.

Abstract

The research proposes a methodology taking the best parts of the CRISP-DM, SEMMA, KDD and TDSP approaches, for this first a systematic review was conducted, it was oriented to a business approach, taking into consideration the guidelines of data mining, in the process of pilot validation was conducted in a public university to assess the satisfaction of the proposed model, obtaining 67%, which implies that the model has many opportunities to improve and mature to achieve a reference model. Despite having been implemented within the Industrial Engineering career, it was determined that the model can achieve the same or better results in a public or private company. The model allows to show the activities to follow with a business approach and to become a reference for Machine Learning implementations.

Keywords

Reference model, Machine Learning, Implementation, CSKT Methodology, Enterprises.

Introducción

A través de los años, el uso de la tecnología nos ha permitido poder recolectar grandes cantidades de datos, procesar la información y visualizar en tiempo real, actualmente con el uso de la inteligencia artificial se busca encontrar patrones o información oculta, que no es posible obtener mediante métodos estadísticos convencionales (Al-Anqoudi, et al., 2021).

Las metodologías se actualizan o se crean a necesidad de las diferentes disciplinas, para Bhardwaj et al., (2020), señala que la metodología hace referencia al conjunto de procedimientos basados en principios lógicos, utilizados para alcanzar una gama de objetivos que rigen una investigación científica. Esta definición aplicada a la explotación de la información permitirá visualizar información para la toma de decisiones.

El aprendizaje en máquina o también conocido como Machine Learning, es un concepto que permite aplicar miles de procesos de las metodologías, para este artículo vamos a mencionar cuatro metodologías (CRISP-DM, SEMMA, KDD y TDSP) que apoyarán la creación de la nueva metodología CSKT la cual, se encontrará orientada a la carrera de la Ingeniería Industrial, pero puede ser utilizada en otros ámbitos. El objetivo es diseñar una nueva metodología basada en CRISP-DM, SEMMA, KDD y TDSP para hacer posible la implementación de proyectos de Machine Learning

en Ingeniería Industrial o en otras disciplinas, además de que este artículo sea de consulta para los estudiantes que inician proyectos de aprendizaje profunda o máquina en relación con el almacenamiento, análisis, procesamiento y visualización de la información.

Para Cabello (2021), se llevó a cabo una encuesta para conocer el uso de la Inteligencia Artificial (IA) en las empresas, y los resultados mostraron que un 62% tiene planes de implementarla, mientras que el 38% ya la está utilizando. Las empresas que ya la aplican han experimentado mejoras en el rendimiento de sus procesos y una mayor innovación, gracias a la eficiencia en la gestión al reducir tareas repetitivas. Asimismo, la transformación digital en las organizaciones e instituciones permite acelerar el uso de la IA, aquello admitirá lograr un impacto en la manera de generar productos o servicios (Camargo, 2022), así como en las metodologías.

Para Dennis & Aizenberg (2022) se destaca que la Inteligencia Artificial está facilitando el proceso de filtrado de los currículums (CV) y la verificación de antecedentes, dos tareas que suelen ser rutinarias en el área de Recursos Humanos (RR.HH.). Otras aplicaciones de la IA es el reconocimiento facial, en la investigación de Lira (2016) se puede visualizar el paso a paso para llevar a cabo la identificación de las personas utilizando Python como herramienta (Camargo et al., 2022).

En base a los datos recabados, dentro de las investigaciones identificadas dentro del presente trabajo, se considera la importancia de emplear la tecnología de Inteligencia Artificial y sus ramas alternas como la minería de datos y Machine Learning para el tema de desarrollo de proyectos de ingeniería en donde esta ha demostrado ser una solución ante los múltiples problemas relacionados a la toma de decisiones y la adecuada selección de opciones que puedan beneficiar a la entidad y a quienes influyen dentro de estos proyectos.

La metodología CRISP-DM es considerada un modelo de referencia estándar en minería de datos, ya que cubre todas las etapas del ciclo de vida de un proyecto en esta área. Posee seis fases estructuradas que permiten obtener soluciones basadas en el análisis de datos que nos brinden resultados de forma eficiente de acuerdo con las demandas actuales (Velásquez & Alejandro, 2019). Con la metodología SEMMA se logra el entendimiento de negocio que permite resolver las dudas sobre los procesos que genera valor, con ello se logra la primera fase, con la segunda se busca comprender los datos para obtener una visión general de lo que se puede conseguir y tener una idea clara de la viabilidad, las siguientes dos fases son las soluciones técnicas como el modelado y evaluación de resultados, logrado obtener los KPI adecuados para la toma de decisiones, y termina con el despliegue o implementación (Venusamy et al., 2020).

El crecimiento exponencial de los datos en la era digital ha impulsado la necesidad de integrar tecnologías avanzadas como Machine Learning (ML) en las empresas. Sin embargo, muchas organizaciones enfrentan dificultades en la implementación de ML debido a la ausencia de un modelo de referencia estandarizado. La metodología CSKT surge como una solución a esta problemática, integrando las mejores prácticas de enfoques previos para ofrecer una guía adaptable y eficiente. Este documento presenta la estructura de la metodología CSKT y su aplicabilidad en diversos sectores empresariales.

Metodología

La metodología empleada tiene contemplado 4 partes esenciales las cuáles se explicarán a continuación.

Modelamiento Empresarial

Este consiste en la comprensión del negocio sobre el que se va a trabajar. Para ello es necesario identificar objetivos estratégicos y fuentes de datos relevantes de la organización. En esta fase se usan técnicas de modelamiento de procesos para estructurar la problemática y definir los requerimientos del proyecto de ML.

Preparación de Datos

Para esta fase, los datos que fueron recopilados de la organizaron pasarán filtros de limpieza, procesos de estructuración y validación. Se realiza ello, con el objetivo de garantizar su calidad y utilidad. Se utilizan herramientas como Power BI, Google Colab y Python para este proceso.

Exploración y Modelado

Luego de haber preparado los datos, se implementan análisis exploratorios y técnicas de modelado para extraer conocimiento de los datos aplicando algoritmos de aprendizaje automático. Los resultados de esta fase serán útiles para optimizar el modelo.

Visualización y Validación

Finalmente, se presentan los resultados mediante dashboards y reportes, facilitando la interpretación de los hallazgos. Se valida la efectividad del modelo con pruebas piloto y feedback de los usuarios finales.

Desarrollo

Fase I: Modelamiento Empresarial

La Fase I es fundamental para el entendimiento del negocio e identificación de las fuentes de datos, a continuación, detallaremos dos etapas:

Exploración del negocio:

A través del modelamiento de procesos o ingeniería empresarial podemos conocer las

diferentes actividades, tareas, equipos, tecnologías que maneja la organización, esta comprensión permitirá priorizar las metas a lograr con relación al análisis, procesamiento y visualización de la información para lograr los objetivos de la entidad. En la tesis Lira (2016), se puede visualizar la manera llevar a cabo el modelamiento de procesos de una entidad pública, de la misma forma se puede utilizar para cualquier organización.

Explotación de los datos:

A través de la identificación de los flujos de actividades, se puede identificar las fuentes de datos, aquellos pueden ser datos no estructurados, semi estructurados o estructurados, esto dará inicio para explorar las diferentes tecnologías para procesar la información.

Fase II: Preparación

Luego de haber identificado el tipo de datos se procede a limpiar la información para ser utilizado, asimismo, se evalúa la estructura de los datos con el flujo de procesos, se valida la redundancia de datos e identificación de campo claves para relacionar con otras tablas o reportes. Existen herramientas que ayudan a este proceso como el Power BI de Excel, Big Query, Google Colab, etc. En ocasiones se requiere de lenguajes de programación para preparar los datos como Python, Google Script, Java Script y otros.

Fase III: Exploración y modelado

Utiliza la fuente de datos para preparar, analizar y limpiar, aquello se repite las veces necesarias hasta lograr describir y predecir la información para la toma de decisiones, para ello se evalúa a través de muestras. El modelo trabaja directamente en combinación entre CRISP-DM, SEMMA, KDD y TDSP, logrando tomar lo mejor de cada una para proponer la metodología CSKT como se ilustra en la siguiente Figura 1.

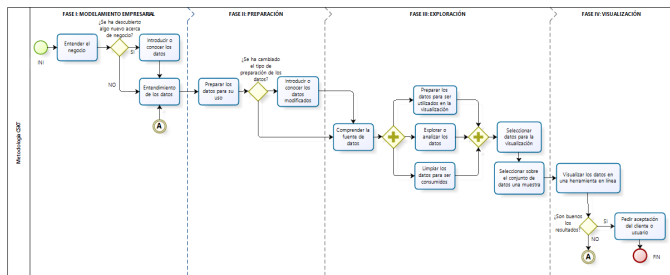


Figura 1. Diagrama de flujo para la implementación de la metodología CSKT

Se introduce una nueva propuesta metodológica con el objetivo de hacerla accesible para cualquier persona, facilitando así su disponibilidad a nivel global. La implementación de esta propuesta consistirá en diseñar una nueva metodología CSKT para adaptar los proyectos de implementación de Machine Learning en Ingeniería Industrial. El propósito principal es que los estudiantes de esta carrera puedan utilizar estas guías en la implementación de proyectos de Machine Learning.

Esta nueva metodología se diseñó teniendo en cuenta las necesidades específicas de los estudiantes, las cuales se identificaron en conjunto con el desarrollo de la metodología Design Thinking. Esta última se emprendió con el fin de tomar las tesis de Ingeniería Industrial como muestra para generar una metodología diseñada según sus necesidades y forma de trabajar. En la siguiente etapa, se lleva a cabo el marco de trabajo Scrum para gestionar eficazmente el diseño de la nueva metodología.

En lo que respecta a la metodología CSKT estará conformada por cuatro fases empezando con el modelamiento empresarial la cual se tendrá en cuenta el entendimiento del negocio considerando los aspectos principales del mismo, seguido de ello está la segunda fase que es la preparación la cual se realiza el entendimiento de los datos, es decir, tener la data que se empleará para el proyecto y que esta es capacitada para realizar los múltiples procesos, he de ahí la parte de preparación de los datos, una vez hecho esta actividad se realiza una reducción de toda la data total, es decir, tomar una parte significativa de esa data y proseguir con esta en la siguiente parte he de ahí la parte de tomar la muestra del conjunto de datos, luego, se realiza la tercera fase de exploración en donde una vez se toma esta parte de la data seleccionada se realiza un análisis completo de estos datos si estos son válidos para una proseguir con el proceso de la metodología.

Es decir, la preparación de esta data para luego realizar un profundo análisis de estos datos si son relevantes o presentan resultados positivos para la investigación, después de ello se procede a realizar filtros descartando y seleccionando la nueva data para emplearla en la minería de datos, por último se realiza la cuarta fase de visualización el cual se presenta este modelo de datos sintetizado al cliente para luego corroborar si se tiene la aprobación de este cumpliendo así, el modelo de esta nueva metodología realizando revisiones constantes de la data empleada para el proyecto (Ver Figura 2).

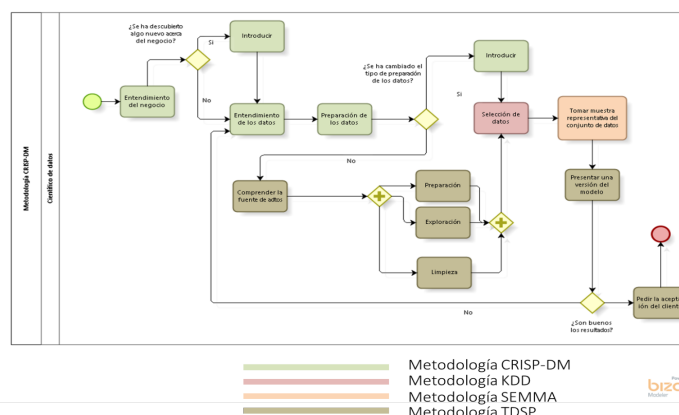


Figura 2. Planificación del Proyecto - Metodología CSKT

Con base en la Figura exhibida, se considera que la metodología comienza con la comprensión del negocio. En este punto, se toma una decisión en la cual, si se identifica algo novedoso para la empresa, se procede a incorporar estos puntos dentro del proyecto. Posteriormente, se avanza hacia la comprensión de los datos. En caso contrario, si no se descubre algo nuevo, se pasa directamente al entendimiento de los datos. Luego, se preparan los

datos y surge una condición: si hay cambios en el tipo de preparación de los datos, se integran estos cambios en el proyecto como registros. Después, se selecciona una muestra significativa de este conjunto de datos. Si no hay cambios, se lleva a cabo una comprensión de la fuente de los datos, realizando preparación, exploración y limpieza para obtener una fuente de datos clara y depurada.

A continuación, se procede a la selección de estos nuevos datos, seguido de la toma de una muestra significativa de esa data. Posterior a este proceso, se presenta la versión del modelo. En caso de que los resultados sean viables, se solicita la aceptación del cliente, con lo cual se concluye el proceso de la metodología. En caso contrario, se vuelve al entendimiento de los datos hasta que se cumplan los requisitos previos.

Fase IV: Validación y diseño experimental

La validación permitirá evaluar la madurez para llevar a cabo su implementación, por ello la metodología CSKT se puso a prueba con estudiantes de Ingeniería Industrial, quienes evaluaron su utilidad en diversas aplicaciones. Por ello, se elaboró un instrumento con 8 categorías para ser sometidas, de manera de evaluar los indicadores de KPI's. A continuación, se muestra en la Tabla 1 las preguntas de evaluación.

Categoría	Evaluación	Observaciones
Automatización y Planificación	Utilización de mecanismos simples para	Alta aceptación

	planificación de tareas.	
Digitalización y Análisis de KPI's	Búsqueda de información para digitalización de KPI's.	Alta utilidad
Optimización de Procesos	Ayuda en la automatización laboral y eliminación de tareas repetitivas.	Beneficio claro
Modelos Predictivos y Costos	Identificación de elementos en modelos predictivos y análisis de costos.	Aplicabilidad destacada
Auditoría y Gestión Empresarial	Ejecución de instrumentos de capacitación y auditoría empresarial.	Relevante para organizaciones
Visualización y Toma de Decisiones	Análisis de predicciones de demanda y satisfacción del cliente.	Soporte clave
Predicción de Precios y Finanzas	Explicación y modelado de predicción de precios del cobre.	Utilidad estratégica
Mejora Continua y Validación	Evaluación de la precisión del modelo predictivo.	Validación confiable

Tabla 1. Prueba piloto

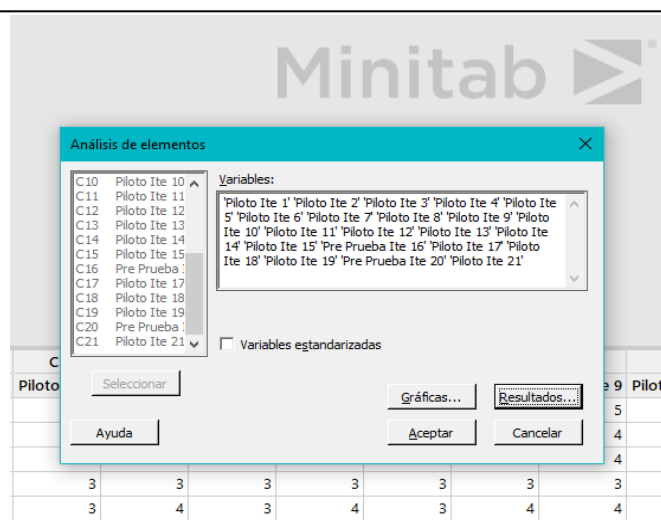


Figura 3. Análisis estadístico de los datos extraídos de la prueba piloto

El análisis de los datos obtenidos de los estudiantes, quienes respondieron a las 8 categorías mediante la herramienta Minutaba, se realiza con base en la Figura 4, obteniendo los siguientes resultados.



Figura 4. Análisis de elementos de Piloto

Rangos	Magnitud
0,81 a 1,00	Muy Alta
0,61 a 0,80	Alta
0,41 a 0,60	Moderada
0,21 a 0,40	Baja
0,01 a 0,20	Muy Baja

Tabla 2. Tabla de rangos para los resultados de confiabilidad

Nota: Según la interpretación basada en Sampieri et al., (2014), al obtener un valor de alfa de Cronbach de 0.9683, se concluye que la encuesta es confiable, lo que permite continuar con esta fase utilizando las 36 encuestas requeridas (Ver Tabla 2).

Resultados y discusión

Bajo el modelo CANVAS se ha conseguido varios objetivos del proyecto. A continuación, puede visualizar el detalle en la Figura 5.

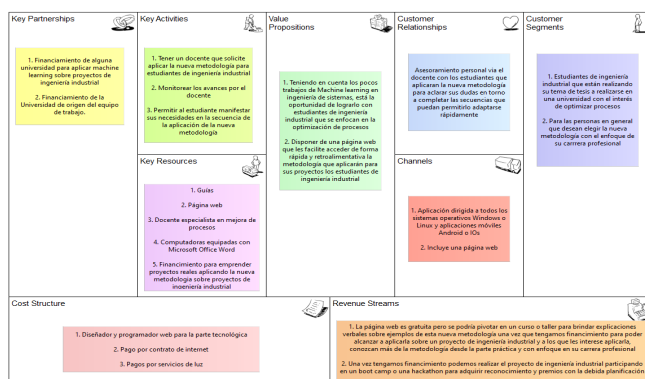


Figura 5. Modelo CANVAS de la metodología propuesta

Luego de la encuesta realizado a los estudiantes de una carrera profesional, se logró identificar que el 67% se encuentra totalmente de acuerdo, y el 22% totalmente en desacuerdo, este último indicador permitirá seguir mejorando la metodología para fortalecer su implementación, esta prueba piloto permite identificar oportunidades de mejora para llevar a cabo la madurez de la metodología.



Figura 6. Modelo CANVAS de la metodología propuesta

El modelo CANVAS de la metodología propuesta permite identificar que dicho modelo puede ser implementado en otros sectores como las instituciones privadas, y a través de las recomendaciones se puede lograr obtener un modelo de referencia para la implementación de la metodología CSKT en Machine Learning.

Conclusiones

Después de completar el análisis de investigación, se destacan como elementos clave las metodologías CRISP-DM, SEMMA, KDD y TDSP. Estas pueden ser adoptadas de manera independiente o, en ciertos casos, se puede elegir una sola para mejorar un proceso utilizando Machine Learning. A pesar de su efectividad, existen otros campos, como la Ingeniería Industrial, que requieren un enfoque más adaptado.

De esta forma, se propone el desarrollo de una metodología innovadora llamada CSKT, que buscar reunir las ventajas de las metodologías mencionadas anteriormente, destinada a facilitar la implementación de proyectos de Machine Learning dentro de la Ingeniería Industrial. Tomando como

base los resultados, se tiene como objetivo evaluar en qué medida la metodología propuesta facilita la integración de Machine Learning en procesos que deben ser automatizados, con el feedback de los alumnos sobre su satisfacción con CSKT, que sigue en proceso de desarrollo.

También, se aplicó el enfoque Design Thinking, que sirvió como base para proponer mejoras en la metodología CSKT, por ejemplo, al utilizarla repetidamente para ayudar a los estudiantes de Ingeniería Industrial a generar nuevas propuestas para sus proyectos de tesis. Al hablar de metodologías, nos referimos también a su uso en el desarrollo y enseñanza de proyectos específicos. Las diferencias entre las metodologías tradicionales de Machine Learning deben ser aprovechadas para beneficiar a todos los estudiantes de Ingeniería Industrial.

Si bien dentro de la investigación se logró emplear esta nueva metodología y se realizaron pruebas mediante encuestas para los estudiantes sobre si hubo una mayor ventaja en emplear esta metodología frente a otras para proyectos de Machine Learning, hubo algunas limitaciones en base al desarrollo y diseño de esta metodología los cuales implicaron las pruebas iniciales y selección de pasos los cuales puedan relacionarse con pasos anteriores haciendo que esta sea más efectiva y eficiente al momento de realizar proyectos de implementación siguiendo el diseño estructurado

por lo que, se tuvo que emplear varias pruebas haciendo un análisis de qué aspectos se puedan extraer de las otras metodologías y se puedan incorporar en la metodología CSKT siendo esta una de las principales limitaciones al momento de proceder con la investigación.

Otra limitación identificada en esta investigación fue la selección de los candidatos a emplear esta nueva metodología puesto que la mayoría de los estudiantes que estaban dispuestos a realizar proyectos de implementación de Machine Learning, empleaban las metodologías tradicionales en base a la experiencia propia lo cual se les hacía más fácil desarrollar dichos proyectos sumado al temor de que al emplear esta nueva metodología, podrían no llegar a entender en algunos de los pasos que se realizan dentro de esta por ello, es que para futuras investigaciones se debe de instruir a la población seleccionada de la investigación sobre la solución que se esté realizando y cuáles son las ventajas que pueda presentar frente a otras soluciones alternas.

Referencias

- Al-Anqoudi, Y., Al-Hamdani, A., Al-Badawi, M., & Hedjam, R. (2021). Using Machine Learning in business process re-engineering. *Big Data & Cognitive Computing*, 5(4), 61.
- Bhardwaj, G., Singh, S. V., & Kumar, V. (2020). An Empirical Study of Artificial Intelligence and its Impact on Human Resource Functions. En 2020 International Conference on Computation, Automation and Knowledge Management

- (ICCAKM) (pp. 47-51). Documento en línea. Disponible. doi: 10.1109/ICCAKM46823.2020.9051544. Disponible. doi: 10.1109/ICISS49785.2020.9315881.
- Cabello, J. (2021) ¿Cuánto cuesta realmente una página web? Descubre su precio justo. Diseñador Web Experto en Funnels
- Camargo, J. L. (2022). Propuesta e implementación de un modelo para el rediseño organizacional y su influencia en la transformación digital en el sector público. Un Camino hacia la transformación digital; Road to digital transformation (p. 5). SESU Editorial.
- Camargo, J. L., Baca, L. D. H., Valencia, E. T., Aquino, R. E. A., Miranda, A. G. R., & Camargo, L. G. L. (2022). Facial recognition proposal with the use of Python. In 2022 17th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-5). IEEE.
- Dennis, M. J., & Aizenberg, E. (2022). The Ethics of AI in Human Resources. Ethics and Information Technology, 24(3). Documento en línea. Disponible. <https://doi.org/10.1007/s10676-022-09653-y>
- Lira Camargo, J. (2016). Rediseño e implementación de escalafón y su influencia en la calidad de atención a los usuarios en una entidad del estado, Periodo 2014-2015.
- Sampieri, R. H., Collado, C. F., Lucio, P. B., Valencia, S. M., & Torres, C. P. M. (2014). Metodología de la investigación. McGraw Hill España.
- Velásquez, F., & Alejandro, D. (2019). Técnicas de Machine Learning aplicado a la predicción de fuga de clientes en una institución de salud previsional bajo la metodología KDD. Universidad Andrés Bello.
- Venusamy, K., Krishnan Rajagopal, N., & Yousoof, M. (2020). A study of Human Resources Development through Chatbots using Artificial Intelligence. En 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) (pp. 94-99). Documento en línea.