

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



Sistema web basado en machine learning como apoyo a la identificación del nivel de aprendizaje de artesanos del CITE Sipán

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Angela Yanixa Guzman Huaman

ASESOR

Luis David Robles Pizarro

<https://orcid.org/0000-0003-0392-3292>

Chiclayo, 2025

**Sistema web basado en machine learning como apoyo a la
identificación del nivel d aprendizaje de artesanos del CITE Sipán**

PRESENTADA POR

Angela Yanixa Guzman Huaman

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Jury Yesenia Aquino Trujillo
PRESIDENTE

Ernesto Ludwin Nicho Cordova
SECRETARIO

Luis David Robles Pizarro
VOCAL

Dedicatoria

Agradezco a Dios y a mi madre por el apoyo incondicional a lo largo de mis estudios.

Agradecimientos

Mis agradecimientos al equipo humano perteneciente al CITE SIPAN quienes me apoyaron con los recursos necesarios en la elaboración de mi proyecto, así mismo a los docentes asesores quienes me guiaron y orientaron.

SISTEMA WEB BASADO EN MACHINE LEARNING COMO APOYO A LA IDENTIFICACIÓN DEL NIVEL DE APRENDIZAJE DE ARTESANOS DEL CITE SIPAN

INFORME DE ORIGINALIDAD

24%

INDICE DE SIMILITUD

22%

FUENTES DE INTERNET

8%

PUBLICACIONES

9%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1

hdl.handle.net

Fuente de Internet

6%

2

tesis.usat.edu.pe

Fuente de Internet

3%

3

Submitted to Universidad Tecnológica
Centroamericana UNITEC

Trabajo del estudiante

1%

4

unividafulp.edu.co

Fuente de Internet

1%

5

www.risti.xyz

Fuente de Internet

1%

6

revistasojs.utn.edu.ec

Fuente de Internet

1%

7

Mohamed El Jihaoui, Oum El Kheir Abra,
Khalifa Mansouri, Moulay El Houssine Ech-
Chhibat. "Towards a Literature Review

Methodology: A Practical Guide in the Context

1%

Índice

Resumen	6
Abstract	7
Introducción.....	8
Revisión de literatura.....	10
Materiales y métodos	14
Resultados y discusión	16
Conclusiones	17
Recomendaciones	18
Referencias.....	19

Resumen

En el contexto de un centro de innovación tecnológica, se abordó la problemática de identificar adecuadamente el nivel de aprendizaje de los artesanos en el CITE Sipán, un centro de innovación en Lambayeque que apoya a los artesanos mediante capacitaciones y asistencia técnica. La falta de un sistema adecuado para caracterizar a los artesanos generaba ineficiencias en la asignación de recursos y en la adaptación de los servicios a las necesidades específicas de cada artesano.

Para resolver esta situación, se implementó un sistema web basado en Machine Learning como apoyo a la identificación del nivel de aprendizaje de los artesanos del CITE Sipán, facilitando así la planificación de actividades y optimizando el uso de recursos.

Se exploraron varios modelos de Machine Learning, entre ellos la regresión logística, el árbol de decisión y el CatBoostClassifier, de los cuales el árbol de decisión mostró ser el más efectivo, alcanzando una precisión del 75%. Además, el modelo fue optimizado mediante validación cruzada y se implementó en una interfaz web que permitió una visualización intuitiva de los resultados.

Finalmente, se evaluó la usabilidad del sistema a través de encuestas, mostrando una alta aceptación y satisfacción por parte de los usuarios, quienes consideraron el sistema útil y fácil de utilizar. Esto demostró que la implementación del modelo de Machine Learning fue una herramienta efectiva para apoyar la gestión y la mejora continua del aprendizaje en el sector artesanal.

Palabras clave: Aprendizaje automático, CRISP-DM, aplicación web, artesano, aprendizaje supervisado.

Abstract

In the context of a technological innovation center, the problem of properly identifying the learning level of artisans at CITE Sipán was addressed. CITE Sipán is an innovation center in Lambayeque that supports artisans through training and technical assistance. The lack of an adequate system to characterize the artisans led to inefficiencies in resource allocation and in adapting services to the specific needs of each artisan.

To resolve this issue, a web-based system utilizing Machine Learning was implemented to help identify the learning level of the artisans at CITE Sipán, thus facilitating activity planning and optimizing resource use. Several Machine Learning models were explored, including logistic regression, decision trees, and the CatBoostClassifier, with the decision tree proving to be the most effective, achieving an accuracy of 75%. Additionally, the model was optimized through cross-validation and was implemented on a web interface that allowed for intuitive visualization of the results.

Finally, the system's usability was evaluated through surveys, showing high acceptance and satisfaction from users, who found the system useful and easy to use. This demonstrated that the implementation of the Machine Learning model was an effective tool to support management and continuous improvement of learning in the artisan sector.

Keywords: Machine learning, CRISP-DM, web application, artisan, supervised learning.

Introducción

Se realizó un estudio en organizaciones gubernamentales [1], donde las decisiones se basan principalmente en la experiencia personal de quienes las gestionan. Los resultados identificaron varios elementos clave que pueden guiar las decisiones basadas en datos que estas instituciones manejan. Por otra parte, se llevó a cabo una investigación sobre la lealtad de clientes en una empresa [2], con el objetivo de determinar los factores que influyen en su permanencia, así como las características de aquellos que optan por la competencia. En Kenia, una empresa de servicios móviles ha experimentado la pérdida de clientes, lo que ha generado un impacto financiero negativo [3]. Estos casos, junto a otros problemas como la desertación de clientes, productos poco exitosos y la falta de identificación del cliente con el producto o servicio, han tenido un efecto adverso en las empresas.

Esta problemática surge cuando la empresa almacena datos históricos, ya sea debido al gran número de clientes que esta gestiona, o al tiempo prolongado trabajando con ellos. Además, es necesario definir los segmentos para caracterizar a personas o elementos clave dentro de la organización, lo cual es fundamental para la toma de decisiones, la mejora de productos o servicios, y, en consecuencia, para optimizar su competitividad y posición en el mercado [4].

A nivel global, el sector artesanal juega un papel clave tanto en la preservación del patrimonio cultural como en la economía, representando más del 3% del PBI (Producto Bruto Interno) global, según la UNESCO. Estudios internacionales resaltan la importancia de caracterizar el nivel de aprendizaje de los artesanos para mejorar su capacitación, aumentar la productividad y garantizar la sostenibilidad del sector. Iniciativas exitosas en países como India demuestran que la formación y adaptación a mercados globales incrementan el valor de los productos artesanales, beneficiando tanto a las economías locales como a la preservación cultural [5].

En el distrito de Túcume un estudio revela que la Asociación de Artesanos Valle de las Pirámides enfrenta limitaciones para exportar sus productos debido a la falta de conocimiento en comercio exterior y la escasa capacitación proporcionada por las instituciones competentes. Esta situación resalta la importancia de la formación y la identificación de las habilidades de los artesanos, ya que mejorar su nivel de aprendizaje permitiría acceder a mejores oportunidades en mercados internacionales, aumentar su competitividad y superar las barreras financieras y operativas que actualmente enfrentan [6].

La investigación se llevó a cabo en el Centro de Innovación Tecnológica en Turismo Artesanal Público de Sipán (CITE), institución gubernamental perteneciente al Ministerio de Comercio Exterior y Turismo. El objetivo de esta entidad es apoyar a los artesanos a través de capacitación

y seguimiento guiado por diseñadores contratados, quienes trabajan en conjunto para mejorar sus procesos productivos e ingresar a nuevos mercados.

Una de las funciones de la institución es la identificación de los niveles de competencia de los artesanos en el manejo del algodón nativo, un material de alto valor que requiere un trato especializado. La caracterización de estos perfiles es fundamental, ya que constituye la base para la toma de decisiones, la planificación de actividades del CITE, la definición de zonas de intervención y la determinación del servicio más adecuado para cada artesano. Por ello, es esencial que los niveles de competencia sean evaluados en función de los aspectos técnicos y del proceso productivo de los beneficiarios.

El CITE, como institución gubernamental, está obligado a presentar informes periódicos que respalden el uso y dónde se están utilizando los recursos proporcionados por el Estado. Estos informes deben estar respaldados por la documentación adecuada, que se produjo a lo largo del mes, incluido el desarrollo de planes de formación y contratos con artistas en función de los servicios prestados. Dado que el número de trabajadores en la institución es reducido, es necesario agilizar los procesos para evitar la sobrecarga de trabajo entre los colaboradores.

El problema abordado en la investigación se da porque los trabajadores se enfocan en la elaboración de las actividades y el almacenamiento de los datos que estas generen. La redundancia de datos de los artesanos es preocupante, al igual que los posibles errores de tipeo, pero sobre todo el problema viene de la identificación de los artesanos haciendo que los servicios brindados a los artesanos no sean los más adecuados para el nivel de cada artesano.

Actualmente, existen aplicaciones surgidas a partir de problemas similares a los mencionados, como en el caso de la empresa AISEC [7]. En esta empresa, se implementó una solución basada en machine learning para elaborar perfiles de consumidores, debido a que su público objetivo no estaba claramente definido. La propuesta consiste en desarrollar una herramienta que permita el almacenamiento y tratamiento de datos de leads, clasificándolos en perfiles y subgrupos según características comunes, con el objetivo de analizar la información y desarrollar productos de consumo personalizados.

En conclusión, a partir de la problemática antes planteada se presentó la siguiente interrogante ¿Cómo se puede contribuir a la identificación del nivel de los artesanos textiles del CITE SIPAN?

Desde el punto de vista económico, se busca beneficiar al CITE SIPAN reduciendo los gastos asociados a la inversión en artículos de oficina destinados al almacenamiento de información sobre los artesanos. Además, se optimizan los tiempos de identificación del nivel de los artesanos, lo que agiliza la toma de decisiones y la planificación de actividades dentro de la

institución. Dado el número actual de empleados, no es necesaria la contratación de personal adicional.

Desde una perspectiva social, se considera la accesibilidad a la información gracias a la colaboración del CITE SIPAN, lo que favorece el desarrollo de la investigación y permite caracterizar los niveles de los artesanos textiles para su detección.

Finalmente, desde un enfoque tecnológico, se utilizarán herramientas como Machine Learning y un lenguaje de programación adecuado para el desarrollo del sistema. Todo esto estará orientado a la implementación de un sistema de apoyo basado en Machine Learning para identificar el nivel de aprendizaje de los artesanos que forman parte de la línea textil del CITE SIPAN.

El objetivo general de esta investigación es implementar un sistema web basado en Machine Learning como apoyo a la identificación del nivel de aprendizaje de los artesanos del Cite Sipán. Para lograrlo, se propone, en primer lugar, establecer la técnica de Machine Learning más adecuada para la identificación de dicho nivel. A continuación, se desarrollará un modelo de Machine Learning basado en la técnica seleccionada, que permita la identificación precisa del nivel de aprendizaje de los artesanos. Luego, se verificará la precisión del modelo para asegurar su eficacia en esta tarea. Finalmente, se evaluará la calidad del sistema web construido sobre este modelo, de manera que pueda cumplir de manera óptima su función de apoyo en la caracterización de los artesanos.

Revisión de literatura

Antecedentes

Bravo [8], indica que la clasificación precisa de distribuciones es esencial en diversas áreas científicas, ya que una identificación incorrecta de la distribución de Weibull puede resultar en interpretaciones erróneas de los datos. Para abordar este desafío, el presente estudio introduce un modelo basado en redes neuronales artificiales (ANN) con el fin de mejorar la identificación de la distribución de Weibull, superando así los métodos tradicionales de bondad de ajuste, como la prueba de chi-cuadrado. Mediante el uso de bibliotecas de Python y técnicas de aprendizaje automático, como Scikit-learn, Keras y TensorFlow, se desarrollaron y evaluaron distintas distribuciones (Weibull, Normal, Uniforme, Exponencial, entre otras) para entrenar el modelo. Las metodologías empleadas abarcan simulaciones diversas, generación y clasificación de estadísticas cuantitativas, así como análisis estadísticos y descriptivos, culminando en el entrenamiento de un sistema ANN multicapa que permite reconocer patrones complejos en los datos. Los resultados obtenidos indican que el modelo ANN alcanzó una precisión del 98 % en el conjunto de datos de prueba, superando significativamente la precisión del 79 % obtenida

con el método chi-cuadrado. Este incremento en la precisión se atribuye a las capacidades de aprendizaje automático de las redes neuronales, que ofrecen flexibilidad, adaptabilidad y la habilidad de manejar grandes volúmenes de datos. En consecuencia, el modelo ANN se presenta como una alternativa más efectiva y eficiente frente a los métodos convencionales, constituyendo una herramienta poderosa para la detección de distribuciones.

Tito [9], expone que el abandono universitario es un problema significativo que afecta a estudiantes en diversas partes del mundo, impulsado por múltiples factores tanto internos como externos. La falta de una clasificación adecuada de los estudiantes en riesgo contribuye a la complejidad de abordar esta problemática. Para enfrentar este desafío, se propone utilizar técnicas de Machine Learning, que permiten predecir la deserción al analizar diversos atributos de los estudiantes. Entre las tecnologías empleadas se incluyen modelos como regresión logística, árboles de decisión, Support Vector Machines y Redes Neuronales, las cuales han demostrado su eficacia en clasificar y prever el éxito académico. Estos métodos se implementan con herramientas como Python y Scikit-Learn, lo que permite optimizar el análisis y la identificación de estudiantes en riesgo, facilitando así intervenciones oportunas y mejorando las tasas de retención en la educación superior.

Aquino [10], narra que el mercado asegurador venezolano ha enfrentado desafíos significativos en las últimas décadas, operando en un entorno limitado donde es crucial conocer a los competidores cercanos. El órgano regulador solo utiliza las primas cobradas como criterio de clasificación, a pesar de publicar otros indicadores. Para abordar esta limitación, el estudio propone el uso de algoritmos de aprendizaje automático que, a través del análisis de 18 variables y la aplicación de técnicas como K-Means y Fuzzy C-Means, permiten una clasificación más robusta y precisa de las aseguradoras, considerando múltiples factores y mejorando la exactitud en comparación con los métodos tradicionales.

Herrera [11], menciona que el sector comercial en Ecuador, pese a su relevancia económica, enfrenta el desafío de la existencia de "empresas zombis", entidades financieramente inviables que continúan operando y absorben recursos que podrían destinarse a empresas más productivas. Esta situación genera una distorsión en la asignación de capital en un país con recursos financieros limitados, lo que afecta la competitividad y el desarrollo económico. Identificar estas empresas es complejo debido a las múltiples definiciones y enfoques tradicionales de clasificación, lo que dificulta su detección efectiva.

Para abordar este problema, el estudio propone el uso de técnicas de Machine Learning como regresión logística, Random Forest y redes neuronales, que han demostrado ser herramientas eficaces para predecir la presencia de empresas zombis durante los años 2019, 2020 y 2021.

Los resultados indican que estos modelos mejoran significativamente la precisión en la identificación de dichas empresas, permitiendo una mejor canalización de recursos tanto desde el ámbito público como privado, y contribuyendo a una mayor eficiencia en la asignación de capital en la economía ecuatoriana.

Lazaro [12], indica el incremento en la deserción universitaria, especialmente tras la pandemia de COVID-19, ha generado preocupación en las instituciones educativas debido a factores económicos, sociales y académicos que afectan a los estudiantes. Esta problemática requiere de sistemas efectivos para identificar y reducir la deserción, considerando su impacto negativo en el desarrollo de habilidades y competencias necesarias para el progreso económico. Como solución, el uso de técnicas de Machine Learning ha emergido como una herramienta eficaz para predecir la deserción académica, permitiendo a las universidades anticiparse a este fenómeno.

Diversos modelos de Machine Learning como los árboles de decisión, redes neuronales, máquinas de vectores de soporte y bosques aleatorios han demostrado ser útiles en la predicción de la deserción académica. Investigaciones recientes han empleado estas técnicas para identificar variables clave como el rendimiento académico, la situación socioeconómica y el entorno familiar del estudiante. Con estos modelos, se han logrado precisiones de hasta 90% en algunos casos, evidenciando su efectividad en la predicción y prevención de la deserción universitaria. Estos modelos no solo permiten mejorar la retención estudiantil, sino que también optimizan la asignación de recursos en las instituciones educativas.

Díaz [13], indica que la deserción universitaria ha aumentado significativamente, especialmente tras la pandemia de COVID-19, debido a factores económicos, sociales y académicos que afectan a los estudiantes y sus familias. Esta problemática se ha vuelto crítica, ya que impacta no solo a los individuos, sino también al progreso económico y social de las instituciones educativas. Para abordar esta situación, se han implementado tecnologías y modelos de Machine Learning, aprovechando la Minería de Datos Educativos (EDM) para predecir y mitigar la deserción académica.

Una revisión sistemática de 30 artículos ha revelado diversas metodologías, técnicas y métricas útiles en esta área. Las técnicas más destacadas incluyen Árboles de Decisión, Redes Neuronales Artificiales y Bosques Aleatorios, entre otros. Asimismo, las variables más influyentes en la predicción de deserción abarcan el rendimiento académico, el contexto socioeconómico y las características demográficas de los estudiantes. Las métricas de evaluación de modelos, como precisión, exactitud y sensibilidad, son esenciales para medir el rendimiento de estos modelos predictivos. Estas iniciativas buscan fortalecer los sistemas

educativos y reducir la tasa de deserción mediante una comprensión más profunda de los factores que la impulsan.

Bases teóricas

Machine learning

El aprendizaje automático, conocido también como machine learning (ML), es una rama de la inteligencia artificial (IA) que otorga a las computadoras la capacidad de analizar datos y hacer predicciones o decisiones sin requerir programas específicos para cada tarea. A través del uso de algoritmos, los sistemas pueden identificar e interpretar patrones en grandes volúmenes de datos, mejorando su desempeño de manera continua para lograr una mayor precisión [14 y 15].

Según Hansen [16], el aprendizaje automático (Machine Learning) se clasifica en tres tipos principales, en función de la naturaleza de los datos disponibles y del objetivo del modelo: aprendizaje supervisado, no supervisado y aprendizaje por refuerzo.

En el aprendizaje supervisado, el modelo se entrena usando un conjunto de datos etiquetado, es decir, cada entrada del conjunto de entrenamiento cuenta con una etiqueta o resultado esperado (variable objetivo). La meta es aprender una función que relacione las entradas con las salidas correctas [17].

CRISP-DM

Es un método para la orientación en proyectos de machine learning cuyas siglas en inglés significan Cross-Industry Standard Process for Data Mining.

Ofrece un ciclo de vida compuesto por seis fases. Es un modelo flexible que permite la creación de un modelo de minería de datos adaptándose a las necesidades concretas. Algunas de las fases son bidireccionales como se muestra en la Fig. 1, haciendo posible la revisión parcial o total de las fases anteriores [18].

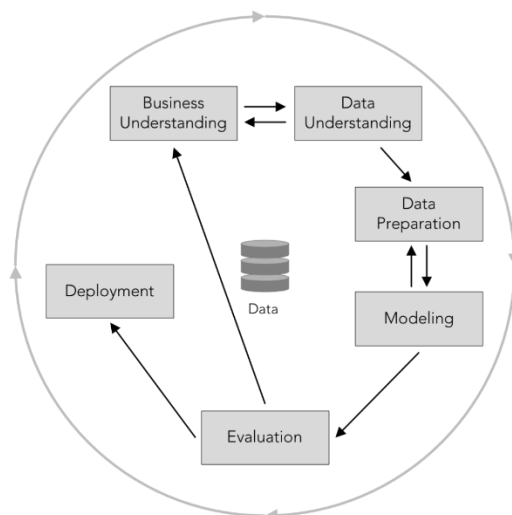


Fig. 1. Interfaz “Etapas de la metodología CRISP-DM”

SCRUM

Es un marco de trabajo que facilita la colaboración en equipo mediante iteraciones. Se lleva a cabo a través de entregas parciales del producto final y está diseñado para proyectos que requieren resultados rápidos. Se fundamenta en la adaptabilidad, el enfoque en las personas, la colaboración constante con el cliente y un desarrollo iterativo que garantiza buenos resultados [19].

Materiales y métodos

Según el manual de Frascati [20] este proyecto corresponde a una investigación tecnológica aplicada cuyos resultados están orientados al desarrollo de un sistema de apoyo al proceso de identificación de los niveles de los artesanos textiles en el CITE SIPAN.

Para esta investigación se consideraron los siguientes métodos propuestos por Bernal [21]:

TABLA I
MÉTODOS DE INVESTIGACIÓN

Método	Sustento por el cual será empleado en la investigación
Analítico	Estudio y análisis del problema
Deductivo	Planteamiento de la situación problemática
Implementación	Implementación de una solución de machine learning para el apoyo al proceso de identificación de los niveles de los artesanos textiles.

Se realiza la integración de métodos CRISP-DM [18] para la aplicación de Machine Learning y la aplicación de algoritmos y SCRUM [19] para el desarrollo y operación del sistema. Los pasos realizados en cada iteración de integración de las metodologías mencionadas con anterioridad se describirán a continuación:

ITERACIÓN #1: COMPRENSIÓN DEL NEGOCIO (CRISP-DM)

Esta iteración desarrollará las actividades siguientes:

- Determinación de los objetivos del negocio.
- Valoración de la situación.
- Determinación de los objetivos DM.
- Realización del plan del proyecto.

ITERACIÓN #2: COMPRENSIÓN DE LOS DATOS (CRISP-DM)

Esta iteración desarrollará las actividades siguientes:

- Recolección de los datos iniciales.
- Descripción de los datos.
- Exploración de los datos.
- Verificación de la calidad de los datos.

ITERACIÓN #3: PREPARACIÓN DE LOS DATOS (CRISP-DM)

Esta iteración desarrollará las actividades siguientes:

- Selección de los datos.
- Limpieza de los datos.
- Estructuración de los datos.
- Integración los datos.
- Formateo de los datos.

ITERACIÓN #4: MODELADO (CRISP-DM)

Esta iteración desarrollará las actividades siguientes:

- Selección de la técnica de modelado.
- Generación del plan de prueba.
- Construcción del modelo.
- Evaluación del modelo.

ITERACIÓN #5: PLANEAMIENTO (SCRUM)

Esta iteración desarrollará las actividades siguientes:

- Análisis de la problemática.
- Definición de los requisitos.
- Definición de objetivos y alcance.

ITERACIÓN #6: EVALUACIÓN (CRISP-DM), PLANIFICACIÓN, DEMOSTRACIÓN Y RETROSPECTIVA (SCRUM)

Esta iteración desarrollará las actividades siguientes:

- Evaluación de los resultados (CRISP-DM).
- Revisión del proceso (CRISP-DM).
- Determinar próximos pasos (CRISP-DM).
- Planificación de los sprints (SCRUM).

- Desarrollo del proyecto (SCRUM).
- Revisión del proyecto (SCRUM).

ITERACIÓN #7: IMPLEMENTACIÓN (CRISP-DM)

Esta iteración desarrollará las actividades siguientes:

- Plan de implementación (CRISP-DM).
- Monitorización y mantenimiento (CRISP-DM).
- Revisión del proyecto (CRISP-DM).
- Informe final (CRISP-DM).

Resultados y discusión

Establecer la técnica de machine learning adecuada para la identificación del nivel de aprendizaje de los artesanos

Para la establecer de la técnica de ML, se utilizó la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), un estándar en minería de datos que facilita el desarrollo de modelos de machine learning mediante un ciclo estructurado de seis fases [5].

Durante la fase de modelado, se probaron varias técnicas de clasificación supervisada. Se evaluaron modelos de regresión logística, árbol de decisión y CatBoostClassifier. Finalmente, el árbol de decisión se seleccionó como la técnica más adecuada debido a su capacidad de interpretar fácilmente los resultados y su precisión, que alcanzó el 75%. Este modelo fue elegido además por su simplicidad y facilidad para identificar relaciones claras entre las características de los datos.

Desarrollar el modelo de machine learning basado en la técnica seleccionada para la identificación del nivel de aprendizaje de los artesanos

Para la implementación del modelo, se construyó un árbol de decisión como algoritmo principal, aprovechando su estructura interpretativa que permite representar decisiones en una serie de condiciones, alineadas con las características y necesidades del proyecto en el CITE Sipán. La técnica de validación cruzada se aplicó para optimizar los hiperparámetros del modelo y asegurar su ajuste a los datos.

En cuanto a la integración, se utilizó la biblioteca Prickle para almacenar el modelo en un archivo .pkl (“ModeloAD”), lo cual facilitó su carga y acceso desde la interfaz web del sistema. Esto permitió al equipo del CITE Sipán utilizar el modelo sin necesidad de

retrain o modificaciones, asegurando la consistencia y rapidez en los resultados generados a partir de los datos de los artesanos.

Verificar la precisión del modelo de machine learning para la identificación del nivel de aprendizaje de los artesanos

La verificación del modelo se realizó utilizando una combinación de métricas de evaluación: precisión (accuracy), recall y F1-score, logrando una precisión final de 75%, la cual superó a los otros modelos evaluados en el estudio (como la regresión logística y el CatBoostClassifier).

El modelo fue sometido a una validación cruzada con cinco pliegues (folds) y múltiples combinaciones de hiperparámetros, resultando en 320 pruebas en total. Esta evaluación minuciosa ayudó a optimizar la capacidad del modelo para clasificar correctamente los niveles de aprendizaje de los artesanos y minimizar tanto los falsos positivos como los falsos negativos en sus predicciones.

Evaluar la calidad del sistema web basado en el modelo de machine learning para apoyar en la identificación del nivel de aprendizaje de los artesanos

Para evaluar la calidad y usabilidad del sistema web, se realizaron encuestas de satisfacción a los usuarios clave, incluyendo a la coordinadora del CITE SIPAN, quien interactuó con el sistema y ofreció retroalimentación sobre su facilidad de uso y utilidad en la identificación de competencias de los artesanos.

Los resultados de las encuestas indicaron un alto grado de satisfacción, con respuestas que variaron entre “Totalmente de acuerdo” y “Bastante de acuerdo” en cuanto a la usabilidad y experiencia del sistema. Esta aceptación mostró que el sistema no solo cumplía con los objetivos técnicos, sino que también proporcionaba una interfaz amigable, lo cual fue esencial para que el equipo del CITE pudiera interpretar los resultados generados por el modelo de machine learning y tomar decisiones basadas en datos confiables.

Conclusiones

El sistema web desarrollado para la identificación del nivel de aprendizaje de los artesanos ha cumplido con el objetivo de selección del modelo de machine learning y para ello se establecieron tres modelos para el análisis: regresión logística, árbol de decisión y CatBoostClassifier. Cada uno de estos modelos fue evaluado en términos de su precisión

(accuracy). De estos, el árbol de decisión destacó con un 75% de precisión, superando a los otros dos modelos. Como resultado, se seleccionó el árbol de decisión como el modelo más adecuado para ayudar en la identificación del nivel de aprendizaje del artesano.

A través de la técnica de validación cruzada, se logró determinar la combinación óptima de hiperparámetros, la cual resultó ser: 'C': 0.01, 'penalty': 'l1', 'solver': 'liblinear'. Esta configuración permitió alcanzar una precisión máxima en la identificación, asegurando que el modelo sea lo más exacto y fiable posible.

Para verificar la precisión del modelo de machine learning para la identificación del nivel de aprendizaje de los artesanos, se obtuvo una exactitud del 75%, superando a los otros dos modelos empleados en el estudio. Este resultado resalta la eficacia y fiabilidad del modelo, posicionándolo como la mejor alternativa dentro del contexto evaluado.

Se llevó a cabo una encuesta, cuyo contenido se detalla en el Anexo 01, dirigida a la coordinadora del CITE SIPAN. En esta encuesta, la coordinadora evaluó su nivel de acuerdo con respecto a la usabilidad del sistema y la calidad de la experiencia de usuario proporcionada. El grado de acuerdo varió entre "Totalmente de acuerdo" y "Bastante de acuerdo", reflejando resultados altamente significativos y una notable aprobación. Este rango de respuestas demuestra un fuerte consenso positivo entre los participantes, indicando una aceptación generalizada y favorable hacia el tema en cuestión.

Recomendaciones

Se recomienda automatizar los procesos restantes mediante la implementación de módulos que permitan registrar las capacitaciones y llevar un control de las ventas de los productos elaborados por los artesanos. Esto facilitará una mejor integración de los datos y optimizará la gestión de la información.

Integrar datos adicionales, como evaluaciones periódicas de desempeño, enriquecería el análisis y aumentaría la precisión del modelo. Este enfoque capturaría mejor el progreso de los artesanos en diversas habilidades, beneficiando la efectividad del sistema.

Incorporar un módulo de generación automática de informes periódicos optimizaría el tiempo y estandarizaría la documentación de los avances de los artesanos, cumpliendo los requerimientos administrativos del CITE con mayor eficiencia.

También se recomienda incluir en el sistema un módulo que permita al modelo continuar entrenándose con nuevos datos de los artesanos, con el objetivo de mejorar su precisión y optimizar su desempeño a lo largo del tiempo.

Referencias

- [1] A. Deligianni, "Data-driven innovation in NGOs," M.S. Thesis, ICT in Business, Leiden Univ., Leiden, The Netherlands, 2018. [Online]. Available: <http://bit.ly/32mYq0f>. [Accessed: Apl 21, 2022].
- [2] J. Naber, "Achieving Customer Loyalty from Email Campaigns by Using Data Mining Techniques," M.S. Thesis, Faculty of Business Economics, Hasselt Univ., Hasselt, Belgium, 2017. [Online]. Available: <http://bit.ly/2nSUtBt>. [Accessed: Apl 21, 2022].
- [3] K. Manero, R. Rimiru and C. Otieno, "Customer Behaviour Segmentation among Mobile Service Providers in Kenya using K-Means Algorithm," International Journal of Computer Science Issues (IJCSI), vol. 15, no. 5, pp. 67-76, Sept. 2018. [Online]. Available: <http://bit.ly/2oDNhJV>.
- [4] J. N. Alarcón et al, "Sistema de Innovación de la Minería de datos: Deserción y Permanencia de los estudiantes del Instituto Tecnológico Superior Guayaquil- Guayas," Revista Ibérica De Sistemas e Tecnologias De Informação, pp. 102-111, 2019. Available: <http://usat.lookproxy.com/scholarly-journals/sistema-de-innovación-la-minería-datos-deserción/docview/2317841852/se-22>.
- [5] D. Ramírez et al, "ARTESANOS EMPRENDEDORES DE LA ARTESANÍA LOCAL EN LA PROVINCIA DE GRANMA, CUBA/ENTREPRENEURS OF THE LOCAL ARTISANS HANDICRAFTS IN THE PROVINCE OF GRANMA, CUBA," Holos, vol. 32, (3), pp. 218-228, 2016. Available: <http://usat.lookproxy.com/scholarly-journals/artesanos-emprendedores-de-la-artesanía-local-en/docview/1813522281/se-2>. DOI: <https://doi.org/10.15628/holos.2016.4323>.
- [6] R. P. Chapoñan Zapata, Plan de negocios para la exportación de cojines de algodón nativo al mercado de Chile de la Asociación de Artesanos Valle de las Pirámides Lambayeque, 2018-2021, Universidad Señor de Sipán, Pimentel, Perú, 2021. Available: <https://repositorio.uss.edu.pe/handle/20.500.12802/9686>.
- [7] W. M. Rioja Curo, «Aplicación web para la elaboración de perfiles de consumidor basada en minería de datos y arquitectura cloud para el apoyo al proceso de conversión de leads en la asociación AIESEC en Perú», 2020, Accedido: 22 de abril de 2022. [En línea]. Disponible en: <http://tesis.usat.edu.pe/handle/20.500.12423/2721>

- [8] V. J. P. Bravo et al, "Modelo basado en redes neuronales para la clasificación de la distribución de Weibull," *Revista Ibérica De Sistemas e Tecnologias De Informação*, pp. 338-351, 2023. Available: <http://usat.lookproxy.com/scholarly-journals/modelo-basado-en-redes-neuronales-para-la/docview/2871352814/se-2>.
- [9] A. E. A. Tito, B. O. H. Condori and Y. P. Vera, "Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria," *Revista Ibérica De Sistemas e Tecnologias De Informação*, (51), pp. 84-98, 2023. Available: <http://usat.lookproxy.com/scholarly-journals/análisis-comparativo-de-técnicas-machine-learning/docview/2903934757/se-2>. DOI: <https://doi.org/10.17013/14811.51.84-98>.
- [10] J. Aquino-Olmos, E. López-Meléndez and L. Lara-Rodriguez, "Clasificación de aseguradoras venezolanas con algoritmos de aprendizaje automático," *Informador Técnico*, vol. 88, (1), pp. 39-55, 2024. Available: <http://usat.lookproxy.com/scholarly-journals/clasificación-de-aseguradoras-venezolanas-con/docview/3071581641/se-2>. DOI: <https://doi-org.usat.lookproxy.com/10.23850/22565035.6219>
- [11] R. A. Herrera et al, "Empresas comerciales zombis ecuatorianas: Una clasificación mediante Machine Learning," *Revista Ibérica De Sistemas e Tecnologias De Informação*, pp. 541-554, 2024. Available: <http://usat.lookproxy.com/scholarly-journals/empresas-comerciales-zombis-ecuatorianas-una/docview/3085714905/se-2>.
- [12] M. Asto-Lázaro S. and H. Bermejo-Terrones, "Revisión Sistemática: Machine Learning en la Predicción de Deserción Académica," *Revista Ibérica De Sistemas e Tecnologias De Informação*, pp. 463-476, 2023. Available: <http://usat.lookproxy.com/scholarly-journals/revisión-sistemática-machine-learning-en-la/docview/2973215570/se-2>.
- [13] B. Díaz et al, "Deserción de estudiantes, factores asociados con árboles de decisión: caso Escuela de Postgrado 1de una Universidad pública en Perú," *Revista Ibérica De Sistemas e Tecnologias De Informação*, pp. 197-211, 2022. Available: <http://usat.lookproxy.com/scholarly-journals/deserción-de-estudiantes-factores-asociados-con/docview/2768748027/se-2>.
- [14] A. Smart and A. Kasirzadeh, "Beyond Model Interpretability: Socio-Structural Explanations in Machine Learning," *ArXiv.Org*, 2024. Available:

<http://usat.lookproxy.com/working-papers/beyond-model-interpretability-socio-structural/docview/3101379017/se-2>. DOI: <https://doi.org/10.1007/s00146-024-02056-1>.

[15] T. M. Mitchell, "Does machine learning really work?," *AI Magazine*, vol. 18, no. 3, p. 11, 1997.

[16] S. Hansen, "APLICACIÓN DEL APRENDIZAJE AUTOMÁTICO AL ANÁLISIS ECONÓMICO Y LA FORMULACIÓN DE POLÍTICAS," *Papeles De Economía Española*, (157), pp. 216-234, 2018. Available: <http://usat.lookproxy.com/scholarly-journals/aplicación-del-aprendizaje-automático-al-análisis/docview/2133762843/se-2>.

[17] M. Segura, J. Mello, and A. Hernández, "Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role?," *Mathematics*, vol. 10, no. 18, p. 3359, 2022. [Online]. Available: <https://doi.org/10.3390/math10183359>.

[18] P. Chapman et al., "CRISP-DM 1.0: Step-by-step data mining guide", 2000, p. 78. [En línea]. Disponible en: <https://api.semanticscholar.org/CorpusID:59777418>

[19] K. Schwaber, *Agile Project Management with Scrum*, 1ra Edición. Microsoft Press, 2004. Accedido: 22 de mayo de 2022. [En línea]. Disponible en: <https://es.b-ok.lat/book/807586/2fcf16>

[20] OECD, "Manual de Frascati 2015". MIC, 2018. [En línea]. Disponible en: <https://www.oecd-ilibrary.org/content/publication/9789264310681-es>

[21] C. Bernal, "Metodología de la investigación", 4ta Edición. Colombia: Pearson Educación de Colombia S.A.S., 2016. Accedido: 22 de mayo de 2022. [En línea]. Disponible en: <https://es.b-ok.lat/book/11810697/ab124>