

Tendencias en la gestión de facturación para ACEROTRANS.S.A en 2024: Un enfoque práctico aplicando técnicas de Inteligencia Artificial

Trends in Billing Management for ACEROTRANS.S.A in 2024: A Practical Approach Applying Artificial Intelligence Techniques

Diego Alexander Reinoso Cueva^{1,2} [0000-0003-2058-0797], Edison Patricio Bedón Salazar² [0000-0002-1142-0196]

¹ Universidad Técnica de Cotopaxi. Carrera de Sistemas de Información. Latacunga, Cotopaxi. Ecuador

² Compañía de transportes pesados Acerotrans S.A. Latacunga, Cotopaxi. Ecuador

¹diego.reinoso4051@utc.edu.ec, ²edison.bedon@utc.edu.ec

CITA EN APA:

Reinoso Cueva, D. A., & Bedón Salazar, E. P. (2024). Tendencias en la gestión de facturación para ACEROTRANS.S.A en 2024: Un enfoque práctico aplicando técnicas de Inteligencia Artificial: Un enfoque práctico aplicando técnicas de Inteligencia Artificial. *Tesla Revista Científica*, 4(2), e395. <https://doi.org/10.55204/trc.v4i2.e395>

Recibido: 2024-07-15

Revisado: 2024-07-22 al 2023-08-11

Corregido: 2024-08-20

Aceptado: 2024-08-26

Publicado: 2024-09-01

TESLA

Revista Científica

ISSN: 2796-9320



Los contenidos de este artículo están bajo una licencia de Creative Commons Attribution 4.0 International (CC BY 4.0)

Los autores conservan los derechos morales y patrimoniales de sus obras. The contents of this article are under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. The authors retain the moral and patrimonial rights of their works.

Resumen.

Introducción: La integración del Machine Learning y el desarrollo de aplicaciones informáticas, está en constante evolución debido a que mejoran la capacidad para gestionar eficazmente requerimientos de una empresa u organización y a la vez apoyar a la toma de decisiones informadas.

Objetivo: La predicción de tendencias en facturación de transportistas de Acerotrans a partir de datos históricos con la finalidad de evaluar cómo se pueden tomar decisiones anticipadas.

Métodos: Recopilar datos históricos, seguidos por el uso de técnicas de Machine Learning para implementar algoritmos de regresión. Se aplicaron procesos ETL para subida masiva de datos y técnicas de BI para que la información se visualice mediante reportes gráficos. Todos estos métodos forman parte de las fases de KDD.

Resultados: Los algoritmos de regresión se encargarán de entrenar, agrupar y predecir la facturación, lo que permitió hacer un análisis sobre los datos originales y los datos predichos a fin de evaluar el comportamiento y ajuste del modelo.

Conclusiones: La integración de la inteligencia artificial, inteligencia de negocios y el desarrollo Web es vital para que un sistema informático se adapte a cambios y necesidades del usuario. La predicción de tendencias en base a datos históricos es esencial para anticiparse a eventos futuros, por lo que el volumen de datos debe ser considerable.

Palabras clave: Desarrollo Web, Machine Learning, KDD, BI, ETL.

Abstract:

Introduction: The integration of Machine Learning and software application development is constantly evolving, as it enhances the ability to effectively manage the requirements of a company or organization while also supporting informed decision-making.

Objective: Predicting billing trends for Acerotrans carriers based on historical data aims to evaluate how proactive decisions can be made.

Methods: Collecting historical data, followed by the use of Machine Learning techniques to implement regression algorithms. ETL processes were applied for bulk data upload and BI techniques were used to visualize the information through graphical reports. All these methods are part of the KDD phases.

Results: Regression algorithms will be responsible for training, clustering, and predicting billing, enabling an analysis of the original data and the predicted data to evaluate the model's behavior and fit.

Conclusions: The integration of artificial intelligence, business intelligence, and web development is vital for a computer system to adapt to user changes and needs. Predicting trends based on historical data is essential to anticipate future events, making it necessary for the data volume to be substantial.

Keywords: Web Development, Machine Learning, KDD (Knowledge Discovery in Databases), BI (Business Intelligence), ETL(extract- transform-load).

1. INTRODUCCIÓN

Actualmente se considera que los datos de una empresa u organización son el activo más valioso, los datos son transformados en información y luego en conocimiento, pero para salvaguardar los datos se debe asegurar la integridad, confidencialidad y disponibilidad de los mismos. Estos datos deben ser concisos por lo que unos datos ruidosos afectaran todo el proceso hacia el conocimiento, los mismos deben ser almacenados de tal manera que su gestión sea rápida y eficiente.

Según (Sánchez et al., 2020) asegura que la gestión de la información es parte fundamental dentro de las organizaciones y afirma que al gestionar la información de una empresa aumenta el valor y la competitividad de una empresa. Y por otro lado (Zambrano et al., 2019) afirma que para gestionar la información de manera correcta se debe recolectar los datos, procesarlos y clasificarlos.

(Amaya, 2015) asegura que todos los humanos tomamos decisiones a diario en cualquier actividad que realizamos. Es necesario razonar de forma continua y concentrada para tomar la mejor decisión. En una empresa el gerente debe tomar decisiones todos los días. Una decisión acertada beneficia de manera positiva a una organización, le ayuda a crecer de manera estratégica y sobresalir de la competencia, además que se gana la fidelidad de los clientes y en algunos casos los mismos recomiendan un producto o servicio, sirviendo de alguna forma de publicidad.

Ciertas empresas medianas y pequeñas (PYMES) gestionan su información de manera física, lo cual hace difícil recuperar o buscar la misma, puede haber perdida e inconsistencias de información, lo que provoca que no se tenga un control de los procesos internos (Useche-Aguirre et al., 2021). Un sistema Informático ayuda a sistematizar procesos, tener un control de los mismos y reducir tiempos de búsqueda de información. Teniendo en cuenta la brecha tecnológica que existe en Ecuador en ciertos casos es un reto el camino hacia la transformación digital, sin embargo, se deben buscar estrategias y alternativas como capacitaciones, cursos prácticos, etc.

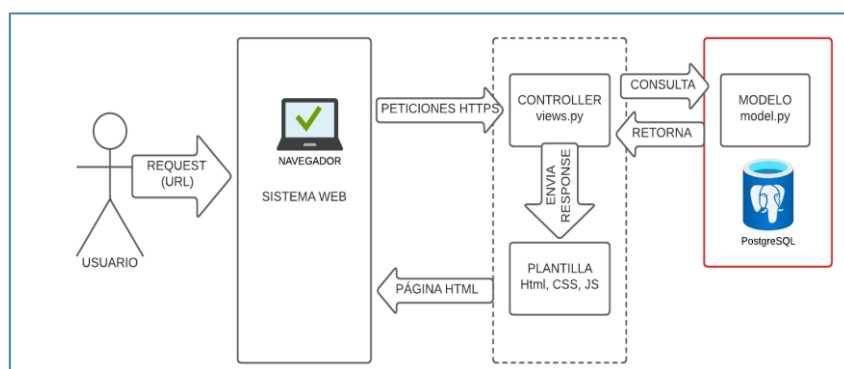
(Osorio, 2023) afirma que un sistema informático permite que muchos procesos que realizan las empresas sean más óptimos en términos de gestión de información, son muchos los beneficiados, desde empresarios hasta clientes finales. Es importante abarcar las necesidades y preferencias tecnológicas de los usuarios y no segmentarlos, sin importar el dispositivo ni el sistema operativo que estos utilicen en su día a día.

Una alternativa acertada es un entorno web ya que no necesita instalarse, es escalable, y con solo un desarrollo, se puede acceder desde cualquier lugar que se tenga un dispositivo con un navegador actualizado y acceso a internet (Rivera et al., 2016). Una aplicación Web es la integración de base de datos, el Backend, Frontend, Web Service, API's entre otros. En cuestión del Backend hace referencia a la lógica de negocio que se implementa con ayuda de lenguajes de programación del lado del servidor; hay un sinnúmero de Frameworks con los cuales trabajar.

Python es un lenguaje muy usado dentro de Data Science por ser es un lenguaje de alto nivel orientado a objetos, multiparadigma y multipropósito, posee un sinnúmero de librerías orientadas hacia la

inteligencia artificial, visualización de la información, minería de datos, desarrollo de software. Según (*TIOBE Index*, s. f.) asegura que Python es el lenguaje más usado a nivel mundial. Para desarrollo Web, Python tiene Frameworks como Django y Flask que permiten trabajar con Bases de Datos, consumir servicios. Algo que destaca a Django es su arquitectura MVT(Model-View-Template), una variante de MVC(Model-View-Controller)(Vidal-Silva et al., 2021). Donde se separa por capas la programación, la capa de acceso a datos o el modelo interactúa con la base de datos o ORM y peticiones, mientras que la capa de lógica de negocio o Vista ayuda a hacer peticiones al servidor, y finalmente la capa de presentación o Template se presenta al usuario final en donde interactúa con la interfaz gráfica. Tal como lo muestra la Figura 1.

Figura 1: Diagrama de arquitectura MVT



Elaborado por: Autores

En los últimos años han surgido estrategias innovadoras dentro del ámbito informático como es la inteligencia artificial, inteligencia de negocios y el desarrollo de software, la unión de estos tres ayuda a tomar decisiones informadas a una empresa u organización. La implementación de algoritmos y técnicas de inteligencia ayudan hacer predicciones en base a datos históricos, la detección de fraudes bancarios, identificación de imágenes, análisis de sentimientos, sugerencias en base a gustos personales, entre otros.

En este contexto, la compañía de transportes pesados Acerotrans S.A., ubicada en Lasso, Tanicuchi, perteneciente a la provincia de Cotopaxi – Ecuador; se dedica a gestionar permisos de operación, a mejorar la planificación logística, la gestión de facturación y cobranza de fletes de los transportistas. Acerotrans nace en el año 2007 y está legalmente constituida en la superintendencia de Compañías. El objetivo principal ayudar a los socios a prestar servicios de transporte a empresas. Un reto que se presenta es que sus procesos han sido manuales en tema de su facturación, liquidación, pagos y registro de compras y ventas, es por ello que se evidencia un sinnúmero de registros históricos almacenados de manera física de hace varios años atrás hasta la actualidad. Sin embargo, para recuperar esta información se dificulta debido a que la información almacenada es considerable en volumen y en archivos.

La ausencia de un sistema informático para la gestión de facturación de los socios es evidente, debido que el almacenamiento, clasificación y recuperación de la información es de forma física, al no tener un control de la gestión de facturas, ocurre: inconsistencias y pérdidas de información, lo que repercute en toma de decisiones incorrectas. Como parte del apoyo a toma de decisiones informadas e innovación es

necesario aplicar técnicas de inteligencia de negocios que permitan generar reportes gráficos como dashboards. La implementación de algoritmos de predicción basados en datos históricos ayudará a determinar: la tendencia de facturación a manera global, por socio y la ganancia por inversión, lo que permitirá evaluar la precisión del modelo y conocer sus métricas.

Debido a que la facturación de los socios depende de las temporadas y fletes realizados, se deben considerar los umbrales que pueden existir como daños mecánicos, ausencia de fletes, y el retraso en el proceso de facturación y liquidación de cada empresa. Es por ello que la facturación es algo compleja de predecir, ya que siempre es variable mes a mes, algo a considerar es el tipo de vehículo, la categoría y el tonelaje. El reto de implementar un algoritmo de inteligencia artificial para predecir la facturación de los socios basados en los datos históricos se hace presente en la integración del sistema informático para la gestión de facturación para Acerotrans, el fin es poder gestionar socios, vehículos, categorías, y facturación.

Los algoritmos de predicción pertenecen al aprendizaje supervisado, es decir requieren clases y etiquetas, dichos algoritmos pertenecen al ámbito de la regresión. Hay diferentes algoritmos que permiten predecir, entre los más conocidos, SVR(Support Vector Regression), Redes Neuronales LSTM, Random Forest Regressor, GradientBoostingRegression, entre otros(Castro & Vega, 2023). Sin embargo, SVR es rápido, fácil de implementar y se puede utilizar empleando técnicas como secuencias, hiperparámetros, Cross Validation. Sin embargo, SVR destaca por epsilon y por sus diversos kernels o núcleos, lo que ayuda en casos de tener datos muy dispersos.

Python es robusto en términos de inteligencia artificial, contiene diversas librerías como Scikit-Learn que ya incluyen los algoritmos de Machine Learning, lo que facilita mucho la implementación de los algoritmos con dicho lenguaje(Mamani Rodriguez, 2022). Uno de estos algoritmos más usados es Support Vector Regression(SVR), tiene diferentes núcleos para adaptarse a los panoramas de los datos, SVR se adapta de buena manera a datos dispersos ya que también considera valores fuera del modelo a esto se lo conoce como epsilon haciendo posible que el modelo tolere también pequeños errores pero siempre centrándose en valores significativos(Sanabria-Castro et al., 2022).

(Sanabria-Castro et al., 2022) asegura que SVR es eficiente en cuestión de tiempos lo que en producción ayudará a que los resultados que genere el algoritmo sean cortos comparado con otros modelos como LSTM. Para evaluar el modelo se puede emplear la técnica de validación cruzada, y los resultados que el algoritmo devuelva deben generar conocimiento hacia el usuario final ya que es complejo que pueda interpretar debido que la mayoría no tendrá conocimiento informático. Por ello se debe visualizar la información a través de gráficos, tablas, reportes, entre otros. Las decisiones que un sistema informático ayuda a tomar es el resultado de una correcta gestión de la información, unas buenas técnicas de inteligencia de Negocio y un buen modelo de AI.

Como caso de estudio la predicción de facturación será en base a los datos históricos. El proceso de recolección de información se basó en entrevista con la Gerenta de la compañía, observación directa e indirecta, y la digitalización de las facturas históricas desde 2021 hasta la actualidad. Debido a los datos

históricos físicos se aplican procesos ETL para subida masiva de datos, debido que se trabaja con regresión es necesario un volumen de datos alto para que SVR genere mejores resultados, en cuestión de las herramientas de desarrollo se emplea para la parte del Backend Python con el Framework Django; por otro lado, en cuestión del Fronted se emplea HTML5, JavaScript para validaciones y manejo de formularios, y diversas librerías como FileInput, DataTables, Chart.JS, entre otros.

2. METODOLOGÍA O MATERIALES Y METODOS

Existen diversas metodologías que se pueden aplicar al ámbito relacionado con Inteligencia Artificial y la minería de datos. Entre las más conocidas son KDD (Knowledge Discovery in Databases), SEMMA y CRISP-DM(García-González et al., 2019). Las metodologías proponen una serie de fases y procedimientos que ayudan a construir algoritmos y modelos eficientes y debidamente ajustados. KDD por su lado es una metodología robusta que abarca fases de carga de datos que posterior a ser procesados serán entrenados y agrupados, luego se aplica técnicas de minería de datos y finalmente se valida y se visualizan los resultados. El proceso de KDD es iterativo por lo que se puede mover entre las diferentes fases dependiendo de la lógica de negocio.

A continuación, se habla acerca de las fases de KDD empleadas como parte de la presente investigación.

2.1. Metodología KDD

La metodología KDD se basa en un proceso automático en el que se combinan el descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones a partir de datos para que el usuario los interprete a esto se lo conoce como preprocesado, minería de datos y presentación de resultados(Timarán Pereira et al., 2016). KDD se puede aplicar en diversos panoramas especialmente a los que combinan el desarrollo de software con otras áreas como inteligencia artificial, inteligencia de negocios, entre otros. Las fases del proceso KDD son:

Selección

En esta etapa se selecciona el conjunto de datos sobre el cual se realiza el proceso de descubrimiento. La selección de datos varía de acuerdo a la lógica y trazabilidad del negocio(Valenzuela-Nunez et al., 2023). En esta etapa es indispensable la selección y extracción de datos con los que se va a trabajar. Se emplea la Base de datos que está integrada al aplicativo de gestión con fin de evaluar los posibles atributos que se pueden emplear al proceso. Python y Django permiten trabajar directamente con el set de datos.

Limpieza y Preprocesado

Se analiza la calidad de los datos, de haber ruido, se aplican técnicas como remover registros que tengan inconsistencias, dar formato a los campos, identificar estrategias como filas duplicadas, información desconocida y técnicas estadísticas para su reemplazo(Valenzuela-Nunez et al., 2023). Parte de la limpieza y preprocesado que se aplica fue separar las fechas de cada factura, identificando así el mes y año por separado, y también se relaciona las facturas con el propietario, con el vehículo y finalmente con la categoría a la que pertenecen.

Transformación

Aquí se busca las características más significativas para representar los datos, con fin de lograr un modelo efectivo. Es necesario considerar que SVR trabaja con variables dependientes e independientes. La variable dependiente Y es el objetivo a predecir y solo abarca un atributo o característica numérica, por otro lado, la variable independiente X abarca características complementarias que ayudan a mejorar o entorpecer la predicción de datos, las variables independientes pueden ser varias a la vez.

Minería de datos

Es el proceso de descubrir patrones y otra información valiosa en los conjuntos de datos, la minería de datos ha mejorado la toma de decisiones por medio del análisis de datos (IBM, 2023). En esta etapa se aplica y evalúa la técnica más apropiada de minería de datos que mejor se adapte al panorama del proyecto. Asociado a la investigación se consideran los parámetros del algoritmo que SVR presenta como el tipo de Kernel, los hiperparámetros y la técnica GridSearchCV mediante Cross Validation ayuda a dividir el porcentaje de datos para entrenamiento y pruebas.

Con ayuda de Inteligencia de Negocios se genera dashboards informativos que mostrará los estados de las facturas: compradas, pagadas y anuladas, así como las ganancias obtenidas. Además, se integra un algoritmo de predicción para analizar tendencias en los pagos. Esto hará posible generar indicadores claves de rendimiento KPI's identificando la variable de análisis y la variable de medición; por ejemplo: el mes de facturación más alto y más bajo, el transportista con la facturación más alta y más baja, la categoría de vehículos que mejor y peor factura, entre otros. Esta información facilitará una rápida toma de decisiones a nivel de gerencia y a nivel de transportistas.

Evaluación

Se interpretan los patrones o resultados del modelo descubiertos, y de ser el caso se retorna a anteriores fases a manera de un proceso de iteraciones, En esta etapa también se visualizan los patrones extraídos, remoción de patrones irrelevantes y la interpretación de los mismos para que sean entendibles para el usuario (Timarán Pereira et al., 2016). Se emplea la regresión para lograr la predicción de los meses históricos, posteriormente se unen los datos originales con los predichos con fin de luego lograr visualizar de forma comparativa los resultados de manera grafica mediante KPI's (Indicador Clave de Rendimiento).

2.2. Métricas de evaluación

Para evaluar el algoritmo se emplea las métricas de R^2 siendo una métrica que indica la precisión del modelo, ya que indica la varianza que tiene la variable dependiente que se puede predecir desde la variable independiente (Lambis-Alandete et al., 2023). Esta métrica se mide en un rango de 0 a 1 siendo el modelo más preciso cuando se aproxime a 1.

El MSE o error cuadrático medio se puede interpretar como diferencia cuadrática entre el valor predicho y el valor real, se calcula como la media de los cuadrados de los errores entre dichos valores, su interpretación es que a un valor bajo indica que el modelo tiene errores pequeños y viceversa. Un buen modelo debe tener el equilibrio entre un R^2 alto y un MSE bajo (Vargas-Zapata et al., 2022).

Por otro lado, el método GridSearchCV ayuda a identificar que parámetros se adaptan mejor al proyecto y lo aplica, por otro lado, el Cross Validation ayuda a dividir esos entrenamientos con los diferentes parámetros obteniendo así las mejores métricas. Se reserva una partición para realizar las predicciones con la técnica de Machine Learning y con el restante de particiones se construye el modelo predictivo, este proceso es iterativo y se lo determina mediante CV(Páez & Ramírez, 2022).

Se analizan los datos obtenidos, y se identificará el porcentaje de datos para entrenamiento y para evaluar el modelo, comúnmente se aplica el principio de Pareto del 80-20(Forero-Corba & Negre Bennasar, 2023). El Cross Validation ayuda a identificar los mejores hiperparámetros que se utilizaron para generar el modelo durante el entrenamiento.

Para evaluar dichas métricas se propone la tabla 1 de escala de fiabilidad:

Tabla 1:

Escala de fiabilidad

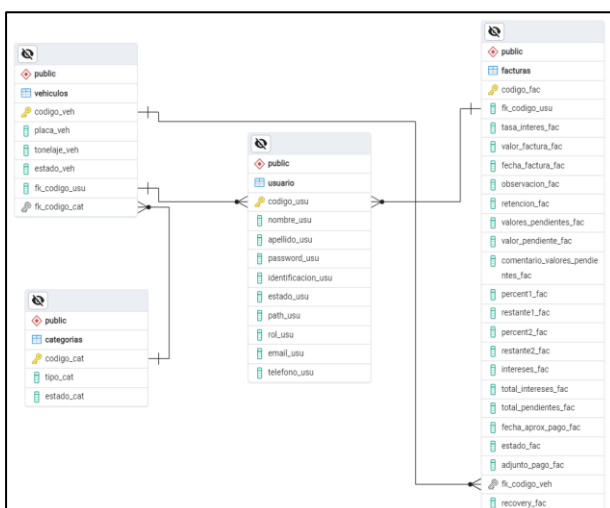
Muy Baja	Baja	Moderada	Buena	Alta
0 – 0.20	0.21 – 0,40	0.41 – 0.60	0.61 – 0.80	0.80 – 1

Elaborado por: Autores

3. RESULTADOS Y DISCUSIÓN

En el proceso de identificar la lógica de negocios y requerimientos de Acerotrans S.A., se evidencio la ausencia de un sistema de gestión de las facturas de los socios transportistas, también se evidencio información histórica almacenada de forma física en archivadores. Como parte previa se aplicaron técnicas de recolección de información como observación y entrevistas, así como el registro de datos históricos al sistema informático realizado mediante ETL. Hay que tener en cuenta que el volumen de datos históricos debe ser considerable para que el modelo realice unas buenas predicciones. Como parte del análisis propuesto se presenta el Diagrama del modelo relacional de los datos que se emplearon para el sistema y para implementar SVR. Dicho modelo se presenta en la figura 2.

Figura 2: Diagrama del Modelo Relacional



Elaborado por: Autores

En cuestión de la limpieza y preprocesado se emplean técnicas estadísticas y matemáticas para identificar las variables más significativas que servirán para el modelo como parte del preprocesado que consiste en dar un formato a las fechas y a las variables seleccionadas, luego se sacan métricas previas como la media, mediana y desviación estándar de los datos antes de entrenarlos. La variable dependiente es única y será el valor que se quiere predecir, por lo general es la variable Y, mientras tanto las variables independientes pueden ser varias a la vez pero que aporten y mejoren el modelo y no lo entorpezcan y se representan en la variable X.

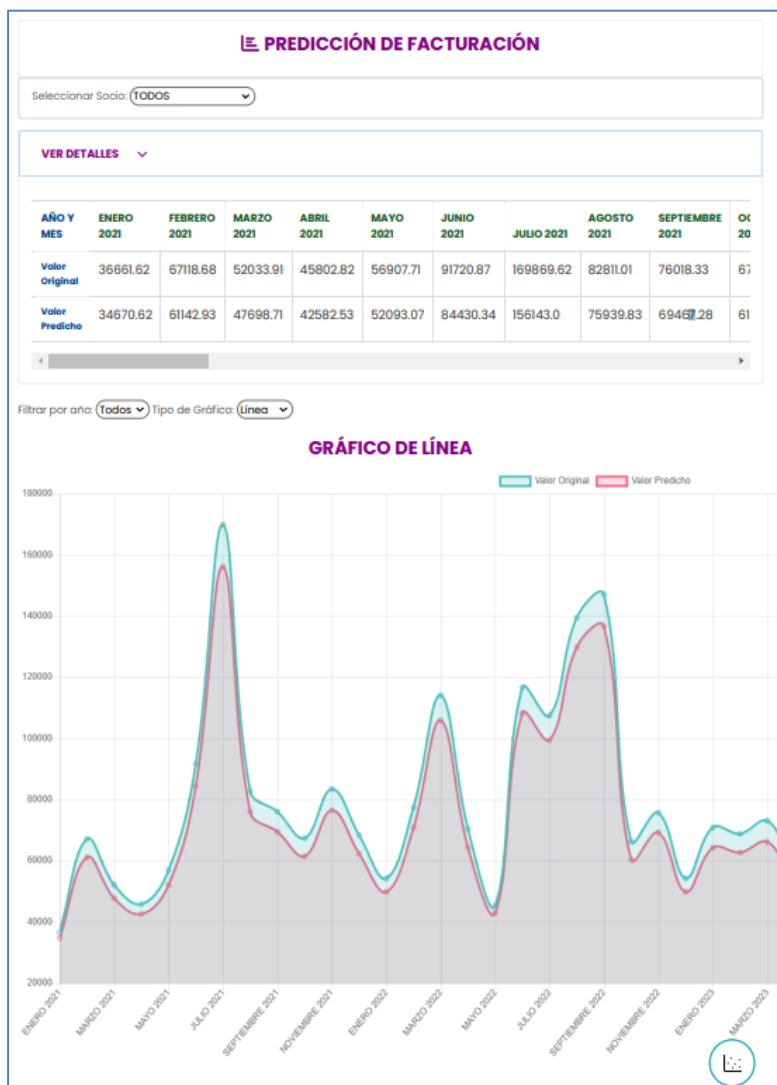
Parte del entrenamiento al modelo se consideran factores como ϵ que define una zona alrededor del modelo para considerar pequeños errores o pérdidas mejorando en parte el modelo, un ϵ bajo es ideal para que no se consideren errores grandes ni exista ruido en los datos. Por otro lado, el parámetro de regularización C, controla el equilibrio entre lograr un entrenamiento más bajo y mantener una regulación adecuada del modelo, un valor de C alto es señal de que el modelo penalice más fuerte a los errores de predicción, haciendo que el modelo sea más preciso. Sin embargo, es necesario recordar que un valor demasiado alto hará que el modelo se sobreajuste. Por último, el parámetro gamma es fundamental dentro del Kernel no lineal de SVR como es RBF mismo que controla el alcance de la predicción. La combinación de estos tres parámetros; es necesaria para evaluar si un modelo está bien ajustado y no existe ruido en las predicciones.

Como parte del algoritmo implementado SVR se considera factores como: la dispersión de los datos, a condiciones que se evidenciaron a lo largo de la carga de información, que se tiene socios que tienen más facturas que otros, esto debido a la categoría del vehículo, la temporada y los umbrales como daños mecánicos, falta de personal, entre otros. El algoritmo es posible aplicarlo gracias a la librería Scikit-Learn que se incluye en Python y cuenta con un sinnúmero de algoritmos de aprendizaje automático.

Con el fin de evaluar el modelo existen técnicas como el coeficiente de determinación R^2 que mide la variabilidad que existe entre la variable dependiente de las variables independientes en el modelo de regresión. Un valor cercano a 1 indica un buen ajuste del modelo y viceversa, un modelo óptimo debería ser mayor a 0.8. Otra técnica a evaluar es MSE, siendo una métrica utilizada para medir la precisión de los modelos de regresión. Representa el promedio de los cuadrados de las diferencias entre los valores predichos y los valores reales. Un MSE más bajo indica que el modelo predice con mayor precisión los valores reales, minimizando las desviaciones grandes.

Finalmente, esas predicciones son necesarias llevarlas a representaciones gráficas y entendibles del usuario final, para que pueda ser analizado e interpretado, aquí entra las técnicas de Business Intelligence que consiste en extraer los resultados del algoritmo y llevarlos a conocimiento, algunas herramientas ofrecen visualización de información dentro de Python es Matplotlib y Seaborn, y también en entorno web se tiene a Chart.JS, Google Charts, HighCharts, entre otros. Como parte de la evaluación e interpretación se indica los resultados obtenidos después de todo este proceso. La Figura 3 muestra el resultado obtenido:

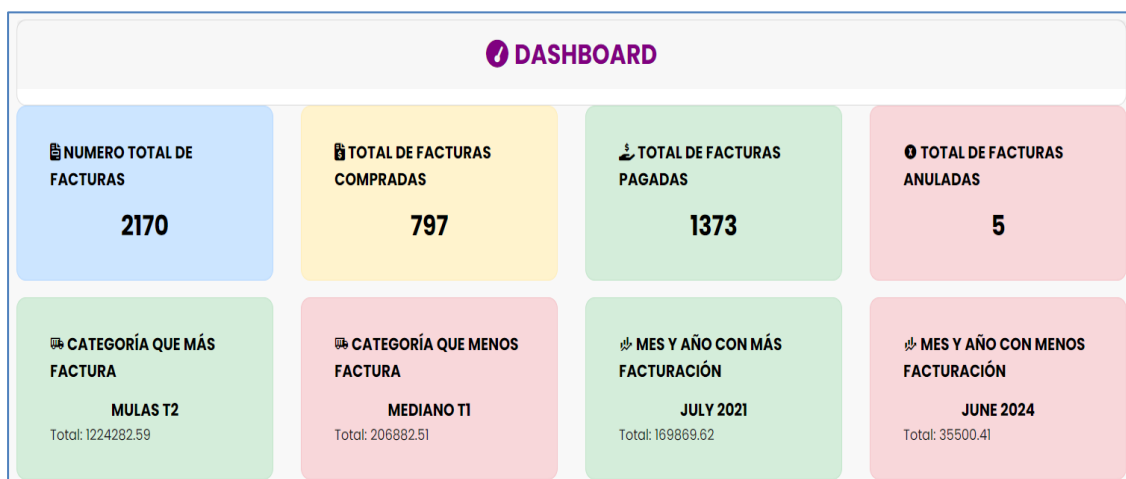
Figura 3: Resultado de las predicciones

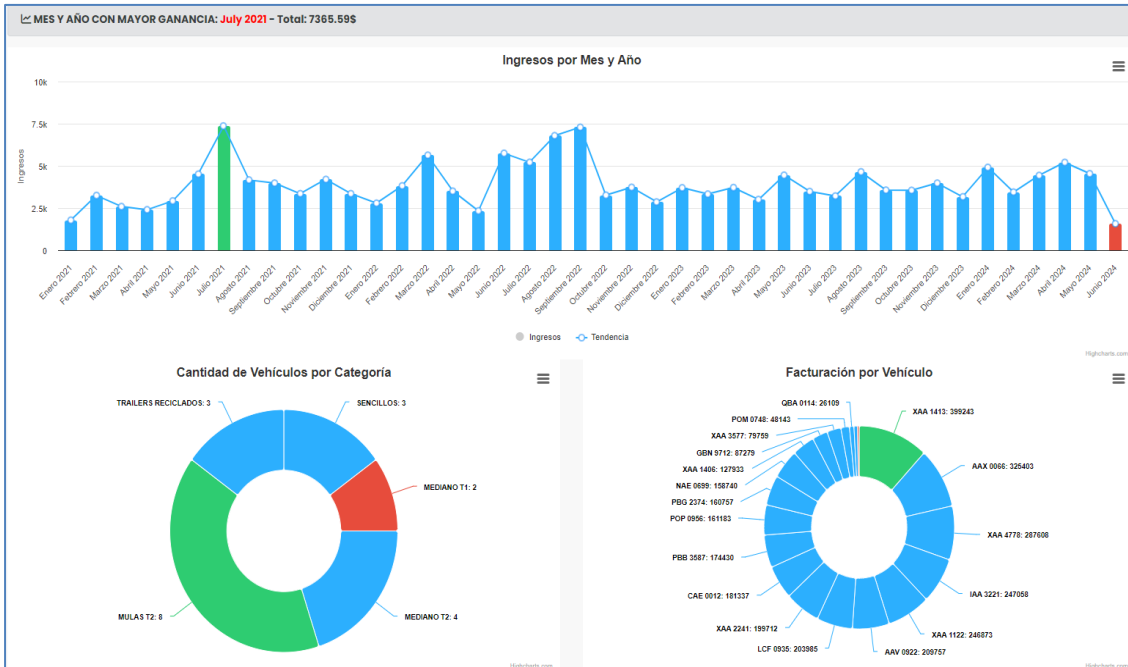
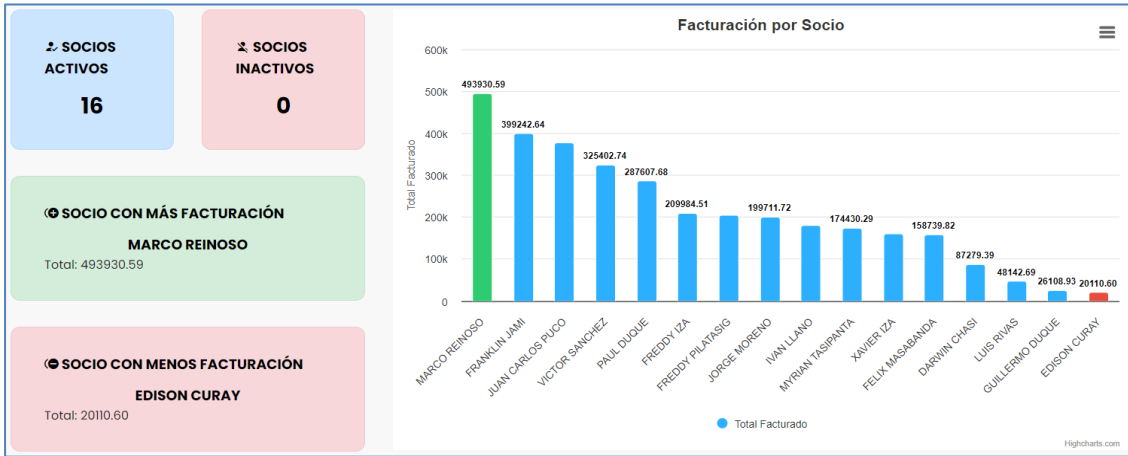


Elaborado por: Autores

Por otro lado, a partir de esas predicciones se diseñan Indicadores Claves de Rendimiento KPI's que en conjunto se denomina Dashboard para ayudar a tomar decisiones basadas en datos históricos, las gráficas son de gran ayuda a nivel de gerencia es por ello que se propone la figura 4 misma que se muestra a continuación:

Figura 4: Dashboard





Elaborado por: Autores

Los reportes gráficos empleando HighCharts ayudan a interactúan de mejor manera con las gráficas, debido que permite descargar los reportes en diversos formatos como imágenes, pdf, xlsx, entre otros. Apoyando a tomar decisiones informadas, sin embargo, para generar unas buenas graficas hay que considerar parámetros como las leyendas, textos y sobre todo los colores que hacen fácil interpretar en que mes y año hubo mayor ganancia por inversión y viceversa.

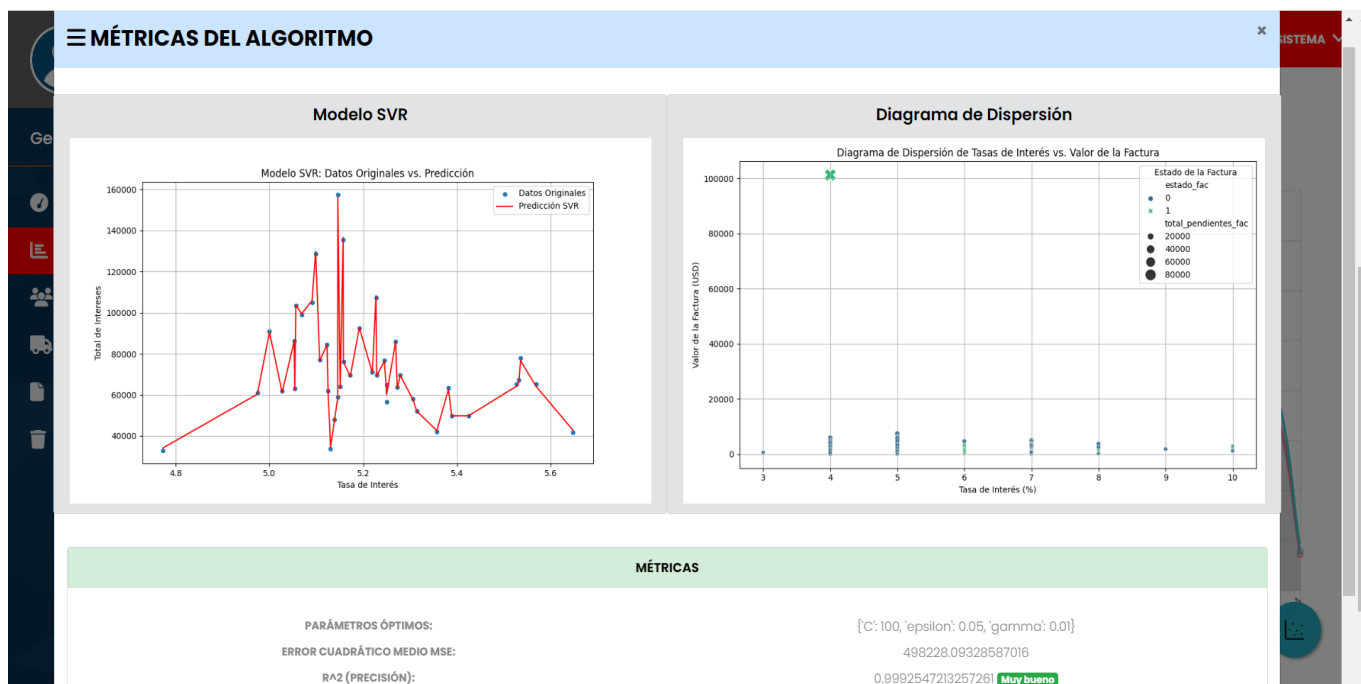
3.1. Resultados de las métricas de Evaluación

Para visualizar las métricas también se aplicaron visualizaciones como diagramas de dispersión y la gráfica del modelo para observar cómo se comporta el modelo con las variables proporcionadas, al tener datos dispersos SVR se adapta bastante bien debido a sus parámetros, su rapidez y su efectividad. Se consideraron las métricas ϵ , C , γ , R^2 , MSE . Esos valores se muestran posterior a mostrar los diagramas de dispersión con la gráfica del modelo que aplica SVR.

Se usa también GridSearchCV para que se halle los mejores parámetros que se pueden adaptar al modelo, es decir se tiene un espacio de hiperparámetros para evaluar y seleccionar la combinación que reduzca el error cuadrático medio y aumente el R^2 . Finalmente, el Cross Validation ayuda a evitar

problemas de sobreajuste y obtener una predicción mas realista del modelo, el CV se emplea en conjunto de GridSearchCV con el fin de ayudar a la selección de los parámetros que equilibren el ajuste del modelo y eviten el sobreajuste. A continuación, la figura 5 muestra las métricas de evaluación de manera gráfica:

Figura 5: Métricas de evaluación



Elaborado por: Autores

Gracias al correcto ajuste del modelo SVR se tiene un R^2 muy alto, y también la grafica del diagrama de dispersión lo confirma, el algoritmo se adapta de una buena forma a la dispersión de los datos, a continuación, se presenta la tabla 2 que interpreta las métricas proporcionadas por el modelo:

Tabla 2: Interpretación de las métricas del modelo

Elemento	Valor	Interpretación
C	100	Un valor alto prioriza un ajuste preciso a los datos de entrenamiento.
epsilon	0.05	Define un margen de tolerancia en el cual los errores no se penalizan.
gamma	0.01	Controla la influencia de cada dato de entrenamiento.
MSE	498,228.09	Cuantifica el error cuadrático promedio de las predicciones del modelo.
R^2	0.99925	Indica que el modelo tiene el 99.93% de precisión, sugiriendo ser alta.

Elaborado por: Autores

3.2. Discusiones:

Los resultados obtenidos reflexionan la importancia del volumen de datos, los algoritmos de predicción requieren gran cantidad de datos para mejorar de mejor forma y ayudar a tomar decisiones informadas. La selección de variables y optimización mediante hiperparámetros son cruciales para mejorar un modelo.

La metodología KDD apoyo a integrar técnicas para extraer información, modelar el algoritmo SVR y aplicar técnicas como GridSearchCV y Cross Validation, por otro lado, también apoya a evaluar el modelo mediante métricas como MSE y R^2 . La integración del algoritmo SVR y técnicas de inteligencia de negocios permitió visualizar los resultados del algoritmo mediante Dashboards.

El modelo SVR fue evaluado utilizando el Error Cuadrático Medio (MSE) para cuantificar la precisión de las predicciones. El MSE obtenido fue 498228.09328587016, lo que indica el promedio de la diferencia cuadrática entre las predicciones del modelo y los valores reales.

Para entender mejor el significado de este valor, consideremos el contexto de los datos. Si las cifras de facturación reales oscilan entre \$500,000 y \$1,500,000, un MSE de este tamaño podría sugerir un error considerable en las predicciones. Por otro lado, si las cifras de ventas están en un rango mucho mayor, el error relativo podría ser menor y, por tanto, el MSE podría ser aceptable. En conjunto con el Coeficiente de Determinación (R^2) de 0.99925, que indica que el modelo explica el 99.93% de la variabilidad en los datos, el MSE proporciona una visión complementaria de la precisión del modelo. Estos resultados reflejan que el modelo SVR tiene un alto grado de precisión en sus predicciones, y se comporta mejor con valores de facturación altos.

La integración del desarrollo de software, inteligencia de negocios e inteligencia artificial es crucial para que un sistema pueda apoyar a tomar decisiones de mejor manera, en el proyecto se aplicaron procesos ETL para subida masiva de información y para poder diseñar Dashboards y presentar información y gráficos amigables con el usuario.

Los dashboards tienen un impacto significativo en Acerotrans debido que se muestra información puntual pero clave para una mejor auditoria contable e informática, se logro un control de facturación, de ganancia por inversión y también sirve para identificar los posibles umbrales con transportistas que tienen baja su facturación.

Para 2024 la tendencia de facturación es menor que los otros años, esto debido a la situación del país, la escasez de viajes y el peligro en las vías. Sin embargo, esto depende el tipo de vehículo y el tonelaje puesto que a mayor tonelaje mayor es el pago, pero también hay que reflexionar que mayor son los gastos de la logística como comida, diésel, provisiones, peajes, entre otros.

Como parte de romper la brecha digital que se evidencia en la compañía se propone hacer capacitaciones a los transportistas y gerente con fin de mostrar las funcionalidades y los roles que tendrá cada uno, por ejemplo, la gerencia tendrá acceso al algoritmo de forma global, sin embargo, el transportista solo de sí mismo.

El impacto del aplicativo en los últimos meses ayudo a rendir cuentas a los transportistas en las reuniones mensuales, los resultados al ser gráficos se logra una mejor interpretación y sobre todo una mejor toma de decisiones anticipadas en 2024 que refleja ser un año no tan bueno para Acerotrans en general.

Como parte de mejora el sistema podría incorporar algoritmos de predicción futura para apoyar a la toma de decisiones anticipadas, pero siempre hay que recordar que el volumen de datos que esto requiere es mucho mayor al volumen de datos que se utilizó en la presente investigación.

Las técnicas del desarrollo del software como el despliegue y la integración continua, la escalabilidad y el desarrollo ágil están presentes en el proyecto a fin de obtener buenos resultados en el menor tiempo posible. Python al ser multiparadigma permite integrar inteligencia artificial con desarrollo

de software mediante Django de forma relativamente sencilla comparada a otros lenguajes de programación.

4. CONCLUSIONES

La implementación de algoritmos de aprendizaje automático juega un papel crucial en los sistemas informáticos, ya que es posible realizar análisis, predicciones y clasificación de los datos, permitiendo así que el apoyo a la toma de decisiones mejor informadas, para el proceso de integrar algoritmos con el desarrollo de software Python es el mejor aliado debido a sus cantidades de librerías y sus Frameworks para desarrollo web y a su vez el sistema implementado tiene gran escalabilidad de desarrollo, integración con más algoritmos y tecnologías innovadoras que ayuden a la organización a optimizar sus procesos y a tomar mejores decisiones.

El algoritmo SVR se adapta de buena forma a los datos dispersos, debido a sus parámetros que ofrece como ϵ mismo que considera pequeños errores que están fuera del modelo como parte del entrenamiento, el Control de regularización C que controla la magnitud de penalización que sufre el modelo respecto a los errores, además SVR cuenta con varios Kernels que se adaptan al panorama de los datos.

Los resultados obtenidos demuestran como el sistema implementado será de gran impacto en Acerotrans S.A. debido a las métricas que brinda el modelo como parte de la evaluación, de esta manera se podrán realizar los análisis respectivos entre los valores reales y predichos con fin de a futuro implementar algoritmos y técnicas que ayuden a predecir valores en la facturación de los transportistas pero a futuro, es decir el modelo se entrena con datos históricos y a partir de ello genere tendencias para los meses posteriores con fin de tomar decisiones mejores informadas.

La importancia de los datos fue crucial en la presente investigación, el modelo se entrena y da mejores resultados con datos en gran volumen, por ello un desafío fue la subida masiva de información, por lo que se emplearon técnicas de inteligencia de negocios como procesos ETL, y luego de los resultados del modelo también se recurrieron a técnicas de BI para generar Dashboards y reportes gráficos capaces de ser amigables e interpretables por el usuario.

Se debe considerar la arquitectura del sistema a la hora de desplegar el sistema, debido a que se esta trabajando con Machine Learning estos algoritmos requieren gran cantidad de recursos en especial una buena GPU y RAM; de esto depende bastante los tiempos de respuesta del algoritmo y las peticiones dentro del servidor.

AGRADECIMIENTOS

A la compañía de transporte pesado Acerotrans S.A. en representación de gerente general a la Ing. Tasipanta Myrian por todo el apoyo brindado para la presente investigación.

FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación

CONFLICTO DE INTERESES

Los Autores declaran que no existe conflicto de intereses con su investigación

CONTRIBUCIÓN DE AUTORÍA

En concordancia con la taxonomía establecida internacionalmente para la asignación de créditos a autores de artículos científicos (<https://credit.niso.org/>). Los autores declaran sus contribuciones en la siguiente matriz:

<i>Participar activamente en:</i>	Reinoso Diego.	Bedón Patricio
Conceptualización	X	
Análisis formal	X	
Adquisición de fondos	X	
Investigación	X	X
Metodología	X	X
Administración del proyecto	X	X
Recursos	X	X
Redacción –borrador original	X	X
Redacción –revisión y edición	X	X
La discusión de los resultados	X	X
Revisión y aprobación de la versión final del trabajo.	X	X

REFERENCIAS

- Castro, R. M. S., & Vega, I. D. L. (2023). Redes neuronales aplicadas al control estadístico de procesos con cartas de control EWMA. *Tecnura*, 27(75), 72-88.
- Forero-Corba, W., & Negre Bennasar, F. (2023). Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: Una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 209-253. <https://doi.org/10.5944/ried.27.1.37491>
- García-González, J. R., Sánchez-Sánchez, P. A., Orozco, M., Obredor, S., García-González, J. R., Sánchez-Sánchez, P. A., Orozco, M., & Obredor, S. (2019). Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia. *Formación universitaria*, 12(4), 55-62. <https://doi.org/10.4067/S0718-50062019000400055>
- Lambis-Alandete, E., Jiménez-Gómez, M., Velásquez-Henao, J. D., Lambis-Alandete, E., Jiménez-Gómez, M., & Velásquez-Henao, J. D. (2023). Comparación de algoritmos de Deep Learning para pronósticos en los precios de Criptomonedas. *Ingeniería y competitividad*, 25(3). <https://doi.org/10.25100/iyc.v25i3.12845>
- Mamani Rodriguez, Z. (2022). Proceso de machine learning para determinar la demanda social de puestos de empleo de profesionales de TI. *Industrial Data*, 25(2), 275-300. <https://doi.org/10.15381/idata.v25i2.21643>
- Osorio, D. A. I. (2023). Sistemas Informáticos: Los factores para mejorar los resultados de las empresas y organizaciones. *Revista Científica Internacional*, 6(1), 98-110. <https://doi.org/10.46734/revcientifica.v6i1.58>
- Páez, A. R., & Ramírez, N. D. G. (2022). Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería. *IE Revista de Investigación Educativa de la REDIECH*, 13. <https://www.redalyc.org/journal/5216/521670731008/html/>
- ¿Qué es la minería de datos? | IBM. (2023, julio 18). <https://www.ibm.com/es-es/topics/data-mining>
- Rivera, R. M., Cámara, F. A., Jiménez, D. E., & Díaz, S. H. (2016). *SISDAM: Web application for processing data according to a Modified Augmented Design*. <https://doi.org/10.13140/RG.2.1.4550.4243>
- Sanabria-Castro, A., Meneses-Guzmán, M., & Chiné-Polito, B. (2022). Uso de regresión de soporte vectorial para el control de espuma metálica. *Revista Tecnología en Marcha*. <https://doi.org/10.18845/tm.v36i1.5891>
- Sánchez, E. R. B., Chilán, J. H. M., & Vásquez, C. A. Á. (2020). Gestión de la información para la toma de decisiones de pequeños y medianos negocios en Jipijapa, Manabí, Ecuador. *Avances*, 22(2), 183-193.
- Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. Universidad Cooperativa de Colombia. <https://doi.org/10.16925/9789587600490>
- TIOBE Index. (s. f.). TIOBE. Recuperado 13 de junio de 2024, de <https://www.tiobe.com/tiobe-index/>
- Useche-Aguirre, M. C., Pereira-Burgos, M. J., & Barragán-Ramírez, C. A. (2021). Retos y desafíos del emprendimiento ecuatoriano, trascendiendo a la pospandemia. *RETOS. Revista de Ciencias de la Administración y Economía*, 11(22), 271-286.

- Valenzuela-Nunez, C. I., Troncoso Espinosa, F. H., & Latorre-Nunez, G. O. (2023). Prediction of absenteeism in medical appointments using Machine Learning. *Universidad Ciencia y Tecnología*, 27(120), 19-30. <https://doi.org/10.47460/uct.v27i120.728>
- Vargas-Zapata, M., Medina-Sierra, M., Galeano-Vasco, L. F., & Cerón-Muñoz, M. F. (2022). Algoritmos de aprendizaje de máquina para la predicción de propiedades fisicoquímicas del suelo mediante información espectral: Una revisión sistemática. *Revista de Investigación, Desarrollo e Innovación*, 12(1), 107-120. <https://doi.org/10.19053/20278306.v12.n1.2022.14212>
- Vidal-Silva, C. L., Sánchez-Ortiz, A., Serrano, J., Rubio, J. M., Vidal-Silva, C. L., Sánchez-Ortiz, A., Serrano, J., & Rubio, J. M. (2021). Experiencia académica en desarrollo rápido de sistemas de información web con Python y Django. *Formación universitaria*, 14(5), 85-94. <https://doi.org/10.4067/S0718-50062021000500085>
- Zambrano, C. A. Y., Campuzano, M. F. P., & Laz, E. M. S. A. (2019). La Gestión De La Información Y El Conocimiento a Partir De Estrategias Formativas Innovadoras. *Revista de Ciencias Humanísticas y Sociales (ReHuSo)*, 4(1), 109-121.