

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



Solución de inteligencia de negocios basada en indicadores de gestión y predicción para apoyar a la toma de decisiones del área comercial de una empresa en Chiclayo

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

AUTOR

Wilber Guerrero Gamarra

ASESOR

Segundo Jose Castillo Zumaran

<https://orcid.org/0000-0001-5613-5519>

Chiclayo, 2024

**Solución de inteligencia de negocios basada en indicadores de
gestión y predicción para apoyar a la toma de decisiones del área
comercial de una empresa en Chiclayo**

PRESENTADA POR
Wilber Guerrero Gamarra

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

Maria Ysabel Aranguri Garcia
PRESIDENTE

Huiler Juanito Mera Montenegro
SECRETARIO

Segundo Jose Castillo Zumaran
VOCAL

Dedicatoria

Esta tesis, que es prueba de mi esfuerzo y dedicación, se la dedico a mis padres. Gracias por creer en mí y brindarme la fortaleza necesaria para seguir adelante, tanto en mi trayectoria académica como en el camino de la vida.

Agradecimientos

A mi asesor, cuyo valioso conocimiento y constante apoyo fueron fundamentales para la realización de esta tesis.

A quienes fueron mis jefes durante mis prácticas, por su orientación y apoyo a lo largo de este proceso.

A mis amigos, quienes con su ánimo me acompañaron en este complicado camino de elaborar esta tesis.

Solución de inteligencia de negocios basada en indicadores de gestión y predicción para apoyar a la toma de decisiones del área comercial de una empresa en Chiclayo

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

17%

FUENTES DE INTERNET

5%

PUBLICACIONES

6%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	6%
2	tesis.usat.edu.pe Fuente de Internet	1%
3	www.coursehero.com Fuente de Internet	1%
4	repositorio.autonoma.edu.pe Fuente de Internet	1%
5	vdocuments.mx Fuente de Internet	1%
6	Submitted to Universidad Católica Santo Toribio de Mogrovejo Trabajo del estudiante	<1%
7	www.slideshare.net Fuente de Internet	<1%
8	www.mysciencework.com Fuente de Internet	<1%

Índice

Resumen	7
Abstract	8
Introducción.....	9
Revisión de literatura.....	11
Antecedentes	11
Antecedentes internacionales.....	11
Antecedentes nacionales	12
Antecedentes locales	13
Toma de decisiones.....	13
Tipos de decisiones organizacionales.....	13
Toma de decisiones basada en datos.....	14
Inteligencia de negocios	14
Sistema de soporte a las decisiones.....	14
Almacenes de datos	15
Sistema ETL	15
Tablero KPI	16
Metodologías para la implementación de sistemas de inteligencia de negocios.....	16
Minería de datos	17
Metodologías para la implementación de soluciones de minería de datos	18
Materiales y métodos	19
Tipo de investigación.....	19
Métodos de investigación.....	19
Técnicas e instrumentos de recolección de datos	20
Metodología de desarrollo	20

Resultados y discusión	21
Conclusiones	26
Recomendaciones	26
Referencias.....	27
Anexos	30
Anexo N° 01. Constancia de aprobación del producto acreditable	30
Anexo N° 02. Entrevista.....	31
Anexo N° 03. Cuestionario de evaluación de calidad de la solución desarrollada 1 .	32
Anexo N° 04. Cuestionario de evaluación de calidad de la solución desarrollada 2 .	33
Anexo N° 05. Pruebas de caja negra realizadas al sistema ETL	34
Anexo N° 06. Pruebas de caja blanca realizadas al sistema ETL.....	36
Anexo N° 07. Pruebas de caja negra realizadas a la aplicación de inteligencia de negocios	38
Anexo N° 08. Prueba de conciliación de datos realizada a la aplicación de inteligencia de negocios.....	40

Resumen

La tesis se enfocó en la implementación de una solución de inteligencia de negocios destinada a mejorar la toma de decisiones en el área comercial de una empresa en Chiclayo. Utilizando una metodología híbrida que combinó elementos de Ralph Kimball y CRISP-DM, se lograron resultados notables. Primero, se completó con éxito la implementación del 100% de los indicadores clave de rendimiento en reportes interactivos, los cuales se integraron al sistema transaccional a través de un módulo personalizado. En segundo lugar, se determinó que el algoritmo "Voting Ensemble" era el más eficaz para predecir ventas, con un coeficiente de determinación (R^2) de 0.60771, el más alto entre los algoritmos probados. Tercero, se realizaron predicciones precisas de ventas basadas en datos históricos, las cuales se integraron en los reportes, brindando información valiosa sobre la rentabilidad futura de los productos. Finalmente, la evaluación de la calidad de la solución según el estándar ISO/IEC 25010 arrojó calificaciones sobresalientes, destacando la adecuación funcional, usabilidad y mantenibilidad.

Palabras clave: inteligencia de negocios, toma de decisiones, data mart, ETL

Abstract

The thesis focused on the implementation of a business intelligence solution aimed at improving decision-making in the commercial area of a company in Chiclayo. Using a hybrid methodology that combined elements of Ralph Kimball and CRISP-DM, remarkable results were achieved. First, the implementation of 100% of the key performance indicators in interactive reports was successfully completed, which were integrated into the transactional system through a customized module. Secondly, it was determined that the "Voting Ensemble" algorithm was the most effective for predicting sales, with a coefficient of determination (R^2) of 0.60771, the highest among the tested algorithms. Thirdly, accurate sales predictions were made based on historical data, which were integrated into the reports, providing valuable information about the future profitability of the products. Finally, the evaluation of the quality of the solution according to the ISO/IEC 25010 standard yielded outstanding ratings, highlighting the functional adequacy, usability and maintainability.

Keywords: business intelligence, decision making, data mart, ETL

Introducción

Toda empresa, en la realización de sus procesos, gestiona información que se convierte en un activo valioso que permite conocer la condición actual de la empresa y así tomar decisiones acertadas. [1] Dada la importancia de la información en la toma de decisiones, el uso de los sistemas de inteligencia de negocios resulta vital, debido a que existe una relación estrecha y directa entre la inteligencia empresarial y la toma de decisiones en un contexto empresarial [2], pues se les proporciona a las personas encargadas de tomar decisiones información tanto significativa como relevante para su análisis y evaluación. [3].

En el ámbito mundial, la empresa Information Builders Ibérica [4], afirmó que, en promedio, el empleado europeo pierde alrededor de 67 minutos cada día en la búsqueda de información relacionada con la compañía, lo que representa un 15.9 % de sus horas laborales. Adicionalmente, una encuesta aplicada por la empresa MicroStrategy [5], halló que el 94 % de los encuestados consideran que la información es importante para el crecimiento empresarial.

En el ámbito nacional, un artículo publicado en la Revista Latinoamericana y del Caribe de la AIS [6], concluye que las empresas que operan en Perú tienen más necesidad de apoyo tecnológico para la toma de decisiones empresariales y este problema afecta en mayor medida a las empresas más pequeñas.

En el contexto de ventas de una empresa, estas deben evaluar la situación actual antes de agregar otro producto a su catálogo, pues existe la necesidad de que generen utilidad y retornar el capital invertido, por lo que se debe conocer la situación actual de las ventas con el fin de tomar decisiones precisas y bien fundamentadas. [7]

En el contexto de la investigación, esta se aplicará a una empresa especializada en la comercialización de productos agropecuarios que cuenta con tres sedes en distintas ciudades del país. Tanto el área de ventas como el área de compras de la empresa utilizan un sistema transaccional, que consiste en una aplicación de escritorio programada en Visual FoxPro que da mantenimiento a las tablas de una base de datos hecha en el motor Postgres. Los reportes generados por el sistema transaccional de la empresa proporcionan información sobre las utilidades por producto, cliente y sede, incluyendo un ranking de los diez productos y clientes más rentables, así como un resumen de los ingresos por sede. No obstante, la falta de detalles en los reportes genera incertidumbre dentro del proceso de toma de decisiones correspondiente al área de ventas. Esto afecta la confiabilidad de los informes y provoca una demora en el proceso de toma de decisiones, lo que causa descontento entre los tomadores de decisiones. En consecuencia, es necesario buscar soluciones que permitan mejorar la calidad y confiabilidad de los reportes generados por el sistema transaccional.

Aunque se ha observado una tendencia positiva en el uso de sistemas de inteligencia de negocios en el país [6], las empresas peruanas aún están en proceso de adaptación en cuanto a la gestión de la información, lo que sugiere que todavía falta conciencia sobre la importancia de la información para alcanzar objetivos y tomar decisiones efectivas [8].

El proceso de toma de decisiones empresariales es uno de los aspectos más críticos en el éxito de una organización, por ello, de resultar que la información disponible no es suficiente para comprender completamente la situación, puede generarse incertidumbre [9]. Esta situación impide un análisis objetivo y aumenta el riesgo de tomar decisiones incorrectas, lo que afecta negativamente el desempeño de la organización [10], pues la persona encargada de la toma de decisiones se enfrenta a mayor presión y puede optar por soluciones que tienen una mayor probabilidad de fracaso, lo que a su vez puede causar problemas económicos, de tiempo y de esfuerzo [11]. Las consecuencias de una mala toma de decisiones pueden ser graves y pueden afectar a toda la organización durante un período prolongado de tiempo, dependiendo de la magnitud de las decisiones tomadas [12].

Para apoyar a la toma de decisiones de empresas se han propuesto distintos tipos de soluciones, por ejemplo, Requejo y Pisfil [13] presentan un modelo de toma de decisiones basado en el modelo racional, que consiste en seis fases a seguir para facilitar las decisiones en las pymes.

Con base en lo expuesto, se establece la pregunta ¿De qué manera una solución de inteligencia de negocios puede apoyar el proceso de toma de decisiones del área comercial en una empresa?

El objetivo principal es implementar una solución de inteligencia de negocios para apoyar a la toma de decisiones en el área comercial de la empresa. Los objetivos específicos son: implementar los indicadores clave de rendimiento del área comercial de la empresa en reportes interactivos para desarrollar una solución a medida; determinar el mejor algoritmo para predecir las ventas de los productos; predecir las ventas de los productos con base en la información histórica para conocer la rentabilidad a futuro de estos; y verificar la calidad de la solución desarrollada mediante la aplicación del estándar ISO/IEC 25010.

En el ámbito científico, esta investigación servirá como antecedente de futuras tesis o investigaciones relacionadas al apoyo a la toma de decisiones, pues se produjo un trabajo con bases sólidas en comparación a muchos trabajos relacionados que no cuentan con citas debidas, haciendo que la investigación se vuelva subjetiva.

En el ámbito tecnológico, se requiere una solución de inteligencia de negocios para apoyar a la toma de decisiones, ya que existe una interrelación directa entre ambos [2] y se ha

demostrado que produce un impacto positivo en el desempeño en organizaciones pequeñas y medianas [14], siendo de utilidad para las decisiones tomadas en el área comercial [15]. También se requiere la utilización de una metodología para asegurar la correcta implementación del *data mart*. Esta solución optimizará el proceso de toma de decisiones en el área comercial y permitirá conocer mejor la realidad actual y futura del área con base en los datos almacenados a través de una solución aplicable a la realidad.

Revisión de literatura

Antecedentes

Antecedentes internacionales

En su artículo, Negro y Mesia [2] llevaron a cabo una revisión bibliográfica con el fin de determinar el grado de conexión e interdependencia directa que existe entre la inteligencia de negocios y la toma de decisiones organizacionales. Tras realizar un análisis de la información encontrada, los autores concluyen que existe una fuerte interdependencia, conexión e interrelación entre la inteligencia de negocios y la toma de decisiones organizacionales en un entorno empresarial global, en el que ambas se influyen mutuamente de manera directa. Este trabajo se tomó en cuenta porque demuestra que el tipo de solución propuesta en esta investigación puede tener efecto en el proceso de toma de decisiones.

En el artículo de Lawton [16], publicado en 2021, se analiza el impacto del uso de datos en decisiones importantes, dando como ejemplo destacable el replaneamiento de las rutas de los buses escolares en la ciudad de Boston, en Estados Unidos, reduciendo así un aproximado de 9.07 toneladas de emisión de dióxido de carbono. Este trabajo aporta a la investigación en la demostración de la importancia de las decisiones orientadas a datos, como puede ser una solución de inteligencia de negocios.

En su investigación, Bolaños *et al.* [17], el problema tratado es el alto número de accidentes de tránsito reportados en la ciudad de Popayán, en Colombia. Los autores implementan un *data warehouse*, para luego generar reportes gráficos que permitan comprender la situación actual y tomar mejores decisiones respecto al problema, para llevar esto a cabo, se siguió la metodología Ralph Kimball. Este trabajo tiene relación con la investigación porque aborda la dificultad para la toma de decisiones y cómo la inteligencia de negocios ayuda a reducirla.

En su artículo, los autores Popovič, Puklavec y Oliveira [14] analizaron los datos obtenidos de encuestas realizadas a ciento ochenta y un empresas pequeñas y medianas siguiendo el modelo de ecuaciones estructurales por el método de mínimos cuadrados parciales (PLS-SEM). Los resultados del análisis mostraron que el uso de sistemas de inteligencia de negocios tiene un impacto positivo en el desempeño general de las empresa pequeñas y medianas. Se tomó en cuenta este trabajo porque demuestra que la solución propuesta en la investigación puede ser aplicada en una empresa mediana, que es el tipo de la empresa en la que se aplicará la investigación.

En su artículo, los autores Guerreiro, Correia y Sousa [15] realizaron un análisis estadístico de una muestra de clientes de un hotel en Portugal con la finalidad de evaluar sus preferencias de consumo en los establecimientos de alimentación y bebida. Los autores concluyen en que es importante un sistema de inteligencia de negocios para el análisis de datos numerosos en la realidad comercial de un hotel. Este trabajo resulta relevante para esta investigación porque muestra que la inteligencia de negocios resulta de utilidad en la realidad comercial, aunque en esta investigación se aplicará en un contexto de ventas de otro sector económico.

En su artículo, los autores Koptilina, Bulgakov y Aleshina [18] hacen un análisis de cómo es usada la inteligencia de negocios en la agroindustria y es de importancia para esta investigación debido a que la solución se implementará en dicho sector económico.

Antecedentes nacionales

Piedra [19] narra la problemática de obtener información para la gestión de forma oportuna en la Oficina Centra de Admisión de una universidad peruana. Con la aplicación de la metodología Timo Elliot, construyó un aplicativo de inteligencia de negocios, el cual ayudó a agilizar el uso de la información para la toma de decisiones. Tanto en dicha tesis como en esta tesis se utiliza una herramienta en común, además de que la inclusión del análisis de los reportes creados ayudó a formular mejor los indicadores clave de rendimiento en esta tesis.

En su investigación aplicada en la ciudad de Lima, Portal y Quispe [20] trataron el problema de la dificultad que conlleva realizar informes debido a la gran cantidad de datos registrados por los distintos medios de atención de la compañía. Ante la situación presentada, los investigadores propusieron una solución de inteligencia de negocios, con ello se mejoró el proceso de toma de decisiones y se redujo el tiempo ocupado en las actividades implicadas en dicho proceso. Para lograr su objetivo, se siguió la metodología Ralph Kimball. Este trabajo es de utilidad para la investigación porque brinda recomendaciones para la implementación del sistema ETL de la solución de inteligencia de negocios.

Apolaya [21] abordó la problemática de la elaboración manual e individual de los reportes en el área comercial de una empresa que brinda servicios de seguridad ocupacional en Lima. Para hacer frente a la problemática, se usó la metodología Ralph Kimball para la implementación de inteligencia de negocios. Como resultado, se mejoró la calidad de la información, se redujo el tiempo para la elaboración de informes y se crearon indicadores de gestión. La relación que se guarda con este trabajo es que se está aplicando al área comercial de una empresa y la utilización de la herramienta Power BI para la implementación de los reportes.

Antecedentes locales

Villanueva [22] habla de la problemática de la toma de decisiones gerenciales, ante la cual, aplicando la metodología de Ralph Kimball, construyó una solución de inteligencia de negocios con la que logró mejorar el proceso de toma de decisiones gerenciales. La relación que guarda con esta tesis es que se aplicó la solución a los procesos de compra y venta, también, se incluyeron predicciones en los reportes elaborados con Power BI.

Valderrama [23], en su investigación, relata sobre la problemática de no tener una herramienta que permita brindar soporte a la información que se maneja en la Gerencia Regional de Agricultura – Lambayeque. Para dar solución a este problema, el autor aplicó una metodología híbrida basada en Ralph Kimball y CRISP-DM, con la cual se llevó a cabo la implementación de una solución de inteligencia de negocios basada en técnicas de minería de datos. Con dicha solución, logró implementar reportes relacionados a siembras y cosechas, así como generar un pronóstico de producción de cultivos. La relación que guarda con esta tesis es la utilización de una metodología híbrida; así, se tuvo en cuenta la organización de las actividades la metodología para la ejecución de esta tesis.

Toma de decisiones

Según el autor Idalberto Chiavenato [24], la toma de decisiones se define como «un procedimiento mediante el cual se lleva a cabo un análisis y se escoge entre diversas alternativas un plan de acción».

Tipos de decisiones organizacionales

El autor anteriormente mencionado [24] clasifica las decisiones organizacionales en dos grupos:

a. Decisiones programadas

Son las decisiones cotidianas, estas se repiten frecuentemente y el problema se resuelve siguiendo procedimientos ya establecidos. Son decisiones de tipo estructurado, pues se cuenta con información y opciones claras para enfrentar la situación.

b. Decisiones no programadas

Al contrario del anterior tipo de decisiones, se presentan de manera ocasional y no se cuenta con procedimientos ya establecidos para seguir, de manera que los tomadores de decisiones no se enfrentan a la incertidumbre.

La solución de inteligencia de negocios que se implementa en la presente investigación está enfocada principalmente a la toma de decisiones programadas, pues se está realizando la implementación de los indicadores clave de rendimiento (KPI), aunque esto no significa que no apoye en absoluto a la toma de decisiones no programadas, simplemente lo hace de manera más limitada, puesto que no envía alertas en tiempo real cuando se detectan desviaciones significativas de los valores esperados.

Toma de decisiones basada en datos

El término toma de decisiones basada en datos describe el proceso de utilizar datos reales de la organización para la toma de decisiones sin basarse solamente en la intuición. [25]

Un artículo publicado por Tableau [26] menciona que la cantidad de información que se recolecta en la actualidad nunca ha sido tan grande y compleja, lo cual dificulta su gestión y análisis. Un hecho que evidencia la importancia de este tema es un estudio publicado en Harvard Business Review [27], donde se menciona que las organizaciones están haciendo grandes inversiones en modernizar sus negocios a través de la tecnología; esto, acompañado de una cultura de datos puede representar una mejoría en la toma de decisiones de una organización.

Inteligencia de negocios

El término inteligencia de negocios fue acuñado por Howard Dresner y el Grupo Gartner en 1989, y hace referencia a un conjunto de conceptos y técnicas diseñados para mejorar la toma de decisiones en el ámbito empresarial mediante el uso de un sistema de apoyo a la toma de decisiones. [28]

Sistema de soporte a las decisiones

Un sistema de apoyo a la toma de decisiones es un sistema de información interactivo que utiliza datos y modelos para ayudar a los gerentes a analizar problemas semiestructurados. [29]

En un comienzo, los sistemas de soporte a decisiones estaban diseñados para apoyar a cualquier tomador de decisiones de la organización, sin embargo, con el paso de los tiempos empezaron a los gerentes intermedios y superiores o también para tipos de organizaciones en específico. [28]

Almacenes de datos

Data warehouse

De manera tradicional, el Grupo Kimball llamaba data warehousing al proceso en general de proveer información para apoyar a la toma de decisiones de negocio, sin embargo, existe un uso del término para referirse al almacén de datos central de todo el entorno de inteligencia de negocios, por lo cual el Grupo Kimball ahora se refiere al concepto tradicional de *data warehousing* como *data warehousing/business intelligence* (DW/BI). [30] En el caso de esta investigación, el término se referirá al almacén de datos central.

Data mart

Un *data mart* es una base de datos centrada en un proceso y representa un subconjunto de la arquitectura de datos general de la empresa. Sin embargo, el término ha sido usado incorrectamente para referirse a soluciones independientes destinadas a un departamento de la empresa. [30]

En esta investigación, será este uno de los componentes de la solución de inteligencia de negocios. La razón por la que no se construye todo un *data warehouse* es porque el producto se centrará en dos procesos principales de la empresa solamente.

Sistema ETL

El proceso de extracción, transformación y carga (ETL) se encarga de recolectar los datos de los sistemas de origen para su posterior presentación y está compuesto por un área de trabajo, una estructura de datos y un conjunto de procesos. [31]

En esta investigación, uno de los componentes del producto final será un sistema ETL personalizado que permitirá transportar los datos del sistema transaccional a la solución de inteligencia de negocios.

Tablero KPI

Según Qlik [32], una reconocida compañía de analítica de negocios, un tablero KPI muestra los indicadores clave de rendimiento a través de tablas y gráficos interactivos que permiten hacer revisión y análisis de manera rápida y organizada.

Esta será la primera pantalla que se mostrará en la aplicación de inteligencia de negocios del producto acreditable, así se puede tener un panorama general de la situación actual del área comercial de la empresa.

Metodologías para la implementación de sistemas de inteligencia de negocios

Bill Inmon

La metodología propuesta por Inmon presenta un ciclo de vida de desarrollo de sistemas (CVDS) que difiere significativamente del CVDS clásico, como se puede apreciar en la siguiente figura. [33]

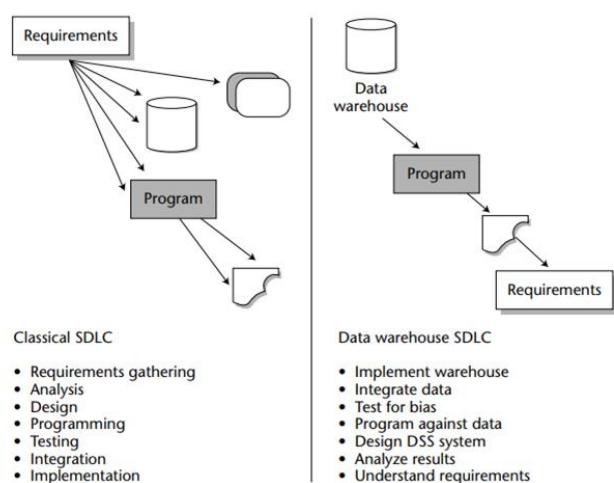


Fig. 1. El ciclo de vida de desarrollo de sistemas (CVDS) para el entorno del data warehouse es casi lo contrario del CVDS clásico [33]

Esta metodología tiene un enfoque up-bottom, es decir, primero se encarga de realizar el *data warehouse*, que será de donde se obtienen los *data mart*. [33]

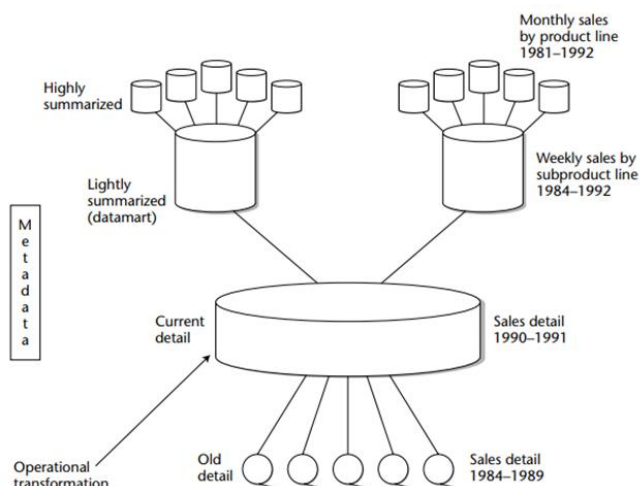


Fig. 2. Estructura del data warehouse [33]

Esta metodología es comparada en muchas investigaciones con la que se mencionará a continuación, pero se suele optar por la segunda debido a que implica menos costos de implementación, misma razón por la que se optó por Ralph Kimball en esta investigación.

Ralph Kimball

La metodología de Kimball se basa en un modelado dimensional, empezando por los *data mart*. La implementación exitosa de un sistema de inteligencia de negocios depende de la integración de varias tareas y componentes, no solo del modelo de datos. La metodología Kimball sigue un ciclo de vida denominado el Ciclo de Vida de Kimball que puede apreciarse en la figura 1. [30]

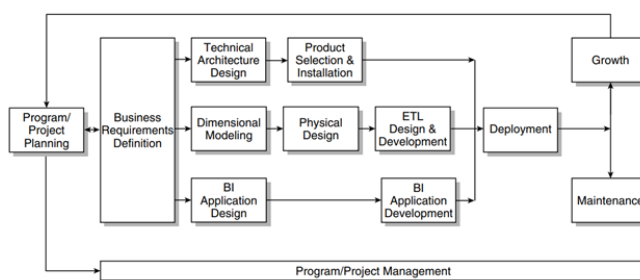


Fig. 3. Diagrama del ciclo de vida de Kimball [30]

Minería de datos

La minería de datos la base matemática del proceso de descubrimiento de conocimiento en bases de datos (conocido como KDD, siglas del inglés *knowledge discovery in databases*), implica el uso de algoritmos de inferencia para examinar los datos, desarrollar modelos matemáticos e identificar patrones significativos tanto implícitos como explícitos. [34]

En este contexto, la minería de datos se utilizó para la creación de un modelo de predicción de ventas de los artículos de la empresa donde se aplicó la presente investigación.

Metodologías para la implementación de soluciones de minería de datos

CRISP-DM

El modelo de proceso de CRISP-DM provee el ciclo de vida de un proyecto de minería de datos, que está constituido por seis fases, cada una con sus respectivas tareas. [35]

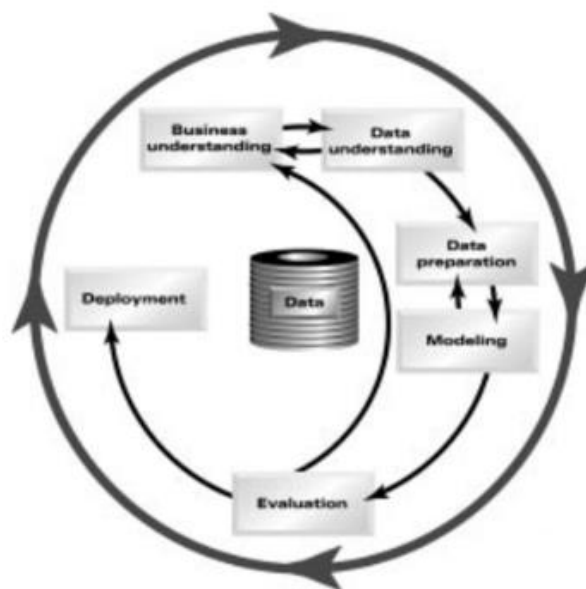


Fig. 4. Fases de CRISP-DM [35]

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives <i>Background Business Objectives Business Success Business Success Criteria</i>	Collect Initial Data <i>Initial Data Collection Report</i>	Select Data <i>Rationale for Inclusion/ Exclusion</i>	Select Modeling Techniques <i>Modeling Technique Modeling Assumptions</i>	Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</i>	Plan Deployment <i>Deployment Plan</i>
Assess Situation <i>Inventory of Resources Requirements, and Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</i>	Describe Data <i>Data Description Report</i>	Clean Data <i>Data Cleaning Report</i>	Generate Test Design <i>Test Design</i>	Review Process <i>Review of Process</i>	Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i>
Determine Data Mining Goals <i>Data Mining Goals Data Mining Success Criteria</i>	Explore Data <i>Data Exploration Report</i>	Construct Data <i>Derived Attributes Generated Records</i>	Build Model <i>Parameter Settings Models Model Descriptions</i>	Determine Next Steps <i>List of Possible Actions Decision</i>	Produce Final Report <i>Final Report Final Presentation</i>
Produce Project Plan <i>Project Plan Initial Assessment of Tools and Techniques</i>	Verify Data Quality <i>Data Quality Report</i>	Integrate Data <i>Merged Data</i>	Assess Model <i>Model Assessment Revised Parameter Settings</i>		Review Project <i>Experience Documentation</i>
		Format Data <i>Reformatted Data Dataset Dataset Description</i>			

Figure 3: Generic tasks (bold) and outputs (italic) of the CRISP-DM reference model

Fig. 5. Tareas genéricas y sus resultados del modelo de referencia CRISP-DM [35]

SEMMA

Para el Instituto SAS [36], la minería de datos es el proceso que abarca el muestreo, la exploración, la modificación, el modelado y la evaluación (abreviado SEMMA, del inglés *sampling, exploring, modifying, modeling, and assessing*) de grandes cantidades de datos con el fin de descubrir patrones que no se conocían anteriormente.

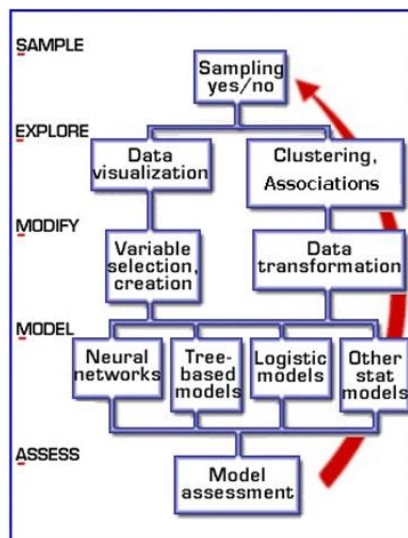


Fig. 6. Diagrama de flujo del proceso SEMMA para minería de datos [36]

La elección de utilizar CRISP-DM en lugar de SEMMA se justifica por la mayor aceptación y flexibilidad de CRISP-DM en la industria, su estructura detallada y su enfoque en comprender las necesidades empresariales. CRISP-DM proporciona una guía más completa que se adapta bien a proyectos integrales y permite una gestión efectiva a lo largo de todo el ciclo de vida del proyecto.

Materiales y métodos

Tipo de investigación

Según el Manual de Frascati 2015 [37], esta investigación es de tipo aplicada, pues se resolverá un problema en un contexto específico.

Métodos de investigación

Los métodos de investigación empleados fueron los que se muestran en la tabla a continuación.

TABLA I MÉTODOS DE INVESTIGACIÓN	
Método	Descripción
Analítico [38]	Se estudió y analizó el problema correspondiente a la toma de decisiones en Sanidad Agropecuaria & Asociados S.A.C.
Implementación [39]	Se implementó una solución de inteligencia de negocios para apoyar la toma de decisiones en Sanidad Agropecuaria & Asociados S.A.C.

Técnicas e instrumentos de recolección de datos

A continuación, se muestra en la siguiente tabla las técnicas e instrumentos que se utilizaron para la recolección de datos

TABLA II
TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Técnicas	Instrumentos	Elementos de la población	Propósito
Entrevista [39]	Anotaciones en papel	Gerente	Se realizó para obtener los indicadores clave de rendimiento (KPI) y los requerimientos de la solución que se presentó
Encuesta [39]	Cuestionario	Desarrolladores	Se realizó encuestas para validar información relacionada a la solución implementada
<i>(ver anexos N° 02, 03 y 04)</i>			

Metodología de desarrollo

Para el desarrollo de esta investigación, se utilizó una metodología híbrida basada en Ralph Kimball [30] y CRISP-DM [35]:

TABLA III
FASES DE LA METODOLOGÍA UTILIZADA

Ralph Kimball	CRISP-DM
Planeación del proyecto Se define la justificación y el alcance del proyecto	-
Definición de requerimientos del negocio Se especifican los requerimientos y se procesan para que sean considerados en el diseño	Comprensión del negocio Determinación de los objetivos del negocio, requerimientos
Diseño de la arquitectura técnica Se define las tecnologías a integrar teniendo en cuenta los requerimientos del negocio, el entorno técnico y las direcciones técnicas estratégicas planificadas.	-
Selección e instalación de productos Se decide los componentes específicos que conformarán la arquitectura técnica diseñada, como la plataforma de hardware, el sistema de gestión de base de datos, la herramienta ETL y la herramienta de reportes.	-

Modelado dimensional	Comprensión de los datos
Se elabora una matriz bus del data warehouse empresarial, que representa los procesos clave de la organización y es tomada en cuenta al elaborar la arquitectura de datos, de manera que la solución de inteligencia de negocios pueda ser integrada en la organización y sea escalable.	Se explora los datos para un mejor entendimiento de cómo están estructurados.
Diseño físico	-
Se selecciona las estructuras necesarias para soportar el diseño físico.	
Diseño y desarrollo del sistema ETL	Preparación de los datos
Se elabora el diseño de los procesos que tendrá el sistema ETL y se procede a implementarlo	Se selecciona y limpia los datos presentes en el dataset
Diseño y desarrollo de aplicación de inteligencia de negocios	-
Se diseña e implementa las interfaces que mostrarán la información al usuario y el flujo de navegación entre estas	
-	Modelado de datos
	Se elige la técnica de modelado y se construirá el modelo
-	Evaluación
	Evaluación de los resultados
Despliegue	
Los productos mencionados anteriormente son llevados a su respectivo alojamiento para que puedan ser usados en un entorno de producción	

Resultados y discusión

Para gestionar su área comercial, la empresa utiliza un sistema transaccional que permite registrar y mantener información precisa sobre compras, ventas, proveedores, clientes, inventario y trabajadores. Gracias a esta solución, la empresa puede capturar y almacenar los datos relevantes de sus operaciones comerciales.

El sistema transaccional actual presenta limitaciones en los reportes, los cuales son estáticos y carecen de interactividad. Estas limitaciones dificultan la comprensión de ciertos datos, ya que la visualización de información solo se presenta en forma de texto o tablas sin gráficos o elementos visuales que faciliten su interpretación. Por lo antes mencionado, se requiere facilitar la comprensión y el análisis de los datos, brindando una visión más clara y detallada de la información empresarial.

A través de la aplicación de una metodología híbrida basada en Ralph Kimball y CRISP-DM, se desarrolló una solución de inteligencia de negocios para apoyar a la toma de decisiones.

Para desarrollar el *data mart*, se utilizó Postgresql, pues se ha definido como la tecnología preferida en el desarrollo de sistemas de la empresa, además de que el sistema transaccional lo usa para almacenar la información. En cuanto al sistema ETL, debido a que no existe un lenguaje a usar en este tipo de solución definido por la empresa, se elige con base en el criterio del ejecutor del proyecto. El sistema ETL está programado en el lenguaje Python debido a las facilidades que proporciona para el manejo de datos y la vasta documentación que existe para el procesamiento de datos utilizando dicho lenguaje. Para la visualización de los datos, se decidió usar Power BI, pues cuenta con herramientas como el lenguaje DAX, que permiten manejar los datos que se tiene en el *data mart*, haciendo posible implementar los KPI con mayor facilidad. Además, Power BI tiene las herramientas necesarias para integrar los reportes creados en otros sistemas existentes, que es lo que desea hacer la empresa en la que se realizará el proyecto.

A través de una entrevista al encargado del área comercial, se recolectaron los indicadores clave de rendimiento (KPI) que se implementaron en su totalidad en reportes interactivos. Los KPI son los siguientes: compras mensual y anual, ventas mensual y anual, utilidad mensual y anual, variedad de productos vendidos, compras año actual vs. año anterior, ventas año actual vs. año anterior, utilidad año actual vs. año anterior, evolución de la utilidad en el año actual según mes y sede, top de clientes según monto vendido, monto vendido por sede, top de productos según monto vendido, top de categorías según monto vendido, evolución de la utilidad en los últimos 3 años, top de zonas por número de clientes, monto de ventas por forma de pago, número de clientes por tipo de documento, top de productos más comprados por cliente seleccionado según monto medido, top de categorías más compradas por cliente seleccionado según monto medido, evolución de las ventas en los últimos 3 años, top de trabajadores según monto vendido, listado de productos sin rotación en el último mes, departamentos con clientes que consumen el producto seleccionado, evolución de las compras en los últimos 3 años, top de proveedores según monto comprado, monto comprado a proveedor por forma de pago, top de productos según monto comprado, monto de compra por sede.

Para realizar una predicción de las ventas de los productos, es necesario entrenar un algoritmo de *machine learning* (ML). Existen muchas maneras de hacer dicho entrenamiento y escribir un programa que pruebe los diversos algoritmos de predicción, evalúe cuál es el mejor para el caso presentado y permita integrar el modelo elegido a los reportes elaborados sería muy demandante de tiempo, esfuerzo y recursos. Por ello, se decidió utilizar la plataforma de computación en la nube (*cloud computing* en inglés) Azure Machine Learning, la cual provee

los recursos necesarios para el entrenamiento de un modelo de ML sin elevar los costos de mantenimiento de servidores de la empresa.

A continuación, se presenta siete de los más de treinta algoritmos probados en la plataforma para entrenar el modelo de predicción, ordenados de forma descendente según su coeficiente de determinación (R^2).

TABLA IV
ALGORITMOS PROBADOS Y SUS RESPECTIVOS PUNTAJES

Nombre del algoritmo	Coficiente de determinación (R^2)
Voting Ensemble	0.73097
Stack Ensemble	0.72111
Seasonal Naive	0.71594
Standard Scaler Wrapper, Random Forest	0.68181
Robust Scaler, Random Forest	0.67276
Standard Scaler Wrapper, Random Forest	0.67185
Min Mx Scaler, Gradient Boosting	0.66848

Como se puede observar en la tabla anterior, el mejor algoritmo con el puntaje R^2 es el aprendizaje por conjuntos por votación (*voting ensemble* en inglés), compuesto por los algoritmos Gradient Boosting, Random Forest y Seasonal Naive, con un puntaje de 0.73097. Por tanto, se determina que voting ensemble es el mejor algoritmo para predecir las ventas de los productos en el contexto de este proyecto.

Si bien la plataforma de computación en la nube utilizada permite probar el algoritmo elegido, esta forma de consumir el modelo no es la óptima, pues supondría un proceso tardado.

Para utilizar el modelo de ML, se realizó un despliegue a través de un servicio web (*web service* en inglés).

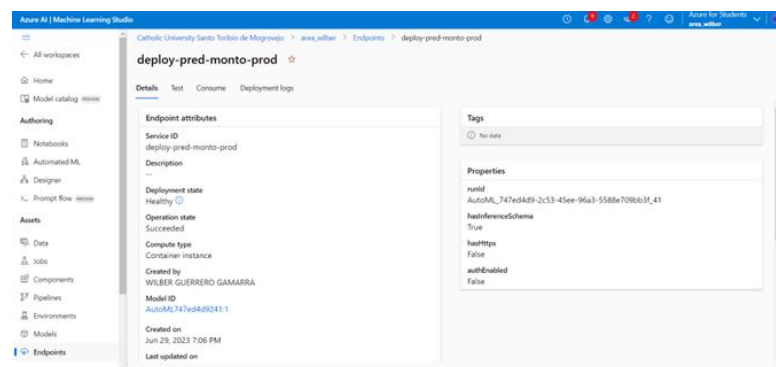


Fig. 7. Pantalla con información del modelo desplegado

Haber desplegado el modelo a través de un *web service* ofrece la posibilidad de utilizarlo desde diferentes sistemas, sin que estos necesariamente usen las mismas tecnologías.

En este proyecto, se consumió el servicio desde los reportes creados y se mostró los resultados en gráficos que contienen los datos de ventas de los últimos meses junto con la

predicción, permitiendo así complementar la información presentada ofreciendo un panorama tanto actual como futuro de las ventas.



Fig. 8. Captura del reporte predictivo de ventas

Para la evaluación de la calidad de la solución desarrollada se aplicará el estándar ISO/IEC 25010. Esta evaluación se llevará a cabo mediante cuestionarios diseñados específicamente para medir diversos aspectos de la solución. Los cuestionarios incluirán preguntas que requerirán respuestas basadas en la escala de Likert, lo que permitirá obtener una visión detallada de la percepción de los usuarios sobre la calidad de la solución en base a las características de adecuación funcional, usabilidad y mantenibilidad del modelo de calidad de producto del estándar ISO/IEC 25010.

La información recopilada a través de los cuestionarios mencionados fue procesada y analizada posteriormente para calcular el índice de satisfacción del usuario.

1. Adecuación Funcional:

Respuestas «totalmente de acuerdo»: 79 %

Respuestas «parcialmente de acuerdo»: 7 %

ISU para Adecuación Funcional = (79 % + 7 %) = 86 %

2. Usabilidad:

Respuestas «totalmente de acuerdo»: 72 %

Respuestas «parcialmente de acuerdo»: 28 %

ISU para Usabilidad = (72 % + 28 %) = 100 %

3. Mantenibilidad:

Respuestas «totalmente de acuerdo»: 58 %

Respuestas «parcialmente de acuerdo»: 34 %

ISU para Mantenibilidad = (58 % + 34 %) = 92 %

En esta investigación, se implementó una solución de inteligencia de negocios con la intención de mejorar la toma de decisiones organizacionales basadas en datos. La relación entre estos términos se determinó en [2], [16] y [17]. El contexto de esta investigación es el área

comercial de la empresa, abarcando los procesos de ventas, como en [15], y compras; pero, a diferencia de [15], no se implementó en el sector hotelero, sino en el sector agropecuario como sucede en [18].

Comparando esta investigación con los estudios de Apolaya [21] y Villanueva [22], observamos una similitud en el enfoque de utilizar la inteligencia de negocios para respaldar la toma de decisiones en los procesos de compra y venta de la organización. No obstante, se diferencia de estos enfoques en el sentido de que se ha aplicado en un tipo de empresa distinto, específicamente en una empresa dedicada a la comercialización de productos agropecuarios. Esto agrega un matiz significativo a la discusión, ya que implica la adaptación de las estrategias de inteligencia de negocios a un entorno empresarial específico y, por lo tanto, brinda una contribución única al campo.

Para abordar la predicción de las ventas, se optó por el algoritmo Voting Ensemble, tras realizar una evaluación exhaustiva de diversos algoritmos en la plataforma de *machine learning* de Azure. Esto se diferencia de la investigación previa citada en [14], que tiene un enfoque más estadístico, además de la elección de utilizar el método de mínimos cuadrados parciales (PLS-SEM). Esta elección se basó en la eficacia del algoritmo en el contexto específico de la empresa agropecuaria, lo que refleja la necesidad de seleccionar métodos y algoritmos apropiados para contextos empresariales particulares.

En cuanto a la implementación de los *key performance indicators* (KPI), se siguió una metodología que requiere una cuidadosa selección de estos indicadores y la provisión de descripciones detalladas para garantizar su comprensión por parte de los lectores. Aunque se ha tomado inspiración de la metodología de Piedra [19], que presenta los KPI en tablas como parte de las historias de usuario de la metodología Scrum, se ha optado por presentarlos en una lista para mayor claridad. Esta decisión destaca la necesidad de adaptar las mejores prácticas de inteligencia de negocios a las necesidades específicas de la empresa agropecuaria.

En lo que respecta al proceso ETL, se ha implementado un sistema utilizando el lenguaje Python para la extracción, transformación y carga de datos desde el sistema transaccional al *data mart*, utilizado para el análisis de la información. A diferencia de Portal y Quispe [20], quienes utilizaron una herramienta ETL preconstruida, nuestra investigación destacó la importancia de recibir recomendaciones y personalizar la implementación de ETL según las necesidades de la empresa agropecuaria.

Finalmente, cabe mencionar que se ha empleado una metodología híbrida basada en las metodologías de Ralph Kimball y CRISP-DM. Aunque Valderrama [23] también utilizó una metodología híbrida, nuestro enfoque difiere al considerar once fases en lugar de las diez

utilizadas por Valderrama. Esta elección se basó en un análisis conjunto con un experto en el campo y permitió una descripción más detallada del proceso de implementación, enfatizando la necesidad de una metodología flexible y adaptable en proyectos de inteligencia de negocios.

Conclusiones

1. Se logró implementar el 100 % de los indicadores clave de rendimiento en reportes interactivos construidos en Power BI, los cuales fueron integrados al sistema transaccional a través de la creación de un módulo. Con esto, se consiguió implementar reportes a medida para la empresa en la que se desarrolló esta investigación.
2. Se halló que el mejor algoritmo para realizar una predicción de ventas en determinado concepto es Voting Ensemble, al ser el algoritmo con coeficiente de determinación (R^2) de 0.73097, el más alto entre los que se probaron.
3. Se predijo las ventas de las ventas de los productos con base en la información histórica provista por la empresa donde se ejecutó la investigación con un porcentaje de precisión de un 73.09 %, asimismo, estas predicciones fueron integradas en los reportes implementados en el primer objetivo específico.
4. La aplicación del estándar ISO/IEC 25010 para verificar la calidad de la solución de inteligencia de negocios ha demostrado resultados excepcionales. Con un 86 % de los usuarios expresando estar total o parcialmente de acuerdo con relación a la adecuación funcional, un 100 % mostrando un alto grado de satisfacción en cuanto a la usabilidad, y un 92 % manifestando acuerdo total o parcial en cuanto a la mantenibilidad, estos datos cuantitativos respaldan de manera sólida la calidad de la solución en términos de estos aspectos cruciales. Estos hallazgos ofrecen un índice de satisfacción del usuario (ISU) significativamente alto, consolidando así la validez de la solución y proporcionando una base confiable para la toma de decisiones informadas en el entorno empresarial.

Recomendaciones

1. Se recomienda que en futuras investigaciones se pueda dar un mayor apoyo a las decisiones no programadas. Una forma de lograrlo sería implementando un sistema de alertas en tiempo real que notifique al usuario por su medio de comunicación preferido de las desviaciones significativas de los valores esperados.
2. Se recomienda que para futuras investigaciones se pueda considerar una mayor cantidad de data histórica, para así mejorar la precisión de las predicciones. Asimismo, se

recomienda que en entre estos datos se encuentren productos que tengan al menos tres años en el mercado para que sea más fácil encontrar patrones en las ventas de estos.

3. Se recomienda que en investigaciones futuras que implementen sistemas transaccionales, permitan que estos registren una mayor cantidad de datos en sus transacciones para realizar un análisis de mayor profundidad.
4. Se recomienda que en futuras investigaciones se automatice la prueba de los diferentes algoritmos de Azure Machine Learning para obtener un modelo de predicción actualizado cada cierto periodo de tiempo.
5. Se recomienda para futuras investigaciones se recomienda integrar data set de fuentes de datos abiertas para ampliar el análisis de los hechos identificados.

Referencias

- [1] R. Lapiedra Alcamí, C. Devece Carañana, y J. Guiral Herrando, *Introducción a la gestión de sistemas de información en la empresa*. Castelló de la Plana: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions, 2011., 2011. Accedido: 29 de marzo de 2022. [En línea]. Disponible en: <https://libros.metabiblioteca.org/handle/001/193>
- [2] A. R. Negro y R. Mesia, «The Business Intelligence and Its Influence on Decision Making», *J. Appl. Bus. Econ.*, vol. 22, n.º 2, Art. n.º 2, 2020.
- [3] S. S. Naidoo, «Business Intelligence Systems Input: Effects on Organizational Decision-Making», D.B.A., Capella University, United States -- Minnesota, 2019. Accedido: 30 de marzo de 2022. [En línea]. Disponible en: <https://www.proquest.com/docview/2191193907/abstract/793E795EF24F41FCPQ/1>
- [4] P. Zúmel, «Reporting Empresarial, el fundamento de la buena toma de decisiones», *GdR*, pp. 20-22, 10 de noviembre de 2008.
- [5] MicroStrategy, «The 2020 Global State of Enterprise Analytics», 2020. Accedido: 7 de abril de 2022. [En línea]. Disponible en: <https://www3.microstrategy.com/getmedia/db67a6c7-0bc5-41fa-82a9-bb14ec6868d6/2020-Global-State-of-Enterprise-Analytics.pdf>
- [6] U. Kulkarni y J. Robles-Flores, «Inteligencia de Negocios en Empresas Peruanas: Una Evaluación de las Dimensiones de Éxito», *RELCASI*, vol. 5, n.º 1, Art. n.º 1, ene. 2012, doi: 10.17705/1relc.00022.
- [7] P. A. Fernández Calle, V. A. G. Velásquez, y L. F. Uribe Cardona, «Problemática en la toma de decisiones financieras en las empresas familiares», *Activos*, vol. 16, n.º 31, Art. n.º 31, 2018, doi: <https://doi.org/10.15332/25005278/5317>.
- [8] L. E. Silva Solano, «Business Intelligence: un balance para su implementación», *InnovaG*, n.º 3, Art. n.º 3, mar. 2017.
- [9] M. Florencio Carrera, «Toma de decisiones y solución de problemas en administración», *gestiopolis*. Accedido: 19 de abril de 2022. [En línea]. Disponible en: <https://www.gestiopolis.com/toma-decisiones-solucion-problemas-administracion/>
- [10] Randstad, «Toma de decisiones en la empresa: ¿Cómo hacerlo? | Randstad». Accedido: 19 de abril de 2022. [En línea]. Disponible en: <https://www.randstad.es/tendencias360/el-proceso-de-toma-de-decisiones/>

- [11] P. Nutt, *Why Decisions Fail: Avoiding the Blunders and Traps That Lead to Debacles*, 1.^a ed. Berrett-Koehler Publishers, 2002.
- [12] P. Senge, *La quinta disciplina: el arte y a práctica de la organización abierta al aprendizaje*, 2.^a ed. Buenos Aires: Granica, 2010.
- [13] A. M. Requejo Paiva y O. S. Sanchez Pisfil, «Sistema de toma de decisiones en las pymes caso : empresa La Casa del Tornillo de la ciudad de Chiclayo», 2019, Accedido: 27 de abril de 2022. [En línea]. Disponible en: <http://tesis.usat.edu.pe/handle/20.500.12423/1780>
- [14] A. Popovič, B. Puklavec, y T. Oliveira, «Justifying business intelligence systems adoption in SMEs: Impact of systems use on firm performance», *Ind. Manag. Data Syst.*, vol. 119, n.º 1, Art. n.º 1, 2019, doi: <https://doi.org/10.1108/IMDS-02-2018-0085>.
- [15] M. da L. Guerreiro, M. B. Correia, y C. M. R. Sousa, «Análise De Consumo Numa Unidade Hoteleira No Algarve – Portugal, Com O Recurso De Ferramentas De Business Intelligence», *Tur. - Visão E Ação*, vol. 19, n.º 3, Art. n.º 3, 2017, doi: <https://doi.org/10.14210/rtva.v19n3.p513-539>.
- [16] J. Lawton, «Toward a Data-Driven Culture», *J. Gov. Financ. Manag.*, vol. 69, n.º 4, Art. n.º 4, 2021.
- [17] M. C. H. Bolaños *et al.*, «Inteligencia de Negocios para el Análisis de la Accidentalidad Vial en la Ciudad de Popayán», *Rev. Ibérica Sist. E Tecnol. Informação*, n.º E38, Art. n.º E38, 2020.
- [18] D. Koptilina, A. Bulgakov, y A. Aleshina, «“Big data” and the “Internet of things”, or what accelerates the development of the agro-industrial complex», *IOP Conf. Ser. Earth Environ. Sci.*, vol. 274, n.º 1, Art. n.º 1, 2019, doi: <https://doi.org/10.1088/1755-1315/274/1/012058>.
- [19] L. H. Piedra Rubio, «Implementación de un aplicativo de inteligencia de negocios (BI) para la mejora en la toma de decisiones en la Oficina Central de Admisión de una universidad peruana», Universidad Peruana Cayetano Heredia, 2023. Accedido: 9 de octubre de 2023. [En línea]. Disponible en: <https://repositorio.upch.edu.pe/handle/20.500.12866/13222>
- [20] H. M. Portal Uipan y D. L. Quispe Alcca, «Implementación de Business Intelligence para mejorar el proceso de toma de decisiones en el área de soluciones de la empresa telefónica del Perú S.A.A», Universidad Autónoma del Perú, 2018. Accedido: 3 de mayo de 2022. [En línea]. Disponible en: <http://repositorio.autonoma.edu.pe/handle/20.500.13067/515>
- [21] H. D. Apolaya Saravia, «Implementación de Inteligencia de negocios para mejorar la toma de decisiones gerenciales del área comercial, para un Centro de Prevención de Salud Ocupacional», Universidad San Ignacio de Loyola, 2019. Accedido: 9 de octubre de 2023. [En línea]. Disponible en: <https://repositorio.usil.edu.pe/entities/publication/5247d41d-dd51-48b5-bccf-eeb145531e97>
- [22] S. M. Villanueva Callirgos, «Implementación de una solución de inteligencia de negocios para apoyar la toma de decisiones en el proceso de compra y venta en una empresa farmacéutica en la ciudad de Chiclayo», Universidad Católica Santo Toribio de Mogrovejo, 2019. Accedido: 9 de octubre de 2023. [En línea]. Disponible en: <http://tesis.usat.edu.pe/handle/20.500.12423/1994>
- [23] O. E. Valderrama Mundaca, «Solución de inteligencia de negocios basada en técnicas de minería de datos, para apoyar la toma de decisiones, en la Gerencia Regional de Agricultura-Lambayeque», Universidad Católica Santo Toribio de Mogrovejo, 2022. Accedido: 9 de octubre de 2023. [En línea]. Disponible en: <http://tesis.usat.edu.pe/handle/20.500.12423/5047>

- [24] I. Chiavenato, *Comportamiento organizacional*, 2.^a ed. McGraw-Hill, 2009. Accedido: 12 de mayo de 2022. [En línea]. Disponible en: <http://148.202.167.116:8080/xmlui/handle/123456789/2873>
- [25] Northeastern University Graduate Programs, «Data-Driven Decision Making: A Primer for Beginners», Northeastern University Graduate Programs. Accedido: 28 de junio de 2022. [En línea]. Disponible en: <https://www.northeastern.edu/graduate/blog/data-driven-decision-making/>
- [26] Tableau, «A Guide To Data Driven Decision Making: What It Is, Its Importance, & How To Implement It», Tableau. Accedido: 28 de junio de 2022. [En línea]. Disponible en: <https://www.tableau.com/learn/articles/data-driven-decision-making>
- [27] B. Tabrizi, E. Lam, K. Girard, y V. Irvin, «Digital Transformation Is Not About Technology», *Harvard Business Review*, 13 de marzo de 2019. Accedido: 28 de junio de 2022. [En línea]. Disponible en: <https://hbr.org/2019/03/digital-transformation-is-not-about-technology>
- [28] D. J. Power, «A Brief History of Decision Support Systems», DSSResources.COM. Accedido: 12 de mayo de 2022. [En línea]. Disponible en: <http://dssresources.com/history/dsshistory.html>
- [29] D. J. Power, *Decision Support Systems: Concepts and Resources for Managers*, 1.^a ed. United States: Quorum Books, 2002. [En línea]. Disponible en: <https://scholarworks.uni.edu/cgi/viewcontent.cgi?article=1066&context=facbook>
- [30] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, y B. Becker, *The Data Warehouse Lifecycle Toolkit*, 2.^a ed. John Wiley & Sons, 2011.
- [31] R. Kimball y M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3.^a ed. Wiley, 2013.
- [32] Qlik, «What is a KPI Dashboard? 4 Key Examples and Best Practices», Qlik. Accedido: 23 de mayo de 2022. [En línea]. Disponible en: <https://www.qlik.com/us/dashboard-examples/kpi-dashboards>
- [33] W. H. Inmon, *Building the Warehouse*, 4.^a ed. Wiley, 2005.
- [34] O. Maimon y L. Rokach, *Data Mining and Knowledge Discovery Handbook*, 2.^a ed. Springer, 2010.
- [35] P. Chapman *et al.*, «CRISP-DM 1.0: Step-by-step data mining guide», 2000, Accedido: 9 de junio de 2022. [En línea]. Disponible en: <https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman-Clinton/54bad20bbc7938991bf34f86dde0babfd2d5a72>
- [36] SAS, «SAS Help Center: Introduction to SEMMA». Accedido: 15 de junio de 2022. [En línea]. Disponible en: <https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jnj8bbjmla2.htm>
- [37] OECD, «Manual de Frascati 2015». 2018. [En línea]. Disponible en: <https://www.oecd-ilibrary.org/content/publication/9789264310681-es>
- [38] C. A. Bernal Torres, *Metodología de la investigación*, 3.^a ed. Bogotá: Pearson, 2010.
- [39] M. Berndtsson, J. Hansson, B. Olsson, y B. Lundell, *Thesis Projects*. London: Springer, 2008. Accedido: 30 de mayo de 2022. [En línea]. Disponible en: <https://link.springer.com/book/10.1007/978-1-84800-009-4>

Anexos

Anexo N° 01. Constancia de aprobación del producto acreditable

CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE

Chiclayo, 20 de octubre del 2023

El gerente general de Sanidad Agropecuaria & Asociados S.A.C., por medio de la presente.

Hace constar:

Que **Wilber Guerrero Gamarra**, estudiante de la Universidad Católica Santo Toribio de Mogrovejo, identificado con DNI 72690243, ha desarrollado la tesis titulada «**Solución de inteligencia de negocios basada en indicadores de gestión y predicción para apoyar a la toma de decisiones del área comercial de una empresa en Chiclayo**» en Sanidad Agropecuaria & Asociados S.A.C.

Asimismo, se informa que **el producto entregado por el estudiante en mención ha sido aprobado** al haber sido entregado satisfactoriamente a quienes corresponde y haber cumplido con los requerimientos solicitados.

Esta constancia se emite a solicitud del estudiante para los fines que estime convenientes.

Atentamente,

SANIDAD AGROPECUARIA & ASOCIADOS S.A.C.

Oscar Iván Córdoba Nuñez del Arco
GERENTE GENERAL
DNI: 71540090

Anexo N° 02. Entrevista

Responsabilidades

1. Describa a la organización
2. ¿Cuáles son sus principales responsabilidades?

Objetivos de negocios e incidentes

1. ¿Cuáles son los objetivos de la organización? ¿Qué objetivos de negocio son prioritarios?
2. ¿Cómo logra saber si el negocio va bien? ¿Con qué mide el éxito?
3. ¿Con qué frecuencia monitorea los factores de éxito?
4. Por favor, comente sobre los incidentes que a los que se enfrenta el negocio. ¿Hay algo que impida lograr los objetivos? ¿Cómo impacta eso a la organización?
5. ¿Cómo identifica la presencia de problemas?
6. Por favor, comente sobre lo que se vende. ¿Cómo organiza los productos si la lista de estos es muy larga?
7. ¿Qué tan común es que se cambie la forma en que categorizan los productos?

Análisis de requerimientos

1. ¿Qué tipo de análisis de rutina realiza a la información? ¿Sobre qué información se realiza el análisis?
2. De haber un análisis de rutina, ¿hay alguna forma de que este sea mejorado?
3. ¿Qué hace con los resultados del análisis? ¿Qué decisiones toma con esta información?
4. ¿Hay decisiones que tome sobre la marcha (sin contar con un análisis de datos)?
5. ¿Con qué reportes dispone ahora? ¿Qué datos de cada reporte le resultan más importante?
6. ¿Con qué fuentes de datos cuenta actualmente para los reportes realizados o que cree útiles para la generación de nuevos reportes?
7. ¿Hay características o funciones en específico que le gustaría ver en la solución con la que contará?

Cierre

1. ¿En qué cree que beneficiaría al negocio si contara con un mejor acceso a la información? ¿Cree que esto tendría un impacto financiero positivo?
2. ¿Qué criterios considera que debe cumplir este proyecto para ser un proyecto exitoso?

Anexo N° 03. Cuestionario de evaluación de calidad de la solución desarrollada 1

Cuestionario de evaluación de calidad de la solución desarrollada

Lea las siguientes afirmaciones y marque la casilla correspondiente para mostrar su acuerdo o desacuerdo. Utilice la escala donde 1 significa «Totalmente en desacuerdo»; 2, «Parcialmente en desacuerdo»; 3, «Ni en acuerdo ni en desacuerdo»; 4, «Parcialmente de acuerdo» y 5, «Totalmente de acuerdo».

	1	2	3	4	5
1. El software proporciona todas las funciones necesarias para satisfacer los requisitos del usuario					
2. El software proporciona todas las funciones necesarias para satisfacer el alcance del proyecto					
3. El software realiza todas sus funciones de manera precisa y exacta, proporcionando resultados correctos y coherentes					
4. El software contiene errores significativos que afectan la precisión y coherencia de los resultados					
5. El software implementa los KPI indicados al desarrollador					
6. El software es altamente adecuado para el propósito previsto y satisface plenamente las necesidades					
7. El software satisface las expectativas del usuario en su contexto de uso					
8. El software es adecuado para el propósito previsto, pero hay algunos aspectos donde no satisfacen completamente las necesidades del usuario					
9. Es sencillo comprender cuáles son las funciones y características de la solución de inteligencia de negocios					
10. Es fácil acceder a los reportes BI desde la interfaz del sistema transaccional					
11. Los reportes BI cuentan con un diseño agradable para el usuario					
12. Los reportes BI tienen una interfaz que facilita la navegación					
13. La solución BI permite encontrar rápidamente la información que necesita					
14. Es fácil filtrar los informes para obtener información específica					
15. Es sencillo aprender a usar el producto (reportes BI)					

Anexo N° 04. Cuestionario de evaluación de calidad de la solución desarrollada 2

Cuestionario de evaluación de calidad de la solución desarrollada

Lea las siguientes afirmaciones y marque la casilla correspondiente para mostrar su acuerdo o desacuerdo. Utilice la escala donde 1 significa «Totalmente en desacuerdo»; 2, «Parcialmente en desacuerdo»; 3, «Ni en acuerdo ni en desacuerdo»; 4, «Parcialmente de acuerdo» y 5, «Totalmente de acuerdo».

	1	2	3	4	5
1. La configuración inicial y la operación del sistema ETL son sencillas y eficientes, a pesar de no tener una interfaz de usuario					
2. Las funciones y acciones dentro del sistema ETL son fáciles de entender y utilizar					
3. Es fácil realizar actualizaciones y mejoras en el sistema ETL programado en Python					
4. Las documentaciones y comentarios en el código del sistema ETL facilitan la comprensión de su funcionamiento					
5. Es sencillo identificar y solucionar problemas o errores en el sistema ETL					
6. El sistema ETL se puede adaptar fácilmente a nuevas fuentes de datos o cambios en los requisitos de negocio					
7. El sistema está modularizado de manera que los mantenedores del software consideran entendible y mantenible					
8. Los reportes BI tienen una estructura adecuada y que facilita su comprensión					
9. Las funciones implementadas dentro de los reportes BI son fáciles de entender					
10. Las funciones implementadas dentro de los reportes BI son fáciles de actualizar					
11. Es sencillo identificar y solucionar problemas o errores en los reportes BI					
12. Los reportes BI se puede adaptar fácilmente a nuevas fuentes de datos o cambios en los requisitos de negocio					

Anexo N° 05. Pruebas de caja negra realizadas al sistema ETL

PCN-01	Extracción de datos
Descripción	Comprobar que el subsistema ETL logra extraer datos de una fuente de datos
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Observar si se muestra el mensaje de confirmación de la extracción
Resultado esperado	Se muestra el mensaje de confirmación de la extracción
Resultado obtenido	Se mostró el mensaje de confirmación de la extracción
Resultado de la prueba	Correcto
<hr/>	
PCN-02	Prueba de carga al <i>data mart</i>
Descripción	Comprobar que el subsistema ETL logra cargar los datos en el <i>data mart</i> .
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - Haber completado exitosamente las fases de carga y transformación
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Observar si se muestra el mensaje de confirmación de la fase de carga 3. Revisar si se encuentran los datos cargados en el <i>data mart</i>
Resultado esperado	Los datos aparecen en el <i>data mart</i>
Resultado obtenido	Los datos aparecieron en el <i>data mart</i>
Resultado de la prueba	Correcto
<hr/>	
PCN-03	Prueba de transformación de datos
Descripción	Comprobar que el subsistema ETL logra cargar datos ya transformados al <i>data mart</i>
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - Haber realizado la carga de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Observar si se muestra el mensaje de confirmación de la carga de datos 3. Verificar si los datos cargados han sido transformados
Resultado esperado	Los datos cargados cumplen con las transformaciones
Resultado obtenido	Los datos cargados cumplieron con las transformaciones
Resultado de la prueba	Correcto
<hr/>	
PCN-04	Prueba de no realización de carga
Descripción	Comprobar que el subsistema ETL logra detectar cuando no hay datos nuevos para cargar al <i>data mart</i>

Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - No hay datos registrados en una fecha posterior a la de la última carga
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Observar si se muestra el mensaje indicando que no hay datos por cargar
Resultado esperado	Se muestra el mensaje indicando que no hay datos por cargar
Resultado obtenido	Se mostró el mensaje indicando que no hay datos por cargar
Resultado de la prueba	Correcto

PCN-05	Prueba de carga progresiva
Descripción	Comprobar que el subsistema ETL logra realizar una carga progresiva
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - Tener datos registrados en una fecha posterior a la última fecha de carga
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Esperar el mensaje de confirmación de carga 3. Verificar en la base de datos que la fecha de carga de las nuevas filas sea la correcta
Resultado esperado	Se cargan solo los datos nuevos
Resultado obtenido	Se cargaron solo los datos nuevos
Resultado de la prueba	Correcto

Anexo N° 06. Pruebas de caja blanca realizadas al sistema ETL

PCB-01	Prueba de optimización de carga
Descripción	Comprobar que el subsistema ETL logra cargar los datos de manera óptima al <i>data mart</i>
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Observar si el tiempo de carga es razonable para el volumen de información
Resultado esperado	Se carga los datos en un tiempo óptimo
Resultado obtenido	Se usó la función COPY de Postgresql para cargar los datos que están en memoria al <i>data mart</i> en lugar de usar INSERT
Resultado de la prueba	Correcto
PCB-02	Acceso a columna por nombre en las filas de un resultado de una consulta
Descripción	Comprobar que el subsistema ETL permite acceder una columna usando su nombre para facilitar el tratamiento de los datos
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa en modo debug 2. Ejecutar hasta el punto en que se guarden los datos extraídos en una variable 3. Verificar que el tipo de dato permita acceder a la columna por su nombre
Resultado esperado	El tipo de dato permite acceder a la columna por su nombre
Resultado obtenido	El tipo de dato es diccionario y cada llave es el nombre de la columna
Resultado de la prueba	Correcto
PCB-03	Extracción de datos necesarios
Descripción	Comprobar que el subsistema ETL extrae solo los datos necesarios del sistema transaccional
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - No hay datos nuevos registrados luego de la última fecha de carga
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa en modo debug 2. Ejecutar hasta el punto en que se guardan los datos extraídos en una variable 3. Observar que no haya datos que no correspondan a la carga a menos que estos sean requeridos por otra parte del programa
Resultado esperado	Se extraen solo los datos necesarios
Resultado obtenido	Se extrajeron solo los datos necesarios
Resultado de la prueba	Correcto
PCB-04	Creación de fila para foráneas incorrectas en las dimensiones donde sea necesario

Descripción	Comprobar que el subsistema ETL realiza la inserción de un id al que se asignará la foránea que corresponda a esa la relación con esa tabla de ser necesario
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 4. Correr el programa 5. Correr el programa 6. Observar si se realiza las inserciones
Resultado esperado	Se realiza las inserciones en la base de datos
Resultado obtenido	Se realizaron las inserciones en la base de datos
Resultado de la prueba	Correcto

PCB-05	Prueba de deduplicación de filas
Descripción	Comprobar que el subsistema ETL logra quitar las filas duplicadas en caso la consulta configurada las genere
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa 2. Esperar a que complete su ejecución 3. Comprobar que no haya filas duplicadas en la base de datos
Resultado esperado	No se encuentran filas duplicadas
Resultado obtenido	No se encontraron filas duplicadas
Resultado de la prueba	Correcto

PCB-06	Prueba de excepción de fase de transformación
Descripción	Comprobar que el subsistema ETL no realiza la transformación de datos si no hay nuevos datos a insertar
Prerrequisitos	<ul style="list-style-type: none"> - Tener instalado Python - Haber instalado las dependencias del proyecto - Haber configurado las consultas a realizar en la fuente de datos - No tener datos registrados luego de la última fecha de carga
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Correr el programa en modo debug 2. Esperar a que se llegue al punto en donde se realiza la transformación 3. Revisar que no existan tablas sin datos nuevos en la cola para la transformación
Resultado esperado	No se transforman las tablas sin datos nuevos
Resultado obtenido	No se transformaron las tablas sin datos nuevos
Resultado de la prueba	Correcto

Anexo N° 07. Pruebas de caja negra realizadas a la aplicación de inteligencia de negocios

PCN-01	Acceder a los reportes desde el dashboard
Descripción	Comprobar que el usuario puede acceder a los demás reportes desde el dashboard de la aplicación de inteligencia de negocios.
Prerrequisitos	<ul style="list-style-type: none"> - Tener acceso al reporte embebido de Power BI - Haber iniciado sesión con una cuenta que tenga acceso a los reportes
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Acceder al dashboard 2. Dar clic en cualquier botón de la parte izquierda del dashboard
Resultado esperado	Acceder al reporte elegido
Resultado obtenido	Se accedió al reporte elegido
Resultado de la prueba	Correcto

PCN-02	Regresar al reporte anterior
Descripción	Comprobar que el usuario puede retroceder una vez accedido un reporte
Prerrequisitos	<ul style="list-style-type: none"> - Tener acceso al reporte embebido de Power BI - Haber iniciado sesión con una cuenta que tenga acceso a los reportes
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Acceder al dashboard 2. Dar clic en cualquier botón de la parte izquierda del dashboard 3. Dar clic en el botón de retroceso
Resultado esperado	Acceder al reporte anterior nuevamente
Resultado obtenido	Se accedió al reporte correspondiente
Resultado de la prueba	Correcto

PCN-03	Realizar comparación de dos periodos de tiempo
Descripción	Comprobar que el usuario puede comparar los datos de dos periodos de tiempo. En este caso, se probará que el usuario puede filtrar las ventas por cliente de este mes con el mes anterior.
Prerrequisitos	<ul style="list-style-type: none"> - Tener acceso al reporte embebido de Power BI - Haber iniciado sesión con una cuenta que tenga acceso a los reportes
Datos de entrada	N/A
Secuencia	<ol style="list-style-type: none"> 1. Acceder al dashboard 2. Dar clic en cualquier botón «Ventas por cliente» 3. Ubicar el filtro de periodo 4. Seleccionar la opción «Este mes (vs. mes pasado completo)»
Resultado esperado	Ver un cambio en los datos mostrados y observar los datos de ambos periodos en los gráficos que soportes ese tipo de comparaciones
Resultado obtenido	Se observó un cambio en la información
Resultado de la prueba	Correcto

PCN-04	Obtener información sobre un cliente específico
Descripción	Comprobar que el usuario puede acceder a los informes de detalles
Prerrequisitos	<ul style="list-style-type: none"> - Tener acceso al reporte embebido de Power BI - Haber iniciado sesión con una cuenta que tenga acceso a los reportes

Datos de entrada	Nombre de cliente
Secuencia	<ol style="list-style-type: none"> 1. Acceder al dashboard 2. Dar clic en el botón «Ventas por cliente» 3. En un listado de clientes, dar clic al nombre de un cliente 4. Seleccionar «Obtener detalles» 5. Hacer clic en «Detalle de cliente»
Resultado esperado	Acceder al reporte detallado de cliente y observar la información del cliente seleccionado
Resultado obtenido	Se accedió al reporte y se visualizó la información del cliente
Resultado de la prueba	Correcto

PCN-05	Obtener los productos más vendidos por un trabajador
Descripción	Comprobar que el usuario puede usar filtros que tienen la capacidad de búsqueda
Prerrequisitos	<ul style="list-style-type: none"> - Tener acceso al reporte embebido de Power BI - Haber iniciado sesión con una cuenta que tenga acceso a los reportes
Datos de entrada	Nombre de vendedor
Secuencia	<ol style="list-style-type: none"> 1. Acceder al dashboard 2. Dar clic en el botón «Ventas por trabajador» 3. Ubicar el filtro de trabajador 4. Dar clic en la lista desplegable 5. Escribir el nombre de un trabajador 6. Seleccionar el nombre del trabajador en la lista
Resultado esperado	Se filtra la información del reporte con la información que corresponde a ese trabajador
Resultado obtenido	Se observó el cambio en la información del reporte
Resultado de la prueba	Correcto

Anexo N° 08. Prueba de conciliación de datos realizada a la aplicación de inteligencia de negocios



> Soluciones Empresariales

Chiclayo, 5 de julio del 2023

CONSTANCIA DE VALIDACIÓN DE RESULTADOS DE APLICACIÓN DE INTELIGENCIA DE NEGOCIOS

Yo, César Wilder Vallejos Dávila, en calidad de Gerente de Cesly Soft & Business S.R.L., por medio de la presente constancia, certifico que se ha llevado a cabo una validación de los resultados obtenidos mediante la aplicación de la inteligencia de negocios desarrollada por Wilber Guerrero Gamarra utilizando la plataforma Power BI.

La mencionada aplicación de inteligencia de negocios ha sido implementada en nuestro sistema transaccional, el cual es utilizado por Sanidad Agropecuaria & Asociados S.A.C., en el contexto de la tesis de Wilber Guerrero Gamarra titulada "Solución de inteligencia de negocios basada en indicadores de gestión y predicción para apoyar a la toma de decisiones del área comercial de una empresa en Chiclayo". Durante el proceso de validación, se han realizado comparaciones entre los reportes generados por la aplicación y los reportes del sistema transaccional utilizado por Sanidad Agropecuaria & Asociados S.A.C., desarrollado por Cesly Soft & Business S.R.L.

Debido a las políticas de seguridad y confidencialidad establecidas por nuestra empresa, así como las de Sanidad Agropecuaria & Asociados S.A.C., no es posible proporcionar copias ni mostrar directamente los reportes generados por el sistema transaccional. Sin embargo, puedo asegurar que los resultados obtenidos mediante la aplicación de inteligencia de negocios coinciden de manera precisa con los reportes internos generados por nuestro sistema transaccional.

Por lo tanto, basándonos en los resultados obtenidos y la validación realizada por el equipo de Cesly Soft & Business S.R.L., confirmamos que los reportes generados por la aplicación desarrollada por Wilber Guerrero Gamarra son consistentes y coinciden con los reportes del sistema transaccional utilizado por Sanidad Agropecuaria & Asociados S.A.C.

La presente constancia se emite a solicitud de Wilber Guerrero Gamarra para respaldar los resultados obtenidos en su tesis y para dar testimonio de la validación realizada por Cesly Soft & Business S.R.L. Quedo a disposición para cualquier consulta o aclaración adicional que se requiera.

Atentamente,

CESLY SOFT & BUSINESS S.R.L.
César Vallejos Dávila
GERENTE