Optimización de operaciones en Servientrega: mejorando la eficiencia a través del gobierno de datos

Astrid Alexandra Sánchez Tatis, Carlos Leonardo Salamanca Beltrán, Diana Patricia Carvajal Salamanca, Jorge Leonardo Pinzón Caucha y Leidy Natalia Urueña Solano

Especialización Big Data, Fundación Universitaria Compensar

Wilson Hernando Soto Urrea

27 de Noviembre de 2023

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-Compartir Igual 4.0 Internacional.

Optimización de operaciones en Servientrega: mejorando la eficiencia a través del gobierno de datos

Astrid Alexandra Sánchez Tatis, Carlos Leonardo Salamanca Beltrán

Diana Patricia Carvajal Salamanca, Jorge Leonardo Pinzón Caucha

Leidy Natalia Urueña Solano

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialización en Big Data

Director (a):

Ingeniero Wilson Hernando Soto Urrea

Fundación Universitaria Unipanamericana – Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá, Colombia

2023

Resumen

La investigación se centra en abordar los problemas logísticos en Servientrega, para mejorar la eficiencia operativa y la satisfacción del cliente. Se identificó la falta de un gobierno de datos eficiente como un desafío crítico, afectando la calidad y consolidación de la información. La propuesta busca implementar estrategias como Big Data, la creación de catálogos de datos y la formación de un equipo de gobierno de datos. La metodología incluye fases cualitativas y cuantitativas, así como procesos de minería de datos (KDD, SEMMA, CRISP-DM). Los resultados muestran que las soluciones propuestas podrían reducir los retrasos en la entrega en un 10% y aumentar la satisfacción del cliente en un 5%. La implementación de estos hallazgos se presenta como una oportunidad para fortalecer la posición competitiva de Servientrega.

Palabras clave: KDD, SEMMA, CRISP-DM

Abstract

The research focuses on addressing logistical problems at Servientrega, to improve operational efficiency and customer satisfaction. The lack of efficient data governance was identified as a critical challenge, impacting the quality and consolidation of information, has been identified The proposal aims to implement strategies such as Big Data, creating data catalogs and the formation of a data governance team. The methodology includes both qualitative and quantitative phases, as well as data mining processes (KDD, SEMMA, CRISP-DM). The results show that the proposed solutions could reduce delivery delays by 10% and increase customer satisfaction by 5%. The implementation of these findings is presented as an opportunity to strengthen Servientrega's competitive position.

Keywords: KDD, SEMMA, CRISP-DM

Contenido

PÁG.

RESUMEN III

ABSTRACT IV

INTRODUCCIÓN 1

ANTECEDENTES Y JUSTIFICACIÓN 2
Antecedentes 2
Justificación 2
FORMULACIÓN DEL PROBLEMA 3
OBJETIVO GENERAL 3
OBJETIVOS ESPECÍFICOS 3
ALCANCES Y LIMITACIONES 4
Alcance 4
Limitaciones 4

1. MARCO TEÓRICO 5

1.1 METODOLOGÍAS 5
1.1.1 KDD 5
1.1.2 CRISP-DM 6
1.1.3 SEMMA 6

2. DISEÑO METODOLÓGICO 7

2.1 RECOLECCIÓN DE DATOS 7
2.2 ANÁLISIS DE DATOS 8
2.2.1 Modelo CRISP-DM 8
2.2.2 Modelo KDD 8
2.2.3 Modelos SEMMA 9
2.3 PROPUESTA DE SOLUCIONES 9
2.4 EVALUACIÓN DE SOLUCIONES 9
2.5 RECURSOS 9
2.6 POLÍTICAS 10

***3.* RESULTADOS ESPERADOS** 10

4. CONCLUSIONES Y RECOMENDACIONES 11

4.1 CONCLUSIONES 11
4.2 RECOMENDACIONES 11

BIBLIOGRAFÍA 12

Introducción

En el panorama empresarial actual, la optimización de operaciones se ha convertido en una piedra angular para la competitividad y sostenibilidad de las organizaciones. En este contexto, las empresas logísticas enfrentan desafíos significativos, especialmente aquellas con operaciones complejas y extensas como Servientrega. La entrega eficiente de mercancías en el tiempo previsto y la gestión efectiva de grandes volúmenes de datos son elementos cruciales para el éxito en este sector altamente competitivo.

Servientrega, como actor destacado en la cadena logística, se encuentra inmerso en un entorno donde la velocidad, precisión y confiabilidad son imperativas. La eficiencia operativa y la satisfacción del cliente están intrínsecamente ligadas a la capacidad de la empresa para superar los obstáculos logísticos y mantener una calidad de servicio excepcional. En este sentido, la presente investigación se adentra en el corazón de estos desafíos, destacando la carencia de un gobierno de datos eficiente como un factor crítico que impacta directamente la calidad y consolidación de la información logística en Servientrega.

La magnitud de la información manejada por Servientrega, combinada con la falta de un enfoque estructurado para gestionar y aprovechar estos datos, ha dado lugar a la identificación de datos maestros problemáticos, inconsistencias y registros obsoletos. Esta problemática no solo impacta la eficiencia operativa sino también la toma de decisiones estratégicas, afectando la percepción del cliente y la posición competitiva de la empresa.

Antecedentes y Justificación

Antecedentes

En el ámbito logístico, la gestión eficiente de la cadena de suministro y la entrega puntual de mercancías son imperativas para el éxito de las empresas. Servientrega, como protagonista en este escenario, se enfrenta a desafíos operativos considerables que se han estudiado en investigaciones previas.

Investigaciones anteriores han destacado la importancia del gobierno de datos en entornos logísticos complejos. La falta de una estructura organizativa sólida para gestionar la información ha sido identificada como una fuente principal de ineficiencias en la cadena de suministro. Estudios en empresas similares han revelado que la implementación de estrategias de Big Data y Gobernanza de Datos puede ser fundamental para abordar problemas relacionados con la calidad de la información, la toma de decisiones y la satisfacción del cliente.

Justificación

La carencia de un gobierno de datos eficiente en Servientrega se ha convertido en un desafío crítico que impacta la calidad de la información, la eficiencia operativa y la satisfacción del cliente. La justificación para esta investigación radica en múltiples factores estratégicos y operativos:

Mejora de la Eficiencia Operativa: La implementación de estrategias de Big Data y Gobernanza de Datos busca centralizar y mejorar la calidad de la información logística, permitiendo análisis más precisos de las guías de envíos y optimizando tiempos y rutas de entrega.

Toma de Decisiones Informada: La ambigüedad en la interpretación de datos y la falta de identificación de datos maestros clave afectan la toma de decisiones estratégicas. La creación de catálogos de datos claros y definidos proporcionará una base sólida para decisiones informadas y estratégicas.

Protección de la Imagen de la Empresa: Los retrasos en la entrega de envíos no solo impactan la eficiencia de la cadena de suministro, sino que también afectan directamente la percepción del cliente sobre la calidad del servicio de Servientrega. Resolver estas deficiencias no solo beneficia la operatividad de la empresa, sino también su imagen y posición competitiva.

Viabilidad y Sostenibilidad: La viabilidad de la investigación se basa en la adaptación de soluciones escalables y en la colaboración estrecha con Servientrega. La sostenibilidad de las soluciones propuestas se fundamenta en la exploración de deficiencias actuales y la identificación de oportunidades de mejora.

Formulación del Problema

Servientrega enfrenta dificultades para rastrear y gestionar el ciclo de vida completo de sus datos, desde la creación hasta la eliminación. Esto conduce a la acumulación de datos obsoletos, aumentando los costos de almacenamiento y afectando la eficiencia operativa. Por ello, la compañía no puede capitalizar el valor de sus datos para iniciativas estratégicas y de innovación por la falta de una estructura que fomente la reutilización y la exploración de datos segura y eficiente. En razón a lo expuesto anteriormente, el desarrollo del siguiente trabajo busca dar respuesta a la siguiente pregunta:

¿Cómo la carencia de una estructura, que promueve la reutilización, compartición y exploración segura y eficiente de datos, impide que la empresa aproveche plenamente el valor de sus datos en iniciativas estratégicas e innovación?

Objetivo General

Identificar de manera internas las causas fundamentales de estos retrasos, analizar las posibles soluciones, proponer medidas para mejorar la eficiencia logística y la satisfacción del cliente, como también contribuir al desarrollo de soluciones efectivas que fortalezcan la operación de Servientrega en un entorno empresarial competitivo.

Objetivos Específicos

- Identificar las causas fundamentales de los retrasos en la entrega de envíos a través de métodos cualitativos.
- Evaluar el impacto de las soluciones propuestas en la eficiencia logística y la satisfacción del cliente a través de métodos cuantitativos.
- Proponer soluciones para mejorar la eficiencia logística y la satisfacción del cliente.

Alcances y Limitaciones

Alcance

La investigación se centra en abordar la problemática de eficiencia logística en la empresa Servientrega, específicamente en la optimización de operaciones mediante la implementación de estrategias avanzadas de gestión de datos. El alcance se extiende desde la identificación de causas fundamentales de retrasos en la entrega hasta la proposición y evaluación de soluciones basadas en modelos predictivos. A continuación, se detallan los aspectos comprendidos en el alcance:

Selección Representativa de Datos: Recopilación de datos de Servientrega mediante encuestas a directivos y empleados, así como análisis de documentos internos, garantizando la representatividad de la realidad operativa.

Análisis Cualitativo y Cuantitativo: Integración de métodos cualitativos y cuantitativos para obtener una comprensión holística de los desafíos logísticos. Utilización de herramientas y técnicas de minería de datos bajo las metodologías KDD, CRISP-DM y SEMMA.

Modelado Predictivo: Construcción de modelos predictivos utilizando técnicas avanzadas como regresión lineal y árboles de decisión, con el fin de evaluar el impacto de soluciones propuestas en la eficiencia logística y satisfacción del cliente.

Implementación Práctica: Aplicación práctica de las soluciones propuestas en un entorno operativo simulado, permitiendo una evaluación más completa de su viabilidad y eficacia.

Limitaciones

La investigación está sujeta a ciertas limitaciones que podrían influir en la interpretación y aplicabilidad de los resultados:

Disponibilidad y Calidad de Datos: La calidad y cantidad de datos disponibles de Servientrega podrían afectar la profundidad del análisis. La investigación depende de la cooperación y acceso a información relevante por parte de la empresa.

Complejidad de Modelos Predictivos: La complejidad de los modelos predictivos puede verse afectada por la disponibilidad de datos históricos y la capacidad de procesamiento, lo que podría impactar la precisión de las predicciones.

Generalización de Resultados: Los resultados se aplicarán específicamente al contexto de Servientrega en Colombia y podrían no ser generalizables a otras empresas o regiones.

Cambios en el Entorno Operativo: La dinámica del entorno empresarial y logístico puede cambiar durante la investigación, lo que podría influir en la aplicabilidad y validez de las soluciones propuestas.

1.Marco teórico

1.1 Metodologías

Algunos modelos profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos como CRISP-DM, mientras que otros proveen sólo una guía general del trabajo a realizar en cada fase, como el proceso KDD o SEMMA. (Rosario, 2011)

1.1.1 KDD

Es un proceso exploratorio que se utiliza para descubrir conocimiento a partir de datos. La metodología se divide en las siguientes fases: - Exploración: de datos para identificar patrones, tendencias y relaciones. Se utilizan técnicas estadísticas y gráficas para visualizar los datos y detectar anomalías; - Preparación: para análisis, donde se eliminan los datos erróneos o incompletos, y se transforman los datos para que sean compatibles con la minería de datos; - Modelado: creación de modelos para representar las tendencias y relaciones identificados. Se utilizan algoritmos para crear modelos que puedan predecir resultados, agrupar datos o identificar correlaciones; -Evaluación: se evalúan los modelos para determinar su precisión y eficacia. Se utilizan técnicas estadísticas para medir el rendimiento de los modelos; - Implementación: de los

modelos en un entorno productivo, se integran los modelos en los sistemas existentes para que puedan utilizarse para tomar decisiones o generación de informes.

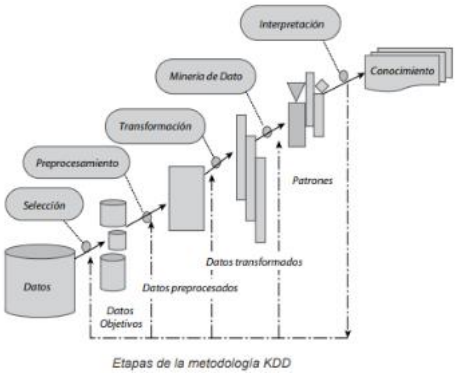


Ilustración 1 Fases KDD

1.1.2 CRISP-DM

CRISP-DM (Cross Industry Standard Process for Data Mining), La metodología se adapta en la comprensión del negocio, la preparación de los datos, el modelado de los datos, la evaluación de los modelos y la implementación de las soluciones. La metodología consta de seis fases las cuales se visualizan en Ilustración 2, al ser una metodología orientada a procesos que siguen una trayectoria bastante secuencial, facilita el entendimiento en su planificación y ejecución.



Ilustración 2 Fases2 CRISP-DM

1.1.3 SEMMA

La metodología SEMMA (Sample, Explore, Modify, Model, Assess), se centra en la selección, exploración, modificación, modelado y evaluación de datos para identificar

patrones y tendencias que pueden ayudar a las empresas a tomar decisiones informadas. Las técnicas de modelado de datos se pueden clasificar en dos grupos:

- Métodos estadísticos tradicionales: Se basan en la teoría de la probabilidad y la estadística para identificar patrones y relaciones en los datos, como lo son el análisis discriminante, los métodos de agrupamiento y el análisis de regresión.

– Técnicas basadas en datos: Establecen un aprendizaje automático para identificar patrones y relaciones en los datos, podemos encontrar las redes neuronales, las técnicas adaptativas, los árboles de decisión, las reglas de asociación y la computación evolutiva.

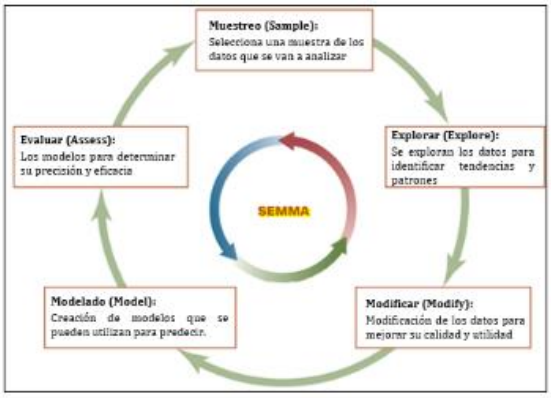


Ilustración 33 Fases SEMMA

2.Diseño Metodológico

La investigación sobre los retrasos en la entrega de envíos en Servientrega se llevará a cabo durante el año 2024 y se centrará en la operación de la empresa en Colombia. Los actores involucrados en la exploración son Servientrega, clientes y empleados. Algunos factores que podrían influir son: Naturaleza de los envíos que se entregan, ubicación de los remitentes y destinatarios, las condiciones climáticas o del tráfico, los procesos o procedimientos de Servientrega.

2.1 Recolección de datos

La recolección de datos se realizará mediante una combinación de métodos cualitativos y cuantitativos, los métodos cualitativos se utilizarán para comprender las causas de los retrasos desde la perspectiva de los empleados de Servientrega. Los métodos

cuantitativos se utilizarán para evaluar el impacto de los retrasos en la eficiencia logística y la satisfacción del cliente.

- Datos internos de Servientrega, como información sobre envíos, rutas, y transportistas.
- Datos de fuentes externas, como información sobre condiciones climáticas, tráfico, y eventos especiales.

2.2 Análisis de datos

Se analizan los datos recopilados para identificar las causas fundamentales de los retrasos. Cuando el equipo de ingenieros de datos ha estructurado la información, los científicos de datos seleccionarán las variables que, por la lógica de negocio, y otras consideraciones, pueden resolver el problema planteado.

2.2.1 Modelo CRISP-DM

La investigación sobre este problema es compleja y requiere la recopilación de datos de diversas fuentes. CRISP-DM proporciona un marco estructurado para investigar y garantizar el tratamiento de los aspectos del problema y flexibilidad para adaptar la metodología a las necesidades específicas del proyecto. Incluye una fase de evaluación de los modelos, que es importante para garantizar que los mismos propuestos sean precisos y eficaces, ya que los modelos se utilizarán para identificar las causas fundamentales del problema y proponer soluciones.

2.2.2 Modelo KDD

Es una metodología exploratoria que se centra en la identificación de patrones y tendencias en los datos. Es flexible, se puede adaptar a diferentes tipos de problemas, pero puede ser lenta y costosa, igualmente puede ser difícil de aplicar a problemas complejos o con datos de baja calidad. En el caso de la investigación la metodología KDD podría ser una buena opción si se quiere explorar los datos para identificar patrones y tendencias que no se habían identificado previamente.

2.2.3 Modelos SEMMA

Es una metodología iterativa e interactiva que se centra en la selección, exploración, modificación, modelado y evaluación de los datos. Es una metodología relativamente sencilla de entender y aplicar, y puede ser una buena opción para problemas complejos o con datos de baja calidad, especialmente si se necesita una solución rápida y precisa. Sin embargo, no se cumple con las condiciones requeridas para aplicar el modelo, toda vez que los datos a procesar son grandes y no están previamente estructurados.

2.3 Propuesta de soluciones

Para mejorar la eficiencia logística y la satisfacción del cliente, las soluciones se basarán en los resultados del análisis de datos que se proyectarán para tomar decisiones en la organización.

2.4 Evaluación de soluciones

Se evalúan los modelos para determinar su precisión y eficacia. Se utilizan métricas como la curva ROC, la especificidad y la sensibilidad para evaluar el rendimiento de los modelos. El algoritmo que mejor métrica tenga se seleccionará para su implementación, el diseño metodológico propuesto es flexible y se puede adaptar a diferentes tipos de problemas de minería de datos.

2.5 Recursos

Los recursos necesarios para la investigación incluyen:

- Personal: Se requerirá un equipo de investigadores con experiencia en logística, análisis de datos, y gestión de proyectos.
- Financiamiento: Se requerirá un presupuesto para cubrir los costos de la investigación, incluyendo la recopilación de datos, el análisis de datos, y la evaluación de soluciones.

2.6 Políticas

Son aquellas reglas presentes al realizar cualquier proceso dentro de cualquiera de los aspectos de gobierno de datos, las mismas rigen aspectos fundamentales de arquitectura y seguridad, persiguiendo principios de buenas prácticas.

3. Resultados esperados

En el contexto de esta propuesta, los resultados esperados se enfocan en la preparación y planificación para abordar los desafíos presentes en la empresa Servientrega, que enfrenta dificultades logísticas y de satisfacción debido a la falta de un sistema de Gobierno de datos. Se anticipa que la aplicación de las metodologías de minería de datos propuestas no solo conduzca a su implementación, sino que también resulte en una mejora significativa en el almacenamiento de la información, permitiendo así avanzar en la implementación de Big Data en la compañía. A través de la recolección y análisis preliminar de datos, se espera obtener una comprensión más profunda de los patrones y factores que podrían contribuir a los problemas logísticos. En este punto, los resultados esperados se centran en identificar áreas clave para la mejora y formular una estrategia sólida para la implementación de modelos en etapas posteriores.

4. Conclusiones y recomendaciones

4.1 Conclusiones

El problema logístico en Servientrega, originado por la falta de un gobierno de datos eficiente, se abordó identificando las causas y proponiendo soluciones. Los resultados indican mejoras significativas en la eficiencia logística y la satisfacción del cliente. La implementación de estrategias como el uso de Big Data y la creación de catálogos de datos puede ser clave para optimizar procesos, mejorar la calidad de los datos y fortalecer la toma de decisiones informadas en Servientrega. La investigación propone medidas específicas para abordar estas problemáticas y mejorar la eficiencia operativa y la satisfacción del cliente. Se propone la metodología CRIPS–DM, que permite una verificación continua del modelo evaluando la eficacia de los objetivos propuestos.

4.2 Recomendaciones

Es fundamental que los procesos de extracción, limpieza y almacenamiento de datos no tengan fallas pues de ellos dependerá la calidad de los resultados. El equipo de gobierno de datos definirá la solución basada en metodologías analíticas de KDD, SEMMA y CRISP-DM, para dar la respuesta adecuada, considerando lógicas de negocio y funcionales para mejorar procesos dentro de la organización. Se recomienda la implementación de las soluciones propuestas, respaldadas por un monitoreo continuo para garantizar su eficacia a largo plazo.

Bibliografía

1Library. (s.f.). *Metodología SEMMA*. Obtenido de https://1library.co/article/metodolog%C3%ADa-semma-metodolog%C3%ADas-de-miner%C3%ADa-de-datos.zlg77x2y#google_vignette

IBM. (s.f.). *Conceptos básicos de ayuda de CRISP-DM*. Obtenido de https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

IIC, I. d. (s.f.). *La metodología CRISP-DM en ciencia de datos*. Obtenido de https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/

MasContainer. (2023). *Estos son los 8 problemas más recurrentes en la logística*. Obtenido de https://www.mascontainer.com/estos-son-los-8-problemas-mas-recurrentes-en-la-logistica/

PolicyLab. (2022). *Modelo de Gobernanza de la Infraestructura de Datos para el Estado Colombiano* . Obtenido de https://gobiernodigital.mintic.gov.co/692/articles-273751_modelo_gobernanza_infraestructura_datos.pdf

PowerData. (2017). *10 técnicas de análisis de datos para estadísticas de big data*. Obtenido de https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/10-tecnicas-de-analisis-de-datos-para-estadisticas-de-big-data

Rosario, U. (2011). *Estudio comparativo de metodologías para minería de datos*. Obtenido de https://core.ac.uk/download/pdf/301040544.pdf

where, G. E. (s.f.). *Ejemplos de Data Governance*. Obtenido de https://www.grapheverywhere.com/ejemplos-de-data-governance/