

Integración y Automatización de Datos para Inteligencia Artificial en el Proyecto: ETL y Modelo Predictivo para Información de Venta de combustible de la Estación de Servicio TuyaSolar

Fernando Roa¹, Cristian Rivera²

¹⁻²Facultad de Ingeniería y Ciencias Básicas,
Universidad Central
Maestría Analítica de Datos
Automatización e Integración de Datos para la IA
Bogotá, Colombia
{¹froam1, ²criveraa}@correo1.com

September 11, 2023

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Integración y Automatización de Datos para Inteligencia Artificial en el Proyecto: ETL y Modelo Predictivo para Información de Venta (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2.1	Objetivos específicos (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	5
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	5
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	6
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.2	¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	6

3.3	¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.4	¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
4	Diseño de integración y Automatización de Datos para IA (Diagrama) (<i>Primera entrega</i>)	8
5	integración de Datos (<i>Segunda entrega</i>)	8
6	Automatización de Datos (<i>Segunda entrega</i>)	8
7	IA (<i>Segunda entrega</i>)	9
8	Proximos pasos (<i>Tercera entrega</i>)	10
9	Lecciones aprendidas (<i>Tercera entrega</i>)	11
10	Bibliografía	12

1 Introducción (Max 250 Palabras) - (Primera entrega)

En la era actual de la información, la capacidad de integrar y automatizar datos de múltiples fuentes es crucial para el desarrollo y la eficiencia de sistemas basados en Inteligencia Artificial (IA). El presente trabajo se enfoca en la implementación de estas capacidades creando un ETL (Extract, Transform, Load), una metodología consolidada que facilita la extracción de datos desde distintos orígenes, su transformación en un formato adecuado y, finalmente, su carga en sistemas de almacenamiento o análisis, además de la ejecución periódica de predicciones de ventas usando modelos de machine learning utilizando recursos de la nube de Oracle.

El núcleo de este proyecto radica en la concepción de un modelo predictivo especialmente diseñado para analizar información de ventas de combustible de una estación de servicio, la cual a través de un sistema POS registra todas las ventas que realiza. Dicho modelo busca no sólo comprender las tendencias y patrones históricos, sino también anticipar futuras variaciones y comportamientos en el mercado. La implementación de la IA en este contexto se vuelve una herramienta invaluable para las empresas que buscan mantenerse a la vanguardia y tomar decisiones basadas en datos precisos y proyecciones confiables.

Esta investigación abordará las técnicas y herramientas utilizadas para la integración y automatización de datos, así como la estructura y funcionamiento del modelo predictivo propuesto, ofreciendo una perspectiva detallada de su aplicabilidad y potencial impacto en el ámbito comercial.

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (Primera entrega)

El proyecto presenta una serie de características distintivas que lo posicionan como una herramienta esencial para el manejo de grandes volúmenes de datos comerciales. En primer lugar, cuenta con una robusta infraestructura de integración que facilita la recolección y consolidación de datos desde páginas web, asegurando la completitud de la información. La automatización incorporada garantiza flujos de trabajo eficientes y reduce el margen de error humano, maximizando la precisión en cada etapa del proceso ETL (Extracción, Transformación y Carga). En el aspecto de transformación, se han implementado algoritmos avanzados que adaptan y preparan los datos para análisis, garantizando su relevancia y utilidad. El corazón del proyecto es su modelo predictivo basado en técnicas de Inteligencia Artificial, diseñado para detectar tendencias y patrones en la información de venta. Este modelo se reentrena constantemente con nuevos datos, mejorando su precisión y adaptabilidad frente a cambiantes escenarios de mercado. Además, presenta interfaces amigables para los usuarios y sistemas de visualización de datos, permitiendo una interpretación intuitiva de los resultados. Asimismo, se ha puesto un especial énfasis en la escalabilidad y flexibilidad del sistema, permitiendo su adaptación a diferentes tamaños y tipos de organizaciones.

Por último, pero no menos importante, el proyecto se rige por estrictas normas de seguridad y confidencialidad, en materia de privacidad y gestión de datos. El compromiso con estas normativas asegura que, además de proporcionar insights valiosos, el proyecto mantiene la integridad y confianza de sus usuarios. La combinación de estas características convierte a este proyecto en una solución integral, capaz de atender las crecientes demandas del análisis de datos en el mundo comercial y ofrecer previsiones acertadas que pueden guiar estrategias de negocio hacia el éxito. Esta innovadora propuesta representa un paso adelante en la confluencia de la Inteligencia Artificial, la gestión de datos y el análisis predictivo, consolidándose como un referente en el ámbito de la tecnología y la toma de decisiones basadas en datos.

2.1 Integración y Automatización de Datos para Inteligencia Artificial en el Proyecto: ETL y Modelo Predictivo para Información de Venta (Max 100 Palabras) - (Primera entrega)

La estación de Servicio TuyaSolar cuya actividad económica es la venta de combustibles como gasolina corriente y diesel cuenta con un sistema Pos mediante el cual se almacena la información de todas las ventas realizadas. Dicha información puede ser consultada y descargada en archivos de formato .xls desde un sitio web suministrado por el proveedor del sistema.

Actualmente dicha información no es aprovechada en ningún sentido, por lo que el presente proyecto es vital para potenciar la toma de decisiones basadas en datos por parte de la estación.

2.2 Objetivo general (Max 100 Palabras) - (Primera entrega)

Desarrollar e implementar un sistema de Integración y Automatización de Datos mediante la creación de un ETL y un Modelo Predictivo para la proyección de ventas, constituyéndose en una herramienta para la toma de decisiones basadas en datos.

2.2.1 Objetivos específicos (Max 100 Palabras) - (Primera entrega)

- Crear un ETL para la recolección y consolidación de datos, garantizando su homogeneidad y consistencia, además de garantizar la adaptabilidad y escalabilidad del sistema a diferentes contextos organizacionales.
- Crear un modelo de machine learning para predecir las ventas de combustible de la estación de servicio entrenado a partir de datos históricos.
- Automatizar el ETL y adaptar el modelo predictivo para que éste se retroalimente y se adapte a nuevos escenarios de mercado, manteniéndolo ajustado a la realidad.

- Crear una interfaz de visualización intuitiva que facilite la comprensión y explotación de los insights derivados del modelo, al tiempo que se asegura el cumplimiento de las normativas de privacidad y gestión de datos.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

Nuestro alcance incluye el diseño de una infraestructura de integración que permita la armonización y consolidación de datos provenientes de páginas web, garantizando la calidad y coherencia de la información. Se dará especial énfasis a la implementación de algoritmos avanzados para la transformación de datos y la adaptabilidad constante del modelo predictivo a las dinámicas cambiantes del mercado. Además, se desarrollará una interfaz de usuario con un sistema de visualización que brindarán interpretaciones claras y accesibles de los resultados generados. Por último, el proyecto se compromete a respetar las normativas vigentes en materia de privacidad y gestión de datos, asegurando la protección y confiabilidad de la información manejada. Si bien el objetivo principal es el análisis de información de venta, la flexibilidad y adaptabilidad del sistema posibilitan futuras expansiones y adaptaciones a otros ámbitos comerciales y de análisis.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cómo la implementación de un sistema de Integración y Automatización de Datos impacta en la toma de decisiones de Pymes en el desarrollo de su actividad comercial?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

La implementación de un sistema de Integración y Automatización de Datos se convertirá en una herramienta robusta para la toma de decisiones como: negociaciones de contratos con proveedores y clientes, estimación de fechas para la compra del combustible e identificación de temporadas de mayor y menor venta. El uso de la herramienta propuesta conducirá a decisiones estratégicas respaldadas en datos, optimizando las operaciones, la identificación de tendencias de mercado y la respuesta a cambios dinámicos, resultando en una ventaja competitiva sustancial para las organizaciones que adoptan dicho sistema.

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (*Primera entrega*)

La información utilizada para el desarrollo del presente proyecto es de tipo estructurado y consiste en una tabla en donde se alojan los registros de todas las ventas realizadas por la estación de servicio. Dentro de los campos más relevantes destacan: Fecha y hora de la venta, vendedor, producto, volumen de venta, total venta y nombre del cliente en caso de contratos.

Se cuenta con la autorización para el uso de la información por parte del dueño de la estación y de mutuo acuerdo se establece que el producto final será aprovecha solamente por personal autorizado de tal empresa.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

Los datos de ventas de la estación son tomados de la página web con url <https://www.imedoil.online/nexus.php> suministrada por la empresa Imedoil S.A.S. proveedora del sistema POS, con un formato de descarga con extensión .xls.

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

Teniendo en cuenta que los datos utilizados para el proyecto no contienen datos personales, no existe la obligación a cumplir con todas las disposiciones de la Ley de Protección de Datos Personales 1581 de 2012.

A pesar de lo anterior, se tendrán en cuenta las siguientes consideraciones éticas:

- Privacidad: Recopilar y utilizar solo la información no personal que sea necesaria para el propósito previsto.
- Transparencia: Informar a los usuarios sobre cómo se recopila y utiliza su información no personal.
- Seguridad: Tomar medidas de seguridad para proteger la información no personal de los usuarios.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (*Primera entrega*)

Es necesario realizar un análisis de valores faltantes, duplicados y posibles inconsistencias en la data descargada del sitio web, teniendo en cuenta que hasta el momento no ha sido utilizada y se desconoce la presencia o ausencia de posibles inconsistencias.

3.4 ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) -
(Primera entrega)

La integración y automatización de datos en este proyecto es el corazón del mismo, ya que se espera una automatización del 100% de los procesos tales como: extracción, transformación y carga de datos en la nube, predicción y visualización de ventas en tableros cuya disponibilidad será únicamente para usuarios autorizados por la estación de Servicio Tuya Solar, evitando al máximo cualquier tipo de mantenimiento o intervención manual.

4 Diseño de integración y Automatización de Datos para IA (Diagrama) (Primera entrega)

El diseño del proyecto desde la extracción de los datos hasta el tablero de visualización se presenta en el siguiente diagrama:

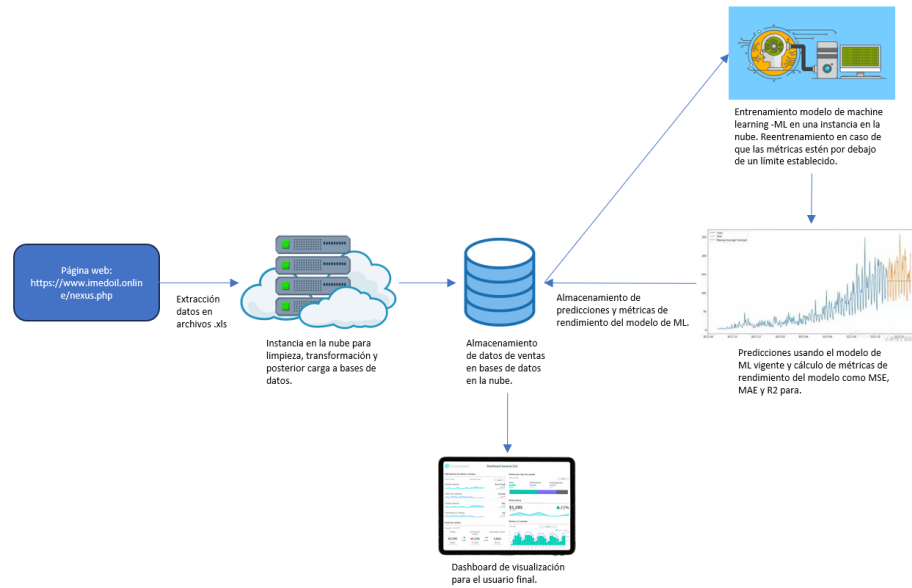


Figure 1: Diagrama del proyecto - End to End

5 integración de Datos (Segunda entrega)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

6 Automatización de Datos (Segunda entrega)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod

mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

7 IA (*Segunda entrega*)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

8 Proximos pasos (*Tercera entrega*)

9 Lecciones aprendidas (*Tercera entrega*)

10 Bibliografía