







**HOME CENTER**<sup>®</sup>  
**SODIMAC corona**

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1  PÁG.: 2 de 1	

MACHINE LEARNING Y WEB SCRAPING APLICADO AL SENTIMIENTO DE  
LOS CLIENTES PARA HOMECENTER

RAFAEL RICARDO SALINAS AMAYA

UNIVERSIDAD POPULAR DEL CESAR  
DEPARTAMENTO DE ADMINISTRACIÓN, CONTADURÍA Y ECONOMIA  
AGUACHICA, CESAR  
2024

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 3 de 1	



MACHINE LEARNING Y WEB SCRAPING APLICADO AL SENTIMIENTO DE LOS  
CLIENTES PARA HOMECENTER

RAFAEL RICARDO SALINAS AMAYA

INFORME PRACTICAS



SONIA ACEVEDO  
ASESOR

UNIVERSIDAD POPULAR DEL CESAR  
DEPARTAMENTO DE ADMINISTRACIÓN, CONTADURÍA Y ECONOMIA  
AGUACHICA, CESAR  
2024



	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1	
		PÁG.: 4 de 1	

## CONTENIDO

CAPÍTULO 1 .....	8
1. ASPECTOS BÁSICOS DE LA EMPRESA.....	8
1.1 ACTIVIDAD ECONÓMICA QUE DESARROLLA LA EMPRESA .....	9
1.1.1 Información adicional relevante:.....	9
1.2 MISIÓN.....	9
1.3 VISIÓN .....	10
1.4 POLÍTICAS.....	10
1.5 VALORES.....	11
1.6 ORGANIGRAMA .....	13
1.7 ÁREA ESPECÍFICA DE LA PRÁCTICA .....	14
CAPÍTULO 2 .....	15
2. ACTIVIDADES ESPECÍFICAS DE LA PRÁCTICA PROFESIONAL. ....	15
2.1 NOMBRE DEL TRABAJO .....	15
2.2 DIAGNÓSTICO .....	15
Fuente. Elaboración Propia.....	15
2.3 JUSTIFICACIÓN .....	15
2.3.1 Teórico .....	15
2.3.2 Práctico .....	16
2.3.3 Metodológico.....	16
2.3.4 Social .....	16
2.4 OBJETIVOS .....	17
2.4.1 Objetivo general .....	17
2.4.2 Objetivos específicos .....	17
2.5 PLAN DE ACTIVIDADES .....	19
2.6 ENTREGABLES .....	20
CAPITULO 3 .....	21
3. Implementación de un proceso de Web Scraping para extraer datos relevantes de diversas fuentes en línea, garantizando la recopilación de información actualizada sobre las opiniones de los clientes de Homecenter. ....	21
3.1 Conceptos Básicos: Estructura de una Página Web .....	21

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1	

3.2	Descripción de las Páginas Web .....	22
3.3	Descripción de los Scripts de Wed Scraping .....	25
4.	Desarrollo de un modelo de machine learning para análisis de sentimiento que clasifique las opiniones en categorías positivas, neutrales y negativas, que resuman los resultados y generen insights valiosos para la empresa. ....	27
4.1	Desarrollo Modelo de Machine Learning .....	27
4.2	Entrenamiento del Modelo y Pipeline de Machine Learning .....	28
4.3	Resultados y Métricas de Evaluación.....	29
4.4	Interpretación de las Métricas:.....	29
5.	Creación de visualizaciones claras que resuman los resultados del análisis de sentimiento, facilitando la interpretación de los datos por parte de los equipos de la empresa. ....	29
5.1	Implementación en Python de las visualizaciones.....	30
5.2	Generación y Compartición de Imágenes .....	31
6.	CONCLUSIONES .....	33
7.	RECOMENDACIONES.....	34
	BIBLIOGRAFÍA .....	35

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 6 de 1	

## GLOSARIO

**Machine Learning:** Campo de la informática que permite a las computadoras aprender sin ser programadas explícitamente.

**Web Scraping:** Técnica automatizada para extraer datos de sitios web.

**Análisis de Sentimiento:** Proceso de identificar y extraer información subjetiva de datos textuales para determinar la opinión o sentimiento expresado.

**TF-IDF (Term Frequency-Inverse Document Frequency):** Técnica de vectorización de texto que asigna un valor numérico a cada palabra en función de su frecuencia en un documento.

**SMOTE (Synthetic Minority Over-sampling Technique):** Técnica de sobre-muestreo que genera ejemplos sintéticos para las clases minoritarias.



**Random Forest Classifier:** Algoritmo de Machine Learning basado en árboles de decisión, que combina múltiples árboles para mejorar la precisión.

**Accuracy:** Métricas de rendimiento que mide la precisión general del modelo.

**Precision:** Métricas de rendimiento que mide la exactitud de las predicciones positivas.

**Recall:** Métricas de rendimiento que mide la capacidad del modelo para identificar correctamente todas las instancias positivas.

**F1 Score:** Métricas de rendimiento que combina la precisión y el recall en una sola cifra.



	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 7 de 1	

## INTRODUCCIÓN

El presente trabajo aborda el resultado de las prácticas en Sodimac Colombia S.A., una empresa líder en el sector del retail especializada en el mejoramiento del hogar y la construcción. Se basa en la implementación de un proceso de análisis de sentimiento que combina técnicas de Machine Learning y Web Scraping. El objetivo principal es obtener información valiosa sobre la percepción de la marca y las tiendas Homecenter, utilizando datos de reseñas y comentarios en línea, para apoyar la toma de decisiones estratégicas de la Gerencia Comercial.

El proyecto se basa en la recopilación de datos de diversas fuentes en línea, incluyendo la página oficial de Homecenter, Google Maps y Facebook. Se implementaron scripts de Web Scraping para extraer información actualizada sobre las opiniones de los clientes, utilizando herramientas como BeautifulSoup y Selenium. Posteriormente, se desarrolló un modelo de Machine Learning para analizar el sentimiento de las opiniones extraídas, utilizando técnicas de preprocesamiento de datos, vectorización de texto, balanceo de clases con SMOTE y entrenamiento con Random Forest Classifier. El modelo se evaluó utilizando métricas de rendimiento como Accuracy, Precision, Recall y F1 Score, demostrando su capacidad para clasificar correctamente el sentimiento de las reseñas.

Finalmente, se crearon visualizaciones interactivas para comunicar los resultados del análisis de sentimiento de manera clara y accesible. Se utilizaron gráficos de barras, nubes de palabras y gráficos de línea para mostrar la distribución del sentimiento, las palabras clave relevantes y la evolución del sentimiento a lo largo del tiempo. Este documento presenta el proceso completo de la práctica profesional, desde la definición de los objetivos hasta la interpretación de los resultados, incluyendo la implementación de las herramientas y técnicas utilizadas. La práctica aporta valiosos insights sobre la percepción del cliente hacia Homecenter, información que puede ser utilizada para optimizar la estrategia comercial de la empresa y mejorar la experiencia del cliente.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 8 de 1	

## CAPÍTULO 1

### 1. ASPECTOS BÁSICOS DE LA EMPRESA



Sodimac Colombia S.A. es una empresa colombiana, cuyos accionistas principales son la Organización Corona S.A. y sus vinculadas con un 51% de participación accionaria e Inversiones Falabella de Colombia S.A. con un 49%. La Sociedad se constituyó legalmente en 1994. A partir de este momento Sodimac inicia su expansión como una gran cadena de retail en el país.

Esta una destacada empresa en el sector del retail, especializada en el mejoramiento del hogar y la construcción. La razón social de la compañía es Sodimac Colombia S.A, y su OAT se encuentra en Carrera 68D # 80 - 70, Bogotá, Colombia. Desde su entrada al mercado colombiano en 1993 con la apertura de su primer almacén, HOMECENTER Américas, Sodimac ha crecido significativamente y se ha establecido como el líder en su categoría. La empresa opera en varias ciudades de Colombia, ofreciendo una amplia gama de productos y servicios diseñados para satisfacer las necesidades de remodelación y construcción tanto para clientes residenciales como empresariales.

Miguel Pardo Brigard es el actual Gerente General de Sodimac Colombia S.A., habiendo asumido el cargo el 2 de noviembre de 2012. Miguel Pardo Brigard cuenta con una sólida formación académica en Administración de Empresas y una destacada trayectoria en el sector retail. Su experiencia incluye roles directivos en importantes empresas de la región, lo que le ha permitido aportar una visión estratégica y un enfoque innovador a Sodimac. Bajo su liderazgo, la empresa ha continuado su expansión y consolidación en el mercado colombiano, destacándose por su compromiso con la calidad del servicio y la satisfacción del cliente.

La empresa cuenta con dos formatos comerciales que permiten atender cualquier requerimiento de los clientes en cuanto a productos y servicios necesarios para construir, remodelar o mejorar su hogar, empresa, oficina o local comercial:

Los hogares colombianos reconocen a Homecenter como el almacén con el mejor surtido de productos y servicios, una opción confiable para proyectos grandes y pequeños de mejoras en la vivienda. La experiencia de compra en Homecenter es excelente, con innovación constante, un ambiente agradable y fácil de recorrer. Los productos están bien organizados y acompañados de información pertinente, y los asesores expertos están disponibles para asistencia adicional. La política de precios bajos siempre asegura precios competitivos durante todo el año, y los clientes tienen acceso a diversas opciones de financiación para facilitar la adquisición de productos. En Homecenter, no solo vendemos productos, sino que ofrecemos proyectos completos.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 9 de 1	

Constructor es un formato diseñado para facilitar la compra a los profesionales del sector, como constructores, arquitectos, maestros de obra y electricistas. Su innovador circuito de compra permite adquirir materiales en las cantidades necesarias de manera rápida y eficiente, con una garantía de precios bajos. El programa Círculo de Especialistas reconoce a los clientes frecuentes de Constructor, ofreciendo beneficios exclusivos que facilitan su trabajo y mejoran su experiencia.

## 1.1 ACTIVIDAD ECONÓMICA QUE DESARROLLA LA EMPRESA

Sodimac Colombia S.A. - Homecenter, es una empresa del sector retail, dedicada a la comercialización de productos para el mejoramiento del hogar, construcción, decoración y muebles. Su objeto social principal es la compra, venta, importación, exportación, representación, distribución y comercialización de todo tipo de productos relacionados con la construcción, remodelación y equipamiento del hogar.

De acuerdo con el Registro Único Tributario (RUT) de la empresa, el código de la actividad económica principal es el 4752, que corresponde a "Comercio al por menor de artículos de ferretería, pinturas, productos de vidrio en establecimientos especializados".



### 1.1.1 Información adicional relevante:

- Forma Jurídica: Sociedad Anónima (S.A.)
  - Tamaño: Empresa grande (considerando su volumen de ventas, número de empleados y activos).
  - Ámbito de Operación: Nacional, con presencia en las principales ciudades de Colombia a través de tiendas físicas y plataforma de comercio electrónico.
  - Composición del Capital: Pertenece al grupo empresarial chileno Falabella, uno de los conglomerados de retail más grandes de Latinoamérica. Y al grupo Corona
- Se caracteriza por ofrecer una amplia variedad de productos y marcas, buscando satisfacer las necesidades de diferentes segmentos de clientes. Ha implementado estrategias omnicanales para brindar una experiencia de compra integral, integrando sus tiendas físicas con su plataforma online. Se destaca por su enfoque en la sostenibilidad, implementando iniciativas para reducir el impacto ambiental de sus operaciones y promover prácticas responsables en su cadena de valor.

## 1.2 MISIÓN

*"Desarrollamos con innovación y sostenibilidad, ofreciendo los mejores productos, servicios y asesoría a través del canal de venta que el cliente escoja, al mejor precio del mercado, para inspirar y construir los sueños y proyectos de nuestros clientes"*

La misión actual de Sodimac Colombia se enfoca en ofrecer productos, servicios y asesoría de alta calidad con un fuerte énfasis en la innovación y sostenibilidad, a través

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 10 de 1	

de múltiples canales de venta. Este enunciado es eficaz en varios aspectos. Primero, refleja el compromiso de la empresa con la innovación y la sostenibilidad, dos factores cruciales en el sector del retail y la construcción. Estos valores son importantes para atraer y retener a los clientes que buscan soluciones modernas y responsables desde el punto de vista ambiental.

En cuanto a la actividad de la empresa, la misión está bien alineada. Sodimac se especializa en el mejoramiento del hogar y la construcción, y la misión menciona la oferta de productos y servicios para estos fines, así como la asesoría, lo que cubre las principales áreas de operación de la empresa. El compromiso con el mejor precio del mercado es también un aspecto relevante, dado que el precio es un factor decisivo para muchos clientes en el sector del retail.

### 1.3 VISIÓN

*“Ser la empresa líder de proyectos para el hogar y construcción, que, mediante el mejoramiento de la calidad de vida, sea la más querida, admirada y respetada por la comunidad, los clientes, los trabajadores y los proveedores en América”*



La visión actual de Sodimac Colombia está bien orientada hacia una ambición a largo plazo y abarca varios aspectos clave. La propuesta de ser la empresa líder en proyectos para el hogar y construcción es adecuada y refleja una meta clara de dominancia en el mercado. El enfoque en mejorar la calidad de vida es relevante y positivo, alineándose con el propósito de ofrecer soluciones que impacten positivamente en los hogares y espacios de trabajo de los clientes.

La visión también se centra en ser la empresa más querida, admirada y respetada por la comunidad, clientes, trabajadores y proveedores. Este aspecto es importante porque refleja un deseo de establecer relaciones sólidas y de confianza con todos los grupos de interés, lo cual es crucial para el éxito a largo plazo.

### 1.4 POLÍTICAS

En el contexto empresarial de Sodimac Colombia S.A., las políticas establecen los principios fundamentales que guían las operaciones y la conducta de la empresa. Estas políticas están diseñadas para asegurar la excelencia en la calidad de productos y servicios, al tiempo que promueven prácticas sostenibles y responsables. La política de calidad de Sodimac se enfoca en ofrecer productos que cumplan con los más altos estándares de seguridad y durabilidad, garantizando así la satisfacción de los clientes. La empresa también pone un fuerte énfasis en la sostenibilidad, implementando prácticas que reducen su impacto ambiental.

En cuanto a la atención al cliente, se busca proporcionar un servicio excepcional, respaldado por procesos eficientes para la resolución de problemas y la retroalimentación

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 11 de 1	

constante. Además, Sodimac Colombia S.A. mantiene un entorno de trabajo basado en la ética y el cumplimiento de las leyes, incluyendo una política rigurosa de protección de datos y de competencia leal. Estas políticas reflejan el compromiso de la empresa con una gestión empresarial íntegra y efectiva, y son fundamentales para el éxito y la reputación de la organización.

- Política de Calidad: "Ofrecemos productos y servicios que cumplen con los más altos estándares de calidad y seguridad, garantizando la satisfacción del cliente a través de rigurosos controles y mejora continua."
- Política de Sostenibilidad: "Implementamos prácticas sostenibles para minimizar nuestro impacto ambiental y promover el uso responsable de recursos, colaborando con proveedores que compartan nuestro compromiso con la sostenibilidad."
- Política de Atención al Cliente: "Brindamos una atención al cliente excepcional, facilitando la resolución de problemas y procesos de devolución, y utilizando la retroalimentación para mejorar continuamente."
- Política de Ética y Cumplimiento: "Fomentamos un ambiente de trabajo ético y transparente, cumpliendo con todas las leyes y regulaciones vigentes y promoviendo la igualdad y la diversidad."
- Política de Protección de Datos: "Protegemos la privacidad y seguridad de los datos personales mediante medidas adecuadas y cumplimiento con las leyes de protección de datos."
- Política General de Leal y Libre Competencia: "Operamos en un entorno de competencia leal y libre, respetando las normativas y promoviendo prácticas comerciales transparentes y justas."

## 1.5 VALORES



En Sodimac Colombia S.A., los valores fundamentales que guían la cultura organizacional se encapsulan en el acrónimo REIR. Este conjunto de valores no solo define el enfoque hacia el trabajo, sino también la manera de interactuar con las personas y con nuestro entorno. Los valores que forman el acrónimo REIR son los siguientes:

Figura 1  
Valores de la Cultura Organizacional de Sodimac

**REIR:** Responsabilidad -  
Excelencia - Integridad - Respeto

Nota. Este gráfico ilustra los valores centrales de Sodimac Colombia S.A.

- R: Respeto – Valorar a cada persona y sus ideas, promoviendo un ambiente

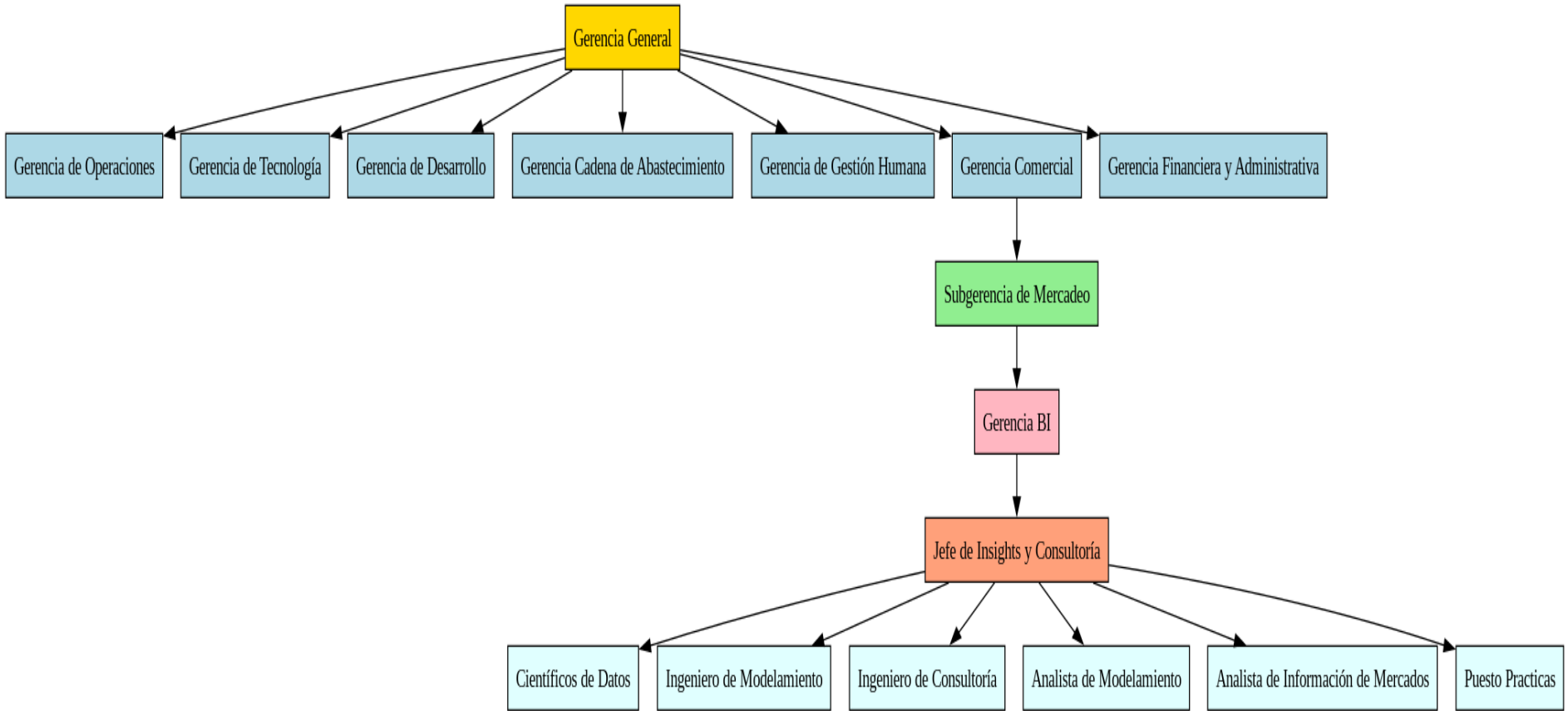
	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 12 de 1	



inclusivo y colaborativo.

- E: Excelencia – Compromiso con la calidad y la mejora continua en todos los aspectos de nuestra operación.
- I: Integridad – Actuar con honestidad y transparencia, manteniendo elevados estándares éticos en todas nuestras acciones.
- R: Responsabilidad – Asumir la responsabilidad de nuestras acciones y decisiones, tanto en el ámbito laboral como en el comunitario.

Los valores de REIR en Sodimac Colombia S.A. están bien alineados con las actividades y objetivos de la empresa. La Excelencia y la Responsabilidad son esenciales para mantener la calidad en los productos y servicios que la empresa ofrece, y para cumplir con sus compromisos hacia los clientes y la comunidad. La Integridad y el Respeto son cruciales para fomentar un ambiente de trabajo ético y colaborativo, lo que refuerza la confianza interna y externa en la empresa.

### 1.6 ORGANIGRAMA



	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 14 de 1	



## 1.7 ÁREA ESPECÍFICA DE LA PRÁCTICA

Actualmente, estoy realizando mis prácticas profesionales en el área de Insight y Consultoría de Sodimac Co. Homecenter. Esta área juega un papel fundamental en la empresa al utilizar los datos recolectados para brindar información estratégica, crear modelos predictivos y apoyar la toma de decisiones de la Gerencia Comercial. El área de Insight y Consultoría se reporta directamente a la Gerencia Comercial, siendo un pilar fundamental para su estrategia. El equipo está compuesto por 6 miembros, liderados por un jefe inmediato, Ernesto Agudelo, quien se desempeña como Cargo de Jefe de Insight y Consultoría.

Mis funciones dentro del área se centran principalmente en la investigación y análisis de datos, buscando insights accionables para la Gerencia Comercial. Esto implica:

- Búsqueda y recopilación de información: Identifico las fuentes de datos relevantes dentro de la empresa, incluyendo bases de datos de clientes, ventas, operaciones, y marketing.
- Análisis y procesamiento de datos: Utilizo herramientas de análisis y software especializado para procesar la información, identificar tendencias, patrones y oportunidades.
- Presentación de información relevante: Elaboro reportes, presentaciones y visualizaciones (dashboards) claras y concisas para comunicar los hallazgos clave al equipo y a la Gerencia Comercial.

Si bien mis funciones aún están en desarrollo, confío en que esta experiencia en Sodimac Co. Homecenter me permitirá fortalecer mis habilidades de análisis, desarrollar mi capacidad de trabajar con grandes volúmenes de datos y comprender la importancia del análisis de datos para la toma de decisiones estratégicas en una empresa líder como Sodimac.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

## CAPÍTULO 2

### 2. ACTIVIDADES ESPECÍFICAS DE LA PRÁCTICA PROFESIONAL.

#### 2.1 NOMBRE DEL TRABAJO

MACHINE LEARNING Y WEB SCRAPING APLICADO AL SENTIMIENTO DE LOS CLIENTES PARA HOMECENTER

#### 2.2 DIAGNÓSTICO

Figura 2  
Matriz Dofa Diagnostico Sodimac





Fuente. Elaboración Propia

#### 2.3 JUSTIFICACIÓN

##### 2.3.1 Teórico

La presente práctica empresarial tiene un fundamento teórico sólido en el campo del Machine Learning y el Web Scraping. Se explorarán las últimas tecnologías de análisis de sentimiento del cliente, con un enfoque en la aplicación de algoritmos de aprendizaje automático, específicamente para la clasificación de opiniones y comentarios. La implementación del Web Scraping permitirá la extracción automatizada de datos textuales relevantes de diversas plataformas digitales, creando un conjunto de datos robusto para el análisis.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 16 de 1	

Este proyecto profundizará en el uso de herramientas y metodologías avanzadas para la obtención y análisis de datos textuales, con un enfoque en el análisis de sentimiento del cliente, un campo de gran relevancia en el contexto actual del marketing digital y la gestión de la experiencia del cliente. La aplicación de estas tecnologías innovadoras permitirá avanzar en el conocimiento del comportamiento del cliente y la percepción de los productos y servicios ofrecidos por Homecenter.

### **2.3.2 Práctico**

La implementación de este proyecto se enfocará en el análisis de sentimiento del cliente, obtenido a través de la recopilación de datos de reseñas y comentarios, proporcionará información valiosa sobre las fortalezas y debilidades de los productos y servicios ofrecidos, permitiendo identificar las áreas de mejora y optimizar la estrategia de la empresa.

Los insights obtenidos a través del análisis de sentimiento contribuirán a la toma de decisiones estratégicas en áreas como el desarrollo de productos, la gestión de marketing, la atención al cliente y la optimización de los procesos operativos. Al comprender las necesidades y expectativas de los clientes, Homecenter estará en una mejor posición para ofrecer productos y servicios que se adapten a sus preferencias, creando una experiencia de compra más satisfactoria y fomentando la fidelización de los clientes.



### **2.3.3 Metodológico**

La práctica empresarial se basará en la aplicación de metodologías de aprendizaje automático y Web Scraping. Se utilizarán herramientas de programación como Python y bibliotecas como BeautifulSoup, Selenium y NLTK para la extracción y el procesamiento de datos textuales. Se implementarán modelos de aprendizaje automático para la clasificación de sentimientos, se evaluará el rendimiento de los modelos y se presentarán los resultados a través de dashboards interactivos.

La metodología del proyecto permitirá explorar la aplicación práctica de las tecnologías de Machine Learning y Web Scraping en el análisis de sentimiento del cliente, con un enfoque en la obtención de datos de fuentes externas, la limpieza y preparación de los datos, la creación y entrenamiento de modelos de aprendizaje automático y la visualización de los resultados. Se utilizarán técnicas de procesamiento del lenguaje natural para el preprocesamiento de los datos textuales, garantizando la calidad y la validez del análisis.

### **2.3.4 Social**

Este proyecto tiene un impacto social positivo, al contribuir a la mejora de la experiencia del cliente de Homecenter. Al comprender las necesidades y preferencias de los clientes,

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11 VERSIÓN: 1	
	<b>PRACTICAS CURRICULARES  INFORME</b>	PÁG.: 17 de 1	

Homecenter podrá ofrecer productos y servicios que se adapten a sus expectativas, lo que generará mayor satisfacción y lealtad. Además, el proyecto permitirá la toma de decisiones más responsables y éticas, considerando la opinión y el feedback de los clientes. El análisis de sentimiento también puede ayudar a Homecenter a identificar y abordar posibles problemas con sus productos o servicios, mejorando su imagen y la confianza del público.

## 2.4 OBJETIVOS

### 2.4.1 Objetivo general

Desarrollar un proceso estructurado de análisis de sentimiento mediante técnicas de web scraping y machine learning, que permita evaluar las opiniones de los clientes sobre la marca y tiendas Homecenter proporcionando insights para mejorar la estrategia comercial.



### 2.4.2 Objetivos específicos

- Implementar un proceso de web scraping para extraer datos relevantes de diversas fuentes en línea, garantizando la recopilación de información actualizada sobre las opiniones de los clientes de tiendas Homecenter.
- Desarrollar un modelo de machine learning para análisis de sentimiento que clasifique las opiniones en categorías positivas, neutrales y negativas, que resuman los resultados y generen insights valiosos para la empresa.
- Crear visualizaciones interactivas y claras que resuman los resultados del análisis de sentimiento, facilitando la interpretación de los datos por parte de los equipos de la empresa.





## 2.5 PLAN DE ACTIVIDADES

Actividades	Agosto					Septiembre					Octubre					Noviembre					Diciembre				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<b>Fase 1: Planificación y Preparación (01 - 07 de Agosto, 2024)</b>																									
01 - 03 de Agosto: Identificación de fuentes de datos: Listar sitios web, foros, redes sociales, y otras fuentes relevantes para la industria.																									
04 - 05 de Agosto: Selección de herramientas de Web Scraping: Elegir herramientas de scraping (e.g., BeautifulSoup, Scrapy, Selenium).																									
06 - 07 de Agosto: Configuración del entorno de desarrollo: Instalar y configurar el software necesario, definir estructura de carpetas y repositorio de código.																									
<b>Fase 2: Web Scraping (08 de Agosto - 31 de Agosto, 2024)</b>																									
08 - 14 de Agosto: Desarrollo del script de Web Scraping: Crear scripts iniciales para extraer datos de las fuentes seleccionadas.																									
15 - 21 de Agosto: Ejecución y prueba de scraping: Ejecutar el scraping y realizar ajustes en los scripts según sea necesario.																									
22 - 28 de Agosto: Validación de los datos obtenidos: Revisar y verificar la relevancia y consistencia de los datos obtenidos.																									
29 - 31 de Agosto: Documentación de resultados de scraping: Documentar el proceso, retos y ajustes realizados durante la fase de scraping.																									
<b>Fase 3: Limpieza y Preprocesamiento de Datos (01 de Septiembre - 20 de Septiembre, 2024)</b>																									
01 - 05 de Septiembre: Análisis exploratorio de los datos: Realizar análisis preliminares para identificar inconsistencias, valores nulos, duplicados, etc.																									
06 - 12 de Septiembre: Limpieza de datos: Eliminar datos irrelevantes, corregir errores, tratar valores faltantes.																									
13 - 20 de Septiembre: Preprocesamiento de datos: Normalización, tokenización, eliminación de stopwords, y otras técnicas necesarias para preparar los datos para el análisis de sentimiento.																									
<b>Fase 4: Análisis de Sentimiento (21 de Septiembre - 20 de Octubre, 2024)</b>																									
21 - 30 de Septiembre: Selección de técnicas de análisis de sentimiento: Decidir el enfoque (e.g., modelos basados en reglas, modelos ML como SVM, RNN, etc.).																									
01 - 10 de Octubre: Desarrollo del modelo de análisis de sentimiento: Entrenamiento y validación del modelo usando datos preprocesados.																									
11 - 17 de Octubre: Pruebas y refinamiento: Ajustar parámetros y mejorar el modelo según los resultados de las pruebas.																									
18 - 20 de Octubre: Evaluación de desempeño del modelo: Medir precisión, recall, F1 score, etc., para asegurar la calidad del modelo.																									
<b>Fase 5: Implementación y Almacenamiento en GCP (21 de Octubre - 15 de Noviembre, 2024)</b>																									
21 - 25 de Octubre: Implementación del modelo: Desplegar el modelo para su uso en un entorno de producción.																									
26 de Octubre - 05 de Noviembre: Integración con GCP: Configurar la tabla para el almacenamiento en GCP para almacenar comentarios, fechas, sitios de extracción,																									
06 - 10 de Noviembre: Pruebas de integración: Verificar que los datos se almacenen correctamente en la tabla de GCP.																									
11 - 15 de Noviembre: Documentación del proceso: Crear la documentación completa de la implementación y uso del modelo y almacenamiento en GCP.																									
<b>Fase 6: Revisión y Entrega Final (16 de Noviembre - 01 de Diciembre, 2024)</b>																									
16 - 25 de Noviembre: Preparación de informe final: Compilación de un informe detallado sobre todo el proceso, hallazgos, y recomendaciones.																									
26 - 30 de Noviembre: Presentación del proyecto al equipo ejecutivo: Mostrar resultados y entregar el informe final a Ernesto A. Ernesto R.																									
01 de Diciembre: Entrega final del proyecto: Hacer la entrega formal del proyecto con todas las documentaciones, código, y acceso al entorno de producción en GCP.																									

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1  PÁG.: 20 de 1	

## 2.6 ENTREGABLES

- Un modelo de Machine Learning que clasifique las opiniones en categorías específicas (por ejemplo, positivas, neutrales, negativas) y así entender mejor las preferencias del mercado o las opiniones de los clientes.
- Visualizaciones claras de los resultados del análisis de sentimiento para comunicar hallazgos de manera efectiva (gráficos y tablas en el júpiter notebook como resumen de la metodología aplicada).

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1  PÁG.: 21 de 1	

## CAPITULO 3

### 3. Implementación de un proceso de Web Scraping para extraer datos relevantes de diversas fuentes en línea, garantizando la recopilación de información actualizada sobre las opiniones de los clientes de Homecenter.

El web scraping es una técnica utilizada para extraer datos de sitios web de manera automatizada. Este proceso permite recopilar información relevante y actualizada, como las opiniones de los clientes sobre productos o servicios, facilitando la toma de decisiones basada en datos. Se refiere al proceso de extraer datos estructurados de páginas web. A través de esta técnica, es posible descargar y procesar datos relevantes de manera eficiente, automatizando lo que de otro modo sería un proceso manual de recopilación. Como se señala en la documentación de BeautifulSoup:

BeautifulSoup es una biblioteca de Python para analizar documentos HTML y XML. Proporciona herramientas para navegar por el árbol de documentos, buscando y extrayendo elementos con facilidad" (Richardson, 2011).

Por otro lado, Selenium es una herramienta útil para interactuar con páginas web que cargan contenido dinámicamente mediante JavaScript. Según su documentación oficial:

Selenium WebDriver es una interfaz para interactuar con navegadores de manera programática. Es útil para automatizar pruebas y también para interactuar con páginas web dinámicas" (Selenium, 2024).

#### **Herramientas utilizadas:**



BeautifulSoup: Utilizado para el scraping de páginas web estáticas, en las que el contenido está disponible directamente en el HTML.

Selenium: Ideal para interactuar con sitios web dinámicos que cargan contenido a través de JavaScript.

### **3.1 Conceptos Básicos: Estructura de una Página Web**

Antes de proceder con el scraping, es crucial entender cómo están estructuradas las páginas web. Las páginas web están construidas principalmente con HTML y CSS, los cuales definen su estructura y estilo visual. Según la documentación de HTML Living Standard:

"El HTML define la estructura de la web, organizando contenidos en una jerarquía de elementos como <div>, <span>, <p>, <h1>, entre otros" (HTML Living Standard, 2023).

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

El DOM (Document Object Model) representa la estructura jerárquica de una página web en un árbol de nodos. Cada nodo del DOM corresponde a un elemento HTML, lo que permite navegar y acceder a los datos de forma eficiente.

En este proceso de web scraping, se identifican los elementos HTML y sus clases o ID, que son utilizados para extraer información específica. Esto es posible gracias a bibliotecas como BeautifulSoup, que permite buscar y seleccionar estos elementos fácilmente a través de selectores CSS o expresiones XPath.

### 3.2 Descripción de las Páginas Web

Para este proyecto, se seleccionaron tres páginas web representativas que contienen opiniones de clientes sobre productos o servicios. Las páginas scrapeadas incluyen:

#### **Página oficial de Homecenter**

La página web oficial de Homecenter es el principal canal de ventas en línea de la empresa, donde los clientes pueden encontrar productos, realizar compras y consultar información sobre servicios. En esta página, los clientes también pueden dejar reseñas sobre los productos adquiridos. En términos generales, la estructura de la página es clara y fácil de navegar, lo que permite a los usuarios dejar comentarios en la sección de productos.

Las reseñas en la página de Homecenter se enfocan principalmente en la calidad de los productos, el proceso de compra y la atención al cliente. Los comentarios son mayormente positivos, con muchas menciones a la facilidad de compra en línea y la rapidez en la entrega. Sin embargo, también se destacan algunas quejas sobre problemas en la atención al cliente y la calidad de ciertos productos.

#### **Ejemplo de reseña:**

"Los productos son de buena calidad y la compra fue fácil. El único inconveniente fue el retraso en la entrega, ya que esperábamos recibir los productos más rápido."

Imagen:

A continuación, se muestra una captura de pantalla de la sección de reseñas en la página oficial de Homecenter:



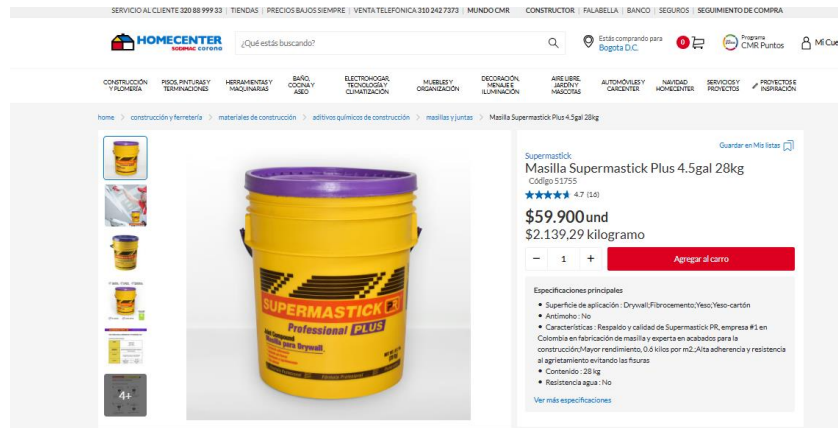
	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1	

Figura 3  
Pagina Inicial Homecenter tienda virtual



Nota. Esta se caracteriza por presentar varios niveles a hora de realizar el scraping.

### Google Maps - Opiniones sobre Homecenter

En Google Maps, los usuarios pueden calificar las tiendas Homecenter ubicadas en diversas ciudades y dejar comentarios sobre su experiencia de compra en tienda física. Esta plataforma es especialmente útil para conocer la satisfacción de los clientes con la atención en los puntos de venta y la calidad de los productos disponibles.

Las reseñas en Google Maps varían según la ubicación, pero comúnmente los clientes mencionan aspectos como la amabilidad del personal y la disponibilidad de productos. También se observan comentarios sobre la limpieza de las tiendas y la facilidad para encontrar los productos deseados.

### Ejemplo de reseña:

"Excelente atención al cliente, encontré todo lo que buscaba. El lugar estaba bien organizado y limpio. ¡Volveré pronto!".

"Tuve problemas para encontrar un empleado disponible para ayuda, pero el lugar es grande y tiene una buena variedad de productos.".



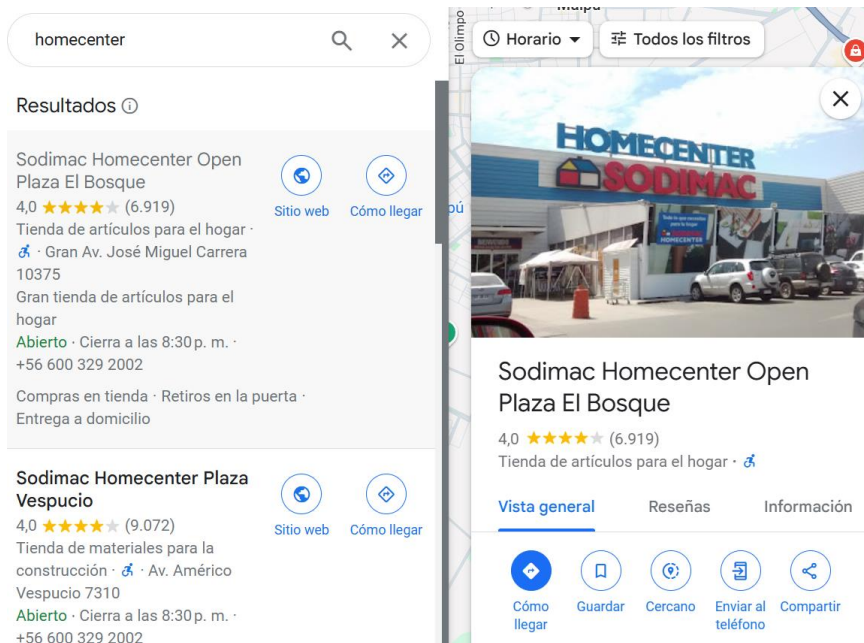
	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1	

Figura 4  
Captura Google Maps reseñas tienda El bosque Bogotá



Nota: captura de pantalla de una reseña de Google Maps sobre una tienda Homecenter:



### Facebook - Opiniones en la Página Oficial de Homecenter

La página oficial de Homecenter en Facebook también es una plataforma activa donde los clientes dejan sus reseñas y comentarios sobre productos, promociones y su experiencia en general. A diferencia de otras plataformas, los comentarios en Facebook a menudo son más interactivos, ya que los usuarios pueden responder a publicaciones y realizar preguntas directamente a la empresa.

Las opiniones en Facebook incluyen tanto comentarios positivos como negativos. Los clientes elogian las promociones y la facilidad para navegar por la tienda en línea, pero también se mencionan problemas con el servicio postventa y la calidad de algunos productos. Algunos usuarios también comparten fotos de sus compras.

### Ejemplo de reseña:

"Me encantaron las promociones en la tienda online, pero la calidad de una lámpara que

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

compré no era la esperada. Tuvieron que hacer un cambio, pero todo salió bien."  
"Excelente atención en la tienda física y variedad de productos. Muy contento con mi compra."

Figura 5  
Página oficial de Facebook de Homecenter





Nota: captura de pantalla de una reseña de Facebook sobre Homecenter:

### 3.3 Descripción de los Scripts de Wed Scraping

El proceso de recolección de datos comienza con la navegación a la página web utilizando la herramienta Selenium. Esta herramienta permite abrir el navegador de manera automatizada y simular las interacciones necesarias en la página, como hacer clic en botones o desplazarse hacia abajo. Este último es especialmente útil cuando el contenido de la página se carga dinámicamente mediante JavaScript, ya que permite que el script se desplace hacia el final de la página, forzando la carga de más reseñas u otros datos que pueden no estar presentes inicialmente en el HTML cargado. Selenium es esencial en este tipo de páginas ya que interactúa con la interfaz de usuario, cargando la información relevante antes de pasar a la siguiente etapa.

Una vez que la página ha sido completamente cargada, se obtiene el HTML completo de la página a través de Selenium, que es el contenido que se utiliza para extraer la información necesaria. En este paso, BeautifulSoup entra en acción. BeautifulSoup es una biblioteca de Python que facilita el análisis del HTML y permite extraer datos específicos de manera eficiente. Utilizando los métodos de búsqueda de BeautifulSoup, es posible localizar las etiquetas HTML correspondientes a las reseñas de los clientes, como el texto de las opiniones y las calificaciones, que luego pueden ser extraídas para su análisis posterior.

Una vez obtenidos los datos relevantes, el siguiente paso es la limpieza de los datos. Los datos extraídos del HTML no siempre son perfectos. Pueden contener etiquetas HTML innecesarias, caracteres especiales o espacios en blanco redundantes. Por ejemplo, las reseñas pueden

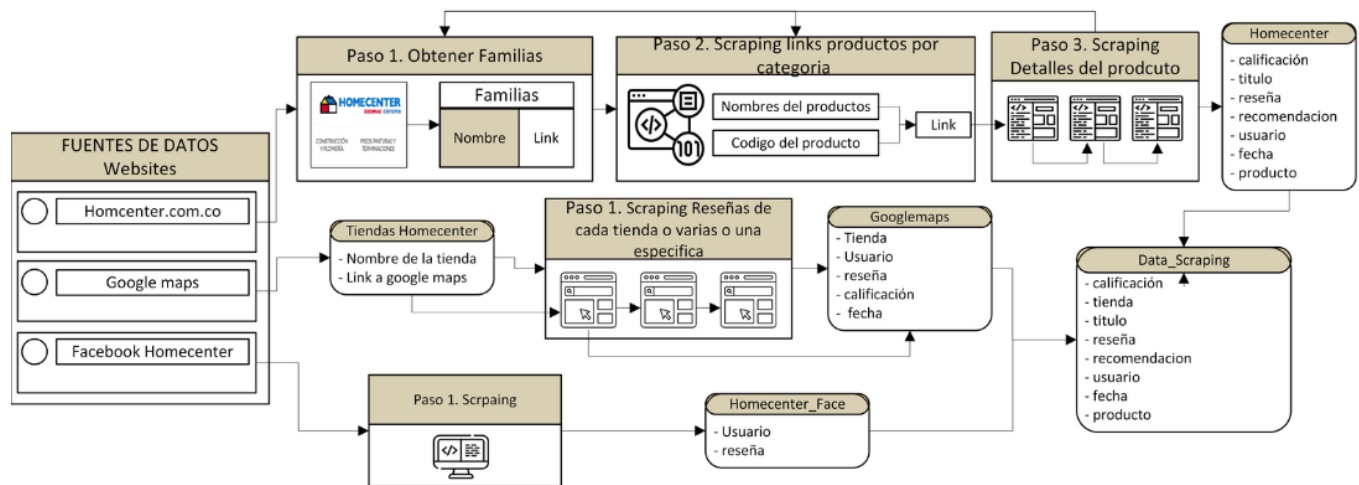
	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	<b>PRÁG.: 26 de 1</b>	

incluir saltos de línea o caracteres especiales que no aportan valor y podrían interferir con el análisis posterior. Para solucionar esto, se utiliza un proceso de limpieza de texto en el cual se eliminan estos elementos no deseados. Esto incluye la eliminación de etiquetas HTML y el uso de expresiones regulares para reemplazar saltos de línea, tabulaciones y otros caracteres no alfanuméricos.



Además, la normalización del texto es otro paso crucial en la limpieza. Este proceso implica transformar el texto en un formato estándar, eliminando acentos o caracteres especiales. Esto es particularmente útil cuando se trabajan con datos multilingües o cuando se desea comparar opiniones en un formato consistente, sin importar las diferencias de escritura o las variaciones de los caracteres. Por ejemplo, en el caso de reseñas en español, se eliminan los acentos para hacer que las palabras sean más fáciles de procesar y analizar, además de convertir todo el texto a minúsculas para evitar que las mismas palabras en mayúsculas o minúsculas se cuenten como diferentes.

Una vez que los datos han sido limpiados, se almacenan en un formato adecuado para su análisis posterior. Un formato común es el CSV (Comma-Separated Values), ya que es fácil de leer y manipular tanto para análisis programáticos como para su visualización en herramientas como Excel o Google Sheets. Almacenar los datos en un archivo CSV permite que los analistas o los sistemas automatizados accedan fácilmente a las reseñas y calificaciones, lo que facilita la generación de informes, estadísticas y otros análisis cuantitativos o cualitativos.

Figura 6  
Representación Funcionamiento de código y almacenamiento de datos.



Nota. Elaboración Propia

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

#### 4. Desarrollo de un modelo de machine learning para análisis de sentimiento que clasifique las opiniones en categorías positivas, neutrales y negativas, que resuman los resultados y generen insights valiosos para la empresa.

Un modelo de machine learning es un algoritmo matemático que aprende patrones a partir de datos, sin ser explícitamente programado para realizar una tarea específica. En el contexto de análisis de sentimientos, el modelo aprende a partir de un conjunto de datos de reseñas etiquetadas (con emociones o sentimientos clasificados como positivos, negativos o neutrales) para predecir el sentimiento de nuevas reseñas no etiquetadas.

El aprendizaje automático se puede dividir en tres tipos principales:

- Aprendizaje supervisado: El modelo aprende de ejemplos etiquetados (en este caso, reseñas clasificadas como satisfecho, insatisfecho o neutral).
- Aprendizaje no supervisado: El modelo intenta identificar patrones en los datos sin etiquetas.
- Aprendizaje por refuerzo: El modelo aprende a través de recompensas o castigos a medida que realiza ciertas acciones.
- En este caso, se está utilizando aprendizaje supervisado, donde las reseñas de los clientes ya están clasificadas en categorías de sentimiento (positivas, negativas o neutrales), y el modelo aprende a partir de esos ejemplos.

##### 4.1 Desarrollo Modelo de Machine Learning



Para desarrollar un modelo de análisis de sentimientos, primero se necesita un conjunto de datos etiquetado que contenga opiniones de clientes previamente clasificadas en categorías como "satisfecho", "insatisfecho" o "neutral". Estos conjuntos de datos son cruciales para el proceso de entrenamiento y evaluación del modelo.

En este caso, el proceso de tratamiento de datos se lleva a cabo en varias etapas:

Primera etapa recopilación de Datos, Se obtienen las reseñas de clientes de diversas fuentes, como redes sociales, páginas de productos o encuestas de satisfacción.

Preprocesamiento: Antes de alimentar los datos en el modelo, es necesario limpiarlos.

Esto incluye, eliminar caracteres especiales, puntuación y otros elementos no alfanuméricos. convertir el texto a minúsculas para evitar que las mismas palabras se traten como diferentes debido a las mayúsculas. Tokenización y lematización: dividir el texto en palabras y reducir esas palabras a su forma básica.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

Segunda etapa división del Conjunto de Datos: Los datos se dividen en dos conjuntos: uno de entrenamiento y otro de prueba. El conjunto de entrenamiento es el que se usa para enseñar al modelo, mientras que el conjunto de prueba se utiliza para evaluar su rendimiento en datos que no ha visto antes. Usualmente, se utiliza una proporción del 70%-80% para el entrenamiento y el 20%-30% restante para las pruebas.

Tercera etapa balanceo de Clases con SMOTE: En los conjuntos de datos de sentimientos, algunas clases pueden estar desbalanceadas (por ejemplo, puede haber muchas más reseñas positivas que negativas). En estos casos, se utiliza SMOTE (Synthetic Minority Over-sampling Technique), una técnica de sobre-muestreo que genera ejemplos sintéticos para las clases minoritarias, equilibrando las clases para evitar que el modelo tenga un sesgo hacia la clase mayoritaria.

#### 4.2 Entrenamiento del Modelo y Pipeline de Machine Learning

Para entrenar el modelo de análisis de sentimientos, se utiliza un pipeline de machine learning, que es un flujo de trabajo secuencial que lleva los datos desde su entrada hasta la predicción final. Este pipeline incluye varias etapas clave:

- **Vectorización de Texto con TFIDF**



La Vectorización es el proceso de convertir el texto en una representación numérica que el modelo pueda entender. En este caso, se utiliza el TFIDF Vectorizer (Term Frequency-Inverse Document Frequency), que asigna un valor numérico a cada palabra en función de su frecuencia en un documento, pero penalizando las palabras que aparecen con demasiada frecuencia en todos los documentos. Esto ayuda a darle mayor peso a las palabras que son más relevantes en cada reseña.

- **Balanceo de Clases con SMOTE**

Como mencionamos anteriormente, SMOTE se utiliza para generar ejemplos sintéticos de las clases minoritarias (en este caso, insatisfecho o neutral) para que el modelo no tenga un sesgo hacia la clase mayoritaria (generalmente, satisfecho). Esta técnica es crucial para mejorar la precisión y la fiabilidad del modelo.

- **Entrenamiento con Random Forest Classifier**

El Random Forest Classifier es un algoritmo de aprendizaje supervisado basado en árboles de decisión. Funciona creando múltiples árboles de decisión (por lo que es un modelo en ensamblaje) y luego promediando sus resultados. Este enfoque reduce el riesgo de sobreajuste y mejora la precisión. Durante el entrenamiento, el modelo aprende a identificar patrones en el texto de las reseñas para predecir si el sentimiento es positivo, negativo o neutral.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1	

Una vez entrenado, el modelo está listo para clasificar nuevas reseñas y predecir su sentimiento (satisfecho, insatisfecho, o neutral).

### 4.3 Resultados y Métricas de Evaluación

El modelo se evalúa utilizando varias métricas de rendimiento, que miden la calidad de las predicciones. Las métricas clave son accuracy (precisión general), precision (precisión), recall (recuperación) y F1 score. Estas métricas ayudan a entender qué tan bien el modelo está clasificando las reseñas en las categorías deseadas.

A continuación, se detallan los resultados obtenidos al evaluar el modelo en los datos de prueba:

Tabla 1  
Resultado modelo Machine learning Random Forest

	Accuracy	Precision	Recall	F1-Score
Insatisfecho	82%	86%	86%	1193
Neutral	80%	78%	78%	1194
Satisfecho	92%	93%	93%	4804

Nota.



Elaboración propia

### 4.4 Interpretación de las Métricas:

El modelo predice correctamente el sentimiento de 89% de las reseñas la precisión la cual mide la exactitud de las predicciones positivas. Por ejemplo, una precisión de 0.94 para "satisfecho" significa que el 94% de las veces que el modelo predijo "satisfecho", realmente fue correcto. Otra parte la Recuperación la cual mide la capacidad del modelo para identificar correctamente todas las instancias positivas. Un recall de 0.89 para "insatisfecho" indica que el modelo pudo identificar el 89% de las reseñas verdaderamente insatisfechas. El F1 Score: es una métrica que combina la precisión y el recall en una sola cifra, buscando un balance entre ambos. Un F1 score de 0.93 para "satisfecho" sugiere que el modelo tiene un buen equilibrio entre precisión y recuperación.

### 5. Creación de visualizaciones claras que resuman los resultados del análisis de sentimiento, facilitando la interpretación de los datos por parte de los equipos de la empresa.

Para facilitar la interpretación de los resultados del análisis de sentimiento por parte de los equipos de la empresa, se crearon visualizaciones claras y concisas. Estas visualizaciones

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

permiten una rápida comprensión de la información y brindan insights valiosos para la toma de decisiones.

Se utilizaron diferentes tipos de gráficos para representar los resultados del análisis. Por ejemplo, se creó un gráfico de barras que muestra la distribución de las opiniones por sentimiento (positivo, negativo y neutral), permitiendo una visión general del panorama general. Además, se diseñó una nube de palabras que muestra las palabras clave más relevantes para cada sentimiento.

Este mapa de calor ayuda a identificar los temas o conceptos que están más relacionados con las opiniones positivas, negativas o neutras, proporcionando información valiosa sobre los aspectos que generan mayor satisfacción o insatisfacción en los clientes. Finalmente, se implementó un gráfico de línea que muestra la evolución del sentimiento a lo largo del tiempo. Esta visualización permite analizar las tendencias en la percepción de la marca y detectar posibles cambios o eventos que puedan influir en la satisfacción de los clientes.

## 5.1 Implementación en Python de las visualizaciones

Para llevar a cabo la creación de las visualizaciones, se utilizó Python debido a su potente ecosistema de bibliotecas para análisis de datos y visualización. A continuación, se describe el proceso utilizado para generar cada uno de los gráficos mencionados anteriormente:

**Gráfico de barras:**

Para representar la distribución de las opiniones por sentimiento (positivo, negativo y neutral), se emplearon bibliotecas como Matplotlib y Seaborn. Estas herramientas permitieron crear un gráfico de barras de fácil lectura, donde se mostraron las cantidades de opiniones clasificadas por sentimiento. Además, se personalizó el gráfico para que los colores y las etiquetas fueran claros y accesibles para cualquier miembro del equipo.

**Nube de palabras:**

Con el objetivo de resaltar las palabras clave asociadas a cada tipo de sentimiento, se utilizó la biblioteca WordCloud. Esta herramienta generó nubes de palabras para cada sentimiento (positivo, negativo y neutral), en las cuales las palabras más frecuentes se mostraron con mayor tamaño, proporcionando una representación visual rápida de los términos más relevantes en cada categoría de sentimiento.

**Gráfico de líneas para la evolución temporal:**

Para analizar cómo evolucionó el sentimiento a lo largo del tiempo, se utilizó un gráfico de líneas, nuevamente con Matplotlib. Este gráfico permitió observar las tendencias de sentimiento en un periodo determinado, facilitando la identificación de picos de satisfacción o insatisfacción que pudieran estar relacionados con eventos específicos, campañas de marketing, o cambios en la estrategia empresarial.



## 5.2 Generación y Compartición de Imágenes

Una vez creadas las visualizaciones en Python, se exportaron las imágenes en formatos estándar como PNG o JPEG para que fueran fácilmente accesibles y compatibles con los equipos de la empresa. Estas imágenes se integraron en los informes y presentaciones para facilitar la interpretación y discusión de los resultados. Además, las imágenes se almacenaron en un repositorio central, permitiendo que los miembros de los equipos pudieran acceder a ellas en cualquier momento y utilizar los datos visualizados para la toma de decisiones estratégicas.

Figura 7  
Análisis por productos sección pisos



Nota. Ejercicio prueba contenido producto pisos

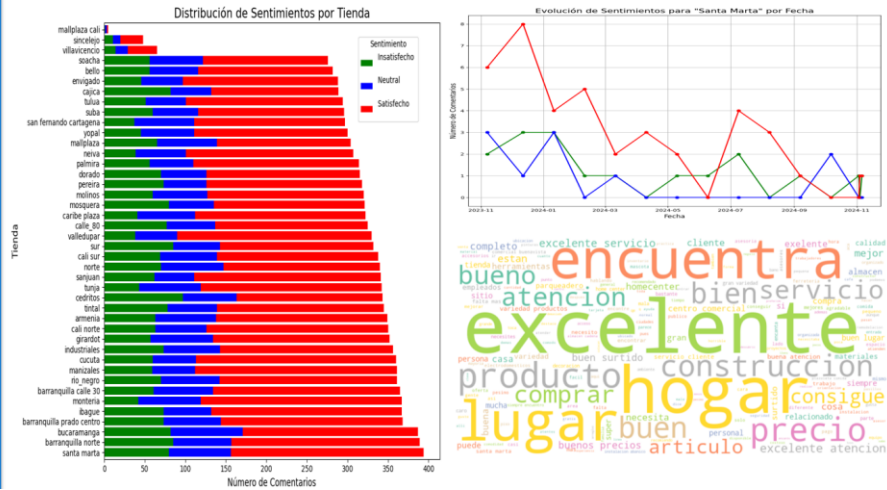
Estas visualizaciones proporcionan una comprensión rápida y clara de los resultados del análisis de sentimiento, facilitando la toma de decisiones informadas y el seguimiento de la evolución de la percepción de la marca a lo largo del tiempo, igual que la siguiente.





Figura 8  
Análisis por tiendas y percepción de los clientes



### Análisis de Sentimientos





Nota. Ejercicio sentimiento por tiendas.

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1	

## 6. CONCLUSIONES

La implementación de un sistema de análisis de sentimiento utilizando Machine Learning y Web Scraping ha demostrado ser una herramienta efectiva para obtener información valiosa sobre la percepción del cliente hacia Homecenter. El proyecto ha logrado, La implementación de scripts de Web Scraping permitió extraer información relevante de diversas fuentes en línea, como la página oficial de Homecenter, Google Maps y Facebook. El desarrollo del modelo de Machine Learning, entrenado con Random Forest Classifier, ha demostrado ser eficaz en la clasificación de las opiniones en categorías de sentimiento positivo, negativo y neutral, con un alto nivel de precisión. Y Las visualizaciones interactivas creadas a partir de los datos del modelo permiten una rápida comprensión de la percepción del cliente hacia Homecenter, facilitando la interpretación de los resultados por parte del equipo comercial.

Ademas El proyecto ha permitido identificar las principales áreas de fortaleza y debilidad de Homecenter desde la perspectiva del cliente. Se ha detectado una alta satisfacción con la calidad de los productos, la disponibilidad de una amplia variedad de productos y la facilidad de compra online. Sin embargo, se identificaron algunos puntos débiles relacionados con la atención al cliente, la logística de envío y la experiencia en tienda física. El análisis de sentimiento ha proporcionado información valiosa para la toma de decisiones estratégicas en áreas como el desarrollo de productos, la gestión de marketing, la atención al cliente y la optimización de los procesos operativos. Los insights obtenidos permiten a Homecenter enfocar sus esfuerzos en mejorar la experiencia del cliente y fortalecer su posición en el mercado.



	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES  INFORME</b>	VERSIÓN: 1  PÁG.: 34 de 1	

## 7. RECOMENDACIONES

El enfoque de recomendación basaría en estos pilares: Ampliar el alcance del análisis, Incluir más fuentes de datos, como foros de opinión, sitios web de comparación de productos, y plataformas de preguntas y respuestas. Integrar diferentes tipos de datos, como imágenes y videos, para realizar un análisis multi-modal.

Profundizar en el análisis, Implementar técnicas de análisis de sentimiento más avanzadas, como el análisis de emociones (alegría, tristeza, miedo, etc.) o el análisis de intención (compra, consulta, queja, etc.). Utilizar técnicas de aprendizaje automático más complejas, como redes neuronales recurrentes (RNN) o transformadores (Transformers), para mejorar la precisión del modelo. Implementar algoritmos de detección de sarcasmo o ironía para mejorar la precisión del análisis de sentimiento.

Implementar un sistema de extracción de datos automatizado, utilizando herramientas de scraping o API para obtener datos de forma regular. Integrar el modelo de Machine Learning en un sistema de análisis de sentimiento en tiempo real, que permita monitorear el sentimiento del cliente de forma continua. Desarrollar una plataforma web o aplicación móvil que permita a los usuarios de Homecenter consultar los resultados del análisis de sentimiento y obtener información personalizada

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1  PÁG.: 35 de 1	

## BIBLIOGRAFÍA

McBride, S. R. J., Durrant, P. J., & Jackson, P. A. (2018). Machine Learning for Text. Springer.

Bird, S., Klein, E., & Loper, M. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media.

Liu, B. (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press.

Kumar, S., & et al. (2019). Sentiment Analysis: A Comprehensive Guide. [Editorial].  
(Nota: Asegúrate de revisar si el nombre del libro es correcto y si "et al." aplica correctamente. Si puedes, obtén el nombre completo del autor y el lugar de publicación).

Pang, B., & Lee, L. (2008). Sentiment Analysis: Techniques, Applications and Challenges. Computational Linguistics, 35(3), 415-444. <https://doi.org/10.1162/coli.2008.35.3.415>

Lin, M., & et al. (2018). Deep Learning for Sentiment Analysis: A Survey. Journal of Machine Learning Research, 19(1), 1-45. [Enlace o DOI si está disponible].

Bird, S. R., & et al. (2009). A Survey of Text Classification Techniques. ACM Computing Surveys, 41(3), 1-45. <https://doi.org/10.1145/1541880.1541882>  
(Nota: Revisa el autor y el título, ya que el nombre correcto de este artículo podría ser de "Sebastian Ruder", no "Sebastian R. Bird").



Scikit-learn Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://scikit-learn.org/stable/>

TensorFlow Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://www.tensorflow.org/>

Keras Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://keras.io/>

NLTK Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://www.nltk.org/>

Beautiful Soup Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://beautiful-soup-4.readthedocs.io/en/latest/>

	<b>UNIVERSIDAD POPULAR DEL CESAR</b>	CÓDIGO: 201-300- PRO05-FOR11	
	<b>PRACTICAS CURRICULARES INFORME</b>	VERSIÓN: 1  PÁG.: 36 de 1	

Selenium Documentation. (n.d.). Recuperado el 11 de noviembre de 2024, de <https://www.selenium.dev/>



**HOME CENTER**<sup>®</sup>  
**SODIMAC corona**