

Análisis de Caso: Definición del pipeline para un proyecto End-to-End.

Escenario de Negocio: NeoSocial Boost es una agencia especializada en marketing digital y creación de contenido para marcas juveniles en TikTok e Instagram. En los últimos 12 meses ha gestionado más de 300 campañas, generando millones de visualizaciones, likes y comentarios. Sin embargo, a pesar de estos resultados en engagement, sus clientes han manifestado una baja satisfacción con las tasas de conversión (ventas, clics, formularios). Por ejemplo, campañas con 100.000 visualizaciones han generado menos de 20 acciones efectivas, lo que impacta negativamente el retorno sobre la inversión (ROI). Ante esta situación, NeoSocial Boost desea implementar un proyecto de ciencia de datos con enfoque en machine learning (ML) para mejorar el rendimiento de sus campañas y ofrecer mejores resultados a sus clientes.

Datos Disponibles: La empresa ha recopilado una base de datos estructurada con más de 40.000 publicaciones de campañas anteriores. A continuación se describen algunas variables clave:

Variables del contenido	Datos del influencer	Métricas de interacción	Variables adicionales
<ul style="list-style-type: none"> Duración del video (segundos) Estilo del video (vlog, efectos, transiciones, etc.) Texto en pantalla (sí/no) Hashtags (hasta 5) Tipo de música (viral, original, sin música) Día y hora de publicación 	<ul style="list-style-type: none"> Tipo de influencer: Nano (0–10K), Micro (10K–100K), Macro (100K+) Industria (belleza, moda, tecnología, etc.) Tasa de engagement promedio (%) 	<ul style="list-style-type: none"> Visualizaciones Tiempo medio de reproducción Porcentaje de visualización completa Likes, comentarios, compartidos, guardados Clics al perfil y al enlace de conversión Conversión realizada (1 = sí, 0 = no) 	<ul style="list-style-type: none"> Tipo de campaña (orgánica o pagada) Segmento objetivo: GenZ (18–25), Millennials (26–35) País de audiencia: Panamá, México, Colombia

Objetivo General del Proyecto: NeoSocial Boost desea comprender mejor los factores que influyen en la conversión de sus campañas y tomar decisiones más informadas a partir del análisis de datos. El objetivo es identificar oportunidades de mejora mediante el uso de ciencia de datos y explorar si el uso de modelos de machine learning puede mejorar la eficacia de sus acciones.

Instrucciones: Redacta una propuesta que desarrolle la Fase I del modelo CRISP-DM: Comprensión del Negocio. Para ello y utilizando tus propias palabras, debes abordar y estructurar tu respuesta en los siguientes cuatro elementos:

- Objetivos Empresariales:** ¿Qué desea lograr NeoSocial Boost desde la perspectiva del negocio?
- Análisis de la Situación Actual:** ¿Qué capacidades, limitaciones, datos y herramientas tiene actualmente la empresa? ¿Qué desafíos enfrenta?
- Objetivos de Minería de Datos:** Desde una perspectiva analítica y de machine learning:
 - ¿Qué tipo de problema de datos se puede modelar?
 - ¿Qué posibles soluciones o aplicaciones de ML podrían considerar?
- Planificación del Proyecto**

Propón una planificación preliminar que incluya:

 - Duración estimada
 - Recursos humanos y técnicos necesarios
 - Posibles riesgos técnicos o de implementación