





Anteproyecto de Machine Learning

Introducción






Imaginen que son consultores contratados por una empresa que necesita tomar mejores decisiones basadas en datos. Su trabajo NO es programar ni analizar datos, sino **diseñar la estrategia completa**: identificar dónde puede ayudar el machine learning, planear cómo conseguir los datos correctos, y explicar qué valor generaría para el negocio.

Este ejercicio simula lo que hacen consultores estratégicos antes de que cualquier científico de datos escriba código: pensar críticamente sobre **si vale la pena** hacer un proyecto de ML, **cómo** se debería hacer, y **por qué** generaría valor.

Lo que NO harán:

-  Programar en Python/R
-  Descargar y analizar datasets
-  Calcular métricas estadísticas
-  Generar modelos o resultados

Lo que SÍ harán:

-  Diagnosticar un problema de negocio real
-  Diseñar una estrategia para evaluar y conseguir datos
-  Justificar qué tipo de ML usar y por qué
-  Explicar qué valor concreto se generaría
-  Planear cómo se implementaría paso a paso

Rol: Consultor estratégico presentando una propuesta a un comité ejecutivo que decidirá si invertir en su proyecto.

Entregable: Documento de 6-8 páginas que convenza a ejecutivos (que NO son técnicos) de que su proyecto es valioso y viable.

Estructura del Anteproyecto

PARTE 1: El Problema y la Oportunidad (1.5 páginas)

¿Qué deben explicar?

a) Contexto de negocio

- Industria específica (retail, salud, finanzas, logística, etc.)

- Tamaño y tipo de organización (startup, corporativo, pymes)
- Descripción del problema actual en lenguaje simple

b) El dolor de negocio

- ¿Qué decisión se toma mal o con información incompleta hoy?
- ¿Quién sufre este problema? (clientes, empleados, la empresa)
- ¿Cuánto está costando este problema? (pérdidas, ineficiencias, oportunidades perdidas)

c) La oportunidad

- ¿Qué cambiaría si se resolviera?
- ¿Por qué resolverlo AHORA? (urgencia, competencia, nuevas posibilidades)

Preguntas guía para investigar:

- ¿Hay empresas similares que ya resolvieron esto? ¿Cómo?
- ¿Qué pasa si no se hace nada?
- ¿Este problema es específico de una empresa o común en la industria?

Ejemplo compacto:

Una cadena de farmacias con 80 locales enfrenta pérdidas del 18% en medicamentos por vencimiento de inventario. Los gerentes de tienda ordenan manualmente basándose en ventas del mes anterior, sin considerar estacionalidad (gripes en invierno) ni eventos locales. Resultado: sobrestocks de algunos productos y quiebres de otros que sí se venden. El problema cuesta ~\$450K anuales y genera quejas de clientes que no encuentran lo que buscan.



PARTE 2: Estrategia Analítica (2 páginas)

¿Qué deben diseñar?

a) Pregunta de investigación clara Debe ser específica y responder a una decisión concreta.

Formato: ¿Qué/Cuál/Cómo + [fenómeno medible] + [contexto específico]?

Ejemplos:

-  Mal: "¿Cómo mejorar las ventas?"
-  Bien: "¿Qué factores (precio, ubicación, temporada, promoción) predicen la demanda semanal de cada categoría de producto?"

b) Selección y justificación del modelo ML

Escoger UNA técnica y explicar POR QUÉ es la correcta para este problema:

OPCIÓN A: Regresión Lineal

- Úsala cuando necesites predecir una CANTIDAD (ventas, precio, tiempo, probabilidad)
- Ventaja clave: puedes decir "si cambio X, ¿cuánto cambia Y?"
- Útil para simular escenarios

OPCIÓN B: Clustering (K-Means)

- Úsalo cuando necesites DESCUBRIR GRUPOS sin saber cuáles son
- Ventaja clave: segmentación automática para personalizar estrategias
- Útil para entender diversidad oculta en clientes/productos

Justificación requerida: Explicar en lenguaje de negocios (NO técnico) por qué eligieron esa técnica:

- ¿Qué tipo de decisión facilita?
- ¿Por qué no la otra opción?
- ¿Qué podrá hacer la empresa con los resultados?

c) Variables clave a usar

Listar 5-8 variables y explicar POR QUÉ cada una importa para el negocio.

Formato de tabla simple:

Variable	¿Por qué importa para el negocio?	¿Se puede medir hoy?
Días_desde_última_compra	Señal temprana de que cliente se está alejando	Sí, en sistema CRM
Temporada_del_año	Productos de invierno vs verano tienen patrones distintos	Sí, fecha de venta
Promoción_activa	Descuentos aumentan demanda, hay que prever inventario	Sí, calendario marketing

Criterio: Incluir solo variables que:

1. Tengan lógica de negocio clara
2. Sean posibles de medir (aunque hoy no las tengan)
3. Puedan influenciarse o usarse para decisiones

d) Estrategia para conseguir datos

NO buscar datasets todavía. En cambio, explicar:

¿Qué tipos de fuentes podrían tener estos datos?

- Sistemas internos (ERP, CRM, punto de venta)
- Datos públicos (gobierno, investigaciones académicas)
- APIs de terceros (redes sociales, clima, tendencias)
- Datasets de ejemplo para aprendizaje (Kaggle, UCI)

¿Cómo evaluarían si un dataset es bueno? Definir sus propios criterios. Ejemplos:

- ¿Tiene suficientes casos? (¿cuántos es "suficiente"?)
- ¿Incluye las variables que necesitamos?
- ¿Es reciente o está desactualizado?
- ¿Representa bien la realidad o tiene sesgos?

¿Qué harían si no encuentran datos perfectos?

- Usar datos aproximados (proxy)
- Ajustar la pregunta de investigación
- Hacer piloto pequeño para generar datos propios
- Documentar limitación y explicar que resultados son preliminares

PARTE 3: Valor para el Negocio (1.5 páginas)

¿Qué deben demostrar?

a) Casos de uso concretos (3 mínimo)

Para cada caso, responder:

¿Cómo se toma esa decisión HOY?

- Proceso actual, quién decide, con qué información

¿Cómo cambiaría CON el modelo?

- Qué información nueva tendrían
- Qué decisión diferente tomarían

¿Qué valor genera ese cambio?

- Estimación razonable de impacto (no tiene que ser exacta, pero sí justificada)

Ejemplo:

Caso de uso: Reordenamiento de inventario en farmacias

Hoy: Gerente ordena basándose en "lo que vendí el mes pasado". Si vendió 50 antigripales en marzo, pide 50 para abril. No considera que las gripes bajan en primavera.

Con modelo: Predicción semanal por categoría considerando temporada, promociones y clima. Sistema sugiere "ordenar 20 antigripales, pero 80 protectores solares".

Valor estimado: Reducción conservadora del 30% en vencimientos (\$135K/año) + captura del 50% de ventas perdidas por quiebres (\$90K/año) = ~\$225K impacto anual.

b) Impacto agregado

Consolidar el valor total:

- ¿Cuánto dinero se ahorraría o ganaría?
- ¿Qué procesos serían más eficientes?
- ¿Cómo mejoraría la experiencia de clientes/empleados?
- ¿Hay beneficios intangibles? (aprendizaje, cultura de datos, ventaja competitiva)

c) Responder la pregunta crítica: "Si yo soy el CFO, ¿por qué debería darte presupuesto para esto en vez de para marketing, ventas o mejorar el local?"

PARTE 4: Plan de Implementación (1.5 páginas)

¿Qué deben planear?

Hoja de ruta en 4-5 fases:

Cada fase debe tener:

- Nombre y objetivo claro
- ¿Qué se hace en esta fase? (3-5 actividades clave)
- ¿Cuánto tiempo tomaría?
- ¿Qué se entrega al final?
- ¿Cómo saben si pueden avanzar a la siguiente fase?

Estructura sugerida:

Fase 1: Validación Confirmar que el proyecto tiene sentido antes de invertir mucho tiempo.

- Verificar que los datos existen o se pueden conseguir
- Hablar con usuarios: ¿realmente necesitan esto?
- Identificar quién aprobaría cambios basados en el modelo
- *Entregable:* Decisión Go/No-Go

Fase 2: Diseño Definir CÓMO se haría el proyecto.

- Establecer criterios de éxito: ¿qué tan preciso debe ser el modelo?
- Diseñar cómo se mostrarían los resultados (dashboard, reporte)
- Definir reglas: si el modelo dice X, ¿qué se hace?
- *Entregable:* Plan metodológico aprobado

Fase 3: Construcción Desarrollar el modelo (aquí trabajaría un data scientist).

- Limpiar y preparar datos

- Entrenar y probar el modelo
- Validar que funciona con casos reales
- *Entregable*: Prototipo funcional

Fase 4: Implementación Integrar el modelo en el día a día.

- Capacitar usuarios sobre cómo usar los resultados
- Modificar procesos de trabajo
- Monitorear que se está usando realmente
- *Entregable*: Modelo en operación

Fase 5 (opcional): Mejora continua Asegurar que sigue siendo útil en el tiempo.

- Actualizar el modelo periódicamente
- Recoger feedback de usuarios
- Expandir a otras áreas si funciona
- *Entregable*: Plan de sostenibilidad

Diagrama simple: Flujo visual mostrando las fases con tiempos estimados.

PARTE 5: Riesgos y Decisión Final (1 página)

¿Qué deben anticipar?

a) Principales riesgos

Identificar 5-6 cosas que podrían salir mal:

¿Qué podría salir mal?	¿Qué tan probable?	¿Qué haríamos?
No encontramos datos de buena calidad	Alta	Usar datos simulados para demostrar metodología
El modelo no es lo suficientemente preciso	Media	Definir mínimo aceptable ANTES de empezar
Los usuarios no confían en el modelo	Alta	Involucrarlos desde el diseño, mostrar cómo funciona
Los hallazgos son obvios (no aportan nada nuevo)	Media	Validar con expertos en fase temprana
El sponsor ejecutivo pierde interés	Media	Construir red de apoyo con varios stakeholders

b) Criterios de éxito

¿Cómo sabrán si el proyecto funcionó?

Éxito mínimo:

- El modelo cumple criterio técnico básico (definirlo)
- Al menos 3 decisiones se tomaron usando el modelo
- Usuarios reportan que es útil

Éxito real:

- KPI principal mejora (especificar cuál y cuánto)
- Más de 70% de usuarios target lo usan regularmente
- El modelo sigue usándose después de 6 meses

Éxito extraordinario:

- ROI positivo en menos de 12 meses
- Otras áreas piden proyectos similares
- La empresa desarrolla cultura de decisiones basadas en datos

c) Recomendación final

Declaración clara: ¿Deberían hacer este proyecto? ¿Por qué sí o por qué no?

Incluir:

- Inversión estimada necesaria (recursos, tiempo, herramientas)
 - Retorno esperado
 - Nivel de riesgo (bajo/medio/alto)
 - Próximo paso concreto
-

Formato de Entrega

Documento de 6-8 páginas:

1. PORTADA

- Título del proyecto (orientado a valor, no a técnica)
- Nombres de estudiantes

2. RESUMEN EJECUTIVO (1/2 página)

- Problema en 2 frases
- Solución propuesta en 2 frases
- Valor esperado en 1 frase
- Recomendación en 1 frase

3. DESARROLLO (5-7 páginas)

- Parte 1: El Problema y la Oportunidad

- Parte 2: Estrategia Analítica
- Parte 3: Valor para el Negocio
- Parte 4: Plan de Implementación
- Parte 5: Riesgos y Decisión Final

4. REFERENCIAS

- Papers, artículos, casos de estudio consultados
-

Criterios de Evaluación

Lo que se evaluará:

- ✓ **Claridad del problema:** ¿Se entiende por qué esto importa?
- ✓ **Lógica de la estrategia:** ¿La técnica elegida tiene sentido para ese problema?
- ✓ **Realismo:** ¿El plan es viable o están prometiendo cosas imposibles?
- ✓ **Conexión con valor:** ¿Queda claro QUÉ se ganaría y POR QUÉ?
- ✓ **Pensamiento crítico:** ¿Anticiparon problemas? ¿Tienen plan B?
- ✓ **Comunicación:** ¿Un ejecutivo sin formación técnica lo entendería?

Lo que NO se evaluará:

- ✗ Complejidad técnica (no necesitan el modelo más sofisticado)
 - ✗ Precisión de estimaciones financieras (está bien aproximar razonablemente)
 - ✗ Datasets perfectos encontrados (es sobre la ESTRATEGIA para buscarlos)
-

Preguntas para Reflexionar

Antes de empezar:

1. ¿Por qué este problema necesita ML y no se resuelve con Excel + sentido común?
2. ¿Quién pagaría por implementar esto?

Durante el trabajo: 3. Si tuviera 3 minutos en un elevador con el CEO, ¿cómo explicaría esto? 4. ¿Qué pasa si el modelo se equivoca? ¿Es grave o tolerable?

Al terminar: 5. ¿Yo invertiría MI dinero en este proyecto? ¿Por qué? 6. ¿Qué aprendí sobre conectar análisis con decisiones de negocio?

Ejemplos de Temas (Solo como Inspiración)

Regresión Lineal:

- Predicción de ventas de una cadena retail por temporada y promoción
- Estimación de precios de apartamentos según zona, tamaño y amenidades
- Pronóstico de demanda de taxis/Uber según hora, clima y eventos

Clustering:

- Segmentación de clientes de banco para personalizar productos
- Tipologías de restaurantes según menú, precio y ubicación
- Perfiles de usuarios de app de fitness por hábitos de ejercicio

Consejo: Elijan algo que les interese o conozcan. Es más fácil pensar estratégicamente sobre algo que entienden.

Pregunta final para autoevaluarse:

"Si un inversionista me da 5 minutos para convencerlo de financiar este proyecto con \$50,000, ¿mi documento me da los argumentos para lograrlo?"

Si la respuesta es sí, están listos. Si es no, revisen qué falta.