Checklist General - Fase 0: Definición del Problema y Comprensión del Contexto

•	echa:
• ***!	
• *** [
• ***	
• *** [
	esponsable ***:
I. Cont	exto General y Objetivo del Proyecto 🎯
• 🗆 P	regunta Principal: ¿Cuál es el objetivo general que impulsa este análisis? Intender un fenómeno, optimizar un proceso, hacer predicciones, descubrir patrones, etc.)

	del Dominio: ¿0 salud, medio a				el proyecto?	
, educacion,	saluu, meulo a	imblence, mac	istria, techolo	igia, ecc.)		
Interesado álisis?	s clave (Stake	holders) : ¿Qu	iénes se verái	n beneficiado	os o tomarán	decisiones con

Resultados : ¿Para olíticas, desarrollo		stigación cient	ífica, etc
: tuales : ¿Cómo se manuales, sistema			

Definición del Problema y Alcance 🌋							
	ta Analítica Esp ⁄ariables explica				remos respo	nder con los	datos?
Tipo de	· Análisis Espera	udo:					
	otivo, Diagnóstico		Prescriptivo	o, Clasificaci	ón, Clusterin	g, Series de t	:iempo,

ntregables espe nforme PDF, dasl		no interactivo	o, modelo en	trenado, API,	artículo cient	ífico)
lcance del Proye	ecto (In/Out): ¿	Qué incluye <u>:</u>	y qué no incl	uye este anál	isis?	

ato	os Requeridos 💾
	Fuentes de Datos: ¿Qué fuentes usarás?
(EJ:	bases SQL, archivos planos, sensores, APIs, encuestas, datos abiertos, etc.)
	Disponibilidad y Acceso: ¿Están disponibles los datos? ¿Hay restricciones de acceso?

alidad de los Datos: ¿Los datos tienen problemas comunes? alores faltantes, duplicados, errores, inconsistencias, desbalance, ruido, etc.)						
	s Éticas y de Pri pedir permisos,					
·	· · ·	•				

Indicado elevante?	res de Impacto (KPIs	o Métricas de	ا Contexto): زCo	ómo sabrás si el p	oyecto fue úti
	Técnicas: ¿Cómo eval , recall, F1, AUC, error			análisis?	

	e Comparación (Baseline): ¿Existe algún valor o método de referencia que se deb
superar?	
: C:	:
anırıcac	ión y Logística 🛗
— D • • • • • • • • • • • • • • • • • • •	sabilidades del Equipo: ¿Quién se encarga de qué fase o componente del proyecto
Respons	
Respons	
Respons	

Pacursos Dispon	ibles : ¿Qué herra	amientas soft	rware infraes	tructura v dato	s sa tianan?
vecursos Dispon	ibles. ¿Que nerre	illielicas, sorc	.ware, iiii aes	cruccura y dato	s se tiellell:
Cronograma y Hi	itos : ¿Cuáles son	las etapas cla	ve y sus fecha	s límite?	

	os Potenciales: ¿Que incompletos, cambie		
Refere	Opcionales) encias o recursos re s, artículos, docume	enlaces web)	

	versiones prev fue intentado a		to	
-		•		

Modelo	Tipo de tarea	¿Para qué se usa?	Características clave
Linear Regression	Regresión	Predecir valores numéricos continuos (precio, temperatura, etc.)	Modelo simple, fácil de interpretar
Logistic Regression	Clasificación binaria	Clasificación entre dos clases (sí/no, 0/1, aprobado/no aprobado)	Rápido, base para clasificación
Decision Tree	Regresión/Clasificación	Tareas con decisiones basadas en condiciones (reglas tipo árbol)	Fácil de visualizar, puede sobreajustar
Random Forest	Regresión/Clasificación	Tareas generales, clasificación multiclase, regresión	Robusto, combina muchos árboles, reduce overfitting
Gradient Boosting (XGBoost, LightGBM, etc.)	Regresión/Clasificación	Problemas complejos con alto rendimiento	Muy potente, tuning delicado, puede tardar más en entrenar

Modelo	Tipo de tarea	¿Para qué se usa?	Características clave
Support Vector Machines (SVM)	Clasificación/Regresión	Clasificación con márgenes máximos, separación no lineal	Bueno para datasets pequeños, puede ser lento en grandes
K-Nearest Neighbors (KNN)	Clasificación/Regresión	Basado en vecinos más cercanos (proximidad)	Fácil de entender, costoso en predicción
Naive Bayes	Clasificación	Clasificación de textos, spam, sentimiento, etc.	Rápido, asume independencia entre variables
DummyClassifier / DummyRegressor	Baseline (referencia)	Comparar con modelos tontos (estrategias simples)	No aprende, útil como base de comparación
Linear Discriminant Analysis (LDA)	Clasificación	Clasificación con reducción de dimensionalidad	Bueno con datos linealmente separables
Quadratic Discriminant Analysis (QDA)	Clasificación	Igual que LDA, pero para datos con covarianzas diferentes entre clases	Más flexible, pero más complejo
Stochastic Gradient Descent (SGD)	Clasificación/Regresión	Casos con grandes volúmenes de datos	Ligero, útil en streaming y textos
Multilayer Perceptron (MLP - Redes neuronales)	Clasificación/Regresión	Problemas no lineales complejos	Aprende patrones no lineales, requiere más datos
K-Means	Clustering (no supervisado)	Agrupar datos similares sin etiquetas	Fácil de usar, no requiere etiquetas
DBSCAN	Clustering (no supervisado)	Agrupar con formas arbitrarias y detectar outliers	Útil para detectar ruido, no requiere especificar K
Hierarchical Clustering	Clustering (no supervisado)	Agrupación por jerarquía (árbol de similitud)	Visualización con dendrogramas
PCA (Análisis de Componentes Principales)	Reducción de dimensionalidad	Visualización, eliminar ruido, acelerar algoritmos	Proyecta los datos a un nuevo espacio
t-SNE / UMAP	Reducción de dimensionalidad	Visualización de datos en 2D o 3D	Útiles en visualización de clusters

Modelo	Tipo de tarea	¿Para qué se usa?	Características clave
ARIMA / SARIMA	Series temporales	Predicción de datos secuenciales con tendencias y estacionalidad	Requiere datos ordenados temporalmente
Facebook Prophet	Series temporales	Predicción de series de tiempo con eventos y estacionalidad	Fácil de usar, útil para negocios
LSTM (Redes recurrentes)	Series temporales / Secuencias	Modelado de secuencias complejas (texto, tiempo, audio)	Potente, usado en deep learning
Autoencoder	Reducción de dimensión / Detección de anomalías	Aprender representación comprimida de datos	Parte de redes neuronales