

OWNML MACHINE LEARNING CANVAS

Designed for: Alejandro Calderon

Designed by: Equipo 10

Date: 11/10/25

Iteration: 1

PREDICTION TASK	DECISIONS	VALUE PROPOSITION	DATA COLLECTION	DATA SOURCES
<p>Tipo de tarea: Regresión supervisada.</p> <p>Entidad: Cada registro horario del sistema de bicicletas (día y hora).</p> <p>Resultado a predecir: Número total de bicicletas rentadas (cnt).</p> <p>Cuándo se observa el outcome: Al cierre de cada hora o día (según granularidad elegida).</p>	<ul style="list-style-type: none"> Cómo se usan las predicciones: <p>Las predicciones alimentan un tablero de operaciones para planificar la reubicación de bicicletas y optimizar la disponibilidad por estación o zona.</p> <p>Si se predice baja demanda → reducir bicicletas disponibles.</p> <p>Si se predice alta demanda → enviar unidades extra o extender horarios.</p> <ul style="list-style-type: none"> Parámetros del proceso: Umbral de ocupación, clima, hora pico, día laboral. 	<ul style="list-style-type: none"> Beneficiarios: <ol style="list-style-type: none"> Equipo de operaciones (mejor logística y menor costo de transporte). Usuarios finales (mayor disponibilidad y menos tiempos de espera). Pain points: <ol style="list-style-type: none"> Sobrecarga o escasez de bicicletas. Costos altos por reubicación manual. Integración / interfaz: Dashboard web con visualización de demanda esperada por hora o zona; alertas automáticas para operaciones. 	<ul style="list-style-type: none"> Origen inicial: Dataset público (UCI). Método: Descarga de CSV. Actualización: En un escenario real, mediante API del sistema de bicicletas o IoT de estaciones. Estrategia de frescura: Actualización diaria con pipeline automatizado. Costo: Bajo (datos abiertos). 	<ul style="list-style-type: none"> Internos: Base de datos de uso histórico del sistema (si se operara real). Externos: No Aplica Formato: CSV o tabular
IMPACT SIMULATION	MAKING PREDICTIONS	BUILDING MODELS	FEATURES	MONITORING
<p>1. Costo/beneficio:</p> <ul style="list-style-type: none"> Falsa predicción de alta demanda → sobrecosto de operación. Falsa predicción de baja demanda → pérdida de ingresos y satisfacción. <p>2. Simulación pre-despliegue: Backtesting sobre datos 2011–2012.</p> <p>3. Criterios de despliegue: MAE ≤ 45 y RMSE ≤ 70 en validación temporal.</p> <p>4. Fairness constraints: Evitar sesgos por temporada o clima extremo (balance de muestras).</p>	<ul style="list-style-type: none"> Modo: Batch (por día). Frecuencia: Cada día. Tiempo máximo: 5 min por batch diario. Recursos: dvc, kubeflow, bucket 	<ul style="list-style-type: none"> Número de modelos: 1 principal (regresión de demanda). Actualización: Semanal o mensual (según drift detectado). Tiempo de entrenamiento: <30 min por pipeline. Recursos: Kubeflow, DVC, sklearn + LightGBM. 	<ul style="list-style-type: none"> Representaciones: Variables temporales, meteorológicas y de contexto. Transformaciones: <ul style="list-style-type: none"> Normalización [0–1] One-hot encoding (season, weathersit, weekday) Features de rezago (cnt_lag_1h, cnt_lag_24h) Variables derivadas (is_weekend, is_peak_hour) 	<p>Métricas técnicas: RMSE, MAE, MAPE, drift (PSI/KS).</p> <p>KPIs de negocio:</p> <ul style="list-style-type: none"> Reducción del 15–20% en redistribuciones ineficientes. Incremento de disponibilidad >95%. Revisión: Semanal (métricas técnicas) y mensual (negocio). <p>Herramientas: MLflow.</p>

