



# Tecnológico de Monterrey

Inteligencia artificial avanzada para la ciencia de datos 2

Gpo 501

## **Docentes**

Dr. Benjamín Valdés Aguirre

Ma. Eduardo Daniel Juárez Pineda

Dr. Ismael Solis Moreno

Dr. José Antonio Cantoral Ceballos

Dr. Carlos Alberto Dorantes Dosamantes

## **Integrantes**

Carlos Rodrigo Salguero Alcántara	A00833341
Diego Perdomo Salcedo	A01709150
Dafne Fernández Hernández	A01369230
José Emiliano Riosmena Castañón	A01704245
Luis Arturo Rendón Iñarritu	A01703572

Querétaro, Querétaro

<b>1. Contexto y Justificación del Proyecto</b>	<b>2</b>
<b>2. Objetivos de negocio y de ciencia de datos</b>	<b>2</b>
2.1 Criterios de éxito	2
2.2 Objetivo de minería de datos	2
2.3 Criterios de éxito	2
<b>3. Evaluación de la situación</b>	<b>2</b>
3.1 Recursos de datos	2
3.2 Recursos computacionales	3
3.3 Infraestructura existente	3
3.4 Profesores / Expertos	3
<b>4. Requerimientos del proyecto</b>	<b>4</b>
4.1 Requerimientos de Negocio	4
4.2 Requerimientos de Ciencia de Datos	4
4.3 Restricciones	4
4.4 Requerimientos de los datos	4
4.5 Supuestos	5
<b>5. Riesgos y Mitigaciones</b>	<b>5</b>
5.1 Análisis de Costo-Beneficio	9
5.2 Apéndice A: Terminología	9
<b>6. Análisis preliminar y exploración de datos</b>	<b>10</b>
6.1 Análisis de Imágenes	10
6.2 Análisis de datos complementarios	10
<b>7. Plan de Proyecto</b>	<b>11</b>

# 1. Contexto y Justificación del Proyecto

El CAETEC (Campo Agro Experimental del Tec de Monterrey) es un laboratorio para el aprendizaje de actividades agropecuarias. Este cuenta con un establo con corrales destinados a las vacas que habitan esta área. Las vacas tienen varias actividades que llegan a realizar a lo largo de un día tales como; socializar con otras vacas, comer, dormir en sus camas, y ser ordeñadas automáticamente por un máquina. En esta área del CAETEC una de sus prioridades es mantener a las vacas en un nivel alto de comodidad. Esto es porque mientras más cómodas se encuentren tienden a rendir una mayor cantidad y calidad de leche.

El sector lácteo enfrenta desafíos continuos en la optimización de sus procesos de ordeño. La eficiencia en este proceso no solo afecta la productividad de la granja, sino también el bienestar animal y la calidad de la leche. Este proyecto de ciencia de datos busca abordar estas cuestiones mediante el análisis avanzado de datos y la implementación de técnicas de visión por computadora.

## 2. Objetivos de negocio y de ciencia de datos

Identificar el número de vacas en cada fila en un periodo de tiempo determinado.

### 2.1 Criterios de éxito

- ❖ Determinar con alta precisión la cantidad de vacas en una imagen. Arturo o Ivo determinarán si la precisión es satisfactoria.

### 2.2 Objetivo de minería de datos

Determinar la cantidad de vacas en cada imagen en cualquier condición.

### 2.3 Criterios de éxito

- ❖ Un modelo para condiciones diurnas con un 80% de precisión.
- ❖ Un modelo para condiciones nocturnas con un 50% de precisión

## 3. Evaluación de la situación

### 3.1 Recursos de datos

- ❖ Dataset con 8,115 imágenes de la fila de ordeño.

### 3.2 Recursos computacionales

Marca	Modelo	CPU	GPU	RAM
Apple	Macbook Pro 14	Apple M3	Apple M3	36
Razer	Blade 15	Intel i7	Nvidia RTX 3060 MAX-Q	64
Razer	Blade 18	Intel i9	Nvidia RTX 4070 MAX-Q	32
Asus	TUF17	Intel i5	Nvidia RTX 3050	32
HP	Spectre x360	Intel i5	Intel(R) Iris(R) Xe Graphics	8
ASUS	Vivobook 15	Ryzen 7	AMD Radeon(TM) Graphics	16
cisco	100 megabits	-	-	-

### 3.3 Infraestructura existente

- ❖ Webcam LOGITECH C920 en el área de ordeño.
- ❖ Raspberry Pi para la captura y transmisión de imágenes.
- ❖ Sistema de almacenamiento de OneDrive
- ❖ Conexión a Internet Telmex.

### 3.4 Profesores / Expertos

- ❖ **Arturo González de Cosío:** Director de CAETEC. Responsable de la supervisión general del proyecto, asegurando que esté alineado con los objetivos estratégicos de CAETEC.
- ❖ **Ivo Neftali Ayala García:** Ingeniero electrónico. Se encargará de la integración y mantenimiento de los sistemas de cámaras instalados en el rancho, además de un segundo supervisor general del proyecto.
- ❖ **Guadalupe López:** Especialista pecuaria. Aporta su experiencia en el manejo del ganado bovino.
- ❖ **Dr. Benjamin Valdes Aguirre:** Profesor de técnicas y arquitecturas de deep learning.
- ❖ **Dr. Ismael Solis Moreno:** Profesor de Big data y cómputo en la nube.
- ❖ **Dr. José Antonio Cantoral:** Profesor de herramientas para el procesamiento del lenguaje natural.

- ❖ **Dr. Carlos Alberto Dorantes:** Profesor de estadística avanzada para la ciencia de datos.
- ❖ **Ma. Eduardo Daniel Juárez Pineda:** Profesor de Metodologías de proyectos de ciencia de datos.

## 4. Requerimientos del proyecto

### 4.1 Requerimientos de Negocio

- ❖ Integración fluida con los sistemas de gestión de la granja existentes.
- ❖ Interfaz de usuario intuitiva para acceder a los insights y métricas clave.
- ❖ Capacidad de generar reportes automatizados sobre eficiencia del ordeño.

### 4.2 Requerimientos de Ciencia de Datos

- ❖ Desarrollo de pipelines de datos robustos para el procesamiento de imágenes.
- ❖ Implementación de técnicas de aumento de datos para mejorar la generalización del modelo.
- ❖ Uso de técnicas de interpretabilidad de modelos para entender las decisiones del sistema.

### 4.3 Restricciones

- ❖ No hay presupuesto
- ❖ Iluminación cambiante
- ❖ Lejos del CAETEC

### 4.4 Requerimientos de los datos

- ❖ Imágenes o videos de la fila de ordeño: Para entrenar el modelo de identificación de vacas.
- ❖ Datos históricos de producción de leche: Para analizar el impacto de la automatización en la producción.
- ❖ Tiempos de espera actuales: Para medir mejoras en la reducción del tiempo de espera de las vacas.
- ❖ Datos sobre el número de vacas ordeñadas por hora: Para evaluar la eficiencia del flujo de ordeño.

## 4.5 Supuestos

### ❖ Datos:

- **Calidad de Datos Consistente:** Se asume que la calidad de las imágenes tomadas (nitidez, resolución, exposición) será suficientemente buena para permitir una identificación precisa de las vacas. Esto incluye tanto imágenes diurnas como nocturnas.
- **Obtención de Datos:** Se espera que podamos solicitar una mayor cantidad de imágenes en el caso de que sea requerido por el equipo de trabajo.
- **Variabilidad de Datos:** Se supone que el dataset incluirá imágenes que reflejan todas las posibles variaciones de luz y clima en el área de ordeño, permitiendo así la generalización del modelo.
- **Estabilidad de la Infraestructura:** Asumimos que el sistema de cámaras y el hardware asociado (Raspberry Pi) no presentarán fallos significativos durante el periodo de recolección de datos y que la conexión a internet será estable.

### ❖ Negocio:

- **Impacto Positivo en la Producción:** Se considera que el proyecto contribuirá de manera significativa a mejorar la eficiencia en el proceso de ordeño, impactando en la producción de leche de manera positiva.
- **Mantenimiento Regular del Sistema:** Se asume que se realizarán mantenimientos preventivos y correctivos necesarios en el equipo de captura de imágenes y en los dispositivos de transmisión para evitar interrupciones en el flujo de datos.
- **Adopción y Escalabilidad del Sistema:** Se prevé que, de obtener resultados satisfactorios, el sistema será adoptado como parte integral de la infraestructura de CAETEC y podrá escalarse a otras áreas.

## 5. Riesgos y Mitigaciones

Riesgo	Categoría	Impacto	Probabilidad	Mitigación
Modificaciones en las imágenes del dataset completo	Datos	Alto	Bajo	Migrar de dataset y preparar los datos nuevamente
Violación de seguridad de los datos o problemas de	Datos	Alto	Bajo	Implementar medidas de seguridad robustas y

privacidad				cumplir con las regulaciones de protección de datos
Corrupción o pérdida de datos durante el almacenamiento o procesamiento	Datos	Alto	Bajo	Implementar un sistema de respaldo regulares y verificación de la integridad de datos
Interrupciones en la conexión a internet que impida la carga de imágenes a OneDrive	Datos	Alto	Medio	Implementar un sistema de almacenamiento local temporal en el Raspberry Pi para retener las imágenes hasta que puedan ser cargadas. Considerar una conexión de internet de respaldo (por ejemplo, un módem 4G) para situaciones críticas.
Desincronización entre las imágenes locales y las almacenadas en OneDrive.	Datos	Medio	Bajo	Implementar verificaciones periódicas de consistencia entre el almacenamiento local y OneDrive. Utilizar registros de carga (logs) para rastrear qué imágenes se han subido exitosamente.
Alcanzar el límite de almacenamiento de la cuenta de OneDrive.	Datos	Alto	Medio	Monitorear regularmente el uso de espacio de almacenamiento. Implementar una política de retención de datos para eliminar o archivar

				imágenes antiguas.
Nuevas normativas que afecten la recopilación o uso de datos de animales	Externos	Alto	Bajo	Mantenerse informado sobre las regulaciones y adaptar el proyecto según sea necesario
Eventos climáticos que afecten el comportamiento normal de las vacas o el funcionamiento del equipo.	Externos	Medio	Medio	Considerar las condiciones climáticas en el análisis de datos y asegurar la protección del equipo.
Falta de cooperación o resistencia al nuevo sistema por parte del personal.	Externos	Medio	Medio	Involucrar al personal desde el inicio, proporcionar capacitación y demostrar los beneficios del proyecto.
Cortes de energía que afecten la captura continua de imágenes.	Operativos	Alto	Bajo	Instalar sistemas de energía de respaldo
Falta de mantenimiento regular que afecte el rendimiento del sistema	Operativos	Medio	Medio	Establecer un calendario de mantenimiento preventivo
Modificaciones en la disposición del área de ordeño que afecten la captura de imágenes.	Operativos	Medio	Bajo	Coordinar con el personal del corral y adaptar el sistema según sea necesario
Fallas en la red donde se realice el proyecto	Operativos	Medio	Medio	Tener un dispositivo alternativo como punto de conexión alternativo
Demoras en la implementación que afecten los plazos del	Proyecto	Medio	Medio	Establecer hitos claros y realizar un seguimiento regular del progreso



proyecto				
Costos imprevistos que excedan el presupuesto planificado	Proyecto	Alto	Medio	Mantener un margen de contingencia en el presupuesto y realizar un seguimiento detallado de los gastos
Modificaciones significativas en el alcance o los objetivos del proyecto.	Proyecto	Alto	Bajo	Establecer un proceso claro para la gestión de cambios y mantener una comunicación constante con los stakeholders.
Empalme con otras entregas de los proyectos	Proyecto	Medio	Alto	Mantener una organización constante con la distribución del tiempo hacia los proyectos individuales y el reto grupal
El socio formador no muestra conformidad con el resultado del proyecto	Proyecto	Crítico	Medio	Mantener una comunicación constante entre el objetivo principal del proyecto con el socio formador, y de los avances realizados en el proyecto. Presentar al socio los posibles márgenes de error y buscar alternativas que satisfagan la necesidad del socio. Robar una vaca
Imágenes borrosas o de baja calidad que dificultan la identificación de vacas	Técnico	Alto	Medio	Ajustar la configuración de la Logitech C920 para optimizar la calidad de imagen en las condiciones

				de iluminación del corral. Considerar la adición de iluminación suplementaria si es necesario, especialmente para las horas nocturnas.
Interrupciones en la captura de imágenes cada 5 minutos.	Técnico	Alto	Medio	Configurar el Raspberry Pi para reiniciarse automáticamente en caso de fallos.

## 5.1 Análisis de Costo-Beneficio

Costo	Beneficio
<b>Tiempo de Desarrollo:</b> <ul style="list-style-type: none"> <li>Horas de trabajo: Supongamos que cada integrante del equipo (5 integrantes) trabaja 20 horas a la semana durante 10 semanas.</li> <li>Total de horas: 5 integrantes x 20 horas/semana x 10 semanas = 1,000 horas</li> </ul>	Incrementar la producción de leche al conseguir información que pueda ser utilizada en el CAETEC

## 5.2 Apéndice A: Terminología

**Detección:** Identificación de una vaca individual en la imagen

**Tiempo de espera:** Duración desde que una vaca entra en la fila hasta que sale para ser ordeñada

**Ciclo de ordeño:** Tiempo total que toma ordeñar a todas las vacas del rebaño

**Producción de leche:** La cantidad de leche que producen las vacas en una determinada cantidad de tiempo.

**Modelo de Aprendizaje Automático:** Algoritmo que aprende patrones a partir de datos y puede hacer predicciones o clasificaciones.

**Red Neuronal Convolucional (CNN):** Tipo de red neuronal especialmente diseñada para procesar datos con una estructura de cuadrícula, como imágenes, utilizando convoluciones para extraer características.

**Aumento de Datos (Data Augmentation):** Técnicas para aumentar el tamaño y la diversidad de un conjunto de datos mediante transformaciones (rotaciones, escalados, recortes, etc.) para mejorar la generalización del modelo.

**Segmentación de Imágenes:** Proceso de dividir una imagen en partes (o segmentos) más significativas para facilitar el análisis.

**Detección de Anomalías:** Identificación de datos que no siguen un patrón esperado, lo que puede indicar problemas en la calidad de los datos o en el proceso.

**Sistema de Respaldo:** Estrategias y tecnologías implementadas para asegurar la recuperación de datos en caso de fallos o pérdida de datos.

## 6. Análisis preliminar y exploración de datos

### 6.1 Análisis de Imágenes

- ❖ Distribución de la cantidad de vacas por imagen.
- ❖ Variación en condiciones de iluminación (día vs noche)
- ❖ Identificación de elementos que pueden causar falsos positivos (trabajadores, equipamientos)

### 6.2 Análisis de datos complementarios

- ❖ Correlación entre tiempos de espera y producción de leche.

## 7. Plan de Proyecto

PLAN DE PROYECTO	SEP		OCTUBRE				NOVIEMBRE			
	3	4	1	2	3	4	1	2	3	4
Entendimiento del negocio										
Entendimiento de los datos										
Preparación de los datos I										
Modelado I										
Preparación de los datos II										
Modelado II										
Evaluación										
Implementación										

1

🇬🇧 PVG - Yeehaw 🤖🐼

<sup>1</sup> Se considera que las fases pueden llegar a ocurrir más de una vez en el ciclo de vida del proyecto.