



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

## **Fase de Entendimiento de Negocio**

**Autores:**

A01368818 Joel Sánchez Olvera

A01661090 Juan Pablo Cabrera Quiroga

A01704076 Adrián Galván Díaz

A01708634 Carlos Eduardo Velasco Elenes

A01709522 Arturo Cristián Díaz López

**TC3007C.501**

Inteligencia artificial avanzada para la ciencia de datos II

**Fecha:**

9 de Octubre del 2024

# **Fase de Entendimiento de negocio**

## **1. Determinación de los objetivos de negocio**

### **Background**

El Centro de Agricultura Experimental y Tecnología (CAETEC) es un laboratorio agropecuario del Tecnológico de Monterrey, ubicado en Querétaro, que utiliza tecnologías de agricultura de precisión e interconexión con la industria. CAETEC se especializa en proyectos de innovación que van desde el monitoreo de cultivos hasta la producción eficiente de leche con robots de ordeña. Además, colaboran con empresas como John Deere y Nestlé para mejorar la sostenibilidad y reducir emisiones. Sus instalaciones incluyen invernaderos, cultivos y establos con tecnología avanzada.

En el CAETEC, se monitorea la producción de leche utilizando robots de ordeña que ofrecen beneficios para el bienestar animal, esto ocurre gracias a que cuentan con tres de los cinco robots que hay en México para ordeñar a las vacas y aunque adaptarse a este nuevo método de ordeña tomó alrededor de tres años, el hecho que las vacas vivan sin estrés ha resultado en un aumento de producción y sobre todo, permite que todo sea medible. Cada vaca tiene su rutina, y el sistema registra su producción de leche, niveles de grasa y horarios de ordeño.

### **Objetivos de Negocio**

1. Determinar la ocupación de los espacios destinados al descanso de las vacas en el rancho, identificando la población de ganado en cada zona y evaluando qué áreas están ocupando la mayor parte del tiempo.

### **Criterios de éxito**

1. Se identifica el número total de vacas dentro de un área específica en la imagen (conteo de población).
2. Se mapea la distribución de vacas en el espacio geográfico asignado para evaluar la ocupación de los espacios.
3. Los registros generados por el sistema son accesibles y disponibles en formatos comunes para la visualización y análisis de datos.
4. Se presenta el análisis de los patrones de arena en las camas, identificando si existe o no patrones que impidan el aprovechamiento eficiente del espacio.

## **2. Asesorar situación**

### **Profesores e interesados en el proyecto**

- 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.
- Dra. 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.

### **Inventario de recursos**

1. Expertos
  - a. Ivo Neftali Ayala García
  - b. Dra. Guadalupe López
  - c. Arturo González de Cosío
  - d. Dr. Benjamín Valdés Aguirre
  - e. Dr. José Antonio Cantor
  - f. Dr. Ismael Solis Moreno
  - g. Dr. Carlos Alberto Dorantes Dosamantes
  - h. Eduardo Juárez Pineda
2. Datos
  - a. 9634 imágenes JPG
  - b. Peso de 600 - 650 kb por imagen
  - c. Resolución individual de 1920 \* 1080 px
  - d. [Images](#)
3. Hardware

- a. Webcam Logitech C920
- b. Raspberry Pi 3 B+ 16 GB
- 4. Software
  - a. Python
  - b. Pandas
  - c. Numpy
  - d. Pytorch \*
  - e. Tensor Flow \*
  - f. PySpark \*

## Requerimientos Funcionales

RF01	Identificación de vacas a través de un modelo inteligente de Bounding Box
Descripción	Generar un modelo inteligente que identifique la o las vacas en un espacio dado.
Módulo	Modelo
Criterios de aceptación	<input type="checkbox"/> El modelo debe identificar la <b>Bounding Box</b> de una o múltiples vacas en un espacio dado. <input type="checkbox"/> El modelo debe obtener una métrica de al menos 90% de <b>precisión</b> tras el entrenamiento. <input type="checkbox"/> El modelo debe obtener una métrica de al menos 80% de <b>mAP 0.5</b> tras el entrenamiento.

RF02	Identificación de la clase de una vaca a través de un modelo inteligente
Descripción	Generar un modelo inteligente <b>clasificador</b> que identifique si una vaca está acostada o parada en un espacio dado.
Módulo	Modelo
Criterios de aceptación	<input type="checkbox"/> El modelo debe identificar la <b>Clase (Parada / Acostada)</b> de una vaca en un espacio dado. <input type="checkbox"/> El modelo debe tener una métrica... <input type="checkbox"/> El modelo debe tener una métrica...

RF03	Análisis de posibles patrones de arena en el área de descanso de las vacas
Descripción	Generar un análisis de las camas vacías con el fin de identificar posibles patrones de arena en las mismas.
Módulo	Modelo

<b>Criterios de aceptación</b>	<input type="checkbox"/> Se obtiene un mapa de calor que identifica posibles patrones de la arena presente en las camas. <input type="checkbox"/> Se obtienen otras gráficas relevantes que pudieran revelar patrones de la arena presente en las camas. <input type="checkbox"/> Se obtienen hallazgos significativos para el cliente, en patrones divididos y categorizados en mañana, tarde y noche.
--------------------------------	---

<b>RF04</b>	<b>Recorte de imagen de Bounding Box generada por modelo inteligente</b>
<b>Descripción</b>	Realizar un recorte de las <a href="#">bounding box</a> identificadas por el modelo inteligente del RF01.
<b>Módulo</b>	Modelo
<b>Criterios de aceptación</b>	<input type="checkbox"/> Se realiza un recorte de la bounding box definida por el modelo inteligente del RF01. <input type="checkbox"/> El recorte resultante es una imagen de formato jpg de aproximadamente 400*800 px.

<b>RF05</b>	<b>Obtención del centroide de Bounding Box</b>
<b>Descripción</b>	Obtener el centroide de las <a href="#">bounding box</a> identificadas por el modelo inteligente del RF01.
<b>Módulo</b>	Modelo
<b>Criterios de aceptación</b>	<input type="checkbox"/> Se obtienen las coordenadas xmax, xmin, ymax, ymin de una bounding box en una imagen dada. <input type="checkbox"/> Se realiza el cálculo del centroide usando las coordenadas de la bounding box. <input type="checkbox"/> Se imprime o registra el valor del centroide obtenido.

<b>RF06</b>	<b>Interconexión entre modelos RF01 y RF02</b>
<b>Descripción</b>	Realizar una interconexión entre los modelos de <a href="#">Bounding Box</a> y <b>Clasificador</b> .
<b>Módulo</b>	Modelo
<b>Criterios de aceptación</b>	<input type="checkbox"/> El sistema identifica el Bounding Box de una vaca en una imagen dada. <input type="checkbox"/> El sistema recorta la Bounding Box identificada, generando una imagen. <input type="checkbox"/> El sistema alimenta la imagen obtenida al modelo clasificador. <input type="checkbox"/> El modelo clasificador arroja label (parada / acostada).

RF07	Generación de registros en base de datos relacional
Descripción	Crear registros en una base de datos a partir de los hallazgos de los modelos RF01 y RF02.
Módulo	Modelo
Criterios de aceptación	<input type="checkbox"/> El sistema registra en:  tabla1: a. id_img_pk: Llave primaria de tabla1. b. timestamp: El timestamp en el que se tomó la imagen. c. cow_count: El número de vacas identificadas en la imagen.  tabla2: a. id_detected_cow_pk: Llave primaria de tabla2. b. id_img_fk: Llave foránea a tabla1. c. cow_centroid: El centroide de la vaca en la imagen. d. prediction_accuracy: La precisión que el modelo arrojó para la clase acostada / parada.

## Requerimientos No Funcionales

RNF01	Rendimiento
Descripción	El sistema debe realizar el procesamiento de una imagen en un tiempo no mayor a 5000 milisegundos.
Módulo	Modelo
Criterios de aceptación	<input type="checkbox"/> El sistema: a. Procesa la imagen de entrada b. Identifica las bounding boxes correspondientes a las vacas existentes en la imagen de entrada. c. Identifica las clases de las vacas existentes en las bounding boxes identificadas. d. Genera registros en la base de datos con los hallazgos.  En un tiempo no mayor a 5000 milisegundos.

RNF02	Escalabilidad
Descripción	El sistema debe ser capaz de manejar mayores volúmenes de datos a los esperados sin degradar significativamente su rendimiento.
Criterios de aceptación	<input type="checkbox"/> El sistema es capaz de procesar hasta 30 imágenes por minuto durante 5 minutos. <input type="checkbox"/> El sistema utiliza no más de un 85% del poder computacional disponible en CPU. <input type="checkbox"/> El sistema utiliza no más de un 85% de la memoria RAM disponible.

RNF03	Seguridad
Descripción	<i>Los datos e imágenes usados por el modelo son protegidos bajo las normas de privacidad y seguridad aplicables.</i>
Criterios de aceptación	<input type="checkbox"/> <i>Se excluyen del dataset usado para el entrenamiento, validación y pruebas las imágenes que contengan individuos en ellas.</i> <input type="checkbox"/> <i>Las imágenes utilizadas por el modelo están protegidas por la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP).</i>

RNF04	Mantenibilidad
Descripción	<i>El sistema debe contar con una documentación clara y concisa que permita futuras modificaciones y mantenimiento por parte de otros desarrolladores o personal técnico.</i>
Criterios de aceptación	<input type="checkbox"/> <i>El sistema cuenta con manual de usuario y manual técnico. Los manuales son validados por el cliente.</i>

## Supuestos

### 1. Disponibilidad y Calidad de Datos

- Las imágenes tienen una resolución mínima adecuada para detectar vacas individuales y camas de arena.
- Existe un sistema de respaldo para el almacenamiento de las imágenes.
- Las imágenes de las camas de arena tienen suficiente calidad y contraste para permitir el análisis de patrones en la superficie

### 2. Infraestructura Tecnológica

- Se cuenta con capacidad de procesamiento suficiente para analizar las imágenes
- La base de datos puede manejar el volumen de registros generados diariamente
- Existe una conexión a internet estable para la transmisión de datos
- Se tienen los sistemas de almacenamiento necesarios para datos históricos

### 3. Aspectos Operativos

- Las áreas de descanso están claramente delimitadas y visibles en las imágenes
- Las condiciones de iluminación permiten la captura de imágenes útiles durante todo el día
- El sistema de monitoreo opera las 24 horas del día
- No existen obstáculos significativos en el campo visual de las cámaras

### 4. Condiciones del Entorno

- Las condiciones climáticas no afectan significativamente la captura de imágenes
- La distribución física de las áreas de descanso se mantiene constante
- Las prácticas de manejo del ganado no interfieren con el sistema de monitoreo

### Riesgos y contingencias

#### Matriz de riesgos

#### Terminología

- Aprendizaje Supervisado

*Es un tipo de aprendizaje automático donde el modelo se entrena con datos de entrada y sus correspondientes salidas etiquetadas. El objetivo es que el modelo aprenda a predecir la salida para nuevos datos basándose en ejemplos previos.*

- Aprendizaje No Supervisado

*En este tipo de aprendizaje, el modelo trabaja con datos que no están etiquetados. El modelo intenta encontrar patrones o estructuras ocultas en los datos, cómo agrupar elementos similares.*



- Bounding Box  
*Es una caja rectangular que se dibuja alrededor de un objeto en una imagen para delimitar su posición y tamaño. Se usa comúnmente en tareas de detección de objetos, como identificar las camas de vacas en tu proyecto.*
- DB  
*Es una colección organizada de datos que pueden ser fácilmente accesibles, gestionados y actualizados.*
- Epochs  
*Es una medida en machine learning que indica cuántas veces el algoritmo de aprendizaje ha visto todo el conjunto de datos de entrenamiento. Cada epoch representa una pasada completa por los datos.*
- Features  
*Son los atributos o propiedades de los datos que el modelo de machine learning utiliza para hacer predicciones. En imágenes, pueden ser bordes, colores, texturas, etc., que el modelo usa para identificar objetos o patrones.*
- Heat Map 3D  
*Es una representación visual en 3D que muestra la intensidad o densidad de datos en diferentes áreas.*
- Labels  
*Son las salidas correctas o categorías asignadas a los datos en un conjunto supervisado.*
- MVP (Producto Mínimo Viable)  
*Es una versión básica de un producto que incluye las funcionalidades esenciales necesarias para que sea funcional y permita a los usuarios interactuar con él.*
- MBI (Indicador de Beneficio Mínimo): Es una métrica o indicador que establece el valor mínimo esperado que un producto o sistema debe ofrecer.
- Modelo de Clasificación  
*Es un tipo de modelo en machine learning que categoriza los datos de entrada en diferentes clases o grupos.*

- Rumia

*Es el proceso digestivo en el cual las vacas y otros animales rumiantes regurgitan y vuelven a masticar su comida.*

### **3. Determinar los objetivos de minería de datos**

#### **Objetivos de minería de datos**

Desarrollar un modelo de detección y clasificación de vacas que permita cuantificar y caracterizar la población ganadera según su postura (parada/acostada) en las diferentes áreas del rancho.

#### **Criterios de éxito**

1. El sistema realiza un conteo de bounding boxes y guarda sus resultados en la tabla:
  - Tabla 1: Registro de imágenes con los campos id\_img\_pk, timestamp, y cow\_count.
2. El sistema identifica la clase de una vaca en un espacio dado y guarda sus resultados en la tabla:
  - Tabla 2: Detalle de las vacas detectadas, incluyendo id\_detected\_cow\_pk, id\_img\_fk, cow\_centroid, y prediction\_accuracy, cow\_class.
3. El sistema permite una conexión a la base de datos a través de una conexión local o por ssh en un cliente que permita exportar la información en formato csv y xlsx.
4. El sistema genera visualizaciones y análisis cuantitativos de los patrones de arena, incluyendo mapas de calor, análisis de bordes y métricas texturales.

### **4. Producir el plan del proyecto**

#### **Duración**

- Etapa de preparación: 2 semanas (semanas 3 y 4)
- Etapa de MVP: 4 semanas (de semana 5 a semana 8)
- Etapa de MBI 1: 1 semana (semana 9 a 10)

#### **Entradas**

- 9634 imágenes JPG (Peso aprox. 600 - 650 kb, Resolución: 1920 \* 1080 px)


## Salidas

- Base de Datos
  - Estado de las camas por diferentes intervalos de tiempo
- Código Fuente
  - Modelos:
    - a. Bounding Box de una o múltiples vacas en una imagen dada.
    - b. Clasificador: Determina si una vaca está acostada o parada en un espacio dado.
    - c. Clasificador: Identifica posibles patrones de arena en las camas vacías.
- Hallazgos significativos (Análisis de Resultados)

## Restricciones

- Tiempo limitado con el socio formador (solo nos vemos una vez a la semana)
- Tiempos de entrega se cruzan con entregas de evidencias individuales
- Traslado al CAETEC es largo
- Tenemos imágenes limitadas a solo un lugar de camas (solo una posición de la cámara)
- 

## Plan del Proyecto

 Plan CAETEC

## Diagrama de Plan del Proyecto

