



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. Eficientar el espacio de descanso utilizado por las vacas. Preguntas asociadas: ¿Cuántas vacas están descansando en las camas? ¿Cuánto tiempo permanecen ahí? ¿Qué camas son más utilizadas que otras? ¿Cuáles son las posiciones de las vacas descansando?

Criterios de éxito

1. Se identifica el tiempo promedio del uso de cama de una vaca
2. Se identifica la presencia de una vaca y su clase (parada / acostada)
3. Se identifica el porcentaje de uso de cada cama al día

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.
- 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. Personal CAETEC: Doctora Veterinaria, Arturo, Ivo.
 - b. Profesores
2. Datos
 - a. 9634 imágenes JPG
 - b. 600 - 650 kb, 1920 * 1080 px
 - c. [Images](#)
3. Hardware
 - a. Webcam Logitech C920
 - b. Raspberry Pi 3
4. Software
 - a. Python
 - b. Pandas
 - c. Numpy
 - d. Pytorch *
 - e. Tensor Flow *
 - f. PySpark *

Requerimientos, supuestos y restricciones

Modelos:

1. Bounding box camas -> Recorte de bounding boxes.
2. Clasificador -> Vaca parada, acostada, o no vaca.
3. Clasificador -> Posibles patrones de arena (Heat Map 3D)

Requerimientos:

- El modelo debe de:
 - Encontrar el bounding box de la imagen original de cada una de las camas
 - Recortar cada bounding box y crear una imagen para cada uno
 - Predecir una clase a cada imagen (No vaca/Acostada/Parada).
 - Calcular la precisión de su predicción.
 - Calcular el error de su predicción.
 - Detectar patrones en la arena de las camas vacías para identificar qué comportamientos en la arena no es el preferido por las vacas.

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

1. Encontrar hallazgos significativos respecto al uso de las camas por las vacas. Este objetivo lo evaluará el personal de CAETEC.

Criterios de éxito

1. El modelo debe alcanzar un porcentaje de precisión satisfactorio (por ejemplo, $\geq 80\%$) al clasificar las vacas en las tres categorías: 'no vaca', 'acostada', y 'parada'.
2. El modelo debe mantener un margen de error bajo en las predicciones (un error $\leq 15\%$).
3. El modelo debe lograr un IoU promedio de al menos 0.75 para la detección de camas y vacas. Esto significa que, en promedio, los bounding boxes predichos deben coincidir con los reales en al menos un 75% de los casos.

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 (Modelo Entrenado)
- Hallazgos significativos (Análisis de Resultados)

Plan del Proyecto

 Plan CAETEC

Diagrama de Plan del Proyecto

