Hacia Gestión sostenible de productos lácteos: un método mejorado de aprendizaje automático para la detección del estro

# Kevin Fauvel

Univ Rennes, Inria, CNRS, IRISA [kevin.fauv](mailto:kevin.fauvel@inria.fr)[el@inria.fr](mailto:el@inria.fr)

# V´eronique Masson

Univ Rennes, Inria, CNRS, IRISA [v](mailto:veronique.masson@irisa.fr)[eronique.masson@irisa.fr](mailto:eronique.masson@irisa.fr)

# MILisa FRomont

Univ Rennes, Inria, CNRS, IRISA [elisa. de](mailto:elisa.fromont@irisa.fr)[t@irisa.fr](mailto:t@irisa.fr)

**RESUMEN**

# Philippe Faverdin

PEGASE, INRA, AGROCAMPUS OUEST

[philippe.fav](mailto:philippe.faverdin@inra.fr)[erdin@inra.fr](mailto:erdin@inra.fr)

# Alexandre Termier

Univ Rennes, Inria, CNRS, IRISA [alexandre.termier@irisa.fr](mailto:alexandre.termier@irisa.fr)

Nuestra investigación aborda el desafío de la eficiencia en el uso de recursos de producción de leche en granjas lecheras con métodos de aprendizaje automático. La reproducción es un factor clave para el rendimiento de la granja lechera, ya que la producción de leche de vaca comienza con el nacimiento de un ternero. Por lo tanto, detectar el estro, el único período en el que la vaca es susceptible de preñez, es crucial para la eficiencia de la granja. Nuestro objetivo es mejorar la detección de celos (rendimiento, interpretabilidad), especialmente en el estro silencioso actualmente no detectado (35% del estro total), y permitir que los agricultores confíen en soluciones automáticas de detección de celos basadas en datos asequibles (actividad, temperatura). En este artículo, primero proponemos un enfoque novedoso con análisis de datos del mundo real para abordar la detección de celos tanto conductuales como silenciosos a través de métodos de aprendizaje automático. En segundo lugar, presentamos LCE, un algoritmo local basado en cascada que supera significativamente a una solución comercial típica para la detección de celos, impulsado por su capacidad para detectar celos silenciosos. Luego, nuestro estudio revela el papel fundamental del despliegue de sensores de actividad en la detección del celo. Finalmente, proponemos un enfoque que se basa en la interpretabilidad del algoritmo (SHAP) global y local (conductual versus silencioso) para reducir la desconfianza en las soluciones de detección de celos.

**CONCEPTOS CCS**

### Metodologías informáticas → Aprendizaje automático; •

**Computación aplicada** → **Agricultura**.

## PALABRAS CLAVE

Manejo sustentable de productos lácteos; Aprendizaje automático; Clasificación; Interpretabilidad

Derechos de publicación con licencia para ACM. ACM reconoce que esta contribución fue escrita o en coautoría por un empleado, contratista o afiliado de un gobierno nacional. Como tal, el Gobierno se reserva el derecho no exclusivo y libre de regalías para publicar o reproducir este artículo, o para permitir que otros lo hagan, solo para fines gubernamentales.

*KDD '19, del 4 al 8 de agosto de 2019, Anchorage, Alaska, EE. UU.*

© 2019 Los derechos de autor pertenecen al propietario / autor (es). Derechos de publicación con licencia para ACM.

ACM ISBN 978-1-4503-6201-6 / 19/08. . . $ 15.00

<https://doi.org/10.1145/3292500.3330712>

**Formato de referencia de ACM:**

Kevin Fauvel, V´eronique Masson, MILisa Fromont, Philippe Faverdin y Alexandre Termier. 2019. Hacia la gestión sostenible de los productos lácteos: un método mejorado de aprendizaje automático para la detección del estro. En la 25ª Conferencia de ACM SIGKDD sobre descubrimiento de conocimientos y minería de datos (KDD '19), del 4 al 8 de agosto de 2019, Anchorage, Alaska, EE. UU. ACM, Nueva York, NY, EE. UU., 9 páginas.<https://doi.org/10.1145/3292500.3330712>

## INTRODUCCIÓN

Como se subraya en el informe Creación de un futuro alimentario sostenible [28], el ganado rumiante (bovinos, ovinos y caprinos), que se utiliza para la producción de productos lácteos y cárnicos, ocupa dos tercios de las tierras agrícolas mundiales y aporta aproximadamente la mitad de las emisiones relacionadas con la producción agrícola. El aumento de la eficiencia del uso de los recursos en las granjas es uno de los pasos más importantes para alcanzar los objetivos ambientales y de producción de alimentos. Como respuesta, la ganadería de precisión (PLF) es una forma prometedora de mejorar el rendimiento de la explotación [30]. PLF es el uso de información continua para optimizar un manejo animal individualizado.

Hoy en día, los datos (por ejemplo, temperatura, actividad, peso corporal, producción de leche) se recopilan en las granjas lecheras a través de diferentes tipos de sensores para apoyar la toma de decisiones de los agricultores en varios aspectos de la gestión (por ejemplo, reproducción, enfermedades, alimentación, medio ambiente). Los métodos de aprendizaje automático pueden ayudar a aprovechar el valor de este volumen de datos en constante crecimiento.

La reproducción es un factor clave para el desempeño de las granjas lecheras. Afecta directamente la producción de leche cuando las vacas comienzan a producir leche después de dar a luz a un ternero; y la productividad de la leche disminuye después de los primeros 3 meses. La razón más común para el sacrificio de vacas, el acto de sacrificar una vaca, es la cuestión de la reproducción (p. Ej., Intervalo largo entre 2 terneros) [3]. Por lo tanto, es crucial detectar el estro, el único período en el que la vaca es susceptible de preñez, para inseminar oportunamente a las vacas y, por lo tanto, aumentar la eficiencia de la granja.

Tradicionalmente, la detección del estro se basa en la observación visual del comportamiento de los animales. La actividad generalmente aumenta marcadamente en las vacas durante el estro [15] a menos que la vaca esté experimentando un estro silencioso (estro sin signos de comportamiento obvios - 35% del estro total). En la práctica, menos del 50% de los celos se detectan visualmente [25] debido a dos razones principales: el primer celo silencioso no se puede detectar visualmente y, en segundo lugar, los comportamientos sexuales son

mayoritariamente expresado por la noche. Se han desarrollado diferentes métodos para ayudar a la detección visual. El método de referencia es la estimación del celo mediante el análisis automático de progesterona en la leche [9]. Sin embargo, el costo de esta solución prohíbe su implementación extensiva.

Como resultado, se considera que los datos asequibles del sensor de temperatura corporal y de actividad tienen potencial para la detección automática del celo [27]. Se encuentran disponibles algunas soluciones basadas en datos de actividad. Sin embargo, su tasa de adopción sigue siendo moderada [30]. Estas soluciones de detección comerciales se enfrentan a dos deficiencias importantes. En primer lugar, las soluciones basadas en la actividad cubren sólo los estro de comportamiento (estro asociados con signos de comportamiento obvios - 65% del estro total). En segundo lugar, las falsas alertas y la falta de explicación detrás de las detecciones generan desconfianza en las soluciones por parte de los agricultores. Por lo tanto, además de un rendimiento mejorado, también se necesitan justificaciones clave para las alertas de celo para expandir la adopción de soluciones de detección automática.

## Nuestras Contribuciones

Nuestra investigación aborda el desafío de la eficiencia en el uso de recursos para la producción de leche en las granjas lecheras con métodos de aprendizaje automático. Nuestro objetivo es mejorar la detección de celos, especialmente en el estro silencioso actualmente no detectado, y permitir que los agricultores confíen en soluciones automáticas de detección de celos basadas en datos asequibles (por ejemplo, actividad, temperatura). Con nuestro análisis de datos del mundo real y un enfoque exhaustivo de etiquetado del estro (conductual, silencioso), este estudio:

* + - Presentar LCE, un algoritmo basado en cascada local para la detección de celos;
    - Demuestre que LCE supera significativamente a una referencia comercial en la detección de celos;
    - Evaluar la relevancia de implementar una combinación de 2 sensores asequibles (actividad y temperatura);
    - Identificar los impulsores clave que respaldan las alertas de celo a nivel global y local (conductuales versus silenciosos) basados ​​en la interpretabilidad del algoritmo y proponer un enfoque

para reducir la desconfianza en las soluciones.

## RELACIONADOS TRABAJA

Las series de tiempo multivariante (MTS) recopiladas de los sensores de actividad y temperatura se etiquetan como estro o anestro, el período de inactividad sexual entre dos períodos de estro. La detección del estro se puede formular como un problema de clasificación binaria. En esta sección, primero discutimos los clasificadores adecuados para nuestro estudio. Luego, examinamos la literatura sobre la interpretabilidad de clasificadores. Finalmente, presentamos el trabajo existente sobre detección de celos a través de métodos de aprendizaje automático.

## Clasificación

Entre los clasificadores binarios de última generación para datos numéricos y en el contexto de nuestro problema, podemos excluir el uso de clasificadores dedicados a MTS. Los clasificadores MTS no se ajustan a nuestras necesidades por diferentes motivos. Primero, la literatura actual en ciencia animal no proporciona información para hacer suposiciones sobre un modelo fisiológico particular. En segundo lugar, los algoritmos de clasificación de intervalos [5] y shapelets [17] se excluyen debido a las breves ventanas de tiempo que consideramos. Los datos de los sensores son

24 horas agregadas (el período relevante para la evaluación del estro); y según los científicos de animales, los datos sobre el día del estro y el día antes del estro podrían ser suficientes para la detección del estro (tamaño de ventana de tiempo de dos). Finalmente, los enfoques de representación de diccionario no nos permiten explotar las interacciones temporales entre variables debido a la representación agregada de series en el tiempo [4]. Nuestro conjunto de datos tiene la misma frecuencia entre las variables, administramos el aspecto del tiempo estableciendo las diferentes marcas de tiempo como variables de columna.

En consecuencia, exploramos clasificadores de última generación en las siguientes clases: k vecinos más cercanos, regresiones logísticas regularizadas, máquinas de vectores de soporte, redes neuronales y métodos de conjunto.

En primer lugar, consideramos la red elástica [31], la regresión logística que combina los métodos de regularización L1 y L2, que constituye la referencia en la regresión logística regularizada.

Luego, dado el menor número de características que el número de muestras en nuestro conjunto de datos, probamos una máquina de vectores de soporte con un núcleo de función de base radial.

Entre las categorías de redes neuronales (perceptor multicapa - MLP, red neuronal convolucional - CNN y red neuronal recurrente - RNN), consideramos las MLP pequeñas. Los MLP profundos, sin capas convolucionales, son difíciles de entrenar debido a la gran cantidad de parámetros y al problema del gradiente de desaparición [23]. Además, el tamaño de nuestro conjunto de datos (18.000 muestras) y la inexistencia de una red CNN previamente entrenada sobre un problema comparable no nos permiten utilizar CNN. Entonces, los RNN no se adaptan a las breves ventanas de tiempo que consideramos. Por último, los enfoques explícito (ensacado y refuerzo) e implícito (aprendizaje de correlación negativa y mezcla de expertos) exhiben sus respectivas fortalezas y limitaciones, por lo que se recomienda un método de conjunto híbrido [20]. Las fortalezas y limitaciones de los enfoques explícitos e implícitos se refieren a su capacidad para generalizar más allá del conjunto de datos de entrenamiento. El rendimiento de la generalización depende del equilibrio encontrado entre un algoritmo que no captura la estructura subyacente del conjunto de datos de entrenamiento (desajuste - alto sesgo) y un algoritmo que está aprendiendo demasiado de cerca el conjunto de datos de entrenamiento (sobreajuste - alta varianza). Este desafío se denomina compensación de sesgo-varianza. El aprendizaje de correlación negativa (NCL) intenta entrenar clasificadores individuales en un conjunto y los combina en el mismo proceso de aprendizaje. En todo el conjunto de entrenamiento, los clasificadores individuales se entrenan de manera simultánea e interactiva a través de los términos de penalización de correlación de sus funciones de error para ajustar la compensación de sesgo-varianza. La desventaja es que todos los clasificadores individuales se preocupan por el error de conjunto completo. Mezcla de expertos (ME) es un método de conjunto basado en el principio de divide y vencerás en el que el espacio del problema se divide entre unos pocos expertos (por ejemplo, clasificadores), supervisado por un esquema de promedio ponderado dinámico (red de control). Permite que cada experto aprenda una parte de los datos de entrenamiento con su correspondiente error individual. Sin embargo, no hay control sobre la compensación sesgo-varianza. Existen combinaciones de enfoques implícitos de NCL y ME [1, 12]. Estos métodos integran un término de penalización por correlación de función de error para fomentar diferentes clasificadores (NCL), a través de un enfoque de divide y vencerás (ME), Mezcla de expertos (ME) es un método de conjunto basado en el principio de divide y vencerás en el que el espacio del problema se divide entre unos pocos expertos (por ejemplo, clasificadores), supervisado por un esquema de promedio ponderado dinámico (red de control). Permite que cada experto aprenda una parte de los datos de entrenamiento con su correspondiente error individual. Sin embargo, no hay control sobre la compensación sesgo-varianza. Existen combinaciones de enfoques implícitos de NCL y ME [1, 12]. Estos métodos integran un término de penalización por correlación de función de error para fomentar diferentes clasificadores (NCL), a través de un enfoque de divide y vencerás (ME), Mezcla de expertos (ME) es un método de conjunto basado en el principio de divide y vencerás en el que el espacio del problema se divide entre unos pocos expertos (por ejemplo, clasificadores), supervisado por un esquema de promedio ponderado dinámico (red de control). Permite que cada experto aprenda una parte de los datos de entrenamiento con su correspondiente error individual. Sin embargo, no hay control sobre la compensación sesgo-varianza. Existen combinaciones de enfoques implícitos de NCL y ME [1, 12]. Estos métodos integran un término de penalización por correlación de función de error para fomentar diferentes clasificadores (NCL), a través de un enfoque de divide y vencerás (ME), no hay control sobre la compensación sesgo-varianza. Existen combinaciones de enfoques implícitos de NCL y ME [1, 12]. Estos métodos integran un término de penalización por correlación de función de error para fomentar diferentes clasificadores (NCL), a través de un enfoque de divide y vencerás (ME), no hay control sobre la compensación sesgo-varianza. Existen combinaciones de enfoques implícitos de NCL y ME [1, 12]. Estos métodos integran un término de penalización por correlación de función de error para fomentar diferentes clasificadores (NCL), a través de un enfoque de divide y vencerás (ME),

para aprender a utilizar diferentes partes de los datos de entrenamiento. Sin embargo, las combinaciones de enfoques implícitos no se benefician de la capacidad de generalización mejorada de crear explícitamente diferentes conjuntos de entrenamiento cambiando probabilísticamente la distribución de los datos de entrenamiento originales (embolsado, refuerzo). Existe un método que combina el enfoque de impulso explícito con el enfoque implícito de divide y vencerás de ME [13]. No obstante, el cambio de distribución de sesgo bajo del impulso no garantiza una compensación de sesgo-varianza.

Por lo tanto, dado el rendimiento más bajo de los MLP pequeños en comparación con los métodos de conjunto en promedio (confirmado por nuestros experimentos), proponemos un nuevo método de conjunto híbrido. Combina un enfoque explícito de embolsado-impulso para manejar la compensación de sesgo-varianza y un enfoque implícito de dividir y conquistar ME para aprender diferentes partes de los datos de entrenamiento.

Como se mencionó anteriormente, no podemos separar el rendimiento de detección de los clasificadores de la interpretabilidad. Esto se explorará en la siguiente sección.

## Interpretabilidad

No existe una definición matemática de interpretabilidad. Una definición propuesta por [21] establece que cuanto mayor sea la interpretabilidad de un algoritmo de aprendizaje automático, más fácil será para alguien comprender por qué se han tomado ciertas decisiones o predicciones.

Nuestro problema requiere conocimientos sobre el tipo de estro (conductual versus silencioso), lo que sugiere explicaciones locales. Además, necesitamos un método capaz de trabajar para los diferentes clasificadores identificados (agnóstico del modelo). Los métodos de última generación que cumplen con estos requisitos (locales, independientes del modelo) son Explicaciones locales interpretables independientes del modelo (LIME)

[26] y explicaciones de aditivos SHapley (SHAP) [19]. Los valores SHAP vienen con las ventajas de estimación local de caja negra de LIME, pero también con garantías teóricas. Por lo tanto, usamos SHAP para interpretar el resultado de nuestro algoritmo de aprendizaje automático. Esta técnica está inspirada en la teoría de juegos, que se utiliza para determinar cuánto ha contribuido cada jugador en un juego colaborativo a su éxito. En nuestro estudio, los valores SHAP miden cuánto impacta una variable de actividad o temperatura en las predicciones del estro. Un valor SHAP absoluto más alto de una variable en comparación con otras variables significa que esta variable tiene un mayor poder predictivo o discriminativo en el algoritmo de detección. Los valores SHAP se calculan mediante la contribución marginal promedio de un valor de característica hacia la predicción sobre todas las coaliciones posibles. Valores de interacción SHAP, una extensión de los valores SHAP basados ​​en el índice de interacción de Shapley [14], captura los efectos de interacción por pares. Además, los valores SHAP están disponibles a nivel local. Lo analizamos para comparar el impacto de las variables en las predicciones del algoritmo en el estro conductual y silencioso.

## Detección automática de estro

Hay un par de estudios sobre la aplicación de métodos de aprendizaje automático en la detección de celos [11, 18, 22]. Ninguno de ellos utiliza el método actualmente reconocido para la identificación de celos conductuales y silenciosos como etiquetas (perfiles de progesterona), por lo que sus métodos de etiquetado de celos no son exhaustivos. Es más,

dos Los estudios utilizan diferentes variables (volumen de leche, orden de ordeño, días desde el último estro) en lugar de las medidas asequibles de actividad o temperatura. Finalmente, ninguno de ellos brinda información sobre las predicciones de algoritmos basados ​​en su interpretabilidad.

[22] basa el estudio en datos de series de tiempo de volumen de leche y orden de ordeño, utilizando la detección visual como la verdad del suelo. Se probaron dos esquemas de aprendizaje: FOIL y C4.5. Los algoritmos detectaron el 69% de los estro identificados por método visual y se produjo una gran cantidad de falsos positivos (74%).

[18] aprende un MLP sobre datos de series de tiempo de actividad y el número de días desde el último estro, utilizando la inseminación exitosa como la verdad básica. El modelo mostró una sensibilidad, una especificidad y una tasa de error de 77,5, 99,6 y 9,1% en 373 estro.

Y por último, [11] basa el estudio en datos de actividad de series de tiempo, utilizando la detección visual como la verdad del terreno (65,6% de todos los estro). Se probaron tres técnicas de aprendizaje automático

- bosque aleatorio, discriminante lineal y MLP. Los algoritmos mostraron una precisión del 91% al 100% en un conjunto de datos limitado de 18 vacas.

1. **LCE: ENSAMBLE DE CASCADA LOCAL** Como se mencionó anteriormente, proponemos un nuevo método de conjunto híbrido que combina un método explícito de embolsado-impulso

enfoque para manejar la compensación sesgo-varianza y un enfoque implícito de dividir y conquistar ME para individualizar el error del clasificador en diferentes partes de los datos de entrenamiento Hemos decidido partir de un enfoque basado en apilamiento existente combinado implícito (NCL y ME) (generalización en cascada [29]): cascada local [16]. El potencial de ensacado / impulso del método de división y conquista del árbol de decisión en cascada local motiva nuestra elección. En esta sección, primero presentamos la cascada local, el enfoque inicial basado en apilamiento implícito. A continuación, explicamos LCE, nuestra versión aumentada (explícita e implícita) de cascada local, y luego comparamos el rendimiento de LCE con la cascada local. La Figura 1 ilustra la presentación de cascada local y LCE.

## Cascada local

En primer lugar, la generalización en cascada utiliza un conjunto de clasificadores secuencialmente y en cada paso agrega nuevos atributos al conjunto de datos original [29]. Los nuevos atributos se derivan de las probabilidades de clase dadas por un clasificador base

(e.g. 𝐻0(𝐷), 𝐻1(𝐷01)

en la Figura 1). Se obtiene la compensación de sesgo-varianza por Aprendizaje de correlación negativa: en cada etapa de la secuencia, se seleccionan clasificadores con diferentes comportamientos. Se recomienda en la generalización en cascada comenzar con un algoritmo de baja varianza para dibujar superficies de decisión estables (𝐻0 en la Figura 1) y luego usar un algoritmo de bajo sesgo para ajustar las más complejas (*H1* en la Figura 1). Local cascade [16] aplica la generalización en cascada localmente siguiendo una estrategia de divide y vencerás basada en el principio de mezcla de expertos. El objetivo de este enfoque es capturar nuevas relaciones que no se pueden descubrir globalmente. El método de dividir y conquistar en cascada local es un árbol de decisiones. Al hacer crecer el árbol, se calculan nuevos atributos (probabilidades de clase de un clasificador - clasificador base) en cada nodo de decisión y se propagan por el árbol. Para ser aplicado como predictor, las tiendas en cascada locales,

Cascada local

Conjunto de datos original D

Conjunto de datos *D*1

LCE

Conjunto de datos original D

HARPILLERA

. . .

Conjunto de datos *D*norte

NCL



*H*0 *D*0=*D*+*H*0(*D*) D01 *D*02

*H*1 *D*11=*D*01+ *H*1 *D*12=*D*02+ *H*1(*D*01) *H*1(*D*02)

*H*2 *H*2 *H*2 *H*2

*H*2(*D*111) H2(*D*112) H2(*D*121) H2(*D*122)



IMPULSAR

*H*B *D*0=*D*1+*H*B(*D*1) *D*01 *D*02

*H*B *D*11=*D*01+ *H*B *D*12=*D*02+ *H*B(*D*01) *H*B(*D*02)

*H*B *H*B *H*B *H*B



IMPULSAR

*H*B *D*0=*D*norte+*H*B(*D*norte) *D*01 *D*02

*H*B *D*11=*D*01+ *H*B *D*12=*D*02+ *H*B(*D*01) *H*B(*D*02)

*H*B *H*B *H*B *H*B

*H*B(*D*111) HB(*D*112) HB(*D*121) HB(*D*122) *H*B(*D*111) HB(*D*112) HB(*D*121) HB(*D*122)

Explícito Implícito

Predicciones

ME

Reducir el sesgo Reducir la varianza

Predicciones

ME

Predicciones

Predicciones

ME

**Figura 1: Cascada local frente a LCE**

## LCE: Conjunto de cascada local

Nuestra contribución interviene en nuestra manera explícita de manejar la compensación sesgo-varianza mientras que el enfoque en cascada local es implícito, alternando entre los comportamientos de los clasificadores base (sesgo, varianza) en cada nivel del árbol.

LCE reduce el sesgo en el árbol de decisión mediante el uso de impulso como clasificador base (𝐻𝑏 en la Figura 1). Impulsar el cambio de distribución de datos iterativos del clasificador base (reponderación) disminuye el sesgo en cada nivel de árbol. Además, el impulso se propaga por el árbol agregando probabilidades de clase del clasificador base al conjunto de datos de entrenamiento (nuevos atributos). Las probabilidades de clase contienen información sobre la capacidad del clasificador base para clasificar correctamente una muestra. En el siguiente nivel de árbol, las probabilidades de clase agregadas al conjunto de datos son explotadas por el clasificador base como un esquema de ponderación para enfocarse más en muestras previamente mal clasificadas.

Luego, el sobreajuste generado por el enfoque de reducción de sesgos de divide y vencerás del árbol de decisión se mitiga mediante el uso de ensacado. El ensacado proporciona una reducción de la varianza mediante la creación de múltiples árboles de decisión a partir de diferentes submuestras del conjunto de datos original (muestreo aleatorio con reemplazo, ver

*𝐷*1.. . 𝐷 en la Figura 1). Los árboles se agregan con un voto de mayoría simple.

El nuevo método de conjunto híbrido de LCE permite equilibrar la compensación de sesgo-varianza sin la necesidad de un aprendizaje interactivo entre clasificadores individuales (NCL), mientras se beneficia de la capacidad de generalización mejorada de crear explícitamente diferentes conjuntos de entrenamiento (embolsado, refuerzo). Además, el método de divide y vencerás de LCE asegura que los clasificadores aprendan sobre diferentes partes de los datos de entrenamiento sin la necesidad de un esquema de supervisión (red de entrada).

Nosotros presente el pseudocódigo LCE en el algoritmo 1. Una función (LCE Tree) construye un árbol y la segunda (LCE) el bosque de árboles mediante ensacado.

Hay 2 criterios de parada durante la fase de construcción de un árbol: cuando un nodo tiene una clase única o cuando el árbol alcanza la profundidad máxima. Establecemos el rango de profundidad del árbol de 0 a 3 en

**Algoritmo 1** LCE: Conjunto de cascada local

**Exigir:** Un conjunto de datos 𝐷, un conjunto de clasificadores 𝐻, profundidad máxima de un árbol 𝑎 𝑑𝑒 ℎ, número de árboles 𝑒𝑒

1: **función** LCE (𝐷, 𝐻,  *𝑒𝑒* ,  *𝑎 𝑑𝑒 ℎ*)

2: *𝐹* ← ∅

3: **por** cada en [1, 𝑒𝑒 ] hacer

4: *𝑆* ← Una muestra de bootstrap de 𝐷

5: ← Árbol LCE (𝑆, 𝐻, 𝑎 𝑑𝑒 ℎ, 0)

6: *𝐹* ← 𝐹 ∪

7: **regreso** *𝐹*

8: **función** Árbol LCE (𝐷, 𝐻, 𝑎 𝑑𝑒 ℎ, 𝑑𝑒 ℎ)

9: **Si**  *𝑎 𝑑𝑒 ℎ* o clase uniforme entonces

10: **regreso** hoja

11: **demás**

12: *𝐷 ′* Concatenar (𝐷, 𝐻𝑑𝑒 ℎ (𝐷))

←

13: Dividir 𝐷 ′ en el atributo que maximiza el criterio de Gini

14: *𝑑𝑒 ℎ 𝑑𝑒 ℎ* + 1

←

15: **por** *𝐷′*( ) (𝐷 ′) **hacer**

∈

dieciséis: *𝑇 𝑒𝑒* = LC Tree (𝐷 ′ ( ), 𝐻,  *𝑎 𝑑𝑒 ℎ*, 𝑑𝑒 ℎ) 17: **regreso** árbol que contiene un nodo de decisión, almacenando el clasificador 𝐻𝑑𝑒 ℎ + 1 (𝐷) y subárboles descendientes

*𝑇 𝑒𝑒*

LCE en lugar de 0 a 5 en cascada local. Este hiperparámetro se utiliza para controlar el sobreajuste. Nuestra elección de clasificadores base de impulso de bajo sesgo justifica el ajuste de profundidad máxima para

3. En este estudio, el conjunto de clasificadores de base de bajo sesgo se limita al algoritmo de refuerzo de última generación (refuerzo de gradiente extremo - XGB [8]).

Además, eliminamos dos reglas implementadas en cascada local para reducir la varianza: la tasa máxima de error del clasificador base y la representación de clase mínima en un nodo. La primera regla requiere detener la propagación hacia abajo del árbol para evitar el sobreajuste si el clasificador base, en un nodo, tenía una tasa de error por debajo de un cierto umbral (0.5). Nuestro enfoque sugiere una reducción de la varianza a través del ensacado, y no durante

una construcción de árbol; así que no cumplimos esta regla. Para restringir la atención a clases bien pobladas, la segunda regla requiere considerar una clase en un nodo si el número de ejemplos que pertenecen a esta clase es mayor que 𝑁 veces (3) el número de atributos. No cumplimos la segunda regla por la misma razón.

## Comparación de rendimiento: local Cascada versus LCE

Nuestra comparación tiene como objetivo subrayar el rendimiento superior de LCE en comparación con una cascada local en nuestro conjunto de datos del mundo real, inducido por sus diferentes enfoques de manejo de la compensación de sesgo-varianza (enfoque explícito versus implícito). LCE se implementa de acuerdo con la descripción dada en la sección anterior. La implementación en cascada local corresponde a la descripción del artículo original y, como se recomienda, utilizamos bayes ingenuos para el clasificador de base de baja varianza. Para ser comparable, el clasificador base de bajo sesgo es XGB. La profundidad se establece en 1 para LCE y la cascada local. Los resultados se presentan en la Tabla 1.

### Mesa 1: Puntuación F1 con intervalo de confianza del 95% de LCE versus cascada local (LC) en nuestro conjunto de datos

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Árboles** | **1** | **5** | **10** | **30** | **50** | **70** | **90** |
| **LCE** | 68,1 | 69,2 | 68,9 | 69,1 | 69,1 | 68,9 | 68,9 |
|  | ± 3,2 | ± 2,6 | ± 2,8 | ± 2,4 | ± 2,5 | ± 2,4 | ± 2,5 |

**LC** 53,2 ± 2,8

Como se esperaba, los resultados muestran una mayor variabilidad entre los pliegues de LCE en comparación con la cascada local cuando el número de árboles se establece en 1 debido a su baja orientación de sesgo (error estándar de 1,6% frente a 1,4% en la puntuación F1, cálculo de rendimiento detallado en la sección Experimentos). Sin embargo, LCE en 1 árbol exhibe un rendimiento de detección más alto que la cascada local (puntuación F1: 68,1% frente a 53,2%).

Además, a través del ensacado, observamos una reducción de la variabilidad de LCE a un nivel más bajo que la cascada local, así como un aumento del rendimiento de detección (puntaje F1, intervalo de confianza del 95%: 68.1 ± 3.2 con 1 árbol versus 68.9 ± 2.4 con 70 árboles

versus 53,2 ± 2,8 con cascada local).

Por tanto, esta comparación afirma la superioridad de nuestro

Enfoque explícito de compensación de sesgo-varianza en comparación con el enfoque implícito de NCL de cascada local en nuestro conjunto de datos. El comportamiento intrínseco diferente de LCE y cascada local se confirma en la sección de resultados y discusiones.

## EXPERIMENTOS

En esta sección, presentamos la composición de nuestro conjunto de datos del mundo real, el preprocesamiento realizado y el entorno experimental.

## Conjunto de datos

Nuestro conjunto de datos está fuera de línea. De 2014 a 2017, se realizó un experimento conducido en la granja lechera INRA M´ejusseaume (4806 '

N, 147 'W, Bretaña, Francia). Este experimento inscribió a 125 vacas Holstein alojadas en puestos gratuitos que representan 153 lactancias. Cada vaca estaba equipada con un medidor de actividad montado en un collar.(HeatPhone - Medria Technologies, Chˆateaubourg, Francia) y un sensor de temperatura en el primer estómago (Thermobo- lus - Medria Technologies, Chaˆteaubourg, Francia). Basándonos en su buen rendimiento en comparación con otras soluciones [7] y su presencia en el mercado internacional, consideramos que el sistema de detección de celosía Medria es una base de comparación razonable. En las siguientes secciones, Medria se denomina solución comercial (CS). El conjunto de datos consta de alertas visuales de estro, alertas de estro de Medria y variables numéricas de Medria con una frecuencia de 5 minutos (rumia, ingestión, descanso, pararse, hiperactividad, otra actividad, temperatura y temperatura corregida). La temperatura corregida tiene en cuenta el efecto refrescante de la ingestión de agua por parte de las vacas. Con respecto a las alertas visuales de celo, el personal de la granja realizó una observación visual. El personal también verificó las alertas de la solución comercial antes de introducir sus registros visuales, por lo que estas alertas de estro visual se muestran como Visual y CS en el estudio. El preprocesamiento aplicado a los datos recolectados es una agregación de 24 horas (actividad: suma, temperatura: media) que corresponde a la ventana relevante tanto para la detección de celos como, desde un punto de vista de alerta, las necesidades de los agricultores. Suponemos que el tratamiento operado por Medria sobre datos brutos para generar variables es estable durante nuestro

experimentar.

Nuestro enfoque novedoso aborda la detección de ambas categorías de celo (conductual y silencioso). Por lo tanto, marcamos el estro midiendo la concentración de progesterona en la leche entera, la referencia actual para una identificación exhaustiva del estro. Este método eficaz en el tiempo y no invasivo para la vaca induce errores comúnmente aceptados (mediciones de progesterona, análisis de perfiles [2]). Marcamos un estro como estro conductual cuando se produce una detección visual o una alerta de Medria. Un estro se considera silencioso cuando no se produce ni la detección visual ni una alerta de Medria. Nuestro conjunto de datos está compuesto por 671 estro con un 37% de estro silencioso que está alineado con la tasa del 35% observado en la literatura [24].

Los días anteriores al estro son una valiosa fuente de información para la detección del estro, lo configuramos como un hiperparámetro. Se evalúan todos los valores en el rango de 1 a 21 días, la duración de un ciclo ovárico regular. Los días pasados ​​de variables se agregan como columnas de características.

* + 1. *Selección de características.* Nosotros realizar la selección de características en este estudio debido a la sensibilidad del método elegido para interpretar el algoritmo de detección (SHAP) a las altas correlaciones entre las características. Realizamos una selección de subconjuntos en pares de características colineales basadas en el coeficiente de correlación de Pearson (umbral 0.8). Un par de características está por encima del umbral(0,9: temperatura corregida, temperatura). Dado que la temperatura se ve afectada por el efecto de enfriamiento de la ingestión de agua, se selecciona la temperatura variable corregida. A partir de este punto, la temperatura corregida se denomina temperatura. Después de esta selección de características, ninguna correlación por pares de Pearson en el caso del conjunto de datos de 21 días pasados ​​está por encima del umbral.
    2. *Estructura del conjunto de datos.* Nosotros hacer una validación cruzada de 5 veces. La división del conjunto de datos se presenta en la Tabla 2.

### Tabla 2: División del conjunto de datos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Doblar1** | **Fold2** | **Doblar3** | **Fold4** | **Doblar5** | **Todas** |
| Estro | 126 | 136 | 118 | 141 | 153 | 671 |
| Silencio % | 33 | 40 | 24 | 40 | 46 | 37 |

La división ha mantenido el mismo número de días en celo en cada pliegue (1,144 días). Tomamos esta decisión para evitar el sobreajuste en un animal en particular. Discutimos el impacto de una división manteniendo el mismo número de animales por pliegue en la sección de rendimiento de detección. Además, no observamos ningún desequilibrio estructural en el porcentaje de celo silencioso en los pliegues.

## Entorno experimental

Presentamos en esta sección los algoritmos y métodos utilizados en nuestros experimentos.

* + 1. *Algoritmos.* Nosotros probamos nuestro conjunto híbrido LCE (enfoques explícitos - implícitos) versus el enfoque implícito inicial (cascada local) y el algoritmo de última generación para cada enfoque explícito (ensacado: bosque aleatorio, impulso: aumento extremo de gradiente) . También se prueban los vecinos más cercanos K, la red elástica, las máquinas de vectores de soporte y los MLP pequeños.
       - k vecinos más cercanos - KNN: usamos la implementación vecinos. KNeighborsClassifier en el paquete scikit-learn para Python1
       - Elastic net - EN: utilizamos el modelo lineal de implementación SGDClassifier en el paquete scikit-learn para

Python1

Support Vector Machine - SVM: usamos la implementación svm.SVC en el paquete scikit-learn para Python1

∙

* + - * Random Forest - RF: usamos el conjunto de implementación- ble.RandomForestClassifier en el paquete scikit-learn

para Python1

Extreme Gradient Boosting - XGB: usamos la implementación en el paquete xgboost para Python2

∙

* + - * Cascada local - LC: el algoritmo se ha vuelto a implementar en Python 2.7 según la descripción del artículo

ya que no hay una versión pública disponible.

* + - * LCE: algoritmo implementado en Python 2.7 Multilayer Perceptron - MLP: usamos la implementación disponible en el paquete Keras para Python3 y

∙

limitar la arquitectura de la red neuronal a 3 capas

* + 1. *Optimización de hiperparámetros.* Los hiperparámetros de los clasificadores se establecen mediante hyperopt, una optimización basada en modelos secuenciales que utiliza un árbol de algoritmo de búsqueda de estimadores Parzen [6]. Hyperopt elige la siguiente decisión de hiperparámetros de las opciones anteriores y un algoritmo de optimización basado en árboles. Los estimadores de Tree of Parzen cumplen o superan la búsqueda de cuadrícula y el rendimiento de búsqueda aleatoria para la configuración de hiperparámetros.

1https://scikit-learn.org/stable/ 2https: //xgboost.readthedocs.io/en/latest/python/ 3https: //keras.io/

Nosotros use la implementación disponible en el paquete de Python hyperopt4 y el contenedor de hyperas para keras. La optimización se lleva a cabo para maximizar la puntuación F1. La elección de esta métrica se debe a dos razones. Primero, no hacemos suposiciones sobre el estilo de manejo de la lechería; los agricultores pueden favorecer una mayor tasa de detección de estros (mayor recuerdo) o menos alertas falsas (mayor precisión) de acuerdo con sus necesidades. En segundo lugar, nos enfrentamos a un desequilibrio de clases (33% de los días de estro) que hace que la métrica de precisión sea irrelevante.

* + 1. *Rendimiento de clasificación.* Nuestros experimentos utilizan perfiles de progesterona como verdad fundamental para la identificación exhaustiva del estro. Los niveles de progesterona nos permiten identificar una ventana de tiempo de 3 días para el estro con una duración de menos de 24 horas, en el esquema estándar. Adoptando un enfoque conservador, decidimos sumar según el máximo de nuestras predicciones diarias sobre el período de estro / anestro para calcular el rendimiento de clasificación. Además, observamos que para umbrales altos (umbral> 0,95), el desempeño de los clasificadores es inestable con una disminución significativa en la tasa de detección de celos (recuerdo por debajo del 70%). Además, para umbrales bajos (umbral

*<* 0.1), los clasificadores son equivalentes a un clasificador aleatorio. Por lo tanto, decidimos adoptar un cálculo de puntuación F1 basado en el promedio de puntuación F1 en el rango de umbral 0.1-0.95. Este cálculo no modifica los resultados de la selección del clasificador ni el resultado de la comparación con la solución comercial. No obstante, corresponde al rango plausible de calibración para el manejo de lechería y muestra un desempeño de detección más cercano a las condiciones reales.

* + 1. *Selección de algoritmos.* Basado en una división de entrenamiento / validación / prueba de validación cruzada de 5 veces 60/20/20, el mejor clasificador se selecciona en función de la puntuación F1 más alta en los conjuntos de validación.
    2. *Prueba estadística.* Como recomienda [10], hemos utilizado una prueba t de validación cruzada de 5 × 2 para determinar la significación estadística de los algoritmos de aprendizaje automático en un conjunto de datos.
    3. *Interpretabilidad.* Como se mencionó en la sección de trabajo relacionada, usamos la implementación SHAP disponible en el paquete de Python shap5.

## RESULTADOS Y DISCUSIONES

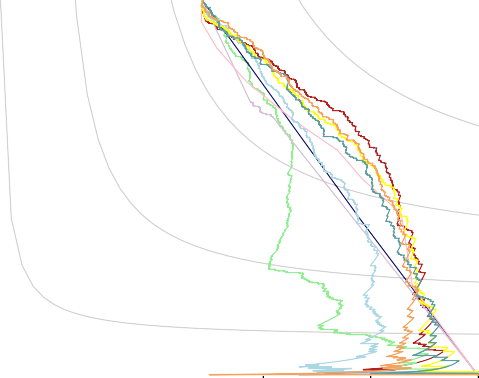
Esta sección está estructurada en dos partes: desempeño e interpretabilidad. La parte de rendimiento de detección compara LCE con otros métodos de detección (clasificadores, solución comercial) y evalúa la relevancia de implementar 2 sensores. Luego, identificamos los impulsores clave (impacto de variables, interacciones temporales) detrás de las alertas de detección de celo a nivel global y local (conductual versus silencioso) basados ​​en la interpretabilidad del algoritmo (SHAP) y proponemos un enfoque para reducir la desconfianza en la solución.

## Rendimiento de detección

Los resultados de la clasificación de los equipos de prueba se presentan en la Figura 2.

4https://github.com/hyperopt/hyperopt 5https: //github.com/slundberg/shap

1*.*0



F1 = 0,8

F1 = 0,6

F1 = 0,4

F1 = 0,2

0*.*8

0*.*6

Recordar

0*.*4

0*.*2

0*.*0

0*.*00 *.*20 *.*40 *.*60 *.*81 *.*0

Precisión

La correlación se explica principalmente por el rendimiento nulo de la solución comercial en la detección silenciosa del celo en todo el rebaño. En el 67% de las vacas, compuesto por una proporción ligeramente mayor de estro silencioso en comparación con el promedio (40% versus 37%), correlación de predicciones de la solución comercial con

LCE es 0,21 ± 0,03.

A continuación, la baja correlación entre LCE y cascada local

(0.41) confirma el valor agregado por la compensación explícita de sesgo-varianza del enfoque LCE. Esta baja correlación se explica por el bajo recuerdo (11%) de la cascada local para una precisión del 78%. La superficie de decisión estable dibujada por bayes ingenuos en la raíz del árbol de decisión en cascada local limita sustancialmente el rango de desempeño del algoritmo en nuestro conjunto de datos (la recuperación cae con una precisión superior al 66%).

Curvas Iso-F1 LCE

LC CS

EN KNN SVM

RF MLP XGB

Observamos esta caída en el rendimiento para una precisión superior al 66%.

en la Figura 2.

Finalmente, el clasificador con el comportamiento más cercano a LCE es XGB (0.8). Sin embargo, la diferencia de correlación permanece por debajo de

### Figura 2: Curvas de recuperación de precisión en conjuntos de prueba de los clasificadores frente a la solución comercial

El mejor clasificador de conjuntos de validación es LCE con los siguientes hiperparámetros: 3 últimos días, la profundidad es igual a 1 y 70 árboles. No observamos un sobreajuste de LCE, el rendimiento observado en los conjuntos de prueba (puntuación F1: 68,9) es estable en comparación con uno de los conjuntos de validación (puntuación F1: 68,1). Además, el desempeño de LCE responde al objetivo de un aumento en el desempeño tanto en la tasa de detección de celos como en menos alertas falsas en comparación con la solución comercial (CS). Con la misma precisión, la recuperación de LCE es constantemente más alta que la recuperación de soluciones comerciales. Con una precisión del 78%, la tasa de precisión de la solución comercial en este estudio,

nuestro algoritmo detecta un 22% más de estro.

* + 1. *Análisis comparativo.* Nosotros compare la correlación de la tasa de error del LCE con la de otros métodos de detección. Esta comparación nos permite:
       - obtener información sobre las deficiencias de la solución comercial y las detecciones de LCE
       - identificar las limitaciones de nuestro enfoque para la implementación

Una correlación baja indica que los clasificadores se equivocan en diferentes regiones del espacio de instancias. La Tabla 3 presenta las correlaciones de Pearson de los errores de predicción de LCE con otros métodos de detección (clasificadores y solución comercial) en equipos de prueba. Para que sea comparable, hemos establecido el umbral de cada clasificador con la misma precisión que la solución comercial (78%).

### Tabla 3: Correlaciones por pares de Pearson de los errores de predicción de LCE con otros métodos de detección en equipos de prueba

**KNN EN SVM MLP RF XGB LC CS**

0.610.190.570.690.730.80.41 0,37

En primer lugar, la solución comercial muestra un comportamiento intrínseco diferente al de LCE (correlación: 0,37). Este bajo

sustancial y se explica por alguna divergencia entre unas pocas vacas. La divergencia, una correlación de la tasa de error por debajo de 0,6, afecta al 12% de las vacas que comprenden una proporción de estro silencioso alineado con el promedio (35%). Por lo tanto, nuestro enfoque de variación de sesgo mejora el rendimiento de XGB en casos estándar (vacas con 35% de estro silencioso). Sin embargo, observamos un pobre desempeño de LCE en el 11% de las vacas que exhiben una alta proporción de estro silencioso (puntuación F1 <55%, proporción de estro silencioso: 54%). El estro silencioso no se distribuye por igual entre las vacas. En nuestro conjunto de datos, el 16% de las vacas representan el 40% del estro silencioso. El rendimiento de LCE por vaca está expuesto a la proporción del tipo de estro animal. Se confirma por la caída del rendimiento de LCE cuando se evalúa en el conjunto de datos de actividad y temperatura generado por un estratificado de 5 veces en animales

(66,3 ± 3,4). La variabilidad del rendimiento de LCE por vaca de acuerdo con la proporción del tipo de estro animal es una limitación de nuestra

solución para implementación; mientras tanto, también es un factor que impulsa la mejora de la detección. Sugerimos una mayor investigación para incorporar características individuales de animales adicionales.

* + 1. *¿Uno o dos sensores?* Para responder a esta pregunta, comparamos el rendimiento de detección en conjuntos de prueba de LCE sobre la temperatura, la actividad y ambas variables. También comparamos los resultados de la detección de LCE con la solución comercial y el método visual.

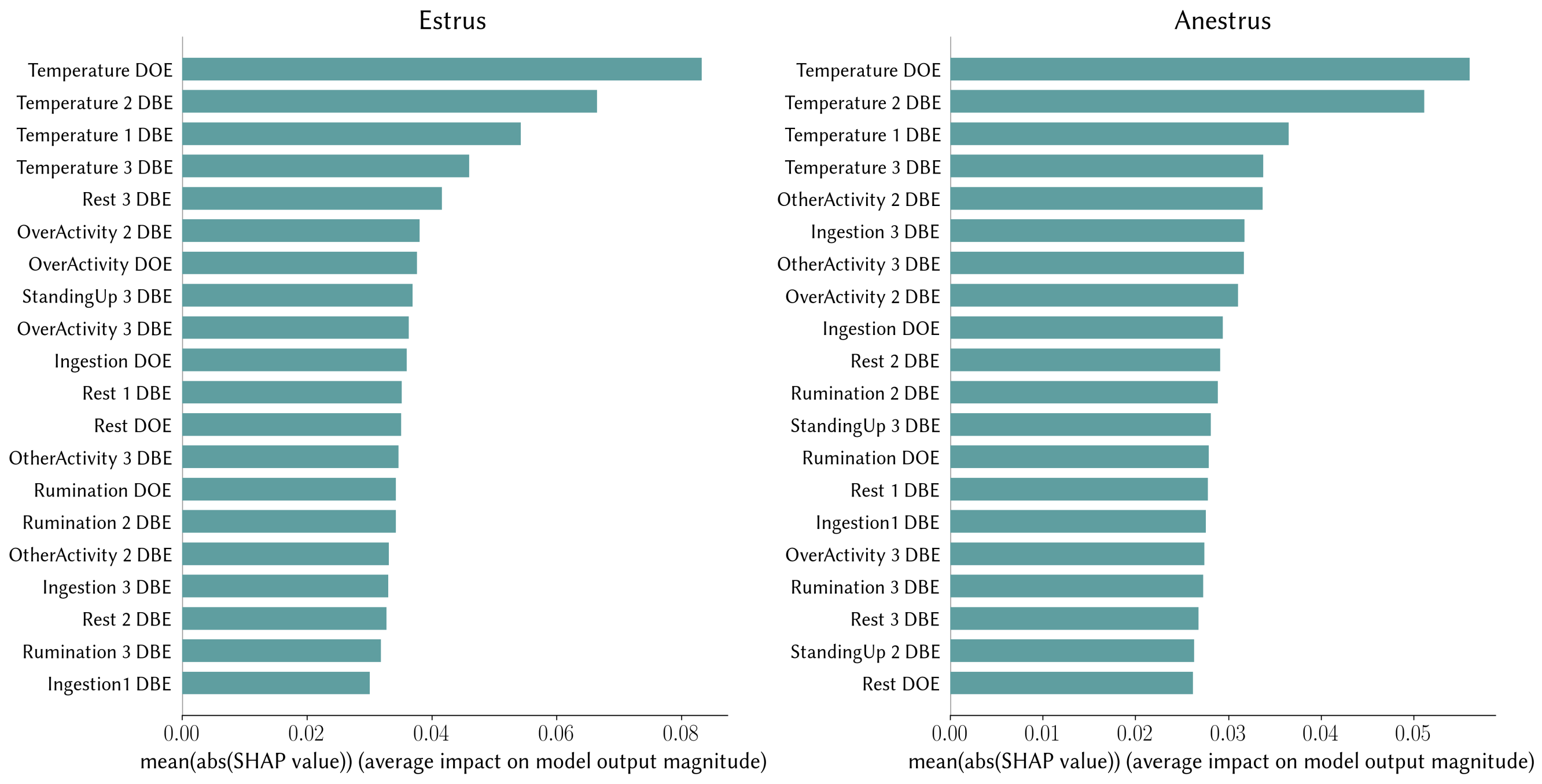
En primer lugar, los resultados confirman el potencial de las técnicas de ciencia de datos para la detección automática del celo frente a la detección visual, como concluye [11]. Observamos que el LCE para la detección de celos tanto conductuales como silenciosos, entrenado en datos de actividad y temperatura, manifiesta un rendimiento significativamente mejor

(Puntuación F1 y menor variabilidad) que Visual y CS (68,9 ± 2,4 frente a 60,4 ± 4,6, 𝑃 <0,05). Nuestro desempeño visual y CS está alineado con el estado de la técnica [25]; la tasa de detección es

ligeramente por debajo del 50% (47%).

En segundo lugar, observamos un mejor rendimiento (puntuación F1 más alta y menor variabilidad) con nuestro algoritmo entrenado en actividad y temperatura que en actividad o temperatura por sí solas.

(68,9 ± 2,4 frente a 67,0 ± 3,0 frente a 55,9 ± 2,3). La diferencia de rendimiento sólo es significativa cuando se compara con el



### Figura 3: Impacto promedio de los atributos en las predicciones del algoritmo para estro y anestro. Abreviaturas: DOE - Día del estro; DBE - Día antes del estro

algoritmo entrenado usando la temperatura. Inferimos que, en las condiciones de nuestro experimento, solo debería desplegarse el sensor de actividad: el rendimiento no es significativamente inferior al entrenado con dos sensores (actividad y temperatura).

No obstante, no se puede excluir la información sobre la temperatura. Observamos una variabilidad marcadamente menor del algoritmo basado en la temperatura a través de los pliegues, lo que permite que el algoritmo basado en la actividad y la temperatura reduzca su variabilidad. Significa que el algoritmo basado en la temperatura es consistente en diferentes datos. Implica un posible poder discriminativo y generalizador superior. Proponemos estudiar más a fondo el potencial de los datos de temperatura para la detección de celos con una heterogeneidad de datos más amplia (raza de vacas, medio ambiente). El siguiente paso consistiría en una asociación con un proveedor de soluciones de detección automática para tener acceso a un conjunto de datos más diverso.

## Interpretabilidad de nuestra solución

En esta sección, presentamos en primer lugar el impacto relativo de las variables en las predicciones de LCE y sus interacciones temporales. Luego, proponemos un enfoque para brindar información sobre la detección de celos a los agricultores en función de estos elementos.

La Figura 3 muestra el impacto promedio de cada variable en las predicciones del algoritmo para el estro y el anestro por orden decreciente.

Estos resultados confirman el poder discriminativo de la temperatura y su potencial para mejorar el rendimiento de detección de celos. La variable con mayor impacto en las predicciones del algoritmo es la temperatura el día del estro tanto para las clases de estro como de anestro.

A continuación, observamos que la clasificación de todas las variables de actividad es diferente con un cambio de clasificación significativo entre estro y anestro. Por lo tanto, el impacto relativo de cada variable de actividad en las predicciones de LCE difiere entre el estro y el anestro. La hiperactividad en el día del estro, una característica típica de la mayoría de los estro (65%), aparece como el tercero más

variable de impacto después del estro por temperatura y no aparece en el top 20 de variables para el anestro.

Al tomar el mismo enfoque de clasificación de impacto a nivel local para el comportamiento frente al celo silencioso, también observamos un cambio significativo en la clasificación de las variables de actividad (75% del cambio de clasificación). La rumia 2 días antes del estro es una variable clave en la detección silenciosa del estro. Es la tercera variable de actividad más impactante para el estro silencioso y aparece en la posición 19 para el estro conductual.

Finalmente, las relaciones temporales entre variables difieren entre el comportamiento y el celo silencioso. Los valores de interacción SHAP revelan que las predicciones del algoritmo se ven más afectadas por las variables de actividad más allá del día del estro para el estro silencioso que el estro conductual. Por ejemplo, la variable de mayor interacción con la rumia en el día del estro es el descanso 3 días antes del estro para el estro silencioso versus el resto 2 días antes del estro para el estro conductual. Esta observación es válida para la sobreactividad, el estar de pie y la ingestión (dos tercios de las variables de actividad).

Por lo tanto, para respaldar las alertas de celo de LCE y facilitar la adopción de la solución, proponemos un enfoque basado en la interpretabilidad de LCE (solo sensor de actividad). Primero, comunique al agricultor la relación de la detección del estro con los casos históricos a través de un indicador de confianza y la amplitud de las diferencias en las 3 variables de actividad más impactantes (descanso 3 días antes del estro, sobreactividad 2 días antes del estro y sobreactividad en el día de estro). El indicador de confianza corresponde al promedio ponderado de las diferencias de valores absolutos de SHAP por la clasificación de las variables de impacto para el estro de nuestra referencia presentada anteriormente. En segundo lugar, en caso de estro, informe al agricultor sobre el tipo de estro (conductual / silencioso) con un nivel de confianza y qué interacciones temporales se satisfacen.

desconfianza. El nivel de confianza se calcula como el anterior pero tomando como referencia el ranking de variables impacto del estro silencioso. Además, las interacciones temporales se comunican en orden decreciente de impacto variable.

## CONCLUSIÓN

Nuestro estudio confirma la mejora significativa del rendimiento de LCE en la detección de celos en comparación con las soluciones comerciales, un resultado impulsado por la detección silenciosa de celos. También demuestra el papel fundamental del despliegue de sensores de actividad en estas detecciones. La interpretabilidad de LCE ofrecida por SHAP, revelando información sobre la relación de las predicciones con casos históricos y la posibilidad de verificar visualmente el estro (conductual versus silencioso), promete reducir la desconfianza de los agricultores. Con respecto al despliegue de nuestra solución, la homogeneidad (raza de vacas, entorno) de nuestro conjunto de datos es una limitación. El siguiente paso consistiría en una asociación con un proveedor de soluciones de detección automática para tener acceso a un conjunto de datos heterogéneo.

## EXPRESIONES DE GRATITUD

WmiAgradecemos a todo el personal técnico de la granja lechera INRA M´ejusseaume que ayudaron en la gestión y seguimiento de esta experimentación a largo plazo. También agradecemos a Medria por su colaboración al proporcionar datos de sensores de actividad y temperatura. Este trabajo fue apoyado por la Agencia Nacional de Investigación de Francia en el marco del Programa Inversiones para el Futuro (ANR-16-CONV-0004), el proyecto nacional francés Deffilait (ANR-15-CE20-0014) y APIS-GENE.

## REFERENCIAS

1. E. Abbasi, M. Shiri y M. Ghatee. 2016. Mezcla raíz-cuatrica de expertos para problemas complejos de clasificación. Sistemas expertos con aplicaciones 53 (2016).
2. I. Adriaens, W. Saeys, T. Huybrechts, C. Lamberigts, L. Franois,

K. Geerinckx, J. Leroy, B. De Ketelaere y B. Aernouts. 2018. Un sistema novedoso para el control de la fertilidad en granjas basado en progesterona de la leche. Revista de ciencia láctea 101, 9 (2018).

1. S. Bascom y A. Young. 1998. Un resumen de las razones por las que los agricultores sacrifican vacas. Revista de ciencia láctea 81, 8 (1998).
2. M. Baydogan y G. Runger. 2014. Aprendizaje de una representación simbólica para la clasificación de series de tiempo multivariante. Minería de datos y descubrimiento de conocimientos 29, 2 (2014).
3. M. Baydogan, G. Runger y E. Tuv. 2013. Marco de conjunto de funciones para clasificar series de tiempo. Transacciones IEEE sobre análisis de patrones e inteligencia de máquinas 35, 11 (2013).
4. J.Bergstra, R. Bardenet, Y. Bengio y B. K´egl. 2011. Algoritmos para la optimización de hiperparámetros. En Advances in Neural Information Processing Systems 24, J. Shawe-Taylor, RS Zemel,

PL Bartlett, F. Pereira y KQ Weinberger (Eds.). Curran Associates, Inc.

1. A. Chanvallon, S. Coyral-Castel, J. Gatien, J. Lamy, D. Ribadud, C. Allain, P. Clment y P. Salvetti. 2014. Comparación de tres dispositivos para la detección automática de estro en vacas lecheras. Theriogenology 82, 5 (2014).
2. T. Chen y C. Guestrin. 2016. XGBoost: un sistema de aumento de árboles escalable. En Actas de la 22ª Conferencia Internacional ACM SIGKDD sobre Descubrimiento de Conocimiento y Minería de Datos.
3. E. Cutullic, L. Delaby, Y. Gallard y C. Disenhaus. 2011. La respuesta reproductiva de las vacas lecheras al nivel de alimentación difiere según

la etapa reproductiva y la raza. Animal 5, 5 (2011).

1. TG Dietterich. 1998. Pruebas estadísticas aproximadas para comparar algoritmos de aprendizaje de clasificación supervisada. Computación neural 10, 7 (1998).
2. KA Dolecheck, WJ Silvia, G.Jr Heersche, YM Chang, DL

Ray, AE Stone, BA Wadsworth y JM Bewley. 2015. Cambios de comportamiento y fisiológicos alrededor de los eventos de celo identificados usando múltiples tecnologías de monitoreo automatizado. Revista de ciencia láctea 98, 12 (2015).

1. R. Ebrahimpour, S. Arani y S. Masoudnia. 2013. Mejora del método de combinación de expertos de NCL utilizando la red de puertas. Computación neuronal y aplicaciones 22, 1 (2013).
2. R. Ebrahimpour, N. Sadeghnejad, S. Arani y N. Mohammadi. 2013. Mezcla precargada de expertos para tareas de clasificación. Computación neuronal y aplicaciones 22, 1 (2013).
3. K. Fujimoto, I. Kojadinovic y JL Marichal. 2006. Caracterizaciones axiomáticas de índices de interacción probabilísticos y cardinal-probabilísticos. Juegos y comportamiento económico 55, 1 (2006).
4. C. Gaillard, H. Barbu, MT Sørensen, J. Sehested, H. Callesen y M. Vestergaard. 2016. Rendimiento de leche y comportamiento estral durante ocho celos consecutivos en vacas Holstein alimentadas con dietas estandarizadas o de alta energía y agrupadas según los cambios de peso vivo en la lactancia temprana. Revista de ciencia láctea 99, 4 (2016).
5. J. Gama y P. Brazdil. 2000. Generalización en cascada. Aprendizaje automático 41, 3 (2000).
6. I.Karlsson, P. Papapetrou y H. Bostr¨om. 2016. Bosques de shapelet aleatorios generalizados. Minería de datos y descubrimiento de conocimientos 30, 5 (2016).
7. J. Krieter. 2005. Detección de celo en vacas lecheras mediante gráficos de control y redes neuronales. En Actas de la 56ª Reunión Anual de la Asociación Europea para la Producción Animal. Comisión de Producción Ganadera.
8. S. Lundberg y S. Lee. 2017. Un enfoque unificado para interpretar predicciones de modelos. Avances en sistemas de procesamiento de información neuronal 30, I. Guyon, UV Luxburg, S. Bengio,

H. Wallach, R. Fergus, S. Vishwanathan y R. Garnett (Eds.). Curran Associates, Inc.

1. S. Masoudnia y R. Ebrahimpour. 2014. Mezcla de expertos: una encuesta bibliográfica. Revisión de inteligencia artificial 42, 2 (2014).
2. T. Miller. 2017. Explicación en Inteligencia Artificial: Insights de las Ciencias Sociales. CoRR abs / 1706.07269 (2017).
3. RS Mitchell, RA Sherlock y LA Smith. 1996. Una investigación sobre el uso del aprendizaje automático para determinar el estro en vacas. Computadoras y electrónica en la agricultura 15, 3 (1996).
4. MA Nielsen. 2015. Redes neuronales y aprendizaje profundo. Determinación Presione.
5. MA Palmer, G. Olmos, LA Boyle y JF Mee. 2010. Detección de celo y características del celo en vacas Holstein-Friesian alojadas y en pastoreo. Theriogenology 74, 2 (2010).
6. OA Peralta, RE Pearson y RL Nebel. 2005. Comparación de tres sistemas de detección de celos durante el verano en un gran hato lechero comercial. Ciencia de la reproducción animal 87, 1 (2005).
7. MT Ribeiro, S. Singh y C. Guestrin. 2016. ¿Por qué debería confiar en ti ?: Explicando las predicciones de cualquier clasificador. En Actas de la 22ª Conferencia Internacional ACM SIGKDD sobre Descubrimiento de Conocimiento y Minería de Datos.
8. M. Saint-Dizier y S. Chastant-Maillard. 2012. Hacia una detección automatizada de estro en ganado lechero. Reproducción en animales domésticos 47, 6 (2012).
9. T. Searchinger, R. Waite, C. Hanson, J. Ranganathan, P. Dumas y E. Matthews. 2018. Creando un futuro alimentario sostenible. Instituto de Recursos Mundiales.
10. M. Sesmero, A. Ledezma y A. Sanchis. 2015. Generación de conjuntos de clasificadores heterogéneos usando Stacked Generalization. Revisiones interdisciplinarias de Wiley: minería de datos y descubrimiento del conocimiento 5, 1 (2015).
11. W. Steeneveld y H. Hogeveen. 2015. Caracterización de granjas lecheras holandesas utilizando sistemas de sensores para el manejo de vacas. Revista de ciencia láctea 98, 1 (2015).
12. H. Zou y T. Hastie. 2005. Regularización y selección de variables vía Elastic Net. Revista de la Royal Statistical Society. Serie B (Metodología estadística) 67, 2 (2005).