

Detección de ganado con disfunciones motoras mediante técnicas de aprendizaje máquina. Propuesta de plan de negocio y viabilidad del Proyecto

Autor: Lucas González Herrera

Tutorizado por:

Moisés Díaz Cabrera

Miguel Ángel Ferrer Ballester

Jorge Marín Rodríguez Díaz

**Programa de Doble Titulación: Grado en Ingeniería Informática
y Grado en Administración y Dirección de Empresas**

Julio 2023

Agradecimientos

A mis tutores, por acompañarme durante todo este proyecto y ayudarme a seguir adelante cuando las circunstancias se tornaban desfavorables.

A mi familia, por ser un pilar fundamental en mi educación y desarrollo personal, ayudándome a convertirme en la persona que soy hoy.

A Daniel, Carlos y Roberto, por compartir conmigo cada momento desde hace ya más de siete años, sin ellos nada de esto hubiera sido posible. Lo conseguimos.

A mi novia, por ser mi apoyo y mi luz en los buenos y los malos momentos, siendo siempre ejemplo e inspiración para mí.

Resumen

El análisis del movimiento ha adquirido relevancia recientemente, gracias a estudios relacionados, en gran parte, con la salud y el rendimiento humano. En el presente trabajo, se analiza el movimiento del ganado bovino para la detección de disfunciones motoras. A través de los datos capturados por un sensor inercial colocado en las patas traseras de los animales y mediante técnicas de aprendizaje máquina, se generará una clasificación automática de las reses en dos grupos diferenciados. A su vez, se desarrollará un plan de negocio que, basado en el modelo *Canvas*, será el primer paso para comercializar el producto desarrollado realizando, además, un análisis de la viabilidad tecnológica, comercial y económico-financiera, entre otros aspectos de interés.

Palabras clave: análisis de movimiento, disfunciones motoras, ganado bovino, sensores inerciales, aprendizaje máquina, plan de negocio, modelo *Canvas*, viabilidad

Abstract

Movement analysis has recently gained relevance thanks to studies largely related to human health and performance. In the present work, the movement of cattle is analyzed for the detection of motor dysfunctions. Using the data captured by an inertial sensor placed on the hind legs of the animals and by means of Machine Learning techniques, an automatic classification of the cattle into two different groups will be generated. In turn, a business plan will be developed which, based on the Canvas model, will be the first step to commercialize the developed product, also carrying out an analysis of the technological, commercial, and economic-financial viability, among other aspects of interest.

Keywords: motion analysis, motor dysfunctions, cattle, inertial sensors, machine learning, business plan, Canvas model, viability, business plan, viability

Guía de lectura

El presente documento se corresponde con la memoria de un Trabajo de Final de Grado del Programa de Doble Titulación del Grado en Ingeniería Informática (GII) y el Grado en Administración y Dirección de Empresas (GADE), por lo que en su desarrollo engloba secciones referentes a una u otra titulación.

Esta guía de lectura se presenta para facilitar la lectura de esta memoria e indicar los epígrafes en los que se exponen los conocimientos sobre cada una de las titulaciones mencionadas. Se recomienda de cualquier manera la lectura completa de este documento para una comprensión más completa del trabajo.

Los cuatro primeros epígrafes, así como el número seis, el número siete y el número 9, son comunes a ambas titulaciones, presentándose por separado los puntos 2.1. referente al GII y el 2.2. al GADE.

Requieren una atención especial con respecto al Grado en Administración y Dirección de Empresas los puntos 5.1 y 5.2 con sus respectivos subapartados, además del punto 8.6.

Requieren una atención especial con respecto al Grado en Ingeniería Informática los puntos 5.4 y 5.5 con sus respectivos subapartados, así como los puntos 8.1, 8.2, 8.3, 8.4 y 8.5.

Índice general

1. Introducción.....	9
2. Competencias cubiertas	12
2.1. Competencias del Grado en Ingeniería Informática	12
2.2. Competencias del Grado en Administración y Dirección	13
3. Aportaciones.....	14
4. Normativa y plan de trabajo	16
4.1. Normativa	16
4.2. Plan de trabajo.....	16
5. Marco teórico.....	18
5.1. Sectores productivos	18
5.1.1. Ganadería.....	20
5.2. Modelo de negocio	23
5.2.1. Modelo <i>Canvas</i>	24
5.3. Desarrollo del negocio	26
5.3.1. Modelo <i>Lean StartUp</i>	27
5.4. Minería de datos.....	28
5.4.1. Análisis de movimiento	32
5.5. Aprendizaje máquina.....	33
5.5.1. Red Neuronal	36
6. Metodología	38
6.1. Comprensión del negocio	40
6.2. Comprensión de los datos	40
6.3. Preparación de los datos	40
6.4. Modelado.....	41
6.5. Evaluación	41
6.6. Despliegue	41

7. Recursos	42
7.1. Datos	42
7.2. Software	44
7.3. Hardware	46
8. Desarrollo de la metodología utilizada	48
8.1. Comprensión del negocio	48
8.2. Comprensión de los datos	49
8.2.1. Contenido de los datos	49
8.2.2. Análisis de los datos	50
8.3. Preparación de los datos	53
8.4. Modelado	55
8.4.1. Modelo completo	56
8.4.2. Modelo filtrado	60
8.5. Evaluación	62
8.5.1. Resultados del modelo completo.....	62
8.5.2. Resultados del modelo filtrado	64
8.6. Despliegue	67
8.6.1. Modelo <i>Canvas</i>	67
8.6.2. Análisis DAFO	71
8.6.3. Análisis de viabilidad y desarrollo del negocio.	73
9. Conclusiones y trabajos futuros	78
10. Fuentes de información	80
11. Anexos	87

Índice de figuras

Figura 1. Características estructurales de las explotaciones bovinas. Fuente: MAPA.....	10
Figura 2. Producción y empleo del sector terciario sobre el total de la economía en España (1970-2015).....	19
Figura 3. Esquema del cálculo de la Renta Agraria. Fuente: Ministerio de Agricultura, Pesca y Alimentación (2018)	21
Figura 4. Porcentaje de usuarios de e-commerce en España 2010-2021.	25
Figura 5. Modelo construir-medir-aprender. Fuente: Bortolini et al. (2021) adaptado de Ries (2011).	28
Figura 6. Crecimiento de usuarios en Internet 2012-2022.....	29
Figura 7. Variación del uso de los datos entre el 2013 y el 2022. Fuente: Data Never Sleeps 10.0.....	29
Figura 8. Volumen de datos digitales (marzo 2021) Fuente: elaboración propia a partir de los datos de Roa (2021).....	30
Figura 9. Transformación de datos para predecir diabetes con SVM. Fuente: Bi et al. (2019, p. 6).....	35
Figura 10. Aproximación del clasificador Naive-Bayes para la predicción de diabetes tipo 2. Fuente: Bi et al. (2019, p. 7)	35
Figura 11. Metodología CRISP-DM. Fuente: Chapman et al. (2000) en Marbán et al. (2009).	39
Figura 12. Niveles jerárquicos del CRISP-DM. Fuente: Chapman et al. (2000).....	39
Figura 13. Descomposición de las fases del CRSIP-DM en los diferentes niveles jerárquicos. Fuente: Marbán et al.(2009).	40
Figura 14. Ficheros que guardan los datos capturados por el sensor inercial. Fuente: elaboración propia.....	42
Figura 15. Ejemplo del desfase que se presenta en una serie de ficheros de datos. Fuente: elaboración propia.....	43
Figura 16. Estructura de carpetas que contienen los datos de las vacas según un ID único. Fuente: elaboración propia.....	43
Figura 17. Índice TIOBE en el mes de mayo de 2023. Fuente: TIOBE.	44
Figura 18. Evolución del índice TIOBE a través de los años . Fuente: TIOBE.....	44
Figura 19. Columnas presentes en los ficheros de datos. Fuente: elaboración propia.	49
Figura 20. Yaw, roll, pitch. Fuente: (Understanding Reference Frames and Device Attitude Apple Developer Documentation (s. f.).	50

Figura 21. Llamada a la función "procesar" sobre cada ventana. Fuente: elaboración propia.	52
Figura 22. Gráficas de los sumatorios de energía de una vaca, con ventanas de 2 y 4 minutos, solapadas 1 y 2 minutos respectivamente. Fuente: elaboración propia.....	52
Figura 23. Gráficas de los sumatorios de energía de una vaca, con ventanas de 10 y 12 minutos, solapadas 5 y 6 minutos respectivamente. Fuente: elaboración propia.	53
Figura 24. Procesado de los datos en crudo. Fuente: elaboración propia.....	54
Figura 25. División de los datos en lotes. Fuente: elaboración propia.....	56
Figura 26. Primera parte del desarrollo de una red neuronal para el ajuste global. Fuente: elaboración propia.....	57
Figura 27. División de los datos de cada ventana estudiada en lotes. Fuente: elaboración propia.	61
Figura 28. Modelo Canvas. Fuente: elaboración propia.	70
Figura 29. Análisis DAFO. Fuente: elaboración propia	71

Índice de tablas

Tabla 1. Censos Ganaderos 2020. Fuente: elaboración propia a partir de los datos del MAPA	10
Tabla 2. Plan de trabajo. Fuente: elaboración propia.....	17
Tabla 3. Evolución histórica de la importancia del sector primario en la economía. Fuente: elaboración propia a partir de los datos de Delgado et al. (1994) y del MAPA.	19
Tabla 4. Tipo de ganado en relación a los sectores ganaderos que recoge el MAPA.	21
Tabla 5. Variación de los componentes de la Renta Agraria en los últimos cinco años. Fuente: Elaboración propia a partir de datos del MAPA.....	22
Tabla 6. Especificaciones ordenador de sobremesa personal. Fuente: elaboración propia...	46
Tabla 7. Equipo Facultad de Tecnologías de la Comunicación. Fuente: elaboración propia..	47
Tabla 8. Lista de características extraídas de los datos del acelerómetro. Fuente: Zebin et al. (2017).....	57
Tabla 9. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste global del modelo completo. Fuente: elaboración propia.....	63
Tabla 10. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste individual sobre los datos de cada vaca, dentro del modelo completo. Fuente: elaboración propia	64
Tabla 11. Resultados y media de 3 ejecuciones de la red neuronal sobre las diferentes ventanas de tiempo para analizar y escoger la que presenta mejores resultados. Fuente: elaboración propia.....	65
Tabla 12. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste global del modelo filtrado. Fuente: elaboración propia.....	65
Tabla 13. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste individual sobre los datos de cada vaca, dentro del modelo filtrado. Fuente: elaboración propia	66
Tabla 14. Variables económicas iniciales del proyecto. Fuente: elaboración propia.....	74
Tabla 15. Ganado bovino por explotación en cada CCAA en 2023. Fuente: elaboración propia a partir de los datos del MAPA.	75
Tabla 16. Desarrollo de los elementos necesarios para el cálculo de payback, VAN y TIR. Fuente: elaboración propia.....	87

1. Introducción

No es un secreto que, desde hace ya varias décadas, el dibujo clásico de una economía con tres sectores bien definidos y que se repartían la importancia a la hora de aportar riqueza a un país se ha desdibujado por completo, como se muestran González-Díez y Moral-Benito (2019). El avance imparable del sector servicios, en un movimiento global de terciarización de la economía ha llevado al sector primario a una presión asfixiante por parte de los compradores o intermediarios.

Ante esta realidad son muchos los ganaderos que han tenido que cerrar sus explotaciones y que han llevado a cabo huelgas y manifestaciones con el propósito de buscar un trato digno a su trabajo. Muchas veces pasa por alto el hecho de que el sector primario es un sector esencial no solo para la economía sino, directamente para nuestra propia supervivencia.

No se debe entender la producción agraria como un elemento destinado únicamente a su comercialización y posterior venta para su uso y consumo, sino como una necesidad básica del ser humano (en cuanto a alimentación se refiere) que, si la situación sigue como hasta ahora, dentro de unos años podría escasear, lo cual tendría unas terribles consecuencias sobre la vida tal y como la conocemos.

Los recursos se agotan, los campos se secan, las explotaciones quiebran y la vida sigue como si no pasara nada, sostenida sobre un actividad consumista muy alejada de la sostenibilidad a día de hoy y que solo se podrá frenar a través de una concienciación eficaz y una acción precisa sobre diferentes factores como bien describe Ávila-Foucat (2017)

Hay que tener en cuenta también que, mientras la economía presentaba esta tendencia a lo largo de los años, se han ido produciendo una serie de avances tecnológicos que han cambiado la perspectiva y la forma en la que entendemos el mundo.

Desde la *World Wide Web* hasta *Chat GPT*, han visto la luz numerosas herramientas que han ayudado al ser humano a realizar nuevas tareas o a automatizar otras ya existentes, además de brindar acceso a un sinfín de conocimientos y componentes que han hecho la vida más fácil en campos muy diversos como, por ejemplo, la innovación y la ciencia, según Andrade (2022).

Es importante tener en cuenta que todos estos avances no están exentos de riesgos, no solo por pasar de procesos conocidos y controlados a otros en los que somos mucho más vulnerables, sino por el hecho de que lo que se concibe como un instrumento positivo, enriquecedor y con

el objetivo de ayudar a las personas puede, y es utilizado, para delinquir y atentar contra la libertad del ser humano. Un ejemplo claro puede ser el caso de la dinamita, creada por Alfred Nobel con la idea de ayudar en tareas mineras y en la construcción de infraestructuras y utilizada muchas veces con fines muy diferentes

También es cierto que, en numerosas ocasiones, estos avances se han orientado o se han desarrollado en exclusiva para actividades del sector servicios, teniendo un impacto mínimo o nulo en el panorama agrario. Esto tiene sentido teniendo en cuenta la expansión de un sector y la retracción del otro, además de otros factores como la dinamicidad o la capacidad de adaptación de los diferentes sectores económicos.

En este sentido, este proyecto busca desarrollar una tecnología que pueda ayudar a las explotaciones ganaderas a conocer de forma precisa el estado de sus animales, pudiendo trabajar anticipadamente buscando su bienestar, a la vez que aumenta la rentabilidad y posibilita la reducción de costes asociados a las visitas veterinarias.

Aunque este trabajo se enfoca desde una fase de prototipado del producto y centra su estudio en el ganado bovino, aporta las claves necesarias para extender su uso a otros animales. La Tabla 1 muestra los datos referentes al número de cabezas de ganado bovino español en el año 2020 así como la variación tanto con el año anterior como con el promedio de los cinco años anteriores, lo que permitirá realizar una primera aproximación a la realidad de este trabajo.

Tabla 1. Censos Ganaderos 2020. Fuente: elaboración propia a partir de los datos del MAPA

N.º de cabezas (miles)	2019	2020	Var.(%) 20/19	Promedio 2015-2019	Var.(%) 2020/ media 15-19
Bovino	6600	6636	0,5	6415	3,4
Vacas ordeño	813	811	-0,2	826	-1,9

Además, la Figura 1 muestra un extracto del Informe Anual de Indicadores de Agricultura, Pesca y Alimentación, donde se visualizan las características estructurales de las explotaciones por orientación tecno-económica. (Ministerio de Agricultura, Pesca y Alimentación, 2022).

Orientación Técnico-Económica	Nº Explotac.	SAU (ha)	SAU/expl.	Producción estándar (M€)	PE(€)/expl.
Bovinos de leche	12.508	435.314	34,8	2.147	171.652
Bovinos de carne y cría	53.302	3.519.973	66,0	2.688	50.432
Bovinos de carne y leche y cría	3.016	84.468	28,0	124	41.209

Figura 1. Características estructurales de las explotaciones bovinas. Fuente: MAPA

La base del proyecto es el análisis de movimiento de vacas lecheras, estudiando los datos capturados por un sensor inercial a través de un proceso de minería de datos que se describirá a lo largo de los siguientes apartados de este documento.

Dentro de este proceso se utilizan técnicas de aprendizaje máquina para poder realizar una clasificación de los animales en dos grupos (clasificación binaria) y conocer de esta forma si una vaca se encuentra sana o enferma.

Además, este proyecto incluye la propuesta de un plan de negocio que pueda ser utilizado para sacar al mercado la tecnología desarrollada, teniendo como premisa la idea de que se trata de un avance que trata de ayudar a mejorar las situaciones que sufren, no solo de los ganaderos, sino también los ganados.

Este plan de negocio se fundamentará en el *Canvas* de Osterwalder (Osterwalder & Pigneur, 2010), una herramienta que permitirá desarrollar una propuesta integral a través del análisis de nueve elementos que se entrelazan para representar la situación de la empresa

Cabe destacar que una empresa no puede funcionar únicamente con un modelo de negocio, sino que necesita desarrollarlo en acciones concretas que le permitan poner en marcha esa propuesta. En este caso se sigue el modelo *Lean StartUp* para este fin, añadiendo también otros elementos de análisis como pueden ser el DAFO, el *payback*, el VAN o la TIR.

En definitiva, se trata de un proyecto que, apoyado en los avances tecnológicos, pretende potenciar una actividad como la ganadería, a través de una mejora del rendimiento y del bienestar animal, tratando de buscar la sostenibilidad sin dejar de lado la rentabilidad tanto del propio producto como de los potenciales clientes, aportando un valor añadido a la sociedad.

Es necesario dejar constancia de que este trabajo ha sido realizado en colaboración con dos proyectos de investigación, por los que se justifica la cotutorización en lo que al Grado en Ingeniería Informática se refiere. Se trata de un proyecto de investigación asociado al Ministerio de Universidades con referencia PID2019-109099RB-C41 por el que se justifica la tutorización del Dr. Miguel Ángel Ferrer Ballester y otro asociado a la Fundación CajaCanarias y Caixa Bank con referencia 2019SP19, que justifica la tutorización del Dr. Moisés Díaz Cabrera.

Además, este proyecto como Trabajo de Final de Grado, abarca de forma global y con desarrollos concretos los conocimientos adquiridos durante seis años en la Universidad de Las Palmas de Gran Canaria, a través del Doble Grado en Ingeniería Informática y Administración y Dirección de Empresas.

2. Competencias cubiertas

2.1. Competencias del Grado en Ingeniería Informática

El documento “*Objetivos y Competencias del GIP*” de la ULPGC, en su punto número dos, recoge las diferentes competencias que se deben adquirir a lo largo de esta titulación (ULPGC, s. f.), comenzando con las competencias generales, siguiendo con las competencias nucleares de la Universidad de Las Palmas de Gran Canaria (ULPGC), para finalizar con las competencias del Título. El citado documento, en el apartado tres del segundo punto, “Competencias del Título”, añade además que las competencias mencionadas se desarrollan en competencias específicas para cada bloque (ULPGC, s. f.).

Se recogen a continuación las competencias específicas relacionadas de forma más directa con el trabajo desarrollado y cómo se han cubierto con este TFG

- » **FB05 y FB06.** Cubierta en la propia naturaleza del trabajo ya que, además de programar y trabajar con diferentes sistemas, se trata directamente la gestión empresarial.
- » **TFG01.** Es la competencia específica relacionada directamente con la realización, presentación y defensa de este trabajo, por lo que, el desarrollo de este la justifica por sí mismo.
- » **CII06 y CII15.** Cubiertas en el modelado, desarrollando algoritmos básicos de sistemas inteligentes.
- » **CP03 y CP07.** Cubiertas en la propia naturaleza del trabajo, así como en la aplicación de técnicas de aprendizaje máquina.

2.2. Competencias del Grado en Administración y Dirección

En lo referente al Grado en Administración y Dirección de Empresas, la propia página de la Facultad de Economía, Empresa y Turismo, en su sección referente al grado mencionado, recoge en el apartado “Proyectos docentes, justificación, objetivos y resultados” un desplegable donde se recogen los objetivos y las competencias del título.

Se presentan en ese espacio, además de las competencias básicas, las competencias del título de Grado en Administración y Dirección de Empresas por la Universidad de Las Palmas de Gran Canaria que, en este caso, se han organizado en torno a cuatro bloques (FEET-ULPGC, 2023).

Al igual que en la Grado en Ingeniería Informática, en primer lugar, se encuentran las competencias nucleares, pasando luego a las generales y, por último, las específicas, de las que se recoge a continuación, aquellas que se han desarrollado de forma más específica a través de este trabajo.

- » **CE1 y CE3.** Competencias directamente relacionadas con la realización y defensa de este trabajo, cubiertas por el desarrollo de las mencionadas fases.
- » **CE7.** Cubierta con la investigación desarrollada en el punto cinco de este trabajo, trabajando sobre las explotaciones ganaderas y sobre las propuestas referentes a la tecnología desarrollada.
- » **CE9.** Cubierta con la propuesta realizada con el fin de mejorar diferentes aspectos económico-sociales de la actividad ganadera.
- » **CE10.** Cubierta con el desarrollo del punto 5.1 del presente documento, donde se recoge la realidad del sector ganadero y se llevan a cabo diferentes valoraciones.

3. Aportaciones

Este trabajo se entiende, no como un trámite para poder superar una etapa académica, sino como una oportunidad para demostrar los conocimientos adquiridos después un largo periodo de estudio. También se busca que puedan plasmarse en una propuesta que, realmente, tenga un valor significativo dentro de la sociedad actual y pueda ayudar a seguir desarrollando herramientas que beneficien a todos. Con esta premisa se presentan una serie de aportaciones que en este trabajo se hacen visibles de manera explícita o implícita, a través de los diferentes elementos desarrollados.

- » **Búsqueda del bienestar animal.** Es importante destacar el hecho de que, en nuestra sociedad, cada día se pone más en valor el paradigma del bienestar animal por lo que nuestra tecnología puede ser una pieza clave a la hora de hacer llegar esta búsqueda, directamente a las explotaciones ganaderas de una manera sencilla y sin suponer un sobrecoste elevado. Aumentar de manera significativa el bienestar animal, no tiene por qué estar relacionado con un gasto elevado en diferentes factores, más bien, en la correcta distribución de los fondos disponibles para alcanzar los mejores resultados posibles. Esta aportación presenta mayor importancia si cabe tras la publicación de la Ley 7/2023, de 28 de marzo, de protección de los derechos y el bienestar de los animales, por parte del gobierno español.
- » **Transformación digital del sector primario.** En el momento de transformación y expansión digital en el que se encuentra nuestra sociedad, la digitalización se presenta casi como una necesidad a la hora de desarrollar cualquier actividad económica. Aunque el sector primario haya ido menguando su importancia con el paso de los años, no es una excepción, por lo que se deben buscar vías adecuadas para llevar a cabo este proceso. Esto es exactamente lo que propone este trabajo, que, a través de acciones sin un impacto elevado, llevan a cabo una transformación capaz de conseguir unos resultados con una importancia capital para este sector y para la sociedad en general.
- » **Aumento de la productividad de las explotaciones ganaderas.** El bienestar animal no supone solo una ventaja para los propios animales, sino que, si estos se encuentran en condiciones óptimas, serán más productivos, lo que beneficiará de igual manera a los ganaderos. Teniendo esto en cuenta, se nos presenta un escenario más que provechoso para el desarrollo de la tecnología propuesta ya que beneficia a la gran mayoría de grupos de interés que está presentes en nuestro marco de actuación. Hay que tener en cuenta de igual

manera, que no supone una dificultad insalvable para los ganaderos, sino que, con unas simples indicaciones podrán comprender su funcionamiento y usarlo para su beneficio sin ningún tipo de problema más allá de las propias deficiencias que pudiera presentar la tecnología en un momento dado.

- » **Desarrollo de la tecnología basada en los sensores inerciales.** Los sensores inerciales son ampliamente utilizados en tecnologías de análisis del movimiento o simplemente en dispositivos electrónicos que los incluyen con diferentes finalidades. Esta tecnología se presenta como un factor clave en una gran cantidad de campos de investigación, donde este trabajo aportará su granito de arena. Además, abre una vía de trabajo sobre la detección de enfermedades a través de los datos capturados con estos sensores, en este caso aplicada a animales de ganado, pero que puede ser extrapolada a otro tipo de animales en un primer momento e incluso a personas en un futuro, buscando una mejora de la calidad de vida.

4. Normativa y plan de trabajo

4.1. Normativa

Este documento recoge el desarrollo de Trabajo de Fin de Grado del Doble Grado en Ingeniería Informática y Administración y Dirección de Empresas por la Universidad de Las Palmas de Gran Canaria. Al tratarse de un documento de carácter oficial, está sujeto a una normativa específica, que en este caso viene dictada por la ULPGC y se refrenda en las normas que presentan, acerca de proyectos de esta índole, tanto la Escuela de Ingeniería Informática, como la Facultad de Economía Empresa y Turismo de la mencionada universidad.

A continuación, se referencian los documentos oficiales que contienen las normas de obligado cumplimiento sobre el desarrollo de este trabajo:

- » Normativa de desarrollo de las especificidades de los Trabajos de Fin de Título, de la Escuela de Ingeniería Informática (EII-ULPGC, 2020).
- » Reglamento para la realización y evaluación de Trabajos de Fin de Título, de la Facultad de Economía, Empresa y Turismo (FEET, 2020).
- » Reglamento General de los Trabajos de Fin de Grado y de Fin de Máster de las Titulaciones Oficiales de la ULPGC (ULPGC, 2019).

Por su parte la capturas de datos previa al desarrollo de este trabajo se rige por la normativa referente a cada proyecto de investigación, así como otra normativa general relacionada con la experimentación en animales que no procede recoger en este apartado.

4.2. Plan de trabajo

A lo largo de todo el proceso de creación de este trabajo se han confeccionado diferentes documentos para cumplir con los requisitos que marca la normativa. En uno de estos documentos, concretamente en el modelo TFT-01, se presentaba una tabla que marcaba de forma estimada las tareas a realizar para la consecución de este proyecto, así como su duración en horas.

Teniendo en cuenta que todo este trabajo se ha ido realizando a lo largo de varios meses, es necesario presentar un ajuste sobre las duraciones de algunas fases que previamente se habían estimado y que ahora se conocen de forma más detallada. Toda esta información viene recogida en la Tabla 2.

Tabla 2. Plan de trabajo. Fuente: elaboración propia.

<i>Fases</i>	<i>Duración Estimada</i>	<i>Duración Real</i>	<i>Tareas</i>
Estudio previo / Análisis	50	40	Familiarización con el proyecto
			Familiarización con los datos y las herramientas
			Estudio del sector
Diseño / Desarrollo / Implementación	110	120	Tratamiento de datos
			Programación de los diferentes modelos
			Diseño del modelo <i>Canvas</i>
Evaluación / Validación / Prueba	70	80	Evaluación del sistema desarrollado
			Análisis de la viabilidad
Documentación / Presentación	70	60	Búsqueda de fuentes bibliográficas
			Redacción de la memoria
			Elaboración de las presentaciones

Como se puede observar, la duración estimada de algunas actividades ha sufrido variaciones, incrementándose las fases de desarrollo y evaluación, en detrimento de las fases de análisis y presentación en este caso.

En cuanto al análisis, se ha centrado sobre todo en la comprensión del problema a tratar, de los datos sobre lo que desarrollar el trabajo y sus posibilidades de tratamiento, así como un estudio profundo de la ganadería en nuestro país. Además, esta actividad está ligada con la de documentación ya que, durante la misma, ha sido necesario consultar diferentes fuentes de información que más tarde se han utilizado en diferentes apartados de este documento.

Por su parte, las fases de desarrollo y evaluación han requerido de una mayor carga horaria debido a la infinidad de modelos que se podían plantear y que era necesario probar para conocer cuál se ajustaba mejor al problema planteado.

5. Marco teórico

A continuación, se desarrollará el marco teórico del trabajo donde se presentará la bibliografía sobre la que se sustenta este Trabajo de Final de Grado. Se debe tener en cuenta que, aunque se trate de una creación personal del autor, se basa en estudios y artículos de otros autores que han trabajado en campos afines a los tratados en este proyecto.

5.1. Sectores productivos

De modo introductorio y con el fin de definir de forma más precisa el objeto de estudio de este trabajo, será necesario presentar las diferentes realidades que lo componen.

Nuestra economía se divide comúnmente en grupos de actividades o sectores económicos que, como se puede observar en la mayoría de literatura que hace referencia a estos y a la utilizada en este trabajo, son tres: sector primario, también denominado sector agrario (Cuadrado Roura, 2016); sector secundario, también denominado sector industrial (Cuadrado Roura, 2016) y sector terciario, también denominado sector servicios. En este caso, nuestra actividad estudiada (ganadería) se enmarca dentro del sector primario, junto con otras como la agricultura o la pesca. Concretamente, este sector comprende según Álvarez et al. (2013) y Frontela Sánchez (2022) aquellas actividades económicas dedicadas a la adquisición y transformación de los recursos que ofrece el medio natural en productos sin procesar, para destinarlos al consumo directo o la industria.

Históricamente el sector primario, también conocido como sector agroalimentario o agrario (no se debe confundir el término agrario y agrícola: el sector agrario abarca tanto al agrícola como el ganadero (Álvarez et al., 2013)), ha sido el más importante dentro de las economías mundiales, pero desde hace 70 años, ha dado paso a una transformación económica que lo ha llevado a una situación bien distinta. Esta transformación se ilustra en la Tabla 3, a través de la evolución del sector primario en su aportación al PIB y a la población activa de nuestro país.

Tabla 3. Evolución histórica de la importancia del sector primario en la economía.
Fuente: elaboración propia a partir de los datos de Delgado et al. (1994) y del MAPA.

AÑOS	PIB agrario respecto al PIB total (%)	Población activa agraria respecto a población activa total (%)
1950	29,9	49,6
1960	23,7	39,7
1970	11,3	24,8
1980	7,4	15,7
1993	3,5	9,0
2021	2,7	3,8

Se puede observar como el peso del sector primario en la economía española ha ido disminuyendo con el paso de los años observándose caídas constantes y medianamente proporcionales en ambos factores. Ha pasado de representar el 30% del PIB total y ocupar casi al 50% de la población a no llegar al 4% en ninguno de estos aspectos.

Este hundimiento del sector primario no ha tenido lugar únicamente en nuestro país, sino que venido siendo una tendencia mundial en los últimos tiempos como consecuencia de una economía cada vez más globalizada. Este fenómeno viene explicado en González-Díez y Moral-Benito (2019), donde se describen diferentes fases comunes en muchas economías, entendiendo la terciarización como un proceso que se replica.

Cabe destacar también como esta caída del sector primario ha tenido una respuesta clara, pero en sentido contrario, por parte del sector terciario o sector servicios que ha visto como el valor de su producción ha incrementado su porcentaje en el PIB nacional a precios corrientes cerca de un 30% en los últimos cincuenta años como muestra Cuadrado Roura (2016).

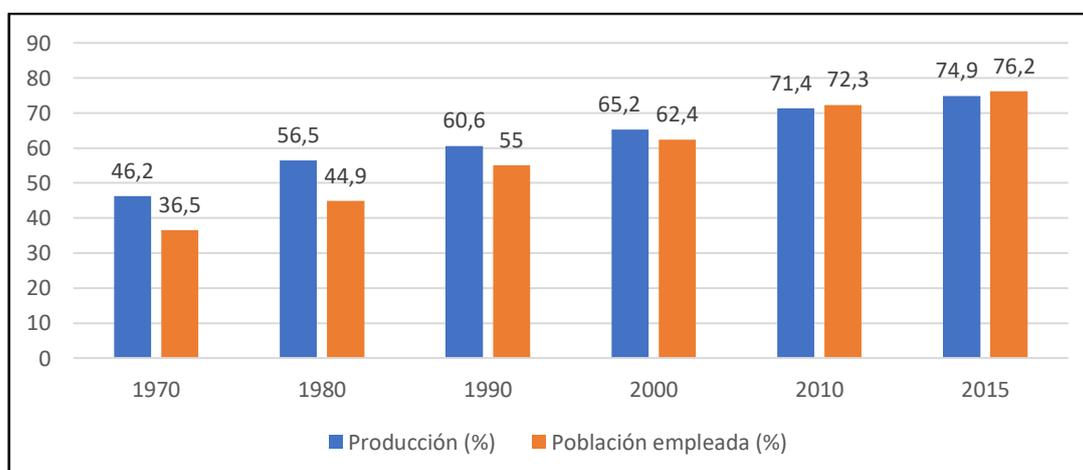


Figura 2. Producción y empleo del sector terciario sobre el total de la economía en España (1970-2015)

Fuente: elaboración propia a partir de los datos de Cuadrado Roura (2016)

En el caso de la Comunidad Autónoma de Canarias, esta terciarización es incluso más alarmante que en el resto del territorio nacional ya que, según datos de Lorenzo Díaz et al. (2021), la Comunidad Autónoma de Canarias es, junto con Baleares y Madrid, la comunidad donde el sector servicios tiene una mayor importancia dentro del PIB (77,7% en los territorios insulares y 76,7% en la capital). Además, como se puede observar también en Lorenzo Díaz et al. (2021), los sectores primario e industrial en Canarias están muy por debajo de la media española, 1,5% y 5,5% respectivamente frente al 2,8% y el 14,6%.

5.1.1. Ganadería

Dentro del sector primario, la actividad sobre la que se ha trabajado es la ganadería. Las primeras evidencias de la capacidad del ser humano de domesticar animales datan aproximadamente de hace unos veinte mil años, durante la era Glaciar Tardía, donde cazadores-recolectores europeos y asiáticos utilizaban lobos para la caza, sin suponer grandes cambios en el modo de vida del ser humano. (Vigne, 2011)

La ganadería comúnmente conocida surgiría después, hace aproximadamente unos doce mil años cuando se empezaron a domesticar otras especies en diferentes partes del mundo, dando lugar a profundos cambios en todos los ámbitos de la sociedad de aquella época, como recoge Vigne (2011).

Teniendo claros los orígenes de esta actividad, es necesario conocer los diferentes tipos de ganadería que se dan en la actualidad, pudiendo realizar clasificaciones según el tipo de ganado y el medio en el que se encuentren.

Según el tipo de ganado, tradicionalmente se ha dividido la ganadería en bovina (vacas, bueyes y toros), porcina (cerdos), ovina (ovejas), caprina (cabras), avícola (aves de corral), equino (caballos y yeguas) y cunícola (conejos). En este sentido, el Ministerio de Agricultura, Pesca y Alimentación (en adelante MAPA) da un paso más dividiendo la ganadería en “sectores ganaderos”, dependiendo de la producción que se lleve a cabo con los distintos animales, por lo que nos encontramos con subdivisiones dentro de los grupos tradicionales como se puede ver reflejado en la Tabla 4.

Tipo de ganado	Sectores Ganaderos
Vacas, toros, bueyes, novillos, etc.	Vacuno de carne
	Vacuno lechero
Ovejas, cabras, carneros, corderos, etc.	Ovino y caprino de carne
	Ovino y caprino de leche
Cerdos	Porcino
Aves de corral	Avícola de puesta
	Avícola de carne
Caballos	Equino
Conejos	Cunícola
Abejas	Apícola
Caracoles	Helicícola

Tabla 4. Tipo de ganado en relación a los sectores ganaderos que recoge el MAPA.
Fuente: elaboración propia a partir de los datos del MAPA.

Según datos del MAPA, la producción ganadera generó en nuestro país 20.915 millones de euros en 2021, un 3,7% más que el año anterior. Con este crecimiento superó la media europea (en este caso de la UE-27), un 2,4%, y la de países como Francia (1,7%) o Alemania (-2,5%). Este dato se debe poner en contexto ya que es parte de un conjunto de macromagnitudes que se utilizan para obtener la Renta Agraria de un territorio, que mide la remuneración de los factores de producción, tierra, capital y trabajo (Ministerio de Agricultura, Pesca y Alimentación, 2022). Esta Renta Agraria (RA) tiene la siguiente fórmula:

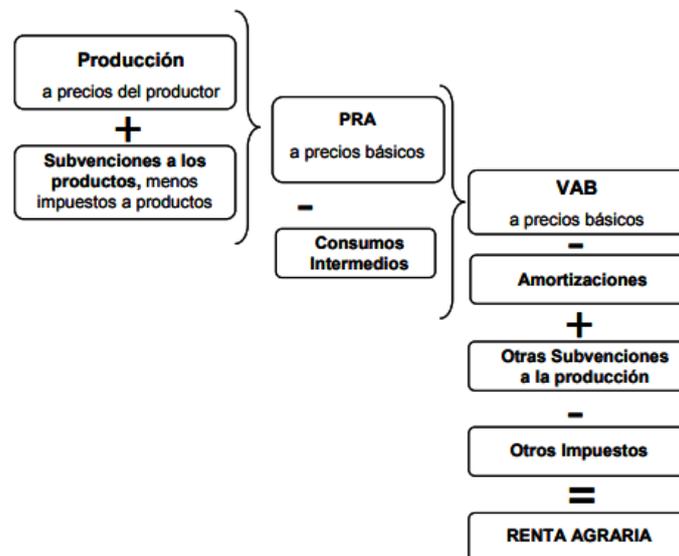


Figura 3. Esquema del cálculo de la Renta Agraria. Fuente: Ministerio de Agricultura, Pesca y Alimentación (2018)

- » **PRA (Producción de la Rama Agraria):** que engloba tanto la producción animal como la producción vegetal.
- » **Consumos Intermedios (CI):** que engloba los insumos y servicios necesarios para alcanzar dicha producción.
- » **VAB (Valor Añadido Bruto):** es el resultado de restar los CI a la PRA

De este modo se puede observar como la producción ganadera no se encuentra como un dato aislado, sino que forma parte de la macroestructura agraria, aportando de forma positiva al crecimiento agrario a través de la subida interanual comentada. Con respecto a años anteriores, la Tabla 5 muestra la variación interanual de todos los componentes de la Renta Agraria.

*Tabla 5. Variación de los componentes de la Renta Agraria en los últimos cinco años.
Fuente: Elaboración propia a partir de datos del MAPA*

	Valores en millones de euros				
	2021	2020	2019	2018	2017
PRA	56.427	52.919	50.657	53.392	49.165
Producción Animal	20.915	20.133	19.597	19.215	18.756
CI	26.774	23.631	24.101	23.205	21.438
VAB	29.653	29.288	26.556	30.187	27.728
RA	28.985	29.123	26.234	30.217	27.831

Como se puede observar, aunque la Producción de la Rama Agraria y la Renta Agraria hayan sufrido variaciones importantes tanto positiva como negativamente, el valor de la producción animal ha mantenido un crecimiento constante, con valores entre el 2% y el 4%, siendo la mayor subida la del año 2021, comentada anteriormente (3,7%).

Hay que tener en cuenta que parte de esta productividad animal se sustenta en el sacrificio de ganado para su consumo como carne. Ante una sociedad cada vez más comprometida con el bienestar de todos los seres sintientes, la ganadería y el consumo de alimentos derivados de esta actividad, se han visto fuertemente criticados por todo lo que suponen. Esta realidad se puede observar en Infante (2019, citado en Luque Cuesta, 2022) que muestra la presión que se lleva a cabo desde diferentes plataformas para minimizar o incluso erradicar, no solo el consumo de carne, sino que buscan una meta vegana.

Este activismo frente a las prácticas ganaderas no se basa únicamente en el rechazo total a las mismas, sino que también existen sectores que buscan “otra forma de hacer las cosas”, por lo que se buscan las claves que permitan obtener una producción más sostenible, teniendo como pilar principal el bienestar animal y que ayude a solucionar problemas ambientales (Garnett et al., 2016 citado en Luque Cuesta, 2022).

5.2. Modelo de negocio

El modelo de negocio es un término empleado en el ámbito de la organización de empresas. Según Teece (2010), todas las empresas emplean explícita o implícitamente un modelo de negocio concreto que describe el diseño de los mecanismos de creación, distribución y captura de valor de esta y que está presente desde su creación. De esta definición podemos extraer una idea clave a la hora de entender la realidad de los modelos de negocio: todas las empresas tienen un modelo de negocio, pero no todas saben que lo tienen o lo tienen formalmente recogido.

Durante años, pequeñas empresas familiares han ido creciendo y convirtiéndose en empresas económicamente estables siguiendo unas ideas o formas de hacer las cosas sin saber que esto hacía referencia a un modelo de negocio. Este modelo de negocio se presenta como una guía mediante la cual una empresa define de qué manera ofrece valor a los clientes, les incita a pagar por ello y convierte esos pagos en beneficios. (Teece, 2010).

El descubrimiento de la importancia que tiene este campo en el éxito de numerosas empresas ha hecho que aumente considerablemente la investigación sobre el mismo, como muestran Foss y Saebi (2017), aunque esta área de estudio se sigue caracterizando por tener una base teórica muy débil y aprovechar ideas de otras áreas para adaptarlas de forma oportunista .

En una situación tan dinámica y cambiante como la que se presenta en la sociedad actual, es sumamente complicado, para la mayoría de las empresas, fijar un modelo de negocio cerrado y que no esté sujeto a modificaciones ya que en gran parte su éxito depende de la dinamicidad de su modelo y de cómo llevan a cabo las adaptaciones necesarias (Bortolini et al., 2021,).

Hay que entender también que una vez que una empresa ha conseguido asentarse y ha encontrado un modelo de negocio funcional, este no estará sujeto a los mismos cambios que pueden darse en empresas más pequeñas o que no han llegado a una situación propicia.

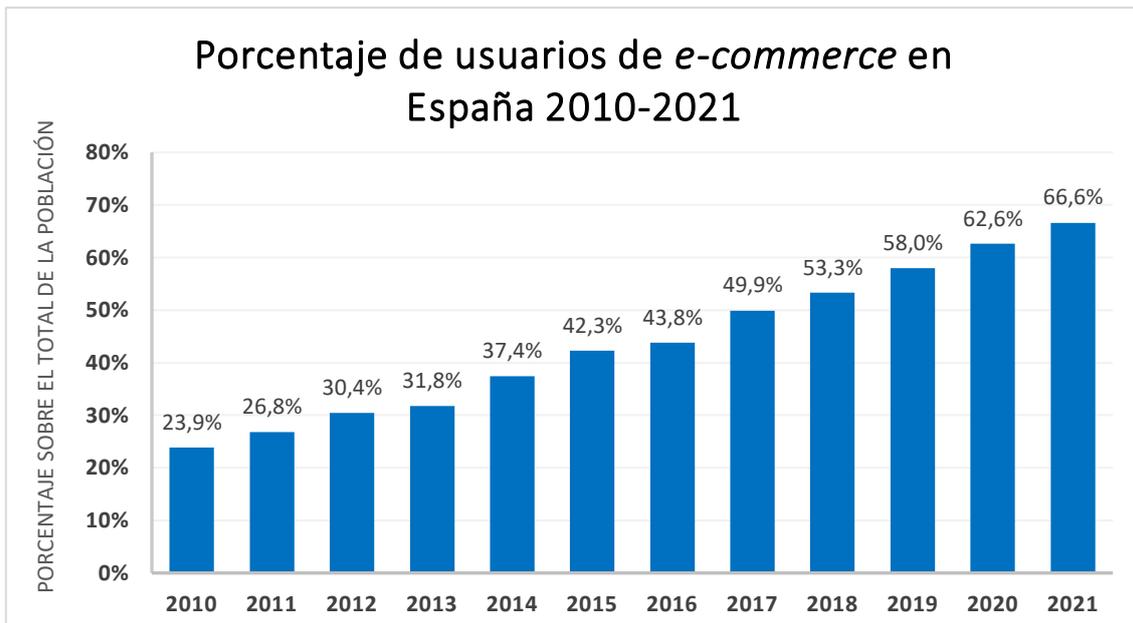
5.2.1. Modelo *Canvas*

En el año 2010, Alexander Osterwalder e Yves Pigneur presentaron el libro *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers* (Osterwalder & Pigneur, 2010) donde se presentaba el Modelo *Canvas*, describiendo el mismo como un lenguaje común sobre el que se pueden analizar diferentes modelos de negocio (Osterwalder & Pigneur, 2010).

En el libro, los autores comienzan describiendo qué es un modelo de negocio (descrito en el punto anterior), pasando luego a presentar las características de su modelo *Canvas* (lienzo). En este sentido, Osterwalder y Pigneur (2010), hablan de un concepto mediante el que es posible no solo analizar el propio modelo de negocio sino también el de competidores o cualquier compañía, a través de un lenguaje común que permite manipular fácilmente los modelos para crear nuevas alternativas.

El Modelo *Canvas* se caracteriza por su composición ya que, según Osterwalder y Pigneur (2010), un modelo de negocio puede explicarse a través de nueve bloques básicos que abarcan las cuatro principales áreas de un negocio: clientes, oferta, infraestructura y viabilidad financiera. Estos bloques son:

- » **Segmento de clientes.** Normalmente un producto no va dirigido únicamente a un cliente y menos aún en una sociedad cada día más globalizada por lo que, este bloque englobará a los diferentes grupos que puedan estar interesados en nuestro producto, teniendo en cuenta las necesidades que pretende satisfacer
- » **Propuesta de valor.** En este caso no se habla de un solo producto o servicio sino de un conjunto de ellos, ya que los autores entienden esta propuesta como una composición de factores que guían a la empresa a la consecución de una propuesta de valor completa y significativa.
- » **Canales.** Engloban los diferentes medios a través de los cuales se lleva a cabo la comunicación, compra-venta y distribución de la propuesta de valor alcanzada por la empresa. A lo largo de los años, los canales han ido sufriendo importantes cambios, guiados por la profunda transformación de la economía y la sociedad en general. Según datos de Statista (2023), el comercio electrónico o *e-commerce* (comercio que se lleva a cabo a través de Internet (Asale, s.f.)) en España, ha aumentado un 42,7% desde el año 2010 como se puede apreciar en la Figura 4.



*Figura 4. Porcentaje de usuarios de e-commerce en España 2010-2021.
Fuente: elaboración propia a partir de los datos de Statista (2023)*

- » **Relación con los clientes.** La relación con los diferentes grupos de clientes es fundamental a la hora de alcanzar acuerdos comerciales provechosos. Es necesario, en muchas ocasiones, diferenciar el perfil de cada cliente con la intención de entablar una relación que lleve no solo a la captación sino, también, a la fidelización de los diferentes grupos.
- » **Fuentes de ingresos.** Al fin y al cabo, el objetivo de cualquier empresa es obtener unos beneficios por lo que las fuentes de ingresos son un elemento clave dentro del modelo de negocio con el fin de asegurar la continuidad de este
- » **Recursos clave.** Toda actividad cuenta con una serie de recursos que son imprescindibles para el correcto funcionamiento de esta y que deben ser marcados como críticos en la planificación de cualquier empresa. Estos recursos no tienen por qué ser solo físicos (materias primas, por ejemplo) sino que puede tratarse de personas, patentes, marcas, etc.
- » **Actividades clave.** Para desarrollar una propuesta de valor que realmente sea significativa, habrá que llevar a cabo una serie de actividades que permitan a la empresa hacer realidad esta propuesta. Estas actividades pueden ir desde el análisis previo a la creación del producto o servicio que se brinda al cliente

- » **Socios clave.** Hoy en día es muy difícil, por no decir imposible, sacar adelante un negocio sin tener apoyos por parte de otros individuos o corporaciones. Estas alianzas pueden crearse buscando economías de escala, acceso a recursos clave, nuevas oportunidades de negocio, etc.
- » **Estructura de costes.** Para que cualquier empresa funcione hay que incurrir en unos costes que serán mayores o menores en función de la actividad a desarrollar. Habrá que tener en cuenta en este caso no solo el valor de estos costes, sino también, si son fijos o variables, por ejemplo.

5.3. Desarrollo del negocio

Una vez se ha definido el modelo de negocio, es necesario conocer la forma en la que una organización va a poner en marcha ese modelo, es decir, que plan va a seguir y que acciones va a desarrollar para llevar a la realidad los diferentes aspectos recogidos en el modelo de negocio.

El lienzo y los diferentes modelos de negocio son un gran punto de partida a la hora de desarrollar la acción empresarial ya que proponen diferentes elementos a los que se debe atender. (Felin et al., 2020). Sin embargo, el primer paso no trata solo de catalogar lo que se ve y hacer un modelo completo, sino que, basándose en una teoría coherente de valor, se debe desarrollar una hipótesis potencialmente valiosa. (Felin et al., 2020).

De lo que no cabe duda es que el modelo de negocio y el desarrollo de este están interconectados dentro de cualquier empresa ya que, un modelo de éxito no sirve de nada si no es probado y validado al igual que tampoco es inteligente hacer pruebas sin tener claros los aspectos claves del negocio, como muestran Felin et al. (2020) dejando claro que aspectos como los grupos de interés son importantes, pero no definen explícitamente la estrategia a seguir por la empresa.

Este desarrollo es un punto clave dentro de cualquier organización, pero lo que está claro es que no en todas se puede dar de la misma manera ya que no es lo mismo tratar con una empresa que ya ha desarrollado un modelo de negocio exitoso, que con nuevas empresas que sólo pueden enfrentarse a una serie de hipótesis no probadas. A la hora de ir probando estas hipótesis, las empresas las revisan continuamente o pivotan hacia otras que se ajusten mejor a su realidad empresarial. (Yang et al., 2019).

Teniendo en cuenta las características del desarrollo que se pretende seguir en este proyecto, será necesario visualizar la forma en que nuevas empresas lanzan al mercado productos innovadores o de nueva creación. Al enfrentarse a hipótesis no probadas, los emprendedores tienen que sortear diferentes obstáculos sobre todo en el momento de creación de nuevas empresas (Souza Silva et al., 2020) que van desde la obtención de capital hasta la progresiva ampliación de la empresa, siguiendo así un camino que no siempre es exitoso.

5.3.1. Modelo *Lean StartUp*

Según Felin et al. (2020), el concepto *lean* (ligero) tiene su origen en el movimiento de calidad total y adquirió una gran importancia gracias al libro de James P. Womack, Daniel T. Jones y Daniel Roos titulado *La máquina que cambió el mundo*. Este libro sobre la producción de automóviles se centra en las técnicas utilizadas por Toyota para mejorar la calidad e implementar un sistema de producción ligero.

Este concepto se ha trasladado después a numerosos sectores dentro de la sociedad, desde el desarrollo de software hasta, como es el caso aquí, el desarrollo de negocios. Cómo ya se ha visto, el desarrollo del negocio se basa en un modelo de negocio que muchas veces es difícil de definir ya que es consecuencia de la aplicación de herramientas que implican ensayos, análisis de datos y pruebas. (Bortolini et al., 2021).

Tradicionalmente, estas fases se han ido desarrollando de manera secuencial desde que surge la idea hasta que se lanza el producto al mercado lo que provoca unos elevados costes sin que se hayan podido obtener datos reales del impacto de este en el segmento de clientes. Además, según Ries (2011), las previsiones solo son precisas si se tienen datos de funcionamiento estables y en un entorno relativamente estático. Estas situaciones definitivamente no se contemplan en una startup

En el año 2011, Eric Ries desarrolló el método conocido como *Lean Startup* (LS), en su libro homónimo, al proponer una metodología de validación de modelos de negocio basada en iteraciones rápidas, aunque ya antes otros autores habían propuesto aproximaciones similares. Esta metodología se centra en un modelo construir-medir-aprender similar al ciclo planificar-hacer-verificar-actuar (PDCA por sus siglas en inglés) de Deming (1986). (Bortolini et al., 2021).

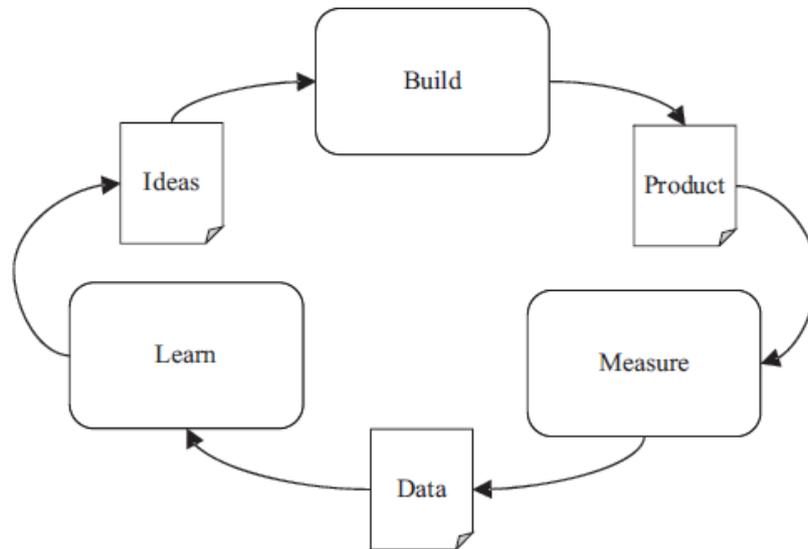


Figura 5. Modelo construir-medir-aprender. Fuente: Bortolini et al. (2021) adaptado de Ries (2011).

5.4. Minería de datos

Según una el científico estadounidense-jordano Usama Fayyad, las bases de datos no paran de crecer, creando la necesidad de nuevas técnicas para analizar estos datos. Aunque son afirmaciones de hace más de un cuarto de siglo, siguen siendo una realidad en la sociedad de hoy, conocida como “sociedad de la información”, término acuñado por el sociólogo estadounidense Daniel Bell en su obra *“El advenimiento de la sociedad post-industrial: un intento de prognosis social.”* (Bell, 1976).

Este hecho se puede observar claramente en la Figura 6, donde se observa cómo no solo los datos sino también el acceso a ellos a través de internet no ha parado de crecer en los últimos 10 años. Los datos recogidos en dicha tabla pertenecen a la infografía *Data Never Sleeps* que realiza la empresa Domo anualmente. En sus publicaciones recoge el uso de los datos en diferentes ámbitos y como estos han ido aumentando, como se puede observar en la Figura 7, donde se comparan la primera y la última edición disponible del mencionado proyecto, quedando claramente evidenciado este crecimiento en el uso y la cantidad de datos necesarios dentro del funcionamiento “normal” de la sociedad.

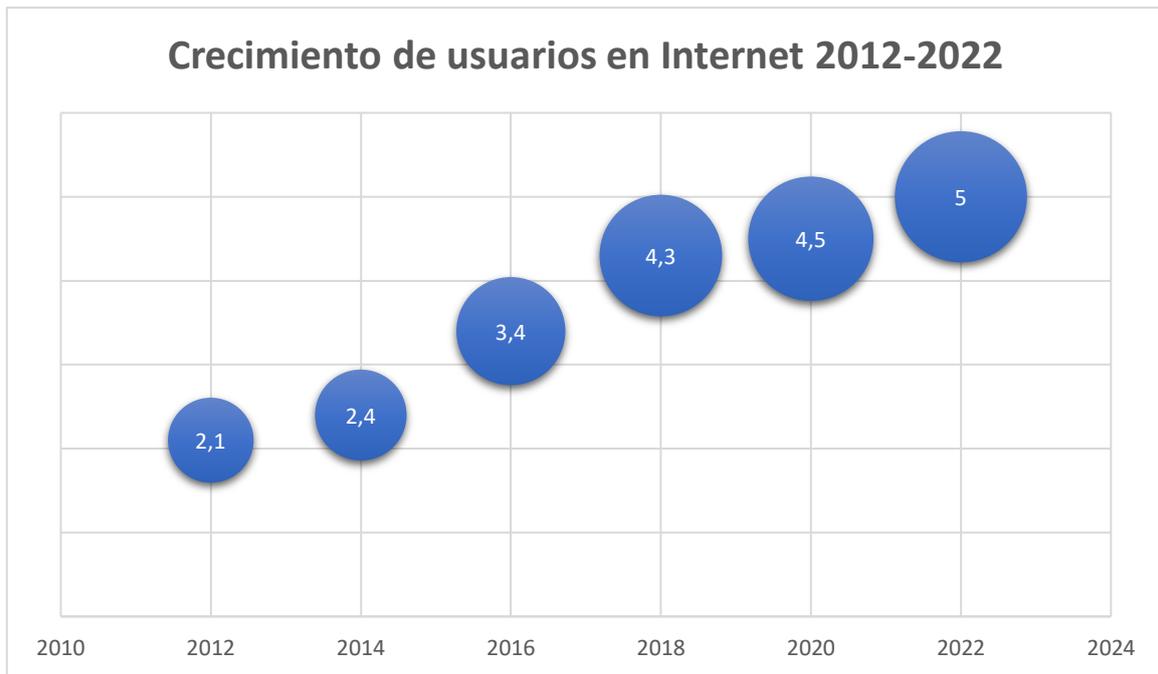


Figura 6. Crecimiento de usuarios en Internet 2012-2022.
Fuente: elaboración propia a partir de los datos de Data Never Sleeps

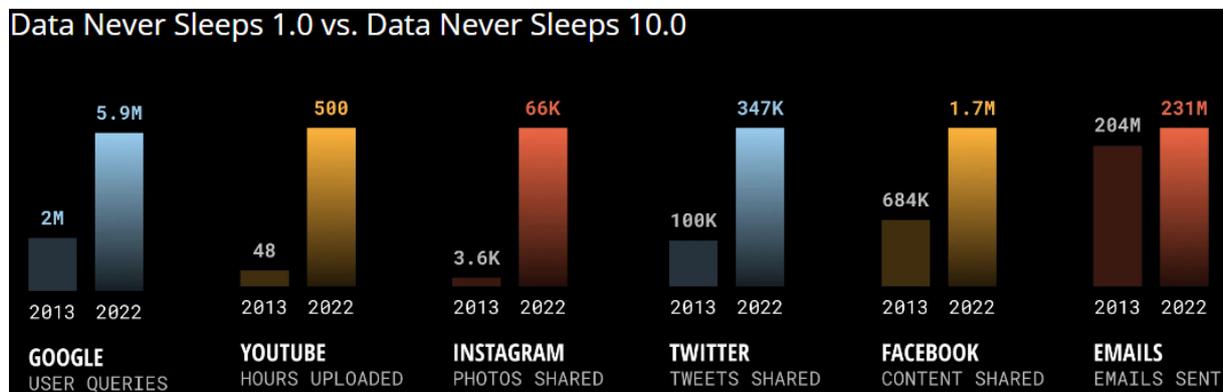


Figura 7. Variación del uso de los datos entre el 2013 y el 2022. Fuente: Data Never Sleeps 10.0

En la Figura 8 se presenta el volumen real y estimado (a fecha de marzo de 2021) de datos digitales creados o replicados en todo el mundo. La medida utilizada en esta medición es el zettabyte que equivale a mil millones de gigabytes, es decir $1 * 10^{18}$ bytes de datos. (Roa, 2021)

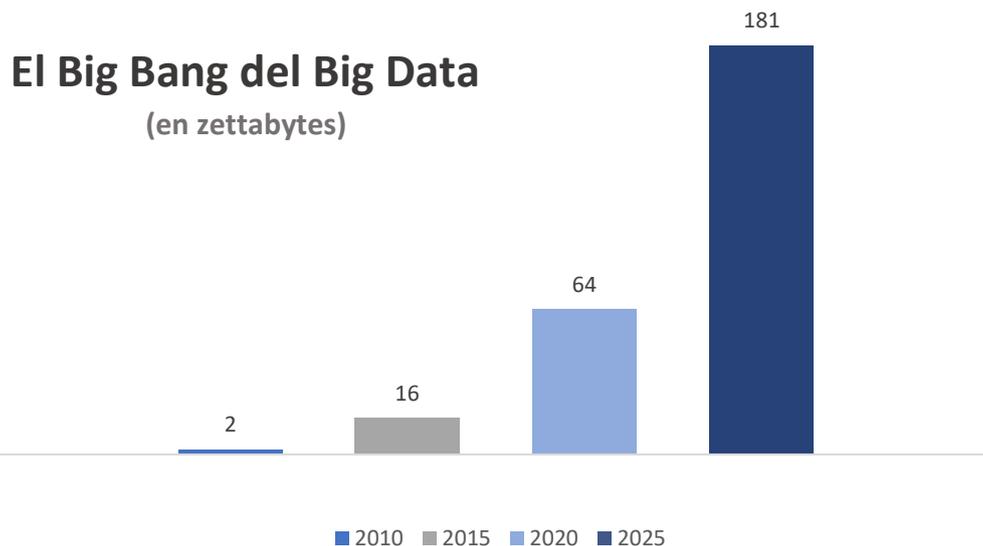


Figura 8. Volumen de datos digitales (marzo 2021) Fuente: elaboración propia a partir de los datos de Roa (2021)

Toda esta realidad se vive a día de hoy y se lleva viviendo años ya, el problema es que se habla de datos y no de información, que son términos parecidos pero muy diferentes. Este hecho se refleja en la posibilidad de acceso a una cantidad de datos pero que después no son aprovechados para convertirlos en un recurso clave a la hora de extraer información relevante.

Hay que tener en cuenta también que, ante esta creciente creación y acumulación de los datos, muchos campos de investigación se han visto en la necesidad de aprender a manejar conjuntos de datos muy grandes cuando antes los datos con los que costaban eran escasos (Maggioni, 2012). Este autor además hace hincapié en como los ordenadores, a través de diferentes algoritmos juegan un papel fundamental a la hora de ayudarnos en numerosas tareas mediante el procesamiento de los datos.

Según Fayyad et al. (1996), la búsqueda de patrones útiles en los datos se conoce con diferentes nombres, muchas veces, en función del ámbito o del área de trabajo de donde provienen los datos que se han de tratar. En este sentido, los autores utilizan el término “KDD”, siglas de *Knowledge Discovery in Databases* (Descubrimiento de Conocimiento en Bases de Datos), para referirse a un proceso para extraer información relevante de los datos, además de afirmar que la minería de datos es solo un paso de este proceso y no se trata de un proceso completo propiamente dicho (Fayyad et al., 1996).

De esta manera se entiende el KDD como un proceso o una metodología que engloba diferentes pasos como la preparación o la limpieza de los datos, interpretación de los resultados, etc. (Fayyad et al., 1996). Por su parte, el minado de datos o *Data Mining*, término que ha sido utilizado sobre todo por estadísticos o analistas de datos, es una actividad dentro de este proceso, mediante la que se aplican algoritmos específicos para extraer patrones de los datos (Fayyad et al., 1996).

Una aproximación muy parecida al *Data Mining* se recoge en Maggioni (2012), quien habla de la búsqueda de patrones como el objetivo principal de esta actividad. Además, apunta también al hecho de que dicho concepto (*Data Mining*), se puede interpretar de maneras diferentes en función del ámbito de trabajo.

Atendiendo a Fayyad y Stolorz (1997), las técnicas de minería de datos pueden dividirse en cinco métodos:

- » **Modelado predictivo.** Fayyad y Stolorz (1997) realizan una distinción clave en función del tipo de dato sobre el que se vaya a realizar la predicción: si se trata de una variable numérica se habla de regresión mientras que, si la predicción se realiza sobre variables categóricas, se trata de un problema de clasificación. Además, apuntan que existen numerosas técnicas tanto de regresión como de clasificación, mediante las cuales es posible resolver gran cantidad de situaciones, aunque el problema es siempre el mismo: encontrar el valor más probable de la variable que se predice (Fayyad & Stolorz, 1997).
- » **Clustering.** Fayyad y Stolorz (1997) se refieren a esta técnica como segmentación, puntualizando que su objetivo es separar los datos en subconjuntos que guarden características comunes. Este análisis se utiliza sobre numerosos campos de estudio, pero con un objetivo común, encontrar los métodos que realicen una segmentación verdaderamente significativa (Hennig, 2015).
- » **Síntesis de datos.** Esta técnica no tiene como objetivo predecir, sino que busca extraer patrones que describan subconjuntos de datos y las relaciones entre ellos (Fayyad & Stolorz, 1997).
- » **Modelado de dependencia.** Esta clase de técnicas buscan obtener estructuras causales entre los datos, ya sean probabilísticos o determinísticos, con el fin de comprender mejor el conjunto de datos (Fayyad & Stolorz, 1997).

- » **Cambio y detección de desviaciones.** En este caso, la característica que distingue a estos métodos tiene que ver con la importancia en el orden de las observaciones (Fayyad & Stolorz, 1997).

5.4.1. Análisis de movimiento

Dentro del *Data Mining*, este proyecto se enfoca en el análisis de datos de movimiento, en concreto, movimiento animal. Echeverry et al. (2018) presentan el término “*Mocap*” (que deriva de *motion capture* o captura del movimiento) como las técnicas para capturar datos espaciotemporales de movimiento y representarlos digitalmente.

Este proceso de captura de datos se debe llevar a cabo a través de un hardware específico para su posterior tratamiento. Este es el caso de los sensores o sistemas de medida inerciales IMU (*Inertial Measurement Units*). Estos dispositivos combinan, y son capaces de captar, los datos de medidores como acelerómetros, giróscopos y magnetómetros (Echeverry et al., 2018).

Estos sistemas de medida o sensores inerciales se pueden encontrar hoy en día en casi cualquier dispositivo electrónico, de ahí el auge que ha tenido el análisis de movimiento, convirtiéndose en un área de interés con numerosas investigaciones y artículos publicados. (Roberts et al., 2017).

Dentro de este movimiento, Roberts et al. (2017) indican que, al fin y al cabo, caminar es la forma más común de moverse, estando presente en multitud de actividades de nuestra vida cotidiana de ahí que muchas de estas investigaciones vayan enfocadas al análisis de la marcha, una actividad que se utiliza para diagnosticar diferentes patologías.

Otra actividad común dentro de este campo de estudio es el análisis biomecánico, sobre todo en deportistas de élite. El objetivo de estudios es, según Pueo y Jimenez-Olmedo (2017), registrar el movimiento del deportista para poder analizar su rendimiento o técnica y de esta manera poder mejorar y prevenir lesiones.

Como se puede observar, la mayor parte de los estudios realizados en este campo se centran en el análisis de movimiento humano, siendo relativamente más difícil encontrar artículos que lleven a cabo este análisis en animales.

Brown et al. (2013) por ejemplo, presentan un estudio sobre animales, pero no se centran en analizar la marcha para detectar posibles patologías, sino que utilizan los datos recogidos por acelerómetros para observar en remoto el comportamiento animal y poder, de esta forma, modelar procesos ecológicos sin pasar por alto sucesos de comportamiento crípticos pero importantes.

5.5. Aprendizaje máquina

El *Machine Learning* o aprendizaje máquina es una rama de la informática, considerada un subgrupo de la inteligencia artificial (Bini, 2018). Tiene su origen en los años 50 con el desarrollo de la inteligencia artificial y busca que una máquina aprenda de forma autónoma, sin ser directamente programada, mejorando su rendimiento en las tareas gracias a la experiencia (Bi et al., 2019).

Hoy en día, después de su rápida extensión en relación con esta era tecnológica, tiene un gran impacto en numerosos campos de la sociedad y numerosas aplicaciones, como muestran Zhou et al. (2017). En relación con el Big Data, las técnicas de *Machine Learning*, aunque no han sido diseñadas específicamente para este fin, permiten hacer predicciones a partir de grandes conjuntos de datos, siendo una parte esencial de la analítica de *Big Data*. (Zhou et al., 2017).

El *Machine Learning* es un campo bastante amplio dentro del que se desarrollan otros subcampos que cuentan con sus propias aplicaciones y enfoques, pero con un objetivo común, que es el buscar que las máquinas aprendan a través de diferentes algoritmos buscando el mejor rendimiento posible a la hora de realizar diferentes tareas.

Uno de los subcampos más extendidos dentro del *Machine Learning* en los últimos años es el *Deep Learning*. Normalmente dentro del aprendizaje máquina, es necesario realizar un tratamiento de los datos para limpiarlo e incluso obtener ciertas relaciones entre los datos de entrada en crudo. En este caso, el Deep Learning combina diferentes módulos que van transformando la información de un nivel hacia el siguiente, lo que permite mantener y amplificar aspectos importantes dentro de (LeCun et al., 2015)

Cabe mencionar también en este punto el *Reinforcement Learning* o aprendizaje por refuerzo, un campo cuyo nombre hace recordar a las técnicas de aprendizaje conductista y que, al fin y al cabo, se basa en esta idea. Es la propia máquina la que explora las diferentes posibilidades

de aprendizaje y recibe una retroalimentación (simulando penalizaciones y recompensas) a través de la cuál va ajustando su comportamiento.

Hay que tener en cuenta también que el *Machine Learning* se puede clasificar en diferentes tipos según como se lleve a cabo el aprendizaje por parte de la máquina, dando lugar al aprendizaje supervisado, no supervisado y semisupervisado.

En este caso se ha utilizado el aprendizaje supervisado que se basa en el entrenamiento del modelo a través de unos datos con la clasificación correcta asignada de antemano (Sathya & Abraham, 2013). En este tipo de aprendizaje, los datos se presentan con etiquetas que clasifican los datos, de manera que el modelo es capaz de aprender a través de las características propias de cada grupo para mejorar la forma en la que clasifica los datos.

En cuanto al aprendizaje no supervisado, es justo lo contrario, los datos se presentan sin etiquetas por lo que es el propio algoritmo el que revisa y aprende de las relaciones subyacentes entre los datos. Por otra parte, el aprendizaje semisupervisado combinan datos etiquetados y sin etiquetar, siendo estos últimos más fáciles de extraer ya que es difícil etiquetar todo un conjunto de datos. Hay que tener en cuenta que hay estructuras que aprenden con éxito utilizando datos etiquetados, pero fallan cuando se entrenan con ambos tipos de datos, como por ejemplo un clasificador Naive-Bayes (Cohen et al., 2004).

Dentro del aprendizaje máquina se pueden utilizar una serie de algoritmos o “modelos” diferentes, que se suelen elegir en función de la tarea que se vaya a realizar (por ejemplo, reconocimiento de imágenes o análisis de movimiento como en este caso).

Algunos modelos populares de aprendizaje automático son las redes neuronales, los árboles de decisión, Naive Bayes y la regresión logística. Estos modelos también pueden combinarse en "conjuntos de modelos", buscando maximizar su precisión. (Burrell, 2016). En la siguiente sección se desarrollará el algoritmo de red neuronal, que ha sido utilizado en el desarrollo de este trabajo por lo que ahora se presentarán brevemente otros modelos ampliamente utilizados en el aprendizaje máquina:

- » **Arboles de decisión.** Se trata de un algoritmo de aprendizaje supervisado que parte de un nodo raíz que se va desagregando de manera jerárquica en nodos internos hasta llegar a los nodos hoja que representan los resultados. Se realiza una búsqueda entre todas las posibilidades para encontrar la rama que presenta mejores resultados (Kotsiantis, 2013).

» **SVM.** Las máquinas de vectores soporte (SVM por sus siglas en inglés), son métodos de aprendizaje supervisado que se pueden utilizar en problemas de regresión o clasificación, sobre todo binaria. Este algoritmo separa las observaciones mediante un límite óptimo llamado hiperplano, aunque algunas veces es necesario transformar estas observaciones para poder realizar la separación.

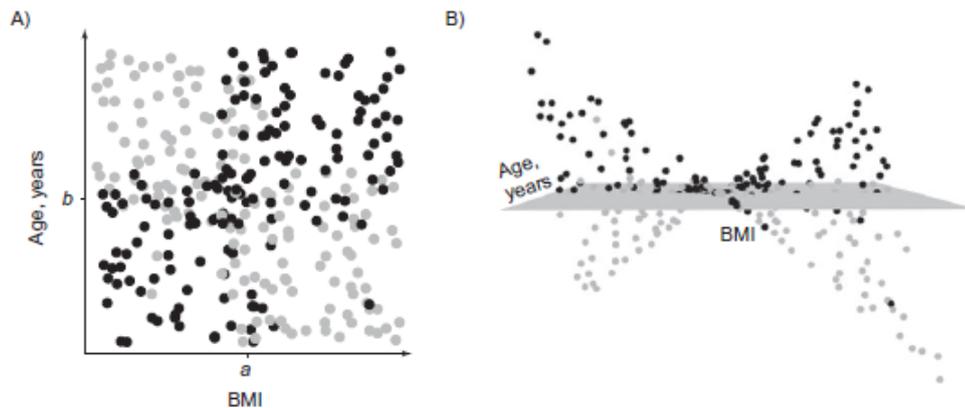


Figura 9. Transformación de datos para predecir diabetes con SVM. Fuente: Bi et al. (2019, p. 6)

» **Naive Bayes.** Este algoritmo se basa en el teorema de Bayes, un método que permite recalcular la probabilidad de hipótesis con la entrada de nuevos datos (Berrar, 2016). Este algoritmo calcula la probabilidad asociada a una clase como el producto de la probabilidad previa y la función de verosimilitud. De esta forma, el clasificador selecciona como clase correcta la de probabilidad más alta. Se suele utilizar sobre todo en clasificación. (Bi et al., 2019).

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Posterior Probability of Diabetes Status	=	Prior Probability	×	Likelihood			
Pr(D+ BMI > 32 and Age > 55 and Sex = F)	=	Pr(D+)	×	Product of <table style="display: inline-table; vertical-align: middle; border-left: 1px solid black; border-right: 1px solid black; border-collapse: collapse;"> <tr><td style="padding: 0 5px;">Pr(BMI > 32 D+)</td></tr> <tr><td style="padding: 0 5px;">Pr(Age > 55 D+)</td></tr> <tr><td style="padding: 0 5px;">Pr(Sex = F D+)</td></tr> </table>	Pr(BMI > 32 D+)	Pr(Age > 55 D+)	Pr(Sex = F D+)
Pr(BMI > 32 D+)							
Pr(Age > 55 D+)							
Pr(Sex = F D+)							
Pr(D- BMI > 32 and Age > 55 and Sex = F)	=	Pr(D-)	×	Product of <table style="display: inline-table; vertical-align: middle; border-left: 1px solid black; border-right: 1px solid black; border-collapse: collapse;"> <tr><td style="padding: 0 5px;">Pr(BMI > 32 D-)</td></tr> <tr><td style="padding: 0 5px;">Pr(Age > 55 D-)</td></tr> <tr><td style="padding: 0 5px;">Pr(Sex = F D-)</td></tr> </table>	Pr(BMI > 32 D-)	Pr(Age > 55 D-)	Pr(Sex = F D-)
Pr(BMI > 32 D-)							
Pr(Age > 55 D-)							
Pr(Sex = F D-)							

Figura 10. Aproximación del clasificador Naive-Bayes para la predicción de diabetes tipo 2. Fuente: Bi et al. (2019, p. 7)

5.5.1. Red Neuronal

Como en otras muchas ocasiones, la ciencia ha tratado de replicar la naturaleza, es por ello por lo que, en 1943, McCulloch y Pitts modelaron una especie de neurona artificial que recibiendo una entrada de otras neuronas se activaba o no (Krogh, 2008). Más tarde, durante los años 60, se demostró que conjuntos de estas neuronas (redes neuronales artificiales o RNA), se comportaban de forma similar al cerebro, reconociendo patrones y funcionando, aunque se destruyeran algunas. (Krogh, 2008).

El funcionamiento de las RNA se basa un proceso de aprendizaje, llamado fase de entrenamiento, que busca calibrar los pesos entre las conexiones de las neuronas para adaptarlas al problema que se debe resolver. (Tostado Sánchez et al., 2016). Las conexiones iniciales se generan al azar y cada salida se compara con la salida correcta establecida en las etiquetas de clase (aprendizaje supervisado). Además, se utiliza una función de pérdida que muestra la diferencia entre valores predichos y valores reales. El objetivo de la red al fin y al cabo es minimizar esta función de pérdida. (Bi et al., 2019)

Las RNA realizan tareas de clasificación de forma obvia y eficiente gracias a su diseño estructural y a sus métodos de aprendizaje. No existe un algoritmo único para diseñar y entrenar modelos de RNA, ya que los algoritmos de aprendizaje difieren entre sí por su capacidad de aprendizaje y su grado de inferencia. (Sathya & Abraham, 2013)

Todos los modelos son únicos por naturaleza y cada uno ofrece ventajas propias además de tener diversas aplicaciones, siendo la clasificación de patrones de las más estudiadas. (Sathya & Abraham, 2013) Una de las clasificaciones más comunes dentro de las redes neuronales tiene que ver con su propiedad de interconexión. En este sentido, nos podemos encontrar con:

- » **Redes *feedforward*:** dentro de la estructura de una red *feedforward*, la señal se extiende de forma unidireccional desde las neuronas de la capa de entrada hasta la capa de salida. La conexión entre estas capas se realiza a través de las capas ocultas (cuantas más capas ocultas, más compleja será la red) (Bi et al., 2019) y en ningún caso se dan conexiones entre neuronas de capas anteriores o de la misma capa. Además, si cada una de las neuronas de una capa, están conectadas con todas las neuronas de la capa siguiente, esta red se conoce como perceptrón multicapa o red totalmente conectada (Zebin et al., 2017).

- » **Redes recurrentes:** según Schuster y Paliwal (1997), las redes feedforward están limitadas ya que solo pueden trabajar con datos estáticos y no pueden enfrentarse a problemas representados por secuencias temporales de datos (reconocimiento del habla, predicción de series temporales, etc.), mientras que las redes recurrentes permiten incorporar esta dinámica temporal. En este caso cuentan con conexiones retroalimentadas que les permiten usar la información que ha recibido el sistema hasta ese instante (Schuster & Paliwal, 1997).

6. Metodología

En el desarrollo de este proyecto se ha utilizado la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), una de las más utilizadas en lo que a proyectos de minería de datos se refiere. Según Martínez-Plumed et al. (2021) esta metodología se origina en la segunda mitad de la década de los noventa, cuando las técnicas de *Data Mining* se volvieron más comunes en la industria y las compañías necesitaban crear estándares para identificar buenas prácticas, así como errores cometidos en el pasado.

En cuanto a los inicios de esta metodología, en Chapman et al. (2000) se describe de qué forma fue concebida, y es que CRISP-DM surgió a finales de 1996, desarrollada por tres veteranos ingenieros expertos en minería de datos. (Chapman et al., 2000). Por su parte, Martínez-Plumed et al. (2021) añade más información al respecto mencionando que la Unión Europea financió el desarrollo de esta metodología de aplicación general que llegaría a ser ampliamente aceptada.

La metodología CRISP-DM se puede utilizar para resolver cualquier problema independientemente de la forma en la que se proveen los datos (Marbán et al., 2009). Además, se ha constituido como una de las más usadas en el desarrollo de proyectos de *Data Mining*.

CRISP-DM define las fases que debe contener un proyecto de *Data Mining* además de definir para cada fase, sus tareas y los entregables de cada tarea. (Marbán et al., 2009). Esta metodología se presenta de forma jerárquica, dividiéndose en cuatro niveles que, de más general a más específico son: fase, tarea genérica, tarea especializada e instancia de proceso (Chapman et al., 2000).

Como se comentaba, la metodología se divide en fases, compuestas a su vez por tareas generales (lo suficiente como para cubrir todas las situaciones posibles del minado de datos). En el tercer nivel se tratan estas tareas en situaciones específicas y el cuarto nivel es un registro de las acciones, decisiones y resultados del proyecto. (Chapman et al., 2000).

CRISP-DM se compone de seis fases iterativas desde la comprensión del negocio hasta su implantación (Schröer et al., 2021) aunque la secuencia de las fases no es rígida. Depende del resultado de cada fase, qué fase o qué tarea concreta de una fase debe realizarse a continuación. (Chapman et al., 2000). Las fases que componen esta metodología están recogidas en la Figura

12 que muestra su modelo de referencia como una visión general del ciclo de vida de un proyecto de minería de datos.

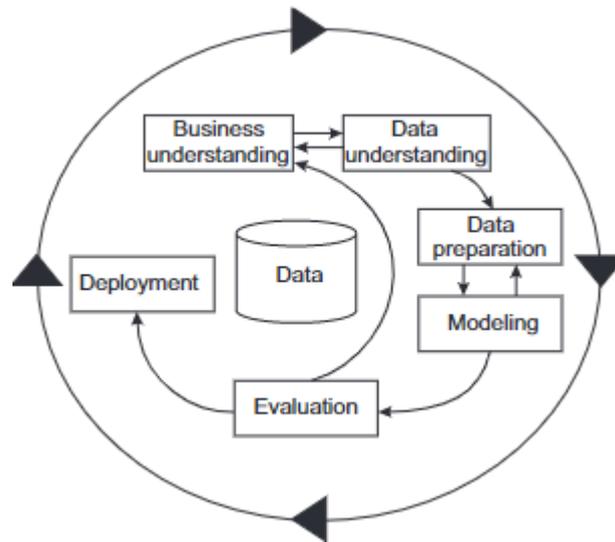


Figura 11. Metodología CRISP-DM. Fuente: Chapman et al. (2000) en Marbán et al. (2009).

Por su parte, la Figura 13 muestra los niveles en los que se descomponen jerárquicamente las diferentes tareas, quedando ambas figuras aunadas en la Figura 14 donde se puede observar un ejemplo de cómo las fases se pueden ir descomponiendo, aunque en este caso solo se muestre una tarea en cada nivel (excepto en el cuarto nivel de la tercera fase).

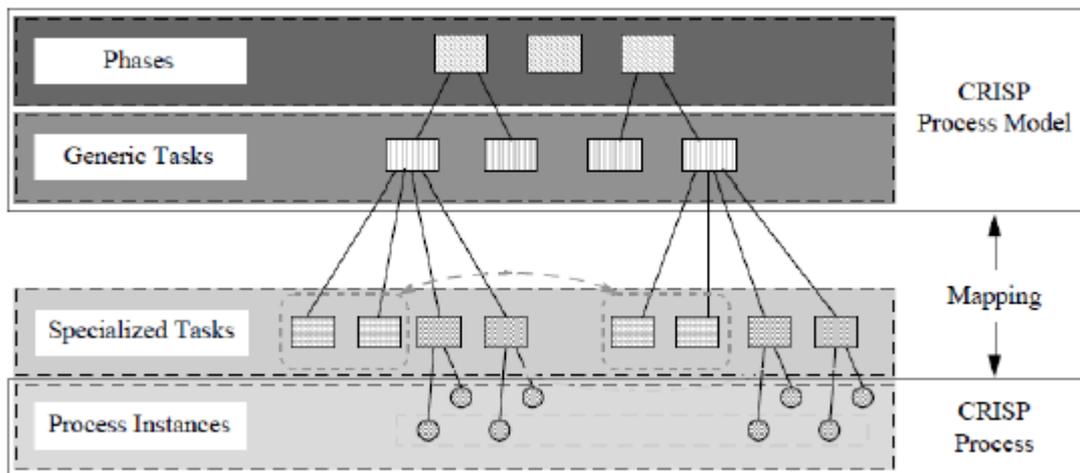


Figura 12. Niveles jerárquicos del CRISP-DM. Fuente: Chapman et al. (2000).

Business understanding	Data understanding	Data preparation	Modeling	Evaluation	Deployment
Determine business objectives	Collect initial data	Select data	Select modeling techniques	Evaluate results	Plan deployment
Assess situation	Describe data	Clean data	Generate test design	Review process	Plan monitoring and maintenance
Determine DM objectives	Explore data	Construct data	Build model	Determine next steps	Produce final report
Produce project plan	Verify data quality	Integrate data Format data	Assess model		Review project

Figura 13. Descomposición de las fases del CRSIP-DM en los diferentes niveles jerárquicos. Fuente: Marbán et al.(2009).

6.1. Comprensión del negocio

El conocimiento o la comprensión del negocio no se centra en conocer de manera inequívoca todos los datos referentes a una actividad productiva sino, más bien, tener un conocimiento amplio de esta y de los procesos que la componen para poder ajustar los objetivos y requisitos del proyecto que se va a realizar definiendo que clase de tratamiento se va a realizar (clasificación, por ejemplo) y el criterio de éxito que se va a seguir (como la precisión) (Schröer et al., 2021).

Por su parte, Marbán et al. (2009) añaden que este conocimiento se debe ajustar este conocimiento a un proyecto de *Data Mining* diseñando de esta forma un plan preliminar para alcanzar los objetivos.

6.2. Comprensión de los datos

La adquisición de los datos se puede llevar a cabo de múltiples maneras, pero a través de estas se obtienen siempre unos datos en bruto que confieren una primera visión acerca de las características del proyecto. Estos datos deberán ser estudiados a través de diferentes actividades a lo largo de esta fase para tener un conocimiento más profundo del proyecto que se va a desarrollar, explorando y comprobando la calidad de los datos. (Schröer et al., 2021).

6.3. Preparación de los datos

En esta fase se recogen todas las actividades que se llevan a cabo sobre los datos obtenidos en primera instancia para “limpiarlos” y dejarlos listos para su procesado. Martínez-Plumed et al.

(2021) hablan también del hecho de que los datos preparados pueden ser el producto final del proyecto, sin necesidad de realizar actividades más allá de esta fase.

6.4. Modelado

Para un mismo problema de *Data Mining* se pueden crear infinidad de modelos, utilizando diferentes técnicas y fijando diferentes parámetros. (Shafique & Qaiser, 2014). Aunque se presenten diferentes posibilidades el fin de esta etapa, a través de la selección de técnicas y la elaboración del modelo o de los modelos, será buscar los mejores resultados posibles ajustándose al problema tratado y los datos con los que se trabaja, que han sido estudiados previamente.

6.5. Evaluación

Esta es una fase clave del proceso porque se evalúan los modelos y los resultados obtenidos para comprobar si cumplen con los objetivos y se puede dar paso a la última fase del proceso. La evaluación no pasa únicamente por decidir si los resultados obtenidos son buenos o malos, sino que se refiere también a la toma de decisiones sobre la manera de utilizar estos resultados. (Shafique & Qaiser, 2014).

Si la evaluación resultara negativa será necesario volver a revisar fases previas con el fin de encontrar y corregir posibles errores o, incluso, plantear de nuevo la forma de abordar la problemática a tratar.

6.6. Despliegue

Numerosos proyectos de *Data Mining* se centran estudiar diferentes fenómenos y obtener unos resultados al respecto, derivando su fase de despliegue en la elaboración de un informe que recoja el proceso y los resultados conseguidos, como es el caso de numerosos trabajos de naturaleza similar a este.

Sin embargo, esta fase también está destinada a la aplicación práctica del proceso mediante tecnologías que utilicen los modelos para dar respuesta a necesidades que se puedan presentar en diferentes ámbitos, en este caso, el de la ganadería.

7. Recursos

7.1. Datos

Los datos utilizados en el trabajo fueron adquiridos por el grupo de investigación del IDETIC - Procesado Digital de la Señal (DPDS) del que forman parte tanto los tutores, D. Miguel Ángel Ferrer Ballester y D. Moisés Díaz Cabrera, como otros investigadores. Estos datos han sido cedidos para la realización del presente trabajo y no cuentan con autorización para su distribución.

Los datos se han extraído de 21 vacas que se encuentran dentro de una granja ubicada en el sur de la isla de Gran Canaria. La extracción se ha llevado a cabo a través del sensor inercial que incorpora un *Apple Watch Series 6* colocado en las patas traseras de estas vacas.

La grabación de los datos se realizaba en intervalos de 90 segundos, con una frecuencia de muestreo de 100 hercios (Hz) y, por lo tanto, un período de 0,01 segundos (cabe recordar que la frecuencia y el periodo son magnitudes inversas), por lo que cada fichero contiene 9.000 muestras. Los ficheros además están nombrados según se muestra en la Figura 15 y están compuestos por cuatro campos diferenciados:

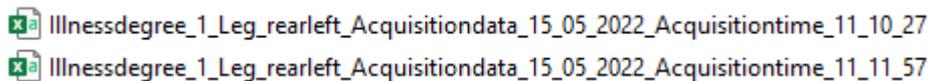


Figura 14. Ficheros que guardan los datos capturados por el sensor inercial. Fuente: elaboración propia.

- Illnessdegree_1:** hace referencia al nivel “nivel de enfermedad” en el que se encuentra un animal. Estos niveles están numerados del 1 al 5 teniendo en cuenta un análisis veterinario previo. De esta forma se podrá entender que las vacas con un valor de “*Illness Degree*” comprendido entre 3 y 5 están enfermas, mientras que se clasificarán como vacas sanas las que cuenten con un valor igual a 1 o 2.
- Leg_rearleft:** este campo hace referencia a la pata del animal donde se coloca el sensor. En este caso se trata de la pata trasera izquierda.

3. **Acquisitiondata_15_05_2022:** como su propio nombre indica en este campo se recoge la fecha en la que se llevó a cabo la captura de datos. Los de este fichero en concreto fueron recogidos hace ya más de un año, el 15 de mayo de 2022.
4. **Acquisitiontime_11_10_27:** en este caso no se recoge la fecha sino la hora exacta en la que se inició la captura de datos. Podemos observar además en la Figura 15 como ambos ficheros están desplazados en el tiempo 90 segundos como ya habíamos comentado. Aunque la toma de datos es muy precisa, tiene un pequeño desfase que se puede observar de manera clara entre algunos ficheros puntuales como, por ejemplo, los que se recogen la Figura 16.

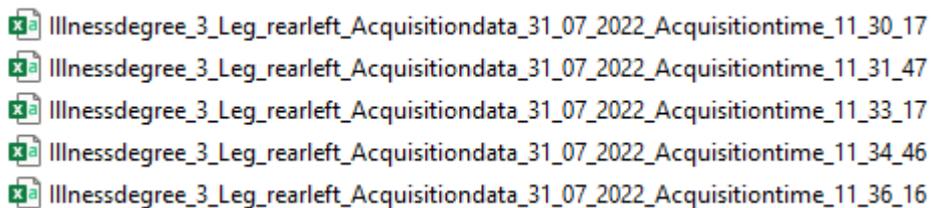


Figura 15. Ejemplo del desfase que se presenta en una serie de ficheros de datos. Fuente: elaboración propia.

Todos los ficheros correspondientes a una vaca se encuentran dentro de una carpeta que tiene por nombre un identificador único para cada animal como se muestra en la Figura 17, incluyendo al final una “h” si la vaca se encuentra sana (*healthy*) o una “l” si la vaca presenta cojera (*lame*), en función de un estudio veterinario, que como ya se ha comentado, se realizó antes de la captura de datos.

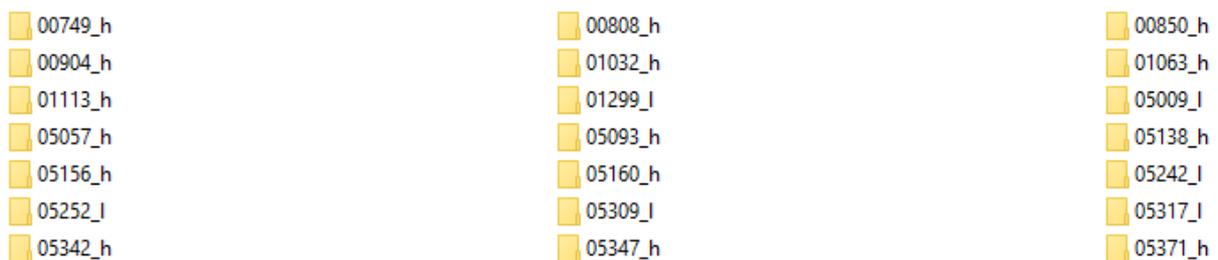


Figura 16. Estructura de carpetas que contienen los datos de las vacas según un ID único. Fuente: elaboración propia.

7.2. Software

El primer punto dentro de los recursos software es el lenguaje de programación utilizado que en este caso ha sido el lenguaje Python. Este lenguaje de programación de alto nivel fue creado por Guido van Rossum a finales de la década de 1980, con el objetivo de desarrollar un lenguaje de programación fácil de leer y entender. Desde entonces, Python ha experimentado un crecimiento exponencial y se ha convertido en uno de los lenguajes más populares en la actualidad como demuestra el índice TIOBE.

El índice TIOBE mide de la popularidad de los diferentes lenguajes de programación existentes. Este índice se mensualmente y se basa en diferentes métricas, como las que extraen de diferentes motores de búsqueda (TIOBE Index - TIOBE, 2022)

May 2023	May 2022	Change	Programming Language	Ratings	Change
1	1		 Python	13.45%	+0.71%
2	2		 C	13.35%	+1.76%
3	3		 Java	12.22%	+1.22%
4	4		 C++	11.96%	+3.13%

Figura 17. Índice TIOBE en el mes de mayo de 2023. Fuente: TIOBE.

Programming Language	2023	2018	2013	2008	2003	1998	1993	1988
Python	1	4	8	7	13	25	19	-
C	2	2	1	2	2	1	1	1
Java	3	1	2	1	1	17	-	-
C++	4	3	4	4	3	2	2	6

Figura 18. Evolución del índice TIOBE a través de los años . Fuente: TIOBE.

Hoy en día el lenguaje Python se utiliza casi a cualquier nivel ya sea educativo, empresarial o particular, tanto por usuarios que se inician en el mundo de la programación como por otros que son expertos en la materia y que desarrollan grandes códigos. García Monsálvez (2017) nos da las claves en el desarrollo de este lenguaje y es que Python es fácil de aprender, cuenta con numerosas librerías y módulos de terceros, es de software libre y tiene una gran versatilidad para adaptarse a diferentes ámbitos y problemáticas.

El siguiente punto a definir dentro de los recursos software será el entorno que se ha utilizado para desarrollar el código Python correspondiente. En este caso se trata del entorno Jupyterlab, un entorno extensible para la computación interactiva y reproducible. (Jupyterlab, s.f.)

Jupyterlab forma parte del denominado Proyecto Jupyter (organización que busca desarrollar software y otras herramientas de código abierto) y es una versión renovada del clásico Jupyter Notebook, ofreciendo las mismas herramientas (terminal, editor de texto, navegador de archivos, etc.) a través de una interfaz flexible. (Jupyterlab, s.f.)

Por último, se debe tener en cuenta los módulos y las librerías utilizadas en el desarrollo del código para poder llevar a cabo las operaciones necesarias para la consecución de los objetivos propuestos y el correcto funcionamiento del software.

Mientras que un módulo de Python nos proporciona diferentes funciones y clases, las librerías son agrupaciones de estos módulos que se utilizan para tareas relacionadas. Los diferentes elementos utilizados se detallan a continuación:

- » **Matplotlib.** Matplotlib se utiliza para crear diferentes visualizaciones de datos (en este caso gráficas) en Python. (Matplotlib documentation — Matplotlib 3.7.1 documentation, s. f.)
- » **NumPy.** NumPy es uno de los paquetes más utilizados en Python ya que proporciona una base para el cálculo y la computación, aportando diferentes objetos y rutinas para realizar operaciones aritméticas, de estadística básica o de álgebra lineal entre otras (NumPy documentation — NumPy v1.24 Manual, s. f.).
- » **Os.** Este módulo provee una manera versátil para interactuar con el sistema operativo y utilizar diferentes funcionalidades dependientes del mismo (crear carpetas, abrir archivos, etc.) (os — Interfaces misceláneas del sistema operativo — documentación de Python - 3.10.11, s. f.).
- » **Pandas.** Pandas es una biblioteca de código abierto que ofrece diferentes estructuras que resultan muy útiles para la manipulación de los datos por lo que es muy efectiva a la hora de resolver problemas que impliquen su tratamiento (pandas documentation — pandas 2.0.2 documentation, s. f.).
- » **Re.** Este módulo se utiliza para realizar diferentes operaciones con expresiones regulares (re — Operaciones con expresiones regulares, s. f.) .
- » **Scikit-learn.** Basada en NumPy, SciPy y Matplotlib, proporciona herramientas sencillas y eficaces para el análisis predictivo de datos.(scikit-learn: Machine Learning in Python — scikit-learn 1.2.2 documentation, s. f.)

- » **SciPy.** Se basa en la librería NumPy, ampliándola a través de nuevos elementos que se pueden aplicar en numerosos ámbitos de la computación (SciPy, s. f.).
- » **Seaborn.** Se basa en la librería matplotlib para proporcionar gráficos novedosos y modernos (seaborn: statistical data visualization — seaborn 0.12.2 documentation, s. f.).
- » **Statmodels.** Se trata de un módulo de Python que proporciona múltiples funciones y clases para realizar pruebas estadísticas y comparar los resultados. (Statsmodels 0.14.0, s.f.)
- » **Tensorflow.** Es una librería de código libre que ofrece diferentes niveles de abstracción, lo que permite que sea fácil aprender a utilizarla, pero también potente y se pueda utilizar en producción. Además, añade diferentes módulos para la evaluación de los modelos o el desarrollo de algoritmos de aprendizaje profundo (TensorFlow, s.f.)

7.3. Hardware

Las diferentes tareas que han sido necesarias dentro del desarrollo de este trabajo se han desarrollado en dos equipos diferentes. Por un lado, un ordenador de sobremesa personal, cuyas especificaciones se recogen en la Tabla 6, para tareas como la redacción de los diferentes documentos a incluir en el trabajo o la búsqueda de fuentes de información;

Por otro lado, un equipo del laboratorio de Procesado Digital de la Señal, ubicado en el pabellón B de la Escuela de Ingeniería de Telecomunicación y Electrónica de la ULPGC. Al contar con mayores prestaciones, era más adecuado a la hora de realizar las tareas referentes a la ejecución del código software. Las especificaciones de este equipo se recogen en la Tabla 7.

Tabla 6. Especificaciones ordenador de sobremesa personal. Fuente: elaboración propia

Marca	Hewlett-Packard (HP)
Sistema Operativo	Windows 10 Home, 22H2
Procesador	Intel(R) Core(TM) i5-7200U
Memoria RAM	8,00 GB
Almacenamiento Interno	480 GB (SSD)
Gráfica	Integrada

Tabla 7. Equipo Facultad de Tecnologías de la Comunicación. Fuente: elaboración propia

TORRE	Marca	Acer
	Sistema Operativo	Windows 10 Enterprise, 1903
	Procesador	Intel(R) Core(TM) i7-4770K
	Memoria RAM	8,00 GB
	Almacenamiento Interno	1TB (repartido en varias particiones)
	Gráfica	Integrada
PANTALLA	Marca	Acer
	Tamaño	22 pulgadas

Cabe mencionar en este apartado que la captura de los datos que se han utilizado para la realización de este trabajo se ha realizado mediante el sensor inercial integrado en un *Apple Watch Series 6*, como ya se ha comentado anteriormente, colocado en las patas traseras de las vacas.

Fue necesario desarrollar una aplicación que tomase las grabaciones de este sensor, y las enviase a la nube para su posterior descarga. Estas actividades no han formado parte del presente trabajo, sino que se han llevado a cabo por los equipos de los diferentes proyectos de investigación anteriormente nombrados.

8. Desarrollo de la metodología utilizada

8.1. Comprensión del negocio

La comprensión del negocio no se basa, únicamente, en las variables socioeconómicas que afectan tanto a las explotaciones ganaderas como al sector tecnológico, en este caso, sino que también entran en juego factores como la cultura, la tradición, el arraigo a la tierra y a toda una vida en el medio rural.

En este sentido, se debe poner en valor el hecho de que el autor se ha criado en este entorno, en contacto directo con actividades agrícolas y ganaderas, conociendo de primera mano las necesidades de algunas de ellas, así como haber desarrollado los estudios en una disciplina tecnológica en pleno auge como es la Informática, lo que ayuda a tener una visión más completa en lo que a la comprensión del negocio se refiere.

Además, es importante atender a todos los factores que se han mencionado en el sección número cinco del presente trabajo. En relación a la actividad ganadera, estos vienen recogidos anualmente en el “*Informe Anual de Indicadores: Agricultura, Pesca y Alimentación*” (Ministerio de Agricultura, Pesca y Alimentación, 2022) y reflejan una imagen fiel de la ganadería nacional, además de añadir comparaciones con gran valor como las interanuales o las referentes a otros países.

Es necesario de igual manera, entender los elementos que subyacen dentro del tratamiento y clasificación de los datos. Por especificar un poco más, y haciendo referencia a lo mencionado en puntos anteriores, podemos dirimir que nuestro problema es un problema de clasificación binaria ya que se dividirá a los animales según dos variables categóricas: sanas y enfermas. Además, se trata de un algoritmo de red neuronal que utiliza aprendizaje supervisado.

Por último, es de vital importancia hacer hincapié en la necesidad de acometer proyectos de este estilo, por dos razones principalmente. Es necesario trabajar y apoyar a esas actividades que, aunque parecen remotas y han ido perdiendo protagonismo, son esenciales para nuestro día a día y nuestra supervivencia, todo esto, sin dejar de mirar al futuro y adaptarnos a los retos e innovaciones que van apareciendo en un mundo que no deja de expandirse.

8.2. Comprensión de los datos

En este apartado se presentarán y explicarán los diferentes campos que componen los datos sobre los que se ha trabajado. Además, se desarrollará un punto explicando el análisis que se ha llevado a cabo sobre estos datos para intentar predecir si una vaca se encuentra en movimiento o no en un instante de tiempo determinado.

8.2.1. Contenido de los datos

Cada fichero contiene 9.000 filas de datos y una cabecera donde se recogen los nombres de las doce columnas de datos que es capaz de capturar el sensor. Dichas columnas se muestran en la Figura 20 y se describen a continuación:

#time (s) lax lay laz gx gy gz mx my mz yaw roll pitch

Figura 19. Columnas presentes en los ficheros de datos. Fuente: elaboración propia.

- » **# time (s):** la primera columna de datos a los intervalos de tiempo en los que se ha capturado cada entrada de datos, Como ya se comentó anteriormente, cada 0,01 segundos, es decir, cada centésima de segundo.
- » **lax, lay, laz:** las siguientes tres columnas recogen los valores de la aceleración en cada uno de los ejes: X, Y y Z. Vienen expresados en aceleración gravitacional antes de su transformación a unidades del Sistema Internacional (m/s^2).
- » **gx, gy, gz:** estas tres columnas hacen referencia a los valores del giróscopo en cada uno de los ejes. Estos valores se miden en
- » **mx, my, mz:** el valor del magnetómetro en cada uno de los ejes está reflejado en estas columnas. Estos valores se miden en
- » **yaw, roll, pitch:** las tres últimas columnas contienen los valores de *yaw*, *roll* y *pitch*. Estas variables, comúnmente utilizadas en aeronáutica, miden los movimientos o rotaciones en los tres componentes eulerianos (Wilson et al., 2018). En el caso del dispositivo utilizado en la captura de datos del presente trabajo, estas tres variables se pueden identificar de forma visual en la Figura 21

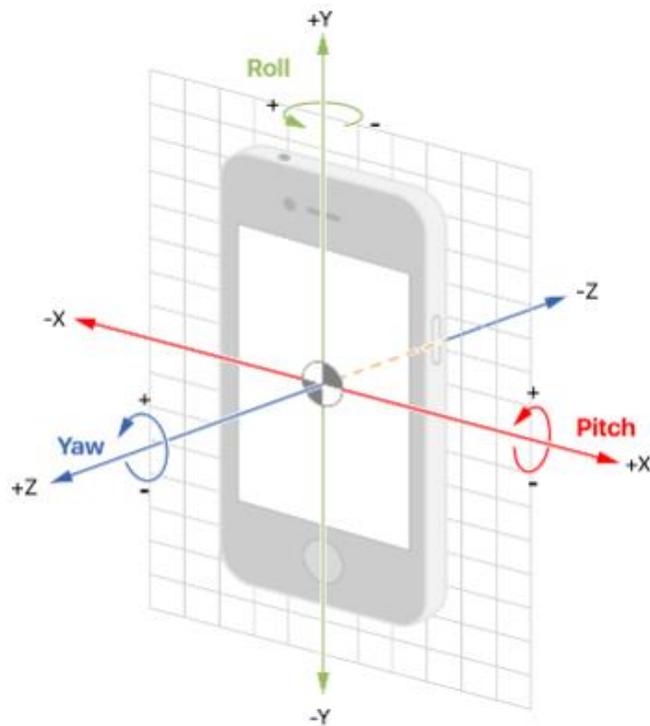


Figura 20. Yaw, roll, pitch. Fuente: (Understanding Reference Frames and Device Attitude | Apple Developer Documentation (s. f.).

8.2.2. Análisis de los datos

Además, es necesario mencionar que los datos se han adquirido durante el tiempo que la vaca pasaba entre la sala de ordeño (donde se le colocaba el sensor) y el corral. Se añade el tiempo que las vacas pasaban en dicha sala, por lo que se incluyen datos donde el animal no se mueve. Para poder aislar los momentos donde la vaca estaba en movimiento se ha realizado un análisis de los datos con los que se ha trabajado, aportando un conocimiento más profundo sobre los mismos.

El primer paso en este análisis ha sido crear un array que contiene el módulo de la energía de cada una de las muestras por cada vaca. Es decir, sabiendo que un fichero contiene 9000 muestras, si tenemos una vaca de a que se han capturado 100 ficheros de datos, hablaríamos de un array de 900000 elementos, cada uno de ellos referente al módulo de la energía de una vaca en un instante, que se obtiene a partir de la siguiente fórmula, donde a_x , a_y y a_z , hacen referencia a la aceleración en cada eje ortogonal.

$$E[i] = \sqrt{a_x[i]^2 + a_y[i]^2 + a_z[i]^2}$$

Sin embargo, estos valores de la energía en instantes tan pequeños de tiempo no son indicadores válidos para conocer si realmente una vaca está en movimiento o no, por lo que el siguiente paso para determinarlo ha sido dividir este array en diferentes ventanas de tiempo para, mediante el sumatorio de los valores de energía contenidos dentro de este subconjunto, obtener valores de energía que nos ayuden a saber si una vaca se mueve o no. Para esto se han llevado a cabo varias acciones.

En primer lugar, se dividen los datos del array en ventanas de tiempo diferentes añadiendo a cada una un solapamiento para evitar sesgos entre los datos. El nombre de las variables utilizadas hace referencia a los subconjuntos de energía y cada una marca el tiempo de la ventana (primer número) y el solapamiento correspondiente (segundo número).

La primera de las variables, “*step*” hace referencia al tamaño del paso que se utilizará para poder dividir el array en subconjuntos. Se establece con el valor 6000. Teniendo en cuenta que los datos se obtienen en centésimas de segundo, este valor es igual a 60 segundos, es decir, un minuto. Después de esto se define una variable por cada tamaño de ventana, multiplicando el valor de “*step*” (1 minuto) por el valor correspondiente para cada ventana. En este caso tenemos ventanas de 2, 4, 9, 10, 12 y 15 minutos.

A continuación, se define cada una de las ventanas que se van a probar como una lista, dividiendo el array según el tamaño de ventana y utilizando como paso el tamaño de la ventana menos el solapamiento. Es decir, si una ventana de 10 minutos tiene un solapamiento de 4 minutos, el paso será de 6 minutos, esto es, el primer subconjunto se tomará desde el valor 0 al valor 60000 y el segundo desde el valor 36000 hasta el valor 96000, con lo que quedarán solapados 24000 valores, lo que equivale a 4 minutos.

Una vez que se han creado todas las ventanas, se añaden a la lista a una lista y se van pasando como parámetros a una función ya definida llamada “procesar”. Se pasan también los valores de diferentes variables: primero una variable auxiliar que se utiliza para determinar sobre que ventana se está trabajando y un valor límite que se utilizará para discriminar si una vaca realmente se está moviendo o no.

La función devuelve una lista a la que se iguala cada ventana y que luego se reinicia, además de actualizar la variable auxiliar utilizada, que se inicia en 0. Este código se presenta en la siguiente figura.

```

for value in prevals:
    procesar(value,x,(vnt[x]*1.5*1000))
    value = aux
    aux = []
    values.append(value)
    x+=1

```

Figura 21. Llamada a la función "procesar" sobre cada ventana. Fuente: elaboración propia.

Al tratarse cada ventana como una lista, estará compuesta por varios elementos que englobarán los diferentes valores de dicha ventana. De esta manera, la función "procesar" se encarga de recorrer valor por valor cada elemento de cada ventana y realizar un sumatorio de cada elemento para añadirlo a la lista "gráfica".

Además, comparará cada valor de esta lista con el límite establecido en cada ventana para saber si una vaca se mueve o no. En caso de que se mueva (el valor de la lista gráfica es mayor que el límite) se añadirá un 1 a la lista que devuelve la función, añadiéndose un 0 en caso contrario. De esta forma, en el código de la figura anterior, cada ventana queda igualada a una lista que representa que elementos se encuentran por encima del límite establecido y nos permitirá saber en qué momento un animal está en movimiento.

Por último, pasa como parámetros a la función graficar el tamaño de la ventana, el valor del solapamiento de esa ventana, la lista "gráfica" que contiene en cada posición los valores del sumatorio de cada subconjunto y el valor límite para saber si una vaca se mueve en un instante o no. A continuación, se muestran algunas gráficas obtenidas sobre las diferentes ventanas:

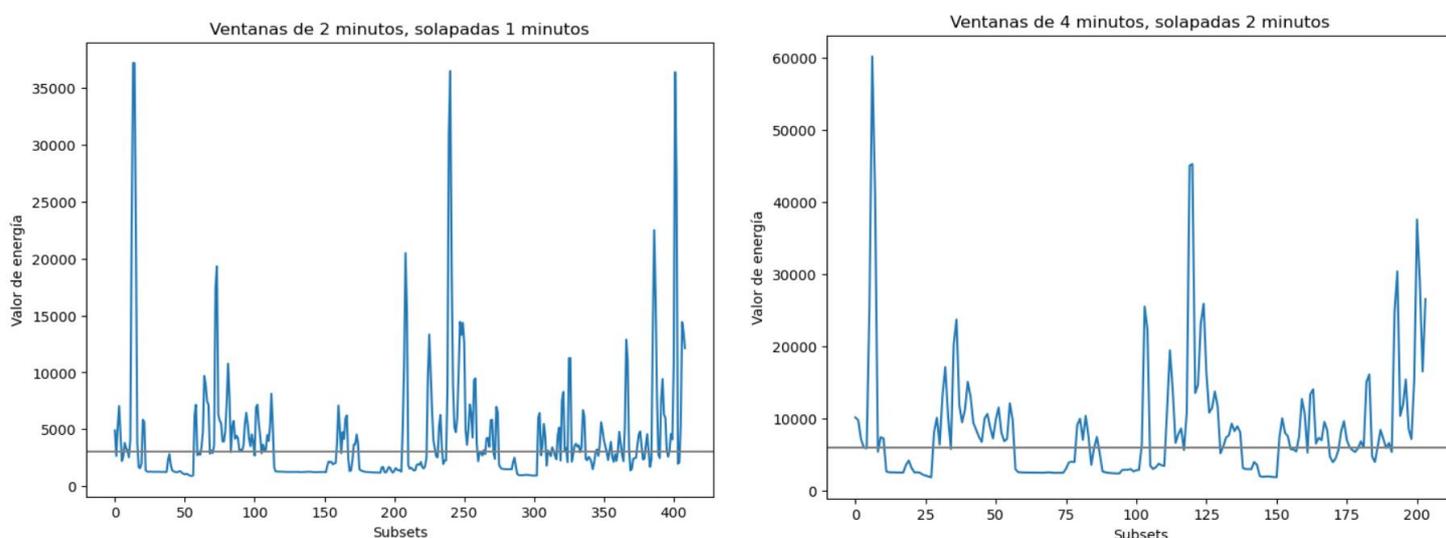


Figura 22. Gráficas de los sumatorios de energía de una vaca, con ventanas de 2 y 4 minutos, solapadas 1 y 2 minutos respectivamente. Fuente: elaboración propia

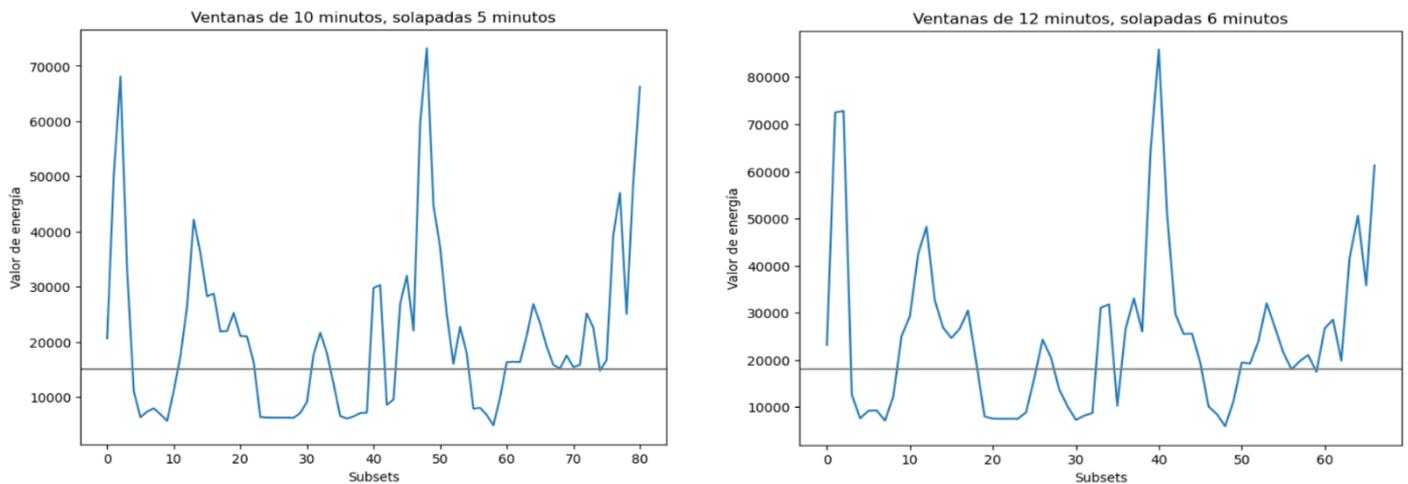


Figura 23. Gráficas de los sumatorios de energía de una vaca, con ventanas de 10 y 12 minutos, solapadas 5 y 6 minutos respectivamente. Fuente: elaboración propia.

Como se puede observar, las dos primeras gráficas, al tratarse de ventanas de tiempo muy pequeñas, muestran numerosos picos de energía, lo que dificulta la tarea de fijar un umbral que sea realmente discriminante a la hora de predecir si una vaca se mueve o no. En cambio, con las ventanas de 10 y 12 minutos, por ejemplo, estas gráficas se suavizan, dejando a la vista zonas donde claramente la vaca está en movimiento y zonas donde no lo está. Por esta razón, se han descartado las ventanas de 2 y 4 minutos y se han utilizado ventanas a partir de 9 minutos, con el fin de mejorar el veredicto sobre el movimiento o no del animal.

8.3. Preparación de los datos

Los datos aportados por la investigación eran “*raw data*”, es decir, datos en crudo o en bruto que necesitaban ser tratados para poder fijar un punto de partida desde el que darle sentido a los resultados que se irían obteniendo.

El primer paso para esto fue obtener las columnas referentes a la aceleración en los tres ejes (X, Y, Z) y multiplicarlas por el valor de la gravedad, 9,81, para obtener estos datos en m/s^2 , unidad del sistema internacional para medir la aceleración. Esto se debe a que los datos capturados utilizaban como unidad de medida la aceleración gravitacional, esto es, la aceleración que sufre un cuerpo al estar sometido a la fuerza de la gravedad (o fuerza G, por lo que comúnmente se conoce a esta medida de aceleración como los “Ges”) que es aproximadamente $9,81 m/s^2$, de ahí la necesidad de llevar a cabo esta transformación.

La Figura 26 ilustra este proceso, donde primero se fija la variable que hace referencia al directorio que guarda los datos en crudo (dir_B), para después referenciar y crear el directorio que contendrá los datos en m/s² (dir_P). A continuación, se listan los elementos contenidos en “dir_B”. Cabe destacar que no es necesario crear los directorios de antemano ya que el código incluye la opción de crearlos directamente a través de la función “mkdir”.

```
dir_B = 'Ruta al directorio inicial de datos'
dir_P = 'Ruta al directorio de datos preprocesados'
os.mkdir(dir_P)

vacas = os.listdir(dir_B)
columnas = list(range(0,13,1))

for vaca in (vacas):
    datos_vaca = os.listdir(dir_B + vaca)
    subdir_B = dir_B + vaca + '/'
    datos_p = pd.DataFrame(columns=['lax', 'lay', 'laz'])

    for i in range(len(datos_vaca)-3):
        datos = pd.read_csv(subdir_B + datos_vaca[i+3], sep=' ', usecols=columnas)
        datos_p = datos_p.append(datos[['lax', 'lay', 'laz']]*9.81)

datos_p.to_csv(dir_P + vaca + '.csv', encoding='utf-8', index=False, sep=' ')
```

Figura 24. Procesado de los datos en crudo. Fuente: elaboración propia.

Seguidamente, dentro de cada elemento de ese directorio (carpetas correspondientes a cada vaca) se listan los elementos que contiene (ficheros correspondientes a cada vaca). Se crea un *dataframe* vacío que contendrá las columnas de la aceleración de los ficheros de cada vaca (que serán los datos que se utilizarán para la clasificación).

Por último, se leen los ficheros y se copian las columnas correspondientes al *dataframe* creado, guardando este como un archivo con extensión “csv” (*comma separated values* o valores separados por comas) dentro del directorio que contiene los datos procesados. Hay que mencionar que se han desestimado los tres primeros ficheros, así como los tres últimos al entender que podían contener movimiento correspondiente a la colocación y la retirada del sensor.

Además de esto, no ha sido necesario llevar a cabo cualquier otro tipo de preprocesado debido a que los datos capturados son de gran calidad debido a los tecnología utilizada y eran válidos para el tratamiento que se iba a realizar. Hay que tener en cuenta también que, al tratarse de datos que reflejan magnitudes físicas, al aplicar técnicas de preprocesado, pueden llegar a perder significatividad o interpretación física.

8.4. Modelado

Dentro de la realización de este proyecto se presentan dos modelos de clasificadores que se han desarrollado siguiendo técnicas de *Machine Learning* y que han sido probados sobre los datos mencionados en el punto anterior.

La diferencia entre ambos modelos radica en que uno de ellos se ha probado teniendo en cuenta el análisis de datos descrito anteriormente y solo utiliza los datos en los que las vacas están en movimiento, por lo que será referenciado como modelo filtrado. Por otro lado, el modelo que no se basa en este análisis y utiliza todo el conjunto de datos se nombrará como modelo completo.

Ambos son modelos de clasificación, cuyo objetivo es predecir el estado más probable de una variable categórica (Fayyad & Stolorz, 1997) y tendrán un desarrollo similar teniendo en cuenta la diferencia que ya se ha comentado. Se parte de una extracción de características se conciben como transformaciones de los datos de entrada en variables con un significado y que pueden ser comparables.

Con los datos de estas características se alimentará una red neuronal que clasificará a las vacas en sanas y enfermas, arrojando diferentes métricas para la evaluación de los diferentes modelos. Esta red se ha concebido como un perceptrón multicapa con una capa de entrada, una capa oculta y una capa de salida con una sola neurona, que, mediante una función de activación sigmoïdal, realiza la clasificación binaria.

En definitiva, se trabaja sobre un perceptrón multicapa con una capa oculta, realizando dos pasos a la hora de validar el modelo: primero se realizan las transformaciones necesarias de los datos (extracción de características), para entrenar y probar el algoritmo a través de un modelado predictivo individualizado sobre los datos de cada vaca. Además, se presenta también un modelado predictivo global que se ha llevado a cabo a modo de prueba, con el fin de obtener diferentes datos que nos sirvan para contrastar ambos modelos.

8.4.1. Modelo completo

Como se ha comentado, este modelo utiliza la totalidad de los datos capturados, sin tener en cuenta si un animal está en movimiento o no. Esto resulta en un tratamiento más sencillo de los datos ya que no hacen falta filtros intermedios, sino que, directamente se extraen las características necesarias y se pasan a la red neuronal.

En cuanto a la extracción de características, se comienza creando el directorio que recogerá estos datos, para después listar los elementos que anteriormente se han guardado en el directorio de datos procesados mediante la función “`listdir`” del módulo “`os`”. Se fijan además a 0 dos contadores que servirán para añadir elementos dentro de dos *dataframes* que se inicializan con los nombres de columna recogidos en la lista “*columns*”.

Estas estructuras de datos recogerán, por un lado; los datos que caracterizan a todas las vacas y que se utilizarán en el ajuste global del modelo y, por otro lado; los datos de cada vaca en ficheros diferenciados, que se utilizarán en el modelado predictivo individualizado. Además, de los elementos que componen estos *dataframes*, el primero hace referencia a las etiquetas de clase y el resto, a las características que se utilizarán en la predicción.

Por último, dentro del bucle *for* que se ha establecido sobre los elementos del directorio de datos procesados, se irán leyendo los ficheros correspondientes a cada una de las vacas en cada iteración para extraer sus características. En este caso, se recogerán en lotes de 10 segundos, de ahí el hecho de realizar la división de la longitud total del fichero de datos de una vaca entre el valor 500 (que como se ha explicado equivale a 5 segundos).

Al realizar la división con dos barras, se mantiene únicamente la parte entera del resultado, dejando fuera los datos que no se ajusten al tamaño del lote. Por ejemplo, si un fichero contiene 10200 filas, el resultado de la división en este caso sería 20.

```
for vaca in datos_mov:
    aux=0
    features=pd.DataFrame(columns=columns)
    datos=pd.read_csv(dir_P + vaca, sep=' ')
    rango = len(datos.index) // 500
```

Figura 25. División de los datos en lotes. Fuente: elaboración propia.

A continuación, dentro del número de iteraciones necesarias en cada fichero, se toman 500 datos que se guardan en la variable “lote”. Sobre estos datos se realizan diversas operaciones

para obtener una serie de características que, en este caso, se han tomado de Zebin et al. (2017), donde se describen y se explican con claridad, lo que es una gran ayuda para el desarrollo de trabajos de la misma índole, como este. Las características extraídas se explican en la Tabla 8.

Tabla 8. Lista de características extraídas de los datos del acelerómetro. Fuente: Zebin et al. (2017)

Nombre de característica	Número de característica	Información adicional
Valor medio (1 cada uno)	Característica (1) Característica (2) Característica (3)	Para los tres componentes de la aceleración (dirección x,y,z)
Valor cuadrático medio (1 cada uno)	Característica (4) Característica (5) Característica (6)	Para los tres componentes de la aceleración
Características de autocorrelación (3 cada uno)	Característica (7:9) Característica (10:12) Característica (13:15)	Altura del pico principal; altura y posición del segundo pico

En el código se pueden visualizar de forma clara cada una de las características recogidas en esta tabla, calculadas a través de sus fórmulas correspondientes y mediante la función “find_peaks” sobre la autocorrelación de cada componente de la aceleración.

Para añadir las etiquetas de clase a los diferentes *dataframes* utilizados, se utiliza una expresión regular que compara la letra que se encuentra justo delante de la extensión del archivo (.csv). Si dicha letra es una h la vaca estará sana y se añadirá como etiqueta un 0. En caso de que sea una l, se añadirá un 1 como etiqueta de clase. Además, atendiendo a la formulación de los diferentes bucles utilizados, se guarda un archivo de características por cada vaca al pasar a formato “csv” la variable utilizada para guardar los datos individuales y un fichero global que incluye las características de todas las vacas.

La Figura 28 muestra las primeras fases en el desarrollo la red neuronal, que se utiliza para realizar un ajuste global del modelo sobre el problema planteado y, obtener de esta forma, un valor global que nos pueda indicar si el modelo planteado puede tener éxito o no.

```
data=pd.read_csv(dir_C + file, header=0, encoding='utf-8', sep=' ')
x = data.drop('Grupo', axis=1)
y = data['Grupo']

X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2)

model = Sequential()
model.add(Dense(32, input_shape=(15,), activation='relu'))
model.add(Dense(16, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```

Figura 26. Primera parte del desarrollo de una red neuronal para el ajuste global. Fuente: elaboración propia.

En primer lugar, se cargan en un *dataframe* las características que se han presentado anteriormente y se separan las etiquetas de clase con las que se compararán las diferentes predicciones. Luego se divide el conjunto de datos en datos de entrenamiento y datos de prueba, siendo el tamaño de los datos de prueba un 20% del total.

A continuación, se crea el modelo y se añaden en este caso tres capas, una de entrada, una oculta y una de salida con 32, 16 y 1 neurona en cada capa y con sus correspondientes funciones de activación. Cabe destacar que, al tener 15 datos de entrada, en la primera capa ha sido necesario añadir el parámetro “*input_shape*” para ajustar la capa de entrada al número de características.

Además, la función de activación escogida para la capa de salida ha sido la función sigmoide ya que se trata de una clasificación binaria. Las funciones de activación se utilizan para devolver valores de salida (*outputs*) en función de los valores de entrada (*inputs*), dotando al modelo de no linealidad. En este caso, la función sigmoide devuelve valores entre 0 y 1 por lo que es adecuada para el problema tratado en este caso.

El siguiente paso es compilar el modelo, especificando la función de pérdida y el optimizador a utilizar (*binary_crossentropy* y Adam) para después entrenarlo con la función “*fit*” definiendo los diferentes hiperparámetros necesarios. El parámetro “*epochs*” define el número de veces que se itera sobre los datos en el entrenamiento, mientras que el parámetro “*batch_size*” hace referencia al número de muestras que actualizan sus pesos en cada iteración.

La función de pérdida, como ya se comentó, se utiliza para comparar los valores predichos con los reales, por lo que el objetivo del modelado es minimizar dicha función. En este caso se ha escogido la función *binary_crossentropy* al tratarse de un problema de clasificación con dos clases. Por su parte, el optimizador va actualizando y optimizando, valga la redundancia, los pesos dentro de la red durante el entrenamiento.

Se ha incluido un “*early stopping*” que traducido literalmente al español sería “parar pronto”, cuya función como bien indica su nombre es la de finalizar la ejecución del modelo si los valores obtenidos no mejoran después de varias épocas.

Una vez se ha entrenado el modelo se realizan las predicciones sobre los datos de prueba. Además, las predicciones que se han obtenido se transforman en binarias igualándose a 0 o a 1 en función de que su valor esté por encima o por debajo de 0,5. Después, se crean las

métricas que se utilizarán para evaluar el desempeño del modelo, comparando las predicciones con los datos de prueba. Las métricas utilizadas son las siguientes:

- » **Accuracy.** Mide la exactitud del modelo, es decir, que proporción de casos se han clasificado con éxito sobre el total de las predicciones.
- » **Precision.** Mide el éxito a la hora de predecir casos positivos y resulta de dividir las predicciones positivas exitosas (denominadas como verdaderos positivos, “VP”, o *true positives*, “TP”, por sus siglas en inglés), entre el total de predicciones positivas, ya sean verdaderos o falsos positivos (“FP”, casos predichos como positivos, que realmente son negativos).
- » **Recall.** Mide la sensibilidad del modelo, es decir, su capacidad para predecir correctamente casos positivos. Resulta de dividir el los “VP” entre la suma de estos mismos valores con los falsos negativos (“FN”, casos predichos como negativos, que realmente son positivos).

Por último, se muestra la matriz de confusión para poder visualizar y estudiar de forma gráfica, los resultados que se han dado durante la ejecución del modelo y se imprimen los valores de las métricas mencionadas. De esta matriz se pueden derivar fácilmente las métricas que se explican anteriormente, convirtiéndolas de esta forma en las métricas más populares a la hora de evaluar diferentes modelos, como recogen Arisholm et al. (2010).

En el caso del modelado predictivo individual, la primera característica distintiva es el hecho de que este se entrena sobre las características de unas vacas concretas y se prueba sobre las restantes. Para esto, se guardan en dos listas lo ficheros individuales de características extraídos en los pasos previos: una lista para los datos de las vacas que se utilizarán en el entrenamiento del modelo y otra para los datos de las vacas sobre las que se realizarán las pruebas.

Dado que el modelo debe realizar predicciones sobre todo el conjunto de datos, se ha desarrollado una estructura que rellena las listas mencionadas de forma aleatoria en cada ejecución. A la hora de insertar ficheros de entrenamiento o prueba, se ha contemplado la necesidad de introducir varios ficheros de vacas enfermas para poder entrenar un modelo con datos de ambas clases.

Para esta variante, los datos de entrenamiento se cargan concatenando en un solo *dataframe* los datos pertenecientes a los ficheros dentro de la lista de datos de entrenamiento, para después crear y compilar el modelo como ya hemos visto anteriormente, además de definir el

early stopping. A parte, se dividen también los datos para separar las etiquetas de clase y se entrena el modelo mediante la función *fit*.

Es necesario mencionar también el hecho de que las clases están claramente desbalanceadas, contando con 15 vacas sanas y solo 6 enfermas, por lo que se ha llevado a cabo un balance de las clases a través de un sobre muestreo de la clase minoritaria.

Como se puede apreciar, ambos ajustes siguen una misma estructura y las diferencias existentes se pueden observar de forma clara y no suponen una dificultad superlativa a la hora de entender el código realizado, sino que se adaptan de forma flexible a la solución desarrollada para este problema de clasificación.

Para llevar a cabo la evaluación del modelo, se debe tener en cuenta que, al predecir de forma individual sobre un fichero de un animal concreto, este presentará una única clase, por lo que:

- » La matriz de confusión presentará una única fila de datos que se asociará a uno de los valores reales dependiendo de la clase a la que corresponda cada fichero.
- » Esto deriva en la imposibilidad de extraer las métricas que se basen en la detección de falsos positivos o falsos negativos (*precision* y *recall*).

Debido a esto, la evaluación del modelo pasará por obtener la predicción asociada a cada fichero y evaluar la probabilidad de que se pueda dar esa predicción.

8.4.2. Modelo filtrado

En el caso del modelo filtrado, el proceso que se ha seguido es muy parecido al del modelo, pero ha sido necesario, separar las características según las diferentes ventanas de tiempo como se ha mostrado en puntos anteriores para poder predecir en que momentos una vaca podía estar en movimiento o no y utilizar estos datos, en detrimento de aquellos correspondientes a los instantes donde los animales no se movían.

La primera diferencia significativa dentro de este modelo pasa por la creación un modelado predictivo global previo para cada ventana de tiempo, para poder obtener sus resultados y decidir qué tipo de ventana se adapta mejor al problema planteado y, de esta forma, construir el modelo individual mediante los datos de esta.

La transformación del código a la hora de extraer las características en este modelo es bastante sencilla ya que se trata de añadir un bucle “*for*” que vaya iterando sobre cada una de

las ventanas y de esta forma, cargar un fichero global de características de cada ventana para pasar como datos de entrada a la red neuronal en cada caso. Esta acción se puede observar en la Figura 29 que se presenta a continuación.

```
for name in names:
    subdir_V = dir_V + name + '/'
    datos_mov = os.listdir(subdir_V)
    energy = []
    aux = 0
    features=pd.DataFrame(columns=columns)

    for vaca in datos_mov:
        datos=pd.read_csv(subdir_V + vaca, sep=' ')
        rango = len(datos.index) // 500
```

Figura 27. División de los datos de cada ventana estudiada en lotes. Fuente: elaboración propia.

Hay que tener en cuenta que, al iterar sobre diferentes ventanas de tiempo, cada una con sus ficheros de datos correspondientes (individuales y globales); será necesario guardar los diferentes ficheros creando subcarpetas específicas para cada ventana y de esta forma, organizar los datos de manera eficiente para facilitar la posterior construcción de los diferentes algoritmos.

En cuanto a la creación del modelo predictivo global, la transformación ha sido exactamente la misma, añadiendo un bucle al inicio del código para recorrer los diferentes ficheros de características y cargar uno de ellos en cada iteración, obteniendo de esta manera el ajuste global de cada una de las ventanas de tiempo y unos resultados que pueden ser comparados para encontrar la ventana de tiempo que mejor se ajusta a este modelo.

Hay que tener en cuenta que iterar de esta forma conlleva un gran coste computacional ya que, explicado de forma sencilla, se trata de realizar la misma acción que en el modelo completo once veces seguidas, lo que supone un consumo de recursos y de tiempo bastante elevados.

La comparación de los resultados obtenidos sobre las diferentes ventanas de tiempo se presentará en el siguiente punto, basándose en las métricas ya mencionadas. Una vez realizada esta comparación es también posible realizar una comparación entre modelos.

Esta primera comparación puede no aportar una información definitiva para elegir sobre uno u otro modelo, pero puede ser de ayuda a la hora de estudiar y observar características tales

como la forma en la que va disminuyendo la función de pérdida o si se llega a dar el caso de que se activa el *early stopping* desarrollado.

El siguiente paso es desarrollar el modelado predictivo individual de la misma forma que se ha desarrollado para el modelo completo ya que, una vez se extraen las características, los modelos desarrollados son similares. Cabe mencionar el hecho de que en ambos modelos se ha añadido una capa oculta más (con 8 neuronas), con el fin de mejorar las predicciones.

8.5. Evaluación

Para evaluar cada uno de los modelos presentados se han utilizado las métricas que ofrece la librería *scikit-learn* y que se han mostrado en el apartado anterior. El valor de estas métricas no es fijo, sino que cambia con cada ejecución de un modelo ya que el aprendizaje no se da siempre de la misma forma.

Por eso para cada una de las fases donde se han obtenido estas métricas, se han llevado a cabo varias ejecuciones del código y se ha realizado una media aritmética sobre los valores obtenidos, para poder conseguir de esta manera unos datos más fiables y contrastados.

8.5.1. Resultados del modelo completo

Al igual que en el punto anterior, se presentan en primer lugar los resultados obtenidos en el modelo completo ya que su tratamiento y explicación resultan más sencillos, y ayudarán al lector a situarse y comprender de manera correcta los contenidos aquí recogidos.

Como se comentó anteriormente, a parte del modelo de ajuste individualizado que se presenta en el código del proyecto, se realizaron pruebas a través de un modelado predictivo global para obtener información adicional acerca de los modelos utilizados.

En la Tabla 9 se recogen los resultados obtenidos a la hora de realizar cinco ejecuciones consecutivas sobre el código en el entorno de desarrollo mencionado con anterioridad. Una vez que se obtenían los resultados, se anotaban y se pasaba a la siguiente ejecución para evitar cualquier problema derivado de la integridad de los datos, las conexiones de red, los elementos del hardware, etc.

Tabla 9. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste global del modelo completo.
Fuente: elaboración propia

	Ejecución 1	Ejecución 2	Ejecución 3	Ejecución 4	Ejecución 5	Media	σ
Precision	0,7584	0,7850	0,7703	0,7543	0,7676	0,7671	0,0107
Recall	0,6647	0,7032	0,6795	0,7188	0,7324	0,6997	0,0248
Accuracy	0,7851	0,8057	0,7969	0,7990	0,8053	0,7984	0,0075

Analizando los resultados obtenidos podemos observar cómo los valores de la desviación estándar en todas las métricas son bastante bajos, no superando en ningún caso el 2,5% y presentando la métrica *accuracy* una desviación muy cercana a 0. Esto muestra que, entre las diferentes ejecuciones del modelo, los resultados son muy parecidos entre sí, lo que indica un ajuste consistente.

Además, la métrica *accuracy* es la que presenta mejores resultados en todas las ejecuciones por lo que el éxito del modelado predictivo global ronda el 80%. Es necesario comentar el hecho de que el valor de una sola métrica no define el éxito o no de un modelo, ya que puede tratarse, por ejemplo, de un problema donde las clases estén desbalanceadas, y mientras una métrica tiene valores muy buenos, otras pueden presentar deficiencias.

En este caso, como se puede observar, aunque la métrica *recall* tenga un valor medio un poco inferior, la desviación estándar entre los tres resultados medios supero en una décima el 4% por lo que no se puede entender como unos resultados con sesgos importantes.

A continuación, se presentan los resultados obtenidos a través del ajuste individualizado. Solo se presentan 3 valores ya que el coste computacional de este modelo es muy alto y tarda bastante en mostrar los resultados. En este caso, como ya se ha comentado anteriormente, no es posible obtener las métricas de *recall* y *precision* ya que se predice sobre un fichero con una clase concreta, en lugar de un conjunto de datos con ambas clases.

Por esta razón, se ha utilizado el método del voto mayoritario que toma como predicción final la que se ha repetido más veces a la hora de predecir sobre cada muestra del fichero de datos. Además, se muestra la probabilidad de esa predicción, valor que resulta al dividir el número de muestras que se han predicho correctamente entre el total de muestras.

Tabla 10. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste individual sobre los datos de cada vaca, dentro del modelo completo. Fuente: elaboración propia

Vaca	00749_h	00808_h	05057_h	05156_h	05317_l
Predicción	Sana	Sana	Sana	Sana	Enferma
Probabilidad	77,67%	91,62%	69,16%	68,30%	72,77%

Vaca	05317_l	00749_h	00808_h	00904_h	005252_l
Predicción	Enferma	Sana	Sana	Sana	Sana
Probabilidad	76,69%	86,29%	92,78%	70,40%	55,18%

Vaca	00749_h	00850_h	005057_h	05138_h	05342_h
Predicción	Sana	Enferma	Sana	Sana	Enferma
Probabilidad	81,06%	58,27%	72,88%	64,50%	52,85%

Como se puede observar nos encontramos con diferentes situaciones. Por un lado, predicciones correctas y con muy buenos resultados, sobre todo de vacas sanas, teniendo en cuenta que es la clase mayoritaria, sobre la que el modelo ha entrenado más muestras diferentes.

Por otro lado, nos encontramos con predicciones entre el 55% y el 85% para ambas clases que se asemejan a los valores obtenidos en el ajuste global y que muestran como el modelo tiene ciertos errores de predicción, aunque siguen obteniendo la clase correcta con porcentajes que se presentan como aceptables, aunque mejorables en ciertas ocasiones.

Por último, se nos presentan casos donde el modelo no ha sido capaz de predecir correctamente y clasifica a las vacas en la clase equivocada. Esto refleja el hecho de que las características utilizadas pueden no ser del todo discriminantes, ya que el modelo no es capaz de diferenciar entre vacas sanas y enfermas.

8.5.2. Resultados del modelo filtrado

En primer lugar, se visualizan los valores de las métricas que se han utilizado para elegir cual era la ventana de tiempo que mejores resultados presentaba, con el fin de utilizarla como ventana de tiempo de referencia a la hora de construir el modelo de ajuste individual.

En este caso, la ventana con mejores resultados ha sido la de 12 minutos con un solapamiento de 4 minutos. Solo se han realizado 3 ejecuciones ya que al tener un número elevado de ventanas el coste computacional era alto como ya se ha mencionado. A continuación, en la

Tabla 11, se presentan los resultados obtenidos en cada ejecución del código, así como el valor medio final y la desviación estándar entre las diferentes ejecuciones.

Tabla 11. Resultados y media de 3 ejecuciones de la red neuronal sobre las diferentes ventanas de tiempo para analizar y escoger la que presenta mejores resultados. Fuente: elaboración propia

	Ejecución 1	Ejecución 2	Ejecución 3	Media	σ
Precision	0,7094	0,7522	0,7301	0,7306	0,0175
Recall	0,6864	0,601	0,69	0,6591	0,041
Accuracy	0,8037	0,803	0,8167	0,8078	0,0063

Aunque para obtener estas métricas solo se hayan hecho tres ejecuciones se puede valorar de igual forma que en el punto anterior el hecho de que las desviaciones estándar siguen siendo bastante bajas, pero llama bastante la atención el caso del *recall*.

El resultado de la segunda ejecución se aleja bastante de los otros, presentando el ajuste una muy leve inconsistencia a la hora de medir la sensibilidad.

A continuación, se presentan, para cada modelo, los valores obtenidos en cuanto al ajuste del modelo globalizado. Como se ha mencionado con anterioridad, estos datos miden el porcentaje de acierto del modelo de forma global en sus predicciones sobre si una vaca se encuentra sana o enferma, pero no arrojan información específica de cada animal.

En este caso, el código se ha ejecutado en cinco ocasiones, obteniéndose por cada métrica, los valores que se muestran en la Tabla 12. Además, se han calculado tanto la media aritmética como la desviación estándar en cada caso.

Tabla 12. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste global del modelo filtrado. Fuente: elaboración propia

	Ejecución 1	Ejecución 2	Ejecución 3	Ejecución 4	Ejecución 5	Media	σ
Precision	0,6989	0,6959	0,716	0,7208	0,7146	0,7092	0,01
Recall	0,6725	0,7049	0,6591	0,6583	0,6528	0,6695	0,0188
Accuracy	0,8009	0,8034	0,7997	0,8005	0,8015	0,8012	0,0012

A la hora de elegir una ventana y realizar el modelado predictivo global sobre ella, se puede observar como la perturbación que se había dado en el *recall* se ha corregido y su desviación estándar es incluso menor que la del modelo completo para este mismo caso.

Si es necesario comentar que la desviación estándar entre los valores medios de las tres métricas se eleva en este caso hasta el 5,5%. Tampoco es un valor demasiado alto, pero

muestra como las métricas del modelo anterior están mejor balanceadas que las de este modelo en lo que al ajuste global se refiere.

Por último, se muestran los resultados obtenidos en el modelo filtrado, pero esta vez trabajando de manera individual sobre cada una de las vacas. Hay que tener en cuenta, como ya se explicó anteriormente, que en este caso las predicciones se realizan sobre diferentes vacas en cada ejecución

Tabla 13. Resultados y media de 5 ejecuciones de la red neuronal que representa un ajuste individual sobre los datos de cada vaca, dentro del modelo filtrado. Fuente: elaboración propia

Vaca	05309_1	05252_1	00808_h	00904_h	05342_h
Predicción	Enferma	Enferma	Sana	Sana	Sana
Probabilidad	78,83%	79,96%	83,03%	59,23%	65,63%

Vaca	05317_1	00808_h	01063_h	01113_h	05009_1
Predicción	Enferma	Sana	Sana	Sana	Sana
Probabilidad	77,58%	80,15%	70,99%	61,22%	52,94%

Vaca	00749_h	01032_h	01063_h	05057_h	05093_h
Predicción	Sana	Sana	Sana	Sana	Sana
Probabilidad	89,27%	81,77%	81,31%	79,03%	76,07%

En este caso se pueden visualizar las mismas situaciones que las mencionadas en el modelo completo, pero cabe destacar el hecho de que solo ha habido una predicción errónea y con un porcentaje muy cercano al 50%, tratándose de una vaca clasificada con un 3 en el estudio veterinario previo, por lo que puede tratarse de una vaca que presente características similares a un animal sano pero clasificado con un 2.

Aun así, se puede confirmar que las características utilizadas en Zebin et al. (2017), no son las óptimas para este trabajo de clasificación binaria.

8.6. Despliegue

A continuación, se detallan los diferentes elementos que se han creado para presentar un plan de negocio que pueda utilizarse para desplegar y desarrollar comercialmente la tecnología que se ha presentado anteriormente

8.6.1. Modelo *Canvas*

- » **Socios clave.** Es importante destacar el hecho de que, en nuestra sociedad, cada día se pone más en valor el paradigma del bienestar animal que, al fin y al cabo, es el objetivo que persigue nuestra tecnología por lo que será clave asociarnos con organizaciones cuyos objetivos se alineen con los nuestros y que puedan percibir nuestro producto como una forma de ayudarles a alcanzar esos objetivos. Al tratarse del lanzamiento de una nueva tecnología, será muy beneficioso compartirla con la comunidad científica donde se podrán encontrar apoyos para futuras modificaciones de este proyecto en busca de nuevos usos. Además, en este caso, el trabajo se ha basado en referencias de otros trabajos sobre problemáticas parecidas, lo que es un claro argumento a favor de incluir a la comunidad científica dentro de este apartado. Otro factor clave son las asociaciones de ganaderos, que se presentan como socios estratégicos que mediante reuniones o a través del boca a boca podrán difundir las ventajas de esta tecnología, además de ofrecer oportunidades para la presentación del producto a nuevas explotaciones interesadas. Por último, al ser necesarios para nuestra actividad los sensores inerciales, sería conveniente la asociación con empresas que produjeran dichos sensores en masa para aprovechar las economías de escala y poder cerrar acuerdos beneficiosos a la hora de adquirir estos dispositivos.
- » **Actividades clave.** Las cinco primeras actividades son los pasos en los que se puede dividir, a grandes rasgos, la tecnología desarrollada por lo que serán primordiales para su correcto funcionamiento. El primer paso será capturar los datos de diferentes animales, teniendo en cuenta que se debe obtener datos tanto de animales sanos como de otros que presenten disfunciones motoras, para que seguidamente sean procesados y se extraigan las características que ayudarán a agrupar a los animales según su estado, para después pasar estos elementos a los modelos para su clasificación. Por último, se procederá a analizar los resultados obtenidos para verificar el acierto que ha tenido el sistema y, si fuera necesario, realizar las correcciones pertinentes (saltando de nuevo al paso 2). A la

hora de implantar el producto será necesario instalar el sensor a los animales, preparar todo el equipo de clasificación y explicar a los responsables su funcionamiento. Además, como en cualquier producto de software, será necesario un mantenimiento para controlar el correcto funcionamiento del producto y también las posibles mejoras que se puedan ir incluyendo en busca de alargar su ciclo de vida.

- » **Recursos clave.** En primer lugar, se deben destacar los dispositivos utilizados en la captura de los datos que, mediante una banda que normalmente se utiliza para introducir los dispositivos móviles a la hora de realizar entrenamientos, se colocan en las patas traseras de las vacas. Es cierto que son dispositivos con una captura de datos muy estable pero no serían rentables en fase de producción debido a su elevado coste por lo que se optaría por sensores específicos para esta tarea con características similares al que posee el dispositivo usado en esta fase de prototipado. Son fundamentales también los equipos que se utilizarán para procesar esos datos ya que serán tareas exigentes y de alto coste computacional. En un entorno rural donde esta tecnología no se usa de forma frecuente se debe destacar la calidad de la información que se ofrece a los clientes, adaptándola en todo momento a las necesidades de cada interlocutor para lo que será necesario además contar con un equipo de expertos que puedan explicar en detalle cada una de las dudas que puedan surgir al cliente durante todo el proceso.
- » **Propuesta de valor.** La propuesta de valor de este producto se basa en ofrecer una innovación real dentro de la actividad ganadera, que la adapte en la medida de lo posible a la era de las Tecnologías de la Información y la Comunicación (TIC). Utilizando los avances tecnológicos existentes, se buscará detectar anticipadamente disfunciones motoras en vacas, ayudando de esta forma a los ganaderos a mejorar la rentabilidad de sus explotaciones. Además, aunque en este caso el producto solo se haya probado en vacas, sus características permiten que se pueda adaptar a otros animales, tratándose en cualquier caso de una técnica poco invasiva y de fácil implementación.
- » **Relación con el cliente.** El correcto uso de la tecnología desarrollada se transmite a los potenciales clientes a través de una demostración del producto por lo que preferimos una relación con el cliente en la que este pueda observar los resultados, ya sea presencial o telemáticamente. Por esta razón se proponen las demos, que serían el último paso para poder cerrar una posible venta. Además, tanto en reuniones, presentaciones y charlas, se presentan las especificaciones y los datos más significativos del producto, así como

resultados obtenidos en pruebas realizadas con anterioridad. Estas se contemplan como un paso previo en el que convencer al cliente para probar el producto y, si todo va bien, adquirirlo tras una demo.

- » **Canales.** En un proyecto de este tipo que trata de implantar una nueva tecnología en el medio rural, será necesaria la atención presencial para poder transmitir a los ganaderos todas las ventajas que se derivan de la implantación del producto, así como las características técnicas que puedan ser de interés para estos. A la hora de resolver dudas puntuales que puedan surgir, durante el desarrollo de las acciones necesarias para esta implantación, se habilitará la comunicación telefónica y por medios electrónicos, con la finalidad de brindar una respuesta rápida ante las necesidades de los clientes.
- » **Segmento de clientes.** La tecnología desarrollada está claramente dirigida hacia las vacas en un primer momento por lo que los principales clientes serán las explotaciones ganaderas de vacuno, para cuidar el estado de estos animales y de este modo maximizar su eficiencia y su productividad. Es cierto también que la base de esta tecnología es extrapolable a otros usos ya sea en otro tipo de animales, de ahí que se incluya en el segmento de clientes las clínicas veterinarias. Del mismo modo se podrá ofrecer la tecnología tanto a centros de investigación como de enseñanza para que puedan aprender sobre su uso y desarrollar otras tecnologías a partir de esta.
- » **Estructura de costes.** Los sensores con los que se van a capturar los datos, además de los equipos de procesamiento necesarios son gastos fundamentales dentro de la actividad. Por otro lado, habrá que crear demos para utilizar en las pruebas con los potenciales clientes lo que supondrá unos costes, al igual que el desplazamiento y el tiempo que nos lleven estas pruebas. Hay que tener en cuenta también que, al tratarse de presentaciones con interés comercial, la organización tendrá un coste. Cabe mencionar que, en los dos últimos casos mencionados, ese coste podrá derivar en la captación de nuevos clientes. No se pueden obviar las actualizaciones ya que la tecnología necesitará mantenimiento y mejoras para no quedar obsoleta. Por último, hay que incluir también los gastos de personal, ya que son una parte fundamental de la estructura.

» **Fuentes de ingreso.** Atendiendo a la estructura del producto desarrollado se deben tener en cuenta tanto la parte hardware como la parte software ya que ambas, en cualquier caso, se podrían comercializar por separado, fijando el precio de un pack en función de los precios de ambos módulos. Está claro que lo más sensato en este caso es ofrecer el producto completo a todo el segmento de clientes, lo que será la mayor fuente de ingresos del proyecto. Se tienen en cuenta también las subvenciones y ayudas que podamos recibir, atendiendo por ejemplo al hecho de que se trata de una tecnología innovadora y que busca el bienestar animal.

<p>SOCIOS CLAVE</p> <p>Asociaciones que busquen el bienestar animal (ONG, partidos políticos, etc.)</p> <p>Comunidad científica</p> <p>Asociaciones de ganaderos</p> <p>Empresas de desarrollo de sensores inerciales</p>	<p>ACTIVIDADES CLAVE</p> <p>Captura de datos</p> <p>Preprocesado de los datos</p> <p>Extracción de características</p> <p>Clasificación automática</p> <p>Análisis de resultados</p> <p>Instalación del producto</p> <p>Mantenimiento y mejora</p>	<p>PROPUESTA DE VALOR</p> <p>Innovación y adaptación del sector agrario a la era TIC</p> <p>Capacidad de detectar anticipadamente posibles disfunciones motoras</p> <p>Capacidad de adaptación del producto a diferentes animales</p>	<p>RELACIÓN CON EL CLIENTE</p> <p>Reuniones</p> <p>Demos</p> <p>Presentaciones y charlas</p> <p>Atención personalizada</p>	<p>SEGMENTO DE CLIENTES</p> <p>Explotaciones ganaderas</p> <p>Clinicas veterinarias</p> <p>Centros de investigación y enseñanza</p>
<p>RECURSOS CLAVE</p> <p>Sensores precisos</p> <p>Equipos de procesado</p> <p>Información clara y precisa</p> <p>Personal especializado</p>	<p>Técnica poco invasiva y de fácil implementación</p>	<p>CANALES</p> <p>Atención telefónica</p> <p>Atención presencial</p> <p>Medios online</p> <p>Revistas especializadas</p>		
<p>ESTRUCTURA DE COSTES</p> <p>I+D+i</p> <p>Equipos de procesado</p> <p>Sensor inercial (IMU)</p> <p>Demos y pruebas con clientes</p> <p>Presentaciones y promociones</p> <p>Gastos de personal</p>		<p>FUENTES DE INGRESOS</p> <p>Venta de los sensores inerciales</p> <p>Subvenciones y ayudas</p> <p>Licencias de uso</p>		

Figura 28. Modelo Canvas. Fuente: elaboración propia.

8.6.2. Análisis DAFO

	INTERNO	EXTERNO
NEGATIVO	<p>DEBILIDADES</p> <p>Tecnología nueva (en versión experimental) Margen de error dentro de los resultados obtenidos Resultados dilatados en el tiempo, no inmediatos</p>	<p>AMENAZAS</p> <p>Rechazo de los ganaderos Falta de información en el sector Incremento de los costes del cereal</p>
POSITIVO	<p>FORTALEZAS</p> <p>Captura de datos estable Detección de enfermedades en sus primeros estadios Mantenimiento, mejora y actualización del software Campo de estudio apoyado en potentes herramientas Mejora de la productividad y ahorro en costes</p>	<p>OPORTUNIDADES</p> <p>Ya se han implantado sensores en animales de ganado Presión hacia la transformación digital de los negocios Mayor importancia del bienestar animal en la sociedad Mayor acceso a la financiación</p>

Figura 29. Análisis DAFO. Fuente: elaboración propia

- » **Debilidades.** La base del producto desarrollado es una tecnología nueva y en versión experimental (este proyecto, como ya se ha comentado, se refiere a la fase de prototipado, no a la de producción) por lo que puede presentar carencias en su desarrollo o no abordar todas las funcionalidades que los clientes esperan o pueden necesitar. Hay que tener en cuenta también que los resultados obtenidos no son inmediatos puesto que el procesado lleva un tiempo y será necesario contrastar e interpretar estos resultados que, además, tienen un margen de error como ya se ha visto, por lo que puede derivar en un sobrecoste a la hora de tratar a un animal, aunque las posibilidades son bastante bajas.
- » **Amenazas.** Dado que el ámbito tecnológico y el rural han estado tradicionalmente separados y en tendencias muy distintas, los ganaderos pueden entender el producto erróneamente como un factor de riesgo para sus animales o solo ver el posible sobrecoste en lugar de los beneficios que puede suponer la implantación de esta tecnología. Esto puede deberse también a una falta de información generalizada dentro del sector que puede llevar a que los ganaderos no conozcan realmente las posibilidades del producto y cómo puede ayudarle tanto a ellos como a sus animales. Además, tal y como se está desarrollando la economía actual, con una inflación disparada que repercute de manera directa en los ganaderos a la hora de comprar piensos y forraje para los animales, puede parecer que no es el momento adecuado para hacer una inversión de este tipo.

- » **Fortalezas.** El producto se ha desarrollado pensando en el beneficio que puede suponer para los ganaderos en cuanto a la detección de enfermedades en los primeros estadios, conllevando esta acción un tratamiento más sencillo, más corto y barato, mejorando la productividad, limitando el sufrimiento animal y ahorrando costes a la explotación, que de otra manera podrían llevarla a la ruina. Cabe destacar también que, aunque se trate de la fase de prototipado, se han utilizado sensores potentes que permiten una captura de datos estables y que demuestran como este tipo de desarrollos se apoyan en herramientas que cada día se desarrollan a una velocidad vertiginosa y que permiten seguir mejorando los productos de este tipo. Esta mejora es un punto clave dentro de cualquier producto software ya que dentro de su ciclo de vida siempre se debe incluir el mantenimiento y las actualizaciones que vayan siendo necesarias, no solo para corregir errores, sino con vistas a nuevas funcionalidades o extensiones que puedan ayudar a la expansión del producto en otros ámbitos.

- » **Oportunidades.** Es necesario mencionar que este proyecto no es el primero que incluye dentro de su desarrollo la utilización de sensores en animales, lo que puede suponer una oportunidad a la hora de su aceptación tanto por parte de los ganaderos como de los propios animales, al tratarse de una técnica poco inclusiva y con la que ya pueden estar familiarizados. Además, se basa en dos pilares fundamentales que hoy en día marcan las pautas de la sociedad en la que nos encontramos: el bienestar animal y la digitalización de los negocios. Al fin y al cabo, es el propio paradigma social el que necesita de iniciativas que aúnen y pongan en valor las vías de evolución que se van presentando. Además, una vez aprobados los fondos de recuperación de la Unión Europea, conocidos como Next Generation, proyectos de este tipo cuentan con una fuente de financiación excepcional ya sea en el marco del programa Horizonte Europa (para la investigación y el desarrollo tecnológico) o dentro del FEADER (Fondo Europeo Agrario de Desarrollo Rural), además de muchas otras subvenciones que pueden ser solicitadas tanto a nivel nacional como regional.

8.6.3. Análisis de viabilidad y desarrollo del negocio.

A la hora de desarrollar el análisis de la viabilidad de este proyecto se han tenido en cuenta 3 indicadores: el plazo de recuperación (o *payback*), el Valor Actual Neto (VAN) y la Tasa Interna de Retorno (TIR). Cada una de estas métricas se explica con mayor detalle a continuación:

- » **Payback:** como su nombre indica, el plazo de recuperación hace referencia al tiempo necesario para recuperar la inversión inicial del proyecto. Se calcula mediante el acumulado de los flujos netos de caja.
- » **VAN:** el valor actual neto, como su nombre indica, actualiza los flujos netos de caja a su valor presente mediante una tasa de descuento. De esta forma se puede saber si un proyecto generará un valor superior a su inversión inicial
- » **TIR:** la tasa interna de retorno es aquella que, al utilizarla como tasa de descuento en el cálculo del VAN, conlleva que este sea igual a cero.

Para calcular estos tres indicadores es necesario un trabajo previo de conocimiento y estimación de las principales variables económicas que entran en juego, tanto en la actividad estudiada, como en el desarrollo del propio proyecto. Muchas de estas variables se han obtenido mediante una comunicación directa con la veterinaria que colabora con los proyectos de investigación mencionados al inicio.

Las diferentes variables económicas que se han tenido en cuenta a la hora de analizar la viabilidad de este proyecto se recogen en la Tabla 14. Teniendo en cuenta que el método elegido para llevar a cabo el desarrollo del negocio es el *Lean Startup*, los datos que se muestran a continuación harán referencia a una primera etapa de este modelo que con el paso del tiempo puede sufrir alteraciones hasta obtener un modelo de negocio estable.

Tabla 14. Variables económicas iniciales del proyecto. Fuente: elaboración propia.

Concepto	Coste	Descripción
Pérdidas derivadas de la cojera de una vaca	50€/mes	De 100€ a 1100€ anuales dependiendo del grado de la afección, que se puntúa de 2 a 5 (ya que un 1 representa ausencia de enfermedad)
Coste medio visita veterinaria y tratamiento	150€	La visita costaría entre 60€ y 80€ mientras que el tratamiento rondaría de 40€ (antibióticos) a 120€ (añadiendo vendaje y taco)
Coste sensor inercial	50€ * 500u. = 25.000€	Aprovisionamiento inicial. Un solo sensor se puede utilizar hasta con 2 vacas en un mismo día
Coste equipo de procesado	2.500€ * 4 = 10.000€	Se estiman 4 equipos para poder llevar a cabo las tareas de procesado de forma simultánea
Coste personal	183.960€/año	4 agentes y 1 encargado de sistemas (28.000€) más las cuotas de la Seguridad Social
Coste por constitución de una sociedad	1.500€	
Coste por primer establecimiento	21.000 + 3.500 + 2.000 + 4.000 = 30.500€	Incluye gastos de personal (3 agentes), pago por servicios y material, alquiler y publicidad
Coste anual total	183.960€ + otros gastos	Las dietas irán en función del desplazamiento de los agentes y en otros gastos se recoge, por ejemplo, la compra de nuevos sensores
Inversión	25.000 + 10.000 + 1.500 + 30.500 = 67.000€	

Aunque a raíz de esta tabla es posible realizar una estimación de los gastos en los que incurre el proyecto, es necesario tener en cuenta también las fuentes de ingreso que se prevé conseguir mediante la comercialización del producto. Partiendo de la base del lienzo desarrollado, y dejando a un lado las subvenciones y ayudas a las que se puedan optar, se separan los casos que se recogen en el apartado de las fuentes de ingresos.

En cuanto a la implantación en las explotaciones ganaderas, se ha pensado en un sistema que permita revisar cada mes un número determinado de cabezas de ganado, manteniendo siempre los mismos animales como objetivos del estudio o rotando el ganado monitorizado cada mes.

De esta forma es posible alargar los contratos en el tiempo y supone para las explotaciones un control continuo sobre los animales. El precio fijado para este servicio es de 5€ por animal al mes, pudiendo añadir o quitar animales de un mes a otro, aumentando o disminuyendo de esta forma los ingresos mensuales.

A raíz de esta estrategia será necesario estudiar en que comunidades autónomas se puede tener acceso a una mayor cantidad de animales, intentando minimizar los desplazamientos entre explotaciones, o lo que es lo mismo, donde haya un mayor número de reses por explotación.

Tabla 15. Ganado bovino por explotación en cada CCAA en 2023. Fuente: elaboración propia a partir de los datos del MAPA.

CCAA	N.º expl.	N.º vacas	N.º medio de reses por expl.
Andalucía	9917	542235	54,7
Aragón	4380	401384	91,6
Asturias	15319	378715	24,7
Baleares	955	27148	28,4
Canarias	1015	20444	20,1
Cantabria	10854	267904	24,7
C-M	4380	498146	113,7
CyL	22974	1456164	63,4
Cataluña	6636	672081	101,3
Extremadura	15736	938635	59,6
Galicia	40610	949484	23,4
Madrid	2081	115288	55,4
Murcia	448	95.540	213,3
Navarra	1941	122.006	62,9
País Vasco	5775	133.568	23,1
La Rioja	439	40.059	91,3
Valencia	1072	56.519	52,7
España	144532	6715320	46,5

En la Tabla 15 se pueden apreciar diferentes datos de interés como, por ejemplo, el hecho de que Galicia es la comunidad autónoma con mayor número de explotaciones de ganado bovino, pero solo cuenta con una media de 24,4 reses por explotación, lo que es una muestra inequívoca del uso de una ganadería más extensiva.

A la hora de elegir las comunidades autónomas sobre las que centrar la fuerza de ventas, será necesario atender no solo a la media de vacas por explotación, sino también, al número de vacas

en el territorio, ya que Murcia presenta la mejor media, pero el número de ganado sobre el que realizar el despliegue es muy reducido.

A la vista de estos datos y atendiendo también al territorio que compone cada comunidad para poder estimar los desplazamientos necesarios entre explotaciones, las comunidades más factibles a la hora de buscar potenciales clientes son: Cataluña, Castilla La Mancha, Aragón y Castilla y León. Hay que tener en cuenta que es sumamente difícil para una empresa que está iniciando su actividad abarcar la totalidad del territorio nacional por lo que, siguiendo el modelo *Lean Startup*, se tendrán en cuenta los mencionados territorios a la hora de plantear una primera hipótesis sobre el desarrollo del negocio.

Si fuera posible obtener contratos mensuales sobre el 0,1% del total de reses de todos estos territorios, esto supondría unos ingresos aproximados de 180000€ durante el primer año. Además, para superar la barrera de los gastos de personal (183.960€) sería necesario obtener contratos mensuales sobre 3.066 animales, menos de un 0,05% del total del ganado bovino español.

En cuanto al uso del producto a través de un pack que incluya tanto el sensor inercial como el acceso al software desarrollado, se ha optado por un sistema de licencias anuales por valor de 100€. Estas licencias se bloquean hasta que se realice el pago correspondiente, buscando de esta manera la renovación de los contratos por parte de los centros de investigación y enseñanza.

En la Tabla 16 que se encuentra en los anexos, se recogen las diferentes cuantías que afectan al proyecto, pudiendo derivar de estas las diferentes métricas que se han comentado anteriormente. Cabe destacar que el préstamo a solicitar tiene un tipo de interés del 7% y un plazo de 5 años. A continuación, se reflejan las fórmulas a través de las que se han obtenido el plazo de recuperación, el valor actual neto y la tasa interna de retorno.

$$payback = 5 + \left(\frac{12.852}{20.773} \right) = 5 + 0,61869 = 5,62 \text{ años}$$

$$VAN = -7.000 + \frac{-9.537}{1,05} + \frac{-8.847}{(1,05)^2} + \frac{-1.797}{(1,05)^3} + \frac{5.443}{(1,05)^4} + \frac{8.883}{(1,05)^5} + \dots = -7.000 + 146.476 = 139.476$$

$$VAN = -7.000 + \frac{-9.537}{1 + TIR} \dots = 0 \rightarrow TIR = 0,3875 = 38,75\%$$

El plazo de recuperación del proyecto es de 5,62 años, lo que quiere decir que, a partir del sexto año, el flujo de caja acumulado pasará a ser positiva y se recuperará el dinero invertido. En cuanto al VAN, su valor es mayor que 0 por lo que el proyecto es rentable ya que, actualizadas al momento inicial, las entradas de dinero son mayores que las salidas. Por su parte, la TIR toma un valor mayor al tipo de interés establecido, por lo que ofrece una rentabilidad positiva, como ya mostraba también el VAN.

Como ya se ha comentado esta sería una primera aproximación dentro del modelo *Lean Startup*, pero se debe tener en cuenta que sufrirá cambios a lo largo del tiempo derivado de situaciones como: captación de clientes deficiente, contratación de más personal para poder abarcar más territorio, obsolescencia o malos resultados del software, etc.

9. Conclusiones y trabajos futuros

En primer lugar, es necesario remarcar la importancia de que en este trabajo se desarrolla una prueba de concepto sobre la que se realizan hipótesis, con el fin de sentar las bases al desarrollo de un proyecto completamente funcional.

En este mismo sentido, cabe mencionar una vez más el hecho de que, dentro del desarrollo de negocio, se parte de esta propuesta como primer paso dentro del método *Lean Startup* propuesto. Este método se irá desarrollando a medida que se prueben y ajusten diferentes elementos a raíz de lo plasmado en este documento.

Aunque se hayan obtenido unos resultados cuantificados a través de los diferentes modelos, este proyecto no abarca únicamente la parte informática, sino que esta es una base importantísima sobre la que desarrollar una propuesta empresarial.

Bien es cierto que, ante la situación de incertidumbre presente en la economía mundial, aunque esta propuesta quede bien acotada y sea posible realizar predicciones sobre su desarrollo, no será del todo fiable hasta que no se lleven a cabo pruebas de mercado que puedan aportar datos reales que permitan obtener conclusiones realmente significativas.

En lo referente al Grado en Administración y Dirección de Empresas, los contenidos presentados se ajustan y demuestran los conocimientos adquiridos y muestran una imagen fiable del posible desarrollo a seguir para el producto planteado.

A futuro sería posible estudiar otros modelos o incluso, la implantación del producto en otros mercados o ámbitos de la sociedad (ya sea en otros animales o incluso en personas), analizando su impacto económico con las nuevas variables que se presenten en cada caso.

Por otra parte, sería interesante desarrollar nuevas funcionalidades que confieran a la tecnología desarrollada una mayor autonomía a la hora de ser utilizada. Herramientas como un portal web para la recepción de datos y visualización de resultados pueden ser el siguiente paso a realizar para acompañar el trabajo realizado hasta ahora.

Es cierto que se han presentado algunos problemas a la hora de utilizar los datos proporcionados ya que, de las 21 vacas monitorizadas, solo 6 de ellas se clasificaban como “vacas enfermas” (grado de afección mayor a dos en los estudios veterinarios realizados).

Este hecho derivaba, como ya se ha comentado, en que las diferentes clases no estuviesen balanceadas, lo que suponía un tratamiento extra a la hora de desarrollar el modelo de clasificación individual que puede llevar a una deficiencia a la hora de obtener los resultados.

Con respecto a esto, y con vistas a la posibilidad de expansión del producto a mercados internacionales, sería necesario realizar una captura de datos sobre vacas que presenten este tipo de disfunciones para poder contar con una base de datos de garantías a la hora de entrenar los diferentes modelos que se utilicen.

En este sentido, se debe tener en cuenta que los modelos desarrollados en este trabajo son únicamente la punta del iceberg, ya que los campos tratados dentro de la informática ofrecen múltiples opciones para la solución del problema planteado, opciones que pueden ser estudiadas en investigaciones posteriores.

A la luz de los resultados obtenidos es necesario hacer constar la importancia de la correcta predicción por parte del modelo escogido. Será necesario encontrar las características que mejor definan las diferencias existentes entre cada una de las clases para, de esta forma, asegurar que las predicciones realizadas se ajusten de forma correcta a la realidad observada.

Aun así, se ha podido constatar como la adaptación de una solución propuesta para un problema similar (en este caso el planteado en Zebin et al. (2017)), puede ser efectiva o, al menos, ser un punto de partida desde el que desarrollar una solución definitiva.

10. Fuentes de información

Álvarez, J. A. M., Martos, D. G., Burgos, A. M., & Arranz, S. (2013). Características del sector agrario español en el marco de la Unión Europea. Papeles de trabajo del Instituto de Estudios Fiscales. Serie economía, 13, 3-39. <http://repositori.uji.es/xmlui/handle/10234/172799>

Arisholm, E., Briand, L. C., & Johannessen, E. (2010). A systematic and comprehensive investigation of methods to build and evaluate fault prediction models. *Journal of Systems and Software*, 83(1), 2-17. <https://doi.org/10.1016/j.jss.2009.06.055>

Asale, R.-. (s.f.). comercio | Diccionario de la lengua española. «Diccionario De La Lengua Española» - Edición Del Tricentenario. <https://dle.rae.es/comercio?m=form2#Go3q3co>

Ávila-Foucat, V. S. (2017). Desafíos del sector primario y políticas públicas sustentables. *Economía Informa*, 402, 29-39. <https://doi.org/10.1016/j.ecin.2017.01.003>

Bell, D. (1976). El advenimiento de la sociedad post-industrial: un intento de prognosis social.

Berrar, D. (2016). Bayes' Theorem and Naive Bayes Classifier. En Elsevier eBooks (pp. 403-412). <https://doi.org/10.1016/b978-0-12-809633-8.20473-1>

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*. <https://doi.org/10.1093/aje/kwz189>

Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *Journal of Arthroplasty*, 33(8), 2358-2361. <https://doi.org/10.1016/j.arth.2018.02.067>

Bortolini, R. F., Cortimiglia, M. N., De Moura Ferreira Danilevicz, A., & Ghezzi, A. (2021). Lean Startup: a comprehensive historical review. *Management Decision*, 59(8), 1765-1783. <https://doi.org/10.1108/md-07-2017-0663>

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in Machine Learning algorithms. *Big Data & Society*, 3(1), 205395171562251. <https://doi.org/10.1177/2053951715622512>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step Data Mining Guide. SPSS. <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>

Cohen, I., Cozman, F. G., Sebe, N., Cirelo, M. C., & Huang, T. S. (2004). Semisupervised learning of classifiers: theory, algorithms, and their application to human-computer interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12), 1553-1566. <https://doi.org/10.1109/tpami.2004.127>

Cuadrado Roura, J. R. (2016). Desindustrialización “versus” terciarización: del aparente conflicto a una creciente integración. *Documentos de Trabajo (IAES, Instituto Universitario de Análisis Económico y Social)*, 150, 1-24. <https://dialnet.unirioja.es/servlet/articulo?codigo=5778022>

Data Never Sleeps 10.0 | Domo. (s.f.). <https://www.domo.com/data-never-sleeps>

Delgado, J. A., Ciudad, C. M., & Balboa, C. A. (1994). La agricultura española en el último tercio del siglo XX: principales pautas evolutivas. *Modernización y cambio estructural en la agricultura española*, 1994, ISBN 84-491-0071-2, págs. 69-126, 69-126.

Echeverry, L. L. G., Henao, A. M. J., Molina, M. A. R., Restrepo, S. M. V., Velásquez, C. D., & Bolívar, G. S. (2018). Human motion capture and analysis systems: a systematic review/Sistemas de captura y análisis de movimiento cinemático humano: una revisión sistemática. *Prospectiva*, 16(2), 24-34. <https://doi.org/10.15665/rp.v16i2.1587>

Escuela de Ingeniería Informática de la Universidad de Las Palmas de Gran Canaria [EII-ULPGC]. (2020). Normativa de desarrollo de las especificidades de los Trabajos de Fin de Título. En <https://www.eii.ulpgc.es/es/informacionacademica/tft/normativa>.

Facultad de Economía, Empresa y Turismo de la ULPGC [FEET-ULPGC]. (2023). Grado en Administración y Dirección de Empresas - Facultad de Economía, Empresa y Turismo de la ULPGC. Facultad De Economía, Empresa Y Turismo De La ULPGC. <https://feet.ulpgc.es/estudios-oficiales/grado-en-administracion-y-direccion-de-empresas/>

Facultad de Economía, Empresa y Turismo de la Universidad de Las Palmas de Gran Canaria [FEET-ULPGC]. (2020). Reglamento para la realización y evaluación de Trabajos de Fin de Título. En <https://feet.ulpgc.es/estudiantes/trabajo-fin-de-grado/>.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Ai Magazine*, 17(3), 37-54. <https://doi.org/10.1609/aimag.v17i3.1230>

Fayyad, U. M., & Stolorz, P. (1997). Data mining and KDD: Promise and challenges. *Future Generation Computer Systems*, 13(2-3), 99-115. [https://doi.org/10.1016/s0167-739x\(97\)00015-0](https://doi.org/10.1016/s0167-739x(97)00015-0)

Felin, T., Gambardella, A., Stern, S., & Zenger, T. (2020). Lean startup and the business model: Experimentation revisited. *Long Range Planning*, 53(4), 101889. <https://doi.org/10.1016/j.lrp.2019.06.002>

Foss, N. J., & Saebi, T. (2017). Business models and business model innovation: Between wicked and paradigmatic problems. *Long Range Planning*, 51(1), 9-21. <https://doi.org/10.1016/j.lrp.2017.07.006>

Frontela Sánchez, P. (2022). Soluciones a las dificultades en el relevo generacional y la creación de empleo joven en el sector primario. [Trabajo de Final de Título]. Universidad de Valladolid.

García Monsálvez, J. C. (2017). Python como primer lenguaje de programación textual en la Enseñanza Secundaria. *Education in the Knowledge Society (EKS)*, 18(2), 147-162. <https://doi.org/10.14201/eks2017182147162>

González-Díez, V., & Moral-Benito, E. (2019). El proceso de cambio estructural de la economía española desde una perspectiva histórica (The Process of Structural Change in the Spanish Economy). Social Science Research Network. <https://doi.org/10.2139/ssrn.3398149>

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, 64, 53-62. <https://doi.org/10.1016/j.patrec.2015.04.009>

Kotsiantis, S. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4), 261-283. <https://doi.org/10.1007/s10462-011-9272-4>

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2), 195-197. <https://doi.org/10.1038/nbt1386>

Lorenzo Díaz, D. J., Díaz Fariña, E., Díaz Hernández, J. J., & Esparza Ferrera, E. T. (2021). El PIB de Canarias y de sus economías insulares. Reflexiones sobre su cálculo y aproximación al impacto de la COVID-19. <https://doi.org/10.18002/jclcreg/v0i12>

Luque Cuesta, F. (2022). Producción industrial, animalismo y contracultura. El álbum ilustrado como reivindicación paleolítica de la cultura rural. *Arte y Políticas de Identidad*, 27, 68-92. <https://doi.org/10.6018/reapi.552651>

Maggioni, M. (2012). WHAT IS... Data Mining? *Notices of the American Mathematical Society*, 59(04), 1. <https://doi.org/10.1090/noti831>

Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach. *Information Systems*, 34(1), 87-107. <https://doi.org/10.1016/j.is.2008.04.003>

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. A. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048-3061. <https://doi.org/10.1109/tkde.2019.2962680>

Matplotlib documentation — Matplotlib 3.7.1 documentation. (s. f.). <https://matplotlib.org/stable/index.html>

Ministerio de Agricultura, Pesca y Alimentación. (2018). Informe Anual de Indicadores: Agricultura, Pesca y Alimentación. En *Indicadores de medio ambiente, agricultura, pesca, alimentación y desarrollo rural* (N.o 013-17-031-X). https://www.mapa.gob.es/es/ministerio/servicios/analisis-y-prospectiva/indicadores_semestre.aspx

Ministerio de Agricultura, Pesca y Alimentación. (2022). Informe Anual de Indicadores: Agricultura, Pesca y Alimentación. En *Indicadores de medio ambiente, agricultura, pesca, alimentación y desarrollo rural* (N.o 003200916). https://www.mapa.gob.es/es/ministerio/servicios/analisis-y-prospectiva/indicadores_semestre.aspx

NumPy documentation — NumPy v1.24 Manual. (s. f.). <https://numpy.org/doc/stable/>

os — Interfaces misceláneas del sistema operativo — documentación de Python - 3.10.11. (s. f.). <https://docs.python.org/es/3.10/library/os.html>

Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley & Sons.

pandas documentation — pandas 2.0.2 documentation. (s. f.). <https://pandas.pydata.org/docs/>

Pueo, B., & Jimenez-Olmedo, J. M. (2017). Application of motion capture technology for sport performance analysis (El uso de la tecnología de captura de movimiento para el análisis del rendimiento deportivo). *Retos: Nuevas Tendencias en Educación Física, Deportes y Recreación*, 32, 241-247. <https://doi.org/10.47197/retos.v0i32.56072>

re — Operaciones con expresiones regulares. (s. f.). Python documentation. <https://docs.python.org/es/3/library/re.html>

Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Currency.

Roa, M. M. (2021). El Big Bang del Big Data. Statista Infografías. <https://es.statista.com/grafico/26031/volumen-estimado-de-datos-digitales-creados-o-replicados-en-todo-el-mundo/>

Roberts, M. F., Mongeon, D., & Prince, F. (2017). Biomechanical parameters for gait analysis: a systematic review of healthy human gait. *Physical therapy and rehabilitation*, 4(1), 6. <https://doi.org/10.7243/2055-2386-4-6>

Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International journal of advanced research in artificial intelligence*, 2(2). <https://doi.org/10.14569/ijarai.2013.020206>

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534. <https://doi.org/10.1016/j.procs.2021.01.199>

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681. <https://doi.org/10.1109/78.650093>

scikit-learn: Machine Learning in Python — scikit-learn 1.2.2 documentation. (s. f.). <https://scikit-learn.org/stable/>

SciPy. (s. f.). <https://scipy.org/>

(seaborn: statistical data visualization — seaborn 0.12.2 documentation, s. f.)

Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International journal of innovation and scientific research*, 12(1), 217-222. <http://www.issr-journals.org/links/papers.php?journal=ijisr&application=pdf&article=IJISR-14-281-04>

Souza Silva, D. A., Ghezzi, A., De Aguiar, R. A., Cortimiglia, M. N., & Caten, C. S. T. (2020). Lean Startup, Agile Methodologies and Customer Development for business model innovation. *International Journal of Entrepreneurial Behaviour & Research*, 26(4), 595-628. <https://doi.org/10.1108/ijebr-07-2019-0425>

Statista. (2023). Porcentaje de usuarios de e-commerce en España 2010-2021. <https://es.statista.com/estadisticas/647143/porcentaje-de-compradores-online-sobre-el-total-de-la-poblacion-espana/>

Statsmodels 0.14.0. (s.f.). <https://www.statsmodels.org/stable/index.html>

Teece, D. J. (2010). Business Models, Business Strategy and Innovation. *Long Range Planning*, 43(2-3), 172-194. <https://doi.org/10.1016/j.lrp.2009.07.003>

TensorFlow. (s.f.). TensorFlow. <https://www.tensorflow.org/?hl=es-419>

TIOBE Index - TIOBE. (2022). TIOBE. <https://www.tiobe.com/tiobe-index/>

Tostado Sánchez, S. E., Ornelas Rodríguez, M., Espinal Jiménez, A., & Puga Soberanes, H. J. (2016). Implementación de Algoritmos de Inteligencia Artificial para el Entrenamiento de Redes Neuronales de Segunda Generación. *JÓVENES EN LA CIENCIA*, 2(1), 6–10. <https://www.jovenesenlaciencia.ugto.mx/index.php/jovenesenlaciencia/article/view/715>

Understanding Reference Frames and Device Attitude | Apple Developer Documentation. (s. f.).

Apple Developer Documentation.

https://developer.apple.com/documentation/coremotion/getting_processed_device-motion_data/understanding_reference_frames_and_device_attitude

Universidad de Las Palmas de Gran Canaria [ULPGC]. (s. f.). Objetivos y Competencias del GII. https://www2.ulpgc.es/archivos/plan_estudios/4008_40/ObjetivosyCompetenciasdelGII.pdf

Universidad de Las Palmas de Gran Canaria [ULPGC]. (2019). Reglamento General de los Trabajos de Fin de Grado y de Fin de Máster de las Titulaciones Oficiales de la Universidad de Las Palmas de Gran Canaria. En <https://www.ulpgc.es/reglamentos/reglamentos-tema/61>.

Vigne, J. (2011). The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. *Comptes Rendus Biologies*, 334(3), 171-181. <https://doi.org/10.1016/j.crv.2010.12.009>

Wilson, J. J., Alexandre, N., Trentin, C., & Tripodi, M. (2018). Three-Dimensional Representation of Motor Space in the Mouse Superior Colliculus. *Current Biology*, 28(11), 1744-1755.e12. <https://doi.org/10.1016/j.cub.2018.04.021>

Yang, X., Sun, S. L., & Zhao, X. (2019). Search and execution: examining the entrepreneurial cognitions behind the lean startup model. *Small Business Economics*, 52(3), 667-679. <https://doi.org/10.1007/s11187-017-9978-z>

Zebin, T., Scully, P., & Ozanyan, K. B. (2017). Inertial Sensor Based Modelling of Human Activity Classes: Feature Extraction and Multi-sensor Data Fusion Using Machine Learning Algorithms. En *Lecture Notes in Computer Science* (pp. 306-314). Springer Science+Business Media. https://doi.org/10.1007/978-3-319-49655-9_38

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine Learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350-361. <https://doi.org/10.1016/j.neucom.2017.01.026>

11. Anexos

Tabla 16. Desarrollo de los elementos necesarios para el cálculo de *payback*, VAN y TIR. Fuente: elaboración propia.

Año	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033
Ingresos		181.980	186.420	192.720	200.460	207.900	218.040	224.580	231.720	243.360	252.900
Gastos		183.960	187.710	186.960	187.460	191.460	189.710	190.710	191.460	192.460	193.710
Beneficios		-1.980	-1.290	5.760	13.000	16.440	28.330	33.870	40.260	50.900	59.190
Amortización		6.700	6.700	6.700	6.700	6.700	6.700	6.700	6.700	6.700	6.700
Inversión	67.000										
Préstamos	60.000										
Devolución prestamos		-14.256,86	-14.256,86	-14.256,86	-14.256,86	-14.256,86					
Flujo neto de caja	-7.000	-9.537	-8.847	-1.797	5.443	8.883	20.773	40.570	46.960	57.600	65.890
FNC acumulado	-7.000	-16.537	-25.384	-27.180	-21.736	-12.852	7.921	48.491	95.451	153.051	218.941