



Universidad Austral de Chile

Facultad de Ciencias de la Ingeniería
Escuela de Ingeniería Civil en Informática

DETECCIÓN DE POSICIÓN JERÁRQUICA (LIDERAZGO) EN OVINOS A TRAVÉS DE ANÁLISIS DE VIDEOS Y MACHINE LEARNING

Proyecto para optar al título de
Ingeniera Civil en Informática

PROFESOR PATROCINANTE:
CLAUDIA AIDA LETELIER VELASQUEZ
MÉDICO VETERINARIO
DOCTORA EN CIENCIAS VETERINARIAS

PROFESOR CO-PATROCINANTE
CHRISTIAN LAZO RAMÍREZ
INGENIERO EN GESTIÓN INFORMÁTICA
DOCTOR EN INGENIERÍA TELEMÁTICA

PROFESOR INFORMANTE
PABLO ANDRÉS HUIJSE HEISE
DOCTOR EN INGENIERÍA ELÉCTRICA

ROXANA EVELYN CAYUL NAVARRO

VALDIVIA – CHILE
2020

ÍNDICE

ÍNDICE DE TABLAS	IV
ÍNDICE DE FIGURAS	V
RESUMEN.....	VII
ABSTRACT	VIII
1. INTRODUCCIÓN	1
1.1. Descripción de la oportunidad	2
1.2. Motivación.....	2
1.3. Definición de Objetivos.....	3
1.3.1 Hipótesis.....	3
1.3.2 Objetivo general	3
1.3.3 Objetivos específicos	3
2 ESTADO DEL ARTE	4
2.1 Inteligencia Artificial.....	4
2.2 Jerarquía en ovinos domésticos	5
2.3 Aplicaciones multidisciplinar de Informática y Veterinaria.....	6
3. MARCO TEÓRICO	8
3.1. Machine learning	8
3.2. Aprendizaje no supervisado.....	9
3.2.1 Ventajas	9
3.2.2 Desventajas	9
3.3. Reducción de dimensionalidad.....	9
3.3.1. Análisis de componentes principales (PCA).....	10
3.3.1 Visualización	12
3.4. Técnicas de agrupamiento	13
3.4.1. K-means	14
3.4.2. Agrupamiento jerárquico	16
3.4.3. GMM	18
4. MATERIALES Y MÉTODOS	21
4.1. Estudio de campo	21
4.1.1. Selección y medición de las características fenotípicas de los individuos involucrados en el estudio.....	29
4.1.2 Recolección de datos	34
4.2. Trabajo de gabinete	42

4.2.1. Dataset	46
4.2.2. Ambiente de desarrollo	48
4.2.3. Análisis estadístico	48
4.2.4. Reducción de dimensionalidad.....	51
4.3. Implementación de modelos de agrupamiento	52
 4.3.1. Implementación y programación del modelo K-means.....	52
 4.3.2. Implementación y programación del modelo de Agrupamiento jerárquico.	54
 4.3.3. Implementación y programación del modelo GMM.....	55
5. RESULTADOS Y DISCUSIÓN.....	57
 5.1. Análisis y comparativa de modelos de agrupamiento	57
 5.2. Validación y evaluación del modelo en condiciones de pastoreo y predación	61
 5.2.1. Pastoreo	62
 5.1.2. Predación	63
6. CONCLUSIONES	66
 6.1. Trabajos futuros	66
7. REFERENCIAS	68

ÍNDICE DE TABLAS

Tabla 1: Distribución de individuos según número de dientes definitivos (edad).....	22
Tabla 2: Equivalencia de edad según cantidad de dientes definitivos.....	30
Tabla 3: Ecuaciones de calibración obtenidas con el plato para cada estación climática en las praderas permanentes del sur de Chile.....	44
Tabla 4: Descripción y tipos de datos de los atributos que componen el dataset.....	47
Tabla 5: Estadísticos de las características fenotípicas de los individuos del rebaño	48
Tabla 6: Porcentaje de varianza según número de componentes principales al aplicar PCA.....	52
Tabla 7: Distribución de individuos por grupo.....	53
Tabla 8: Distribución de individuos por grupo en Agrupamiento jerárquico	55
Tabla 9: Distribución de individuos por grupo en GMM.....	56
Tabla 10: Individuos seleccionados por cada grupo del modelo K- means	61

ÍNDICE DE FIGURAS

Figura 1: Crecimiento de las publicaciones sobre inteligencia artificial en China, Europa y USA.	5
Figura 2: Muestra de imágenes capturadas por el drone	6
Figura 3: Multiplicación de una matriz por un vector.	10
Figura 4: Conjunto de datos de dos dimensiones X_1 y X_2	11
Figura 5: Conjunto de datos y sus dos componentes principales.....	12
Figura 6: Reducción de dimensionalidad dataset iris.	13
Figura 7: Gráficas comparativas entre diferentes técnicas de agrupamiento.....	14
Figura 8: Método del codo para encontrar el valor de K óptimo.....	15
Figura 9: Fórmula de asignación de datos a su respectivo centroide.	15
Figura 10: Dendrograma Agrupamiento jerárquico aglomerativo.....	17
Figura 11: Dendrograma Agrupamiento jerárquico divisivo.	18
Figura 12: Modelo de mezclas de Gaussianas con 3 componentes.....	19
Figura 13: Gráfica del resultado al aplicar GMM y el algoritmo de EM.	20
Figura 14: Localización geográfica de la estación experimental agropecuaria UACH.	22
Figura 15: Esquema del potrero y sus sectores.....	23
Figura 16: Esquema de la locación de estudio	24
Figura 17: Vista desde abajo de la cámara 1	24
Figura 18: Manga de manejo para ovinos.....	25
Figura 19: Pruebas de vuelo con drone	25
Figura 20: Imagen tomada por el drone a una altura de 10 metros.	26
Figura 21: Primer mecanismo de identificación.	27
Figura 22: Segundo mecanismo de identificación.	27
Figura 23: Visualización a través de las cámaras del segundo método de identificación.....	28
Figura 24: Ovejas captadas por cámara estática, individualizadas con el tercer método de identificación.	29
Figura 25: Ovino con un par de dientes definitivos.	30
Figura 26: método utilizado de medir la condición corporal de ovinos.....	31
Figura 27: Escala de medición de la condición corporal de ovinos.	32
Figura 28: Esquema de mediciones de la altura de la cruz y la grupa.....	33
Figura 29: Esquema de medición de la distancia entre los ojos.	33
Figura 30: Esquema de medición de la distancia hocico-ojos.....	34
Figura 31: Ovejas en la manga de manejo.....	35
Figura 32: Toma de características fenotípicas	35
Figura 33: Estación de monitoreo	36
Figura 34: Cámara 1: Pasillo	37
Figura 35: Cámara 2: detrás del corral.....	37
Figura 36: Cámara 3: comederos.	38
Figura 37:Plato medidor de forraje	38
Figura 38: Esquema de medición sector 1	39
Figura 39: Drone Dji Mavic Air 2.....	40
Figura 40: Aplicación DJI GO 4	40
Figura 41:Captura fotográfica desde el drone a 10 metros de altura.	41
Figura 42:Diagrama del recorrido del drone por el potrero.....	41
Figura 43:Línea referencial para obtener los lugares de los individuos en dirección a pastar.....	42

Figura 44: Línea referencial para obtener los lugares de los individuos en dirección al corral.....	43
Figura 45: Kilogramos de materia seca por hectárea para la primera medición.....	44
Figura 46: Kilogramos de materia seca por hectárea para la segunda medición.....	45
Figura 47: Kilogramos de materia seca por hectárea para la tercera medición.....	45
Figura 48: Gráfica de las mediciones por potrero.....	46
Figura 49: Gráfica de distribución de la edad según el número de dientes definitivos.....	49
Figura 50: Gráfica de distribución de la condición corporal al comenzar el estudio.....	50
Figura 51: Gráfica de distribución de la condición corporal al finalizar el estudio.....	50
Figura 52: Matriz de correlación.....	51
Figura 53: Método del codo.....	53
Figura 54: Implementación del modelo K-means en python.....	53
Figura 55: Dendrograma Agrupamiento aglomerativo.....	54
Figura 56: Implementación del modelo de agrupamiento jerárquico aglomerativo en python	55
Figura 57: Implementación del modelo GMM en python.....	55
Figura 58: Agrupamiento K-means	58
Figura 59: Agrupamiento jerárquico aglomerativo	59
Figura 60: Agrupamiento GMM	60
Figura 61: Potrero por sectores.....	62
Figura 62: Mapa de calor de la disponibilidad de materia seca del potrero 3.....	62
Figura 63: Individuos 20 y 22 frente a la predación.....	63
Figura 64: Individuo 29 frente a la predación	64
Figura 65: Individuo 50 frente a la predación	64
Figura 66: Individuos 35 y 37 frente a la predación.....	65

RESUMEN

Los individuos de un rebaño presentan una tendencia natural a formar relaciones cooperativas interdependientes que permiten una comunicación recíproca que va más allá de la actividad sexual, constituyendo este tipo de agrupaciones lo que se conoce como rebaño. La forma en que se organiza un rebaño tiene aspectos positivos como una mejor protección frente al ataque de predadores y negativos como que se enfrenten entre ellos por alimentos, territorio y aspectos reproductivos, principalmente. Estudiar la jerarquía de los rebaños ovinos es fundamental para entender las relaciones de dominancia y subordinación que se presentan en el rebaño. Existe distintas formas de estudiar el comportamiento animal y las tecnologías de agricultura 4.0 han permitido grandes avances para comprender y desarrollar tecnologías asociadas a los manejos ganaderos. El objetivo de este estudio es identificar dominancia entre los individuos basado en atributos fenotípicos y comportamiento animal utilizando técnicas de machine learning (Clustering) para establecer niveles de jerarquía en un rebaño ovino doméstico.

Se utilizaron 50 ovejas de raza Austral sobre las cuales se determinó atributos fenotípicos (edad, condición corporal, peso vivo y medidas morfométricas) y el comportamiento animal durante dos semanas, realizando el seguimiento a través de cámaras estáticas y visión aérea (drone) durante el arreo (am y pm) y el pastoreo. Esto se realizó en el potrero de estudio dos veces al día por tres observaciones de 10 minutos cada una. Se utilizó estadística descriptiva y analítica para conocer el comportamiento de los datos e identificar relaciones a priori entre las mediciones, utilizando librerías de Python. Los datos fueron analizados utilizando tres algoritmos de clustering: K-means, Agrupamiento jerárquico aglomerativo y Modelo de mezclas gaussianas (GMM). La validación del modelo de jerarquías más ajustado a los datos fue probada en situación de pastoreo y arreo.

El modelo que logró el mejor desempeño fue K-means, demostrando que las características y comportamiento que afectan la dominancia en un rebaño fueron el peso vivo corporal, la condición corporal y el lugar que ocupaban los individuos durante el arreo hacia y desde el potrero de pastoreo. Sin embargo, el patrón de comportamiento observado durante el pastoreo no estableció relaciones de dominancia que permitieran observar una estructura jerárquica del rebaño durante esos momentos, a diferencia de lo que ocurre durante la simulación de un ataque por un depredador carnívoro (perro) donde fue posible determinar una fuerte estructura jerárquica que coincide con la establecida durante el arreo de los animales, siendo los mismos animales identificados por el modelo de agrupamiento que ocupaban la cúspide de la gráfica, confrontando y alertando al resto del rebaño, ubicándose en la periferia del rebaño en el lugar donde se producía el ataque del predador. Por otro lado, fue posible establecer otro tipo de relaciones jerárquicas durante el ataque, donde los animales del eslabón más bajo ocupaban el espacio perimetral opuesto durante el ataque. Además, los animales que se encontraban en el tramo medio del modelo de agrupamiento mantuvieron su espacio entre el ataque y los animales que defendían el rebaño. Con lo cual se puede concluir que la agrupación de algunas variables del estudio utilizando el algoritmo K-means permitieron establecer la estructura jerárquica de un rebaño por atributos fenotípicos y el comportamiento del arreo hacia y desde el potrero.

ABSTRACT

Individuals in a herd have a natural tendency to form interdependent cooperative relationships that allow reciprocal communication which goes beyond sexual activity, this type of group constituting what is known as a herd. The way in which the herd is organized has positive aspects such as better protection against attacks of predators and negative aspects such as competition with each other for food, territory and reproductive aspects, mainly. Studying the hierarchy of sheep herds is essential to understand the dominance and subordination relationships that occur in the herd. There are different ways of studying animal behavior, and agriculture 4.0 technologies have allowed great advances to understand and develop technologies associated with livestock management. The objective of this study is to identify dominance among individuals based on phenotypic attributes and behavior using machine learning techniques (Clustering) to establish levels of hierarchy in a domestic sheep herd.

Fifty Austral breed sheep were used, on which phenotypic attributes (age, body condition, live weight and morphometric measurements) and animal behavior were registered for two weeks, monitoring through static cameras and aerial vision (with the help of a drone) during the herding (am and pm) and grazing periods. This was conducted in the study paddock twice a day for three observations of 10 minutes each. Descriptive and analytical statistics were used to know the behavior of the data and identify a priori relationships between the measurements, using Python libraries. Data was analyzed using three clustering algorithms: K-means, Hierarchical Agglomerative Clustering, and Gaussian Mixture Model (GMM). The validation of the hierarchies model more adjusted to the data was tested in grazing and herding situations. The model that achieved the best performance was K-means, demonstrating that the characteristics and behavior that affect dominance in a herd were live body weight, body condition and the position occupied by individuals during herding to and from the paddock. However, the behavior pattern observed during grazing did not establish dominance relationships that would allow observing a hierarchical structure of the herd during those moments, unlike what occurs during the simulation of an attack by a carnivorous predator (dog) where it was possible to determine a strong hierarchical structure that coincides with that established during the herding of the animals, being the same animals identified by the grouping model that occupied the top of the graph, confronting and alerting the rest of the herd, being located on the periphery of the herd in the place where the predator's attack took place. On the other hand, it was possible to establish other types of hierarchical relationships during the attack, where the animals of the lowest link were standing at the opposite perimeter during the attack. In addition, the animals that were in the middle section of the grouping model maintained their space between the attack and the animals defending the herd. With which it can be concluded that the grouping of some variables of the study using the K-means algorithm allowed to establish the hierarchical structure of a herd by phenotypic attributes and the behavior of the herd to and from the paddock.

1. INTRODUCCIÓN

El ganado ovino se domesticó hace más de diez mil años, con el fin de ser utilizado por el hombre para satisfacer necesidades primarias (Sáenz, 2007). Su origen geográfico se localiza en Europa y en las regiones frías de Asia. En el caso de Chile, los primeros ovinos llegaron con la colonización española y pasaron a ser parte fundamental de los sistemas productivos y económicos, existiendo actualmente cerca de 4 millones de ovinos en el país (INE, 2007). Gran parte de esta población se localiza en la zona Sur Austral y es parte importante de la economía familiar campesina.

La mayoría de los animales y en especial los ovinos, se caracterizan por su comportamiento social, es decir, interactúan con otros individuos bajo pautas de dominancia y subordinación (Ungerfeld y Núñez, 2011). Por una parte, la dominancia es el atributo que brinda a un animal el acceso a recursos alimenticios, sexuales y territoriales, mientras que para el subordinado esto significa menor acceso a estos recursos. De esta manera, se puede indicar que la raza ovina está entre los animales productivos más sociales. Es una especie altamente gregaria con distancias sociales relativamente pequeñas, lo cual le garantiza mejores condiciones para criar y protección frente a ataques de predadores. Sin embargo, el requisito de contacto estrecho que existe en un rebaño trae consigo efectos negativos como el incremento de la competencia por recursos y parejas. además, de un aumento de las probabilidades de ser partícipe de enfrentamientos con otro miembro del rebaño (Espejo, 2015). Desde esta perspectiva, se hace necesario entender el contexto social y las relaciones existentes dentro de un grupo de ovinos domésticos para tener un mejor manejo productivo.

En los últimos años, algunos etólogos (Lynch et al 1992) han puesto gran parte de su atención en el estudio del comportamiento que tienen los rebaños de ovejas en diferentes tipos de situaciones, tales como pastoreo, procedimientos de arreo e incluso predación. Sin embargo, y pese a todas las investigaciones relacionadas con estos animales, no se ha logrado determinar con claridad la jerarquía que poseen los ovinos dentro de un rebaño doméstico, ni tampoco las características que son claves para determinarla y mucho menos cuál es el nivel de importancia que tiene cada una de estas características al momento de utilizar esta información. Con el fin de aprovechar la estructura social en favor de los sistema productivo se hace necesario el estudio del comportamiento para mejorar los sistemas productivos tal como se realiza en otras especies de interés zootécnico. Considerando estos aspectos se hace esencial el estudio de estas características, siendo el uso de tecnologías tales como cámaras de captura de fotos y videos, sensores de movimiento y dispositivos de posicionamiento espacial (GPS), los cuales permiten obtener información relevante para realizar estudios de comportamiento animal, ya que brindan una mirada objetiva y completa del entorno que comparten los animales.

El objetivo principal de este estudio es establecer relaciones entre características fenotípicas de un rebaño ovino estándar para la región y otros comportamientos que serán medidos a través de una metodología de toma y procesamiento de datos y la implementación de técnicas de machine learning, específicamente, métodos de clusterización para el estudio de los datos obtenidos de mediciones morfológicas y análisis de videos con la finalidad de encontrar grupos jerárquicos y establecer patrones de comportamiento asociados a esta conducta.

1.1. Descripción de la oportunidad

El establecimiento de vida en grupo y sistemas sociales ha estado presente en diversas especies animales como un mecanismo de supervivencia. Los ovinos domésticos no quedan exentos de esto, ya que son animales sociables y dóciles que dependen de otras ovejas para sobrevivir. Si bien, este comportamiento trae consigo beneficios como la posibilidad de encontrar pareja, reducir los riesgos de depredación y conseguir mejores fuentes de alimento; También existen riesgos de enfrentamientos por la dominancia y la permanencia en el grupo.

En la actualidad, no están claras cuáles son las características ni cómo influyen en la posición jerárquica de un individuo en un rebaño doméstico. Hasta ahora, la metodología utilizada ha sido a través de observaciones y mediciones a criterio del investigador, como por ejemplo: tamaño de ornamentos, masa corporal, Frame Score (relación numérica entre la altura y la edad de un animal), edad, distancia anogenital, entre otras.

Entender y determinar la jerarquía de los rebaños de ovinos ayudaría a reducir los problemas de enfrentamientos entre miembros del grupo, prevención de enfermedades por efectos nutricionales y disminución de la depredación. Debido a que las ovejas siguen a un líder, sólo bastaría con reconocer ese rol y monitorearlo para comprender el actuar del rebaño.

Debido a todas estas características y factores, se hace necesario el estudio de los comportamientos que sustentan los lazos sociales que mantienen las ovejas.

Dado este contexto y la enorme cantidad de datos que se podrían extraer de las características y comportamiento de estos animales, parece natural pensar que existan una gran cantidad de estudios que relacionan inteligencia artificial con ovinos, sin embargo, la mayoría de los trabajos se enfocan en relacionar la industria de los bovinos y porcinos con esta disciplina. Esto se debe principalmente a que se invierte mucho más en estas dos industrias porque tienen una mayor demanda de productos, como la carne y la leche junto con sus derivados. Si bien, la demanda de la industria ovina está muy por detrás de las mencionadas anteriormente, esto no impide que el estudio de su comportamiento desde la mirada de otra disciplina sea importante, interesante y una importante oportunidad para generar conocimiento en un área escasamente explorada. Dado que las problemáticas asociadas a la crianza de ganado ovino son muchas, es relevante tomar medidas preventivas que logren disminuir los factores de riesgo a los cuales se ven sometidos los ovinos. Por ende, es importante conocer la jerarquía que forman estos grupos de animales y las características que influyen en que algunos estén en la cúspide de la pirámide de liderazgo y cuánto es su grado de exposición.

1.2. Motivación

Una de las principales motivaciones para realizar este estudio es aportar en el conocimiento del comportamiento animal de animales domésticos incorporando tecnologías de agricultura 4.0 que utilizando dispositivos para la toma de datos e inteligencia artificial permiten detectar patrones de comportamiento de interés zootécnico. Si bien lo que se conoce sobre jerarquía está enfocado principalmente en especies silvestres (panales de abejas, colmenas de hormigas,

manadas de leones, etc), en el caso de los ovinos domésticos existe poca información que haya permitido transferir este conocimiento a los sistemas de manejo productivo, tal y como se ha desarrollado el uso de estas tecnologías en bovinos de leche. Por ello, es interesante generar conocimiento en esta rama desde los aportes que se pueden realizar desde otras áreas profesionales lo que incentiva la multidisciplinariedad en la resolución de problemas, para generar un impacto en la industria ganadera moderna y sostenible.

1.3. Definición de Objetivos

1.3.1 Hipótesis

- Los modelos de agrupamiento basados en inteligencia artificial permiten encontrar niveles de jerarquía en base a características fenotípicas y observaciones de comportamiento de los ovinos domésticos que inciden en la dominancia que tiene el individuo sobre el rebaño.

1.3.2 Objetivo general

Utilizar técnicas de machine learning (Clustering) para identificar niveles de dominancia basado en características, comportamiento y sus influencias para hallar la jerarquía en rebaños de ovinos domésticos.

1.3.3 Objetivos específicos

1. Generar el conjunto de datos para armar el dataset del comportamiento jerárquico de los ovinos en condiciones de arreo y pastoreo.
2. Estudiar métodos de agrupamiento basados en machine learning que permitan cuantificar y asignar valoración al dataset para determinar liderazgo en las diferentes situaciones de estudio.
3. Implementar los métodos estudiados en el punto dos y ajustar los modelos a los datos del punto uno.
4. Evaluar el desempeño del modelo por medio de observaciones en dos situaciones de jerarquía: acceso a la fuente de alimento y defensa frente a un ataque de predadores.

2 ESTADO DEL ARTE

En esta sección, se presentan los avances relacionados a la inteligencia artificial, el estudio de comportamiento animal, la jerarquía en rebaños de ovinos domésticos y finalmente las aplicaciones que existen del trabajo multidisciplinario entre la informática y la veterinaria.

2.1 Inteligencia Artificial

La inteligencia artificial o comúnmente conocida por sus siglas en español como IA, es probablemente una de las ramas de la informática que más crecimiento ha tenido en la actualidad.

Hasta hace pocos años, existía una limitada capacidad de cómputo que provocaba que la inteligencia artificial proporcionara resultados muy acotados a problemas en los que se aplicaba, lo que inspiraba una falta de interés por parte de investigadores y descontento por parte de la industria al no contar con resultados en tiempo real. Sin embargo, en la actualidad esto ha cambiado producto del avance tecnológico y de la capacidad de cómputo de las máquinas que hoy se pueden encontrar en el mercado, lo que genera que los resultados sean mejores y que diversos modelos sean implementados a pequeña y gran escala en distintas áreas de investigación.

Dada la creciente demanda que ha tenido la inteligencia artificial alrededor del mundo, es que países desarrollados como Alemania, Estados Unidos y Japón han generado sus propias políticas públicas y códigos de ética que buscan regular y promover el buen uso de la inteligencia artificial y la robótica en sus países (Torres, 2019) promoviendo la inversión en investigación y desarrollo de estas tecnologías. Siguiendo el ejemplo de estos países se encuentra Chile, ya que, ha decidido promover políticas mediante la construcción de la Política Nacional de Inteligencia Artificial, documento que contendrá los lineamientos estratégicos que debe seguir el país en esta materia durante los próximos 10 años. El plan de trabajo para la elaboración de esta política está siendo ejecutada por un comité de expertos con el apoyo de tres equipos interministeriales. En la actualidad, el proceso se encuentra en la etapa de consulta ciudadana que busca abrir el debate sobre el primer proceso de redacción que incluyó discusiones a lo largo de todo Chile a través de mesas autoconvocadas (Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2020)

En la Figura 1 se puede observar la curva de crecimiento de los diferentes artículos científicos relacionados con inteligencia artificial en China, Europa y Estados Unidos.

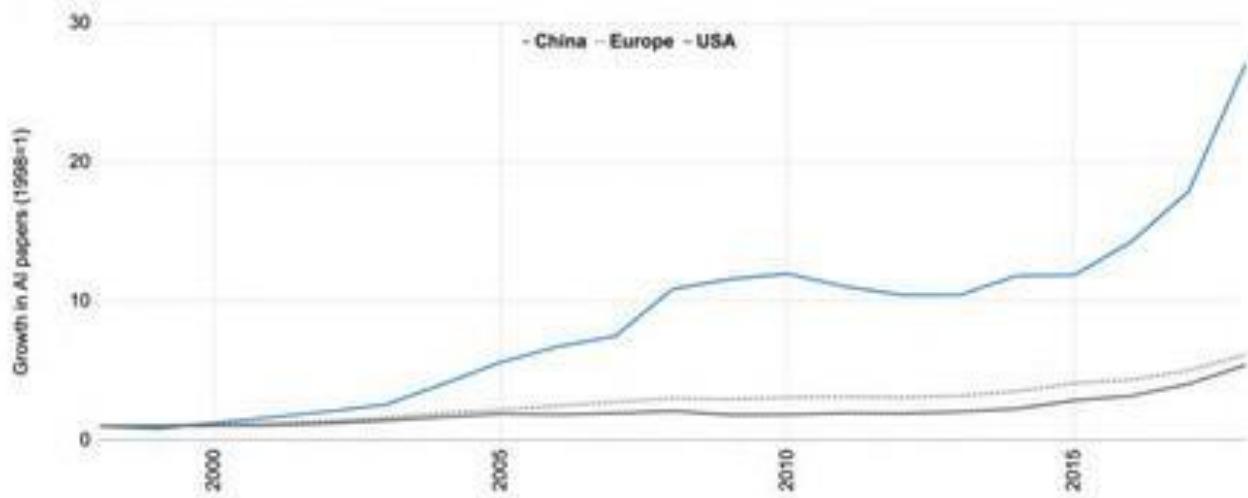


Figura 1: Crecimiento de las publicaciones sobre inteligencia artificial en China, Europa y USA. [FUENTE: Artificial Intelligence Index Report 2019]

La gráfica nos muestra que China lidera las publicaciones sobre inteligencia artificial. Sin embargo, según *Artificial Intelligence Index Report 2019*, Estados Unidos lidera en cuanto a relevancia, lo que se mide en base al impacto de las citas que reciben sus trabajos, con un Field-Weighted Citation Impact un 50% superior al de China.

2.2 Jerarquía en ovinos domésticos

En las últimas décadas, ha habido un auge en los países desarrollados por el estudio de la etología de animales domésticos y silvestres en cautiverio, como una herramienta científica para un mayor entendimiento de los comportamientos de estos individuos y como herramienta para mejorar su manejo y sistema productivo. En el caso de los ovinos, esto no es una excepción. Existen diversos estudios que se han volcado a entender las relaciones de dominancia que establecen los individuos sobre otros del mismo rebaño. Por ejemplo, el estudio de Ungerfeld y Núñez (2011) establece que la estructura social jerárquica de los carneros está estrechamente relacionada con los resultados reproductivos de cada uno. Además, evidencia que el comportamiento de los individuos es influenciado tanto por genes como por el ambiente. Por otra parte, también establece que animales de un mismo grupo si o sí deben competir entre ellos por los recursos muchas veces limitados y que de esta forma los animales de mayor rango jerárquico lograrán desplazar a otros de menor rango. Según Galindo (2004), los individuos de menor rango en la pirámide jerárquica suelen comer menos, se echan menos tiempo, o permanecen más tiempo en los lugares menos preferidos. Dadas estas condiciones, suelen ser más susceptibles a problemas de salud derivados de la mala alimentación.

Otro hallazgo importante que se ha logrado establecer con respecto a la jerarquía en ovinos según Ungerfeld y González (2008) es que la jerarquía se establece desde edades tempranas, y que la misma puede afectar el desarrollo reproductivo de corderos jóvenes.

Otro estudio interesante, fue el realizado por Espejo (2015) el cual concluyó que las ovejas no presentan jerarquías lineales y que el orden jerárquico se modifica al cambiar las condiciones

de grupo. Además, establece que la formación de jerarquías en ovinos domésticos ocurre por la acción conjunta tanto de atributos a priori como del contexto social donde se desenvuelven los individuos.

2.3Aplicaciones multidisciplinar de Informática y Veterinaria

Uno de los conceptos más innovadores en cuanto a utilización de recursos tecnológicos para la recopilación y análisis de datos sobre el campo, es la Agricultura 4.0. Esta se basa en la incorporación de tecnología que permita la automatización y la eficiencia de procesos que antes dependían de mano de obra humana para ser llevados a cabo. Hoy en día, existen variadas implementaciones que se le pueden atribuir a la agricultura 4.0, entre ellas encontramos control de características de suelo y agua con drones, tractores y sistemas de optimización por medio de posicionamiento satelital, monitoreo de animales a través de dispositivos GPS, entre otras.

Un ejemplo, interesante sobre la implementación de tecnología en el monitoreo de animales es el desarrollado por Rivas, Chamoso, González y Corchado (2018), donde utilizan técnicas de inteligencia artificial, en específico, redes neuronales convolucionales para la detección de bovinos por medio de la utilización de imágenes obtenidas de drones para un monitoreo en tiempo real. En la Figura 2 se pueden observar algunas imágenes que conforman el conjunto de datos de entrenamiento utilizados por la red neuronal.



(a) Sample of the target dataset



(b) Sample of the background dataset

Figura 2: Muestra de imágenes capturadas por el drone. [FUENTE: Detection of Cattle Using Drones and Convolutional Neural Networks]

Los resultados obtenidos por esta aplicación tuvieron una alta tasa de éxito en la identificación de ganado a partir de imágenes captadas por un drone, lo que ayudó en el monitoreo e identificación de situaciones anómalas que requerían de intervención humana.

Otra aplicación interesante en el campo de la agricultura 4.0 fue la realizada por Owoeye y Hailes (2018) porque estudiaron las actividades de los animales desde la mirada de rebaño y no de individuos por sí solo como lo venían haciendo estudios previos. Para ello, utilizaron redes neuronales profundas y descubrieron que podían aprender las actividades colectivas con buena precisión, incluso cuando estas actividades estaban sesgadas.

3. MARCO TEÓRICO

3.1. Machine learning

El término Machine Learning o aprendizaje automático, en español. Es una disciplina que se ha venido desarrollando desde los años cincuenta y que está asociada a la inteligencia artificial que hace referencia a la detección automática de patrones significativos en un conjunto de datos y que permite el análisis instantáneo de información (Shalev-Shwartz, 2014). De esta manera, en las últimas dos décadas estas técnicas se han convertido en una herramienta común en tareas que requieren de extracción de información a partir de grandes volúmenes de datos.

En los métodos tradicionales de la informática clásica, la única manera de conseguir que un algoritmo realice una tarea específica es escribiendo líneas de código que definan el contexto y el detalle de cada una de las acciones y posibilidades que se quieren manejar, dando paso de esta manera a un escenario determinista donde ninguna acción es producto del azar y todo está determinado por la ejecución del algoritmo. En cambio, las técnicas de Machine Learning analizan las distintas opciones desde un enfoque probabilístico, realizando la mayoría de las acciones por cuenta propia producto de la complejidad de los patrones que se deben detectar. Con estas técnicas, el programador normalmente no proporciona especificaciones explícitas sobre las tareas que se deben realizar, siendo las herramientas de Machine Learning las encargadas de dotar a los programas de la capacidad de aprender y adaptarse a los distintos escenarios y conjuntos de datos (Shalev-Shwartz, 2014).

Una buena manera de resolver problemas relacionados a grandes cantidades de datos y técnicas de Machine Learning es utilizar herramientas basadas en la teoría de la probabilidad. La teoría de la probabilidad se puede aplicar a cualquier problema que implique algún nivel de incertidumbre. En el caso del aprendizaje automático, la incertidumbre se expresa de muchas maneras, como, por ejemplo: la elección del mejor modelo para los datos, encontrar la mejor predicción sobre el futuro dados algunos datos del pasado, entre otras. (Murphy, 2012).

Existen tres tipos de aprendizaje automático; por una parte, tenemos el aprendizaje supervisado, a este sistema se le proporciona un volumen de datos junto con sus etiquetas u objetivos para luego proporcionar nuevos datos sin necesidad de etiquetas para que en base a lo aprendido el sistema pueda clasificar los datos entregados. También, existe el aprendizaje no supervisado que se basa en proporcionar al modelo datos sin etiquetas para que éste logre encontrar patrones de información de manera directa, a este tipo de modelos se les conoce como clustering o agrupamiento. Finalmente, encontramos el aprendizaje por refuerzo, donde los sistemas aprenden a partir de experiencia en base a penalizaciones cuando se comete un error y a recompensas cuando se toma una buena decisión.

3.2. Aprendizaje no supervisado

El aprendizaje no supervisado está conformado por modelos que tienen como principal objetivo aumentar el conocimiento estructural de los datos, esto se traduce en agrupar estos datos sin etiquetas en función de sus características y similitudes con un mínimo de supervisión humana (Gonzalez, 2020). Este último punto es importante porque facilita la tarea del programador ya que en ocasiones nos enfrentamos a grandes volúmenes de datos en los que es difícil encontrar patrones a simple vista.

Normalmente, es más sencillo adquirir datos que no están etiquetados, ya que, etiquetarlos es costoso porque debe realizarse de forma manual por un experto en el área del cual provienen los datos o porque sencillamente no se conocen las etiquetas a priori.

3.2.1 Ventajas

- Es útil en fases exploratorias donde no se tiene conocimientos previos sobre el comportamiento de los datos porque no utiliza datos etiquetados.
- Los métodos de aprendizaje no supervisado ayudan a encontrar características que pueden ser útiles para la categorización.
- Alta capacidad de adaptación a los datos.
- No necesita entrenamiento previo como en métodos de aprendizaje supervisado.
- El aprendizaje no supervisado encuentra todo tipo de patrones en los datos.

3.2.2 Desventajas

- No se puede obtener información precisa con respecto a la clasificación de datos, ya que la etiqueta es desconocida.
- Es difícil validar los resultados obtenidos, ya que no se conoce la verdadera agrupación de los datos
- Es complejo decidir qué método de agrupación es el adecuado para un determinado dataset ya que se tienen que analizar múltiples parámetros. Además, no existen parámetros preestablecidos para comparar resultados como en el aprendizaje supervisado (recall, precision, accuracy, etc)
- Se necesita evaluación externa de los resultados, de parte de un profesional en el tema o se puede recurrir a una función objetiva sobre agrupación.

3.3. Reducción de dimensionalidad

Muchos de los algoritmos de reducción de dimensionalidad utilizan el aprendizaje no supervisado de métricas de distancias. Según Van der Maaten, Postma y Van den Herik (2009) el problema de la reducción de dimensionalidad se define como una matriz X de $n \times D$

compuesta por n vectores de datos x_i donde $i \in \{1, 2, 3, \dots, n\}$ con dimensionalidad D . Además, asume que el conjunto de datos posee una dimensión intrínseca donde esta dimensión intrínseca denotada d cumple que $d < D$. En términos matemáticos, la dimensionalidad intrínseca hace referencia a que los puntos en el conjunto de datos X se encuentran cerca o en una variedad dimensional contenida en el espacio D -dimensional. De esta manera, las técnicas de reducción de dimensionalidad pueden transformar el conjunto de datos X con dimensionalidad D en un nuevo conjunto de datos Y con dimensionalidad d , mientras se conserva la geometría de los datos.

3.3.1. Análisis de componentes principales (PCA)

El método de análisis de componentes principales o PCA por sus siglas en inglés, es una técnica lineal que se basa en la eliminación de la redundancia de los datos a partir de una transformación lineal ortogonal. PCA es ampliamente usado en todas las formas de análisis ya que es un método simple no paramétrico de extraer información de los datos, sin embargo, su mayor limitante es que se basa en el supuesto de linealidad (Shlens, 2005).

La técnica de análisis de componentes principales construye una representación de baja dimensión del conjunto de datos que describe la mayor parte de la varianza en sus componentes, esto se logra encontrando una base lineal de dimensionalidad reducida para los datos en la que la cantidad de varianza de los datos sea máxima (Van der Maaten et al. 2009).

Para entender el funcionamiento de PCA debemos tener en cuenta dos conceptos matemáticos imprescindibles que se deben aplicar para la reducción de dimensionalidad, vectores y valores propios. Los vectores propios son un caso particular de multiplicación entre una matriz y un vector (Amat, 2017). En la Figura 3 el vector resultante de la multiplicación es un múltiplo entero del vector original.

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} x \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 x \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Figura 3: Multiplicación de una matriz por un vector.

Por otra parte, los vectores propios de una matriz son todos aquellos vectores que, al multiplicarlos por la matriz, resultan en el mismo vector o en un múltiplo del mismo vector. Las propiedades de los vectores propios son las siguientes:

- Si se escala un vector propio antes de multiplicarlo por la matriz, el resultado será un múltiplo del mismo vector propio. Esto se debe básicamente a que si se escala un vector propio lo único que se está haciendo es aumentar su longitud, pero la dirección es la misma.
- Los vectores propios sólo existen para matrices cuadradas y no para todas. Por ejemplo, si se tiene una matriz de $n \times n$ que tenga vectores propios, estos corresponden a n vectores propios.
- Todos los vectores propios de una matriz son perpendiculares entre sí.

Dada la propiedad que al escalar un vector propio sólo cambia su magnitud y no su dirección, es usual escalarlos de tal forma que su longitud sea 1. De esta manera se consigue que todos estén estandarizados (Amat, 2007).

Cuando se multiplica una matriz por alguno de sus vectores propios se obtiene un múltiplo del vector original, este múltiplo es conocido como valor propio. A todo vector propio le corresponde un valor propio y viceversa. En PCA, cada uno de los componentes principales se compone de un vector propio y el orden de estos componentes viene dado por orden decreciente de valores propios. De esta manera, el primer componente principal es el vector propio con el mayor valor propio porque es el componente que más retiene la varianza de los datos. Para entender mejor cómo funciona PCA, se pueden analizar de forma geométrica los componentes principales. Supongamos que tenemos los datos que se muestran en la Figura 4, pertenecientes a un conjunto de datos de dos dimensiones, X_1 y X_2 .

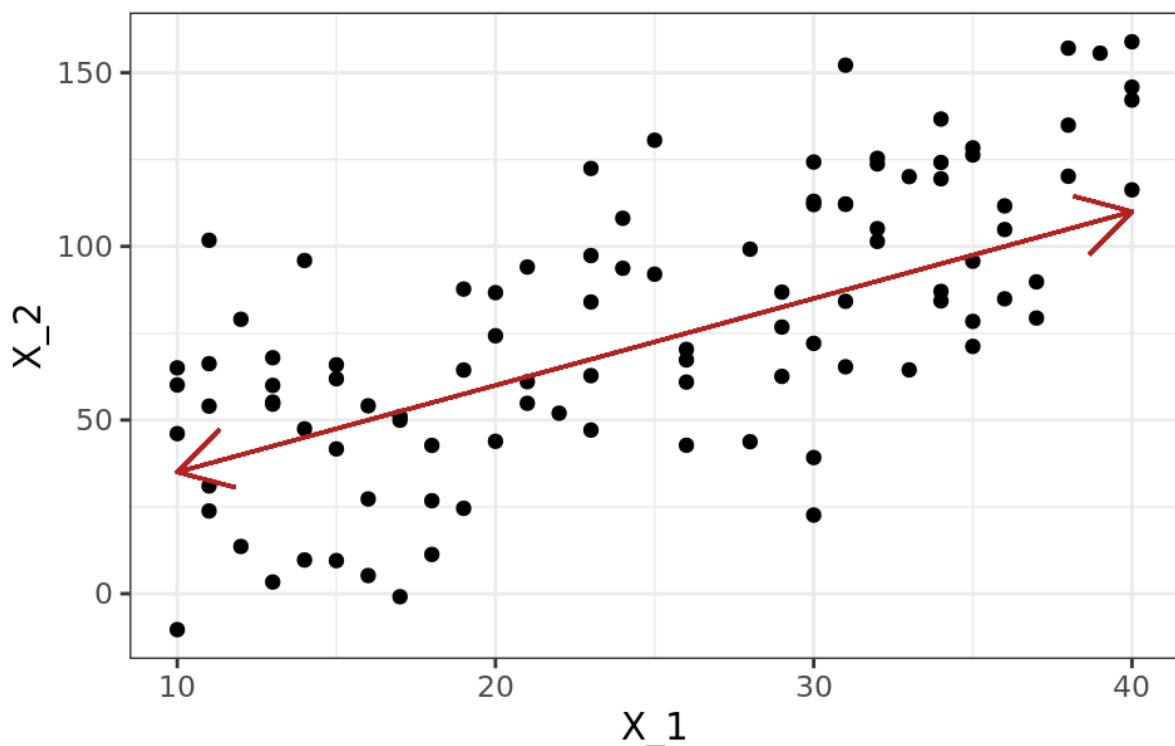


Figura 4: Conjunto de datos de dos dimensiones X_1 y X_2 . [FUENTE: www.cienciadedatos.net]

Las muestras de la Figura 5 forman a grandes rasgos una elipse, por lo que podríamos plantearnos trazar un vector que se aproxime al eje mayor de la elipse, de esta manera obtendremos la mayor cantidad de varianza en los datos obtenida en nuestra primera componente principal (línea roja). El segundo componente principal, sigue la segunda dirección donde los datos muestran mayor varianza y que no está correlacionada con el primer componente, esto equivale a decir que nuestras componentes deben ser ortogonales entre sí (Amat, 2007). En la Figura 3 se muestra la segunda proyección (línea verde) que se puede identificar con los datos dados.

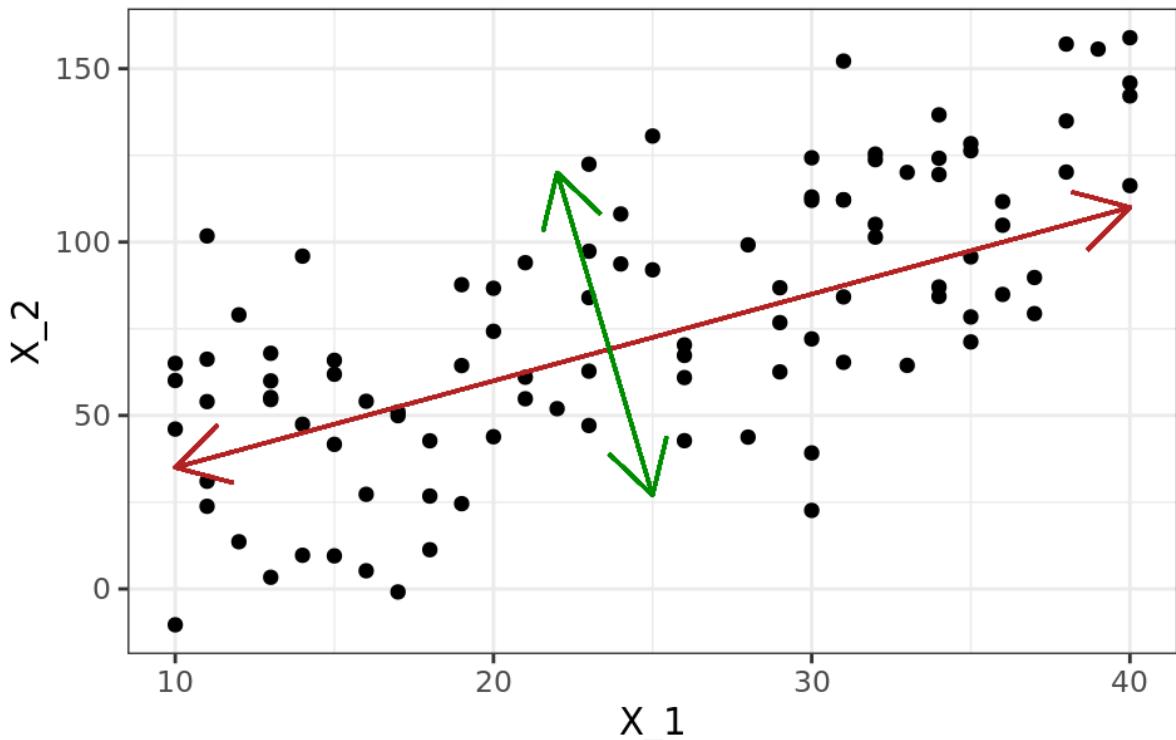


Figura 5: Conjunto de datos y sus dos componentes principales. [FUENTE: www.cienciadedatos.net]

3.3.1 Visualización

Una aplicación derivada de la reducción de dimensionalidad es la utilidad que dan al ser capaces de mostrar en una gráfica de dos o tres dimensiones la estructura de un dataset multidimensional y las proyecciones de las dimensiones originales. En la Figura 6 podemos ver cómo se aplica reducción de dimensionalidad a la base de datos Iris (Iris, hace referencia a un género de plantas), la cual consta de 3 diferentes tipos de iris (Setosa, Versicolor y Virginica), donde las columnas del dataset representan longitud del sépalo, ancho del sépalo, longitud del pétalo y ancho del pétalo (Scikit-learn: Machine Learning in Python). Este dataset es mundialmente conocido porque es el más utilizado en la literatura de reconocimiento de patrones. En la Figura 4 se puede ver que la dimensionalidad disminuye a dos y que esto hace que los grupos se puedan ver de manera más clara que en tres dimensiones o más.

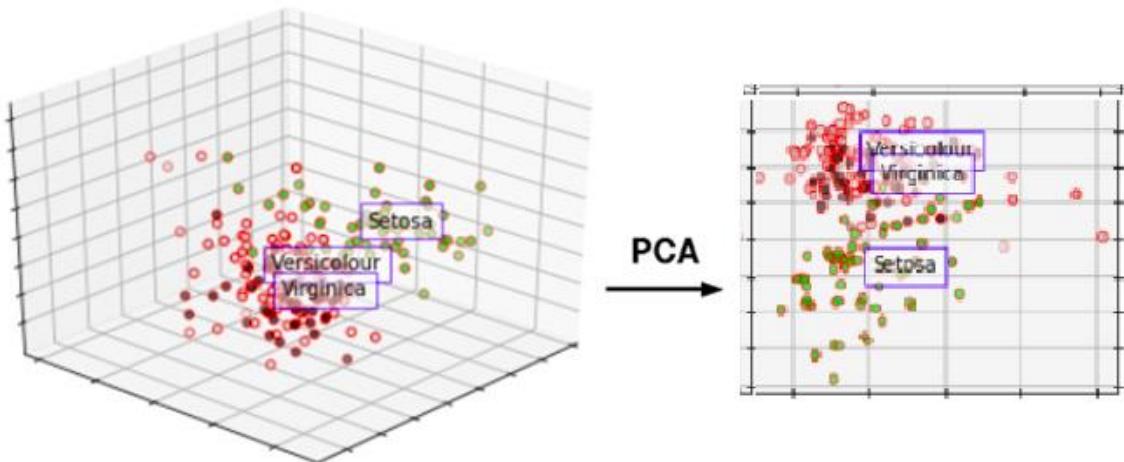


Figura 6: Reducción de dimensionalidad dataset iris.[FUENTE: www.deeplearningitalia.com]

3.4. Técnicas de agrupamiento

El agrupamiento o clustering es un tipo de aprendizaje no supervisado que consiste en dividir un conjunto de datos en grupos llamados clusters. Los datos pertenecientes a un cluster son más similares entre sí que a datos pertenecientes a otros clusters (Sammut y Webb, 2011). Es por este motivo, que las técnicas de agrupamiento son de las más utilizadas para el análisis exploratorio de los datos, ya que, puede ser aplicado a cualquier disciplina para obtener una primera intuición de los datos, siendo capaz de identificar grupos de interés a partir de los datos.

Los métodos de agrupamiento pueden ser clasificados de la siguiente manera:

- Agrupamiento por particiones: suele fijarse el número de clusters (k-means, CLARANS)
- Métodos basados en densidad: un cluster se ubica en una zona densa de datos y es separada de otro cluster por zonas menos densas de datos, estos métodos son útiles cuando los cluster tienen formas irregulares (DBSCAN)
- Agrupamiento jerárquico: no se fija el número de clusters (BIRCH, ROCK)

En la Figura 7, se pueden observar comparaciones entre diferentes tipos de clustering con datos de prueba. Cabe destacar que los parámetros de cada algoritmo y los conjuntos de datos se han ajustado para producir buenos resultados a modo de ejemplo ilustrativo (Scikit-learn: Machine Learning in Python).

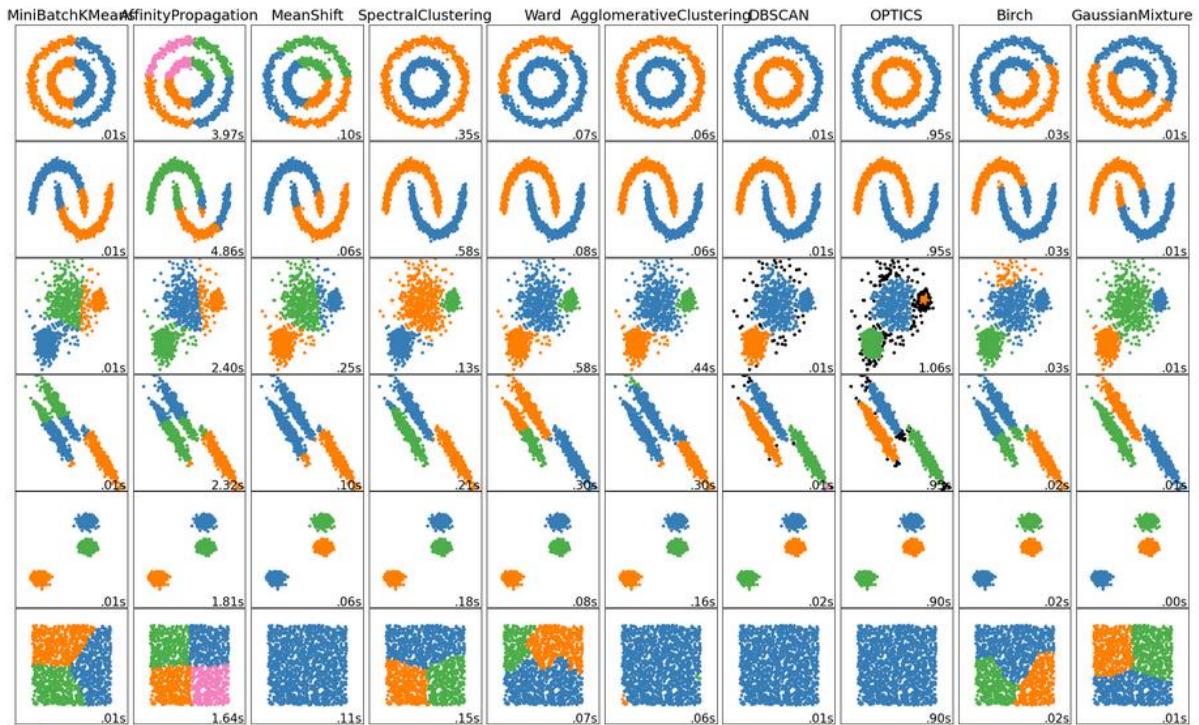


Figura 7: Gráficas comparativas entre diferentes técnicas de agrupamiento. [FUENTE: scikit-learn.org]

Un punto importante a tener en consideración para implementar técnicas de agrupamiento es la normalización de los datos, en especial cuando se utilizan agrupamientos basados en distancias como k means o cualquier tipo de cluster jerárquico porque evitamos que atributos que estén en una escala mayor influencien de sobremanera a los que se encuentran en una escala menor y que esto nos pueda traer como consecuencia un agrupamiento erróneo.

3.4.1. K-means

Es una técnica de agrupamiento en la cual se segmentan los datos en K grupos o clusters indicados por el programador. El número de grupos no se conoce a priori, pero existen técnicas para elegir este parámetro K, uno de ellos es el método del codo que utiliza la distancia media de las observaciones pertenecientes a un centroide, esto quiere decir que el método se centra en la distancias o inercia intra-cluster (Scikit-learn: Machine Learning in Python). En la Figura 8 se puede ver un ejemplo gráfico del método, en ella se puede apreciar que a medida que se aumenta el valor de K (número de clusters) , la inercia intra-cluster tiende a disminuir. Cuanto menor sea esta distancia, es mejor, ya que se infiere que los cluster son más compactos. Sin embargo, también hay que tener en cuenta que debemos seleccionar el número de cluster en el punto donde se evidencie una inflexión significativa o codo en relación a la inercia anterior, en el caso del ejemplo este valor corresponde a K=5, pero además hay que considerar que dependiendo del caso y la naturaleza del problema el valor de K ya podría estar pre definido por diversas situaciones, como por ejemplo: Consideremos que una empresa desea establecer cuatro estrategias de captación de clientes en base a su comportamiento de compras durante un año, en este caso el número de cluster debiese ser cuatro, ya que corresponde a la cantidad de grupos diferentes de interés para la empresa para definir una estrategia comercial para cada uno de los cluster.

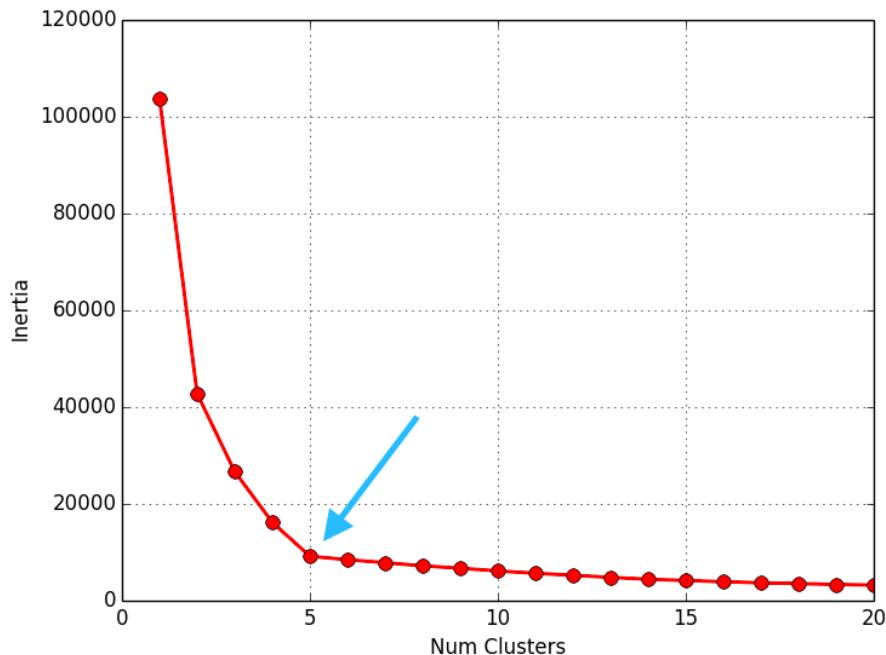


Figura 8: Método del codo para encontrar el valor de K óptimo.

Una vez definido el valor K, el algoritmo de k means sigue los siguientes pasos (IArtificial, 2020):

1. Se inicializan los centroides de los K grupos de manera aleatoria en el espacio de los datos.
 2. Se asigna cada dato al centroide más cercano en base a la fórmula de la Figura 9. Al aplicar esta fórmula, estamos calculando las distancias entre los puntos y los centroides y asignando los puntos al centroide más próximo.

Obtener las asignaciones, S , que minimizan la fórmula

$$\arg \min_S \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

Por cada punto asignado al grupo i

Cantidad de grupos k Centroide del grupo i

Figura 9: Fórmula de asignación de datos a su respectivo centroide.

3. Finalmente se actualizan las posiciones de los centroides, quedando en la media de las muestras más cercanas.

Este proceso se repite en forma iterativa hasta que la asignación no cambie. El resultado final representa el ajuste que minimiza la distancia entre elementos del mismo cluster y maximiza la distancia entre los clusters.

K-means es una de las técnicas de agrupamiento más sencillas y veloces en cuanto a cómputo. Sin embargo, posee varias limitantes que hay que tener en cuenta al momento de implementarlo. Por una parte, las agrupaciones resultantes van a variar dependiendo de la inicialización de los centroides, provocando que muchas veces lleguemos a un mínimo local y no a uno global. Para minimizar este problema es recomendable repetir el proceso de clustering varias veces, un número razonable es sobre 25 (Amat, 2017). Otro problema visible en este tipo de agrupamiento es que se presentan problemas de robustez frente a datos atípicos (outliers, en inglés), la única solución a este problema es eliminar estos datos o implementar un algoritmo de agrupamiento más robusto.

3.4.2. Agrupamiento jerárquico

Es una técnica de agrupamiento que no requiere que se defina el número de clusters a priori. Los métodos que componen el agrupamiento jerárquico pueden ser subdivididos en dos tipos (Amat, 2007), ambos pueden ser representados por un dendrograma que se define como un diagrama de árbol que muestra los cluster de manera gráfica por medio de líneas que se fusionan para formar grupos con las observaciones en cada paso. Los dos tipos de Agrupamiento jerárquico son:

- Cluster aglomerativo o bottom-up: El agrupamiento se inicia con cada observación como un clúster individual, y se van agrupando en base a las distancias entre pares de clúster, esta distancia y el enlace (linkage en inglés) deben ser definidas por el programador con anterioridad. Finalmente, los dos clusters más cercanos o similares se combinan creando uno. En la Figura 10 se puede ver un ejemplo de dendrograma formado por un Agrupamiento jerárquico aglomerativo.

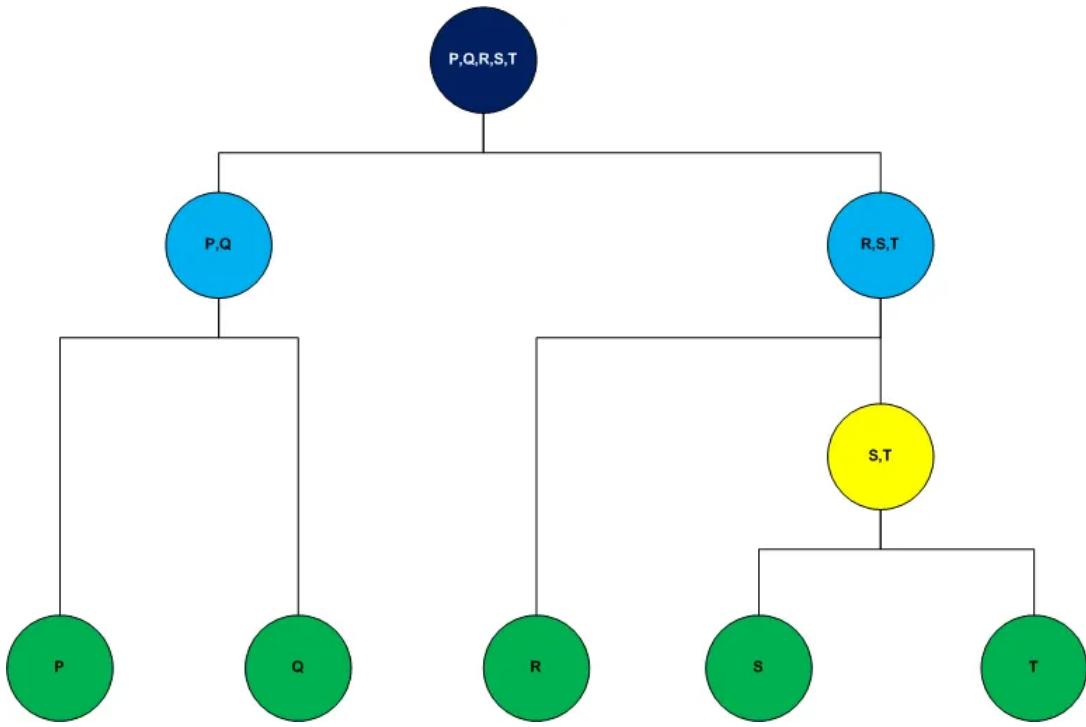


Figura 10: Dendrograma Agrupamiento jerárquico aglomerativo.
 [FUENTE:www.jacobsoft.com.mx]

- Cluster divisivo o top-down: Este algoritmo inicia con un único cluster y calcula para cada cluster la mayor de las distancias entre pares de observaciones, luego selecciona el cluster con mayor diámetro y calcula la distancia media de cada observación con respecto a las demás. Para finalmente, tomar la observación más distante para que inicie un nuevo cluster. Luego, se reasignan las observaciones restantes al nuevo cluster o al viejo dependiendo de cuál se encuentre más cercano (Amat, 2017). En la Figura 11 podemos observar el dendrograma correspondiente a un cluster divisivo.

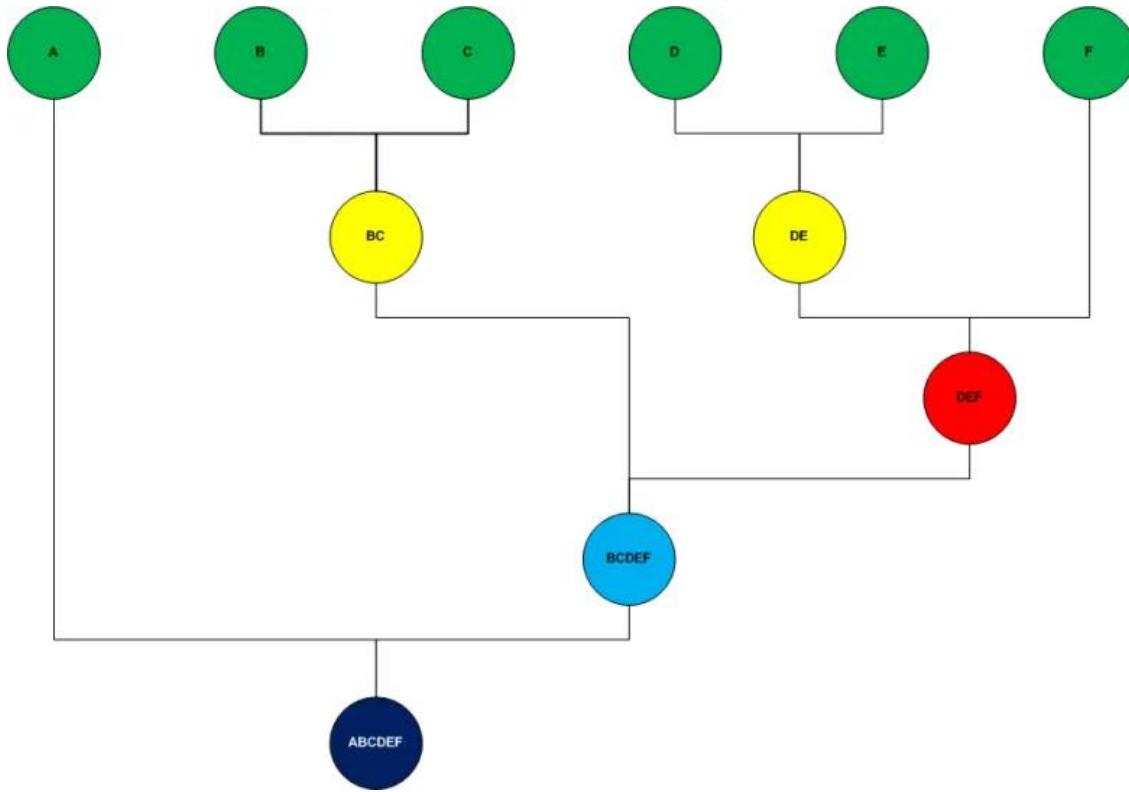


Figura 11: Dendrograma Agrupamiento jerárquico divisivo.
 [FUENTE:www.jacobsoft.com.mx]

El dendrograma resultante de aplicar cualquiera de los dos métodos de agrupamiento jerárquico, puede ser trazado por una línea horizontal en cualquiera de sus niveles y se puede obtener el número de clúster que hay a esa altura del dendrograma. Por lo tanto, la altura del corte al dendrograma representa lo mismo que el valor K de k-means (número de grupos). Las ventajas del agrupamiento jerárquico es que resulta fácil de entender y de implementar. Además, su dendrograma brinda una manera gráfica y sencilla de encontrar el número óptimo de clusters. Sin embargo, el agrupamiento jerárquico no es útil para grandes volúmenes de datos, ya que las operatorias de cálculo de distancias se realizan para cada observación.

3.4.3. GMM

El Modelo de Mezcla Gaussiana, corresponden a una técnica de agrupamiento perteneciente al grupo de modelos probabilísticos, que se utiliza para representar la presencia de subpoblaciones dentro de una misma población, el objetivo es calcular la probabilidad de pertenencia de los datos a cada cluster, asignando cada observación al cluster más probable.

Debido a que no es posible explicar las distribuciones de los datos mediante una sola distribución estadística, es necesario utilizar una composición de ellas, las que se pueden describir mediante modelos de mixturas, las que se definen por los parámetros de cada componente y las proporciones en las que cada una aporta a la distribución general (Díaz, 2018). Los parámetros que describen los modelos pueden ser estimados en base a varias

técnicas: Aproximaciones bayesianas, métodos gráficos, Expectation-Maximization (EM), entre otras.

El algoritmo EM es una herramienta iterativa que soluciona el problema de incertidumbre de no saber qué puntos provienen de qué componente latente. Primero, se asumen componentes aleatorios y se calcula para cada punto una probabilidad de ser generado por cada componente del modelo. Luego, se ajustan los parámetros para maximizar la probabilidad de los datos dadas las asignaciones. La repetición de este proceso garantiza la convergencia a un óptimo local (Scikit-learn: Machine Learning in Python).

En la Figura 12 se puede ver un ejemplo ilustrativo compuesto por 24 observaciones que según GMM se pueden agrupar a partir de 3 gaussianas diferentes o cluster. Además, se observan las medias de las gaussianas (μ_1 , μ_2 y μ_3)

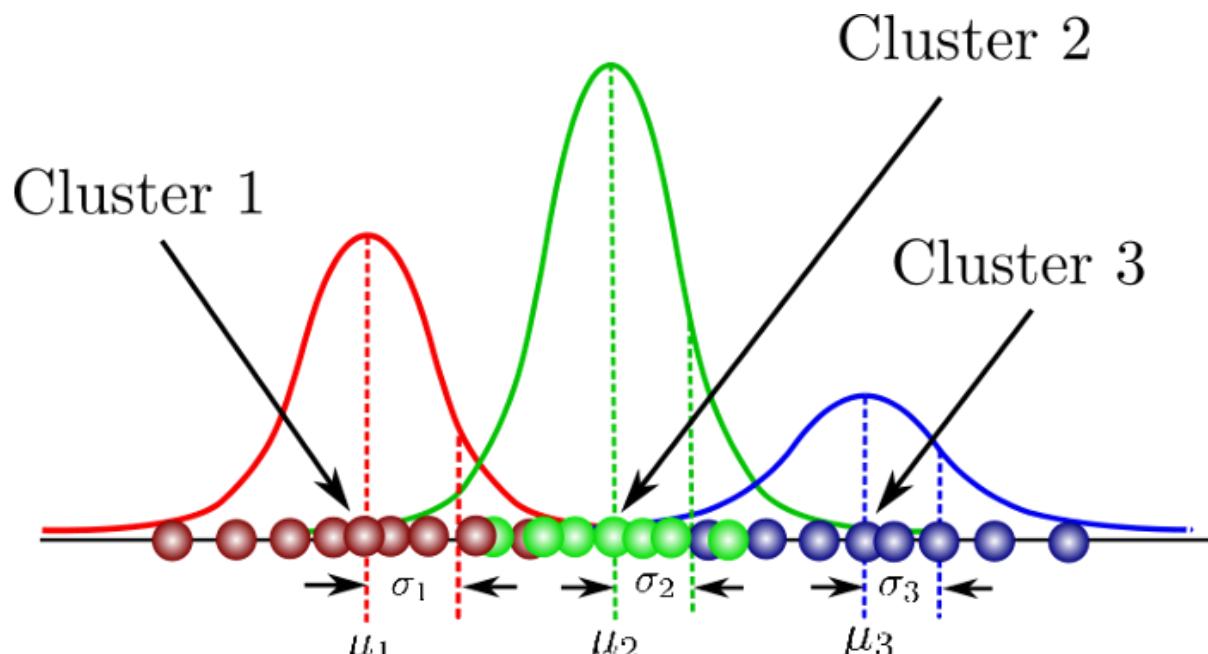


Figura 12: Modelo de mezclas de Gaussianas con 3 componentes. [FUENTE: towardsdatascience.com]

El resultado de haber agrupado con el método de GMM y haber ajustado los parámetros con el algoritmo de EM se puede ver en la Figura 13, donde se logró converger a un buen resultado.

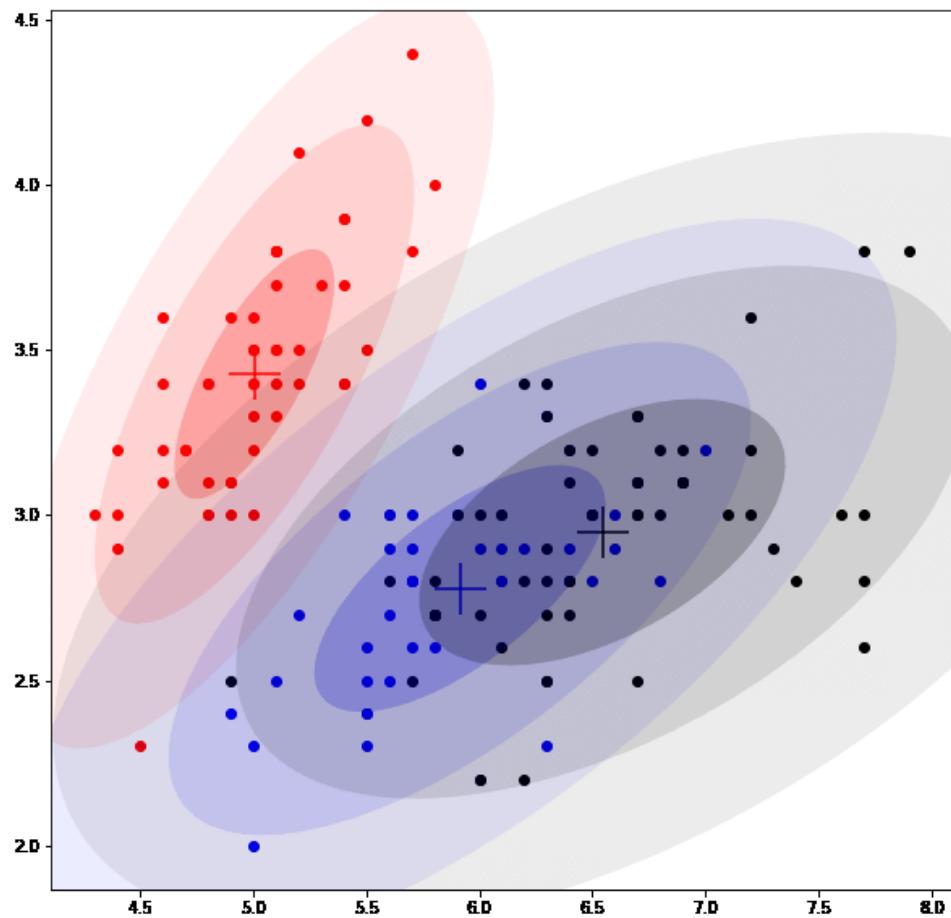


Figura 13: Gráfica del resultado al aplicar GMM y el algoritmo de EM.
[FUENTE:towardsdatascience.com]

4. MATERIALES Y MÉTODOS

El estudio para establecer los patrones jerárquicos en base a atributos fenotípicos y comportamiento animal a través de machine learning se llevó a cabo en 3 etapas, comenzando por los estudios de campo realizados en un rebaño ovino de la Universidad Austral de Chile, continuando con el trabajo de gabinete que implica la digitalización de los datos obtenidos de las mediciones de los animales en estudio y los registros visuales de los individuos para el preprocesamiento de los datos, y finalmente la etapa de implementación de los algoritmos de agrupamiento que explicaría la conducta jerárquica.

4.1. Estudio de campo

Se utilizaron 50 hembras ovinas de raza Austral de 1 a 6 años de edad seleccionadas al azar del rebaño de la ovejería de la Estación Experimental Agropecuaria Austral (EEAA) de la Universidad Austral de Chile, ubicado en la salida norte de la ciudad de Valdivia, sector de Cabo Blanco (Ver Figura 14). El rebaño de estudio consideró replicar un rebaño modelo de la agricultura familiar campesina que contiene menos de 60 animales. La estructura del rebaño fue realizada mediante la edad de los animales, con una composición por edad de 5 categorías animales entre adultas y jóvenes según la Tabla 1, resguardando mantener el equilibrio etario en la composición del rebaño de estudio. Los animales pastoreaban durante el día un potrero de 2 hectáreas asignadas para el estudio con agua a libre disposición. Durante las tardes los animales eran resguardados del ataque de predadores en un corral especial para esos fines.

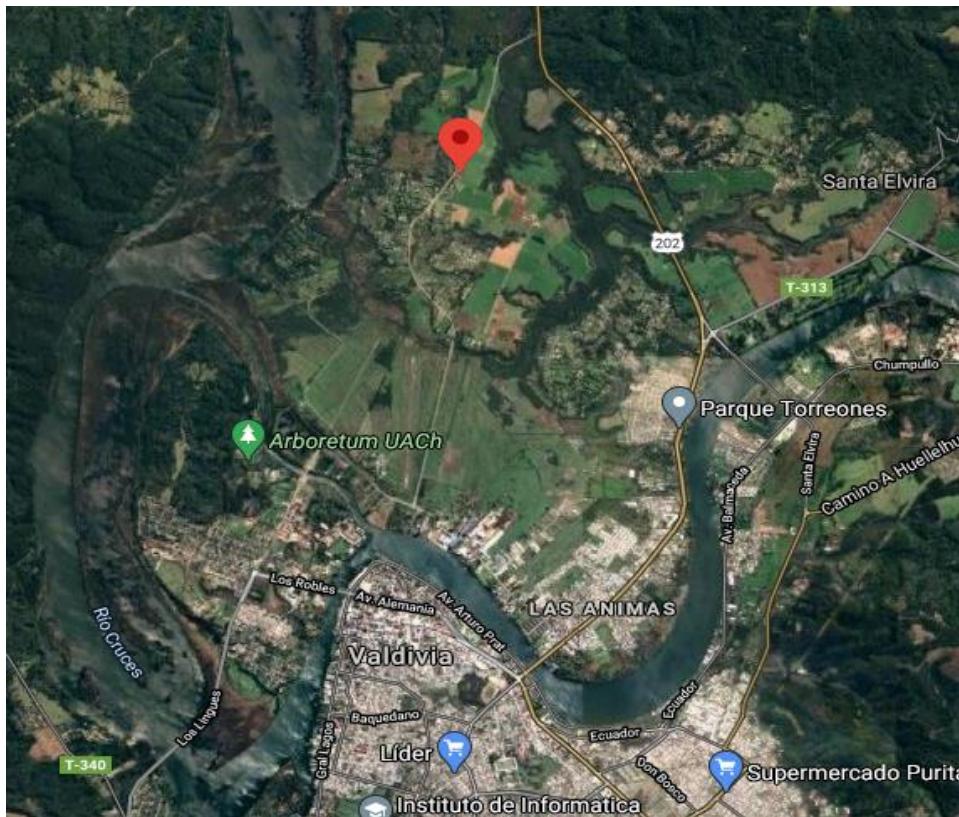


Figura 14: Localización geográfica de la estación experimental agropecuaria UACH.
[FUENTE: www.google.cl/maps]

Tabla 1: Distribución de individuos según número de dientes definitivos (edad)

Número de dientes definitivos	Número de individuos	Porcentaje
0	13	26%
2	7	14%
4	9	18%
6	10	20%
8	11	22%
Total	50	100%

El estudio fue llevado a cabo durante los meses de junio y julio del año 2020, durante la gestación del rebaño, realizando preliminarmente visitas al predio y su plantel para conocer y comprender el manejo del rebaño, la infraestructura, así como la identificación individual de los animales. Además, durante las visitas del rebaño se aprovechó de determinar los mejores lugares para instalar los aparatos que capturarían los registros visuales de manera fija. Las distintas actividades de preparación de las condiciones ideales del estudio se realizaron en paralelo con estas actividades:

- Conocer los animales

- Seleccionar el rebaño que constituirá el grupo de estudio
- Seleccionar el potrero para el estudio y dividirlo en sectores
- Identificación de sectores estratégicos para la recolección de datos
- Instalación de cámaras DVR
- Instalación de la estación de monitoreo
- Adquisición de un drone
- Pruebas de identificación para individualizar a las ovejas
- Elaboración del diseño experimental

Las actividades de preparación fueron realizadas durante las visitas a la EEAA por una semana (del 15 al 19 de junio). Además de la selección de los animales durante esos días se seleccionó el potrero a utilizar durante el estudio. Dadas sus dimensiones y la disponibilidad de forraje para mantener a las ovejas no fue necesario aportar alimento extra. En la Figura 15 se puede visualizar un esquema con sus dimensiones y la división de sectores que se realizó para facilitar el monitoreo del drone haciendo estudios preliminares de las formas en que debía ser monitoreado desde el aire el rebaño y la disponibilidad de forraje.

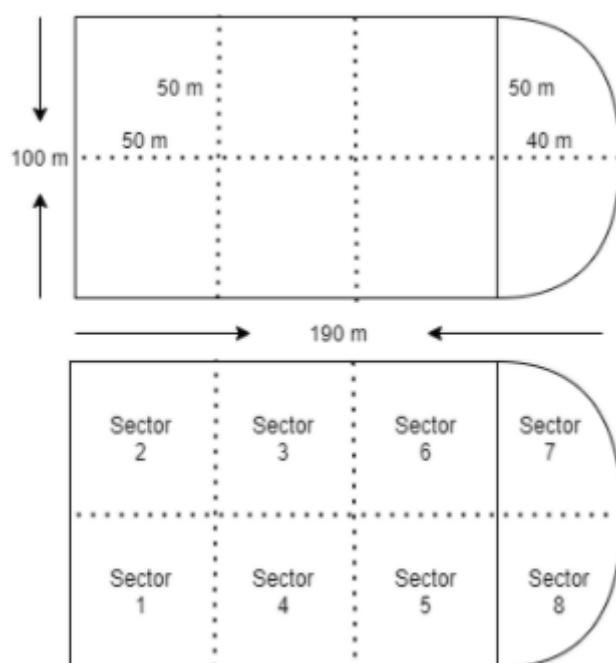


Figura 15: Esquema del potrero y sus sectores.

Además, se identificaron los lugares estratégicos para la instalación de las cámaras fijas para garantizar que el movimiento del rebaño durante el día fuera registrado de manera eficiente (Figura 16). Permitiendo el registro del recorrido de mañana y tarde del rebaño desde y hacia el potrero (Cámara 1 y 2) y la cámara 3 se ubicó sobre los comederos.

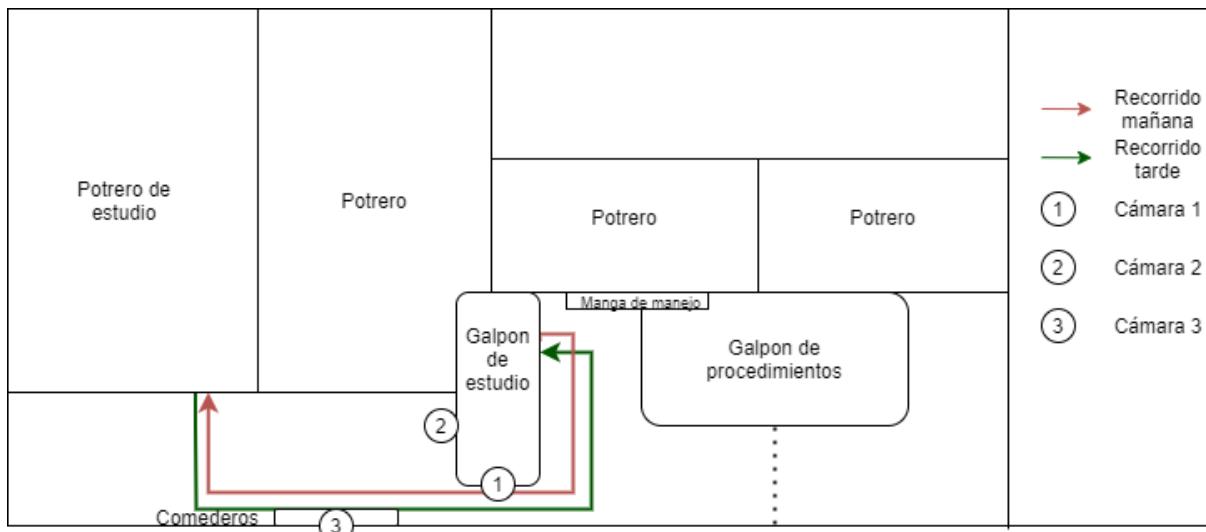


Figura 16: Esquema de la locación de estudio

La instalación de las cámaras tuvo desafíos asociados, ya que se tuvieron que hacer pruebas para corroborar que estén enfocando bien el objetivo y que no sufriera daños con las condiciones meteorológicas propias del invierno. Es por esto, que las cámaras fueron protegidas con una carcasa plástica para evitar que la lluvia moje el lente y evite la toma de datos (Ver Figura 17).



Figura 17: Vista desde abajo de la cámara 1

El manejo de los animales se realizó dentro del galpón en una manga portátil (Figura 18) que contenía una pesa. Una vez constituido el estudio se debió considerar un tiempo preexperimental para acostumbrar al rebaño a mi presencia y al ruido que genera el drone volando sobre ellas con la finalidad de que su comportamiento no se altere, condicionado por el miedo que pudiese provocar el sonido, la cercanía del dispositivo, y el operador/observador. En la Figura 20 se muestra una imagen de las tomas de prueba que se realizaron para examinar

el comportamiento de las ovejas frente a la presencia del drone, donde además se puede apreciar por la postura corporal que la presencia del drone no interfiere en su comportamiento habitual.



Figura 18: Manga de manejo para ovinos.



Figura 19: Pruebas de vuelo con drone



Figura 20: Imagen tomada por el drone a una altura de 10 metros.

Además de la selección del potrero, las pruebas de vuelo y la selección de individuos, se realizaron las pruebas de identificación visual para poder individualizar a cada ovino a través de cámaras estáticas y tomas realizadas por el drone. Para ello, se utilizó una metodología de ensayo y error, confeccionando números de diferentes materiales para adherirlos al lomo de las ovejas y probar su adherencia y visualización a través de las cámaras. La primera prueba se realizó con números elaborados sobre una base de cartón (15 x 10 cm) con un número impreso plastificado con cinta aisladora y adherido a la oveja con pequeñas cantidades de silicona de baño (Ver Figura 21). Si bien, la silicona pegó bien al principio, los números comenzaron a caerse después de algunas horas, además este mecanismo de identificación tuvo problemas porque a veces no era visible a través de las cámaras.



Figura 21: Primer mecanismo de identificación.

Luego del primer intento fallido, se construyó un segundo mecanismo de identificación con números de dos dígitos fabricados con huincha aisladora negra en sobre plásticos haciendo uso de la lana para fijarlos al lomo como se muestra en la Figura 22, este mecanismo tampoco fue exitoso porque al término del día ya se había perdido la identificación de al menos el 50% de las ovejas, sumado a esto, la visualización a través de las cámaras muchas veces era mala porque al ser una base plástica la luz del sol se reflejaba y no se lograba ver el número (Ver Figura 23)



Figura 22: Segundo mecanismo de identificación.



Figura 23: Visualización a través de las cámaras del segundo método de identificación.

Finalmente, se terminó por confeccionar un método de identificación que siguiera la misma construcción del primer método, pero de mayores dimensiones (20 x 12 cm) y adherido a la oveja haciendo uso de su propia lana, la numeración de la identificación estaba comprendida entre 01 y 50, cada número quedó asociado al número de arete de la oveja para evitar pérdida de información en caso de extravío del número asignado (Ver Figura 24). Este método fue exitoso porque los números permanecieron en su lugar y además eran visibles a través de las cámaras estáticas, así como de la cámara del Dron.



Figura 24: Ovejas captadas por cámara estática, individualizadas con el tercer método de identificación.

4.1.1. Selección y medición de las características fenotípicas de los individuos involucrados en el estudio.

Al finalizar la semana de observaciones preliminares, se identificaron las características que serían parte del dataset de este estudio. En este aspecto, se involucraron características fenotípicas convencionales y otras no convencionales. A continuación, se presentan y describen las diferentes características de interés:

- Cantidad de dientes definitivos: La determinación del número de dientes de un ovino está estrechamente ligada con la edad. Los ovinos carecen de incisivos superiores, por ende, la observación se centra en los incisivos inferiores, donde se presta especial atención al cambio de dientes de leche por dientes definitivos o permanentes (Romero, 2015).
Los dientes de leche son más pequeños y amarillentos que los dientes definitivos. En la Figura 25 se puede observar a un ovino con sus dos primeros incisivos definitivos, en la imagen se puede notar la diferencia de tamaño y coloración entre los dientes de leche y los permanentes.



Figura 25: Ovino con un par de dientes definitivos.

La estimación de edad de los ovinos según la cantidad de dientes definitivos se puede ver en la Tabla 2.

Tabla 2: Equivalencia de edad según cantidad de dientes definitivos

Cantidad de dientes definitivos	Rango de edad
0	Entre 0 y 12 meses
2	Entre 12 y 18 meses
4	Entre 18 y 24 meses
6	Entre 24 y 36 meses
8	36 meses o más

Una vez pasados los 4 años de edad, sólo se puede obtener un estimado de la edad por medio del desgaste de los dientes.

- Peso vivo: este atributo es usado comúnmente como una medida estándar de la composición corporal y es medido en kilos de peso vivo mediante una balanza electrónica, dispuesta al final de una manga.
- Condición corporal: La evaluación de la condición corporal es una técnica utilizada para determinar el estado nutricional de un animal. Esta medición se realiza mediante palpaciones a nivel lumbar (Ver Figura 26), que permiten estimar la cantidad de grasa que existe bajo la piel.
La escala utilizada para la medición de la condición corporal va de 1 a 5 y representa distintos estados nutricionales (Ver Figura 27)

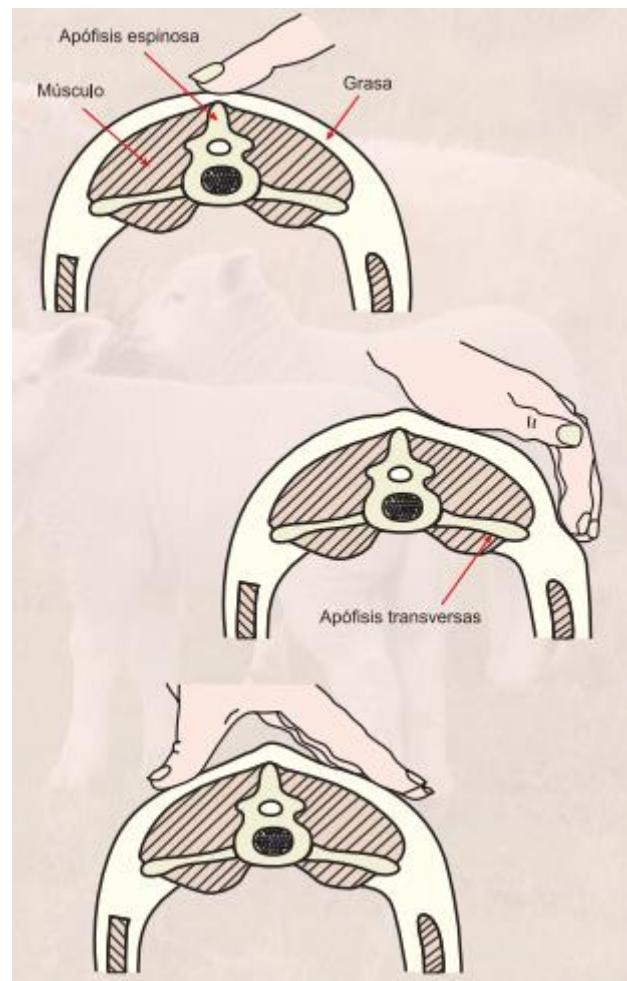


Figura 26: método utilizado de medir la condición corporal de ovinos [FUENTE:
www.oviespana.com]

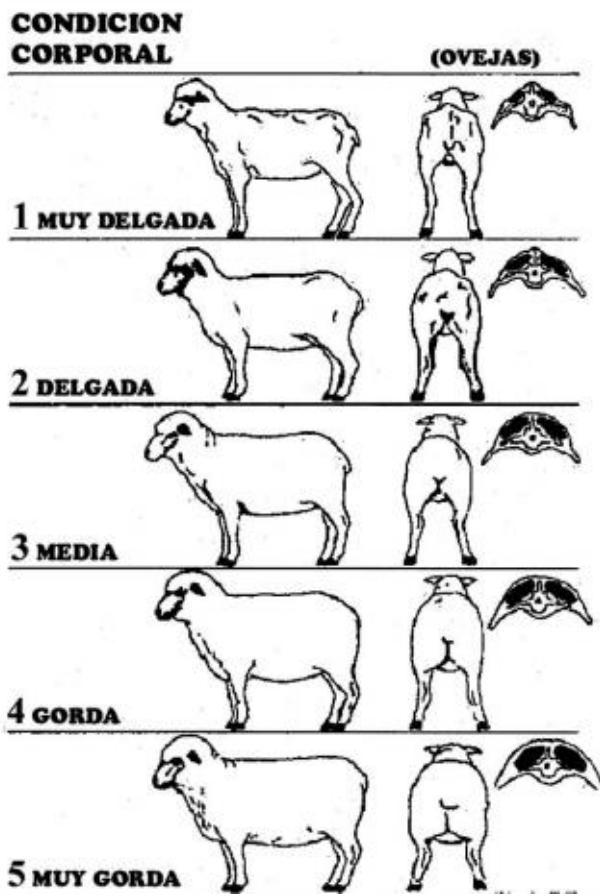


Figura 27: Escala de medición de la condición corporal de ovinos.[FUENTE:inta.gob.ar]

- Altura a la cruz: Estatura del ovino desde el suelo hasta la parte más alta de la cruz (Ver Figura 28)
- Altura a la grupa: Estatura del ovino desde el suelo hasta el punto de unión de la región del lomo con la grupa (Ver Figura 28)

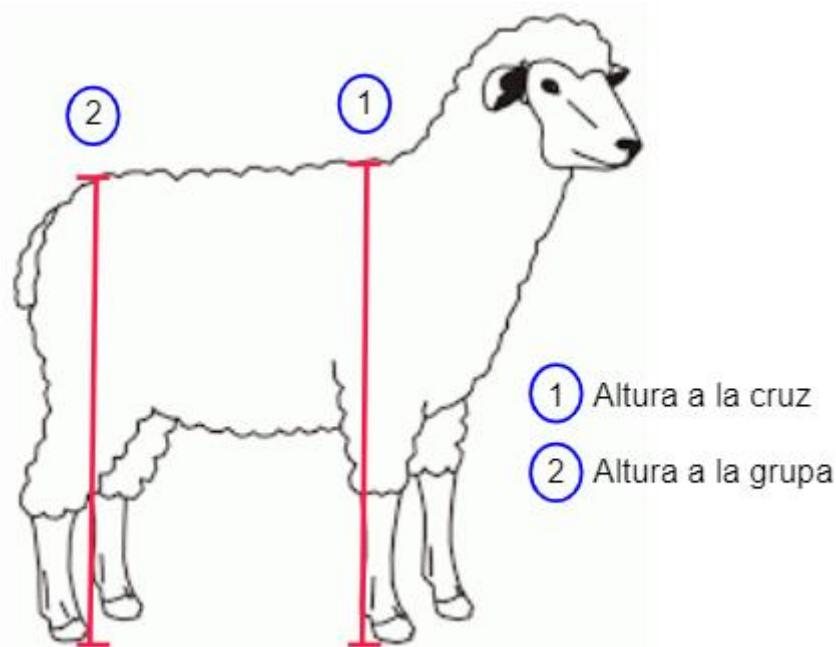


Figura 28: Esquema de mediciones de la altura de la cruz y la grupa.

- Distancia entre los ojos: Corresponde a una medida no convencional que mide la distancia entre los ojos como lo muestra el esquema de la Figura 29.

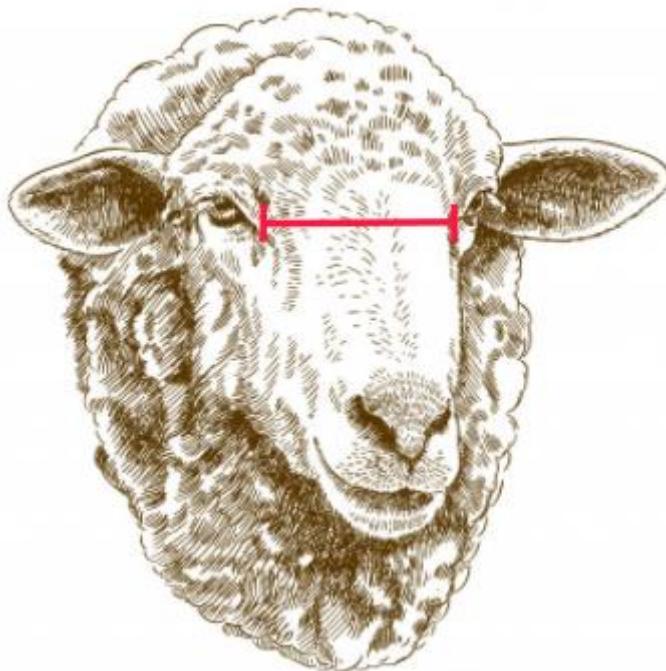


Figura 29: Esquema de medición de la distancia entre los ojos.

- Distancia entre el hocico y los ojos: Corresponde a una medida no convencional que mide la distancia entre el hocico en línea recta hasta la altura de los ojos como se muestra en el esquema de la Figura 30.

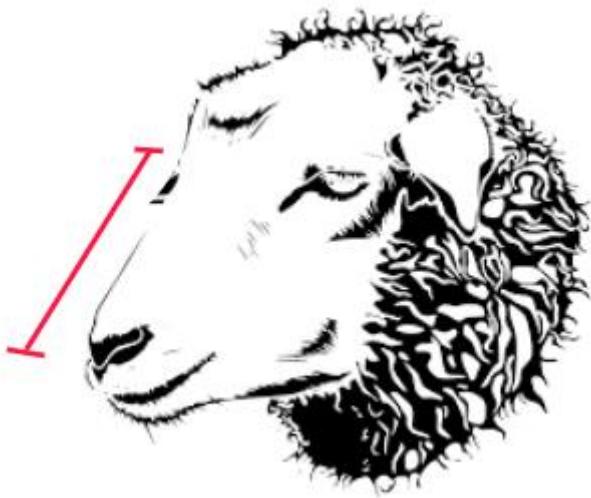


Figura 30: Esquema de medición de la distancia hocico-ojos.

4.1.2 Recolección de datos

El período experimental del trabajo contempló 15 días de estudio en terreno, entre el 22 de junio y el 06 de Julio del 2020, periodo en el cual se efectuó la recolección de datos. En esta etapa se utilizaron cuatro mecanismos con distinta periodicidad de captura, destinados a obtener diferentes tipos de datos de los ovinos: características de los individuos, comportamiento y disponibilidad de pasto.

➤ Caracterización del rebaño

Las características de los individuos fueron medidas dos veces a lo largo del período experimental: el primer día y el último.

El primer día de estudio alrededor de las 8.30 am, antes de que el rebaño saliera a pastorear se realizó la caracterización de los individuos para que las medidas sean lo más fidedigna posible, ya que atributos como el peso puede variar considerablemente si son medidas luego de la ingestión de alimentos. Para la toma de estos datos, el rebaño fue conducido a la manga de manejo para ovinos, donde se realizaron las mediciones de cada una de las características presentadas en el punto 4.1.1 (Ver Figura 31 y 32).



Figura 31: Ovejas en la manga de manejo.



Figura 32: Toma de características fenotípicas

Al finalizar el estudio se volvieron a realizar las tomas de datos, pero esta vez sólo se consideraron los que podían variar en el transcurso de 15 días, es por esto que sólo se volvió a medir peso vivo y condición corporal.

➤ Cámaras

Para la captura de los videos de las cámaras estáticas se utilizó un DVR (Digital Video Recorder) como el que muestra la Figura 8. Su principal función es tomar los videos captados por las cámaras y almacenarlos en un disco duro en formato digital. Además, este dispositivo brinda diversas configuraciones para las cámaras y la periodicidad de grabación. La instalación del DVR se hizo en el corral de las ovejas (Ver Figura 33) y se configuró para que grabe todo

el tiempo. Además, se instaló un monitor para poder verificar durante el estudio que las cámaras estuvieran grabando de manera correcta, para no perder datos.



Figura 33: Estación de monitoreo

Las cámaras se ubicaron en tres puntos de interés: la cámara 1 se instaló sobre un pasillo de 3 metros de ancho por 6 metros de largo por donde las ovejas transitaban dos veces al día, la primera, para ir a comer en las mañanas entre 8.30 am y 9.00 am y la segunda para volver al corral luego de haber pasado todo el día en el sector de pastoreo, entre 17.30 pm y 18.00 pm (Ver Figura 34).

La segunda cámara se instaló detrás del corral donde las ovejas pasaban la noche. Este lugar tiene una vista panorámica de la ruta que las ovejas siguen para ir a comer y el sector de comederos donde se les da de comer en las tardes (Ver Figura 35).

La tercera cámara se ubicó arriba de los comederos para monitorear el comportamiento del rebaño cuando se acercaban al lugar (Ver Figura 36).



Figura 34: Cámara 1: Pasillo



Figura 35: Cámara 2: detrás del corral



Figura 36: Cámara 3: comederos.

➤ Medición de pasto

La medición de disponibilidad de pasto se realizó con un plato medidor de praderas (Rising plate meter, en inglés) como el de la Figura 37, esta herramienta desarrollada en Nueva Zelanda permite medir una pradera de manera sencilla. Para medir, se debe posicionar el plato en un lugar donde no se observan grandes cambios de pendiente en el terreno y hacer presión sobre el dispositivo, de esta manera se obtiene la medición de la altura comprimida del forraje.

Al tratarse de una herramienta manual, se tuvo que tomar nota de cada medición con el plato para luego aplicar la fórmula que calcula la cantidad de kilogramos de materia seca por hectárea.



Figura 37:Plato medidor de forraje

La medición con el plato se realizó tres veces durante el periodo experimental. La primera fue el primer día de estudio, la segunda el octavo día de estudio y la última al finalizar el estudio. Cada una de las tres mediciones se realizaron a todo el potrero con una distancia de 3 metros entre cada medición dando un promedio de 80 mediciones por cuadrante, siguiendo el esquema de la Figura 38.

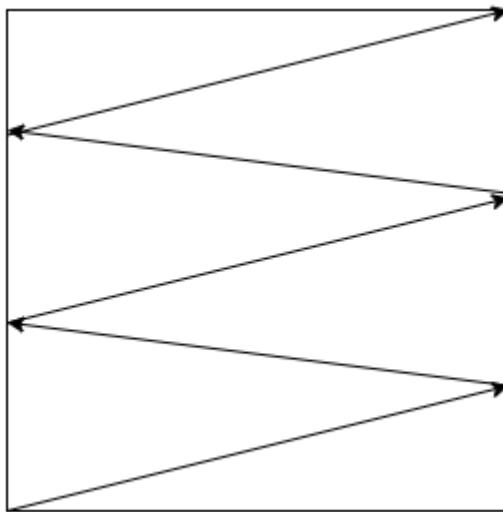


Figura 38: Esquema de medición sector 1

➤ Drone

La captura de videos con el drone fueron realizadas durante los 15 días de estudio, para ello se utilizó un drone marca Dji modelo Mavic Air 2 como el de la Figura 39, el cual cuenta con 3 baterías, de 10 minutos de autonomía cada una, lo cual nos permitía realizar 3 vuelos en la mañana entre 9 a.m y 11 a.m y 3 en la tarde entre 15.00 pm y 17.30 pm, registrando un total de 1 hora de grabación al día del comportamiento de los ovinos en la pradera.

Las tomas con el drone, se realizaron por medio de la aplicación DJI GO 4, que es la aplicación oficial de la marca DJI que se puede encontrar tanto para Android como iOS (Ver Figura 40), posee una interfaz fácil de usar con distintos indicadores que permiten realizar un vuelo seguro.

La captura de videos se realizó a una altura de 10 metros (Ver Figura 41), esta altura se considera óptima porque se podían identificar a los individuos por medio del número en su lomo y además evitamos cambios de conducta en los ovinos debido al sonido o a la proximidad del drone.

La razón por la cual se grabó dos veces al día fue porque en esos periodos las ovejas presentan mayor actividad (Quichiyao, 2003).

En la Figura 42 se puede ver cuál fue el recorrido del drone cada vez que se capturaban los videos.



Figura 39: Drone DJI Mavic Air 2

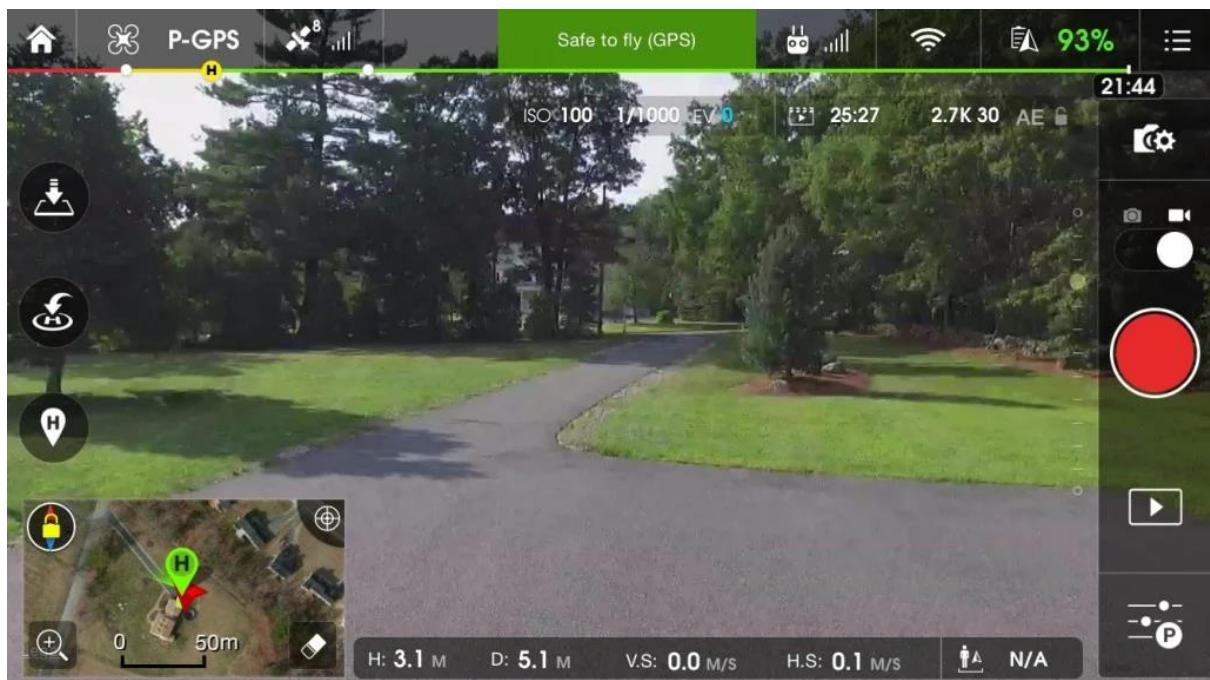


Figura 40: Aplicación DJI GO 4



Figura 41: Captura fotográfica desde el drone a 10 metros de altura.

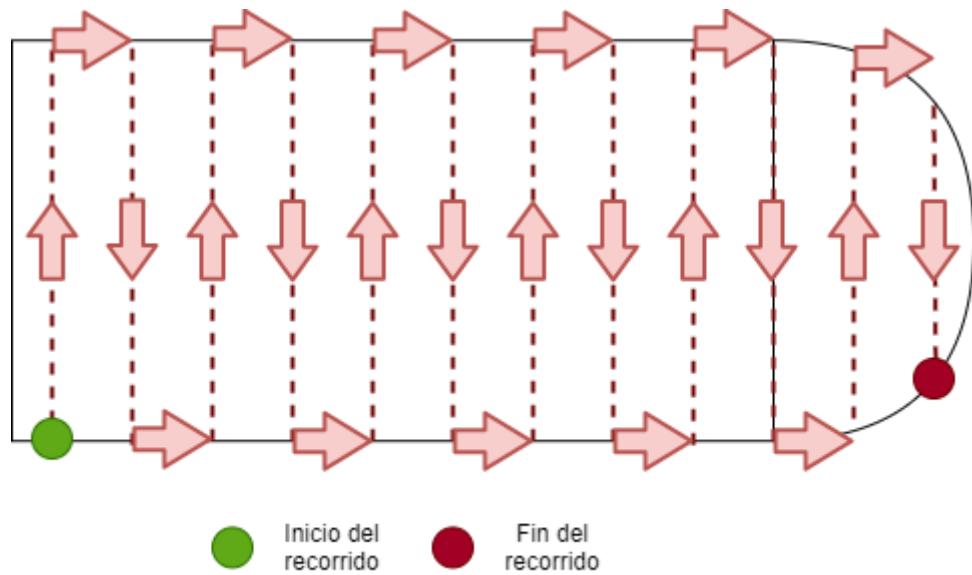


Figura 42: Diagrama del recorrido del drone por el potrero.

4.2. Trabajo de gabinete

Al finalizar el periodo de estudio en terreno, comenzó la fase de digitalización y extracción de datos para la construcción del dataset que sería utilizado para trabajar los modelos de agrupamiento. La primera tarea que se realizó en esta fase fue la digitalización de los datos obtenidos de la caracterización de cada individuo del rebaño, ya que la toma de datos utilizando lápiz y papel hacía más expedito el trabajo en terreno debido al clima y a las interacciones que se tenían con los individuos. Esto es tanto para la toma de características fenotípicas de los ovinos como en las mediciones de pasto. Ambas mediciones de características se digitalizaron en un archivo excel haciendo distinción entre las dos tomas de datos.

En cuanto a los videos, estos se analizaron y sólo se extrajo información de la cámara 1 y 2 para establecer lugares de salida y llegada al potrero. En la Figura 43 se puede ver un ejemplo y la línea referencial que se trazó para identificar el lugar de las ovejas cuando se dirigían a pastar.



Figura 43:Línea referencial para obtener los lugares de los individuos en dirección a pastar.

En la Figura 44 se puede apreciar la misma situación, pero en sentido contrario.



Figura 44: Línea referencial para obtener los lugares de los individuos en dirección al corral.

Ambas tomas se digitalizaron en el mismo archivo excel de las características fenotípicas, donde los datos fueron agrupados como la suma de veces que un individuo llegaba en los siguientes intervalos sin hacer distinción entre llegada y salida:

- ❖ Lugar 1 - 10
- ❖ Lugar 11 - 20
- ❖ Lugar 21 - 30
- ❖ Lugar 31 - 40
- ❖ Lugar 41 - 50

En cuanto a las mediciones de pasto, cada una de las tres mediciones realizadas durante el estudio se digitalizaron en un archivo excel separados por número de sector y segmento de medición para posteriormente obtener la cantidad de materia seca disponible en cada uno de los sectores y el promedio general del potrero, con el propósito de brindarle al rebaño el alimento suficiente durante el estudio.

La cantidad de materia seca, expresada en kilogramos de materia seca por hectárea (kg MS/ha) corresponde al estimado cuantitativo de la masa vegetal que posee una muestra de pasto sin considerar el agua. La disponibilidad de materia seca es dinámica y cambia permanentemente en función de las estaciones del año, el tipo de pradera y del consumo de los animales. Es por esto, que los datos de altura obtenidos con el plato medidor de praderas deben ser transformados a kilogramos de materia seca por medio de una ecuación lineal de la forma $y=ax+b$, adaptada a las condiciones propias de la pradera estudiada. En la Tabla 3 se presentan las ecuaciones lineales de calibración para el plato medidor para la zona sur de Chile en praderas permanentes según las estaciones del año (Canseco et al. 2006), donde la variable Y representa la cantidad de materia seca y la variable x corresponde al promedio de mediciones de altura comprimida con el plato medidor de praderas.

Tabla 3: Ecuaciones de calibración obtenidas con el plato para cada estación climática en las praderas permanentes del sur de Chile.

Estación	Ecuación de calibración
Invierno	$Y = 95x + 400$
Primavera	$Y = 100x + 400$
Verano	$Y = 160x + 250$
Otoño	$Y = 120x + 350$

Para la conversión de los datos de altura comprimida se utilizó la ecuación de otoño para cada uno de los 8 sectores que componen el potrero. Dadas las condiciones de la pradera, la tasa de crecimiento del forraje y la cantidad de ovinos estudiados se consideró dejar como límite inferior la cantidad de 900 kg MS/ha, es decir, la conversión de las mediciones del potrero no debía estar bajo ese valor porque en ese caso los individuos no estarían recibiendo la cantidad óptima de nutrientes durante el estudio. En la Figura 45, 46 y 47 se muestran los kilos de materia seca por hectárea al aplicar la fórmula de otoño a cada medición.

SECTOR 2 1117,948718	SECTOR 3 1385,185185	SECTOR 6 1616,666667	SECTOR 7 1613,207547
SECTOR 1 1275,641026	SECTOR 4 1303,797468	SECTOR 5 1537,804878	SECTOR 8 1684,142857

Figura 45: Kilogramos de materia seca por hectárea para la primera medición.

SECTOR 2 1164,935065	SECTOR 3 1217,283951	SECTOR 6 1383,333333	SECTOR 7 1331,914894
SECTOR 1 1219,230769	SECTOR 4 1200	SECTOR 5 1265,060241	SECTOR 8 1313,72549

Figura 46: Kilogramos de materia seca por hectárea para la segunda medición.

SECTOR 2 960,5263158	SECTOR 3 1032,142857	SECTOR 6 1060,91954	SECTOR 7 1150
SECTOR 1 1052	SECTOR 4 1008,235294	SECTOR 5 1060,91954	SECTOR 8 1238,181818

Figura 47: Kilogramos de materia seca por hectárea para la tercera medición

En la Figura 48 se puede observar una gráfica con la evolución de los sectores por medición realizada.

Medición 1, Medición 2 y Medición 3

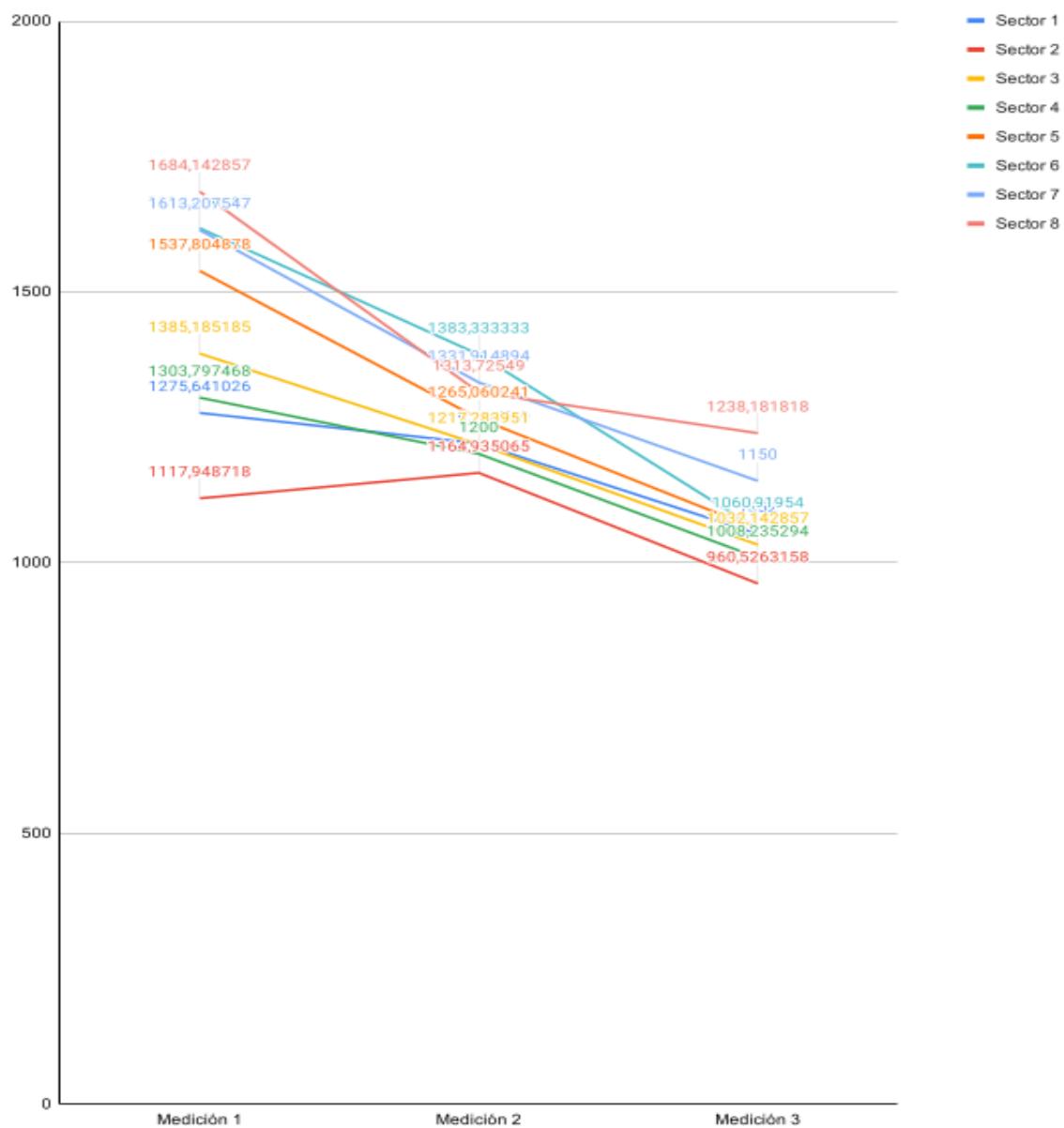


Figura 48: Gráfica de las mediciones por potrero

4.2.1. Dataset

Para generar el dataset se utilizaron los datos de caracterización de cada individuo del rebaño antes y después del estudio y los datos referentes a los lugares de salida y llegada al corral.

Finalmente, el dataset quedó compuesto como se muestra en la Tabla 4.

Tabla 4: Descripción y tipos de datos de los atributos que componen el dataset.

Atributo	Descripción	Tipo de dato
Altura	Distancia de la cruz al suelo en centímetros	Número entero (int)
Grupa	Distancia del suelo a la grupa en centímetros	Número entero (int)
dist. ojos (cm)	Distancia entre los ojos en centímetros	Número entero (int)
Dist. hocico-ojos (cm)	Distancia desde el hocico hasta la distancia media entre los ojos en centímetros	Número entero (int)
dientes	Número de dientes definitivos	Número entero (int)
delta condición	Diferencia de condición corporal durante el estudio (Condición corporal 2 - condición corporal 1)	Número real (float)
delta peso	Diferencia de peso durante el estudio (Peso 2 - Peso 1)	Número real (float)
lugar 1-10	Cantidad de veces que el individuo llegó entre los lugares 1 y 10	Número entero (int)
lugar 11-20	Cantidad de veces que el individuo llegó entre los lugares 11 y 20	Número entero (int)
lugar 21-30	Cantidad de veces que el individuo llegó entre los lugares 21 y 30	Número entero (int)
lugar 31-40	Cantidad de veces que el individuo llegó entre los lugares 31 y 40	Número entero (int)
lugar 41-50	Cantidad de veces que el individuo llegó entre los lugares 41 y 50	Número entero (int)

4.2.2. Ambiente de desarrollo

Para implementar los diferentes modelos de agrupamiento y la estadística descriptiva se utilizó el lenguaje de programación Python 3.8 debido a su facilidad de implementación y la amplia gama de librerías que ofrece en relación con data science y matemáticas. El desarrollo se realizó utilizando el intérprete IPython (*IPython*, 2007) junto con el ambiente de cómputo interactivo Jupyter. Además, se utilizaron las siguientes librerías:

- Pandas: Es una herramienta de análisis y manipulación de datos construido sobre el lenguaje de programación Python.
- Matplotlib: Es una librería completa para la generación de visualizaciones estáticas, animadas e interactivas en Python.
- NumPy: Es una librería que es usada para la creación y manipulación de matrices multidimensionales y una gran cantidad de funciones matemáticas de alto nivel.
- Scikit-learn: Es una librería de software de aprendizaje automático para el lenguaje de programación Python.
- SciPy: Es un ecosistema de software orientado a las matemáticas, ciencias e ingeniería, se compone de algoritmos y herramientas matemáticas para Python.

4.2.3. Análisis estadístico

Para estudiar el comportamiento de los datos se utilizaron diferentes medidas de resumen como tablas y gráficos. En la Tabla 5 se puede observar un resumen de los estadísticos como media, desviación estándar, valor máximo y valor mínimo para cada característica fenotípica que compone el dataset.

Tabla 5: Estadísticos de las características fenotípicas de los individuos del rebaño

Característica	Media	Desviación estándar	Valor máximo	Valor mínimo
Condición corporal inicial	3.1	0.404061	4	2.5
Condición corporal final	3.09	0.436708	4	2
Peso inicial	75.788	14.354017	104	42.7
Peso final	80.188	14.486061	105	46.5
Altura	71.7	5.10078	83	58
Grupa	74.78	4.35323	82	65
distancia hocico-ojos	20.08	1.45462	22	16
Distancia entre ojos	12.16	0.865672	14	10

En la Figura 49 se puede ver la distribución de edad según la cantidad de dientes definitivos de cada individuo.

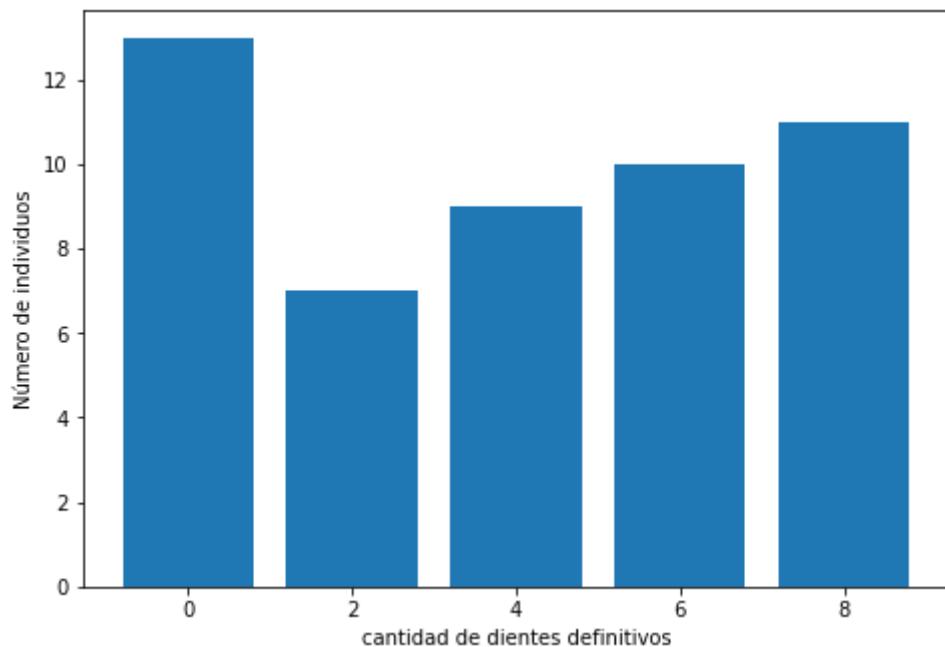


Figura 49: Gráfica de distribución de la edad según el número de dientes definitivos.

En la Figura 50 se puede observar la distribución de la condición corporal al comenzar el estudio. La condición corporal predominante del rebaño se encuentra entre 2.5 y 3.0, con un total de 38 individuos equivalente al 76%. Además, se evidencia que sólo un 24% de los individuos se encuentra en el rango 3.5 y 4.0 de condición corporal antes del estudio.

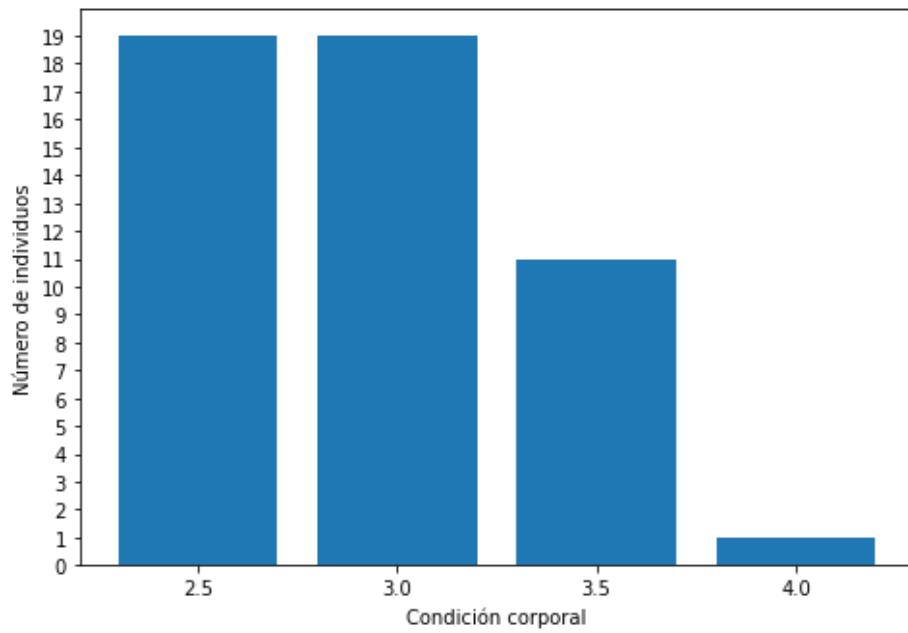


Figura 50: Gráfica de distribución de la condición corporal al comenzar el estudio.

Por otra parte, en la Figura 51 se puede ver la distribución de condición corporal al culminar el estudio, el cual evidencia una baja considerable en la condición corporal de los individuos ya que un 46% del total de los individuos bajaron a 2.0 (condición corporal que antes del estudio no se evidenciaba).

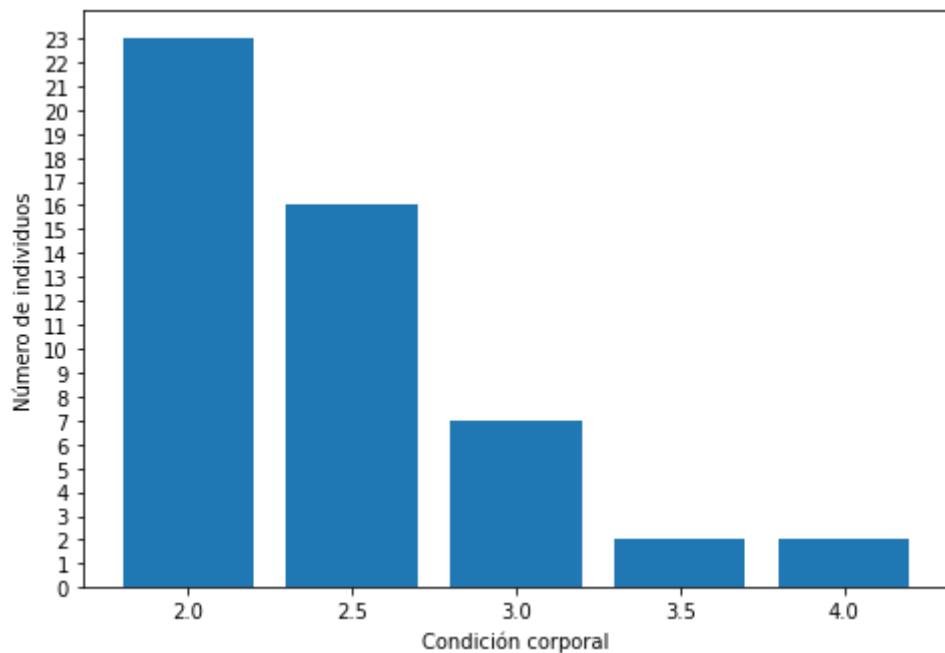


Figura 51: Gráfica de distribución de la condición corporal al finalizar el estudio.

También, se realizó una matriz de correlación para el dataset como se ve en la Figura 52, para esta matriz se utilizó el coeficiente de correlación de Pearson que nos entrega la relación lineal existente entre dos variables cuantitativas.

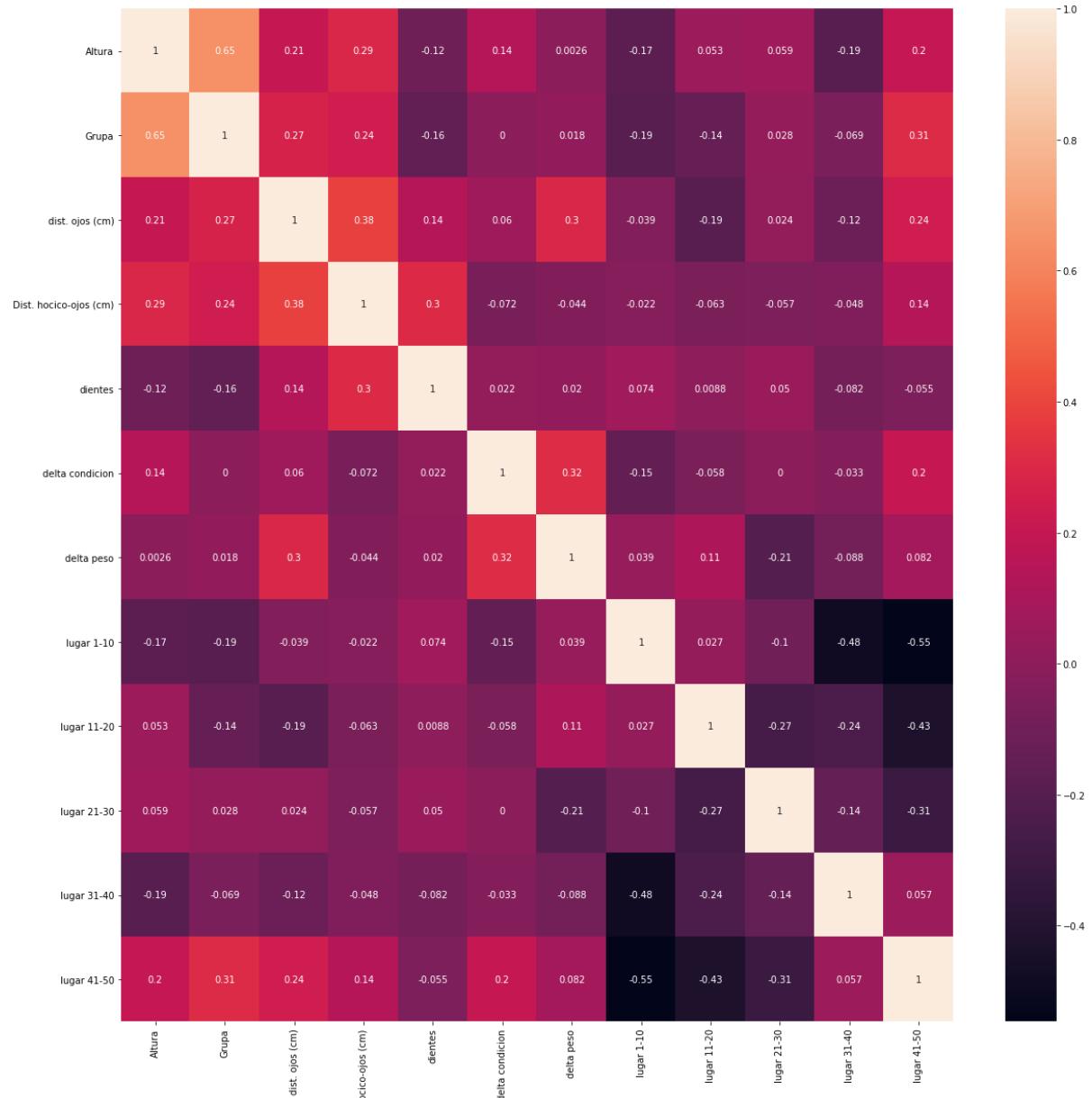


Figura 52: Matriz de correlación

4.2.4. Reducción de dimensionalidad

Para comenzar a aplicar el Análisis de Componentes Principales y los métodos de agrupamiento lo primero que se realizó fue la normalización de los datos, esto consiste en aplicar una transformación a los datos para que la media sea igual a 0 y su desviación estándar igual a 1, con la finalidad de que todos los atributos de los dataset queden en los mismos rangos.

Con esto se evita que atributos que estén en una escala mayor tengan más influencia que atributos que tengan una escala menor.

En la Tabla 6 se puede observar una comparativa del porcentaje de varianza capturado al aplicar PCA al dataset.

Tabla 6: Porcentaje de varianza según número de componentes principales al aplicar PCA.

Nº de componentes principales	Dataset
2 componentes	35.5%
3 componentes	47.9%
4 componentes	59.5%
5 componentes	69.4%
6 componentes	77.7%

De la Tabla 6 se puede concluir que implementar una reducción de dimensionalidad no es conveniente, ya que el dataset está compuesto por pocos atributos. Además, la varianza descrita por las componentes principales en el mejor de los casos representa un poco más del 50% de la varianza que poseen los datos originales. Además, con cada aumento en el número de componentes principales el aporte de varianza del componente se va reduciendo considerablemente.

Sin embargo, se puede usar la reducción de dimensionalidad para graficar en dos dimensiones los resultados con los grupos resultantes al aplicar los métodos de agrupamiento y las proyecciones de los ejes para comprender de manera gráfica los resultados y sus tendencias.

4.3. Implementación de modelos de agrupamiento

Se utilizaron los modelos de agrupamiento K-means, agrupamiento jerárquico aglomerativo y GMM para agrupar los datos en base a patrones de similitud, las cuales pueden resultar de diferentes formas dependiendo del modelo utilizado.

4.3.1. Implementación y programación del modelo K-means

Para aplicar el método K-means es esencial establecer el número de grupos o clusters a considerar en el estudio. En la Figura 53 podemos ver una gráfica que nos muestra la tendencia de la inercia con distintos valores de K , lo que se conoce como método del codo.

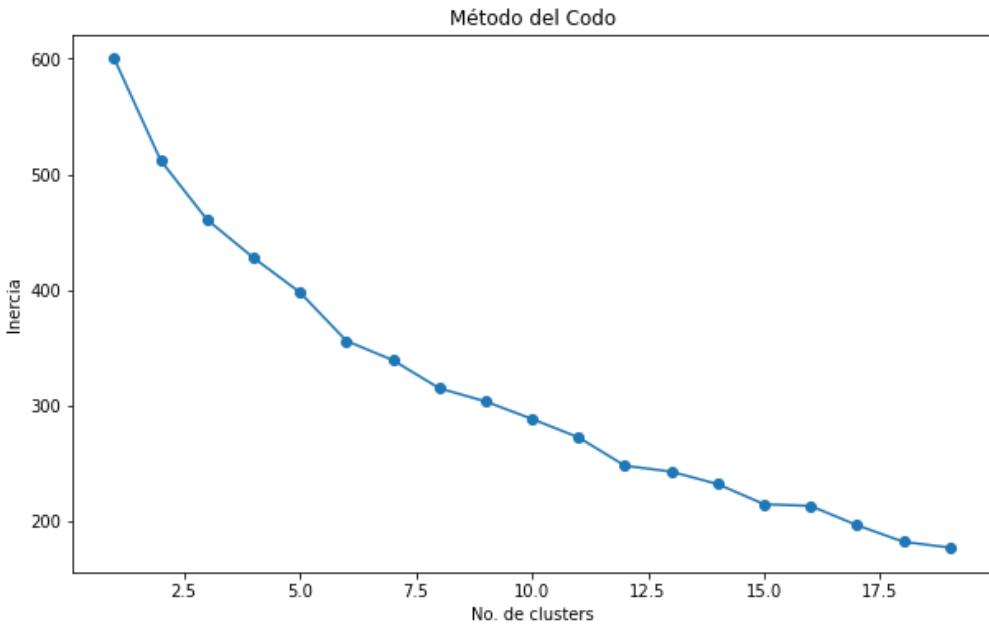


Figura 53: Método del codo.

Este método nos da nociones sobre qué valor de K utilizar dependiendo de donde se encuentre el codo o qué valor de K creamos que es conveniente para nuestro problema, en nuestro caso utilizaremos K=3 porque según nuestra problema nos interesa identificar a 3 grupos.

A continuación, en la Figura 54 se muestra la implementación del método K-means, donde se fijó el número de clusters en 3 con un total de 30000 iteraciones que se ejecutarán 2000 veces con distintas semillas de centroides con una tolerancia de 1e-6.

```

1 kmeans = KMeans(n_clusters=3, max_iter=30000, n_init=2000,
2                      tol=1e-6).fit(ovejas)
3 arr = [m for m in range(1,51)]
4 centroids = kmeans.cluster_centers_
5 labels = kmeans.labels_
6 print(labels)
7 print(centroids)
8 pred_y = kmeans.predict(ovejas)

```

Figura 54: Implementación del modelo K-means en python

En la Tabla 7 se muestra la distribución de individuos por agrupamiento al aplicar el modelo K-means.

Tabla 7: Distribución de individuos por grupo.

Grupo	Número de individuos
1	9
2	23
3	18

4.3.2. Implementación y programación del modelo de Agrupamiento jerárquico.

Para implementar el cluster jerárquico se utilizó el agrupamiento aglomerativo con el método ward de varianzas mínimas utilizando la métrica euclíadiana de distancia por pares. El dendrograma resultante se puede ver en la Figura 55.

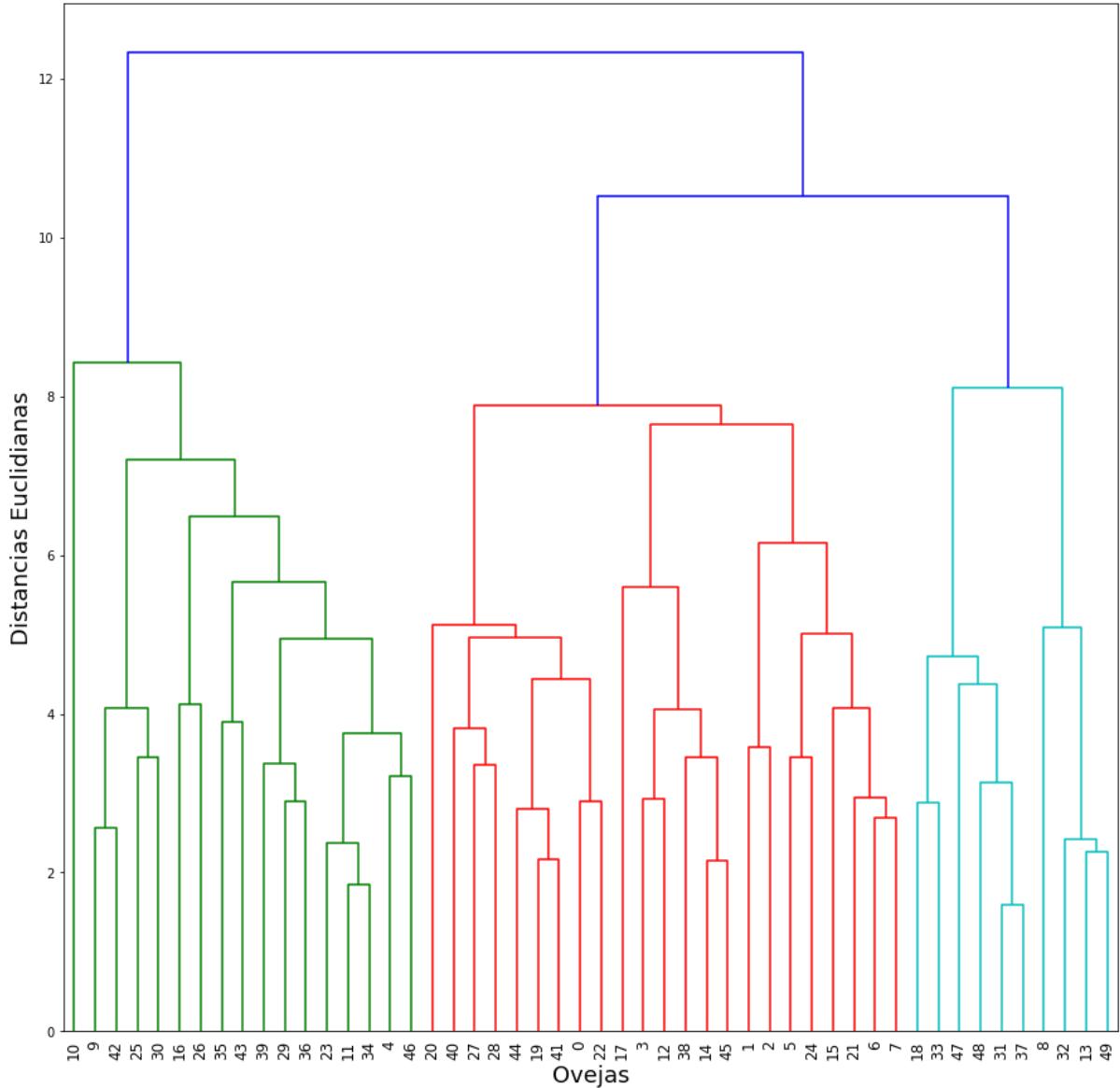


Figura 55: Dendrograma Agrupamiento aglomerativo.

Como se puede ver en la Figura 55 el dendrograma comienza con 50 grupos, es decir, en principio cada individuo es un grupo en sí mismo y se van fusionando con otros cluster a medida que las distancias euclidianas y la varianza van siendo mínimas hasta llegar a componer un único grupo.

En la Figura 56 se muestra la implementación del agrupamiento jerárquico con $n_clusters = 3$ con la finalidad de que el resultado sea comparable con el de K-means.

```

1 | hc = AgglomerativeClustering(n_clusters = 3,
2 |                               affinity = 'euclidean',
3 |                               linkage = 'ward')
4 | y_hc = hc.fit_predict(data_cj)

```

Figura 56: Implementación del modelo de agrupamiento jerárquico aglomerativo en python

En la Tabla 8 se observa la distribución de individuos por cluster al aplicar agrupamiento jerárquico aglomerativo.

Tabla 8: Distribución de individuos por grupo en Agrupamiento jerárquico

Grupo	Número de individuos
1	23
2	10
3	17

4.3.3. Implementación y programación del modelo GMM

Para el Modelo Mezcla Gaussiana también se normalizaron los datos, a pesar de que para este modelo no era necesario ya que trabaja en base a probabilidades de pertenencia a los grupos y no en base a medidas de distancia para encontrar agrupamientos. Sin embargo, esto se hace para fines netamente de convergencia del modelo.

En la Figura 57 se puede observar la implementación del modelo GMM con $n_components=3$ y un número de inicializaciones igual a 10, de la cual el modelo se queda con la mejor.

```

1 | gmm = GaussianMixture(n_components=3, n_init=10).fit(x)
2 | labels = gmm.predict(x)

```

Figura 57: Implementación del modelo GMM en python

la variable labels contiene la etiqueta del grupo al cual pertenece cada individuo en base a la mayor probabilidad obtenida del modelo con la función $gmm.predict_proba(x)$ que nos entrega una matriz de probabilidades de pertenencia de cada individuo a los grupos. En la Tabla 9 observamos la distribución de individuos por grupo al aplicar el modelo GMM.

Tabla 9: Distribución de individuos por grupo en GMM

Grupo	Número de individuos
1	19
2	7
3	24

5. RESULTADOS Y DISCUSIÓN

En esta sección se analizan y comparan los diferentes resultados obtenidos por cada modelo de agrupamiento aplicado; así como también la validación de los resultados en base al comportamiento de los individuos más representativos de cada grupo en situaciones de pastoreo y de la posición que ocupan los integrantes del rebaño cuando son expuestos a condiciones de predación, con una situación simulada del ataque de un perro asilvestrado.

5.1. Análisis y comparativa de modelos de agrupamiento

En la Figura 58 se pueden observar de manera gráfica los resultados cuando se aplicó el modelo k-means al dataset, según las proyecciones de los ejes.

El grupo de animales bajo el color rojo, que resultaron del modelo de agrupamiento corresponden a los individuos dominantes, ya que estos se diferenciaron del resto del rebaño debido a el 95% de las veces se ubicaban a la cabeza del rebaño al momento de ser arreados desde y hacia el potrero. Además, estos se caracterizan principalmente por ser de mayor edad, donde un 73,9% de este grupo tiene más de 18 meses de vida y un 39% de estos lo componen ovejas viejas (< 3 años). A pesar de que el grupo está compuesto en su mayoría por ovejas de mayor edad, también se pudieron encontrar 3 borregas que se caracterizaron por ser las hembras que bajaron entre 0.5 y 1.0 su condición corporal durante el estudio.

Con respecto al grupo de animales que quedan agrupados en la gráfica bajo el color azul, se caracterizan principalmente por ser de menor edad. En este aspecto, un 88.9% de los individuos son borregas, es decir, tienen menos de un año de vida. Además, se caracterizan por ser más pequeños y menos pesados y haber subido menos su peso vivo o incluso haber bajado durante el estudio. Otro aspecto para tener en consideración es que los individuos de este grupo en su mayoría llegaron más veces en el rango de lugares 31 – 40 cuando eran arreados desde y hacia el potrero.

Finalmente, el grupo verde se caracteriza en su mayoría, por ser ovejas corpulentas (mayor peso y altura) que durante el estudio subieron o mantuvieron su condición corporal. Además, en este grupo hay presencia equilibrada de individuos con respecto a la edad, a excepción de dos ovejas viejas que se caracterizan por tener un peso más elevado que sus congéneres de la misma edad. Este último grupo se considera el menos dominante, ya que son las últimas del rebaño.

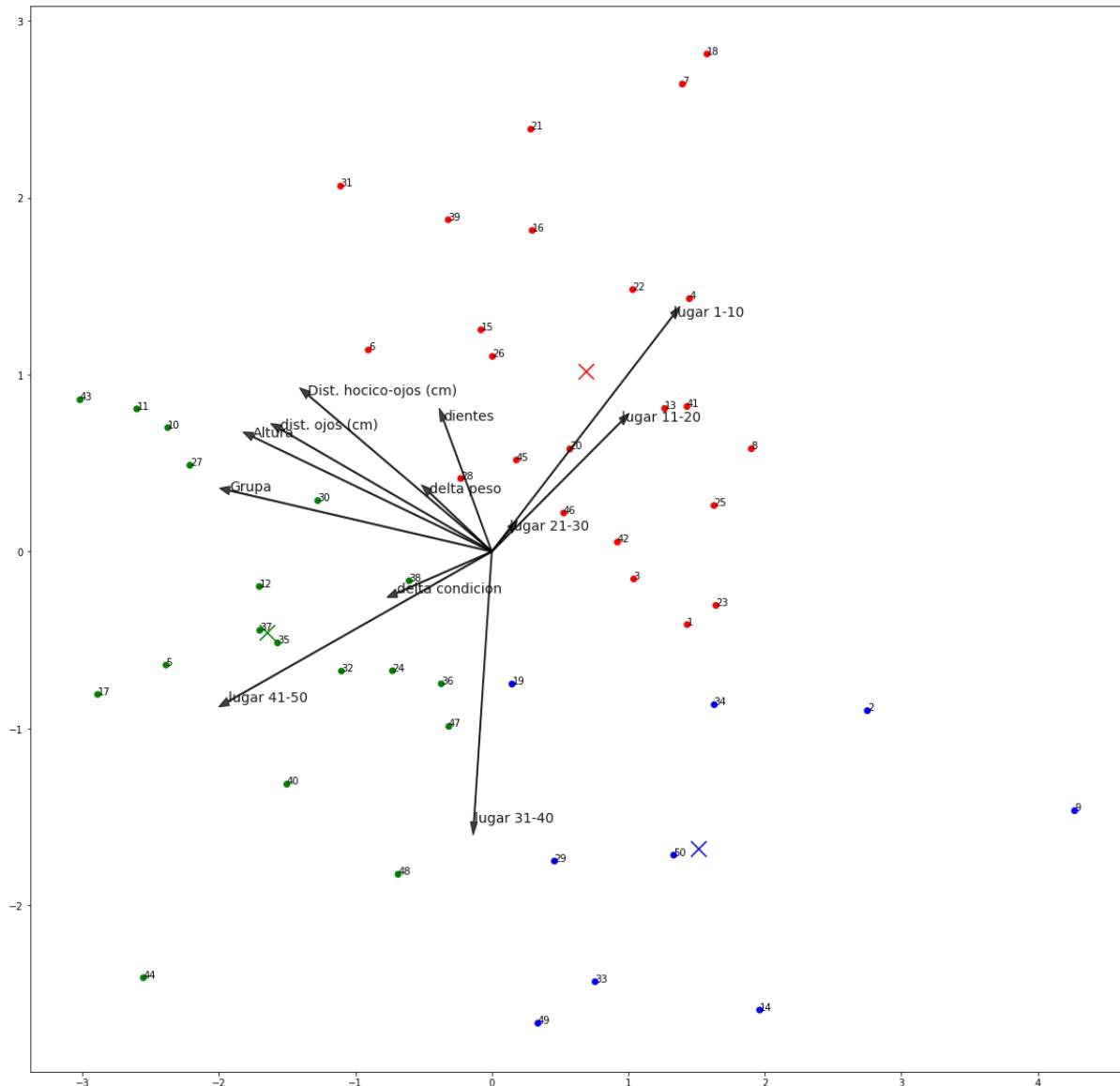


Figura 58: Agrupamiento K-means

En la Figura 59 se pueden observar de manera gráfica los resultados cuando se aplicó el modelo de agrupamiento jerárquico aglomerativo al dataset, según las proyecciones de los ejes.

El grupo de ovejas que quedó bajo el color verde quedó constituido por 9 de las 11 ovejas viejas del rebaño. Además, 60% de los individuos del grupo se caracterizaron por posicionarse entre los primeros 30 lugares del rebaño. Sin embargo, el grupo también consideró un 40% de individuos que llegaron en lugares más altos.

El grupo constituido bajo el color rojo quedó conformado en un 77% por borregas pequeñas que aumentaron poco más de 3 kilos su peso vivo. Un 90% de las ovejas que componen este grupo llegaron entre los lugares 31 a 40 en el arreo desde y hacia el corral.

El grupo constituido bajo el color azul quedó conformado principalmente por ovejas que en el arreo desde y hacia el potrero ocupaban los lugares 41 a 50. Además, mantuvieron una condición corporal que se mantuvo o subió 0.5 puntos con relación a la primera medición de características fenotípicas. También se evidenció que un 24% de los ovejas de este grupo tienen una altura de grupa y cruz mayor en comparación a otros individuos del mismo grupo.

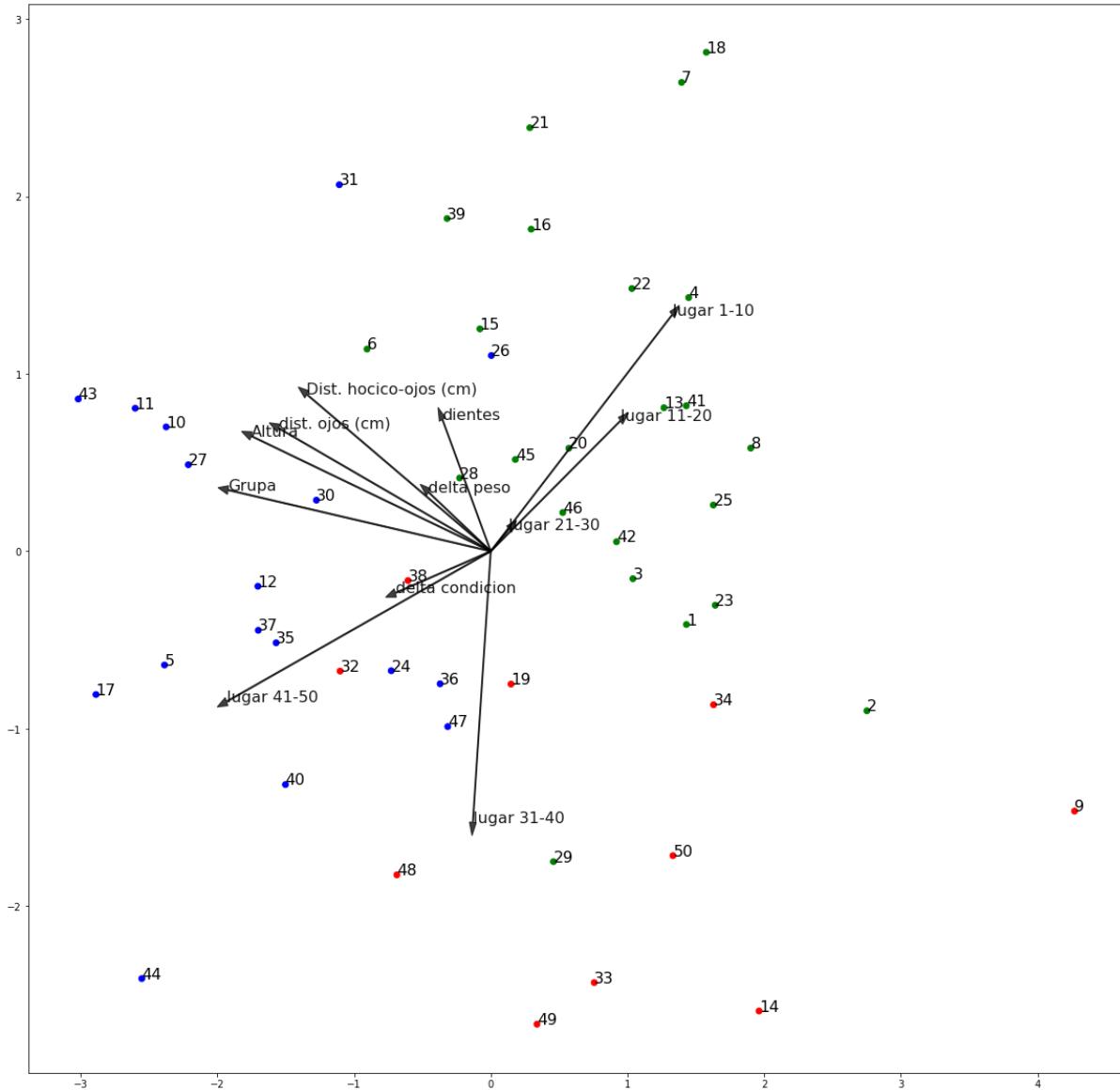


Figura 59: Agrupamiento jerárquico aglomerativo

En la Figura 60 se pueden observar de manera gráfica los resultados cuando se aplicó el modelo de GMM al dataset, según las proyecciones de los ejes.

Con respecto al grupo de animales que quedaron agrupados bajo el color rojo, se puede observar que se encuentra disperso y no sigue un patrón claro de agrupamiento jerárquico, un ejemplo de ello es que por ejemplo el individuo 19 pertenece al grupo rojo pero tiene características muy distintas a otros individuos del mismo agrupamiento, entre ellas la edad porque es una borrega y los elementos del grupo están en un 80% constituido por individuos de mayor edad (< 3 años), y el delta de condición corporal que indica que el individuo 19 es más

corpulento que sus pares del mismo grupo. Una situación similar se da con elementos del mismo agrupamiento.

En relación con el agrupamiento de individuos que fueron agrupados bajo el grupo azul, se caracterizan por estar compuesto en un 65% por ovejas jóvenes y pequeñas con una variación nula o negativa de peso (oscilación entre -3 y 0).

Finalmente, el grupo compuesto por ovejas bajo la agrupación verde al igual que el rojo se encuentra disperso con individuos de todas las edades, con variaciones positivas y negativas del peso vivo durante el estudio. Además, las ovejas que componen este grupo están ubicadas en distintos puntos del plano de proyecciones de lugares, tanto de los últimas como de las primeras en el orden de llegada del arreo desde y hacia el potrero.

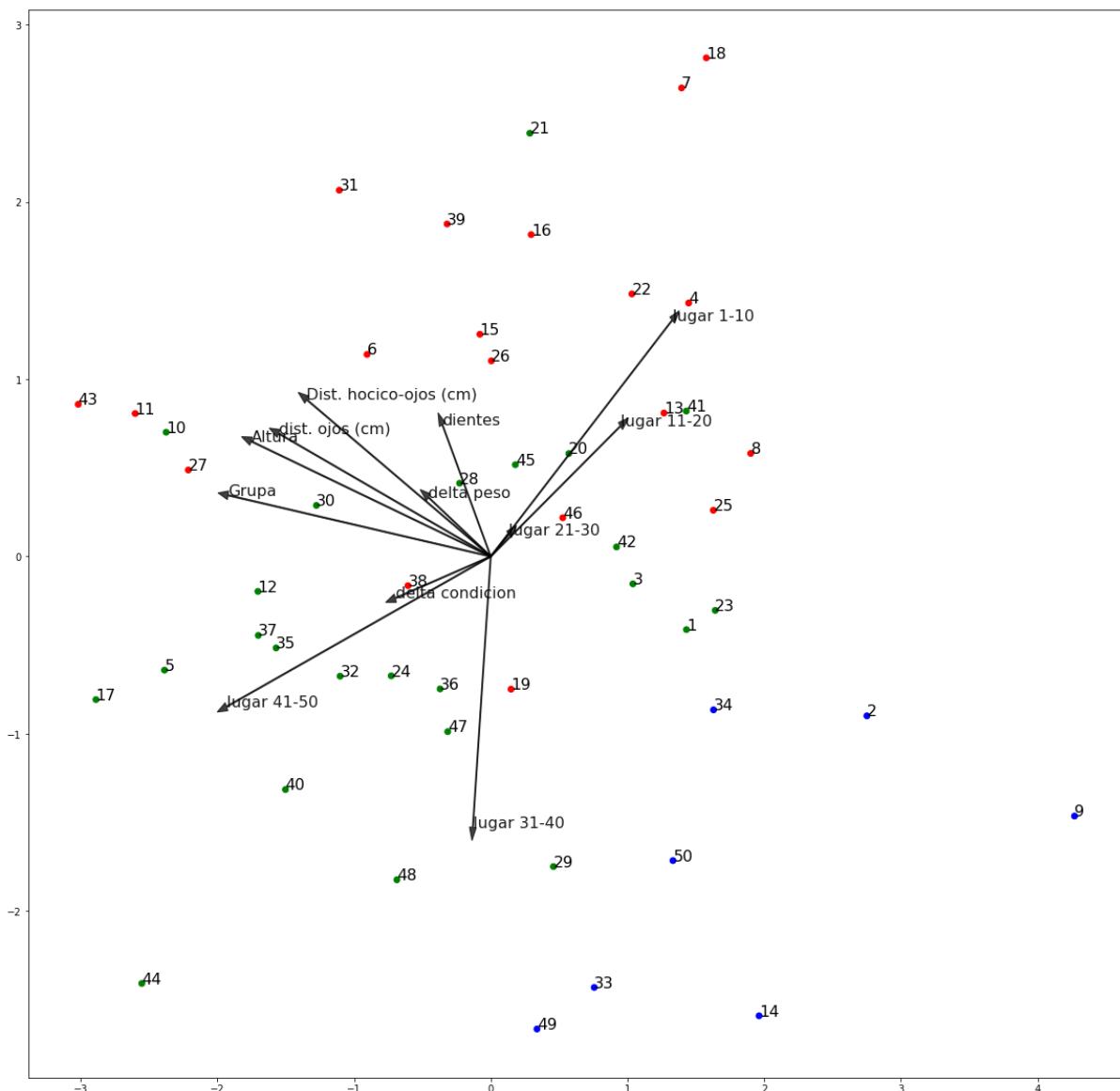


Figura 60: Agrupamiento GMM.

Es importante tener en cuenta que, dado el carácter exploratorio del estudio, sólo se analizaron los tres modelos más utilizados en fases exploratorias que dada la naturaleza de los datos pudiesen dar mejores resultados, como lo son k-means, agrupamiento jerárquico y GMM.

Luego de haber analizado los grupos resultantes de cada modelo se llegó a la conclusión de que k-means es el modelo que entrega el mejor agrupamiento, es más equilibrado en cuanto a número de individuos por cluster y los individuos de un cluster tienen más características en común que en GMM y agrupamiento jerárquico. Sin embargo, se puede notar una tendencia en los tres modelos, ya que el lugar más denso de cada grupo coincide en todos los modelos pero difieren en los bordes, esto se debe al tipo de agrupamiento y la flexibilidad de cada modelo, por ejemplo en el caso de k-means los agrupamientos son circulares mientras que para GMM este agrupamiento es más flexible, pudiendo ser capaz de agrupar de forma elíptica los datos.

En general, las características que más influyen y diferencian un cluster de otro, en los agrupamientos realizados en los tres métodos son el lugar que ocupa el animal cuando son llevados o traídos desde el potrero, la edad y la condición corporal. Mientras que las variables morfométricas, no convencionales como distancia entre los ojos y distancia hocico-ojos no tienen relación directa con los niveles jerárquicos encontrados porque según la matriz de correlaciones, estas características poseen una correlación cercana a 0 con respecto a los demás atributos fenotípicos. Además, en la gráfica de proyecciones con dos componentes principales se puede observar que el ángulo que forman las características no convencionales es cercano a 90° con respecto a las demás características fenotípicas, lo que indica que no están relacionadas con el establecimiento de la estructura jerárquica del rebaño de ovinos domésticos.

5.2. Validación y evaluación del modelo en condiciones de pastoreo y predación

La validación de los grupos resultantes del modelo del cual se obtuvo un mejor agrupamiento según las características entregadas, fue el algoritmo k-means. Para ello se analizó el comportamiento de los individuos más representativos de cada grupo en la situación de pastoreo y predación, con la finalidad aceptar o rechazar la hipótesis planteada para el estudio. En la Tabla 10 se pueden ver el número de los individuos seleccionados para la validación según sus grupos.

Tabla 10: Individuos seleccionados por cada grupo del modelo K- means

Grupo	Nº de individuos
1 (Rojo)	22 y 20
2 (Azul)	50 y 29
3 (Verde)	37 y 35

5.2.1. Pastoreo

En el pastoreo no se observaron grandes diferencias entre los grupos, ya que se trataba de un potrero de grandes dimensiones con abundante presencia de materia seca, que no daba paso a la competencia por alimento. Es por estas razones que no se pudieron observar diferencias de comportamiento entre grupos de jerarquía en el rebaño. Sin embargo, se pudo observar que el comportamiento del rebaño en sí variaba dependiendo del momento del día en el que se realizaban las tomas con el drone. En la Figura 61 se puede observar el potrero y sus sectores para entender mejor el comportamiento del rebaño durante el pastoreo y en la Figura 62 se puede observar un mapa de calor según la disponibilidad de materia seca con las imágenes del drone captadas a 100 metros de altura.

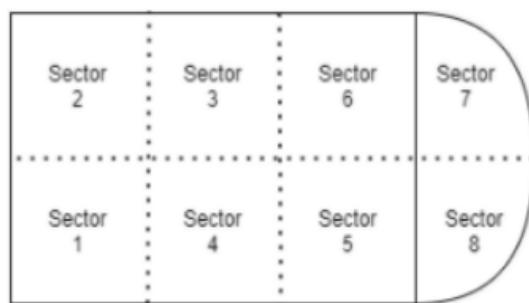


Figura 61: Potrero por sectores.

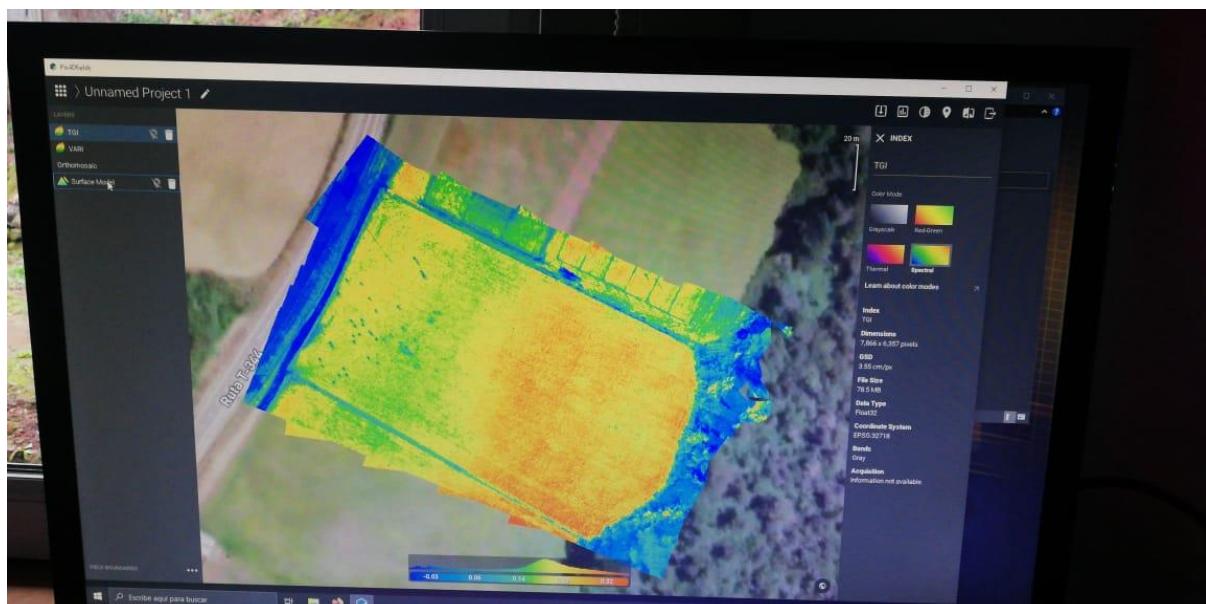


Figura 62: Mapa de calor de la disponibilidad de materia seca del potrero 3.

En las mañanas el rebaño completo se dirigía hacia los sectores 5, 6, 7 y 8 que correspondían a los más alejados y las que tenían mayor abundancia de materia seca. Para en la tarde trasladarse a los sectores 1, 2, 3 y 4 que eran los más cercanos a la entrada del potrero.

Si bien, podemos evidenciar una tendencia de pastoreo en el comportamiento del rebaño gracias a las tomas hechas por el drone, no podemos garantizar que los individuos dominantes pertenecientes al primer grupo de k-means hayan accedido a las mejores fuentes de alimento.

5.1.2. Predación

Para este estudio, se realizaron tomas con el drone simulando una predación ficticia en un ambiente controlado con un perro de raza Pastor Suizo con el cual las ovejas no estaban familiarizadas con la finalidad de que su comportamiento sea lo más parecido posible a una predación real. A continuación, se presenta el análisis del comportamiento de los individuos de cada grupo:

- Grupo 1: Al realizarle seguimiento a los individuos 20 y 22 del grupo 1 durante la predación, se pudo establecer que son individuos que durante los ataques se mantienen alerta en los exteriores del rebaño. Normalmente forman la primera línea frente al ataque del predador y alertan al resto de los miembros del grupo en caso de peligro. En la Figura 63 se pueden observar los individuos y su posición dentro del rebaño frente a un ataque.



Figura 63: Individuos 20 y 22 frente a la predación.

- Grupo 2: El comportamiento de los individuos seleccionados del grupo 2 es muy diferente de los individuos del grupo 1, por una parte en la Figura 64 se puede observar que la oveja 29 se mantiene lo más lejos posible del predador y sigue pastando porque se siente fuera de peligro. Mientras que la oveja 50 que se puede apreciar en la Figura 65 se mantiene alerta, pero ubicada siempre en el centro del rebaño.



Figura 64: Individuo 29 frente a la predación.



Figura 65: Individuo 50 frente a la predación.

- Grupo 3: Al analizar el comportamientos de los individuos seleccionados del grupo 3, se puede ver en la Figura 66 que se encuentran al final del rebaño, alertas y lejos del predador. La totalidad del tiempo que dura el ataque mantienen esa posición.



Figura 66: Individuos 35 y 37 frente a la predación.

Se puede concluir que en base a la validación, existe una relación de jerarquía clara en situaciones de predación. Sin embargo, no se puede concluir lo mismo en situaciones de pastoreo, ya que no se evidencia una tendencia de comportamiento individual en base a la jerarquía hallada.

6. CONCLUSIONES

El presente estudio bajo las condiciones mencionadas permite concluir que:

Se logró recolectar suficiente información para generar un conjunto de datos basado en características fenotípicas y comportamiento de ovinos domésticos en condiciones de arreo y pastoreo.

Se pudieron estudiar métodos y algoritmos basados en machine learning que permitieron cuantificar y asignar valoración al dataset para determinar liderazgo en diferentes situaciones de estudio como pastoreo, arreo y ataque de predadores.

Se logró implementar los algoritmos estudiados y ajustar los modelos al dataset construido mediante observaciones de atributos convencionales y no convencionales, además de patrones de comportamiento establecidos durante el arreo desde y hacia el potrero.

Finalmente se logró evaluar las agrupaciones obtenidas por el método de agrupamiento K-means que demostró ser el modelo que más se ajustaba a los datos y logró un mejor desempeño en el agrupamiento de los individuos por nivel de jerarquía.

La validación del modelo arrojó que no se evidenció una fuerte estructura jerárquica durante el pastoreo. Sin embargo, se observó un patrón de movimiento del rebaño durante el día, ya que en la mañana el rebaño se dirigía a los sectores 5, 6, 7 y 8 para comer y al finalizar el día, las ovejas se trasladaban a los sectores más próximos a la salida del potrero (sectores 1 y 2).

En cuanto a las validaciones en situaciones de predación, fue posible determinar una fuerte estructura jerárquica que coincidía con lo establecido por el modelo K-means, ya que los animales agrupados por el modelo en la cúspide de la pirámide jerárquica confrontaban y alertaban al resto del rebaño, ubicándose próximas a las periferias del rebaño en dirección al predador. Por otro lado, el grupo identificado como el eslabón más bajo de la pirámide según el modelo fue el que se mantuvo más alejado del predador, ubicándose en el espacio perimetral opuesto de los individuos del eslabón más alto. Además, los individuos que se encontraban en el eslabón medio de la pirámide jerárquica construida a partir del modelo de agrupamiento se mantuvieron en el espacio entre el ataque y las ovejas que defendían el rebaño.

A partir de lo expuesto anteriormente, se puede concluir que la agrupación realizada a partir del método de agrupamiento K-means en base a atributos fenotípicos y comportamiento en arreo permite establecer la estructura jerárquica dentro de un rebaño de ovinos domésticos.

6.1. Trabajos futuros

Determinar condiciones fisiológicas (enfermedad) estudiando a los individuos que llegan últimos.

Esta línea de investigación resulta interesante porque hace uso de cámaras y detección de patrones en los videos para la identificación de individuos que presenten alguna enfermedad que sea visible a este tipo de dispositivos, como por ejemplo la cojera, que es muy común en los ovinos. Esto permitiría que se hiciera una detección temprana de anomalías y se pueda actuar de manera oportuna.

Hacer uso de drones para medir la disponibilidad de materia seca en los potreros.

Este trabajo sería interesante porque permitiría hacer un monitoreo de las zonas de pastoreo mucho más expeditas y sin la necesidad de pasar horas haciendo mediciones manuales con el plato medidor de forraje o utilizando otras técnicas igual de engorrosas.

7. REFERENCIAS

Amat, J. (2017). *Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE.*

Amat, J. (2017). *Clustering y heatmaps: aprendizaje no supervisado.* Obtenido de https://www.cienciadatos.net/documentos/37_clustering_y_heatmaps

Canseco, C., Dermanet, R., Balocchi, O., Parga, J., Anwandter, V., Abarzúa, A., Teuber, N., Lopetegui, J. (2006) *Determinación de la disponibilidad de materia seca de praderas en pastoreo.*

Díaz, P. (2018). *Aplicación De Mezclas Finitas Para Segmentación De Clientes De Una Entidad Bancaria (CRM).*

Espejo, C. (2015). *Atributos fenotípicos v/s contexto social. ¿Cómo se establece la posición jerárquica en ovinos domésticos?*

Fernando Pérez, Brian E. Granger (2007), *IPython: A System for Interactive Scientific Computing in Science and Engineering.* <https://ipython.org/>

Galindo, FA. (2004). *Introducción a la etología aplicada.*

Gonzalez, L (2020). *Aprendizaje no supervisado.* Obtenido de <https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/>

IArtificial (2020). *Clustering (Agrupamiento), K-Means con ejemplos en python.* Obtenido de <https://www.iartificial.net/clustering-agrupamiento-kmeans-ejemplos-en-python/>

INE (2007). *Cambios estructurales en la agricultura Chilena. Análisis intercensal 1976 -1997 - 2007*

Lynch, J.J., Hinch,G.N.,Adams, D.B (1992). *The Behaviour of Sheep: Biological Principles and Implications for Production.*

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Accedido el 02 de Enero, 2020, desde <https://www.minciencia.gob.cl/consultaIA>

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective.*

Oriella, R (2015). *Evaluación de la condición corporal y edad de los ovinos.*

Owoeye, K., Hailes, S (2018). *Online Collective Animal Movement Activity Recognition*

Pérez, I., Arco, L. (2016). *Una revisión sobre aprendizaje no supervisado de métricas de distancia.*

Quichiyao, R. (2003). *Efecto de la suplementación con concentrado sobre la respuesta productiva y comportamiento ingestivo de oveja Latxas en lactancia en pastoreo continuo controlado.*

Rivas, A., Chamoso, P., González, B., Corchado., J (2018). *Detection of Cattle Using Drones and Convolutional Neural Networks.*

Sáenz, A. (2007). *OVINOS Y CAPRINOS*

Sammut, C., Webb, G. (2011). *Encyclopedia of machine learning.* Springer Science & Business Media.

[Scikit-learn: Machine Learning in Python](#), Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011

Shalev-Shwartz, S., Ben-David, S. (2014). *Understanding Machine Learning From Theory to Algorithms.*

Shlens, J. (2005). *A Tutorial on Principal Component Analysis.*
<https://projects.ics.forth.gr/mobile/pca.pdf>

Torres, R. (2019). *Políticas Públicas para la Robótica y la Inteligencia Artificial*

Ungerfeld R, González Pensado SP. (2008). *Social rank affects reproductive development in males lambs.*

Ungerfeld, R., Núñez, M.L. (2011). *Jerarquía y dominancia en grupos de carneros: establecimiento y efectos sobre la reproducción.*

Van der Maaten, L., Postma, E. y Van den Herik, H. (2009). *Dimensionality Reduction: A Comparative Review.*