



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



Instituto Tecnológico de Ciudad Guzmán



**INSTITUTO TECNOLÓGICO DE CD. GUZMÁN**

**TESIS**

**MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

**TEMA:**

**ESTIMACIÓN DE GRADOS BRUX EN CULTIVOS  
DE AGAVE UTILIZANDO IMÁGENES  
ADQUIRIDAS DESDE UN VEHÍCULO AÉREO NO  
TRIPULADO**

**QUE PARA OBTENER EL TÍTULO DE:  
MAESTRA EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:**

**ING. FRIDA MAYELA FLORIÁN PINTO**

**DIRECTORES:**

**DRA. MARÍA GUADALUPE SÁNCHEZ CERVANTES  
DR. HIMER ÁVILA GEORGE**

**CD. GUZMÁN JALISCO, MÉXICO, AGOSTO DE 2022**



Instituto Tecnológico de Ciudad Guzmán  
SUBDIRECCIÓN ACADÉMICA

Ciudad Guzmán, Jalisco, **15/agosto/2022**  
**Asunto:** Autorización de impresión de Tesis

**ING. FRIDA MAYELA FLORIÁN PINTO**  
**CANDIDATA AL GRADO DE MAESTRA EN CIENCIAS DE LA COMPUTACIÓN**  
**PRESENTE**

De acuerdo con los Lineamientos para la Operación de los Estudios de Posgrado en el Tecnológico Nacional de México y las disposiciones en este Instituto, habiendo cumplido con todas las indicaciones que la Comisión Revisora realizó con respecto a su Trabajo de Tesis titulado "**Estimación de grados Brix en cultivos de agave utilizando imágenes adquiridas desde un Vehículo Aéreo No Tripulado**", la División de Estudios de Posgrado e Investigación de este Instituto, concede la Autorización para que proceda a la impresión del mismo.

Sin otro particular, quedo de Usted.

**ATENTAMENTE**

*Excelencia en Educación Tecnológica*

"2022, AÑO DEL CINCUENTA ANIVERSARIO DEL INSTITUTO TECNOLÓGICO DE CIUDAD GUZMÁN"

**CARLOS RUVALCABA MÁRQUEZ**  
**SUBDIRECTOR ACADÉMICO**

ccp. Archivo  
CRM/MGSS/megg



<https://cdguzman.tecnm.mx/sgcicg/>

Av. Tecnológico #100 C.P. 49100 Ciudad Guzmán, Jal. Tel. (341) 5752050

tecnm.mx | cdguzman.tecnm.mx



**2022 Flores**  
Año de Magón  
PRELUSOR DE LA REVOLUCIÓN MEXICANA

# Resumen

El *Agave tequilana* Weber variedad azul es una planta xerófila de hojas azul verdoso con una punta en forma terminal y una cabeza mejor conocida como piña de agave, la cual es rica en contenido de fructanos. El proceso desde la siembra hasta la cosecha de un *Agave tequilana* tiende a ser un proceso largo y costoso, ya que deberán de pasar entre 6 a 8 años para que éste pueda ser jimado. Es importante que un agave cosechado cumpla con el peso y el nivel de azúcares adecuados, el nivel de azúcar contenido se mide a través de la escala de grados Brix, siendo éste la materia prima principal en la elaboración del tequila, cabe mencionar la importancia de la industria tequilera para la economía del país.

Se conoce que en la actualidad, la tecnología ha ayudado a mejorar diversos procesos productivos, tal es el caso del sector agrícola. Las diversas herramientas tecnológicas que actualmente se ofrecen, permiten hacer de las labores cotidianas un proceso más ágil, preciso y en cierto punto económico. El uso de VANTS (Vehículos Aéreos No Tripulados) en la agricultura han permitido explorar nuevas oportunidades de mediciones y evaluaciones más precisas y rápidas.

En el presente proyecto se aborda una propuesta para estimar el contenido de grados Brix de un lote de *Agave tequilana* a través del uso de imágenes capturadas por un vehículo aéreo no tripulado. En el cual, la primera parte consiste en capturar las imágenes de los predios de *Agave tequilana* y a la par realizar un muestreo de la cantidad de grados Brix reales contenidos en este. Posteriormente las imágenes obtenidas son preprocesadas, es decir, se resaltan las características esenciales de importancia en la imagen para después calcular un índice de vegetación. En esta tesis se ha utilizado el *índice de vegetación normalizada* (NDVI). Para realizar la estimación se utiliza un algoritmo de regresión, el cual muestra parámetros de correlación y finalmente muestra la predicción de grados Brix para un predio de *Agave tequilana*.

# Agradecimientos

En este proceso de aprendizaje, agradezco a los docentes que conforman el posgrado en Ciencias de la Computación del Tecnológico Nacional de México Campus Ciudad Guzmán, por haber compartido sus conocimientos conmigo, por su guía y apoyo en áreas del conocimiento nuevas para mi. En especial, a la Dra. Ma. Guadalupe Sánchez Cervantes por su apoyo y dedicación como mi directora de tesis, de igual forma a la Dra. Rosa María Michel Nava y al Dr. Marco Antonio Meza Aguilar por su tiempo dedicado a la revisión de mi trabajo.

Al Dr. Himer Ávila George, por su apoyo, su dedicación y conocimientos transmitidos en el área de ciencias de la computación como mi codirector de tesis.

Al Ing. Omar Hernández Calvario por su tiempo y apoyo en transmitir conocimientos nuevos y su gran compañerismo en este proceso.

Al Ing. Ricardo Velasco Vázquez por su apoyo en la investigación de algoritmos de regresión y su funcionamiento.

Al Consejo Nacional de Ciencia y Tecnología por otorgarme la beca 522785 para el desarrollo de mis estudios de maestría.

Y por su puesto, quiero agradecer a mi esposo Edgar García por su apoyo para hacer posible mis estudios de posgrado y a mi mamá y a mi hija por su apoyo durante este proceso.

# Dedicatoria

*A mi hija, que siempre ha sido mi inspiración para realizar mis metas personales y profesionales.*

*A mis Directores de Tesis que confiaron siempre en que podía con todos los retos que se me presentaran.*

*A mi esposo, por ser siempre quien me impulsa a nunca desfallecer en mis metas y estar ahí apoyándome a cumplirlas.*

# Índice general

Resumen	I
Agradecimientos	II
Dedicatoria	III
Índice de figuras	VIII
Índice de tablas	IX
Lista de siglas y acrónimos	X
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	2
1.2. Planteamiento del problema . . . . .	4
1.3. Justificación . . . . .	4
1.4. Objetivos . . . . .	5
1.4.1. Objetivo general . . . . .	5
1.4.2. Objetivos específicos . . . . .	5
1.5. Hipótesis . . . . .	5
<b>2. Marco Teórico</b>	<b>6</b>
2.1. El Tequila en México . . . . .	6
2.2. Agave tequilana Weber . . . . .	6
2.3. Grados Brix . . . . .	7
2.4. Fotogrametría . . . . .	7
2.4.1. Fotogrametría digital . . . . .	8
2.4.2. Uso de drones en la fotogrametría . . . . .	8
2.5. Índice de vegetación normalizada ( <i>NDVI</i> ) . . . . .	9
2.6. Técnicas de regresión . . . . .	10
2.6.1. Regresión lineal y correlación . . . . .	10
2.6.2. Regresión Múltiple . . . . .	12
2.6.3. Regresión Logística . . . . .	13

2.7.	Algoritmos de regresión . . . . .	13
2.7.1.	Algoritmo de Regresión Lineal . . . . .	13
2.7.2.	Vectores de Soporte de Regresión (SVR) . . . . .	14
2.7.3.	Regresión de cuadrados mínimos parciales (PLSR) . . . . .	15
2.8.	Reconocimiento de patrones . . . . .	15
2.9.	Machine Learning . . . . .	16
2.9.1.	Aprendizaje por refuerzo . . . . .	17
2.9.2.	Aprendizaje Supervisado . . . . .	17
2.9.3.	Aprendizaje no supervisado . . . . .	17
<b>3.</b>	<b>Revisión sistemática de la Literatura</b>	<b>19</b>
3.1.	Planificación de la revisión . . . . .	19
3.2.	Desarrollo de la revisión . . . . .	21
3.3.	Análisis de los resultados . . . . .	22
<b>4.</b>	<b>Materiales y métodos</b>	<b>28</b>
4.1.	Equipo a utilizar para el desarrollo del proyecto . . . . .	28
4.1.1.	VANT Parrot BlueGrass . . . . .	28
4.1.2.	Especificaciones del fabricante del VANT Parrot BlueGrass . . . . .	29
4.1.3.	Refractómetros . . . . .	30
4.2.	Metodología propuesta . . . . .	32
4.2.1.	Adquisición de imágenes y muestreo de grados Brix . . . . .	32
4.2.2.	Preprocesamiento de imágenes y cálculo de índice NDVI . . . . .	36
4.2.3.	Identificación de las técnicas de regresión para estimación . . . . .	37
4.2.4.	Diseño e implementación del algoritmo de estimación de grados Brix . . . . .	38
4.2.5.	Análisis del algoritmo de estimación empleado . . . . .	40
<b>5.</b>	<b>Resultados</b>	<b>41</b>
5.1.	Zona de estudio . . . . .	41
5.2.	Adquisición de imágenes . . . . .	42
5.2.1.	Plan de vuelo y adquisición de imágenes . . . . .	42
5.2.2.	Muestreo de grados Brix . . . . .	45
5.3.	Preprocesamiento de imágenes y cálculo de índice NDVI . . . . .	46
5.3.1.	Eliminación de ruido de la imagen . . . . .	46
5.3.2.	Cálculo del índice NDVI . . . . .	47
5.4.	Identificación de las técnicas de regresión para estimación . . . . .	49
5.5.	Diseño e implementación del algoritmo de estimación de grados Brix . . . . .	50
5.5.1.	Análisis del algoritmo de regresión empleado . . . . .	52
<b>6.</b>	<b>Conclusiones</b>	<b>54</b>
	<b>Referencias</b>	<b>58</b>

<b>Anexos</b>	<b>59</b>
<b>A. Publicación derivada de este trabajo de tesis</b>	<b>61</b>
<b>B. Estancia académica</b>	<b>62</b>

# Índice de figuras

2.1. <i>Agave tequilana</i> Weber variedad azul. . . . .	7
2.2. VANT utilizado para captura de imágenes. . . . .	9
2.3. Rango e interpretación de valores del índice de vegetación normalizada Barrenechea (2020). . . . .	10
2.4. Ejemplo y componentes del diagrama de dispersión. . . . .	11
2.5. Intensidad y dirección del coeficiente de correlación. . . . .	12
2.6. Representación gráfica del algoritmo SVM y sus componentes. . . . .	15
3.1. Publicaciones por año. . . . .	23
3.2. Vegetación estudiada utilizando VANT. . . . .	24
4.1. VANT utilizado Parrot BlueGrass . . . . .	28
4.2. Refractómetro PEN-PRO . . . . .	30
4.3. Refractómetro Pocket. . . . .	31
4.4. Metodología propuesta. . . . .	32
4.5. Muestreo de grados Brix en agave utilizando refractómetro Pocket. . . . .	34
4.6. Calibración refractómetro PEN-PRO. . . . .	35
4.7. Calibración realizada refractómetro PEN-PRO. . . . .	35
4.8. Calibración realizada refractómetro Pocket. . . . .	36
5.1. Mapa del estado de Jalisco resaltando la región Valles. . . . .	41
5.2. VANT Parrot Bluegrass iniciando vuelo para captura de imágenes. . . . .	42
5.3. Imagen 1 formato RGB. . . . .	43
5.4. Imagen 2 formato RGB. . . . .	43
5.5. Imagen 3 formato RGB. . . . .	44
5.6. Imagen formato GREEN. . . . .	44
5.7. Imagen formato NIR. . . . .	45

5.8. Imagen formato RED. . . . .	45
5.9. Imagen formato REG. . . . .	46
5.10. Imagen RGB antes de ser procesada. . . . .	48
5.11. Imagen binarizada. . . . .	48
5.12. Imagen de la banda infrarrojo cercano (NIR). . . . .	49
5.13. Imagen de la banda roja (RED). . . . .	49
5.14. Imagen resultado del cálculo de NDVI. . . . .	50
5.15. Gráfica de regresión lineal de la relación entre NDVI y grados Brix. . . . .	51
5.16. Modelo final de regresión lineal. . . . .	52
A.1. Publicación de artículo. . . . .	61

# Índice de tablas

5.1. Resultados muestreo de grados Brix. . . . .	47
--	----

# Lista de siglas y acrónimos

**CRT** Consejo Regulador del Tequila.

**ML** Machine learning.

**MSE** Mean Square Error.

**NDVI** Normalized Difference Vegetation Index.

**NIR** Infrarrojo cercano.

**PLSR** Partial Least Squares Regression.

**RED** Banda ROJO del espectro electromagnético.

**REG** Banda Red Edge del espectro electromagnético.

**RGB** Red, Green, Blue.

**RMSE** Root Mean Square Error.

**SAR** Synthetic Aperture Radar.

**SVM** Support Vector Machine.

**SVR** Support Vector Regression.

**UAV** Unmanned Aerial Vehicle.

**VANT** Vehículo aéreo no tripulado.

# Capítulo 1

## Introducción

La implementación de las nuevas tecnologías en las industrias de producción de bienes y servicios actualmente está en auge, puesto que su implementación no sólo permite precisión en los procesos productivos, sino que genera un ahorro de recursos económicos, materiales y en algunos casos humanos.

La agricultura inteligente tiene como objetivo brindar una herramienta al agricultor para poder procesar la información del cultivo y con ésta poder predecir el comportamiento que va a tener la cosecha, esto entonces le permitirá al productor poder mejorar la toma de decisiones haciendo que la agricultura sea más eficiente y eficaz. La agricultura inteligente se ha venido desarrollando en los últimos años, y con el fin de dar un panorama de avance de esta nueva área de conocimiento, se realizó una revisión en la cual no sólo se busca mostrar cuales son las tendencias en el área, si no también ver los referentes teóricos Gómez (2020).

Al hablar de tequila no sólo se habla de una bebida que representa la cultura y tradición de un país; sino de un producto que genera ganancias importantes y genera cientos de empleos en su producción. Actualmente la industria del tequila aporta divisas por 1.3 billones de dólares por concepto de exportaciones, pasando de 64.6 millones de litros en 1995 a 339.4 millones de litros exportados en el año 2021. Es por esto que la industria del tequila es un aliado estratégico y fundamental para el desarrollo económico del país Forbes (2018).

Al ser un pilar importante en el desarrollo económico del país, esta industria poco a poco va implementando nuevas tecnologías a sus procesos. En el presente proyecto se desarrolla una propuesta de un modelo capaz de estimar los grados Brix de un predio de *Agave tequilana* Weber, el cual es la materia prima principal en la producción del tequila,

haciendo uso de técnicas de fotogrametría y aprendizaje máquina.

## 1.1. Antecedentes

El tequila es una de las bebidas más famosas a nivel mundial y que sin duda alguna, representa a México en el mundo. Actualmente, según cifras del CRT (Consejo Regulador del Tequila), su consumo ha aumentado considerablemente en los últimos años. Esto entonces, ha llevado a profundizar un poco principalmente en conocer su proceso y por ende las características del *Agave tequilana* como su principal materia prima.

De acuerdo con Montañez-Soto, Venegas-González, Vivar-Vera, y Ramos-Ramírez (2011), se obtuvieron como resultados que el peso promedio de una planta de *Agave tequilana* pesa en promedio 91.1 kg y está comprendido entre el intervalo de 59 a 128 kg, mientras que el peso promedio de las cabezas (piñas del agave) fue de 48.9 kg y que estas representan el 53.7% del peso total de la planta mencionan los autores. Estudios como este sirven de referencia para el presente trabajo en cuestión de poder conocer la composición y estructura del *Agave tequilana*. Cabe destacar que ningún trabajo analizado se centra en estimar los grados Brix del mismo.

Algo importante a esclarecer, es que, en cuanto a estimación se trata, se encontró que diversos autores tienen algo en común: el uso de ecuaciones de regresión, algoritmos de tales como el algoritmo *Support Vector Machine*, *Random Forest* y el *Método de Cuadrados Mínimos* entre otras técnicas estadísticas.

La combinación de diversas metodologías tales como técnicas de percepción remota como lo es el uso de imágenes satelitales, en conjunto con un algoritmo de como lo aplican en el trabajo de “*Estimación de biomasa aérea de Eucalyptus grandis y Pinus spp. usando imágenes Sentinel1A y Sentinel2A en Colombia*” Tovar Blanco, Lizarazo Salcedo, y Rodríguez Eraso (2020), en el cual se hace uso de estas técnicas ya mencionadas para estimar la biomasa aérea de 2 tipos de árboles en Colombia, como resultados obtuvieron un coeficiente de determinación de 0.27 y de error cuadrado promedio 42.75 para el Eucalipto gradis. En este trabajo los autores demuestran que algoritmos de aprendizaje de máquina como Random Forest, en combinación con índices espectrales y datos SAR (Radar de Apertura Sintética) como variables predictoras, se generan mejores resultados para la estimación de biomasa aérea.

En cuanto a índices espectrales en la búsqueda de trabajos relacionados, se encontra-

ron diversos trabajos que utilizan diferentes tipos de imágenes para estimar la biomasa, tales como imágenes satelitales e imágenes adquiridas desde vehículos aéreos no tripulados. “El empleo de VANT para la toma de imágenes aéreas de una elevada resolución posee múltiples beneficios sobre el uso de aviones tripulados y satélites con iguales fines, ya que estos pueden brindar una calidad superior en los mapas, operan en el ámbito local y logran sobrevolar áreas inaccesibles” mencionan los autores del trabajo titulado “Análisis de imágenes multiespectrales adquiridas con vehículos aéreos no tripulados” Kharuf-Gutierrez, Hernández-Santana, Orozco-Morales, Aday Díaz, y Delgado Mora (2018). En este trabajo se establece que la captura de imágenes en el terreno de algún plantío específico se realiza simultáneamente en varios rangos del espectro electromagnético llamados bandas. “La necesidad de su adquisición radica en las propiedades espectrales (reflectancia, transmitancia y absorbancia, entre otras) del suelo y de las plantas en sus diferentes etapas de desarrollo, las que van ligadas estrechamente a factores como actividad fotosintética, cantidad de agua, clorofila, biomasa, entre otros” mencionan los autores.

## 1.2. Planteamiento del problema

Al planear la producción para un producto en específico, es importante planear los requerimientos de la materia prima necesaria y así cumplir con las metas estimadas de producción.

Siendo el *Agave tequilana* la materia prima principal en la elaboración del tequila se estima que para producir 1 litro de esta bebida son requeridos de entre 7 y 8 kilogramos de éste. Los cuales se convierten en aproximadamente 5 kilogramos de bagazo (residuo que es la base húmeda) una vez que se ha extraído el jugo. A su vez, para que un *Agave tequilana* pueda ser cosechado para la producción de tequila deben pasar de 6 a 8 años desde que éste fue cultivado. Algunos de los parámetros para determinar el valor de la planta de *Agave tequilana* es el peso y la cantidad de azúcares contenidos en ella.

## 1.3. Justificación

La producción de tequila es para el país una industria que anualmente genera importantes ganancias. Es por esto por lo que contar con un control de su materia prima es de suma importancia para asegurar su producción estipulada para cierto periodo de tiempo.

Debido al alto costo que el proceso de cultivo hasta cosecha que un *Agave tequilana* genera, es importante para el productor así como para la industria del tequila el poder estimar la cantidad de grados Brix que un determinado predio de *Agave tequilana* contiene, ya que esto va directamente relacionado con su producción y hacerlo de forma efectiva y óptima reduce costos de producción y operación.

El realizar un método que permita estimar los grados Brix de un predio de *Agave tequilana* permite a su vez tener un mayor control de la materia prima destinada a producción. Es por eso que se decide diseñar un algoritmo basado en las técnicas de fotogrametría, reconocimiento de patrones y aprendizaje computacional que permita estimar el nivel de azúcares contenidos.

## 1.4. Objetivos

### 1.4.1. Objetivo general

Desarrollar un algoritmo que estime los grados Brix de un cultivo de *Agave tequilana* a través de imágenes adquiridas desde un vehículo aéreo no tripulado.

### 1.4.2. Objetivos específicos

- Identificar técnicas de regresión utilizadas para la estimación de parámetros agronómicos
- Comparar las técnicas de regresión identificadas.
- Aplicar la técnica de regresión seleccionada para la estimación de grados Brix en cultivos de *Agave tequilana*
- Analizar los resultados obtenidos

## 1.5. Hipótesis

Los grados Brix de un predio de agaves destinados a la producción de tequila pueden ser estimados mediante el uso de imágenes adquiridas desde un vehículo aéreo no tripulado, el cálculo de un índice de vegetación y un modelo que incluye un algoritmo de regresión.

# Capítulo 2

## Marco Teórico

### 2.1. El Tequila en México

El tequila se define como la bebida alcohólica regional obtenida de la planta *Agave tequilana* Weber también conocido como *agave azul*. Su obtención es por medio de la destilación de mostos (jugos de agave) preparados, directa y originalmente del material extraído en las instalaciones de un productor autorizado. Derivados de las cabezas de *Agave tequilana* previa o posteriormente hidrolizadas o cocidas, y sometidos a fermentación alcohólica con levaduras, cultivadas o no, siendo susceptibles los mostos de ser enriquecidos y mezclados conjuntamente en la formulación con otros azúcares hasta en una proporción no mayor de 49 % de azúcares reductores totales expresados en unidades de masa, en los términos establecidos por la Norma Oficial del Tequila NOM-006-SCFI-2012, se menciona que no están permitidas las mezclas en frío, de acuerdo al Consejo Regulador del Tequila (CRT) que es la organización dedicada a inspeccionar y certificar el cumplimiento de esta norma así como de salvaguardar la calidad, cultura y prestigio del tequila CRT (2020).

### 2.2. Agave tequilana Weber

También conocido como *agave azul*, es una planta xerófila (es decir, que crece en zonas áridas y cálidas) de hojas color azul verdoso, delgadas y casi planas (ver Figura 2.1). Mide aproximadamente 1.25 metros de largo, 10 cm de ancho y tiene una espina terminal de color rojo oscuro de 2 cm Granados Sánchez (1993).

El *Agave tequilana* se cultiva ampliamente en México en las zonas de denominación



Figura 2.1: *Agave tequilana* Weber variedad azul.

de origen (las cuales son Jalisco, Michoacan, Nayarit, Guanajuato y Tamaulipas) para la producción de tequila. En tiempos de cosecha se recolectan las cabezas o piñas y se llevan a las fábricas para la producción de tequila Cervantes Rendón y Montano Armendáriz (2018).

### 2.3. Grados Brix

Grados Brix (símbolo ° Bx) es el contenido de azúcar de una solución acuosa. Un grado Brix es 1 gramo de sacarosa en 100 gramos de solución y representa la fuerza de la solución.

Los grados Brix (°B) indican el porcentaje de sólidos solubles en agua, en el jugo de frutas y puede verse afectado por muchos factores incluyendo variedad, región de crecimiento, año de crecimiento y nivel de madurez de la fruta o planta Türkmen y Ekşi (2011).

### 2.4. Fotogrametría

La fotogrametría es la técnica que permite crear modelos 3D a partir de imágenes 2D. Es un proceso cuyo objetivo principal es el convertir datos bidimensionales (tales como imágenes planas) en información tridimensional.

La Sociedad Estadounidense de Fotogrametría y percepción Remota por sus siglas en inglés (ASPRS), define la fotogrametría como “el arte, la ciencia y la tecnología de obtener información confiable sobre objetos físicos y el medio ambiente, a través de procesos de registro, medición e interpretación de imágenes y representaciones digitales de energía patrones derivados de sistemas de sensores sin contacto”.

Etimológicamente la palabra fotogrametría se deriva del griego *photos* que significa luz, *gramma*, que significa lo que está dibujado o escrito y *metrón* que significa medir. Esto significa pues, juntando las palabras “*medir gráficamente por medio de la luz*”.

El principio en el que se basa la fotogrametría consiste en proyectar en forma ortogonal sobre un plano de referencia, la imagen registrada en una fotografía, la cual ha sido proyectada sobre el negativo mediante la proyección central, que es la usada por las lentes. En fotogrametría se asume que la proyección central es perfecta, lo cual implica que no existe desviación de los rayos de luz que atraviesan los lentes de la cámara. La imagen se proyecta sobre una superficie perfectamente plana y la relación matemática que une al objeto y su imagen se conoce con el nombre de principio de colinealidad, Jauregui (2006).

### 2.4.1. Fotogrametría digital

Actualmente en auge, surge como consecuencia del gran desarrollo de la computación, que permitió realizar todos los procesos fotogramétricos mediante el uso de computadoras.

Con la fotogrametría digital crecen las posibilidades de explotación de las imágenes, a la vez que se simplifican las tecnologías, permitiendo con ello la generación automática de modelos de elevación del terreno, orto imágenes y estereortoimágenes, generación y visualización de modelos tridimensionales entre otros, Jauregui (2006).

Para llevar a cabo la restitución digital, las imágenes digitales son ingresadas en la computadora, y mediante visualización en pantalla de las mismas, el operador ingresa los puntos necesarios para realizar el proceso de orientación en forma matemática. La restitución puede ser un proceso iterativo con el operador o ser realizada en forma automática por correlación de imágenes. La salida en la fotogrametría digital puede ser en formato raster o formato vectorial, Jauregui (2006).

### 2.4.2. Uso de drones en la fotogrametría

Actualmente, el uso de drones también conocidos como Vehículo Aéreo No Tripulado (VANT), han hecho que el uso de la fotogrametría aérea se generalice en diversos campos.

Las técnicas que hoy en día se utilizan, tienen sustento en las técnicas tradicionales; sin embargo, el equipamiento necesario, los requerimientos y la complejidad de las operaciones, se han reducido sustancialmente. Junto con ello el costo de operaciones de adquisición de imágenes tradicionales y costo de operaciones para los cuales la fotogrametría con vehículos aéreos proporciona una solución mas rápida, óptima y de mejor calidad. La fotogrametría con drones permite conocer propiedades geométricas de un objeto o de una superficie, partiendo de varias imágenes adquiridas con información redundante. Con la finalidad de que un objeto logre ser reconstruido fielmente, este debe aparecer en una cantidad suficiente de imágenes en repetidas ocasiones. Pues es así, como la información repetitiva permite extraer su estructura.

Las ventajas de la fotogrametría con drones son la rapidez, al tenerse en poco tiempo el resultado esperado (después de que un software especializado realice la interpretación), costos de operación relativamente bajos y un mayor control en el resultado final Aerial Insights (2018). En la Figura 2.2, se muestra el VANT utilizado en el presente proyecto.



Figura 2.2: VANT utilizado para captura de imágenes.

## 2.5. Índice de vegetación normalizada (*NDVI*)

En teledetección la relación entre las bandas del infrarrojo cercano y el rojo constituyen un test que comprueba el vigor de la vegetación o por el contrario su débil vitalidad Manrique (1999). El índice NDVI es un indicador de la salud de las plantas que se basa en cómo se reflejan las ondas de luz. Las plantas con bastante clorofila y buena estructura celular absorben activamente la luz roja (RED) y reflejan la luz del infrarrojo cercano (NIR),

mientras que con una planta enferma sucede lo contrario. Este índice de vegetación tiene como ventaja el tener rangos de valores entre -1 y 1. Los valores de -1 a 0 corresponden a superficies como nieve, agua, arena, piedras, caminos, casas, etc. Su cálculo está dado por la ecuación 2.1. Se utiliza para estimar la cantidad, calidad y desarrollo de la vegetación con base a la medición de la intensidad de la radiación de ciertas bandas del espectro electromagnético que la vegetación refleja. Para el cálculo de los índices de vegetación es necesaria la información que se encuentra en las bandas roja e infrarroja de ese espectro electromagnético.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2.1)$$

Donde:

- NIR: banda del infrarrojo cercano del espectro electromagnético
- RED: banda roja del espectro electromagnético

En la Figura 2.3 se muestran los rangos de NDVI y su interpretación.

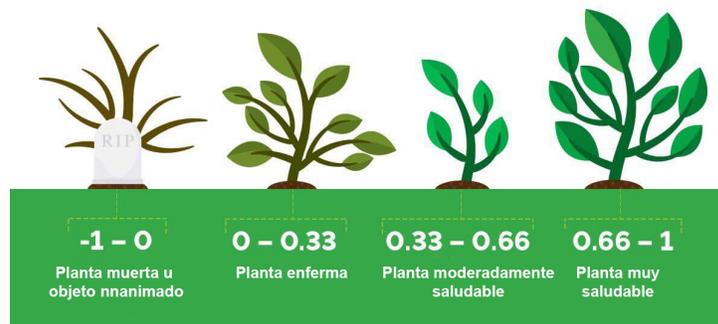


Figura 2.3: Rango e interpretación de valores del índice de vegetación normalizada Barrenechea (2020).

## 2.6. Técnicas de regresión

### 2.6.1. Regresión lineal y correlación

Existen muchas técnicas de regresión en función del tipo de variables y de la forma funcional supuesta entre ellas. Las más elementales son las lineales. La regresión lineal supone

que la relación entre dos variables tiene una forma lineal o linealizable. La regresión lineal tiene una versión simple que estudia la relación entre dos variables y consiste en generar un modelo que basándose en la relación existente entre ambas variables, permita predecir el valor de una a partir de la otra. La regresión también muestra la magnitud de la correlación; el incremento marginal, el valor de una variable cuando la otra es cero y si dicha relación puede considerarse significativa o no significativa Granados (2016). En la ecuación 2.2 se muestra la fórmula para la ecuación de la regresión lineal.

$$y = \beta_0 + \beta_1 X \quad (2.2)$$

**Diagrama de dispersión:** Es la gráfica que representa la relación entre dos variables, Levin y Rubin (2004). Los diagramas de dispersión usan una colección de puntos colocados usando coordenadas cartesianas para mostrar valores de dos variables, se muestra una variable en cada eje y se puede detectar si existe una relación o correlación entre las dos variables. Son en sí una herramienta gráfica en la regresión lineal. En la Figura 2.4 se muestra un ejemplo de un diagrama de dispersión así como las partes que la componen.

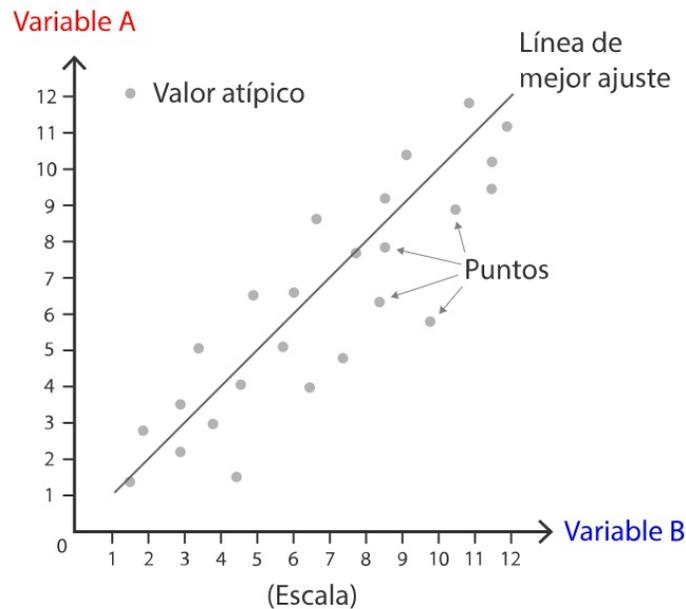


Figura 2.4: Ejemplo y componentes del diagrama de dispersión.

## Análisis de correlación

Se le conoce como análisis de correlación al conjunto de técnicas estadísticas que se emplean para medir la intensidad en la que dos variables se asocian. El principal objetivo de este análisis es determinar qué tanto se relaciona una variable dependiente con la variable independiente y así lograr realizar una predicción Levin y Rubin (2004).

## Coefficiente de correlación

El coeficiente de correlación describe la intensidad de la relación entre dos variables de intervalo. Se le denota con  $r$  y puede tomar cualquier valor de  $-1.00$  a  $+1.00$ . Entonces, se puede especificar que el coeficiente de correlación es la medida de la intensidad de la relación lineal entre dos variables Levin y Rubin (2004). En la Figura 2.5 se muestra que intensidad y dirección tiene la correlación dependiendo de sus valores.

## Coefficiente de determinación

El coeficiente de determinación es la porción de la variación total en la variable dependiente  $Y$ , que se explica por la variación en la variable independiente  $X$ . Este coeficiente se calcula elevando al cuadrado el coeficiente de correlación, se expresa mediante  $r^2$  y determina una proporción o porcentaje, este será la variación total Levin y Rubin (2004).

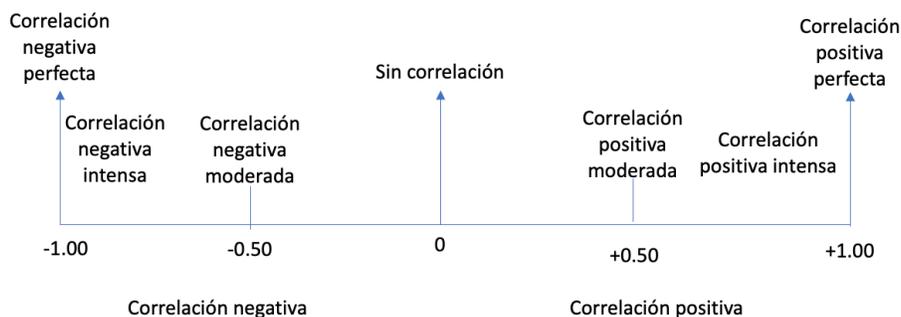


Figura 2.5: Intensidad y dirección del coeficiente de correlación.

### 2.6.2. Regresión Múltiple

La regresión múltiple se caracteriza por tener dos o más variables independientes que pueden influir sobre la variable dependiente. Está dada por la ecuación 2.3, el análisis

de regresión múltiple permite establecer la relación entre la variable dependiente con las variables independientes. El análisis de regresión lineal múltiple, se aproxima más a situaciones de análisis real puesto que los fenómenos, hechos y procesos sociales, suelen ser de mayor complejidad. Cuando se aplica un análisis de regresión múltiple lo más frecuente es que la variable dependiente y las independientes sean variables continuas que son medidas en escala de intervalo o razón. Además, es posible aplicar este análisis cuando se relacione una variable dependiente continua con un conjunto de variables categóricas. Es posible de igual forma aplicar un análisis de regresión múltiple en el caso de que se relacione una variable dependiente nominal con un conjunto de variables continuas Rodríguez-Jaume, Mora Catalá, y cols. (2001).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 \dots \beta_p X_p + \epsilon \quad (2.3)$$

### 2.6.3. Regresión Logística

La regresión logística es un tipo de análisis de regresión que se utiliza para predecir el resultado de una variable categórica, esto en función de la variable predictora. Suele ser utilizada para predicciones en eventos cuya función depende de otros factores.

La regresión logística es un instrumento estadístico de análisis multivariado. Resulta útil su empleo cuando se tiene una variable dependiente dicotómica (un atributo cuya ausencia o presencia hemos puntuado con los valores cero y uno, respectivamente) y un conjunto de variables predictoras o independientes, que pueden ser cuantitativas o categóricas Chitarroni (2002).

## 2.7. Algoritmos de regresión

### 2.7.1. Algoritmo de Regresión Lineal

La regresión lineal es un algoritmo de aprendizaje supervisado que es utilizado en machine learning y en estadística. Regresión lineal es una aproximación para modelar la relación entre una variable escalar dependiente “Y” y una variable explicativas “X”.

Los algoritmos de machine learning Supervisados aprenden por sí mismos, en este caso el algoritmo aprenderá a obtener automáticamente esa “recta” que buscamos con la tendencia de predicción. Para hacerlo se mide el error con respecto a los puntos de entrada

y el valor “Y” de salida real. El algoritmo deberá minimizar el costo de una función de error cuadrático y esos coeficientes corresponderán con la recta óptima.

### 2.7.2. Vectores de Soporte de Regresión (SVR)

Las máquinas de vectores de soporte (SVM) se utilizan para resolver problemas de clasificación en el aprendizaje automático. Support Vector Regression (SVR) utiliza el mismo principio que SVM, pero para problemas de regresión.

#### Máquinas de Soporte Vectorial (SVM)

Retomando brevemente como funcionan las máquinas de soporte vectorial, se tiene que es un enfoque para la clasificación que se desarrolló en la década de 1990 y que ha ganado popularidad desde entonces. Se ha demostrado que las SVM funcionan bien en una variedad de entornos y, a menudo, se las considera uno de los mejores clasificadores "listos para usar". El objetivo del algoritmo SVM es encontrar un hiperplano que separe de la mejor forma posible dos clases diferentes de puntos de datos. Esto implica el hiperplano con el margen más amplio entre las dos clases. El margen se define como la anchura máxima de la región paralela al hiperplano que no tiene puntos de datos interiores. El algoritmo sólo puede encontrar este hiperplano en problemas que permiten separación lineal. Las Máquinas de Vector Soporte se fundamentan en el *Maximal Margin Classifier*, que a su vez, se basa en el concepto de hiperplano.

**Clasificador de margen máximo:** En un espacio p-dimensional, un hiperplano es un subespacio plano y un subespacio de dimensión p - 1. Por ejemplo, en dos dimensiones, un hiperplano es un subespacio unidimensional plano, en otras palabras, una línea. En tres dimensiones, un hiperplano es un subespacio bidimensional plano, es decir, un plano Cormen, Leiserson, Rivest, y Stein (2022). En la Figura 2.6 se muestra la representación gráfica de como trabaja el algoritmo SVM, los vectores de soporte, el hiperplano de separación y el clasificador de margen máximo.

La definición matemática de un hiperplano es bastante simple. En el caso de dos dimensiones, el hiperplano se describe acorde a la ecuación de una recta que se muestra en la ecuación 2.4.

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 = 0 \quad (2.4)$$

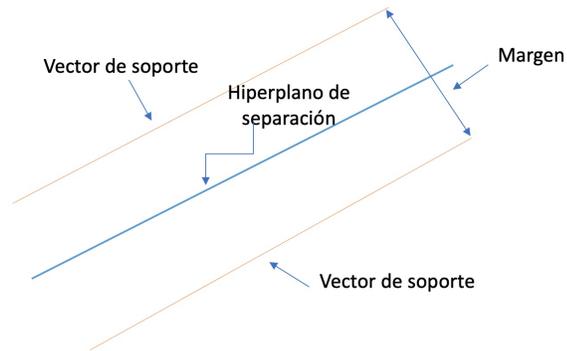


Figura 2.6: Representación gráfica del algoritmo SVM y sus componentes.

### 2.7.3. Regresión de cuadrados mínimos parciales (PLSR)

El Algoritmo de regresión de cuadrados mínimos parciales (PLSR) es una técnica que trata de reducir los predictores a un conjunto más pequeño de componentes no correlacionados para así realizar una regresión de mínimos cuadrados en estos componentes. Esta regresión es útil cuando los predictores con los que se cuenta son muy colineales o en el caso donde existen más predictores que observaciones y la regresión de cuadrados mínimos ordinarios produce coeficientes con errores estándar altos. La regresión de mínimos cuadrados parciales (PLSR) es un método lineal clásico y ampliamente utilizado para el modelado de datos espectrales Cheng y Sun (2017). Esta técnica se desarrolla en sí para evitar el efecto de la multicolinealidad en las estimaciones de los parámetros de una regresión. El problema de la multicolinealidad o relación de dependencia extrema entre variables explicativas en un modelo de regresión, produce situaciones de inestabilidad de los coeficientes de regresión. Los coeficientes de regresión pueden ser no significativos cuando las variables explicativas están muy correlacionadas con la variable explicativa, produciendo dificultades Morales y Domínguez (2009).

## 2.8. Reconocimiento de patrones

Un patrón es una entidad a la que se le puede dar un nombre y que está representada por un conjunto de propiedades medidas y las relaciones entre ellas (vector de características) Alonso Romero y Calonge Cano (2011).

Es la representación de una relación estocástica (es decir de elementos aleatorios)

entre señales; el cual se obtienen mediante análisis matemático de señales que se adquieren previamente. Por lo cual, un patrón representa una clase de señales, que a su vez representa una clase de entidades individuales (objetos, acciones, eventos, procesos entre otros).

Dichos patrones son obtenidos a partir de extraer características distintivas o atributos de estas entidades, usualmente del medio ambiente, por lo que estas señales suelen contener ciertos detalles no relevantes y es necesario reducir la información adquirida, de tal manera que se representen sólo aquellos rasgos o características significativas de la clase.

Un sistema de reconocimiento de patrones tiene uno de los siguientes objetivos: identificar el patrón como miembro de una clase ya definida (clasificación supervisada) y asignar el patrón a una clase todavía no definida (clasificación no supervisada, agrupamiento o clustering).

El diseño de un sistema de reconocimiento de patrones se lleva a cabo normalmente en tres fases que se mencionan a continuación Alonso Romero y Calonge Cano (2011).

- Adquisición y preproceso de datos
- Extracción de características
- Toma de decisiones o agrupamiento

## 2.9. Machine Learning

Machine learning como en su traducción al español significa “máquina que aprende” y se considera que es un maestro en el reconocimiento de patrones, en el cual la estadística es su base fundamental. Machine learning es una rama de la inteligencia artificial que permite que las máquinas aprendan sin ser expresamente programadas para ello Bonaccorso (2017).

Su objetivo principal es estudiar, reinventar, y mejorar modelos matemáticos los cuales pueden ser entrenados (una vez o continuamente) con la información de contexto relativa (proporcionada por un ambiente genérico), para interferir en el futuro y tomar decisiones sin el conocimiento completo de todos los elementos influyentes (factores externos). En otras palabras, un agente (que es una entidad de software que recibe información de un entorno, selecciona la mejor opción para alcanzar un objetivo específico, y observa los resultados de eso) adopta un aprendizaje estadístico intentando determinar las distribuciones de probabilidad exactas y usarlas para computarizar la acción Bonaccorso (2017).

Se conocen actualmente tres enfoques de machine learning que son:

- *Reinforcement Learning*
- *Supervised machine learning*
- *Unsupervised machine learning*

### 2.9.1. Aprendizaje por refuerzo

Este tipo de aprendizaje es particularmente eficiente cuando el ambiente no es completamente determinístico, cuando es frecuentemente dinámico y cuando es imposible tener una medida de error precisa. Incluso sin haber supervisión, el aprendizaje por refuerzo está basado en la retroalimentación proporcionada por el ambiente Bonaccorso (2017).

El aprendizaje por refuerzo se basa en un aprendizaje por práctica, en el cual, conforme se va entrenando, se va adquiriendo cierta experiencia para poder predecir los eventos futuros. En el aprendizaje por refuerzo se utiliza un principio de la psicología conductiva: el condicionamiento operante, aplicado a la inteligencia artificial. Este principio consiste en realizar una práctica para obtener una acción deseada en particular, y todo depende de la práctica que se adquiera.

### 2.9.2. Aprendizaje Supervisado

El aprendizaje supervisado proporciona una herramienta para clasificar y procesar datos utilizando el lenguaje de máquina. Con el aprendizaje supervisado, se suele utilizar datos etiquetados (los cuales son un conjunto de datos que se han clasificado), para inferir un algoritmo de aprendizaje. El conjunto de datos se utiliza como base para predecir la clasificación de otros datos sin etiquetar mediante el uso de algoritmos de aprendizaje automático.

Entre las técnicas importantes conocidas del aprendizaje supervisado se encuentra la regresión lineal y las técnicas de clasificación, Ryan M. (2015).

### 2.9.3. Aprendizaje no supervisado

Este tipo de aprendizaje es una técnica de machine learning en el cual el usuario no necesita supervisar el modelo al momento de trabajar, en lugar de eso permite que el modelo trabaje por si mismo para descubrir patrones e información que no haya sido

detectada con anterioridad. El aprendizaje no supervisado trabaja principalmente con los datos no etiquetados.

El aprendizaje no supervisado puede ser más impredecible en comparación con otros métodos de aprendizaje naturales. Los algoritmos de aprendizaje no supervisados incluyen agrupamiento, detección de anomalías, redes neuronales, etc.

## Capítulo 3

# Revisión sistemática de la Literatura

La revisión sistemática de la literatura tiene como objetivo principal el recolectar información de campo acerca de trabajos relacionados al propio y comparar posibles metodologías y resultados para verificar la viabilidad del estudio. Posteriormente de recolectar los artículo e información de interés; se analiza y compara la videncia que aportan con la de otros similares.

En el presente capítulo se aborda como se trabajó la revisión sistemática de la literatura, la cual fue basada en el método propuesto por Keele y cols. (2007). La cual consta de tres facetas principales que son: planificación de la revisión, desarrollo de la revisión y reporte de resultados. Y la revisión sistemática que se ha realizado por Corona (2019). Cabe resaltar que la presente revisión sistemática se ha realizado buscando estudios que estimaran la biomasa.

### 3.1. Planificación de la revisión

#### **(1) Identificación de la necesidad para realizar la revisión sistemática**

El uso de VANT hoy en día está en auge, gracias a sus menores costos y tecnología avanzada, estos están siendo utilizados en el área de agricultura para la adquisición de imágenes tanto de vegetación como de áreas forestales y con ello extraer características de interés. Mientras que la estimación de la biomasa por medio de imágenes permite tener una aproximación cercana al total de la masa que se tiene por cultivo, vegetación o área forestal. Es por eso que este trabajo se centra en revisar la literatura existente de trabajos realizados acerca de estimación de biomasa utilizando VANTS. Esto con la finalidad de

conocer técnicas de procesamiento de imágenes así como las distintas técnicas estadísticas y tecnológicas para la estimación de la biomasa.

## (2) Establecimiento de las preguntas de investigación

- (PI01) ¿Cómo ha evolucionado el interés en el uso de VANTS para estudios de estimación?
- (PI02) ¿Qué tipo de vegetación han sido las más estudiadas utilizando imágenes adquiridas desde un VANT?
- (PI03) ¿Qué tipo de imágenes se han utilizado en estudios relacionados con vegetación adquiridas desde un VANT?
- (PI04) ¿Qué métodos estadísticos o de aprendizaje máquina se han utilizado para estimar la biomasa?
- (PI05) ¿Qué tecnologías se utilizan para estimación?

## (3) Creación de la cadena de búsqueda

Para definir la cadena de búsqueda, se identificaron las palabras clave de las preguntas de investigación. Como resultado se obtuvo: “Forest, Tree, Plant, Crop, biomass, estimation, UAV, Unamed aerial vehicle”. La cadena de búsqueda se obtuvo al agrupar las palabras clave y el uso de los operadores lógicos AND y OR; como resultado, se obtuvo la siguiente cadena:

*(Forest OR Tree OR Plant OR Crop) AND (Biomass) AND (Estimation) AND (UAV OR UAV'S OR Unmanned Aerial Vehicle).*

## (4) Identificación de las principales fuentes de búsqueda

Para la realización de este estudio, se decidió utilizar las fuentes de búsquedas sugeridas por Keele y cols. (2007) las cuales son: (1) ACM Digital Library, (2) IEEE Xplore Digital Library, (3) ScienceDirect, y (4) Scopus.

## 3.2. Desarrollo de la revisión

La finalidad de esta fase consiste en encontrar los trabajos relacionados que respondan a las preguntas de investigación tantos como sean posibles. Para llevar a cabo el desarrollo de la presente revisión sistemática fue necesario realizar las siguientes dos actividades: (1) seleccionar los estudios primarios y (2) extraer y sintetizar la información Corona (2019).

### (1) Seleccionar los estudios primarios

Para garantizar la imparcialidad de esta búsqueda, se han definido algunos criterios de inclusión y exclusión, siguiendo un procedimiento definido.

#### *Criterios de inclusión*

- (CI01) Estudios que hablen de estimación de biomasa o de grados Brix
- (CI02) Estudios en idioma inglés
- (CI03) Estudios que en su solución de estimación hayan sido resueltos con alguna técnica estadística o de aprendizaje máquina
- (CI04) Estudios que en su abstrac contengan las palabras claves: estimación, biomasa o Brix y UAV o UAV'S

#### *Criterios de exclusión*

- (CE01) Estudios que no se basen en el uso de vehículos aéreos no tripulados
- (CE02) Estudios publicados anterior al año 2010
- (CE03) Estudios que no hayan sido publicados en revistas científicas

*Procedimiento.* Para seleccionar los estudios primarios se realizaron los siguientes pasos mencionados a continuación.

- I Adaptar la cadena de búsqueda al motor de la fuente de datos seleccionada.
- II Aplicar los criterios de inclusión CI01 y CI02
- III Aplicar los criterios de inclusión CI03 y CI04
- IV Aplicar los criterios de exclusión CE01, CE02 y CE03

- V Leer título, resumen, introducción y conclusiones y aplicar los criterios de inclusión y exclusión mencionados
- VI Eliminar y excluir aquellos trabajos duplicados

Durante el proceso de búsqueda y clasificación varios estudios aparecieron duplicados. Para gestionar las referencias y eliminar los duplicados se utilizó la herramienta Rayyan®. Tras aplicar el procedimiento definido para la selección de los estudios primarios se seleccionaron 73 estudios.

## (2) Extracción y síntesis de la información

Después de haberse realizado la búsqueda, y haber eliminado elementos duplicados y excluir trabajos que no cumplían con los criterios de exclusión establecidos, en la herramienta Rayyan® se procedió a dar la segunda lectura para extraer la información, la cual fue vaciada a una hoja de calculo en la cual estaba separada por columnas la información a extraer. La información es la siguiente: fuente, título, tipo de vegetación, tipo de imágenes utilizadas, métodos estadísticos utilizados y tipo de algoritmo utilizado.

## 3.3. Análisis de los resultados

En esta sección se muestran los resultados obtenidos en la presente revisión sistemática de la literatura. Estos resultados permiten entonces conocer en general cual es el tipo de vegetación que se estudia usando imágenes adquiridas desde un VANT y cuales son las técnicas estadísticas y algoritmos de aprendizaje máquina que más se han utilizado para la estimación. Los resultados han sido organizados por pregunta de investigación.

(PI01) ¿Cómo ha evolucionado el interés en el uso de VANTS para estudios de estimación?

Los estudios identificados se han analizado durante la última década, en el periodo de tiempo que comprende desde el año 2010 hasta el año 2021. Así se ha podido identificar la frecuencia y la evolución del número de publicaciones. En la Figura 3.1, se muestran los resultados de este proceso de análisis. Se puede observar que recientemente se han incrementado la cantidad de estudios relacionados con el uso de VANTS para estudios de estimación y que este interés ha ido incrementando desde el año 2017. Cabe mencionar

que la presente revisión sistemática se concluyó en Enero del 2022, razón por la cual no se incluyen estudios del año en curso.

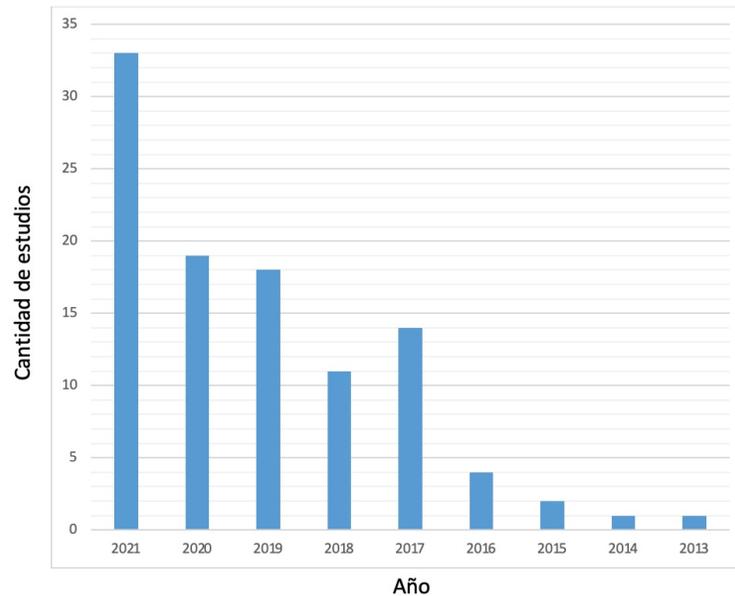


Figura 3.1: Publicaciones por año.

(PI02) ¿Qué tipo de vegetación han sido las más estudiadas utilizando imágenes adquiridas desde un VANT?

En la presente revisión sistemática se pudo observar que para estimación, los vehículos aéreos no tripulados son comúnmente utilizados para estimar la biomasa aérea de ciertos tipos de árboles y bosques, siendo estos el 38 % de la vegetación más estudiada para estimar biomasa utilizando VANT. De igual forma, se encontró que entre la vegetación estudiada están el maíz, el trigo de invierno, arroz y manglares, en la Figura 3.2 se puede observar el porcentaje de estudios analizados para estos tipos de vegetación.

(PI03) ¿Qué tipo de imágenes han utilizado en estudios relacionados con vegetación adquiridas desde un VANT?

En esta revisión se pudo observar que los tipos de imágenes populares para realizar estudios de estimación de biomasa, son las imágenes satélites e imágenes capturas con VANT ya sea utilizando cámaras RGB, múltiplespectrales, hiper espectrales térmicas o sistemas lidar. Cabe mencionar que para esta revisión sistemática, los estudios que utilizaban

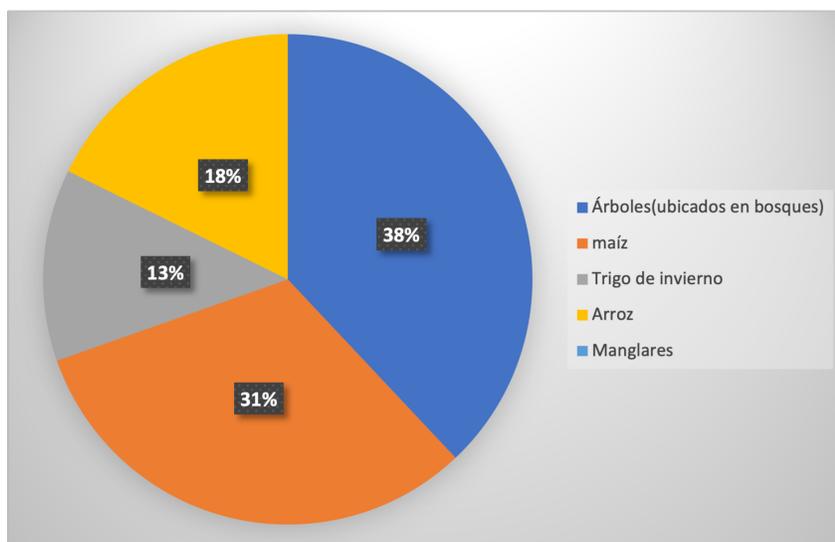


Figura 3.2: Vegetación estudiada utilizando VANT.

imágenes satélites fueron descartados por el criterio de exclusión CE01.

(PI04) ¿Qué métodos estadísticos o de aprendizaje máquina se han utilizado para estimar la biomasa?

Entre las técnicas estadísticas que se observaron en la presente revisión sistemática, destacan los modelos de regresión lineal y regresión lineal múltiple, así como las métricas que evalúan modelos de regresión que se mencionan a continuación.

- Índice de determinación ( $r^2$ )
- Error cuadrático medio (MSE)
- Raíz del error cuadrático medio (RMSE)
- Error absoluto medio (MAE)
- Raíz del error cuadrático medio normalizada (NRMSE)

Cabe destacar que de las métricas mencionadas, el coeficiente de determinación y la raíz de error cuadrado medio, son utilizadas en 70, de los 79 estudios analizados.

Referente a los métodos de aprendizaje máquina que se observaron, se han utilizado en los estudios analizados fueron los siguientes:

- Diameter at Breast Height (DBH)

- Multiple Linear Regression (MLR)
- Linear Regression (LR)
- Partial Least Square (PLSR)
- Random Forest (RF)
- Support Vector Machine (SVM)
- Support Vector Regression (SVR)

Los algoritmos mencionados fueron los que más se repiten en los estudios analizados.

(PI05) ¿Qué tecnologías se utilizan para estimación?

Esta pregunta de investigación refiere como tecnologías a todas aquellas técnicas y métodos modernos utilizados para estimación. Entre las tecnologías utilizadas en los estudios analizados en la presente revisión sistemática se consideraron los siguientes: (1) Índices de vegetación y (2) Algoritmos o redes neuronales utilizados. Se considera a estas tecnologías para la estimación, ya que ambos en conjunto han logrado realizar la estimación del objetivo de cada estudio. A continuación se describen estas tecnologías.

(1) Índices de vegetación

Se puede definir a los índices de vegetación como combinaciones de las bandas espectrales registradas por los sistemas de teledetección, cuya función es realzar la cubierta vegetal en función de su respuesta espectral Gilabert, González-Piqueras, y García-Haro (1997). Entre los índices de vegetación más utilizados en los estudios analizados, se encontraron los mencionados a continuación

- Índice de Vegetación Normalizada (NDVI)
- Vegetación de Diferencia Normalizada Verde (GNDVI)
- Índice de Vegetación Ajustado al Suelo 8S(SAVI)

(2) Algoritmos y redes neuronales utilizadas

Los algoritmos que se observaron en los estudios analizados son mencionados en la pregunta de investigación PI04. Referente a las redes neuronales se pudo observar que en

algunos estudios utilizaron las *redes neuronales convolucionales (CNN)* y las *redes neuronales de propagación hacia atrás (backpropagation neuronal network)*. Cabe mencionar que de los estudios analizados se observó que fueron muy pocos los estudios que utilizaron redes neuronales para solucionar la estimación de biomasa.

*Conclusiones*

Se observó que desde el año 2010, el uso de vehículos aéreos no tripulados ha ido en aumento para solucionar diversos problemas tales como detección de vegetación y estimación de biomasa, agregando a esto los diferentes tipos de dispositivos adjuntos a estas plataformas para la captura de imágenes. Los distintos tipos de imágenes capturadas permiten a su vez realizar un cálculo de índices de vegetación, los cuales mediante operaciones algebraicas permiten realzar la vegetación y con ello hacer el proceso de estimación (en el caso de esta revisión sistemática enfocada a biomasa) pueda realizarse. Se pudo observar de igual forma que la mayoría de los estudios analizados tenían como métrica de su estimación el coeficiente de determinación y la raíz del error cuadrado medio y que cuando de estimación se trata, una técnica estadística de regresión es la adecuada a utilizar apoyándose en algún algoritmo computacional que resuelva la regresión. Esta revisión sistemática es la referencia de las cual se ha tomado las técnicas estadísticas y algoritmos de regresión que se mencionan en los capítulos posteriores de la presente tesis, ya que, aunque haya sido sobre "estimación de biomasa", las técnicas de estimación utilizadas son las que ayudan de referencia. En esta búsqueda de igual forma se ingreso la palabra "Brix" para buscar trabajos relacionados a la estimación a grados Brix, sin embargo, la no encontrarse resultados, se optó por continuar solamente con la cadena de búsqueda original que hace mención de estimar biomasa.

# Capítulo 4

## Materiales y métodos

### 4.1. Equipo a utilizar para el desarrollo del proyecto

#### 4.1.1. VANT Parrot BlueGrass

Para la captura de imágenes aéreas se empleó el VANT Parrot BlueGrass mostrado en la Figura 4.1, puesto que este equipo utiliza tecnología que está orientada al sector agrícola. Este equipo cuenta con seis sensores, una cámara multispectral, una cámara RGB, sus accesorios y componentes de vuelo.



Figura 4.1: VANT utilizado Parrot BlueGrass  
Fuente:Corona (2019).

### 4.1.2. Especificaciones del fabricante del VANT Parrot BlueGrass

Fuente: Corona (2019)

#### ■ Sensores de vuelo

- Sistema de Posicionamiento Global (GPS, por sus siglas en inglés) + GLO-NASS integrados
- Sistema de navegación inercial (INS, por sus siglas en inglés)
- Altímetro
- Ultrasonidos
- Cámara vertical de flujo óptico

#### ■ Sensor de luz solar

- 4 sensores espectrales (filtros idénticos al cuerpo)
- Unidad de medida inercial (IMU, por sus siglas en inglés) y magnetómetro
- Ranura para tarjeta tipo SD o seguro digital (SD por sus siglas en inglés)
- Potencia: 1 Watt
- Sistema de posicionamiento global o GPS

#### ■ Cámara RGB

- Foto: cámara gran angular de 14MP
- Video: 1080p Full HD
- Transmisión de vídeo: 360p / 720p
- Memoria interna de vídeo: De 32 GB

#### ■ Cámara Multiespectral

- Resolución: 1,2 Mpx, 1280x960 píxeles
- HFOV: 61.9
- VFOV: 48.5
- DFOV: 73.7

#### ■ Bandas separadas

- Verde: 550 NM +/- 40 NM
- Rojo: 660 NM +/- 40 NM
- Red Edge: 735 NM +/- 10 NM
- Infrarrojo cercano (NIR): 790 NM +/- 40 NM

#### ■ Segunda cámara RGB

- Resolución: 16 MP, 4608 x 3456 píxeles
- HFOV: 63,9°
- VFOV: 50.1°
- DFOV 73.5°

### 4.1.3. Refractómetros

Un refractómetro es un instrumento de laboratorio utilizado en esta investigación para medir la cantidad de grados Brix contenidos en un agave. Los refractómetros utilizados en este estudio fueron el refractómetro digital de mano PEN-PRO (tipo lápiz) que se muestra en la Figura 4.2 y el refractómetro Pocket mostrado en la Figura 4.3.



Figura 4.2: Refractómetro PEN-PRO

Especificaciones técnicas del fabricante:

- Rango de medida Brix 0.0 a 85.0 %
- Exactitud de la medición Brix +/- 0.2 %

- Resolución Brix 0.1 %
- Temperatura de trabajo 10 a 40°C
- Rango de compensación de temperatura 10 a 100°C
- Volumen de la muestra 0.3 ml
- Clase de protección internacional IP65 (protección al polvo y chorros de agua) Prisma: esta protegido contra el polvo y la inmersión temporal en agua
- Fuente de poder: 1 batería alcalina tipo AAA
- Tiempo de medición: 2 segundos
- Dimensiones y peso: 16(V)x 3.8(D)x 1.8(H)cm, 70g



Figura 4.3: Refractómetro Pocket.

## 4.2. Metodología propuesta

La metodología propuesta en este trabajo de investigación consta de 5 etapas (ver Figura 4.4). En la etapa 1 se incluyen actividades correspondientes a la adquisición de imágenes y muestreo de para obtener los grados Brix en los *Agaves tequilana*. En la segunda etapa se realiza el preprocesamiento de imágenes y el cálculo de un índice de vegetación. Dentro de la etapa 3 se realiza la identificación de las técnicas de regresión más comunes utilizadas para la estimación. En la etapa 4 se diseña y se implementa el algoritmo de regresión seleccionado y, finalmente en la etapa 5 se realiza el análisis del algoritmo implementado.

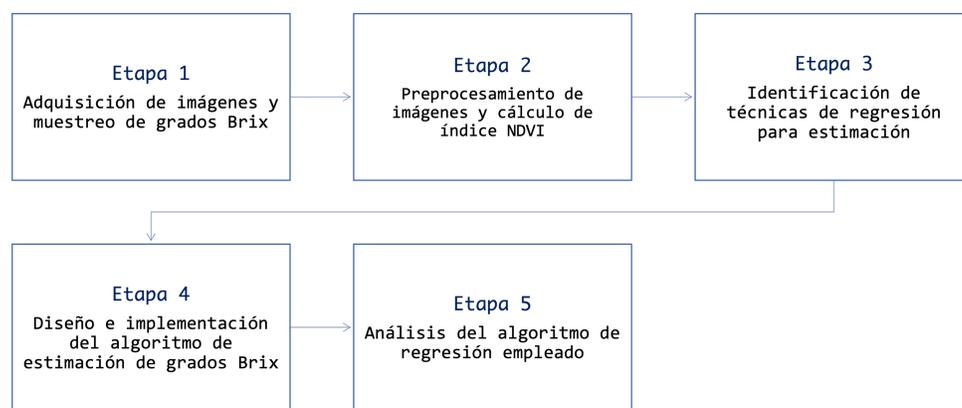


Figura 4.4: Metodología propuesta.

La metodología empleada se describe a continuación.

### 4.2.1. Adquisición de imágenes y muestreo de grados Brix

Este primer paso consistió en establecer un protocolo de vuelo para la adquisición de imágenes, su objetivo principal es definir la trayectoria de vuelo, altura de vuelo, cantidad de imágenes a capturar y posición de la cámara. Todo esto con el propósito de realizar la captura de imágenes de manera óptima. Para facilitar el vuelo y la toma de imágenes aéreas es necesario que el vuelo se realice de manera regular con orientación de este a oeste, sin importar si el terreno sea o no regular, con el objetivo de garantizar que la adquisición de imágenes sea de buena calidad.

Es importante de igual forma las condiciones de vuelo, donde las condiciones geográficas y ambientales son el principal problema. Para poder realizar el vuelo es necesario que se cumplan las condiciones mencionadas a continuación.

- Poco viento y turbulencia
- El cielo debe estar despejado de nubes, polvo, neblina y humo
- Se debe evitar el vuelo en lluvia o tormenta eléctrica.
- La elevación del sol influye en las sombras proyectadas en la superficie
- Es importante seleccionar que el día que se realizará el vuelo, existan condiciones climatológicas óptimas para garantizar que éste pueda ser realizado y adquirir imágenes de buena calidad

El plan de vuelo se tiene que configurar considerando la norma oficial Mexicana NOM-107-SCT3-2019, donde se establecen los requerimientos para operar un sistema de aeronave pilotada a distancia (RPAS, por sus siglas en inglés) en el espacio aéreo mexicano. En la sección 6.2 de la norma oficial Mexicana NOM-107-SCT3-2019 se encuentran las recomendaciones técnicas, preventivas y de seguridad relacionadas con los VANT.

### **Muestreo de grados Brix en *Agaves tequilana***

Para la realización de este paso fueron necesario seguir la secuencia que se describe a continuación:

- Calibrar el refractómetro
- Seleccionar los agaves del lote a muestrear
- Cortar la hoja (penca) del agave a muestrear
- Retirar la corteza de la penca en el lugar que fue cortada y dejar expuesta la parte húmeda.
- Medición de grados Brix con refractómetro PEN-PRO: tocar la cabeza del prisma en la muestra (penca de agave en este caso)
- Medición de grados Brix con refractómetro Poket: extraer un poco de pulpa de la penca de agave y colocarlo sobre el cabezal principal del refractómetro hasta obtener la medición (ver Figura 4.5)
- Registrar los datos obtenidos de los grados Brix de cada agave muestreado



Figura 4.5: Muestreo de grados Brix en agave utilizando refractómetro Pocket.

### Calibración de los refractómetros

Referente al muestreo de grados Brix en *Agaves tequilana* es necesario contar con los instrumentos necesarios para realizarlo. En esta etapa se requiere en primera estancia calibrar el refractómetro. Para el refractómetro pen-pro se realiza sumergiendo el prisma que es la parte donde se encuentra el sensor de los grados Brix en un vaso con 10 ml. de agua a temperatura ambiente y presionar el botón de encendido hasta que en el display todos los dígitos marquen cero (ver Figura 4.6). Posteriormente secar el exceso de agua con un paño seco, en la Figura 4.7 se muestra la calibración realizada a este refractómetro. Para el caso del refractómetro Pocket, su calibración requiere de colocar en el cabezal central (parte del sensor) una gota de agua de aproximadamente 5 ml. presionar el botón de encendido hasta que todos los dígitos marquen cero en el display y secar el agua con un paño, (ver Figura 4.8).

Una vez calibrado el refractómetro que se utilizará, para poder medir los datos es necesario tomar una hoja del agave, retirar su corteza y en la parte húmeda de la hoja

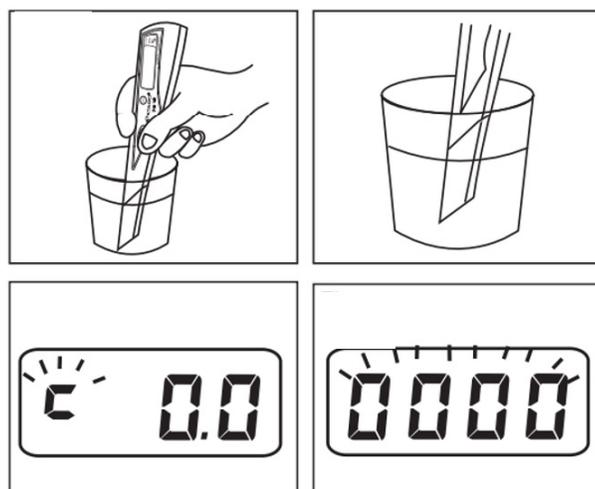


Figura 4.6: Calibración refractómetro PEN-PRO.



Figura 4.7: Calibración realizada refractómetro PEN-PRO.

colocar el prisma por 2 segundos y capturar el dato obtenido de los grados Brix.



Figura 4.8: Calibración realizada refractómetro Pocket.

#### 4.2.2. Preprocesamiento de imágenes y cálculo de índice NDVI

El preprocesamiento de imágenes consiste en extraer las características esenciales de una imagen que serán útiles en el presente trabajo y de igual forma eliminar aquellas características que no son relevantes para este estudio por medio de algoritmos. El preprocesamiento consistió de dos fases: 1) eliminar el ruido de la imagen, es decir, discriminar todas aquellas partes de la imagen que no sean *Agave tequila* y 2) resaltar sólo lo que sí es. El algoritmo 1, muestra el pseudocódigo de la fase de la eliminación del ruido de la imagen. Para ello se utilizaron las librerías *numpy* y *cv2* de Python.

En la segunda parte se realiza el cálculo del NDVI. Para realizarlo se necesita utilizar las imágenes de las bandas rojo (RED) e infrarrojo cercano (NIR) del espectro electromagnético. Para este cálculo fue necesario implementar el algoritmo *cálculo del índice*

---

**Algorithm 1** Eliminación de ruido de la imagen RGB.

---

**Entrada:** Imagen RGB de un cultivo de agave.

**Salida:** Imagen Binarizada.

- 1: cargar imagen RGB
  - 2: convertir imagen RGB a HSV
  - 3: umbralizar imagen HSV (obtener solo color verde)
  - 4: **return** imagen binarizada
- 

---

**Algorithm 2** Cálculo del índice NDVI.

---

**Entrada:** Imagen NIR e imagen RED de un cultivo de agave.

**Salida:** Imagen NDVI y Valores NDVI.

- 1: NIRimage  $\leftarrow$  cargar imagen (NIR)
  - 2: cargar imagen (banda RED)
  - 3: filas  $\leftarrow$  getImageRowCount(NIRimage)
  - 4: columnas  $\leftarrow$  getImageColsCount(NIRimage)
  - 5: geoTrans  $\leftarrow$  getGeographicTransformation(NIRimage)
  - 6: nirArray  $\leftarrow$  transformDatasetToArray(NIRdataset)
  - 7: nirArray  $\leftarrow$  transformDatasetToArray(REDdataset)
  - 8: numeratorArray  $\leftarrow$  subtractArrays(nirArray, redArray)
  - 9: denominatorArray  $\leftarrow$  addArrays(nirArray, redArray)
  - 10: resultArray  $\leftarrow$  divideArrays(numeratorArray, denominatorArray)
  - 11: resultArray  $\leftarrow$  removeOutboundsElements(resultArray)
  - 12: resultImage  $\leftarrow$  writeArrayToImage(resultImage)
  - 13: resultImage  $\leftarrow$  applyGeoTransformation(resultImage, geoTrans)
  - 14: **return** resultImage y resultArray
- 

*NDVI*, (ver algoritmo 2). Este algoritmo fue implementado en el lenguaje de programación Python.

### 4.2.3. Identificación de las técnicas de regresión para estimación

En esta etapa se realizó una investigación previa que consistió en identificar las técnicas de regresión estadística más utilizadas en proyectos de investigación que tuvieran como objetivo lograr una estimación estadística. Así mismo investigar el algoritmo computacional que diera como salida esa estimación.

Como resultado de esta búsqueda se obtuvo que referente a la estimación las técnicas y métricas más utilizadas son las listadas a continuación y se explican detalladamente una por una en la sección 2.6 y 2.7 respectivamente.

**Modelos de regresión:**

- Regresión Lineal
- Regresión Múltiple
- Regresión Logística

**Métricas:**

- índice de correlación ( $r$ )
- índice de determinación ( $r^2$ )

**Algoritmos de regresión:**

- Regresión lineal (ver algoritmo 4)
- Vectores de soporte de regresión SVR (ver algoritmo 3)
- Regresión de mínimos cuadrados parciales (PLSR)

---

**Algorithm 3** Vectores de soporte de regresión (SVR)

---

**Entrada:** Dataset a utilizar para el análisis SVR.**Salida:** Regresión con el algoritmo SVR.

- 1: Cargar el dataset
  - 2:  $X_{svr} \leftarrow \text{dataset}[:, \text{np.newaxis}, n]$
  - 3:  $Y_{svr} \leftarrow \text{dataset.target}$
  - 4: implementar los vectores de soporte de regresión
  - 5: separar los datos de entrenamiento y de prueba
  - 6:  $\text{svr} \leftarrow \text{SVR}(\text{kernel}=\text{linear}, C=1.0, \text{épsilon}=0.2)$
  - 7:  $\text{svr.fit}(X_{\text{train}}, Y_{\text{train}})$
  - 8:  $Y_{\text{pred}} \leftarrow \text{svr.predict}(X_{\text{test}})$
  - 9: **return**  $Y_{\text{pred}}$
- 

#### 4.2.4. Diseño e implementación del algoritmo de estimación de grados Brix

El diseño e implementación del algoritmo de estimación, consiste en seleccionar la técnica que mejor se adapte al caso de estudio. El diseño final del algoritmo que dará como resultado una estimación de grados Brix es un algoritmo de regresión lineal, que se muestra en

el algoritmo 3, el cual funciona ingresando los datos que corresponden a la variable dependiente "X" y a la variable independiente "Y" para finalmente realizar una estimación. El algoritmo de regresión es un algoritmo de machine learning que pertenece al *aprendizaje supervisado*. Este algoritmo fue programado en el lenguaje de programación Python. La implementación de este algoritmo será mediante el uso de los datos obtenidos de los valores del índice NDVI como variable dependiente "X" y los grados Brix obtenidos del muestro que serán la variable independiente "Y". Los parámetros que resultan del análisis de regresión son los mencionados a continuación:

- Gráfica de relación entre los grados Brix y los valores de NDVI
- Ecuación de la línea de regresión
- Coeficiente de correlación
- Coeficiente de determinación
- Predicción 1 para valor de NDVI ingresado
- Predicción 2 para valor de NDVI ingresado

---

**Algorithm 4** Regresión Lineal

---

**Entrada:** Grados Brix y valores NDVI.

**Salida:** Regresión lineal y predicción de grados Brix.

- 1: cargar el dataset
  - 2:  $X \leftarrow \text{data}$  (Valor NDVI)
  - 3:  $Y \leftarrow \text{data}$ (Grados Brix)
  - 4: calcular regresión lineal
  - 5: calcular el parámetro beta 1
  - 6: calcular el parámetro beta 0
  - 7: calcular el índice de correlación
  - 8: realizar predicción
  - 9: valor NDVI para realizar primera predicción
  - 10: valor NDVI para realizar segunda predicción
  - 11: **return** regresión lineal y predicción de grados Brix
-

#### 4.2.5. Análisis del algoritmo de estimación empleado

En esta etapa se realiza el análisis y de igual forma la evaluación del modelo aplicado. Al ser una análisis de regresión lineal fue evaluado mediante dos parámetros obtenidos del mismo modelo: el coeficiente de correlación y coeficiente de determinación. El análisis del modelo consiste en interpretar los resultados obtenidos, para determinar que tan bueno es así como su precisión.

# Capítulo 5

## Resultados

### 5.1. Zona de estudio

Para la realización de este estudio se ha seleccionado la región de los valles del estado de Jalisco (ver Figura 5.1), particularmente los municipios de Ameca y San Martín de Hidalgo.

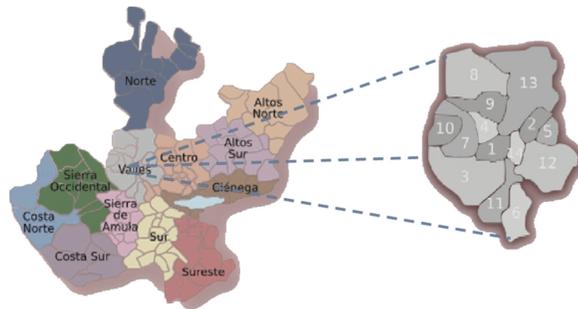


Figura 5.1: Mapa del estado de Jalisco resaltando la región Valles.

De acuerdo con el Sistema de Información de Jalisco (SIJ), la región valles, se localiza en la parte centro-oeste, los 14 municipios que la conforman son: Cocula, San Martín Hidalgo, Ameca, Tala, Ahualulco de Mercado, Etzatlán, El Arenal, Magdalena, Hostotipaquillo, Tequila, San Marcos, Teuchitlán, Amatitán y San Juanito Escobedo. En esta región destaca en dos actividades económicas: la agricultura y la minería. En la primera existen municipios como Tequila, Arenal y Amatitán con alta producción de agave siendo a la vez importantes productores de tequila. Por otro lado, existe una especialización en la producción de caña de azúcar en los municipios de Tala y Ameca, mientras que en

otros municipios se produce de manera relevante el maíz y otros bienes agrícolas que son demandados por la capital del estado Bautista y Garrafa (s.f.).

## 5.2. Adquisición de imágenes

### 5.2.1. Plan de vuelo y adquisición de imágenes

Para realizar el plan de vuelo se utilizó el software Pix4DCapture. La altura de vuelo fue configurada a 30 metros de altura debido a que se favorece la detección de objetos y la percepción de colores a esta altura, mientras que una configuración de mayor altura captura regiones que no son de interés para el trabajo a realizar. El traslape entre fotografías fue de 70 %. Capturando un total de 1204 fotografías de tamaño 2507 x 2309 píxeles. En la Figura 5.2 se muestra el inicio de un vuelo para la captura de imágenes.



Figura 5.2: VANT Parrot Bluegrass iniciando vuelo para captura de imágenes.

Las imágenes obtenidas de los vuelos realizados fueron imágenes en formato RGB y en las bandas del espectro electromagnético NIR, RED, GREEN y REG. Las Figuras 5.3,

5.4 y 5.5 son ejemplos de las imágenes RGB capturadas.



Figura 5.3: Imagen 1 formato RGB.



Figura 5.4: Imagen 2 formato RGB.



Figura 5.5: Imagen 3 formato RGB.

Las Figuras 5.6, 5.7, 5.8 y 5.9 son ejemplos de las imágenes capturadas en formato GREEN, NIR, RED y REG respectivamente.

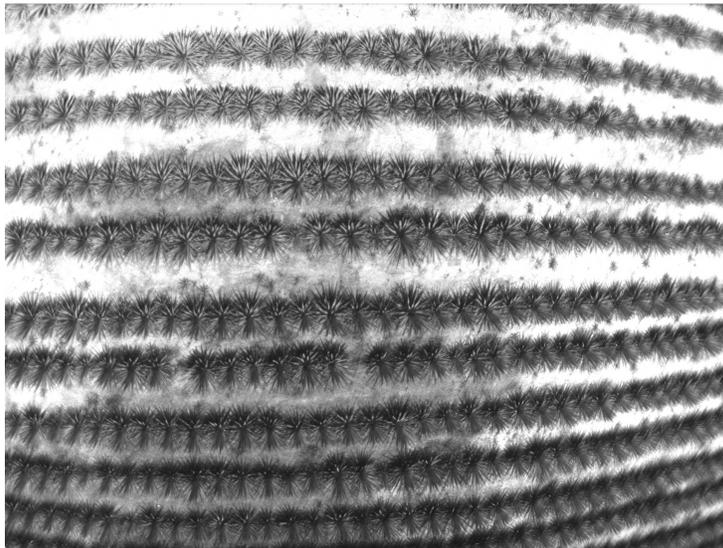


Figura 5.6: Imagen formato GREEN.

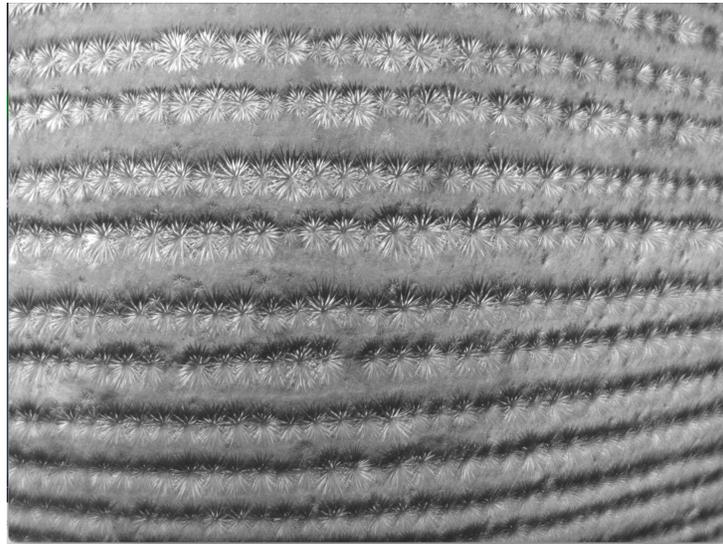


Figura 5.7: Imagen formato NIR.

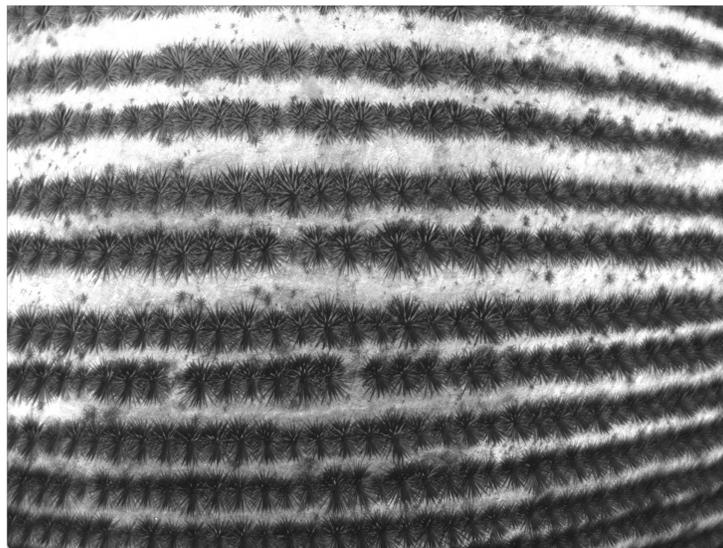


Figura 5.8: Imagen formato RED.

### 5.2.2. Muestreo de grados Brix

El muestreo de grados Brix se realizó de forma aleatoria en 24 *Agaves tequilana*. Este muestreo se realizó en los predios ubicados en la carretera que conduce de Ameca a San Martín de Hidalgo. Para la realización de este paso fueron necesario seguir la secuencia mostrada en el capítulo 4 sección 4.2.1 (Protocolo de vuelo, adquisición de imágenes y muestreo de grados Brix).

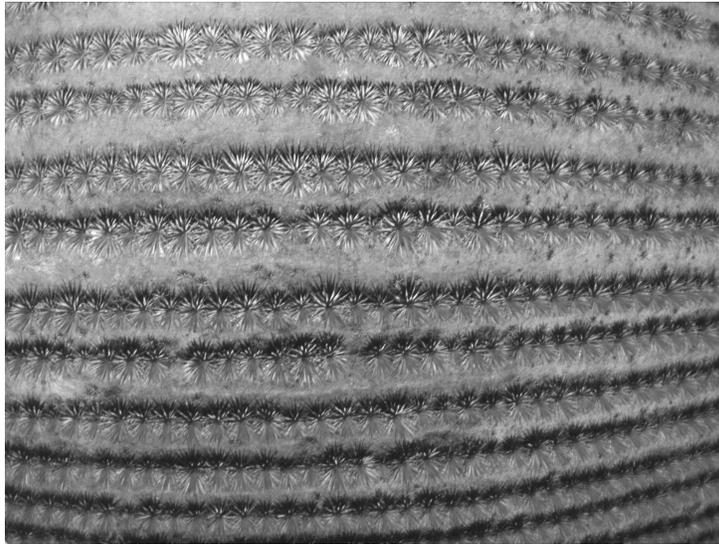


Figura 5.9: Imagen formato REG.

En la tabla 5.1 se muestran los resultados obtenidos de las mediciones realizadas de grados Brix a los 24 *Agaves tequilana* del predio anteriormente mencionado.

### 5.3. Preprocesamiento de imágenes y cálculo de índice NDVI

#### 5.3.1. Eliminación de ruido de la imagen

La imagen ingresada al algoritmo 1 que se encuentra en la sección 4.2.2, fue programado en el lenguaje Python para la eliminación de ruido de la imagen. La imagen ingresada fue en formato RGB. La imagen antes de ser procesada es mostrada en la Figura 5.10 y la imagen después de su procesamiento (eliminación de ruido/binarización) en la mostrada en la figura 5.11.

Tabla 5.1: Resultados muestreo de grados Brix.

No. de agave	Grados Brix
1	23.1
2	13.1
3	11.2
4	13.3
5	23.0
6	7.0
7	11.2
8	17.1
9	8.2
10	10.0
11	5.6
12	8.7
13	5.2
14	9.9
15	8.0
16	8.2
17	7.2
18	16.6
19	6.2
20	10.2
21	5.4
22	8.1
23	8.2
24	8.5

### 5.3.2. Cálculo del índice NDVI

En la Figura 5.12 se muestra un ejemplo de una imagen que captura el infrarrojo cercano (NIR) y en la figura 5.13 se muestra la imagen de la banda roja (RED).



Figura 5.10: Imagen RGB antes de ser procesada.

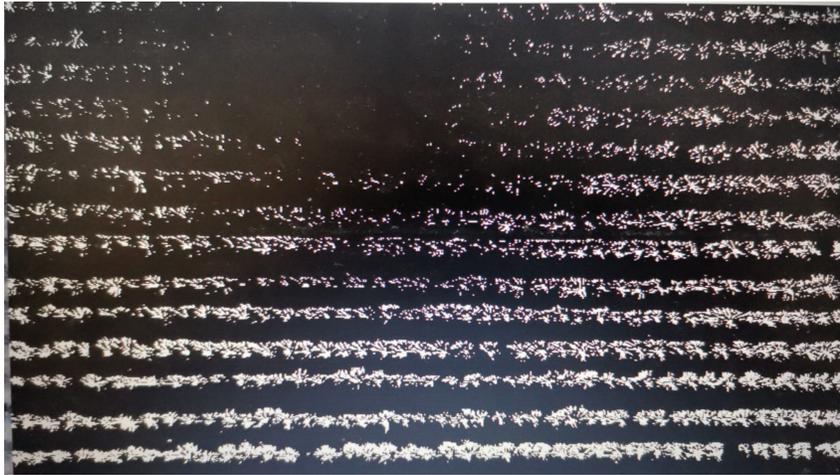


Figura 5.11: Imagen binarizada.

En la Figura 5.14 se muestra el primer resultado del cálculo del índice de vegetación normalizada NDVI.

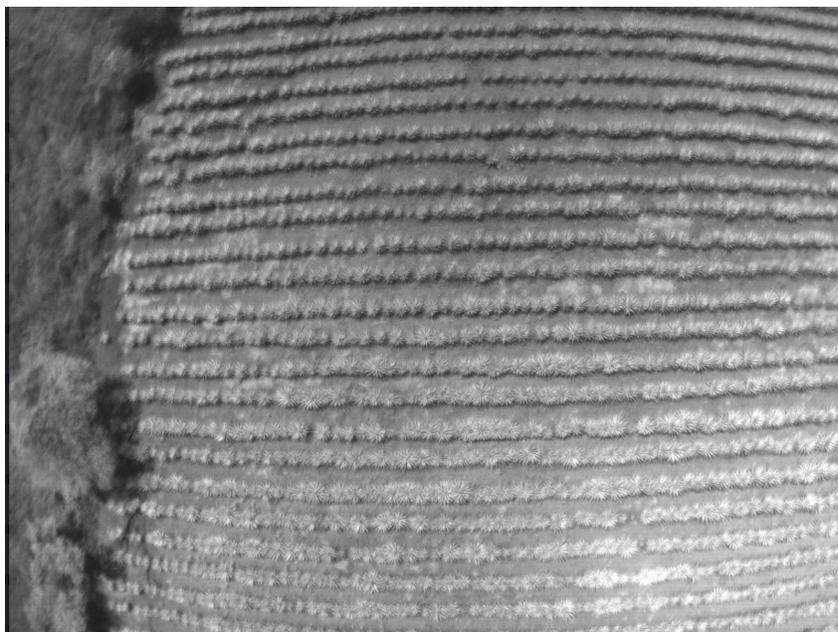


Figura 5.12: Imagen de la banda infrarrojo cercano (NIR).

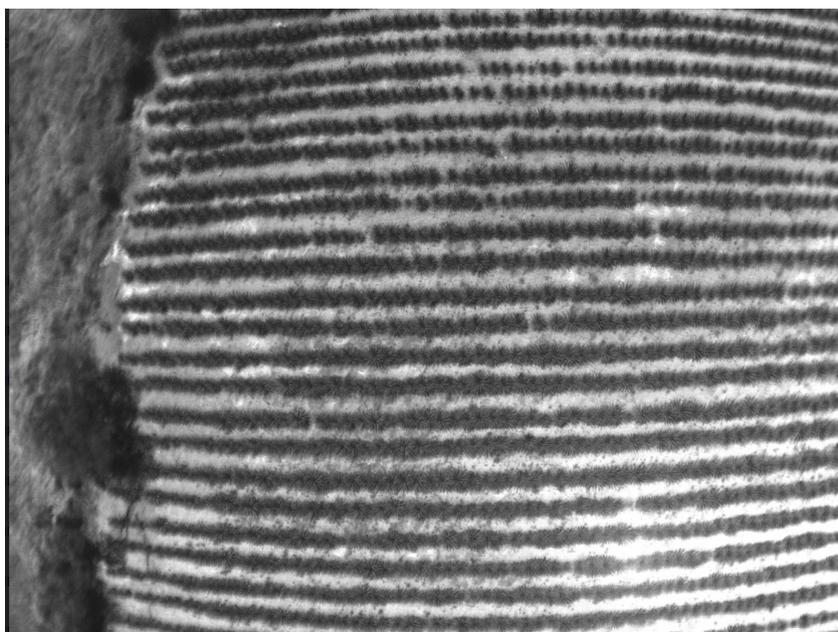


Figura 5.13: Imagen de la banda roja (RED).

## 5.4. Identificación de las técnicas de regresión para estimación

Después de realizar una revisión sistemática de la literatura (ver capítulo 3), se observó que los métodos para estimación más utilizados se encuentran el análisis de regresión

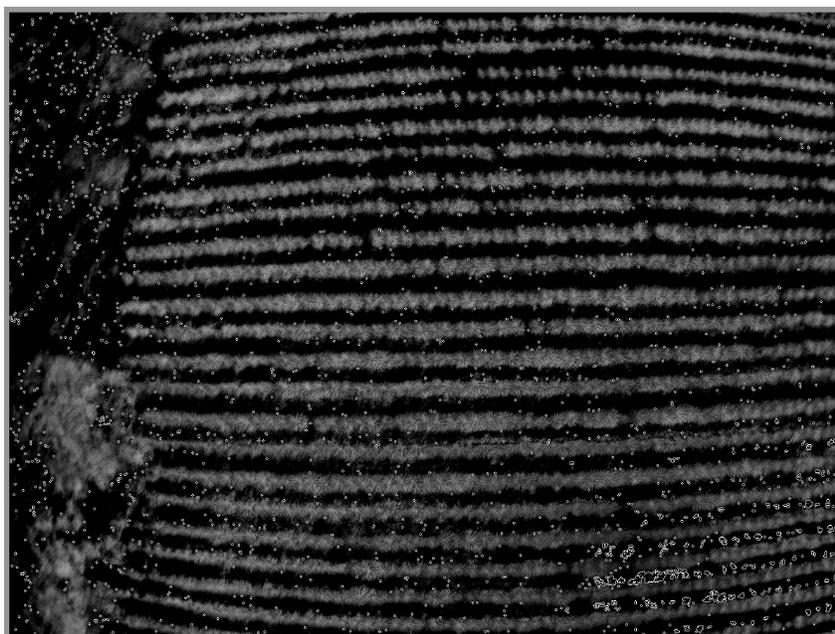


Figura 5.14: Imagen resultado del cálculo de NDVI.

lineal y análisis de regresión múltiple. En cuanto a métricas, se observó que las más utilizadas fueron el error cuadrado medio (MSE, por sus siglas en inglés), la raíz del error cuadrado medio (RMSE, por sus siglas en inglés) así como los coeficientes de correlación y el coeficiente de determinación. De igual forma, entre los algoritmos utilizados para la estimación se encuentran los algoritmos PLSR, SVR y Random Forest por mencionar algunos. En la sección 3.3 (análisis de los resultados de la revisión sistemática de la literatura) se describen los resultados para este punto.

## 5.5. Diseño e implementación del algoritmo de estimación de grados Brix

El algoritmo que fue seleccionado para la estimación de grados de Brix fue el algoritmo de regresión lineal (algoritmo 4 mostrado en la sección 4.2.4), ya que, por sus características se adecúa bien para realizar una predicción precisa teniendo como valor de entrada dos variables.

Como resultado de haber ingresado los datos de la variable dependiente "Y" grados Brix y de la variable independiente "X" siendo estos los valores de NDVI, se obtuvo un

## 5.5. DISEÑO E IMPLEMENTACIÓN DEL ALGORITMO DE ESTIMACIÓN DE GRADOS BRIX51

diagrama de regresión lineal con los datos graficados en su coordenada correspondiente (ver Figura 5.15). La ecuación de la recta de regresión resultante es mostrada a continuación.

$$y = 5,951 + 21,214X \quad (5.1)$$

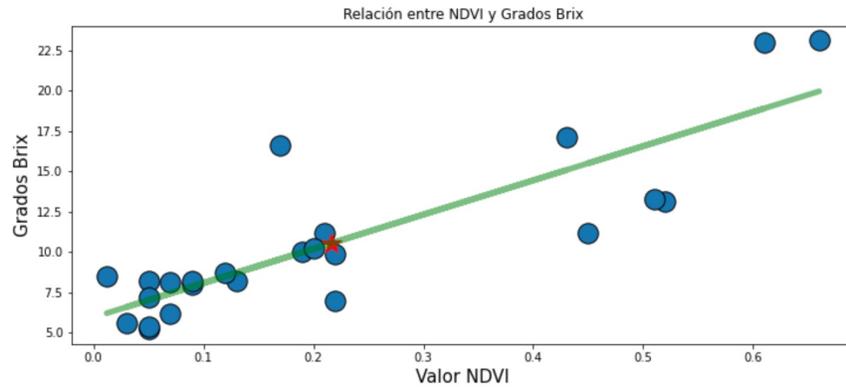


Figura 5.15: Gráfica de regresión lineal de la relación entre NDVI y grados Brix.

Este algoritmo de igual forma calcula los siguientes parámetros pertenecientes a un análisis de regresión lineal:

- Coeficiente de correlación : 0.85
- Coeficiente de determinación : 0.72

El valor de 0.85 para el coeficiente de correlación indica que existe una relación fuerte entre el valor de NDVI y los grados Brix. Mientras que el valor del coeficiente de determinación es de 0.72 indica también que la predicción es fuerte, de acuerdo a la figura 2.5 de la sección 2.6.1 que muestra la intensidad y dirección de coeficiente de correlación.

El modelo desarrollado finalmente realiza una predicción con base a la regresión lineal calculada. La predicción se realiza ingresando valores de NDVI y el modelo en automático calcula los grados Brix estimados de acuerdo a los valores ingresados. En esta prueba, se ingresaron los dos siguientes valores de NDVI y su estimación fue la mostrada a continuación.

- Valor NDVI 0.66: grados Brix estimados: 19.95

- Valor NDVI 0.20: grados Brix estimados: 10.19

En la Figura 5.16 se muestra el modelo final, con todos los parámetros resultantes del mismo. Cabe destacar que el modelo que representa la ecuación 5.2, se llevó a cabo en el lenguaje de programación Python, mismo que maneja el paquete *linnear regression* y contiene esta formula, en la cual  $\beta$  que multiplica a  $\beta_1$  es el valor de "X" que se muestra en la ecuación 2.2 de la sección 2.6.1.

$$y = \beta_0 + \beta_1\beta \quad (5.2)$$

› Ecuación de la línea de regresión:  $y = 5.95183103778812 + 21.214\beta$   
 Coeficiente de correlación: 0.851058179153913  
 Coeficiente de determinación: 0.7243000243047739  
 Los grados Brix estimados para 0.66 NDVI son 19.95 grados Brix  
 Los grados Brix estimados para 0.20 NDVI son 10.19 grados Brix

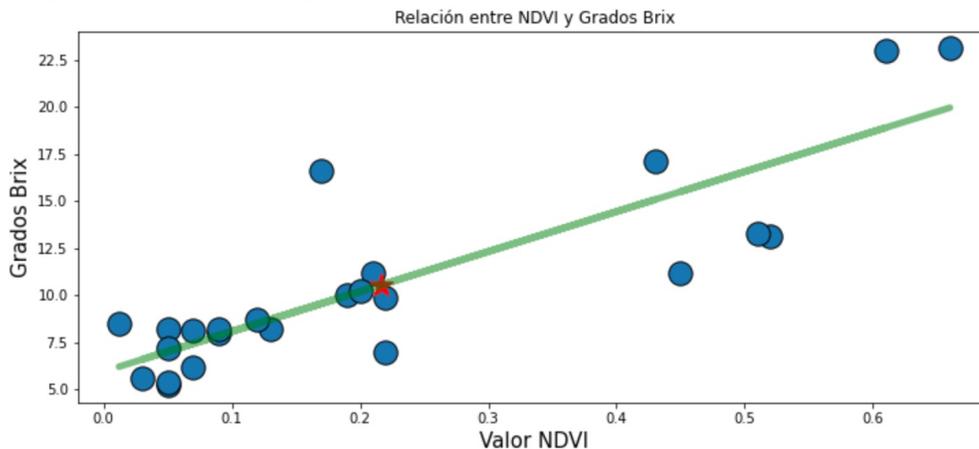


Figura 5.16: Modelo final de regresión lineal.

### 5.5.1. Análisis del algoritmo de regresión empleado

En la gráfica de la Figura 5.16, se puede observar que los puntos azules corresponden a los valores de grados Brix y valores de NDVI no se encuentran muy dispersos con una variación no muy amplia. Para este modelo, se tomó en cuenta un muestreo de grados Brix en el predio de *Agaves tequilana* que corresponde al ubicado sobre la carretera Ameca a San Martín de Hidalgo Jal. Los valores de NDVI ingresados al modelo por su parte son de igual forma adquiridos mediante el algoritmo 2 mostrado en la sección 4.2.2, los cuales corresponden a una fotografía del mismo predio y de la misma sección de los *Agaves*

## 5.5. DISEÑO E IMPLEMENTACIÓN DEL ALGORITMO DE ESTIMACIÓN DE GRADOS BRIX53

*tequilana* muestreados, es así como se ha logrado realizar el modelo. Dados los valores de las métricas utilizadas (coeficiente de correlación y coeficiente de determinación igual a 0.85 y 0.72 respectivamente) es posible afirmar que si existe relación entre los grados Brix y los valores NDVI.

# Capítulo 6

## Conclusiones

En la presente tesis se propuso un modelo que fuera capaz de predecir el total de grados Brix de un predio de *Agaves tequilana* todo esto mediante el uso de imágenes adquiridas desde un vehículo aéreo no tripulado (VANT, por sus siglas en inglés). Para la realización del modelo fue necesario desarrollar una metodología, misma que constó de cinco pasos, que van desde definir un plan de vuelo para adquirir imágenes de los predios de *Agave tequilana*, muestrear los grados Brix contenidos en los mismos, preprocesar las imágenes, pasando por calcular el índice NDVI hasta seleccionar el algoritmo de regresión que sería utilizado.

Al desarrollar este proyecto de tesis, se tuvo la oportunidad de conocer diferentes términos relacionados al mismo, tales como conocer acerca del *Agave tequilana* Weber que es la materia prima en la elaboración del tequila, los conceptos de fotogrametría, las bandas con las que trabaja el espectro electromagnético, terminología de aprendizaje máquina (machine learning, por su siglas en inglés) así como términos estadísticos referentes al análisis de regresión.

Después de tomar las imágenes de campo y muestrear los grados Brix de los *Agaves tequilana* se desarrolló el prototipo de algoritmo que fue capaz de discriminar características no relevantes en las imágenes para así poder obtener el índice de vegetación normalizada. De esta manera se pudo obtener finalmente el modelo de regresión dando como resultado una ecuación de la recta  $y = 5.951 + 21.214X$ , un coeficiente de correlación igual a 0.85 y un coeficiente de determinación igual a 0.72, los cuales indican que sí existe una correlación significativa entre el índice NDVI y los grados Brix.

El modelo desarrollado fue capaz de predecir en dos ocasiones los grados Brix ingresando valores de NDVI igual a 0.66 estimando 19.95 grados Brix y 10.19 grados Brix para

el valor NDVI de 0.20. Se puede concluir así, que los grados Brix de un predio de *Agave tequilana* sí puede ser estimado mediante el uso de imágenes adquiridas desde un vehículo aéreo no tripulado y el cálculo de un índice de vegetación.

Finalmente, es posible concluir que tanto la hipótesis como los objetivos establecidos en el presente trabajo se cumplieron.

# Referencias

- Aerial Insights. (2018). *¿Qué es la fotogrametría con drones?* (Accedido el 14/12/2020: <https://www.aerial-insights.co/blog/fotogrametria-con-drones/>)
- Alonso Romero, L., y Calonge Cano, T. (2011). *Redes neuronales y reconocimiento de patrones*. Universidad de Salamanca, Universidad de Valladolid.
- Barrenechea, L. (2020, Apr). *Ndvi: Todo sobre la tecnología que está revolucionando el agro*. El Agrario. Descargado de <https://www.elagrario.com/agricultura-ndvi-todo-sobre-la-tecnologia-que-esta-revolucionando-el-agro-10512.html>
- Bautista, J. J. R., y Garrafa, D. J. (s.f.). Desarrollo y potencialidades de la region valles, jalisco, méxico. *DE AQUÍ, DE ALLÁ.*, 289.
- Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.
- Cervantes Rendón, E., y Montano Armendáriz, G. (2018). *Medio ambiente y sociedad: Potencial de la energía renovable en México*. El Colegio de Chihuahua. (ISBN: 978-607-8214-52-5)
- Cheng, J.-H., y Sun, D.-W. (2017). Partial least squares regression (pls) applied to nir and hsi spectral data modeling to predict chemical properties of fish muscle. *Food engineering reviews*, 9(1), 36–49.
- Chitarroni, H. (2002). La regresión logística.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., y Stein, C. (2022). *Introduction to algorithms*. MIT press.
- Corona, J. E. P. (2019). *Algoritmo para el conteo de agaves usando imágenes aéreas tomadas desde un vehículo aéreo no tripulado*. (Tesis de Master no publicada). Instituto Tecnológico Ciudad Guzmán.
- CRT. (2020). *¿Qué es el tequila?* OnLine. (Accedido el 14/12/2020: <https://www.crt.org.mx/>)
- Forbes. (2018, oct). *El tequila ya es pieza clave del desarrollo económico de México*. Descargado de <https://www.forbes.com.mx/el-tequila-ya-es-pieza-clave-del>

-desarrollo-economico-de-mexico/

- Gilabert, M. A., González-Piqueras, J., y García-Haro, J. (1997). Acerca de los índices de vegetación. *Revista de teledetección*, 8(1), 1–10.
- Gómez, C. A. R. (2020). Aplicación del machine learning en agricultura de precisión. *Revista CINTEX*, 25(2), 14–27.
- Granados, R. M. (2016). Modelos de regresión lineal múltiple. *Granada, España: Departamento de Economía Aplicada, Universidad de Granada*.
- Granados Sánchez, D. (1993). *Los agaves en México* (Inf. Téc.).
- Jauregui, L. (2006). *Introducción a la fotogrametría*. OnLine. (Accedido el 14/12/2020: <http://webdelprofesor.ula.ve/ingenieria/iluis/publicaciones/Fotogrametr\%EDa/CAPITULO1.pdf>)
- Keele, S., y cols. (2007). *Guidelines for performing systematic literature reviews in software engineering* (Inf. Téc.). Technical report, ver. 2.3 ebse technical report. ebse.
- Kharuf-Gutierrez, S., Hernández-Santana, L., Orozco-Morales, R., Aday Díaz, O. d. l. C., y Delgado Mora, I. (2018). Análisis de imágenes multiespectrales adquiridas con vehículos aéreos no tripulados. *Ingeniería Electrónica, Automática y Comunicaciones*, 39(2), 79–91.
- Levin, R. I., y Rubin, D. S. (2004). *Estadística para administración y economía*. Pearson Educación.
- Manrique, E. (1999). Índice de vegetación. aplicación del ndvi. En *Teledetección. avances y aplicaciones. VIII congreso nacional de teledetección. albacete, España* (pp. 217–219).
- Montañez-Soto, J., Venegas-González, J., Vivar-Vera, M., y Ramos-Ramírez, E. (2011). Extracción, caracterización y cuantificación de los fructanos contenidos en la cabeza y en las hojas del agave tequilana weber azul. *Bioagro*, 23(3), 199–206.
- Morales, G. M.-A., y Domínguez, A. J. C. (2009). La regresión por mínimos cuadrados parciales: orígenes y evolución. *Historia de la probabilidad y la estadística (IV)*, 441–448.
- Rodríguez-Jaume, M.-J., Mora Catalá, R., y cols. (2001). Análisis de regresión múltiple.
- Ryan M., D. K., Talabis. (2015). Supervised learning. *Science Direct*.
- Tovar Blanco, A. L., Lizarazo Salcedo, I. A., y Rodríguez Eraso, N. (2020). Estimación de biomasa aérea de eucalyptus grandis y pinus spp usando imágenes sentinel1a y sentinel2a en Colombia. *Colombia forestal*, 23(1), 79–93.
- Türkmen, İ., y Ekşi, A. (2011). Brix degree and sorbitol/xylitol level of authentic pome-

granate (punica granatum) juice. *Food chemistry*, 127(3), 1404–1407.

# Anexos



# Apéndice A

## Publicación derivada de este trabajo de tesis

La Revista Ibérica de Sistemas y Tecnologías de Información esta indexada en las principales bases de datos internacionales, entre las cuales destacan Web of Science y Scopus.



Recebido/Submission: xx/12/2021  
Aceitação/Acceptance: xx/02/2021

### **Conteo de plantas de agave usando redes neuronales convolucionales e imágenes adquiridas desde un vehículo aéreo no tripulado**

Omar Hernández-Calvario<sup>1</sup>, Frida Florián<sup>1</sup>, María Guadalupe Sánchez<sup>1</sup>,  
Himer Ávila-George<sup>2</sup>

omar.calvario@cusur.udg.mx; m20291046@cdguzman.tecnm.mx;  
himer.avila@academicos.udg.mx; msanchez@itcg.edu.mx

<sup>1</sup> Departamento de Sistemas y Computación, TecNM - Instituto Tecnológico de Ciudad Guzmán, Ciudad Guzmán 49100, Jalisco, México.

<sup>2</sup> Departamento de Ciencias Computacionales e Ingenierías, Universidad de Guadalajara, Ameca 46600, Jalisco, México.

DOI: 10.17013/risti.45.64-76

Figura A.1: Publicación de artículo.

# Apéndice B

## Estancia académica

Durante la estancia en el Centro Universitario de los Valles con apoyo del Dr. Rodolfo Omar Domínguez García, se desarrollaron actividades relacionadas con la presente investigación.



UNIVERSIDAD DE GUADALAJARA

Centro Universitario de los Valles  
Maestría en Ingeniería de Software

Ameca, Jalisco a 1 de abril de 2022  
Asunto: CARTA DE FINALIZACIÓN

**M.C. ADA MABEL VÁZQUEZ PAZ**  
**COORDINADORA DE LA MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**  
**INSTITUTO TECNOLÓGICO DE CIUDAD GUZMÁN**  
**PRESENTE:**

Por este conducto, me permito informar a usted que la alumna **Frida Mayela Florián Pinto** con número de control **M20291046**, que actualmente cursa el **cuarto Semestre** de "**la Maestría en Ciencias de la Computación**", cumplió satisfactoriamente su **Estancia en Línea**, del **1 de Marzo al 1 de abril del 2022**.

Durante su estancia en esta institución la alumna continuó con el desarrollo de las actividades pertinentes a su tema de tesis "**Estimación de grados Brix en cultivos de agave utilizando imágenes adquiridas desde un vehículo aéreo no tripulado**", dirigido y supervisado por el Dr. Rodolfo Omar Domínguez García.

Sin otro particular me despido de usted con un cordial saludo.

**A T E N T A M E N T E**

**Dr. Rodolfo Omar Domínguez García.**  
**Profesor investigador titular C y**  
**Coordinador de la Maestría en Ingeniería de Software**



UNIVERSIDAD DE GUADALAJARA  
C UNIVRSITARIO  
DE LOS VALLES  
Coordinación de Investigación y  
Posgrado  
Maestría en Ingeniería de Software