



UNIVERSIDAD TÉCNICA DE COTOPAXI
FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS
CARRERA DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS
COMPUTACIONALES

PROYECTO DE INVESTIGACIÓN

**“REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS
EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”**

Proyecto de Titulación presentado previo a la obtención del Título de Ingenieras en
Informática y Sistemas Computacionales.

Autoras:

Molina Cajas Jhoanna Elizabeth

Salazar Segovia Thalía Maricruz

Tutor Académico:

Ing. Mg. Karla Susana Cantuña Flores

LATACUNGA – ECUADOR

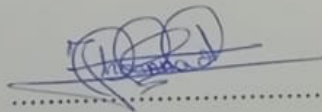
2022

DECLARACIÓN DE AUTORÍA

Nosotras, Jhoanna Elizabeth Molina Cajas con C.I: 050406902-2 y Thalia Maricruz Salazar Segovia con C.I: 050395280-6, por ser las autoras del presente proyecto de Titulación: **“REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”**, siendo la Ing. Mg. Karla Susana Cantuña Flores, tutor del presente trabajo, eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

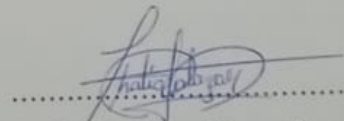
Además, certificamos que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de nuestra exclusiva responsabilidad.

Atentamente,



Jhoanna Elizabeth Molina Cajas

CI: 050406902-2



Thalia Maricruz Salazar Segovia

CI: 050395280-6

DECLARACIÓN DE AUTORÍA

Nosotras, Jhoanna Elizabeth Molina Cajas con C.I: 050406902-2 y Thalía Maricruz Salazar Segovia con C.I: 050395280-6, por ser las autoras del presente proyecto de Titulación: **“REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”**, siendo la Ing. Mg. Karla Susana Cantuña Flores, tutor del presente trabajo, eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certificamos que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de nuestra exclusiva responsabilidad.

Atentamente,

.....

Jhoanna Elizabeth Molina Cajas

CI: 050406902-2

.....

Thalía Maricruz Salazar Segovia

CI: 050395280-6



Ingeniería
Informática y Sistemas
Computacionales

AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutor del Trabajo de Investigación sobre el título: “REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”, de las señoritas Molina Cajas Jhoanna Elizabeth y Salazar Segovia Thalía Maricruz, de la carrera de Ingeniería en Informática y Sistemas Computacionales, considero que dicho Informe Investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos a la evaluación del Tribunal de Validación de Proyecto que el Consejo Directivo de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga, Marzo, 2022

Tutor de Titulación
Ing. Mg. Karla Susana Cantuña Flores
C.C.: 0502305113

AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutor del Trabajo de Investigación con el título: **“REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”**, de las estudiantes: Jhoanna Elizabeth Molina Cajas y Thalía Maricruz Salazar Segovia de la Carrera de Ingeniería en Informática y Sistemas Computacionales, considero que dicho Informe Investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos a la evaluación del Tribunal de Validación de Proyecto que el Honorable Consejo Académico de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga, marzo, 2022

.....
Ing. Mg. Karla Susana Cantuña Flores
C.C.: 0502305113



Anexo 19. Aprobación del Tribunal de Titulación

APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotacachi, y por la Facultad de Ciencias de la Ingeniería y Aplicadas; por cuanto, las postulantes: Molina Cajas Johanna Elizabeth y Salazar Segovia Thalia Maricruz con el título de Proyecto de titulación: **"REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES"**, han considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, marzo 2022

Para constancia firman:

Lector 1 (Presidente)

Nombre: Ing. Jorge Rubio

CC: 050222229-2

Lector 2

Nombre: Ing. Mg Manuel Villa

CC: 180338695-0

Lector 3

Nombre: Ing. MSc. José Cadena

CC: 050155279-8

APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la Facultad de **CIENCIAS DE LA INGENIERÍA Y APLICADAS**; por cuanto, los postulantes: **JHOANNA ELIZABETH MOLINA CAJAS** y **THALIA MARICRUZ SALAZAR SEGOVIA**, con el título del proyecto de investigación: **“REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”**, ha considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación del Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, marzo, 2022

.....

Lector 1 (Presidente)

Nombre: Ing. Jorge Rubio

CI: 050222229-2

.....

Lector 2

Nombre: Ing. Mg Villa Quishpe

Manuel **CI:** 180338695-0

.....

Lector 3

Nombre: Ing. MSc. José Augusto Cadena

Moreano

CI: 050155279-8

AGRADECIMIENTO

Dicen que la mejor herencia que nos pueden dejar los padres son los estudios, sin embargo, no creo que sea el único legado del cual yo particularmente me siento muy agradecida, mis padres me han permitido trazar mi camino y caminar con mis propios pies. Ellos son mis pilares fundamentales de vida, con este trabajo de titulación, quiero agradecer a Giovanni, Mery y a mi Hermana Sol, por creer en mí y por el esfuerzo que hicieron al brindarme la educación.

Gracias también a mis maestros quienes me formaron en esta carrera con paciencia y dedicación, sobre todo agradecer a mi tutora de tesis la Ing. Karla Cantuña quien nos supo guiar con perseverancia hasta culminar nuestro propósito y paso por la prestigiosa UTC, así también gracias a mi amiga Thalia por ser ese eje y complemento en las buenas y malas durante el procesos de tesis, ya que por motivos de mi trabajo se me hacían difíciles cumplir, pero siempre fue incasable su apoyo incondicional.

Jhoanna Elizabeth Molina Cajas

Mis sinceras palabras de agradecimiento las dedico a todas las personas que de uno u otro modo contribuyeron para alcanzar esta meta y de manera más especial a mi hermana Maricruz pues siempre supo extenderme su mano no solo durante mi carrera sino también en incontables aspectos de la vida, te quiero inmensamente hermana.

Thalia Maricruz Salazar Segovia

DEDICATORIA

A Dios y a mis ángeles protectores que están en el cielo mis abuelitos Aurora Tovar y Segundo Cajas quienes han sido mi protección diaria y fuerza absoluta para salir adelante sin rendirme.

A mis padres Giovanni y Mery, a mi hermana Soledad, quienes con su amor, paciencia y esfuerzo me han permitido llegar a cumplir hoy un sueño más y esta meta universitaria que fue muy dura pero no imposible, gracias por inculcar en mí el ejemplo de esfuerzo y valentía, de no temer a las adversidades porque Dios y el Manuelito de Isinche siempre están conmigo.

A, mi abuelito Luis, a mis tíos Amada, Hugo y Diego, a mis primos, Ashly, Luis, Nicolás, Eduardo y Bryan por su cariño y apoyo incondicional, durante todo este proceso, al estar conmigo en todo momento, gracias por sus oraciones, consejos y palabras de aliento que hicieron de mí una mujer perseverante y luchadora, ya que siempre me acompañan en todos mis sueños y metas.

A mi enamorado Cristian, por ser ese amigo incondicional de lucha constante quien no ha permitido que me derrote y me detenga para cumplir mis sueños, siendo la inspiración de perseverancia y constancia, a mi hijita de 4 patas Lía por ser mi paño de lágrimas en los momentos más difíciles al final de mi carrera y a todos quienes formaron parte de esta etapa de mi vida, gracias por confiar en mí.

Jhoanna Elizabeth

DEDICATORIA

El presente proyecto de tesis lo dedico a mis hijos Brisa y Aron, pues ellos fueron la motivación más fuerte que encontré a lo largo de este duro pero incansable camino, no hay nadie en este mundo que merezca más reconocimiento que mis amados pequeños que hoy ven cumplir nuestro sueño.

Thalía Maricruz

UNIVERSIDAD TÉCNICA DE COTOPAXI
FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS

TÍTULO: “REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES”

Autores:

Jhoanna Elizabeth Molina Cajas

Thalía Maricruz Salazar Segovia

RESUMEN

El presente proyecto de investigación, describe el análisis de revisión sistemática literaria (RLS) con técnicas de aprendizaje de máquinas en especies vegetales. En RLS se realizó un análisis de diferentes artículos, en base: reconocimiento de especies vegetales durante los últimos 5 años y su frecuencia de uso, para ello se utilizó términos de búsqueda, selección de protocolos de búsqueda y recursos de búsqueda en: Science Direct, Google académico y Scielo. Se aplicó la selección de búsqueda bajo criterios de calidad a los 23 artículos de los cuales fueron aprobados 7, los cuales serán empleados en la investigación. Como resultado se eligió las especies vegetales menta y orégano en base a características físicas y a su vez la técnica de aprendizaje de máquina regresión logística. Para el desarrollo de la aplicación móvil, se utilizó una base de datos de 500 fotografías entre menta y orégano, para la aplicación del algoritmo de regresión logística se utilizó como datos de entrada a 25 atributos resultantes de la segmentación de imágenes (Convolución, Difuminación, Luminosidad, Densidad, Umbralización) utilizando los filtros Blur, Canny, Dilate, Eroded y la imagen original para cada uno de ellos y dos datos binarios de salida para la determinación del nombre de la planta. Para el reconocimiento de la planta se utilizó la función $(y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ para } i = 1, 2, n(2,29))$. Finalmente, se realiza las pruebas de funcionamiento para la verificación del funcionamiento de la aplicación móvil.

Palabras claves: aprendizaje de máquinas, especies vegetales, regresión logística, RLS, aplicación móvil.

TECHNICAL UNIVERSITY OF COTOPAXI

FACULTY OF ENGINEERING SCIENCES AND APPLIED

THEME: “REVIEW OF MACHINE LEARNING TECHNIQUES USED IN THE RECOGNITION OF PLANT SPECIES”

Authors:

Jhoanna Elizabeth Molina Cajas

Thalía Maricruz Salazar Segovia

ABSTRACT

This research project describes the analysis of systematic literature review (RLS) with machine learning techniques in plant species. In RLS, an analysis of different articles was carried out, based on: recognition of plant species during the last 5 years and their frequency of use, for which search terms, selection of search protocols, and search resources were used in Science Direct, Academic Google and Scielo. The search selection was applied under quality criteria to the 23 articles of which 7 were approved, which will be used in the investigation. As a result, the mint and oregano plant species were chosen based on physical characteristics and, in turn, the logistic regression machine learning technique. For the development of the mobile application, a database of 500 photographs between mint and oregano was used, for the application of the logistic regression algorithm, 25 attributes resulting from the segmentation of images were used as input data (Convolution, Diffusion, Luminosity, Density, Threshold) using the filters Blur, Canny, Dilate, Eroded and the original image for each of them and two binary output data for determining the name of the plant. For the recognition of the plant, the function $(y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i=1,2,n(2,29)$ was used. Finally, the performance tests are carried out to verify the operation of the mobile application.

Keywords: machine learning, plant species, logistic regression, RLS, mobile application.

AVAL DE TRADUCCIÓN

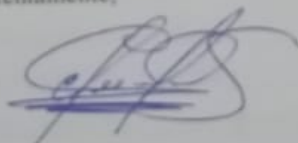
En calidad de Docente del Idioma Inglés del Centro de Idiomas de la Universidad Técnica de Cotopaxi; en forma legal CERTIFICO que:

La traducción del resumen al idioma Inglés del trabajo de titulación cuyo título versa: **"REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS, USADAS EN EL RECONOCIMIENTO DE ESPECIES VEGETALES"** presentado por **Jhoanna Elizabeth Molina Cajas** y **Thalia Maricruz Salazar Segovia**, estudiantes de la Carrera de **Ingeniería en Informática y Sistemas Computacionales** perteneciente a la **Facultad de Ciencias de la Ingeniería y Aplicadas** lo realizaron bajo mi supervisión y cumple con una correcta estructura gramatical del Idioma.

Es todo cuanto puedo certificar en honor a la verdad y autorizo a las peticionarias hacer uso del presente aval para los fines académicos legales.

Latacunga, 22 marzo del 2022

Atentamente,



CENTRO
DE IDIOMAS

Mg. Marco Paúl Beltrán Semblantes

DOCENTE CENTRO DE IDIOMAS-UTC
CI: 0502666514

Contenido

PORTADA.....	i
DECLARACIÓN DE AUTORÍA.....	ii
AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN	iv
APROBACIÓN DEL TRIBUNAL DE TITULACIÓN	vi
<i>AGRADECIMIENTO</i>	viii
RESUMEN	xi
ABSTRACT.....	xii
ÍNDICE DE TABLAS	xvii
ÍNDICE DE FIGURAS.....	xviii
ÍNDICE DE ANEXOS	xx
1. INFORMACIÓN GENERAL.....	7
2. INTRODUCCIÓN	8
2.1. PROBLEMA.....	8
2.1.1. Situación problemática.....	8
2.1.2. Formulación del problema	9
2.2. OBJETO Y CAMPO DE ACCIÓN.....	9
2.2.1. Objeto de estudio	9
2.2.2. Campo de acción.....	9
2.3. BENEFICIARIOS	9
2.3.1. Beneficiarios Directos.....	9
2.3.2. Beneficiarios Indirectos	9
2.4. JUSTIFICACIÓN	10
2.5. HIPÓTESIS	10
2.6. OBJETIVOS	11
2.6.1. Objetivo General.....	11
2.6.2. Objetivos Específicos.....	11
2.7. SISTEMA DE TAREAS	12
3. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA	13
3.1. ANTECEDENTES	13

3.2. FUNDAMENTACIÓN TEÓRICA	14
3.2.1. Especies vegetales.....	14
3.2.2. Plantas vasculares	14
3.2.3. Clasificación de las plantas.....	15
3.2.4. Clasificación de las plantas por su utilidad.....	17
3.2.5. Aprendizaje automático	19
3.2.6. Reconocimiento	22
3.2.7. Aplicación móvil.....	29
3.2.8. Herramientas de software	30
3.2.9. Metodología SRL.....	31
3.2.10. Metodología KDD	32
3.2.11. Metodología Mobile-D	34
4. MATERIALES Y MÉTODOS	35
4.1. TIPOS DE INVESTIGACIÓN.....	35
4.1.1. Investigación bibliográfica.....	35
4.1.2. Investigación explicativa	35
4.1.3. Investigación aplicada.....	35
4.2. MÉTODO TEÓRICO	35
4.2.1. Método Hipotético deductivo	35
4.3. MÉTODO EMPÍRICO	36
4.3.1 Técnicas de Investigación	36
4.3.2. Instrumentos de investigación.....	36
4.3.3. Población y muestra.....	36
4.3.4. Cálculo de la muestra.....	37
4.4. MÉTODO ESPECÍFICO	37
4.4.1. Metodología RSL.....	37
4.4.2. Metodología KDD	37
4.4.3. Metodología Mobile-D	37
5. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	38
5.1. Metodología de revisión sistemática literaria (RSL).....	38
5.1.1. Definición de la búsqueda.....	38
5.1.2. Ejecución de búsqueda.....	40
5.1.3. Discusión de resultados.....	44
5.2. Metodología KDD	45

5.3. Metodología Mobile-D	58
5.4. Comprobación de la hipótesis	67
5.5. Análisis de impacto	68
5.5.1. Impacto social	68
5.5.2. Impacto tecnológico	68
5.5.3. Impacto ambiental.....	68
5.5.4. Impacto económico	68
6. CONCLUSIONES Y RECOMENDACIONES	71
6.1. Conclusiones	71
6.2. Recomendaciones	71
7. BIBLIOGRAFÍA	72
8. ANEXOS	75

ÍNDICE DE TABLAS

Tabla 1. Planificación de las actividades	12
Tabla 2. Característica de las plantas vasculares	14
Tabla 3. Características de plantas sin flor	15
Tabla 4. Características de las plantas con flores.....	16
Tabla 5. Características de las plantas angiospermas	16
Tabla 6. Características de las plantas angiospermas	17
Tabla 7. Características de las plantas medicinales	18
Tabla 8. Características de las plantas alimenticias	18
Tabla 9. Recursos utilizados para la RSL	39
Tabla 10. Búsqueda de palabras claves en las bases de datos.....	39
Tabla 11. Número de documentos obtenidos por cadena de búsqueda	41
Tabla 12. Criterios de calidad	41
Tabla 13. Evaluación de artículos según los criterios de calidad.....	42
Tabla 14. Artículos con técnicas de aprendizaje de maquinas.....	43
Tabla 15. Frecuencia de algoritmos utilizadas en los artículos.....	44
Tabla 16. Plan de proyecto de la metodología KDD	46
Tabla 17. Valores del resultado de Iteración 1.....	54
Tabla 18. Valores del resultado de Iteración 2.....	55
Tabla 19. Valores del resultado de Iteración 3.....	56
Tabla 20. Valores del resultado de Iteración 4.....	57
Tabla 21. Valores del resultado de Iteración 5.....	58
Tabla 22. Historia de usuario capturar imagen	59
Tabla 23. Historia de usuario buscar en la galería	59
Tabla 24. Historia de usuario detención a tiempo real.....	59
Tabla 25. CU001-Capturar imagen.....	60
Tabla 26. CU002-Ingresa a la galería	61
Tabla 27. CU003 Detención a Tiempo Real	62
Tabla 28. Caso de prueba capturar imagen	65
Tabla 29. Caso de prueba seleccionar imagen	66
Tabla 30. Caso de prueba capturar imagen	66

ÍNDICE DE FIGURAS

Figura 1. Clasificación de las plantas	15
Figura 2. Clasificación de las plantas por su utilidad	17
Figura 3. Tipos de aprendizaje automático [17]	20
Figura 4. Aprendizaje supervisado [15].....	20
Figura 5. Aprendizaje no supervisado [3].....	21
Figura 6. Aprendizaje profundo [3]	21
Figura 7. Aprendizaje por refuerzo [3]	22
Figura 8. Borde de una imagen digital en escala de grises [18]	24
Figura 9. Detección de bordes en imágenes en escala de grises [18]	24
Figura 10. Filtro de imagen por dilatación y erosión [18]	25
Figura 11. Filtro GaussianBlur [18].....	25
Figura 12. Clasificación de aprendizaje supervisada.....	26
Figura 13. Clasificación de aprendizaje supervisado-clasificación.....	26
Figura 14. Aprendizaje supervisado por regresión [15]	28
Figura 15. Clasificación de aprendizaje no supervisado [20].....	29
Figura 16. Proceso de la metodología SRL	31
Figura 17. Etapas de la Metodología KDD.....	32
Figura 18. Base de datos de plantas medicinales.....	47
Figura 19. Transformación de datos	48
Figura 20. Utilización de función cv.threshold.....	48
Figura 21. Obtención de valores de orégano	49
Figura 22. Valores de segmentación de datos.....	49
Figura 23. Leer data set.....	50
Figura 24. Creación de datos de entrenamiento y salida	50
Figura 25. Gráfica de la segmentación	50
Figura 26. Creación del Modelo Keras.....	50
Figura 27. Creación del gradiente descendiente	50
Figura 28. Comportamiento de pérdida y precisión	51
Figura 29. Modelo de regresión logística	51
Figura 30. Comportamiento de la regresión logística.....	52
Figura 31. K -fold Iteración N° 1.....	53
Figura 32. Comprobación de pérdida y precisión Iteración 1.....	53
Figura 33. Resultado ejecutado de la Iteración 1	53
Figura 34. K -fold Iteración N° 2.....	54
Figura 35. Comprobación de pérdida y precisión Iteración 2.....	54
Figura 36. Resultado ejecutado de la Iteración 2.....	54
Figura 37. K -fold Iteración N° 3.....	55
Figura 38. Comprobación de pérdida y precisión Iteración 3.....	55
Figura 39. Resultado ejecutado de la Iteración 3.....	55
Figura 40. K -fold Iteración N° 4.....	56
Figura 41. Comprobación de pérdida y precisión Iteración 4.....	56

Figura 42. Resultado ejecutado de la Iteración 4.....	56
Figura 43. K -fold Iteración N° 5.....	57
Figura 44. Comprobación de pérdida y precisión Iteración 5.....	57
Figura 45. Resultado ejecutado de la Iteración 5.....	57
Figura 46. Diagrama de cas de uso general	60
Figura 47. Diagrama de secuencia de captura de imagen.....	62
Figura 48. Diagrama de secuencia de ingresar galería	63
Figura 49. Diagrama de secuencia de detección a tiempo real.....	63
Figura 50. Diagrama de clases del aplicativo	64
Figura 51. Interfaz gráfica de la pantalla principal de la aplicación.....	64
Figura 52. Interfaz gráfica detección a tiempo real	64
Figura 53.Caso de prueba-Capturar imagen	65
Figura 54. Caso de prueba seleccionar imagen.....	66
Figura 55. Caso de prueba capturar imagen.....	67
Figura 56. Base de datos Science Direct.....	76
Figura 57. Base de datos Google académico	76
Figura 58. Base de datos Scielo	76

ÍNDICE DE ANEXOS

Anexo A. Ficha bibliográfica	75
Anexo B. Selección de artículos en base a criterios de calidad.....	76
Anexo C. Ficha técnica de la aplicación	77
Anexo D. Juicio de expertos.....	79
Anexo E. Costo del desarrollo de la investigación teórica y técnica.....	80

1. INFORMACIÓN GENERAL

Título del proyecto: “Revisión de técnicas de aprendizaje de máquinas, usadas en el reconocimiento de especies vegetales”.

Fecha de inicio: 25 de octubre del 2021.

Fecha de finalización: 23 de febrero del 2022

Lugar de ejecución: Cotopaxi- Latacunga

Facultad que auspicia: Facultad en Ciencias de la Ingeniería y Aplicadas.

Carrera que auspicia: Carrera de Ingeniería en Informática y Sistemas Computacionales.

Proyecto de investigación vinculado: Análisis del reconocimiento de imágenes en tiempo real de especies medicinales como el orégano y la menta mediante dispositivos móviles.

Equipo de trabajo:

Coordinador:

Nombre: Ing. Mgs. Karla Susana Cantuña Flores

E-mail: karla.cantuna@utc.edu.ec

Estudiantes:

Nombre: Jhoanna Elizabeth Molina Cajas

Correo: jhoanna.molina2@utc.edu.ec

Nombre: Thalía Maricruz Salazar Segovia

Correo: thalia.salazar2806@utc.edu.ec

Área del conocimiento: Tecnologías de la Información y Comunicación.

Línea de investigación: Tecnologías de la información y comunicación (TICS) y Diseño Gráfico.

Sub-línea de investigación de la carrera: Inteligencia artificial e inteligencia de negocios.

2. INTRODUCCIÓN

2.1. PROBLEMA

2.1.1. Situación problemática

El sector agrícola cumple un rol muy importante en el desarrollo económico del mundo junto con la tecnología han sido piezas claves para el incremento de ganancias económicas, con la optimización del tiempo y los recursos. Esta tecnología “Machine Learning” o conocida como aprendizaje de máquinas ha tomado gran relevancia en la agricultura [1]. Por lo tanto, con el transcurrir del tiempo la utilización de técnicas de Machine Learning en comparación con los métodos manuales, permiten aumentar la productividad de los cultivos y así cumplir con la alta demanda de alimentos a nivel mundial generada por el rápido crecimiento de la población y las zonas urbanas. Otros aspectos son los sistemas de identificación actuales en la agricultura los cuales identifican a la planta en base a un conjunto de rasgos, generalmente las características pertenecen a un solo grupo: texturales, cromáticas, geométricas o de venación. Estos sistemas de identificación son eficientes, sin embargo, en la literatura actual, no existe un estudio que ilustre la importancia de cada grupo de características y su factor discriminante.

A nivel mundial el reconocimiento de especies vegetales con , Machine Learning ha sido un desafío por los tipos de características (geométricas, texturales y cromáticas) y la precisión de clasificación [2]. Como el estudio realizado en México con el nombre de “Análisis Comparativo de las técnicas utilizadas en un sistema de reconocimiento de hojas de planta”. Se enfocan en comparar el desempeño de varios clasificadores de Machine Learning mediante la extracción de características y la clasificación, centrado en la agricultura.

Otro estudio realizado en Colombia denominado como: “Machine Learning en la detección de enfermedades en planta” [1]. A través del procesamiento de imágenes que permitan detectar en tallos, ramas, hojas y flores características propias de las diferentes enfermedades que pueden afectar a una planta y con ello generar una red de entrenamiento con una gran variedad de muestras para hacer más precisa la detección de enfermedades.

En el Ecuador [3] la Universidad Católica realizaron un estudio de “Aplicación móvil para la detección y tratamiento de daños de los cultivos de la parroquia Taura del cantón Durán, mediante el uso de software de análisis de imagen basado en técnicas de Machine Learning” se basó el estudio en el potencial bajo en la producción se debe a que los agricultores no cuentan con un conocimiento técnico para enfrentar la presencia de plagas, por lo que

proceden a comprar cualquier tipo de insumos sin el debido asesoramiento, los mismos que al ser aplicados en el cultivo, en lugar de combatir las plagas van matando el suelo, dejándolo estéril.

Con los estudios anteriormente analizados es importante debido al beneficio al agricultor a través de la tecnología que es usada para la agricultura, el lento crecimiento, la producción, cuidado, mantenimiento y cosecha que conservan técnicas manuales que no abastecen la cantidad de productos necesario para un rango de tiempo [3]. Contar con una aplicación móvil utilizando una agricultura de precisión donde se hace uso de la tecnología. Por ello, en base a estos problemas identificados en el sector agrícola en Latacunga se ha optado en la revisión literaria sobre la utilización de aprendizaje de máquinas para el reconocimiento de especies vegetales.

2.1.2. Formulación del problema

¿Cuáles son las técnicas de aprendizaje de máquinas que se utilizan en el reconocimiento de especies vegetales en el área de la agricultura?

2.2. OBJETO Y CAMPO DE ACCIÓN

2.2.1. Objeto de estudio

Técnicas de aprendizaje de máquina usadas en el reconocimiento de especies vegetales

2.2.2. Campo de acción

Aprendizaje de máquina en la agricultura.

2.3. BENEFICIARIOS

2.3.1. Beneficiarios Directos

Alumnos y profesores de la Universidad Técnica de Cotopaxi

2.3.2. Beneficiarios Indirectos

Investigadores en el área de Machine Learning y de la agricultura.

2.4. JUSTIFICACIÓN

En la actualidad la tecnología se usa de acuerdo a las necesidades y requerimientos por lo cual se optó hacer uso de tecnología de aprendizaje de máquinas en el sector agrícola específicamente en reconocimiento de especies vegetales a través de la utilización de técnicas de aprendizaje de máquinas. La actividad agrícola puede tecnificar y aprovechar las herramientas informáticas, internet y los servicios de procesamiento de datos a través de técnicas de Machine Learning (aprendizaje automático), por ende, la inclusión de la tecnología será pilar fundamental para la evolución en el área agrícola.

Reconocer de manera manual 370 especies vegetales es muy extenso y dificulta al agricultor en tiempo y costo, motivo por el cual se busca optimizar este proceso de manera automática a través de una revisión sistemática literaria, el cual sirva de guía con relación a cuáles son las técnicas de aprendizaje de máquinas en el reconocimiento de especies vegetales a través de palabras claves y criterios de selección para la elección de artículos.

Esta investigación indudablemente beneficiará a los investigadores en el área de Machine Learning y de agricultura que buscan información detallada sobre la tecnología que se utiliza en la agricultura, la cual mostrará datos minuciosos sobre las diferentes técnicas de Machine Learning y la selección de la especie vegetal que fue aplicado. Otro motivo fue que no existe una revisión literaria sistemática sobre los avances de reconocimiento de imágenes en especies vegetales, es por ello que existen cada vez menos aplicaciones que ayudan al agricultor a la supervisión y gestión automática de los procesos del crecimiento y desarrollo de las plantas.

El objetivo principal del proyecto de investigación es tener un análisis de revisión sistemática literaria sobre las diferentes técnicas de aprendizaje de máquinas utilizadas en la agricultura mediante esto se irá mostrando el análisis, diseño, desarrollo y validación de técnicas de aprendizaje de máquinas que permitan al agricultor obtener los resultados de las imágenes y modelado de las especies vegetales de acuerdo al color que se vaya destinando a cada estimación del proceso de cobertura vegetal dentro de los cultivos.

2.5. HIPÓTESIS

La aplicación de la técnica de regresión logística del aprendizaje de máquina, permitirá garantizar el reconocimiento de imágenes de las especies vegetales en el área de la agricultura.

2.6. OBJETIVOS

2.6.1. Objetivo General

- Determinar las principales técnicas de aprendizaje de máquinas por medio de la revisión sistemática literaria para el reconocimiento de especies vegetales en el área de la agricultura durante los cinco últimos años.

2.6.2. Objetivos Específicos

- Analizar la literatura científica de los últimos cinco años relacionados con técnicas de aprendizaje de máquina y reconocimiento de especies vegetales mediante fuentes bibliográficas que sirvan de base teórica para la investigación.
- Distinguir las características y frecuencia de uso de las técnicas de aprendizaje de máquinas usadas en el reconocimiento de especies vegetales mediante la revisión sistemática literaria.
- Ejemplificar una de las principales técnicas de aprendizaje de máquina en el reconocimiento de plantas medicinales mediante la metodología KDD para la demostración de eficiencia del algoritmo.

2.7. SISTEMA DE TAREAS

Tabla 1. Planificación de las actividades

OBJETIVOS ESPECÍFICOS	ACTIVIDADES	RESULTADO DE LAS ACTIVIDADES	DESCRIPCIÓN (TÉCNICAS E INSTRUMENTOS)
Analizar la literatura científica de los últimos cinco años relacionados con técnicas de aprendizaje de máquina y reconocimiento de especies vegetales mediante fuentes bibliográficas que sirva de base teórica para la investigación.	-Búsqueda de información en revistas científicas. -Búsqueda de información en artículos científicos. -Búsqueda de información en libros y proyectos similares, redacción del marco teórico.	Marco Teórico	Técnica: Análisis bibliográfico Instrumento: Ficha bibliográfica
Distinguir las características y frecuencia de uso de las técnicas de aprendizaje de máquinas usadas en el reconocimiento de especies vegetales mediante la revisión sistemática literaria.	-Selección de la técnica de aprendizaje de máquinas. -Selección del algoritmo.	Técnica de aprendizaje de máquina.	Técnica: Análisis bibliográfico Instrumento: Ficha bibliográfica Fotografía
Ejemplificar una de las principales técnicas de aprendizaje de máquina en el reconocimiento de plantas medicinales mediante la metodología KDD para la demostración de eficiencia del algoritmo.	-Aplicación de la metodología KDD -Desarrollo del código fuente en base a las herramientas de software libre. -Creación de un plan de pruebas.	Desarrollo de la aplicación móvil.	Metodología KDD- Aplicación móvil

Elaborado por: Equipo de trabajo

3. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA

3.1. ANTECEDENTES

a. Aplicación móvil para la detección y tratamiento de daños de los cultivos de la parroquia Taura del cantón Durán, mediante el uso de software de análisis de imagen basado en técnicas de Machine Learning.

Este proyecto se basa en identificar las plagas y enfermedades más conocidas de la parroquia Taura del cantón Durán, obtener una base de datos de imágenes de cada plaga, enfermedad o daño para entrenar una red neuronal con algoritmos de redes convolucionales usados en Machine Learning [3]. La aplicación muestra un detector visual a través de la cámara que identifica si la planta enfocada posee una plaga presente consultando con su red neuronal entrenada con ejemplos de un número seleccionado de plagas por tipo de cultivo. y generar un modelo que sea capaz identificar las plagas, enfermedades y daños que estén presentes en la planta capturado por el sensor de imagen de los dispositivos móviles, al completarla.

Está desarrollada una aplicación móvil que utiliza librería nativas de las plataformas, con el propósito de mostrar como resultado las plaga que posean presencia en la planta con su descripción, características y recomendaciones para tratarlas y poder cuidar los cultivos de una manera más eficiente, ahorrando en dispositivos de altos costes, mano de obra especializada o compañías que ofrecen servicios similares, instruyendo a los agricultores novatos o experimentados las posibilidades del uso de la tecnología en la actividad agrícola.

b. Reconocimiento automático de plantas medicinales usando Técnicas de aprendizaje automático

Este proyecto se basa en el uso de técnicas de visión artificial y aprendizaje automático donde se recogieron hojas de 24 especies diferentes de plantas medicinales [4]. Este proceso se realizó mediante la recopilación de fotografías con un teléfono inteligente en un laboratorio. Se extrajeron una gran cantidad de características de cada hoja. Como su longitud, ancho, perímetro, área, número de vértices, color, perímetro y área del casco. Varias características derivadas fueron luego calculadas a partir de estos atributos. Los mejores resultados fueron obtenidos de un clasificador de bosque aleatorio utilizando una técnica de validación cruzada de 10 veces. Dando como resultado una precisión del 90,1%. Este proyecto utiliza un clasificador forestal funcionó mejor que otros tipos de aprendizaje automático.

3.2. FUNDAMENTACIÓN TEÓRICA

3.2.1. Especies vegetales

Una parte que se encarga del estudio de las plantas es la botánica [5]. En la botánica se desarrolla el conocimiento de las plantas y la observación de cómo crecen y dónde, cómo se desarrollan y se reproducen sus raíces, pétalos, hojas, lugares, temporadas, climas, en definitiva, conocimiento del material vegetal. Además, las plantas proporcionan una gran multitud de recursos naturales para la humanidad. Ellos proporcionan la base de los alimentos, la mayor parte de los medicamentos y muchos otros materiales esenciales para la vida diaria.

A nivel mundial se estima que existen actualmente 370.000 especies conocidas [6]. Las especies vegetales en Ecuador se registran un alto número de especies de plantas y flores. Ecuador ocupa el séptimo lugar en el mundo en lo que respecta a biodiversidad de plantas vasculares. Ecuador recoge 25.000 especies de plantas vasculares. De igual manera, aquí se concentra el 10% de todas las especies de plantas que hay en el planeta. De este porcentaje, la mayor cantidad crece en la cordillera de los Andes, en la zona noroccidental, donde se calcula que hay aproximadamente 10.000 especies. Los Andes comprende 1.050 especies, mientras que 850 fueron inventariadas en la región amazónica y a lo largo de la Costa [7].

3.2.2. Plantas vasculares

Las plantas vasculares disponen de hojas, tallo y raíces y tienen un sistema vascular formado por vasos conductores. También son conocidas por dos nombres: plantas cormofitas o superiores [8].

De las plantas vasculares 15.306 son especies nativas confirmadas, un total de 4.176 especies se registró según el Catálogo como “endémicas al Ecuador”, es decir, no registradas en ningún otro país como especie nativa. En la Tabla 2 detalla las características principales de las plantas vasculares.

Tabla 2. Característica de las plantas vasculares

Características	Definición
Alimentación	Son seres autótrofos (generan su propio alimento)
Fotosíntesis	Sintetizan la energía química a partir de la luz. Este produce oxígeno
Sésiles	Carecen de capacidad de locomoción.

Elaborado por: Los investigadores

3.2.3. Clasificación de las plantas

Para la clasificación de las plantas se basa en sus órganos fundamentales, como: raíces, tallos, hojas, flores y frutos. Esta clasificación está basada en las estructuras morfológicas y características particulares de la planta lo que permite identificar hasta el nivel de especie [8].

Se clasifican en dos grandes grupos: plantas sin flor y plantas con flor como se muestra en la **Figura 1**.

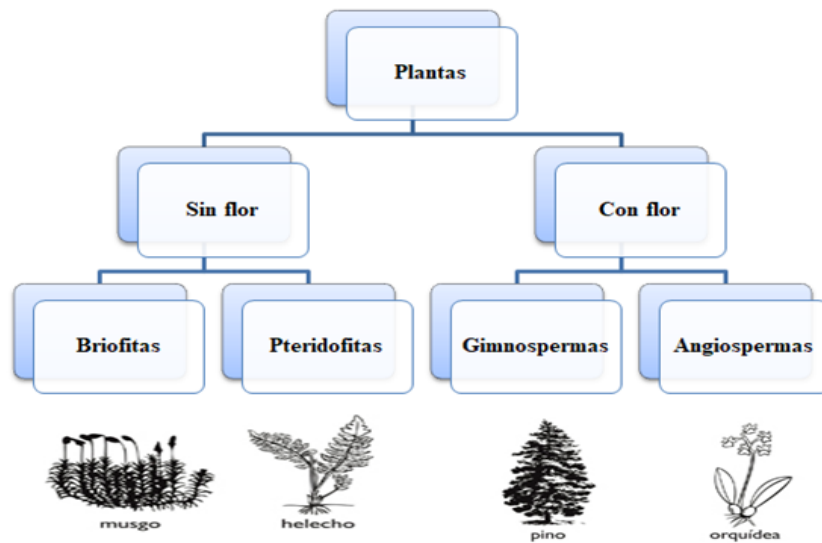


Figura 1. Clasificación de las plantas [8]

3.2.3.1. Plantas sin flor

Las plantas sin flor comúnmente están en sitios húmedos, estos se caracterizan por que no tienen flores ni semillas y se reproducen a través de esporas [9].

Tabla 3. Características de plantas sin flor

Características	
Reproducción	Por esporas
Ubicación	Sitios húmedos

Elaborado por: Los investigadores

En la clasificación de plantas sin flor tiene dos grupos: briofitas y pteridofitas

a. Briofitas

Las briofitas son plantas terrestres vasculares, son representadas por los musgos. Estos se observan como: suaves alfombras de color verde en bosques y jardines húmedos. Las características son: la cutícula toman agua y sales minerales debido a que carecen de vasos conductores, tienen estructuras similares denominadas filoides, cauloides y rizoides [9].

b. Pteridofitas

Las pteridofitas son plantas de mayor tamaño a diferencia de los musgos, prefieren un hábitat de bosques sombríos o las orillas de ríos y lagos. Entre sus características están: vasos conductores para transportar agua y salir los minerales, el cuerpo presenta tallo, hojas y raíz verdaderos.

3.2.3.2. Plantas con flor

Son las plantas que han desarrollado flores y semillas [9]. En la flor no solo se produce la fecundación. El cuerpo de las plantas con flores posee órganos vegetativos, que se encargan de la nutrición y el mantenimiento de la planta, y órganos reproductores, que llevan a cabo la función de reproducción. Posee órganos vegetativos como: raíz, tallo y hojas. Entre las características principales en la Tabla 4.

Tabla 4. Características de las plantas con flores

Características	
Órganos vegetativos	Tallo: transporta sustancias entre raíz y tallo. Hojas: Se realiza la fotosíntesis. Raíz: Absorbe agua y sales minerales.
Órganos reproductores	Flores: Se encargan de la reproducción sexual.

Elaborado por: Los investigadores

En la clasificación de las plantas con flor son: angiospermas y gimnospermas

a. Gimnospermas

Las plantas gimnospermas sus semillas no están recubiertas por el fruto. Su nombre procede del griego: “gymnos”, cuyo significado es desnudo” y “sperma”, su significado es semilla”. Entre los ejemplos de las plantas angiospermas tenemos los siguientes: abeto, ginkgo, cedro, ciprés. En la Tabla 5 nos detalla las características principales de las plantas angiospermas [10].

Tabla 5. Características de las plantas angiospermas

Características	
Semillas	No se encuentran encerradas en fruto
Flores	Carecen de cáliz y corona.
Reproducción	Los óvulos que están desnudos sobre las escamas del cono femenino y se produce la fecundación
Hojas	Son aciculares, puntiagudas o en forma de aguja

Elaborado por: Los investigadores

En el Ecuador actualmente existe 18 especies de plantas gimnospermas, por lo cual no es una población considerable para elegir este tipo de planta para el reconocimiento de especies

vegetales. También otro aspecto a considerar es la forma de su hoja (puntiaguda), lo dificulta el reconocimiento.

b. Angiospermas

Las plantas angiospermas sus semillas están rodeadas de tejidos denominados fruto. Su nombre procede del griego: angión, “ánfora” y sperma, “semilla” [10]. La mayoría de las plantas de nuestro planeta son angiospermas. Estas plantas comparten las siguientes características en la Tabla 6.

Tabla 6. Características de las plantas angiospermas

Características	
Semillas	Están encerradas en fruto
Flores	Presentan los estambres y el pistilo en la misma flor.
Óvulos	Están protegidos por un ovario transformándose en un fruto
Hojas	Tienen una gran variedad de formas.

Elaborado por: Los investigadores

Se clasifican en monocotiledóneas y dicotiledóneas, en función de que su semilla presenta un embrión: monocotiledóneas; los embriones tienen una única hoja o cotiledón en su semilla, los nervios de sus hojas son paralelos y las envueltas florales suelen tener tres o seis partes [10] y dicotiledóneas; los embriones poseen dos hojas o cotiledones en su semilla, los nervios de las hojas son ramificados y las envueltas florales suelen tener cuatro o cinco partes.

3.2.4. Clasificación de las plantas por su utilidad

Las plantas son consideradas de manera necesaria en la vida del ser humano porque son extremadamente complejas y diversas [5]. Existen millones de especies diferentes, algunas de las cuales aún no han sido descubiertas ni estudiadas por completo. En el catálogo de Ecuador lo clasifican de la siguiente manera:



Figura 2. Clasificación de las plantas por su utilidad

3.2.4.1. Plantas medicinales

Son aquellos vegetales que elaboran unos metabolitos secundarios, llamados “principios activos”, sustancias que ejercen una acción farmacológica, beneficiosa o perjudicial, sobre el organismo vivo [11]. La utilidad de estas plantas es servir como droga o medicamento que ayude aliviar las enfermedades. Estas plantas constituyen la séptima parte de las especies existentes.

Tabla 7. Características de las plantas medicinales

Nombre	Tamaño	Hojas	Utilidad
Tomillo	30 cm	Las ramas rectas y leñosas reúnen en su extremidad pequeñas flores blancas dispuestas en espiga	Afecciones de las vías respiratorias
Valeriana	1.5m	Las hojas opuestas y pinadas.	Utilizado como sedante
Romero	45cm	Las hojas lineales opuestas de color verde en la parte superior y blancas en la inferior	La producción de jugos gastrointestinales
Ruda	40cm	Las hojas alternas y lisas.	Estimulación de la función biliar, mejora la digestión.

Elaborado por: Los investigadores

En la Tabla 7 muestra las principales plantas medicinales de acuerdo a la hoja existe la variedad de formas y texturas por lo que sería una alternativa para el reconocimiento de especies vegetales. En el Ecuador se registran 255 especies, de las cuales 199 (78%) son nativas, 43 (16.7%) introducidas y 13 (5.1%) endémicas [12]. Un motivo es la cantidad de población ya que es amplia y factible en el desarrollo del reconocimiento de la planta por medio de la hoja.

3.2.4.2. Plantas alimenticias

Las plantas alimenticias son aquellas que el hombre cultiva y explota para su alimentación o nutrición. Entre las plantas medicinales se clasifican como muestra en la Tabla 8.

Tabla 8. Características de las plantas alimenticias

Características	
Cereales	Son plantas cultivadas por su granos
Leguminosas	A las plantas con flores y semillas encerradas en un fruto
Hortalizas	Sus hojas son comestibles y una variedad amplia.
Condimentarías	Producen sustancias especiales que le confieren a los alimentos.
Fruta	Son aquellas frutas donde la semilla es comestible.

Elaborado por: Los investigadores

El número total de plantas alimenticias registradas para el Ecuador es de 1561 especies, que pertenecen a 160 familias y 461 géneros [13]. El 95% de las plantas alimenticias son catalogadas como angiospermas y los alimentos derivados de las mismas son muy variados. Estas plantas abarcan un gran número de especies por lo que la población para el reconocimiento no es apta, por el motivo de tiempo para el desarrollo de la investigación. Cuando la muestra es grande el proceso ocupa mayor tiempo de desarrollo.

3.2.4.3. Plantas ornamentales

Una planta ornamental es aquella que se cultiva y se comercializa con la finalidad de mostrar su belleza. Las plantas ornamentales generalmente son cultivadas en el aire libre, viveros, invernaderos estos son controlados su temperatura. Entre las plantas ornamentales en el Ecuador se tiene: árboles, arbustos, trepadores, acuáticas, palmeras y helechos. En el Ecuador existe una variedad de 27 especies nativas que son potencialmente ornamentales. Este es un número menor de población por lo que no es una alternativa para el reconocimiento de especies vegetales y la mayoría se centra en sus flores [14].

3.2.4.4. Plantas industriales

Las plantas industriales proporcionan materia prima utilizada en la industria. Por ejemplo, en el caso del coco y la gran variedad que existen de palmeras, se llegan a obtener frutos, grasas y fibras. La mayoría de estas plantas se basa en el uso de su fruto y tallo para desarrollar productos alimenticios, y sus hojas no son utilizadas. Por este motivo se descartado esta opción para el desarrollo de la investigación [6]

3.2.5. Aprendizaje automático

El aprendizaje automático se puede considerar como modelo matemático que adquiere conocimientos a través de la formación en la forma de los predicados y utilizando el conocimiento adquirido en más problemas [15]. Tiene un papel importante un algoritmo de aprendizaje automático porque propone el método de entrenamiento o cómo se lleva a cabo la formación. Es importante tener en cuenta que para resolver un problema en el aprendizaje automático es programar explícitamente, delegar un modelo entrenable y este modelo se capacita hasta que brinde los resultados requeridos [16]. Otra definición más explícita se considera como “aprendizaje de máquinas donde busca algoritmos y heurísticas mediante la conversión de muestras de datos en programas de computadoras” [17]. Al referirse del aprendizaje automático es usar algoritmos para aprender de los patrones de datos.

Ejemplos claros sobre este tipo de aprendizaje automático son: las redes sociales “Facebook” o “Google”. Existen dos tipos de aprendizaje automático como se muestra en la **Figura 3**.

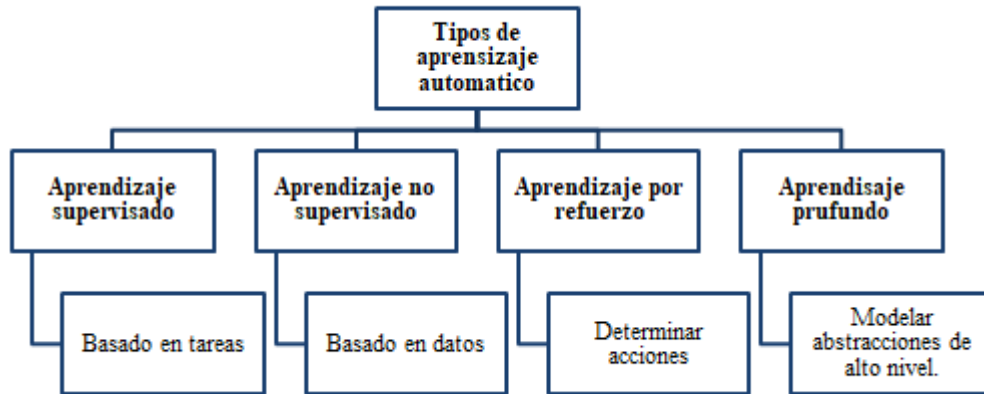


Figura 3. Tipos de aprendizaje automático [17]

3.2.5.1. Aprendizaje supervisado

En el aprendizaje supervisado el conjunto de datos debe estar etiquetado. Los algoritmos usan datos que ya han sido etiquetados u organizados previamente para indicar cómo tendría que ser categorizada la nueva información. Con este método, se requiere la intervención humana para proporcionar retroalimentación [15].

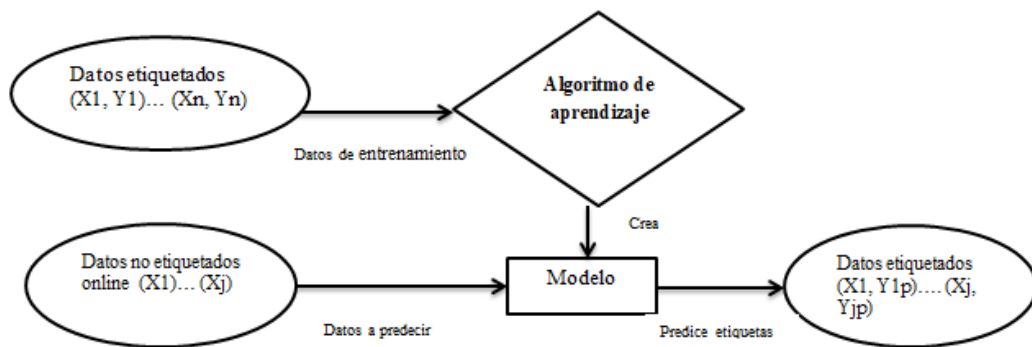


Figura 4. Aprendizaje supervisado [15]

En la **Figura 4** tiene un conjunto de datos, etiquetado y clasificado. Cuenta con un conjunto de muestra llamado “dato de entrenamiento”, se realiza el ajuste al modelo inicial planteado. De esta forma es como el algoritmo va “aprender” a clasificar las muestras de entrada comparando el resultado del modelo, y la etiqueta real de la muestra [3].

3.2.5.2. Aprendizaje no supervisado

Los algoritmos no usan ningún dato etiquetado u organizado previamente para indicar cómo tendría que ser categorizada la nueva información, sino que tienen que encontrar la manera de clasificarlas ellos mismos. Por tanto, este método no requiere la intervención humana [3]

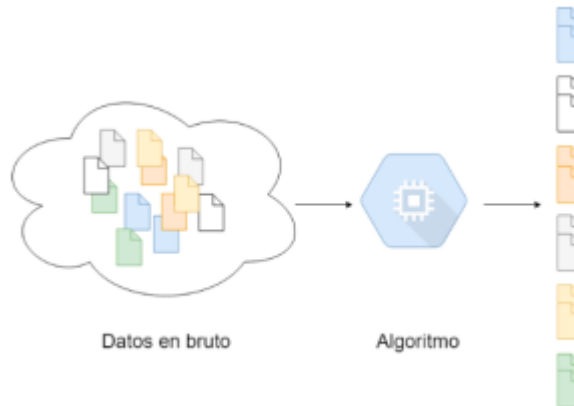


Figura 5. Aprendizaje no supervisado [3]

3.2.5.3. Aprendizaje profundo

Se trata de un sub-campo del aprendizaje automático que se utiliza para resolver problemas muy complejos y que normalmente implican grandes cantidades de datos [3]. A todo esto se le llama aprendizaje profundo porque se lleva a cabo a través de un gran número de capas. En cada una de estas se realizan continuos ajustes hasta alcanzar un determinado punto final.

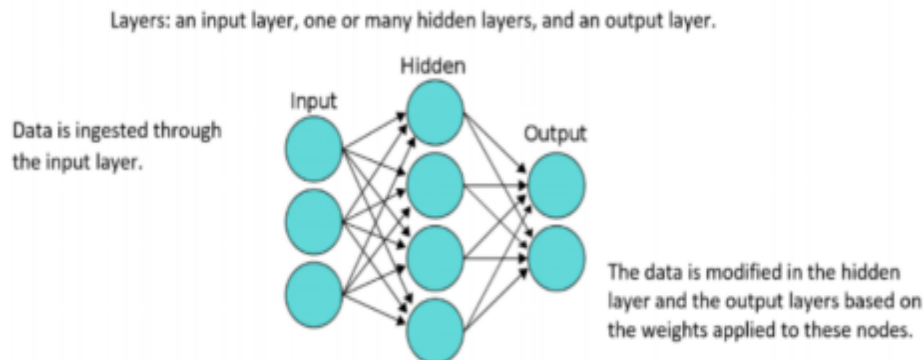


Figura 6. Aprendizaje profundo [3]

4.2.5.4. Aprendizaje de refuerzo

Los algoritmos de aprendizaje por refuerzo definen modelos y funciones enfocadas en maximizar una medida de “recompensas”, basados en “acciones” y al ambiente en el que el agente inteligente se desempeñará. Es decir, son algoritmos que se basan en el modelo de prueba o error que les permite analizar y aprender del entorno en el que se encuentra

realizando determinadas acciones mediante las cuales será capaz de llegar al objetivo final. A lo largo de todo el proceso se recibirán recompensas o penalizaciones en función de la acción realizada hasta conseguir la mejor puntuación posible y resolver el problema [3].

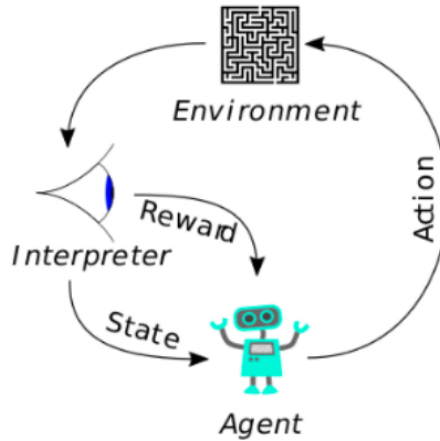


Figura 7. Aprendizaje por refuerzo [3]

3.2.6. Reconocimiento

El reconocimiento de plantas no es un reto fácil debido a que existen muchas características foliares que los botánicos emplean para identificar las plantas, además de la enorme cantidad de diferentes plantas, muchas hojas poseen y/o comparten una o varias propiedades como: forma, tamaño, textura, color, aun cuando pertenecen a plantas diferentes [2].

En el reconocimiento de imágenes se basa en “el procesamiento digital de imágenes”, con el objetivo de procesar las imágenes del mundo real de manera digital por medio de una computadora [18]. Mediante el procesamiento digital de imágenes se agrupa en nivel bajo y nivel medio el nivel bajo se requiere a operaciones básicas como reducción de ruido, aumento de contraste y mejora de nitidez y el nivel medio se requiere a operaciones como reconocimiento o extracción de características como la segmentación (regiones, objetos), caracterización de objetos, clasificación y la salida son atributos de objetos (bordes, contornos, identidades de objetos individuales).

3.2.6.1. Importancia de reconocimiento de especies vegetales

Las plantas juegan un importante papel para la vida y el desarrollo humano, no solo son de interés en investigaciones de botánica, sino también en otras ramas, tales como la agricultura [2]. El hombre se ha preocupado en diferenciar plantas y flores comestibles de las venenosas y tratar de aprovechar semillas, frutos, flores, tallos, hojas y raíces para fines diversos. Un sector beneficiado es la botánica ciencia encargada de estudiar las plantas, se desarrolla el

conocimiento de las plantas y la observación de cómo crecen y dónde, cómo se desarrollan y se reproducen sus raíces, pétalos, hojas, lugares, temporadas, climas, en definitiva conocimiento del material vegetal [2]. Con el objetivo de la botánica permite identificar y ordenar las plantas se han clasificado y organizado en grupos de modo que se pueda identificar y establecer relaciones entre ellas. De hecho, todavía no sabemos el número exacto de especies de plantas en el mundo (se estima actualmente en 370.000 especies conocidas). Sin embargo, se prevé que hasta dos tercios de las especies vegetales del mundo están en peligro de extinción en la naturaleza en el transcurso del siglo XXI. El reconocimiento de especies vegetales sería una alternativa para los investigadores debido la información sobre la clase que pertenece y tendría conocimiento cual planta estaría en peligro de extinción. Otro aspecto a considerar que es la identificación de especies vegetales es fundamental para prevenir la pérdida de rendimiento de los productos agrícolas. Su importancia radica en distinguir la especie vegetal, influye el monitoreo de las enfermedades de las plantas manualmente en cada etapa, lo que requiere más esfuerzo y tiempo [19].

En el proyecto de investigación de reconocimiento de especies vegetales, una vez obtenida la información sobre los diferentes tipos de plantas y cómo se clasifica ingresa la importancia del reconocimiento de especies vegetales por las características principales de cada clasificación de plantas, también el número de plantas que existe. Tener la información sobre la planta sería efectivo para cualquier investigador.

3.2.6.2. Segmentación de imágenes

La segmentación de imágenes en [18] detalla como una técnica de procesamiento digital de imágenes el cual permite obtener una descripción precisa de la forma de los objetos. La segmentación puede considerarse como la partición de una imagen (conjuntos de píxeles, también conocidos como objetos de imagen). Tiene la finalidad de simplificar o cambiar la representación de una imagen en algo más significativo y de esta manera sea más fácil de analizar

a. Discontinuidad:

Consiste en los bordes de regiones son suficientemente diferentes del fondo lo que permite la detección de los mismos basándose en cambios bruscos de nivel de intensidad.

Detección de puntos aislados: imagen tiene un tono de gris que difiere significativamente de los tonos de gris de sus píxeles vecinos, es decir, de los ocho píxeles de su entorno 3×3 .

Borde de una imagen digital en escala de grises: una imagen digital en escala de grises, un píxel pertenece a un borde si se produce un cambio brusco entre niveles de grises con sus vecinos. Por tanto, como ya hemos comentado, el fundamento para la detección de bordes está en la aplicación del operador derivada en un entorno de vecindad.

$$E = \frac{\sum_{i=1}^{I_D} \left(\frac{1}{1 + \alpha d(i)^2} \right)}{\max(I_D, I_I)}$$

Figura 8. Borde de una imagen digital en escala de grises [18]

En la **Figura 8** muestra la ecuación I_D = cantidad de píxeles que devuelve el detector de bordes, I_I = cantidad de píxeles reales de un borde en la imagen, $d(i)$ = distancia entre el píxel i -ésimo del detector y el píxel más próximo del verdadero borde de la imagen, α es una constante que se usa para re-escalado. Un ejemplo como se aplica detección de bordes de imágenes se muestra en la **Figura 9**.

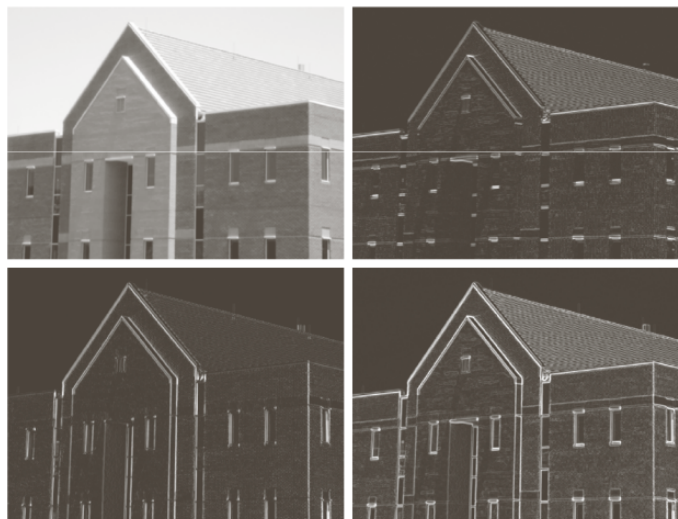


Figura 9. Detección de bordes en imágenes en escala de grises [18]

b. Similitud:

Se divide la imagen basándose en la búsqueda de zonas que tengan valores similares, conforme a unos criterios prefijados.

3.2.6.3. Filtros

Otro punto importante a tratar es el hecho de que las imágenes, una vez salen del dispositivo de captación, no suelen permitir una segmentación fácil. Ello puede suceder por diferentes motivos. Por ejemplo, la existencia de ruido de fondo, que nos hace confundir manchas con

objetos. Otras veces, por la existencia de unos márgenes difusos en los objetos. En otras ocasiones por existir un contacto entre objetos.

a. Filtro por dilatación y erosión

Estos filtros de dilatación y erosión son conocidos también con el nombre de filtros de morfología matemática. La erosión, es el resultado de la aplicación de la llamada regla de erosión: un pixel "blanco" (generalmente de objeto) pasa a "negro" si en su entorno inmediato existe al menos un pixel "negro". La dilatación, es el resultado de la aplicación de la regla de dilatación: un pixel "negro" (generalmente de fondo) pasa a "blanco" si en su entorno inmediato existe al menos un pixel "blanco". La dilatación produce un aumento en el tamaño de los objetos por adición de pixel marginal. Como se muestra en la **Figura 10**.

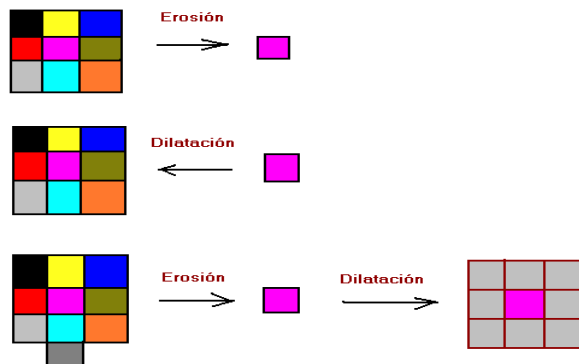


Figura 10. Filtro de imagen por dilatación y erosión [18]

b. Filtro GaussianBlur

El filtro gaussiano se usa para emborronar imágenes y eliminar ruido. Es similar al filtro de media, pero se usa una máscara diferente, modelizando la función gaussiana. Presentan ventajas como: es separable: es decir, en lugar de realizar una convolución bidimensional, podemos realizar dos convoluciones unidimensionales. Una en sentido horizontal y otra en sentido vertical como se muestra en la **Figura 11**.

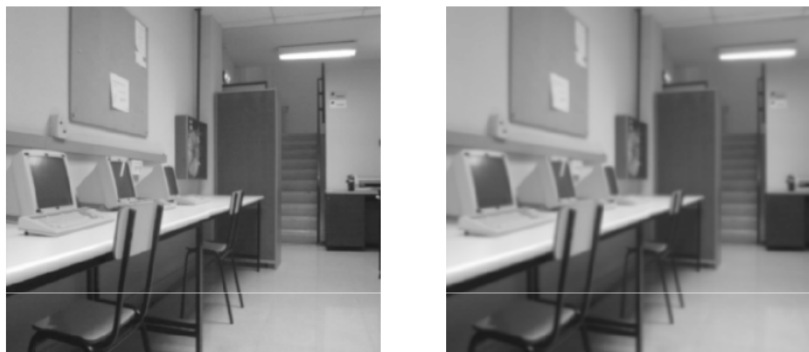


Figura 11. Filtro GaussianBlur [18]

3.2.6.4. Técnicas de reconocimiento supervisado

Entre los principales problemas de aprendizaje supervisado [20] se encuentran el aprendizaje supervisado por clasificación, que consiste en predecir una etiqueta o categoría, y el aprendizaje supervisado por regresión, que consiste en predecir un valor numérico.

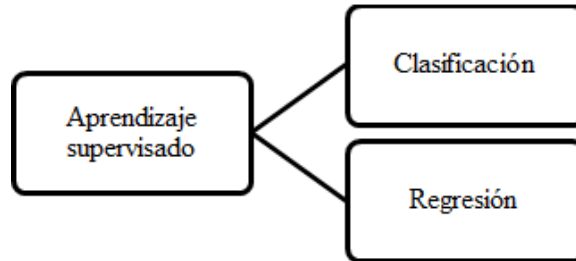


Figura 12. Clasificación de aprendizaje supervisada

En el aprendizaje supervisado [16] por el método de clasificación tienen los siguientes algoritmos detallados en la **Figura 13**.

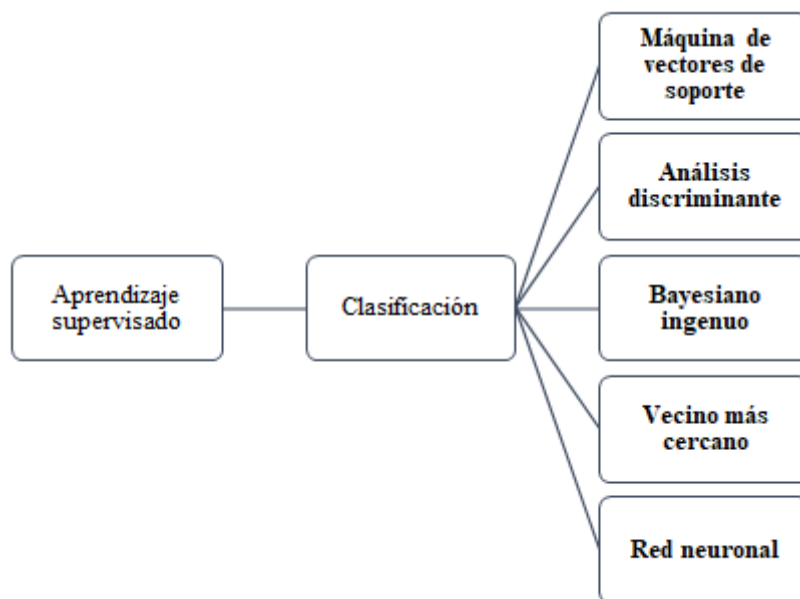


Figura 13. Clasificación de aprendizaje supervisado-clasificación

a. Máquina de vectores de soporte (support vector machine):

Los puntos de datos pertenecientes a diferentes clases están separados por planos divisorios. La clasificación es realizada en base al margen entre los planos de separación. Los puntos de datos de prueba se mapean en el mismo espacio y la predicción se realiza en base a un grupo de puntos [21].

b. Análisis discriminante (discriminant analysis):

El análisis discriminante es una técnica estadística multi-variante cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados [15].

c. Bayesiano ingenuo (naive bayes):

El CBI reduce significativamente la complejidad del clasificador bayesiano en espacio y tiempo de cálculo [16]. En cuanto a espacio de memoria, se requiere la probabilidad previa de las m clases (vector de $1 \times m$), y las n probabilidades condicionales de cada atributo dada la clase (si suponemos que los atributos son discretos con k posibles valores, esto implica n matrices de $m \times k$). Básicamente el espacio requerido aumenta linealmente con el número de atributos. También el cálculo de la probabilidad posterior se vuelve muy eficiente, ya que se requieren del orden de n multiplicaciones para calcular la probabilidad posterior de cada clase dados los atributos (complejidad lineal).

d. Vecino más cercano (nearest neighbor):

El algoritmo vecino más cercano (KNN) es un proceso para clasificar datos basados en ejemplos de entrenamiento cercanos en el conjunto de funciones. KNN es un tipo de aprendizaje basado en instancias o aprendizaje perezoso donde la clasificación se realiza en base al voto mayoritario para etiquetar sus k vecinos más cercanos [15]. KNN convierte la imagen en un vector de longitud fija con valores reales, usamos distancia euclidiana para encontrar la distancia por ecuación.

e. Red neuronal (neural network):

Una red neuronal artificial (ANN) es un esquema de computación distribuida inspirada en la estructura del sistema nervioso de los seres humanos. La arquitectura de una red neuronal es formada conectando múltiples procesadores elementales, siendo éste un sistema adaptativo que posee un algoritmo para ajustar sus pesos (parámetros libres) para alcanzar los requerimientos de desempeño del problema basado en muestras representativas [3].

En el aprendizaje supervisado [15] por el método de regresión tienen los siguientes algoritmos detallados en la **Figura 14**.

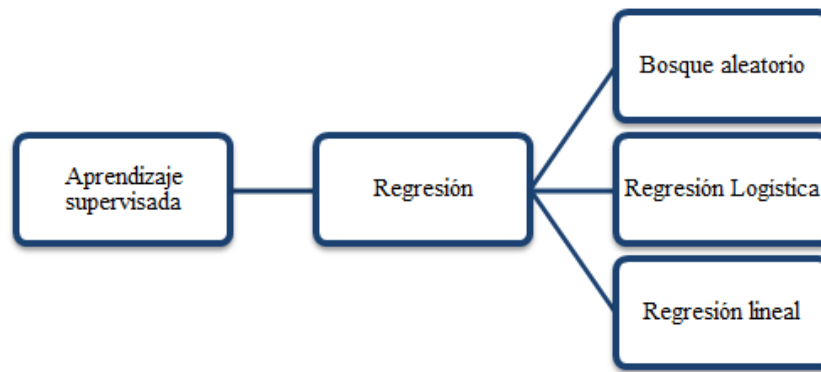


Figura 14. Aprendizaje supervisado por regresión [15]

a. Bosque aleatorio

Es un conjunto de clasificadores. $h_1(x)$, $h_2(x)$... $h_k(x)$ y el conjunto de entrenamiento se extrae de forma aleatoria distribución del vector (X, Y) y función de margen [21]. La función de margen da la medida en que el promedio número de votos en X, Y para la clase correcta es más que promedio para cualquier otra clase. Mayor margen produce mejor clasificación. El error generalizado viene dado por:

b. Regresión Lineal

Un modelo de regresión lineal permite describir cómo influye una variable X sobre otra variable Y .

- X : Variable independiente o explicativa
- Y : Variable dependiente o respuesta

El objetivo es obtener estimaciones razonables de Y , es para distintos valores de X a partir de una muestra de n pares de valores $(x_1, y_1), \dots, (x_n, y_n)$. En general, la regresión lineal intenta establecer una relación lineal entre las variables independientes y un resultado una variable, o variable dependiente, mediante el uso de datos de entrenamiento para construir el modelo.

c. Regresión Logística

Es una técnica estadística multi-variante que nos permite estimar la relación existente entre una variable dependiente no métrica (en particular dicotómica), y un conjunto de variantes independientes métricas o no métricas. Esta regresión es parecida a la regresión lineal multivalente [22].

3.2.6.5. Técnicas de reconocimiento no supervisado

Los tipos de aprendizaje no supervisado más destacables son el denominado clustering (o agrupamiento en español) y las reglas de asociación [20].

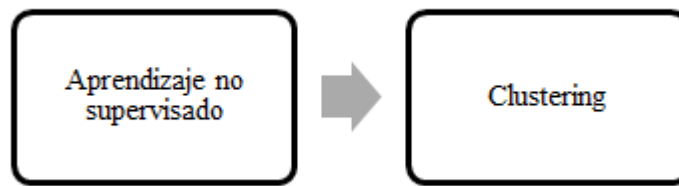


Figura 15. Clasificación de aprendizaje no supervisado [20]

El clustering es un tipo de problema que consiste en agrupar porciones de instancias en función de ciertas características, como podría ser una serie de clientes de un banco que tienen una edad aproximada.

Las reglas de asociación se utilizan para destacar ciertas normas o tendencias que aparecen en los conjuntos de datos. Un ejemplo sería destacar una tendencia que muestre que un tipo de cliente, que haya comprado un producto específico, sea propenso a comprar más adelante otro producto distinto. Este tipo de información es muy interesante para que los negocios y las empresas puedan predecir los intereses de sus clientes.

Los algoritmos utilizados en el método de clustering son: Fuzzy y K-means

a. Fuzzy

En muchas situaciones cotidianas ocurre el caso que un dato está lo suficientemente cerca de dos clúster de tal manera que es difícil etiquetarlo en uno o en otro, esto se debe a la relativa frecuencia con la cuál un dato particular presenta características pertenecientes a clúster distintos y como consecuencia no es fácilmente clasificado; fuzzy [23].

b. K-means

Es un algoritmo [19] basado en clustering. Lo que quiere decir que su principal objetivo es optimizar la partición de la imagen en áreas conforme a unas características dadas. Este tipo de algoritmos ha sido usado comúnmente para una segmentación básica de imágenes ya que existen ciertas restricciones como la necesidad de que la imagen sea generalmente homogénea, sin texturas.

3.2.7. Aplicación móvil

Una aplicación móvil o conocida como app móvil, es un tipo de aplicación que se ejecuta en un dispositivo móvil. Por lo general se encuentran en plataformas de distribución, operadas por las compañías propietarias de los sistemas operativos móviles como: GooglePlayStore de Google para Android, APPStore de Apple para iOS, BlackBerry OS, WindowsStore de Microsoft para Windows Phone, entre otros [3].

3.2.7.1. Tipos de aplicaciones móviles

En la actualidad [3] se manejan tres aplicaciones móviles como:

a. Aplicación nativa

Una aplicación nativa es la que se desarrolla de forma específica para un determinado sistema operativo, llamado Software Development Kit o SDK.

b. Aplicación web

Una aplicación web o conocida como “webapp” es la desarrollada en HTML, JavaScript y CSS. Cuya ventaja con respecto a la nativa es la posibilidad de programar independiente del sistema operativo en el que se usará la aplicación.

c. Aplicación híbrida

Las apps híbridas están basadas en webapp, es decir, HTML, Javascript y CSS lo cual permite su uso en diferentes plataformas, adicional a esto la posibilidad de acceder a gran parte de las características del hardware del dispositivo.

3.2.8. Herramientas de software

Para el desarrollo de la aplicación móvil [24] se basó en las siguientes herramientas que se detalla a continuación:

3.2.8.1. Dart

Es el lenguaje de programación de código abierto que se utiliza para la creación de aplicaciones con Flutter, ayuda a crear aplicaciones de múltiples plataformas con un único código. Flutter es un framework de desarrollo de aplicaciones móviles en Android y iOS.

3.2.8.2. Android Studio

Es un nuevo entorno de desarrollo integrado del sistema operativo Android, diseñado para el desarrollo de aplicaciones está basado en un entorno llamado IntelliJ IDEA, con el que comparte el editor de texto y otras herramientas para desarrolladores.

3.2.8.3. TensorFlow lite

Es un conjunto de herramientas que tiene como objetivo ejecutar sus modelos en dispositivos incorporados, móviles o de IoT, y les permite implementar el aprendizaje automático integrado en el dispositivo.

3.2.8.4. Librería OpenCV

Es considerada la biblioteca de visión por computadora más grande en términos de funciones poseídas. Está contiene implementaciones de más de 2500 algoritmos. Esta gran biblioteca tiene interfaces para múltiples lenguajes, incluidos Python, Java y C++.

3.2.9. Metodología SRL

Las revisiones sistemáticas de literatura (RSL), el objetivo es identificar, evaluar y combinar la evidencia de estudios primarios usando un método riguroso. Este método ha sido ampliamente implementado en algunas disciplinas de ingeniería de software. Desde la publicación del artículo de Ingeniería de software basado en evidencia (EBSE), las revisiones sistemáticas de literatura se han vuelto una importante metodología de EBSE, y desde entonces numerosas RSL han sido realizadas y reportadas [25].

Se pueden utilizar para identificar la literatura disponible. Antes de emprender SLR convencionales. Usan métodos de búsqueda y extracción de datos como los convencionales SLR, pero se basan más en la tabulación de los estudios primarios en categorías específicas.

En la revisión literaria es necesario cumplir con los siguientes aspectos para evitar errores durante la investigación.

- Tener el contexto claro sobre la investigación
- Identificar los conceptos y las variables a investigar.
- Tener claro la relación que existe entre concepto y variable.
- Finalmente establecerla necesidad de la investigación.

Se basa en tres etapas fundamentales para la investigación como se muestra en la **Figura 16**.

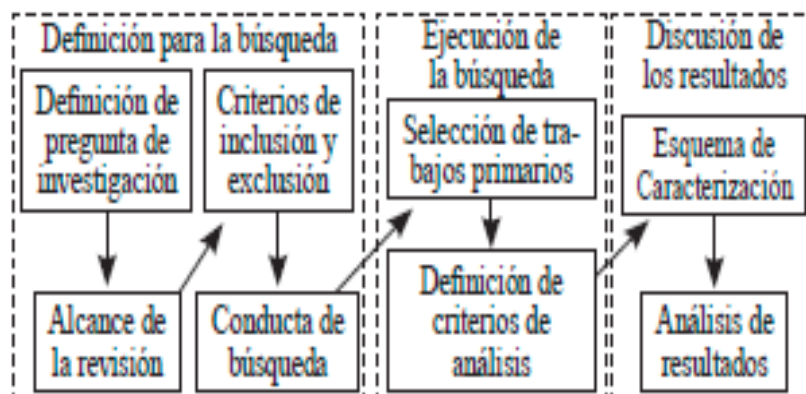


Figura 16. Proceso de la metodología SRL

3.2.9.1. Definición de la búsqueda

En esta etapa, es donde se define la pregunta de investigación, el alcance de la revisión, los criterios de inclusión y exclusión, y finalmente la cadena de búsqueda.

3.2.9.2. Ejecución de la búsqueda

En la etapa de búsqueda, es donde se define la selección de trabajos primarios y difusión de criterios de análisis.

3.2.9.3. Discusión de los resultados

En la etapa de discusión de resultados, se definen los esquemas de caracterización y se analizan los resultados.

3.2.10. Metodología KDD

El Descubrimiento de conocimiento en bases de datos KDD , en inglés *Knowledge Discovery in Databases*, básicamente un proceso automático en el que se combinan descubrimiento y análisis [26]. Mediante el KDD cumple con el proceso de extraer conocimiento a partir de grandes volúmenes de datos y ha sido reconocido por muchos investigadores como un tópico de investigación principal en los sistemas de bases de datos.

La metodología CRIS-KDD consiste en un conjunto de tareas que están organizadas en niveles de abstracción: fases, tareas generales, tareas especializadas e instancias de proceso.

3.2.10.1. Etapas de la metodología KDD

El proceso KDD, es interactivo e iterativo, involucra numerosos pasos donde interviene el usuario en la toma de mayoría de decisiones. Se detallan las siguientes fases en la **Figura 17**.

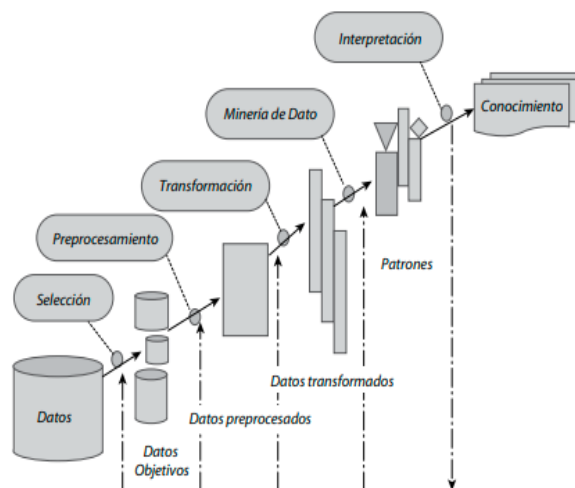


Figura 17. Etapas de la Metodología KDD

a. Etapa de selección

En la etapa de selección, el conocimiento debe ser identificada y definidas las metas del proceso KDD, con la participación del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento [26]. Con esta etapa los datos varían de acuerdo con los objetivos del negocio.

b. Etapa de pre-procesamiento

En la etapa de pre-procesamiento de la metodología KDD, se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos (desconocidos, nulos, duplicados y técnicas estadísticas) para su reemplazo. En esta etapa, es de suma importancia la interacción con el usuario o analista [26].

c. Etapa de transformación

En la etapa de transformación se utiliza los métodos de reducción de dimensiones pueden simplificar una tabla de una base de datos (horizontal o vertical). A través de dos reducciones como: la reducción horizontal implica la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo por otro de alto nivel, en una jerarquía definida de valores categóricos y la reducción vertical implica la eliminación de atributos que son insignificantes o redundantes con respecto al problema, como la eliminación de llaves, la eliminación de columnas [26].

d. Etapa de minería de datos – modelado

En la etapa de minería de datos se crean modelos (predictivos o descriptivos). Los modelos predictivos pretenden estimar valores futuros o desconocidos. El proceso de descubrimiento de conocimiento en bases de datos como: variables objetivo, dependientes o clases, usando otras variables (llamadas predictivas). Entre las tareas predictivas están la clasificación y la regresión [26].

e. Etapa de interpretación

En la etapa de interpretación/evaluación, son descubiertos los patrones y posiblemente se retorna a las anteriores etapas para posteriores iteraciones.

Puede incluir a través de la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos [26].

3.2.11. Metodología Mobile-D

Esta metodología consiste en ciclos de desarrollo muy rápidos en equipos muy pequeños. Mobile-D es una mezcla de muchas técnicas.

Tiene distintas fases: exploración, inicialización, fase de producto, fase de estabilización y la fase de pruebas. Cada una tiene un día de planificación y otro de entrega [27].

a. Fase de exploración

En esta fase se define el alcance del proyecto y su establecimiento con las funcionalidades donde se quiere llegar. Esta fase se dedica al establecimiento del plan de proyecto, así como los conceptos básicos. Por lo tanto, se puede separar del ciclo principal de desarrollo.

b. Fase de iniciación

En esta fase se configura el proyecto identificando y preparando todos los recursos necesarios como hemos comentado anteriormente en esta fase la dedicaremos un día a la planificación y el resto al trabajo y publicación.

c. Fase de producto

En esta fase se basa en el desarrollo dirigido por pruebas (TDD), antes de iniciar el desarrollo de una funcionalidad debe existir una prueba que verifique su funcionamiento.

En esta fase podemos decir que se lleva a cabo toda la implementación.

d. Fase de estabilización

En esta fase se llevan a cabo las últimas acciones de integración para asegurar que funcione correctamente el sistema completo. Una vez finalizada esta fase debe cumplir con: la funcionalidad puesta en funcionamiento en todo software del proyecto y la documentación del producto terminado.

e. Fase de pruebas

Finalmente, en la fase de pruebas cumple con la estabilidad y funcionamiento para que los clientes los utilicen. La aplicación una vez terminada se la integra y la prueba en base a los requisitos del cliente y suprimen todos los errores descubiertos.

4. MATERIALES Y MÉTODOS

4.1. TIPOS DE INVESTIGACIÓN

4.1.1. Investigación bibliográfica

Se aplicó la investigación bibliográfica para la recopilación de información relacionada con antecedentes de estudios realizados con aprendizaje de máquinas en especies vegetales mediante definiciones básicas, herramientas de software libre y el procedimiento para la aplicación de la metodología KDD, a través de una búsqueda en distintas fuentes bibliográficas primarias y secundarias como artículos científicos, libros, revistas y proyectos similares que permitió una redacción adecuada de la fundamentación teórica del proyecto de investigación.

4.1.2. Investigación explicativa

A través de esta investigación se detallará las principales técnicas de aprendizaje de máquinas con la finalidad de identificar las posibles causas y efectos en base al planteamiento del problema "¿Cuáles son las técnicas de aprendizaje de máquinas que se utilizan en el reconocimiento de especies vegetales en el área de la agricultura?", describiendo cada uno de los estudios realizados en diferentes países a través de la validación de porcentaje del algoritmo utilizado.

4.1.3. Investigación aplicada

La investigación aplicada busca la generación de conocimiento en base al problema "¿Cuáles son las técnicas de aprendizaje de máquinas que se utilizan en el reconocimiento de especies vegetales en el área de la agricultura?", este en base al sector agrícola. En donde se basa fundamentalmente en los hallazgos tecnológicos (técnicas de machine learning). Con el resultado de un reconocimiento de las especies vegetales.

4.2. MÉTODO TEÓRICO

4.2.1. Método Hipotético deductivo

Se aplicó el método hipotético deductivo debido a que se planteó una hipótesis al inicio del proyecto, el cual permite demostrar resultados positivos en base al desarrollo de la aplicación móvil para reconocimiento de especies vegetales con técnicas de machine learning. El método hipotético-deductivo se basa en pasos esenciales: fenómeno a estudiar (técnicas de aprendizaje de máquinas y especies vegetales), con hipótesis de "Las principales técnicas de

aprendizaje de máquinas que se utilizan para el reconocimiento de especies vegetales en el área de la agricultura se caracterizan por tener mayor rendimiento y frecuencia de uso.”, de esta manera explicar dicha investigación

4.3. MÉTODO EMPÍRICO

4.3.1 Técnicas de Investigación

4.3.1.1. Análisis bibliográfico

Se aplicó esta técnica para realizar la búsqueda de documentos primarios y secundarios sobre técnicas de aprendizaje de máquinas y especies vegetales que brinden información relevante que permitan una adecuada fundamentación teórica durante el desarrollo de estudios de 5 años para el análisis documental del proyecto de investigación. Así como las investigaciones realizadas para obtener información verídica y confiable.

4.3.1.2. Investigación de campo

A través de la investigación de campo se va recolectar las imágenes de las hojas de las plantas de especies vegetales (orégano y menta) en su entorno natural. Se tomó la imagen en base a parámetros de iluminación, de esta manera las fotos servirán para construir la base de datos con la cual trabajará la regresión logística.

4.3.2. Instrumentos de investigación

4.3.2.1. Ficha bibliográfica

Mediante este instrumento de investigación se realizó un análisis del contenido de los documentos encontrados y de esa forma captar información que se relacione al desarrollo de la aplicación móvil sobre el reconocimiento de imágenes.

4.3.2.2. Fotografía

Se utiliza la fotografía en la investigación para recabar y armar la base de datos de imágenes necesarias para el proceso de aprendizaje.

4.3.3. Población y muestra

La población con la que se realiza la investigación son los invernaderos en la provincia de Cotopaxi, debido a que se encuentra la mayor cantidad de especies vegetales.

4.3.4. Cálculo de la muestra

Para el cálculo de la población se optó por el muestreo por conveniencia debido a que es una técnica de muestreo no probabilística porque se eligió porque están convenientemente disponibles para los investigadores, este caso en especies vegetales (menta y orégano). Estas muestras se seleccionan solo porque son fáciles de reclutar y porque no se consideró seleccionar una muestra que represente a toda la población

4.4. MÉTODO ESPECÍFICO

4.4.1. Metodología RSL

En el desarrollo de la revisión literaria sistemática, se hace análisis de los artículos de cinco años de publicación relacionado a reconocimiento de especies vegetales con aprendizaje de máquinas, se enfoca en la frecuencia de validación y el uso de técnica de aprendizaje con la finalidad de conocer cuál técnica es utilizada con mayor frecuencia de uso en los artículos. Se basa en tres etapas: definición de la búsqueda (relacionado a las palabras claves), ejecución de la búsqueda en base a las palabras claves realizar la búsqueda en base de datos), finalmente la discusión de resultados el cual se detalla los artículos seleccionados al tema de proyecto de investigación, la especie vegetal utilizada, y la técnica de aprendizaje de máquinas empleada.

4.4.2. Metodología KDD

En el desarrollo de la metodología KDD se va dar la elección de la base de datos de imágenes de plantas de menta y orégano a través de las variables (Convolución, Difuminación, Luminosidad, Densidad y Umbralización) que se van a utilizar y así como los resultados. Donde va cumplir con las siguientes etapas: elegir un conjunto de datos, en la etapa de pre-procesamiento se analiza la calidad de datos, en la etapa de transformación se hará la eliminación de tuplas idénticas como producto de la sustitución del valor de un atributo, en la etapa de minería de datos y finalmente la etapa de transformación en cual son descubiertos los patrones.

4.4.3. Metodología Mobile-D

En el desarrollo de la metodología Mobile-D se utiliza debido que es factible por el uso de ciclos de desarrollo y estos son cortos lo que ayuda a progresar en el desarrollo de la aplicación móvil de reconocimiento de especies vegetales con aprendizaje de máquinas. Contará con las 5 fases (exploración, inicialización, producto, estabilización y pruebas) para la verificación del correcto funcionamiento de la aplicación móvil.

5. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

5.1. Metodología de revisión sistemática literaria (RSL)

La metodología de revisión sistemática de literatura se basa en la metodología propuesta por Bárbara Kitchenham: Para la realización de RSL se basó en tres etapas. Es necesario definir preguntas de investigación, para llevar a cabo la investigación se plantean las siguientes preguntas de investigación, con la finalidad de encontrar términos claros y precisos, para identificar los componentes principales.

- ¿Qué técnicas se utilizaron para el reconocimiento de especies vegetales durante los últimos 5 años?
- ¿Cuál fue la técnica con mayor frecuencia de uso en los estudios?

5.1.1. Definición de la búsqueda

Para la fundamentación teórica, se inició la búsqueda: selección de las palabras claves, establecer cadenas de búsqueda, búsqueda de la información en base de datos electrónicas con enfoque académico, mediante los criterios de exclusión e inclusión de documentos (idioma, año, título, resumen, especie vegetal, técnica de aprendizaje de máquinas, porcentaje de validación, etc.) y se muestra el proceso, con la finalidad que el filtro sea más eficiente. En las secciones posteriores se describe con claridad el proceso que se realizó para la recolección y síntesis de la información. En la estrategia de búsqueda: se realizó una búsqueda a partir de la selección de las palabras clave, en bases de datos científicas electrónicas, de forma automatizada, posteriormente se describe los protocolos de búsqueda y las bases de datos usadas

a. Términos de búsqueda

Para la búsqueda de la información se identificó términos específicos con relación al tema de técnicas de aprendizaje de máquinas y especies vegetales, con la ayuda del Tesoro de la IEEE, estos términos fueron traducidos al idioma inglés, los mismos que se presentan a continuación: Machine learning, Vegetable species, Detection Plant, Techniques.

b. Diseño de los protocolos de búsqueda

Una vez identificados los términos de búsquedas, se combinan con operadores booleanos “AND” y “OR”, cuya finalidad es mejorar la estrategia de búsqueda y maximizar los resultados, el diseño de los protocolos de búsqueda podemos divisar a continuación: “Machine learning”

AND “Techniques” OR Techniques AND Machine learning AND Vegetable species OR Detection Plant AND Machine learning OR Techniques, Detection Plant OR Machine learning AND Techniques AND Vegetable species.

c. Recursos utilizados

La búsqueda de información se realizó en Internet, mediante herramientas de búsqueda especializadas como son las bases de datos, ya que los datos que contienen son selectivos, confiables y de temas diversos. A continuación, se presenta la Tabla 9 con las bases de datos usadas:

Tabla 9. Recursos utilizados para la RSL

Código	Base de datos	Url
BD1	Science Direct	https://www.sciencedirect.com/browse/journals-and-books
BD2	Google académico	https://scholar.google.es/schhp?hl=es
BD3	Scielo	https://www.scielo.br

Elaborado por: Los investigadores

En este proceso se eligió las bases de datos detalladas en la Tabla 10. Para ello se utilizó el protocolo de búsqueda que ayuda de manera rápida para la elección de artículos.

Tabla 10. Búsqueda de palabras claves en las bases de datos

Código	Base de datos	Año	Palabras claves	Cantidad de artículos	Cantidad total de artículos	Artículos considerados
BD1	Science Direct		Machine learning	149,052	267,486	395
			Machine learning techniques	116,492		
			Machine learning techniques vegetable species	1,039		
			Machine learning techniques recognition vegetable species	508		
			Review of used machine learning techniques recognition of vegetable species	395		
BD2	Google academic	2018 2019 2020 2021	Machine learning- Aprendizaje de maquinas	1.050.000	1,051,496	200
			Machine learning techniques	280.000		
			Machine learning techniques vegetable species	16.000		
			Machine learning	1.000		

		2022	techniques recognition vegetable species			
			Review used machine learning techniques recognition vegetable species	200		
BD3	Scielo		Machine learning- Aprendizaje de maquinas	407	417	10
			Machine learning- plantas medicinales	10		
TOTAL					1,319,399	605

Elaborado por: Los investigadores

Se realizó la búsqueda general del tema de proyecto de investigación de acuerdo a las palabras claves en la cual nos da una respuesta del valor de artículos desarrollados, también se consideró artículos mediante el filtro de año 2018-2022. Dando un total de 605 artículos con el cual se va evaluar cuales cumplen con técnicas de aprendizaje de máquinas.

- **Criterios de selección:** en este proceso, se presentan los criterios aplicados para la selección de los estudios, la estrategia de exploración y los métodos para la recolección y síntesis de la información.
- **Criterios de inclusión:** Filtro por publicación: artículos seleccionados en el área sector agrícola, especies vegetales; y que fueran publicados en el periodo comprendido entre los años 2018 a 2022.

Filtro por palabras: tomando como palabras claves: Machine learning, Vegetable species, Detection Plant, Techniques y los operadores booleanos empleados: OR y AND.

- **Criterios de exclusión:** se establecen los criterios de exclusión, para que los documentos encontrados sean de mayor precisión. Se tomó en cuenta el idioma: se consideraron los documentos en idioma inglés.

5.1.2. Ejecución de búsqueda

Una vez realizado los procesos anteriores se detalla los primeros resultados de la búsqueda de información se presenta en la Tabla 11, con el contenido total de los documentos obtenidos para posteriormente analizarlo y preseleccionado.

Tabla 11.Número de documentos obtenidos por cadena de búsqueda

Código	Base de datos	Cadena de búsqueda	Número de documentos
BD1	Science Direct	“Machine learning” AND “Techniques AND Reconocimiento de Hojas de Planta AND Plantas medicinales	4
BD2	Google académico	Detection Plant OR Machine learning AND Techniques AND Vegetable species OR Automatic Recognition	15
BD3	Scielo	Techniques AND Machine learning AND vegetable species	4
TOTAL			23

Elaborado por: Los investigadores

Los primeros estudios obtenidos no fueron muy diversos, en su gran mayoría fueron artículos, para evaluar la calidad de los diferentes documentos obtenidos, los documentos se clasificaron en función del tipo de estudio, técnica de utilización, técnica de aprendizaje de máquina, porcentaje de validación del algoritmo, y finalmente un conjunto genérico de preguntas para evaluar su rigurosidad, credibilidad y relevancia, y de esta manera obtener los documentos finales.

Se evaluó la calidad de cada estudio se eligió los 23 artículos (ver Anexo 2) lo que permitió la clasificación según 11 criterios diferentes recogidos en la Tabla 12, cada una de las preguntas tenía tres respuestas opcionales: “Si”, “parcialmente” y “No”, estas respuestas se puntuaron de la siguiente forma: “Si” = 1, “parcialmente” = 0,5, y “No” = 0. La evaluación de calidad de los artículos se calcula sumando las puntuaciones de las respuestas a las preguntas previamente definidas.

Tabla 12.Criterios de calidad

Id	Criterio de calidad	VALOR		
		Si = 1	No = 0	Parcialmente = 0.5
CC1	¿Los objetivos de la investigación están claramente especificados?	Si = 2	No = 1	Parcialmente = 0.5
CC2	¿El estudio fue diseñado para lograr esos objetivos?	Si = 3	No = 2	Parcialmente = 0.5
CC3	¿El documento encontrado, responden a las preguntas planteadas para la investigación?	Si = 1	No = 0	Parcialmente = 0.5
CC4	¿Existe una descripción adecuada del contexto en el que se llevó a cabo la investigación?	Si = 1	No = 0	Parcialmente = 0.5
CC5	¿Los métodos de recolección de datos se describen adecuadamente?	Si = 1	No = 0	Parcialmente = 0.5
CC6	¿Está claro el propósito del análisis de los datos?	Si = 1	No = 0	Parcialmente = 0.5
CC7	¿La metodología de investigación fue apropiada para abordar los objetivos de la investigación?	Si = 1	No = 0	Parcialmente = 0.5
CC8	¿Las técnicas estadísticas utilizadas para	Si = 1	No = 0	Parcialmente = 0.5

	analizar los datos se describen adecuadamente y se justifica su uso?			
CC9	¿Los investigadores solucionan algún problema y lo reflejan en sus resultados?	Si = 1	No = 0	Parcialmente = 0.5
CC10	¿Está clara la relación entre los datos, la discusión y las conclusiones?	Si = 1	No = 0	Parcialmente = 0.5
CC11	¿Están las conclusiones respaldadas por los resultados que se presentan?	Si = 1	No = 0	Parcialmente = 0.5

Elaborado por: Los investigadores

En la primera selección una vez realizado en base a criterios de calidad de la Tabla 13. En el cual se habían seleccionado 23 artículos de acuerdo a la base de datos. En cual nos da como resultados de documentos aprobados 7 que tienen un total entre 9.5-11 y de documentos descartados 16 cuyo valor del total es de 4.5 – 7.5 detallada en la Tabla 12. Con este resultado se pasa a la segunda selección para determinar qué artículos son necesarios para el desarrollo del proyecto de investigación.

Tabla 13. Evaluación de artículos según los criterios de calidad

Artículo	Criterios Calidad											TOTAL	Valor
	CC 1	CC 2	CC 3	CC 4	CC 5	CC 6	CC 7	CC 8	CC 9	CC 10	CC 11		
CA1	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1	7.5	Reprobado
CA2	0,5	1	1	1	0	1	0,5	1	1	0	0	7	Reprobado
CA3	1	0.5	0.5	1	1	1	1	1	1	1	1	10	Aprobado
CA4	0,5	0,5	1	0,5	1	0,5	0,5	0,5	1	1	0,5	7.5	Reprobado
CA5	1	1	1	1	0.5	1	1	1	1	1	1	10.5	Aprobado
CA6	0.5	0.5	0	0.5	0.5	0.5	1	0.5	1	1	1	7	Reprobado
CA7	1	1	1	1	0.5	0.5	1	1	0.5	0.5	1	9	Aprobado
CA8	0,5	0,5	1	0,5	0,5	0,5	0,5	0,5	1	1	0,5	7	Reprobado
CA9	0,5	0,5	1	0,5	1	0,5	0,5	0,5	0,5	0,5	0,5	6.5	Reprobado
CA19	1	1	1	1	1	1	1	1	1	1	1	11	Aprobado
CA11	0,5	1	1	1	1	1	0,5	0,5	0,5	0	0	7	Reprobado
CA12	1	1	0.5	1	1	1	1	1	1	1	0.5	10	Aprobado
CA13	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	1	7.5	Reprobado
CA14	0,5	0,5	1	0,5	0,5	0,5	0,5	0,5	1	1	0,5	7	Reprobado
CA15	0,5	0,5	1	0,5	1	0,5	0,5	1	1	1	0,5	5.5	Reprobado
CA16	1	0.5	0.5	1	1	1	1	1	1	0.5	1	9.5	Aprobado
CA17	1	0.5	0,5	1	0,5	1	0,5	0,5	0,5	1	0,5	7.5	Reprobado
CA18	1	1	0,5	0,5	0,5	0,5	1	0,5	0	1	0,5	7	Reprobado
CA19	1	1	1	1	1	1	1	1	1	1	1	11	Aprobado
CA20	1	0.5	0,5	0,5	1	0,5	0,5	0,5	1	0,5	0,5	7.5	Reprobado
CA21	0.5	0.5	0,5	1	0,5	1	0,5	1	0,5	1	0,5	7	Reprobado
CA22	1	0.5	0,5	1	0,5	1	0,5	0,5	0,5	1	0,5	7.5	Reprobado
CA23	0,5	0,5	1	0,5	1	0,5	0,5	0,5	0,5	0,5	0,5	6.5	Reprobado
Documentos aprobados												7	
Documentos descartados												16	

Elaborado por: Los investigadores

Una vez realizado la primera selección de datos nos dio un resultado de 7 artículos el mismo que serán utilizados para la segunda selección, la cual consiste en base a dos preguntas de investigación que son:

- ¿Qué técnicas se utilizaron para el reconocimiento de especies vegetales durante los últimos 5 años?
- ¿Cuál fue la técnica con mayor frecuencia que se utilizó en los estudios?

Estos artículos serán evaluados en base al título de los artículos donde debe contener al menos las palabras claves como: técnicas de máquinas de aprendizaje, especies vegetales y reconocimiento. En Tabla 14, detalla de los 7 artículos seleccionados 5 artículos tienen como enfoque de especie a las plantas medicinales por lo que se optó por la selección de estas plantas básicamente en menta y orégano. En respecto al desarrollo de los artículos están entre el año 2018- 2019 por lo que cumple con el problema de proyecto de investigación en la metodología de revisión sistemática literaria.

Tabla 14. Artículos con técnicas de aprendizaje de máquinas

Número de artículo	Nombre	Base de datos	Especie vegetal	Año	Técnica de aprendizaje de máquinas	Algoritmo	Porcentaje de validación
CA3	Detection of nutrition deficiencies in plants using proximal images and machine learning: A review	Science Direct	Plantas medicinales	2019	Aprendizaje no supervisado. Aprendizaje supervisado	Fuzzy Regresión logística	84% 97%
CA5	Automatic Recognition of Medicinal Plants using Machine Learning Techniques	Google académico	Plantas medicinales	2018	Aprendizaje supervisado	Regresión logística Bosque aleatorio	Un ponderado de 90.1%
CA7	Real-Time Identification of Medicinal Plants using Machine Learning Techniques	Google académico	Plantas medicinales	2019	Aprendizaje supervisado	Regresión logística	93.3%
CA10	Aplicación móvil para la detección y tratamiento de daños de los cultivos de la parroquia taura del cantón Durán, mediante el uso de software de análisis	Google académico	Planta medicinal	2018	Aprendizaje supervisado	Red neuronal	92%

	de imagen basado en técnicas de machine learning						
CA12	Identifying Green Vegetables Using Machine Learning Approach	Google académico	Vegetal verde	2020	Aprendizaje supervisado	Red neuronal	98%
CA16	Apr Machine learning for LC-MS medicinal plants identification	Google académico	Planta medicinales	2018	Aprendizaje supervisado	Regresión logística Máquina de vectores de soporte	99.74% 94.11%
CA19	Machine Learning Methods Analysis For Calories Measuremnt of Fruits and Vegetables	Google académico	Frutas y vegetales	2021	Aprendizaje supervisado	Bosque aleatorio	96%

Elaborado por: Los investigadores

En la Tabla 14, la mayoría de estudios en especies vegetales se basan en las plantas medicinales debido a la variedad de textura que existen.

5.1.3. Discusión de resultados

En base a la Tabla 15, se detalla según la frecuencia de los años de publicación los artículos están aptos para el desarrollo de investigación donde pasaron por dos revisiones las cuales arrojan los siguientes resultados:

Tabla 15. Frecuencia de algoritmos utilizadas en los artículos

Ítem	Algoritmo de técnica de aprendizaje de máquinas	Frecuencia	Porcentaje por artículo	Porcentaje de validación
1	Red neuronal	2	98% 92%	95%
2	Regresión logística	4	97% 99.74% 90.1% 93.3%	95.03%
3	Bosque aleatorio	2	96% 90.1%	93.05
4	Máquina de vectores de soporte	1	94.11%	94.04
5	Fuzzy	1	84%	84%

Elaborado por: Los investigadores

En la Tabla 15, se detalla las frecuencias de utilización del algoritmo de técnicas de aprendizaje de máquinas con respecto a la utilización de los 9 artículos seleccionados. Se demuestra que la mayoría se basan a través de la regresión logística se toma en cuenta para

desarrollar las investigaciones lo que nos permite la elección de este algoritmo debido al porcentaje de validación es de 95.03% con respecto a los demás algoritmos utilizados.

5.2. Metodología KDD

En la realización de la metodología se optó por este, debido a que cuenta por tareas iterativas la cual permite el cambio en cualquier fase de desarrollo y adicional a esto permite la verificación de la base de datos, así como el algoritmo.

5.2.2.1. Etapa de selección

En esta etapa se detalla las tareas que constan en la primera fase de esta metodología, en donde se van a determinar los objetivos de la minería de datos, requisitos y un plan de proyecto.

a. Reconocimiento de las especies vegetales

El objetivo que se va aplicar en este modelo de negocio es reconocer en tiempo real si existe una especie vegetal sea menta u orégano a partir de una imagen .

b. Medición de Altura de la Planta de Maíz

Aplicar un algoritmo de aprendizaje de máquinas que permita reconocer en tiempo real a la especie vegetal.

c. Evaluación de la Situación

Para el desarrollo del modelo de reconocimiento de la planta se cuenta con un total de 500 imágenes digitales entre plantas de orégano y menta las cuales fueron tomadas en diferentes estados de iluminación.

d. Costes y beneficios

Las imágenes se recopilaron desde una parcela propia. Este proyecto de investigación no genera un beneficio económico a la universidad, sin embargo, beneficia en el ámbito educativo a los estudiantes y profesores de la Carrera de Ingeniería en informática y sistemas computacionales de la Universidad Técnica de Cotopaxi.

e. Determinar los Objetivos de la Minería de Datos

- Recolectar imágenes digitales de plantas vegetales (orégano y menta) en diferentes estados de iluminación.
- Obtener un modelo de reconocimiento en tiempo real de plantas vegetales en imágenes digitales.

- Conseguir un modelo que permita el reconocimiento en tiempo real de las plantas vegetales.

f. Plan del proyecto

El proyecto de investigación se basa en las siguientes fases para el desarrollo de aprendizaje de máquinas como se muestra en la Tabla 16.

Tabla 16. Plan de proyecto de la metodología KDD

Nombre de la tarea	Duración	Comienzo	Fin
Fase-Comprensión del negocio	5 días	06-12-2021	10-12-2021
Fase-Comprensión de datos	10 días	13-12-2021	24-12-2021
Fase-Preparación de datos	10 días	27-12-2021	7-01-2022
Fase-Modelado	15 días	10-01-2022	28-01-2022
Fase-Evaluación	7 días	31-01-2022	08-02-2022

Elaborado por: Los investigadores

5.2.2.2. Etapa de pre-procesamiento

En esta fase se realiza la recolección inicial de los datos para determinar su calidad de las imágenes tomadas y así establecer un contacto con el problema, además con familiarizarse.

a. Recolectar los datos iniciales

En la recolección de datos iniciales se eligió 100 imágenes para el modelo de detección y algoritmo de reconocimiento de aprendizaje de máquinas con un objeto de referencia, tomadas las imágenes desde un dispositivo móvil, en diferentes estados del día y clima, en un formato JPG y con diferentes ángulos para tener una mejor calidad de pixeles de la imagen.

b. Descripción de datos

La base de datos para el modelo de detección está constituida por 25 atributos resultantes de la segmentación (Convolución, Difuminación, Luminosidad, Densidad y Umbralización) y la aplicación de filtros (Blur, Canny Dilate, Eroded e Imagen Original), y los resultados están dados en binario (0 y 1).

5.2.2.3. Etapa de transformación

En esta fase de transformación de los datos se preparó los datos para adecuarlos a las técnicas de minería de datos. Esto conlleva a seleccionar los datos para limpiarlos, así mejorar su calidad y darles el formato adecuado y requerido para cada modelo.

segundo se denomina elemento estructurante o núcleo y cv2.erode se utiliza para realizar la erosión en la imagen como se muestra en la **Figura 19**.

```
import cv2
import numpy as np

img = cv2.imread('2.jpg')
img = cv2.resize(img, (300, 250))

kernel = np.ones((5, 5), np.uint8)

imgGray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
imgBlur = cv2.GaussianBlur(imgGray, (5, 5), 1)
imgCanny = cv2.Canny(img, 200, 200)
imgDialation = cv2.dilate(imgCanny, kernel, iterations=1)
imgEroded = cv2.erode(imgDialation, kernel, iterations=1)
```

Figura 19. Transformación de datos

La función cv.threshold se utiliza para aplicar el umbral. Mediante argumentos: el primer argumento es la imagen de origen, que debe ser una imagen en escala de grises y el segundo argumento es el valor umbral que se utiliza para clasificar los valores de píxeles. El tercer argumento es el valor máximo que se asigna a los valores de píxeles que superan el umbral. Esto se detalla en la **Figura 20**.

```
import cv2
import numpy as np
from matplotlib import pyplot as plt

img = cv2.imread('img.jpg', 0)

u, th1 = cv2.threshold(img, 0, 255, cv2.THRESH_BINARY+cv2.THRESH_OTSU)
u, th2 = cv2.threshold(img, 0, 255, cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)
u, th3 = cv2.threshold(img, 0, 255, cv2.THRESH_TRUNC+cv2.THRESH_OTSU)
u, th4 = cv2.threshold(img, 0, 255, cv2.THRESH_TOZERO+cv2.THRESH_OTSU)
u, th5 = cv2.threshold(img, 0, 255, cv2.THRESH_TOZERO_INV+cv2.THRESH_OTSU)

imagenes = [img, th1, th2, th3, th4, th5]
titulo = ['im1', 'im2', 'im3', 'im4', 'im5', 'im6']

for i in range(6):
    plt.subplot(3,2,i+1)
    plt.imshow(imagenes[i], 'gray', vmin=0, vmax=255)
    plt.title(titulo[i])
    plt.xticks([], plt.yticks([]))

plt.show()
print(u)
```

Figura 20. Utilización de función cv.threshold

Una vez realizada las funciones anteriormente mencionadas nos dan valores de números para la obtención de datos de la base de datos. Como se muestra en la **Figura 21**.

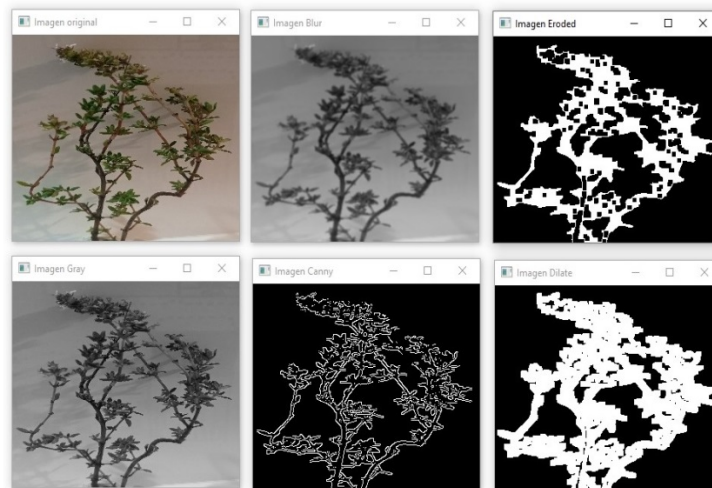


Figura 21. Obtención de valores de orégano

Una vez seleccionado los datos de entrada para la aplicación de la regresión logística nos dan valores como se muestra en la **Figura 22.**

```

C:\Windows\System32\cmd.exe - python data1.py
img_original_convolucion : 0,232682686
img_blur_convolucion : 0,115909407
img_canny_convolucion : 0,914621976
img_dilate_convolucion : 0,20361477
img_eroded_convolucion : 0,312435217
img_original_difuminacion : 0,389714002
img_blur_difuminacion : 0,01262828

```

Figura 22. Valores de segmentación de datos

5.2.2.4. Etapa de minería de datos-modelado

En esta fase se escogerá la técnica adecuada para cumplir los objetivos mencionado en la fase primera una vez aplicado las técnicas de aprendizaje de máquinas se eligió el algoritmo regresión logística el cual se va evaluar si el modelo cumple con los criterios de éxitos o no.

a. Técnica de modelado

Para la realización del reconocimiento de la planta se utilizó el algoritmo de regresión logística con un modelo pre - entrenado para la detección de objetos. Para alojar el entrenamiento se utilizó la herramienta Keras, esta herramienta guarda el proceso mediante códigos entre 0 y 1 formando matrices lo que nos permite realizar el reconocimiento.

Para ello se realiza la lectura del data set y dibujar la nube de puntos de categoría como se muestra en la **Figura 23.**

```
datos = pd.read_csv('dataset.csv', sep=",")
print(datos)
```

Figura 23. Leer data set

Se crean los datos de entrenamiento x y la categoría de salida, estos datos de tipo numérico y la salida de tipo varchar y como se muestra en la **Figura 24**.

```
X = datos.values[:,0:2]
Y = datos.values[:,2]
```

Figura 24. Creación de datos de entrenamiento y salida

Creando los datos se realiza la gráfica de la especie de orégano y menta a través de la obtención de los datos numéricos y la salida correspondiente como muestra en la **Figura 25**.

```
idx0 = np.where(Y==0)
idx1 = np.where(Y==1)

plt.scatter(X[idx0,0],X[idx0,1],color='red',label='Categoría: 0')
plt.scatter(X[idx1,0],X[idx1,1],color='gray',label='Categoría: 1')
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
plt.legend(bbox_to_anchor=(0.765,0.6),fontsize=8,edgecolor='black')
plt.title('Datos originales')
plt.show()
```

Figura 25. Gráfica de la segmentación

Se realiza la creación del modelo de Keras el cual el `input_dim` es 2 pues que cada dato tiene 25 características, `output_dim` pues se tiene una de dos posibles categorías 0 y 1. Se utilizó la activación por sigmoidea en la **Figura 26**.

```
np.random.seed(1) # para reproducibilidad
input_dim = X.shape[1]
output_dim = 1

modelo = Sequential()
modelo.add(Dense(output_dim, input_dim = input_dim, activation='sigmoid'))
```

Figura 26. Creación del Modelo Keras

Se realiza la optimización el cual se usa el gradiente descendente, la función de error entropía y la métrica para el desempeño. Así como el entrenamiento del algoritmo para un mejor resultado de reconocimiento de la especie vegetal como se muestra en la **Figura 27**.

```
sgd = SGD(lr=0.2)
modelo.compile(loss='binary_crossentropy', optimizer=sgd, metrics=['accuracy'])

# Entrenamiento
#
num_epochs = 1000
batch_size = X.shape[0]
historia = modelo.fit(X, Y, epochs=num_epochs, batch_size=batch_size, verbose=2)
```

Figura 27. Creación del gradiente descendente

Se realiza la gráfica de comportamiento de la pérdida y de la precisión y la gráfica de frontera de decisión **Figura 28**.

```
# Graficar comportamiento de la pérdida y de la precisión
plt.subplot(1,2,1)
plt.plot(historia.history['loss'])
plt.ylabel('Pérdida')
plt.xlabel('Epoch')
plt.title('Comportamiento de la pérdida')

plt.subplot(1,2,2)
plt.plot(historia.history['accuracy'])
plt.ylabel('Precisión')
plt.xlabel('Epoch')
plt.title('Comportamiento de la precisión')

ax = plt.gca()
ax.yaxis.set_label_position("right")
ax.yaxis.tick_right()

plt.show()

#
# Graficar frontera de decisión
#
def dibujar_frontera(X,Y,modelo,titulo):
    # Valor mínimo y máximo y rellenado con ceros
    x_min, x_max = X[:, 0].min() - .5, X[:, 0].max() + .5
    y_min, y_max = X[:, 1].min() - .5, X[:, 1].max() + .5
    h = 0.01
```

Figura 28.Comportamiento de pérdida y precisión

Una vez realizado el comportamiento de pérdida y de precisión se realiza la grilla de puntos con el objetivo de predecir categorías para cada punto en la grilla y se aplica el modelo de regresión logística para el reconocimiento de menta y orégano **Figura 29**.

```
# Grilla de puntos
xx, yy = np.meshgrid(np.arange(x_min, x_max, h), np.arange(y_min, y_max, h))

# Predecir categorías para cada punto en la grilla
Z = modelo.predict_classes(np.c_[xx.ravel(), yy.ravel()])
Z = Z.reshape(xx.shape)

fig = plt.figure()
plt.contourf(xx, yy, Z, cmap=plt.cm.Set1, alpha=0.8)

idx0 = np.where(Y==0)
idx1 = np.where(Y==1)
plt.scatter(X[idx0,0],X[idx0,1],color='red', edgecolor='k', label='Categoría: 0')
plt.scatter(X[idx1,0],X[idx1,1],color='gray',edgecolor='k', label='Categoría: 1')
plt.legend(bbox_to_anchor=(0.765,0.6),fontSize=8,edgecolor='black')

plt.xlim(xx.min(), xx.max())
plt.ylim(yy.min(), yy.max())
plt.title(titulo)

plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
plt.show()

dibujar_frontera(X,Y,modelo,'Frontera de decisión después del entrenamiento')
```

Figura 29.Modelo de regresión logística

Finalmente se puede observar el comportamiento del algoritmo regresión logística **Figura 30**.

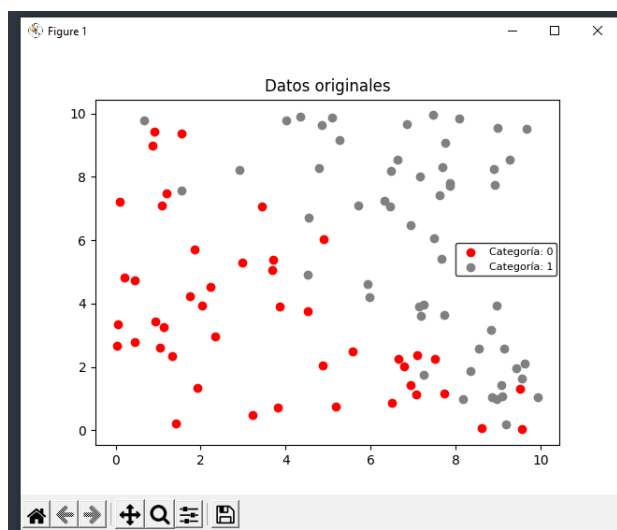


Figura 30. Comportamiento de la regresión logística

5.2.2.5. Etapa de interpretación

Finalmente, en la interpretación se verifica si la clasificación se realizó correctamente en base al uso del algoritmo de regresión logística para lo que se utilizó la validación cruzada para la verificación e interpretación de los resultados.

a. Validación cruzada

Se utilizó la técnica de la validación cruzada para el cual se separó en conjunto de datos inicial de 500 registros en 2: conjunto de entrenamiento (train) y conjunto de pruebas (test). Por lo general se divide haciendo “400-100”. Y se toman muestras aleatorias, no en secuencia, si no, mezcladas.

Cross-Validation: K-fold con 5 splits

Se aplicó K-Folds con 5 splits en el cual del total de 500 registros, se seleccionó 400 registros para cada conjunto de Train y 100 registros para el conjunto de test de forma aleatoria como se explica a continuación:

- Iterar 5 veces:
 1. Apartaremos el 400 del total de los registros de forma aleatoria para el Train.
 2. Apartaremos 100 de los registros de forma aleatoria para el test.
 3. Entrenamos al modelo con los registros seleccionados.
 4. Mediremos el accuracy obtenido sobre los 500 registros que se seleccionaron aleatoriamente.

- Se realizará 5 entrenamientos independientes ya que se va a seleccionar los datos de forma aleatoria 5 veces.
- El accuracy final será el promedio de las 5 accuracy anteriores

K -fold Iteración N° 1

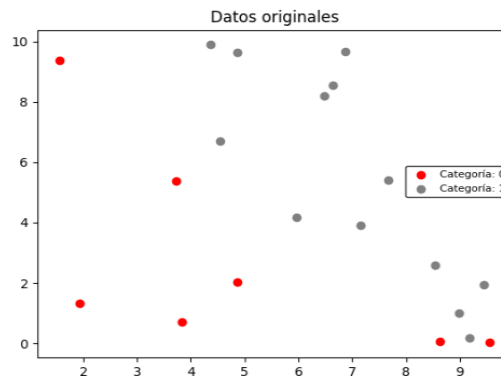


Figura 31. K -fold Iteración N° 1

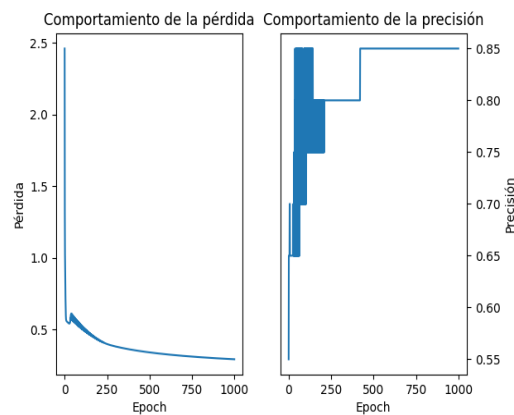


Figura 32. Comprobación de pérdida y precisión Iteración 1

```
Epoch 995/1000
1/1 - 0s - loss: 0.2915 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 996/1000
1/1 - 0s - loss: 0.2914 - accuracy: 0.8500 - 2ms/epoch - 2ms/step
Epoch 997/1000
1/1 - 0s - loss: 0.2914 - accuracy: 0.8500 - 2ms/epoch - 2ms/step
Epoch 998/1000
1/1 - 0s - loss: 0.2913 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 999/1000
1/1 - 0s - loss: 0.2912 - accuracy: 0.8500 - 2ms/epoch - 2ms/step
Epoch 1000/1000
1/1 - 0s - loss: 0.2912 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
```

Figura 33. Resultado ejecutado de la Iteración 1

Finalmente se detallan los valores obtenidos de la Iteración 1 en la Tabla 17.

Tabla 17. Valores del resultado de Iteración 1

Resultados Iteración N° 1:	
Steps/Epoch:	1000
Accuracy:	0.8500
Loss:	0.2912

K -fold Iteración N° 2

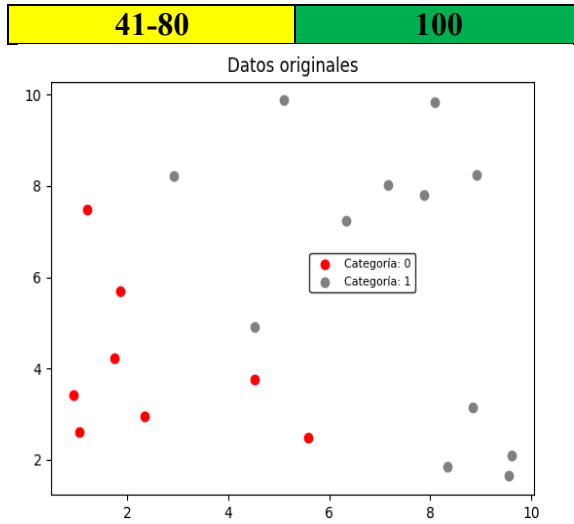


Figura 34. K -fold Iteración N° 2

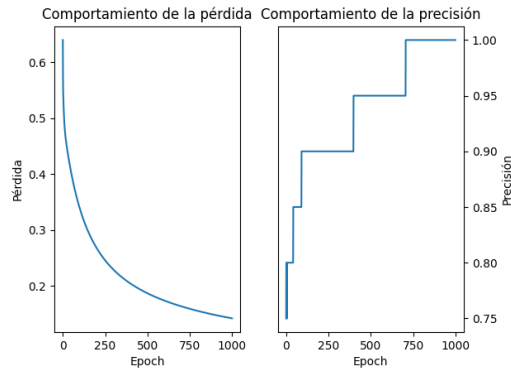


Figura 35. Comprobación de pérdida y precisión Iteración 2

```
Epoch 995/1000
1/1 - 0s - loss: 0.1426 - accuracy: 1.0000 - 2ms/epoch - 2ms/step
Epoch 996/1000
1/1 - 0s - loss: 0.1426 - accuracy: 1.0000 - 2ms/epoch - 2ms/step
Epoch 997/1000
1/1 - 0s - loss: 0.1425 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
Epoch 998/1000
1/1 - 0s - loss: 0.1425 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
Epoch 999/1000
1/1 - 0s - loss: 0.1424 - accuracy: 1.0000 - 2ms/epoch - 2ms/step
Epoch 1000/1000
1/1 - 0s - loss: 0.1423 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
```

Figura 36. Resultado ejecutado de la Iteración 2

Finalmente se detallan los valores obtenidos de la Iteración 2 en la Tabla 18.

Tabla 18. Valores del resultado de Iteración 2

Resultados Iteración N° 2	
Steps/Epoch:	1000
Accuracy:	1.0000
Loss:	0.1423

K -fold Iteración N° 3

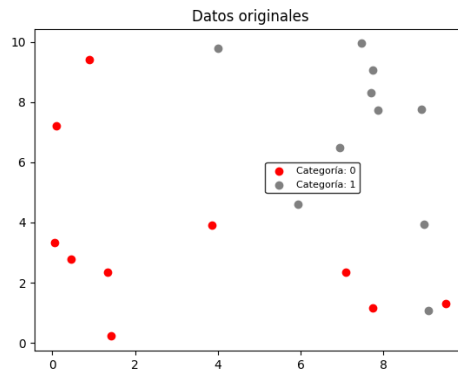


Figura 37. K -fold Iteración N° 3

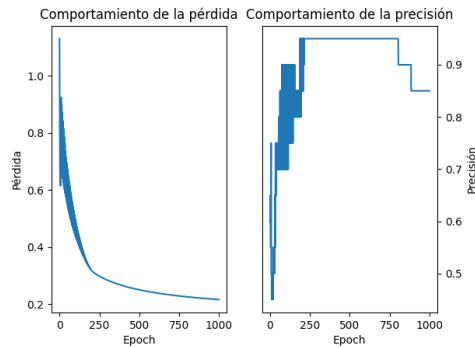


Figura 38. Comprobación de pérdida y precisión Iteración 3

```

Epoch 995/1000
1/1 - 0s - loss: 0.2160 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 996/1000
1/1 - 0s - loss: 0.2160 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 997/1000
1/1 - 0s - loss: 0.2159 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 998/1000
1/1 - 0s - loss: 0.2159 - accuracy: 0.8500 - 2ms/epoch - 2ms/step
Epoch 999/1000
1/1 - 0s - loss: 0.2158 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
Epoch 1000/1000
1/1 - 0s - loss: 0.2158 - accuracy: 0.8500 - 3ms/epoch - 3ms/step
    
```

Figura 39. Resultado ejecutado de la Iteración 3

Finalmente se detallan los valores obtenidos de la Iteración 3 en la Tabla 19.

Tabla 19. Valores del resultado de Iteración 3

Resultados Iteración N° 2	
Steps/Epoch:	1000
Accuracy:	0.8500
Loss:	0.2158

K -fold Iteración N° 4

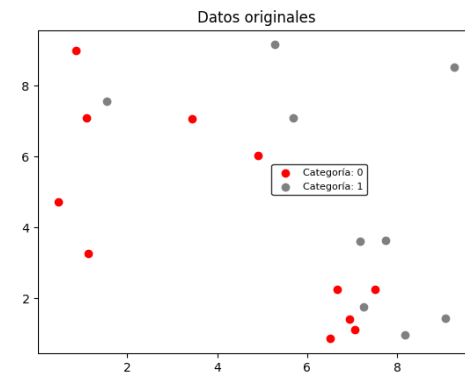


Figura 40. K -fold Iteración N° 4

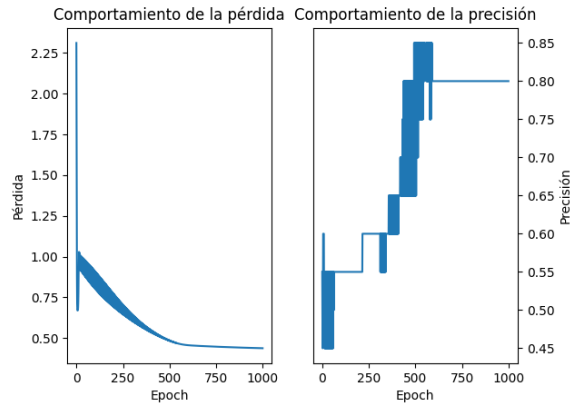


Figura 41. Comprobación de pérdida y precisión Iteración 4

```
Epoch 995/1000
1/1 - 0s - loss: 0.4382 - accuracy: 0.8000 - 4ms/epoch - 4ms/step
Epoch 996/1000
1/1 - 0s - loss: 0.4381 - accuracy: 0.8000 - 4ms/epoch - 4ms/step
Epoch 997/1000
1/1 - 0s - loss: 0.4381 - accuracy: 0.8000 - 3ms/epoch - 3ms/step
Epoch 998/1000
1/1 - 0s - loss: 0.4381 - accuracy: 0.8000 - 5ms/epoch - 5ms/step
Epoch 999/1000
1/1 - 0s - loss: 0.4381 - accuracy: 0.8000 - 2ms/epoch - 2ms/step
Epoch 1000/1000
1/1 - 0s - loss: 0.4380 - accuracy: 0.8000 - 3ms/epoch - 3ms/step
```

Figura 42. Resultado ejecutado de la Iteración 4

Finalmente se detallan los valores obtenidos de la Iteración 4 en la **Tabla 20**.

Tabla 20. Valores del resultado de Iteración 4

Resultados Iteración N° 2	
Steps/Epoch:	1000
Accuracy:	0.8000
Loss:	0.4380

K -fold Iteración N° 5

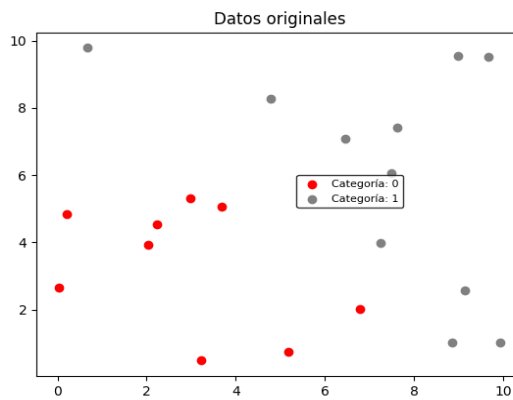


Figura 43. K -fold Iteración N° 5

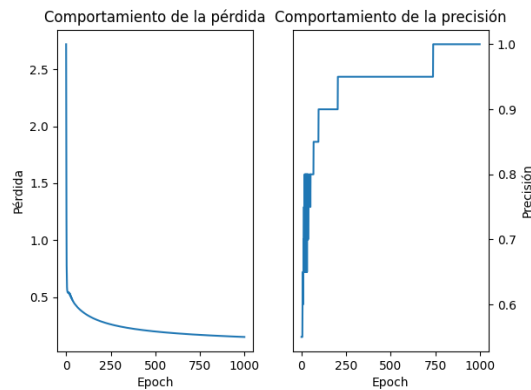


Figura 44. Comprobación de pérdida y precisión Iteración 5

```
Epoch 995/1000
1/1 - 0s - loss: 0.1494 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
Epoch 996/1000
1/1 - 0s - loss: 0.1493 - accuracy: 1.0000 - 2ms/epoch - 2ms/step
Epoch 997/1000
1/1 - 0s - loss: 0.1493 - accuracy: 1.0000 - 4ms/epoch - 4ms/step
Epoch 998/1000
1/1 - 0s - loss: 0.1492 - accuracy: 1.0000 - 2ms/epoch - 2ms/step
Epoch 999/1000
1/1 - 0s - loss: 0.1492 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
Epoch 1000/1000
1/1 - 0s - loss: 0.1491 - accuracy: 1.0000 - 3ms/epoch - 3ms/step
```

Figura 45. Resultado ejecutado de la Iteración 5

Finalmente se detallan los valores obtenidos de la Iteración 5 en la **Tabla 21**.

Tabla 21. Valores del resultado de Iteración 5

Resultados Iteración N° 2	
Steps/Epoch:	1000
Accuracy	1.0000
Loss:	0.1491

Finalmente se da un exactitud del 1% lo que equivale que el algoritmo de regresión logística cumple con el 99% de validez para el reconocimiento de la planta de menta y orégano.

5.3. Metodología Mobile-D

Para la construcción de esta aplicación móvil, se seleccionó la plataforma de Android Studio y se aplicó la metodología MOBILE-D, esta metodología permite desarrollar un proyecto de manera ágil para software de menos complejidad y en un tiempo corto, utilizando menos recursos para aplicaciones móviles.

En la metodología Mobile-D encontramos cinco fases aplicables que son:

5.2.3.1. Exploración

En esta primera fase se delimita el alcance de la aplicación móvil que se desarrolla como prototipo para ejemplificar la investigación realizada en este documento, la aplicación permite la detección de especies vegetales mediante capturas fotográficas instantáneas o imágenes de la galería, la determinación del nombre entre una planta de menta y otra de orégano con una breve descripción de cada una, también nos da una detención a tiempo real de cualquiera de las dos plantas.

El proceso de detección de imagen se basa en la técnica de regresión perteneciente a la rama de machine learning con el algoritmo regresión logística. Este proyecto está diseñado para estudiantes de carreras afines a las temáticas tratadas, así como también para profesores que requieran hacer uso de estos temas dentro del grupo de investigación de la Universidad Técnica de Cotopaxi de la ciudad de Latacunga donde se desarrolla toda la investigación.

5.2.3.2. Inicialización

Los requerimientos necesarios para el desarrollo de esta aplicación se determinaron y planearon en esta fase del proyecto.

a) Historias de usuario

Tabla 22. Historia de usuario capturar imagen

HISTORIA DEL USUARIO
Numero: 1
Nombre: capturar imagen.
Usuario: Usuario
Descripción: El usuario tiene el acceso a la interfaz del aplicativo donde captura una imagen desde la cámara del dispositivo y una vez detectada la planta se visualiza una breve descripción de la misma.
Fuente: Grupo de Investigadores

Tabla 23. Historia de usuario buscar en la galería

HISTORIA DEL USUARIO
Numero: 2
Nombre: Buscar en la Galería
Usuario: Usuario
Descripción: El usuario tiene acceso a la interface de la app donde tiene la opción de elegir el botón Galería e ingresa a la galería de fotos y elige la una. Una vez detectada la planta se visualiza un abreve descripción de la misma.
Fuente: Grupo de Investigadores

Tabla 24. Historia de usuario detención a tiempo real

HISTORIA DEL USUARIO
Numero: 3
Nombre: Detección a Tiempo Real
Usuario: Usuario
Descripción: El usuario desde la interfaz de la app tiene acceso a la opción Tiempo Real donde ingresa a la cámara del dispositivo y buscara imágenes al instante de las plantas elegidas. El aplicativo muestra en la parte inferior el número de plantas detectadas.

b. Requerimientos funcionales

RF1: Capturar la imagen con el dispositivo móvil para el reconocimiento de imagen.

RF2: Seleccionar Imagen desde la galería

RF3: Detectar imágenes a tiempo real.

RF4: Visualizar nombre y descripción de la planta.

Requerimientos no Funcionales

RNF1: Android Studio

RNF2: SQL lite

RNF4: Flutter framework multiplatform

RNF4: OPENCV

De la manera más eficaz se seleccionó los requerimientos funcionales y no funcionales para la programación de la aplicación móvil y de esta manera optimizar recursos y tiempo en el desarrollo. Para la construcción de esta aplicación se eligió el entorno de desarrollo de Android Studio junto con el SDK Flutter de código abierto que no necesita la utilización de componentes intermedios y trabaja bajo el lenguaje de programación Dart, para el alojamiento del modelo se utilizó la librería OPENCV para detección de imágenes de plantas y SQLite como base de datos local.

5.2.3.3. Producto

Es la fase en la que se lleva a cabo la implementación de la aplicación, para lo cual se debe construir diseños sencillos de utilizar y fáciles de comprender para la comodidad del usuario.

a) Diagramas de casos de uso

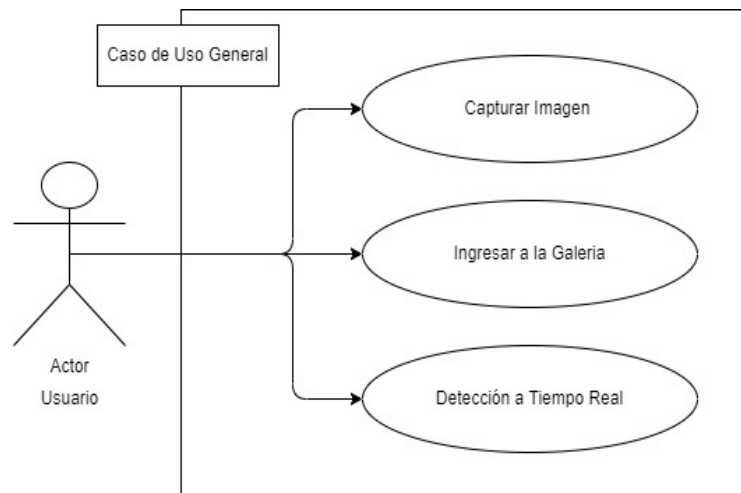


Figura 46. Diagrama de cas de uso general

b) Diagrama a detalle

Tabla 25. CU001-Capturar imagen

Núm.:	CU001
Nombre:	Capturar Imagen
Autores:	Tesistas investigadoras.
Descripción:	El usuario realiza la Detección de la imagen desde la cámara.
Actor:	Usuario
Precondiciones:	El usuario debe tener la aplicación.

Flujo normal:	<ol style="list-style-type: none"> 1.-El usuario ingresa a la aplicación. 2.-El usuario selecciona el botón “Capturar”. 3. La aplicación muestra la interfaz para tomar la fotografía. 4.-La aplicación activa la cámara del dispositivo. 5.-El usuario procede a tomar la foto. 6.- La aplicación detecta la planta de “menta” u “orégano”. 7.- El aplicativo muestra el nombre de la planta en la parte inferior y un cuadro de descripción.
Flujo alternativo 1:	<p>OPCIÓN INCORRECTA.</p> <ol style="list-style-type: none"> 3.1- El usuario no selecciona el botón de capturar. 3.2- El usuario regresa al paso 1. 6.1 La aplicación no detecta la planta. 6.1.1- La aplicación muestra un mensaje con “No se detectaron plantas”regresa al paso 5.
Post-condiciones:	El usuario realiza la detección de la planta con éxito.

Tabla 26.CU002-Ingresar a la galería

Núm.:	CU002
Nombre:	Ingresar a la galería
Autores:	Tesistas investigadoras.
Descripción:	El usuario realiza la Detección de la plata desde una imagen previamente guardada en la galería del dispositivo.
Actor:	Usuario
Precondiciones:	El usuario debe tener la aplicación y tener una imagen de la planta en la galería.
Flujo normal:	<ol style="list-style-type: none"> 1.-El usuario ingresa a la aplicación. 2.-El usuario selecciona el icono “Galería”. 3. La aplicación muestra la interfaz para seleccionar la imagen de la galería. 4.-El usuario selecciona la imagen. 5.- La aplicación detecta si corresponde a la planta de “menta” u “orégano”. 6.- El aplicativo muestra el nombre de la planta y la descripción
Flujo alternativo 1:	<p>OPCIÓN INCORRECTA.</p> <ol style="list-style-type: none"> 2.1- El usuario no selecciona el botón de Galería. 2.2- El usuario regresa al paso 1. 5.1 La aplicación no detecta la planta. 5.1.1- La aplicación muestra un mensaje con “No se detectaron plantas”regresa al paso 3.
Post-condiciones:	El usuario realiza la Detección de la planta con éxito.

Tabla 27.CU003 Detención a Tiempo Real

Núm.:	CU003
Nombre:	DetECCIÓN a Tiempo Real
Autores:	Tesistas investigadoras.
Descripción:	El usuario puede realizar la detección a tiempo real mediante la cámara del dispositivo.
Actor:	Usuario
Precondiciones:	El usuario debe tener la aplicación.
Flujo normal:	<ol style="list-style-type: none"> 1.-El usuario ingresa a la aplicación. 2.-El usuario selecciona el botón “Tiempo Real”. 3. La aplicación abre la cámara del dispositivo directamente. 4.-El usuario enfoca el objeto para su detección. 5.- La aplicación detecta si corresponde a la planta de “menta” u “orégano”. 6.- El aplicativo muestra el nombre en la parte inferior.
Flujo alternativo 1:	<p>OPCIÓN INCORRECTA.</p> <ol style="list-style-type: none"> 5.1 La aplicación de detecta ninguna de las dos plantas. 5.1.1 La aplicación muestra un mensaje “No se encontraron coincidencias”.
Post-condiciones:	El usuario realiza la Detección de la planta con éxito.

c) Diagrama de secuencia

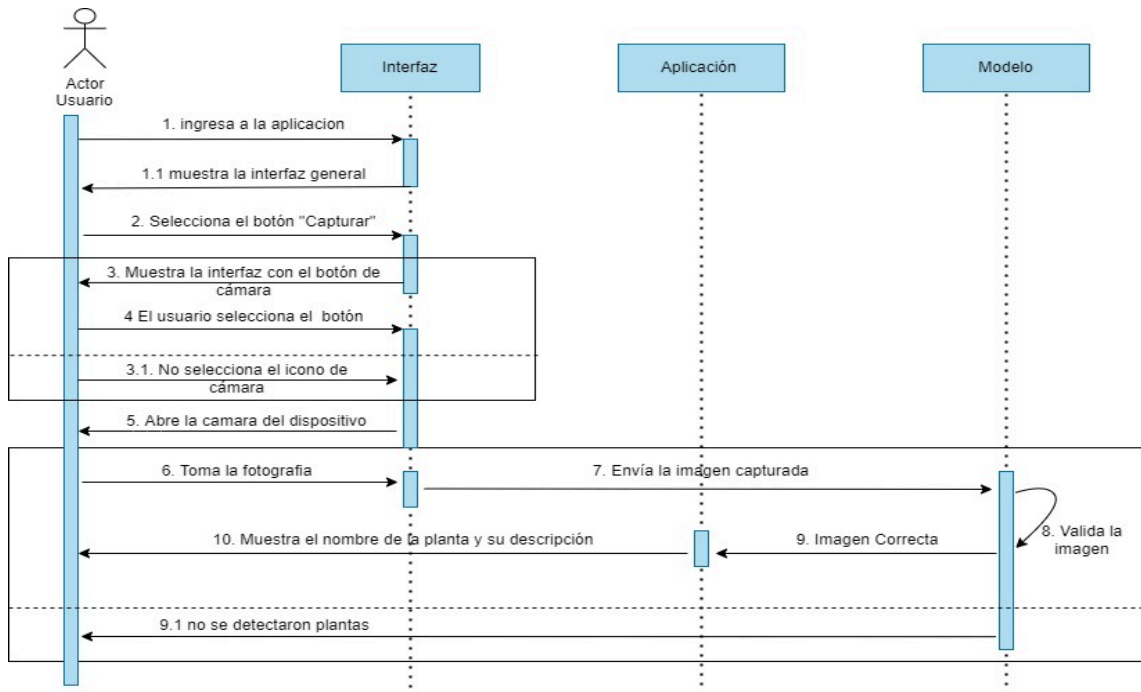


Figura 47.Diagrama de secuencia de captura de imagen.

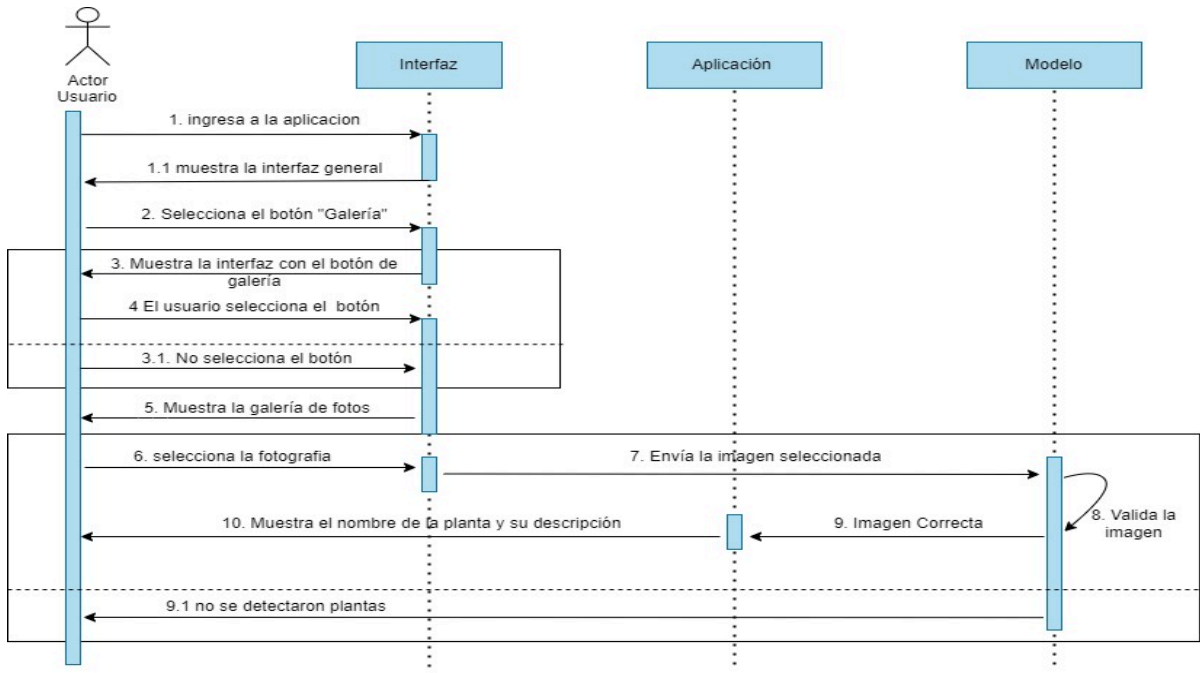


Figura 48. Diagrama de secuencia de ingresar galería

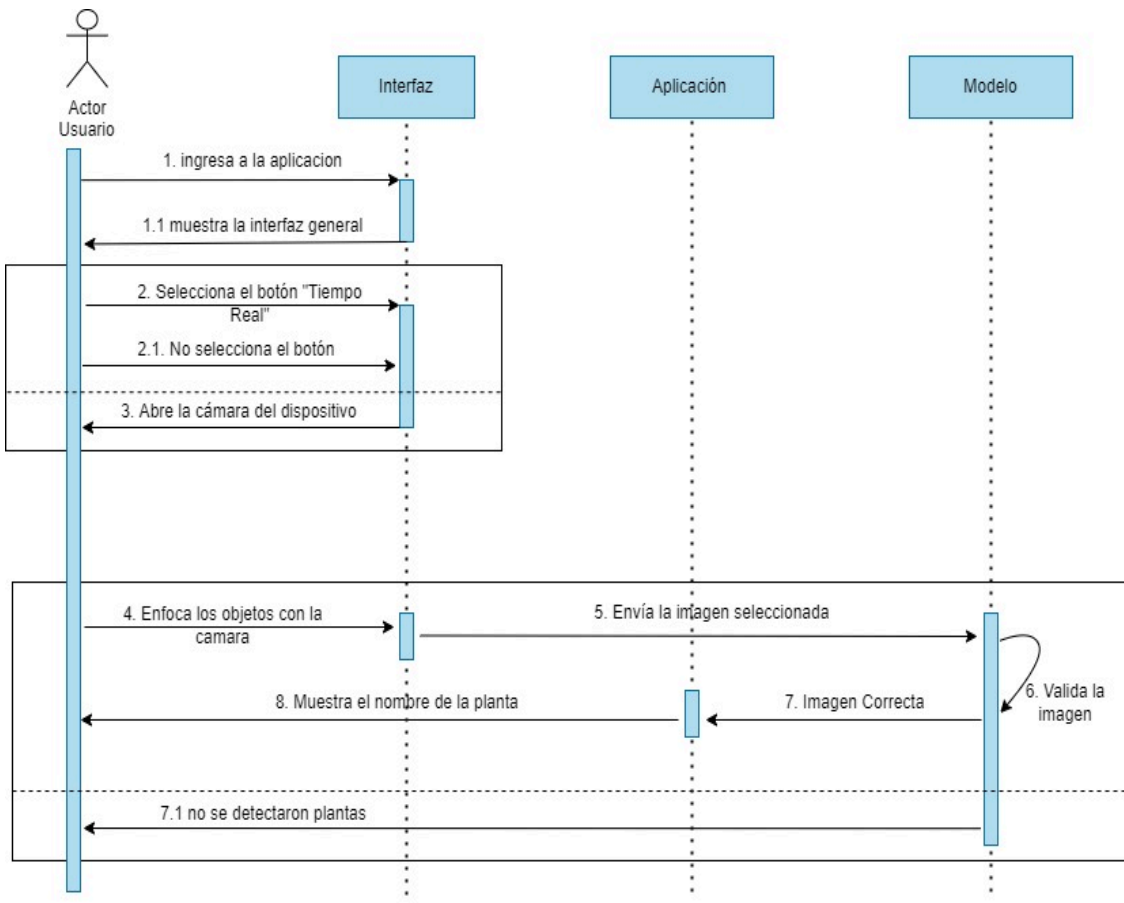


Figura 49. Diagrama de secuencia de detección a tiempo real

d) Diagrama de clases de la aplicación – servidor

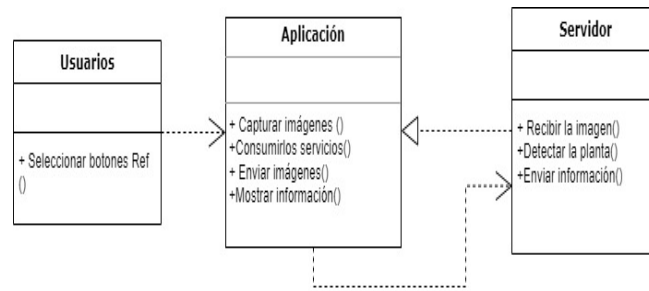


Figura 50.Diagrama de clases del aplicativo

e) Diseño de la aplicación

Pantalla principal del aplicativo



Figura 51. Interfaz gráfica de la pantalla principal de la aplicación

Pantalla para la captura de imagen.

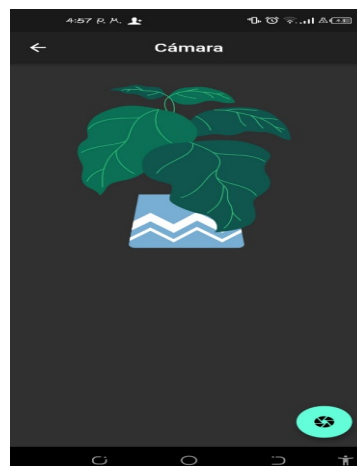


Figura 52. Interfaz gráfica detección a tiempo real

5.2.3.4. Pruebas

En esta etapa se realizaron las pruebas correspondientes para la verificación de manera correcta de las funcionalidades de la APP dentro del sistema.

a) Pruebas Unitarias

Mediante esta fase de pruebas se verificará el correcto funcionamiento del código verificando cada una de sus partes, ya que sus funcionamientos se fueron creando en base a los requerimientos que se iba obteniendo durante el proceso de la creación de la APP, esto con la finalidad de sus opciones sean de forma funcional, y que se visualice en la consola de Android Studio el aplicativo funcionando sin error alguno.

b) Pruebas de Sistemas

En este proceso se detalla todas las pruebas del sistema para verificar que no exista ningún error y esta se encuentre funcionando.

Tabla 28.Caso de prueba capturar imagen

CASO DE PRUEBA 1	SE ESPERA	RESULTADOS
Capturar la imagen mediante la cámara del dispositivo móvil para poder ingresar en el sistema y así poder detectar con el objeto de referencia.	Que la aplicación permite capturar la imagen mediante la cámara del dispositivo móvil para ingresar y que se así pueda determinar el tipo de planta (orégano o menta) de acuerdo a la objeto de referencia que se obtenga.	La aplicación permitió capturar la imagen con la cámara del dispositivo y así verificar la imagen y determinar el tipo de planta de manera satisfactoria.
PRUEBAS:	<p>Prueba 1: mediante el dispositivo móvil realizaremos la captura de una imagen la cual el sistema determinará si es (orégano o menta), de acuerdo a su textura.</p> <p>Prueba 2: verifica la imagen.</p>	
		
Evaluación de pruebas:		Aprobado

Tabla 29.Caso de prueba seleccionar imagen

CASO DE PRUEBA 2	SE ESPERA	RESULTADOS
<p>Seleccionar la imagen por medio de la galería de la APP para ingresar en el sistema.</p>	<p>Que la aplicación permite seleccionar la imagen de la galería para ingresar al sistema y así pueda determinar el tipo de planta (orégano o menta) de acuerdo a la imagen de referencia que se obtenga.</p>	<p>La aplicación permitió la agregación de la imagen a verificar y determinó el tipo de planta sin error alguno.</p>
<p>PRUEBAS:</p>	<p>Prueba 1: mediante el dispositivo móvil seleccionaremos una imagen de la galería de fotos la cual el sistema determinará si es (orégano o menta), de acuerdo a su textura.</p> <p>Prueba 2: verifica la imagen</p>	
		
<p>Figura 54. Caso de prueba seleccionar imagen</p>		
<p>Evaluación de pruebas:</p>		<p>Aprobado</p>

Tabla 30.Caso de prueba capturar imagen

CASO DE PRUEBA 3	SE ESPERA	RESULTADOS
<p>Capturar la imagen en tiempo real en la APP para ingresar en el sistema.</p>	<p>Que la aplicación permite identificar la imagen a través de la cámara del aplicativo para poder ingresar al sistema y así pueda determinar el tipo de planta (orégano o menta) de acuerdo a la imagen de referencia que se obtenga.</p>	<p>La aplicación permitió la verificación de la imagen en tiempo real para poder determinar el tipo de planta sin ninguna equivocación.</p>
<p>PRUEBAS:</p>	<p>Prueba 1: mediante la cámara del dispositivo móvil identificamos una imagen en tiempo real, la cual el sistema determinará si es (orégano o menta), de acuerdo a su textura.</p> <p>Prueba 2: verifica la imagen.</p>	

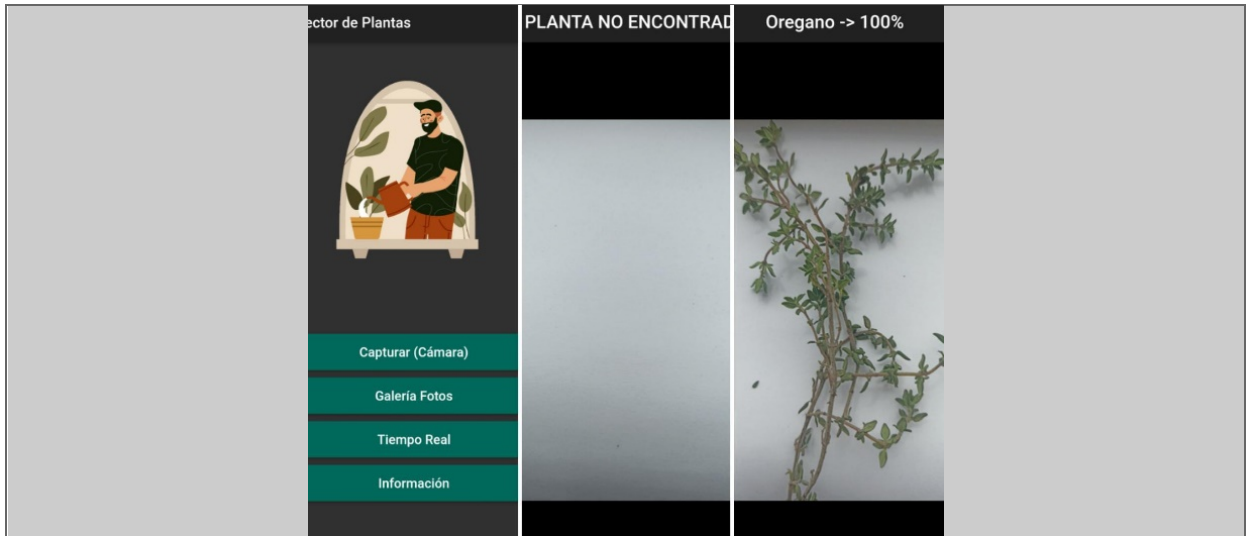


Figura 55. Caso de prueba capturar imagen

Evaluación de pruebas:	Aprobado
------------------------	----------

5.4. Comprobación de la hipótesis

Como hipótesis se planteó la siguiente: La aplicación de la técnica de regresión logística del aprendizaje de máquina, permitirá garantizar el reconocimiento de imágenes de las especies vegetales en el área de la agricultura. Se basó en una estrategia de evaluación el cual permite la posibilidad de obtener una amplia y pormenorizada información sobre el objeto de estudio y la calidad de las respuestas por parte de los jueces.

Se realizó la consulta a varios expertos, donde un ingeniero en sistemas verifica el funcionamiento de la aplicación móvil y un ingeniero agrónomo que permita verificar la valides de los datos de entrada para el reconocimiento de la planta vegetal (orégano y menta).

En base a seis preguntas que demuestran el funcionamiento de la aplicación móvil se detalla en el **Anexo D** y la calificación que da cada uno de los expertos para evaluar la aplicación móvil. Tomando en cuenta su trayectoria para que la evaluación tenga mayor porcentaje de validación del producto.

El resultado de la aplicación del juicio de expertos fue el primero concuerda que la aplicación cumple con los objetivos planteados y cumple con las funcionalidades entonces aprueba la aplicación móvil, mientras el ingeniero agrónomo verifica los datos y el funcionamiento de la aplicación móvil aprueba la aplicación móvil.

5.5. Análisis de impacto

5.5.1. Impacto social

El presente proyecto de investigación tiene un impacto social, porque da una información detallada para los investigadores sobre técnicas de aprendizaje de máquinas que son utilizados específicamente en el sector agrícola y adicional a esto beneficia al sector agrícola y sector botánico con la aplicación móvil de reconocimiento automático de especies vegetales usando regresión logística porque imita la forma en que los seres humanos toman decisiones de manera automática con respecto a que planta pertenece sea orégano o menta, ya que en el lenguaje natural se suele mencionar conceptos vagos o imprecisos, por tal razón se considera de gran ayuda esta aplicación.

5.5.2. Impacto tecnológico

Contar con información actualizada sobre las diferentes técnicas de aprendizaje de máquina tiene un impacto sobre alternativas de hacer uso de nuevas tecnologías y manipular tecnologías nuevas e innovadoras que ayuden a mejorar el reconocimiento de especies vegetales, por ello para demostración se creó una aplicación móvil que permita demostrar la investigación estudiada aplicando el algoritmo de regresión logística utilizado mayormente en los estudios realizados lo que permite al agricultor evitar la intervención de los trabajadores en el reconocimiento de las especies vegetales de forma rápida y precisa.

5.5.3. Impacto ambiental

Tiene un impacto ambiental porque al tener una aplicación donde me permita reconocer en tiempo real que especie vegetal a la que pertenece, ayuda al botánico en sus investigaciones que pueda tener.

5.5.4. Impacto económico

En el impacto económico contar con una aplicación móvil que ayude a reconocer en tiempo real las especies vegetales indudablemente permite la reducción de mano de obra debido a que se suplanta el reconocimiento manual por el reconocimiento automático mediante aprendizaje de máquinas, así como también el impacto que se consideró para el desarrollo de la investigación escrita y técnica como se detalla en el **Anexo E**.

5.5.4.1. Costo de software

Para tener el costo de software de la aplicación móvil se basó mediante el método de COCOMO, la utilidad de este modelo es utilizado en proyectos de software para estimar los costes. El cual se detalla una estimación de la cantidad de meses del programador se tomará para el desarrollo.

a. Estimación de la cantidad de instrucciones

Para la estimación se basó en la siguiente fórmula (1) significa L=cantidad de líneas de código, FD E/S Flujo de Entrada + Flujo de Salida que tiene el sistema.

En la ecuación (2) significa ML=Miles de código fuente que tendrá el sistema.

$$L = 1265 * FD E/S \quad (1)$$

$$L = 1265 * 5 = 6325$$

$$ML = \frac{F}{10000} \quad (2)$$

$$ML = \frac{6325}{10000} = 0.632$$

b. Estimación del esfuerzo

En la fórmula (3) significa el 3= modo de desarrollo semiencajado y el ML=miles de línea de código del sistema

$$ESF = 3 * ML^{1.12} \quad (3)$$

$$ESF = 3 * 0.632^{1.12} = 1.76 \approx 2 \text{ personas}$$

c. Estimación del tiempo de desarrollo

En la fórmula (4) TDE= tiempo de desarrollo del proyecto, 2.5= modo semiencajado dependiendo del problema, el grupo puede incluir una mezcla de personas experimentadas y no experimentadas, 0.35= el tiempo de desarrollo en el modo semiencajado.

$$TDES = 2.5 * ESF^{0.35} \quad (4)$$

$$TDES = 2.5 * 1.76^{0.35} = 3.05 \approx 3 \text{ meses}$$

d. Estimación del personal necesario

En la fórmula (5) detalla ESF=estimación del esfuerzo y TDES= estimación de tiempo de desarrollo.

$$CP = \frac{ESF}{TDES} \quad (5)$$

$$CP = \frac{1.76}{3.05} = 0.57 \approx 1$$

e. Estimación de productividad

En la fórmula (6) detalla ESF=estimación del esfuerzo y TDES= estimación de tiempo de desarrollo.

$$P = \frac{L}{ESF} \quad (6)$$

$$P = \frac{6325}{1.76} = 3593 \frac{\text{instrucciones}}{\text{personas_mes}}$$

f. Estimación del coste

En la fórmula (7) detalla ESF=estimación del esfuerzo y CHM= sueldo del personal del proyecto.

$$P = ESF * CHM \quad (7)$$

$$P = 1.76 * 400 = 704\$$$

Para la realización del presupuesto del prototipo de riego es la programación fue de un valor de 704\$ la estimación de esfuerzo de 1.76 y el personal requerido para realizar este proyecto es de 2 personas.

6. CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- La revisión sistemática literaria de las técnicas de aprendizaje de máquinas se basan en un aprendizaje este sea supervisado o no supervisado y a su vez estos aprendizajes tienen algoritmos matemáticos. También la mayoría de estudios realizados se enfocaron en las especies vegetales.
- Se utilizó para el estudio de la investigación las especies vegetales de acuerdo a su forma y textura de la hoja, y en la frecuencia de los estudios analizados coinciden que el algoritmo de regresión logística tiene un porcentaje de validación de 99% con respecto al análisis de la revisión sistemática literaria.
- El algoritmo de regresión logística es la más alta, porque tiene un margen de error 1% analizado por medio de la metodología KDD, la cual permitió la realización por selección de la forma de la hoja y se utilizó la validación cruzada para la eficiencia del algoritmo.

6.2. Recomendaciones

- Para el desarrollo de la investigación del marco teórico se defina correctamente los criterios de selección de los artículos para evitar contradicciones sobre los artículos seleccionados. Y adicional a esto los artículos cumplan con los criterios de calidad.
- Se debe definir correctamente la especie vegetal que se desea utilizar debido a que la mayoría de los artículos solo detallan de manera general y no se basan en una especie en específica. Por lo que provoca que el análisis tome más tiempo de investigación.
- Tener conocimiento sobre diferentes algoritmos en un lenguaje de programación específico lo que ayuda a la realización de la investigación sea rápidamente implementada en una aplicación móvil.

7. BIBLIOGRAFÍA

- [1] A. Calderon and H. D. H. Cortes, “Machine learning en la detección de enfermedades en plantas,” *Tecnol. Investig. y Acad.*, vol. 7, no. 2, pp. 55–61, 2019.
- [2] J. Cervantes, J. Taltempa, F. García-Lamont, J. S. R. Castilla, A. Y. Rendon, and L. D. Jalili, “Análisis comparativo de las técnicas utilizadas en un Sistema de Reconocimiento de hojas de planta,” *RIAI - Rev. Iberoam. Autom. e Inform. Ind.*, vol. 14, no. 1, pp. 104–114, 2018, doi: 10.1016/j.riai.2018.09.005.
- [3] J. A. Andrade Vera, “Aplicación móvil para la detección y tratamiento de daños de los cultivos de la parroquia Taura del cantón Durán, mediante el uso de software de análisis de imagen basado en técnicas de machine learning.,” p. 78, 2018, [Online]. Available: <http://repositorio.ucsg.edu.ec/handle/3317/11769>.
- [4] A. Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, “Automatic Recognition of Medicinal Plants using Machine Learning Techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, 2017, doi: 10.14569/ijacsa.2017.080424.
- [5] ANAYA, “El reino Plantas,” *Univ. del pacifico*, vol. 1, pp. 172–196, 2017, [Online]. Available: file:///C:/Users/Master/Downloads/BIOLOGIA_GEOLOGIA_1_ESO_U08_ReinoPlantas.pdf.
- [6] Secretaría del Convenio de Diversidad Biológica, *Informe sobre la Conservación de las Especies Vegetales: s: Una revisión de los progresos realizados en la aplicación de la Estrategia Mundial para la Conservación de Plantas (GSPC)*. 2019.
- [7] Santiago Viteri y Adriana Nuñez., “Ocupa el séptimo lugar en el mundo en cuanto a biodiversidad de plantas vasculares,” *Ministerio de agricultura y aganderia.*, 2019. <https://www.agricultura.gob.ec>.
- [8] Aina S. Erice, “La invención del reino vegetal,” in *Editorial Ariel*, 2017, p. 448.
- [9] Autoridad Nacional del Servicio Civil, “¿Cuántas especies nativas de plantas vasculares hay en Ecuador?,” *Angew. Chemie Int. Ed.* 6(11), 951–952., pp. 2013–2015, 2021.

- [10] S. A. Silva, “Biologías de las Plantas I,” *Esc. Norm. Super. Argentinos*, vol. 1, p. 96, 2018, [Online]. Available: <https://red.infed.edu.ar/blog/wp-content/uploads/2018/11/SilvaLibro-digital-Botánicapdf-1.pdf>.
- [11] J. T. Roig Mesa, “Plantas medicinales y aromáticas,” *Rev. Cuba. farm*, vol. 2, no. 3, pp. 89–95, 2018.
- [12] C. E. C. Martínez, “Artículo Plantas medicinales de los Andes ecuatorianos,” 2017.
- [13] V. Van Den Eynden and E. Cueva, “Las plantas en la alimentación,” *Sierra*, no. September, pp. 62–66, 2018.
- [14] ESTec, “USO DE ESPECIES NATIVAS COMO,” *CienciAmérica*, vol. 10, 2021.
- [15] M. O. García, “Técnicas de aprendizaje máquina para la optimización de características de una red de datos,” 2020.
- [16] R. S. Kumar and R. Sunder, “A Comparative Study on Disease Detection of Plants using Machine Learning Techniques,” pp. 1937–1941, 2021.
- [17] S. K. Fernández Estrella, L. M. Olivares Gutiérrez, and F. J. Zamora Rasgado, “Inteligencia Artificial,” *XIKUA Boletín Científico la Esc. Super. Tlahuelilpan*, vol. 3, no. 5, 2017, doi: 10.29057/xikua.v3i5.1271.
- [18] L. Dominguez, “Desarrollo de un sistema de identificación y reconocimiento de plantas usando técnicas de aprendizaje máquina,” Universidad Autónoma del Estado de México, 2018.
- [19] S. Poornima, S. Kavitha, S. Mohanavalli, and N. Sripriya, “Detection and classification of diseases in plants using image processing and machine learning techniques,” *AIP Conf. Proc.*, vol. 2095, no. April, 2019, doi: 10.1063/1.5097529.
- [20] D. Cabañas González, “Trabajo fn de grado Entrenamiento de un mini-robot para realizar tareas mediante aprendizaje automático www.uam.es Universidad Autónoma de Madrid,” 2021, [Online]. Available: www.uam.es.
- [21] M. Mittal, G. Dhingra, and V. Kumar, “Machine Learning Methods Analysis for Calories Measuremnt of Fruits and Vegetables,” *Proc. IEEE Int. Conf. Signal Process. Control*, vol. 2019-Octob, pp. 212–219, 2019, doi: 10.1109/ISPPCC48220.2019.8988487.

- [22] D. V. Nazarenko, P. V. Kharyuk, I. V. Oseledets, I. A. Rodin, and O. A. Shpigun, “Machine learning for LC-MS medicinal plants identification,” *Chemom. Intell. Lab. Syst.*, vol. 156, pp. 174–180, 2018, doi: 10.1016/j.chemolab.2016.06.003.
- [23] J. G. A. Barbedo, “Detection of nutrition deficiencies in plants using proximal images and machine learning: A review,” *Comput. Electron. Agric.*, vol. 162, no. May, pp. 482–492, 2019, doi: 10.1016/j.compag.2019.04.035.
- [24] V. Vázquez Rodríguez, “Desarrollo de aplicaciones móviles multiplataforma con Flutter,” 2019.
- [25] D. Carrizo and C. Moller, “Estructuras metodológicas de revisiones sistemáticas de literatura en Ingeniería de Software: un estudio de mapeo sistemático,” *Ingeniare. Rev. Chil. Ing.*, vol. 26, pp. 45–54, 2018, doi: 10.4067/s0718-33052018000500045.
- [26] J. Reyes Saldaña and R. García Flores, “El proceso de descubrimiento de conocimiento en bases de datos,” *Ediciones Univ. Coop. Colomb.*, vol. 8, no. 26, pp. 37–47, 2018.
- [27] C. Muñoz, “Aplicación de la metodología Mobile- D en el desarrollo de una app móvil para gestionar citas médicas del centro Jel Riobamba,” 2020.
- [28] L. Guadalima-inga, R. Lojano-chavez, and M. Cabrera-sarango, “Revisión sistemática de literatura : Inteligencia Artificial en sanidad vegetal,” pp. 63–69, 2020.

8. ANEXOS

Anexo A. Ficha bibliográfica

No. FICHA BIBLIOGRAFICA		AREA	UBICACIÓN	
A1		Investigación en Ciencias de la Computación	Artículo 978-1-6654-0521-8/21/\$31.00 https://ieeexplore.ieee.org/document/9441844	
TEMA		TITULO	AUTOR	BASE DE DATOS
A Comparative Study on Disease Detection of Plants using Machine Learning Techniques-Un estudio comparativo sobre la detección de enfermedades de las plantas Uso de técnicas de aprendizaje automático		Machine learning	Vishnu S Babu-R. Satheesh Kumar-R. Sunder	Google academic
RESUMEN			PALABRAS CLAVES	
<p>La agricultura era la columna vertebral de la India ahora lo es enfrentando varias dificultades que incluyen enfermedades, selección de semillas de calidad, escasez de agua, etc. Uno de los principales problemas del campo de la agricultura es enfermedades de las plantas que causan a los agricultores una enorme pérdida ya sea por la pérdida de la cosecha o por el uso innecesario de medicamentos. Temprano La detección de una enfermedad de las plantas puede prevenir su propagación, por lo tanto, la pérdida de rendimiento. La detección de enfermedades se puede hacer usando una máquina, técnicas de aprendizaje y procesamiento de imágenes. Un modelo entrenado puede identifique la planta enferma fácilmente. Se requiere un gran conjunto de datos para formación que será un gran inconveniente ya que la falta de disponibilidad de conjunto de datos Esta propuesta de investigación es desarrollar un sistema de aprendizaje automático modelo basado en la detección y clasificación de enfermedades adecuado para sector agrícola indio.</p>			plant disease,machine learning	

Anexo B. Selección de artículos en base a criterios de calidad

The screenshot shows the ScienceDirect search interface. At the top left is the ScienceDirect logo. The search bar contains the text "Machine learning techniques vegetable species". Below the search bar, it indicates "965 results". On the left side, there is a "Refine by:" section with a "Years" filter. The years listed are: 2022 (161), 2021 (382), 2020 (219), 2019 (153), 2018 (124), 2017 (87), 2016 (92), and 2015 (65). The main search results area shows two articles. The first is a "Review article" titled "Vegetable waste and by-products to feed a healthy gut microbiota: Current evidence, machine learning a design novel microbiome-targeted foods" by Carlos Sabater, Inés Calvete-Torre, and Lorena Ruiz. The second is a "Research article" titled "Machine learning applied to canopy hyperspectral image data to support biological control of soil-borne vegetables" by Catello Pane, Gelsomina Manganiello, and Corrado Costa.

Figura 56. Base de datos Science Direct

The screenshot shows the Google Académico search interface. The search bar contains "machine learning". Below the search bar, it indicates "Aproximadamente 1.050.000 resultados (0,04 s)". On the left side, there are filters for "Cualquier momento" (Desde 2022, Desde 2021, Desde 2018, Intervalo específico...), "Ordenar por relevancia", "Ordenar por fecha", and "Cualquier idioma" (Buscar sólo páginas en español). The main search results area shows two articles. The first is "[LIBRO] Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2" by S. Raschka and V. Mirjalili. The second is "A guide to machine learning for biologists" by J.G. Greener, S.M. Kandathil, and L. Moffat.

Figura 57. Base de datos Google académico

The screenshot shows the Scielo search interface. The search bar contains "machine learning". Below the search bar, it indicates "Resultados: 417". On the left side, there is a "Filtros" section with a "Colección" filter. The main search results area shows one article titled "Determining the geographical origin of lettuce with data mining applied to micronutrients and soil properties" by Malone, Camilla; Araujo, Eloá Moura; Santos-Araujo, Sabrina Novaes dos; Boim, Alexys Giorgia Friol; Barbosa, Rommel Melgaço; Alleoni, Luis Reynaldo Ferracú.

Figura 58. Base de datos Scielo

Anexo C. Ficha técnica de la aplicación

FICHA TÉCNICA DE APLICACIÓN MÓVIL

I. DATOS GENERALES	
Nombre de la aplicación	Health Plant
Descripción	App móvil para el reconocimiento de plantas medicinales como son menta y orégano
Sistema Operativo	Android
Fecha de creación del documento	4 de noviembre del 2021
Fecha de última actualización	04 de febrero del 2022

II. REFERENTE TECNICO	
Nombre:	Tesistas
E-mail	jhoanna.molina2@utc.edu.ec - thalia.salazar2806@utc.edu.ec

III. REQUERIMIENTOS DEL PRODUCTO										
Tipo de puesta	x	Pruebas						Producción		
Tipo de aplicación		Web				x	Móvil			
Código ejecutado	x	Dart		C		ASP		PHP	Otro ¿Cuál?	
¿Requiere base de datos?	x	Si	SQL Server		Oracle	x	SQLite	PostgreSQL	Otro ¿Cuál?	
		No								
¿Utiliza manejador de reportes?		Si	¿Cuál?							
	x	No								
Interactúa con otra aplicación		Si			X		No			
		¿Cuál?								

IV. REQUISITOS DE SOFTWARE DE LA APLICACIÓN MÓVIL			
Sistema operativo	Android	Versión	8
Requisitos		Características	
Cámara		>3.0 MP	
Permisos		Escritura y Lectura de datos - Acceso	

V. REQUISITOS DE HARDWARE DE LA APLICACIÓN MÓVIL	
Plataformas de hardware que están soportadas	Android 8 o más.
Velocidad mínima de procesador	1.2 GHz
Requisito núcleo de procesador	Quad Core
Requisito de memoria	2GB RAM
Pantalla	>= 5'' (por comodidad del usuario)
Conectividad a internet	NO
Cámara de fotos	>3.0 MP
Marca del celular	Dispositivos Android
Observaciones	No funciona en dispositivos con iOS

Anexo D. Juicio de expertos

**UNIVERSIDAD
TECNICA DE
COTOPAXI**

**FACULTAD DE
CIENCIAS DE LA
INGENIERIA Y APLICADAS**

INGENIERIA EN INFORMATICA
y
SISTEMAS COMPUTACIONALES

TEMA:
"REVISIÓN DE TÉCNICAS DE APRENDIZAJE DE MÁQUINAS,
USADAS EN EL RECONOCIMIENTO DE ESPECIES
VEGETALES"

ESTUDIANTES:
Molina Cajas Joanna Elizabeth
Salazar Segovia Thalia Maricruz

TUTORA:
Ing. Mrt Cantuía Flores Karla Susana

Anexo E. Costo del desarrollo de la investigación teórica y técnica

GASTOS DIRECTOS

Propuesta para la Elaboración					
Ítem	Recursos	Cantidad	Unidad	Valor Unitario	Valor Total
1	Esferos	5	Unidades	0,3	1,5
3	Impresiones	450	Hojas	0,2	90
4	Internet	350	Horas	0,5	175
5	Cuadernos	2	Unidades	0,75	1,5
6	Anillados	6	Unidades	2,5	15
7	Empastados	2	Empastados	20	40
8	Plantas	400	Unidades	0,25	100
9	CDs	2	Unidades	1,5	3
Total					426

GASTOS INDIRECTOS

Propuesta para la Elaboración					
Ítem	Recursos	Cantidad	Unidad	Valor Unitario	Valor Total
1	Alimentación	50	Meriendas	2,5	125
2	Transporte	50	Pasajes	2,5	125
3	Comunicación	70	Recargas	3	210
5	Visitas Viveros	20	Ingresos	1	20
Total					480

COSTO DEL SOFTWARE

Estimación del costo de desarrollo de la aplicación web	
Estimación de cantidad de instrucciones (1) y (2)	L=6325 ML=0.632
Estimación de esfuerzo (3)	ESF=1.76≈ 2 personas
Estimación de tiempo de desarrollo (4)	TDES=3.05≈3 meses
Estimación del personal necesario (5)	CP=0.57≈1
Estimación de productividad (6)	P=3593 instrucciones/personas_mes
Estimación de coste (7)	EC=\$704

GASTOS GENERALES

Propuesta General	
Recursos	Valor Total
Gastos Directos	\$426
Gastos Indirectos	\$480
Costo del software	\$704
Imprevistos	\$45,3
Total	\$1655.3